

# Family structure & Heritability

Thomas E. Nichols  
University of Warwick

# Why Families?

- NIH HCP RFA specified 300-400 subjects
  - State of the art imaging + genetic data
  - N=400 is minuscule genetic sample size
    - Single SNP would need to explain 10% variance to survive genome-wide significance

**Power Analysis: Variance explained by one genetic variant for 80% power\***

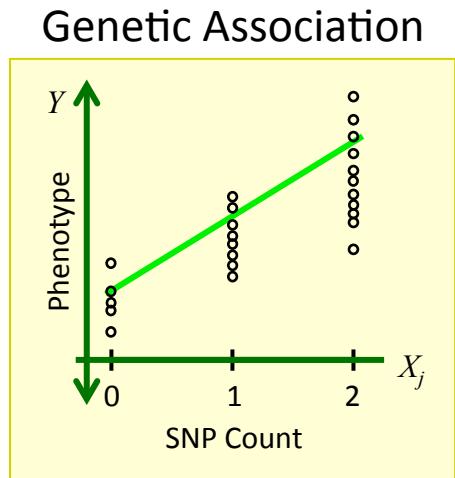
	Genome-Wide Testing 1 pheno., $\alpha = 5 \times 10^{-7}$	Candidate SNP 200 phenos., $\alpha = 0.05/200$
400 unrelated subjects	10%	6%
200 twin pairs	5%	3%
300 sibships, size 4	1%	1%

\* Countless details omitted

- N=400 twins (200 pairs) halved the effect size needed
- 300 families, target family size of 4 (N=1,200) used
  - Brings effect size needed (for 80% power) down to 1%

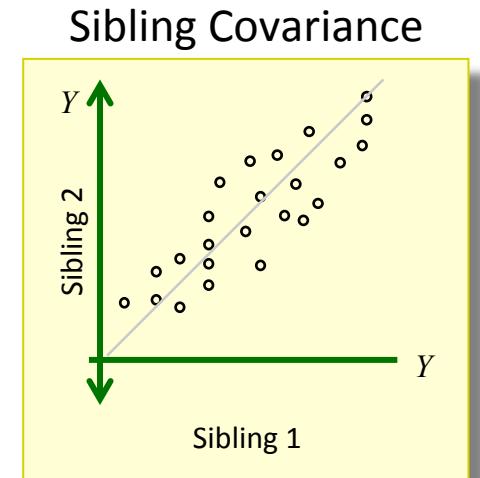
# How can this be?

- Families have a special advantage for genetic association
  - Evidence comes from *both*
    - The effect of a SNP over subjects
    - The correlation it induces between siblings



Usual source of genetic signal (in unrelated subjects)

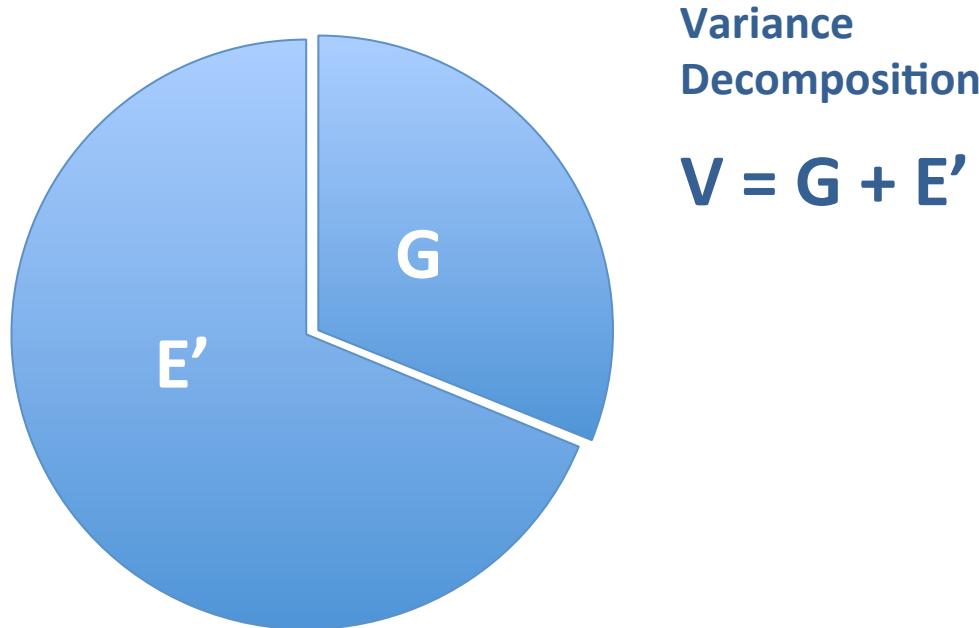
Boost! When a SNP is associated, it explains **some fraction** of the sibling correlation



# Heritability basics

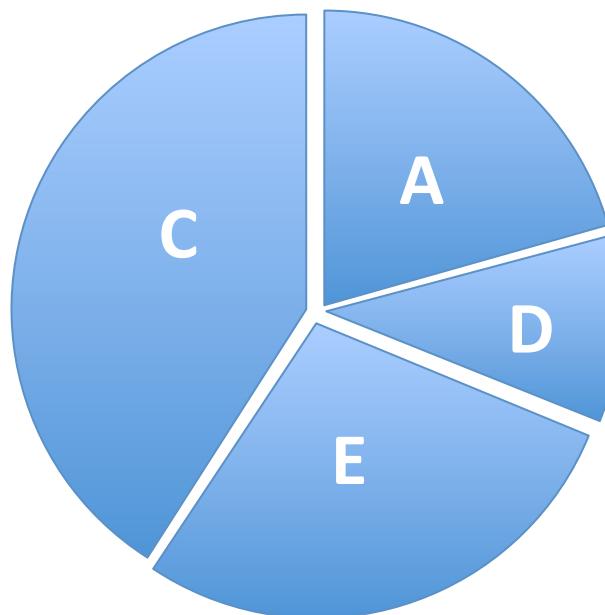
# Partitioning Variance

- All phenotypic variance can be decomposed
  - Due to genetic sources **G**
  - Due to (non-genetic) environmental sources **E'**
- But need family structure or genotypes to estimate



# Partitioning Variance

- With family data, further divisions possible
  - $G = A + D$ 
    - A additive genetic variance
    - D dominant genetic variance (hard to estimate)
  - $E' = C + E$ 
    - C environmental effects “C”ommon to twins/family members
    - E measurement “E”rrors, or other effects unique to each individual



