Scaling of gene transcriptional gradients with brain size across mouse development

Lau Hoi Yan Gladys[1,2], **Alex Fornito**[1], **and Ben D. Fulcher**[2,✉]

[1]The Turner Institute for Brain and Mental Health, School of Psychological Sciences, and Monash Biomedical Imaging, Monash University, Victoria, Australia.
[2]School of Physics, The University of Sydney, NSW 2006, Australia.

The structure of the adult brain is the result of complex physical mechanisms acting through development. Accordingly, the brain's spatial embedding plays a key role in its structural and functional organization, including the gradient-like patterning of gene expression that encodes the molecular underpinning of functional specialization. But we do not understand how this transcriptional heterogeneity is spatially organized across the major alterations in brain geometry that occur through development. Here we investigate the spatial embedding of transcriptional patterns of over 1800 genes across seven time points through mouse-brain development using data from the Allen Developing Mouse Brain Atlas. We find that transcriptional similarity decreases exponentially with distance across all developmental time points, with a correlation length scale that satisfies a power-law scaling relationship with a linear dimension of brain size. This scaling suggests that the mouse brain achieves a characteristic spatial balance between local transcriptional similarity (within functionally specialized brain areas) and longer-range diversity (between functionally specialized brain areas) throughout its development. Extrapolating this mouse-developmental scaling relationship to predict the correlation length of gene expression in the human cortex yields a slight overestimate, consistent with the human cortex being more molecularly diverse and functionally specialized than the mouse brain. We develop a simple model of brain growth as spatially autocorrelated gene-expression gradients that expand through development, which captures key features of the mouse developmental data. Complementing the well-known exponential distance rule for structural connectivity, our findings thus characterize an exponential distance rule for transcriptional gradients that scales across mouse-brain development, providing new understanding of the molecular patterns underlying the functional specialization in the brain.

gene expression | transcription | gradients | brain development
Correspondence: ***ben.fulcher @sydney.edu.au***

Introduction

The brain's structure is the result of physical mechanisms playing out through development. For example, as short-range axonal connections incur a lower metabolic cost than long-range connections, the brain's structural connectome encodes a balance between efficient brain function against this physical cost (1–5). The connection probability between pairs of neural elements decays with their separation distance in *C. elegans* (6), mouse (7, 8), rat (9), zebrafish (10), non-human primate (11), and human (12), and is frequently characterized as an exponential distance rule (13–15). Indeed, very simple spatial rules can explain much of the high-order statistics of connectome topology, including its modularity, long-tailed degree distribution, and existence of a network core (12, 14, 16–19). Spatial proximity also plays a strong role in gradients of gene expression, with nearby areas exhibiting more similar gene-expression profiles than more distant areas in the head neurons of *C. elegans* (6), the mouse brain (7, 20), and the human cortex (21–26).

The spatial embedding of molecular patterning provides important clues about how the brain's functional specialization is topographically organized. Macroscopic functional gradients across the human cortex maximize the spatial distance between areas functionally involved in sensory perception and those involved in integrative cognition (27, 28). These hierarchical functional gradients are underpinned by a corresponding variation in structural microarchitecture (29–31). But given the constraint of a fixed brain size, how does the brain strike the right balance local molecular similarity—similar gene-expression patterns enable the functional specialization within a given brain area—while still 'fitting in' many differentially specialized areas into the brain (longer-range molecular diversity). Furthermore, how does the spatial embedding of transcriptional gradients emerge through development: do different rules hold early in development compared to later? Although macroscopic brain organization in adult is often attributed to gradients (e.g., in transcriptional factors) set up during development (32), to our knowledge, no study has analyzed the spatial embedding of gene expression through development.

The Allen Developing Mouse Brain Atlas (ADMBA) (33) is a comprehensive database of gene-expression across mouse-brain development. The atlas contains

in situ hybridization measurements for approximately 2100 genes at seven developmental time points, available in a three-dimensional reference space registered to an anatomical reference atlas (for each developmental time point). Many analyses of gene transcription through mouse-brain development have shed light on how the spatiotemporal variation of specific genes through development shape different aspects of brain structure (33–36). But these developmental atlas data have not previously been analyzed from the viewpoint of spatially embedded gradients of transcriptional similarity. By characterizing gradients, we can address global questions about transcriptional patterning through development in a way that that does not require tracking specific, anatomically defined areas.

Here we study how the correlated gene expression (CGE), defined as the similarity in the gene-expression signature of two points in the brain, is spatially embedded across mouse-brain development. Matching the well-known exponential distance rule for connectivity, we find an exponential decay in CGE at each developmental time point, with a spatial correlation length that scales with a linear dimension of brain size. Extrapolating the relationship fitted on developing mouse-brain data, allows us to obtain a slight overestimate for the correlation length of CGE in the human cortex. We then use a simple model to show that the patterns are consistent with brain growth in which spatially autocorrelated gene-expression gradients expand with brain size over time.

Methods

Data Retrieval. We analyze an open dataset from the ADMBA, which contains *in situ* hybridization (ISH) expression estimates for approximately 2100 genes in C57Bl/6J mice across seven developmental stages, spanning both embryonic (E11.5, E13.5, E15.5, and E18.5) and postnatal (P4, P14, and P28) stages (33). These genes were selected to encompass relevant transcription factors, neurotransmitters and their receptors, cell-type or regional marker genes, genes relevant to developmental signaling pathways, and genes corresponding to common drug targets or implicated in neurodevelopmental disorders (33). The raw imaging data, encompassing 434 946 brain sections of high-resolution ISH, are registered to a three-dimensional mouse-brain atlas at each time point. We accessed this registered data at the voxel level via the Allen Software Development Kit (SDK, 2015). Our code for retrieving data (in python) and for data processing and analysis (Matlab) is available at `https://github.com/NeuralSystemsAndSignals/DevelopingMouse`. At each time point, we retrieved expression data for all available genes. Expression data for each gene was obtained as 'expression energy' (ISH signal intensity) in each voxel in the three-dimensional grid. Distances between pairs of voxels were computed as Euclidean distances in three-dimensional space. This computation required transforming between (integer-valued) coordinate space and physical space using the spatial resolution used at each time point: E11.5 (0.08 mm), E13.5 (0.1 mm); E15.5 (0.12 mm), E18.5 (0.14 mm), P4 (0.16 mm), P14 (0.2 mm), and P28 (0.2 mm).

