scientific reports

OPEN

Unique scales preserve self-similar integrate-and-fire functionality of neuronal clusters

Anar Amgalan[1,2,7], Patrick Taylor[3,7], Lilianne R. Mujica-Parodi[1,2,4✉] & Hava T. Siegelmann[3,5,6✉]

Brains demonstrate varying spatial scales of nested hierarchical clustering. Identifying the brain's neuronal cluster size to be presented as nodes in a network computation is critical to both neuroscience and artificial intelligence, as these define the cognitive blocks capable of building intelligent computation. Experiments support various forms and sizes of neural clustering, from handfuls of dendrites to thousands of neurons, and hint at their behavior. Here, we use computational simulations with a brain-derived fMRI network to show that not only do brain networks remain structurally self-similar across scales but also neuron-like signal integration functionality ("integrate and fire") is preserved at particular clustering scales. As such, we propose a coarse-graining of neuronal networks to ensemble-nodes, with multiple spikes making up its ensemble-spike and time re-scaling factor defining its ensemble-time step. This fractal-like spatiotemporal property, observed in both structure and function, permits strategic choice in bridging across experimental scales for computational modeling while also suggesting regulatory constraints on developmental and evolutionary "growth spurts" in brain size, as per punctuated equilibrium theories in evolutionary biology.

Neuronal clustering is observed far more frequently than by chance[1]. The most common types of neuronal clustering believed to play a role in neuronal activity, development, and modularity[2,3] are *dendritic bundles* (neurons clustered together at their apical ends, with axons terminating at the same target[4]), *minicolumns* (radially clustered cell bodies, often around 80–100 neurons), *columns* (radially grouped minicolumns, often around 60–80 minicolumns[5,6]), *cluster-columns* (single minicolumn surrounded by one circular layer of columns, often seven or eight columns total[7], also defined by their functional role: e.g. color processing), as well as small *lego blocks* (of one order of magnitude only)[1,8–10]. Similarly, temporal clustering of spikes is observed more frequently than predicted by spiking rates alone[11–13]. Yet the brain's "least functional unit" above the neuron, or *monad* is still undefined. The potential utility of defining a monad is that the "least functional unit" may actually be a "sufficient functional unit." In practical terms, identifying neural monads tells us which modules are cognitive or constitute computational functions worth modeling. This is of direct relevance for choosing the most strategic scales to measure and model in computational neuroscience, as well as for building intelligent computation[14].

The diverse experimental modalities utilized in brain sciences probe scales that can overlap little and obtain information through different physiological aspects of brain. Microscopic techniques resolve individual neurons, but lack coverage; fMRI captures an image of the entire cortex, but at a resolution of million or so neurons per voxel; EEG and NIRS are limited by depth and coarse spatial resolution, electrode readings can reach spatial extent of >1 mm of tissue, but miss on the biophysical details of individual neurons. Here, through our computational coarse-graining of spiking network activity data, we provide justification for a procedure of joining distinct scales, and therefore information from some of the distinct imaging modalities available for a fuller picture of brain's workings. We establish an instance of scale-invariance of spiking network's functionality at the transition from a single neuron to an ensemble of handful of neurons and suggest that a similar procedure can

---

[1]Physics and Astronomy Department, Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY, USA. [2]Laboratory for Computational Neurodiagnostics, Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY, USA. [3]College of Information and Computer Sciences, University of Massachusetts, Amherst, MA, USA. [4]Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital/Harvard Medical School, Charlestown, MA, USA. [5]Neuroscience and Behavior Program, University of Massachusetts, Amherst, MA, USA. [6]Center for Data Science, University of Massachusetts, Amherst, MA, USA. [7]These authors contributed equally: Anar Amgalan and Patrick Taylor. ✉email: Lilianne.Strey@stonybrook.edu; hava@cs.umass.edu

detect a pair of experimental scales and techniques (e.g. calcium imaging combined with multi-electrode array recording) that are amenable to simultaneous modeling by virtue of sharing functional organizing principle. Such an estimate of paired similarly-behaving scales would guide the choice of imaging modalities to inform multi-scale modeling efforts aimed at maximal coverage of brain's dynamical repertoire.

Some understanding of the dynamical repertoire of brain's larger than neuron structures, and particularly their repetitive and self-similar nature has been obtained: synfire chain[11,12] describes the sequentially wired sub-assemblies propagating synchronous activities; avalanche models describe the power-law statistics of sizes of events in multielectrode recordings[15]; power-law noise statistics in brain signals have been robustly characterized[16-18]. Also reliably documented is the scale-free structural properties of brain network[19]. Yet the rules unifying structures and processes in brain at distinct scales are not fully described. Existing computational methods for re-scaling the neuronal network focus on enforcing known first and second order statistics of a fine-scale simulation to its coarse-grained version[20,21]. These approaches provide insight into statistical conditions that need to be satisfied for improved simulation runtime complexity.

The most basic functionality of brain is that of spiking neurons. Neurons, however variable, act as non-linear summing integrators, with a decay in potential and a threshold for firing; they spike when excited frequently and strongly enough. Here, we propose a scale-invariant dynamic, where a pair of similar views of the same object are separated by a specific re-scaling factor that transforms one into another. A simple analogy is provided by Sierpinski triangle[22], where a finer stage is obtained from the previous coarser stage by halving the spatial dimension of each triangle and tripling the number of triangles, thereby characterizing each stage of re-scaling of the fractal object by two integers: (2, 3). Here, we propose a functional scale-invariance or "fractality," such that a redefinition of the spiking network's basic primitives: node, edge, and spike as a particular multiple of themselves allows one to observe the neuron-like signal-integration functionality re-emerge in the coarser structure, albeit only at unique re-scaling. The consequence of this proposition is that multi-scale models can coarse-grain for computational expediency while strategically choosing a unique set of scales specifically determined to permit translation. Thus, we expand on previous studies, which have demonstrated the feasibility of hierarchical clustering[23] and networks' scale-free features[24-26] with regard to brain *structure* to introduce self-similarity of *dynamics* of clusters of neurons, replicating integrate-and-fire-like functionality across scales.

To computationally test whether dynamics of signal transmission are preserved in a manner akin to "functional harmonics" of the original scale, we construct a spiking network that follows actual brain connectivity. The starting point of our study is an all-to-all resting-state human fMRI-derived functional connectivity matrix, extracted from 91,282 voxels, and providing full coverage of the human cortex[27,28]. The fMRI data describe the temporal correlation of each of these voxels with every other voxel, allowing an inference about the degree to which they are functionally connected. Such inferences have been externally validated by comparison to structural and DTI data[29]. Each voxel of this matrix is $(2 \text{ mm})^3$, and thus measures a compensatory hemodynamic (blood oxygen level dependent, or BOLD) response across ~1000 cortical minicolumns, or about ~1 M neurons. Each voxel is represented as a node in our initial network, and in our computational experiment it follows the functionality of a leaky integrate-and-fire neuron. We start our computational experiment from an already non-neuronal scale of data acquisition, as we suggest that the observations made on fMRI scale may be close to one of the stages of the potential network re-scalings, along which the signal integration property is preserved.

