



Statistical methods and challenges in connectome genetics

Dustin Pluta^a, Zhaoxia Yu^a, Tong Shen^a, Chuansheng Chen^b, Gui Xue^c,
Hernando Ombao^{d,*}

^a Department of Statistics, University of California, Irvine, USA

^b Department of Psychology and Social Behavior, University of California, Irvine, USA

^c State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, PR China

^d Statistics Program, King Abdullah University of Science and Technology, Saudi Arabia



ARTICLE INFO

Article history:

Available online 10 March 2018

Keywords:

Connectivity
Genetics

ABSTRACT

The study of genetic influences on brain connectivity, known as *connectome genetics*, is an exciting new direction of research in imaging genetics. We here review recent results and current statistical methods in this area, and discuss some of the persistent challenges and possible directions for future work.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The human brain is one of the most intensely studied objects in modern science, yet much about its structure and function is still poorly understood. Despite the efforts of generations of researchers, and the creation of numerous disciplines specializing in particular aspects of the brain, the complexity of neurophysiology and human behavior have stymied attempts to develop a comprehensive model to describe brain structure, brain function, and cognition. Imaging genetics is a relatively recent direction of study, which seeks to determine the genetic factors that significantly influence the brain and behavior. Initially motivated by the study of neurological diseases such as Alzheimer's disease (AD) and schizophrenia, the scope of imaging genetics has been expanded to consider nearly all aspects of the brain and cognition.

We here review recent developments in statistical methods for imaging genetics, with a specific focus on the analysis of genetic influences on brain connectivity, sometimes referred to as “connectome genetics”. To date, relatively few studies have considered the genetic influences of brain connectivity, although there is growing interest in this area due to continued findings supporting the importance of brain connectivity for many aspects of cognition. An important survey of this topic was given a few years ago in [Thompson et al. \(2013\)](#), which introduces the essential concepts and modeling approaches for connectome genetics, and summarizes the key results in the field up to that point. Following the comprehensive exposition given there, this review presents a selection of recent methods and results in connectome genetics. Interested readers may also refer to the contemporaneous review by [Nathoo et al. \(2017\)](#) which considers other general statistical methods for imaging genetics.

2. Brain connectivity

Commonly used technologies in current neuroimaging studies include structural and functional magnetic resonance imaging (MRI and fMRI), electroencephalogram (EEG), magnetoencephalogram (MEG), and diffusion tensor imaging (DTI).

* Corresponding author at: Statistics Program, King Abdullah University of Science and Technology, Saudi Arabia.
E-mail address: hernando.ombao@kaust.edu.sa (H. Ombao).

Since each of these imaging modalities measures a different set of neural features or phenomena, they may differ in their effectiveness in answering particular scientific questions. For instance, fMRI measures changes in the blood oxygen level over time in over 5×10^5 voxels throughout the brain (Lindquist, 2008), with very high spatial resolution (typically less than 3 mm), but poor temporal resolution (around 2s). In contrast, EEG measures the collective cortical activity of populations of neurons via 64–256 electrodes placed on the scalp (Ombao et al., 2016), allowing for a high temporal resolution (1 kHz) at the cost of spatial resolution. Instead of measuring dynamic brain activity, structural imaging methods provide information about the physical relationship of brain regions, such as DTI, which measures the structure and orientation of the brain's white matter fiber tracks via the diffusion of water molecules through the entire brain volume (Basser et al., 1994).

It is now generally understood that higher-level cognitive processing (e.g., in memory retrieval, decision making) critically depends on the interaction and transfer of information between many localized regions. Numerous different methods for calculating brain connectivity have been proposed, but the reliability, interpretation, and relationship of these different measures is not well established (Fiecas et al., 2013). There are three concepts of brain “connectivity” that have been of interest, namely *structural*, *functional*, and *effective* connectivities. Structural connectivity refers to the anatomical connections between brain regions, measured using DTI or structural MRI. Functional connectivity is a symmetric and undirected measure of concordant activity between brain regions, commonly calculated as Pearson's correlation, partial correlation, coherence, or partial coherence between the activation signals of two regions (Fiecas and Ombao, 2011). Modalities used for functional connectivity studies include fMRI, EEG, and MEG. By contrast, effective connectivity is a directed measure of how past activity in one region may influence the future activity of another region. Effective connectivity is closely related to Granger causality and is often estimated with a vector autoregressive (VAR) model (Gorrostieta et al., 2013; Chiang et al., 2017).

