Resampling genes with mean correlation as a test statistic inflates false positive rates in associations between transcriptional profiles and imaging-derived phenotypes

Zhipeng Cao[1], Renata B. Cupertino[1], Peter Callas[2], Jonatan Ottino-Gonzalez[1], Alistair Murphy[1], Devarshi Pancholi[1], Sage Hahn[1], Dekang Yuan[1], Scott Mackey[1], Hugh Garavan[1]

1. Department of Psychiatry, University of Vermont College of Medicine, Burlington VT, 05401, USA
2. Department of Mathematics and Statistics, University of Vermont College of Engineering and Mathematical Sciences, Burlington VT, 05401, USA

Address correspondence to Zhipeng Cao, Department of Psychiatry, University of Vermont College of Medicine, Burlington 05401, VT, USA. Email: zhipeng30@foxmail.com.

Short Tittle: False positives in imaging transcriptomics

Number of pages: 33

Number of figures: 2

Number of tables: 0

Number of words for abstract: 283

Number of words for main text: 3462 (with numeric style)

Abstract

Correlating transcriptional profiles with imaging-derived phenotypes has the potential to reveal possible molecular architectures associated with brain development and disorders. A permutation-based statistical test has been widely used to determine the significance of the association wherein the mean value derived from an empirical distribution of correlations is tested against null models. However, because of the potential functional dependence among transcriptional profiles, this statistical test may fail to optimally control the false positive correlation rate. In the present study, we examined the performance of nine non-parametric approaches consisting of three types of null models (i.e., resampling genes, shuffling brain regions, and spinning brain regions) and three test statistics (i.e., mean correlation, mean absolute correlation, number of significant correlations) in determining the spatial association between transcriptional profiles of genes of interest (GOIs) and interregional variations of single brain phenotypes. Simulated brain maps with different profiles of spatial autocorrelation, simulated GOIs and realistic GOIs were used to calculate the probability of significance (Psig) for each statistical test. We found that a commonly used approach, resampling genes with the mean correlation as the test statistic, was associated with an inflated Psig that could be as high as 0.68, suggesting previous findings based on this method may need to be revisited. Moreover, null models built by spinning brain regions were superior to resampling genes or shuffling brain regions. The choice of test statistics also affected Psig with sign-insensitive test statistics such as mean absolute correlation or the number of significant correlations superior to mean correlation. The present study advocates for the consideration of co-expression, spatial autocorrelation, and test statistics when employing a non-parametric approach to test spatial associations between multiple transcriptional profiles and a single imaging derived phenotype.

Key words: transcriptional profiles; imaging-derived phenotypes; false positive rate; genes of interest

Introduction

The whole-brain gene expression maps of the Allen Human Brain Atlas (AHBA) have enabled investigation into the spatial association between ex vivo transcriptional and in vivo imaging derived patterns (Hawrylycz, et al., 2012). Linking regional transcriptional profiles and interregional variations in cortical measures has informed our understanding of the putative molecular architectures (e.g., biological process, cell-type specificity) underlying cortical phenotypes related to health and disease (Arnatkeviciute, et al., 2021; Norbom, et al., 2021). For instance, a popular integrative analysis that correlates a set of transcriptional profiles specific to molecular features of interest with a single imaging-derived phenotype shows the potential to reveal possible molecular architectures associated with brain development and disorders (Hess, et al., 2018; Parker, et al., 2020; Patel, et al., 2021; Patel, et al., 2020; Patel, et al., 2019; Pecheva, et al., 2020; Romme, et al., 2017; Shin, et al., 2018; Vidal-Pineiro, et al., 2020).

Non-parametric tests, in which a test statistic derived from an empirical model is tested against null models, have been widely used in examining the spatial correspondence between transcriptional and imaging data. Previous studies have shown that the presence of statistical dependence in spatial data (i.e., spatial autocorrelation of adjacent regions) and transcriptional data (i.e., co-expression of genes) largely contribute to the false-positive associations when employing the non-parametric statistical test (Fulcher, et al., 2021; Markello and Misic, 2021; Selvaggi, et al., 2021; Wei, et al., 2022). For instance, the spatial autocorrelation in brain maps contributed to the false positive findings when examining the topographical correspondences between two imaging maps (Markello and Misic, 2021). In addition, the co-expression among the transcriptional profiles can lead to false-positive associations between transcriptional and imaging profiles (Fulcher, et al., 2021). However, no study has examined the impact of these two factors on the non-parametric test used for determining the significance of correlations between multiple transcriptional profiles and a single imaging derived phenotype. Furthermore, it is unclear whether the averaged statistic, a commonly used test statistic in a non-parametric test, is optimal for this scenario.

The present study utilized the AHBA transcriptomic data that were mapped to the left Desikan-Killiany cortical regions and examined the performance of 9 types of statistical tests in assessing the spatial association between a set of transcriptional profiles and an imaging derived brain map. Specifically, we examined three types of null models (i.e., resampling genes, shuffling brain

regions, and spinning brain regions) in combination with three test statistics (i.e., mean correlation, mean strength of absolute correlations, number of significant correlations) that could be derived from the empirical distribution and compared against that of the null models. Two types of simulated brain maps (1,000 for each) with different profiles of spatial autocorrelation, and 9000 sets of simulated GOIs of varied set sizes were used to quantify the Psig for each statistical test. According to the previous studies (Fulcher, et al., 2021; Markello and Misic, 2021; Wei, et al., 2022), we anticipated that null models that failed to account for the co-expression and spatial information would inflate the Psig.

Methods

Transcriptional profiles

The Allen Human Brain Atlas (AHBA, http://www.brain-map.org) provides human postmortem brain gene expression maps employing whole-brain microarray surveys on six donors (Hawrylycz, et al., 2012). The transcriptional profiles were mapped to the Desikan-Killiany atlas using Freesurfer (French and Paus, 2015). As the right-hemisphere data were available for 2 donors only, the analysis in the present study was restricted to the 34 Desikan cortical regions of the left hemisphere. A conservative 2-stage filtering process was applied to ensure the consistency of gene-expression profiles (details and rationale for the threshold can be found in Shin, et al., 2018). First, the genes showing consistent transcriptional profiles across the 34 Desikan regions across the donors with donor-to-median correlation > 0.446 were retained. Second, the profiles of the genes that passed the first-stage filtering in AHBA were correlated to that of an independent atlas (i.e., BrainSpan) across 11 approximately matched cortical regions. Only genes that showed a correlation > 0.52 between the two atlases were retained, which resulted in a final set of 2511 genes used for the analysis. The transcriptional profiles, genes passing 2-stage filtering, and detailed description of the preprocessing are available in the previous studies (French and Paus, 2015; Shin, et al., 2018).

Statistical tests

Figure 1A shows one of the typical scenarios in which we examined the spatial association between a derived brain map and the transcriptomic profiles of a set of GOIs by correlating the brain phenotype with the transcriptional profile of each gene in the set. The empirical correlations were tested against different null models (i.e., resampling genes, shuffling brain regions, and spinning brain regions). For the null models built by resampling genes (**Figure 1B**), each null model (i.e., iteration) generated a distribution of correlations between the original brain phenotype and the transcriptional profiles of randomly selected genes. The number of random genes in each set was kept the same as that of the original GOIs. The hypothesized concern with this approach is that the randomly selected genes are unlikely to have expression profiles that are as similar to one another as those of the original GOIs. For the null models built by shuffling brain regions (**Figure 1C**), the regions in the brain map were randomly shuffled and then correlated with the transcriptomic profiles of the original GOIs. In this approach, the hypothesized concern is that the relationships among adjacent regions (i.e., spatial autocorrelation) are not preserved in the random shuffles. For the null models built by spinning brain regions (**Figure 1D**), the regions in the brain map were shuffled with spatial autocorrelation preserved (i.e., spinning) (Váša, et al., 2018), and then correlated with the transcriptional profiles of the original GOIs. These procedures were repeated 10,000 times for each type of null model, resulting in 10,000 null distributions of correlations respectively.