Variance  
Decomposition  
 $V = A + D +$   
 $C + E$

# Heritability

- Generally work with narrow sense heritability

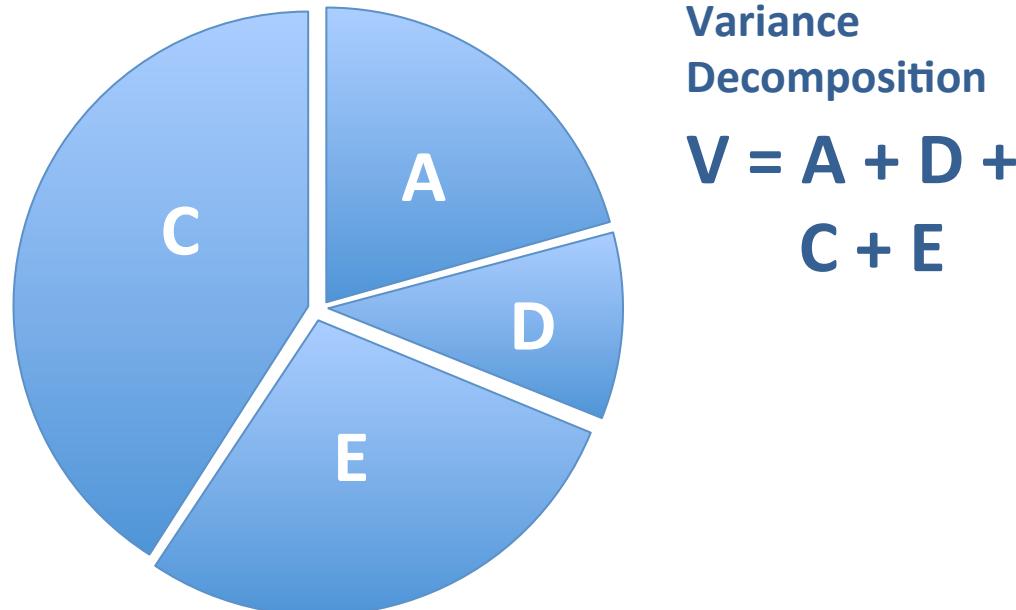
$$h^2 = A/V$$

- Proportion of all variance attributed to additive genetic variance

- Common environment

$$c^2 = C/V$$

- Proportion of all variance attributable shared environmental variance



# Estimating Heritability

- Standard approach is studying twins
  - Because twin-pair covariance reveals the variance decomposition!

$$\text{Cov}(Y_{\text{Twin1}}, Y_{\text{Twin2}}) =$$

MZ Twins	<i>Raised Together</i>	DZ Twins
$A + D + C + E$	$A + D + C$	$A + D + C + E$
$A + D + C$	$A + D + C + E$	$\frac{1}{2}A + \frac{1}{4}D + C$

MZ Twins	<i>Raised Apart</i>	DZ Twins
$A + D + C + E$	$A + D$	$A + D + C + E$
$A + D$	$A + D + C + E$	$\frac{1}{2}A + \frac{1}{4}D$

- Variance isn't helpful – always the same
- But **Covariance** has 4 unique expected values and 4 unknowns
  - But only you have twins raised apart!
- With only MZ & DZ twins, can only estimate 2 unknowns!
  - Must either assume D=0 or C=0; generally assume D=0

# ACE Model

$$\text{Cov}(Y_{\text{Twin}1}, Y_{\text{Twin}2}) =$$

$$\begin{array}{ccc} \text{MZ Twins} & \text{\textit{Raised Together}} & \text{DZ Twins} \\ \left[ \begin{array}{cc} A+C+E & A+C \\ A+C & A+C+E \end{array} \right] & & \left[ \begin{array}{cc} A+C+E & \frac{1}{2}A+C \\ \frac{1}{2}A+C & A+C+E \end{array} \right] \end{array}$$

- ACE Model
  - Comes from assuming D=0
  - Takes the form of variance components model
  - Also can be cast as a Structural Equation Model
  - In terms of correlations...

# ACE Model Correlations

$$\text{Corr}(Y_{\text{Twin1}}, Y_{\text{Twin2}}) =$$

$$\begin{array}{ccc} \text{MZ Twins} & \text{\textit{Raised Together}} & \text{DZ Twins} \\ \left[ \begin{array}{cc} 1 & h^2 + c^2 \\ h^2 + c^2 & 1 \end{array} \right] & & \left[ \begin{array}{cc} 1 & \frac{1}{2}h^2 + c^2 \\ \frac{1}{2}h^2 + c^2 & 1 \end{array} \right] \end{array}$$

- This immediately leads to Falconer's Estimate

$$h_F^2 = 2(r_{\text{MZ}} - r_{\text{DZ}})$$

- Where
  - $r_{\text{MZ}}$  is MZ intra-class correlation
  - $r_{\text{DZ}}$  is DZ intra-class correlation
- Falconer's is easy but not optimal
  - E.g. it can be negative!
  - Requires iterative maximum likelihood methods

# Illustrations of Twin Correlations

- Example of trait that is heritable,  $h^2=40\%$ 
  - 40% of variability attributable to genetic variation
  - No common environment,  $c^2=0\%$

Family of 4,  
1 MZ twin pair

	MZ Twin 1	MZ Twin 2	Sib 3	Sib 4
MZ Twin 1		0.4	0.2	0.2
MZ Twin 2	0.4		0.2	0.2
Sib 3	0.2	0.2		0.2
Sib 4	0.2	0.2	0.2	

Correlation Matrix

Family of 4,  
1 DZ twin pair

	DZ Twin 1	DZ Twin 2	Sib 3	Sib 4
DZ Twin 1		0.2	0.2	0.2
DZ Twin 2	0.2		0.2	0.2
Sib 3	0.2	0.2		0.2
Sib 4	0.2	0.2	0.2	

Correlation Matrix

3 Unrelated  
Individuals

	Unrel. 1	Unrel. 2	Unrel. 3
Unrel. 1		0	0
Unrel. 2	0		0
Unrel. 3	0	0	

Correlation Matrix

# Illustrations of Twin Correlations

- Example of trait that is not heritable,  $h^2=0$
- But common family env. induces corr,  $c^2=20\%$

Family of 4  
1 MZ twin pair

	MZ Twin 1	MZ Twin 2	Sib 3	Sib 4
MZ Twin 1		0.2	0.2	0.2
MZ Twin 2	0.2		0.2	0.2
Sib 3	0.2	0.2		0.2
Sib 4	0.2	0.2	0.2	

Correlation Matrix

Family of 4  
1 DZ twin pair

	DZ Twin 1	DZ Twin 2	Sib 3	Sib 4
DZ Twin 1		0.2	0.2	0.2
DZ Twin 2	0.2		0.2	0.2
Sib 3	0.2	0.2		0.2
Sib 4	0.2	0.2	0.2	

Correlation Matrix

3 Unrelated  
Individuals

	Unrel. 1	Unrel. 2	Unrel. 3
Unrel. 1		0	0
Unrel. 2	0		0
Unrel. 3	0	0	

Correlation Matrix

# Illustrations of Twin Correlations

- Example of trait that is heritable,  $h^2 = 20\%$
- But also common family env.,  $c^2 = 20\%$

Family of 4  
1 MZ twin pair

	MZ Twin 1	MZ Twin 2	Sib 3	Sib 4
MZ Twin 1		0.4	0.3	0.3
MZ Twin 2	0.4		0.3	0.3
Sib 3	0.3	0.3		0.3
Sib 4	0.3	0.3	0.3	

Correlation Matrix

Family of 4  
1 DZ twin pair

	DZ Twin 1	DZ Twin 2	Sib 3	Sib 4
DZ Twin 1		0.3	0.3	0.3
DZ Twin 2	0.3		0.3	0.3
Sib 3	0.3	0.3		0.3
Sib 4	0.3	0.3	0.3	

Correlation Matrix

3 Unrelated  
Individuals

	Unrel. 1	Unrel. 2	Unrel. 3
Unrel. 1		0	0
Unrel. 2	0		0
Unrel. 3	0	0	

Correlation Matrix

# Illustrations of Twin Correlations

- Example of trait that is not heritable,  $h^2=0$
- Nor exhibiting any common env. effect.