In studying the etiology of neurological diseases such as Alzheimer's, schizophrenia, and autism spectrum disorders, significant links with these forms of brain connectivity have been consistently shown (Woodward and Cascio, 2015). There is also evidence that dynamic characteristics of brain connectivity, and not just global measures of connectivity over an entire experiment, are also important for understanding brain function. Formally defining and modeling dynamic connectivity has motivated the development of new statistical methods. Current approaches for modeling dynamic connectivity include sliding window methods (e.g., Chang and Glover, 2010), graphical Bayesian modeling (Warnick et al., 2017) change-point detection via VAR models (Kirch et al., 2015), and hidden Markov switching-VAR models (Samdin et al., 2017). A full treatment of current dynamic connectivity methods is outside the scope of this review, but we briefly describe recent work in this area by our collaborators. Ting et al. (2017) proposed modeling dynamic connectivity with a so-called “regime-switching” factor model (SVAR). A three-step procedure for fitting this model (for a single subject) is described. (1) Compute initial estimates of connectivity subspaces shared across regimes using a stationary factor model; (2) apply a factor SVAR model, and identify regime boundaries with a switching Kalman filter to partition the neural signal into a small number of distinct states; (3) finally, use the low-dimensional factor representations for each regime to estimate within-regime effective connectivity. This model is able to detect abrupt changes in mental state, such as might occur in an experiment with varied cognitive demands, and is also able to estimate recurring states over the course of the experiment.

3. Imaging genetics

Human traits, including neural, cognitive, and behavioral traits, are shaped by genetics and environmental factors through a number of complex biological processes. In the past few decades, single nucleotide polymorphisms (SNPs) have become one of the most commonly measured forms of genetic data, as they are abundant (about 10 million in the human genome), stable, and easy to measure. Each SNP is a single DNA location (nucleotide) that exhibits relatively high variation between individuals, and are measured as *alleles*, which are typically encoded as a 0, 1, and 2 respectively, and treated numerically. Many complex traits are believed to depend on higher-order interactions between collections of SNPs, but determining these relationships is often difficult, and in practice it is often necessary to restrict attention to only the additive effects of SNPs. The proportion of variance of an observed trait that is accounted for by the additive effect of genome-wide SNPs is referred to as *narrow-sense heritability*, denoted h^2 .

Some hallmark successes in genetics have come from the study of prominent neurological and psychological diseases, such as Parkinson's disease, Alzheimer's disease, and schizophrenia. For each of these diseases, many significant genetic risk factors of relatively large effect have been found, which have in turn led to improved understanding of the mechanisms of action for these diseases, and have even resulted in promising gene therapies for Parkinson's disease (Palfi et al., 2014). Initially motivated by these promising results, the field of imaging genetics has expanded to now consider a number of different neural disorders and many facets of cognition in healthy individuals. We mention a few recent results to illustrate the breadth of current research. The study by Bohlken et al. (2016) found that global reductions in white matter, which underpin structural connectivity, are largely explained by genetic risk factors for schizophrenia. In very recent work by Sudre et al. (2017), a study of families with histories of attention-deficit/hyperactivity disorder (ADHD) found significant genetic heritability for the default mode, cognitive control, and ventral attention networks. An analysis of 161 twin pairs found significant heritability of anatomical connectivity measured with DTI, and identified significant differences in heritability across subnetworks (with average heritability over all white matter tracts estimated at around 30% (Shen et al., 2014). With data from 1320 unrelated, young, healthy adults, Ge et al. (2016) demonstrated that several characteristics of brain structure are genetically heritable, such as brain volume and neuroanatomical shape. Using resting-state fMRI data from twins and non-twin siblings, Vidaurre et al. (2017) found evidence of significant heritability for certain characteristics of dynamic connectivity patterns; this is one of the few studies to date to have considered the heritability of dynamic connectivity.

