Beyond excitation/inhibition imbalance in multidimensional models of neural circuit changes in brain disorders

Cian O'Donnell[1,2,*], J. Tiago Gonçalves[3,4], Carlos Portera-Cailliau[3] and Terrence J. Sejnowski[2,5,*]

[1]Department of Computer Science, University of Bristol, Bristol, UK
[2]Howard Hughes Medical Institute, Salk Institute for Biological Studies, La Jolla CA, USA
[3]Departments of Neurology and Neurobiology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA
[4]Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA
[5]Division of Biological Sciences, University of California at San Diego, La Jolla, CA, USA
*Correspondence: cian.odonnell@bristol.ac.uk (C.O'D.), terry@salk.edu (T.J.S.)

Abstract

A leading theory holds that neurodevelopmental brain disorders arise from imbalances in excitatory and inhibitory (E/I) brain circuitry. However, it is unclear whether this one-dimensional model is rich enough to capture the multiple neural circuit alterations underlying brain disorders. Here we combined computational simulations with analysis of *in vivo* 2-photon Ca^{2+} imaging data from somatosensory cortex of *Fmr1* knock-out (KO) mice, a model of Fragile-X Syndrome, to test the E/I imbalance theory. We found that: 1) The E/I imbalance model cannot account for joint alterations in the observed neural firing rates and correlations; 2) Neural circuit function is vastly more sensitive to changes in some cellular components over others; 3) The direction of circuit alterations in *Fmr1* KO mice changes across development. These findings suggest that the basic E/I imbalance model should be updated to higher-dimensional models that can better capture the multidimensional computational functions of neural circuits.

Introduction

The nervous system shows complex organization at many spatial scales: from genes and molecules, to cells and synapses, to neural circuits. Ultimately, the electrical and chemical signaling at all of these levels must give rise to the behavioral and cognitive processes seen at the whole-organism level. When trying to understand prevalent brain disorders such as autism and schizophrenia, a natural question to ask is: where is the most productive level of neuroscientific investigation? Traditionally, most major disorders are diagnosed entirely at the behavioral level, whereas pharmaceutical interventions are targeted at correcting alterations at the molecular level. However even for the most successful drugs, we have little understanding of how pharmaceutical actions at the molecular level percolate up the organizational ladder to affect behavior and cognition. This classic bottom-up approach may even be further confounded if phenotypic heterogeneity in disorders such as autism turn out not to reflect a unique cellular pathology, but rather "a perturbation of the network properties that emerge when neurons interact" (Belmonte et al., 2004). These considerations imply that a more promising level of analysis might be at the level of neural circuits, since the explanatory gap between circuits and behavior is smaller than the gap between molecules and behavior. This circuit-level viewpoint argues for a reverse-engineering approach to tackling brain disorders: rather than start at the molecular level and working up, we should instead start by asking how cognitive and behavioral symptoms manifest as alterations at the circuit level, then interpret these changes at the levels of cells, synapses, and molecules as appropriate.

One prominent circuit-level hypothesis for brain disorders has been the idea of an imbalance in excitatory and inhibitory signaling. First proposed as a model for autism (Rubenstein and Merzenich, 2003), the concept has since been applied to many other brain disorders, including Schizophrenia, Rett syndrome, fragile-X syndrome, tuberous sclerosis, and Angelman Syndrome. However, a major drawback of this model is that it only considers overall activity, which is 1-dimensional. It implies that either too much excitation or too much inhibition is unhealthy (Figure 1A). Although several studies have found evidence that the E/I balance is indeed upset in multiple brain disorders (Bateup et al., 2011; Dani et al., 2005; Gibson et al., 2008; Kehrer et al., 2008; Wallace et al., 2012), a model's usefulness should not be judged on

whether it is nominally true or false, but on its explanatory and predictive powers as compared with competing alternative models. In this study we argue that that even if the E/I imbalance model proves correct, its unidimensionality might ultimately limit its applicability, for three reasons.

First, by placing all disorders on the same single axis, the E/I imbalance model implicitly lumps together some vastly different disorders, such as epilepsy, schizophrenia and autism (Figure 1A) because they share an excess of excitation. By extension it implies that the symptoms of diverse disorders could be normalized solely by either enhancing or reducing the level of, say, GABAergic signaling as appropriate. Although clinical trials for such GABAergic-based interventions are ongoing (Braat and Kooy, 2015), no treatment for a neurodevelopmental disorder based on this principle has yet been approved.

A second issue with the unidimensionality of the E/I imbalance model is that it lumps together all excitatory and inhibitory neural circuit components. In Figure 1B we show a schematic diagram of a generic neural circuit with excitatory components colored red and inhibitory components colored blue. The E/I imbalance model implies that varying any of the excitatory components, such as the strength of recurrent excitatory synapses or the input resistances of excitatory neurons, would have the same overall effect on circuit function. In contrast, theorists have found that these equivalences often do not hold even in very simple circuit models (Wilson and Cowan, 1972).

Third, because the standard E/I imbalance model is given in terms of circuit components, not circuit function, it does not specify which aspect of a neural circuit's activity should be maintained for healthy performance. For example, it leaves unclear which of neuronal firing rates, synchrony, or reliability of responses might be altered if E/I balance is upset.

To motivate our study, we began by investigating which circuit activity properties are altered in a model brain disorder. We re-analyzed published *in vivo* 2-photon Ca^{2+} imaging data we previously recorded from somatosensory cortex in *Fmr1* knockout mice (Gonçalves et al., 2013), a well-studied animal model for fragile-X syndrome (The Dutch-Belgian Fragile X Consortium, 1994). We compared the data from wild-type (WT) mice with *Fmr1* KO mice,

across three different developmental time points: just before (P9—11) and after (P14—16) the critical period for heightened activity-dependent synaptic plasticity in L2/3 barrel cortex, and a more mature timepoint (P30—40). Example ΔF/F raster plots from each group are shown in Figure 1C, top left. We binned the data into 1 s timebins (originally imaged at 4 Hz), then transformed each neuron's timeseries of ΔF/F values into a probabilistic sequence of binary ON/OFF values by assuming a Poisson firing model (Methods). We then summarized the neural population activity from each animal with three statistics: the mean ON probability across all recorded neurons, the standard deviation (s.d.) in ON probability across neurons, and the mean correlation between all pairs of neurons (Figure 1C, bar charts right and scatter plots lower left). Together these measures capture both the statistics of the bulk population activity and some indication of the heterogeneity across neurons.

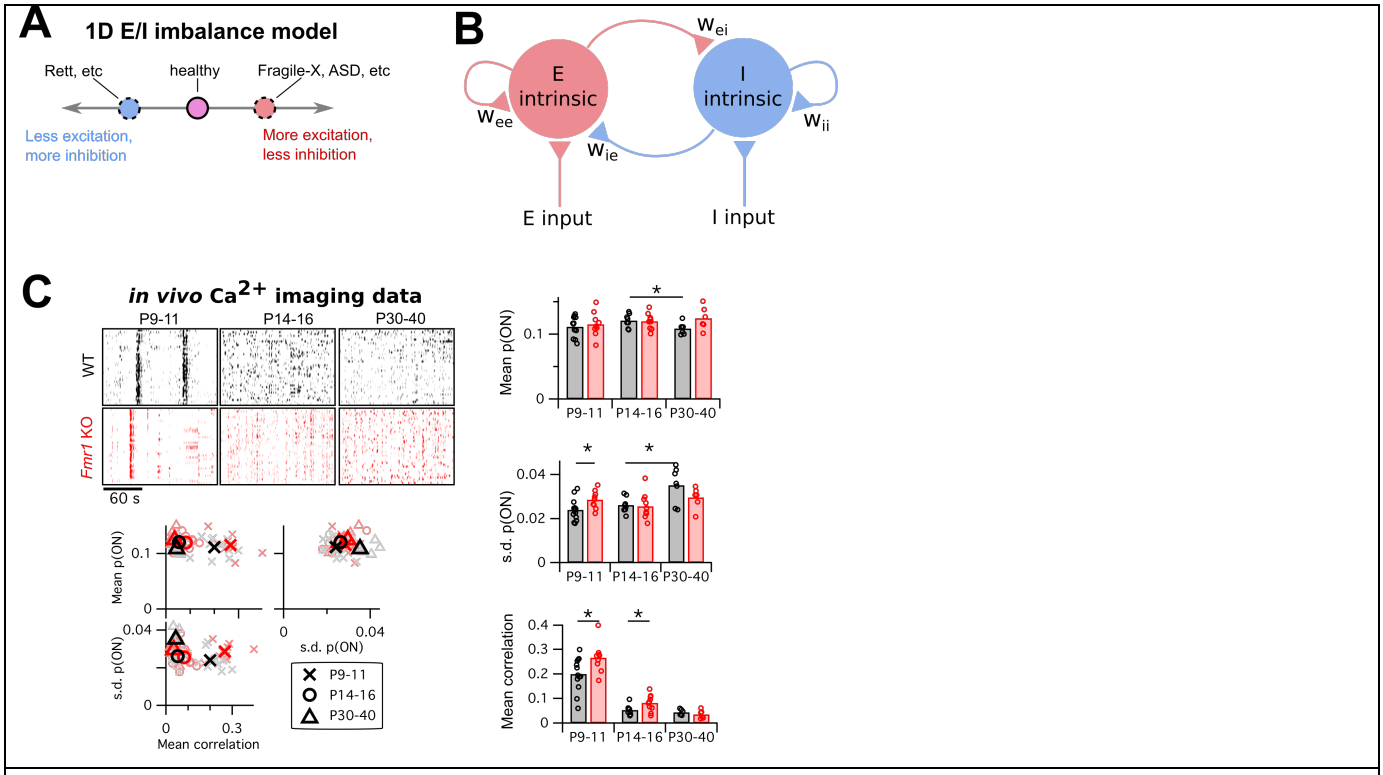

Figure 1. Mismatch between the E/I imbalance model's unidimesionality and the multiple changes in circuit activity in Fragile-X mouse models.

A: Schematic of standard E/I imbalance model as a unidimensional axis.

B: Diagram of a generic neural circuit, showing an excitatory and an inhibitory population of neurons and their interconnections. Although the E/I imbalance model implicitly groups all

components as either excitatory (red) or inhibitory (blue), in principle any component could separately be altered in brain disorders, and may have a distinct effect on circuit function.
C: Upper left, example Ca^{2+} imaging dF/F raster plots from a single animal from each of two genotypes, WT and *Fmr1* KO, and three age groups, P9—11, P14—16 and P30—40. In each case 3 minutes of data are shown from 40 neurons. Right and lower left, mean firing probability, standard deviation of firing probabilities, and mean pairwise correlation across all neurons. Same data in scatter plots lower left and bar charts right. * indicates significant difference in group means at $p<0.05$, by bootstrapping.

For mean firing rates, the only change we detected was a decrease in firing probability in WT between P14—16 and P30—40 ($p = 0.027$), which was coupled with an increased s.d. of firing rates ($p = 0.015$). We also detected a higher firing rate s.d. in P9—11 KO animals than WT ($p = 0.031$). Finally, as previously reported (Golshani et al., 2009; Gonçalves et al., 2013; Rochefort et al., 2009), we found a substantial decrease in pairwise correlations in both genotypes across development, with slightly higher correlations in KO animals than WT at P9—11 ($p = 0.029$) and P14—16 ($p = 0.047$).

These results show that multiple statistics of cortical circuit activity are altered in *Fmr1* KO mice. However, two questions remain: 1) Which circuit components are responsible for these activity alterations? 2) What is the consequence for circuit computation? In the remainder of this study, we used computational simulations and further data analysis to ask whether the E/I imbalance model could help address these questions.

We first built a detailed spiking neural circuit model of mouse L2/3 somatosensory cortex to explore how its various excitatory and inhibitory components affect the circuit's spiking output, and found that the E/I imbalance model was not flexible enough to capture key aspects of this relationship. We then derived an abstract 2-dimensional circuit model that captures more features of the circuit function than the 1-dimensional E/I imbalance model did. Using this new model, we found that certain sets of circuit components have redundant effects on circuit function, and that circuit function is vastly more sensitive to changes in some components over others. To ask how this 2-D model could help interpret brain circuit abnormalities in a particular test case, we fit a version of the model to the Ca^{2+} imaging data from Fragile-X mouse models

presented above. We found that the model predicts opposite changes in *Fmr1* KO circuit properties at different developmental ages. Finally, we applied a new large-scale neural population analysis method (O'Donnell et al., 2017) to the same Ca^{2+} imaging data, and found systematic shifts in the distribution of neural circuit activity patterns in Fragile-X that were not predictable from neural firing rates or correlations alone.

Results

Neural circuits consist of many components that typically interact non-linearly to generate complex circuit activity dynamics. Although many properties of cortical circuit components have been found to be altered in animal models of brain disorders, it remains extremely difficult to predict the net effect of varying any one particular parameter on circuit activity. The E/I imbalance model seeks to simplify this problem by projecting all circuit alterations onto a one-dimensional axis (Figure 1A) where the goal is to achieve a 'healthy' balance of excitation and inhibition. Under this model, either too much excitation or too much inhibition leads to improper circuit function.

To explicitly test whether the E/I imbalance model can account for the effects of cellular component changes on circuit function, we built a detailed computational model of layer (L) 2/3 mouse somatosensory cortex. This circuit has been studied in detail by neurophysiologists, and several of its properties have been found to be altered during development in mouse models of Fragile-X syndrome, including parvalbumin-positive interneuron number (Selby et al., 2007), GABA receptor reversal potential (He et al., 2014), dendritic spine dynamics (Cruz-Martin et al., 2010), and L4 excitatory input (Bureau et al., 2008), reviewed by Contractor et al. (2015). We used numerical computer simulations to perform hypothetical experiments where we perturbed various parameters of the circuit model and observed the resulting changes in circuit-level activity. Although we focused on this particular brain circuit for tractability, our general conclusions and methodology should be readily applicable to other brain circuits (Frye and Maclean, 2016).

The L2/3 computational model we built (Figure 2A, see Methods for details) consisted of four interconnected populations of leaky integrate-and-fire neurons: one group of 1700 excitatory (E) pyramidal neurons and three groups of inhibitory neurons: 115 $5HT_{3A}R$-expressing neurons, 70 parvalbumin-expressing (PV) neurons, and 45 somatostatin-expressing (SOM) neurons. This L2/3 circuit was driven by a separate population of 1500 L4 excitatory neurons. Cellular numbers, intrinsic properties, synaptic strengths, and connectivity statistics were taken from published *in vitro* data from P17—22 wild-type mice (Avermann et al., 2012; Lefort et al., 2009;

Tomm et al., 2014). We chose this level of detail for the model in order to relate experimentally measureable biophysical properties of neurons to their putative role in the circuit at large.

