Behavior Genetics (2019) 49:112–121
https://doi.org/10.1007/s10519-018-9936-9

ORIGINAL RESEARCH

A Fast Method for Estimating Statistical Power of Multivariate GWAS in Real Case Scenarios: Examples from the Field of Imaging Genetics

Baptiste Couvy-Duchesne[1,2,3] · Lachlan T. Strike[2] · Katie L. McMahon[4,5] · Greig I. de Zubicaray[6] · Paul M. Thompson[7] · Nicholas G. Martin[3] · Sarah E. Medland[3] · Margaret J. Wright[2,5]

Received: 7 December 2017 / Accepted: 19 October 2018 / Published online: 15 November 2018

Abstract

In GWAS of imaging phenotypes (e.g., by the ENIGMA and CHARGE consortia), the growing number of phenotypes considered presents a statistical challenge that other fields are not experiencing (e.g. psychiatry and the Psychiatric Genetics Consortium). However, the multivariate nature of MRI measurements may also be an advantage as many of the MRI phenotypes are correlated and multivariate methods could be considered. Here, we compared the statistical power of a multivariate GWAS versus the current univariate approach, which consists of multiple univariate analyses. To do so, we used results from twin models to estimate pertinent vectors of SNP effect sizes on brain imaging phenotypes, as well as the residual correlation matrices, necessary to estimate analytically the statistical power. We showed that for subcortical structure volumes and hippocampal subfields, a multivariate GWAS yields similar statistical power to the current univariate approach. Our analytical approach is as accurate but ~ 1000 times faster than simulations and we have released the code to facilitate the investigation of other scenarios, may they be outside the field of imaging genetics.

Keywords Twin models · Statistical power · Multivariate · Univariate · GWAS · MRI imaging

Introduction

The ENIGMA and CHARGE initiatives have been outstanding in uniting researchers from all over the world to perform large scale GWAS and identifying genetic variants contributing to head size (intracranial volume, ICV) and volumes of subcortical structures (Thompson et al. 2014). The sample size reached since the consortia began makes them the most powerful meta-analytic samples to study the genetics of the brain [see (Strike et al. 2015) for a review]. For example, the ENIGMA sample grew from ~ 8 k scans in the first GWAS on ICV

Edited by Tinca Polderman.

Sarah E. Medland and Margaret J. Wright have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10519-018-9936-9) contains supplementary material, which is available to authorized users.

✉ Baptiste Couvy-Duchesne
b.couvyduchesne@uq.edu.au

1 Institute of Molecular Bioscience, The University of Queensland, St. Lucia, Brisbane 4072, Australia

2 Queensland Brain Institute, The University of Queensland, Brisbane 4072, Australia

3 QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia

4 Herston Imaging Research Facility (HIRF), Queensland Institute of Technology, Brisbane 4006, Australia

5 Centre for Advanced Imaging, The University of Queensland, Brisbane 4072, Australia

6 Institute of Health and Biomedical Innovations, Queensland Institute of Technology, Brisbane 4059, Australia

7 Imaging Genetics Center, Keck School of Medicine, University of Southern California, Marina del Rey, Los Angeles, CA 90292, USA

and hippocampal volume (Stein et al. 2012), to ~ 40 k in the most recent analyses (combined with CHARGE and UK biobank) (Hibar et al. 2015; Adams et al. 2016) and we can expect the sample size to increase again in future analyses with a greater contribution from the biobanks (e.g. UK (Miller et al. 2016) and German biobanks (Greiser et al. 2014; Schram et al. 2014; Breteler et al. 2014). At the same time, the number of phenotypes considered has also increased (from 2 to 7 subcortical volumes to 68 cortical measurements) and we can expect even more phenotypes to be included in future projects (e.g. DTI or voxel-wise cortical morphology).

This raises the question of statistical power of GWAS analyses of magnetic resonance imaging (MRI) derived phenotypes, as the number of tests increases quickly. This led us to compare the statistical power of univariate and multivariate GWAS using realistic scenarios, to illustrate the potential and limitations of each approach.

Multivariate GWAS analyses often consist of a series of MANOVAs, which results in only one test of association per SNP, over all outcome variables considered. Such tests have been implemented in popular GWAS software such as PLINK (Ferreira and Purcell 2009; Purcell et al. 2007). Other multivariate approaches include MultiPhen (O'Reilly et al. 2012) that uses likelihood ratio test, and GEMMA, which further allows the modelling of participant relatedness (Zhou and Stephens 2014). In addition, Bayesian methods are available (Marchini et al. 2007; Stephens 2013), but controlling for familial or cryptic relatedness in these models may not be straightforward. Finally, Medland et al., (Medland and Neale 2010) proposed an integrated model that allows testing SNP effects on the common factor as well as on variable-specific factors, but it remains unclear which factors to test and how to handle the correction for multiple testing. The statistical power of the different multivariate approaches (see Supplementary Section 1 for a quick review) depends on the covariance structure of the phenotypes (van der Sluis et al. 2013; Galesloot et al. 2014; Cole et al. 1994; Minica et al. 2010), and, as nicely summarised by Zhou and Stephens (2014): "[…] in a GWAS setting, no single test will be the most powerful in detecting the many different types of genetic effects that could occur. It is possible to manufacture simulations so that any given test is most powerful".

Here, we aimed to calculate the statistical power of real-case multivariate GWAS of brain phenotypes, by integrating genetic and environmental correlations estimated from twin data. We chose to focus on MANOVA models as they are the direct multivariate equivalent of the univariate GWAS approach. In addition, unlike any other multivariate method, we can calculate the power analytically and avoid time-consuming simulations (Muller and Peterson 1984).

Materials and methods

MANOVA power calculation

Power calculation of multivariate linear models such as MANOVA may be seen as an extension of the power calculation in the univariate case (Muller and Peterson 1984). However, unlike the univariate case, we can choose between four test statistics: Roy's largest root (RLR), Hotelling-Lawley trace (HLT), Wilks' lambda (WL) [implemented in PLINK (Ferreira and Purcell 2009)] and Pillai-Bartlett trace (PB). We focused on the last three as they show the highest sensitivity (Olson 1974, 1976, 1979; Stevens 1979, 1980; Pillai and Jayachandran 1967) and can be well approximated by an F-distribution (Muller and Peterson 1984; Pillai 1956; Nagarsenker and Suniaga 1983; Rao 1973). Multivariate power calculations require specification of the sample size (N), design (X, SNP and covariates), the vector of effect sizes (β, effects of SNP on phenotypes) and a matrix of residual variance–covariance of the phenotypes (Σ, "residual" means here that the SNP effect has been removed) (Muller and Peterson 1984). From there, power can be estimated using non-central F-approximation whose degrees of freedom depend on the choice of test-statistic (Muller and Peterson 1984).

