nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-021-01082-z

Check for updates

Mapping gene transcription and neurocognition across human neocortex

Justine Y. Hansen[1], Ross D. Markello[1], Jacob W. Vogel[2], Jakob Seidlitz[2,3], Danilo Bzdok[1,4,5] and Bratislav Misic[1]✉

Regulation of gene expression drives protein interactions that govern synaptic wiring and neuronal activity. The resulting coordinated activity among neuronal populations supports complex psychological processes, yet how gene expression shapes cognition and emotion remains unknown. Here, we directly bridge the microscale and macroscale by mapping gene expression patterns to functional activation patterns across the cortical sheet. Applying unsupervised learning to the Allen Human Brain Atlas and Neurosynth databases, we identify a ventromedial–dorsolateral gradient of gene assemblies that separate affective and perceptual domains. This topographic molecular-psychological signature reflects the hierarchical organization of the neocortex, including systematic variations in cell type, myeloarchitecture, laminar differentiation and intrinsic network affiliation. In addition, this molecular-psychological signature strengthens over neurodevelopment and can be replicated in two independent repositories. Collectively, our results reveal spatially covarying transcriptomic and cognitive architectures, highlighting the influence that molecular mechanisms exert on psychological processes.

The human brain is an integrated system, involving interactions across multiple scales[1]. At the molecular level, fluctuations in gene expression and protein synthesis in neurons drive single-cell activity[2,3]. The waxing and waning of cellular activity promotes synaptic remodelling[4,5], shaping the wiring of nested and increasingly polyfunctional neural circuits[6,7]. Anatomical connections among mesoscopic neuronal populations promote functional interactions[8], manifesting as patterned neural activity that drives psychological processes[9,10]. The regulation of gene expression is therefore naturally intertwined with the brain's structure and function[11–15]. How molecular dynamics map onto mental states remains a key question in neuroscience.

Modern technological and analytical advances, in concert with global data-sharing initiatives, have created fundamentally new opportunities to link molecular dynamics and psychological processes. High-resolution functional neuroimaging has informed comprehensive meta-analytical atlases of how brain areas selectively respond over a spectrum of perceptual, cognitive and affective experimental manipulations[16–19]. At the same time, high-throughput microarray profiling has yielded precise genome-wide maps of transcript distributions over the brain[20–22], allowing inferences about the spatial distribution of cellular processes and types[11–13,23–27]. Altogether, the concurrent emergence of global functional genomic and brain-mapping initiatives offers an unprecedented chance to identify spatial correspondences between the brain's genetic and cognitive architectures.

Here, we directly relate microscale molecular processes to the macroscale functional architecture of the human brain. We apply partial least-squares (PLS) analysis to gene expression maps (Allen Human Brain Atlas (AHBA)[20]) and probabilistic functional association maps (Neurosynth[16]) to identify molecular signatures related to psychological processes (for a conceptually similar approach, see ref. [28]). We reveal distinct sets of functionally interrelated genes that span a gradient of psychological processes, and show that this molecular signature corresponds to systematic variation in cell type composition, microstructure and large-scale functional system affiliation. Finally, we perform extensive cross-validation, sensitivity testing and robustness analysis using two independent datasets (BrainMap[17] and BrainSpan[22]).

Results

To establish a relationship between gene expression and functional activity, we used the AHBA for estimates of gene expression in the brain[20] and Neurosynth for probabilistic measures of whether specific terms (such as attention, emotion and sleep) are functionally related to specific brain regions[16]. This probability, which we call functional association, describes how often specific terms and voxel coordinates are published in conjunction with one another. To facilitate comparison with other reports, only genes with a differential stability greater than 0.1 were retained for analysis (see Methods[21,26]) and the term set was restricted to those in the intersection of terms reported in Neurosynth and in the Cognitive Atlas[29]. Gene expression data and probabilistic measures were parcellated into 111 left hemisphere cortical regions of interest[30,31]. The resulting gene expression matrix was composed of normalized expression levels of 8,825 stable genes across 111 target brain regions[31], and the functional association matrix represented the functional relatedness of 123 terms to the same 111 brain regions (see Supplementary Table 1 for a full list of terms).

Molecular signatures of psychological processes. We related gene expression to functional associations using PLS analysis, a multivariate statistical technique that extracts optimally covarying patterns from two data domains[32–34] (Fig. 1a; for results when using canonical correlation analysis (CCA) instead, see Supplementary Fig. 1). PLS analysis revealed a single statistically significant latent

[1]McConnell Brain Imaging Centre, Montréal Neurological Institute, McGill University, Montréal, Québec, Canada. [2]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [4]Biological and Biomedical Engineering, McGill University, Montréal, Québec, Canada. [5]Mila, Quebec Artificial Intelligence Institute, Montréal, Québec, Canada. ✉e-mail: bratislav.misic@mcgill.ca

variable relating gene expression to corresponding functional association across the brain ($P_{spin}=0.0228$; bootstrap-estimated confidence interval (CI) = (50, 70); one tailed), where significance was assessed using a permutation test that preserves spatial autocorrelation (spin test)[35,36]. The spin test embodies the null hypothesis that genes and terms are spatially correlated with each other because of inherent spatial autocorrelation. Note that the *P* value does not index the contribution of individual terms or genes to the multivariate pattern; this is assessed by bootstrap resampling (see Methods). This latent variable represents a pattern of gene expression (gene weights) and a pattern of functional association (term weights) that together capture 65% of the covariance between gene expression and functional association (Fig. 1d; for the PCA decompositions of the individual data matrices, see Supplementary Fig. 2). Projecting the gene expression and functional association matrices back onto the gene weights and term weights, respectively, reflects how well a brain area exhibits the gene and term pattern, which we refer to as gene scores and term scores (Fig. 1b,e). These projections also indicate covariation patterns of gene expression and functional associations. Genes and terms whose activity correlates positively with the score pattern covary with one another in the positively scored regions, and vice versa for negatively scored regions. The pattern of gene and term scores across the brain revealed a dorsolateral to ventromedial gradient, in which dorsolateral regions were scored more negatively and ventromedial regions more positively.

Next, we cross-validated the correlation between gene and term scores. Due to inherent spatial autocorrelation, proximal regions exhibit similar gene expression profiles and functional activity[37,38]. Thus, randomly dividing brain regions into training and testing sets may result in interdependencies between the two sets (Supplementary Fig. 3). To ensure that the correlation between gene and term scores is not inflated due to spatial autocorrelation, we selected the 75% of brain regions closest in Euclidean distance to a randomly chosen source node as the training set, and the remaining 25% of brain regions as the testing set. This procedure was repeated 100 times and distributions of Pearson's correlations for the training and testing sets are shown in Fig. 1f. The mean out-of-sample Pearson's correlation between gene and term scores was $r(26)=0.4770$ ($P_{spin}=0.0420$; bootstrap-estimated 95% CI = (0.43, 0.52)).

A genomic signature of cognitive processes. The significance of the first latent variable and the cross-validation of score correlations demonstrates there is a relationship between gene expression and functional association. The relationship itself is determined by the terms and genes that contribute most to the latent variable. The loading of each term was computed as the Pearson's correlation between the term's functional association across brain regions and the PLS analysis-estimated scores (Fig. 1c). Loadings reflect the shared variance between original variables and the PLS score pattern and can therefore be interpreted as indexing the degree of contribution of each variable to the PLS analysis-defined latent variable. The 25% most positively and negatively correlated terms were retained as terms that most contribute to the latent variable (Fig. 2a; for the loadings of all reliable terms, see Supplementary Fig. 4). Terms with large positive loadings were related to affective processes, including emotion, stress, fear, anxiety and mood. Terms with large negative loadings were related to perceptual and attentional processes. Examples include attention (of which visual attention, spatial attention and simply attention were all weighed very highly), visual perception and imagery. This latent variable thus represents a putative perceptual–affective gradient of functional associations. Note that while negatively scored brain regions are dominant in sensorimotor regions, terms related to attention and perception consistently score higher than sensorimotor terms such as 'multisensory' and 'coordination'. Moreover, although the spatial pattern involves prominent contributions from both the visual and somatomotor cortex, the ranking of Neurosynth terms emphasizes psychological processes associated with simultaneous activation in both; hence, terms associated with perception and attention are ranked higher than terms associated with sensorimotor function per se.

Gene contribution was analogously assessed by computing spatial correlations (loadings) between an individual gene's expression pattern and the PLS analysis-derived scores (Fig. 1c). Genes were segregated based on the sign of their loadings, where positive genes refer to genes with positive loading and negative genes refer to those with negative loading. Strongly contributing genes were defined as those among the top 50% of their respective positive/negative gene set, resulting in 2,544 strongly contributing positive genes and 1,869 strongly contributing negative genes. The pattern of positive genes covarying with affective terms is strongest in positively scored brain regions (Fig. 2a) and the pattern of negative genes covarying with perceptual terms is strongest in negatively scored brain regions (Fig. 2b).

To better understand the biological significance of the positive and negative gene sets, we adapted analyses from the gene set enrichment analysis toolbox (https://github.com/benfulcher/GeneSetEnrichmentAnalysis)[38]. We explored the biological processes with which the reliable positive and negative genes are significantly involved (see Methods for details and Supplementary Tables 4 and 5 for a full machine-readable list of biological processes and respective *P* values). A selection of the significant categories most related to brain structure and function are visualized as word clouds in Fig. 2b. In general, affect-related gene sets show enrichment for processes related to neurogenesis and differentiation, while perception-related gene sets are enriched for processes related to synaptic signalling.