The initial fine-grain network is reduced into clusters using a streaming hierarchical clustering algorithm[30]. We chose a streaming clustering algorithm for its computational expediency and reduced memory requirement. In deciding which nodes combine to form a cluster, we chose full-linkage clustering, requiring that only sets of nodes that have all-to-all connection weights higher than the cut-off c become a cluster of nodes (*an ensemble-node*). All pairs of neurons i, j belonging to the same cluster satisfy: $w_{ij} \geq c$, where w_{ij} is the connection weight between nodes i and j. By adjusting cut-off c one can flexibly control the average number of neurons in an ensemble-node. This clustering procedure allows re-scaling to a continuous spectrum of size of networks (Fig. 1A–C).

The set of edges running between two neighboring ensemble-nodes aggregate to give the *ensemble-edge* connecting these ensemble-nodes. The weight of the ensemble-edge is given by the mean weight of the edges defining it: $w_{kl}^{coarse} = mean(w_{ij})$ where i runs over all neurons in cluster k and j runs over all neurons in cluster l. For each ensemble-node we now define the *node strength* as the sum of weights of edges connecting to it: $s_k^{coarse} = \sum_l w_{kl}^{coarse}$, where superscript "*coarse*" indicates that the quantity pertains to the coarse-grained network.

We next redefine the activity of the coarse-grained network by temporally clustering firings of an ensemble-node into an *ensemble-spike*. An ensemble-node consisting of N_N nodes produces an ensemble-spike if its nodes produce a combined burst of above-threshold (N_S) number of spikes in the time-step, and no ensemble-spike occurs otherwise. This characterizes the number of ensemble-spikes required to arrive in rapid succession into an ensemble-node prior to activation of its own ensemble-spike. Intuitively, only areas dense enough in spikes in fine level spike raster (Fig. 1D) become ensemble-spikes in the coarse network's ensemble-spike raster (Fig. 1E) in the corresponding ensemble-node and time step. A spiking neuron's refractory period also re-emerges in the coarse-grain network, and concomitantly guides our choice of *ensemble-step*: the time step re-scaling parameter. Further details are provided in the Methods section. The terms "cluster" and "ensemble-node" will be used interchangeably as well as the terms "burst and "ensemble-spike."