Filtering and Normalization. At each developmental time point, we first filtered out voxels that were labeled as spinal cord or were unannotated. We kept genes with a valid expression value for at least 70% of the remaining voxels (i.e., not labeled as missing). A total of 1861 genes satisfied these quality control criteria at all seven time points. To ensure temporal consistency of our analyzed genes, we restricted our analysis to these 1861 genes. We then retained only voxels with valid expression data for at least 70% of these genes. For voxels and genes that survived these filtering steps, we generated a voxel × gene expression matrix. The number of voxels varied across time as 5034 (E11.5), 9473 (E13.5), 11 314 (E15.5), 11 313 (E18.5), 19 754 (P4), 21 579 (P14), and 24 822 (P28).

Raw expression values measured using *in situ* hybridization are not comparable between genes (37). We accounted for this at each time point by normalizing expression values of each gene using a standard sigmoidal transformation, $S(x) = [1+\exp(-x)]^{-1}$, followed by linear rescaling to the unit interval (for visualization) (7, 21). The resulting voxel × gene matrix, G_t, at a given time point, t, was used for further analysis.

Quantifying the spatial embedding of CGE. To understand the spatial embedding of gene-expression patterns, we analyzed the distance-dependence of correlated gene expression (CGE) at each developmental time point. CGE is a measure of gene-expression similarity between a pair of voxels and was computed as the Pearson correlation coefficient between normalized gene-expression vectors (rows of a given voxel × gene matrix, G_t, defined above). At each time point, we computed CGE for every pair of voxels. Because the distribution of pairwise distances is concentrated near lower distances, we wanted to prevent the exponential fitting from being biased towards fitting these smaller length scales, but to instead fit well across the full range of distances. To do this, we first performed an equiprobable distance binning of the data using twenty bins, and computed the mean CGE value in each bin (and summarizing each bin as its center). To this binned data, we then fitted a three-parameter exponential function, $\mathrm{CGE}(d) = A\exp(-d/\lambda) + f_0$, using nonlinear least squares and identified the fitted parameters as: spatial correlation length, λ, strength A, offset, f_0. Adjusted R^2 was used as a goodness-of-fit statistic. Note that $\mathrm{CGE}(d=0) = 1$ corresponds to self-correlations, and thus fitting data that includes self-correlations should have a fixed strength parameter $A = 1$. However, here we excluded self-correlations and allowed $A < 1$. When characterizing distances relative to brain size, as $d_{\mathrm{rel}} =$

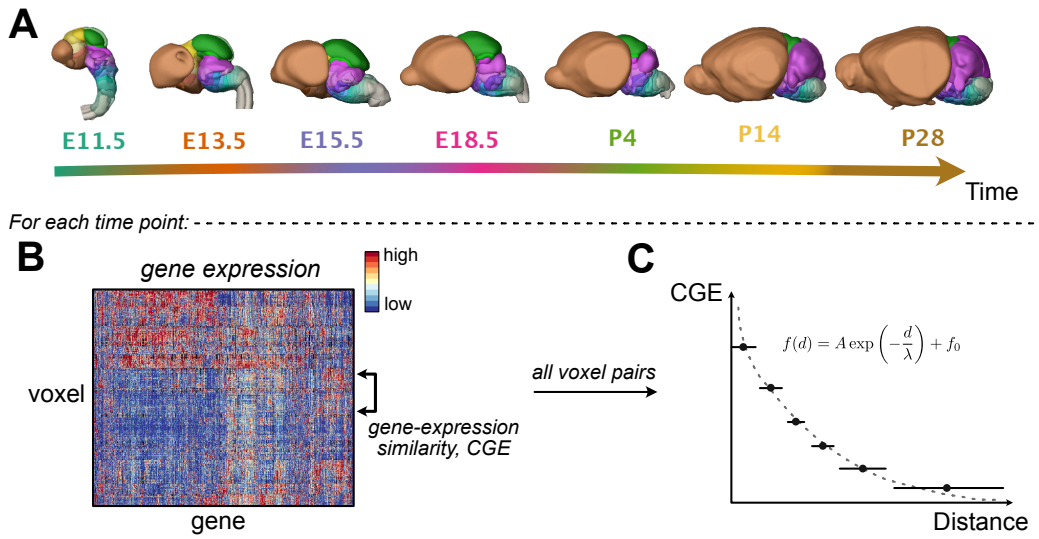

Fig. 1. Schematic analysis pipeline. **A** We analyze gene-expression atlases across seven developmental time points. **B** At each time point, we created a voxel × gene expression matrix (data shown for E11.5). We then compared the correlated gene expression (CGE) between each pair of voxels as a function of their physical separation distance, d. **C** Plotting $\mathrm{CGE}(d)$ across equiprobable distance bins allowed us to analyze the spatial embedding of gene expression at each time point. We fitted an exponential form $\mathrm{CGE}(d) = A\exp(-d/\lambda) + f_0$. Here we focus on the correlation length, λ, and how it scales with brain development.

$d/d_{\max}$, we defined $d_{\max}$ as the maximum extent along the anterior–posterior axis of the brain. For computational efficiency, we restricted our analysis to a random subset of 1000 voxels at each developmental time point. Our results do not strongly depend on this spatial downsampling—the variance in our statistical estimate of λ across 100 random subsamples saturated well before the level of 1000 voxels (and was less than 0.01 at 1000 voxels), as shown in Fig. S1.

Note that our definition of λ as a length scale (m) is consistent with its conventional use to denote wavelength in physics (units of length), but differs by an inverse relationship from its use as a decay rate with units of inverse length (m^{-1}) used to characterize the exponential distance rule for structural connectivity (9, 13–15).

