HHS Public Access

Simultaneous Covariance Inference for Multimodal Integrative Analysis

Yin Xia[a], **Lexin Li**[b], **Samuel N. Lockhart**[c], **William J. Jagust**[d]

[a]Department of Statistics, School of Management, Fudan University, Shanghai, China

[b]Department of Biostatistics and Epidemiology, Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA

[c]Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC

[d]Lawrence Berkeley National Laboratory and School of Public Health, HelenWills Neuroscience Institute, University of California at Berkeley, Berkeley, CA

Abstract

Multimodal integrative analysis fuses different types of data collected on the same set of experimental subjects. It is becoming a norm in many branches of scientific research, such as multi-omics and multimodal neuroimaging analysis. In this article, we address the problem of simultaneous covariance inference of associations between multiple modalities, which is of a vital interest in multimodal integrative analysis. Recognizing that there are few readily available solutions in the literature for this type of problem, we develop a new simultaneous testing procedure. It provides an explicit quantification of statistical significance, a much improved detection power, as well as a rigid false discovery control. Our proposal makes novel and useful contributions from both the scientific perspective and the statistical methodological perspective. We demonstrate the efficacy of the new method through both simulations and a multimodal positron emission tomography study of associations between two hallmark pathological proteins of Alzheimer's disease.

Keywords

1. Introduction

Multimodal integrative analysis is now becoming a norm in many branches of scientific research. It uses different physical and physiological sensitivities of machines and technologies, and acquires different types of data for a common set of experimental subjects.

Integrative analysis fuses such diverse but often complementary information, borrows strength across multiple datasets, and renders an integrated data resolution that would otherwise not be available with any single data type. One example of multimodal analysis is multi-omics study, where gene expressions, DNA copy number alternations, DNA methylation changes, and other genetic information are jointly collected and analyzed from the same biological samples (Shen, Wang, and Mo 2013; Liu et al. 2014; Li et al. 2014; Cai, Cai, and Zhang 2016). Another example is multimodal neuroimaging analysis, where distinct brain characteristics, from brain structure and function to numerous chemical constituents, as well as genetic information, are simultaneously measured and aggregated from the same study subjects (Zhang et al. 2011; Uludag and Roebroeck 2014; Lin et al. 2014; Nathoo, Kong, and Zhu 2017).

In this article, we address the problem of simultaneous covariance inference of associations between multiple modalities, which is of a vital interest in multimodal integrative analysis. Our motivation is a multimodal positron emission tomography (PET) study. The goal is to understand the associations between two hallmark pathological proteins of Alzheimer's disease (AD), beta-amyloid (Aβ), and tau. Both proteins are believed to be part of the driving mechanism of AD, and are common in the brains of not only AD subjects but also those in late life in the absence of dementia. The two proteins are thought to be highly associated in terms of the precise spatial and temporal patterns of their accumulation (Braak and Braak 1991). With the advent of PET radioligands, such as Pittsburgh compound-B (PiB) and AV-1451, it is now feasible to detect both Aβ and tau, respectively, in the brains of living cognitively normal and demented older adults (Chien et al. 2014). Studies combining such tracers suggest that globally increasing Aβ burden is associated with greater neocortical accumulation of tau (Brier et al. 2016; Lockhart et al. 2017). However, the specific associations of regional Aβ and tau deposition remain unclear, and require further investigation to identify brain regions with significant Aβ and tau correlations. In particular, a pressing challenge is the examination of nonlocal associations, that is, correlations between different regions of the brain across different radioligands.

The data consist of multiple subjects. For each subject, two imaging modalities are measured, one for Aβ and the other for tau. Each modality is represented by a vector of quantities, and each entry measures the amount of the protein at a location, or voxel, of the brain. The brain is further parcellated into a set of regions-of-interest following a prespecified brain atlas. Brain parcellation is particularly useful to facilitate the interpretation, and has been frequently employed in brain imaging analysis (Fornito, Zalesky, and Breakspear 2013; Ahn et al. 2015; Kang et al. 2016, among many others). Numerous brain atlases are available, and we adopt the most commonly used automated anatomical labeling atlas (Tzourio-Mazoyer et al. 2002). Furthermore, multiple brain regions are often grouped into functional modules. Each module possesses a relatively autonomous functionality, and complex brain tasks are believed to perform through coordinated collaborations among the modules (Smith et al. 2009; Yeo et al. 2011). In our multimodal study, we focus on the region level analysis, that is, we seek significantly correlated region pairs by examining correlations between voxels from two regions. We also carry out a module level analysis, that is, we seek significantly correlated functional module pairs by examining correlations between regions from two modules.

The search of brain regions or modules where multiple imaging modalities are significantly correlated with each other can be formulated as a simultaneous covariance inference problem. Specifically, for subject $k = 1,\ldots, n$, let X_k denote a p-dimensional vector representing one imaging modality, for example, Aβ, and Y_k denote another p-dimensional vector for the second modality, for example, tau. For notational simplicity, we assume the same dimensionality p for X_k and Y_k, but our test method does not require this assumption. For the region level analysis, the entry of X_k and Y_k represents the protein measure at each voxel, whereas for the module level analysis, the entry represents a summary protein measure at each brain region. Let $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_L\}$ denote the partition of the brain into L regions or modules. The corresponding sizes of the partitions are $p_1, p_2, \ldots, p_L$, respectively, and $\sum_{l=1}^{L} p_l = p$. Stack the two vectors together, $Z_k = (X_k^\mathsf{T}, Y_k^\mathsf{T})^\mathsf{T} \in \mathbb{R}^{2p \times 2p}$, and denote its mean by $\mu \in \mathbb{R}^{2p}$, and the covariance $\Sigma = (\sigma_{i,j})_{2p \times 2p}$. Our goal is to identify the pairs of brain regions or functional modules where (X_k, Y_k) are significantly correlated. That is, we aim at simultaneous testing of hypotheses,

$$H_{0,l,g} : \Sigma_{\mathcal{S}_l^X, \mathcal{S}_g^Y} = 0, \quad \text{versus} \quad H_{1,l,g} : \Sigma_{\mathcal{S}_l^X, \mathcal{S}_g^Y} \neq 0, \tag{1}$$
$$\text{for all } 1 \leq l, g \leq L,$$

where $\Sigma_{\mathcal{S}_l^X, \mathcal{S}_g^Y}$ represents the covariance matrix between the subvector of the modality X_k corresponding to the region $\mathcal{S}_l$ and the subvector of Y_k corresponding to $\mathcal{S}_g$. We propose a simultaneous testing procedure for the simultaneous inference problem (1), with a proper false discovery control. Recognizing that there is no readily available solution in the literature when the region sizes $\{p_1, p_2, \ldots, p_L\}$ are much larger than the sample size n and can diverge along with n, we first construct a test statistic based on the maximum of individual pairs of entries of two regions. We next derive the limiting distribution of the test statistic, and show that, for the test of a single pair of regions based on this limiting distribution, the asymptotic power is minimax rate optimal. We further introduce a novel correction method so to overcome the slow convergence rate of the limiting distribution, as well as a normal quantile transformation so to unify the cut-off values for different pairs of regions. Both enhance the test, and are crucial for the subsequent multiple testing. Finally, we develop a new simultaneous testing procedure for multiple pairs of regions under dependence, and show that it can control false discovery asymptotically.

Our proposal makes novel and useful contributions from both the scientific perspective and the statistical methodological perspective. Scientifically, the multimodal association analysis such as in our motivating example is biologically very useful. For instance, for our multimodal PET study, a rigorous quantification of associations between Aβ and tau would offer solid confirmation and support to prior biological hypotheses about how pathological proteins of AD interact in the aging or demented human brain. In turn, it could enable more accurate prediction of individual subjects demonstrating in vivo neuropathology, and allow better design and subject recruitment of clinical trials to potentially block the spread of AD. More generally, such an association analysis is of a broad scientific interest for a wide range of multimodal integrative analysis, for example, joint analysis of functional magnetic resonance imaging and diffusion tensor imaging (Zhu et al. 2014), joint analysis of

functional magnetic resonance imaging and electroencephalography (Singh, Kim, and Kim 2003), as well as multi-omics association analysis (Richardson, Tseng, and Sun 2016). On the other hand, there is almost no existing solution in the literature that can produce a rigorous and explicit significance quantification for this important type of inference problem. Our proposal, thus, offers a timely response, and would clearly benefit such analyses, by providing a much improved detection power as well as a rigid false discovery control. In addition, by obtaining the asymptotic properties of the test procedure, instead of resorting to a more expensive permutation type test, our proposal offers a computationally sensible solution for the extremely high-dimensional neuroimaging and biomedical data.