Results

Coarse-grained structure is self-similar.
Coarse-graining at various cutoffs produces networks with a range of ensemble-node sizes and number of ensemble-nodes (Fig. 2A). Importantly, cutoff has little impact on the tail of ensemble-nodes' size distribution, which is invariant with respect to mean count and power-law behavior, suggesting self-similarity in agreement with a body of previous work[19,31,32] (Fig. 2B). The original functional brain network exhibits a self-similar property when it comes to its new coarsened structure: the
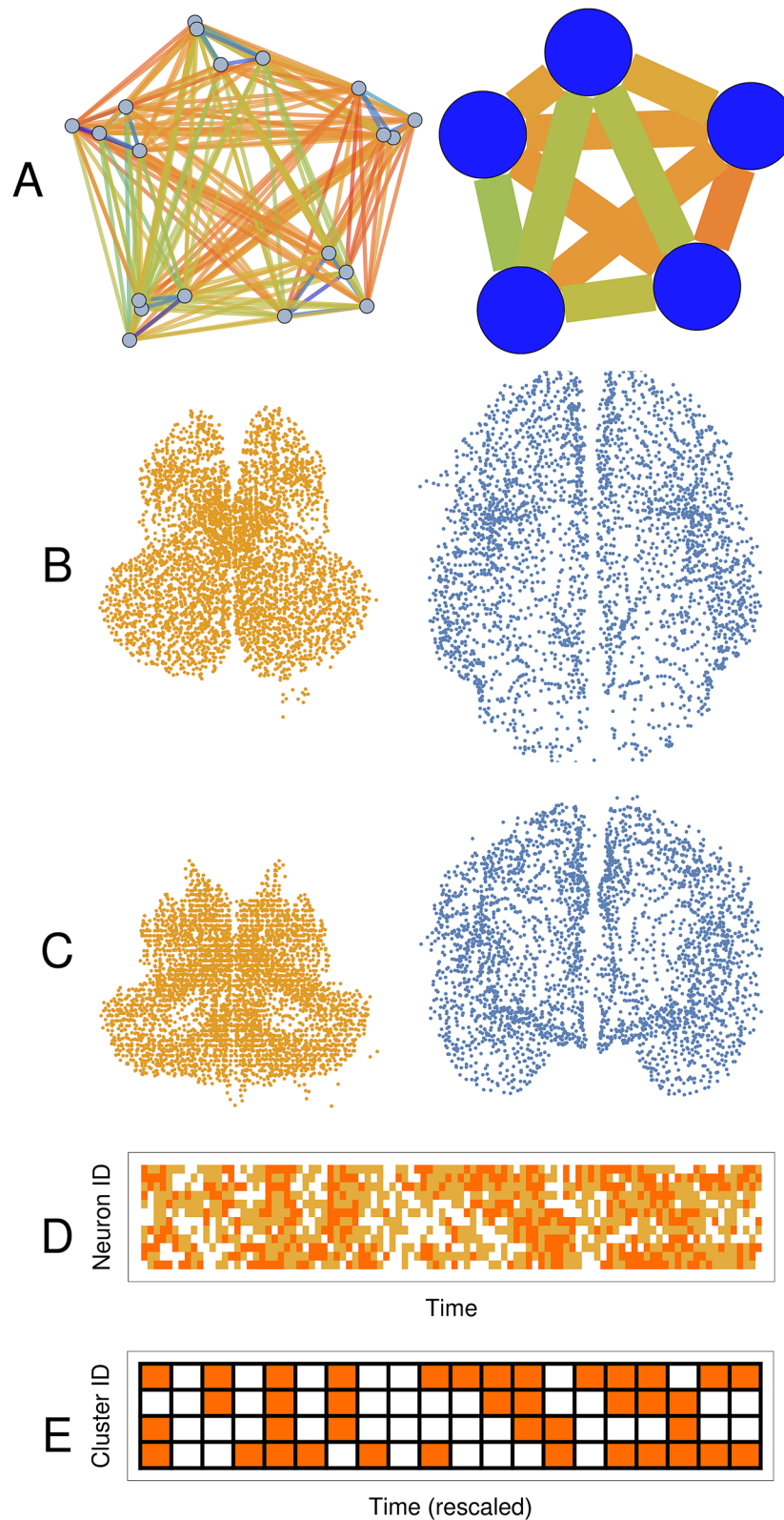
Figure 1. Re-scaling procedure for neuronal clusters and spike clusters demonstrates how proximate regions cluster together, reproducing organ-scale brain morphology. **(A)** We depict a fine-grain network of neurons (left) and its coarser, re-scaled version (right). Each ensemble-node, at a coarser level, is a collection of nodes from finer resolution level, clustered based on their connection weights (thin lines, left panel) with all edges between a pair of clusters averaged (thick lines, right panel). **(B)** Axial view of coarse-grained brain, after the procedure, at a cutoff of 0.98, re-scaling the original network of 91,282 nodes down to 7784 clusters. Each ensemble-node is located at the mean of coordinates of nodes comprising it. **(C)** Similarly, a coronal view. **(D)** The spike-train of fine-grain network is re-scaled into **(E)** ensemble-spike-train of coarser network.
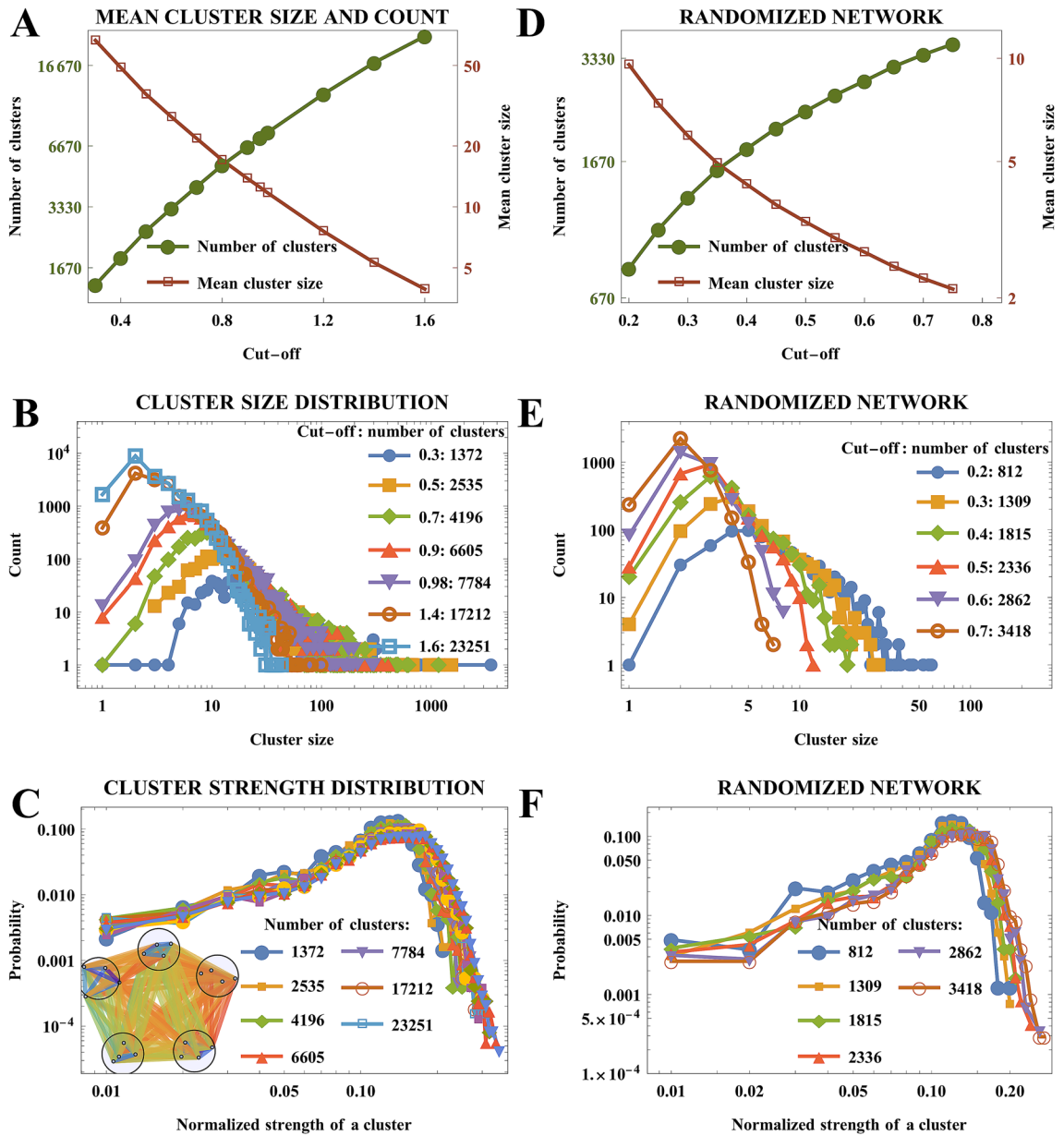
3

Figure 2. The brain network naturally clusters into a spectrum of system sizes while preserving its degree distribution shape. **(A)** Cutoffs are imposed on intra-cluster connection weights. **(B)** The network of 91,282 nodes is coarse-grained to obtain number of clusters ranging from ~ 1300 to ~ 23,000, yet it retains a non-trivial most frequent cluster size and a power-law drop off in distribution of cluster sizes. **(C)** Node strength distribution of the network stays largely invariant when it is clustered to well-connected clusters (inset) and inter-cluster edge weights are averaged. We define node strength as sum of all edges of a node and treat as weighted network's version of "degree" in unweighted network. **(D)** The same as **(A)**, but for a network that was randomized while preserving the strength (degree)-sequence of the nodes. **(E)** The same as **(B)**, but with a similarly randomized network: tail of the cluster size distribution is sensitive to the size of system, suggesting departure from scale-invariance. **(F)** The same as **(C)**, but for a randomized network: the cluster strength distribution is invariant under re-scaling.