L2/3 neurons of the rodent somatosensory neocortex respond only sparsely to sensory stimulation *in vivo*. For example, twitching a whisker activates, on average, only ~20% of L2/3 neurons in its corresponding barrel, each of which typically emits only one action potential (Clancy et al., 2015; Kerr et al., 2007; Sato et al., 2007). Hence any individual neuron carries very little information about the stimulus on its own, implying that information must instead be encoded at the circuit level as the identities of the subset of neurons that respond.

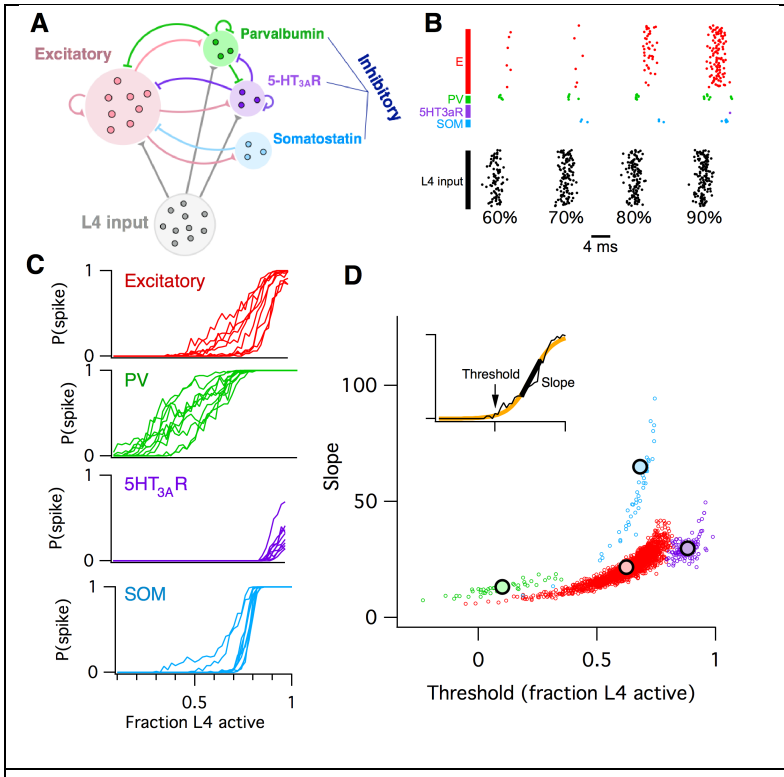

Figure 2. Computational model of L2/3 mouse somatosensory cortex.

A: Schematic diagram of computational circuit model.

B: Example raster plots of spiking responses from a subset of neurons from each cell type (colors as in panel A), for varying fractions of L4 activated (black).

C: Probability of spiking as a function of the fraction of L4 neurons activated. Each curve represents the response probability of a single neuron, averaged over multiple trials and multiple permutations of active L4 cells.

D: Each circle plots the fitted logistic slope and threshold values for a single neuron in the simulation. Circle color indicates cell type: red is excitatory, green is PV inhibitory, purple is $5HT_{3A}R$ inhibitory, blue is SOM inhibitory. Large black circles indicate mean for each cell-type. Inset shows an example fitted logistic response function (orange) to the noisy simulation results from a single excitatory neuron (black).

To model whisker stimulation, we simulated a volley of spikes arriving from L4 as input to the population of L2/3 cells. We chose a random subset of L4 neurons as ON, then sent a single spike from each of these L4 cells to their target neurons in L2/3, and recorded the responses of all neurons in L2/3, some of which spiked and some of which did not. We repeated this identical stimulation multiple times, in order to get an average response probability for each L2/3 neuron, given the probabilistic vesicle release at synapses in the model. Then we chose a different random subset of L4 neurons as ON, and repeated the entire procedure. Finally, we varied the fraction of L4 cells active and plotted the probability of response for each individual L2/3 neuron as a function of L4 activity level (Figure 2B—C).

We found that the mean response probability of each neuron in the simulation increased from zero to one monotonically with increasing L4 activity level. This sigmoidal-shaped response profile of simulated L2/3 neurons mimics the spiking response of mouse L2/3 pyramidal cells to extracellular L4 stimulation *in vitro* (Elstrott et al., 2014), while the sparse, noisy and distributed network responses were reminiscent of *in vivo* activity following whisker stimulation (Clancy et al., 2015; Kerr et al., 2007). Neurons of all four cell types in the simulation responded to the L4 stimulus, including the SOM interneurons which did not receive direct L4 input, but were instead activated by disynaptic connections via L2/3 excitatory neurons. The detailed shape of the response curve varied systematically across cell types, and was heterogeneous for different neurons of a given cell type. To quantify these differences, we used logistic regression to fit the response profile of each neuron with a sigmoid function (Figure 2D, inset), which has just two parameters: the slope (representing the steepness of the response curve) and threshold (representing the minimal fraction of L4 neurons needed to activate the cell). When we plotted the fitted slope and threshold values for each neuron against each other, we found that each cell type falls into a distinct cluster in this 2-dimensional space. For example,

all PV inhibitory neurons had a low slope and low threshold, whereas SOM inhibitory neurons had a steep slope and moderate threshold. We then used these slope-threshold measurements to summarize the circuit-level input-output function of this 'default' model of L2/3 somatosensory cortex. This 2-D logistic model has two benefits over the 1-D E/I imbalance model: first, its extra degree of freedom allows for richer and more flexible fits to data, and second, by describing an input-output mapping for the L2/3 circuit it can capture some aspects of the computation that the circuit performs for the animal. In contrast, the E/I imbalance model is specified purely in terms of circuit components, and so is agnostic to the circuit's computational function.

The biophysical circuit model contained 100 parameters (Methods). How sensitive is the circuit's macroscopic input-output function to alterations in its low-level components? To test this, we varied 76 of the model parameters in turn by ±20%, and repeated the entire set of simulations for each case (Methods). For each individual parameter alteration, we fit a new logistic response function for each model neuron. We summarize the results by plotting the logistic slope and threshold parameters and comparing their values to those found with the default model. The outcomes were hugely varied. We show three examples from the set of seventy-six in Figure 3A, chosen to illustrate three qualitatively different effects that neural parameter changes can have on circuit function. First, when we increased the amplitude of postsynaptic potentials (PSPs) of excitatory synapses from L4 to L2/3 excitatory neurons, we found that the logistic threshold parameter of all cell types shifted leftwards to lower values (Figure 3A left), implying that fewer L4 neurons were needed to activate the entire L2/3 circuit. In contrast, when we increased the PSP amplitude of a different excitatory synapse, the recurrent connections between L2/3 excitatory neurons, we found (Figure 3A center) that excitatory and SOM inhibitory neurons had increased slope parameters relative to default, with little change in their threshold parameters. $5HT_{3A}R$ inhibitory neurons had decreased slopes and thresholds, while PV neurons had little change at all. As a third example we increased the probability of inhibitory synaptic connections from L2/3 PV interneurons to L2/3 excitatory neurons (Figure 3A right). In this case we found that excitatory neurons had a lower slope and increased threshold, SOM inhibitory neurons had a lower slope, $5HT_{3A}R$ inhibitory neurons had

both an increased slope and threshold, and PV neurons showed little change, even though their outgoing synapses were the parameter that was altered.

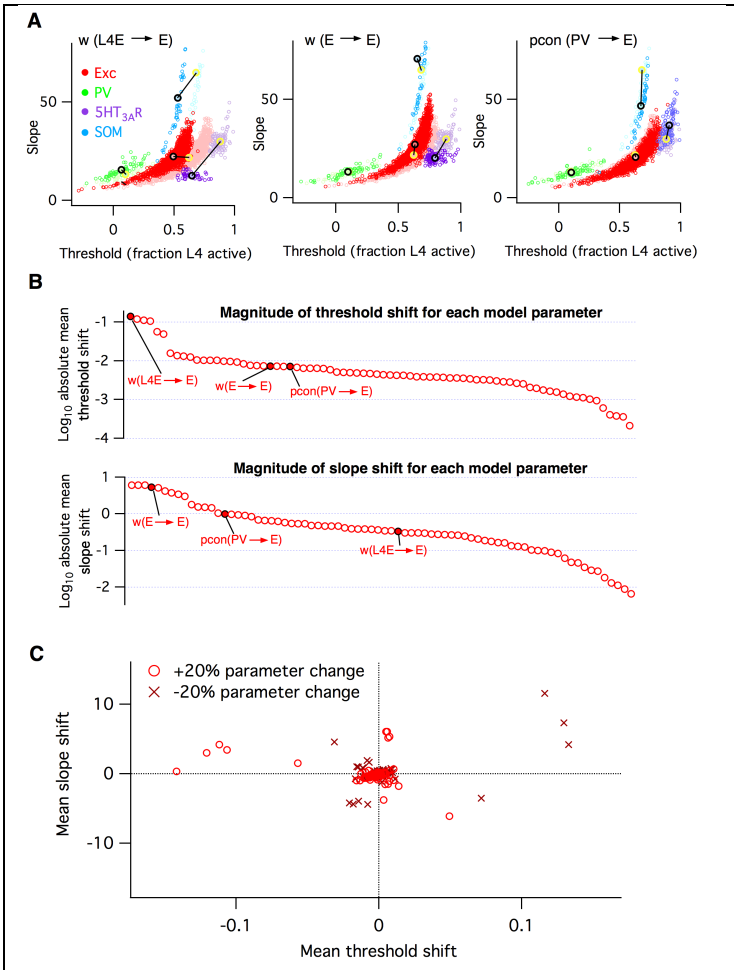

Figure 3. Heterogeneous effects of varying L2/3 parameters on the circuit input-output function.

A: Shifts in the distribution of fitted slope and threshold parameters as a result of increasing the strength of synapses from L4 to L2/3 E neurons (left), increasing the strength of recurrent synapses between L2/3 E neurons (center), or increasing the connection probability between L2/3 PV interneurons and E neurons (right). Transparent circles represent values for default network, heavy circles for altered network. The default and altered group means are large yellow and black open circles, respectively.

B: Absolute values of mean shifts in threshold (upper plot) and slope (lower plot) for Excitatory neurons arising from increasing the value of each parameter by +20%. The three example

parameters from panel A are labeled and indicated as filled red circles. Note that data are presented on a log10 scale.

C: The shift in mean slope-threshold parameter values for E neurons from the default network values, in response to each of the 76 circuit parameter alterations. Light red circles indicate +20% increase in parameter value; dark red crosses indicate a -20% decrease in parameter value.

To synthesize the findings from all simulations, for each altered parameter we plotted the shift in mean slope-threshold values for L2/3 excitatory neurons from the mean values found with the default model (Figure 3B). We focused on excitatory neurons because they constitute 90% of the neurons in this layer (Lefort et al., 2009) and are the primary output to downstream circuits (Mao et al., 2011; Petreanu et al., 2007). Overall, we found a very heterogeneous picture. First, the magnitude of the shift in circuit response varied from parameter to parameter (Figure 3B—C). Varying some parameters, such as the first two examples given above, had large effects, whereas varying other parameters such as w_{I5htE} (the strength of synapses from $5HT_{3A}R$ inhibitory neurons to E neurons) or τm_{Isom} (the membrane time constant of SOM inhibitory neurons) had little effect. Second, the direction of shift in 2-D slope-threshold space also depended on parameter (Figure 3C). Increasing some parameters changed either circuit slope or threshold in isolation, while other parameters changed both slope and threshold together. All four quadrants of the slope-threshold plane could be reached by various subsets of the model parameters. Third, basic knowledge of whether a component was 'excitatory' or 'inhibitory' was insufficient to predict the direction of slope-threshold change. For example, the two glutamatergic projections considered in Figure 3A had distinct effects on circuit function.

In summary, these simulations indicate that the L2/3 somatosensory cortex circuit has extremely varied sensitivities to changes in its cellular components, and that the eventual circuit-level consequences cannot be predicted from knowledge of the class of the perturbed neurotransmitter alone. Since the E/I imbalance model groups all excitatory and inhibitory components as respective equals, it cannot account for these results.

Firing rates and correlations from the logistic model

In the above analysis, we investigated how low-level circuit components affect a high-level circuit input-output function, as parameterized by the slope and threshold of fitted logistic functions. But how is this logistic input-output function related to more common measures of neural population activity, such as firing rates and pairwise correlations between neurons? To investigate this, we considered the following reduced statistical model of cortical activity. We assumed for simplicity that the magnitude of the total input to the L2/3 circuit can be described by a Gaussian distributed random variable, with zero mean and unit standard deviation (Figure 4A lower left). Then we described each L2/3 neuron's input-output as a logistic function as before (Figure 4A upper left), with threshold and slope defined relative to the Gaussian input's mean and standard deviation, respectively. Given this model, we can numerically calculate the probability distribution over a neuron's firing probability, which in general is skewed and non-Gaussian (Figure 4A upper right). From this function we compute (Methods) both the neuron's mean firing probability and the pairwise correlation of two identical neurons following this profile (Figure 4A lower right). Example samples from the model are illustrated in Figure 4B.

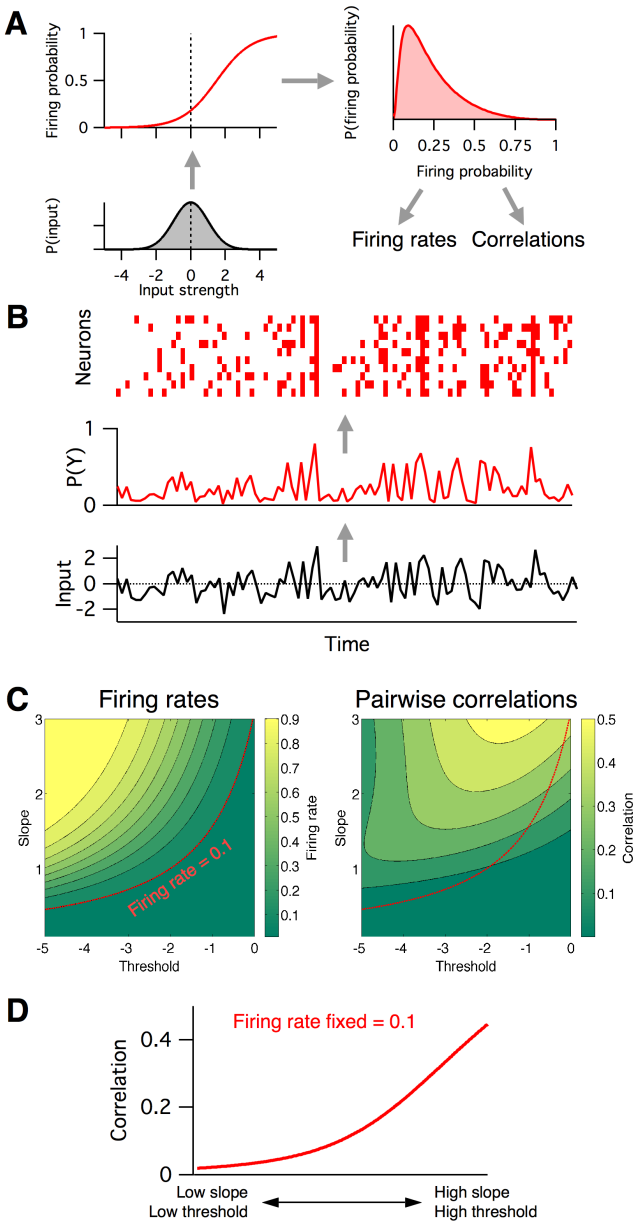

Figure 4. Firing rates and pairwise correlations from the logistic response model.