To obtain realistic β and Σ for the set of brain phenotypes considered, we used a Cholesky decomposition of a multivariate twin model (Fig. 1). This is arguably one of the most general multivariate models and allows

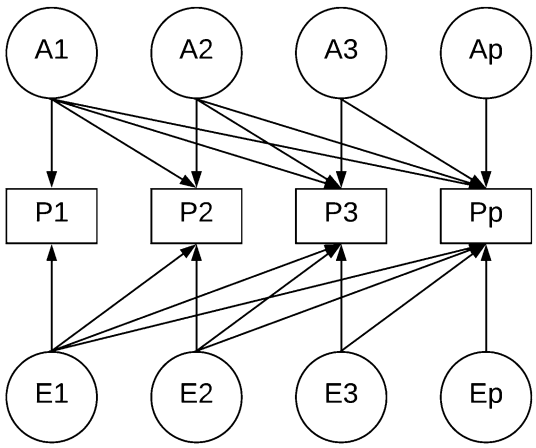

Fig. 1 General Cholesky decomposition of the genetic and environmental variance of the brain phenotypes. The *p* phenotypes are labels, P1 … Pp; we also model additive genetic (A) and environmental (E) sources of variance, with latent variables, A1 … Ap, E1 … Ep. In this model, all of the genetic variance is accounted for by the p additive genetic latent factors and the SNP can only affect one of the independent genetic factors (here, A2). All of the genetic and environmental factors are independent

as many independent genetic or environmental factors as phenotypes, while minimising the number of paths. In this model, all the genetic variance is accounted for by the p additive genetic latent factors and the SNP can only affect one of the independent genetic factors (e.g. Fig. 1). Using a twin sample, we can estimate the matrices of path coefficients a and e that describe the relationship between latent variables A, E and phenotypes P:

$$P = aA + eE. \tag{1}$$

In the example below (Fig. 1), the SNP affects the latent factor A2: $\mathrm{A2} = b \times \mathrm{SNP} + \varepsilon$. Thus, our vector of effect size can be written as:

$$\beta = \left(0,\ \ b \times a_{22},\ \ b \times a_{23},\ \ \ldots,\ b \times a_{2p}\right) \tag{2}$$

with $a_2 = (a_{22} \ldots a_{2p})$ the path coefficients of the factor A2, estimated from the twin model, where SNP is the categorical variable of the observed genotype at a particular locus (e.g., 0: genotype aa, 1: aA, 2: AA). We can then choose b based on the maximum path coefficient and the SNP minor allele frequency (MAF) so that we express the power as a function of maximal phenotypic variance explained (R^2) by the SNP, independently of the SNP MAF. For example, if a_{23} is the largest path coefficient (in absolute value) and we set that the SNP explains at most R^2 of the variance of one trait, we can use:

$$b = \frac{\sqrt{R^2}}{\mathrm{SD}\ (\mathrm{SNP}) \times a_{23}} \tag{3}$$

with $\mathrm{SD(SNP)} = \sqrt{2\mathrm{MAF}\ (1 - \mathrm{MAF})}$, and a_{23} the standardised path coefficient. Specifying the SNP effect size in terms of R^2 (variance explained) simplifies the analysis as it integrates the effect size b and the SNP MAF. Finally, we can calculate the residual phenotypic variance–covariance matrix Σ, using the path coefficients from which we have subtracted the SNP effect. We used the fact that $\mathrm{cov\ (P)} = \mathrm{a} \times \mathrm{t(a)} + \mathrm{e} \times \mathrm{t(e)}$ and that the phenotypic variance–covariance due to the SNP effect is $\mathrm{cov}\left(\mathrm{P_{SNP}}\right) = \frac{\sqrt{R^2}}{a_{23}} \times \mathrm{t}(a_2) \times a_2$. Thus, the residual variance–covariance matrix of phenotypes (after removing the SNP effect) is

$$\mathrm{cov}\left(\mathrm{P_{-SNP}}\right) = \mathrm{a} \times \mathrm{t(a)} + \mathrm{e} \times \mathrm{t(e)} - \mathrm{cov}\left(\mathrm{P_{SNP}}\right) \tag{4}$$

As the estimated β and Σ are specific to the genetic additive factor on which the SNP loads (Fig. 1), we have to calculate the statistical power for each genetic factor. As the genetic factors are independent we aggregated the factor-specific power by taking the mean or the weighted mean (using the % of variance explained by each genetic factor as weights).

Real case scenarios: subcortical volumes and hippocampal subfields

First, we considered seven subcortical volumes (summed over left and right hemisphere structures) processed using FreeSurfer 5.3 (Fischl 2012) using the ENIGMA protocols and QC (http://enigma.ini.usc.edu/protocols/imaging-protocols/). Then, we analysed volumetric data from 12 hippocampal subfields segmented using FreeSurfer 6.0 (Iglesias et al. 2015).

We fitted multivariate twin models with Cholesky decomposition in OpenMx (Boker et al. 2011) and extracted the standardised path coefficients to calculate the multivariate power. We calculated the power varying the sample size (up to N = 60,000) and the maximal SNP effect size (0.05–1% of variance explained).

For the univariate approach, we used a significance threshold of 5e−8/*neff* for univariate GWAS, with *neff* being the number of effectively independent phenotypes (Li and Ji 2005; Li et al. 2012), (code available at http://neurogenetics.qimrberghofer.edu.au/matSpDlite/). Such an approach has been used in prior MRI GWAS (Hibar et al. 2015) to ensure a FWER < 5% without over-correcting when performing tests over correlated variables. Here, we estimated *neff* to be six for the seven subcortical volumes (significance threshold of 8.3e−9) and 7 for the 12 hippocampal subfields (significance threshold of 7.1e−9), after regressing out all of the covariates (including ICV). For MANOVA power, we used the standard genome wide significance threshold for European populations (5e−8).

Multivariate twin modelling

We included 424 complete twin pairs (178 MZ, 246 DZ) with available volumes for the subcortical structures and hippocampal subfields. Participants included a slightly higher proportion of females (63.0%) and were on average 21.9 years old at scanning (SD = 3.3, range 15–29). Missingness (after QC) was less than 1% in all the variables considered.

T1 weighted structural scans were collected as part of the QTIM study (de Zubicaray et al. 2008), with TR = 1500 ms, TE = 3.35 ms, TI = 700 ms, flip angle = 8°, 256 or 240 (coronal or sagittal) slices, FOV = 240 mm, 256 × 256 × 256 (or 256 × 256 × 240) matrix, slice thickness = 0.9 mm and voxel size 0.9 mm^3.

Prior to twin modelling, age, age^2, age^3, sex, acquisition direction (coronal/sagittal), and ICV were regressed from the subcortical and hippocampal subfield volumes, and we used the residuals in the following analyses. Multivariate twin models were run in OpenMx (Boker et al. 2011) using Full Information Maximum Likelihood, which allows for missing observations.

Supplementary Tables 1 and 2 report the estimated standardised path coefficients used in the power calculation. Using the path coefficients, one can calculate genetic and environmental correlations between the subcortical volumes or hippocampal subfields (Figs. 2, 3). The subcortical volumes and hippocampal subfields are genetically correlated (rG values between 0.10 and 0.54 for subcortical volumes, −0.22 and 0.86 for hippocampal subfields); this supports considering them in multivariate GWAS. In addition, the path coefficients indicate multiple independent sources of additive variance (large path coefficients outside of the first genetic factor), which suggests that a GWAS of the first PC would not capture all of the relevant information. These results are consistent with the genetic correlations previously reported (Renteria et al. 2014), even if our estimates are systematically lower, which we attribute to controlling for ICV.

Sensitivity of power calculation

In our power calculation, we used estimated parameters (path coefficients), which asks about the robustness of our results across the range of plausible path coefficients. We performed sensitivity analyses, varying the path coefficients in our power calculations to reflect the uncertainty in the estimation. Thus, we randomly drew path coefficients values from their 95% confidence intervals and evaluated the impact of the new coefficients on our power calculation. We repeated this operation 100 times for each effect size, and visualised the uncertainty in power that results from the uncertainty in path coefficients. In addition, we checked if the power calculation remained the same after changing the order of the variables in the OpenMx model.