Alongside biological process, we asked whether psychologically relevant genes are preferentially expressed in specific cell types (Fig. 2c). Cell type deconvolution was performed using cell-specific aggregate gene sets across five human adult postmortem single-cell and single-nucleus RNA sequencing studies[39–44], as presented previously (Supplementary Table 5 from ref. [23]). Specifically, we calculated the ratio of genes in each gene set that are preferentially expressed in one of seven cell types: astrocytes, microglia, oligodendrocyte precursors, oligodendrocytes, endothelial cells, excitatory neurons and inhibitory neurons (Fig. 2c). Gene sets were thresholded to include the top 50% of genes with the greatest loadings (note that although the threshold is arbitrary, the results are highly consistent across a range of thresholds, from 2.5% to no threshold; Supplementary Fig. 6). Statistical significance was assessed against a null distribution of ratios constructed by repeating the process 10,000 times on a set of random genes (two tailed; false discovery rate corrected). Dominant positive genes (related to affect) are significantly more expressed in astrocytes ($P=2.3\times10^{-4}$), microglia ($P=2.3\times10^{-4}$) and oligodendrocyte precursors ($P=0.0160$) and significantly less expressed in excitatory neurons ($P=0.0052$), oligodendrocytes ($P=2.3\times10^{-4}$) and endothelial cells ($P=0.0052$). Dominant negative genes (related to perception) are significantly more expressed in excitatory neurons ($P=0.0023$) and inhibitory neurons ($P=0.0017$) and significantly less expressed in astrocytes ($P<0.001$), microglia ($P=0.0114$), oligodendrocytes ($P=0.0019$) and oligodendrocyte precursors ($P=0.0114$). Broadly, we find evidence that areas associated with affect are enriched for genetic signal of cells involved in neuron support (astrocytes and microglia); areas associated with perception are enriched for genetic signal of neurons themselves (inhibitory and excitatory). This dichotomy also matches the intuition derived from biological process enrichment analysis (Fig. 2b).

The gene–cognition gradient reflects cortical hierarchies. Having identified a gradient of covarying gene expression and functional association, we next investigated whether these topographic patterns reflect variation in other microstructural and functional attributes[45,46].

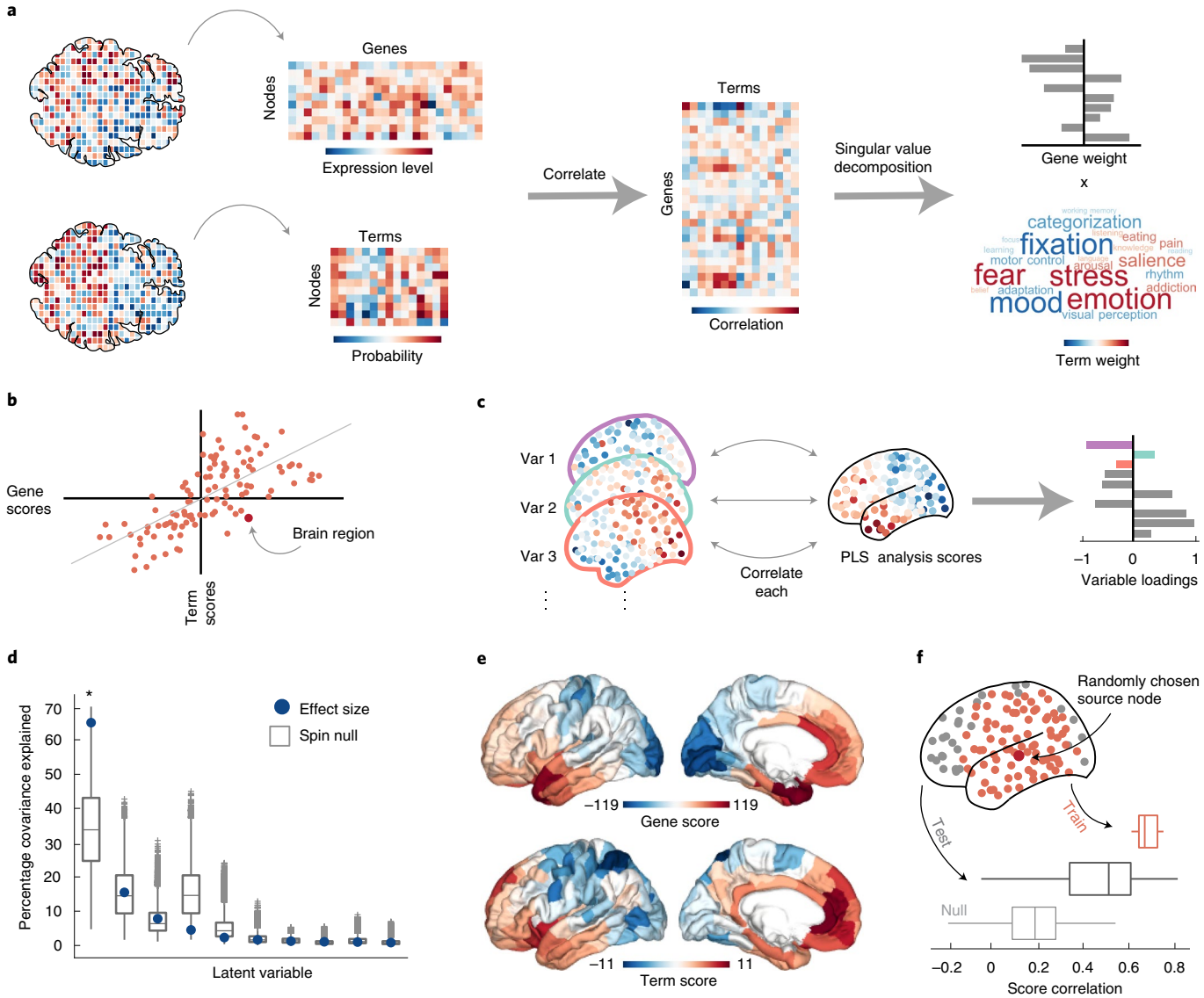

Fig. 1 | Relating gene expression to functional association. PLS analysis was used to identify spatially covarying patterns of gene expression (AHBA) and functional association (Neurosynth). **a**, PLS analysis relates two data domains by correlating the variables across brain regions and subjecting this to singular value decomposition. This results in multiple latent variables: linear weighted combinations of the original variables (gene weights and term weights) that maximally covary with each other. **b**, Gene (term) scores are defined as the product or projection of the original gene expression (functional association) matrix onto the gene (term) weights, such that each brain region is associated with a gene and term score. By design, gene and term scores correlate highly. **c**, Gene (term) loadings are defined as the Pearson's correlation between each gene's expression (term's functional associations) across brain regions and the PLS analysis-derived score pattern. Var, variable. **d**, Latent variables are ordered according to effect size (the proportion of covariance explained between gene expression and functional association they account for) and shown as blue dots. Statistical significance was assessed with respect to the spatial autocorrelation-preserving null model[76], shown in grey (10,000 repetitions). Only the first latent variable was statistically significant, accounting for 65% of the covariance between gene expression and functional associations ($P_{spin} = 0.0228$; bootstrap-estimated 95% CI = (50, 70)). **e**, Projecting the original data back onto the PLS analysis-defined gene/term weights results in gene/term scores for each brain region, indexing the extent to which a brain region expresses covarying gene/term patterns. **f**, The correlation between gene scores and term scores was cross-validated by constructing the training set with 75% of brain regions closest in Euclidean distance to a randomly chosen source node (red) and with the testing set as the remaining 25% of brain regions (dark grey; 100 repetitions). The out-of-sample mean was significant against a permuted null model ($r(26) = 0.4770$; $P_{spin} = 0.0420$; bootstrap-estimated 95% CI = (0.43, 0.52); 1,000 repetitions; null model shown in light grey). See Supplementary Fig. 3 for a comparison with completely random splits. The boxplots in **d** and **f** represent the first, second (median) and third quartiles, the whiskers represent the non-outlier end points of the distribution, and the crosses represent outliers.

To address this question, we averaged measures of cortical thickness and T1w/T2w ratios (a widely used proxy for intracortical myelin[47]) from the left hemisphere cortex across 417 unrelated participants from the Human Connectome Project (HCP; see Methods). We then computed Pearson's correlations of mean cortical thickness and T1w/T2w maps with gene score and term score maps (Fig. 3a). We found a strong positive correlation between cortical thickness and PLS analysis scores ($r(109) = 0.82$, $P_{spin} < 0.001$ and bootstrap-estimated 95% CI = (0.75, 0.88) for gene scores; $r(109) = 0.52$, $P_{spin} = 0.0426$ and bootstrap-estimated 95% CI = (0.40, 0.62) for term scores; two tailed)