connectivity network preserves the geometry of its node strength distribution after the coarse-graining procedure (Fig. 2C). We also provide a comparison to the coarse-graining of a randomly re-wired network (while preserving the strengths/degrees of nodes), and note a major difference: the distribution of cluster sizes as the randomly re-wired version of the original network is coarsened quickly loses the heavy tail of the power-law distribution within half an order of magnitude (Fig. 2E). See Methods, Re-wired Controls section for details of the randomization. The strength distribution of the network, on the other hand preserves its shape (Fig. 2F) in agreement with the original network's behavior under coarse-graining. The dependencies of the mean cluster size and the number of clusters as a function of the cut-off threshold (Fig. 2D) are included for validation and comparison with Fig. 2A.
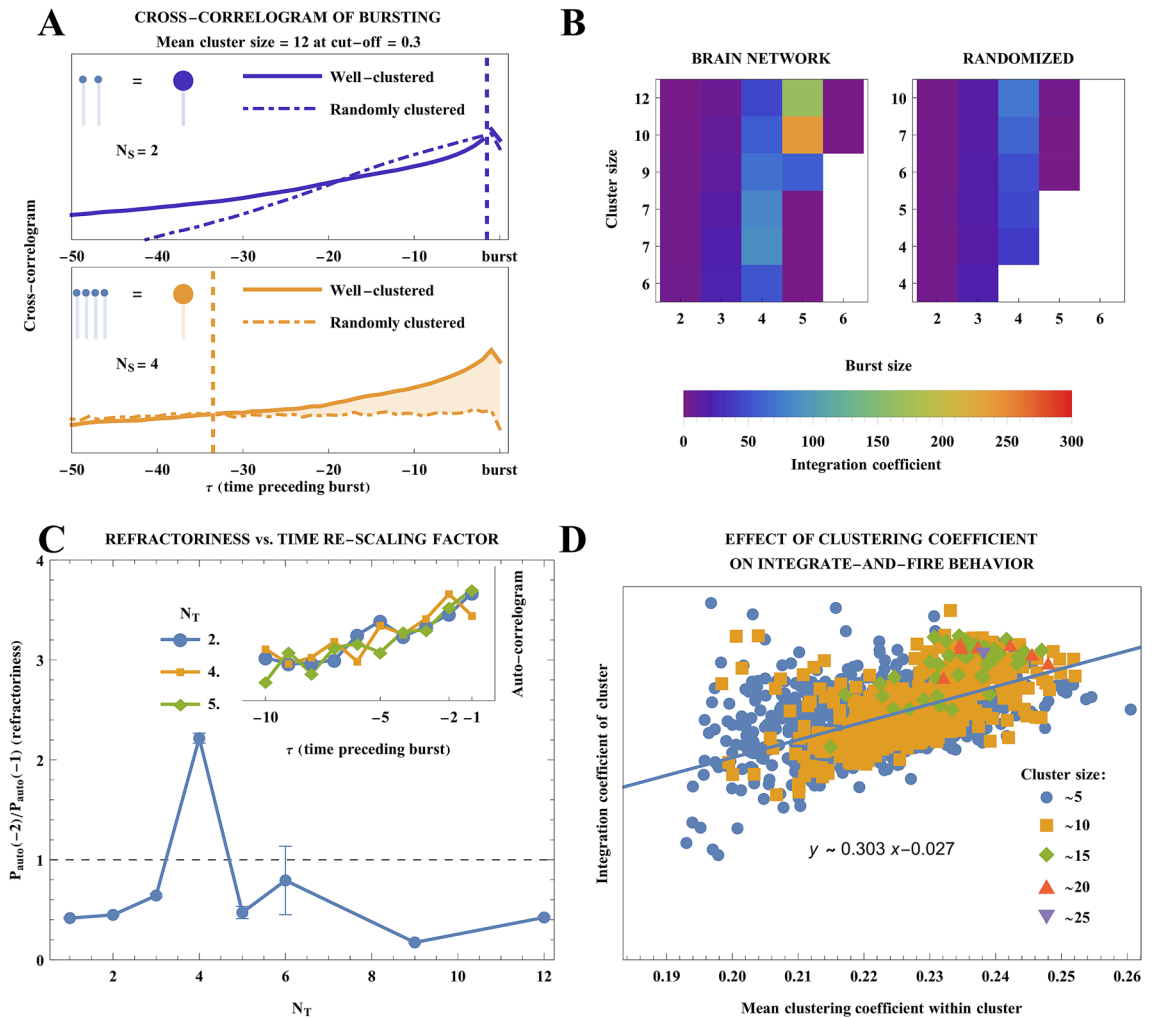
Figure 3. Ensemble-edge-weighted cross-correlogram $P(\tau)$ of cluster-spiking behavior identifies discrete scaling factors, at which integrate-and-fire behavior re-emerges in actual brains. **(A)** The intuition for integration coefficient (IC)—the degree, to which cross-correlogram of well-clustered brain's ensemble-spiking activity $P(\tau)$ exceeds that of randomly clustered brain $P_{random}(\tau)$ is indicated (shaded area) and acts to quantify the neuron-like behavior of ensemble-nodes (neuronal clusters). See Methods for definition of *IC* and cross-correlograms. Shown for ensemble-spikes defined as $N_s = 2$ spikes (top sub-panel) and $N_s = 4$ spikes (bottom sub-panel). **(B)** (Left) Heatmap of the Integration coefficient (IC) determines the ensemble-neuron size N_N and ensemble-spike size N_S used in re-scaling that best reproduces neuron-like behavior when clustered starting from a brain network. (Right) As a control, we show that the same signal integration behavior is lost when brain network is coarse-grained after a random shuffling (see "Methods" for procedure) that preserves strengths of all nodes (strength—sum of node's edge weights). This rules out the node strength sequence of brain network as the driver of its neuron-like behavior at the coarse scale. **(C)** The temporal re-scaling factor N_t is obtained by calculating the auto-correlogram $P_{auto}(\tau)$ (see Methods for definition) and maximizing the ratio $P_{auto}(-2)/P_{auto}(-1)$ to reveal refractory period in coarse-grained network. **(D)** Local clustering coefficient among nodes within an ensemble-node corresponds with its integrate-and-fire behavior. Positive dependence indicates that highly interconnected set of nodes can join into an ensemble-node with more neuron-like integration of input.

Coarse-grained dynamics is self-similar.

Re-scaling our network to a spectrum of average ensemble-node sizes (N_N) reveals signature of an integrate-and-fire property to a varying degree depending on the re-scaling factors (N_N and N_S) (Fig. 3A, *solid lines*), but with three key features. First, leading up to an outgoing ensemble-spike, the cross-correlogram of ensemble-spiking activity weighted by the ensemble-edge weights (see Methods for full description) displays integration of a faster-than-exponentially rising amount of inputs from neighbors. Second, the incoming inputs peak at the time-step immediately preceding ensemble-spike output. Finally, incoming activity from a given ensemble-node's neighbors abruptly drops at the time of an outgoing ensemble-spike.

Self-similar dynamics is preserved only at specific scaling factors displaying "functional harmonics" of the original scale.

