Supplemental Figures

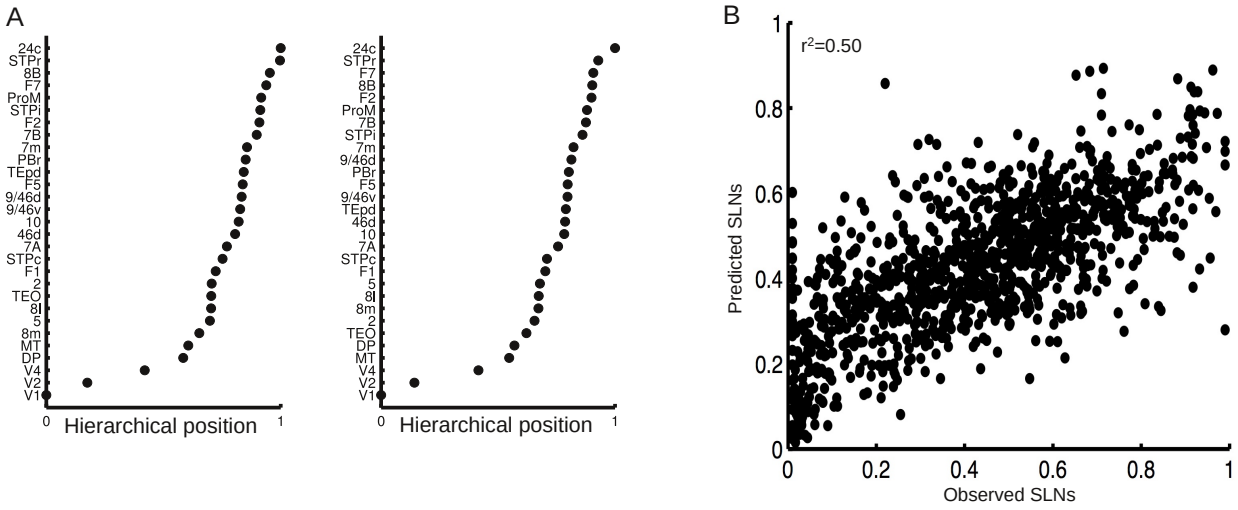

Figure S1. Related to Figure 2. Hierarchy fitted from pairwise SLN relationships. (A) Left panel: Hierarchy fitted from logistic regression (and used in main text). The hierarchical position of an area is normalized to lie between 0 and 1. Right panel: Hierarchy fitted from beta regression (Cribari-Neto and Zeileis, 2010). (B) SLN values predicted from logistic regression compared to observed SLNs.