Modeling. To better understand patterns found in the empirical data, we simulated simple spatial models of gene expression patterning and brain growth. At each developmental time point, we first simulated gene-expression patterns for each of the 1861 genes. We first defined the geometry by approximating the brain as a rectangular prism with dimensions (anterior–posterior × superior–inferior × left–right) equal to that of the developmental reference atlas at that time point: $3.52 \times 4.72 \times 1.20\,\mathrm{mm}$ (E11.5), $5.70 \times 3.30 \times 1.20\,\mathrm{mm}$ (E13.5), $6.60 \times 3.72 \times 1.68\,\mathrm{mm}$ (E15.5), $7.84 \times 4.06 \times 2.10\,\mathrm{mm}$ (E18.5), $10.56 \times 5.44 \times 3.04\,\mathrm{mm}$ (P4), $12.80 \times 6.60 \times 4.20\,\mathrm{mm}$ (P14), and $13.60 \times 7.40 \times 4.20\,\mathrm{mm}$ (P28). We formed a grid over this idealized geometry by dividing each dimension into 50 equally spaced points. We then generated a random spatially autocorrelated map independently for each of 1861 genes using a spatial-lag model (29, 38, 39). The model defines a dependence between samples using an exponential kernel as $W_{ij} = \exp(-d_{ij}/d_0)$, for pairwise distances, d_{ij}, and a characteristic spatial scale, d_0. From this weighting matrix, W_{ij}, an autocorrelated spatial map can be generated as $x_i = (I + \rho W_{ij})u_j$, where I is the identity and $u_j \sim \mathcal{N}(0,1)$ is i.i.d. Gaussian noise. This defines the second model parameter: the spatial autocorrelation strength, ρ. By setting d_0 as a fixed fraction of brain size, $d_0 = d_{\max}/d_0^{\mathrm{scale}}$, we simulated a brain-size scaling rule consistent with a linear rescaling of space (spatial expansion through development). Values for ρ and d_0^{scale} were set through a gradient descent optimization procedure to best fit the empirical values of λ and A, yielding $\rho = 0.16$ and $d_0^{\mathrm{scale}} = 8.41$. Some illustrative examples of the expression patterns generated are plotted in Fig. 5A.

Note that at small distances, all grid points are included in the CGE calculation. But due to the grid's finite size, beyond a critical distance, only a subset of grid points are included in the CGE calculation (e.g., in the extreme case of the maximum distance, only the corners of the cube are included). This finite-size spatial sampling effect can bias the $\mathrm{CGE}(d)$ curve at large distances. For example, distance bins that are affected by this are labeled pink in Fig. 5B.

For each simulated gene-expression dataset, we performed the same processing and analysis methodology as was applied to the empirical data: (i) a random subsample of 1000 points was taken; (ii) voxel × gene expression data were normalized using a scaled sigmoid; (iii) CGE was computed; (iv) $\mathrm{CGE}(d)$ data were binned using 20 equiprobable bins; (v) exponential decay was fit. We repeated the process 50 times to estimate error bars on each parameter estimate under the sources of stochasticity in the model: the generation of spatially autocorrelated gene-expression maps, and the random spatial subsampling. We were then able to compare the parameter estimates from this model with the empirical data. Code for the modeling is available at https://github.com/NeuralSystemsAndSignals/DevelopmentalExpressionModeling.

Results

Our approach to investigating the spatial embedding of transcriptional gradients across mouse-brain development is shown schematically in Fig. 1. At each of the seven time points in the ADMBA (33), shown in Fig. 1A, we obtained the expression of 1861 genes across voxels of the mouse brain (Fig. 1B). We then computed the correlated gene expression (CGE) for each pair of voxels as the Pearson correlation coefficient between their expression signatures (rows of the matrix in Fig. 1B). We fit an exponential function to the decay of CGE(d) as a function of separation distance, d, shown in Fig. 1C, to quantify how transcriptional similarity is spatially embedded. Variation in the parameters of this fit, particularly that of the characteristic length scale over which genes are correlated, λ, tell us how the spatial embedding of gene transcription varies through mouse-brain development.

Spatial embedding of gene transcription through development. Binned CGE(d) data with exponential fits are shown for each of the seven developmental time points in Figs 2A–G. The exponential form captured the data well at all time points (all adjusted $R^2 > 0.94$). Deviations from the exponential form are seen as a 'hump' at moderate relative distances which is most clearly seen at E15.5–P14.

To better understand how CGE varies as a function of brain size through development, we plotted the fitted CGE(d) exponential curves at all time points in Fig. 3A. The plot reveals an increase in the correlation length of CGE through development. When distances are proportionally rescaled by brain size, $d_{\max}$ (maximal anterior–posterior distance), as $\hat{d} = d/d_{\max}$, shown in Fig. 3B, the curves approximately collapse to a common scale of spatial correlation. This suggests that the correlation length scale may be approximately conserved with a characteristic rescaling of brain size.