This scale invariant dynamics results from the brain's organizational structure. Comparing the sharp increase in signal integration preceding the ensemble-spike observed in the clustered

network to that of a control network that was coarse grained to an identical average size of ensemble-nodes but through random clustering (ignoring the edge weights), we see a control behavior that is only exponential (Fig. 3A, dashed lines), indicating its weak input integration. Shaded areas in panels of Fig. 3A indicate the extent of the coarse network's input integration beyond that of an equivalently sized randomly clustered network. The top panel (ensemble-spike requiring ≥ 2 spikes) shows little difference from the random case, while the bottom panel (ensemble-spike requiring ≥ 4 spikes) shows that inputs of ≥ 4 spikes are integrated much more rapidly leading up to an ensemble-spike (solid line) than in a randomly clustered network (dashed line). From the starting network of 7784 nodes, by searching the space of re-scaling factors and maximizing the *integration coefficient* (see "Methods" section for details), we find ensemble-node size of ~ 10 nodes and ensemble-spike size of ~ 5 spikes to maximally recover integration of inputs leading up to ensemble-spikes (Fig. 3B left). *Integration coefficient* measures the degree to which the cross-correlogram of the coarse network exceeds what is randomly expected. We take the integral of the ratio $P(\tau)/P_{random}(\tau)$ weighted by the inverse of absolute value of time delay τ integrated over the range of delays where the ratio exceeds 1. Here $P(\tau)$ and $P_{random}(\tau)$ are the burst cross-correlogram of the coarse-grained network for well clustered (full-linkage clustered) and randomly clustered cases, respectively. In searching for the scale, to which the network coarse-grains to while preserving the most integrate-and-fire behavior, we sought to establish the triplet of re-scaling factors that maximize our measure of similarity to the integrate-and-fire functionality: (number of nodes in a cluster, number of spikes in a burst and the time rescaling factor). We establish the time-axis re-scaling factor, or ensemble-step, of ~ 4 by maximizing the refractoriness present in auto-correlogram (Fig. 3C). The refractoriness is defined as the ratio $P_{auto}(-2)/P_{auto}(-1)$ of the auto-correlogram of the coarse-grained network at two time-steps prior to the burst, to that at one time-step prior to the burst. Thus, the network coarse-graining procedure returns a triplet of values extracted by maximizing the integration coefficient and the refractoriness, consisting of (i) ensemble-node = 10 nodes ($N_N = 10$), (ii) ensemble-spike = 5 spikes ($N_S = 5$), and (iii) ensemble-step = 4 time steps ($N_T = 4$) that recovers properties of spiking nodes in a network of ensemble-nodes. Searching for the culprit of the integrate-and-fire behavior in the coarse-grained network, we extracted for each cluster after re-scaling i) the clustering coefficient of the fine-grain neuron-level network and ii) the integration coefficient from the neurons in the cluster. We observe that the within-cluster clustering coefficient is positively correlated with the signal integration behavior of the clusters (Fig. 3D). Thus, ensemble-nodes with greater local clustering coefficient display greater signal-integration behavior, and the dependence does not deviate from a linear trend.

Discussion

We started the research looking for the cluster size that acts as a minimal cognitive building block, and our fMRI-driven interrogation of simulated spiking network data suggests that brain organization has a dense multi-scale property in structure and functionality. Neurons naturally cluster across many scales, and such clusters create new networks with fewer functional units, but similar algorithmic features. This suggests a more complex answer. Our coarse-graining procedure attempts to capture only one such transition between a pair of scales by detecting similarity of dynamics, but suggests that the brain potentially preserves its dynamics across scales at many distinct levels, recycling already existing functional building blocks. It is therefore feasible that brain's least functional unit larger than a single neuron might be a cluster of neurons (an "ensemble-neuron" in our terminology) displaying a neuron-like integrate-and-fire behavior.

Testing whether neuronal clusters act as integrating neurons and, if so, at what re-scaling factor, can be established by a targeted experiment of large-scale calcium imaging and other optical recording modalities. Such an experiment would, having simultaneously recorded from large number of neurons, apply the machinery we describe, identifying a pairs of coarse and fine scales that display the most self-similar behavior. Such a constraint on the similarly behaving pair of scales would drastically narrow the space of multi-scale models of brain dynamics.

Support to our work comes from previous renormalization-inspired models of neuron-level data coarse-graining. Bialek and team proposed a renormalization procedure that groups variables (neurons) into larger-scale lumped variables based on their observed pairwise correlations[33]. Their procedure reveals the self-similarity of neuron level activities in live animal recordings, but does so without identifying the re-scaling factor that separates similarly behaving scales or whether such a factor exists. They instead estimate a set of renormalization-inspired scaling parameters from the data. Shew and team carry out a renormalization of binarized membrane potential signal from both simulation and rodent data for a coarse-graining similar to ours: their procedure is also partially limited by a single renormalization, rather than a sequence, as the authors point out[34]. In addition, their procedure yields scaling laws, but not constraints on discrete ratios between similarly behaving scales. Instead of finding the scale at which coarse-graining reveals the most self-similar behavior, their procedure probes the dependence of the outcome of coarse-graining on the simulation parameters, presenting an insight complementary to ours.

In clarifying our results, we note that the slope at the tail of the distribution of cluster sizes is not fully unchanged as a function of the system size. However, as a comparison to a case where no signature of similarity of scales is present, we show that for a 1.23 decade ($log_{10}(23251/1372) \approx 1.23$) of system sizes, the slope changes to a noticeably smaller degree than in a randomly clustered case, where the slope change is much larger during for the 0.62 decade ($log_{10}(3418/812) \approx 0.62$) of sizes. We use the term "scale-invariance" in the sense of a pair of spatiotemporal scales that are displaying similar dynamics, rather than a stricter definition in which a system displays identical behavior across all accessed scales.

Our primary finding that there exists a particular re-scaling factor that best recovers the integrate-and-fire principle the most dynamic should not be confused with detection of the same dynamic across all accessed scales. The term scale-invariance is used in this paper to describe a pair of spatiotemporal scales that are displaying

similar dynamics, instead of the sometimes used stricter definition where the system displays identical behavior at all scales accessible to it. We propose a future direction to strengthen the results. In the work reported here, the slope at the tail of the distribution of cluster sizes is not fully unchanged as a function of the system size. Rather as a comparison to a case where no signature of similarity of scales is present, we show that for a 1.23 decade ($\log_{10}(23251/1372) \approx 1.23$) of system sizes, the slope changes to a noticeably smaller degree than in a randomly clustered case, where the slope change is much larger during for the 0.62 decade ($\log_{10}(3418/812) \approx 0.62$) of sizes. An additional limitation of our work is partially a matter of terminology. Here we use the term scale-invariance, in the sense of a pair of spatiotemporal scales that are displaying similar dynamics, instead of the stricter definition where the system displays identical behavior at all scales accessible to it.

