MGH/HST Athinoula A. Martinos Center for Biomedical Imaging

MASSACHUSETTS GENERAL HOSPITAL | HST Harvard-MIT Health Sciences & Technology

Massively Expedited Genome-wide Heritability Analysis (MEGHA)

CGR CENTER FOR HUMAN GENETIC RESEARCH

Tian Ge[1,2,3], Thomas E. Nichols[4], Phil H. Lee[2,3], Avram J. Holmes[5], Joshua L. Roffman[6], Randy L. Buckner[1,6,7], Mert R. Sabuncu[1,8,*], Jordan W. Smoller[2,3,*]

[1]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital / Harvard Medical School, USA; [2]Psychiatric & Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, USA; [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, USA; [4]Department of Statistics & Warwick Manufacturing Group, The University of Warwick, UK; [5]Department of Psychology, Yale University, USA; [6]Department of Psychiatry, Massachusetts General Hospital / Harvard Medical School, USA; [7]Department of Psychology and Center for Brain Science, Harvard University, USA; [8]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA; *MRS and JWS contributed equally

Introduction

With the dramatic expansion of available phenotypic data, practical tools for high-dimensional heritability-based screening have become invaluable for prioritizing phenotypes for genetic studies. Classical estimates of heritability require twin or pedigree data, which can be costly and difficult to acquire. Genome-wide complex trait analysis (GCTA) [1] is an alternative tool to compute heritability estimates from unrelated individuals, using genome-wide data that are increasingly ubiquitous, but is computationally demanding and becomes difficult to apply in evaluating very large numbers of phenotypes. Here we present a novel and practical statistical method for high-dimensional heritability analysis using genome-wide SNP data from unrelated individuals, termed massively expedited genome-wide heritability analysis (MEGHA) [2], and accompanying nonparametric sampling techniques that enable fast, accurate and flexible inferences for arbitrary statistics of interest.

Methods

MEGHA builds on the kernel machines framework [3], which subsumes GCTA as a special case, and utilizes a non-iterative variance component score test for efficient statistical inferences, making it possible to analyse millions of phenotypes and develop sampling techniques that produce accurate inferences and accommodate complex correlation structures within phenotypic data.

References and Acknowledgements

[1] Yang et al. (2011), *AJHG* 88(1): 76-82.
[2] Ge et al. (2015), *PNAS* 112(8): 2479-2484.
[3] Liu et al. (2007), *Biometrics* 63(4): 1079-1088.

This research was funded by P41EB015896, S10RR023043, S10RR023401, and also by NIH grants R01EB015611-01 and U54MH091657-03 (TEN), K99MH101367 (PHL), K01MH099232 (AJH), R01NS083534, R01NS070963 and 1K25EB013649-01 (MRS), K24MH094614 and R01MH101486 (JWS), Wellcome Trust grants 100309/Z/12/Z and 098369/Z/12/Z (TEN), and a BrightFocus grant AHAF-A2012333 (MRS). JWS is a Tepper Family MGH Research Scholar.

Figures and Tables

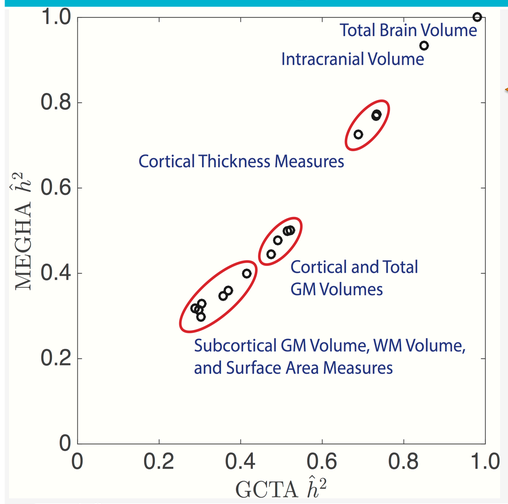
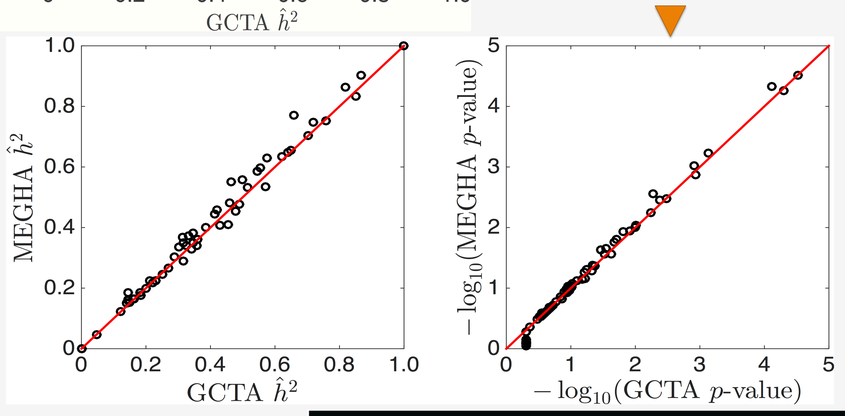
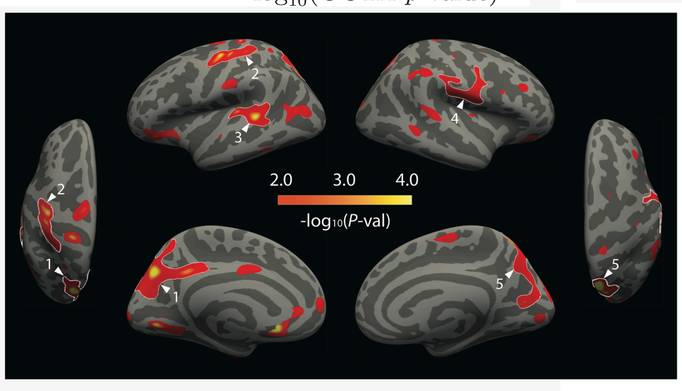