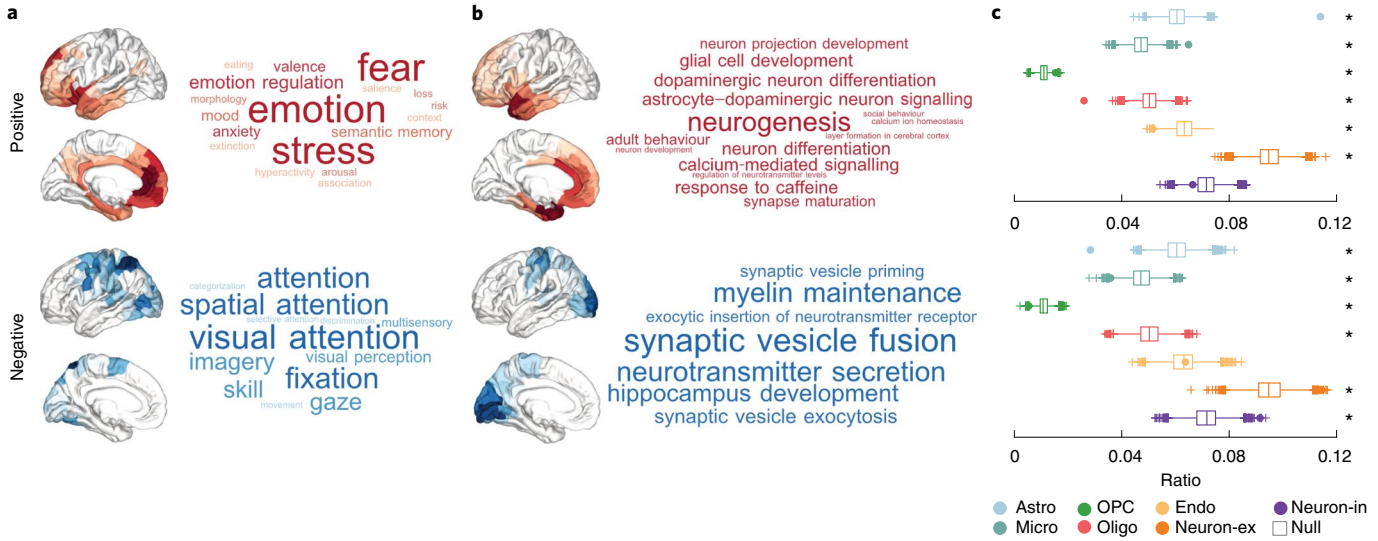

Fig. 2 | Gene sets underlying psychological processes. Genes and terms that contribute most to the latent variable were analysed further. **a**, The contributions of positively and negatively weighed ontological terms to the latent variable were estimated with term loadings: Pearson's correlations between a term's functional association across brain regions and the PLS analysis-derived score pattern. The 25% most positively (red) and negatively (blue) correlated terms reveal a perceptual–affective gradient. Word size represents the relative size of the term loading. **b**, Gene contribution was estimated with gene loadings: Pearson's correlations between a gene's expression across brain regions and the PLS analysis-derived score pattern. Biological processes in which the top 50% of genes with positive and negative loadings are most involved were identified using gene set enrichment analysis (see Methods) and tested against a spatial autocorrelation-preserving null model[38]. **c**, Cell-type deconvolution was used to identify cell-type enrichment in the gene sets identified by PLS analysis[23]. The ratio of genes in each gene set preferentially expressed in seven distinct cell types is shown against a null model of a random selection of all genes (boxplots; 10,000 repetitions). $^*P < 0.05$. The boxplots represent the first, second (median) and third quartiles, whiskers represent the non-outlier end points of the null distribution, and crosses represent outliers. Astro, astrocyte; Micro, microglia; OPC, oligodendrocyte precursor; Oligo, oligodendrocyte; Endo, endothelial; Neuro-ex, excitatory neurons; Neuro-in, inhibitory neurons; Null, empirically derived null distribution.

and a strong negative correlation between T1w/T2w ratio and PLS analysis scores ($r(109) = -0.86$, $P_{spin} < 0.001$ and bootstrap-estimated 95% CI = (−0.90, −0.81) for gene scores; $r(109) = -0.58$, $P_{spin} = 0.0039$ and bootstrap-estimated 95% CI = (−0.67, −0.46) for term scores; two tailed). Altogether, the gene association gradient mirrors microstructural attributes[13,26,48,49].

Given that the score pattern resembles the differentiation between unimodal and transmodal cortex[50,51], we sought to relate the score pattern to the principal functional gradient reported by Margulies and colleagues[52]. For this purpose, we applied diffusion map embedding on a group-averaged functional connectivity matrix computed from the 1,003 HCP participants with complete resting-state functional magnetic resonance imaging (fMRI) data and extracted the first principal gradient[53,54]. This gradient situates brain regions on a continuous axis from unimodal primary sensory and motor cortex to transmodal higher association cortex (Fig. 3a). We found that the gene score and term score patterns significantly correlated with this gradient ($r(109) = 0.55$, $P_{spin} = 0.0054$ and bootstrap-estimated 95% CI = (0.38, 0.65) for gene scores; $r(109) = 0.68$, $P_{spin} < 0.001$ and bootstrap-estimated 95% CI = (0.58, 0.75) for term scores; two tailed; Fig. 3a). This implies that negatively scored regions tend to be more closely aligned with unimodal cortex and positively scored regions tend to be predominantly aligned with transmodal cortex.

As a final step, we sought to understand how well the gene and term score maps conform to other major structural and functional partitions of the human cerebral cortex. We stratified gene and term scores in several complementary ways: (1) within seven intrinsic functional brain networks as defined by Yeo and colleagues[9]; (2) within seven Von Economo classes of cortical cytoarchitecture[15,55]; and (3) within four Mesulam levels of laminar differentiation across the cortex[56,57] (Fig. 3b–d). Consistent with the notion that the gene expression–functional association gradient reflects a differentiation between perceptual and affective psychological domains, we observed a separation between limbic/paralimbic and somatomotor/idiotypic networks across all three partitions.

Gene–cognition convergence over development. Given the continuous development of cognitive processes over the lifespan, we sought to track the gene expression–functional association signature through human development. We used BrainSpan, a dataset that provides gene expression estimates from brain tissue samples aged 8 post-conception weeks to 40 years, across 16 unique cortical regions[22]. To best replicate the original analyses, genes were selected if they were included in the original list of 20,503 genes from the AHBA and if their differential stability was greater than 0.1, as defined on the 34-node parcellation. A region by gene expression matrix was constructed for five different life stages: foetal, infant, child, adolescent and adult (see Methods for details). Gene scores were estimated as the projection of the BrainSpan-defined gene expression matrices onto the PLS analysis-defined gene weights, yielding an estimated gene score per region and life stage (Fig. 4a). The molecular signature, represented by gene scores, increases with development, suggesting that the gene expression–functional association gradient becomes more pronounced with maturation. In other words, the genetic signal captured by the original PLS analysis is specific to adult-derived cells.

We also used the BrainSpan dataset to assess the robustness of the original PLS model. For each life stage, we correlated the 16 estimated gene scores with the PLS analysis-defined term scores, as determined by the 34-node parcellation. To achieve this, we manually defined a region-to-region mapping, defining each of the 34 regions as a child of a region from the 16-node parcellation. We then averaged the term scores across sibling nodes to yield a single term score for each of the 16 BrainSpan-defined cortical regions (Fig. 4b). Estimated gene and term scores correlated significantly in the infant ($r(10) = 0.61$; $P = 0.0352$; bootstrap-estimated 95% CI = (0.03, 0.83)),

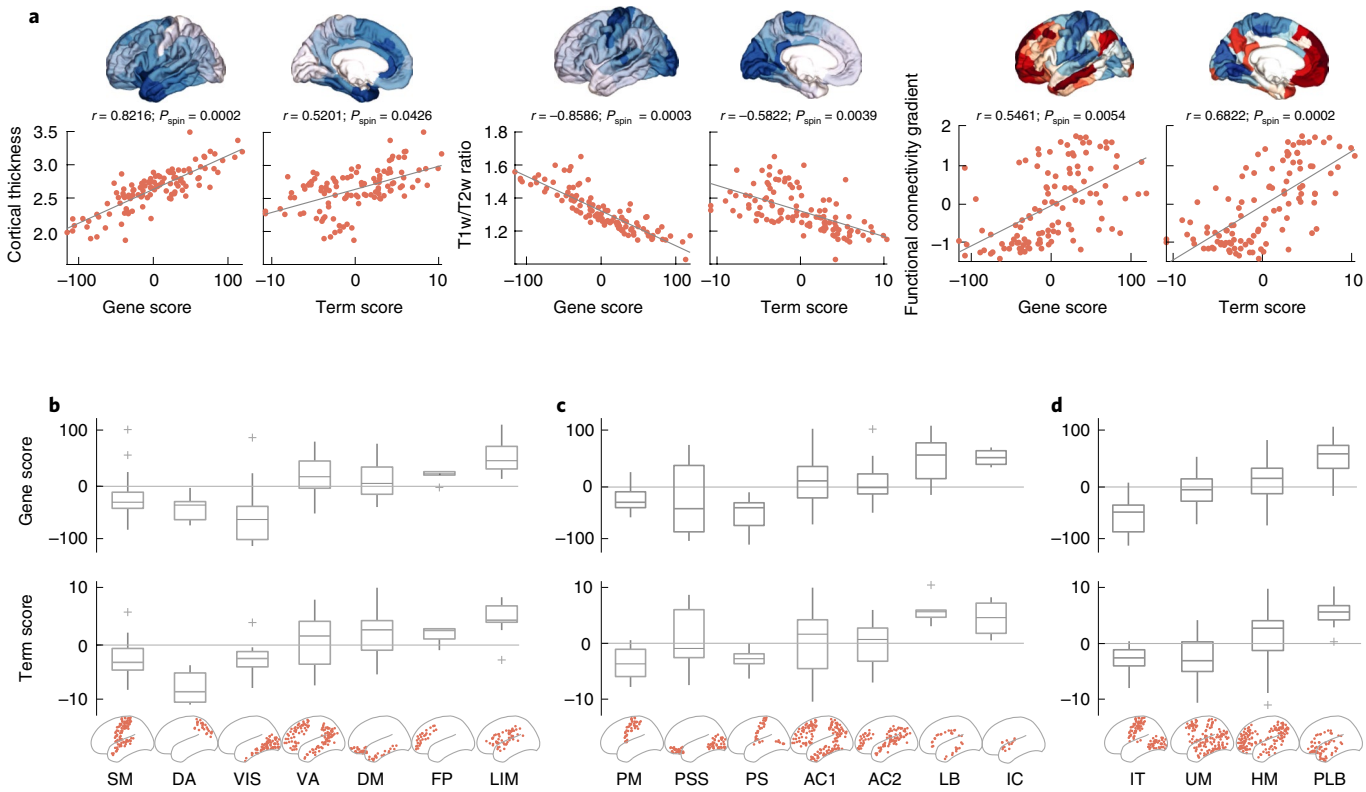

Fig. 3 | The gene expression–functional association gradient is organized around microscale and macroscale hierarchies. a, PLS analysis-derived score patterns are positively correlated with cortical thickness, negatively correlated with intracortical myelin (measured by T1w/T2w ratio) and positively correlated with the principal gradient of functional connectivity (Pearson's *r*; $n = 111$ brain regions; gradients *z*-scored). **b**, Distribution of scores across the seven intrinsic resting-state functional networks defined by Yeo and colleagues[9]. **c**, Distribution of scores across the seven Von Economo cytoarchitectonic classes[15,55]. **d**, Distribution of PLS analysis-derived gene and term scores across the four Mesulam levels of laminar differentiation[56,57]. The boxplots in **b**–**d** represent the first, second (median) and third quartiles, whiskers represent the non-outlier end points of the distribution, and crosses represent outliers. AC1/2, association cortex 1/2 ($n = 40$ and 24, respectively); DA, dorsal attention ($n = 7$); DM, default mode ($n = 33$); FP, frontoparietal ($n = 6$); HM, heteromodal ($n = 37$); IC, insular cortex ($n = 4$); IT, idiotypic ($n = 23$); LB, limbic regions ($n = 7$); LIM, limbic ($n = 13$); PLB, paralimbic ($n = 18$); PM, primary motor cortex ($n = 10$); PS, primary sensory cortex ($n = 9$); PSS, primary/secondary sensory cortex ($n = 17$); SM, somatomotor ($n = 23$ regions); UM, unimodal ($n = 33$); VA, ventral attention ($n = 14$); VIS, visual ($n = 15$).

adolescent ($r(10) = 0.85$; $P < 0.001$; bootstrap-estimated 95% CI = (0.70, 0.96)) and adult life stages ($r(10) = 0.77$; $P = 0.0035$; bootstrap-estimated 95% CI = (0.54, 0.95)).

