A Comparison of Neuroimaging Software and a Contour Inference Method for analysis of Task-fMRI data

by

Alexander Bowring

St Catherine's College

Submitted to the University of Oxford

for the degree of

Doctor of Philosophy

Nuffield Department of Population Health

October 2019

Contents

Acknowledgments

Declarations

I, Alexander Bowring, hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in these, or any other Universities. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

- The work presented in Chapter 3 has been published in the *NeuroImage* journal (Bowring et al., 2018). This work was presented at the *Organization for Human Brain Mapping* (OHBM) Annual Meetings in 2017 and 2018. At the OHBM 2018 Annual Meeting, this work was the recipient of an oral presentation and Merit Abstract Award.
- The work presented in Chapter 4 has been published in the *NeuroImage* journal (Bowring et al., 2018). This work was presented at the OHBM Annual Meeting in 2017, where it was the recipient of an oral presentation.
- The work presented in Chapter 5 is based on a pre-printed manuscript.

Alexander Bowring
September 2019

Abstract

Over the last three decades, Functional Magnetic Resonance Imaging (fMRI) has rapidly progressed to become the primary tool for human brain mapping. Recently however, considerable attention within the field has been directed towards data-sharing and open science initiatives. This has been driven by a growing apprehension about the reproducibility of findings within the neuroimaging literature, amid concerns that current inference procedures are often misused or misinterpreted such that the overall scientific conclusions become distorted. One aspect specific to neuroimaging pinpointed as a cause for poor reproducibility is the high flexibility of a typical fMRI workflow. In the first part of this thesis, we investigate how the choice of software package used to conduct a statistical analysis can influence the group-level results of a task-fMRI study. We use publicly shared data from three published task-fMRI studies, and reanalyze each study within the three main neuroimaging software packages, AFNI, FSL and SPM, using parameteric and nonparametric inference. All information on how to process, analyze, and model each dataset we obtain from the publications. We use a variety of quantitative and qualitative comparison methods to gauge the scale of variability in our results and assess fundamental differences between each software package. While qualitatively we find broad similarities between packages, we also discover marked differences, such as Dice similarity coefficient values ranging from 0.000 to 0.743 in comparisons of thresholded statistic maps between software. We discuss the challenges involved in our replication attempt, while also utilizing open science tools in an effort to make our own research reproducible. In the second part of this thesis, we extend a contour inference method initially proposed by *Sommerfeld*, *Sain*, *and Schwartzman (2018) (SSS)* to develop spatial confidence sets (CSs) on clusters found in thresholded blood-oxygen-level dependent (BOLD) effect size maps. While traditional inferences based on hypothesis testing indicate where the null, i.e. an effect size of zero, can be rejected, the CSs give statements about where effect sizes exceed a *positive* threshold analogous to confidence intervals simultaneously across the entire brain. We make advancements to theoretical aspects and implementation of contour inference to improve the method's finite-sample performance. We extend the wild bootstrap theory presented in *SSS*, proposing a method based on the t-bootstrap, and recommend that the bootstrapped residuals are multiplied by Rademacher variables instead of Gaussian variables. We also develop a linear interpolation method for computing the topological boundary over which the bootstrap is applied. Notably, we demonstrate that

the framework used in *SSS* for assessing simulations manifests considerable positive bias in the simulation results, and propose our own novel construction to solve this issue. In the final part of this thesis, we make further theoretical developments to contour inference so that the method can operate on the Cohen's *d* and partial R^2 effect sizes commonly reported at the end of a neuroimaging study. For the second and third parts of this thesis, we carry out intensive Monte Carlo simulations on synthetic 3D data to investigate the accuracy of contour inference on signals representative of fMRI activation clusters. We also demonstrate the method on two 'big' fMRI datasets, obtaining confidence sets to localize activation in functional data from the Human Connectome Project and UK Biobank.

CHAPTER 1

Introduction

Since its inception at the end of the twentieth century, functional Magnetic Resonance Imaging (fMRI) has experienced a meteoric rise to become the primary tool for human brain mapping. While many forms of the technique exist, introduction of the particular method based on the Blood Oxygenization Level Dependant (BOLD) effect has ultimately been the catalyst in elevating fMRI to such stature within the neuroimaging community. Taking advantage of the magnetic properties of oxygen-rich red blood cells, BOLD fMRI measures changes in blood oxygenization alongside cerebral blood flow and volume as a proxy to identify brain areas where elevated neuronal activity has occurred in response to a stimulus. While the relationship between the BOLD effect and neuronal activity is complex and remains controversial, it is the unique attributes of BOLD fMRI – in particular, its capacity for non-invasive recording of signals across the entire brain at a high spatial resolution – that set the technique apart from other scanning methods.

However, BOLD fMRI is also a *noisy* process. The MR signals researchers set out to measure during a scanning session are corrupted by artefacts from both the imaging hardware and the physiology of the participant. Examples of scanner noise include inhomogeneities of the magnetic field that can cause spatial distortion or blurring in the MR image, and scanner drift characterized by temporal degradation of the signal. Physiological noise induced by subject motion, respiration, and heartbeat exacerbate the problem.

Because of the low signal-to-noise, researchers must apply a series of statistical techniques to find meaning in the data. This usually entails carrying out a number of preprocessing, modelling and analysis steps that together constitute the fMRI processing pipeline. The fundamental objectives of preprocessing are to standardize brain locations across participants, to apply methods ensuring that the data conform to statistical assumptions required for analysis, and to reduce the influence of the

aforementioned noise artefacts present in the data. This is achieved by conducting a number of steps, including slice-timing correction, motion correction, normalization, registration of the functional data to an anatomical template, and spatial smoothing.