The implications of our work are particularly important for neural mass models (NMMs). We add to existing experimental and analytical motivations for NMMs[35–40] and propose a mechanism for constraining scales to be modeled with NMMs. NMMs provide an attractive procedure for reducing brain dynamics to just a few quantities at the population level. Further supporting the treatment of neuronal systems as NMMs, we provide computational justification for approximating the fine-scale neural dynamics (namely integrate-and-fire rule) by a similar dynamical rule, but at a particular coarser scale out of all scales empirically available for imaging. One implication of such an approximation is that brain dynamics may be modeled using experimental data from several non-overlapping scales, as long as those scales are chosen strategically to retain self-similar dynamics. In addition, our analysis coarse-grains a network of neurons while preserving integrate-and-fire dynamics at network sizes exceeding 1000 neurons and ensemble-neurons, highlighting the potential of NMM-like models to describe interactions of a large number of neuronal populations.

From a biological perspective, the periodicity of scale invariance may offer clues into developmental and evolutionary processes. The dynamic structure of the action potential (gating, excitation, inhibition, return to baseline) reflects, at its very basis, a mechanistic solution to the biologically ubiquitous problem of allostatic regulation in response to noisy inputs. In growing the brain from unicellular to multicellular to agglomerate structure—while continuing to maintain allostatic regulation—it therefore makes sense that dynamic signatures associated with such regulation (i.e., those associated with negative feedback loops) would also be preserved. As such, the discontinuity of scales at which allostatic regulation continues to hold may impose key functional constraints on "growth spurts" in brain development[41–43], as per punctuated equilibrium theories in evolutionary biology[44,45].

Methods

Network structure.
The starting point of our study is an all-to-all resting-state fMRI-derived (N = 68, repeated 4 times) functional connectivity matrix, extracted from 91,282 voxels, and providing full coverage of the human cortex[27,28]. Each voxel is $(2\text{ mm})^3$, and thus measures a compensatory hemodynamic (*blood oxygen level dependent*) response across ~ 1,000 cortical minicolumns, or about ~ 1 M neurons (in primates and most mammals, the cortex contains a 20/80 ratio of neurons to glia[46]). Each voxel is represented as a node in a graph.

Nodes.
The nodes follow a leaky integrate-and-fire neuronal model, updated via change in membrane potential: $d\dot{V}_k/dt = -\tau_v \dot{V}_k + \sum_j w_{jk} \times I_j + \sigma_k$ with spike conditions: if $V_k(t) > V_\theta$, then the node initiates a spike by setting: $V_k(t) = V_r$. For intuitive order of magnitude estimation, the quantities tracked and the parameters are implemented in dimensionless units, with the range from the reset membrane potential and the threshold potential equal to 1. Other scaling parameters follow the same units. Here:

$V_k(t)$ is the time-dependent membrane potential (voltage) of a neuron, with k indexing the neuron integrating signal from its neighbors indexed by j.

I_j indicates the input signal from pre-synaptic neurons, indexed by j. $I_j = 1$ during the time-step neuron j fires and 0 otherwise.

w_{jk} is weight connecting the spiking neuron j to the signal-integrating neuron k. Equal to the gain (7.7×10^{-6}) times the connectivity weight.

V_r is resting membrane potential or reset value, to which the neuron returns after spiking Equal to 0 due to the dimensionless units.

V_θ is firing threshold for an action potential, upon reaching which the neuron initiates a spike. Equal to 1 due to the dimensionless units.

τ_v is rate of leakage in membrane potential from $V_k(t)$ towards V_r. Equal to 10^{-4}.

σ_k is the normally distributed noise input to neuron k.

Integration coefficient and normalization by random controls.
We compare the well-clustered coarse network to a randomly clustered control to reveal effects of strong intra-cluster connectedness on coarse network's input integration. In order to quantify the effects of integration of incoming ensemble-spikes from neighbors in a coarse-grained network, we obtain the cross-correlogram of ensemble-spikes:

- For a particular ensemble-node, for each observed ensemble-spike, we collect the most recent ensemble-spike of each of its neighboring ensemble-nodes according to their time lag τ, by which they preceded the ensemble-spike. For each time lag τ, we then sum ensemble-edge weights for all ensemble-spikes that preceded our ensemble-spike of interest by τ time steps. Any neighboring ensemble-node contributes only one ensemble-spike to this list of sums—its most recent ensemble-spike preceding the ensemble-spike of interest, and to the sum of weights at observed time-lag τ. This tells us how much information was integrated by the ensemble-node and from how many time steps prior to producing the ensemble-spike.

- For each time lag τ, we then sum the ensemble-edge weights collected over all ensemble-spikes of all ensemble-nodes, to obtain the (ensemble-edge-weighted) cross-correlogram $P(\tau)$ dependent only on time-delay τ. $P(\tau)$ tells us how much input (ensemble-spikes weighted by ensemble-edge weights) had to be received from neighboring ensemble-nodes τ time steps prior to a typical ensemble-spike in the coarse network.
- The cross-correlogram $P(\tau)$ is then normalized (divided) by the cross-correlogram $P_{random}(\tau)$ for the case if the original network was clustered randomly, and we integrate the ratio over interval of τ, on which the ratio exceeds 1, weighing by the inverse of the time lag τ to discount inputs received early, at large absolute value τ. We term this quantity *integration coefficient* (*IC*):

$$IC = \int_{P/P_{random} > 1} \frac{1}{|\tau|} \frac{P(\tau)}{P_{random}(\tau)} d\tau$$

The integration coefficient then acts as a metric of the coarse network's similarity to a network of spiking nodes. The random clustering used as a baseline coarse-graining step is done by selecting a random partitioning of the nodes into the desired number of ensemble-nodes to match the average size of ensemble-nodes observed under the full-linkage clustering used in our coarse-graining. We note that this is distinct from randomly rewiring the network prior to any coarse-graining is applied in order to dissociate the network features responsible for the observed property.

Re-wired controls. As *node degree sequence* frequently drives emergent properties of a complex network[47], we next examine the role of isolating the fine-grain network's *node strength sequence* in determining whether integrate-and-fire behavior emerges at a coarser scale. *Node strength sequence* is the equivalent of degree sequence for a *weighted network*, and is given by the sum of weights of all edges of a given node. To quantify the contribution of node strength sequence to our observations, we provide a comparison with an original network shuffled while preserving each node's strength. The shuffling follows a local rewiring rule: given 4 nodes (A, B, C and D), edges AB, CD, AC, and BD are shuffled from $AB = CD = w_1$ and $AC = BD = w_2$ to: $AB = CD = w_2$ and $AC = BD = w_1$. This preserves the local strength of node A as: $AC + AB = w_1 + w_2$ before and after the rewiring, while dissolving the network's cluster-forming topology. The value of integration coefficient is lower for a shuffled network than for original brain network (Fig. 3B, right panel). It is also notable that the shuffled network, when re-scaled, is unable to produce the same average cluster size at the same connection threshold as original brain network, nor to maintain ensemble-spikes consisting of large numbers of spikes (Fig. 3B, right panel, bottom right corner where data is absent—area indicated in white).