Next, we checked our analytical results against simulated statistical power, for the two main scenarios considered. We expect highly similar estimates of power as the F approximations used in the analytical approach should be accurate (Muller and Peterson 1984). In addition, we checked that MANOVA power in the particular case of a single phenotype was equivalent to those from a univariate model. Finally, we reported the computing time required for simulations in comparison to our analytical approach.

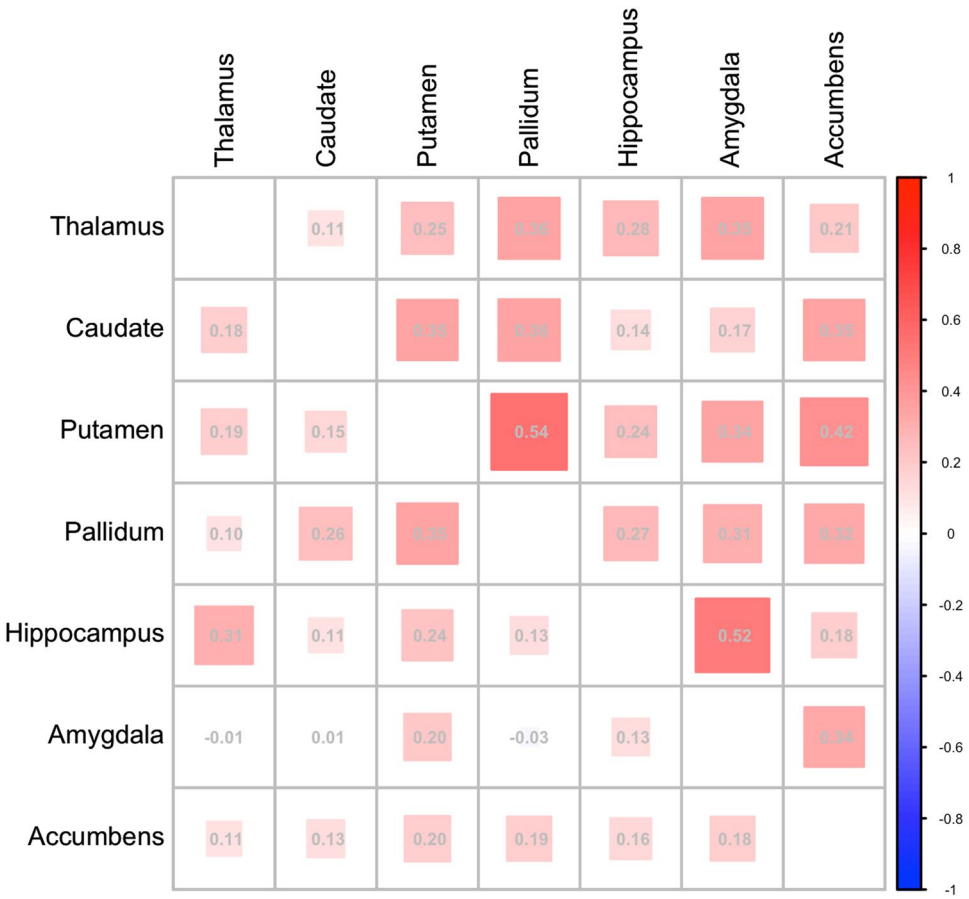

Fig. 2 Genetic (above diagonal) and environmental (below) correlations between subcortical volumes. Colour and size of the coloured squares indicate the strength of the correlations. No thresholding was applied to the matrix based on significance or correlation strength

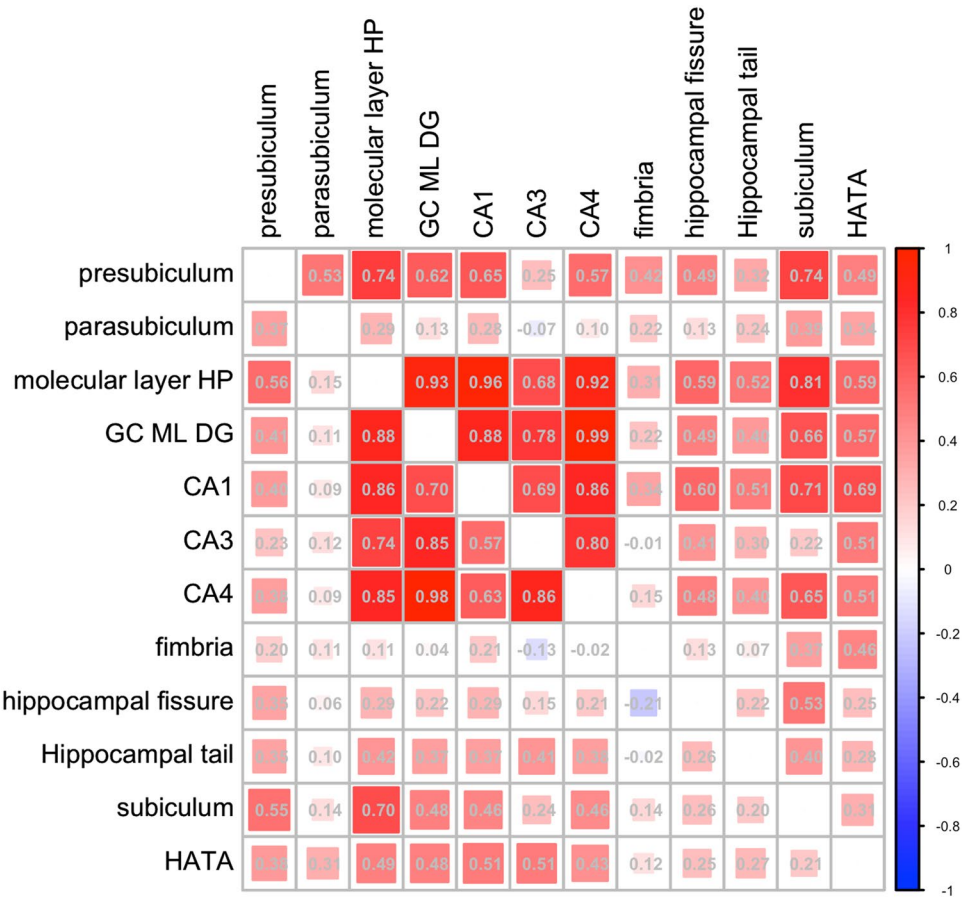

Fig. 3 Genetic (above diagonal) and environmental (below) correlations between hippocampus subfields. Colour and size of the coloured squares indicate the strength of the correlation. No thresholding was applied to the matrix based on significance or correlation strength. GC_ML_DG: Granule Cells of the Molecular Layer of the Dentate Gyrus. CA1-4: Cornu Ammonis areas 1–4. CA2 and CA3 are merged, as FreeSurfer 6.0 does not differentiate between them (Iglesias et al. 2015). HATA: Hippocampus-amygdala transition area

Results

Multivariate versus univariate power

The statistical power of subcortical GWAS approaches is summarised in Fig. 4. The multivariate approach should be comparable or even slightly less powerful in identifying genome-wide significant variants. For example, the current ENIGMA sample (N ~ 30,000) would detect an association approximately four out of ten times (power = 0.38) with the univariate approach (for a SNP explaining 0.1% of the variance in any subcortical volume). For the same parameters, a multivariate approach should detect the association approximately three out of ten times. In other words, 80% statistical power would be achieved for a sample size of 48,000 in a multivariate analysis, compared to 43,000 in the univariate case.