Methodologically, this subcovariance inference problem we study is itself of independent interest, but has so far received little attention in the statistics literature. A related family of solutions is sparse canonical correlation analysis (CCA) (Witten, Tibshirani, and Hastie 2009; Chen et al. 2013). Though it is useful to help identify highly correlated regions from multiple modalities, sparse CCA focuses on the estimation, rather than the inference, aspect of the problem. As such, it does not provide an explicit significance quantification of the correlation between a pair of regions, and does not explicitly control the false discovery. In Section 4, we numerically compare our testing method with sparse CCA. For the inference based solutions, although the multiple testing problem under a dependence structure has been extensively studied (Benjamini and Yekutieli 2001; Efron 2007; Sun and Cai 2009; Sun et al. 2015), the existing methods were either too conservative, or required some stringent dependency assumptions, and thus are not suitable to test the covariance structure. There are only a handful of related solutions about covariance inference. The most relevant method to our proposal is Xie and Kang (2017), who also studied simultaneous inference of sub-covariance matrices. However, their procedure only guaranteed the family wise error rate control, instead of false discovery control, and was less powerful than our solution, as we show numerically in Section 4. Another related solution is Xia, Cai, and Cai (2018), who proposed a test for the submatrices inside a precision matrix. However, a sum-of-square type test statistic was employed and the dimensions of the submatrices had to remain fixed. By contrast, we aim at submatrices of the covariance matrix, use a maximum type test statistic, and allow the submatrix dimensions to diverge. It is noteworthy that there are also a category of testing procedures that are relevant but *not* addressing exactly the same type of inference problem as ours. One family in this category targeted covariance or correlation matrices. Specifically, Cai, Liu, and Xia (2013) tested the equality of two covariance matrices globally, while Cai and Liu (2016) tested the equality of individual entries of the two correlation matrices with false discovery control. Xia (2017) further extended to more than two populations. We recognize that the test statistic we develop in Section 2 and its limiting distribution in Theorem 1 are closely related to those of Cai, Liu, and Xia (2013). This limiting distribution, however, has a slow convergence rate, and would induce an approximation error. This error is negligible in Cai, Liu, and Xia (2013), since their focus is testing two covariance matrices. Our target, on the other hand, is the covariances of the pairs of regions between the two modalities, and there are many, and sometimes diverging number of pairs to test simultaneously. As a result, the approximation errors would accumulate, which in turn would severely affect the subsequent multiple testing error rate control. To address this issue, we propose a novel correction method that leads to a more accurate

approximation of the limiting distribution. Our proposal is thus notably different from Cai, Liu, and Xia (2013), and the theoretical investigation involved is far from a trivial extension. Similarly, another family in this category targeted precision or partial correlations matrices (Liu 2013; Xia, Cai, and Cai 2015; Xia and Li 2017), but again they can not address our testing problem. Our proposal makes a useful addition to the general toolbox of covariance inference.

The rest of the article is organized as follows. Section 2 develops the test for a given pair of regions, and Section 3 develops the simultaneous testing procedure for all pairs. For each, we begin with the proposed testing procedure, then study the corresponding theoretical properties. Section 4 reports the simulation results, and Section 5 revisits the motivating multimodal PET study. Section 6 concludes with a discussion. The supplementary appendix collects additional simulation results and the proofs.

2. Test for a Pair of Regions

We first develop a test for the hypotheses, $H_{0,l,g} : \mathbf{\Sigma}_{\mathcal{S}_l^X, \mathcal{S}_g^Y} = 0$, versus $H_{1,l,g} : \mathbf{\Sigma}_{\mathcal{S}_l^X, \mathcal{S}_g^Y} \neq 0$, for a given pair of regions $1 \le l, g \le L$. We next study its theoretical properties in terms of the limiting null distribution and the asymptotic power of the test. We also propose an intermediate correction and normal quantile transformation, both of which are novel and important for the subsequent simultaneous covariance inference.

2.1. Testing Procedure

For two sets of p-dimensional random samples $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^n$, define the sample covariance matrix as $\widehat{\mathbf{\Sigma}} = (\hat{\sigma}_{i,j})_{p \times p} = n^{-1} \sum_{k=1}^n (\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{Y}_k - \bar{\mathbf{Y}})^\mathsf{T}$, where $\bar{\mathbf{X}} = n^{-1} \sum_{k=1}^n \mathbf{X}_k$, and $\bar{\mathbf{Y}} = n^{-1} \sum_{k=1}^n \mathbf{Y}_k$. To account for heteroscedasticity of the estimates $\{\hat{\sigma}_{i,j}\}_{i,j=1}^p$, define $\theta_{i,j} = \mathrm{var}\{(X_{k,i} - \mu_{1,i})(Y_{k,j} - \mu_{2,j})\}$. The variance of $\hat{\sigma}_{i,j}$ can be estimated by

$$\hat{\theta}_{i,j} = \frac{1}{n} \sum_{k=1}^n \left\{ (X_{k,i} - \bar{X}_i)(Y_{k,i} - \bar{Y}_i) - \hat{\sigma}_{i,j} \right\}^2, \quad 1 \le i, j \le p,$$

where $\bar{X}_i = n^{-1} \sum_{k=1}^n X_{k,i}$, $\bar{Y}_i = n^{-1} \sum_{k=1}^n Y_{k,i}$. Correspondingly, we consider the standardized statistic,

$$T_{i,j} = \frac{\hat{\sigma}_{i,j}}{(\hat{\theta}_{i,j} / n)^{1/2}}, \quad 1 \le i, j \le p.$$

To test the null hypothesis $H_{0,l,g} : \mathbf{\Sigma}_{\mathcal{S}_l^X, \mathcal{S}_g^Y} = 0$, it is equivalent to testing whether the entries in the submatrix $\mathbf{\Sigma}_{\mathcal{S}_l^X, \mathcal{S}_g^Y}$ are all equal to 0, which, in turn, is equivalent to testing whether $\max_{i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y} \sigma_{i,j}^2 = 0$. We then construct the following test statistic,

$$M_{l,g} = \max_{i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y} T_{i,j}^2 .$$

We show in the next section that, under the null hypothesis, $M_{l,g} - 2\log(p_l p_g) + \log\log(p_l p_g)$ converges weakly to a Gumbel random variable with the distribution function $\exp\{-\pi^{-1/2}e^{-t/2}\}$. Accordingly, we define the α-level test by,

$$\Psi_\alpha^{(l,g)} = I\{M_{l,g} \geq q_\alpha + 2\log(p_l p_g) - \log\log(p_l p_g)\},$$

where q_α is the $1 - \alpha$ quantile of the Gumbel distribution. We reject the null hypothesis $H_{0,l,g}$ if $\Psi_\alpha^{(l,g)} = 1$.

2.2. Theory

We next study the limiting null distribution of the test statistic $M_{l,g}$, and the asymptotic power of the test $\Psi_\alpha^{(l,g)}$. We first introduce some notations and technical assumptions.

Denote by $\mathbf{X}_{k,\mathcal{S}_l}$ the subset of X_k for the region $\mathcal{S}_l$, and $\mathbf{Y}_{k,\mathcal{S}_g}$ the subset of Y_k for $\mathcal{S}_g$. Let $\mathbf{Z}_k^{(l,g)} = (\mathbf{X}_{k,\mathcal{S}_l}^\mathsf{T}, \mathbf{Y}_{k,\mathcal{S}_g}^\mathsf{T})^\mathsf{T}$. Denote by $\boldsymbol{\mu}^{(l,g)}$ and $\boldsymbol{\Sigma}^{(l,g)} = \boldsymbol{\Sigma}_{\mathcal{S}_{l,g},\mathcal{S}_{l,g}} \in \mathbb{R}^{p_{l,g} \times p_{l,g}}$ the mean and covariance matrix of $\mathbf{Z}_k^{(l,g)}$, where $\mathcal{S}_{l,g} = \mathcal{S}_l \cup \mathcal{S}_g$ and $p_{l,g} = p_l + p_g$. Define the cardinality of the set of indices that are highly correlated with $i \in \mathcal{S}_{l,g}$ by $s_i^{(l,g)}(\alpha_0) = \mathrm{card}\{i' \in \mathcal{S}_{l,g} : |\rho_{i,i'}| \geq (\log p_{l,g})^{-1-\alpha_0}\}$ for some $\alpha_0 > 0$. Denote the cardinality of a set $\mathcal{A}$ by $|\mathcal{A}|$. We introduce the following assumptions.

(A1) *Weak Dependence Condition:* There exists a subset D_0 of $\mathcal{S}_{l,g}$, $1 \leq l, g \leq L$, with $|D_0| = o(\min\{p_l, p_g\})$ and a constant $\alpha_0 > 0$, such that, for all $\gamma > 0$, $\max_{i \in \mathcal{S}_{l,g} \setminus D_0} s_i^{(l,g)} = o(p_{l,g}^\gamma)$. Furthermore, assume that $|\rho_{i,i'}| \leq r < 1$ for some $r > 0$ and $i, i' \in \{\mathcal{S}_{l,g}, 1 \leq l, g \leq L\}$.

(A2) *Moment Condition:* Assume one of the following two conditions hold.

    a. *Sub-Gaussian Tail:* Assume that $\max_{1 \leq l,g \leq L} \log(p_{l,g}) = o(n^{1/5})$. There exist some constants $\eta > 0$ and $C > 0$, such that, $\mathbf{E}\exp[\eta\{Z_{k,i}^{(l,g)} - \mu_i\}^2 / \sigma_{i,i}] \leq C$, for all $1 \leq l, g \leq L$ and $i \in \mathcal{S}_{l,g}$.

    b. *Polynomial Tail:* Assume that $\max_{1 \leq l,g \leq L} p_{l,g} \leq c_1 n^{\gamma_0}$ for some constants $\gamma_0, c_1 > 0$. Moreover, assume that $\mathsf{E}|(Z_{k,i}^{(l,g)} - \mu_i)/(\sigma_{i,i})^{1/2}|^{4\gamma_0+4+\epsilon} \leq C$, for some constants $\epsilon, C > 0$, for all $1 \leq l, g \leq L$ and $i \in \mathcal{S}_{l,g}$.

(A3) *Elliptically Contoured Condition:* Assume that for some κ $1/3$, for any 1 l, g L and $i, j, i', i', j' \in \mathcal{S}_{l,g}$,

$$\mathsf{E}(Z_{k,i}^{(l,g)} - \mu_i)(Z_{k,j}^{(l,g)} - \mu_j)(Z_{k,i'}^{(l,g)} - \mu_{i'})(Z_{k,j'}^{(l,g)} - \mu_{j'}) = \kappa(\sigma_{i,j}\sigma_{i',j'} + \sigma_{i,i'}\sigma_{j,j'} + \sigma_{i,j'}\sigma_{j,i'})$$

Assumption (A1) is a mild condition on the number of highly correlated entries in $\mathcal{S}_{l,g}$, and this condition is satisfied if the eigenvalues of $\Sigma^{(l,g)}$ are bounded from above. Without this weak dependency assumption, we can still approximately control the size of our test; see our discussion in Section 6. Assumption (A2) is the commonly used moment condition. Assumption (A3) holds for the elliptically contoured distributions (Anderson 2003). The following theorem shows that, for each 1 l, g L, we have $M_{l,g} - 2 \log(p_l p_g) + \log$ $\log(p_l p_g)$ converges weakly to a Gumbel random variable with the distribution function $F(t)$ $= \exp\{-\pi^{-1/2}e^{-t/2}\}$.