To quantify this relationship, we plotted the fitted transcriptional correlation length, λ, as a function of brain size, $d_{\max}$, in Fig. 4A. We find that λ increases with $d_{\max}$ through development (Pearson's $r = 0.98$, $p = 10^{-4}$). Although there are only seven points across only $\sim$ a decade of brain sizes, $d_{\max}$ (preventing rigorous inference of power-law scaling (40)), we have physical reasons to expect a scaling relationship across brain expansion (15). To explore this possibility, we fitted a power-law, $\lambda = 0.0632 \times d_{\max}^{1.45}$, shown in Fig. 4B. We then tested whether this relationship could give an accurate prediction of the correlation length of transcriptional coupling in human cortex, which has $d_{\max} \approx 148\,\mathrm{mm}$ (maximal anterior–posterior distance in the Glasser parcellation (41)) (21). The power-law function fitted to mouse development predicted $\lambda^{\mathrm{pred}}_{\mathrm{human}} = 86.9\,\mathrm{mm}$. Despite differences in data type (ISH for < 2000 genes in mouse, microarray for $> 20\,000$ genes in human), spatial discretization (at the level of voxels in mouse and a multimodal parcellation in human (41)), and anatomical extent (whole brain in mouse versus the left cortical hemisphere in human), our prediction is close to the measured value in human, $\lambda^{\mathrm{meas}}_{\mathrm{human}} = 61.4\,\mathrm{mm}$ (95% confidence interval: 54.1–71.0 mm (21)). The extrapolation of the relationship fitted to the mouse-development data along with the measured value, $\lambda^{\mathrm{meas}}_{\mathrm{human}}$, is shown in Fig. 4C. The predictions from mouse development give a slight overestimate, but reasonable prediction, $\lambda^{\mathrm{pred}}_{\mathrm{human}} = 86.9\,\mathrm{mm}$, for the correlation length of human cortical transcriptional gradients. This overestimation, $\lambda^{\mathrm{pred}}_{\mathrm{human}} > \lambda^{\mathrm{meas}}_{\mathrm{human}}$, suggests that, for its size, transcriptional patterns in the human cortex have a shorter correlation distance than in the mouse brain. This is consistent with a greater degree of molecular diversity (and functional specialization) in the human cortex compared to the mouse brain. It will be important for future work to verify the results of this extrapolation using more comprehensive spatiotemporal transcriptomic data across species.

The spatial embedding strength, A, controls the strength of the exponential relationship in CGE relative to other effects such as noise. As shown in Fig. 4D, this strength is similar across development, $A \approx 0.6$ (Pearson's $r = 0.69$, $p = 0.08$). The offset, f_0, corresponds to the CGE in the limit of large distances, d. As plotted in Fig. 4E, f_0 decreases with brain size (Pearson's $r = -0.80$, $p = 0.03$), approaching zero as the brain approaches adult size (P14 and P28). This is consistent with a base level of transcriptional coupling at the longest spatial scales at earlier time points (where the most spatially distant areas have positive CGE), compared a more transcriptionally diverse brain at P14 and P28 (where average CGE ≈ 0 at the longest spatial scales).

Model of brain growth. To understand potential mechanisms underlying the scaling of transcriptional gradients through brain growth, we developed a simple spatial model of transcriptional patterning and tested its predictions against the empirical data characterized above. As described in Methods, we generated a random spatially autocorrelated map independently for each gene's expression pattern, evaluated in an idealized prism geometry that matches the brain's three-dimensional spatial extent at each developmental time point. To test whether the observed patterns are consistent with a simple rescaling of space in these expression maps, we rescaled the spatial autocorrelation length, d_0, as a function of brain size, $d_{\max}$. The model is therefore similar to uniform, isotropic stretching of three-dimensional gene-expression patterns that exist at the earliest developmental time point (it reduces to a model of this case when all brains have the same relative dimensions). The types of spatially autocorrelated patterns that were generated are illustrated in Fig. 5A. We first verified that our model reproduces an approximately exponential dependence of CGE with distance at a given time point,

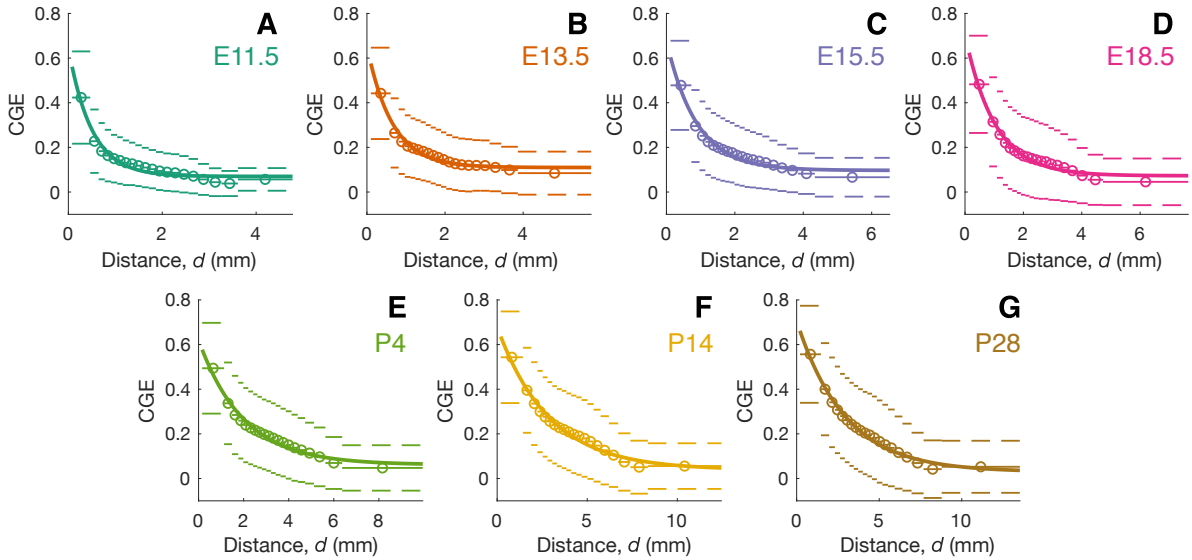

Fig. 2. Correlated gene expression (CGE) exhibits an approximately exponential decay with distance at all developmental time points. A–G: Across twenty equiprobable distance bins, mean CGE of all pairs of voxels in each bin is plotted as circles, with horizontal lines showing the extent of each bin. Dashed lines indicate one standard deviation either side of the mean of each bin. Fitted exponential curves, $\mathrm{CGE}(d) = A\exp(-d/\lambda) + f_0$, are shown solid.