To determine if the brain map and the transcriptional profiles of the GOIs were significantly correlated, a test statistic can be derived from the empirical distribution and compared against that of the null models. Here, as shown in **Figure 1E**, we compared three test statistics: mean correlation, mean absolute correlation, and the number of significant correlations. Specifically, three relevant hypotheses were examined: 1) if the mean correlation in the empirical distribution exceeded 95% (greater than 97.5^{th} or smaller than 2.5^{th} percentile) of the mean correlations in the null models; 2) if the mean strength of absolute correlations (i.e., mean absolution correlation) was greater than that of 95% of null models; 3) if the number of significant correlations (FDR-corrected) was greater than that of 95% of null models. In total, nine statistical test approaches (i.e., three types of null models × three test statistics) were examined.

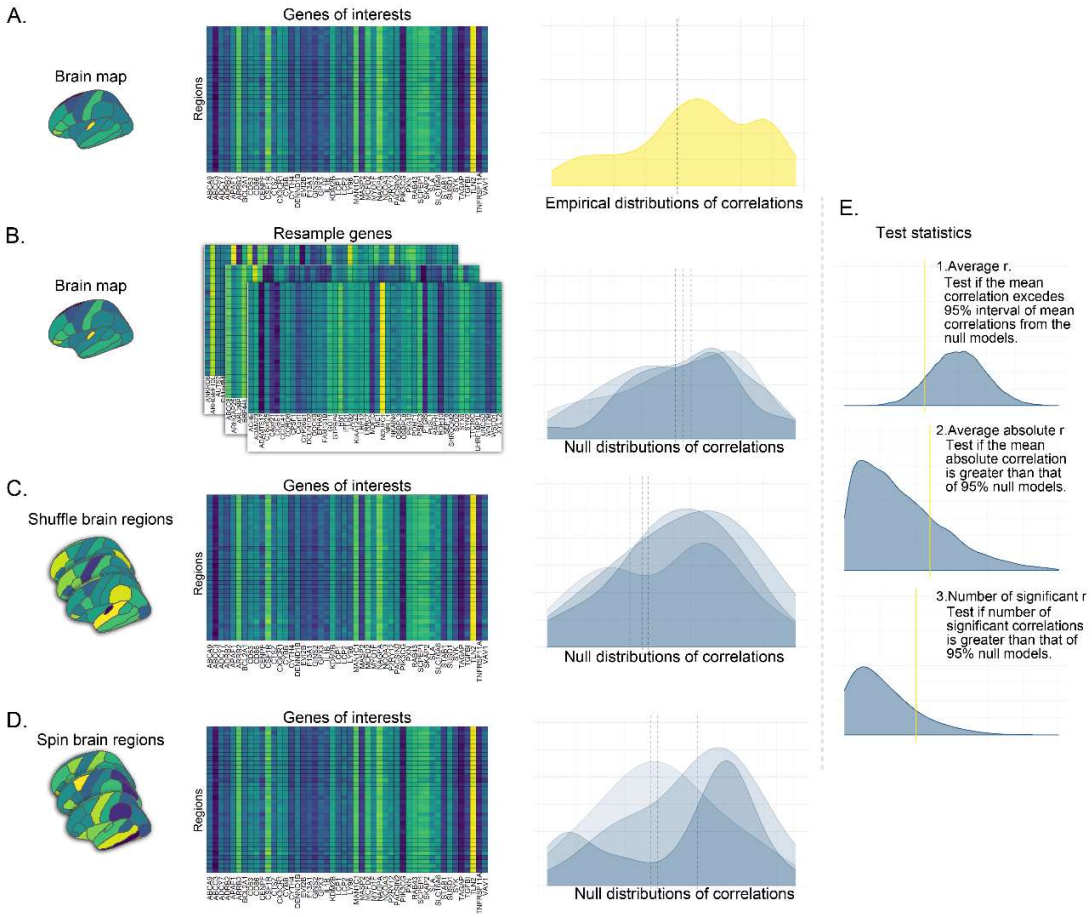


Figure 1. A. A typical scenario in which researchers examine the spatial association between a derived brain map and the transcriptomic profiles of a set of genes of interest (GOIs) by correlating the brain map with the transcriptional profile of each gene in the set. The empirical correlations were tested against different null models (i.e., resampling genes, shuffling brain regions, and spinning brain regions). **B.** Null models built by resampling genes. Each null model (i.e., iteration) generated a distribution of correlations between the original brain phenotype with the transcriptional profiles of randomly selected genes. The number of the random genes was kept the same as that of the original GOIs. **C.** Null models built by shuffling brain regions. The regions in the brain map were randomly shuffled and then correlated with the transcriptional profiles of the original GOIs. **D.** Null models built by spinning brain regions. The regions in the brain map were shuffled with spatial information preserved and then correlated with the transcriptional profiles of the original GOIs. **E.** Three test statistics were derived from the empirical distribution and compared against that of the null models.

Probability of significance

To quantify the performance of the statistical tests, the probability of observing significant correlations (Psig) between a set of GOIs and randomly generated brain maps was calculated as the proportion of the simulated brain maps with a significant correlation. Two types of random brain maps (Type1-Maps and Type2-Maps) with each containing 1,000 maps (as described below) were used to calculate the Psig of statistical tests. For each type, a set of simulated GOIs was tested against the 1,000 random brain maps using a certain statistical test, which resulted in a Psig value for the statistical test. For example, if a set of simulated GOIs was found significantly correlated with 680 out of 1,000 simulated brain maps using a certain statistical test (e.g., resampling genes with mean correlation as the test statistic), the Psig associated with the statistical test was calculated as 680/1000=0.68. The procedure was repeated for 9000 different sets of simulated GOIs (as described below), which yielded 9000 values of Psig per statistical test and type of random brain maps.

Simulation of GOIs

Sets of GOIs containing a varied number of genes (5 to 150 with an increment of 5, thus 30 sets in total) were created by randomly selecting genes from the full list of available genes (i.e., 2511 genes). To avoid sampling bias, the process was repeated 300 times, which resulted in 9000 different simulated sets of GOIs in total (i.e., 30 sets of simulated GOIs × 300 times). These randomly simulated GOIs may not represent realistic GOIs typically selected for their relation to specific molecular components (e.g., biological processes, cell types or the genetic risk of disorders) based on researchers' interests. Therefore, 79 sets of realistic GOIs that were specific to synaptic functions were tested in a supplementary analysis. The realistic GOIs were taken from the SynGO database where gene annotations were manually curated (Koopmans, et al., 2019).