Family of 4  
1 MZ twin pair

MZ Twin 1	MZ Twin 2	Sib 3	Sib 4
MZ Twin 1	0	0	0
MZ Twin 2	0	0	0
Sib 3	0	0	0
Sib 4	0	0	0

Family of 4  
1 DZ twin pair

DZ Twin 1	DZ Twin 2	Sib 3	Sib 4
DZ Twin 1	0	0	0
DZ Twin 2	0	0	0
Sib 3	0	0	0
Sib 4	0	0	0

3 Unrelated  
Individuals

Unrel. 1	Unrel. 2	Unrel. 3
Unrel. 1	0	0
Unrel. 2	0	0
Unrel. 3	0	0

Correlation Matrix

Correlation Matrix

Correlation Matrix

# Estimating & Inferring on Heritability

- Inferring heritability ( $H_0: h^2=0$ ,  $H_a: h^2>0$ )
  - Requires estimation of family-induced correlation
    - Parameterized by  $h^2$  & (possibly)  $c^2$
  - Computation of  $h^2$  &  $c^2$  P-values
- Iterative maximization of likelihood
  - E.g. Model when  $C=0$ 
$$Y = X\beta + \epsilon \quad \text{Var}(\epsilon) = \Sigma = \sigma_A^2(2\Phi) + \sigma_E^2 I$$
    - Must jointly estimate  $\beta$  and  $A$  and  $E$  variances
  - Make inference on  $H_0: h^2=0$  with likelihood ratio test,  $\chi^2$  statistic

# Estimating Heritability

- “Standard” (non-imaging) tools
  - SOLAR
    - Powerful, comprehensive program, standard for family studies
    - But slow; complex, has serious learning curve
      - Overkill if all you want to do is ignore heritability
  - OpenMx
    - R-based implementation of Mx, standard for twin studies
    - Also slow, and if not box-standard twin data, hard to use
      - Requires posing heritability as an SEM problem
    - Doesn’t provide inference on regressors (e.g. IQ)
      - Tiny coding exercise to use  $h^2$  &  $c^2$  estimates in separate GLS
- Alternatives:
  - SOLAR-eclipse – Imaging-oriented SOLAR under development [http://www.nitrc.org/projects/se\\_linux/](http://www.nitrc.org/projects/se_linux/)
  - Linear Regression with Squared Differences (LR-SD)...

# Linear Regression on Squared Differences

- Heritability inference without iteration (Grimes & Harvey, 1980)

- Relate squared differences of data pairs to variance components A,C,E:

$$\mathbb{E} [(MZ_1 - MZ_2)^2] = 2E$$

$$\mathbb{E} [(DZ_1 - DZ_2)^2] = A + 2E$$

$$\mathbb{E} [(I_1 - I_2)^2] = 2A + 2C + 2E$$

... Estimate with linear regression!

- Modification of Grimes and Harvey's method:  $n(n - 1)/2$  obs.  
 $\rightarrow (n_{MZ} + n_{DZ})/2$  obs. (50,721 vs. 141)

- Permutation Inference

- Under  $H_0: h^2 = 0$ , MZ and DZ twin pairs are exchangeable
  - $\binom{(n_{MZ}+n_{DZ})/2}{n_{MZ}/2}$  possible permutations

# LR-SD Evaluations

## Simulation Setting

- 10,000 simulations
- Sample sizes: 10+10, 50+50
- 15 ACE parameter settings:

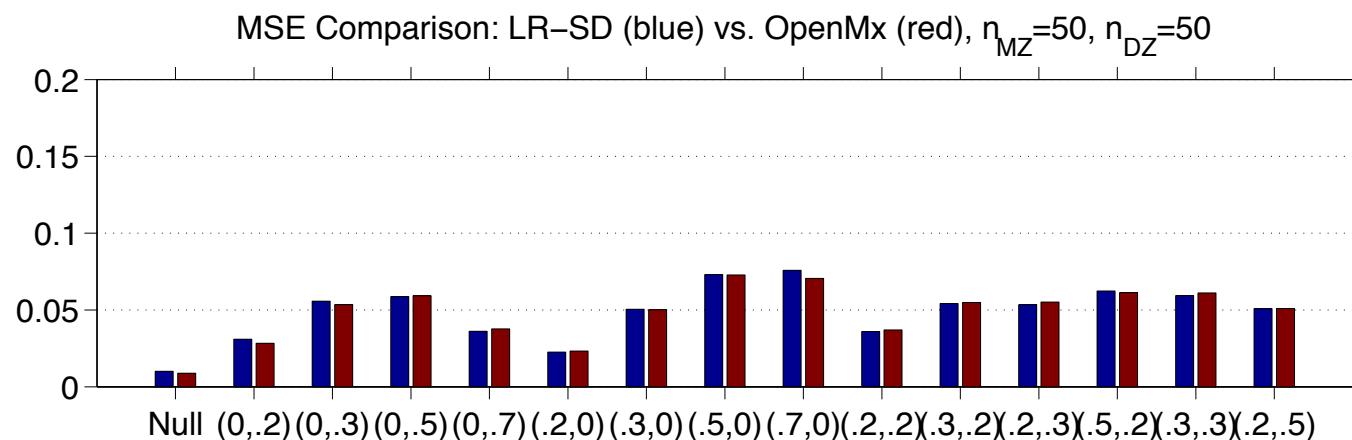
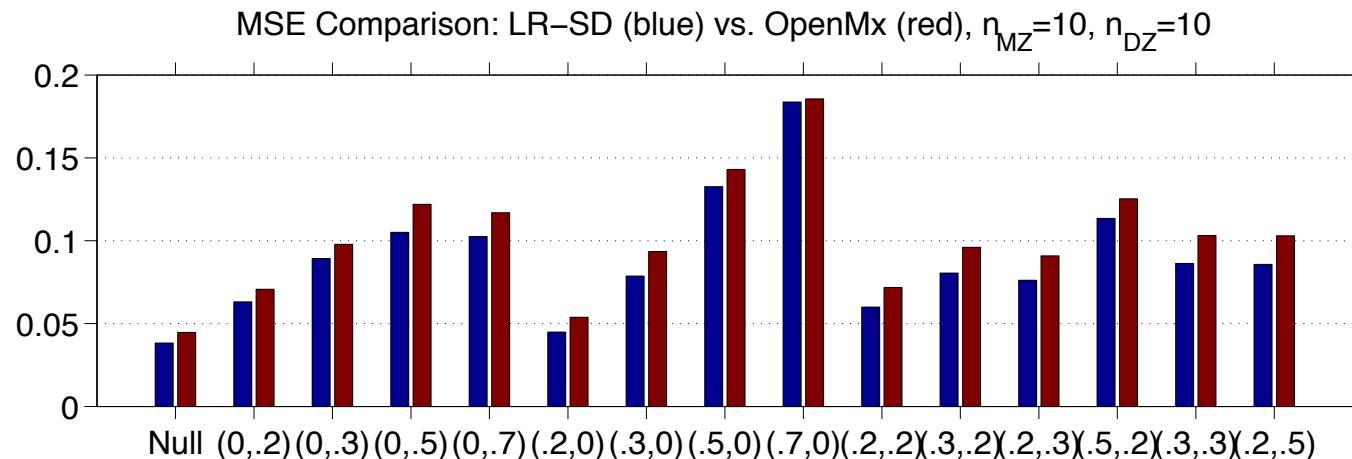
	E	CE				AE			
A	0	0	0	0	0	0.2	0.3	0.5	0.7
C	0	0.2	0.3	0.5	0.7	0	0	0	0
E	1	0.8	0.7	0.5	0.3	0.8	0.7	0.5	0.3