Theorem 1. Assume (A1)–(A3) hold. Then under the null hypothesis $H_{0,l,g}$, we have, as n, $p_{l,g} \to \infty$, for any $t \in \mathbb{R}$,

$$\mathsf{P}(M_{l,g} - 2\log(p_l p_g) + \log\log(p_l p_g) \le t) \to \exp\{-\pi^{-1/2}e^{-1/2}\}.$$

Furthermore, the above convergence holds uniformly for all 1 l, g L and all $\mathbf{X}_{k,\mathcal{S}_l}, \mathbf{Y}_{k,\mathcal{S}_g}$ satisfying (A1)–(A3) and $\Sigma_{\mathcal{S}_l^X, \mathcal{S}_g^Y} = 0$.

We outline the key steps of the proof of this theorem. Specifically, we truncate the random variables and use a normal approximation to translate the problem into deriving the limiting null distribution of normal random variables under the same dependence structure. We further divide the sets of pairs of random variables $\{(i, j) : i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y\}$ into small subsets for separate analyses. We show that the behavior of weakly correlated random variables dominates the rest, and the corresponding extreme value behavior is asymptotically the same as the maximum of independent normal random variables. A detailed proof is given in the supplementary appendix.

We next study the asymptotic power and the optimality of the test $\Psi_\alpha^{(l,g)}$ based on the above theorem. Define the class, $\mathcal{U}(c) = \{\Sigma : \max_{i \in \mathcal{S}_l, j \in \mathcal{S}_g} |\sigma_{i,j}| / \theta_{i,j}^{1/2} \ge c\sqrt{\log p_{l,g}/n}\}$, which includes all covariance matrices with one of the entries in the submatrix $\Sigma_{\mathcal{S}_l, \mathcal{S}_g}$ having the standardized magnitude exceeding $c\sqrt{\log p_{l,g}}$. Denote by $\mathcal{P}$ the collection of distributions satisfying (A2). Let $\mathcal{T}_\alpha$ be the set of α-level tests over $\mathcal{P}$, that is, $\mathsf{P}(T_\alpha = 1)$ α under $H_{0,l,g}$ over all distributions in $\mathcal{P}$ for any $\mathcal{T}_\alpha \in \mathcal{T}_\alpha$. Let $\alpha, \beta > 0$ and $\alpha + \beta < 1$. We show in the next theorem that, the null parameter set, in which $\Sigma_{\mathcal{S}_l^X, \mathcal{S}_g^Y} = 0$, is asymptotically distinguishable from $\mathcal{U}(4)$ by the test $\Psi_\alpha^{(l,g)}$. Namely, we can reject the null with asymptotic full power if $\Sigma \in \mathcal{U}(4)$.

Theorem 2. Assume (*A2*) holds. Then we have,

$$\inf_{\Sigma \in \mathcal{U}(4)} P(\Psi_\alpha = 1) \to 1, \qquad \text{as } n, p_{l,g} \to \infty.$$

Furthermore, there exists a constant $c_0 > 0$ such that for all large n and $p_{l,g}$,

$$\inf_{\Sigma \in \mathcal{U}(c_0)} \sup_{T_\alpha \in \mathcal{T}_\alpha} P(T_\alpha = 1) \le 1 - \beta.$$

This result shows that we only need one entry in the submatrix having the magnitude of the order $\sqrt{\log p_{l,g} / n}$, to reject the null with full power asymptotically, and this rate $\sqrt{\log p_{l,g} / n}$ is minimax optimal.

2.3. Correction and Normal Quantile Transformation

We note that, when the dimension of the submatrix is not sufficiently large, the limiting null distribution derived in Theorem 1 could potentially induce serious size distortion. This is due to the slow convergence rate in the distribution of the extreme value type statistics (Hall 1991; Liu, Lin, and Shao 2008; Birnbaum and Nadler 2012). Specifically, as we show in the proof of Theorem 1, we approximate the Gaussian tail probability, that is, $1 - \Phi(t_{l,g}^{1/2})$ with $t_{l,g} = 2 \log(p_l p_g) = \log \log(p_l p_g) + t$, by $1/(2\pi t_{l,g})^{1/2} \exp(-t_{l,g}/2)$, for each of the $p_l p_g$ random variables. Thus, we obtain the limit $\exp\{-\pi^{-1/2} e^{-t/2}\}$ in Theorem 1. However, there is a gap between the tail probability and its approximation. This gap can be large, and the critical value derived from the limiting distribution in Theorem 1 may not be accurate. Moreover, the approximation errors would accumulate, especially when we have a diverging number of submatrices to evaluate simultaneously. To address this issue, we propose an intermediate correction for the critical value derived from the asymptotic distribution in Theorem 1. Following the proof of Theorem 1, we have the next result.

Proposition 1. Assume (A1)–(A3) hold. Then under $H_{0,l,g}$, as $n, p_{l,g} \to \infty$, we have, for any $t \in \mathbb{R}$,

$$P(M_{l,g} \le t_{l,g}) \to \exp\left[-2p_l p_g \left\{1 - \Phi\left(t_{l,g}^{1/2}\right)\right\}\right],$$

where $t_{l,g} = 2 \log(p_l p_g) - \log \log(p_l p_g) + t$.

Based on this result, we have $P(M_{l,g} \le t_{l,g}(\alpha)) \to \alpha$ under $H_{0,l,g}$, where $t_{l,g}(\alpha)$ is the critical value such that $\exp[-2p_l p_g\{1 - \Phi(t_{l,g}^{1/2}(\alpha))\}] = 1 - \alpha$. Then we define the corrected α-level test by,

$$\Psi'_\alpha = I\{M_{l,g} \ge t_{l,g}(\alpha)\}.$$

To further illustrate the above bias correction, we show in Figure 1 the empirical cumulative distribution of the test statistic $M_{l,g}$, its limiting distribution without bias correction from Theorem 1, and with bias adjustment from Proposition 1. The left panel is for the cumulative

distribution function value between 0 and 0.2, and the right panel for the value between 0.9 and 1. These regions are the most relevant for the testing problem. It is clearly seen from the plot that the cumulative distribution after correction is much closer to the empirical distribution than the one without correction.

We also note that, the critical values for testing different pairs of regions are usually different, since the region sizes are different. Thus, we further introduce a normal quantile transformation in order to unify the cut-off values for different pairs of regions, so that we can perform the simultaneous inference in the next section. Toward that end, denote by $F_{l,g}^{*}(t_{l,g}) = \exp[-2p_{l}p_{g}\{1 - \Phi(t_{l,g}^{1/2})\}]$, the corrected cumulative distribution function of $M_{l,g}$. For each $1 \leq l, g \leq L$, we introduce the normal quantile transformation,

$$N_{l,g} = \Phi^{-1}\{F_{l,g}^{*}(M_{l,g})\}, \tag{2}$$

where $\Phi(t)$ is the standard normal cumulative distribution function, and $M_{l,g}$ is the maximum test statistic calculated from the original data. Based on Proposition 1, it can be easily shown that $N_{l,g}$ follows a standard normal distribution asymptotically, and we reject the null when $M_{l,g}$ is large, so $N_{l,g}$ is large. Thus, for each $1 \leq l, g \leq L$, we define the following α-level test,

$$\Psi_{\alpha}^{(l,g)*} = I(N_{l,g} \geq \Phi^{-1}(1-\alpha)).$$

We reject the null $\Sigma_{\mathcal{S}_{l}^{X}, \mathcal{S}_{g}^{Y}} = 0$ whenever $\Psi_{\alpha}^{(l,g)*} = 1$.

We comment that both the intermediate correction and the normal quantile transformation proposed here are essential, and they distinguish our test from the existing covariance based tests. The intermediate correction is to better approximate the empirical distribution, and to avoid the accumulation of approximation errors among the submatrices. This is because we have many submatrices to test simultaneously, and each has a high dimensionality. This is different from the existing covariance inference solutions, such as Cai, Liu, and Xia (2013) and Cai and Zhang (2016), where the Gaussian approximation error in the extreme value distribution is negligible and no correction is needed. The normal quantile transformation is to ensure that the tail of the maximum test statistics $M_{l,g}$ after the transformation behaves the same asymptotically under the null for various submatrices simultaneously. Both steps are important for the covariance inference developed in the next section.

3.   Simultaneous Test for Multiple Pairs of Regions

We next develop the simultaneous testing procedure for the inference problem (1), so to identify the pairs of regions that are significantly correlated with each other. We study its theoretical properties in terms of false discovery control.

3.1.   Testing Procedure

Since the sample size is often small in multimodal analysis, we build our multiple testing procedure based upon the finite sample corrected and normal quantile transformed test

statistic, $N_{l,g}$, as defined in Equation (2). Let $\mathscr{H} = \{(l, g) : 1 \leq l, g \leq L\}$ denote the set of all pairs of regions from the two data modalities, let $\mathscr{H}_0 = \{(l, g) : \mathbf{\Sigma}_{\mathscr{S}_l \mathscr{S}_g} = 0\}$ denote the set of true null hypotheses, and $\mathscr{H}_1 = \mathscr{H} \setminus \mathscr{H}_0$ the set of alternatives. For any given thresholding level t, we define the false discovery proportion (FDP) and the false discovery rate (FDR) as,

$$\text{FDP}(t) = \frac{\sum_{(l, g) \in \mathscr{H}_0} I(N_{l, g} \geq t)}{\sum_{(l, g) \in \mathscr{H}} I(N_{l, g} \geq t) \vee 1}, \quad \text{and}$$
$$\text{FDR}(t) = \text{E}\{\text{FDP}(t)\}.$$

Algorithm 1 Simultaneous testing procedure

---

1. Calculate the test statistics $M_{l, g}$ and the corresponding finite sample corrected and normal quantile transformed statistics $N_{l, g}$, for all pairs $1 \leq l, g \leq L$.

