Fast Accurate Heritability Screening Using Whole-genome Data

Tian Ge[1], Erin Dickie[2], Amir Tahmasebi[2], Tobias Banaschewski[3], Gareth Barker[4], Arun Bokde[5], Christian Büchel[6], Patricia Conrod[4], Herta Flor[3], Andreas Heinz[7], Hugh Garavan[8], Penny Gowland[8], Bernd Ittermann[9], Claire Lawrence[10], Karl Mann[3], Jean-Luc Martinot[11], Frauke Nees[3], Mark Lathrop[12], Eva Loth[4], Zdenka Pausova[13], Marcella Rietschel[3], Michael Smolka[14], Andreas Ströhle[7], Jürgen Gallinat[7], Gunter Schumann[4], Tomas Paus[2], Thomas Nichols[15], and The IMAGEN Consortium

[1]Athinoula A. Martinos Center for Biomedical Imaging, USA; [2]Rotman Research Institute, Canada; [3]Central Institute of Mental Health, Germany; [4]King's College London, United Kingdom; [5]Trinity College Dublin, Ireland; [6]University Medical Center Hamburg-Eppendorf, Germany; [7]Charité - Universitätsmedizin Berlin, Germany; [8]University of Vermont, USA; [9]Physikalisch-Technische Bundesanstalt, Germany; [10]University of Nottingham, United Kingdom; [11]UMR INSERM-CEA U1000, France; [12]Centre National de Génotypage, France; [13]The Hospital for Sick Children, Canada; [14]Technische Universität Dresden, Germany; [15]University of Warwick, United Kingdom.

Introduction

Heritability estimation is an important tool for imaging genetics, as it can be used to prioritize brain phenotypes for analysis. Classical estimates of heritability require twin or pedigree data, which are difficult to acquire. Genome-wide Complex Trait Analysis (GCTA) [1] uses unrelated individuals to estimate the variance explained by all SNPs in the genome, and thus can provide heritability estimates without twin data. However, GCTA is based on linear mixed effect models and employs the restricted maximum likelihood (ReML) approach to estimate the heritability. The time-consuming iterative optimization procedure involved in ReML constraints the applications of GCTA to more than a handful of a priori selected phenotypes or regions of interest (ROIs). Here we present a fast and accurate statistical test for SNP-based heritability.

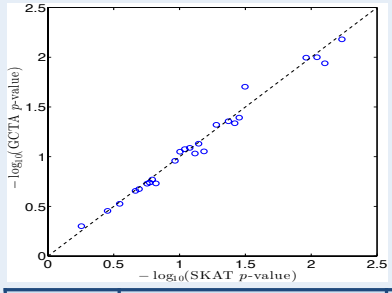
We show the excellent concordance between GCTA results and our approximation, and apply our approach to voxel-wise heritability analysis of per cent BOLD signal change (%BSC) while adolescents viewed dynamic videos of ambiguous faces.

Methods

We propose the sequence kernel association test (SKAT) based on the least square kernel machines [2] for heritability analysis. The semi-parametric linear regression model is easy to adjust for nuisance variables. Each entry of the kernel matrix, also known as the genetic relationship matrix (GRM) in GCTA, is a measure of genetic similarity between pairs of individuals computed from all SNPs over the genome to match the definition of narrow-sense heritability. Most importantly, the SKAT makes use of a non-iterative score test, giving a fast test for significant heritability. The test statistic is approximated chi-squared distributed, whose degrees of freedom capture the effective number of independent SNPs on the genome. The efficient computation of the test statistic allows for the use of standard permutation procedures. Family-wise error (FWE) corrected *p*-values are computed with the maximum statistic over all ROIs.

We applied our approach to fMRI data collected from 1,620 adolescents as part of the IMAGEN project [3]. During fMRI scanning, participants viewed videos clips of ambiguous facial movements, and non-biological motion (control). The GRM was computed from approximately half a million autosomal SNPs. We tested the heritability of the %BSC (ambiguous vs. control contrast) extracted from 25 predefined ROIs with high-probability response to faces [4], and compared their significance with the GCTA results obtained from a previous study [5]. We further performed voxel-wise heritability analysis to show the efficiency of our approach and provide 3D heritability profiles in these ROIs. Moreover, permutation allows us to consider arbitrary summaries of the results, like voxel-wise average heritability, and a variance-weighted voxel-wise average, to summarize heritability into a single number and provide an overall significance.