A: Logistic model components. We assume a normally distributed input drive (gray distribution, bottom left), which is passed through the neuron's probabilistic spike input-output function (red curve, top left), which results in a distribution of spike probabilities (top right) that are determined by the input-output function's slope and threshold parameters. From the output distribution we can directly calculate the mean firing rate and correlation between a pair of such neurons (Methods).

B: Example spikes from the logistic model. The bottom trace (black) shows examples inputs over time drawn randomly from the same normal distribution. This is transformed to spike probability at each time point (red trace). Example spike trains can then be generated from the spike probability trace by drawing Bernoulli samples with the specified probabilities (red ticks, top). If each neuron's spike train is conditionally independent given the same spike probabilities, we can see correlations in their spike trains.

C: Calculated mean firing rate (left) and pairwise correlation (right) color maps as a function of the logistic threshold (x-axis) and slope (y-axis) parameters. Contours indicate lines of fixed firing rate or correlation in the 2D slope-threshold space.

D: Pairwise correlation values along the slope-threshold contour for firing rate = 0.1.

Neural firing rates and correlations had qualitatively different dependencies on the underlying logistic model's slope and threshold. Neural firing rate was greatest when threshold was low and slope was high (top left of phase plot, Figure 4C left), whereas correlations were greatest when both threshold and slope were high (top right of phase plot, Figure 4C right). This implies that any change in the circuit's input-output function slope or threshold will in general have distinct effects on firing rate versus correlations, and so could not be captured by a 1-dimensional E/I balance model that sought to for example normalize firing rates alone. To illustrate this fact, we plot the calculated correlation values along a contour where firing probability is fixed at 0.1 (Figure 4D). In the region of parameter space where both the slope and threshold are low (Figure 4C bottom left), correlations are low, ~0.01. However, as we move along the contour for firing rate = 0.1 towards the region of parameter space where slope and threshold are high (Figure 4C top right), the pairwise correlations increase to ~0.4. This shows that a 1-dimensional E/I balance rule that exclusively sought to normalize neural firing rates would leave neural correlations free to achieve arbitrary values.

Previous studies have found evidence for an E/I imbalance in ASD (Lee et al., 2017; Nelson and Valakh, 2015). Fragile-X syndrome is the leading inherited cause of ASD, and also carries alterations in excitability (Contractor et al., 2015). We aimed to interpret our *Fmr1* KO Ca^{2+} imaging data (Figure 1C) via the 2D logistic model. Since our earlier analysis found important differences in the heterogeneity in firing rates across neurons in *Fmr1* KOs (Figure 1C), we

extended the 2-D logistic model for single neurons to a 5-D neural population version that captured cell-to-cell heterogeneity. The three extra parameters represented the standard deviations and correlation in slope and threshold parameters across the neural population (see Methods for details). We fit the parameters of the logistic model to reproduce the same neural population Ca^{2+} imaging data presented in Figure 1C. Given the three summary statistics from each animal (firing rate mean and s.d., and mean pairwise correlation), we used a gradient descent algorithm to find the five parameters of the population-level version of the logistic model that best matched the activity statistics (see Methods). The output statistics of the fitted models matched well that of the target data (Figure 5 – figure supplement 1). Example neural population activity patterns drawn from the mean model fits for each group are shown in Figure 5A, along with the fitted slope-threshold functions (Figure 5A insets), to be compared with the Ca^{2+} imaging rasters in Figure 1C. For each animal, we also plot the full 5-D parameter fits for all animals in Figure 5 – figure supplement 2. For the rest of the analysis, we focus on the mean slope and mean threshold parameters, which showed the most prominent changes. However we also note an increase in the slope s.d. between P14-16 and P30-40 WT animals that was not observed in *Fmr1* KO, mirroring the increased heterogeneity in firing rates in the same animals (Figure 1C). In Figure 5B we plot the mean slope and mean threshold fits on top of the previously calculated (Figure 4C) 2D slope-threshold maps of firing rate and correlation. We found that in young animals, P9—11, most points were scattered at high values of both slope and threshold (Figure 5B left). With age, the parameter fits for both genotypes moved south-west towards the low slope and low threshold region of parameter space (Figure 5B center and right). The mean location of the cloud of points at each developmental age differed between WT and KO. We plot the direction of shift in group mean from WT to KO in Figure 5C. In young animals, P9—11 and P14—16, the KO group had both higher slope and higher threshold than WT, whereas in adult animals, P30—40, the KO group had a lower slope and lower threshold than WT. These results demonstrate an opposite direction of circuit parameter change in young Fragile-X mice compared to adults, which was not be uncovered by measures of neural firing rates and correlations (Figure 1C).

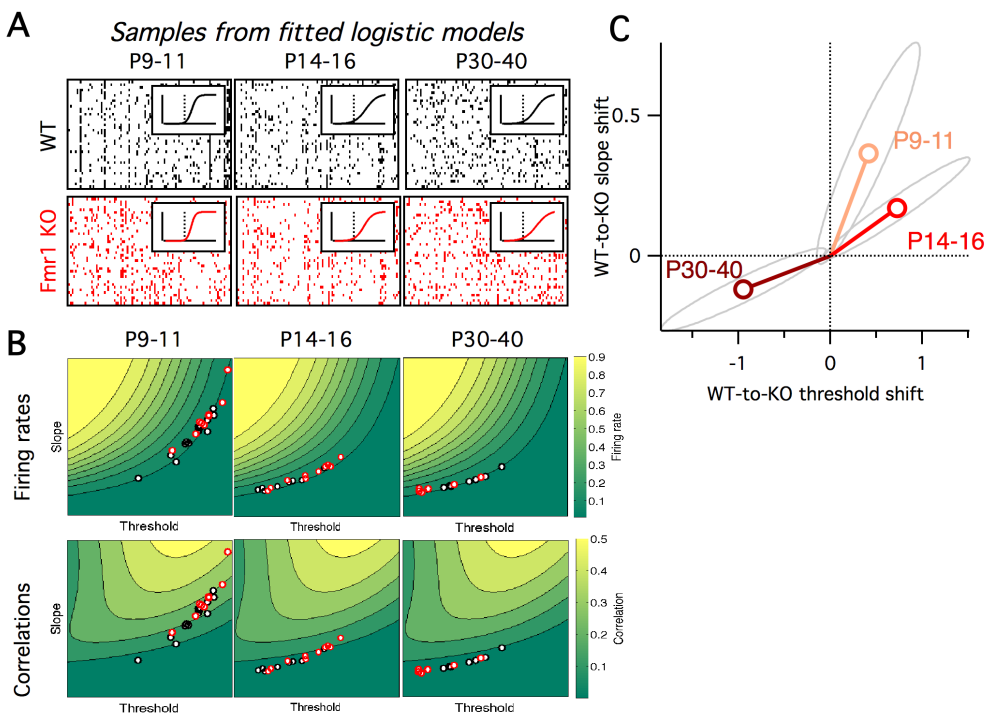

Figure 5. Fragile-X fits from logistic model.

A: Example samples from the fitted logistic models, corresponding to the six groups shown in panel A. Inset shows group mean fitted logistic function, dashed vertical line represents zero.

B: Fitted logistic mean slope and mean threshold values for data from each WT (black circles) and *Fmr1* KO (red circles) animal. Values overlaid on same firing rate (top) and correlation (bottom) maps from Figure 4C.

C: Shift in mean logistic slope and threshold values from WT to KO for P9—11 (orange), P14—16 (red) and P30—40 (brown). Grey ellipses represent 95% confidence intervals (Methods).

Earlier we asked how sensitive the logistic model slope and threshold parameters were to alterations in the many underlying neural circuit components (Figure 3). In a similar way, we can also ask how sensitive the neural firing rates and correlations are to alterations in the logistic slope and threshold parameters. This is important since inspection of the 2-dimensional maps in Figure 3C shows that these sensitivities will differ depending on starting location within the slope-threshold space. To quantify this effect, we calculated the sensitivity of both the firing rate and correlations to small changes in the slope and threshold (Figure 6A-B, see Methods), quantified as the partial derivatives local to the fitted logistic parameter values for each animal

(black and red circles in Figure 5B). In general, increasing the slope or decreasing the threshold always increased both firing rates and correlations, as can be predicted from Figure 5B. However, the magnitude of sensitivities varied across animals. We found only minor differences in sensitivities between genotypes (Figure 5 – figure supplement 2), and as a result we pooled the sensitivity measurements across genotypes to test for statistical differences in sensitivity with developmental age. In young animals, P9—11, changes in the logistic threshold (solid bars in Figure 6) had substantial effects on both firing rates and correlation. This sensitivity decreased with age ($p \leq 0.013$ for firing rates, $p < 0.01$ for correlations from P9—11 to P14—16), so that in adult animals, P30-40, changes in threshold had relatively little effect on neural activity statistics. A different picture emerged for the logistic slope parameter (striped bars in Figure 6). There, the firing rate sensitivity increased with from P9—11 to P14—16 ($p <$ 1e-6) (Figure 6A), while correlation sensitivity stayed approximately constant ($p \geq 0.18$) (Figure 6B). These results show that the quantitative relationships between neural activity statistics and the underlying circuit parameters are not fixed across development.

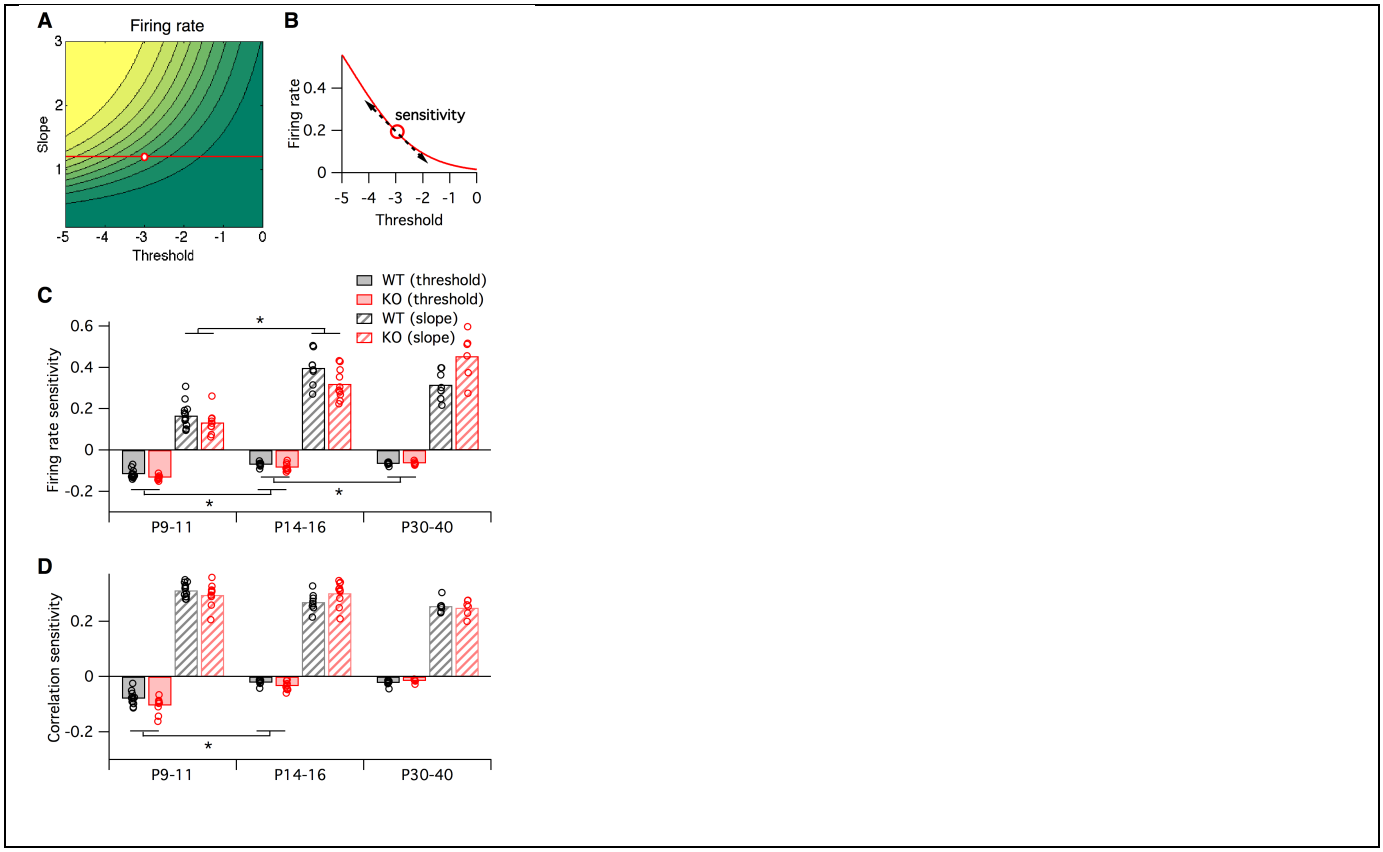

Figure 6. Sensitivity of firing rate and correlations with respect to logistic model parameters, local to the parameter fit for each animal.

A-B: Sensitivity to a parameter is calculated about a given point in parameter space. In this hypothetical example, we plot a slope-threshold parameter fit at the red circle on the firing rate contour map (A). The firing rate varies non-linearly if the threshold is varied away from this point (B). Sensitivity is calculated as the local derivative, or slope of the tangent, about the target point.

C-D: Sensitivity of firing probability (**C**) and pairwise correlations (**D**) to change in threshold (solid bars) and slope (striped bars) parameters of logistic model, about the fitted parameter values for each animal (circles) displayed in Figure 5B. Bars represent group means. Each statistical test compares the mean values between adjacent pairs of age groups, where the data were pooled between genotypes.