Sensitivity and robustness analysis. All analyses presented thus far were conducted on a particular parcellation of brain regions and a predefined set of genes. To ensure the observed results are not dependent on these methodological choices, we compared results when analyses were repeated across different node resolutions and gene sets. Furthermore, we replicated the results using a second dataset for the construction of the functional association matrix.

To ascertain our findings against different choices of parcellation resolution, gene expression and functional association matrices were parcellated into three resolutions: a 34-node parcellation, a 57-node parcellation and a 111-node parcellation[31]. Importantly, the 57- and 111-node parcellations were derived by dividing the 34-node parcellation into smaller parcels, such that each node in the 57- and 111-node parcellations was a child of a node in the 34-node parcellation. Furthermore, since the original analyses were conducted on differentially stable genes (see Methods) and the calculation of differential stability depends on parcellation resolution, the number of stable genes retained across parcellations varied (between 8,825 and 11,560 genes retained). The gene expression and functional association matrices at these three resolutions were subjected to PLS analysis and the gene and term scores were computed. In the two finer resolutions, the mean score of sibling nodes was computed such that one score for each of the 34 parent nodes was available for all three resolutions. Correlating gene scores and term scores across resolutions revealed an almost one-to-one relationship, indicating that node resolution has little impact on gene and term scores (Supplementary Fig. 7).

Likewise, we asked whether the gene sets contributing most to the latent variable would be altered based on which genes were included in the gene expression matrix. Since the two gene sets underlying perception and affect are defined based on gene loadings, we compared loadings of genes across six variations of the gene expression matrix. For each of the three resolutions introduced above, one matrix includes all 20,323 genes and another includes only differentially stable genes, as defined by the specific parcellation. Each of the six gene expression matrices, alongside their corresponding Neurosynth functional association matrix, were subjected to PLS analysis and loadings were computed for each gene, where genes with the top 50% of positive and negative loadings were considered reliable. When compared across the six different gene expression matrices, we found that reliable gene sets were highly consistent (Supplementary Fig. 8).

Next, we replicated the original results using a different data source for construction of the functional association matrix. BrainMap is a manually curated database of published voxel coordinates from neuroimaging studies that are significantly activated or

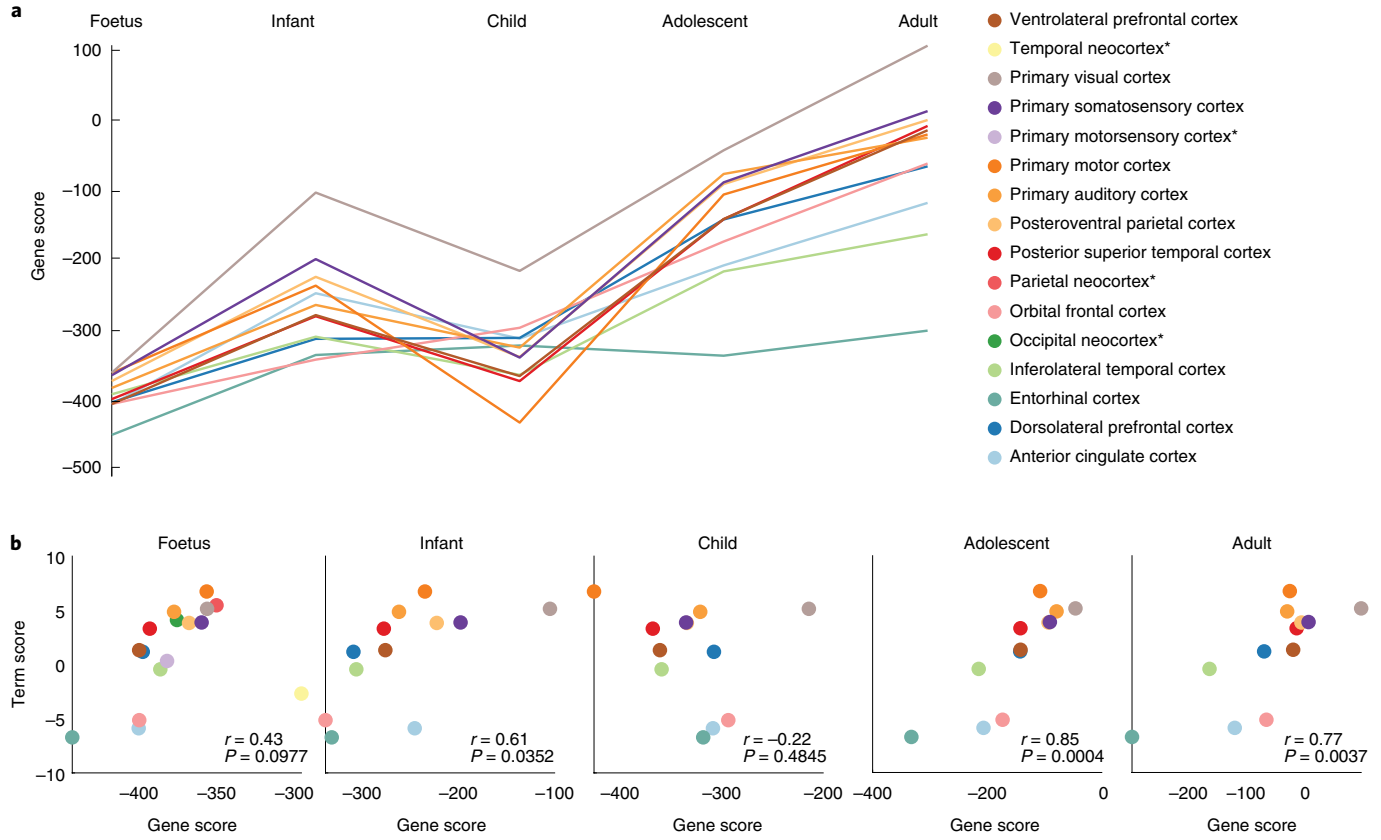

Fig. 4 | The molecular signature of psychological function strengthens with development. The BrainSpan database was used to replicate the results and to compare how the isolated genetic signature develops over the lifespan[22]. **a**, Gene scores for 12 unique brain regions gradually increase with development and peak in adulthood. Brain regions with gene expression levels only available in the foetal stage are indicated with an asterisk, and no corresponding curve is shown. **b**, The Pearson's correlation between estimated gene scores and PLS analysis-derived term scores is strongest in adolescence and adulthood ($n = 16$ brain regions in the foetus; $n = 12$ regions in all other developmental stages).

deactivated during tasks[17,58–60]. Using the analytic pipeline we previously applied to Neurosynth, we converted BrainMap data into a functional association matrix of probabilities, which included 66 terms (see Methods for details and Supplementary Table 2 for a full list of terms).

PLS analysis on the original gene expression matrix with this BrainMap-derived functional association matrix again revealed a single statistically significant latent variable that captured 51% of the covariance between gene expression and functional association ($P_{spin} = 0.0034$). The gene and term score distributions again followed a ventromedial–dorsolateral gradient (Supplementary Fig. 9a) and gene weights were highly correlated with the original Neurosynth-derived gene weights (Supplementary Fig. 9b). Term loadings were computed and the reliable positively and negatively correlated terms are shown in Supplementary Fig. 9c. Unlike the terms used in the Neurosynth-derived functional association matrix, some terms in BrainMap were pharmacological in nature. Interestingly, positive pharmacology terms were primarily depressants (such as alcohol and marijuana) and negative pharmacology terms were primarily stimulants (such as caffeine).

Discussion

Here, we identify spatially covarying gradients of gene expression and functional association across the neocortex. Collectively, these patterns delineate a ventromedial–dorsolateral axis, separating gene sets related to perceptual versus affective function. The spatial patterning of gene and term scores follows a hierarchical organization and is closely related to multiple structural and functional attributes. We assess the robustness of our results in two distinct datasets and show that the gene association signature strengthens with human development. Our results directly bridge microscale gene expression to macroscale functional processes and highlight the influence that molecular mechanisms have on cognition and behaviour.

The present findings build on previous reports that link gene expression to the structural and functional architecture of the brain. Gene expression profiles have been linked to cortical folding[61], cortical shrinkage during adolescence[24], subcortical connectivity[62] and patterns of long- and short-distance neural communication[12,15,63]. In particular, intracortical myelin distribution, as measured by the T1w/T2w ratio, is correlated with regional transcription levels, potentially reflecting a hierarchical axis of cytological properties, including cytoarchitecture and cell density[13,26]. Additionally, the extracted gene pattern closely resembles a number of previously reported transcriptional maps, including the first principal component of gene expression[20] and the gene expression maps associated with magnetization transfer[24], disease vulnerability[23] and structural and functional network topology[15]. While these reports demonstrate a link between transcription and multiple structural features, the functional and behavioural manifestation of such trends remains unknown. Our results fill in this critical gap, showing a statistical correspondence between transcription and psychological function that is likely to be mediated by microstructure. Collectively, the topographic alignment of molecular and structural features appears to reflect a dominant underlying organizational axis, shaping regional functional specialization.

What do the present findings show us about the regional specialization for psychological functions? Although we have summarized the gene association gradient as one primarily differentiating perceptual and affective process, greater nuance is warranted. In particular, the posterior/dorsal system is more specifically related to perception, orienting and attention, whereas the anterior/ventral system is more specifically related to emotion and evaluation. Thus, the axis differentiates attentional and evaluation functions and may be more aptly termed an affective–attentive or evaluation–perception axis. Interestingly, many of the intermediate terms that do not load highly on either end of the axis are integrative in nature (for example, consciousness, integration, episodic memory, communication and so on; Supplementary Fig. 4), suggesting that these more complex functions lie at the intersection of the two systems.