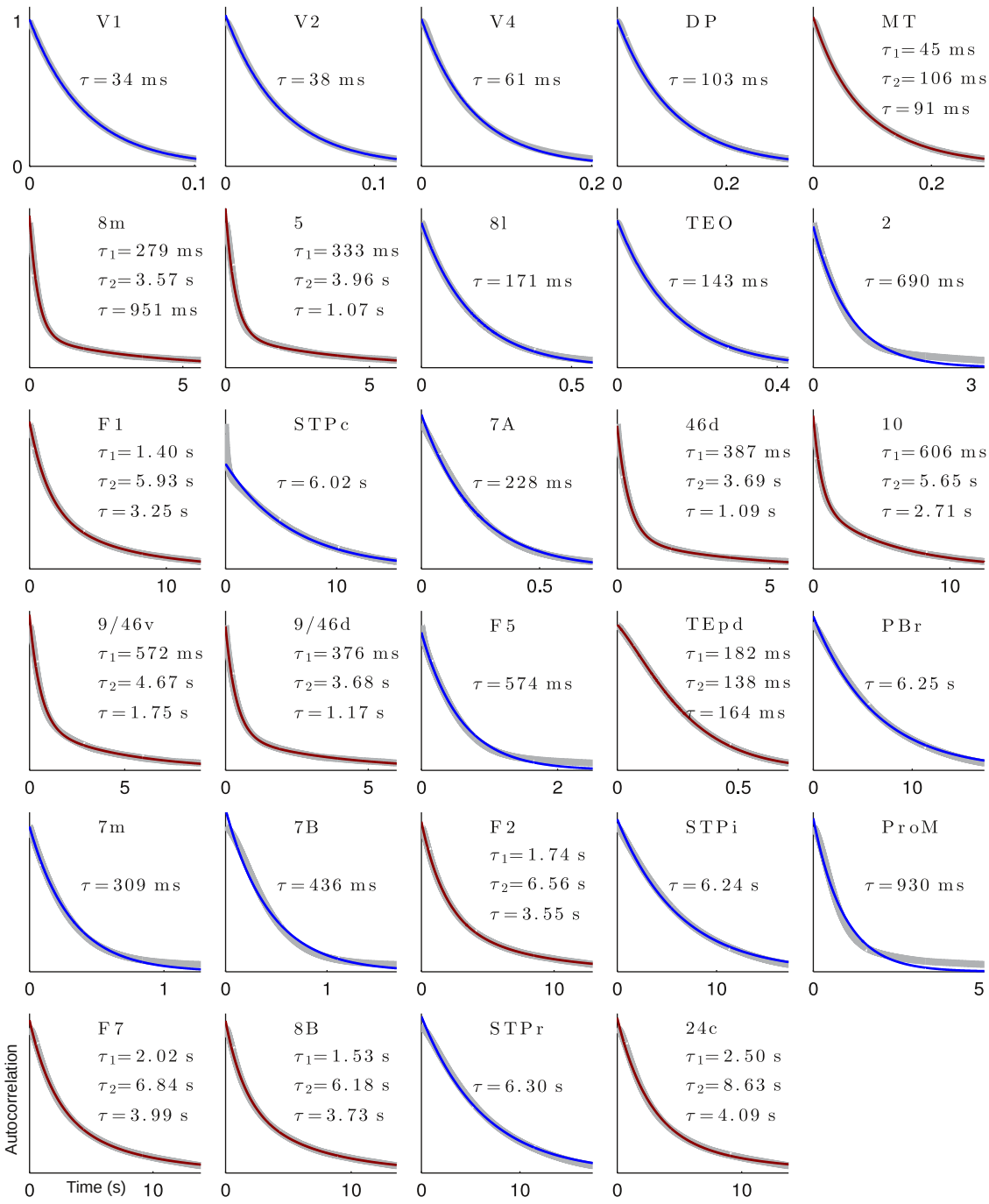

Figure S2. Related to Figure 3. Timescales in response to white-noise input to V1. Data shown in grey, single exponential fits in blue and double exponential fits in dark red. For double exponential fits, τ_1 and τ_2 are the time-constants of individual exponentials, and τ is a weighted average of τ_1 and τ_2, with weights given by the amplitudes of the exponentials.

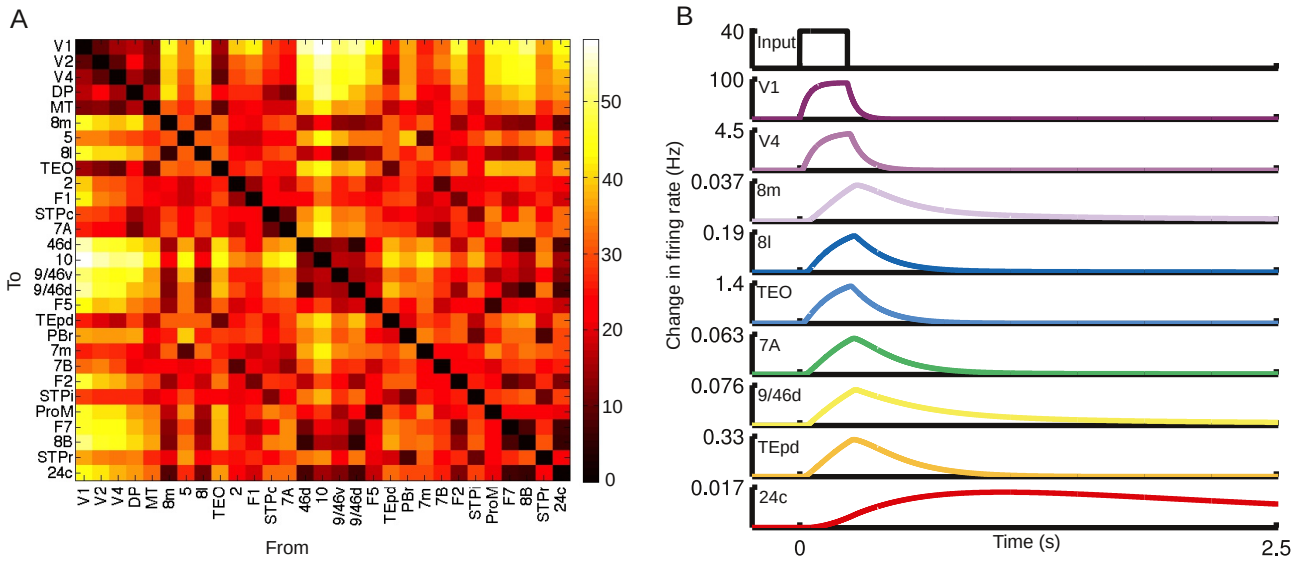

Figure S3. Related to Figure 3. Response of a network with inter-areal conduction delays. (A) Distances (in mm) between the nodes of the network (Ercsey-Ravasz et al., 2013). (B) Response of the network to a pulse of input to area V1. Conduction delays between nodes are imposed using the distances in panel A and a conduction velocity of 1.5 m/s (Deco et al., 2009).