Similarly, multivariate GWAS should lead to similar statistical power than the univariate approach using hippocampal subfields (Fig. 5). For a SNP explaining 0.1% of the variance in the volume of one of the subfields, 80% power can be achieved with a sample of 46,000 using multivariate GWAS, against 44,000 in the univariate case.The choice of test statistic (HLT, WL or PB) did not impact the power calculation, consistent with prior research highlighting their equivalence in large samples (for $N > 10 \times p$, with p the number of dependent variables) (Muller and Peterson 1984).

The different weighting of factor-wise power in the multivariate calculation resulted in similar conclusions for the subcortical volumes scenario (Supplementary Figs. 1, 2). However, for the hippocampal subfields, the power of each genetic factor varied greatly and the multivariate power depended on the weights used to combine the factors (Fig. 6, Supplementary Figs. 3, 4, 5). Thus, the MANOVA does worse than in the univariate approach at detecting SNPs associated with the first additive genetic factor, but may provide more powerful at identifying SNPs influencing only a subset of hippocampus subfields (Fig. 6, Supplementary Fig. 5).

Sensitivity analysis

To evaluate the effect of uncertainty around path estimates on our power calculation, we re-ran the analysis 100 times randomly drawing path coefficients from their 95% confidence intervals. Variations in path coefficients resulted in small variations of multivariate power, over a hundred iterations (Supplementary Figs. 6, 7), with no consequences on

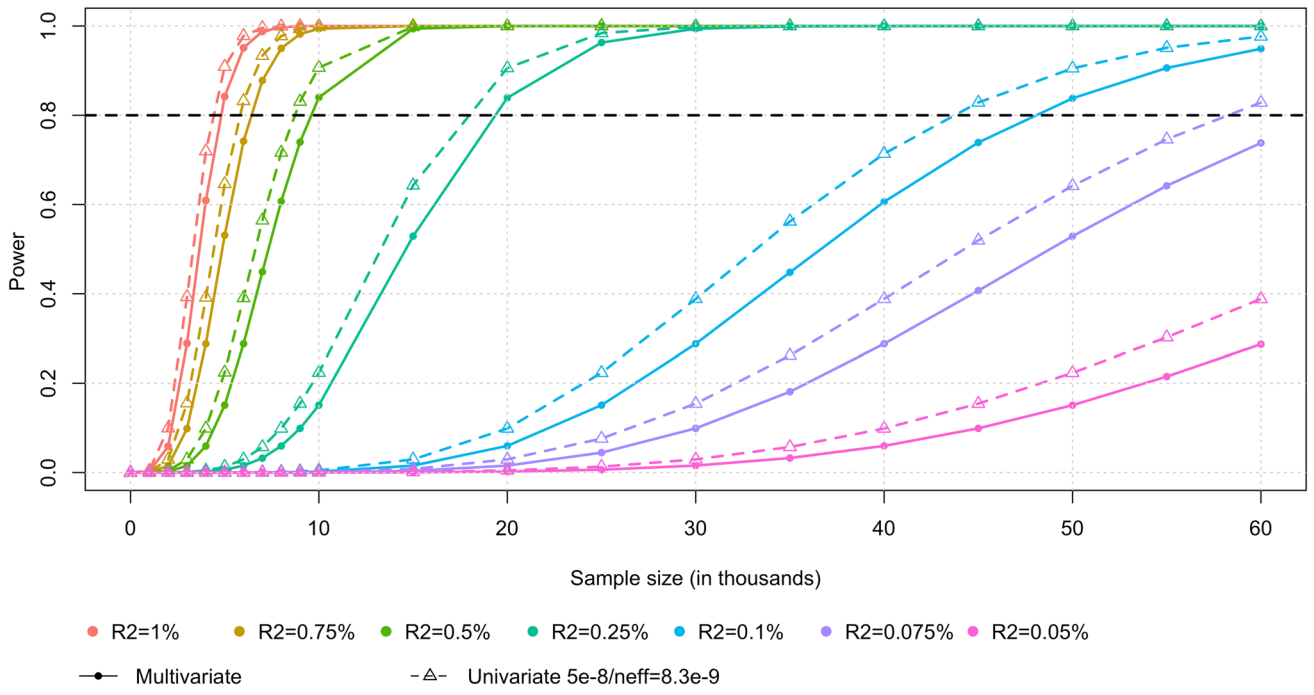

Fig. 4 Statistical power as a function of sample size for multivariate (solid lines) and univariate (dashed/dotted lines) GWAS on subcortical volumes. For all effect sizes and sample sizes, the multivariate approach confers greater statistical power than univariate approach. Multivariate power presented here corresponds to the variance-weighted power of genetic factor. Simply averaging the power across all factors did not change the conclusions (Supplementary Fig. 1)

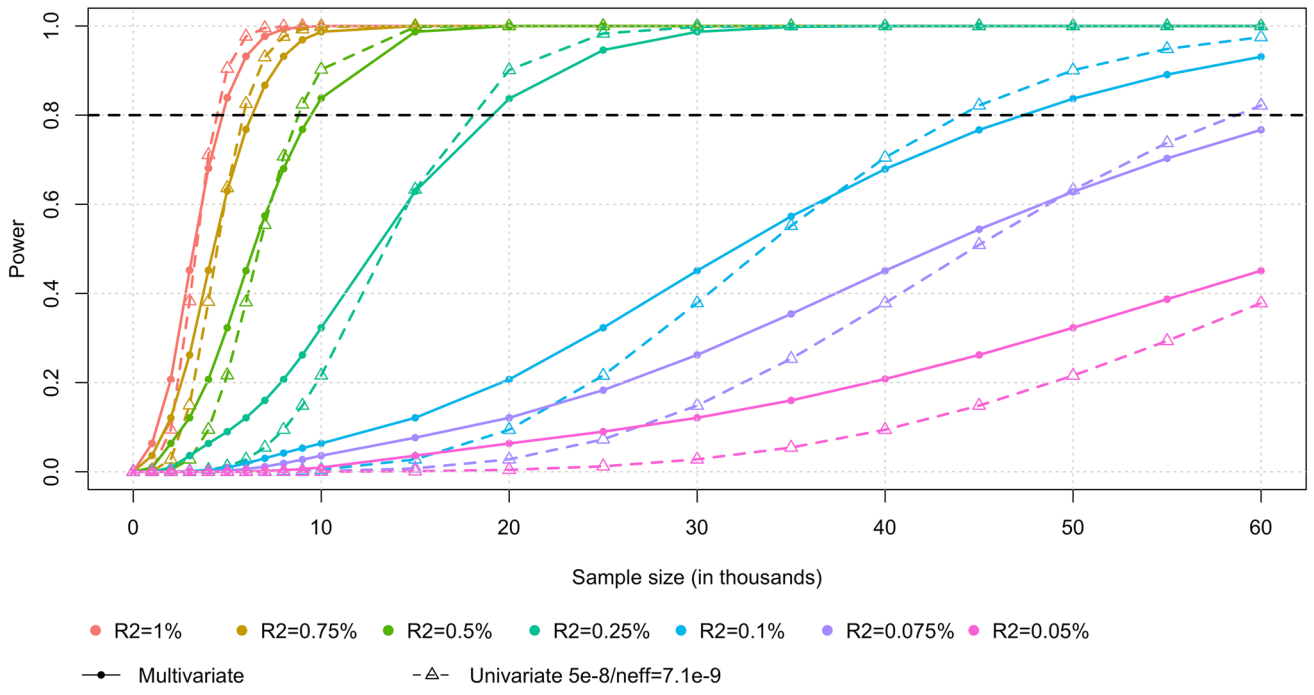

Fig. 5 Statistical power as a function of sample size for multivariate (plain lines) and univariate (dashed lines) GWAS on hippocampus subfields. For all effect sizes and sample sizes, the multivariate approach confers greater statistical power than univariate approach. Multivariate power presented here corresponds to the variance-weighted power of genetic factors. Averaging the power across factors suggested a possible increase of power (Supplementary Fig. 3, Fig. 6)