2. Estimate the FDP by,

$$\widehat{\text{FDP}}(t) = \frac{L^2(1 - \Phi(t))}{\sum_{(l, g) \in \mathscr{H}} I(N_{l, g} \geq t) \vee 1}.$$

3. For a given $0 \leq \alpha \leq 1$, calculate

$$\hat{t} = \inf \{0 \leq t \leq (4 \log L - 2 \log \log L)^{1/2} : \widehat{\text{FDP}}(t) \leq \alpha\}.$$

If $\hat{t}$ does not exist, set $\hat{t} = 2(\log L)^{1/2}$.

4. For $1 \leq l, g \leq L$, reject $H_{0, l, g}$ if and only if $N_{l, g} \geq \hat{t}$, and set

$$\widehat{\mathscr{H}}_1 = \{(l, g) : N_{l, g} \geq \hat{t}\}.$$

We summarize our proposed multiple testing procedure in Algorithm 1, and make a few remarks regarding the procedure. First, we note that, if the set of null hypotheses $\mathscr{H}_0$ were known, one should reject as many hypotheses as possible while controlling the true FDP at the pre-specified error rate. In practice, however, neither the null set $\mathscr{H}_0$, nor the number of false rejections $\sum_{(l, g) \in \mathscr{H}_0} I(N_{l, g} \geq t)$ is known. As such, we first set to estimate the FDP.

Second, for the applications such as multimodal neuroimaging analysis, we are especially interested in the scenario when most of the pairs of the regions between the two modalities are not correlated with each other. This is reflected in that, under such a sparse scenario, $|\mathscr{H}_0|$ is close to the total number of hypotheses L^2. By construction, $N_{l,g}$ is close to a standard normal random variable. Consequently, the FDP can be estimated by $\widehat{\text{FDP}}(t)$ in Step 2. Third, it is critical to restrict t on the range $[0, (4 \log L - 2 \log \log L)^{1/2}]$ in Step 3. When t $(4 \log L - 2 \log \log L)^{1/2}$, $L^2\{1 - \Phi(t)\} \to 0$, and it is no longer a consistent estimator of the number of false rejections. Then the algorithm may not able to control the FDP with a positive probability. Finally, it is also important to threshold the test statistics at $2 (\log L)^{1/2}$ instead of $(4 \log L - 2 \log \log L)^{1/2}$. When t does not exist in the range, thresholding the test statistics at $(4 \log L - 2 \log \log L)^{1/2}$ would cause too many false rejections, and consequently, the FDR cannot be controlled asymptotically at level α.

3.2. Theory

We next investigate the error rate control of the above simultaneous testing procedure. For any $1 \leq l \leq L$, define

$\Lambda_l(\gamma) = \{g : 1 \leq g \leq L, \exists i \in \mathcal{S}_l^X \cup \mathcal{S}_l^Y, j \in \mathcal{S}_g^X \cup \mathcal{S}_g^Y, \text{ s.t. } |\sigma_{i,j}| \geq (\log L)^{-2-\gamma}\}$. Furthermore, define $\mathcal{L}_\rho = \{(l,g) : \exists i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y, \text{ such that } |\sigma_{i,j}|/(\theta_{i,j}/n)^{1/2} \geq (\log p)^{1/2+\rho}\}$. The next theorem shows that, under some mild conditions, the simultaneous testing procedure in Algorithm 1 controls both FDR and FDP at level α asymptotically. That is, for the sparse scenario when $|\mathcal{H}_0| / L^2 \to 1$, we have both FDR and FDP converge to the pre-specified level α asymptotically, as long as there exist a few pairs of regions in which their voxel-wise correlations exceed the threshold $(\log p)^{1/2+\rho}/n^{1/2}$.

Theorem 3. Assume (A1)–(A3) hold. In addition, assume that there exists some $\gamma > 0$, such that $\max_{1 \leq l \leq L} |\Lambda_l(\gamma)| = o(L^\nu)$ for any $\nu > 0$. Assume that $p \leq cn^a$ for some $c > 0$ and $a > 0$. Then we have,

$$\limsup \text{FDR}(\hat{t}) \leq \alpha |\mathcal{H}_0| / L^2, \quad \text{and}$$
$$\lim P\{\text{FDP}(\hat{t}) \leq \alpha |\mathcal{H}_0| / L^2 + \epsilon\} = 1,$$

for any $\epsilon > 0$, where the limit is taken as $n, L, (p_{l,g})_{l,g=1}^L \to \infty$. Furthermore, if for some $\rho, \delta > 0$, $|\mathcal{L}_\rho| \geq (1 / (\sqrt{8\pi}\alpha) + \delta)\sqrt{\log L}$, then we have,

$$\frac{\text{FDR}(\hat{t})}{\alpha |\mathcal{H}_0| / L^2} \to 1, \quad \text{and} \frac{\text{FDP}(\hat{t})}{\alpha |\mathcal{H}_0| / L^2} \to 1 \text{ in probability,}$$
$$\text{as } n, L, (p_{l,g})_{l,g=1}^L \to \infty.$$

We again outline the key steps of the proof. We first show that if $\hat{t}$ does not exist in the range $[0, \{2 \log(L^2) - 2 \log \log L\}^{1/2}]$, then the thresholding $\hat{t}$ at $2 (\log L)^{1/2}$ would lead to no false rejection with probability tending to 1. As such we focus the analysis on the event $\mathcal{A} = \{\hat{t} \text{ exists in the range } [0, \{2 \log(L^2) - 2 \log \log L\}^{1/2}]\}$. We then divide the pairs of submatrices into various subsets, among which the weakly correlated pairs play the dominating role, and consequently, the FDR and FDP get controlled conservatively. With some mild condition on $|\mathcal{L}_\rho|$, we show that the event $\mathcal{A}$ occurs with probability tending to 1, and hence the FDR and FDP converge to $\alpha |\mathcal{H}_0| / L^2$ asymptotically. A detailed proof is given in the supplementary appendix.

4. Simulations

4.1. Setup

In this section, we examine the finite sample performance of the proposed testing procedure. We generate two sets of p-dimensional vectors $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^n$, where each component of X_k and Y_k is from a certain random distribution, and the covariance between X_k and Y_k is governed by some particular structure. We consider two random distributions, and three

covariance structures. Moreover, we examine two scenarios of distribution of the significantly correlated regions.

Specifically, we fix the total number of regions at $L = 100$, then randomly choose the size of each region p_l between 50 and 100, $l = 1,\ldots, L$. This results in the total dimension p ranging between 5000 and 10,000. We then generate p_l random variables of the first modality X_k for each region. Each component follows a random distribution. Note that our procedure is robust to the choice of this distribution as long as the moment condition in (A2) is satisfied. To verify that, we consider two specific distributions:

> Distribution 1: Exponential tail: a standard normal distribution;

> Distribution 2: Polynomial tail: a t-distribution with 10 degrees of freedom.

We remark that, in Cai, Liu, and Xia (2013), a t-distribution with 12 degrees of freedom was considered for global testing of the entire covariance matrix, and the distribution considered in our case deviates more from the normal distribution than Cai, Liu, and Xia (2013). We consider two sample sizes, $n = 100$ and $n = 150$. After generating n copies of the first modality, $\{\mathbf{X}_k\}_{k=1}^{n}$, we then generate n copies of the second modality, $\{\mathbf{Y}_k\}_{k=1}^{n}$, where $\mathbf{Y}_k = \mathbf{\Sigma}_{1,2}^{\top}\mathbf{X}_k + \mathbf{E}_k$, with $\{\mathbf{E}_k\}_{k=1}^{n}$ randomly generated from a standard Gaussian distribution and independent of $\{\mathbf{X}_k\}_{k=1}^{n}$. Accordingly, $\operatorname{cov}(\mathbf{X}_k, \mathbf{Y}_k) = c\mathbf{\Sigma}_{1,2}$, where $c = \operatorname{var}(X_{k,i})$, with $c = \operatorname{var}\{N(0,1)\} = 1$ for Distribution 1, and $c = \operatorname{var}\{t(10)\} = 1.25$ for Distribution 2. We consider three covariance structures for $\mathbf{\Sigma}_{1,2}$:

> Structure 1: $\mathbf{\Sigma}_{1,2} = (\sigma_{i,j})$ where $\sigma_{i,j} = 0.8$ Bernoulli$(1, 2/p)$ for $1 \le i, j \le p$;

> Structure 2: $\mathbf{\Sigma}_{1,2} = (\sigma_{i,j})$ where $\sigma_{i,j} =$Uniform$(0.5, 2)$, $\sigma_{i,j} = 0.8$ for $5(k-1) + 1 \le i, j \le 5k$, with $k = 1,\ldots, [p/5]$, and $\sigma_{i,j} = 0$ otherwise;

> Structure 3: $\mathbf{\Sigma}_{1,2} = (\sigma_{i,j})$ where $\sigma_{i,i} =$Uniform$(0.5, 2)$, $\sigma_{i,i+1} = \sigma_{i+1,i} = 0.8$ for $i = 1,\ldots, p - 1$, and $\sigma_{i,j} = 0$ otherwise.

Structure 1 is essentially a random covariance structure. Structures 2 and 3 imply that the same locations in the two modalities are more likely to be highly correlated with each other. Moreover, Structure 2 has a block covariance structure and Structure 3 has a banded structure. For those pairs of regions with fewer or equal than 3 nonzero correlations, we set the corresponding submatrix equal to zero. We consider two scenarios for the distribution of those significantly correlated regions:

> Scenario 1: The significantly correlated regions span all L regions;

> Scenario 2: The significantly correlated regions concentrate on randomly chosen 60 regions out of 100 for the first modality, and on 30 regions out of 100 for the second modality. This setting mimics the real example in Section 5.

