Genetics of the Young Adult Human Connectome Project

Ganjgahi H[1], Bijsterbosch J[1], Daw EW[2], Donohue B[3], Fieremans E[4], Glahn D[5], Glasser M[2], Harms M[2], Hodge M[2], Jbabdi S[1], Kochunov P[3], Marchini J[1], Novikov D[4], Smith S[1], Sotiropoulos S[1,6], Van Essen D[2], Veraart J[4], Warrington S[6], Winkler A[7], and Nichols T [1]

[1]University of Oxford [2]Washington University [3]University of Maryland [4]New York University [5]Yale University [6]Nottingham University [7]National Institutes of Health

Introduction

The Young Adult Human Connectome Project (HCP) is a transformative project that includes cutting edge multimodal MRI imaging of structure, function, and connectivity. It enables investigation of the impact of environmental and genetic factors on human brain structure and function using state of the art neuroimaging, collected on a population of over 1000 subjects using an extended twin design. In addition to imaging data, detailed cognitive, health, lifestyle and mental health data were collected. All HCP data are publicly available, including genome-wide genotype data to be distributed through dbGAP. We present detailed genetic interrogation – heritability and genetic association – of HCP phenotypes.

Methods

Sample

The HCP is based on a population sample targeting families of twins born to Missouri residents from 1975 to 1991. The HCP recruited nuclear families with at least 4 offspring, with twin siblings aged 21-35 years. The extended twin design was selected to improve power to detect genetic associations, but the ethnically mixed sample presents a challenge for genetic association studies that can be confounded by samples with heterogeneous ancestry.

Genetics

1142 subjects were genotyped using two chips, a custom Illumina 2.5 chip with 2.8m markers and the Neuro Consortium chip. 1,580,642 SNPs passed QC and were used for a 2-stage imputation using the Haplotype Reference Consortium panel (HRC), followed by additional SNPs from the the merged panel from the 1000 Genomes and UK10K projects, ultimately producing genotypes for 26,957,241 SNPs. Of these, 9,828,572 had minimum allele frequency >1% and imputation accuracy >0.3; these were used for GWAS with 766 subjects of European ancestry (CEU, based on PCA). Heritability was computed using genetics-confirmed twin and sibling information using all available subjects.

Imaging

Heritability was computed on a wide variety of imaging and non-imaging measures, and GWAS was performed for measures for which the majority of CEU samples (>600) were available but not for voxel/vertex-wise measures. (See Fig 1).

Modelling

GWA and heritability analyses were performed using NINGA [1], a fast approximate method that accounts for kinship between individuals. In brief, the genetic relationship matrix is used to diagonalise the data and model, thus replacing a correlated-data problem with one that can be solved with weighted least squares; a score test allows fast estimation of heritability, which in turn allows valid estimation of association while accounting for heritability-induced dependence. This efficiency allows computation of familywise error (FWE) corrected thresholds via permutation accounting for dependence over SNPs and phenotypes. The phenotypes are transformed by inverse Gaussian transformation to ensure normality.

Modality	Phenotype	Parcellation	Sample Size	Dimension	Heritability	GWA
behavior	NIH Toolbox Behavioral Tests		907		Y	N
	Other behavioral tests (Non-NIH)		1187	278	Y	N
sMRI	Freesurfer: Thick/Area/Vol	Desikan-Killiany + Destrieux (L,R, L+R Hemispheres)	1133	976	Y	Y (718)
	Freesurfer: Area/Vol	Glasser Parcels	449	2x360	Y	N
	Freesurfer: Thickness & Myelin (unsmoothed)		1094	2 x 59412	Y	N
rfMRI	NetMats d=15, Partial			15x14/2	Y	Y (653)
	NetMats d=25, Partial			25x24/2	Y	N
	NetMats d=50, Partial			50x49/2	Y	N
	NetMats d=200, Partial			200x199/2	Y	N
	Amplitude d=15, Partial (AmpPar)and Full Correlation(Amp)		1003	15 x 2	Y	Y (653)
	Amplitude d=25, Partial and Full (AmpPar)and Full Correlation(Amp)			25 x 2	Y	N
	Amplitude d=50, Partial and Full (AmpPar)and Full Correlation(Amp)			50 x 2	Y	N
	Amplitude d=200, Partial and Full (AmpPar)and Full Correlation(Amp)			200 x 2	Y	N
	NetMats d=360, Partial	Glasser Parcels	449	360 x 359/2	Y	N
tfMRI	Emotion: Faces-Shapes	Desikan-Killiany Atlas	1044		Y	Y (686)
	Gambling: Punish-Reward		1082		Y	Y (705)
	Language: Math-Story		1049		Y	Y (688)
	Motor: Cue-Avg		1080		Y	Y (702)
	Relational: Match-Rel		1040		Y	Y (683)
	Social: TOM		1049		Y	Y (690)
	Working Memory: 2BK-0BK		1080	7 x 87	Y	Y (705)
	tfMRI: Vertex-wise, above contrasts			7 x 91282	Y	N
dMRI	TBSS: FA	JHU Atlass. (L,R, L+R Hemispheres)	1052	63	Y	Y (686)
	Tractography: FA & MD, volume, mean path probability	AutoPtx	954	33 x 7	Y	Y (625)