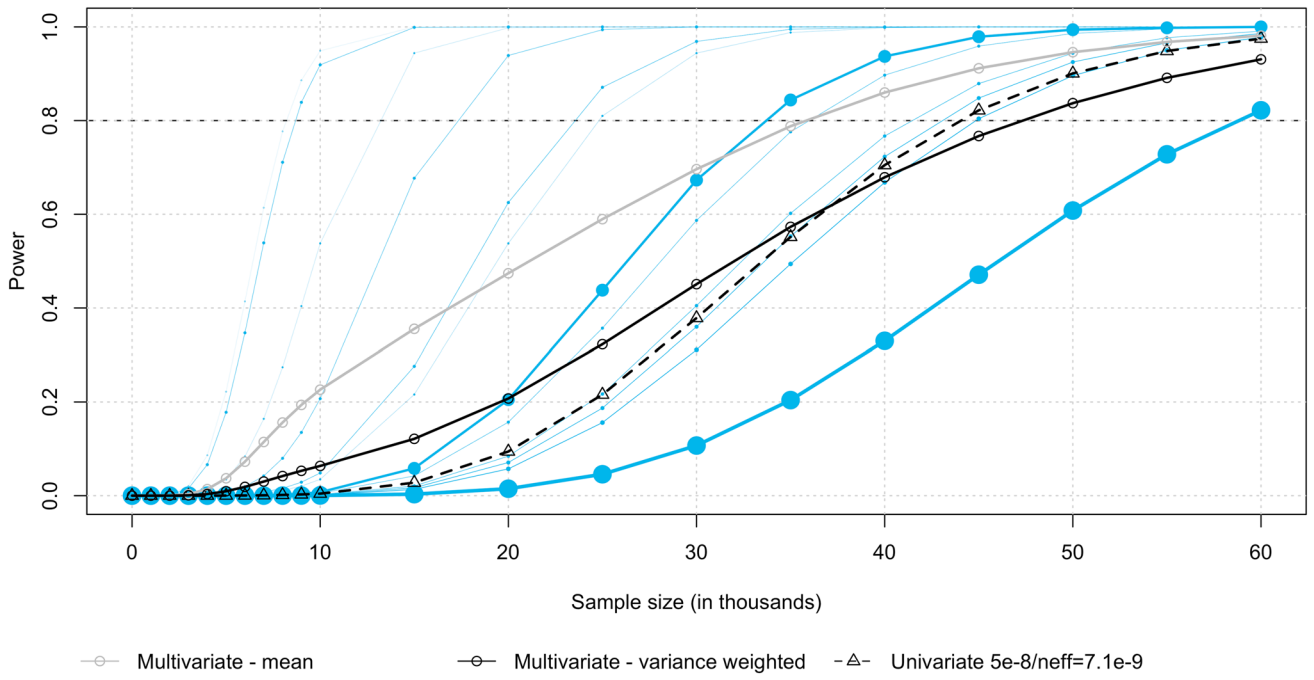

Fig. 6 Power of each genetic factor ($R^2 = 0.001$) for a multivariate GWAS of hippocampal subfield volumes. The power for each genetic factor is represented in blue. For comparison, we plotted the univariate power (controlling for multiple testing) in black (triangles), the multivariate MANOVA power assuming that all factors are as likely to exhibit a SNP effect (grey line), the MANOVA power weighting factors by the total phenotypic variance they explain (black line, circles). See Supplementary Fig. 5 for a brief comparison of the SNP effect from each additive genetic factor

the conclusions of this study. In addition, varying the order of the variables in the OpenMx modelling did not change the power estimation (Supplementary Figs. 8, 9).

We confirmed that our analytical derivations were correct by comparing the MANOVA power to those obtained through simulations (Supplementary Figs. 10, 11, 12). We further confirmed that reducing the MANOVA dimension to one phenotype yielded the same statistical power as of a univariate model (Supplementary Fig. 13).

As expected, relying on simulation to estimate statistical power was more time consuming. For our seven subcortical volumes it took 0.47 s to calculate one power curve analytically versus 21 min using simulations. For the twelve hippocampal subfields, the difference was even greater: 0.83 s analytically versus 1 h 2 min using simulations.

Discussion and conclusions

Here, we describe a method to calculate multivariate power of GWAS of real-case scenarios using the outputs from a multivariate twin modelling. Our approach is fast (less than a second), accurate (Supplementary Figs. 10, 11, 12) and computationally economical, as it does not rely on simulation but on analytical power calculation using the F-approximation (Muller and Peterson 1984). Thus, we could efficiently estimate the power of very large GWAS studies (N = 60,000+), which are realistic sample sizes in the field of imaging genetics nowadays. Our method provides a power estimate that corresponds to the multivariate test implemented in PLINK (Ferreira and Purcell 2009; Purcell et al. 2007) and which is equivalent to the one used in GEMMA (Zhou and Stephens 2014) (when working on related participants) (O'Reilly et al. 2012; van der Sluis et al. 2013; Galesloot et al. 2014).

Using this method, we showed that performing a multivariate GWAS of the volumes for both the subcortical structures and hippocampal subfields confers overall similar power to the standard univariate approach. However, the statistical power may vary greatly depending on the additive genetic source of variance considered (Fig. 6, Supplementary Figs. 2, 4). For example, in both scenarios, the multivariate power was the lowest for SNPs associated with the first additive genetic factor (that contributes to all phenotypes), consistent with previous bivariate simulations (Stephens 2013). For the hippocampal subfields, multivariate GWAS may boost power of detecting SNPs associated with specific subsets of phenotypes (Fig. 6, Supplementary Fig. 5). Indeed, eight additive genetic factors (accounting for 44% of the total additive genetic variance) showed greater power than the univariate approach while four returned similar or lower power. For each additive genetic factor, we

have described the SNP effects on the hippocampus subfields (Supplementary Fig. 6) but more work is needed to understand where the increase/decrease of power comes from. Thus, if overall MANOVA and univariate show similar statistical power, they could lead to the discovery of different genetic variants.

Our results are limited to the two scenarios considered and it would be of interest to extend our analyses, for example considering left and right volumes (instead of the average between left and right), cortical, or voxel-wise measurements. We have released the code used in this analysis to all researchers interested (https://baptistecd.github.io/PowerMultivariateGWAS/). It may be applied to variables and scenarios outside the field of brain imaging.

There are some limitations to this study, the main one being that the power calculation relies on estimated parameters (standardised path coefficients from twin modelling). However, we showed that the uncertainty around path coefficients resulted in limited variability in the power calculation, and did not change the conclusions of the analysis (Supplementary Figs. 6, 7). In absence of twin data, path coefficients may be derived from matrices of genetic correlations estimated from SNPs as well as phenotypic correlations. Another limitation concerns the stability of our results when using a different twin sample, possibly with different demographic characteristics. We did not have the data to investigate this limitation. Furthermore, our conclusions are limited to the MANOVA GWAS and there may be another competing multivariate approach that could yield greater power. We recommend to the interested readers the simulation framework of Porter and O'Reilly (Porter and O'Reilly 2017), that allows comparing the power across the different multivariate tests.