How covarying patterns of gene expression and functional association emerge over the course of ontogeny and phylogeny remains an open challenge. Patterns of gene expression are involved in cortical reorganization during neurodevelopment, including folding[61], pruning[24] and establishment of corticocortical connectivity[25]. In the BrainSpan dataset, we found that the gene–cognition signature has a protracted trajectory over development, gradually becoming most prominent in adulthood (Fig. 4). This external confirmation of our results suggests continued refinement and differentiation of perceptual–affective processes during maturation, but more research is necessary to understand the behavioural consequences of these processes at the individual participant level. A related question is how the association between transcription and functional activation evolves across phylogeny. In particular, evolutionary expansion of the cortical mantle is thought to have altered the relationship between molecular gradients and microcircuitry, promoting increasingly complex cognitive function[63,64]. Thus, the present work could be extended by comparing psychologically relevant expression patterns, biological processes and cell type composition across species.

Indeed, one possible explanation for the relative ordering of terms in the PLS model could be that processes at the ends of the spectrum, related to affect and perception/attention, are putatively more basal and thus retain a stronger link with gene expression[64]. Recent work on inter-individual differences in resting-state fMRI connectivity is consistent with this notion, with lower inter-individual variability and lower fingerprinting identifiability in somatomotor and visual cortex, as well as in ventromedial regions[65,66], suggesting that activations in these areas may be more readily elicited in experimental settings. At the same time, many of the high-scoring regions in the present PLS model show considerable (although not maximal) inter-individual variability, particularly the medial and dorsolateral prefrontal cortex. How genomic–neurocognitive links vary across individuals is a key question for future studies.

We mapped whole-genome transcription patterns to a spectrum of psychological functions across multiple brain areas, but the relationship between gene expression and behaviour has previously been approached from different directions. One approach is to focus on a region of interest. For instance, variations in cognitive function reflect a transcriptional gradient across the long axis of the hippocampus[67]. An alternative approach is to map single functions of interest to single genes or gene modules. For instance, Fox and colleagues[28] used Neurosynth and the AHBA to identify multiple gene–cognition associations in the subcortex, including previously established associations between dopamine receptor genes and reward functions in the basal ganglia, as well as novel associations. The results reported here open new possibilities for mapping high-dimensional transcriptional readouts to neurocognitive function in a data-driven and multivariate analysis framework[68], broadening the scope of inquiry to multiple gene sets and comprehensive neurocognitive profiles.

Mapping transcription to cognition opens a new way to generate top-down hypotheses about genes involved in a specific psychological process, or bottom-up hypotheses about the psychological correlates of specific genes. We envision that these data will motivate future mechanistic studies of transcriptomic vulnerability in disease[23]. For instance, modern psychiatry increasingly emphasizes complex multidimensional profiles of individual patients that include diverse aspects of cognitive performance and affect[69]. Thus, to design therapies tailored to the psychological profiles of individual patients, it is necessary to first map gene expression to psychological processes. In addition, the mapping between genes and cognition offers a powerful biological lens that can reveal multiple targets for therapeutic intervention, from single genes to gene assemblies to molecular pathways to cell types.

A salient example of how this mapping can inform and conceptually link previous findings is in the domain of mood disorders. First, our findings show that meta-analytical activation related to emotion/affect is strongly associated with support cells (astrocytes and microglia) rather than neurons. This is consistent with the inflammation hypothesis, which posits that repeated systemic exposure to microglia-mediated inflammation is a significant risk factor for multiple mental illnesses[70]. Second, this affect-related genomic signature is centred on the anterior cingulate cortex, which is disproportionately vulnerable in mental illnesses[71–73]. Interestingly, this signature is enriched for neurogenesis and differentiation, rather than signalling, suggesting a potential developmental origin of this vulnerability. Altogether, these findings provide the potential to integrate multiple scales of description of mood disorders.

Here, we demonstrate a statistical correspondence between two modalities that are both structurally embedded in the brain. As with any statistical analysis related to biological mechanisms, caution is warranted when interpreting the results. We triangulate towards a statistical mapping between gene expression and functional association using conservative null models, cross-validation and external datasets across a range of parcellations and gene sets, but each avenue alone does not yield an overwhelmingly strong result. Furthermore, as we discuss above, the structural architecture of the brain is probably a mediating factor between these modalities. Although we report an association between gene expression and functional association that is statistically unexpected based on autocorrelation alone (65% covariance), autocorrelation may still account for a considerable portion of the covariance explained between both datasets (34% covariance). Future research is required to identify how the structural architecture of the brain ties transcription to psychology, and is needed to demonstrate a clear and causal biological link between these two modalities.

The present work should be understood alongside some important methodological considerations. First, the main analysis involved two singular datasets, potentially limiting the generalizability of the results. Despite extensive robustness analyses, the present findings are based on small samples of postmortem brains and more comprehensive microarray gene expression datasets are necessary for future studies. Furthermore, replication of results was only conducted on one external dataset at a time and external validation using BrainMap and BrainSpan together could not be done due to the limited number of cortical genetic samples. Second, all analyses were performed in the left cerebral cortices of the six donors, precluding any tests of lateralized brain function, such as language. Third, due to well-documented differences in the transcriptional signatures of the cortex, subcortex and cerebellum[74], the present investigation focused only on the cortex. How gene expression and functional coactivation covary in subcortical structures should be investigated in future work[28]. Fourth, the mapping of functional activation to psychological terms in Neurosynth cannot distinguish activations from deactivations[16]. Thus, the present results identify

gene assemblies whose expression covaries with functional associations, but do not isolate the direction of effect.

In summary, we demonstrate that patterns of gene expression influence cognition and emotion. Organized across a spatially ordered ventromedial–dorsolateral gradient, we show that this genetic signature shapes the composition of cell types and microstructure, ultimately manifesting as a large-scale axis differentiating affective and perceptual processes. Collectively, these results highlight a statistical link between molecular dynamics and psychological function.

Methods

Microarray gene expression. Regional microarray expression data were obtained from six postmortem brains provided by the AHBA (http://human.brain-map.org/)[20]. Since only two of the six brains included samples from the right hemisphere, analyses were conducted on the left hemisphere only. All processing was performed using the abagen toolbox (https://github.com/netneurolab/abagen). These data were processed and mapped to parcellated brain regions at three increasingly finer resolutions, from 34 to 111 left hemisphere cortical grey matter nodes according to the Lausanne anatomical atlas[30,31].

Microarray probes were reannotated using data provided by Arnatkevičiūtė et al.[27]. A single microarray probe with the highest differential stability, $\Delta_S(p)$, was selected to represent each gene[21], where differential stability was calculated as:

$$\Delta_S(p) = \frac{1}{\binom{N}{2}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} r[B_i(p), B_j(p)] \quad (1)$$

Here, r is the Spearman's rank correlation of the expression of a single probe p across regions in two donor brains, B_i and B_j, and N is the total number of donor brains. This procedure retained 20,232 probes, each representing a unique gene. Note that probe selection can alternatively be achieved using RNA sequencing data[75], which produce very similar PLS results (gene scores, term scores and term weights all highly correlated across methodology; $r > 0.99$).

Next, samples were assigned to brain regions, using Montreal Neurological Institute (MNI) coordinates generated via nonlinear registrations (https://github.com/chrisfilo/alleninf), by finding the nearest region, up to 2 mm away. To reduce the potential for misassignment, sample-to-region matching was constrained by hemisphere and cortical/subcortical divisions[27]. If a brain region was not assigned any sample based on the above procedure, the sample closest to the centroid of that region was selected to ensure that all brain regions were assigned a value.

Inter-individual variation was addressed by normalizing tissue sample expression values for each donor across genes using a scaled robust sigmoid function[12]:

$$x_{\text{norm}} = \frac{1}{1 + \exp(-\frac{(x_g - \langle x_g \rangle)}{\text{IQR}_x})} \quad (2)$$

where $\langle x_g \rangle$ is the median and IQR is the normalized interquartile range of the expression value of a single gene across regions. Normalized gene expression values were then rescaled to a unit interval:

$$x_{\text{scaled}} = \frac{x_{\text{norm}} - \min(x_{\text{norm}})}{\max(x_{\text{norm}}) - \min(x_{\text{norm}})} \quad (3)$$

Gene expression values were normalized across tissue samples using the same procedure. Samples assigned to the same brain region were then averaged separately for each donor. Scaled regional expression profiles were finally averaged across donors, resulting in a single matrix $\boldsymbol{X}$ with r rows corresponding to brain regions and g columns corresponding to the retained 20,232 genes. A threshold of 0.1 was imposed on the differential stability of each gene, such that only stable genes were retained for future analysis. Genes with large differential stability have previously been shown to be related to biological relevance, disease, drug targets and literature citations[21]. Furthermore, this stability threshold minimizes the influence of unreliable gene expression measurements by preferentially selecting genes whose expression is consistent across the six sampled brains. At the 34-, 57- and 111-node resolutions, the ensuing numbers of stable genes retained were 11,560, 10,453 and 8,825, respectively.

Term functional associations. Probabilistic measures of the association between voxels and terms were obtained from Neurosynth, a meta-analytical tool that synthesizes results from more than 15,000 published fMRI studies by searching for high-frequency key words (such as pain and attention) that are published alongside fMRI voxel coordinates (https://github.com/neurosynth/neurosynth, using the volumetric association test maps[16]). This measure of association is the probability that a given term is reported in the study if there is activation observed at a given voxel. Note that the tool does not distinguish between areas that are activated or deactivated in relation to the term of interest, nor the degree of activation, only that certain brain areas are frequently mentioned in conjunction with certain words. Although more than 1,000 terms are reported in Neurosynth, we focus primarily on cognitive function and therefore limit the terms of interest to cognitive and behavioural terms. These terms were selected from the Cognitive Atlas, a public ontology of cognitive science[29] that includes a comprehensive list of neurocognitive terms and has previously been used in conjunction with Neurosynth[76]. We used $t = 123$ terms, ranging from umbrella terms (attention and emotion) to specific cognitive processes (visual attention and episodic memory), behaviours (eating and sleep) and emotional states (fear and anxiety). The coordinates reported by Neurosynth were parcellated into 111 left hemisphere cortical regions. The probabilistic measure reported by Neurosynth can be interpreted as a quantitative representation of how regional fluctuations in activity are related to psychological processes. For simplicity, we refer to these probabilities as functional associations throughout the present report. The full list of terms is shown in Supplementary Table 1.