Figure 1: The excellent concordance between SKAT and GCTA



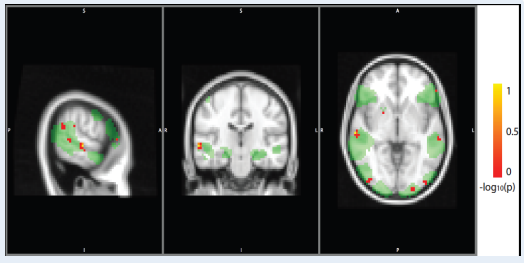


Figure 2: Voxel-wise heritability estimates for the %BSC (ambiguous vs. control). Transparent green areas indicate ROIs.

ROI	ROI averaged data		Uniform weighting		Weighted by variance	
	Uncorrected	FWE-corrected	Uncorrected	FWE-corrected	Uncorrected	FWE-corrected
MVLFC-L	0.0750	0.6471	0.0386	0.4467	0.0511	0.5319
MVLFC-R	0.0858	0.6754	0.0867	0.6926	0.1468	0.8480
MDLFC-L	0.1101	0.7552	0.0543	0.5499	0.0621	0.5921
MDLFC-R	0.0129	0.1789	0.0041	0.0776	0.0030	0.0607
PMC-L	0.0420	0.4735	0.0145	0.2242	0.0144	0.2292
PMC-R	0.0374	0.4452	0.0035	0.0675	0.0014	0.0305
PreSMA-R	0.0078	0.1381	0.0042	0.0793	0.0042	0.0818
RhinalSulcus-L	0.5600	0.9989	0.5578	0.9982	0.6330	0.9993
RhinalSulcus-R	0.0932	0.7053	0.0726	0.6377	0.1947	0.9116
Amygdala-L	0.1541	0.8416	0.1502	0.8494	0.2079	0.9239
Amygdala-R	0.0660	0.5971	0.0568	0.5644	0.0562	0.5595
AntSTS-L	0.0695	0.6252	0.0734	0.6413	0.0686	0.6231
AntSTS-R	0.0089	0.1556	0.0015	0.0323	0.0013	0.0284
PostSTS-L	0.1007	0.7311	0.0300	0.3793	0.0148	0.2339
PostSTS-R	0.0345	0.4216	0.0028	0.0562	0.0045	0.0864
FFA-L	0.1636	0.8564	0.0349	0.4192	0.0842	0.6878
FFA-R	0.2058	0.9096	0.1658	0.8746	0.1921	0.9090
LOC-L	0.2194	0.9208	0.2851	0.9653	0.1884	0.9045
LOC-R	0.1792	0.8791	0.0347	0.4179	0.0843	0.6883
V2V3-L	0.0077	0.1078	0.0031	0.0613	0.0020	0.0425
V2V3-R	0.1706	0.8684	0.0836	0.6818	0.0800	0.6719
Cerebellum-L	0.0354	0.3945	0.0197	0.2818	0.0258	0.3526
Cerebellum-R	0.3547	0.9789	0.3983	0.9892	0.4280	0.9934
Putamen-L	0.2837	0.9611	0.3099	0.9733	0.2988	0.9710
Putamen-R	0.0517	0.5335	0.0213	0.2987	0.0275	0.3674

Results

The *p*-values obtained by the SKAT show excellent concordance with the GCTA results (Fig. 1). Efficient voxel-wise analysis provides 3D profiles of heritability within the 25 ROIs, and localizes the most heritable regions (Fig. 2). "Compute-heritability-then-average" (Table, center and right) has better sensitivity than the "average-then-compute-heritability" approach (Table, left). The variance-weighted average has the best significance, likely due to its up-weighting of more variable voxels that are also more active.

Conclusion

We presented a fast and accurate statistical test for SNP-based heritability. The approach has excellent concordance with GCTA results, and can provide voxel- or vertex-wise heritability estimates for measures extracted from structural or functional brain images. We found that averages of voxel-wise heritability are more sensitive than heritability on averaged data. Our approach has the potential for large-scale heritability screening and for optimally choosing brain phenotypes under genetic control.

References

[1] Yang et al. (2011), *AJHG* 88: 76-82
[2] Liu et al. (2007), *Biometrics* 63: 1079-1088; Ge et al. (2012), *NI* 63: 858-873
[3] Schumann et al. (2010), *MP* 15: 1128-1139
[4] Tahmasebi et al. (2012), *HBM* 33: 938-957
[5] Dickie et al. (2013), OHBM abstract; (2014), *PLoS Genetics*, in press

Emails: tge1@mgh.harvard.edu, t.e.nichols@warwick.ac.uk