4.2. Results

We apply the simultaneous testing procedure in Algorithm 1 to identify the significant pairs of regions. The intermediate correction is employed for the limiting distribution of the test statistic $M_{l,g}$ for $(l, g) \in \mathcal{H}$, with $\mathcal{H} = \{(l, g) : 1 \le l, g \le L\}$ in Scenario 1, and

$\mathcal{H} = \{(l, g) : l \in R_1, j \in R_2\}$ in Scenario 2, where R_1 and R_2 are the randomly chosen 60 and 30 regions for the first and second modality, respectively. The normal quantile transformed test statistics $N_{l,g}$ are calculated based on the corrected distribution. We set the significant level at $\alpha = 0.05$. We compare our procedure with that of Xie and Kang (2017). We also numerically compare with sparse CCA. We use the implementation and the recommended tuning method in the R package PMA by Witten, Tibshirani, and Hastie (2009). Since sparse CCA is applied on the region-wise rather than the voxelwise data, we first average the data within each region, then apply sparse CCA. After obtaining two sets of nonzero region locations, denoted as u_1 and u_2, we have the significant pairs of regions identified by sparse CCA as $\{(i, j) : i \in u_1, j \in u_2\}$. We report the empirical FDR and power, both in percentages, based on 100 data replications, for $n = 100$ in Table 1, and for $n = 150$ in Table 2. Here the empirical power is calculated as

$$\frac{1}{100} \sum_{d=1}^{100} \frac{\sum_{(l, g) \in \mathcal{H}_1} I(N_{l, g, d} \geq \hat{t})}{|\mathcal{H}_1|},$$

where $N_{l,g,d}$ denotes the test statistic for the dth data replication.

From Table 1, we observe that, the empirical FDR for our testing procedure is close to the nominal level $\alpha = 0.05$ in all situations. In comparison, the test of Xie and Kang (2017) is extremely conservative, while sparse CCA has severe FDR distortion in most of the cases. We also observe that, for the empirical power, our testing procedure is powerful, and substantially outperforms both Xie and Kang (2017) and sparse CCA in all situations. Additionally, the performances under the two distributions are close, which reflects the robustness of our proposal with respect to the tail behaviors of the random quantities. From Table 2, we observe a similar pattern in empirical FDR. Meanwhile, the empirical power has considerably improved when the sample size increases.

In the above setting, the variables in modality X_k are generated independently. We also consider the case when the variables within the modality are correlated. We report the results in the supplementary appendix.

5. Multimodal PET Analysis

5.1. Scientific Background and Dataset

We revisit the motivating multimodal PET study introduced in Section 1. AD is an irreversible neurodegenerative disorder and the leading form of dementia in elderly subjects. It is characterized by progressive impairment of cognitive functions and inability to perform activities of daily living. With aging of the worldwide population, the number of affected people is drastically increasing, and it becomes an international imperative to understand, diagnose, and treat this disorder. Beta-amyloid (Aβ) and tau are two hallmark pathological proteins, and are believed to be part of the driving mechanism of AD. These proteins are commonly found in pathological forms not only in the brains of AD subjects, but also the brains of older adults without AD. Our primary goal is to find brain regions where Aβ and tau are statistically significantly correlated with each other. We also seek to identify

significantly correlated brain functional modules. Such analyses would be extremely useful from a biological perspective, as prevailing theories of the development of AD pathology in the aging human brain suggest that increased Aβ in some regions, for example, parietal cortex, may initiate the increase and propagation of tau pathology in other nonlocal regions, for example, medial temporal lobe. In turn, it would allow for the prediction of the earliest stages of in vivo neuropathology, and help generate better designed AD therapeutic clinical trials. For instance, clinical trials aiming at testing anti-Aβ or anti-tau agents would need to know not only that participants have both Aβ and tau in the brain, but also exactly how the relative levels of each pathological protein might predict incident cognitive impairment and dementia.

The dataset we analyzed is part of the ongoing Berkeley Aging Cohort Study. It consists of 81 cognitively normal older adult subjects, aged 77.5 ± 6.2 years, and 32/49 male/female. For each subject, an AV-1451 PET scan for measuring tau, a Pittsburgh Compound B (PiB) PET scan for measuring Aβ protein, and a 1.5T structural MRI scan were acquired. For PiB PET, native-space voxelwise distribution volume ratio (DVR) images were generated using Logan graphical analysis (35–90 min postinjection, cerebellar gray matter [GM] reference region) (Logan et al. 1996). For AV-1451 PET, native-space standardized uptake value ratio (SUVR) images were created for each participant (80–100 min postinjection, inferior cerebellar GM reference region). Native-space PET images were coregistered to each participant's MRI scan. All MRI scans were normalized to the FSL MNI152 2 mm space template via a study-specific intermediate template (Lockhart et al. 2017), and transformations were concatenated and applied to the coregistered AV-1451 and PiB PET images to generate MNI-space PET images. A mask representing voxels likely to accumulate cortical amyloid and tau pathology was created. To do this, we intersected a cortical brain mask from the Automated Anatomical Labeling atlas (Tzourio-Mazoyer et al. 2002) with a mask of high-probability GM voxels from the SPM12 tissue probability map. It was also masked to include only voxels where the coefficient of variation in the signal (DVR or SUVR) across participants in either PET modality was   0.5, and the mean signal across participants was   0.8. MNI-space PiB and AV-1451 PET images were masked by this cortical mask before analysis. From a FreeSurfer segmentation of the MNI152 template structural MRI, a set of MNI-space testing regions were created that encompassed Braak I–IV stage regions for AV-1451, and Braak I–V stage regions for PiB, excluding basal ganglia and thalamus (Schöll et al. 2016). These testing regions were also designed to focus on areas of possible tau and Aβ accumulation in normal elders. Moreover, these regions belong to seven known functional modules, including visual, somatomotor, dorsal attention, ventral attention, limbic, frontoparietal, and default mode (Yeo et al. 2011).

5.2. Analysis and Results

We first applied the proposed testing procedure at the region level. For the two modalities, PiB and AV-1451, the total dimension is 54,540 and 18,194, and the number of regions is 60 and 26, respectively. The number of voxels in the smallest region is 104 and 203, and in the largest region is 2982 and 1729, respectively. For each pair of the regions, the standardized statistic $T_{i,j}$ was calculated and the normal quantile transformed statistic $N_{l,g}$ was constructed based on the corrected limiting distribution of the test statistics $M_{l,g}$, which was

calculated according to the 1% voxels with the largest magnitudes. We set the significance level $\alpha = 0.05$, and identified 385 significantly correlated brain regions between the two modalities. We report those identified significant pairs of inter-tracer region-of-interest (ROI) in Figure 2. The top panel shows the average correlations between the two imaging modalities for those significant pairs of regions, and the bottom panel the corresponding p-values. We plot the average correlation in the top panel to reflect the average behavior of the correlations of those identified significant pairs of ROIs, whereas the p-values in the bottom panel reflects the behavior of our test statistic, that is, the maximum of absolute correlations. Figure 3 further shows the average correlation values for those AV-1451 or PiB regions imposed on a template brain image.

From the plots and the identified significantly correlated brain regions, we have observed extensive inter-tracer associations both across the AV-1451 ROIs (Figure 2, rows), and across PiB ROIs (Figure 2, columns). For AV-1451 ROIs, strong region-specific correlations with PiB uptake across the brain were observed, particularly for AV-1451 accumulation in temporal lobe regions, such as bilateral entorhinal cortex, parahippocampal cortex, fusiform gyrus, middle temporal gyrus, and inferior temporal cortex. In other words, increased PiB uptake in areas of the brain that correspond to stages I–V in a postmortem staging scheme devised by Braak and Braak (1991) was associated with increased AV-1451 uptake in these temporal lobe regions. We also observed less consistent but still numerous associations of PiB ROIs with AV-1451 across the brain, such that increased PiB uptake in regions like bilateral amygdala, isthmus cingulate, and insula correlated with increased AV-1451 in numerous brain regions. In general, increased PiB binding to temporal and extra-temporal regions was correlated with AV-1451 binding in areas such as bilateral entorhinal cortex, parahippocampal gyrus, fusiform gyrus, middle temporal gyrus, and inferior temporal cortex. These findings support previous research, for example, Lockhart et al. (2017) and Sepulcre et al. (2016), in suggesting that there exist complex regional associations between $A\beta$ and tau deposition in normal aging. These results particularly support the concept that there is a regional tau accumulation vulnerability of brain regions like temporal lobes in association with $A\beta$ accumulation across numerous brain regions. While some previous research has identified associations between $A\beta$ and tau PET tracers, many of those studies did not explore regional associations like the current study (e.g., Johnson et al. 2016; Ossenkoppele et al. 2016; Schöll et al. 2016; Schwarz et al. 2016), and none have developed any formal statistical inferential methods for assessing relationships between multimodal neuroimaging data. Our study is able to rigorously and explicitly evaluate how patterns of pathological protein accumulation are associated with one another in brains potentially impacted by preclinical AD.

We also carried out the correlation analysis at the module level, mostly for the illustration purpose. We note that, for the region level analysis, there are potential correlations among the voxels within the same region. However, for the module level analysis, there is no obvious correlation structure among the regions in the same module, as those regions can scatter at distant locations of the brain. Both scenarios are common in a multitude of scientific applications.

Specifically, the ROIs in both modalities belong to all seven functional modules as defined in Yeo et al. (2011), and the seven modules have 8, 10, 4, 6, 16, 6, and 22 ROIs, respectively. Since this level of analysis focuses on the region-wise correlations, we first averaged all the voxels in the same regions, then applied the proposed testing procedure to identify significantly correlated PiB and Tau modules. We again set the significance level $\alpha = 0.05$, and identified 19 significantly correlated functional modules between the two modalities. We report those pairs in Figure 4, with the left panel showing the average correlations between the two imaging modalities for those significant pairs of modules, and the right panel the corresponding p-values.

We have observed that, all significant patterns of module level associations were limited to the AV-1451 modules of dorsal attention, limbic, and default mode networks, while patterns were more distributed across PiB modules. This is consistent with a greater spatial spread of amyloid than tau pathology across brain regions in the cognitively normal older adults in this study; for example, tau in normal aging is rarely present in visual or somatomotor areas. Furthermore, the visual and somatosensory modules do not correlate with each other across tracers; these sensory networks are relatively functionally distinct, and again are often among the last regions where amyloid and particularly tau pathology are observed in the brains of cognitively normal elders. The modules that self-correlate between tracer modalities, dorsal attention, limbic, and default mode, are brain areas where both amyloid and tau pathology tend to colocalize.