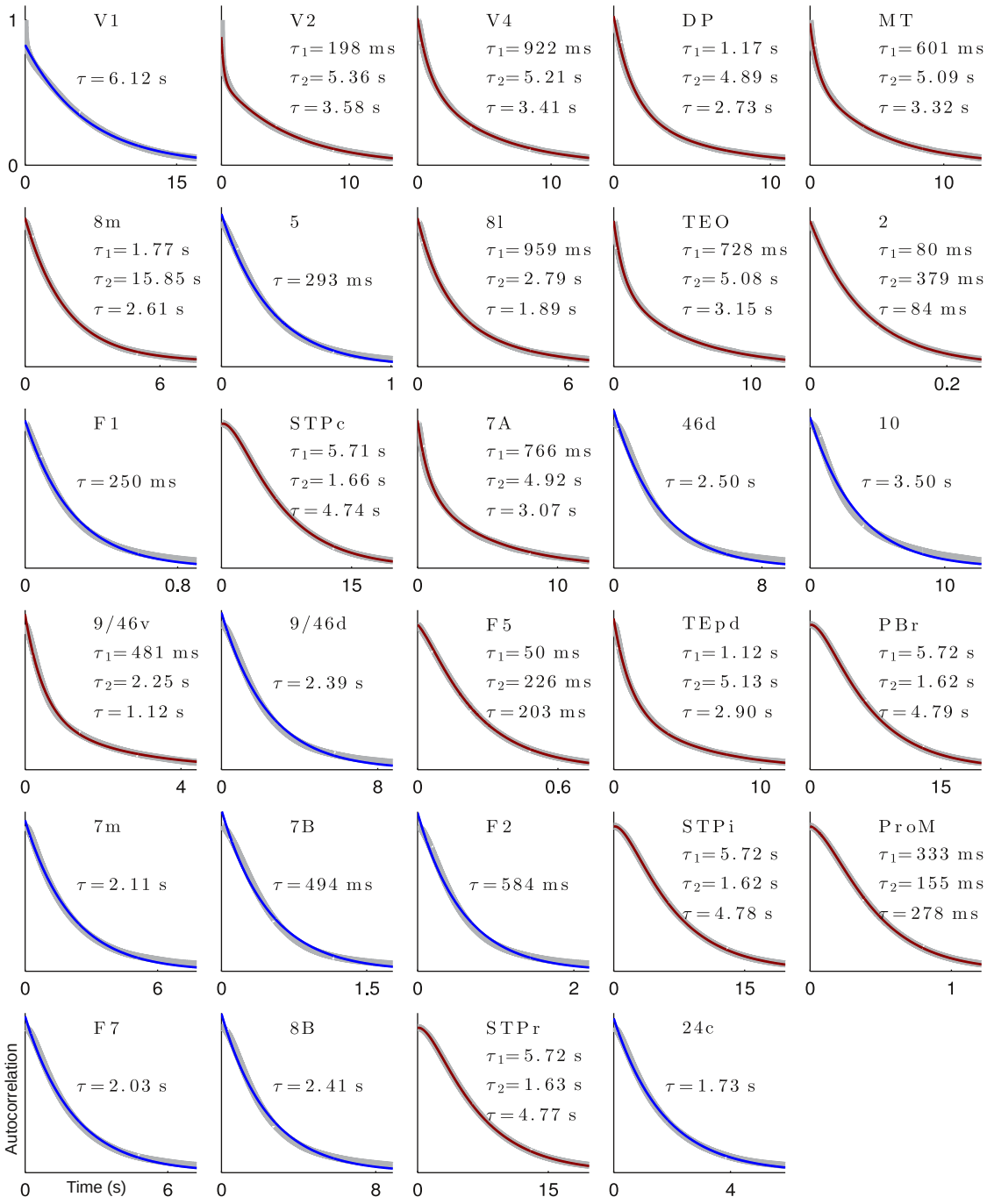

Figure S4. Related to Figure 4. Timescales from exponential fits of activity in response to white-noise input to Area 2. Colors as in Figure S2.