For task-based fMRI, a mass-univariate approach is utilized to model the data. During the scanning session, functional data are acquired in the form of voxels – cubic intensity units that partition the brain comparable to the way in which pixels partition a computer screen. Each voxel's time-series is considered independently within the general linear model framework as a combination of signal components. To evaluate the effect of an experimental task condition relative to a baseline condition, hypothesis testing is performed at each voxel to compute a statistical parametric map of t-statistic values. Here, the behaviour of the signal under the null hypothesis of no activation is estimated using either a parametric approach, appealing to the body of mathematics known as Random Field Theory, or a nonparametric approach, where permutation methods are applied to estimate the null-distribution directly from the data. Finally, the statistical parametric map is thresholded to localize brain function.

While we have provided a brief overview of the fMRI analysis pipeline, it is notable that there is not a general consensus as to how each particular analysis step should be carried out. Consequently, researchers have the freedom to make many choices during an analysis, such as how much smoothing is applied to the data, or how the heamodynamic response of blood flow to active neuronal tissues is modelled. However, this 'methodological plurality' comes with a drawback. While conceptually similar, two different analysis pipelines applied on the same dataset may not produce the same scientific results, and mathematical modelling has shown that the high analytic flexibility associated with fMRI can potentially distort the final scientific findings of an investigation (Ioannidis, 2005, why most research findings are false). The problem is, with so many statistically valid methodological strategies available, if you try them all you are likely to find *something*. Combined with further issues such as p-hacking and publication bias – where there has been evidence to suggest that studies finding a significant effect are disproportionately represented in the fMRI literature – these conditions have created the perfect storm: In recent years, many attempts to replicate the results of published fMRI studies have been unsuccessful, in what has been deemed as an ongoing reproducibility crisis within the field.

The degree to which varying methodological decisions can lead to discrepancies in observed results has been investigated extensively. Choices for each individual procedure in the analysis pipeline (for example, head-motion regression (Lund et al., 2005), temporal filtering (Skudlarski et al., 1999), and autocorrelation correction (Woolrich et al., 2001)) alongside the order in which these procedures are conducted

(Carp, 2013) can all deeply influence the final determined areas of brain activation. In perhaps the most comprehensive of such studies (Carp, 2012a), a single publicly available fMRI dataset was analyzed using over 6,000 unique analysis pipelines, generating 34,560 unique thresholded activation images. These results displayed a substantial degree of flexibility in both the sizes and locations of significant activation.

Alongside issues concerned with the flexibility of the analysis workflow, the statistical procedures carried out for fMRI inference have also come under intense scrutiny. Because statistical tests are conducted at each brain voxel independently, the p-values used to threshold the statistical parametric map are corrected to account for the large number of simultaneous comparisons being carried out and limit the expected number of voxels falsely declared as significant. This is almost always done using a false discovery rate correction procedure or a Bonferonni correction to limit the family-wise error rate of making at least one significant finding.

The importance of such statistical correction methods were made prominent within the neuroimaging community using a humorous example, where one author identified significant activation in the brain of a dead salmon after applying inference with uncorrected p-values. However, in recent times they have been a source of major controversy. In 2016, a shocking paper by *Eklund et. al* discovered that many fMRI software packages were incorrectly carrying out the multiple-correction procedures for clusterwise inferences, inflating the false-positive rate to up to 70%. In a damning blow to the field, the implications of this study brought into question the validity of thousands of published fMRI results. While the relevant software packages have been patched, deeper conceptual problems have been raised regarding the fMRI approach to inference.

CHAPTER 2

Background

2.1 The Study of Brain Function

2.2 Magnetic Resonance Imagery (MRI)

2.3 Task-based functional Magenetic Resonance Imagery (t-fMRI)

2.3.1 Overview

2.3.2 Pre-processing

2.4 Statistical Analysis: Subject-level

2.4.1 Parametric Methods

2.4.2 Nonparametric Methods

2.5 Statistical Analysis: Group-level

2.5.1 Parametric Methods

2.5.2 Nonparametric Methods

2.6 Reproducibility of fMRI Results

CHAPTER 3

Exploring the Impact of Analysis Software on Task-fMRI Results

3.1 Data and Analysis Methods

3.1.1 Study Description and Data Source

3.1.2 Data Analyses

3.1.3 Comparison Methods

3.1.4 Permutation Test Methods

3.2 Results

3.2.1 Cross-Software Variability for Parametric Inference

3.2.2 Cross-Software Varaibility for Non-Parametric Inference

3.2.3 Intra-Software Variability, Parametric vs Non-Parametric

3.3 Reproducibility

3.3.1 Scripting of Analysis and Figures

3.3.2 Results Sharing

3.4 Discussion

3.4.1 Limitations

3.5 Conclusion

CHAPTER 4

Spatial Confidence Sets for Task-fMRI Inference

CHAPTER 5

Contour Inference for Cohen's *d*

5.1 Theory

5.1.1 Transforming the Residual Field

5.2 Method

5.2.1 2D Simulations

5.2.2 3D Simulations

5.2.3 Application to UK Biobank Data

5.3 Results

5.3.1 2D Simulations

5.3.2 3D Simulations

5.3.3 UK Biobank Data

5.3.4 Comparison to Traditional Inference Procedures

5.4 Discussion

5.4.1 Limitations

5.5 Conclusion

CHAPTER 6

Conclusion and Future Work

Bibliography

A. P. Dempster, D. B. R. N. M. Laird (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1), 1–38.

Abney, M. (2015). Permutation testing in the presence of polygenic variation. *Genetic Epidemiology*, **39**(4), 249–258.