	ACE						
A	0.2	0.3	0.2	0.5	0.3	0.2	
C	0.2	0.2	0.3	0.2	0.3	0.5	
E	0.6	0.5	0.5	0.3	0.3	0.3	

# LR-SD Evaluations

## Simulations: MSE Comparison

Mean squared error comparison between LR-SD and OpenMx

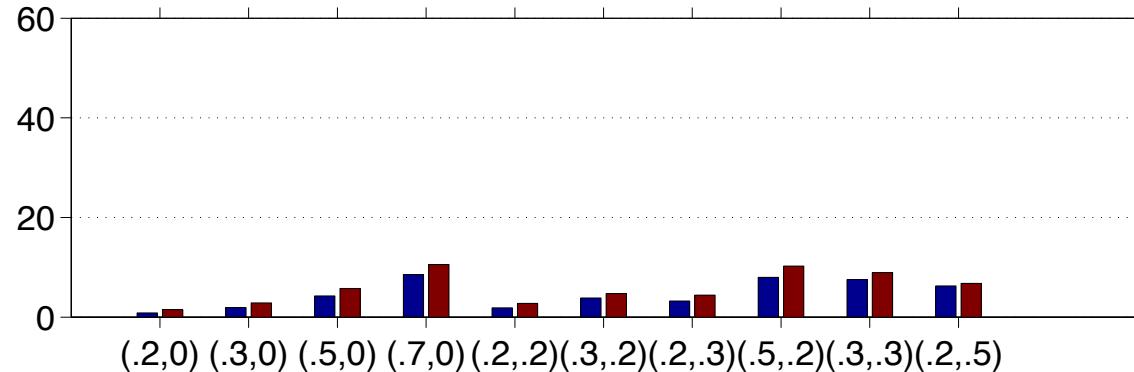


# SD-LR Evaluations

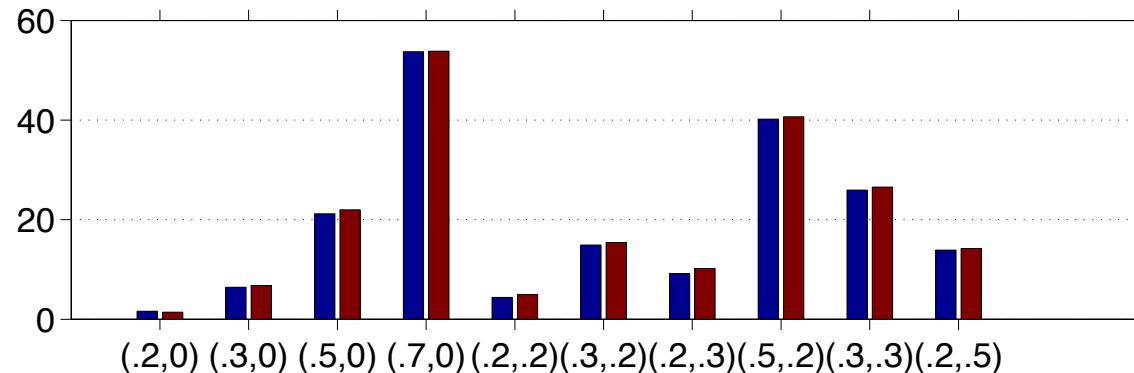
## Simulations: Power Comparison

Statistical power comparison between LR-SD and OpenMx

Power for LRT (100%): LR-SD (blue) vs. OpenMx (red),  $n_{MZ}=10$ ,  $n_{DZ}=10$



Power for LRT (100%): LR-SD (blue) vs. OpenMx (red),  $n_{MZ}=50$ ,  $n_{DZ}=50$

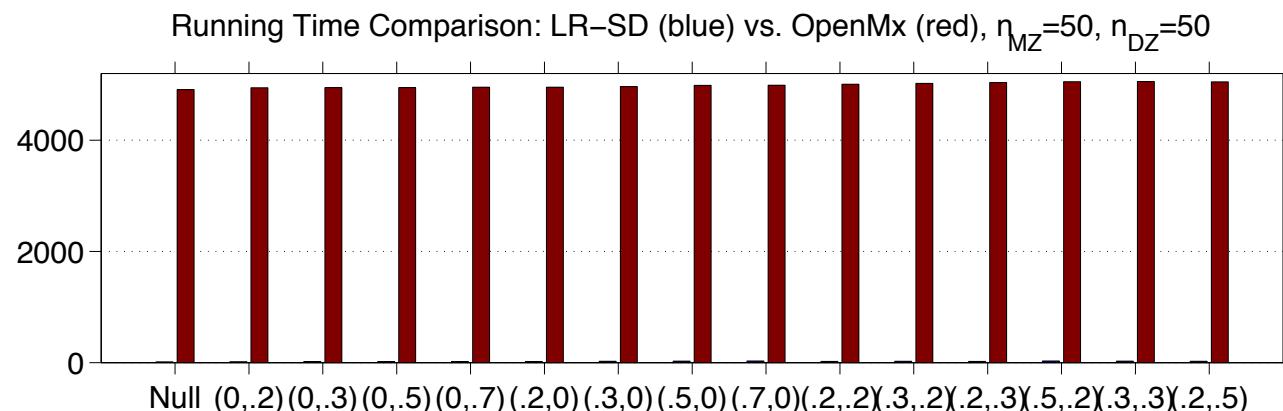
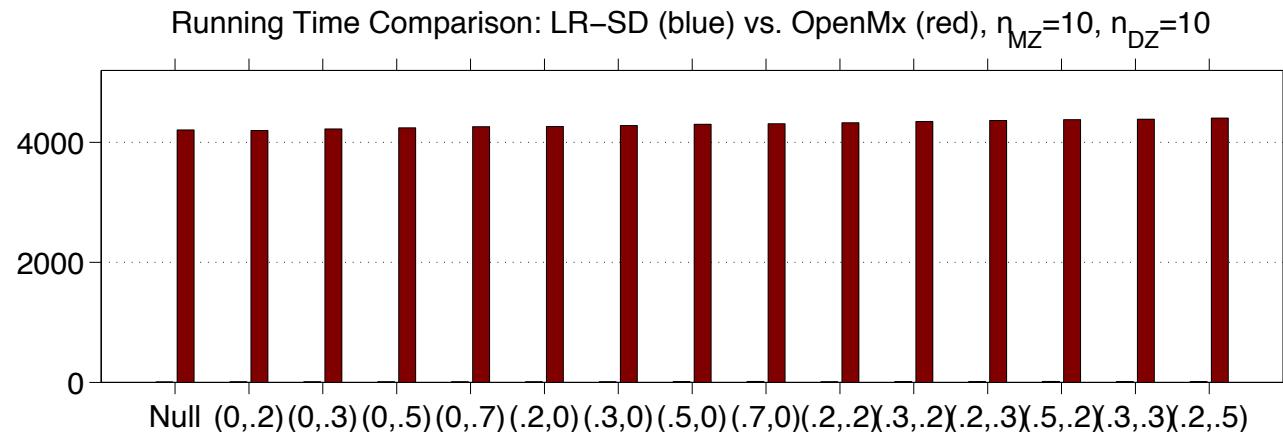