4. Statistical methods for imaging genetics

4.1. Association testing

For many cognitive phenotypes, there is currently little prior information on heritability estimates, important genetic factors, and functional relationships. Given this lack of prior knowledge regarding the likely presence of significant genetic influences, global association testing is a convenient approach to quickly evaluate the strength and significance of genetic factors on a particular observed trait. In genome-wide association studies (GWAS), the random effects (or variance components) model has become a standard approach to estimating the heritability of complex traits (Liu et al., 2007), popularized in recent years by the Genome-wide Complex Trait Analysis (GCTA) software (Yang et al., 2010). In this approach, the effect of an individual genetic variant is treated as a random quantity, rather than a fixed parameter, so that the genetic effects of a large number of SNPs are parameterized by a small number of variance components. As a result of this, the number of genetic variants studied is not required to be less than the sample size. This method is closely related to semi-parametric kernel regression methods, which can flexibly model non-additive effects of SNPs. Motivated by imaging genetics applications, this has also been extended to model the heritability of multi-dimensional phenotypes, such as brain activation or shape (Ge et al., 2016). For simple association testing with the variance components model, a score test is often used since it does not require fitting the model, and generally performs well when effect sizes are small.

Distance-based testing methods, which include the Mantel test, RV coefficient, and the distance-covariance test, are a common class of methods to test for global association between two sets of features using the pairwise distance or similarity between observations in each feature set. These methods can easily accommodate data of arbitrary dimension, and are generally straightforward to implement and interpret. In our recent work, we have found that Mantel's test, using the L_2 inner product to measure similarity, is equivalent to the global association test from the random effects model (Pluta et al., 2017). By adopting other similarity measures, such as weighted inner products or non-linear kernels, the Mantel test can also be used to conduct global association testing corresponding to other types of models or other proposed approaches, e.g. multivariate distance matrix regression, pseudo F -tests, and kernel-based tests. The flexibility of this method makes it possible to test a wide variety of relationships and data types in a common framework. For details on generalizations and applications of distance-based methods, see Schaid (2010), Pan (2011), Zapala and Schork (2012) and Ge et al. (2016).

Although distance-based methods have many attractive features, there is evidence that they are not suitable for high-dimensional data in some settings. Alternative methods referred to as the “sum of powered correlation” (SPC) and “sum of powered score” (SPU) tests have been proposed as more appropriate for the high-dimensional testing encountered in imaging genetics. Versions of these tests have been shown to perform better than distance-based methods in simulation (Xu et al., 2017), but it is not clear if these tests are superior in all settings.

4.2. Inferential modeling

Inferential modeling for imaging genetics is statistically challenging, in part due to the extremely high-dimension of the feature spaces, thus application of dimension reduction methods is often necessary for effective statistical analysis of imaging and genetics data. Commonly used data-driven methods include principal component analysis (PCA), and independent component analysis (ICA) (Liu et al., 2009). Although these methods have proven effective for many applications, that extracted leading components do not necessarily capture the most significant associations. In contrast, simultaneous dimension reduction methods, such as partial least squares (PLS), canonical correlation analysis (CCA), reduced rank regression (RRR), and parallel ICA (Ahn et al., 2015), perform dimension reduction jointly on responses and covariates, consider the joint relationships of the feature sets, and may be preferred for analysis of imaging genetics data. Regularization methods developed from ridge regression and the LASSO have been widely used in these contexts as well (Tibshirani, 1996).

Tensor regression has been developed as a formal model for multi-array data that utilizes a low-rank tensor decomposition (such as PARAFAC) to reduce dimensionality (Zhou et al., 2013; Guhaniyogi et al., 2016). This approach has a number of advantages. Because the low-rank representation is estimated from the model, the extracted features are more likely to retain significant relationships in the data. Furthermore, the tensor decomposition can directly incorporate some of the inherent structure in the data, such as the spatial organization in neuroimaging data. This is an important distinction from the commonly used PCA, which is a similar dimension reduction method, but which requires the imaging data to be vectorized, thereby destroying the spatial structure. Incorporating regularization methods, such as the LASSO, is also straightforward. Tensor regression is potentially well-suited to modeling connectivity, which is naturally written as a matrix, or possibly even dynamic connectivity, although these applications of tensor regression have not yet been thoroughly studied.