Allison, D. B., Neale, M. C., Zannolli, R., Schork, N. J., Amos, C. I., and Blangero, J. (1999). Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *American journal of human genetics*, **65**(2), 531–44.

Almasy, L. and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics*, **62**(5), 1198–211.

Amemiya, T. (1977). A note on a heteroscedastic model. *Journal of Econometrics*, **6**(3), 365 – 370.

Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *American journal of human genetics*, **54**(3), 535–43.

Ashburner, J. and Friston, K. J. (2000). Voxel-based morphometry?the methods. *NeuroImage*, **11**(6), 805 – 821.

B. Devlin, K. R. (1999). Genomic control for association studies. *Biometrics*, **55**(4), 997–1004.

Balding, D. (2006). A tutorial on statistical methods for population association studies. *Nat Rev Genet*, **7**(10), 781–791.

Basser, P., Mattiello, J., and LeBihan, D. (1994). {MR} diffusion tensor spectroscopy and imaging. *Biophysical Journal*, **66**(1), 259 – 267.

Blangero, J. and Almasy, L. (1997). Multipoint oligogenic linkage analysis of quantitative traits. *Genetic Epidemiology*, **14**(6), 959–964.

Blangero, J., Diego, V. P., Dyer, T. D., Almeida, M., Peralta, J., Kent, J. W., Williams, J. T., Almasy, L., and Göring, H. H. H. (2013). *A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees.*, volume 81. Academic Press.

Blokland, G. A., McMahon, K. L., Hoffman, J., Zhu, G., Meredith, M., Martin, N. G., Thompson, P. M., de Zubicaray, G. I., and Wright, M. J. (2008). Quantifying the heritability of task-related brain activation and performance during the n-back working memory task: A twin fmri study. *Biological Psychology*, **79**(1), 70 – 79.

Brouwer, R. M., Mandl, R. C., Peper, J. S., van Baal, G. C. M., Kahn, R. S., Boomsma, D. I., and Pol, H. E. H. (2010). Heritability of {DTI} and {MTR} in nine-year-old children. *NeuroImage*, **53**(3), 1085 – 1092.

Burggren, A., Zeineh, M., Ekstrom, A., Braskie, M., Thompson, P., Small, G., and Bookheimer, S. (2008). Reduced cortical thickness in hippocampal subregions among cognitively normal apolipoprotein e e4 carriers. *NeuroImage*, **41**(4), 1177 – 1183.

Buse, A. (1973). Goodness of fit in generalized least squares estimation. *The American Statistician*, **27**(3), 106–108.

Buse, A. (1979). Goodness-of-fit in the seemingly unrelated regressions model: a generalization. *Journal of Econometrics*, **10**.

Buse, A. (1984). Tests for additive heteroskedasticity: Goldfeld and quandt revisited. *Empirical Economics*, **9**(4), 199–216.

Cao, J. (1999). The size of the connected components of excursion sets of χ^2, t and F fields. *Advances in Applied Probability*, **31**(3), 579–595.

Cardon, L. R. and Palmer, L. J. (2003). Population stratification and spurious allelic association. *The Lancet*, **361**(9357), 598 – 604.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, **25**(3), pp. 573–578.

Chiang, M.-C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., Klunder, A. D., Toga, A. W., McMahon, K. L., de Zubicaray, G. I., Wright, M. J., Srivas-

tava, A., Balov, N., and Thompson, P. M. (2009). Genetics of brain fiber architecture and intellectual performance. *The Journal of Neuroscience*, **29**(7), 2212–2224.

Chiang, M.-C., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Hickie, I., Toga, A. W., Wright, M. J., and Thompson, P. M. (2011). Genetics of white matter development: A {DTI} study of 705 twins and their siblings aged 12 to 29. *NeuroImage*, **54**(3), 2308 – 2317.

Cho, Y. S., Go, M. J., Kim, Y. J., Heo, J. Y., Oh, J. H., Ban, H.-J., Yoon, D., Lee, M. H., Kim, D.-J., Park, M., *et al.* (2009). A large-scale genome-wide association study of asian populations uncovers genetic factors influencing eight quantitative traits. *Nature genetics*, **41**(5), 527–534.

Crainiceanu, C. (2008). Likelihood ratio testing for zero variance components in linear mixed models. In D. Dunson, editor, *Random Effect and Latent Variable Model Selection*, volume 192 of *Lecture Notes in Statistics*, pages 3–17. Springer New York.

Crainiceanu, C. M. and Ruppert, D. (2004a). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, **91**(1), 35 – 52.

Crainiceanu, C. M. and Ruppert, D. (2004b). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**(1), 165–185.

Crainiceanu, C. M. and Ruppert, D. (2004c). Restricted likelihood ratio tests in nonparametric longitudinal models. *Statistica Sinica*, **14**(3), 713–730.

den Braber, A., Bohlken, M. M., Brouwer, R. M., van 't Ent, D., Kanai, R., Kahn, R. S., de Geus, E. J. C., Hulshoff Pol, H. E., and Boomsma, D. I. (2013). Heritability of subcortical brain measures: A perspective for future genome-wide association studies. *NeuroImage*, **83C**, 98–102.

Dominicus, A., Skrondal, A., Gjessing, H., Pedersen, N., and Palmgren, J. (2006). Likelihood ratio tests in behavioral genetics: Problems and solutions. *Behavior Genetics*, **36**(2), 331–340.