What are the functional implications of these alterations in firing rates and correlations in Fragile-X mice across development? To address this, we calculated the entropy of the neural population activity for the data from each animal. Entropy is a quantity from information theory, measured in bits, that puts a hard upper bound on the amount of information that can be represented by any coding system (Cover and Thomas, 2006). Intuitively, the entropy measures how uniform the neural population activity pattern distribution is: it is large if the circuit exhibits many different activity patterns over time, and small if only a few activity patterns dominate. Entropy is an appealing measure for the present problem because it is sensitive both to neural firing rates and to correlations at all orders. It is typically highest when firing rates are high and correlations are low. Although entropy is notoriously difficult to calculate for large neural populations because most estimation methods require impractically long data recordings (Quian Quiroga and Panzeri, 2009), we recently developed a new statistical method for this purpose, called the *population tracking model*, that scales well to large numbers of neurons, even for limited data (O'Donnell et al., 2017). This model matches both the synchrony distribution for the number of neurons simultaneously active, and the variations in individual cell-to-cell firing rates. We fit this population tracking model to the same Ca^{2+} imaging data as analyzed above (Figure 7). An intermediate step in estimating the neural entropy involves calculating a low-parameter approximation of the entire probability distribution

over all 2^N neural population activity patterns, where N is the number of neurons. The cumulatives of these probability distributions calculated for 50-neuron subsets of the recordings are shown in Figure 7A. In young animals P9—11, a small number of activity patterns accounts for a large fraction of the probability mass (Figure 7A left). For example, based on these curves, 50% of the time we would expect to see the same 1000—10,000 patterns out of a possible total $2^{50} \approx 10^{15}$ patterns. In contrast, in older animals P14—16 and P30—40 the cumulative distributions shift rightwards so that more patterns are typically observed (Figure 7A center, right). In these cases, around 1,000,000 patterns are needed to account for 50% probability mass.

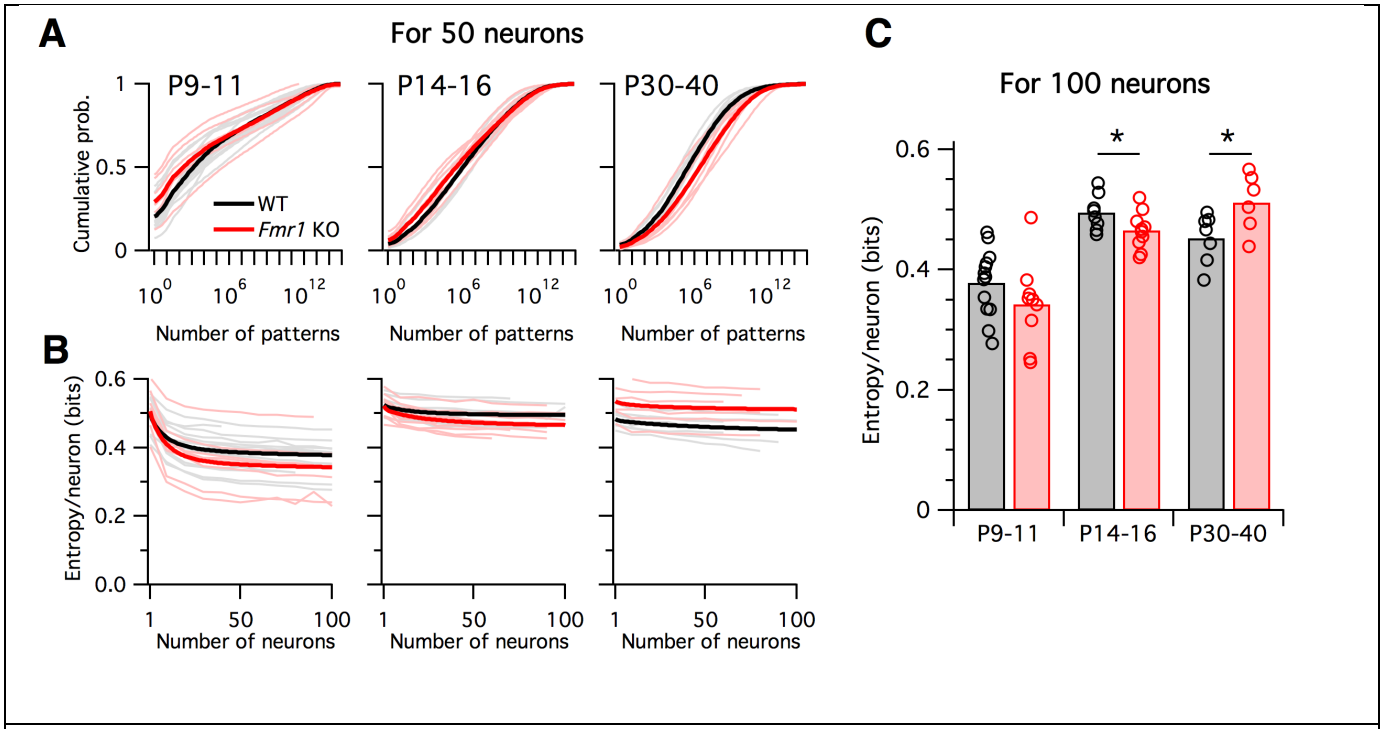

Figure 7. Differing trajectories of WT and KO entropy across development. **A:** Cumulative probability mass as a function of the number of patterns. Patterns ordered from most probable to least probable. Thin lines are mean across many randomly-chosen 50-neuron subsets from a given animal, and thick lines represent means across all animals of a given genotype.

B: Entropy per neuron as a function of the number of neurons analyzed. Thin lines are mean across many randomly chosen subsets for a given animal, thick lines are group mean of

double exponential fits to the data (see Methods). Age groups (left to right) are as in panel A. **C:** Estimated entropy/neuron for 100 neuron populations. Circles represent individual animals, bars are group means.

343

344 Instead of attempting to quantify these shifts by asking how many patterns are needed to cross
345 an arbitrary threshold of probability mass, we instead calculated the entropy, which takes into
346 account the shape of the entire probability distribution. The entropy depends on the number of
347 neurons analyzed, so we normalized all estimates to calculate the entropy per neuron (Figure
348 7B—C). Since we are treating neurons as binary, the entropy/neuron was bounded between 0
349 and 1 bits. For all age groups, and for both WT and *Fmr1* KO animals, entropy/neuron
350 progressively decreased with the number of neurons analyzed (Figure 7B). Because each
351 imaging session captured a different number of neurons (range 40—198, median 97), we fit
352 the entropy/neuron versus number of neurons data with a double exponential function (see
353 Methods) and use the fit to provide a standardized estimate of the entropy/neuron for 100-
354 neuron populations (Figure 7C). In WT animals, entropy/neuron showed a non-monotonic
355 trajectory across development (O'Donnell et al., 2017). At P9—11 it was low, 0.38 bits (95% c.i.
356 [0.35:0.41]), before increasing at P14—16 ($p < 0.001$) to 0.50 bits (95% c.i. [0.48:0.52]), before
357 decreasing again at P30—40 ($p = 0.028$) to 0.45 bits (95% c.i. [0.42:0.48]). We found a
358 different entropy trajectory in *Fmr1* KO animals. There, although entropy/neuron also began
359 low at 0.34 bits (95% c.i. [0.30:0.39]), not different from WT ($p= 0.19$), when it increased at
360 P14—16 ($p < 0.001$) to 0.465 bits (95% c.i. [0.45:0.48]) it remained lower than for WT ($p =$
361 0.048). Finally, instead of decreasing as in the WT case, entropy continued to increase in
362 P30—40 *Fmr1* KO animals ($p = 0.033$) to 0.51 bits (95% c.i. [0.47:0.55]), higher than WT ($p =$
363 0.034). These entropy values estimated directly from Ca^{2+} imaging data agreed well with
364 entropy estimates for synthetic data sampled from the previously fit logistic models (Figure 7 –
365 figure supplement 1). In summary, unlike WT animals, *Fmr1* KO mice showed a monotonically
366 increasing entropy/neuron from P9—11 to P30—40. Furthermore, the direction of change in
367 entropy between P14—16 and P30—40 was opposite for WT and *Fmr1* KO animals,
368 decreasing in the former and increasing in the latter.

Discussion

The 1-dimensional E/I imbalance model has been widely used for interpreting neural circuit changes observed in animal models of diverse brain disorders (Bateup et al., 2011; Dani et al., 2005; Gibson et al., 2008; Kehrer et al., 2008; Wallace et al., 2012). In the case of Fragile-X syndrome, the hyperexcitability prediction of the E/I imbalance model is consistent with many of the symptoms of the disease (e.g. seizures, hyperarousal, hyperactivity, hypersensitivity to sensory stimuli) and the known pathogenic defects implicated in *Fmr1* KO mice (diminished GABA signaling, exaggerated intrinsic excitability, increased neuronal firing rates; reviewed by Contractor et al., 2015). Here we tested the hypothesis that the E/I imbalance model can account for alterations in other neural activity statistics beyond the mean firing rates; however, our results demonstrated that it was inadequate. The model was too inflexible to account for the joint alterations in both neural firing rates and correlations observed in Fragile-X model mice. This suggests that future studies of brain disorders may need to consider higher-dimensional models of neural circuit dysfunction.

To test how cellular components affect their circuit function, we built computational models of mouse L2/3 somatosensory cortex at two levels of abstraction: a detailed, 100-parameter biophysical model, and a 2-parameter logistic response model. The purpose of the detailed model was to build a representation of the circuit where each parameter has a one-to-one mapping with something that could be experimentally measured in a real animal - indeed many of these parameters have been shown to be altered in *Fmr1* KO mice. The purpose of the logistic model was different: it simple enough to be both derivable from the complex model, and provide a direct link with measurable activity variables in our *in vivo* Ca2+ imaging data, firing rates and correlations. The disadvantages of detailed models are that they contain many parameters, and so are hard to constrain to data – in this case it was only possible because of the large dataset from Petersen et al., for P17-22 WT mice (Avermann et al., 2012; Lefort et al., 2009; Tomm et al., 2014). The disadvantages of the simple 2D model is that its logistic input-output structure implies a very strong and specific assumption about the functional purpose of the circuit – to generate single spikes across a subset of neurons. Although this may be a physiologically relevant computation for this particular brain circuit (Clancy et al., 2015; Kerr et

al., 2007; Sato et al., 2007), it is not immediately obvious how to extend this approach to include temporal correlations, for example, or to apply it to other brain circuits where we may have less insight into their natural computations. Nevertheless, our approach demonstrates a new way to tackle such problems.

After building the detailed computational model of L2/3 of mouse somatosensory cortex (Figure 2), and asked how sensitive the spiking responses of the overall circuit were to changes in its underlying neural components, many of which are known to be altered in *Fmr1* KO mice (Bureau et al., 2008; Gibson et al., 2008; Gonçalves et al., 2013; Harlow et al., 2010; Hays et al., 2011; Paluszkiewicz et al., 2011; Patel et al., 2013; Testa-Silva et al., 2012). We found that while changing some neural parameters did have a large effect, changing other parameters had little or no effect on circuit function (Figure 3B). This redundancy property has been reported as widely prevalent in computational models of biological systems (Gutenkunst et al., 2007; O'Leary et al., 2015). Its existence has two important implications for studies of brain disorders: first, many of the physiological component changes discovered in animal models may be entirely benign at the circuit level. Second, any treatment designed to correct circuit function is free to push the system by arbitrary amounts along insensitive directions in parameter space without consequence, as long as it makes the correct perturbations along the sensitive directions. The insensitive directions form a null space, which is a subspace of the parameter space.

An important caveat to our parameter sensitivity analysis is that it was linear and local to a particular point in the high-dimensional model parameter space, corresponding to WT P17-22 mice. Since the circuit dynamics are nonlinear, it is likely that the particular parameter sensitivities would be different in other parts of parameter space, especially near bifurcations where qualitatively different dynamics emerge (Hirsch et al., 2013). However, as long as the redundancy property is widely preserved, as suggested by studies on computational models of other biological systems (Fisher et al., 2013; Gutenkunst et al., 2007; Machta et al., 2013; Panas et al., 2015), then our conclusions for brain disorders remain valid.

425 In addition to the varying magnitudes of circuit components' effect on circuit function, we also
426 found that different components shifted the circuit input-output function in different directions,
427 as defined by our 2D logistic response model (Figures 1 and 2). Even circuit parameters that
428 are nominally of the same type, such as the strength of glutamatergic synapses between
429 excitatory (E) neurons in L4 to E neurons in L2/3 or synaptic strength between E neurons
430 within L2/3, had qualitatively different effects on the circuit response to stimulation (Figure 3).
431 According to the standard E/I imbalance model (Rubenstein and Merzenich, 2003), both of
432 these parameters should have similar effects on circuit function; but according to the logistic
433 response model we studied, their differing effects on slope and threshold parameters must
434 necessarily lead to different magnitudes of change in neural firing rates and correlations
435 (Figure 4C). Indeed, no 1-dimensional model of circuit function could ever capture the
436 heterogeneity in parameter sensitivities that we observed (Figure 3B).

437 Next, we fit the parameters of the logistic response model to match the *in vivo* firing statistics
438 of neural populations from WT and *Fmr1* KO mice of varying age (Figure 5). Previous studies
439 had found that neural correlations decrease during development (Golshani et al., 2009;
440 Rochefort et al., 2009), and that early postnatal *Fmr1* KO mice had higher correlations and
441 firing rates than WT mice (Gonçalves et al., 2013; La Fata et al., 2014). Circuit hypersynchrony
442 may be a general defect in autism disorders, as it is also found in mouse models of Rett
443 syndrome (Lu et al., 2016). However, the relationship between these changes in firing statistics
444 and the underlying neural circuit components were unclear. Our logistic model helps bridge
445 this gap, leading to two findings: first, the direction of circuit parameter change from WT to KO
446 was opposite in young (P9—11 and P14—16) versus mature (P30—40) animals (Figure 5C).
447 Similar opposing switches in sensory cortex properties with age were also recently reported in
448 *Fmr1* KO and WT rats (Berzhanskaya et al., 2016). Second, we found that the sensitivity of
449 neural firing rates and correlations to changes in underlying circuit components depends on
450 developmental age (Figure 6). Taken together, these findings imply that qualitatively different
451 interventions may be needed at different stages of development in Fragile-X, and perhaps
452 other neurodevelopmental disorders, to shift cortical circuit function towards typical wild-type
453 operation.

454 Spontaneous, intrinsic activity is ubiquitously present in mammalian cerebral cortex. It is highly
455 structured at multiple spatiotemporal scales (Mitra et al., 2015; Ringach, 2009) and interacts
456 strongly with the signals evoked by sensory stimulation (Ringach, 2009). Cellular-resolution
457 recordings in animals have shown that the patterns of spontaneous activity in neural
458 populations are representative of the ensemble of activity patterns used by the brain to
459 represent sensory stimuli (Berkes et al., 2011; Luczak et al., 2009; Miller et al., 2014). Here we
460 found that the entropy of spontaneous activity in WT mouse somatosensory cortex follows an
461 inverted-U shaped trajectory across development, and that this trajectory is dramatically
462 altered in the *Fmr1* KO mouse model of Fragile-X (Figure 7). Although we saw no reliable
463 differences across genotypes in early postnatal animals (P9—11), *Fmr1* KO animals showed
464 lower entropy than WT after the second postnatal week (P14—16), while surprisingly switching
465 to show higher entropy than WT in adult (P30—40). Notably, this switch in the direction of
466 entropy change from WT to KO during development mirrors the reversing we saw in logistic
467 model parameter changes in Figure 5C. Together, these findings suggest a perturbed
468 trajectory of cortical development during the critical period in *Fmr1* KO mice (Meredith et al.,
469 2012). However, our results cannot distinguish whether the observed perturbation in L2/3
470 activity statistics reflects a developmental delay, or a permanently altered developmental
471 trajectory. Further studies at later developmental time points are needed.