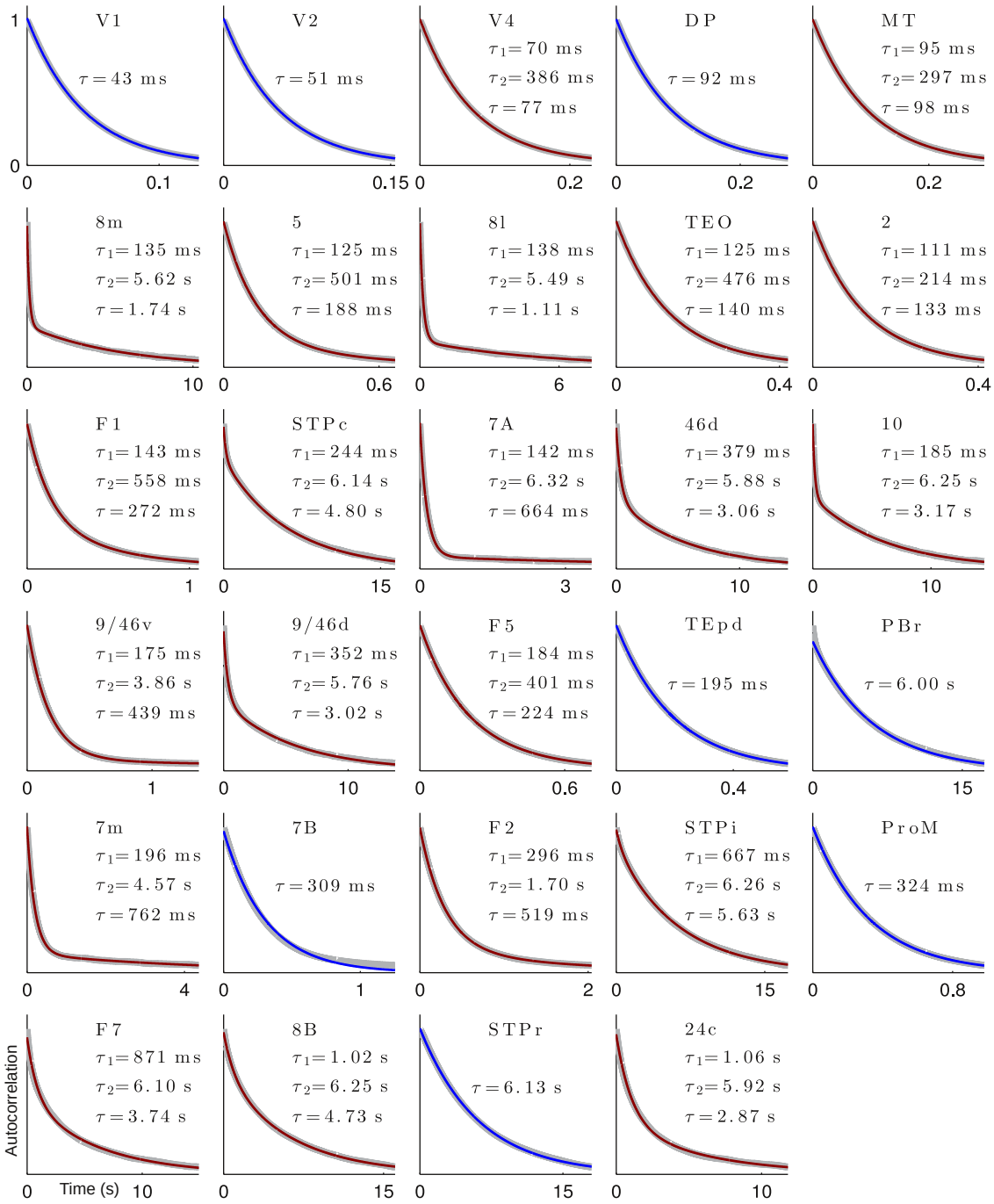

Figure S5. Related to Figure 8. Timescales from exponential fits of resting-state activity (i.e., equal white-noise input to all areas). Colors as in Figure S2.