Fig 1. List of phenotypes in HCP that are used for heritability and GWA analyses. We have used age, age^2, sex, their interactions, BMI, Weight, Height, ICV, Acquisition, Modality-specific head movement and 10 genotype principal components as nuisance covariates. Numbers in GWA column corresponds to the total GWA sample size in each modality. GWA was perfomed using genetic relationship matrix calculated from imputed SNPs.

Results

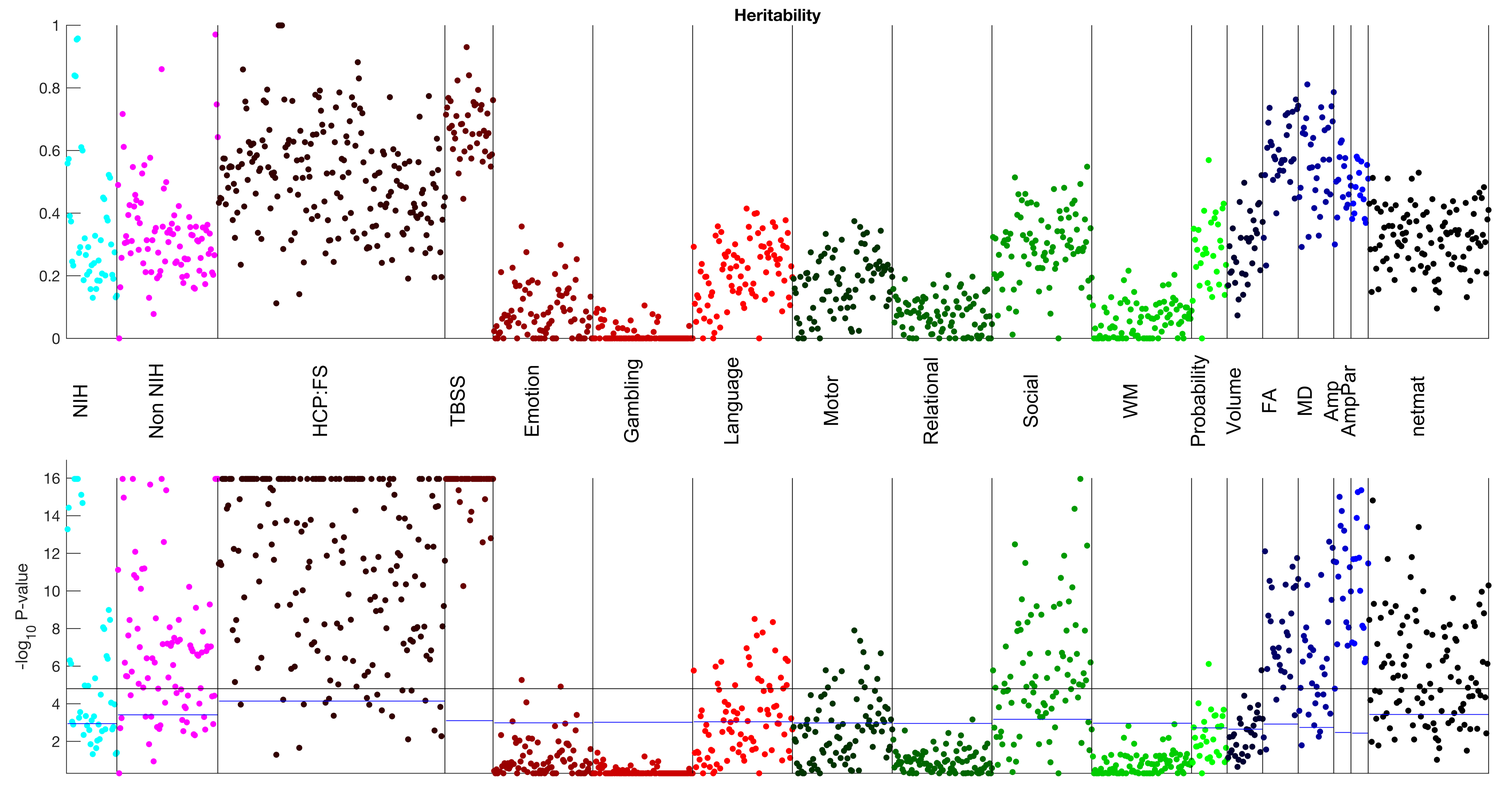
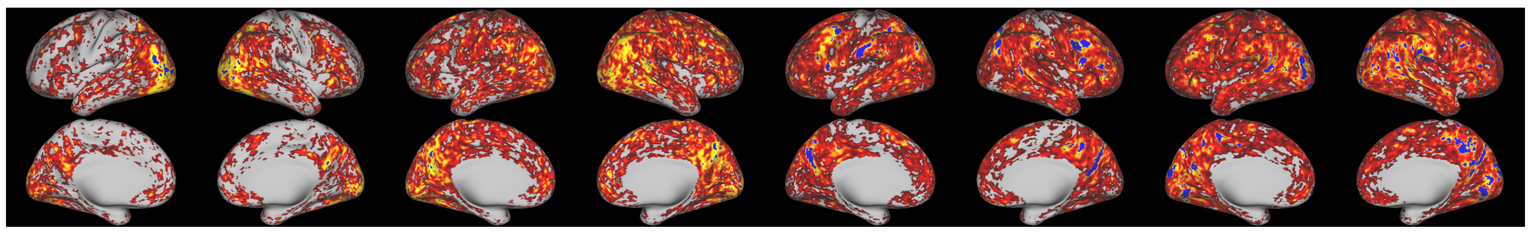