Draper, N. and Stoneman, D. (1966). Testing for the Inclusion of Variables in Linear Regression by a Randomisation Technique. *Technometrics*, **8**(4), 695–699.

Drikvandi, R., Verbeke, G., Khodadadi, A., and Partovi Nia, V. (2013). Testing multiple variance components in linear mixed-effects models. *Biostatistics (Oxford, England)*, **14**(1), 144–59.

Ernst, M. D. (2004). Permutation Methods: A Basis for Exact Inference. *Statistical Science*, **19**(4), 676–685.

Falconer, D. and Mackay, T. (1996). *Introduction to Quantitative Genetics*. Longman.

Fischl, B. and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, **97**(20), 11050–11055.

Fisher, R. (1935). *The Design of Experiments*. Oliver and Boyd.

Fitzmaurice, G. M., Lipsitz, S. R., and Ibrahim, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics*, **63**(3), 942–6.

Freedman, D. and Lane, D. (1983). A nonstochastic interpretation of reported significance levels. *Journal of Business and Economic Statistics*, **1**(4), 292–298.

Freedman, D. a. (2007). How Can the Score Test Be Inconsistent? *The American Statistician*, **61**(4), 291–295.

Friston, K., Penny, W., Phillips, C., Kiebel, S., Hinton, G., and Ashburner, J. (2002). Classical and bayesian inference in neuroimaging: Theory. *NeuroImage*, **16**(2), 465 – 483.

Friston, K. J., Worsley, K. J., Frackowiak, R. S. J., Mazziotta, J. C., and Evans, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, **1**(3), 210–220.

Ganjgahi, H., Winkler, A. M., Glahn, D. C., Blangero, J., Kochunov, P., and Nichols, T. E. (2015). Fast and powerful heritability inference for family-based neuroimaging studies. *NeuroImage*, **115**, 256 – 268.

Genovese, C. R., Lazar, N. A., and Nichols, T. E. (2002). Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage*, **15**(4), 870–878.

Glahn, D. C., Thompson, P. M., and Blangero, J. (2007). Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Human Brain Mapping*, **28**(6), 488–501.

Goldfeld, S. and Quandt, R. (1965). Some Tests for Homoscedasticity. *Journal of the American Statistical …*, **60**(310), 539–547.

Good, P. (2005). *Permutation, Parametric and Bootstrap Tests of Hypotheses*. Springer.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**(2), 383–385.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**(358), 320–338.

Helgason, A., Yngvadóttir, B., Hrafnkelsson, B., Gulcher, J., and Stefánsson, K. (2005). An icelandic example of the impact of population structure on association studies. *Nature genetics*, **37**(1), 90–95.

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivieres, S., Jahanshad, N., Toro, R., Wittfeld, K., Abramovic, L., Andersson, M., Aribisala, B. S., Armstrong, N. J., Bernard, M., Bohlken, M. M., Boks, M. P., Bralten, J., Brown, A. A., Mallar Chakravarty, M., Chen, Q., Ching, C. R. K., Cuellar-Partida, G., den Braber, A., Giddaluru, S., Goldman, A. L., Grimm, O., Guadalupe, T., Hass, J., Woldehawariat, G., Holmes, A. J., Hoogman, M., Janowitz, D., Jia, T., Kim, S., Klein, M., Kraemer, B., Lee, P. H., Olde Loohuis, L. M., Luciano, M., Macare, C., Mather, K. A., Mattheisen, M., Milaneschi, Y., Nho, K., Papmeyer, M., Ramasamy, A., Risacher, S. L., Roiz-Santianez, R., Rose, E. J., Salami, A., Samann, P. G., Schmaal, L., Schork, A. J., Shin, J., Strike, L. T., Teumer, A., van Donkelaar, M. M. J., van Eijk, K. R., Walters, R. K., Westlye, L. T., Whelan, C. D., Winkler, A. M., Zwiers, M. P., Alhusaini, S., Athanasiu, L., Ehrlich, S., Hakobjan, M. M. H., Hartberg, C. B., Haukvik, U. K., Heister, A. J. G. A. M., Hoehn, D., Kasperaviciute, D., Liewald, D. C. M., Lopez, L. M., Makkinje, R. R. R., Matarin, M., Naber, M. A. M., Reese McKay, D., Needham, M., Nugent, A. C., Putz, B., Royle, N. A., Shen, L., Sprooten, E., Trabzuni, D., van der Marel, S. S. L., van Hulzen, K. J. E., Walton, E., Wolf, C., Almasy, L., Ames, D., Arepalli, S., Assareh, A. A., Bastin, M. E., Brodaty, H., Bulayeva, K. B., Carless, M. A., Cichon, S., Corvin, A., Curran, J. E., Czisch, M., de Zubicaray, G. I., Dillman, A., Duggirala, R., Dyer, T. D., Erk, S., Fedko, I. O., Ferrucci, L., Foroud, T. M., Fox, P. T., Fukunaga, M., Raphael Gibbs, J., Goring, H. H. H., Green, R. C., Guelfi, S., Hansell, N. K., Hartman, C. A., Hegenscheid, K., Heinz, A., Hernandez, D. G., Heslenfeld, D. J., Hoekstra, P. J., Holsboer, F., Homuth, G., Hottenga, J.-J., Ikeda, M., Jack Jr, C. R., Jenkinson, M., Johnson, R., Kanai, R., Keil, M., Kent Jr, J. W., Kochunov, P., Kwok, J. B., Lawrie, S. M., Liu, X., Longo, D. L., McMahon, K. L., Meisenzahl, E., Melle, I., Mohnke, S., Montgomery, G. W., Mostert, J. C., Muhleisen, T. W., Nalls, M. A., Nichols, T. E., Nilsson, L. G., Nothen, M. M., Ohi, K., Olvera, R. L., Perez-Iglesias, R., Bruce Pike, G., Potkin, S. G., Reinvang, I., Reppermund, S., Ri-