We also compared the results from the region level analysis and the module level analysis. We converted the significant regions found from the region level analysis to modules according to their corresponding module membership. This converted 385 significantly correlated region pairs to 38 unique module pairs. We then ranked these module pairs by the number of their occurrences. We first found that all 19 significant module pairs found in the module level analysis were also found in the significant module pairs in the region level analysis. Moreover, among the 11 pairs of modules with more than 10 occurrences from the region level analysis, our module level analysis found 9 of them, which are also the ones with the largest 9 occurrences. The two analyses thus yielded consistent findings.

6. Discussion

In this article, we have developed a new testing procedure for simultaneous covariance inference of associations between multiple modalities. The proposed test offers both a timely response to an important type of scientific problem of multimodal integrative analysis, as well as a useful addition to the toolbox of statistical covariance inference. Next, we make a few remarks about our proposal.

We measure the association of the two modalities by Pearson correlation, which is commonly employed in neuroimaging analysis (Fornito, Zalesky, and Breakspear 2013). Other correlations measures, in principle, can be cast into our framework too, though it would require new derivations of the asymptotic properties of the test.

The inference procedure we develop allows the sizes of the regions, that is, $\{p_1, p_2, \ldots, p_L\}$, to be much larger than the sample size n, and also to grow with n. This is a unique feature of our method, and is especially useful for multimodal analysis type applications, where the sample size is often limited.

We are particularly interested in the sparse scenario when the majority of the submatrices in $\{\boldsymbol{\Sigma}_{\mathscr{S}_l^X, \mathscr{S}_g^Y}, 1 \leq l, g \leq L\}$ are zero matrices. Practically, this type of sparsity is to facilitate the scientific interpretation. Technically, it permits a precise approximation of the false discoveries. On the other hand, even if this sparsity is not satisfied, our inference procedure remains valid, only with a relatively conservative false discovery control. This is due to the fact that, when the sparsity does not hold, the procedure still controls the FDR asymptotically at the level $\alpha |\mathscr{H}_0| / L^2$, which is smaller than α.

Our test statistic $M_{l,g}$ is constructed as the maximum of the standardized covariance statistic $T_{i,j}$, for $i \in \mathscr{S}_l^X$, $j \in \mathscr{S}_g^Y$, between the two modalities. Alternatively, one may consider a test statistic as the sum of $T_{i,j}^2$, that is,

$$S_{l,g} = \sum_{i \in \mathscr{S}_l^X, j \in \mathscr{S}_g^Y} T_{i,j}^2.$$

We have chosen the maximum type test statistic over the sum type, for several reasons. First, $M_{l,g}$ is particularly suitable for the case when one or a few entries of the submatrix $\boldsymbol{\Sigma}_{\mathscr{S}_l^X, \mathscr{S}_g^Y}$ have large values. But it also works for the case when many entries of $\boldsymbol{\Sigma}_{\mathscr{S}_l^X, \mathscr{S}_g^Y}$ are large. By contrast, $S_{l,g}$ is not suitable for the case when only a few entries are large, since in this case, most entries would act as "noise" and the sum can fail to be large enough to distinguish between the null and alternative hypotheses. In a multitude of applications, including our multimodal PET study, both scenarios are common. Second, the test based on $M_{l,g}$ has a stable theoretical performance when the dimension $p_l p_g$ diverges. However, the test based on $S_{l,g}$ may become intractable when $p_l p_g$ diverges. This is because, to derive the limiting null distribution of $S_{l,g}$, one needs to estimate the covariance structure of $\{T_{i,j}, i \in \mathscr{S}_l^X, j \in \mathscr{S}_g^Y\}$, which is of dimension $p_l p_g \times p_l p_g$. When $p_l p_g$ diverges, it becomes extremely difficult to estimate such a covariance matrix. Furthermore, since there are L^2 such covariance structures to estimate simultaneously, the FDR of the multiple testing procedure based on $S_{l,g}$ can be easily distorted, if some of the null distributions of those L^2 submatrices are mis-specified. Normalization of $S_{l,g}$ offers a possible alternative solution (Schott 2007; Li and Chen 2012). But the false discovery control based on a normalized $S_{l,g}$ imposes completely new challenges.

Finally, we remark on the spatial dependency and its effect on our test method. The standardized covariance statistics $T_{i,j}$'s in the regions $\mathscr{S}_l^X$ and $\mathscr{S}_g^Y$ can be correlated; for instance, in our multimodal PET example, the brain voxels are spatially correlated, and as such $T_{i,j}$'s are correlated. First of all, we clarify that, for our proposed test, we do *not* impose the spatial independence assumption. In Section 2.2, we have introduced Assumption (A1),

which requires the number of highly correlated entries in $\mathcal{S}_{l,g}$ not to be too many. This condition is easily satisfied, for example, when the eigenvalues of $\mathbf{\Sigma}^{(l,g)}$ are bounded from above. With this weak dependency condition, we can derive the exact limiting distribution for our test statistic. On the other hand, we can relax this weak dependency condition, and can still approximately control the size of the test. Specifically, we have the following result.

Proposition 2. Assume the moment condition (A2) holds. Then for $0 < \alpha < 1$, we have that

$$P(\text{Type I error}) = P_{H_{0,l,g}}(\Psi_\alpha = 1) \leq -\log(1-\alpha) + o(1).$$

The proof of this proposition is based on the relaxation of $P_{H_{0,l,g}}\{M_{l,g} \quad q_\alpha + 2\log(p_l p_g) - \log\log(p_l p_g)\}$ to $\sum_{i \in \mathcal{S}_l^X, j \in \mathcal{S}_g^Y} P_{H_{0,l,g}}\{T_{i,j}^2 \geq q_\alpha + 2\log(p_l p_g) - \log\log(p_l p_g)\}$, and the normal approximation by Lemma 5 in the supplementary appendix. It is interesting to observe that, $-\log(1-\alpha) \approx \alpha$ for a small α. For instance, for the commonly used significance level $\alpha = 0.05$, $-\log(1-\alpha) = 0.05129$, and for $\alpha = 0.01$, $-\log(1-\alpha) = 0.01005$. Therefore, we can still *approximately* control the size of our test even without any dependency assumption. Second, the construction of our test statistic $M_{l,g}$ does not explicitly incorporate potential correlations among the samples $\{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^n$. On one hand, in real scientific applications, both scenarios, where such correlations are present or absent, are not uncommon. In our multimodal PET example, for the region level analysis, there are potential correlations among the voxels within the same region. But for the module level analysis, there is no obvious correlation structure among the regions in the same module. Our test has been designed for both scenarios. On the other hand, we recognize that, by incorporating potential correlations, it may further enhance the testing procedure. This is a very challenging problem though, and we leave it to our future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

References

Ahn M, Shen H, Lin W, and Zhu H (2015), "A Sparse Reduced Rank Framework for Group Analysis of Functional Neuroimaging Data," Statistica Sinica, 25, 295–312. [PubMed: 26405427]

Anderson TW (2003), An Introduction to Multivariate Statistical Analysis (3rd ed.), New York: Wiley-Interscience.

Benjamini Y, and Yekutieli D (2001), "The Control of the False Discovery Rate in Multiple Testing Under Dependency," Annals of Statistics, 29, 1165–1188.

Birnbaum A, and Nadler B (2012), "High Dimensional Sparse Covariance Estimation: Accurate Thresholds for the Maximal Diagonal Entry and for the Largest Correlation Coefficient," Technical report.

Braak H, and Braak E (1991), "Neuropathological Staging of Alzheimer-Related Changes," Acta Neuropathologica, 82, 239–259. [PubMed: 1759558]

Brier MR, Gordon B, Friedrichsen K, McCarthy J, Stern A, Christensen J, Owen C, Aldea P, Su Y, Hassenstab J, Cairns NJ, Holtzman DM, Fagan AM, Morris JC, Benzinger TLS, and Ances BM (2016), "Tau and Aβ Imaging, CSF Measures, and Cognition in Alzheimer's Disease," Science Translational Medicine, 8, 338ra66–338ra66.

Cai T, Cai TT, and Zhang A (2016), "Structured Matrix Completion With Applications to Genomic Data Integration," Journal of the American Statistical Association, 111, 621–633. [PubMed: 28042188]

Cai TT, and Liu W (2016), "Large-Scale Multiple Testing of Correlations," Journal of the American Statistical Association, 111, 229–240. [PubMed: 27284211]

Cai TT, Liu W, and Xia Y (2013), "Two-Sample Covariance Matrix Testing and Support Recovery in High-Dimensional and Sparse Settings," Journal of the American Statistical Association, 108, 265–277.

Cai TT, and Zhang A (2016), "Inference for High-Dimensional Differential Correlation Matrices," Journal of Multivariate Analysis, 143, 107–126. [PubMed: 26500380]

Chen M, Gao C, Ren Z, and Zhou H (2013), "Sparse CCA Via Precision Adjusted Iterative Thresholding," arXiv Preprint arXiv:1311.6186.

Chien D, Szardenings A, Bahri S, Walsh J, Mu F, Xia C, Shankle W, Lerner A, Su M, Elizarov A, and Kolb H (2014), "Early Clinical PET Imaging Results With the Novel PHF-tau Radioligand [f18]–t808," Journal of Alzheimer's Disease, 34, 457–468

Efron B (2007), "Correlation and Large-Scale Simultaneous Significance Testing," Journal of the American Statistical Association, 102, 93–103.

Fornito A, Zalesky A, and Breakspear M (2013), "Graph Analysis of the Human Connectome: Promise, Progress, and Pitfalls," NeuroImage, 80, 426–444. [PubMed: 23643999]

Hall P (1991), "On Convergence Rates of Suprema," Probability Theory and Related Fields, 89, 447–455.