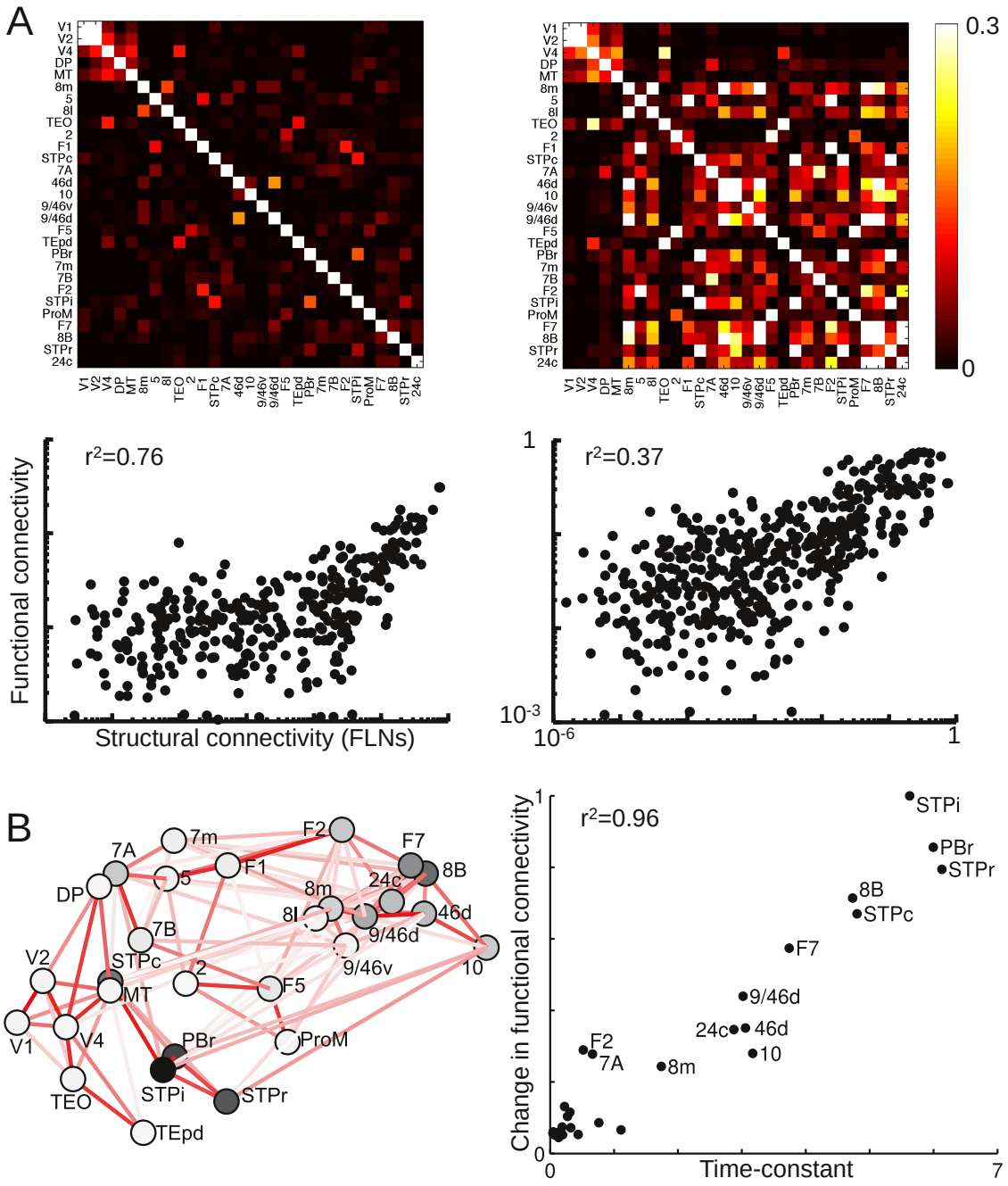

Figure S6. Related to Figure 8. Functional connectivity of simulated BOLD signal. (A) As in Figure 8A, the network on the left has the same local properties at each node, while the network on the right has a gradient of local recurrent strengths. Firing rate is convolved with a gamma function to generate a simulated BOLD signal (Boynton et al., 1996). Top panel: functional connectivity in response to background white noise input to each node. Bottom panel: functional connectivity (correlations in BOLD) vs. structural connectivity (FLN) for non-zero projections. (B) Effect of lesioning areas on functional connectivity measured via simulated BOLD signal. Plots are as in Figure 8B.

Supplemental Experimental Procedures

Several of these sections are expanded versions of the corresponding sections in Experimental Procedures in the main text. To make these descriptions self-contained, the relevant portions from the main text are repeated here.

Anatomical data

The connectivity data are from an on-going project to quantitatively measure all connections between cortical areas in the macaque cortex, with areas defined according to a 91 area parcellation scheme (Markov et al., 2014a). Descriptions of data collection can be found in Markov et al. (2011; 2014a). Briefly, connection strengths between areas are measured by counting the number of neurons labeled by retrograde tracer injections. The number of neurons labeled in a projection ranges from a few neurons to on the order of 100,000 neurons. To control for injection size, these counts are then normalized by the total number of neurons labeled in the injection, yielding a fractional weight or FLN (Fraction of Labeled Neurons) for each pathway, defined as

$$FLN_{B\to A} = \frac{\#\text{ neurons projecting to area A from area B}}{\text{total neurons projecting to area A from all areas}}$$

The corresponding weights span 5 orders of magnitude. So far, 29 areas have been injected and we use the subnetwork consisting of these areas. In this network the presence or absence of all connections is known bidirectionally, and 66% of possible connections exist in the network, though with widely varying strengths.

We also use data on the fraction of neurons in each projection that originate in the upper layers of the source area, which we call the SLN, for Supragranular Layer Neurons (Markov et al., 2014b). These are defined as

$$SLN_{B\to A} = \frac{\#\text{ supragranular neurons projecting to area A from area B}}{\#\text{ neurons projecting to area A from area B}}.$$

The data are included in Table S1, and all data can be downloaded from www.core-nets.org.

Hierarchy and low-dimensional connectivity embedding

In the visual system, projections directed from early visual areas to higher-order areas (i.e. increasing size of receptive field, position-invariance, and so on) tend to originate in the supragranular layers of the cortex and terminate in layer 4 (Felleman and Van Essen, 1991; Barone et al., 2000). Conversely, projections from higher-order areas to early visual areas

originate in the infragranular layers and terminate outside of layer 4. This observation was systematized by Felleman and Van Essen (1991), who used these anatomical constraints to place cortical areas in a hierarchical ordering.

Felleman and Van Essen used a discrete classification of projections: in their framework projections are either feedforward, feedback or lateral depending on where the majority of projections originate and terminate. However, such binary relations are typically insufficient to specify a unique hierarchy (Hilgetag et al., 1996). Subsequently, it was observed that rather than classifying a projection as feedforward, feedback or lateral, the fraction of neurons in a projection originating in the supragranular layers (the SLN) could be used as a continuous measure of hierarchical displacement: the difference of the SLN from 50% is positive for feedforward projections and negative for feedback projections, and its magnitude gets larger as a projection moves further away from lateral (Barone et al., 2000). For example, a projection with an SLN of 90% would be very strongly feedforward, while a projection with an SLN of 65% would be only moderately feedforward. Using these values, the Felleman and van Essen hierarchy could be reproduced using observations of connections to only two areas (V1 and V4) (Barone et al., 2000).

To construct the hierarchy we follow a similar framework to Markov et al. (2014b) and use a generalized linear model. We assign hierarchical values to each area such that the difference in values predicts the SLN of a projection. Specifically, we assign a value H_i to each area A_i such that

$$SLN_{A_j \to A_i} \approx g^{-1}(H_i - H_j). \tag{1}$$

We choose g^{-1} to be a logistic function (logistic regression), which is standard for probabilities and fractional values, but we note that other functions yield similar values (Figure S1A). We have one such constraint for each projection (536 in total), and we find the set of hierarchical values that best fit these constraints. In the fit we weight the contribution of each projection by the log of its FLN to preferentially match stronger and less noisy projections. The resulting best fit hierarchy is shown in the left panel of Figure S1A. We then normalize by the maximum hierarchical value yielding $h_i = H_i/H_{max}$.

We extract the spine counts in Figure 2B from Elston (2007) and plot the areas in common with our data set. The parcellation in that paper is coarser than the parcellation we use, so we report the results in terms of that parcellation. For area 7 we average together the hierarchical positions of 7A, 7B and 7m; for 6 we average F2, F5 and F7; and for 46 we average together 46d, 9/46d and 9/46v.