etschel, M., Romanczuk-Seiferth, N., Rosen, G. D., Rujescu, D., Schnell, K., Schofield, P. R., Smith, C., Steen, V. M., Sussmann, J. E., Thalamuthu, A., Toga, A. W., Traynor, B. J., Troncoso, J., Turner, J. A., Valdes Hernandez, M. C., van 't Ent, D., van der Brug, M., van der Wee, N. J. A., van Tol, M.-J., Veltman, D. J., Wassink, T. H., Westman, E., Zielke, R. H., Zonderman, A. B., Ashbrook, D. G., Hager, R., Lu, L., McMahon, F. J., Morris, D. W., Williams, R. W., Brunner, H. G., Buckner, R. L., Buitelaar, J. K., Cahn, W., Calhoun, V. D., Cavalleri, G. L., Crespo-Facorro, B., Dale, A. M., Davies, G. E., Delanty, N., Depondt, C., Djurovic, S., Drevets, W. C., Espeseth, T., Gollub, R. L., Ho, B.-C., Hoffmann, W., Hosten, N., Kahn, R. S., Le Hellard, S., Meyer-Lindenberg, A., Muller-Myhsok, B., Nauck, M., Nyberg, L., Pandolfo, M., Penninx, B. W. J. H., Roffman, J. L., Sisodiya, S. M., Smoller, J. W., van Bokhoven, H., van Haren, N. E. M., Volzke, H., Walter, H., Weiner, M. W., Wen, W., White, T., Agartz, I., Andreassen, O. A., Blangero, J., Boomsma, D. I., Brouwer, R. M., Cannon, D. M., Cookson, M. R., de Geus, E. J. C., Deary, I. J., Donohoe, G., Fernandez, G., Fisher, S. E., Francks, C., Glahn, D. C., Grabe, H. J., Gruber, O., Hardy, J., Hashimoto, R., Hulshoff Pol, H. E., Jonsson, E. G., Kloszewska, I., Lovestone, S., Mattay, V. S., Mecocci, P., McDonald, C., McIntosh, A. M., Ophoff, R. A., Paus, T., Pausova, Z., Ryten, M., Sachdev, P. S., Saykin, A. J., Simmons, A., Singleton, A., Soininen, H., Wardlaw, J. M., Weale, M. E., Weinberger, D. R., Adams, H. H. H., Launer, L. J., Seiler, S., Schmidt, R., Chauhan, G., Satizabal, C. L., Becker, J. T., Yanek, L., van der Lee, S. J., Ebling, M., Fischl, B., Longstreth Jr, W. T., Greve, D., Schmidt, H., Nyquist, P., Vinke, L. N., van Duijn, C. M., Xue, L., Mazoyer, B., Bis, J. C., Gudnason, V., Seshadri, S., Ikram, M. A., Initiative, T. A. D. N., Consortium, T. C., EPIGEN, IMAGEN, SYS, Martin, N. G., Wright, M. J., Schumann, G., Franke, B., Thompson, P. M., and Medland, S. E. (2015). Common genetic variants influence human subcortical brain structures. *Nature*, **520**(7546), 224–229.

Holmes, A. P., Blair, R. C., Watson, J. D., and Ford, I. (1996). Nonparametric analysis of statistic images from functional mapping experiments. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, **16**(1), 7–22.

Hopper, J. L. and Mathews, J. D. (1982). Extensions to multi- variate normal models for pedigree analysis. *Annals of Human Genetics*, **46**, 373–383.

Hua, X., Leow, A. D., Parikshak, N., Lee, S., Chiang, M.-C., Toga, A. W., Jr, C. R. J., Weiner, M. W., and Thompson, P. M. (2008). Tensor-based morphometry as a neuroimaging biomarker for alzheimer's disease: An {MRI} study of 676 ad, mci, and normal subjects. *NeuroImage*, **43**(3), 458 – 469.

Jahanshad, N., Kochunov, P. V., Sprooten, E., Mandl, R. C., Nichols, T. E., Almasy, L., Blangero, J., Brouwer, R. M., Curran, J. E., de Zubicaray, G. I., Duggirala, R., Fox, P. T., Hong, L. E., Landman, B. A., Martin, N. G., McMahon, K. L., Medland, S. E., Mitchell, B. D., Olvera, R. L., Peterson, C. P., Starr, J. M., Sussmann, J. E., Toga, A. W., Wardlaw, J. M., Wright, M. J., Pol, H. E. H., Bastin, M. E., McIntosh, A. M., Deary, I. J., Thompson, P. M., and Glahn, D. C. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the enigma-dti working group. *NeuroImage*, **81**(0), 455 – 469.

Jennrich, R. and Sampson, P. (1976). Newton-raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics*, **18**(1), 11–17.

Kadri, N. K., Guldbrandtsen, B., SÃÿrensen, P., and Sahana, G. (2014). Comparison of genome-wide association methods in analyses of admixed populations with complex familial relationships. *PLoS ONE*, **9**(3), e88926.

Kang, H. M., Zaitlen, N. a., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. **178**(3), 1709–23.

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. a., Kong, S.-Y., Freimer, N. B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, **42**(4), 348–54.

Kochunov, P., Glahn, D., Lancaster, J., Winkler, A., Smith, S., Thompson, P., Almasy, L., Duggirala, R., Fox, P., and Blangero, J. (2010). Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *NeuroImage*, **53**(3), 1109 – 1116.