PLS analysis. PLS analysis was used to relate gene expression to functional association. PLS analysis is an unsupervised multivariate statistical technique that decomposes relationships between two datasets (in our case, gene expression, $\boldsymbol{X}_{n\times g}$ and functional association, $\boldsymbol{Y}_{n\times t}$) into orthogonal sets of latent variables with maximum covariance, which are linear combinations of the original data. This was done by z scoring both data matrices columnwise and applying singular value decomposition on the matrix $Y'X$ such that:

$$(Y'X)' = USV' \quad (4)$$

where $\boldsymbol{U}_{g\times t}$ and $\boldsymbol{V}_{t\times t}$ are orthonormal matrices consisting of left and right singular vectors and $\boldsymbol{S}_{t\times t}$ is a diagonal matrix of singular values (Fig. 1a)[77]. The ith columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ constitute a latent variable, and the ith singular value in $\boldsymbol{S}$ represents the covariance between singular vectors. The ith singular value is proportional to the amount of covariance between gene expression and functional association captured by the ith latent variable, where the effect size can be estimated as the ratio of the squared singular value to the sum of all squared singular values. In the present study, the left singular vectors (that is, the columns of $\boldsymbol{U}$) represent the degree to which each gene contributes to the latent variable and demonstrate the extracted association between gene expression and functional associations (gene weights; see Supplementary Table 6 for a complete list of gene weights). The right singular vectors (that is, the columns of $\boldsymbol{V}$) represent the degree to which the cognitive terms contribute to the same latent variable (term weights). Positively weighed genes covary with positively weighed terms, and negatively weighed genes covary with negatively weighed terms. Gene and term scores at each brain region for each latent variable can be computed by projecting the original data onto the singular vector weights (Fig. 1b,e). Positively scored brain regions are regions that demonstrate the covariance between the expression of positively weighted genes and the association of positively weighted cognitive terms (and vice versa for negatively scored brain regions):

$$\text{Gene scores for latent variable } i = X_{n\times g} U(i)_{g\times 1}$$

$$\text{Term scores for latent variable } i = Y_{n\times t} \mathbf{V}(i)_{t\times 1}.$$

Loadings for each variable were computed as the Pearson's correlation between each individual variable's activity (gene expression and term functional associations) and the PLS analysis-derived gene score pattern (Fig. 1c and Supplementary Fig. 4). Note that, alternatively, term loadings could have been computed as the correlation between an individual term's functional associations and the PLS analysis-derived term score pattern (Supplementary Fig. 5); term loadings are highly correlated between methodologies ($r(121) = 0.95$; $P < 0.001$; bootstrap-estimated 95% CI = (0.9426, 0.9687)). Squaring the loading (a correlation) equals the percentage variance shared between an original variable (a specific gene or term) and the PLS analysis-derived latent variable. Terms and genes with high absolute loadings are variables that are highly correlated to the score pattern, indicating a large amount of shared variance between the individual variable and the latent variable. The reliability of gene and term loadings was assessed using bootstrap resampling (see the following subsection).

Note that, by design, PLS analysis does not pre-whiten the two data matrices before the decomposition, such that the model takes into account both within-set and between-set relationships[78,79]. As a result, the data matrix with the stronger correlation structure may drive the solution. CCA is a similar reduced-rank regression that seeks to match two sets of variables by extracting patterns in the original dataset that maximally correlate (instead of covary, as in PLS analysis)[80]. Unlike PLS analysis, CCA removes within-set correlations by pre-whitening the data, thus theoretically eliminating the possibility that the correlation structure in one dataset would dominate the solution. In other words, CCA is designed such that one data domain cannot bias the results away from the other data domain. To investigate whether CCA and PLS analysis would yield different results, we applied CCA to the gene expression and functional association matrices and found similar results (Supplementary Fig. 1).

The robustness of the PLS model was assessed by cross-validating the correlation between gene scores and term scores (Pearson's *r*). Since our observations are brain areas and therefore not independent, we designed the cross-validation such that the training and testing set were composed of spatially distant brain regions. To achieve this, a random source node and the 75% of brain regions closest in Euclidean distance composed the training set, and the remaining 25% of brain regions composed the testing set (Fig. 1d). PLS analysis was used to compute gene and term scores from the training set, as well as the correlation between the two (Corr($\boldsymbol{X}_{\text{train}}\boldsymbol{U}_{\text{train}}$, $\boldsymbol{Y}_{\text{train}}\boldsymbol{V}_{\text{train}}$)). The test set was projected onto the training-derived singular vector weights to generate predicted gene and term scores, and the correlation between predicted scores was computed (Corr($\boldsymbol{X}_{\text{test}}\boldsymbol{U}_{\text{train}}$, $\boldsymbol{Y}_{\text{test}}\boldsymbol{V}_{\text{train}}$)). This procedure was repeated 100 times, yielding a distribution of score correlations for the training and testing sets (Fig. 1d). The significance of the mean out-of-sample correlation was assessed against a permuted null model, which was constructed by repeating the cross-validation on spatial autocorrelation-preserving permutations of the functional association matrix (1,000 repetitions; Fig. 1f).

Bootstrap resampling. The reliability of the gene (term) weights was assessed using bootstrap resampling. The rows (brain regions) of the gene expression (functional association) matrix were randomly selected with replacement 10,000 times. PLS analysis was performed using this new bootstrapped gene expression (functional association) matrix and the original functional association (gene expression) matrix. This was repeated 10,000 times to estimate a sampling distribution of gene (term) weights. The ratio of the true gene (term) weight to the standard error of its bootstrap-resampled distribution (termed the bootstrap ratio) was computed[81]. Genes and terms with large bootstrap ratios have large weights (that is, contribute greatly to the relationship captured by the latent variable) and small standard error (that is, are reliable). Note that bootstrap ratios are highly correlated with gene loadings ($r(8{,}823) = 0.94$; $P < 0.001$; bootstrap-estimated 95% CI = (0.9406, 0.9441)).

Null model. Spatial autocorrelation-preserving permutation tests, termed spin tests, were used to assess the statistical significance of associations across brain regions[76]. We created a surface-based representation of the parcellation on the Free Surfer fsaverage surface, via files from the Connectome Mapper toolkit (https://github.com/LTS5/cmp). We used the spherical projection of the fsaverage surface to define spatial coordinates for each parcel by selecting the coordinates of the vertex closest to the centre of the mass of each parcel[35]. These parcel coordinates were then randomly rotated and original parcels were reassigned the value of the closest rotated parcel (10,000 repetitions). Parcels for which the medial wall was closest were assigned the value of the next most proximal parcel instead. The procedure was performed at the parcel resolution rather than the vertex resolution to avoid upsampling the data, and only for the left hemisphere.

In PLS analysis, the spin test is applied to the singular values (or equivalently, the covariance explained) of the latent variables, producing a null distribution of singular values. This is done applying PLS analysis to the original gene expression matrix and a spun term association matrix. The spin test embodies the null hypothesis that genes and terms are spatially correlated with each other because of inherent spatial autocorrelation. The *P* value is computed as the proportion of null singular values that are greater in magnitude than the empirical singular values. Thus, the *P* value represents the probability that the observed spatial correspondence between genes and terms could occur by randomly correlating maps with comparable spatial autocorrelation (Fig. 1d; grey boxplots; the normality of the data distributions was not formally tested). *P* values for all latent variables can be found in Supplementary Table 3 and examples of null gene and term scores can be found in Supplementary Fig. 10.

Gene set analysis. To determine the biological processes in which the gene sets identified by PLS analysis are most involved, we adapted analyses from the gene category enrichment analysis toolbox (originally available at https://github.com/benfulcher/GeneCategoryEnrichmentAnalysis[38]; for the original approach, see ref. [82]). Each biological process category is associated with a subset of genes, annotations of which are provided by Gene Ontology (geneontology.org). To determine the biological process categories in which the strongly contributing positive and negative genes are most involved, we ran our enrichment analysis separately on the positive genes (genes with the 50% most positive loadings) and negative genes (genes with the 50% most negative loadings). For each category, we define the category score as the mean loading of the genes in the category. The null model was constructed by permuting the rows (brain areas) of the functional association matrix while preserving spatial autocorrelation using the spherical projection and rotation procedure (spins) described in the previous subsection. We then subjected the original gene expression matrix and the permuted functional association matrix to PLS analysis and recomputed the null gene loadings and null category scores (10,000 repetitions). Significance was assessed using a two-tailed permutation test.

Next, cell type deconvolution was performed using cell-specific aggregate gene sets across five human adult postmortem single-cell and single-nucleus RNA sequencing studies[39–44], as presented previously (Supplementary Table 5 in ref. [23]). Briefly, cortical cell classes were determined based on hierarchical clustering of regional topographies across all study-specific cell types in the AHBA, resulting in seven major canonical cortical cell classes: astrocytes, endothelial cells, microglia, excitatory neurons, inhibitory neurons, oligodendrocytes and oligodendrocyte precursors (see Fig. 2 in ref. [23]).

HCP dataset. Data from the HCP (S1200 release)[83,84] were used for measures of cortical thickness, T1w/T2w ratios, functional connectivity and behavioural tests, as approved by the WU-Minn HCP Consortium. The 417 unrelated participants (age range = 22–37 years; 193 males) with available resting-state fMRI data had individual measures of cortical thickness and T1w/T2w ratios. These structural modalities were acquired on a Siemens Skyra 3T scanner and included a T1-weighted MPRAGE sequence and a T2-weighted SPACE, both at an isotropic resolution of 0.7 mm. Details on imaging protocols and procedures are available at http://protocols.humanconnectome.org/HCP/3T/imaging-protocols.html. Image processing includes correcting for gradient distortion caused by nonlinearities, correcting for bias field distortions and registering the images to a standard reference space. Measures of cortical thickness are estimated as the geometric distance between the white and grey matter surfaces, and intracortical myelin as the T1w/T2w ratio. The cortical thickness and T1w/T2w ratios for each participant were made available in the surface-based CIFTI file format and parcellated into 219 cortical regions according to the Lausanne anatomical atlas[31]. Only the left hemisphere regions were retained for analysis.