Do clusters show neuron-like refractory periods? By means of an auto-correlogram $P_{auto}(\tau)$ (instead of binning ensemble-spike inputs from neighboring ensemble-nodes, we bin the ensemble-node's own most recent ensemble-spike preceding the ensemble-spike of interest), we search for the time-axis re-scaling parameters that result in the strongest degree of refractory behavior, which we measure as the ratio of ensemble-spike auto-correlogram at $\tau = -2$ to that at $\tau = -1$: $P_{auto}(-2)/P_{auto}(-1)$. This helps us resolve whether a pair of consecutive ensemble-spikes of an ensemble-node had any systematic and noticeable time-gap in between, when ensemble-spiking was suppressed (Fig. 3C).

Further variations. A particular coarse-grain version of a spiking network is realized by choosing multiple parameters: (1) c (edge-weight cut-off for full-linkage clustering), which directly influences N_N—average number of nodes in an ensemble-node, (2) N_S—number of spikes in an ensemble-spike and (3) N_T time-axis re-scaling factor. The latter is formally an integer factor, by which the time bin used for the coarse level statistics is smaller than the mean inter-spike interval of a single node of the fine-grain network. Re-scaling the time-axis so that the number of spikes in an ensemble-node expected from the neuronal spiking frequency alone is ~ 1 (i.e. setting $N_T = N_N$) permits singling out time bins with spiking activity elevated by a factor of N_S as compared to the level expected from mean spiking frequency. These time-bins in the coarse ensemble-spike raster become the basis of the cross-correlogram and calculation of integration coefficient. Maximizing the integration coefficient with respect to N_T after the combination of spatial (N_N) and activity (N_S) re-scaling parameters have been estimated, allows us to establish the full set of three parameters of re-scaling. This completes the analogy between spiking neurons and ensemble-spiking clusters of neurons. In search of the integrate-and-fire behavior for coarse network, we cover the space of spatial and activity re-scaling factors as follows: N_N from minimum of ~ 3 neuron clusters up to ~ 12 in increments of 1; N_S from minimal burst of two spikes to the maximum that can be sustained by the fine spiking network. The coarse-graining procedures described can also be iterated more than once, on already clustered network, by treating it as the fine-grain network and operating with the same machinery enabling a principled dynamics-preserving hierarchical reduction of functional models.

We note that our coarse-graining procedure only resolves a single step in what is potentially a multiple step process from neurons all the way to fMRI voxels. We consider the coarse-graining step demonstrated in the work as only one of the stages of potential renormalizations allowed in the brain with transitions between all pairs of consecutive scales preserving the dynamical principle. Therefore, we consider implanting the integrate-and-fire principle at the voxel level as a reasonable assumption, given the possibility of multiple scales with the principle emerging above the neuron level.

Data availability

The data and code that support the findings of this study are available from www.lcneuro.org/tools.