Figure 1: MEGHA SNP-based heritability estimates for 16 global morphometric measurements plotted against GCTA results.

Figure 2: Evaluation of MEGHA using average cortical thickness measures in 68 regions of interest (ROI) defined by the Desikan-Killiany atlas. MEGHA SNP-based heritability estimates (left) and *p*-values (right) are plotted against GCTA results for each ROI.





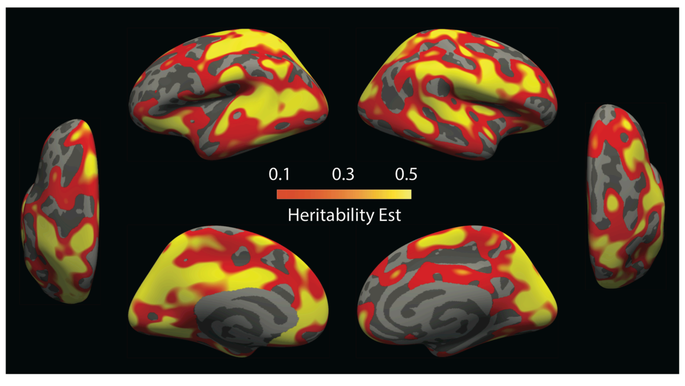


Figure 3: Vertex-wise surface maps for SNP-based heritability significance (left) and estimates (right) of cortical thickness measures constructed by MEGHA. Five family-wise error corrected significant clusters identified by permutation-based cluster inference are white outlined and annotated.

Case	Type of analysis	Effective no. of phenotypes	GCTA	MEGHA
1	Analysis of global morphometric measurements	16	120 s	0.65 s
2	ROI analysis	68	400 s	0.75 s
3	Vertex-wise heritability mapping	299,881	39.01 d	90 s
4	ROI analysis with permutation inference (one million permutations)	68,000,000	24.24 y*	3.5 h
5	Vertex-wise analysis with cluster inference (one thousand permutations)	299,881,000	106.88 y*	6.8 h

*Estimated computational time, using the average computational time, ~11.24 s, for each phenotype in case 3.

Table: Comparison of the computational time of GCTA and MEGHA.

Application

We applied MEGHA to the heritability analyses of global, regional and vertex-wise morphometric measurements derived from brain structural MRI scans, using genome-wide SNP data from 1,320 unrelated young (18-35 y) healthy adults of non-Hispanic European ancestry as part of the Harvard/MGH Brain Genomics Superstruct Project (GSP).

Results

Global morphometric measurements show moderate to high heritability (Fig. 1). MEGHA heritability estimates and *p*-values show excellent concordance with GCTA results in a region of interest analysis (Fig. 2). Vertex-wise MEGHA of cortical thickness measures produces high-resolution surface maps (over 300,000 vertices) for heritability significance and estimates (Fig. 3). Surface-based clustering on the significance map with permutation inferences empirically localized significantly heritable regions of cortical thickness, highlighting bilateral association regions of the parietal cortex extending into precuneus (Fig. 3). The computational efficiency of MEGHA is dramatically improved compared to GCTA, making ultra-high dimensional heritability screening and mapping tractable for the first time (Table).

Conclusion

We presented MEGHA, a novel, fast and accurate statistical method for heritability analysis using genome-wide SNP data from unrelated individuals, with a demonstration of application to imaging measurements.
With excellent computational efficiency and modeling flexibility, MEGHA has the potential for large-scale heritability profile construction, and for facilitating heritability analyses of vast phenotypic repositories in electronic health record systems and population-based biobanks.

Email: tge1@mgh.harvard.edu