472 What is the functional significance of these shifts in population entropy? Previous work
473 suggested that the entropy of neural circuit activity may be optimally tuned at intermediate
474 levels as a trade-off between maximizing representational capacity at high entropy, versus
475 maintaining error correction and regularization at low entropy (Schneidman et al., 2006). These
476 properties can also be thought of as trading off between discrimination and generalization,
477 respectively (Qian and Lipkin, 2011). If we assume that WT mice are optimally tuned, our
478 findings predict that young *Fmr1* KO mice should show poorer somatosensory discrimination in
479 behavioral tasks than wild-type animals, while in contrast adult *Fmr1* KO mice should perform
480 more poorly on tasks involving generalization across somatosensory stimuli.

481 If the unidimensional E/I imbalance model is not sufficiently rich to capture the circuit changes
482 observed in neurodevelopmental disorders, what should replace it? How many dimensions or

483 degrees of freedom should a working model for a brain disorder have? Theoretical
484 neuroscientists have long studied E/I balance in generic models of recurrent neural circuits
485 (Brunel, 2000; Tsodyks and Sejnowski, 1995). These models have uncovered important
486 distinctions between 'loose' balanced regimes, where E and I inputs to a neuron are equal only
487 on average, and fine-tuned 'tight' balanced regimes where E and I inputs to a neuron track
488 each other closely on fast timescales (Deneve and Machens, 2016; Hennequin et al., 2017). In
489 principle these generic models could be used to investigate multidimensional E/I imbalances in
490 brain disorders (Vogels and Abbott, 2007). However, it is currently difficult to directly fit these
491 many-parameter network models to data (although see Arakaki et al., 2017; Fisher et al., 2013;
492 Stringer et al., 2016), and they are agnostic to circuit function. Instead we suggest an
493 alternative, complementary approach: start by assuming a computational function for the
494 particular neural circuit under study, then work backwards to design a model that is both
495 sophisticated enough to capture the key information processing features of the circuit, but
496 simple enough to interpret and link to physiological data. In this study we considered a 2-
497 parameter model of L2/3 somatosensory cortex's input-output function, which could account
498 for both neural firing rates and correlations. Other brain circuits may demand models with more
499 degrees of freedom. Crucially, the most informative models need not be those that include the
500 highest level of physiological detail. All models are ultimately wrong in the sense that they
501 make abstractions about their underlying parts, and detailed models carry the additional
502 burden of fitting many parameters, which may be difficult to adequately constrain (O'Leary et
503 al., 2015). Nonetheless, some models are useful (Box, 1979).

504 One potential use of simple parametric circuit models such as the ones we employed here may
505 be as a tool for rationally designing candidate intervention compounds and then screening their
506 effects on neural population activity. For example, the current study could have been extended
507 to fit the logistic model to neural activity data from another cohort of *Fmr1* KO mice that had
508 received a candidate treatment, then ask if the fitted model parameters were closer in value to
509 those from WT animals or *Fmr1* KO controls. Approaches like this could complement the
510 traditional strategy of designing drugs based on reversing molecular deficits and then
511 assessing the drug's impact on animal model behavior. Indeed, our results suggest that given

the multi-dimensionality of circuit properties, it may prove difficult or impossible to find a single compound that can correctly reverse deficits at any age. This scenario might require a combination of drugs chosen to push circuit-level properties towards the 'correct' region of parameter space. The framework we have introduced in this study can facilitate this type of high-dimensional intervention analysis for diverse neurodevelopmental disorders.

517 ## Materials and Methods

Key Resources Table			
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**
strain, strain background (mus musculus)	c57bl/6J strain of wild type mice	Jackson Labs	IMSR_JAX:000664
genetic reagent (mus musculus)	Fmr1 knockout mouse on a c57 background	William Greenough (originally from Dutch-Belgian Fragile X Consortium)	RRID:MGI:2665400
chemical compound, drug	OGB1 AM (Oregon Green BAPTA-1 AM)	Molecular Probes (ThermoFisher Scientific)	
Software, algorithm	ImageJ	NIH	RRID:SCR_003070
software, algorithm	BRIAN Simulator	http://briansimulator.org	RRID:SCR_002998
software, algorithm	MATLAB	http://www.mathworks.com/products/matlab	RRID:SCR_001622

518 ***Mouse in vivo calcium imaging recording***

519 All Ca^{2+} imaging data were published previously (Gonçalves et al., 2013). Briefly, data were
520 collected from male and female C57Bl/6 wild-type and *Fmr1* KO mice at P9–40. For each
521 group the animal numbers were: P9-11, n=13 WT and n=9 *Fmr1* KO; P14-16, n=8 WT and
522 n=10 Fmr1 KO; P30-40, n=7 WT and n=6 *Fmr1* KO. There were variations in the number of
523 cells recorded from each animal. The range of cell numbers for each group were: P9-11, 49-
524 198 cells in WT and 84-144 cells in *Fmr1* KO; P14-16, 65-119 cells in WT and 40-149 cells in
525 *Fmr1* KO; P30-40, 60-114 cells in WT and 69-105 cells in *Fmr1* KO. Mice were anesthetized
526 with isoflurane, and a cranial window was fitted over primary somatosensory cortex by
527 stereotaxic coordinates. Mice were then transferred to a two-photon microscope and headfixed
528 to the stage while still under isoflurane anesthesia. 2—4 injections of the Ca^{2+} sensitive
529 Oregon-Green BAPTA-1 (OGB) dye and sulforhodamine-101 (to visualize astrocytes) were
530 injected 200 um below the dura. Calcium imaging was performed using a Ti-Sapphire
531 Chameleon Ultra II laser (Coherent) tuned to 800 nm. Imaging in unanesthetized mice began
532 within 30-60 min of stopping the flow of isoflurane after the last OGB injection. Images were
533 acquired using ScanImage software (Pologruto et al., 2004) written in MATLAB (MathWorks;

534 RRID:SCR_001622). Whole-field images were collected using a 20× 0.95 NA objective
535 (Olympus) at an acquisition speed of 3.9 Hz (512 × 128 pixels). Several 3-minute movies were
536 concatenated and brief segments of motion artifacts were removed (always <10 s total). Data
537 were corrected for x—y drift. Cell contours were automatically detected and the average ΔF/F
538 signal of each cell body was calculated at each time point. Each ΔF/F trace was low-pass
539 filtered using a Butterworth filter (coefficient of 0.16) and deconvolved with a 2 s single-
540 exponential kernel (Yaksi and Friedrich, 2006). To remove baseline noise, the standard
541 deviation of all points below zero in each deconvolved trace was calculated, multiplied by two,
542 and set as the positive threshold level below which all points in the deconvolved trace were set
543 to zero. Estimated firing rates of the neurons, r_i(t), were then obtained by multiplying the
544 deconvolved trace by a factor previously derived empirically from cell-attached recordings in
545 vivo (Golshani et al., 2009).

546 **Computational methods**

547 Data analysis and logistic model calculations were done using MATLAB (Mathworks;
548 RRID:SCR_001622). All simulation code is available online at
549 https://github.com/cianodonnell/ODonnelletal_2017_imbalances, and the population tracking
550 model code (O'Donnell et al., 2017) is available at
551 https://github.com/cianodonnell/PopulationTracking.

552 **Detailed layer 2/3 model simulations**

553 Layer 2/3 model simulations (Figures 1 and 2) were implemented with the Python-based
554 simulator Brian 2 (http://briansimulator.org; RRID:SCR_002998) (Goodman and Brette, 2009),
555 and results analyzed with MATLAB (Mathworks; RRID:SCR_001622). The model consisted of
556 four populations of reciprocally connected leaky integrate-and-fire neurons representing a L2/3
557 somatosensory barrel circuit: 1700 excitatory neurons, 70 PV inhibitory neurons, 115 $5HT_{3A}R$
558 inhibitory neurons, and 45 SOM inhibitory neurons, driven by a separate population of 1500
559 excitatory spike sources representing input from L4. Cell numbers were estimated by
560 combining layer-specific excitatory and inhibitory cell count information from (Lefort et al.,
561 2009) with the approximate percentages of the three inhibitory cell groups given by (Petersen
562 and Crochet, 2013). The voltage V of each neuron evolved as

$$\frac{dV}{dt} = (R_{in}(g_e(E_{rev,e} - V) + g_i(E_{rev,i} - V)) - (V - V_{rest}))/\tau_m$$

where R_{in} is the input resistance, $E_{rev,e}$ and $E_{rev,i}$ are the excitatory and inhibitory synaptic reversal potentials respectively, τ_m is the membrane time constant, and g_e and g_i are the summed excitatory and inhibitory synaptic input conductances respectively. Between input events the total excitatory synaptic condunctance g_e evolved in time according to the equation

$$\frac{dg_e}{dt} = -g_e/\tau_{syn,e}$$

where $\tau_{syn,e}$ is the excitatory synaptic time constant. Similar equations governed the inhibitory conductances. When a spike arrived at a synapse, a Bernoulli random number was drawn with release probability set according to the particular synaptic connection type. If this number was equal to one, then the total synaptic conductance for that neuron was instantaneously incremented by the specific amplitude of the chosen conductance for that individual synapse, indexed *j*: $g_e \rightarrow g_e + \bar{g}_j$.

All synaptic connections were formed probabilistically by drawing independent random Bernoulli variables with connection type-specific probabilities. Synaptic PSP amplitudes were drawn independently for each synapse from a log-normal distribution constrained by the experimentally reported mean and median values for each particular connection type. The maximum post-synaptic potential amplitude was set to 8 mV. Synapses in the model were conductance-based, but since synaptic strengths reported in the literature were typically in terms of EPSP/IPSP amplitude, in accordance with how the experiments were performed (Avermann et al., 2012), we set each maximal synaptic conductance as the value needed to generate a PSP of the desired amplitude when the target neuron started at resting potential in the case of EPSPs or -55mV in the case of IPSPs, which we computed analytically. Refractory periods were calculated as the inter-spike-interval corresponding to the maximal experimentally reported firing rate. Release probability and synaptic strength values for unconnected neurons are excluded from Table 1. Excitatory synaptic time constants were set at 2 ms, which is typical for the fast component of AMPA receptor responses, but could not be estimated from the PSP statistics in (Avermann et al., 2012) because of masking by the slower

membrane time constant. The mathematical form of our model meant that inhibitory synaptic time constants needed to be equal for all incoming inhibitory synapses to a neuron. We set these to 40 ms for E, $5HT_{3A}R$ and SOM neurons and 16 ms for PV neurons, which were the typical values of the IPSP decay time constants in the (Avermann et al., 2012) dataset. Due to lack of direct data for this circuit, connection probabilities for synapses from L4 E neurons to E, PV and SOM L2/3 neurons was set to a reasonable cortical value of 0.15, while $5HT_{3A}R$ neurons did not receive any input from L4 (Gentet et al., 2012). Similarly due to a lack of direct data, we set synaptic release probabilities for connections from L4 to L2/3 neurons to a typical cortical value of 0.25, while mean and median L4 excitatory PSP amplitudes onto L2/3 PV and SOM were set to 0.8 and 0.48 mV, respectively, to match reported data for L4 EPSP amplitudes onto L2/3 E neurons (Lefort et al., 2009). The differential equations were solved using the forward Euler method with an integration timestep of 0.01 ms. Each simulation run was 50 ms long, during which we recorded whether or not each neuron responded. In the rare cases where a neuron spiked more than once, we disregarded the extra spikes. L4 neuron dynamics were not explicitly simulated, but instead modeled only as a set of output spike trains. After selecting the subset of active L4 neurons, spike times were drawn randomly from a Gaussian distribution with standard deviation of 2 ms. We repeated the simulations 10 times for this identical input pattern to average over the noise due to probabilistic vesicle release. We repeated this procedure further 10 times for different random allocations of the 'ON' inputs. Then, a neuron's ON probability was defined as the fraction of these $10 \times 10 = 100$ simulations for which it responded with one or more spikes. Finally, we repeated the entire procedure for varying levels of L4 input sparsity.

For the simulations presented in Figure 3 we varied only 76 model parameters, which is 24 less than the total number of 100 model parameters listed in Table 1. We excluded the four neuronal refractory periods (because in almost all simulations each neuron spiked a maximum of once, making the refractory period irrelevant), and the six connection probabilities that were fixed at zero. Finally, we grouped together the mean and median PSP amplitudes for each of the fourteen non-zero synaptic connections, so that both parameters were increased or

decreased by the same fraction in tandem. Together these choices reduced the number of test parameters from 100 to 76.

For all parameters that naturally range from 0 upwards, such as the number of neurons or release probability, we increased or decreased their values during testing in the most intuitive way, by adding +/- 20% of the baseline value. However, this method was less useful for other parameters, such as cell resting voltage, for which we reasoned it made more sense to scale relative to another parameter, such as spike threshold. As a result, we varied 1) resting voltage relative to its difference from spike threshold; 2) spike threshold relative to its difference with resting voltage; 3) excitatory synaptic reversal potentials relative to resting voltage; 4) inhibitory synaptic reversal potentials relative to spike threshold.

Table 1. L2/3 computational circuit model parameters and mean slope & threshold shifts.