# SD-LR Evaluations

## Simulations: Running Time Comparison

Overall running time comparison between LR-SD and OpenMx

→ On average, our LR-SD is around 300 times faster than OpenMx



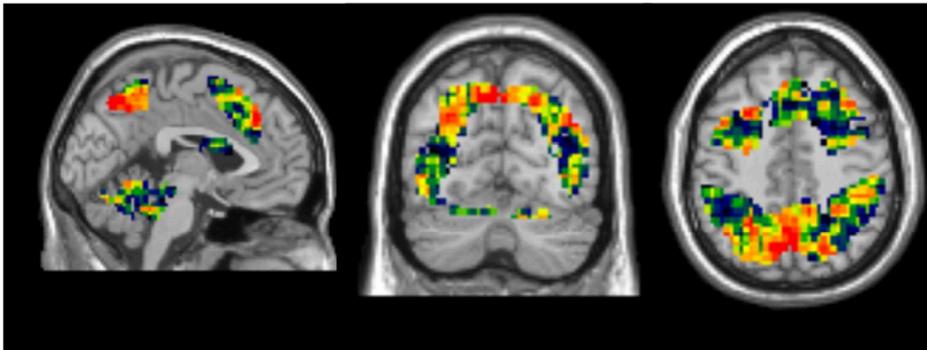
# fMRI Work Memory Example

- Blokland et al. 2011
- N-back, 2-back vs. 0-back contrast
- $n = 319$ :  $n_{MZ} = 150$ ,  $n_{DZ} = 132$ ,  $n_{UR} = 37$ 
  - 199F/120M, Age 20 – 28 (mean  $\pm$  SD:  $23.6 \pm 1.8$  yr)
- Age, gender & 2-back performance covariates
- Results
  - Running time:
    - One perm: LR-SD **6 min**, Mx **2 days**
    - 1000 perms, 10x parallelisation: LR-SD 15.5 hours

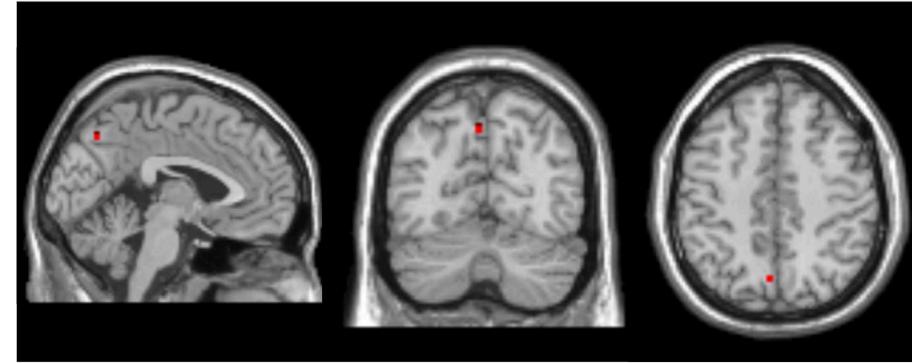
# fMRI Work Memory Example: Results

- Voxel-wise: Min  $P_{\text{FWE}}$  0.006, 3 voxels at 5% FWE
- Clusters-wise:  
Min  $P_{\text{FWE}}$  0.003  
3 clusters at 5% FWE (127, 201, & 210 voxels)