Kochunov, P., Glahn, D., Nichols, T., Winkler, A., Hong, E., Holcomb, H., Stein, J., Thompson, P., Curran, J., Carless, M., Olvera, R., Johnson, M., Cole, S., Kochunov, V., Kent, J., and Blangero, J. (2011a). Genetic analysis of cortical thickness and fractional anisotropy of water diffusion in the brain. *Frontiers in Neuroscience*, **5**(120).

Kochunov, P., Glahn, D., Lancaster, J., Thompson, P., Kochunov, V., Rogers, B., Fox, P., Blangero, J., and Williamson, D. (2011b). Fractional anisotropy of cerebral white matter and thickness of cortical gray matter across the lifespan. *NeuroImage*, **58**(1), 41 – 49.

Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sproote, E., Nichols, T., Hong, L., Behrens, T., Andersson, E. J.and Yacoub, Ugurbil, K., Brouwer, C., Landman, B.,

Braber, A., Almassy, L., Fox, P., Olvera, R., Blangero, J., DC., G., and Van Essen, D. (2014a). Heritability of fractional anisotropy in human white matter: A comparison of human connectome project and enigma-dti data. *NeuroImage*. In Review.

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T. E., Mandl, R. C., Almasy, L., Booth, T., Brouwer, R. M., Curran, J. E., de Zubicaray, G. I., Dimitrova, R., Duggirala, R., Fox, P. T., Hong, L. E., Landman, B. A., Lemaitre, H., Lopez, L. M., Martin, N. G., McMahon, K. L., Mitchell, B. D., Olvera, R. L., Peterson, C. P., Starr, J. M., Sussmann, J. E., Toga, A. W., Wardlaw, J. M., Wright, M. J., Wright, S. N., Bastin, M. E., McIntosh, A. M., Boomsma, D. I., Kahn, R. S., den Braber, A., de Geus, E. J., Deary, I. J., Pol, H. E. H., Williamson, D. E., Blangero, J., van 't Ent, D., Thompson, P. M., and Glahn, D. C. (2014b). Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: Comparing meta and megaanalytical approaches for data pooling. *NeuroImage*, **95**(0), 136 – 150.

Koten, J. W., Wood, G., Hagoort, P., Goebel, R., Propping, P., Willmes, K., and Boomsma, D. I. (2009). Genetic contribution to variation in cognitive function: An fmri study in twins. *Science*, **323**(5922), 1737–1740.

Kremen, W. S., Prom-Wormley, E., Panizzon, M. S., Eyler, L. T., Fischl, B., Neale, M. C., Franz, C. E., Lyons, M. J., Pacheco, J., Perry, M. E., Stevens, A., Schmitt, J. E., Grant, M. D., Seidman, L. J., Thermenos, H. W., Tsuang, M. T., Eisen, S. a., Dale, A. M., and Fennema-Notestine, C. (2010). Genetic and environmental influences on the size of specific brain regions in midlife: the vetsa mri study. *NeuroImage*, **49**(2), 1213–23.

Lange, K. (2003). *Mathematical and statistical methods for genetic analysis*. Springer, 2nd edition.

Lange, K., Westlake, J., and Spence, M. A. (1976). Extensions to pedigree analysis. III. variance components by the scoring method. *Annals of Human Genetics*, **39**(4), 485–91.

Lee, O. E. and Braun, T. M. (2012). Permutation tests for random effects in linear mixed models. *Biometrics*, **68**(2), 486–493.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. (2011a). FaST linear mixed models for genome-wide association studies. *Nature Methods*, **8**(10), 833–837.

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., Heckerman, D., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., and Heckerman, D. (2011b). Improved

linear mixed models for genome-wide association studies. *Nature methods*, **8**(6), 833–5.

Listgarten, J., Lippert, C., and Heckerman, D. (2013). FaST-LMM-Select for addressing confounding from spatial structure and rare variants. *Nature Genetics*, **45**(5), 470–471.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, **74**(4), 817–827.

MacCluer, J. W., Blangero, J., Dyer, T. D., and Speer, M. C. (1997). GAW10: simulated family data for a common oligogenic disease with quantitative risk factors. *Genetic epidemiology*, **14**(6), 737–42.

Matthews, S. C., Simmons, A. N., Strigo, I., Jang, K., Stein, M. B., and Paulus, M. P. (2007). Heritability of anterior cingulate response to conflict: An fmri study in female twins. *NeuroImage*, **38**(1), 223 – 227.

McKay, D., Knowles, E., Winkler, A., Sprooten, E., Kochunov, P., Olvera, R., Curran, J., Kent, J. JackW., Carless, M., GÃűring, H., Dyer, T., Duggirala, R., Almasy, L., Fox, P., Blangero, J., and Glahn, D. (2014). Influence of age, sex and genetic factors on the human brain. *Brain Imaging and Behavior*, **8**(2), 143–152.

Meyer, K. (1985). Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, pages 153–165.

Molenberghs, G. and Verbeke, G. (2007). Likelihood Ratio, Score, and Wald Tests in a Constrained Parameter Space. *The American Statistician*, **61**(1), 22–27.

Morgan, B. J. T., Palmer, K. J., and Ridout, M. S. (2007). Negative Score Test Statistic. *The American Statistician*, **61**(4), 285–288.

Mumford, J. and Nichols, T. (2006). Modeling and inference of multisubject fmri data. *Engineering in Medicine and Biology Magazine, IEEE*, **25**(2), 42–51.