Parameter	Value	Source	+20% effect on slope, thresh	Parameter	Value	Source	+20% effect on slope, thresh
N_E	1700	[1]	slope: 6.4 x10^{-3} thresh: 5.16	$pcon_{I5htE}$	0.465	[2]	slope: -3.25 x10^{-3} thresh: -0.103
N_{Ipv}	70	[1]	slope: 4.3 x10^{-3} thresh: -1.51	$pcon_{I5htIpv}$	0.38	[2]	slope: -7.13 x10^{-3} thresh: -0.298
N_{I5ht}	115	[1]	slope: -0.0104 thresh: -0.58	$pcon_{I5htI5ht}$	0.38	[2]	slope: 6.07 x10^{-4} thresh: 0.175
N_{Isom}	45	[1]	slope: -0.001 thresh: -0.83	$pcon_{I5htIsom}$	0	No data	Not tested
N_{EL4}	1500	[1]	slope: -0.106 thresh: 3.43	$pcon_{IsomE}$	0.5	[5]	slope: -6.29 x10^{-3} thresh: -0.8919
$Vrest_E$	-68 mV	[2]	slope: 0.049 thresh: -6.09	$pcon_{IsomIpv}$	0	No data	Not tested
$Vrest_{Ipv}$	-68 mV	[2]	slope: -0.016 thresh: -0.963	$pcon_{IsomI5ht}$	0	No data	Not tested
$Vrest_{I5ht}$	-62 mV	[2]	slope: -9.48 x10^{-4} thresh: 0.205	$pcon_{IsomIsom}$	0	No data	Not tested
$Vrest_{Isom}$	-57 mV	[3]	slope: 3.58 x10^{-3} thresh: 0.034	$prel_{EL4E}$	0.25	No data	slope: -0.112 thresh: 4.21
Vth_E	-38 mV	[2]	slope: -3.76 x10^{-3} thresh: -0.381	$prel_{EL4Ipv}$	0.25	No data	slope: 0.0103 thresh: 0.682
Vth_{Ipv}	-37.4 mV	[2]	slope: 3.87 x10^{-3} thresh: 0.221	$prel_{EL4Isom}$	0.25	No data	slope: -2.06 x10^{-3} thresh: -0.537
Vth_{I5ht}	-36 mV	[2]	slope: -1.12 x10^{-3} thresh: 0.013	$prel_{EE}$	0.25	No data	slope: 4.99 x10^{-3} thresh: 6.074
Vth_{Isom}	-40 mV	[3]	slope: -3.26 x10^{-3} thresh: -0.083	$prel_{EIpv}$	0.25	No data	slope: -4.04 x10^{-4} thresh: 0.163
Rin_E	160 MΩ	[2]	slope: 1.95 x10^{-3} thresh: -0.283	$prel_{EI5ht}$	0.25	No data	slope: -9.32 x10^{-3} thresh: -0.532
Rin_{Ipv}	100 MΩ	[2]	slope: -8.38 x10^{-3} thresh: -0.283	$prel_{EIsom}$	0.25	No data	slope: -7.32 x10^{-3} thresh: -0.3847
Rin_{I5ht}	200 MΩ	[2]	slope: -3.87 x10^{-3} thresh: -0.653	$prel_{IpvE}$	0.25	No data	slope: 1.03 x10^{-2} thresh: -0.941
Rin_{Isom}	250 MΩ	[4]	slope: 7.55 x10^{-3} thresh: 0.465	$prel_{IpvIpv}$	0.25	No data	slope: 1.21 x10^{-3} thresh: -0.061

τm_E	28 ms	[2]	slope: -9.59 x10^{-3} thresh: 0.268	$prel_{lpvl5ht}$	0.25	No data	slope: -4.73 x10^{-3} thresh: 0.011
τm_{lpv}	21 ms	[2]	slope: 4.82 x10^{-3} thresh: 0.027	$prel_{l5htE}$	0.25	No data	slope: -1.63 x10^{-3} thresh: -0.379
τm_{l5ht}	10 ms	[2]	slope: -5.09 x10^{-3} thresh: -0.302	$prel_{l5htlpv}$	0.25	No data	slope: -3.41 x10^{-3} thresh: -0.262
τm_{lsom}	30 ms	[4]	slope: -2.89 x10^{-3} thresh: -0.159	$prel_{l5htl5ht}$	0.25	No data	slope: 3.05 x10^{-3} thresh: 0.123
$tref_E$	55.5 ms	[2]	Not tested	$p_{rel,lsomE}$	0.25	No data	slope: -2.13 x10^{-4} thresh: -0.65
$tref_{lpv}$	5.4 ms	[2]	Not tested	$w_{EL4E,mean}$	0.8 mV	[1]	slope: -0.142 thresh: 0.342
$tref_{l5ht}$	21.3 ms	[2]	Not tested	$w_{EL4E,median}$	0.48 mV	[1]	slope: -0.142 thresh: 0.342
$tref_{lsom}$	20 ms	[3]	Not tested	$w_{EL4lpv,mean}$	0.8 mV	$=w_{EL4E}$	slope: 3.61 x10^{-3} thresh: 6.54 x10^{-3}
$\tau syn_{E,e}$	2 ms	Typical	slope: 1.37 x10^{-2} thresh: -1.79	$w_{EL4lpv,median}$	0.48 mV	$=w_{EL4E}$	slope: 3.61 x10^{-3} thresh: 6.54 x10^{-3}
$\tau syn_{E,i}$	40 ms	[2]	slope: -7.29 x10^{-3} thresh: 0.48	$w_{EL4lsom,mean}$	0.8 mV	$=w_{EL4E}$	slope: 5.05 x10^{-3} thresh: -0.329
$\tau syn_{lpv,e}$	2 ms	Typical	slope: -9.79 x10^{-3} thresh: -0.477	$w_{EL4lsom,median}$	0.48 mV	$=w_{EL4E}$	slope: 5.05 x10^{-3} thresh: -0.329
$\tau syn_{lpv,i}$	16 ms	[2]	slope: 1.56 x10^{-3} thresh: -0.097	$w_{EE,mean}$	0.37 mV	[2]	slope: 7.44 x10^{-3} thresh: 5.34
$\tau syn_{l5ht,e}$	2 ms	Typical	slope: 4.52 x10^{-3} thresh: -0.047	$w_{EE,median}$	0.2 mV	[2]	slope: 7.44 x10^{-3} thresh: 5.34
$\tau syn_{l5ht,i}$	40 ms	[2]	slope: -3.82 x10^{-3} thresh: -0.387	$w_{Elpv,mean}$	0.82 mV	[2]	slope: -3.77 x10^{-4} thresh: -0.297
$\tau syn_{lsom,e}$	2 ms	Typical	slope: -0.0126 thresh: -8.82 x10^{-3}	$w_{Elpv,median}$	0.68 mV	[2]	slope: -3.77 x10^{-4} thresh: -0.297
$\tau syn_{lsom,i}$	40 ms	[2]	slope: -4.88 x10^{-3} thresh: -0.301	$w_{El5ht,mean}$	0.39 mV	[2]	slope: -6.81 x10^{-3} thresh: -0.46
$Erev_e$	0 mV	Typical	slope: -0.056 thresh: 1.53	$w_{El5ht,median}$	0.19 mV	[2]	slope: -6.81 x10^{-3} thresh: -0.46
$Erev_{Ei}$	-68 mV	$=Vrest_E$	slope: 3.2x10^{-3} thresh: -3.77	$w_{Elsom,mean}$	0.5 mV	No data	slope: -3.61 x10^{-3} thresh: -0.359
$Erev_{lpvi}$	-68 mV	$=Vrest_{lpv}$	slope: -0.010 thresh: -0.617	$w_{Elsom,median}$	0.4 mV	No data	slope: -3.61 x10^{-3} thresh: -0.359
$Erev_{l5hti}$	-62 mV	$=Vrest_{l5ht}$	slope: 3.8x10^{-3} thresh: 0.132	$w_{lpvE,mean}$	0.52 mV	[2]	slope: 6.41 x10^{-3} thresh: -1.47
$Erev_{lsomi}$	-57 mV	$=Vrest_{lsom}$	slope: -4.2x10^{-3} thresh: -0.088	$w_{lpvE,median}$	-0.29 mV	[2]	slope: 6.41 x10^{-3} thresh: -1.47
$pcon_{EL4E}$	0.15	No data	slope: -0.121 thresh: 2.99	$w_{lpvlpv,mean}$	-0.56 mV	[2]	slope: -2.52 x10^{-3} thresh: -0.345
$pcon_{EL4lpv}$	0.15	No data	slope: 1.38x10^{-3} thresh: 0.029	$w_{lpvlpv,median}$	-0.44 mV	[2]	slope: -2.52 x10^{-3} thresh: -0.345
$pcon_{EL4l5ht}$	0	No data	Not tested	$w_{lpvl5ht,mean}$	-0.83 mV	[2]	slope: -4.24 x10^{-3} thresh: -0.266
$pcon_{EL4lsom}$	0.15	No data	slope: 7.62 x10^{-3} thresh: -0.245	$w_{lpvl5ht,median}$	-0.6 mV	[2]	slope: -4.24 x10^{-3} thresh: -0.266
$pcon_{EE}$	0.17	[2]	slope: 5.86 x10^{-3} thresh: 6.084	$w_{l5htE,mean}$	-0.49 mV	[2]	slope: -3.61 x10^{-4} thresh: -0.018
$pcon_{Elpv}$	0.575	[2]	slope: -1.17 x10^{-3} thresh: -0.099	$w_{l5htE,median}$	-0.3 mV	[2]	slope: -3.61 x10^{-4} thresh: -0.018
$pcon_{El5ht}$	0.24	[2]	slope: -6.44 x10^{-3} thresh: -0.541	$w_{l5htlpv,mean}$	-0.49 mV	[2]	slope: -1.77 x10^{-3} thresh: -0.187
$pcon_{Elsom}$	0.5	[5]	slope: -4.37 x10^{-3} thresh: -0.27	$w_{l5htlpv,median}$	-0.15 mV	[2]	slope: -1.77 x10^{-3} thresh: -0.187

$pcon_{lpvE}$	0.6	[2]	slope: 7.16 x10^{-3} thresh: -1.049	$w_{l5htl5ht,mean}$	-0.37 mV	[2]	slope: -4.12 x10^{-3} thresh: -0.416
$pcon_{lpvlpv}$	0.55	[2]	slope: -2.61 x10^{-3} thresh: -0.0456	$w_{l5htl5ht,median}$	-0.23 mV	[2]	slope: -4.12 x10^{-3} thresh: -0.416
$pcon_{lpvl5ht}$	0.24	[2]	slope: -2.81 x10^{-3} thresh: -0.458	$w_{lsomE,mean}$	-0.5 mV	No data	slope: -0.013 thresh: -0.984
$pcon_{lpvlsom}$	0	No data	Not tested	$w_{lsomE,median}$	-0.4 mV	No data	slope: -0.013 thresh: -0.984

Source [1] is (Lefort et al., 2009), [2] is (Avermann et al., 2012), [3] is (Fanselow et al., 2008), [4] is (Kinnischtzke et al., 2012), [5] is (Fino and Yuste, 2011). *N* is number of neurons, *Vrest* is resting potential, *Vth* is spike voltage threshold, *Rin* is input resistance, *tref* is refractory period, *τm* is the membrane time constant, *τsyn* is the synaptic time constant with the first subscript indicating the postsynaptic neuron type and the second subscript the neurotransmitter type of the presynaptic neuron (*e* or *i*), *Erev* is the synaptic reversal potential, *pcon* is the synaptic connection probability, *prel* is the synaptic release probability, *w* is the mean or median post-synaptic potential amplitude as indicated. For all neuronal parameters, the subscript indicates the neuron type: *E* is L2/3 excitatory neurons, *lpv* is PV neurons, *l5ht* is 5HT$_{3A}$R neurons, *lsom* is SOM neurons, and *EL4* is L4 excitatory neurons. For synaptic parameters, the first and second subscripts indicate the pre- and post-synaptic neuron types, respectively.

An important caveat is that although this model may be considered detailed by some measures, it also simplifies many aspects of L2/3 circuit. For example, we assumed that all 5HT$_{3A}$R cells were homogeneous, even though they likely separate into different subclasses with type-specific connectivity (Gentet, 2012; Petersen and Crochet, 2013). Layer 2 and layer 3 may also consist of distinct cell populations (Petersen and Crochet, 2013). Not all likely connections were included in the model (Dalezios et al., 2002; Pfeffer et al., 2013), and connectivity was assumed to be random, even though it is likely non-random (Tomm et al., 2014). Although these choices will likely not affect the conclusions of the current study, they may be important to consider for future work that seeks to understand the biological function of the L2/3 somatosensory microcircuit.

Logistic model

From the L2/3 circuit model simulations, we numerically estimated the probability *q* that each neuron in the model fires a spike as a function of the fraction of L4 inputs that were active, *f*. We then used the generalized linear model regression tool 'glmfit' in MATLAB to find the best

fit of the two logistic model parameters for each neuron:

$$q(f) = \frac{1}{1+\exp\left(-\beta(f-f_{1/2})\right)},$$

where the parameter β represents the slope, and the parameter $f_{1/2}$ represents the fraction of active L4 neurons at which the response probability $q = 0.5$. For clarity of presentation, in the main text we converted this $f_{1/2}$ parameter to what we termed the 'threshold', f_{thresh}, which we defined as the fraction of L4 neurons needed to reach a specified spike probability, q_{thresh}. Throughout the study we fixed $q_{thresh} = 0.01$. The threshold is related to $f_{1/2}$ via the inverse of the logistic function

$$f_{thresh} = f_{1/2} + \log\left(\frac{q_{thresh}}{1-q_{thresh}}\right)/\beta.$$

We computed firing rates and pairwise correlations from the logistic model (Figures 3—4) in the following way. First, we assumed that the fraction of active L4 neurons is described by a normally distributed random variable with zero mean and unit variance:

$$p(f) = \frac{\exp\left(-f^2/2\right)}{\sqrt{2\pi}} = \mathcal{N}(0,1).$$

We defined the β and $f_{1/2}$ parameters relative to the mean and standard deviation of the input distribution. Since q is a monotonically increasing function of f, the probability distribution for q is

$$p(q) = p(f(q))\left|\frac{df}{dq}\right|$$

where $f(q)$ is the inverse of the logistic function $q(f)$ and

$\frac{df}{dq} = \frac{\left(\exp\left(-\beta(f-f_{1/2})\right)+1\right)^2}{\beta\exp(-\beta(f-f_{1/2}))}$. We calculate a neuron's mean firing rate μ as the expectation of q,

$$\mu = \mathbb{E}[q] = \int_0^1 [q \times p(q)]dq = \int_0^1 \left[q \times p(f(q))\left|\frac{df}{dq}\right|\right] dq.$$

We calculate the pairwise covariance of two homogeneous neurons driven by a common input f as

$\mathrm{cov} = \mathbb{E}[q^2] - (\mathbb{E}[q])^2 = \mathbb{E}[q^2] - \mu^2 = \int_0^1 \left[q^2 \times p\left(f(q)\right)\left|\frac{df}{dq}\right|\right] dq - \mu^2$, then find the pairwise correlation by normalizing the covariance by the neurons' shared variance, $\mathrm{var} = \mu(1-\mu)$.