More generally, there are a few caveats when performing multivariate GWAS. Firstly, the effect of the SNP is not specific to one variable but rather the set of variables considered. Thus, the discovery analysis would identify SNPs associated with hippocampal volume in general, without testing which subfields the SNP has an effect on. This could be overcome by performing univariate GWAS follow-ups of the multivariate hits to more accurately identify the location of the SNP effect. Another potential consideration is the computing time of multivariate models, compared to univariate ones. Simulation reveals that multivariate GWAS using PLINK would be faster than running sequentially multiple univariate GWAS [see Supplementary Table 3 in (Porter and O'Reilly 2017)]. For multivariate GWAS using mixed models, GEMMA is currently a viable approach for analyses limited to 50,000 participants and ~ 10 phenotypes (Zhou and Stephens 2014). Both programs require little to no reformatting of the data, compared to univariate analyses.

To conclude, more work is needed to identify multivariate GWAS scenarios in imaging genetics that would yield greater power than standard univariate analyses. Our code allows to efficiently calculate such multivariate power, thus to quickly evaluate new scenarios without time and resource consuming simulations.

Acknowledgements We are very grateful to the twins for their generosity of time and willingness to participate in our studies. We thank research nurses Marlene Grace and Ann Eldridge for twin recruitment, Richard Parker for data collection, research assistants Lachlan Strike, Kori Johnson, Aaron Quiggle, and Natalie Garden, and radiographers Matthew Meredith, Peter Hobden, Kate Borg, Aiman Al Najjar, and Anita Burns for data acquisition, David Butler, Daniel Park, David Smyth and Anthony Conciatore for IT support.

Funding The MRI imaging of QTIM was supported by grants from National Institute of Health (NIH) (R01 HD050735, U54 EB020403) and the National Health and Medical Research Council (NHMRC) (496682, 1009064). PT and MJW are supported in part by NIH Grant to the ENIGMA Center for Worldwide Medicine, Imaging & Genomics. SEM is supported by an NHMRC fellowship (APP1103623). The funding was supported by Australian Research Council (Grant Nos. A7960034, A79906588, A79801419, DP0212016).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Informed consent Written informed consent was obtained from all participants including a parent or guardian for those aged less than 18 years.

Research involving human and animal rights The QTIM study was approved by the ethics review boards of the Queensland Institute of Medical Research, the University of Queensland, and Uniting Health Care, Wesley Hospital, Brisbane. QTIM participants received an honorarium in appreciation of their time. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Adams HH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Renteria ME, Trompet S, Arias-Vasquez A, Seshadri S, Desrivieres S, Beecham AH, Jahanshad N, Wittfeld K, Van der Lee SJ, Abramovic L, Alhusaini S, Amin N, Andersson M, Arfanakis K, Aribisala BS, Armstrong NJ, Athanasiu L, Axelsson T, Beiser A, Bernard M, Bis JC, Blanken LM, Blanton SH, Bohlken MM, Boks MP, Bralten J, Brickman AM, Carmichael O, Chakravarty MM, Chauhan G, Chen Q, Ching CR, Cuellar-Partida G, Braber AD, Doan NT, Ehrlich S, Filippi I, Ge T, Giddaluru S, Goldman AL, Gottesman RF, Greven CU, Grimm O, Griswold ME, Guadalupe T, Hass J, Haukvik UK, Hilal S, Hofer E, Hoehn D, Holmes AJ, Hoogman M, Janowitz D, Jia T, Kasperaviciute D, Kim S, Klein M, Kraemer B, Lee PH, Liao J, Liewald DC, Lopez LM, Luciano M, Macare C, Marquand A, Matarin M, Mather KA, Mattheisen M, Mazoyer B, McKay DR, McWhirter R, Milaneschi Y, Mirza-Schreiber N, Muetzel RL, Maniega SM, Nho K, Nugent AC, Loohuis LM, Oosterlaan J, Papmeyer M, Pappa I,

Pirpamer L, Pudas S, Putz B, Rajan KB, Ramasamy A, Richards JS, Risacher SL, Roiz-Santianez R, Rommelse N, Rose EJ, Royle NA, Rundek T, Samann PG, Satizabal CL, Schmaal L, Schork AJ, Shen L, Shin J, Shumskaya E, Smith AV, Sprooten E, Strike LT, Teumer A, Thomson R, Tordesillas-Gutierrez D, Toro R, Trabzuni D, Vaidya D, Van der Grond J, Van der Meer D, Van Donkelaar MM, Van Eijk KR, Van Erp TG, Van Rooij D, Walton E, Westlye LT, Whelan CD, Windham BG, Winkler AM, Woldehawariat G, Wolf C, Wolfers T, Xu B, Yanek LR, Yang J, Zijdenbos A, Zwiers MP, Agartz I, Aggarwal NT, Almasy L, Ames D, Amouyel P, Andreassen OA, Arepalli S, Assareh AA, Barral S, Bastin ME, Becker DM, Becker JT, Bennett DA, Blangero J, van Bokhoven H, Boomsma DI, Brodaty H, Brouwer RM, Brunner HG, Buckner RL, Buitelaar JK, Bulayeva KB, Cahn W, Calhoun VD, Cannon DM, Cavalleri GL, Chen C, Cheng CY, Cichon S, Cookson MR, Corvin A, Crespo-Facorro B, Curran JE, Czisch M, Dale AM, Davies GE, De Geus EJ, De Jager PL, de Zubicaray GI, Delanty N, Depondt C, DeStefano AL, Dillman A, Djurovic S, Donohoe G, Drevets WC, Duggirala R, Dyer TD, Erk S, Espeseth T, Evans DA, Fedko IO, Fernandez G, Ferrucci L, Fisher SE, Fleischman DA, Ford I, Foroud TM, Fox PT, Francks C, Fukunaga M, Gibbs JR, Glahn DC, Gollub RL, Goring HH, Grabe HJ, Green RC, Gruber O, Gudnason V, Guelfi S, Hansell NK, Hardy J, Hartman CA, Hashimoto R, Hegenscheid K, Heinz A, Le Hellard S, Hernandez DG, Heslenfeld DJ, Ho BC, Hoekstra PJ, Hoffmann W, Hofman A, Holsboer F, Homuth G, Hosten N, Hottenga JJ, Pol HE, Ikeda M, Ikram MK, Jack CR Jr, Jenkinson M, Johnson R, Jonsson EG, Jukema JW, Kahn RS, Kanai R, Kloszewska I, Knopman DS, Kochunov P, Kwok JB, Lawrie SM, Lemaitre H, Liu X, Longo DL, Longstreth WT Jr, Lopez OL, Lovestone S, Martinez O, Martinot JL, Mattay VS, McDonald C, McIntosh AM, McMahon KL, McMahon FJ, Mecocci P, Melle I, Meyer-Lindenberg A, Mohnke S, Montgomery GW, Morris DW, Mosley TH, Muhleisen TW, Muller-Myhsok B, Nalls MA, Nauck M, Nichols TE, Niessen WJ, Nothen MM, Nyberg L, Ohi K, Olvera RL, Ophoff RA, Pandolfo M, Paus T, Pausova Z, Penninx BW, Pike GB, Potkin SG, Psaty BM, Reppermund S, Rietschel M, Roffman JL, Romanczuk-Seiferth N, Rotter JI, Ryten M, Sacco RL, Sachdev PS, Saykin AJ, Schmidt R, Schofield PR, Sigurdsson S, Simmons A, Singleton A, Sisodiya SM, Smith C, Smoller JW, Soininen H, Srikanth V, Steen VM, Stott DJ, Sussmann JE, Thalamuthu A, Tiemeier H, Toga AW, Traynor BJ, Troncoso J, Turner JA, Tzourio C, Uitterlinden AG, Hernandez MC, Van der Brug M, Van der Lugt A, Van der Wee NJ, Van Duijn CM, Van Haren NE, Van TED, Van Tol MJ, Vardarajan BN, Veltman DJ, Vernooij MW, Volzke H, Walter H, Wardlaw JM, Wassink TH, Weale ME, Weinberger DR, Weiner MW, Wen W, Westman E, White T, Wong TY, Wright CB, Zielke HR, Zonderman AB, Deary IJ, DeCarli C, Schmidt H, Martin NG, De Craen AJ, Wright MJ, Launer LJ, Schumann G, Fornage M, Franke B, Debette S, Medland SE, Ikram MA, Thompson PM (2016) Novel genetic loci underlying human intracranial volume identified through genome-wide association. Nat Neurosci 19(12):1569–1582