Nan M. Laird, J. H. W. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**(4), 963–974.

Neyman, J. and Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **231**, pp. 289–337.

Nichols, T. E. and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, **12**(5), 419–446.

Nichols, T. E. and Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, **15**(1), 1–25.

Nichols, T. E. and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, **15**(1), 1–25.

Ober, C., Abney, M., and McPeek, M. S. (2001). The genetic dissection of complex traits in a founder population. *The American Journal of Human Genetics*, **69**(5), 1068–1079.

Olvera, R., Bearden, C., Velligan, D., Almasy, L., Carless, M., Curran, J., Williamson, D., Duggirala, R., Blangero, J., and Glahn, D. (2011). Common genetic influences on depression, alcohol, and substance use disorders in mexican-american families. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, **156**(5), 561–568.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**(3), 545–554.

Pievani, M., Rasser, P., Galluzzi, S., Benussi, L., Ghidoni, R., Sabattoli, F., Bonetti, M., Binetti, G., Thompson, P., and Frisoni, G. (2009). Mapping the effect of {APOE} ?4 on gray matter loss in alzheimer's disease in vivo. *NeuroImage*, **45**(4), 1090 – 1098.

Pirinen, M., Donnelly, P., and Spencer, C. C. A. (2013). Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. *Ann. Appl. Stat.*, **7**(1), 369–390.

Polk, T. A., Park, J., Smith, M. R., and Park, D. C. (2007). Nature versus nurture in ventral visual cortex: A functional magnetic resonance imaging study of twins. *The Journal of Neuroscience*, **27**(51), 13921–13925.

Potkin, S. G., Turner, J. A., Guffanti, G., Lakatos, A., Fallon, J. H., Nguyen, D. D., Mathalon, D., Ford, J., Lauriello, J., Macciardi, F., and FBIRN (2009a). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia Bulletin*, **35**(1), 96–108.

Potkin, S. G., Guffanti, G., Lakatos, A., Turner, J. A., Kruggel, F., Fallon, J. H., Saykin, A. J., Orro, A., Lupoli, S., Salvi, E., Weiner, M., Macciardi, F., and for the Alzheimer's

Disease Neuroimaging Initiative (2009b). Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for alzheimer's disease. *PLoS ONE*, **4**(8), 1–15.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, **38**(8), 904–909.

Price, A. L., Zaitlen, N. A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews*, **11**(June), 459–463.

Pritchard, J. K., Stephens, M., Rosenberg, N. A., and Donnelly, P. (2000). Association mapping in structured populations. *The American Journal of Human Genetics*, **67**(1), 170 – 181.

Rao, C. R. (2008). *Linear Statistical Inference and its Applications*. John Wiley & Sons, Inc.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., Lavery, T., Kouyoumjian, R., Farhadian, S. F., Ward, R., and Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

Rimol, L. M., Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Fischl, B., Franz, C. E., Hagler, D. J., Lyons, M. J., Neale, M. C., Pacheco, J., Perry, M. E., Schmitt, J. E., Grant, M. D., Seidman, L. J., Thermenos, H. W., Tsuang, M. T., Eisen, S. a., Kremen, W. S., and Dale, A. M. (2010). Cortical thickness is influenced by regionally specific genetic factors. *Biological psychiatry*, **67**(5), 493–9.

Sabatti, C., Service, S. K., Hartikainen, A.-L., Pouta, A., Ripatti, S., Brodsky, J., Jones, C. G., Zaitlen, N. A., Varilo, T., Kaakinen, M., *et al.* (2009). Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature genetics*, **41**(1), 35–46.

Salimi-Khorshidi, G., Smith, S. M., and Nichols, T. E. (2010). Adjusting the effect of nonstationarity in cluster-based and TFCE inference. *NeuroImage*, **54**(3), 2006–2019.

Samuh, M. H., Grilli, L., Rampichini, C., Salmaso, L., and Lunardon, N. (2012). The use of permutation tests for variance components in linear mixed models. *Communications in Statistics - Theory and Methods*, **41**(16-17), 3020–3029.

Searle, S. (1971). *Linear Models*. Number v. 1 in Wiley Publication in Mathematical Statistics. John Wiley & Sons.

Searle, S. R., Casella, G., and McCulloch, C. E. (2009). *Variance components*, volume 391. John Wiley & Sons.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**(398), pp. 605–610.

Servin, B. and Stephens, M. (2007). Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLoS Genet*, **3**(7).

Shephard, N. (1993). Maximum likelihood estimation of regression models with stochastic trend components. *Journal of the American Statistical Association*, **88**(422), pp. 590–595.

Shephard, N. G. and Harvey, A. C. (1990). On the probability of estimating a deterministic component in the local level model. *Journal of Time Series Analysis*, **11**(4), 339–347.

Silvapulle, M. J. (1992). Robust wald-type tests of one-sided hypotheses in the linear model. *Journal of the American Statistical Association*, **87**(417), 156–161.

Silvapulle, M. J. and Silvapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association*, **90**(429), 342–349.

Smith, S. and Nichols, T. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, **44**(1), 83–98.

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., and Behrens, T. E. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, **31**(4), 1487 – 1505.

Stein, J. L., Hua, X., Morra, J. H., Lee, S., Hibar, D. P., Ho, A. J., Leow, A. D., Toga, A. W., Sul, J. H., Kang, H. M., Eskin, E., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M. J., Craig, D. W., Gerber, J. D., Allen, A. N., Corneveaux, J. J., Stephan, D. A., Webster, J., DeChairo, B. M., Potkin, S. G., Jr., C. R. J., Weiner, M. W., and Thompson, P. M. (2010a). Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in alzheimer's disease. *NeuroImage*, **51**(2), 542 – 554.