For the two-dimensional circular embedding of Figure 2C, we convert the FLN to a measure of dissimilarity according to

$$d(A_i, A_j) = \begin{cases} -\log(\mathrm{FLN}(A_i, A_j)) & \text{for } \mathrm{FLN}(A_i, A_j) > 0 \\ -\log(\mathrm{FLN}_{\min}) & \text{for } \mathrm{FLN}(A_i, A_j) = 0. \end{cases} \tag{2}$$

Here, A_i is the ith area, and $\mathrm{FLN}_{\min}$ is some value less than the smallest FLN in the network.

We use $\mathrm{FLN}_{\min} = 10^{-7}$ but the results are robust to the precise choice of this value. We then assign angles θ_i to each area such that $d(A_i, A_j) \approx R \min(|\theta_i - \theta_j|, 2\pi - |\theta_i - \theta_j|)$, where R is a single free parameter. We fix area V1 to have $\theta = 0$, but choosing any other area to fix would simply rotate the plot. Finally, we plot the areas on a 2-dimensional polar plot with $\theta(A_i) = \theta_i$ and $R(A_i) = \sqrt{1 - h_i}$.

Model architecture

Each of the 29 nodes consists of an excitatory and an inhibitory population, which summarize the effective dynamics of the area. Populations are described by

$$\begin{aligned} \tau_E \frac{d}{dt}\nu_E &= -\nu_E + \beta_E \left[I_E\right]_+ \\ \tau_I \frac{d}{dt}\nu_I &= -\nu_I + \beta_I \left[I_I\right]_+ . \end{aligned} \tag{3}$$

ν_E is the firing rate of the excitatory population, with intrinsic time constant τ_E and input current I_E, and for which the f-I curve has slope β_E. $[I_E]_+ = \max(I_E, 0)$. The inhibitory population has corresponding parameters τ_I, I_I and β_I. Values for τ_E, τ_I, β_E and β_I are given below and are taken from Binzegger et al. (2009).

At each node, the input currents have a component from within the area (i.e. local input) and another that comes from other areas:

$$\begin{aligned} I_E^i &= (1 + \eta h_i)\big(w_{EE}\nu_E^i + I_{lr,E}^i\big) - w_{EI}\nu_I^i + I_{ext,E}^i \\ I_I^i &= (1 + \eta h_i)\big(w_{IE}\nu_E^i + I_{lr,I}^i\big) - w_{II}\nu_I^i + I_{ext,I}^i. \end{aligned} \tag{4}$$

w_{EE} and w_{EI} are couplings to the excitatory population from the local excitatory and inhibitory population respectively, $I_{lr,E}^i$ is the long-range input to the excitatory population, and $I_{ext,E}^i$ is external input (both stimulus input and any noise we add to the system). w_{IE}, w_{II}, $I_{lr,I}^i$ and $I_{ext,I}^i$ are corresponding parameters for the inhibitory population.

Following Binzegger et al. (2009), we write $w_{ij} = \alpha_j S_{ij}$, where i and j can be E or I. α_E (α_I) measures charge introduced per excitatory (inhibitory) spike times transmitter release probability; both are slightly modified from Binzegger et al. (2009). S_{ij} is the number of synapses from cells of type j to cells of type i, taken from the counts for layer 2/3 cells in Binzegger et al. (2004). Inhibitory values are weighted averages of basket, double bouquet and chandelier cells, with weights chosen according to their projections to the excitatory population.

We scale the excitatory inputs to an area, both local and long-range, by its position in the hierarchy, h_i. h_i is normalized between 0 and 1, and η is a scaling parameter that controls the effect of hierarchy. By setting $\eta = 0$ we remove intrinsic differences between areas. Note

that we scale both local and long-range projections with hierarchy, rather than just local projections, in accordance with the observations of Markov et al. (2011), who find that the proportion of local to long-range connections is approximately conserved across areas.

Long-range input is modeled as excitatory current to both excitatory and inhibitory cells:

$$I^i_{lr,E} = \mu_{EE} \sum_{j=1}^{N} FLN_{ij} \nu^j_E$$
$$I^i_{lr,I} = \mu_{IE} \sum_{j=1}^{N} FLN_{ij} \nu^j_E. \quad (5)$$

Here j ranges over all areas. $I^i_{lr,E}$ and $I^i_{lr,I}$ are the inputs to the excitatory and inhibitory populations, ν^j_E is the firing rate of the excitatory population in area j and FLN_{ij} is the FLN from area j to area i. μ_{EE} and μ_{IE} are scaling parameters that control the strengths of long-range input to the excitatory and inhibitory populations, respectively, and do not vary between connections; all the specificity comes from the FLNs. Long-range connectivity is thus determined by three parameters: μ_{EE} and μ_{IE} control the connection strengths of long-range projections, and η maps the hierarchy into excitatory connection strengths.

We can choose the excitatory to inhibitory ratio of an input current, $\gamma = I_{inp,E}/I_{inp,I}$, such that the steady-state firing rate of the excitatory population does not change when the current is present. Given input of $I_{inp,E}$ to the excitatory population, an input of $\gamma I_{inp,E}$ to the inhibitory population increases the inhibitory firing rate sufficiently to cancel out the additional input to the excitatory population. We call such inputs balanced. We choose μ_{EE} and μ_{IE} with a ratio slightly above this value so that projections are weakly excitatory.