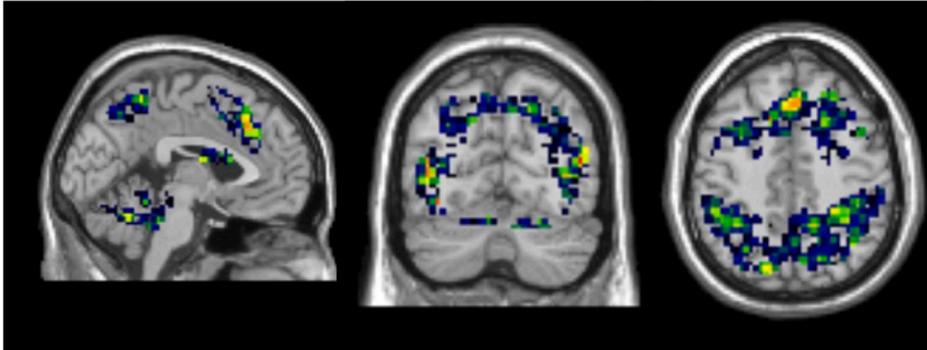
(1) MZ twin correlation



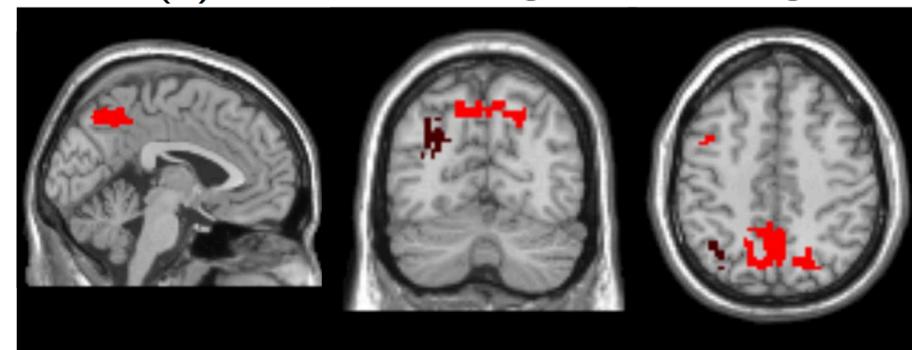
(1) Voxel-wise significance image



(2) DZ twin correlation



(2) Cluster-based significance image



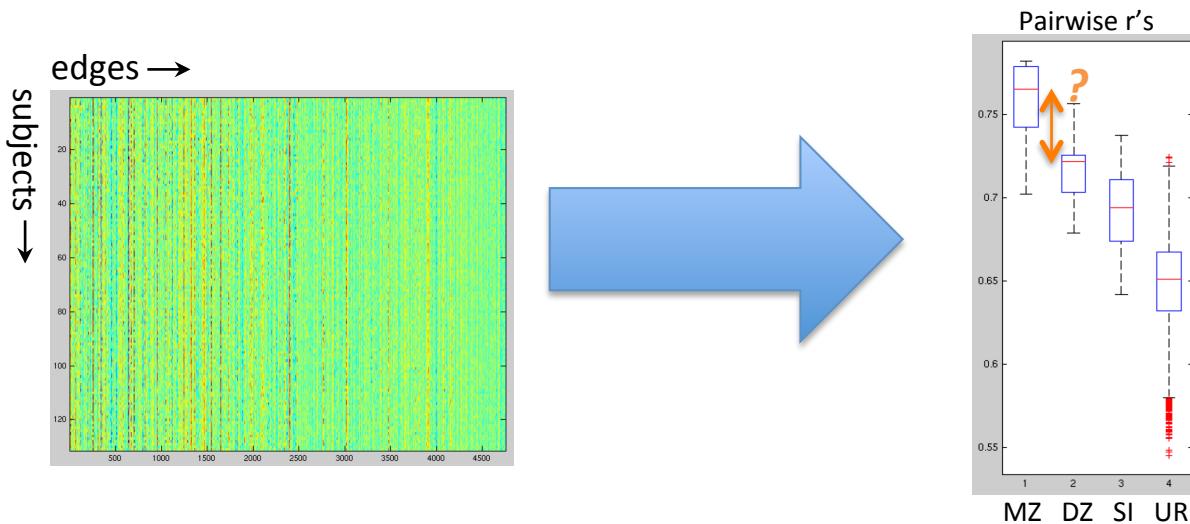
# Linear Regression on Squared Differences

- LR-SD method so fast allows permutation
- Allows standard imaging statistics (on LRT image)
  - FWE corrected voxel/peak, cluster size, cluster mass, etc.
- Limitations
  - No permutation test for  $c^2$ ; never any CI's
  - Use bootstrap for CI's
- Software
  - APACE: Accelerated Permutation Inference for ACE model
  - <http://warwick.ac.uk/tenichols/apace>

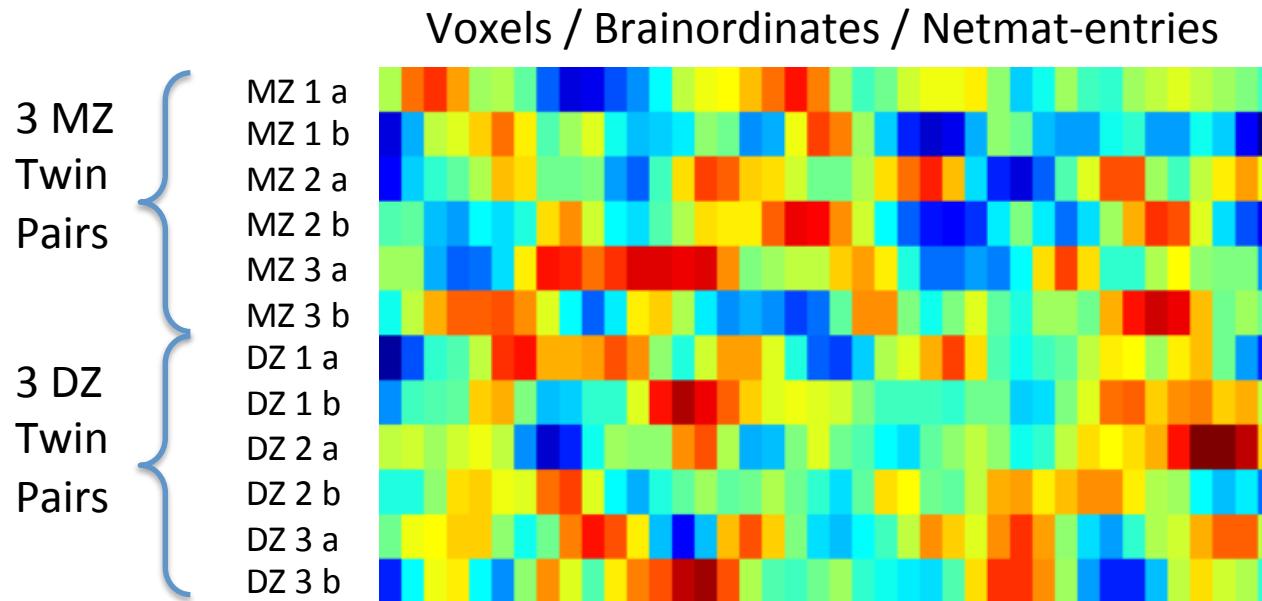
# Aggregate Heritability

- For high-dimensional phenotype, even a fast heritability method can be slow
  - e.g. dense connectomes
- Aggregate Heritability
  - Quick and dirty way to estimate average heritability over all elements
  - Based on subject-pair (**instead** of twin-pair) correlations

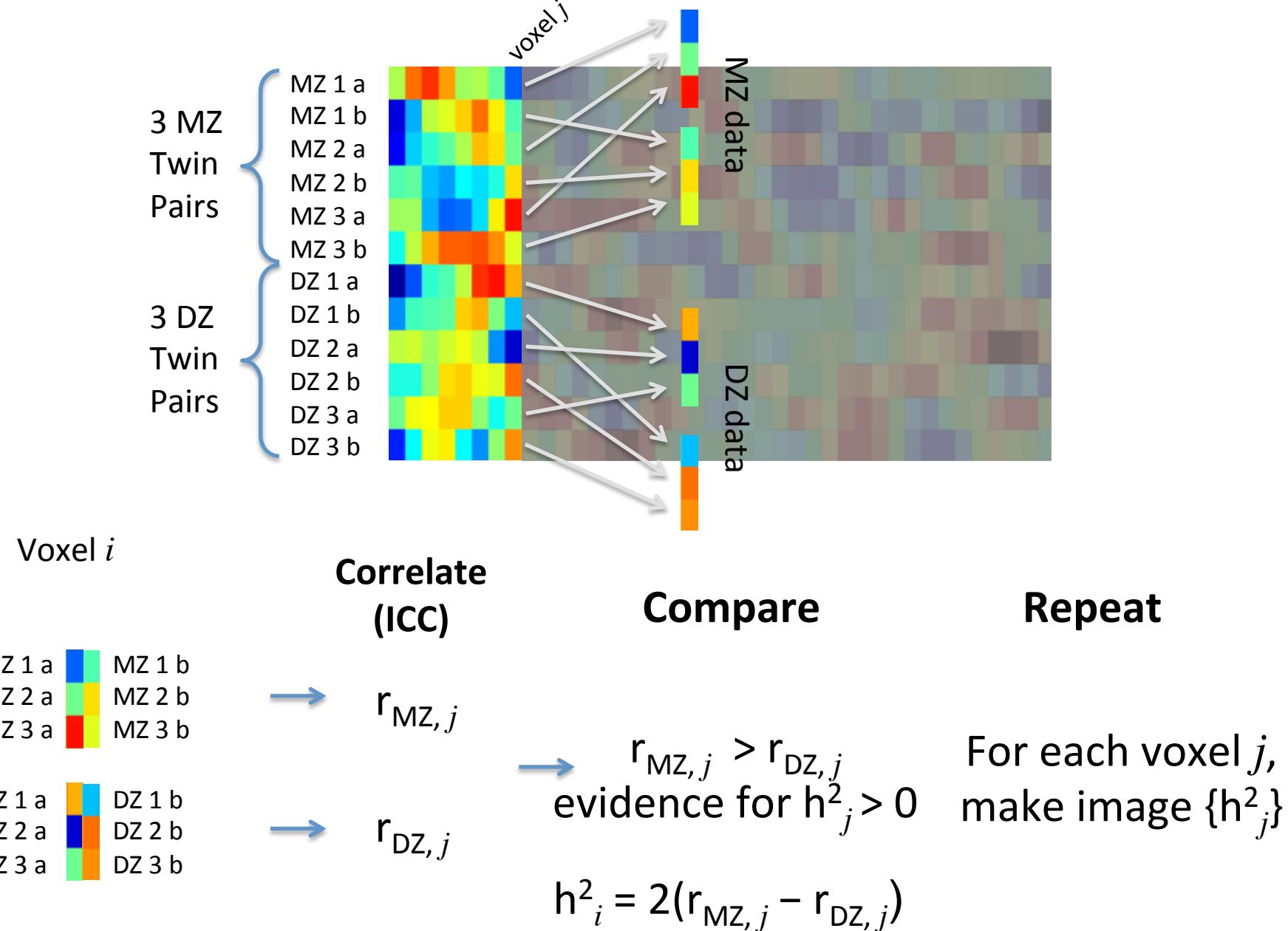
# Understanding Subject-Pair Correlations