For fitting the logistic model to the recorded neural firing rates and correlations (Figure 5), we considered a population model where the joint probability distribution across threshold and slope was specified by a 2D Gaussian, which has five parameters: threshold mean and s.d., slope mean and s.d., and slope-threshold correlation. The three constraint statistics we considered from the neural population data were the mean neural ON probability, the s.d. of neural ON probabilities, and the mean pairwise correlations. We found the best-fit model parameters for each dataset using stochastic gradient descent (code available at https://github.com/cianodonnell/ODonnelletal_2017_imbalances). Briefly, the fitting procedure followed: 1) initialize the parameters at a starting guess points, 2) compute the predicted three output firing statistics via numerical integration over the model's probability distributions, 3) compute the fitting error as the summed squared difference between the model output predictions and the target values, 4) generate a new set of parameter values by adding a small perturbation of a zero-mean Gaussian random number to each parameter, 5) compute the new output statistics, 5) recompute the fitting error, 6) if the new error is smaller than the old error, accept the updated parameter values, otherwise reject them and revert to the old parameters, 7) return to step 4 unless the error is lower than the desired tolerance. We checked for fit convergence by sampling a large number of logistic model parameters from the fitted 2D Gaussian, drawing binary samples from these logistic model neurons and computing the ON probability mean and s.d., and mean pairwise correlation from the synthetic binary samples, and comparing the computed statistical values to the original data statistics (Figure 5 – supplemental figure 1). For the sensitivity analysis presented in Figure 6, we numerically computed the partial derivative in mean firing rate and pairwise correlation with respect to the mean slope and mean threshold parameters in the population logistical model, using standard finite difference methods.

Statistical tests

To avoid parametric assumptions, all statistical tests were done using standard bootstrapping methods with custom-written MATLAB scripts. For example when assessing the observed difference between two group means $\Delta\mu_{obs}$ we performed the following procedure to calculate a p-value. First we pool the data points from the two groups to create a null set S_{null}. We then

construct two hypothetical groups of samples S_1 and S_2 from this by randomly drawing n_1 and n_2 samples with replacement from S_{null}, where n_1 and n_2 are the number of data points in the original groups 1 and 2 respectively. We take the mean of both hypothetical sets $\Delta\mu_1$ and $\Delta\mu_2$ and calculate their difference $\Delta\mu_{null} = \Delta\mu_1 - \Delta\mu_2$. We then repeat the entire procedure 10^7 times to build up a histogram of $\Delta\mu_{null}$. This distribution is always centered at zero. After normalizing, this can be interpreted as the probability distribution $f(\Delta\mu_{null})$ for observing a group mean difference of $\Delta\mu_{null}$ purely by chance if the data were actually sampled from the same null distribution. Then the final p-value for the probability of finding a group difference of at least $\Delta\mu_{obs}$ in either direction is given by $p = \int_{-\infty}^{-\Delta\mu_{obs}} f(\Delta\mu_{null})d\Delta\mu_{null} + \int_{\Delta\mu_{obs}}^{\infty} f(\Delta\mu_{null})d\Delta\mu_{null}$. For Figure 5C we estimated 2-dimensional 95% confidence ellipses for the shift in mean slope-threshold parameters between Fmr1 KO and WT by computing the sample error variances and covariance through bootstrapping. Then the 95% confidence ellipse can be computed using the Chi-squared distribution. We plotted the confidence interval ellipse using the MATLAB function error_ellipse.m, downloaded from https://www.mathworks.com/matlabcentral/fileexchange/4705-error-ellipse.

Conversion from firing rate to ON/OFF probabilities for Ca^{2+} imaging data

For the Ca^{2+} imaging data, we began with estimated firing rate time series $r_i(t)$ for each neuron i recorded as part of a population of N neurons. For later parts of the analysis we needed to convert these firing rates to binary ON/OFF values. This conversion involves a choice. One option would be to simply threshold the data, but this would throw away information about the magnitude of the firing rate. We instead take a probabilistic approach where rather than deciding definitively whether a given neuron was ON or OFF in a given time bin, we calculate the probability that the neuron was ON or OFF by assuming that neurons fire action potentials according to an inhomogeneous Poisson process with rate $r_i(t)$. The mean number of spikes $\lambda_i(t)$ expected in a time bin of width Δt is $\lambda_i(t) = r_i(t)\Delta t$. We choose $\Delta t = 1$ second. Under the Poisson model the actual number of spikes m in a particular time bin is a random variable that follows the Poisson distribution $P(m=k) = \lambda^k \exp(-\lambda) / k!$. We considered a neuron active (ON) if it is firing one or more spikes in a given time bin. Hence the probability that a neuron is ON is $p_{on}(t)=1-P(m=0)=1-\exp(\lambda)$. This approach has two advantages over thresholding: 1) it

preserves some information about the magnitude of firing rates, and 2) it acts to regularize the probability distribution for the number of neurons active by essentially smoothing nearby values together.

Entropy estimation for large numbers of neurons

Entropy was estimated by fitting a statistical model we recently developed, called the population tracking model (O'Donnell et al., 2017), to the binarized Ca^{2+} imaging data. Briefly, the population tracking model fits two aspects of the data: the probability distribution for the number of neurons synchronously active in the population, and also the conditional firing probability that each individual neuron is active given the population count. Hence the model captures both some aggregate statistics of the population activity, and some aspects of the heterogeneity across neurons. See (O'Donnell et al., 2017) for complete details and validation of the method. Code for fitting the model to data is available at https://github.com/cianodonnell/PopulationTracking.

The entropy/neuron generally decreased with the number of neurons considered as result of the population correlations (Figure 7B), so we needed to control for neural population size when comparing data from different experimental groups. On the one hand, we would like to study as large a number of neurons as possible, because we expect the effects of collective network dynamics to be stronger for large population sizes and this may be the regime where differences between the groups emerge. On the other hand, our recording methods allowed us to sample only typically around ~100 neurons at a time, and as few as 40 neurons in some animals. Hence we proceeded by first estimating the entropy/neuron in each animal by calculating the entropy of random subsets of neurons of varying size from 10 to 100 (if possible) in steps of 10. For each population size we sampled a large number of independent subsets, calculated the entropy of each. Finally, for each dataset we fit a double exponential function to the estimated entropy/neuron as a function of the number of neurons: $H/N = A*\exp(-b*N) + C*\exp(-d*N) + e$, and used this fit to estimate H/N for 100 neurons.

Acknowledgements

We thank Timothy O'Leary, Hannes Saal, and Alex Williams for comments on earlier versions

763 of the manuscript. This study was supported by funding from FRAXA Research Foundation,
764 Howard Hughes Medical Institute, Sloan-Swartz Foundation, the Dana Foundation,
765 Developmental Disabilities Translational Research Program grant #20160969 (The John
766 Merck Fund), SFARI grant 295438 (Simons Foundation) and the NIH (NICHD R01HD054453
767 and NINDS RC1NS068093).

768 **Competing Interests**

769 The authors have no competing interests to declare.

References

Arakaki, T., Barello, G., Ahmadian, Y., 2017. Capturing the diversity of biological tuning curves using generative adversarial networks. arXiv.

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., Petersen, C.C.H., 2012. Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. J Neurophysiol 107, 3116–3134. doi:10.1152/jn.00917.2011

Bateup, H.S., Takasaki, K.T., Saulnier, J.L., Denefrio, C.L., Sabatini, B.L., 2011. Loss of Tsc1 in vivo impairs hippocampal mGluR-LTD and increases excitatory synaptic function. J Neurosci 31, 8862–8869. doi:10.1523/JNEUROSCI.1617-11.2011

Belmonte, M.K., Allen, G., Beckel-Mitchener, A., Boulanger, L.M., Carper, R.A., Webb, S.J., 2004. Autism and abnormal development of brain connectivity. J Neurosci 24, 9228–9231. doi:10.1523/JNEUROSCI.3340-04.2004

Berkes, P., Orbán, G., Lengyel, M., Fiser, J., 2011. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. Science 331, 83–87. doi:10.1126/science.1195870

Berzhanskaya, J., Phillips, M.A., Shen, J., Colonnese, M.T., 2016. Sensory hypo-excitability in a rat model of fetal development in Fragile X Syndrome. Sci Rep 6, 30769. doi:10.1038/srep30769

Box, G. E. P., 1979. Robustness in the strategy of scientific model building, in Launer, R. L.; Wilkinson, G. N., Robustness in Statistics, *Academic Press*, pp. 201–236.)

Braat, S., Kooy, R.F., 2015. The GABAA Receptor as a Therapeutic Target for Neurodevelopmental Disorders. Neuron 86, 1119–1130. doi:10.1016/j.neuron.2015.03.042

Brunel, N., 2000. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. J Comput Neurosci 8, 183–208.

Bureau, I., Shepherd, G.M.G., Svoboda, K., 2008. Circuit and plasticity defects in the developing somatosensory cortex of FMR1 knock-out mice. J Neurosci 28, 5178–5188. doi:10.1523/JNEUROSCI.1076-08.2008

Clancy, K.B., Schnepel, P., Rao, A.T., Feldman, D.E., 2015. Structure of a single whisker representation in layer 2 of mouse somatosensory cortex. J Neurosci 35, 3946–3958.

doi:10.1523/JNEUROSCI.3887-14.2015

Contractor, A., Klyachko, V.A., Portera-Cailliau, C., 2015. Altered Neuronal and Circuit Excitability in Fragile X Syndrome. Neuron 87, 699–715. doi:10.1016/j.neuron.2015.06.017

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory, 2nd ed. Wiley-Interscience.

Cruz-Martin, A., Crespo, M., Portera-Cailliau, C., 2010. Delayed Stabilization of Dendritic Spines in Fragile X Mice. J Neurosci 30, 7793–7803. doi:10.1523/JNEUROSCI.0577-10.2010

Dalezios, Y., Luján, R., Shigemoto, R., Roberts, J.D.B., Somogyi, P., 2002. Enrichment of mGluR7a in the presynaptic active zones of GABAergic and non-GABAergic terminals on interneurons in the rat somatosensory cortex. Cereb Cortex 12, 961–974.

Dani, V.S., Chang, Q., Maffei, A., Turrigiano, G.G., Jaenisch, R., Nelson, S.B., 2005. Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. Proc. Natl. Acad. Sci. U.S.A. 102, 12560–12565. doi:10.1073/pnas.0506071102

Deneve, S., Machens, C.K., 2016. Efficient codes and balanced networks. Nature Publishing Group 19, 375–382. doi:10.1038/nn.4243

Elstrott, J., Clancy, K.B., Jafri, H., Akimenko, I., Feldman, D.E., 2014. Cellular mechanisms for response heterogeneity among L2/3 pyramidal cells in whisker somatosensory cortex. J Neurophysiol 112, 233–248. doi:10.1152/jn.00848.2013

Fanselow, E.E., Richardson, K.A., Connors, B.W., 2008. Selective, state-dependent activation of somatostatin-expressing inhibitory interneurons in mouse neocortex. J Neurophysiol 100, 2640–2652. doi:10.1152/jn.90691.2008

Fino, E., Yuste, R., 2011. Dense inhibitory connectivity in neocortex. Neuron 69, 1188–1203. doi:10.1016/j.neuron.2011.02.025

Fisher, D., Olasagasti, I., Tank, D.W., Aksay, E.R.F., Goldman, M.S., 2013. A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. Neuron 79, 987–1000. doi:10.1016/j.neuron.2013.06.041

Frye, C.G., Maclean, J.N., 2016. Spontaneous activations follow a common developmental course across primary sensory areas in mouse neocortex. J Neurophysiol 116, 431–437. doi:10.1152/jn.00172.2016

829 Gentet, L.J., 2012. Functional diversity of supragranular GABAergic neurons in the barrel
830 cortex. Front Neural Circuits 6, 52. doi:10.3389/fncir.2012.00052
831 Gentet, L.J., Kremer, Y., Taniguchi, H., Huang, Z.J., Staiger, J.F., Petersen, C.C.H., 2012.
832 Unique functional properties of somatostatin-expressing GABAergic neurons in mouse
833 barrel cortex. Nat Neurosci 15, 607–612. doi:10.1038/nn.3051
834 Gibson, J.R., Bartley, A.F., Hays, S.A., Huber, K.M., 2008. Imbalance of neocortical excitation
835 and inhibition and altered UP states reflect network hyperexcitability in the mouse model of
836 fragile X syndrome. J Neurophysiol 100, 2615–2626. doi:10.1152/jn.90752.2008
837 Golshani, P., Gonçalves, J.T., Khoshkhoo, S., Mostany, R., Smirnakis, S., Portera-Cailliau, C.,
838 2009. Internally mediated developmental desynchronization of neocortical network activity.
839 J Neurosci 29, 10890–10899. doi:10.1523/JNEUROSCI.2012-09.2009
840 Gonçalves, J.T., Anstey, J.E., Golshani, P., Portera-Cailliau, C., 2013. Circuit level defects in
841 the developing neocortex of Fragile X mice. Nat Neurosci 16, 903–909.
842 doi:10.1038/nn.3415
843 Goodman, D.F.M., Brette, R., 2009. The brian simulator. Front Neurosci 3, 192–197.
844 doi:10.3389/neuro.01.026.2009
845 Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., Sethna, J.P., 2007.
846 Universally sloppy parameter sensitivities in systems biology models. PLoS Comput Biol 3,
847 1871–1878. doi:10.1371/journal.pcbi.0030189
848 Harlow, E.G., Till, S.M., Russell, T.A., Wijetunge, L.S., Kind, P., Contractor, A., 2010. Critical
849 period plasticity is disrupted in the barrel cortex of FMR1 knockout mice. Neuron 65, 385–
850 398. doi:10.1016/j.neuron.2010.01.024
851 Hays, S.A., Huber, K.M., Gibson, J.R., 2011. Altered neocortical rhythmic activity states in
852 Fmr1 KO mice are due to enhanced mGluR5 signaling and involve changes in excitatory
853 circuitry. J Neurosci 31, 14223–14234. doi:10.1523/JNEUROSCI.3157-11.2011
854 He, Q., Nomura, T., Xu, J., Contractor, A., 2014. The developmental switch in GABA polarity is
855 delayed in fragile X mice. J Neurosci 34, 446–450. doi:10.1523/JNEUROSCI.4447-13.2014
856 Hennequin, G., Agnes, E.J., Vogels, T.P., 2017. Inhibitory Plasticity: Balance, Control, and
857 Codependence. Annu. Rev. Neurosci. 40, annurev–neuro–072116–031005.
858 doi:10.1146/annurev-neuro-072116-031005

Hirsch, M.W., Smale, S., Devaney, R.L., 2013. Differential Equations, Dynamical Systems, and an Introduction to Chaos, 3rd ed. Academic Press.