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011) OpenMx: an open source extended structural equation modeling framework. Psychometrika 76(23258944):306–317

Breteler MMB, Stöcker T, Pracht E, Brenner D, Stirnberg R (2014) MRI in the Rhineland study: a novel protocol for population neuroimaging. Alzheimer's Dement J Alzheimer's Assoc 10(4):P92

Cole DA, Maxwell SE, Arvey R, Salas E (1994) How the power of manova can both increase and decrease as a function of the intercorrelations among the dependent-variables. Psychol Bull 115(3):465–474

de Zubicaray GI, Chiang MC, McMahon KL, Shattuck DW, Toga AW, Martin NG, Wright MJ, Thompson PM (2008) Meeting the challenges of neuroimaging genetics. Brain Imaging Behav 2(4):258–263

Ferreira MA, Purcell SM (2009) A multivariate test of association. Bioinformatics 25(1):132–133

Fischl B (2012) FreeSurfer. NeuroImage 62(2):774–781

Galesloot TE, van Steen K, Kiemeney LA, Janss LL, Vermeulen SH (2014) A comparison of multivariate genome-wide association methods. PLoS ONE 9(4):e95923

Greiser KH, Linseisen J, Schmidt B, Wichmann HE, Consortium GNCG (2014) The German National Cohort: aims, study design and organization. Eur J Epidemiol 29(5):371–382

Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N, Toro R, Wittfeld K, Abramovic L, Andersson M, Aribisala BS, Armstrong NJ, Bernard M (2015) Common genetic variants influence human subcortical brain structures. Nature 520:224

Iglesias JE, Augustinack JC, Nguyen K, Player CM, Player A, Wright M, Roy N, Frosch MP, McKee AC, Wald LL, Fischl B, Van Leemput K, Neuroimaging AD (2015) A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. NeuroImage 115:117–137

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95(3):221–227

Li MX, Yeung JM, Cherny SS, Sham PC (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet 131(5):747–756

Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39(7):906–913

Medland SE, Neale MC (2010) An integrated phenomic approach to multivariate allelic association. Eur J Hum Genet 18(2):233–239

Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, Bartsch AJ, Jbabdi S, Sotiropoulos SN, Andersson JL, Griffanti L, Douaud G, Okell TW, Weale P, Dragonu I, Garratt S, Hudson S, Collins R, Jenkinson M, Matthews PM, Smith SM (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 19(11):1523–1536

Minica CC, Boomsma DI, van der Sluis S, Dolan CV (2010) Genetic association in multivariate phenotypic data: power in five models. Twin Res Hum Genet 13(6):525–543

Muller KE, Peterson BL (1984) Practical methods for computing power in testing the multivariate general linear-hypothesis. Comput Stat Data Anal 2(2):143–158

Nagarsenker BN, Suniaga J (1983) Distributions of a class of statistics useful in multivariate-analysis. J Am Stat Assoc 78(382):472–475

O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin MR, Coin LJM (2012) MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS ONE 7(5):e34861

Olson CL (1974) Comparative robustness of 6 tests in multivariate-analysis of variance. J Am Stat Assoc 69(348):894–908

Olson CL (1976) Choosing a test statistic in multivariate-analysis of variance. Psychol Bull 83(4):579–586

Olson CL (1979) Practical considerations in choosing a Manova test statistic—a rejoinder to Stevens. Psychol Bull 86(6):1350–1352

Pillai KCS (1956) On the distribution of the largest or the smallest root of a matrix in multivariate analysis. Biometrika 43(1–2):122–127

Pillai KCS, Jayachandran K (1967) Power comparisons of tests of 2 multivariate hypotheses based on 4 criteria. Biometrika 54:195–210

Porter HF, O'Reilly PF (2017) Multivariate simulation framework reveals performance of multi-trait GWAS methods. Sci Rep 7:38837

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a

tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3):559–575

Rao CR (1973) Linear statistical inference and its applications. Wiley, New York

Renteria ME, Hansell NK, Strike LT, McMahon KL, de Zubicaray GI, Hickie IB, Thompson PM, Martin NG, Medland SE, Wright MJ (2014) Genetic architecture of subcortical brain regions: common and region-specific genetic contributions. Genes Brain Behav 13(8):821–830

Schram MT, Sep SJ, van der Kallen CJ, Dagnelie PC, Koster A, Schaper N, Henry RM, Stehouwer CD (2014) The Maastricht Study: an extensive phenotyping study on determinants of type 2 diabetes, its complications and its comorbidities. Eur J Epidemiol 29(6):439–451

Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, Toro R, Appel K, Bartecek R, Bergmann O, Bernard M, Brown AA, Cannon DM, Chakravarty MM, Christoforou A, Domin M, Grimm O, Hollinshead M, Holmes AJ, Homuth G, Hottenga JJ, Langan C, Lopez LM, Hansell NK, Hwang KS, Kim S, Laje G, Lee PH, Liu X, Loth E, Lourdusamy A, Mattingsdal M, Mohnke S, Maniega SM, Nho K, Nugent AC, O'Brien C, Papmeyer M, Putz B, Ramasamy A, Rasmussen J, Rijpkema M, Risacher SL, Roddey JC, Rose EJ, Ryten M, Shen L, Sprooten E, Strengman E, Teumer A, Trabzuni D, Turner J, van Eijk K, van Erp TG, van Tol MJ, Wittfeld K, Wolf C, Woudstra S, Aleman A, Alhusaini S, Almasy L, Binder EB, Brohawn DG, Cantor RM, Carless MA, Corvin A, Czisch M, Curran JE, Davies G, de Almeida MA, Delanty N, Depondt C, Duggirala R, Dyer TD, Erk S, Fagerness J, Fox PT, Freimer NB, Gill M, Goring HH, Hagler DJ, Hoehn D, Holsboer F, Hoogman M, Hosten N, Jahanshad N, Johnson MP, Kasperaviciute D, Kent JW Jr, Kochunov P, Lancaster JL, Lawrie SM, Liewald DC, Mandl R, Matarin M, Mattheisen M, Meisenzahl E, Melle I, Moses EK, Muhleisen TW, Nauck M, Nothen MM, Olvera RL, Pandolfo M, Pike GB, Puls R, Reinvang I, Renteria ME, Rietschel M, Roffman JL, Royle NA, Rujescu D, Savitz J, Schnack HG, Schnell K, Seiferth N, Smith C, Steen VM, Valdes Hernandez MC, Van den Heuvel M, van der Wee NJ, Van Haren NE, Veltman JA, Volzke H, Walker R, Westlye LT, Whelan CD, Agartz I, Boomsma DI, Cavalleri GL, Dale AM, Djurovic S, Drevets WC, Hagoort P, Hall J, Heinz A, Jack CR Jr, Foroud TM, Le Hellard S, Macciardi F, Montgomery GW, Poline JB, Porteous DJ, Sisodiya SM, Starr JM, Sussmann J, Toga AW, Veltman DJ, Walter H, Weiner MW, Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, Consortium IMAGEN, Saguenay Youth Study Group, Bis JC, Ikram MA, Smith AV, Gudnason V, Tzourio C, Vernooij MW, Launer LJ, DeCarli C, Seshadri S, Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium, Andreassen OA, Apostolova LG, Bastin ME, Blangero J, Brunner HG, Buckner RL, Cichon S, Coppola G, de Zubicaray GI, Deary IJ, Donohoe G, de Geus EJ, Espeseth T, Fernandez G, Glahn DC, Grabe HJ, Hardy J, Hulshoff Pol HE, Jenkinson M, Kahn RS, McDonald C, McIntosh AM, McMahon FJ, McMahon KL, Meyer-Lindenberg A, Morris DW, Muller-Myhsok B, Nichols TE, Ophoff RA, Paus T, Pausova Z, Penninx BW, Potkin SG, Samann PG, Saykin AJ, Schumann G, Smoller JW, Wardlaw JM, Weale ME, Martin NG, Franke B, Wright MJ, Thompson PM, Enhancing Neuro Imaging Genetics through Meta-Analysis Consortium (2012) Identification of common variants associated with human hippocampal and intracranial volumes. Nat Genet 44(5):552–561