Simulation of brain maps

The analysis was performed separately on two types of random brain maps. Each type contained 1,000 random brain maps and the spatial autocorrelation profiles differed between the two types. The first type of random brain map (Type1-Maps; n=1,000) was simulated by drawing each regional value from a uniform distribution ranging from -1 to 1. In this simulation, regional values

were independently sampled resulting in the Moran's I (a measure of spatial autocorrelation) centered around the expected value of no spatial autocorrelation (-0.03; see below). However, Moran's I derived from 12 effect size maps of cortical thickness differences associated with psychiatric and neurological disorders comprising a total of 14,886 cases and 20,962 controls from seven ENIGMA disease-related working groups as well as a cortical map of standard loadings of the first principal component derived from cortical thickness of 24,750 adult participants returned an average Moran's I of 0.03 (see **Table S1**), suggesting realistic brain maps were more likely to have a positive spatial autocorrelation. Therefore, the second type of random brain maps (Type2-Maps; n=1,000) with a more realistic spatial autocorrelation was simulated by randomly drawing a subset of 1,000 brain maps with Moran's I centered at 0.03 (standard deviation = 0.01) from a set of 500,000 uniformly random brain maps.

Spatial autocorrelation

The spatial autocorrelation in the brain map was calculated using Moran's I (Paradis and Schliep, 2019):

$$I = \frac{N}{W} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

$$E(I) = \frac{-1}{N-1}$$

, where *N* is the number of the regions (i.e., 34 Desikan regions) indexed by *i* and *j*, *x* is the variable of interest, w_{ij} is the matrix of spatial weights, and *W* is the sum of all w_{ij}. The weight matrix was set as the inverse Euclidean distance between the centroids of regions *i* and *j*. *E(I)* is the expected value of Moran's I under the null hypothesis of no spatial autocorrelation. In our cases, the expected value of Moran's I where there is no spatial autocorrelation was -1/34-1=-0.03. A Moran's I greater or smaller than the expected value indicates a positive or negative spatial autocorrelation.

Analytical strategies

Standard deviation (SD) and maximum value (MAX) of the Psig were reported for each statistical test. Pairwise F-tests on the variance of Psig were performed to quantify the differences between statistical tests. We anticipated that the null models built by resampling randomly selected genes would fail to account for the co-expression of GOIs in the empirical model. Therefore, supplementary analysis of the relationship between co-expression of GOIs and Psig was performed for the resampling genes approach. Co-expression was assessed by the averaged pairwise correlations among the transcriptional profiles in a set of GOIs. Linear models were used to examine the association between the co-expression and Psig. The *t* statistics associated with the co-expression were reported to quantify the contribution of the co-expression to the Psig.

The analysis was performed using R 3.6 and the Moran's I was calculated using R package *ape* (Paradis and Schliep, 2019). Computations were performed, in part, on the Vermont Advanced Computing Core. The code and data used for the analysis are available on https://github.com/zh1peng/paper_code.

Results

Type1-Maps

For Type1-Maps with Moran's I distributed around -0.03 (see left panel of **Figure 2**), the mean correlation yielded the largest SD (0.06) and MAX (0.55) of Psig for the resampling genes approach, and these were the largest statistics across the nine tests. Further, the Psig obtained by the resampling genes approach with mean correlation as the test statistic showed a strong positive correlation (t=48.06, p<0.001) with the magnitude of co-expression among the GOIs (see left panel of **Figure 2C**). When the mean absolute correlation was used as the test statistic for the resampling genes approach, the SD and Max of Psig decreased to 0.02 and 0.25, respectively. Testing the number of significant gene expression associations in the empirical distribution against the null model distributions yielded the smallest Psig SD and MAX. For the shuffling brain approach, the mean correlation yielded a smaller SD of 0.01 and MAX of 0.08 when compared to that of resampling genes. These two statistics were further decreased when the mean absolute correlation or number of significant correlations was used as the test statistic. Compared to the

shuffling brain approach, the spinning brain approach showed similar performance in controlling the Psig for the same test statistic.

Type2-Maps

For the Type2-Maps with Moran's I distributed around 0.03 and simulated GOIs (see right panel of **Figure 2**), the mean correlation for the resampling genes approach yielded the largest SD (0.08) and MAX (0.68) of Psig which, as was observed for Type1-Maps, were the largest statistics across the nine tests. As above, the Psig showed a strong positive correlation (t=43.07, p<0.001) with the co-expression of the GOIs (see right panel of **Figure 2C**), suggesting that GOIs with more co-expression tended to show more spurious but significant associations with a brain map when resampling genes and mean correlation were used. When the mean absolute correlation was used as the test statistic for the resampling genes approach, the SD and Max of Psig decreased to 0.03 and 0.32, respectively. Testing the number of significant gene expression associations in the empirical distribution against the null model distributions yielded the smallest Psig SD and MAX. For the shuffling brain approach, the mean correlation yielded an SD of 0.04 and a MAX of 0.25. These two values were increased if the mean absolute correlation or number of significant correlations was used as the test statistic. For the spinning brain approach, each test statistic showed a smaller SD and MAX of Psig compared to the same test statistic in the shuffling brain approach, suggesting that the spinning brain approach was better at controlling Psig than the shuffling brain approach for the same test statistic when there was more realistic spatial autocorrelation present in the random brain maps.

Descriptive summaries of the results are shown in **Figure S1** and **Table S2**. Results of the pairwise F tests on the variance are shown in **Figure S1**. The t statistics quantifying the contribution of the co-expression to the Psig of the resampling genes approach with mean correlation as the test statistic are shown in **Table S3**. In contrast to the common approach (i.e., randomly selected GOIs) which inflated false positives the present results showed that false positives could be much more successfully controlled with the spinning brain approach. The choice of test statistics also played a role in determining the Psig.

These observed patterns in simulated GOIs were replicated in realistic GOIs (see Supplementary analysis with realistic GOIs), suggesting the simulated GOIs were able to represent the realistic GOIs that were typically selected due to being related to some molecular architectures.