Johnson KA, Schultz A, Betensky RA, Becker JA, Sepulcre J, Rentz D, Mormino E, Chhatwal J, Amariglio R, Papp K, Marshall G, Albers M, Mauro S, Pepin L, Alverio J, Judge K, Philiossaint M, Shoup T, Yokell D, Dickerson B, Gomez-Isla T, Hyman B, Vasdev N, and Sperling R (2016), "Tau Positron Emission Tomographic Imaging in Aging and Early Alzheimer's Disease," Annals of Neurology, 79, 110–119. [PubMed: 26505746]

Kang J, Bowman FD, Mayberg H, and Liu H (2016), "A Depression Network of Functionally Connected Regions Discovered Via Multi-Attribute Canonical Correlation Graphs," NeuroImage, 141, 431–441. [PubMed: 27474522]

Li J, and Chen SX (2012), "Two Sample Tests for High-Dimensional Covariance Matrices," The Annals of Statistics, 40, 908–940.

Li Q, Wang S, Huang C-C, Yu M, and Shao J (2014), "Meta-Analysis Based Variable Selection for Gene Expression Data," Biometrics, 70, 872–880. [PubMed: 25196635]

Lin J-A, Zhu H, Mihye A, Sun W, and Ibrahim JG (2014), "Functional Mixed Effects Models for Candidate Genetic Mapping in Imaging Genetic Studies," Genetic Epidemiology, 38, 680–691. [PubMed: 25270690]

Liu J, Huang J, Zhang Y, Lan Q, Rothman N, Zheng T, and Ma S (2014), "Integrative Analysis of Prognosis Data on Multiple Cancer Subtypes," Biometrics, 70, 480–488. [PubMed: 24766212]

Liu W (2013), "Gaussian Graphical Model Estimation With False Discovery Rate Control," Annals of Statistics, 41, 2948–2978.

Liu W-D, Lin Z, and Shao Q-M (2008), "The Asymptotic Distribution and Berry—Esseen Bound of a New Test for Independence in High Dimension With an Application to Stochastic Optimization," Annals of Applied Probability, 18, 2337–2366.

Author Manuscript   Author Manuscript   Author Manuscript   Author Manuscript

Lockhart SN, Schöll M, Baker SL, Ayakta N, Swinnerton KN, Bell RK, Mellinger TJ, Shah VD, O'Neil JP, Janabi M, and Jagust WJ (2017), "Amyloid and Tau PET Demonstrate Region-Specific Associations in Normal Older People," NeuroImage, 150, 191–199. [PubMed: 28232190]

Logan J, Fowler JS, Volkow ND, Wang G-J, Ding Y-S, and Alexoff DL (1996), "Distribution Volume Ratios Without Blood Sampling From Graphical Analysis of PET Data," Journal of Cerebral Blood Flow & Metabolism, 16, 834–840. [PubMed: 8784228]

Nathoo FS, Kong L, and Zhu H (2017), "Inference on High-Dimensional Differential Correlation Matrix," arXiv preprint arXiv:1707.07332.

Ossenkoppele R, Schonhaut DR, Schll M, Lockhart SN, Ayakta N, Baker SL, O'Neil JP, Janabi M, Lazaris A, Cantwell A, Vogel J, Santos M, Miller ZA, Bettcher BM, Vossel KA, Kramer JH, Gorno-Tempini ML, Miller BL, Jagust WJ, and Rabinovici GD (2016), "Tau PET Patterns Mirror Clinical and Neuroanatomical Variability in Alzheimer's Disease," Brain, 139, 1551. [PubMed: 26962052]

Richardson S, Tseng GC, and Sun W (2016), "Statistical Methods in Integrative Genomics," Annual Review of Statistics and Its Application, 3, 181–209.

Schöll M, Lockhart S, Schonhaut D, ONeil J, Janabi M, Ossenkoppele R, Baker S, Vogel J, Faria J, Schwimmer H, Rabinovici G, and Jagust W (2016), "PET Imaging of Tau Deposition in the Aging Human Brain," Neuron, 89, 971–982. [PubMed: 26938442]

Schott JR (2007), "A Test for the Equality of Covariance Matrices When the Dimension is Large Relative to the Sample Sizes," Computational Statistics & Data Analysis, 51, 6535–6542.

Schwarz AJ, Yu P, Miller BB, Shcherbinin S, Dickson J, Navitsky M, Joshi AD, Devous MD Sr, and Mintun MS (2016), "Regional Profiles of the Candidate Tau PET Ligand 18 f-av-1451 Recapitulate Key Features of Braak Histopathological Stages," Brain, 139, 1539. [PubMed: 26936940]

Sepulcre J, Schultz AP, Sabuncu M, Gomez-Isla T, Chhatwal J, Becker A, Sperling R, and Johnson KA (2016), "In vivo Tau, Amyloid, and Gray Matter Profiles in the Aging Brain," Journal of Neuroscience, 36, 7364–7374. [PubMed: 27413148]

Shen R, Wang S, and Mo Q (2013), "Sparse Integrative Clustering of Multiple Omics Data Sets," Annals of Applied Statistics, 7, 269–294.

Singh M, Kim S, and Kim T-S (2003), "Correlation Between Bold-FMRI and EEG Signal Changes in Response to Visual Stimulus Frequency in Humans," Magnetic Resonance in Medicine, 49, 108–114. [PubMed: 12509825]

Smith SD, Fox PT, Miller K, Glahn D, Fox P, Mackay CE, Filippini N, Watkins KE, Toro R, Laird A, and Beckmann CF (2009), "Correspondence of the Brain; Functional Architecture During Activation and Rest," Proceedings of the National Academy of Sciences of the United States of America, 106, 13040–5. [PubMed: 19620724]

Sun W, and Cai T (2009), "Large-Scale Multiple Testing Under Dependence," Journal of the Royal Statistical Society, Series B, 71, 393–424.

Sun W, Reich BJ, Tony Cai T, Guindani M, and Schwartzman A (2015), "False Discovery Control in Large-Scale Spatial Multiple Testing," Journal of the Royal Statistical Society, Series B, 77, 59–83.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, and Joliot M (2002), "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," NeuroImage, 15, 273–289. [PubMed: 11771995]

Uludag K, and Roebroeck A (2014), "General Overview on the Merits of Multimodal Neuroimaging Data Fusion," NeuroImage, 102, 3–10. [PubMed: 24845622]

Witten DM, Tibshirani R, and Hastie T (2009), "A Penalized Matrix Decomposition, With Applications to Sparse Principal Components and Canonical Correlation Analysis," Biostatistics, 10, 515–534. [PubMed: 19377034]

Xia Y (2017), "Testing and Support Recovery of Multiple High-Dimensional Covariance Matrices With False Discovery Rate Control," Test, 26, 782–801.

Xia Y, Cai T, and Cai TT (2015), "Testing Differential Networks With Applications to the Detection of Gene-Gene Interactions," Biometrika, 102, 247–266. [PubMed: 28502988]

_____ (2018), "Multiple Testing of Submatrices of a Precision Matrix With Applications to Identification of Between Pathway Interactions," Journal of the American Statistical Association, 113, 328–339. [PubMed: 29881130]

Xia Y, and Li L (2017), "Hypothesis Testing of Matrix Graph Model With Application to Brain Connectivity Analysis," Biometrics, 73, 780–791. [PubMed: 27959470]

Xie J, and Kang J (2017), "High-Dimensional Tests for Functional Networks of Brain Anatomic Regions," Journal of Multivariate Analysis, 156, 70–88. [PubMed: 28413234]

Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zllei L, Polimeni JR, Fischl B, Liu H, and Buckner RL (2011), "The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity" Journal of Neurophysiology, 106, 1125–1165. [PubMed: 21653723]

Zhang D, Wang Y, Zhou L, Yuan H, Shen D, and the Alzheimers Disease Neuroimaging Initiative (2011), "Multimodal Classification of Alzheimer's Disease and Mild Cognitive Impairment," Neuroimage, 55, 856–867. [PubMed: 21236349]