Stein, J. L., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., Huentelman, M. J., Craig, D. W., Gerber, J. D., Allen, A. N., Corneveaux, J. J., DeChairo, B. M., Potkin, S. G., Weiner, M. W., and Thompson, P. M. (2010b). Voxelwise genome-wide association study (vgwas). *NeuroImage*, **53**(3), 1160 – 1174. Imaging Genetics.

Stein, J. L., Medland, S. E., Vasquez, A. A., Hibar, D. P., Senstad, R. E., Winkler, A. M., Toro, R., Appel, K., Bartecek, R., Bergmann, O., Bernard, M., Brown, A. A., Cannon, D. M., Chakravarty, M. M., Christoforou, A., Domin, M., Grimm, O., Hollinshead, M., Holmes, A. J., Homuth, G., Hottenga, J.-J., Langan, C., Lopez, L. M., Hansell, N. K., Hwang, K. S., Kim, S., Laje, G., Lee, P. H., Liu, X., Loth, E., Lourdusamy, A., Mattingsdal, M., Mohnke, S., Maniega, S. M., Nho, K., Nugent, A. C., O'Brien, C., Papmeyer, M., Putz, B., Ramasamy, A., Rasmussen, J., Rijpkema, M., Risacher, S. L., Roddey, J. C., Rose, E. J., Ryten, M., Shen, L., Sprooten, E., Strengman, E., Teumer, A., Trabzuni, D., Turner, J., van Eijk, K., van Erp, T. G. M., van Tol, M.-J., Wittfeld, K., Wolf, C., Woudstra, S., Aleman, A., Alhusaini, S., Almasy, L., Binder, E. B., Brohawn, D. G., Cantor, R. M., Carless, M. A., Corvin, A., Czisch, M., Curran, J. E., Davies, G., de Almeida, M. A. A., Delanty, N., Depondt, C., Duggirala, R., Dyer, T. D., Erk, S., Fagerness, J., Fox, P. T., Freimer, N. B., Gill, M., Goring, H. H. H., Hagler, D. J., Hoehn, D., Holsboer, F., Hoogman, M., Hosten, N., Jahanshad, N., Johnson, M. P., Kasperaviciute, D., Kent, J. W., Kochunov, P., Lancaster, J. L., Lawrie, S. M., Liewald, D. C., Mandl, R., Matarin, M., Mattheisen, M., Meisenzahl, E., Melle, I., Moses, E. K., Muhleisen, T. W., Nauck, M., Nothen, M. M., Olvera, R. L., Pandolfo, M., Pike, G. B., Puls, R., Reinvang, I., Renteria, M. E., Rietschel, M., Roffman, J. L., Royle, N. A., Rujescu, D., Savitz, J., Schnack, H. G., Schnell, K., Seiferth, N., Smith, C., Steen, V. M., Valdes Hernandez, M. C., Van den Heuvel, M., van der Wee, N. J., Van Haren, N. E. M., Veltman, J. A., Volzke, H., Walker, R., Westlye, L. T., Whelan, C. D., Agartz, I., Boomsma, D. I., Cavalleri, G. L., Dale, A. M., Djurovic, S., Drevets, W. C., Hagoort, P., Hall, J., Heinz, A., Jack, C. R., Foroud, T. M., Le Hellard, S., Macciardi, F., Montgomery, G. W., Poline, J. B., Porteous, D. J., Sisodiya, S. M., Starr, J. M., Sussmann, J., Toga, A. W., Veltman, D. J., Walter, H., Weiner, M. W., Bis, J. C., Ikram, M. A., Smith, A. V., Gudnason, V., Tzourio, C., Vernooij, M. W., Launer, L. J., DeCarli, C., and Seshadri, S. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet*, **44**(5), 552–561.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, **50**(4), pp. 1171–1177.

Svishcheva, G. R., Axenovich, T. I., Belonogova, N. M., van Duijn, C. M., and Aulchenko,

Y. S. (2012). Rapid variance componentsâĂŞbased method for whole-genome association analysis. *Nature Genetics*, **44**(10), 1166–1170.

ter Braak, C. J. (1992). *Permutation Versus Bootstrap Significance Tests in Multiple Regression and Anova*, volume 376 of *Lecture Notes in Economics and Mathematical Systems*. Springer Berlin Heidelberg.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**(2), pp. 254–262.

Verbeke, G. and Molenberghs, G. (2007). What Can Go Wrong With the Score Test? *The American Statistician*, **61**(4), 289–290.

Voight, B. F. and Pritchard, J. K. (2005). Confounding from cryptic relatedness in case-control association studies. *PLoS Genet*, **1**(3).

Weir, B. S., Anderson, A. D., and Hepler, A. B. (2006). Genetic relatedness analysis: modern data and new challenges. *Nature Reviews Genetics*, **7**(10), 771–780.

Widmer, C., Lippert, C., Weissbrod, O., Fusi, N., Kadie, C., Davidson, R., Listgarten, J., and Heckerman, D. (2014). Further improvements to linear mixed models for genome-wide association studies. *Scientific reports*, **4**, 6874.

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., and Glahn, D. C. (2010). Cortical thickness or grey matter volume? the importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, **53**(3), 1135–1146.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., and Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, **92C**, 381–397.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, **12**(6), 900–918.

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics*, **46**(2), 100–6.

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., and Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, **38**(2), 203–8.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. a., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., and Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*, **42**(4), 355–360.

Zhou, X. and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature genetics*, **44**(7), 821–4.