A group-averaged dense functional connectivity matrix was constructed from the 1,003 participants with all four 15-min resting-state fMRI runs. For details on how the dense functional connectivity matrix was constructed, see https://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP1200-DenseConnectome+PTN+Appendix-July2017.pdf. The cortical subset of the matrix was parcellated into 219 nodes according to the Lausanne anatomical atlas[31]. Following Margulies and colleagues[52], a principal functional gradient was computed by applying diffusion map embedding to the functional connectivity matrix, using the dimensionality reduction toolbox (https://lvdmaaten.github.io/drtoolbox/). The procedure yielded an eigenvector map representing the differentiation of unimodal and transmodal cortical regions. Only the left hemisphere was retained for comparison with gene and term scores.

Robustness analysis using BrainMap. BrainMap is a manually created and curated data repository of results from published functional and structural neuroimaging studies[17,58–60]. Specifically, BrainMap includes the brain coordinates that are significantly activated during thousands of different experiments. All of the experiments conducted on unhealthy participants were excluded, as well as all experiments without a defined behavioural domain. This resulted in 8,703 experiments organized into 66 unique behavioural domains (Supplementary Table 2). To enable more direct comparability with results using Neurosynth, the BrainMap data were subjected to the Neurosynth meta-analytical pipeline (https://github.com/neurosynth/neurosynth). Note that Neurosynth terms and BrainMap behavioural domains differ considerably, making term comparison challenging (see Supplementary Table 1 versus Supplementary Table 2). This approach resulted in a region by term matrix of probabilistic measures that certain terms are published in conjunction with certain brain regions.

Robustness analysis using BrainSpan. BrainSpan is a database of gene expression in the brain across development, available at https://www.brainspan.org/static/download.html[22]. Gene expression levels were quantified in specific tissue samples from postmortem brains ranging from 8 post-conception weeks to 40 years of age. Ages were binned into five life stages: foetus (8–37 post-conception weeks), infant (4 months–1 year), child (2–8 years), adolescent (11–19 years) and adult (21–40 years)[85]. For each age category, a gene expression matrix was constructed by averaging the expression of every gene across identical regions, ignoring the genes without any expression values (0.057% of the data). Genes with no estimated expression level in a particular brain region across all samples in a life stage were imputed using the median gene expression across all genes in that brain region; this interpolation occurred infrequently (0.022% of the time). Of the 16 unique cortical brain regions with gene expression levels, four regions did not have gene expression estimates in any of the age categories besides the foetal stage. Therefore, only gene scores in the 12 consistent brain regions were tracked across life stages. For comparison with PLS analysis results derived from the AHBA, we selected the 8,607 available genes with differential stability greater than 0.1, as defined on the 34-node parcellation, and complete sample data across all brain regions for each life stage. Gene scores for the 12 regions with expression levels available for all age stages were estimated by multiplying the gene expression matrices and the PLS analysis-derived gene weights (columns of $\boldsymbol{U}$).

To relate estimated gene scores with term scores, we defined a region-to-region correspondence map from the 16-node parcellation to the 34-node parcellation. Term scores were averaged across sibling nodes such that a single term score was available for all 16 regions available in BrainSpan. Note that brain regions in BrainSpan are not organized by hemisphere; therefore, regional expression levels in the 16 regions are not necessarily measured from the left hemisphere only. Furthermore, due to the small sample of cortical brain regions, gene expression estimates are much less specific compared with the AHBA dataset.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The AHBA is available at https://human.brain-map.org/static/download. The Neurosynth database is available at https://neurosynth.org/. The HCP database is available at https://db.humanconnectome.org/data/projects/HCP_1200. The BrainSpan database is available at https://www.brainspan.org/static/download.html. Processed data as used in this report are available at https://github.com/netneurolab/hansen_genescognition. BrainMap data are available upon reasonable request to B.M.

Code availability

Code used to conduct the reported analyses is available at https://github.com/netneurolab/hansen_genescognition.

Received: 17 July 2020; Accepted: 18 February 2021;
Published online: 25 March 2021

References

1. Betzel, R. F. & Bassett, D. S. Multi-scale brain networks. *NeuroImage* **160**, 73–83 (2017).
2. Tripathy, S. J. et al. Transcriptomic correlates of neuron electrophysiological diversity. *PLoS Comput. Biol.* **13**, e1005814 (2017).
3. Cadwell, C. R. et al. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotech.* **34**, 199–203 (2016).
4. Lee, K. F. H., Soares, C., Thivierge, J.-P. & Béïque, J.-C. Correlated synaptic inputs drive dendritic calcium amplification and cooperative plasticity during clustered synapse development. *Neuron* **89**, 784–799 (2016).
5. Frémaux, N. & Gerstner, W. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Front. Neural Circuits* **9**, 85 (2016).
6. Blankenship, A. G. & Feller, M. B. Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nat. Rev. Neurosci.* **11**, 18–29 (2010).
7. Calhoon, G. G. & Tye, K. M. Resolving the neural circuits of anxiety. *Nat. Neurosci.* **18**, 1394–1404 (2015).
8. Suárez, L. E., Markello, R. D., Betzel, R. F. & Misic, B.Linking structure and function in macroscale brain networks. *Trends Cogn. Sci.* **24**, 302–315 (2020).
9. Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
10. Damoiseaux, J. S. et al. Consistent resting-state networks across healthy subjects. *Proc. Natl Acad. Sci. USA* **103**, 13848–13853 (2006).
11. Richiardi, J. et al. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
12. Fulcher, B. D. & Fornito, A. A transcriptional signature of hub connectivity in the mouse connectome. *Proc. Natl Acad. Sci. USA* **113**, 1435–1440 (2016).
13. Fulcher, B. D., Murray, J. D., Zerbi, V. & Wang, X.-J. Multimodal gradients across mouse cortex. *Proc. Natl Acad. Sci. USA* **116**, 4689–4695 (2019).
14. Betzel, R. F. et al. Structural, geometric and genetic factors predict interregional brain connectivity patterns probed by electrocorticography. *Nat. Biomed. Eng.* **3**, 902–916 (2019).
15. Vértes, P. E. et al. Gene transcription profiles associated with intra-modular and inter-modular hubs in human fMRI networks. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 735–769 (2016).
16. Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C. & Wager, T. D.Large-scale automated synthesis of human functional neuroimaging data. *Nat. Meth.* **8**, 665–670 (2011).
17. Fox, P. T. & Lancaster, J. L. Mapping context and content: the BrainMap model. *Nat. Rev. Neurosci.* **3**, 319–321 (2002).
18. Gorgolewski, K. J. et al. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinform.* **9**, 8 (2015).
19. Dockès, J. et al. NeuroQuery, comprehensive meta-analysis of human brain mapping. *eLife* **9**, e53385 (2020).
20. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
21. Hawrylycz, M. et al. Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).
22. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
23. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurodevelopmental disorders. *Nat. Commun.* **11**, 3358 (2020).
24. Whitaker, K. J. et al. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. *Proc. Natl Acad. Sci. USA* **113**, 9105–9110 (2016).
25. Váša, F. et al. Conservative and disruptive modes of adolescent change in human brain functional connectivity. *Proc. Natl Acad. Sci. USA* **117**, 3248–3253 (2020).
26. Burt, J. B. et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).
27. Arnatkevičiūtė, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage* **189**, 353–367 (2019).
28. Fox, A. S., Chang, L. J., Gorgolewski, K. J. & Yarkoni, T. Bridging psychology and genetics using large-scale spatial analysis of neuroimaging and neurogenetic data. Preprint at *bioRxiv* https://doi.org/10.1101/012310 (2014).
29. Poldrack, R. A. et al. The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Front. Neuroinform.* **5**, 17 (2011).
30. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
31. Cammoun, L. et al. Mapping the human connectome at multiple scales with diffusion spectrum MRI. *J. Neurosci. Methods* **203**, 386–397 (2012).
32. McIntosh, A. R., Bookstein, F. L., Haxby, J. V. & Grady, C. L. Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage* **3**, 143–157 (1996).
33. Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. Partial least squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage* **56**, 455–475 (2011).
34. McIntosh, A. R. & Mišić, B. Multivariate statistical analyses for neuroimaging data. *Annu. Rev. Psychol.* **64**, 499–525 (2013).
35. Vázquez-Rodríguez, B. et al. Gradients of structure–function tethering across neocortex. *Proc. Natl Acad. Sci. USA* **116**, 21219–21227 (2019).
36. Markello, R. & Misic, B. Comparing spatially-constrained null models for parcellated brain maps. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.13.249797 (2020).
37. Burt, J. B., Helmer, M., Shinn, M., Anticevic, A. & Murray, J. D.Generative modeling of brain maps with spatial autocorrelation. *NeuroImage* **220**, 117038 (2020).
38. Fulcher, B. D, Arnatkevičiūtė, A. & Fornito, A. Overcoming bias in gene-set enrichment analyses of brain-wide transcriptomic data. Preprint at *bioRxiv* https://doi.org/10.1101/2020.04.24.058958 (2020).
39. Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).
40. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
41. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
42. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).
43. Li, M. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
44. McKenzie, A. T. et al. Brain cell type specific gene expression and co-expression network architectures. *Sci. Rep.* **8**, 8868 (2018).
45. Huntenburg, J. M., Bazin, P.-L. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
46. Wang, X.-J.Macroscopic gradients of synaptic excitation and inhibition in the neocortex. *Nat. Rev. Neurosci.* **21**, 169–178 (2020).
47. Glasser, M. F. & Van Essen, D. C. Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011).
48. Wagstyl, K., Ronan, L., Goodyer, I. M. & Fletcher, P. C. Cortical thickness gradients in structural hierarchies. *NeuroImage* **111**, 241–250 (2015).
49. Huntenburg, J. M. et al. A systematic relationship between functional connectivity and intracortical myelin in the human cerebral cortex. *Cereb. Cortex* **27**, 981–997 (2017).
50. Jones, E. G. & Powell, T. P. S. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* **93**, 793–820 (1970).
51. Mesulam, M.-M. From sensation to cognition. *Brain* **121**, 1013–1052 (1998).
52. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl Acad. Sci. USA* **113**, 12574–12579 (2016).
53. Coifman, R. R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl Acad. Sci. USA* **102**, 7426–7431 (2005).
54. Van Der Maaten, L., Postma, E. & Van den Herik, J. Dimensionality reduction: a comparative. *J. Mach. Learn. Res.* **10**, 66–71 (2009).
55. Von Economo, C. & Koskinas, G. N. *Die Cytoarchitektonik der Hirnrinde des Erwachsenen Menschen* (Springer, 1925).
56. Mesulam, M.-M. *Principles of Behavioral and Cognitive Neurology* 2nd edn (Oxford Univ. Press, 2000).