Kehrer, C., Maziashvili, N., Dugladze, T., Gloveli, T., 2008. Altered Excitatory-Inhibitory Balance in the NMDA-Hypofunction Model of Schizophrenia. Frontiers in Molecular Neuroscience 1, 6. doi:10.3389/neuro.02.006.2008

Kerr, J.N.D., de Kock, C.P.J., Greenberg, D.S., Bruno, R.M., Sakmann, B., Helmchen, F., 2007. Spatial organization of neuronal population responses in layer 2/3 of rat barrel cortex. J Neurosci 27, 13316–13328. doi:10.1523/JNEUROSCI.2210-07.2007

Kinnischtzke, A.K., Sewall, A.M., Berkepile, J.M., Fanselow, E.E., 2012. Postnatal maturation of somatostatin-expressing inhibitory cells in the somatosensory cortex of GIN mice. Front Neural Circuits 6, 33. doi:10.3389/fncir.2012.00033

La Fata, G., Gärtner, A., Domínguez-Iturza, N., Dresselaers, T., Dawitz, J., Poorthuis, R.B., Averna, M., Himmelreich, U., Meredith, R.M., Achsel, T., Dotti, C.G., Bagni, C., 2014. FMRP regulates multipolar to bipolar transition affecting neuronal migration and cortical circuitry. Nat Neurosci. doi:10.1038/nn.3870

Lee, E., Lee, J., Kim, E., 2017. Excitation/Inhibition Imbalance in Animal Models of Autism Spectrum Disorders. Biol. Psychiatry 81, 838–847. doi:10.1016/j.biopsych.2016.05.011

Lefort, S., Tomm, C., Floyd Sarria, J.-C., Petersen, C.C.H., 2009. The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. Neuron 61, 301–316. doi:10.1016/j.neuron.2008.12.020

Lu, H., Ash, R.T., He, L., Kee, S.E., Wang, W., Yu, D., Hao, S., Meng, X., Ure, K., Ito-Ishida, A., Tang, B., Sun, Y., Ji, D., Tang, J., Arenkiel, B.R., Smirnakis, S.M., Zoghbi, H.Y., 2016. Loss and Gain of MeCP2 Cause Similar Hippocampal Circuit Dysfunction that Is Rescued by Deep Brain Stimulation in a Rett Syndrome Mouse Model. Neuron 91, 739–747. doi:10.1016/j.neuron.2016.07.018

Luczak, A., Bartho, P., Harris, K.D., 2009. Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron 62, 413–425. doi:10.1016/j.neuron.2009.03.014

Machta, B.B., Chachra, R., Transtrum, M.K., Sethna, J.P., 2013. Parameter space compression underlies emergent theories and predictive models. Science 342, 604–607.

doi:10.1126/science.1238723

Mao, T., Kusefoglu, D., Hooks, B.M., Huber, D., Petreanu, L., Svoboda, K., 2011. Long-range neuronal circuits underlying the interaction between sensory and motor cortex. Neuron 72, 111–123. doi:10.1016/j.neuron.2011.07.029

Meredith, R.M., Dawitz, J., Kramvis, I., 2012. Sensitive time-windows for susceptibility in neurodevelopmental disorders. Trends Neurosci 35, 335–344. doi:10.1016/j.tins.2012.03.005

Miller, J.-E.K., Ayzenshtat, I., Carrillo-Reid, L., Yuste, R., 2014. Visual stimuli recruit intrinsically generated cortical ensembles. 111, E4053–E4061. doi:10.1073/pnas.1406077111

Mitra, A., Snyder, A.Z., Blazey, T., Raichle, M.E., 2015. Lag threads organize the brain's intrinsic activity. Proc. Natl. Acad. Sci. U.S.A. 112, E2235–E2244. doi:10.1073/pnas.1503960112

Nelson, S.B., Valakh, V., 2015. Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. Neuron 87, 684–698. doi:10.1016/j.neuron.2015.07.033

O'Donnell, C., Gonçalves, J.T., Whiteley, N., Portera-Cailliau, C., Sejnowski, T.J., 2017. The Population Tracking Model: A Simple, Scalable Statistical Model for Neural Population Data. Neural Comput 29, 50–93. doi:10.1162/NECO_a_00910

O'Donnell, C., 2017. PopulationTracking. Github. https://github.com/cianodonnell/PopulationTracking. 2c9eada.

O'Donnell, C., 2017. ODonnelletal_2017_imbalances. Github. https://github.com/cianodonnell/ODonnelletal_2017_imbalances. 2383a26.

O'Leary, T., Sutton, A.C., Marder, E., 2015. Computational models in the age of large datasets. Curr Opin Neurobiol 32C, 87–94. doi:10.1016/j.conb.2015.01.006

Paluszkiewicz, S.M., Olmos-Serrano, J.L., Corbin, J.G., Huntsman, M.M., 2011. Impaired inhibitory control of cortical synchronization in fragile X syndrome. J Neurophysiol 106, 2264–2272. doi:10.1152/jn.00421.2011

Panas, D., Amin, H., Maccione, A., Muthmann, O., van Rossum, M., Berdondini, L., Hennig, M.H., 2015. Sloppiness in spontaneously active neuronal networks. J Neurosci 35, 8480–8492. doi:10.1523/JNEUROSCI.4421-14.2015

919 Patel, A.B., Hays, S.A., Bureau, I., Huber, K.M., Gibson, J.R., 2013. A target cell-specific role
920 for presynaptic Fmr1 in regulating glutamate release onto neocortical fast-spiking inhibitory
921 neurons. J Neurosci 33, 2593–2604. doi:10.1523/JNEUROSCI.2447-12.2013
922 Petersen, C.C.H., Crochet, S., 2013. Synaptic computation and sensory processing in
923 neocortical layer 2/3. Neuron 78, 28–48. doi:10.1016/j.neuron.2013.03.020
924 Petreanu, L., Huber, D., Sobczyk, A., Svoboda, K., 2007. Channelrhodopsin-2-assisted circuit
925 mapping of long-range callosal projections., Nature neuroscience. doi:10.1038/nn1891
926 Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., Scanziani, M., 2013. Inhibition of inhibition in visual
927 cortex: the logic of connections between molecularly distinct interneurons. Nat Neurosci.
928 doi:10.1038/nn.3446
929 Pologruto, T.A., Yasuda, R., Svoboda, K., 2004. Monitoring neural activity and [Ca2+] with
930 genetically encoded Ca2+ indicators. J Neurosci 24, 9572–9579.
931 doi:10.1523/JNEUROSCI.2854-04.2004
932 Qian, N., Lipkin, R.M., 2011. A learning-style theory for understanding autistic behaviors. Front
933 Hum Neurosci 5, 77. doi:10.3389/fnhum.2011.00077
934 Quian Quiroga, R., Panzeri, S., 2009. Extracting information from neuronal populations:
935 information theory and decoding approaches. Nat Rev Neurosci 10, 173–185.
936 doi:10.1038/nrn2578
937 Ringach, D.L., 2009. Spontaneous and driven cortical activity: implications for computation.
938 Curr Opin Neurobiol 19, 439–444. doi:10.1016/j.conb.2009.07.005
939 Rochefort, N.L., Garaschuk, O., Milos, R.-I., Narushima, M., Marandi, N., Pichler, B.,
940 Kovalchuk, Y., Konnerth, A., 2009. Sparsification of neuronal activity in the visual cortex at
941 eye-opening. 106, 15049–15054. doi:10.1073/pnas.0907660106
942 Rubenstein, J.L.R., Merzenich, M.M., 2003. Model of autism: increased ratio of
943 excitation/inhibition in key neural systems. Genes Brain Behav. 2, 255–267.
944 Sato, T.R., Gray, N.W., Mainen, Z.F., Svoboda, K., 2007. The functional microarchitecture of
945 the mouse barrel cortex. PLoS Biol 5, e189. doi:10.1371/journal.pbio.0050189
946 Schneidman, E., Berry, M.J., Segev, R., Bialek, W., 2006. Weak pairwise correlations imply
947 strongly correlated network states in a neural population. Nature 440, 1007–1012.
948 doi:10.1038/nature04701

Selby, L., Zhang, C., Sun, Q.-Q., 2007. Major defects in neocortical GABAergic inhibitory circuits in mice lacking the fragile X mental retardation protein. Neurosci. Lett. 412, 227–232. doi:10.1016/j.neulet.2006.11.062

Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., Lesica, N.A., 2016. Inhibitory control of correlated intrinsic variability in cortical networks. elife 5, 91. doi:10.7554/eLife.19695

Testa-Silva, G., Loebel, A., Giugliano, M., de Kock, C.P.J., Mansvelder, H.D., Meredith, R.M., 2012. Hyperconnectivity and slow synapses during early development of medial prefrontal cortex in a mouse model for mental retardation and autism. Cereb Cortex 22, 1333–1342. doi:10.1093/cercor/bhr224

The Dutch-Belgian Fragile X Consortium, 1994. Fmr1 knockout mice: a model to study fragile X mental retardation. The Dutch-Belgian Fragile X Consortium. Cell 78, 23–33.

Tomm, C., Avermann, M., Petersen, C., Gerstner, W., Vogels, T.P., 2014. Connection-type-specific biases make uniform random network models consistent with cortical recordings. J Neurophysiol 112, 1801–1814. doi:10.1152/jn.00629.2013

Tsodyks, M.V., Sejnowski, T., 1995. Rapid state switching in balanced cortical network models. Network: Computation in neural systems 111–124.

Vogels, T.P., Abbott, L.F., 2007. Gating deficits in model networks: a path to schizophrenia? Pharmacopsychiatry 40 Suppl 1, S73–7. doi:10.1055/s-2007-992130

Wallace, M.L., Burette, A.C., Weinberg, R.J., Philpot, B.D., 2012. Maternal loss of Ube3a produces an excitatory/inhibitory imbalance through neuron type-specific synaptic defects. Neuron 74, 793–800. doi:10.1016/j.neuron.2012.03.036

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12, 1–24. doi:10.1016/S0006-3495(72)86068-5

Yaksi, E., Friedrich, R.W., 2006. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging. Nat. Methods 3, 377–383. doi:10.1038/nmeth874

Figure Supplements

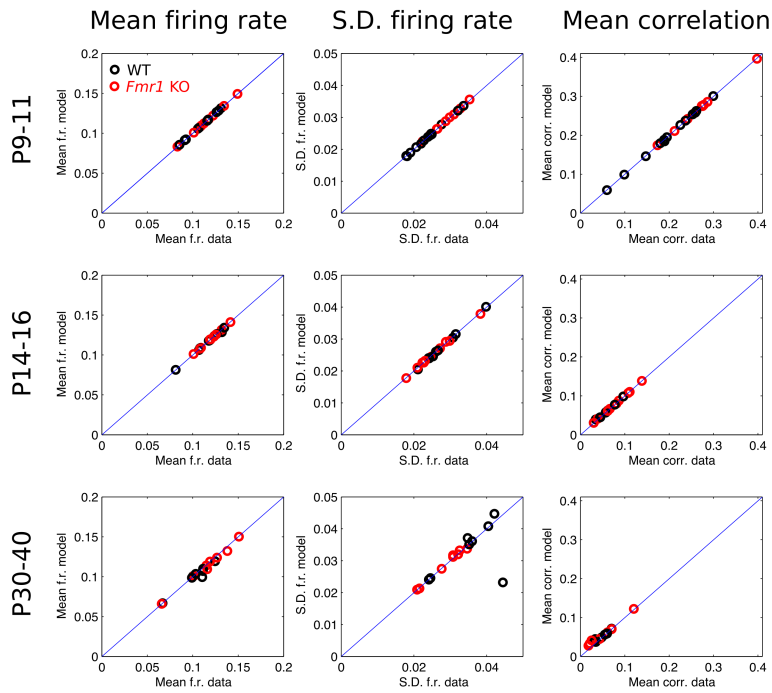

Figure 5 – figure supplement 1. Agreement between population logistic model activity statistics and raw data statistics.

Black circles are WT, red are *Fmr1* KO. Each data point corresponds to a recording from a single animal. Each plot shows model prediction versus raw data target value. Plots in each row correspond to data from a different age group (P9—11, P14—16, and P30—40), and each column corresponds to one of the three target activity statistics (mean firing rate, s.d. in firing rate, and mean pairwise correlation). Blue line is identity.

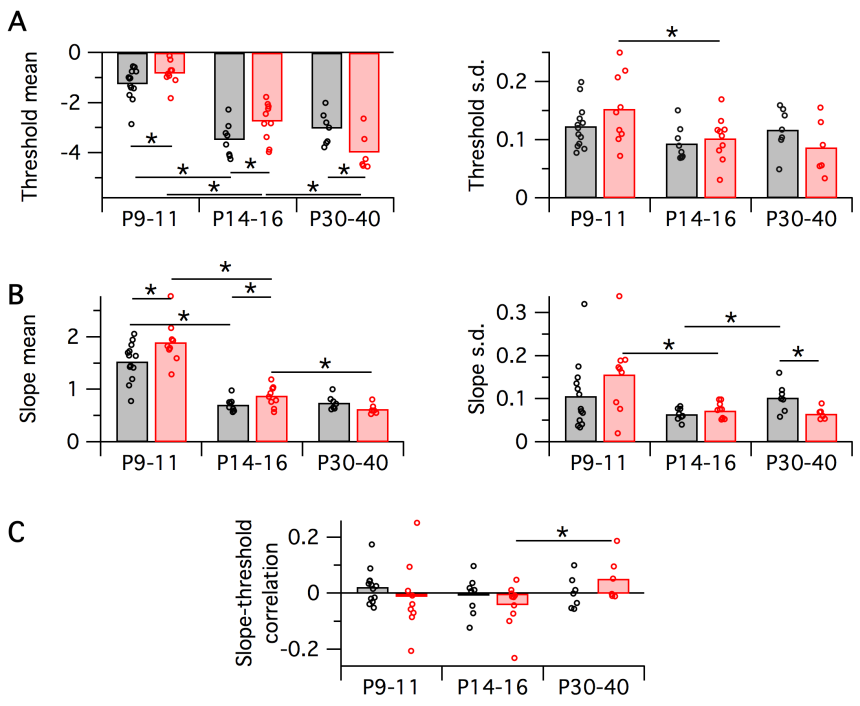

Figure 5 – figure supplement 2. Variation in population logistic model parameter fits with developmental age group and genotype.

Black symbols are WT, red are *Fmr1* KO. Each data point corresponds to a recording from a single animal, bars correspond to group means. In all cases, horizontal bar with asterisk indicates a significant difference in group means ($p<0.05$ via bootstrapping).

A: Logistic threshold parameter mean (left) and s.d. (right).

B: Logistic slope parameter mean (left) and s.d. (right).

C: Slope-threshold correlation parameter.

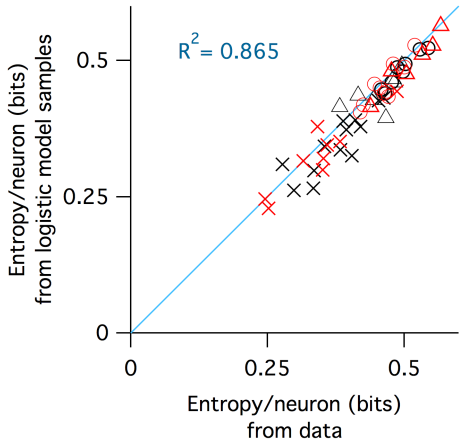

Figure 6 – figure supplement 1. Agreement between entropy estimated from raw data with entropy estimated from samples from fitted logistic models.

Black symbols are WT, red are *Fmr1* KO. Crosses are P9—11, circles P14—16, triangles P30—40. Blue line is identity, the R^2 value of which is reported in the inset.

980