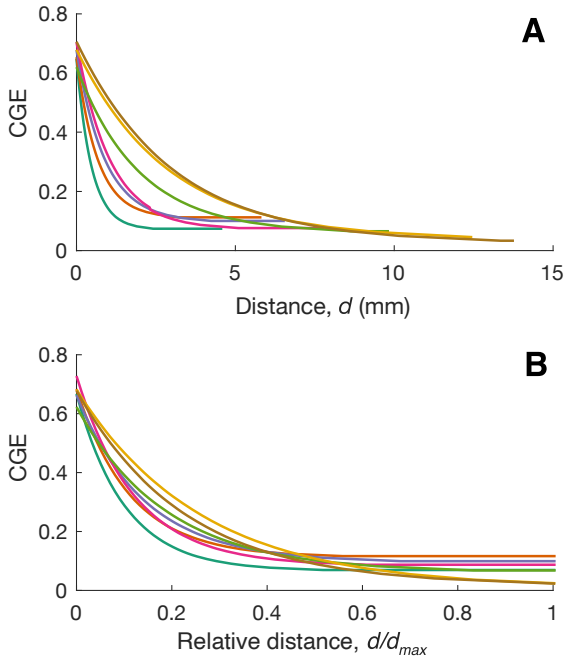

Fig. 3. Rescaling distances by a linear measure of brain size, $d_{\max}$, reveals a similarity in the spatial embedding of correlated gene expression (CGE). Fitted exponential curves are plotted for all seven time points (colored as in Fig. 2) for: **A** Distance, d (mm), and **B** Relative distance, $d/d_{\max}$. The measure of brain size, $d_{\max}$, is taken as the maximum extent along the anterior–posterior axis.

shown in Fig. 5B for E11.5. Apart from a slight decrease of $\mathrm{CGE} < 0$ at moderate distances, and an increase back to $\mathrm{CGE} \approx 0$ at the highest distances (especially where finite-size effects appear, labeled pink, see Methods), we find a good fit to an exponential form. This suggests that the exponential distance dependence of CGE observed in the ADMBA is compatible with an ensemble of genes with independent random, spatially autocorrelated transcriptional patterns.

After fitting the two parameters of our model, ρ and d_0^{scale}, to the empirical data (see Methods), the model approximately reproduces the scaling of the correlation length, λ, with brain size, $d_{\max}$, seen in empirical data (Fig. 5C), and the strength, A (Fig. 5D). The model did not reproduce the positive offsets, f_0, corresponding to the CGE at the longest separation distances, but instead predicted negative offsets, $f_0 < 0$ (Fig. 5E, see also Fig. 5B). This may be in part due to the idealized prism brain geometry used in the simulation, and the lack of anatomical specificity in the simulated expression gradients.

Discussion

In this work we used data from the ADMBA to characterize how gene transcriptional gradients are spatially embedded across brain development. Complementing the well-known exponential distance rules for structural connectivity in mammals (8, 9, 13, 14) that are connected by a scaling rule (15), we show that correlated gene expression is also well characterized by an exponential distance rule at each point in mouse-brain development. Furthermore, the characteristic correlation length, λ, of this embedding was found to scale with a linear dimension of brain size, $d_{\max}$, as $\lambda \sim d_{\max}^{1.45}$. More functionally specialized brains may be expected to encode more molecularly distinctive brain areas within the constraints of its brain size, corresponding to a shorter relative λ. We may therefore interpret how λ scales with brain size as measure of the spatial trade-off between local molecular coordination (uniform gene expression within a specialized brain area) and longer-range molecular differentiation (between molecularly distinct, functionally specialized areas). The scaling of λ seen across mouse-brain development provided a slight overesti-

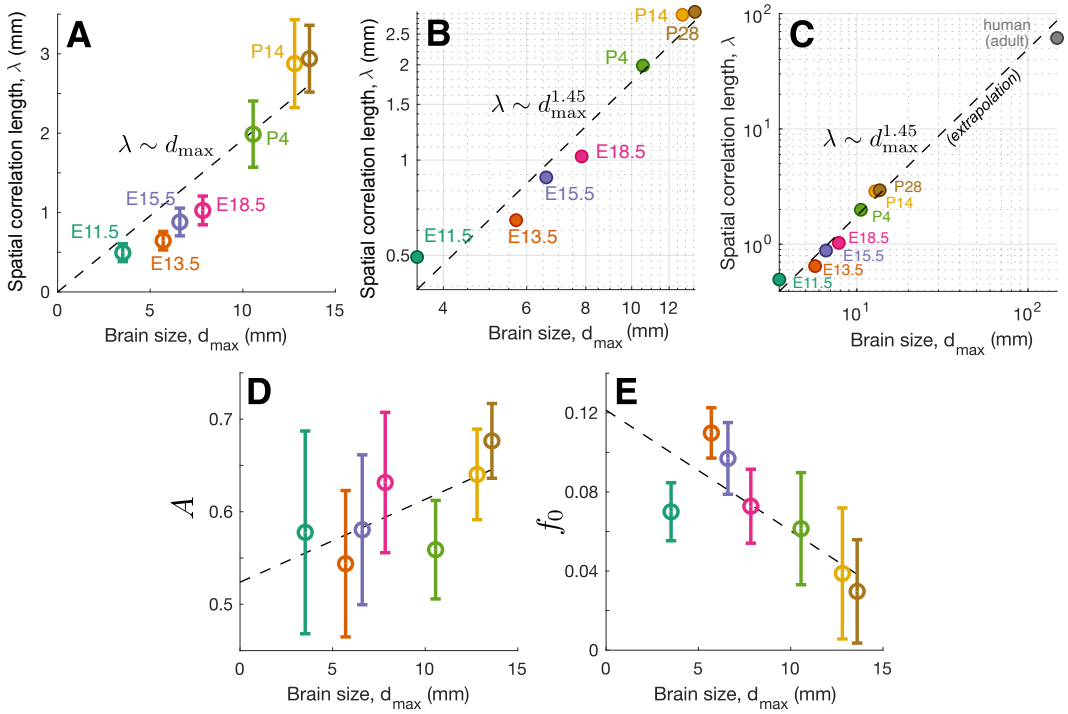

Fig. 4. The correlation length of transcriptional coupling, λ, increases with brain size. The three fitted parameters of the exponential form $\mathrm{CGE}(d) = A\exp(-d/\lambda) + f_0$, vary through brain development, plotted here as a function of brain size, $d_{\max}$ (computed along the anterior–posterior dimension): **A** transcriptional correlation length, λ (mm) (with linear fit); **B** $d_{\max}$–λ as in A but plotted on logarithmic axes with the best power-law fit, $\lambda = 0.0632 d^{1.45}$, annotated; **C** As in B but extrapolated out to the human cortex; **D** spatial embedding strength, A (with linear fit); and **E** offset, f_0 (with linear fit). Each data point represents the fitted parameter value from a given time point (labeled in **A**–**C**). Error bars as 95% confidence intervals are included in linear-axis plots, A, D, and E.