Parameters are τ_E=20 ms, τ_I=10 ms, β_E=0.066 Hz/pA, β_I=0.351 Hz/pA, w_{EE} = 24.3 pA/Hz, w_{IE} = 12.2 pA/Hz, w_{EI} = 19.7 pA/Hz, w_{II} = 12.5 pA/Hz, μ_{EE} = 33.7 pA/Hz, μ_{IE} = 25.3 pA/Hz and $\eta = 0.68$.

Network with conduction delays

In our simulations we ignore conduction delays between areas. While these will be important for oscillations, synchronization and other fine temporal structure, the timescales we consider are typically slow enough that conduction delays do not play an important role.

In Figure S3 we demonstrate that our results hold in a network with realistic conduction delays. We use distances from the same data set as the connectivity strengths (Ercsey-Ravasz et al., 2013) and, to ensure a fair comparison, assume a relatively low conduction velocity of

1.5 m/s (Deco et al., 2009). As shown in Figure S3B, the response of this network to a pulse of input to area V1 is almost identical to that of a network without conduction delays.

Scrambled connectivity

For the simulations shown in Figure 5B, we scramble the connectivity matrix by permuting all entries of the matrix randomly. For Figure 5C, we preserve the absent entries and permute the non-zero entries. Note that the connectivity data show specificity both in terms of which projections exist and in their strengths, and both the probability of a connection and its strength decay exponentially with distance between areas (Markov et al., 2011; 2013; 2014a; Ercsey-Ravasz et al., 2013). In particular, nearby areas tend to be strongly connected and to have similar timescales (see Fig. 2C); thus scrambling projections should reduce the separation of timescales.

We examine the response of these scrambled networks to a pulse of input to all areas, similar to the "resting-state" condition. In the intact network, areas are dominated by a few timescales and are well fit by one or two summed exponentials. However, a number of the scrambled networks show responses that consist of many mixed timescales and are not well described by two exponentials. Thus we use a non-parametric measure of timescale: we compute the time taken after pulse offset for the area's activity to decay to within 5% of its value at baseline. Scrambling the connection strengths makes about 20% of networks unstable, meaning that responses to input grow instead of decaying, and we exclude these networks. We then compute the median and the 5th, 10th, 90th and 95th percentile of the decay time distribution for each area, and contrast it with values for the intact network.

Functional connectivity for a linear network

If a linear network is driven by white noise input then, away from the threshold, it evolves according to the equation

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + \boldsymbol{I} + B\boldsymbol{\xi}(t), \tag{6}$$

where $\boldsymbol{I}$ is the mean of the noise, B is its covariance matrix and A is the coupling matrix, which includes any intrinsic leak of activity.

In the steady-state the covariance, C, of this matrix is the solution to the equation (Gardiner, 1985)

$$AC + CA^{\dagger} + BB^{\dagger} = 0 \tag{7}$$

This equation can be solved given the eigenvector basis (Deco et al., 2013). In the eigenvector decomposition, $A = V\Lambda V^{-1}$, where Λ is the diagonal matrix of eigenvalues and the columns

of V are the right eigenvectors of A. Define

$$\tilde{Q} = V^{-1}BB^{\dagger}V^{-\dagger} \tag{8}$$
$$M_{ij} = -\frac{\tilde{Q}_{ij}}{(\lambda_i + \lambda_j^*)}$$

Then $C = VMV^{\dagger}$.

As an aid to intuition, assume that A is a normal matrix so that $V^{-1} = V^{\dagger}$. Then $\tilde{Q} = V^{\dagger}BB^{\dagger}V$, and the covariance matrix of the network is a rescaled version of the covariance structure of the input noise.

If, as in the simulations of Figure 8, the input noise is independent and identical at each node, then the covariance matrix of the noise is diagonal with constant entries (and all correlations come from the structure of the network). If this has the value σ^2 at each node then, for a normal matrix, $\tilde{Q}_{ij} = \sigma^2\delta_{ij}$, and M is diagonal with ith entry $\tau_i\sigma^2/2$, where $\tau_i = -1/\lambda_i$. Hence the covariance of the ith eigenmode is proportional to its corresponding timescale.

Now $C = VMV^{\dagger}$, meaning that the matrix is rotated out of the eigenvector basis giving a non-diagonal matrix. Thus eigenvectors that are more broadly shared contribute more to the functional connectivity. In this case $C \propto A^{-1}/2$.

We also note that Baria et al. (2013) conduct a similar analysis on a linear network with nodes having identical properties and binary connectivity, and find that nodes with more anatomical connections and, consequently, higher functional connectivity show greater activity at low frequencies (i.e., slower timescales).

Functional connectivity with hemodynamic response function

For Figure S6, we convolve the firing rates of the excitatory population at each node with a hemodynamic response function of the form

$$H(t) = \frac{(t-d)e^{-(t-d)/\tau_h}}{\tau_h^2},$$

with timescale $\tau_h = 1.25$ s and delay $d = 2.25$ s (Boynton et al., 1996). This yields a simulated BOLD signal, and we calculate the functional connectivity as the correlation matrix of this activity.

Nonlinear network

The single area model is a variant of the model developed in Wong and Wang (2006) as a simplified mean-field version of the spiking network of Wang (2002). There the dynamics were assumed to be dominated by the slow time-constant of NMDA synapses, and the activity of the inhibitory population was incorporated into the effective connection strengths between the excitatory populations. As in that study, we assume that the dynamics of the excitatory population are modeled by a dimensionless gating variable, s_N, reflecting the fractional activation of the NMDA conductance, with timescale set by the slow NMDA time-constant. However, we also consider an inhibitory population, modeled with a threshold-linear differential equation (as in the previous sections).