57. Paquola, C. et al. Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS Biol.* **17**, e3000284 (2019).
58. Fox, P. T. et al. BrainMap taxonomy of experimental design: description and evaluation. *Hum. Brain Mapp.* **25**, 185–198 (2005).
59. Vanasse, T. J. et al. BrainMap VBM: an environment for structural meta-analysis. *Hum. Brain Mapp.* **39**, 3308–3325 (2018).
60. Laird, A. R., Lancaster, J. J. & Fox, P. T. BrainMap. *Neuroinformatics* **3**, 65–77 (2005).
61. Alexander-Bloch, A. F. et al. Imaging local genetic influences on cortical folding. *Proc. Natl Acad. Sci. USA* **117**, 7430–7436 (2020).
62. Forest, M. et al. Gene networks show associations with seed region connectivity. *Hum. Brain Mapp.* **38**, 3126–3140 (2017).
63. Krienen, F. M., Yeo, B. T. T., Ge, T., Buckner, R. L. & Sherwood, C. C. Transcriptional profiles of supragranular-enriched genes associate with corticocortical network architecture in the human brain. *Proc. Natl Acad. Sci. USA* **113**, E469–E478 (2016).
64. Buckner, R. L. & Krienen, F. M. The evolution of distributed association networks in the human brain. *Trends Cogn. Sci.* **17**, 648–665 (2013).
65. Mueller, S. et al. Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**, 586–595 (2013).
66. Jo, Y., Faskowitz, J., Esfahlani, F. Z., Sporns, O. & Betzel, R. F. Subject identification using edge-centric functional connectivity. Preprint at *bioRxiv* https://doi.org/10.1101/2020.09.13.291898 (2020).
67. Vogel, J. W. et al. A molecular gradient along the longitudinal axis of the human hippocampus informs large-scale behavioral systems. *Nat. Commun.* **11**, 960 (2020).
68. Mišić, B. & Sporns, O. From regions to connections and networks: new bridges between brain and behavior. *Curr. Opin. Neurobiol.* **40**, 1–7 (2016).
69. Kirschner, M. et al. Latent clinical-anatomical dimensions of schizophrenia. *Schizophr. Bull.* **46**, 1426–1438 (2020).
70. Amor, S., Puentes, F., Baker, D. & Van Der Valk, P. Inflammation in neurodegenerative diseases. *Immunology* **129**, 154–169 (2010).
71. Shafiei, G. et al. Spatial patterning of tissue volume loss in schizophrenia reflects brain network architecture. *Biol. Psychiatry* **87**, 727–735 (2020).
72. Bush, G., Luu, P. & Posner, M. I. Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* **4**, 215–222 (2000).
73. Goodkind, M. et al. Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305–315 (2015).
74. Patania, A. et al. Topological gene expression networks recapitulate brain anatomy and function. *Network Neurosci.* **3**, 744–762 (2019).
75. Miller, J. A. et al. Improving reliability and absolute quantification of human brain microarray data by filtering and scaling probes using RNA-seq. *BMC Genomics* **15**, 154 (2014).
76. Alexander-Bloch, A. F. et al. On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage* **178**, 540–551 (2018).
77. Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218 (1936).
78. Kovacevic, N., Abdi, H., Beaton, D. & McIntosh, A. R. in *New Perspectives in Partial Least Squares and Related Methods* 159–170 (Springer, 2013).
79. Helmer, M. et al. On stability of canonical correlation analysis and partial least squares with application to brain–behavior associations. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.25.265546 (2020).
80. Hotelling, H. The most predictable criterion. *J. Educ. Psychol.* **26**, 139–142 (1935).
81. McIntosh, A. R. & Lobaugh, N. J. Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage* **23**, S250–S263 (2004).
82. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
83. Essen, D. C. Van et al. The WU-Minn Human Connectome Project: an overview. *NeuroImage* **80**, 62–79 (2013).
84. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).
85. Werling, D. M. et al. Whole-genome and RNA sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex. *Cell Rep.* **31**, 107489 (2020).

Acknowledgements

We thank V. Bazinet, L. Suarez, G. Shafiei, B. Vazquez-Rodriguez and Z.-Q. Liu for comments and suggestions on the manuscript. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative. B.M. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant RGPIN 017-04265) and Canada Research Chairs Program. J.W.V. was funded by NIH grant T32MH019112. D.B. was supported by the Healthy Brains for Healthy Lives initiative (Canada First Research Excellence Fund), CIFAR Artificial Intelligence Chairs program (Canada Institute for Advanced Research), Google (Research Award) and NIH grant R01AG068563A. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

J.Y.H. and B.M. conceived of the study idea. J.Y.H. performed the formal analysis with contributions from R.D.M. R.D.M., J.W.V., J.S. and D.B. contributed data. J.Y.H. and B.M. wrote the manuscript with valuable revision by R.D.M., J.W.V., J.S. and D.B. B.M. was the project administrator.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01082-z.

Correspondence and requests for materials should be addressed to B.M.

Peer review information *Nature Human Behaviour* thanks Andre Altmann, Håkon Grydeland and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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.

nature research

Corresponding author(s): Bratislav Misic

Last updated by author(s): Feb 9, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Preprocessing and parcellating of the Allen Human Brain Atlas was done using abagen: https://github.com/netneurolab/abagen. Neurosynth and BrainMap data was converted to a probability matrix using the Neurosynth toolbox: https://github.com/neurosynth/neurosynth.
Data analysis	All analyses were conducted on Matlab version 9.8.0.1359463 (R2020a) Update 1. PLS analysis was done using pls_analysis.m from http://pls.rotman-baycrest.on.ca/source/Pls.zip.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The Allen Human Brain Atlas is available at https://human.brain-map.org/static/download, the Neurosynth database is available at https://neurosynth.org/, the Human Connectome Project S1200 database is available at https://db.humanconnectome.org/data/projects/HCP_1200, the BrainMap database is available at https://www.brainmap.org/, and the BrainSpan database is available at https://www.brainspan.org/static/download.html. Processed data as used in this report are available at https://github.com/netneurolab.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculations were performed. Only the 417 unrelated HCP S1200 subjects were used in analyses to rule out confounding contributions from genetic correlations among individuals due to family structure.
Data exclusions	Of 20,232 genes, a subset of differentially stable genes (defined as genes with differential stability greater than 0.1) were retained for analysis. Genes with differential stability less than 0.1 were excluded to minimize confounding effects caused by genes with high variation across the six brain samples. Analyses were repeated without any gene exclusions to demonstrate similar results. Only unrelated subjects from the Human Connectome Project were retained for analyses to rule out confounding contributions from genetic correlations among individuals due to family structure. Only terms in the intersection of the Neurosynth database and the Cognitive Atlas were retained. Neurosynth terms include a very wide variety of terms including regions ("dorsolateral") and clinical terms ("ADHD"), many of which are not relevant to the research question.
Replication	Data was replicated in using two independent data repositories: BrainMap, for a second probability matrix of terms (used to replace Neurosynth), and BrainSpan, for measures of gene expression. The PLS-derived score pattern was replicated in each dataset.
Randomization	No randomization was performed as this study does not include experimental groups.
Blinding	Blinding is not relevant to this study because it does not include experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Healthy unrelated adults (ages 22-37, 193/224 male/female split) from the HCP S1200 data release.
Recruitment	See Van Essen et al., 2012 for recruitment of HCP subjects.
Ethics oversight	The WU-Minn HCP Consortium (consortium of US and European intitutions led by Washington University and the University of Minnesota) approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Resting-state fMRI data.

Design specifications	Each subject had four 15-minute fMRI runs (1200 volumes).
Behavioral performance measures	No behavioural measures were recorded during the fMRI runs.

Acquisition

Imaging type(s)	Functional and structural MRI.
Field strength	3T
Sequence & imaging parameters	Multi-band sequence; functional images have a 2-mm isotropic signal resolution, structural modalities were acquired on a Siemens Skyra 3T scanner and included a T1-weighted MPRAGE sequence at an isotropic resolution of 0.7mm, and a T2-weighted SPACE at an isotropic resolution of 0.7mm. More details on imaging protocols and procedures are available at http://protocols.humanconnectome.org/HCP/3T/imaging-protocols.html.
Area of acquisition	Whole-brain.

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	We used the HCP data that was previously preprocessed. This preprocessing was done using FSL 5.0.6, FreeSurfer 5.3.0-HCP, and Connectome Workbench v1.1.1.
Normalization	Image processing includes correcting for gradient distortion caused by non-linearities, correcting for bias field distortions, and registering the images to a standard reference space.
Normalization template	fs_LR_32k surface mesh
Noise and artifact removal	FMRIB's ICA-based X-noisefier (FIX) and global signal regression
Volume censoring	No volume censoring was performed.

Statistical modeling & inference

Model type and settings	Cortical thickness (estimated as the geometric distance between the white and grey matter surfaces) and T1w/T2w ratios were estimated for each subject to produce a map across the brain. These maps were used in correlational analyses only. The multivariate modeling in this study is not performed on MRI data. Resting-state functional data was used to extract a principal gradient of functional connectivity using diffusion map embedding.
Effect(s) tested	No effects were being tested, only resting-state data was used.

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s)	Analyses were conducted using the Lausanne anatomical atlas, which is a subdivision of the Desikan-Killiany atlas. See Cammoun et al., 2012.
Statistic type for inference (See Eklund et al. 2016)	NA
Correction	NA

Models & analysis

n/a	Involved in the study
☐	☒ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	Functional connectivity comes from the group-averaged dense connectome from the 1003 HCP subjects with all four fMRI runs, available at https://db.humanconnectome.org/data/projects/HCP_1200.
Multivariate modeling and predictive analysis	No multivariate modeling was conducted on the HCP dataset - however, partial least squares analysis was involved in the study on other datasets (Allen Human Brain Atlas, Neurosynth, BrainSpan, BrainMap).