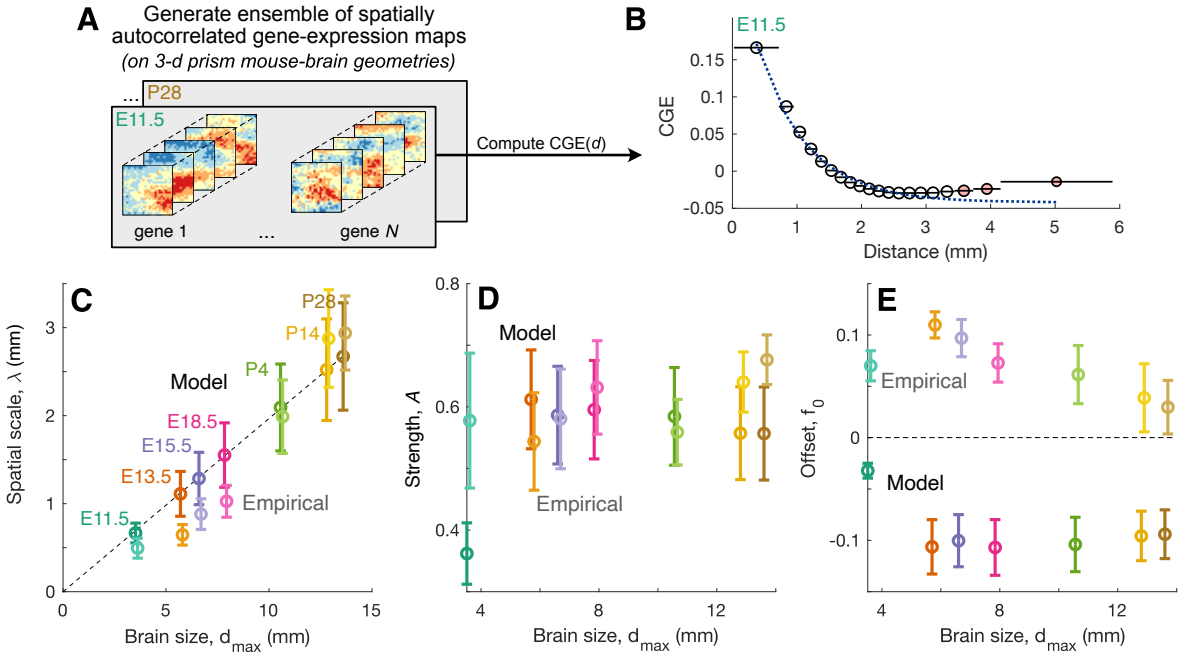

Fig. 5. A simple spatial model for gene expression through development reproduces key features of the empirical data. A Illustrative ensemble of independent, spatially autocorrelated gene-expression maps evaluated on three-dimensional prisms matching the dimensions of the brain at each point in development. Examples are shown for two genes in a three-dimensional volume (expression levels shown as color). **B** Each ensemble exhibits an approximately exponential decay of correlated gene expression (CGE) with distance, d, as $\mathrm{CGE}(d) = A\exp(-d/\lambda) + f_0$. Binned, model-simulated data (black) and fit (dotted blue) are plotted here for E11.5. Distance bins for which not all grid points were sampled (due to finite-size effects) are filled pink. Exponential fit parameters of the model-simulated data (shown with dark colors) and empirical data (shown with light colors) are plotted across development for: **C** spatial correlation length, λ; **D** strength, A; and **E** offset, f_0. To aid visibility, empirical values are slightly offset horizontally.

mate of the transcriptional correlation length in the adult human cortex, $\lambda^{\mathrm{pred}} = 86.9\,\mathrm{mm} > \lambda^{\mathrm{meas}} = 61.4\,\mathrm{mm}$ (Fig. 4C). The overestimation of λ_{human} when predicted from mouse data is consistent with the more functionally specialized human cortex exhibiting a greater level of molecular diversity (for its brain size) than mouse.

However, this difference could also be due to the limitations of performing such a wide extrapolation from a power-law function fitted to less than a decade of mouse brain sizes, or uncertainties related to mouse–human differences in the measurement of gene expression, number of genes compared, the discretization of space, and different anatomical scope (whole brain in mouse versus cortex in human). These differences notwithstanding, Given the recent finding that the brain-related genes exhibit strong consistency in their hierarchical patterning between mouse and human (31), it will be interesting for future work to further investigate these cross-species correspondences, including characterizing whether specific classes of genes drive similarities or differences in the spatial organization of gene expression across species.

Our analysis was performed at the voxel level using *in situ* hybridization data from the ADMBA. This allowed us to explore CGE(d) at a much finer spatial resolution than previous analyses of correlated gene expression, performed at the level of parcellations in adult mouse brain (7) and human cortex (21). Nevertheless, each spatial element in our analysis still contains transcriptional contributions from a large number of molecularly distinct brain cells. The advent of three-dimensional *in situ* single-cell transcriptome profiling (42–45), preserves spatial and transcriptomic information at the cellular level, allowing future analyses of correlated gene expression to be performed at even finer spatial scales. For example, this could allow testing of whether the bulk scaling rule for CGE characterized here applies at the finer scale of neuronal circuits. As transcriptomic atlas data become available for other species, such as *Drosophila melanogaster* (46, 47) and zebrafish (48), we will also be able to more comprehensively compare the spatial organization of the brain's molecular diversity, and assess the cross-species validity of the scaling rule proposed here.