Stephens M (2013) A unified framework for association analysis with multiple related phenotypes. PLoS ONE 8(7):e65245

Stevens J (1979) Choosing a test statistic in multivariate-analysis of variance—comment. Psychol Bull 86(2):355–360

Stevens JP (1980) Power of the multivariate-analysis of variance tests. Psychol Bull 88(3):728–737

Strike LT, Couvy-Duchesne B, Hansell NK, Cuellar-Partida G, Medland SE, Wright MJ (2015) Genetics and brain morphology. Neuropsychol Rev 25(1):63–96

Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Renteria ME, Toro R, Jahanshad N, Schumann G, Franke B, Wright MJ, Martin NG, Agartz I, Alda M, Alhusaini S, Almasy L, Almeida J, Alpert K, Andreasen NC, Andreassen OA, Apostolova LG, Appel K, Armstrong NJ, Aribisala B, Bastin ME, Bauer M, Bearden CE, Bergmann O, Binder EB, Blangero J, Bockholt HJ, Boen E, Bois C, Boomsma DI, Booth T, Bowman IJ, Bralten J, Brouwer RM, Brunner HG, Brohawn DG, Buckner RL, Buitelaar J, Bulayeva K, Bustillo JR, Calhoun VD, Cannon DM, Cantor RM, Carless MA, Caseras X, Cavalleri GL, Chakravarty MM, Chang KD, Ching CR, Christoforou A, Cichon S, Clark VP, Conrod P, Coppola G, Crespo-Facorro B, Curran JE, Czisch M, Deary IJ, de Geus EJ, den Braber A, Delvecchio G, Depondt C, de Haan L, de Zubicaray GI, Dima D, Dimitrova R, Djurovic S, Dong H, Donohoe G, Duggirala R, Dyer TD, Ehrlich S, Ekman CJ, Elvsashagen T, Emsell L, Erk S, Espeseth T, Fagerness J, Fears S, Fedko I, Fernandez G, Fisher SE, Foroud T, Fox PT, Francks C, Frangou S, Frey EM, Frodl T, Frouin V, Garavan H, Giddaluru S, Glahn DC, Godlewska B, Goldstein RZ, Gollub RL, Grabe HJ, Grimm O, Gruber O, Guadalupe T, Gur RE, Gur RC, Goring HH, Hagenaars S, Hajek T, Hall GB, Hall J, Hardy J, Hartman CA, Hass J, Hatton SN, Haukvik UK, Hegenscheid K, Heinz A, Hickie IB, Ho BC, Hoehn D, Hoekstra PJ, Hollinshead M, Holmes AJ, Homuth G, Hoogman M, Hong LE, Hosten N, Hottenga JJ, Hulshoff Pol HE, Hwang KS, Jack CR Jr, Jenkinson M, Johnston C, Jonsson EG, Kahn RS, Kasperaviciute D, Kelly S, Kim S, Kochunov P, Koenders L, Kramer B, Kwok JB, Lagopoulos J, Laje G, Landen M, Landman BA, Lauriello J, Lawrie SM, Lee PH, Le Hellard S, Lemaitre H, Leonardo CD, Li CS, Liberg B, Liewald DC, Liu X, Lopez LM, Loth E, Lourdusamy A, Luciano M, Macciardi F, Machielsen MW, Macqueen GM, Malt UF, Mandl R, Manoach DS, Martinot JL, Matarin M, Mather KA, Mattheisen M, Mattingsdal M, Meyer-Lindenberg A, McDonald C, McIntosh AM, McMahon FJ, McMahon KL, Meisenzahl E, Melle I, Milaneschi Y, Mohnke S, Montgomery GW, Morris DW, Moses EK, Mueller BA, Munoz Maniega S, Muhleisen TW, Muller-Myhsok B, Mwangi B, Nauck M, Nho K, Nichols TE, Nilsson LG, Nugent AC, Nyberg L, Olvera RL, Oosterlaan J, Ophoff RA, Pandolfo M, Papalampropoulou-Tsiridou M, Papmeyer M, Paus T, Pausova Z, Pearlson GD, Penninx BW, Peterson CP, Pfennig A, Phillips M, Pike GB, Poline JB, Potkin SG, Putz B, Ramasamy A, Rasmussen J, Rietschel M, Rijpkema M, Risacher SL, Roffman JL, Roiz-Santianez R, Romanczuk-Seiferth N, Rose EJ, Royle NA, Rujescu D, Ryten M, Sachdev PS, Salami A, Satterthwaite TD, Savitz J, Saykin AJ, Scanlon C, Schmaal L, Schnack HG, Schork AJ, Schulz SC, Schur R, Seidman L, Shen L, Shoemaker JM, Simmons A, Sisodiya SM, Smith C, Smoller JW, Soares JC, Sponheim SR, Sprooten E, Starr JM, Steen VM, Strakowski S, Strike L, Sussmann J, Samann PG, Teumer A, Toga AW, Tordesillas-Gutierrez D, Trabzuni D, Trost S, Turner J, Van den Heuvel M, van der Wee NJ, van Eijk K, van Erp TG, van Haren NE, van 't Ent D, van Tol MJ, Valdes Hernandez MC, Veltman DJ, Versace A, Volzke H, Walker R, Walter H, Wang L, Wardlaw JM, Weale ME, Weiner MW, Wen W, Westlye LT, Whalley HC, Whelan CD, White T, Winkler AM, Wittfeld K, Woldehawariat G, Wolf C, Zilles D, Zwiers MP, Thalamuthu A, Schofield PR, Freimer NB, Lawrence NS, Drevets W, Alzheimer's Disease Neuroimaging Initiative ECICSYSG (2014) The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav 8(2):153–182

van der Sluis S, Posthuma D, Dolan CV (2013) TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genetics 9(1):e1003235

Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 11(4):407–409