Zhu D, Zhang T, Jiang X, Hu X, Chen H, Yang N, Lv J, Han J, Guo L, and Liu T (2014), "Fusing DTI and FMRI Data: A Survey of Methods and Applications," NeuroImage, 102, 184–191. [PubMed: 24103849]
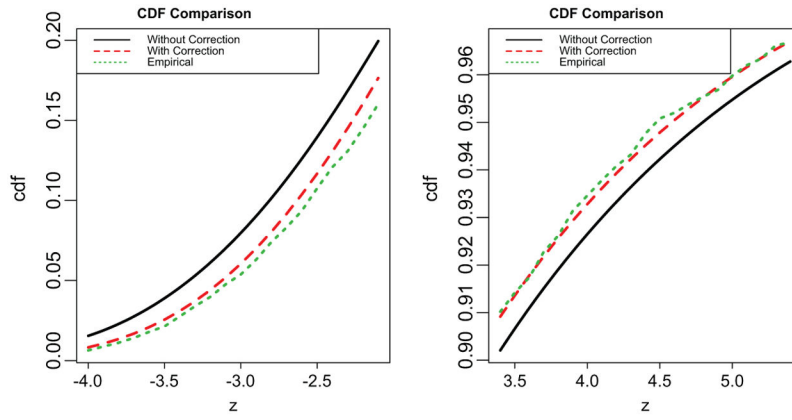
Figure 1.
Comparison of the empirical cumulative distribution and the limiting distribution with and without the correction. The size of the submatrix is 80×100 and $n = 100$.
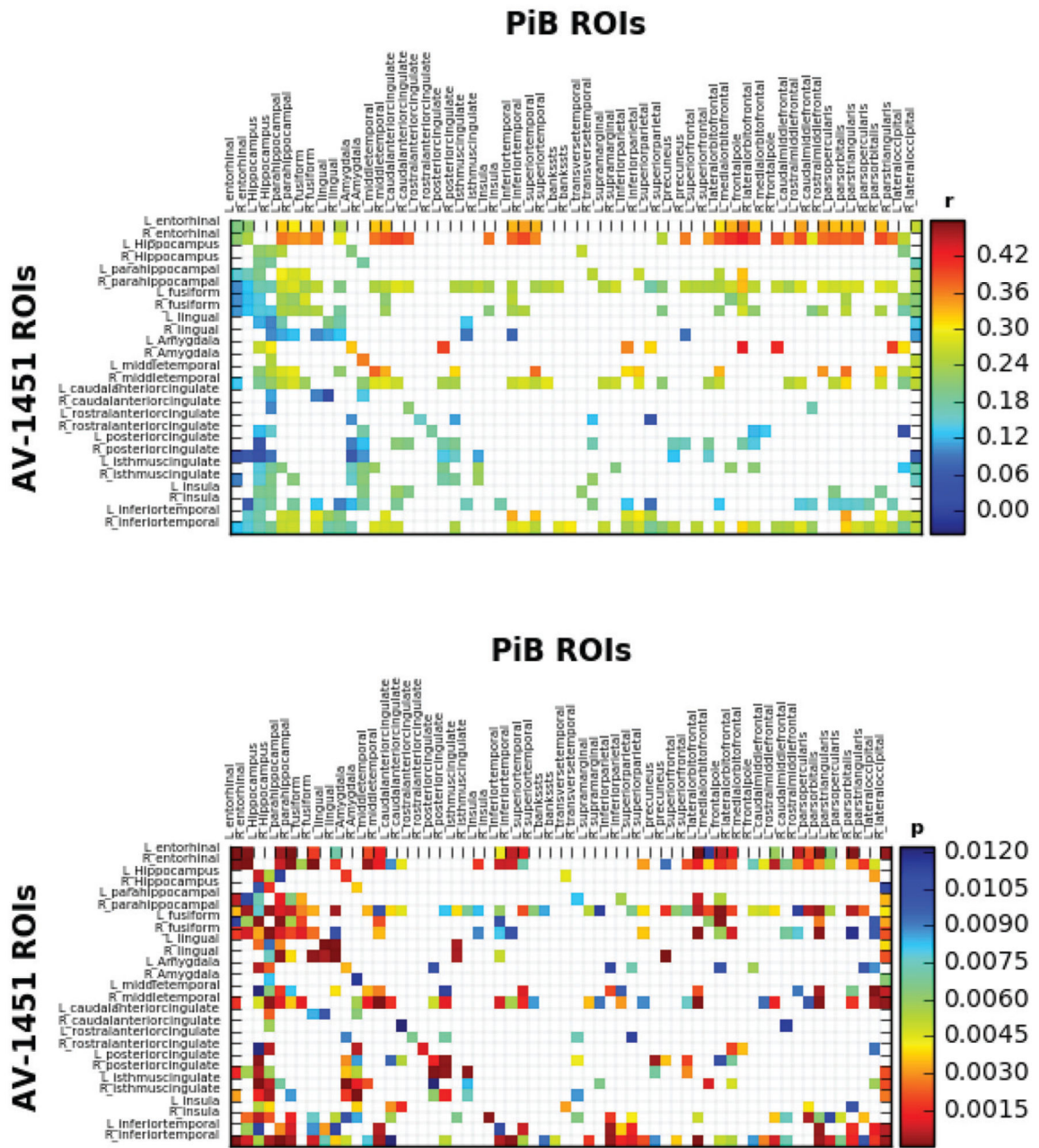
Figure 2.
PiB and AV-1451 intra-tracer correlation matrix (top panel) and *p*-value matrix (bottom panel) for the region level analysis.
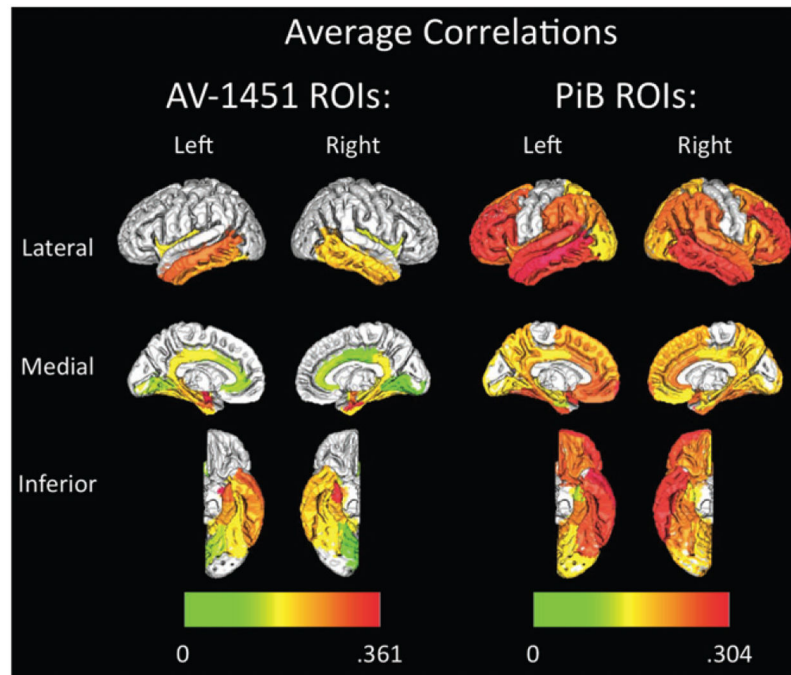
Figure 3.
AV-1451 (left) and PiB (right) average inter-tracer region-of-interest correlation map imposed on a template brain. Color bars at bottom indicate average of all significant ($p <$ 0.05) correlation coefficients, when testing the associations between a given region in one modality (e.g., left inferior temporal AV-1451) with all tested brain regions in the other modality (e.g., all Braak I-V PiB).
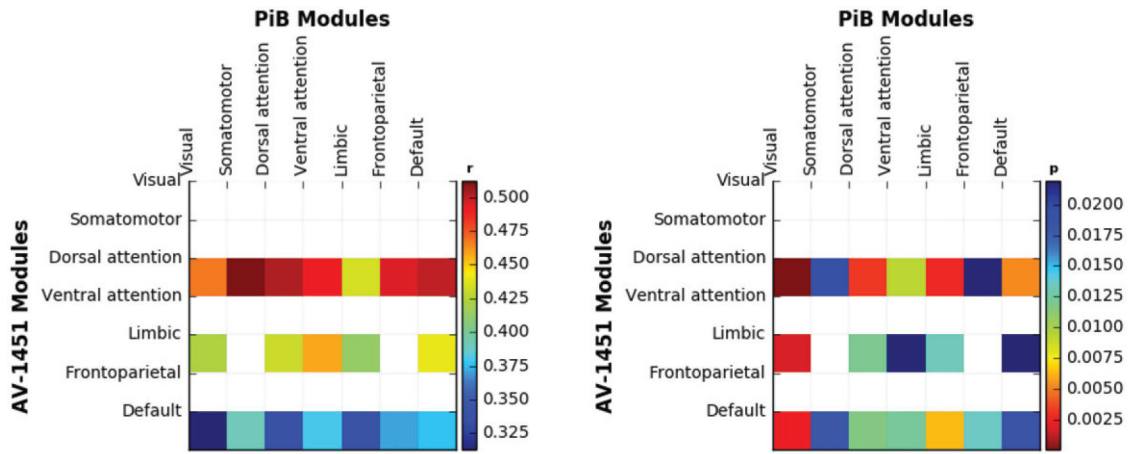
Figure 4.
PiB and AV-1451 intra-tracer correlation matrix (left panel) and *p*-value matrix (right panel) for the module level analysis.

Table 1.

Empirical FDR and empirical power, in percentages, for the proposed testing procedure.

Normal distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
Empirical FDR						
Xie and Kang	0.0	0.0	0.0	0.1	0.0	0.0
Sparse CCA	92.3	18.9	15.8	91.5	17.3	11.1
Our test	4.8	3.3	4.5	4.6	3.6	4.7
Empirical power						
Xie and Kang	18.8	3.4	13.1	35.2	4.4	16.8
Sparse CCA	22.3	11.2	3.8	24.0	1.5	0.8
Our test	96.3	40.3	57.2	96.3	41.0	57.0

t-distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
Empirical FDR						
Xie and Kang	0.0	0.0	0.0	0.0	0.0	0.0
Sparse CCA	91.6	18.0	11.0	93.6	14.4	8.8
Our test	2.8	2.3	3.0	3.0	2.3	2.7
Empirical power						
Xie and Kang	7.1	3.0	11.6	16.5	3.1	13.7
Sparse CCA	21.4	8.4	3.1	24.0	1.8	0.9
Our test	85.2	23.2	37.3	85.5	21.8	38.4

NOTES: It is also compared with the testing method of Xie and Kang (2017) and sparse CCA. The results are based on 100 data replications. The significance level is set at $\alpha = 5\%$. The sample size is $n = 100$.

Table 2.

Empirical FDR and empirical power, in percentages, for the proposed testing procedure.

Normal distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
Empirical FDR						
Xie and Kang	0.0	0.0	0.0	0.0	0.0	0.0
Sparse CCA	92.7	15.2	9.2	91.5	12.3	10.2
Our test	4.1	2.9	4.1	4.0	3.1	4.5
Empirical power						
Xie and Kang	88.5	10.8	37.4	94.7	20.6	49.8
Sparse CCA	27.5	7.3	3.0	26.4	1.1	0.8
Our test	100.0	90.2	95.3	100.0	91.5	95.5

t-distribution	Scenario 1			Scenario 2		
Covariance	Struct-1	Struct-2	Struct-3	Struct-1	Struct-2	Struct-3
Empirical FDR						
Xie and Kang	0.0	0.0	0.0	0.0	0.0	0.0
Sparse CCA	91.8	15.8	3.5	89.8	7.5	8.1
Our test	2.6	2.1	2.8	2.7	2.2	2.7
Empirical power						
Xie and Kang	59.9	6.2	23.7	76.8	11.4	32.4
Sparse CCA	29.2	10.6	0.2	25.8	0.5	0.9
Our test	99.8	81.4	88.0	99.8	82.1	88.3

NOTES: It is also compared with the testing method of Xie and Kang (2017) and sparse CCA. The results are based on 100 data replications. The significance level is set at $\alpha = 5\%$. The sample size is $n = 150$.