A simple model of independent, spatially autocorrelated transcriptional maps that scale with brain size (similar to the isotropic expansion of transcriptional patterns set up early in development) reproduced key features of CGE(d) scaling observed empirically. We kept the spatial autocorrelation scale for our simulated gene-expression maps at a fixed proportion of brain size, consistent with linear, isotropic brain growth (but allowing us to use realistic brain dimensions from empirical measurements). Our results are thus consistent with molecular specialization occurring very early in development, with brain expansion stretching the correlation length through development. While our idealized model could reproduce some key elements of the data, there is much room for improvement, including the inability of our model to underestimate baseline levels of CGE (predicting CGE < 0 at moderate–high distances). The discrepancy may be in part due to the idealized prism brain geometry used in the simulation, or the simplistic lack of anatomical specificity in the random spatially autocorrelated expression gradients that we simulate. While our model reproduces much of the CGE(d) relationship, our model is only evaluated against these coarse-grained properties, having blurred the contribution of the many individual genes that display precise developmental spatiotemporal programs. It will be important for future work to precisely investigate and model these important properties of brain development that are not considered here. By characterizing a generic variation in CGE, the current work may have a role to play in this research effort. For example, genes that most strongly buck the generic trends in CGE(d) could be considered as candidates for playing a more specific, targeted role in shaping brain development, facilitating data-driven identification of genes that play important developmental roles.

ACKNOWLEDGEMENTS
The authors would like to thank Aurina Arnatkevičiūtė for assisting with the mouse–human comparison.

References

1. C. Cherniak. Neural component placement. *Trends Neurosci.* **18**, 522 (1995).
2. V. A. Klyachko and C. F. Stevens. Connectivity optimization and the positioning of cortical areas. *Proc. Natl. Acad. Sci. USA* **100**, 7937 (2003).
3. C. Cherniak, Z. Mokhtarzada, R. Rodriguez-Esteban, and K. Changizi. Global optimization of cerebral cortex layout. *Proc. Natl. Acad. Sci. USA* **101**, 1081 (2004).
4. B. L. Chen, D. H. Hall, and D. B. Chklovskii. Wiring optimization can relate neuronal structure and function. *Proc. Natl. Acad. Sci. USA* **103**, 4723 (2006).
5. D. S. Bassett, D. L. Greenfield, A. Meyer-Lindenberg, et al. Efficient Physical Embedding of Topologically Complex Information Processing Networks in Brains and Computer Circuits. *PLoS Comp. Biol.* **6**, e1000748 (2010).
6. A. Arnatkeviciūtė, B. D. Fulcher, R. Pocock, and A. Fornito. Hub connectivity, neuronal diversity, and gene expression in the Caenorhabditis elegans connectome. *PLoS Comp. Biol.* **14**, e1005989 (2018).
7. B. D. Fulcher and A. Fornito. A transcriptional signature of hub connectivity in the mouse connectome. *Proc. Natl. Acad. Sci. USA* **113**, 1435 (2016).
8. R. Gămănuț, H. Kennedy, Z. Toroczkai, et al. The Mouse Cortical Connectome, Characterized by an Ultra-Dense Cortical Graph, Maintains Specificity by Distinct Connectivity Profiles. *Neuron* **97**, 698 (2018).
9. H. R. Noori, J. Schöttler, M. Ercsey-Ravasz, et al. A multiscale cerebral neurochemical connectome of the rat brain. *PLoS Biol.* **15**, e2002612 (2017).
10. M. Kunst, E. Laurell, N. Mokayes, et al. A Cellular-Resolution Atlas of the Larval Zebrafish Brain. *Neuron* **103** (2019).
11. N. T. Markov, M. Ercsey-Ravasz, D. C. Van Essen, et al. Cortical High-Density Counterstream Architectures. *Science* **342**, 1238406 (2013).
12. J. A. Roberts, A. Perry, A. R. Lord, et al. The contribution of geometry to the human connectome. *NeuroImage* **124**, 379 (2016).
13. S. Horvát, R. Gămănuț, M. Ercsey-Ravasz, et al. Spatial Embedding and Wiring Cost Constrain the Functional Layout of the Cortical Network of Rodents and Primates. *PLoS Biol.* **14**, e1002512 (2016).
14. M. Ercsey-Ravasz, N. T. Markov, C. Lamy, et al. A Predictive Network Model of Cerebral Cortical Connectivity Based on a Distance Rule. *Neuron* **80**, 184 (2013).
15. P. Theodoni, P. Majka, D. H. Reser, et al. Structural attributes and principles of the neocortical connectome in the marmoset monkey. *bioRxiv* p. 2020.02.28.969824 (2020).
16. J. A. Henderson and P. A. Robinson. Geometric effects on complex network structure in the cortex. *Phys. Rev. Lett.* **107**, 018102 (2011).
17. J. A. Henderson and P. A. Robinson. Relations Between the Geometry of Cortical Gyrification and White-Matter Network Architecture. *Brain Conn.* **4**, 112 (2014).
18. J. Stiso and D. S. Bassett. Spatial Embedding Imposes Constraints on Neuronal Network Architectures. *Trends Cog. Sci.* (2018).
19. P. A. Robinson. Physical brain connectomics. *Phys. Rev. E* **99**, 012421 (2019).
20. L. French and P. Pavlidis. Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comp. Biol.* **7**, e1001049 (2011).
21. A. Arnatkeviciūtė, B. D. Fulcher, and A. Fornito. A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage* **189**, 353 (2019).
22. F. M. Krienen, B. T. T. Yeo, T. Ge, R. L. Buckner, and C. C. Sherwood. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc. Natl. Acad. Sci. USA* **113**, E469 (2016).
23. M. J. Hawrylycz, A. Bernard, C. Lau, et al. Areal and laminar differentiation in the mouse neocortex using large scale gene expression data. *Methods* **50**, 113 (2010).
24. S. P. Pantazatos and X. Li. Commentary: BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. Science 348, 1241-4. *Front. Neurosci.* **11**, 412 (2017).
25. J. Richiardi, A. Altmann, and M. Greicius. Distance Is Not Everything In Imaging Genomics Of Functional Networks: Reply To A Commentary On Correlated Gene Expression Supports Synchronous Activity In Brain Networks. *bioRxiv* p. 132746 (2017).
26. A. Fornito, A. Arnatkeviciūtė, and B. D. Fulcher. Bridging the Gap between Connectome and Transcriptome. *Trends Cog. Sci.* **23**, 34 (2019).