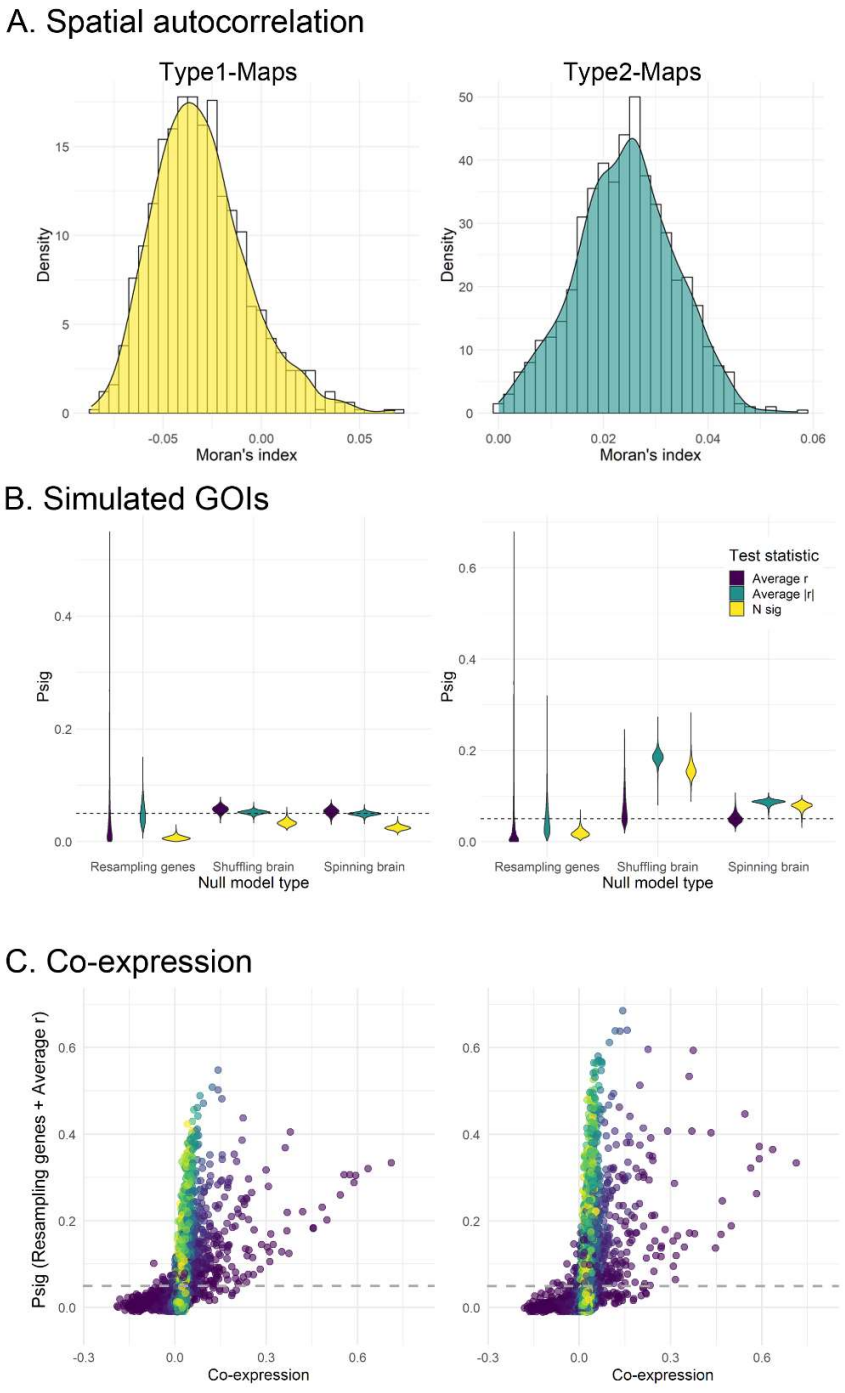


Figure 2. Simulation results. A. Spatial autocorrelation as measured by Moran's I for random brain maps (Left: Type1-Maps with Moran's Is distributed around 0.03; Right: Type2-Maps with Moran's Is distributed around -0.03) used for quantifying the probability of significance (Psig). **B.** Psig of the spatial correspondence between simulated genes of

interest (GOIs) and the random brain maps and for nine statistical tests (three types of null model × three test statistics). **C.** Scatter plot for Psig obtained using the resampling genes approach with mean correlation as the test statistic against the co-expression of the simulated GOIs. Each dot denotes the Psig of correlations between the random brain maps and each set of simulated GOIs, with the lighter color denoting the larger size of GOIs. Horizontal dashed line denotes Psig=0.05. To alleviate overplotting, a small random noise is added to the dots. Scatter plots for other statistical tests are provided in **Figure S2**.

Discussion

The present study examined the performance of nine possible non-parametric test approaches (three types of null models x three types of test statistics) in determining the spatial association between multiple transcriptional profiles and a single brain map using simulated GOIs, realistic GOIs, random brain maps, and random brain maps with more realistic profiles of spatial autocorrelation. Of note, resampling genes with mean correlation (the typical approach employed in the extant literature) yielded the largest Psig for both Type1-Maps and Type2-Maps. In the resampling genes approach, the empirical correlations with the original GOIs were tested against null distributions of correlations with randomly selected genes. The co-expression of the randomly resampled genes can be treated as sampled from an overall distribution of co-expression among all genes (i.e., 2511 genes in the current analysis), which cannot be guaranteed to match the co-expression in the original GOIs even though the number of genes was matched. A mismatch in co-expression between empirical and null models explained the inflation in Psig, which was supported by the positive correlation between the co-expression and Psig associated with the resampling genes and mean correlation approach. Testing the mean absolute correlation alleviated the problem of inflated Psig compared to the mean correlation, but the associated MAX of Psig remained as high as 0.32 in Type2-Maps. This suggests that testing the mean absolute correlation does not compensate for the failure to account for the co-expression in the null models. Interestingly, testing the number of significant correlations was less affected by the co-expression in the resampling genes, and yielded the lowest Psig. It is possible that only strong effects could survive the test when using the number of significant correlations, making it a conservative test statistic for the resampling genes approach. These results indicate that the resampling genes approach was suboptimal because it did not control for the co-expression of genes in null models. Apart from that, how the resultant multiple correlations were combined and tested (i.e., the choice of test statistics) affected the performance of resampling genes in controlling the Psig, especially testing mean correlation that was sensitive to the co-expression of

GOIs inflated the Psig. Therefore, findings obtained using the resampling genes and mean correlation approach need to be viewed with caution and possibly merit re-investigation.

A previous study has proposed to preserve the same degree of co-expression in the resampled genes by iteratively trying different gene sets until the co-expression was matched (Wei, et al., 2022). However, the iterative process can be computationally expensive. A good alternative to preserve the co-expression is to build null models based on null brain maps (e.g., shuffling brain regions) such that the GOIs are kept the same for each iteration of the null models, which has been used and proven effective in a previous study (Fulcher, et al., 2021). For the Type1-Maps with less spatial autocorrelation, both shuffling and spinning brain regions better controlled the Psig than resampling genes when the mean correlation or mean absolute correlation was tested. However, in the presence of the more realistic positive spatial autocorrelation in the Type2-Maps, shuffling brain regions without considering the spatial autocorrelation in the null models yielded greater Psig than that of spinning brain regions, indicating that accounting for the spatial autocorrelation in the null models using the spinning brain approach is vital in term of controlling Psig. These results complement previous studies showing the importance of controlling spatial autocorrelation when correlating two spatial maps (Markello and Misic, 2021). In addition, the differences in performance between the shuffling and spinning brain approaches only emerged in the Type2-Maps where positive spatial autocorrelation was present, which highlights the necessity of incorporating spatial autocorrelation when simulating imaging phenotypes.

Our results demonstrated that the statistical approach that bases null models on spinning brain regions can control for the co-expression and spatial autocorrelation in the null models and that this is superior to resampling genes or shuffling brain regions for controlling Psig. This result is consistent with previous findings showing that failure to account for statistical dependencies in the variable (e.g., co-expression and spatial autocorrelation) when building null models can violate the exchangeability assumption underlying the non-parametric test and lead to inflated Psig (Fulcher, et al., 2021; Markello and Misic, 2021; Wei, et al., 2022). In addition to these two factors, we demonstrated that the choice of the test statistic also played a role in determining Psig in the spinning brain approach. Compared with Type1-Maps, when realistic spatial autocorrelation was present in Type2-Maps, the Psig associated with the sign-insensitive test statistics (i.e., mean strength of correlations or number of significant correlations) showed a shift towards a mean Psig of 0.08, indicating they could be more sensitive to the spatial profiles of the

random brain maps. A previous study has demonstrated that Psig of spatial association between two brain maps increased as a function of spatial autocorrelation even if it was controlled for in the null models built by spinning brain regions (Markello and Misic, 2021). Such a result suggests that the overall increased Psig associated with the sign-insensitive test statistics in the spinning brain approach could be attributed to the presence of positive spatial autocorrelation. By contrast, testing mean correlation in the spinning brain approach appeared less sensitive to spatial profiles of the random brain maps, which could make it conservative for brain maps with high spatial autocorrelation.