References

1. Markram, H. & Perin, R. Innate neural assemblies for Lego memory. *Front. Neural Circuit* https://doi.org/10.3389/fncir.2011.00006 (2011).
2. Hill, S. *et al.* The Blue Brain Project: Calibrating the neocortical column. *BMC Neurosci.* **8**, P110–P110. https://doi.org/10.1186/1471-2202-8-S2-P110 (2007).
3. Schürmann, F., Hill, S. & Markram, H. The Blue Brain Project: Building the neocortical column. *BMC Neurosci.* **8**, P109–P109. https://doi.org/10.1186/1471-2202-8-S2-P109 (2007).
4. Innocenti, G. M. & Vercelli, A. Dendritic bundles, minicolumns, columns, and cortical output units. *Front. Neuroanat.* https://doi.org/10.3389/neuro.05.011.2010 (2010).
5. Mountcastle, V. B. The columnar organization of the neocortex. *Brain* **120**(Pt 4), 701–722 (1997).
6. Buxhoeveden, D. P. & Casanova, M. F. The minicolumn hypothesis in neuroscience. *Brain* **125**, 935–951 (2002).
7. Opris, I. & Casanova, M. F. Prefrontal cortical minicolumn: From executive control to disrupted cognitive processing. *Brain* **137**, 1863–1875. https://doi.org/10.1093/brain/awt359 (2014).
8. Perin, R., Berger, T. K. & Markram, H. A synaptic organizing principle for cortical neuronal groups. *Proc. Natl. Acad. Sci. USA* **108**, 5419–5424. https://doi.org/10.1073/pnas.1016051108 (2011).
9. Perin, R., Telefont, M. & Markram, H. Computing the size and number of neuronal clusters in local circuits. *Front. Neuroanat.* **7**, 1. https://doi.org/10.3389/fnana.2013.00001 (2013).
10. Song, S., Sjostrom, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *Plos Biol.* **3**, 507–519. https://doi.org/10.1371/journal.pbio.0030068 (2005).
11. Ikegaya, Y. *et al.* Synfire chains and cortical songs: Temporal modules of cortical activity. *Science* **304**, 559–564. https://doi.org/10.1126/science.1093173 (2004).
12. Abeles, M. Analysis of local cortical circuits. *Israel J. Med. Sci.* **16**, 222–222 (1980).
13. Shadlen, M. N. & Movshon, J. A. Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron* **24**, 67–77. https://doi.org/10.1016/S0896-6273(00)80822-3 (1999).
14. Horton, J. C. & Adams, D. L. The cortical column: A structure without a function. *Philos. Trans. R. Soc. B* **360**, 837–862. https://doi.org/10.1098/rstb.2005.1623 (2005).
15. Beggs, J. M. & Plenz, D. Neuronal avalanches in neocortical circuits. *J. Neurosci.* **23**, 11167–11177 (2003).
16. He, B. J., Zempel, J. M., Snyder, A. Z. & Raichle, M. E. The temporal structures and functional significance of scale-free brain activity. *Neuron* **66**, 353–369. https://doi.org/10.1016/j.neuron.2010.04.020 (2010).
17. Radulescu, A. & Mujica-Parodi, L. R. Network connectivity modulates power spectrum scale invariance. *Neuroimage* **90**, 436–448. https://doi.org/10.1016/j.neuroimage.2013.12.001 (2014).
18. Rubin, D., Fekete, T. & Mujica-Parodi, L. R. Optimizing complexity measures for fMRI data: Algorithm, artifact, and sensitivity. *PLoS ONE* https://doi.org/10.1371/journal.pone.0063448 (2013).
19. Roberts, J. A., Boonstra, T. W. & Breakspear, M. The heavy tail of the human brain. *Curr. Opin. Neurobiol.* **31**, 164–172. https://doi.org/10.1016/j.conb.2014.10.014 (2015).
20. van Albada, S. J., Helias, M. & Diesmann, M. Scalability of asynchronous networks is limited by one-to-one mapping between effective connectivity and correlations. *Plos. Comput. Biol.* **11**, 1004490. https://doi.org/10.1371/journal.pcbi.1004490 (2015).
21. Romaro, C., Araujo Najman, F., Lytton, W. W., Roque, A. C. & Dura-Bernal, S. NetPyNE implementation and rescaling of the Potjans-Diesmanncortical microcircuit model. *arXiv:2005.03764.* https://ui.adsabs.harvard.edu/abs/2020arXiv200503764R (2020).
22. Sierpinski, M. Sur une courbe dont tout point est un point de ramification. *C. R. Hebdomadaires Sèance l'Acad. Sci. Paris* **160**, 302–305 (1915).
23. Meunier, D., Lambiotte, R., Fornito, A., Ersche, K. D. & Bullmore, E. T. Hierarchical modularity in human brain functional networks. *Front. Neuroinform.* **3**, 37. https://doi.org/10.3389/neuro.11.037.2009 (2009).
24. Chialvo, D. R. Critical brain networks. *Phys. A* **340**, 756–765. https://doi.org/10.1016/j.physa.2004.05.064 (2004).
25. Hardstone, R. *et al.* Detrended fluctuation analysis: A scale-free view on neuronal oscillations. *Front. Physiol.* https://doi.org/10.3389/fphys.2012.00450 (2012).
26. Shin, C. W. & Kim, S. Self-organized criticality and scale-free properties in emergent functional neural networks. *Phys. Rev. E* https://doi.org/10.1103/PhysRevE.74.045101 (2006).
27. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *Neuroimage* **80**, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041 (2013).
28. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127 (2013).
29. Brown, J. A., Rudie, J. D., Bandrowski, A., Van Horn, J. D. & Bookheimer, S. Y. The UCLA multimodal connectivity database: A web-based platform for brain connectivity matrix sharing and analysis. *Front. Neuroinform.* **6**, 28. https://doi.org/10.3389/fninf.2012.00028 (2012).
30. Sofer, T., Schifano, E. D., Hoppin, J. A., Hou, L. & Baccarelli, A. A. A-clustering: A novel method for the detection of co-regulated methylation regions, and regions associated with exposure. *Bioinformatics* **29**, 2884–2891. https://doi.org/10.1093/bioinformatics/btt498 (2013).
31. Ponce-Alvarez, A., Jouary, A., Privat, M., Deco, G. & Sumbre, G. Whole-brain neuronal activity displays crackling noise dynamics. *Neuron* **100**, 1446-1459 e1446. https://doi.org/10.1016/j.neuron.2018.10.045 (2018).
32. Daffertshofer, A., Ton, R., Kringelbach, M. L., Woolrich, M. & Deco, G. Distinct criticality of phase and amplitude dynamics in the resting brain. *Neuroimage* **180**, 442–447. https://doi.org/10.1016/j.neuroimage.2018.03.002 (2018).
33. Meshulam, L., Gauthier, J. L., Brody, C. D., Tank, D. W. & Bialek, W. Coarse graining, fixed points, and scaling in a large population of neurons. *Phys. Rev. Lett.* **123**, 178103. https://doi.org/10.1103/PhysRevLett.123.178103 (2019).
34. Agrawal, V., Chakraborty, S., Knopfel, T. & Shew, W. L. Scale-change symmetry in the rules governing neural systems. *iScience* **12**, 121–131. https://doi.org/10.1016/j.isci.2019.01.009 (2019).
35. Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M. & Friston, K. J. The Dynamic brain: From spiking neurons to neural masses and cortical fields. *Plos Comput. Biol.* https://doi.org/10.1371/journal.pcbi.1000092 (2008).
36. Jansen, B. H. & Rit, V. G. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* **73**, 357–366 (1995).
37. Wendling, F., Bellanger, J. J., Bartolomei, F. & Chauvel, P. Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. *Biol. Cybern.* **83**, 367–378. https://doi.org/10.1007/s004220000160 (2000).
38. Stefanescu, R. A. & Jirsa, V. K. A low dimensional description of globally coupled heterogeneous neural networks of excitatory and inhibitory neurons. *Plos. Comput. Biol.* **4**, e1000219. https://doi.org/10.1371/journal.pcbi.1000219 (2008).
39. David, O. & Friston, K. J. A neural mass model for MEG/EEG: Coupling and neuronal dynamics. *Neuroimage* **20**, 1743–1755 (2003).

40. Moran, R. J. *et al.* A neural mass model of spectral responses in electrophysiology. *Neuroimage* **37**, 706–720. https://doi.org/10.1016/j.neuroimage.2007.05.032 (2007).
41. Jekely, G., Keijzer, F. & Godfrey-Smith, P. An option space for early neural evolution. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* https://doi.org/10.1098/rstb.2015.0181 (2015).
42. Bullmore, E. & Sporns, O. The economy of brain network organization. *Nat. Rev. Neurosci.* **13**, 336–349. https://doi.org/10.1038/nrn3214 (2012).
43. Gonzalez-Forero, M. & Gardner, A. Inference of ecological and social drivers of human brain-size evolution. *Nature* **557**, 554–557. https://doi.org/10.1038/s41586-018-0127-x (2018).
44. Bak, P. & Sneppen, K. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.* **71**, 4083–4086. https://doi.org/10.1103/PhysRevLett.71.4083 (1993).
45. Gould, S. J. & Eldredge, N. Punctuated equilibrium comes of age. *Nature* **366**, 223–227. https://doi.org/10.1038/366223a0 (1993).
46. Herculano-Houzel, S. Brains matter, bodies maybe not: The case for examining neuron numbers irrespective of body size. *Ann. N. Y. Acad. Sci.* **1225**, 191–199. https://doi.org/10.1111/j.1749-6632.2011.05976.x (2011).
47. Liu, Y. Y., Slotine, J. J. & Barabasi, A. L. Controllability of complex networks. *Nature* **473**, 167–173. https://doi.org/10.1038/nature10011 (2011).

Acknowledgements

Author contributions

Conceptualization, P.T., H.T.S., A.A., and L.R.M.P.; Investigation, A.A.; Writing—Original Draft, P.T. and A.A.; Writing—Critical Review & Editing, L.R.M.P. and H.T.S.; Funding Acquisition, H.T.S. and L.R.M.P.; Resources, H.T.S. and L.R.M.P.; Supervision, H.T.S. and L.R.M.P.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.R.M.-P. or H.T.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.