5. Discussion

Despite substantial technological and methodological progress in recent years, imaging genetics studies are still difficult and expensive to design, implement, and analyze. Perhaps the biggest statistical challenge in practice is that, while the dimensionality of the feature space is potentially in the hundreds of thousands, sample sizes for these studies are typically around 1000 subjects, and sometimes are much smaller. False positives, spurious results, and underpowered studies are serious issues when considering the reliability and reproducibility of imaging genetics results. As the number of relevant

studies continues to grow, meta-analyses may be valuable in validating and refining existing results. Similarly, as we develop a better understanding of the types of relationships exhibited by neuroimaging and genetic data, this prior information can be used to improve the efficiency of the analysis. For example, results from the genetics literature suggest including linkage disequilibrium or SNP minor allele frequencies to boost the signal from significant genetic markers (Lasky-Su, 2017).

Data integration methods within an individual study offer another strategy for improving efficiency and power, as different imaging modalities often measure complementary characteristics of brain function. One possibility is to simultaneously acquire neuroimaging measures, such as fMRI and EEG, and leverage the advantages of both in the analysis. However, there are technical issues with this approach. Alternatively, one could consider combining information from separately (non-contemporaneously) recorded data on the same subjects. Examples of this include fusing fMRI and DTI data in estimating brain connectivity, and using structural connectivity from DTI to construct informative priors when estimating resting state functional connectivity from fMRI (Kang et al., 2017; Chiang et al., 2017).