In conclusion, with simulated GOIs and brain maps with different profiles of spatial autocorrelation, we found that resampling genes together with the usage of mean correlation as a test statistic was associated with an inflated Psig that could be as high as 0.68. This suggests that previous findings based on this method need to be viewed with caution. Moreover, null models built by spinning brain regions against which the original GOIs are tested, thereby accounting for both gene co-expression and spatial autocorrelation, were superior to resampling genes or shuffling brain regions. The choice of test statistics played a role in determining the Psig, in which sign-insensitive test statistics such as the mean absolute correlation or the significant number of significant correlations were superior to mean correlation. The present study advocates for the consideration of gene co-expression, spatial autocorrelation in the brain phenotype, and choice of test statistics when employing a non-parametric approach to test the spatial association between multiple transcriptional profiles and a single imaging-derived phenotype.

Acknowledgments

This work received support from NIDA grant R01DA047119.

Conflict of interest

All authors have no conflict of interest.

References

Arnatkeviciute, A., Fulcher, B.D., Bellgrove, M.A., Fornito, A. (2021) Imaging transcriptomics of brain disorders. Biological Psychiatry Global Open Science.

French, L., Paus, T. (2015) A FreeSurfer view of the cortical transcriptome generated from the Allen Human Brain Atlas. Frontiers in neuroscience, 9:323.

Fulcher, B.D., Arnatkeviciute, A., Fornito, A. (2021) Overcoming false-positive gene-category enrichment in the analysis of spatially resolved transcriptomic brain atlas data. Nature Communications, 12:1-13.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., Van De Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. Nature, 489:391-399.

Hess, J.L., Akutagava-Martins, G.C., Patak, J.D., Glatt, S.J., Faraone, S.V. (2018) Why is there selective subcortical vulnerability in ADHD? Clues from postmortem brain gene expression data. Molecular Psychiatry, 23:1787-1793.

Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijsouw, T., Coba, M.P., Cornelisse, L.N., Farrell, R.J., Goldschmidt, H.L., Howrigan, D.P. (2019) SynGO: an evidence-based, expert-curated knowledge base for the synapse. Neuron, 103:217-234. e4.

Markello, R.D., Misic, B. (2021) Comparing spatial null models for brain maps. NeuroImage, 236:118052.

Norbom, L.B., Ferschmann, L., Parker, N., Agartz, I., Andreassen, O.A., Paus, T., Westlye, L.T., Tamnes, C.K. (2021) New insights into the dynamic development of the cerebral cortex in childhood and adolescence: Integrating macro-and microstructural MRI findings. Progress in Neurobiology, 204:102109.

Paradis, E., Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35:526-528.

Parker, N., Patel, Y., Jackowski, A.P., Pan, P.M., Salum, G.A., Pausova, Z., Paus, T. (2020) Assessment of neurobiological mechanisms of cortical thinning during childhood and adolescence and their implications for psychiatric disorders. JAMA Psychiatry, 77:1127-1136.

Patel, Y., Parker, N., Shin, J., Howard, D., French, L., Thomopoulos, S.I., Pozzi, E., Abe, Y., Abé, C., Anticevic, A. (2021) Virtual histology of cortical thickness and shared neurobiology in 6 psychiatric disorders. JAMA Psychiatry, 78:47-63.

Patel, Y., Shin, J., Drakesmith, M., Evans, J., Pausova, Z., Paus, T. (2020) Virtual histology of multi-modal magnetic resonance imaging of cerebral cortex in young men. Neuroimage, 218:116968.

Patel, Y., Shin, J., Gowland, P., Pausova, Z., Paus, T., consortium, I. (2019) Maturation of the human cerebral cortex during adolescence: myelin or dendritic arbor? Cerebral Cortex, 29:3351-3362.

Pecheva, D., Lee, A., Poh, J.S., Chong, Y.-S., Shek, L.P., Gluckman, P.D., Meaney, M.J., Fortier, M.V., Qiu, A. (2020) Neural transcription correlates of multimodal cortical phenotypes during development. Cerebral Cortex, 30:2740-2754.

Romme, I.A., de Reus, M.A., Ophoff, R.A., Kahn, R.S., van den Heuvel, M.P. (2017) Connectome disconnectivity and cortical gene expression in patients with schizophrenia. Biological Psychiatry, 81:495-502.

Selvaggi, P., Rizzo, G., Mehta, M.A., Turkheimer, F.E., Veronese, M. (2021) Integration of human whole-brain transcriptome and neuroimaging data: Practical considerations of current available methods. Journal of Neuroscience Methods, 355:109128.

Shin, J., French, L., Xu, T., Leonard, G., Perron, M., Pike, G.B., Richer, L., Veillette, S., Pausova, Z., Paus, T. (2018) Cell-specific gene-expression profiles and cortical thickness in the human brain. Cerebral Cortex, 28:3267-3277.

Váša, F., Seidlitz, J., Romero-Garcia, R., Whitaker, K.J., Rosenthal, G., Vértes, P.E., Shinn, M., Alexander-Bloch, A., Fonagy, P., Dolan, R.J. (2018) Adolescent tuning of association cortex in human structural brain networks. Cerebral Cortex, 28:281-294.

Vidal-Pineiro, D., Parker, N., Shin, J., French, L., Grydeland, H., Jackowski, A.P., Mowinckel, A.M., Patel, Y., Pausova, Z., Salum, G. (2020) Cellular correlates of cortical thinning throughout the lifespan. Scientific Reports, 10:1-14.

Wei, Y., de Lange, S.C., Pijnenburg, R., Scholtens, L.H., Ardesch, D.J., Watanabe, K., Posthuma, D., van den Heuvel, M.P. (2022) Statistical testing in transcriptomic-neuroimaging studies: A how-to and evaluation of methods assessing spatial and gene specificity. Human Brain Mapping, 43:885-901.

Supplementary Materials

Supplementary Tables

Table S1. Moran's Is for 13 cortical thickness relevant maps from previous studies.

Brain maps	DOI	Source	Moran's I
SCZ vs. Control	https://doi.org/10.1016/j.biopsych.2018.04.023	Table S4a	0.033
Adult: OCD vs. Control	https://doi.org/10.1176/appi.ajp.2017.17050485	Table S4	-0.033
Adult: BD vs. Control	https://doi.org/10.1038/mp.2017.73	Table 1	0.037
Adult: MDD vs. Control	https://doi.org/10.1038/mp.2016.60	Table 1	0.020
CHR vs. Control	https://doi.org/doi:10.1001/jamapsychiatry.2021.0638	Table S9	0.006
EPI vs. control	https://doi.org/10.1093/brain/awx341	Table S31	0.025
Variance explained by Age	https://doi.org/10.1002/hbm.25364	Table S2	0.048
Age-CT Correlation (3-29 years)	https://doi.org/10.1002/hbm.25365	Table S3	0.050
Age-CT Correlation (30-59 years)	https://doi.org/10.1002/hbm.25366		0.061
Age-CT correlation (60-90 years)	https://doi.org/10.1002/hbm.25367		0.069
ALC vs. Control (ENIGMA)	Unpublished work	N/A	-0.001
ALC vs. Control (UK biobank)	Unpublished work	N/A	0.018
PC1 standard loadings derived from CT	Unpublished work	N/A	0.059

Average	0.030

CT: cortical thickness; BD: bipolar disorder; MDD: major depression; OCD: obsessive-compulsive disorder; EPI: epilepsy; CHR: clinical high risk for psychosis; ALC: alcohol dependence; PC1: the first principal component.