27. D. S. Margulies, S. S. Ghosh, A. Goulas, et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. USA* **113**, 12574 (2016).
28. J. M. Huntenburg, P.-L. Bazin, P. L. Bazin, and D. S. Margulies. Large-Scale Gradients in Human Cortical Organization. *Trends Cog. Sci.* **22**, 21 (2017).
29. J. B. Burt, M. Demirtas, W. J. Eckner, et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **27**, 889 (2018).
30. C. Paquola, R. V. De Wael, K. Wagstyl, et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS Biol.* **17**, e3000284 (2019).
31. B. D. Fulcher, J. D. Murray, V. Zerbi, and X.-J. Wang. Multimodal gradients across mouse cortex. *Proc. Natl. Acad. Sci. USA* **116**, 4689 (2019).
32. L. C. Greig, M. B. Woodworth, M. J. Galazo, H. Padmanabhan, and J. D. Macklis. Molecular logic of neocortical projection neuron specification, development and diversity. *Nature Reviews Neuroscience* **14**, 755 (2013).
33. C. L. Thompson, L. Ng, V. Menon, et al. A High-Resolution Spatiotemporal Atlas of Gene Expression of the Developing Mouse Brain. *Neuron* **83**, 309 (2014).
34. A. A. Dillman, D. N. Hauser, J. R. Gibbs, et al. mrna expression, splicing and editing in the embryonic and adult mouse cerebral cortex. *Nat. Neurosci.* **16**, 499 (2013).
35. S. Fertuzinhos, M. Li, Y. I. Kawasawa, et al. Laminar and temporal expression dynamics of coding and noncoding rnas in the mouse neocortex. *Cell Rep.* **6**, 938 (2014).
36. M. Mody, Y. Cao, Z. Cui, et al. Genome-wide gene expression profiles of the developing mouse hippocampus. *Proc. Natl. Acad. Sci. USA* **98**, 8862 (2001).
37. C.-K. Lee, S. M. Sunkin, C. Kuan, et al. Quantitative methods for genome-scale analysis of in situ hybridization and correlation with microarray data. *Genome Biol.* **9**, R23 (2008).
38. L. Anselin. *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media (2013).
39. J. B. Burt, M. Helmer, M. Shinn, A. Anticevic, and J. D. Murray. Generative modeling of brain maps with spatial autocorrelation. *bioRxiv* (2020).
40. A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review* **51**, 661 (2009).
41. M. F. Glasser, T. S. Coalson, E. C. Robinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171 (2016).
42. S. Codeluppi, L. E. Borm, A. Zeisel, et al. Spatial organization of the somatosensory cortex revealed by osmfish. *Nat. Methods* **15**, 932 (2018).
43. J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
44. X. Wang, W. E. Allen, M. A. Wright, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
45. C. Xia, J. Fan, G. Emanuel, J. Hao, and X. Zhuang. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. USA* **116**, 19490 (2019).
46. C. A. Brunet, R. Bruggmann, and S. Sprecher. Single cell transcriptome atlas of the drosophila larval brain. *eLife* **8** (2019).
47. K. Davie, J. Janssens, D. Koldere, et al. A single-cell transcriptome atlas of the aging drosophila brain. *Cell* **174**, 982 (2018).
48. B. Raj, D. E. Wagner, A. McKenna, et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotech.* **36**, 442 (2018).

Supplementary Information

Robustness to voxel resampling. This number was chosen based on computation of variance in exponential decay constant (see later). Briefly, we sampled different numbers of voxels (100 to 1000 in increments of 100), each 100 times, to calculate the inverse decay constant, λ^{-1}, and then determined the variance of λ^{-1} for each sample size. We found that variance in λ^{-1} plateaued at 1000 voxels (Fig. S1), indicating that 1000 was a sufficiently large sample size to obtain reliable λ^{-1} estimates.

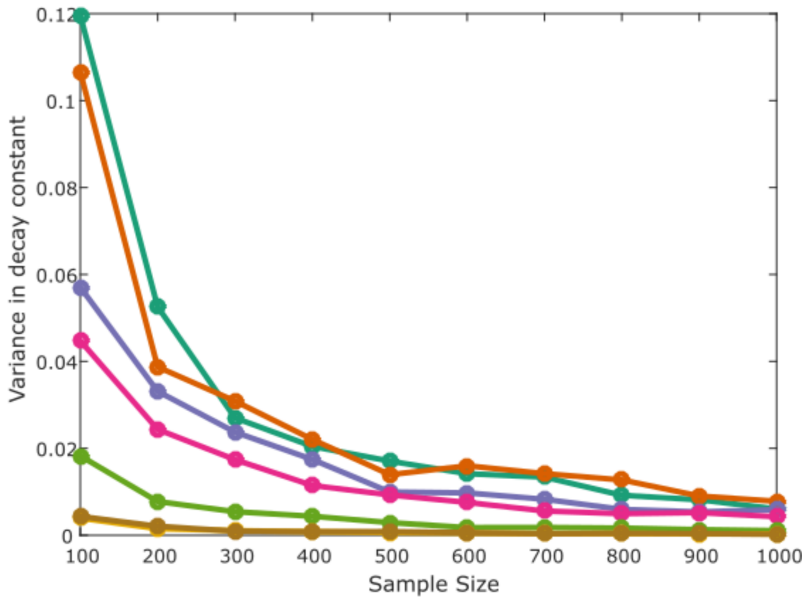

Fig. S1. Variance in λ^{-1} estimates due to random voxel sampling. Variance was estimated across 100 trials each of different voxel sample sizes throughout development. Curves are plotted for each developmental time point using colors from the main text (e.g., Fig. 2). The variance in λ^{-1} estimates across random samples plateaus to a low value < 0.01 by the 1000 voxels used here.