References

- Ahn, M., Shen, H., Lin, W., Zhu, H., 2015. A sparse reduced rank framework for group analysis of functional neuroimaging data. *Statist. Sinica* 25 (1), 295.
- Basser, P.J., Mattiello, J., LeBihan, D., 1994. Estimation of the effective self-diffusion tensor from the NMR spin echo. *J. Magn. Reson. B* 103 (3), 247–254.
- Bohlsen, M.M., Brouwer, R.M., Mandl, R.C., Van den Heuvel, M.P., Hedman, A.M., De Hert, M., Cahn, W., Kahn, R.S., Pol, H.E.H., 2016. Structural brain connectivity as a genetic marker for schizophrenia. *JAMA Psychiatry* 73 (1), 11–19.
- Chang, C., Glover, G.H., 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 50 (1), 81–98.
- Chiang, S., Guindani, M., Yeh, H., Haneef, Z., Stern, J., Vannucci, M., 2017. A Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Hum. Brain Mapp.* 38, 1311–1332.
- Fiecas, M., Ombao, H., 2011. The generalized shrinkage estimator for the analysis of functional connectivity of brain signals. *Ann. Appl. Stat.* 5, 1102–1125.
- Fiecas, M., Ombao, H., van Lunen, D., Baumgartner, R., Coimbra, A., Feng, D., 2013. Quantifying temporal correlations: a test–retest evaluation of functional connectivity in resting-state fMRI. *Neuroimage* 65, 231–241.
- Ge, T., Reuter, M., Winkler, A.M., Holmes, A.J., Lee, P.H., Tirrell, L.S., Roffman, J.L., Buckner, R.L., Smoller, J.W., Sabuncu, M.R., 2016. Multidimensional heritability analysis of neuroanatomical shape. *Nat. Commun.* 7, 13291.
- Gorroitieta, C., Fiecas, M., Ombao, H., Burke, E., Cramer, S.C., 2013. Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. *Front. Comput. Neurosci.* 7 (159), 1–11.
- Guhaniyogi, R., Qamar, S., Dunson, D.B., 2016. Bayesian tensor regression. *arXiv:1509.06490*.
- Kang, H., Ombao, H., Fonnesebeck, C., Ding, Z., Morgan, V.L., 2017. A Bayesian double fusion model for resting-state brain connectivity using joint functional and structural data. *Brain Connectivity* 7, 219–227.
- Kirch, C., Muhsal, B., Ombao, H., 2015. Detection of changes in multivariate time series with application to eeg data. *J. Amer. Statist. Assoc.* 110, 1197–1216.
- Lasky-Su, J., 2017. Chapter 19 - statistical techniques for genetic analysis. In: Robertson, D., Williams, G.H. (Eds.), *Clinical and Translational Science*, second ed. Academic Press, pp. 347–362.
- Lindquist, M., 2008. The statistical analysis of fMRI data. *Statist. Sci.* 439–464.
- Liu, D., Lin, X., Ghosh, D., 2007. Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* 63 (4), 1079–1088.
- Liu, J., Pearson, G., Windemuth, A., Ruano, G., Perrone-Bizzozero, N.I., Calhoun, V., 2009. Combining fMRI and SNP data to investigate connections between brain function and genetics using parallel ICA. *Hum. Brain Mapp.* 30 (1), 241–255.
- Nathoo, F.S., Kong, L., Zhu, H., 2017. A review of statistical methods in imaging genetics. *arXiv preprint arXiv:1707.07332*.
- Ombao, H., Lindquist, M., Thompson, W., Aston, J., 2016. *Handbook of Statistical Methods for Neuroimaging*. CRC Press.
- Palfi, S., Gurruchaga, J.M., Ralph, G.S., Lepetit, H., Lavis, S., Buttery, P.C., Watts, C., Miskin, J., Kelleher, M., Deeley, S., et al., 2014. Long-term safety and tolerability of ProSavin, a lentiviral vector-based gene therapy for Parkinson's disease: a dose escalation, open-label, phase 1/2 trial. *The Lancet* 383 (9923), 1138–1146.
- Pan, W., 2011. Relationship between genomic distance-based regression and kernel machine regression for multi-marker association testing. *Genet. Epidemiol.* 35 (4), 211–216.
- Pluta, D., Shen, T., Xue, G., Chen, C., Yu, Z., Ombao, H., 2017. Adaptive mantel test for penalized inference, with applications to imaging genetics. *arXiv preprint arXiv:1712.07270*.
- Samdin, S.B., Ting, C.-M., Ombao, H., Salleh, S.-H., 2017. A unified estimation framework for state-related changes in effective brain connectivity. *IEEE Trans. Biomed. Eng.* 64 (4), 844–858.
- Schaid, D.J., 2010. Genomic similarity and kernel methods i: advancements by building on mathematical and statistical foundations. *Hum. Heredity* 70 (2), 109–131.
- Shen, K.-K., Rose, S., Frripp, J., McMahon, K.L., de Zubicaray, G.I., Martin, N.G., Thompson, P.M., Wright, M.J., Salvado, O., 2014. Investigating brain connectivity heritability in a twin study using diffusion imaging data. *NeuroImage* 100, 628–641.
- Sudre, G., Choudhuri, S., Szekely, E., Bonner, T., Goduni, E., Sharp, W., Shaw, P., 2017. Estimating the heritability of structural and functional brain connectivity in families affected by attention-deficit/hyperactivity disorder. *JAMA Psychiatry* 74 (1), 76–84.
- Thompson, P.M., Ge, T., Glahn, D.C., Jahanshad, N., Nichols, T.E., 2013. Genetics of the connectome. *Neuroimage* 80, 475–488.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 267–288.
- Ting, C.-M., Ombao, H., Samdin, S.B., Salleh, S.-H., 2017. Estimating time-varying effective connectivity in high-dimensional fMRI data using regime-switching factor models. *arXiv preprint arXiv:1701.06754*.
- Vidaurre, D., Smith, S.M., Woolrich, M.W., 2017. Brain network dynamics are hierarchically organized in time. *Proc. Natl. Acad. Sci.* 114 (48), 12827–12832.
- Warnick, R., Guindani, M., Erhardt, E., Allen, E., Calhoun, V., Vannucci, M., 2017. A Bayesian approach for estimating dynamic functional network connectivity in fMRI Data. *J. Amer. Statist. Assoc.* <http://dx.doi.org/10.1080/01621459.2017.1379404>.
- Woodward, N.D., Cascio, C.J., 2015. Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry* 72 (8), 743–744.
- Xu, Z., Xu, G., Pan, W., 2017. Adaptive testing for association between two random vectors in moderate to high dimensions. *Genet. Epidemiol.*
- Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al., 2010. Common snps explain a large proportion of the heritability for human height. *Nat. Genet.* 42 (7), 565–569.
- Zapala, M.A., Schork, N.J., 2012. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Front. Genet.* 3.
- Zhou, H., Li, L., Zhu, H., 2013. Tensor regression with applications in neuroimaging data analysis. *J. Amer. Statist. Assoc.* 108 (502), 540–552.