Table S2. Descriptive statistics for the probability of significant correlation between random brain maps and simulated genes of interest.

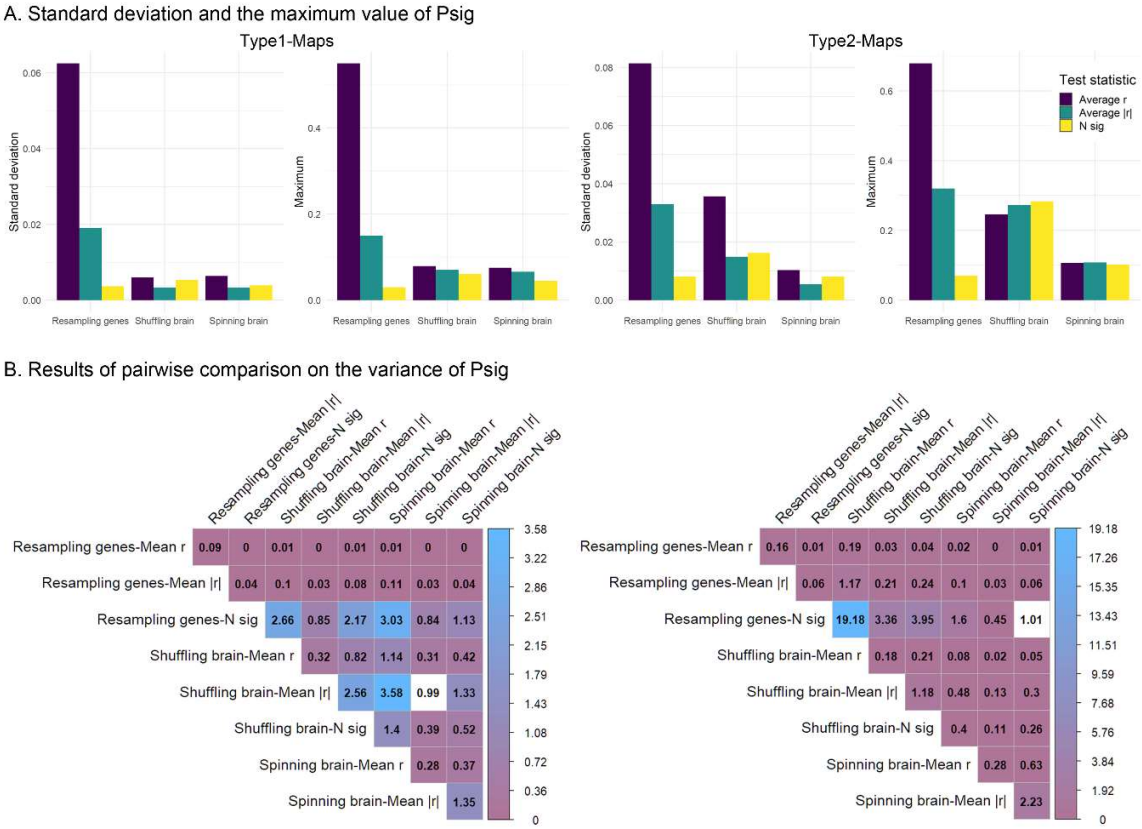
	Null Models	Resampling genes			Shuffling brain regions			Spinning brain regions		
	Test statistics	**Average r**	**Average |r|**	**Sig N**	**Average r**	**Average |r|**	**Sig N**	**Average r**	**Average |r|**	**Sig N**
Type1-Maps	**Mean**	0.05	0.05	0.01	0.06	0.05	0.03	0.05	0.05	0.02
	Standard Deviation	0.06	0.02	0.00	0.01	0.00	0.01	0.01	0.00	0.00
	Maximum	0.55	0.15	0.03	0.08	0.07	0.06	0.08	0.07	0.05
Type2-Maps	**Mean**	0.05	0.05	0.02	0.08	0.18	0.16	0.05	0.09	0.08
	Standard Deviation	0.08	0.03	0.01	0.04	0.01	0.02	0.01	0.01	0.01
	Maximum	0.68	0.32	0.07	0.25	0.27	0.28	0.11	0.11	0.10

Table S3. Correlation between co-expression and the probability of significant correlation between random brain maps and simulated genes of interest.

	Null Models	Resampling genes		
	Test statistics	**Average r**	**Average |r|**	**Sig N**
Type1-Maps	**t value**	48.06	5.59	0.90
	p value	0.00	0.00	0.37
Type2-Maps	**t value**	43.07	-2.80	-4.24
	p value	0.00	0.01	0.00

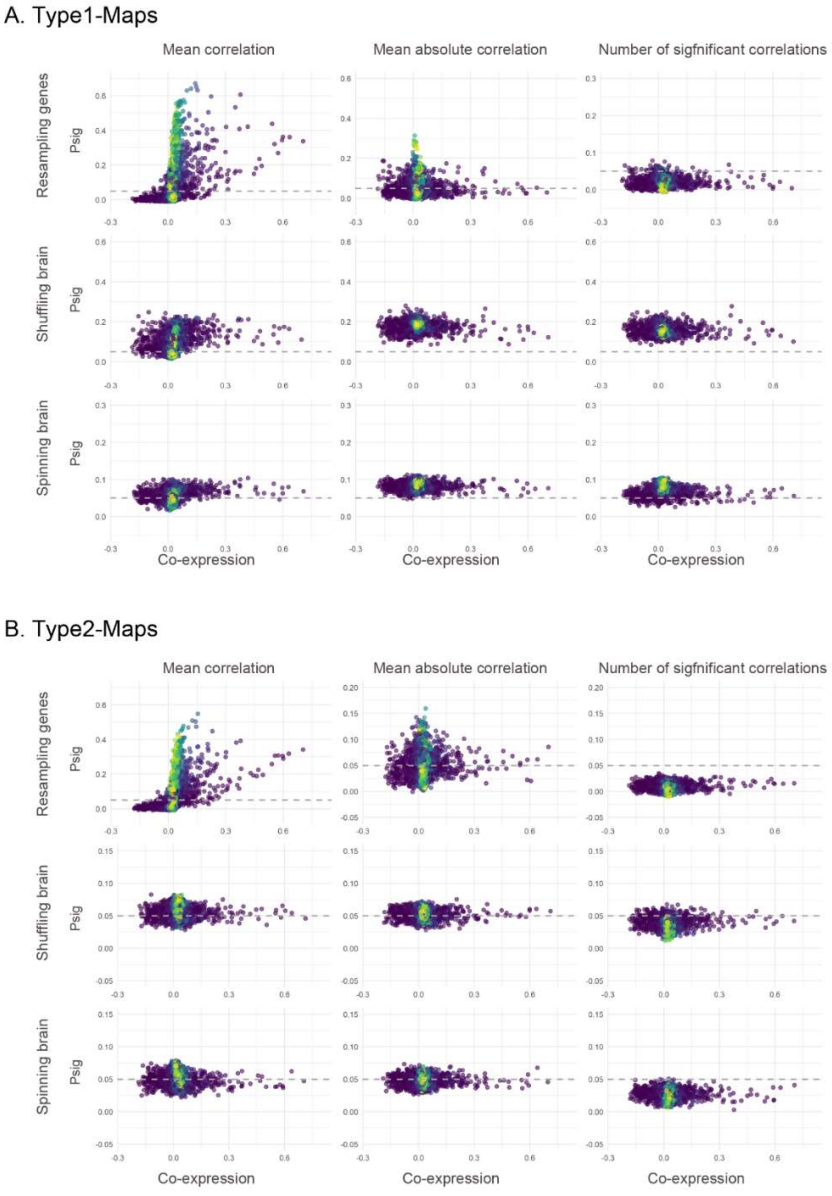
Supplementary Figures

Figure S1. A. Standard deviation and the maximum value of the probability of significant (Psig) correlation between random brain maps (Left: Type1-Maps; Right: Type2-Maps) and simulated genes of interest (GOIs). **B.** Results of pairwise comparison on the variance of Psig. The numbers in the cell indicate the *F* values resulted from an F test with the variance of left items set as the reference. An *F* value greater or smaller than 1 indicates increased or decreased variance in Psig



compared to that of the reference. Significant differences in the variance (Bonferroni-corrected $p<0.05$) are colored.