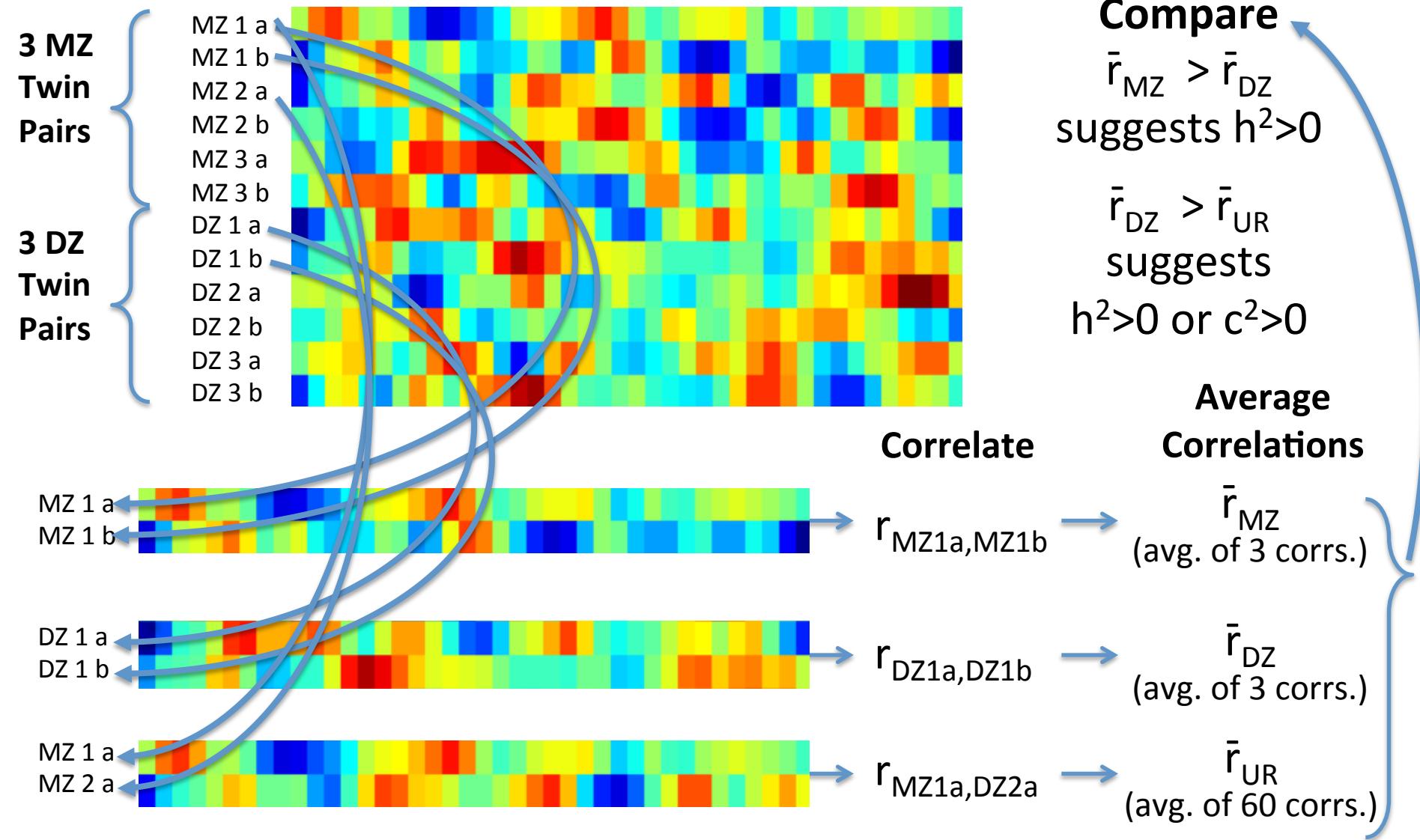
Consider toy version of netmat data:



# Traditional (Per-Phenotype) Heritability



# Subject-Pair Correlations



# What exactly do these mean!?

$$E(r_{MZ}) \approx \frac{\text{Var}(\mu)/\bar{\sigma}^2 + \tilde{h}^2 + \tilde{c}^2 - \widetilde{\text{ERV}}}{\text{Var}(\mu)/\bar{\sigma}^2 + 1 - \tilde{\rho}^P}$$

$\text{Var}(\mu)$  Variance of mean  
 $\mu_j$  of voxel  $j$

$$E(r_{DZ}) \approx \frac{\text{Var}(\mu)/\bar{\sigma}^2 + \frac{1}{2}\tilde{h}^2 + \tilde{c}^2 - \frac{1}{2}\widetilde{\text{ERV}}}{\text{Var}(\mu)/\bar{\sigma}^2 + 1 - \tilde{\rho}^P}$$

$\mu_j$  Mean of  
voxel  $j$

$$E(r_{UR}) \approx \frac{\text{Var}(\mu)/\bar{\sigma}^2}{\text{Var}(\mu)/\bar{\sigma}^2 + 1 - \tilde{\rho}^P}$$

$\bar{\sigma}^2 = \frac{1}{J} \sum_j \sigma_j^2$   
Average of  
voxel  
variance

- Huge influence of phenotype mean
  - Variance of *mean* constant effect
  - *Demean*, then  $\text{Var}(\mu) = 0$