Fig2a. Heritability of imaging measures, each color indicating a different imaging modality. Top figure shows heritability estimates and bottom figure plots uncorrected parametric P-values; horizontal lines represents permutation based FWE thresholds using 5000 permutations (blue: per modality FWE; black: full FWE). Phenotypes correspond to Freesurfer, TBSS, Social task, Netmat amplitude have the most significant and heritable findings.



Emotion: Faces - Shapes | Motor: Cue - Average | Language: Math - Story | Social: TOM

Fig 2b. Vertex-wise maps with any FWE-significant surface elements, among all 7 tfMRI contrasts, cortical thickness and myelin maps. Red shows non-zero heritability, blue indicates 5% FWE-significant heritability.

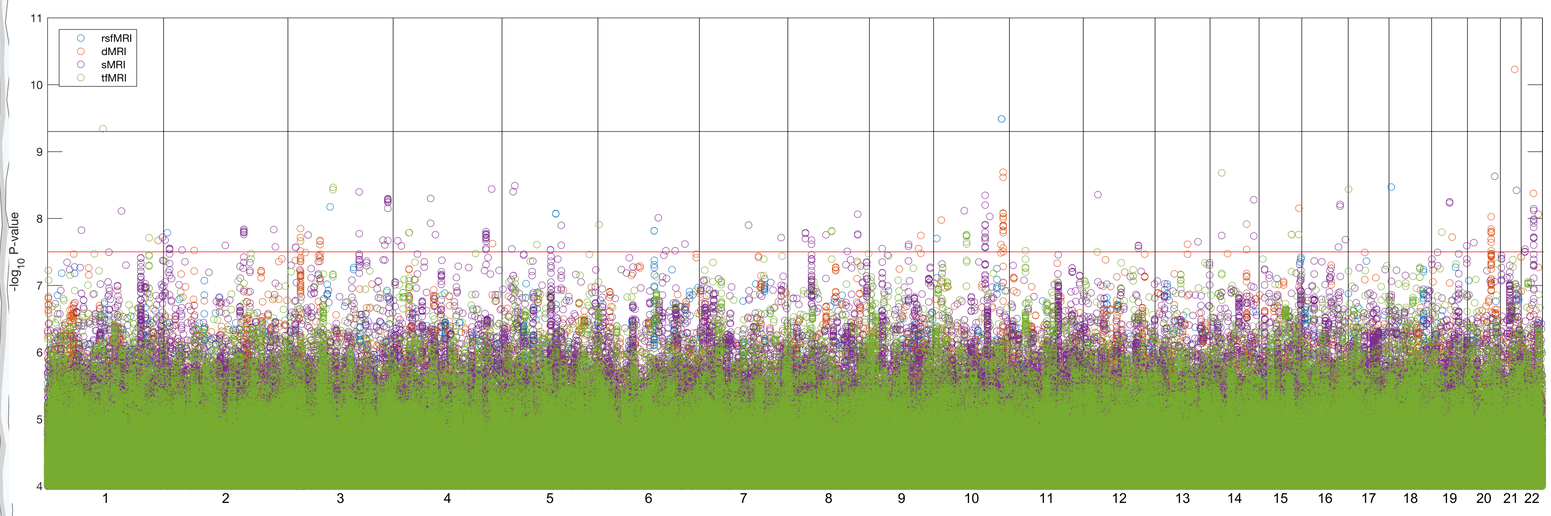


Fig2c. GWAS results for 2,014 phenotypes and all 9,828,572 SNPs where each dot represents maximum $-\log_{10}$ P-values per SNP between different imaging modalities. 294 SNPs passed the usual GWA P-value corrected for number of SNPs being tested. Top hits include associations for TBSS FA, Netmat edges and tfMRI activations. Among 19.8 billion tests, 3 passed the FWE threshold for all SNPs and ROIs. Upper (blue) line is 5% FWE threshold $-\log_{10}$ P threshold (9.4), and lower (red) threshold is usual GWA threshold of 7.5 $-\log_{10}$ P.

Modality	Phenotype	Associations	Max Parametric -log P-value
rfMRI	Netmat edges	10	9.49