Figure S2. Scatter plots for the probabilities of significance (Psig) of correlations between random brain maps (A: Type1-Maps; B: Type2-Maps) and simulated genes of interest (GOIs) against the co-expression of the simulated GOIs. Each horizontal panel shows the results of different null models, and each vertical panel shows the results of different test statistics. Each dot denotes the Psig of correlations between random brain maps and each set of simulated GOIs, with the lighter color denoting the larger size of GOIs. Horizontal dashed line indicates Psig=0.05. To alleviate overplotting, a small random noise is added to the dots.



Supplementary analysis with realistic GOIs

The randomly simulated GOIs may not represent realistic GOIs typically selected due to being related to some molecular components based on researchers' interests. Therefore, a set of realistic GOIs that were specific to synaptic functions was tested. The realistic GOIs were taken from the SynGO database where gene annotations were manually curated (Koopmans, et al., 2019). The GOIs in the SynGO database with at least 5 genes available from the processed AHBA dataset (i.e., 2511 genes) were entered into the analysis, resulting in 79 sets of realistic GOIs that were entered into the analysis. Details of these GOIs are shown in **Table S4**.

The supplementary analysis with realistic GOIs showed similar patterns that were reported in the main analysis with simulated GOIs. The descriptive summaries are shown in **Figure S3, Figure S4 and Table S5**. Results of the pairwise F tests on the variance are shown in **Figure S3**. The t statistics quantifying the contribution of the co-expression to the Psig of the resampling genes approach with mean correlation as the test statistic are shown in **Table S6**. These results suggest the simulated GOIs were able to represent realistic GOIs that were typically selected due to being related to some molecular components.

Table S4. Details for 79 sets of genes of interest from the SynGO. CC: Cellular Component; BP: Biological Process; Orig N: number of genes in the original SynGO term; Avail N: number of genes available for the analysis.

GO term ID	GO domain	GO term name	Orig N	Avail N
GO:0097060	CC	synaptic membrane	17	5
GO:0098793	CC	presynapse	536	152
GO:0099523	CC	presynaptic cytosol	34	10
GO:0048786	CC	presynaptic active zone	91	29
GO:0098831	CC	presynaptic active zone cytoplasmic component	23	5
GO:0048787	CC	presynaptic active zone membrane	57	24
GO:0099059	CC	integral component of presynaptic active zone membrane	45	20
GO:0008021	CC	synaptic vesicle	120	37
GO:0030672	CC	synaptic vesicle membrane	102	29
GO:0030285	CC	integral component of synaptic vesicle membrane	52	19
GO:0098992	CC	neuronal dense core vesicle	32	14

GO:0042734	CC	presynaptic membrane	137	58
GO:0099026	CC	anchored component of presynaptic membrane	8	6
GO:0099056	CC	integral component of presynaptic membrane	117	47
GO:0043083	CC	synaptic cleft	17	9
GO:0098794	CC	postsynapse	624	160
GO:0099524	CC	postsynaptic cytosol	26	6
GO:0099571	CC	postsynaptic cytoskeleton	24	8
GO:0099572	CC	postsynaptic specialization	311	86
GO:0099634	CC	postsynaptic specialization membrane	45	18
GO:0099060	CC	integral component of postsynaptic specialization membrane	41	17
GO:0014069	CC	postsynaptic density	251	63
GO:0099092	CC	postsynaptic density,	42	9

		intracellular component		
GO:0098839	CC	postsynaptic density membrane	101	40
GO:0099061	CC	integral component of postsynaptic density membrane	82	34
GO:0045211	CC	postsynaptic membrane	115	47
GO:0099055	CC	integral component of postsynaptic membrane	97	42
SYNGO:postsyn_ribosome	CC	postsynaptic ribosome	70	5
SYNGO:presynprocess	BP	process in the presynapse	269	86
GO:0099509	BP	regulation of presynaptic cytosolic calcium levels	28	15
GO:0099626	BP	voltage-gated calcium channel activity involved in regulation of presynaptic cytosolic calcium levels	9	6

GO:0099505	BP	regulation of presynaptic membrane potential	44	26
GO:0099507	BP	ligand-gated ion channel activity involved in regulation of presynaptic membrane potential	24	10
GO:0099508	BP	voltage-gated ion channel activity involved in regulation of presynaptic membrane potential	17	13
GO:0099504	BP	synaptic vesicle cycle	189	54
GO:0016079	BP	synaptic vesicle exocytosis	73	33
GO:2000300	BP	regulation of synaptic vesicle exocytosis	35	15
GO:0031629	BP	synaptic vesicle fusion to presynaptic active zone membrane	8	6

GO:0031630	BP	regulation of synaptic vesicle fusion to presynaptic active zone membrane	7	5
GO:0016082	BP	synaptic vesicle priming	13	7
GO:0048488	BP	synaptic vesicle endocytosis	52	7
GO:0099525	BP	presynaptic dense core vesicle exocytosis	19	5
SYNGO:postsynprocess	BP	process in the postsynapse	218	72
GO:0099566	BP	regulation of postsynaptic cytosolic calcium levels	12	5
GO:0060078	BP	regulation of postsynaptic membrane potential	55	23
GO:1904315	BP	transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential	42	16

GO:0098962	BP	regulation of postsynaptic neurotransmitter receptor activity	18	8
GO:0099072	BP	regulation of postsynaptic membrane neurotransmitter receptor levels	121	36
GO:0099645	BP	neurotransmitter receptor localization to postsynaptic specialization membrane	44	19
GO:0098696	BP	regulation of neurotransmitter receptor localization to postsynaptic specialization membrane	22	5
GO:0098884	BP	postsynaptic neurotransmitter receptor endocytosis	38	10
GO:0099149	BP	regulation of postsynaptic neurotransmitter receptor endocytosis	30	8

GO:0099536	BP	synaptic signaling	193	67
GO:0099537	BP	trans-synaptic signaling	185	63
GO:0007268	BP	chemical synaptic transmission	160	56
GO:0050804	BP	modulation of chemical synaptic transmission	90	27
GO:0099531	BP	presynaptic process involved in chemical synaptic transmission	46	23
GO:0099171	BP	presynaptic modulation of chemical synaptic transmission	46	23
GO:0099565	BP	postsynaptic process involved in chemical synaptic transmission	30	7
GO:0099170	BP	postsynaptic modulation of chemical synaptic transmission	28	7
GO:0050808	BP	synapse organization	306	77
GO:0099173	BP	postsynapse organization	71	10

GO:0099175	BP	regulation of postsynapse organization	54	8
GO:0099563	BP	modification of synaptic structure	34	8
GO:0099010	BP	modification of postsynaptic structure	33	8
GO:0098885	BP	modification of postsynaptic actin cytoskeleton	20	6
GO:0050807	BP	regulation of synapse organization	29	6
GO:0098918	BP	structural constituent of synapse	35	14
GO:0099186	BP	structural constituent of postsynapse	25	10
GO:0099560	BP	synapse adhesion between pre- and post-synapse	37	17
GO:0007416	BP	synapse assembly	93	24
GO:0099054	BP	presynapse assembly	17	5
GO:0051963	BP	regulation of synapse assembly	50	12
GO:1905606	BP	regulation of presynapse assembly	31	9

GO:0098698	BP	postsynaptic specialization assembly	32	8
GO:0098883	BP	synapse disassembly	8	5
GO:1905806	BP	regulation of synapse disassembly	8	5
SYNGO:metabolism	BP	metabolism	94	7
SYNGO:transport	BP	transport	36	7

475

476

Table S5. Descriptive statistics for the probability of significant correlation between 79 realistic genes of interest taken from the SynGO database and random brain maps.