The equation for the excitatory population is

$$\begin{aligned} \nu_E^i &= \phi\left(I_E^i\right) = \phi\left((1+\eta h_i)\left(w_{EE}s_N^i + I_{lr,E}^i\right) - w_{EI}\nu_I^i + I_{ext,E}^i\right) \\ \tau_N \frac{d}{dt} s_N^i &= -s_N^i + \gamma\tau_N(1-s_N^i)\nu_E^i \end{aligned} \tag{9}$$

Here ν_E is the excitatory firing rate and s_N is the NMDA gating variable, which is bounded between 0 and 1. ϕ models the firing rate-current dependence of a leaky integrate-and-fire neuron (Abbott and Chance, 2005) and is defined as

$$\phi(I_{syn}) = \frac{aI_{syn} - b}{1 - \exp(-d(aI_{syn} - b))}$$

with $a = 0.27$ Hz/pA, $b = 108$ Hz and $d = 0.154$ s.

The inhibitory population is described with a threshold-linear equation as before.

$$\tau_I \frac{d}{dt}\nu_I^i = -\nu_I^i + \beta_I\left[I_I^i\right]_+ = -\nu_I^i + \beta_I\left[(1+\eta h_i)\left(w_{IE}s_N^i + I_{lr,I}^i\right) - w_{II}\nu_I^i + I_{ext,I}^i\right]_+ .$$

Parameter values are: $\tau_N = 60$ ms, $\tau_I = 10$ ms, $\gamma = 0.641$, $w_{EE} = 250.2$ pA, $w_{EI} = 8.110$ pA/Hz, $w_{IE} = 303.9$ pA and $w_{II} = 12.5$ pA/Hz.

Supplemental References

Abbott L.F., and Chance F.S. (2005). Drivers and modulators from push-pull and balanced synaptic input. Prog. Brain Res. *149*, 147–155.

Baria A.T., Mansour A., Huang L., Baliki M.N., Cecchi G.A., Mesulam M.M., and Apkarian A.V. (2013). Linking human brain local activity fluctuations to structural and functional network architectures. Neuroimage *73*, 144–155.

Barone P., Batardiere A., Knoblauch K., and Kennedy H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. J. Neurosci. *20*, 3263–3281.

Binzegger T., Douglas R.J., and Martin K.A. (2004). A quantitative map of the circuit of cat primary visual cortex. J. Neurosci. *24*, 8441–8453.

Binzegger T., Douglas R.J., and Martin K.A. (2009). Topology and dynamics of the canonical circuit of cat V1. Neural Netw *22*, 1071–1078.

Boynton G.M., Engel S.A., Glover G.H., and Heeger D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. *16*, 4207–4221.

Cribari-Neto F., and Zeileis A. (2010). Beta regression in R. J. Stat. Softw. *34*, 1–24.

Deco G., Jirsa V., McIntosh A.R., Sporns O., and Kotter R. (2009). Key role of coupling, delay, and noise in resting brain fluctuations. Proc. Natl. Acad. Sci. U.S.A. *106*, 10302–10307.

Deco G., Ponce-Alvarez A., Mantini D., Romani G.L., Hagmann P., and Corbetta M. (2013). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. J. Neurosci. *33*, 11239–11252.

Elston G.N. (2007). Specialization of the neocortical pyramidal cell during primate evolution. In Evolution of Nervous Systems: A Comprehensive Reference, J.H. Kass, and T.M. Preuss, eds. (New York: Elsevier), vol. 4, pp. 191–242.

Ercsey-Ravasz M., Markov N.T., Lamy C., Van Essen D.C., Knoblauch K., Toroczkai Z., and Kennedy H. (2013). A predictive network model of cerebral cortical connectivity based on a distance rule. Neuron *80*, 184–197.

Felleman D.J., and Van Essen D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex *1*, 1–47.

Gardiner C.W. (1985). Handbook of stochastic methods, vol. 3 (Springer Berlin).

Hilgetag C.C., O'Neill M.A., and Young M.P. (1996). Indeterminate organization of the visual system. Science *271*, 776–777.

Markov N.T., Ercsey-Ravasz M., Van Essen D.C., Knoblauch K., Toroczkai Z., and Kennedy H. (2013). Cortical high-density counterstream architectures. Science *342*, 1238406.

Markov N.T., Ercsey-Ravasz M.M., Ribeiro Gomes A.R., Lamy C., Magrou L., Vezoli J., Misery P., Falchier A., Quilodran R., Gariel M.A., et al. (2014a). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cereb. Cortex *24*, 17–36.

Markov N.T., Misery P., Falchier A., Lamy C., Vezoli J., Quilodran R., Gariel M.A., Giroud P., Ercsey-Ravasz M., Pilaz L.J., et al. (2011). Weight consistency specifies regularities of macaque cortical networks. Cereb. Cortex *21*, 1254–1272.

Markov N.T., Vezoli J., Chameux P., Falchier A., Quilodran R., Huis-soud C., Lamy C., Misery P., Giroud P., Ullman S., et al. (2014b). The anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. J. Comp. Neurol. *522*, 225–259.

Wang X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955–968.

Wong K.F., and Wang X.J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328.