dMRI	TBSS	10	10.30
	Tractography	60	8.69
sMRI	Thickness	62	8.63
	Area	93	8.40
	Volume	32	8.49
tfMRI	Working Mem	10	8.44
	Gambling	8	9.34
	Language	1	7.80
	Social	4	7.90
	Relational	4	8.68

Fig2d. Break down of 294 SNPs significant with conventional GWA threshold.

chr	rsid	maf	Imputation Accuracy	Parametric -log P-values	Region on Interest	Nearest Gene
21	rs757364414	0.13	0.69	10.23	TBSS: Right Sagittal Stratum FA	TTC3
10	rs10887069	0.23	0.78	9.49	Netmat edge 28	TACC2
1	rs116662163	0.01	0.88	9.34	Gambling: Left Inferior Temporal Gyrus	AKNAD1, AL449266.1

Fig2e. Details on three SNPs significant over all phenotypes and SNPs.

Conclusion

We have produced an initial view of the genetics of the Young Adult HCP. This includes the heritability estimation and inference for HCP phenotypes with high heritability for structural and diffusion brain measures. We demonstrated that NINGA can be used for heritability and GWA of high dimensional neuroimaging data while providing data-adaptive, permutation based FWE corrected thresholds. While the GWA results were limited in power we have found 294 associations significant at a conventional GWA threshold and 3 significant at a genome-phenome-wide FWE threshold.

References

1. H Ganjgahi et al, (2017). Fast and Powerful Genome Wide Association Analysis of Dense Genetic Data with High Dimensional Imaging Phenotypes. doi: https://doi.org/10.1101/179150