	Null Models	Resampling genes			Shuffling brain regions			Spinning brain regions		
	Test statistics	**Average r**	**Average |r|**	**Sig N**	**Average r**	**Average |r|**	**Sig N**	**Average r**	**Average |r|**	**Sig N**
Type1-Maps	**Mean**	0.17	0.07	0.01	0.06	0.05	0.04	0.05	0.05	0.03
	Standard Deviation	0.14	0.03	0.01	0.01	0.01	0.01	0.01	0.00	0.00
	Maximum	0.57	0.16	0.03	0.07	0.07	0.05	0.06	0.06	0.04
Type2-Maps	**Mean**	0.17	0.07	0.03	0.10	0.17	0.16	0.06	0.08	0.07
	Standard Deviation	0.16	0.04	0.01	0.04	0.03	0.03	0.01	0.01	0.01
	Maximum	0.63	0.17	0.07	0.27	0.23	0.23	0.10	0.10	0.09

Table S6. Correlation between co-expression and probability of significant correlation between 79 realistic genes of interest taken from the SynGO database and random brain maps.

	Null Models	resampling genes		
	Test statistics	**Average r**	**Average |r|**	**Sig N**
Type1-Maps	**t value**	5.62	2.88	3.87
	p value	0.00	0.01	0.00
Type2-Maps	**t value**	4.74	2.21	1.65
	p value	0.00	0.03	0.10

Figure S3. A. Probability of significant correlation between 79 realistic genes of interest (GOIs) taken from the SynGO database and random brain maps (Left: Type1-Maps; Right: Type2-Maps). **B.** Standard deviation and the maximum value of the Psig. **C.** Results of pairwise comparison on the variance of Psig. The numbers in the cell indicate the *F* values resulted from an F test with the variance of left items set as the reference. An *F* value greater or smaller than 1 indicates increased or decreased variance in Psig compared to that of the reference. Significant differences in the variance (Bonferroni-corrected $p<0.05$) are colored.

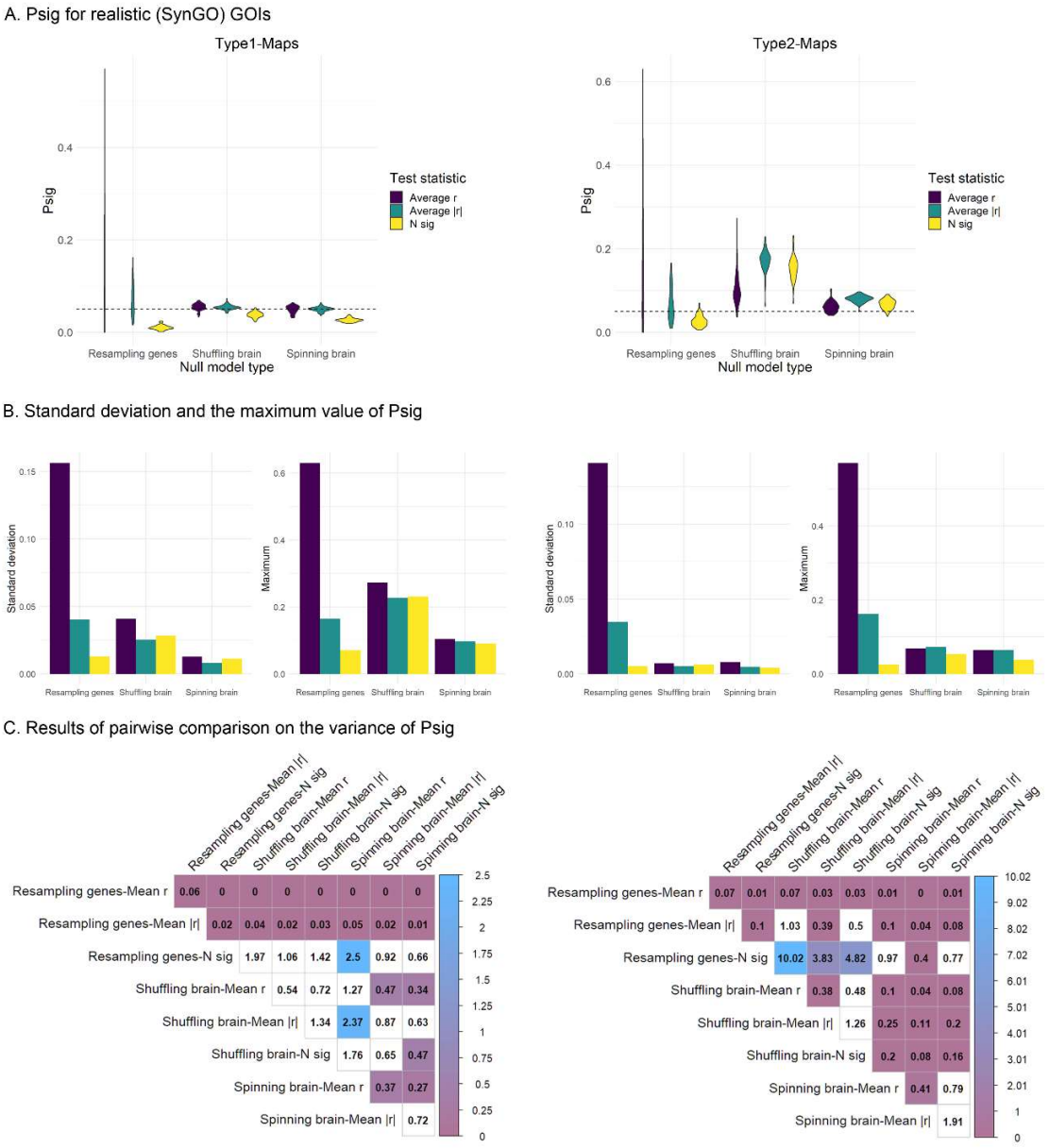


Figure S4. Scatter plots for the probabilities of significance (Psig) of correlations between random brain maps (A: Type1-Maps; B: Type2-Maps) the SynGO genes of interest (GOIs) against the co-expression of the SynGO GOIs. Each horizontal panel shows the results of different null models, and each vertical panel shows the results of different test statistics. Each dot denotes the Psig of correlations between random brain maps and each set of SynGO GOIs, with the lighter color denoting the larger size of GOIs. Horizontal dashed line denotes Psig=0.05. To alleviate overplotting, a small random noise is added to the dots.