Note...  
 voxel = brainordinate  
 = netmat element

# What do these mean, w/out the mean?

$$E(r_{MZ}) \approx \frac{\widetilde{h^2} + \widetilde{c^2} - \widetilde{\text{ERV}}}{1 - \widetilde{\rho^P}}$$

$$E(r_{DZ}) \approx \frac{\frac{1}{2}\widetilde{h^2} + \widetilde{c^2} - \frac{1}{2}\widetilde{\text{ERV}}}{1 - \widetilde{\rho^P}}$$

$$E(r_{UR}) = 0$$

- So a group comparison gives...

$$E(r_{MZ} - r_{DZ}) \approx \frac{1}{2} \frac{\widetilde{h^2} - \widetilde{\text{ERV}}}{1 - \widetilde{\rho^P}}$$

Variance-weighted average heritability

$$\widetilde{h^2} = \frac{1}{J} \sum_j \left( \frac{\sigma_j^2}{\overline{\sigma^2}} \right) h_j^2$$

Variance-weighted avg. common var.

$$\widetilde{c^2} = \frac{1}{J} \sum_j \left( \frac{\sigma_j^2}{\overline{\sigma^2}} \right) c_j^2$$

Variance-weighted ERV

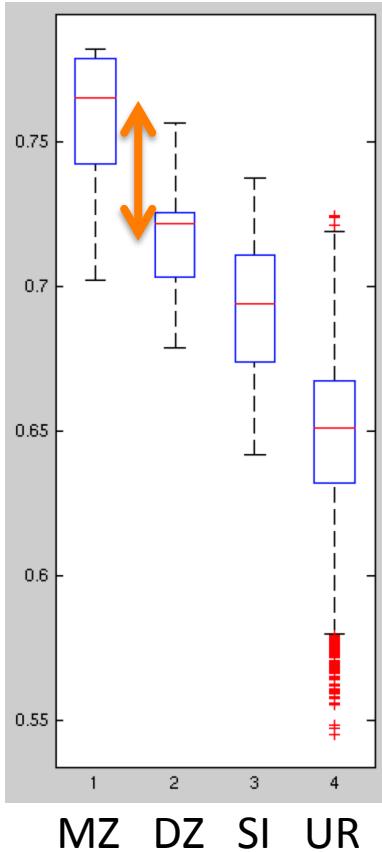
$$\widetilde{\text{ERV}} = \frac{2}{J(J-1)} \sum_{j>j'} \left( \frac{\sigma_j \sigma_{j'}}{\overline{\sigma^2}} \right) \text{ERV}_{jj'}$$

ERV: Heritability  $\times$  genetic correlation

$$\text{ERV}_{jj'} = h_j h_{j'} \rho_{jj'}^G$$

# What does a difference in means mean?

Pairwise r's



- What if this effect is significant?

$$E(r_{MZ} - r_{DZ}) \approx \frac{1}{2} \frac{\widetilde{h^2} - \widetilde{ERV}}{1 - \rho^P} > 0$$

- Indicates significant heritability (-ish)
  - Shifted by  $\widetilde{ERV}$  attenuated by  $(1 - \rho^P)^{-1}$

Variance-weighted average inter-voxel correlation

$$\rho^P = \frac{2}{J(J-1)} \sum_{j>j'} \left( \frac{\sigma_j \sigma_{j'}}{\sigma^2} \right) \rho_{jj'}^P$$

- But! It's a valid test for any heritability!

$$h_j^2 = 0 \quad \forall i \Rightarrow ERV_{jj'} = 0 \quad \forall j, j' \Rightarrow E(r_{MZ} - r_{DZ}) = 0$$

# Applications?

- “Aggregate Heritability” (AgHe)

$$\text{AgHe} = 2(\langle r_{\text{MZ}} \rangle - \langle r_{\text{DZ}} \rangle) \approx \frac{\widetilde{h^2} - \widetilde{\text{ERV}}}{1 - \widetilde{\rho^P}} \approx \widetilde{h^2}$$

demeaned result

- Biased estimate of variance-weighted heritability

- Regardless of AgHe, Suggests Approach to *High-dimensional Phenotype Heritability Ranking*

- Mean  $h^2$ ,  $\overline{h^2}$ 
    - $h^2$  computed at each element/voxel, then averaged
  - Var-Weighted Mean  $h^2$ ,  $\widetilde{h^2}$ 
    - For BOLD phenotypes, not so crazy!
    - Most active voxels most variable

$$\widetilde{h^2} = \frac{1}{J} \sum_j \left( \frac{\sigma_j^2}{\overline{\sigma^2}} \right) h_j^2$$

# HCP Phenotype Ranking

- 22 HCP  
Phenotypes...
  - nElm = 3k-60k

## Structural

Sulcal Depth, Cortical Thickness, Myelin, Cortical location  
Areal registration measures (Area ratio, displacement)

## Resting-State Functional

ICA-based 100- & 200-dimensional  
Dual-regression vs. Direct PCA  
Partial correlation vs Full correlation with Mean GM corr.  
Partial correlation: various regularization parameters

## Task-Based Functional

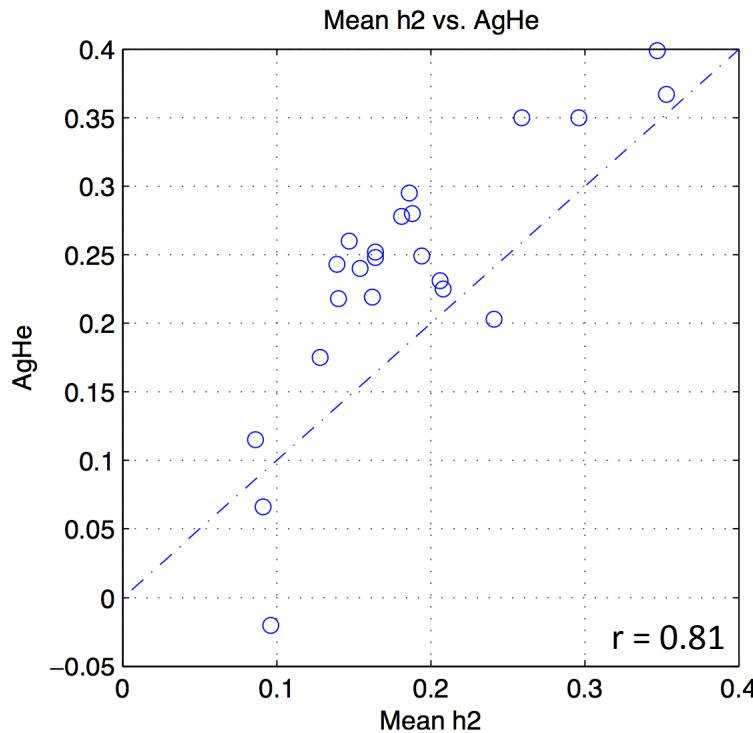
All task data (86 contrasts on 7 tasks), on 100 ROI set

- For each
  - Compute AgHe,  $\overline{h^2}$  &  $\widetilde{h^2}$ 
    - APACE used to find P-values & CI's
- Hypothesis:
  - Ranking will be similar between the 3 methods
  - AgHe most similar to  $\widetilde{h^2}$  (Var-Wt mean  $h^2$ )

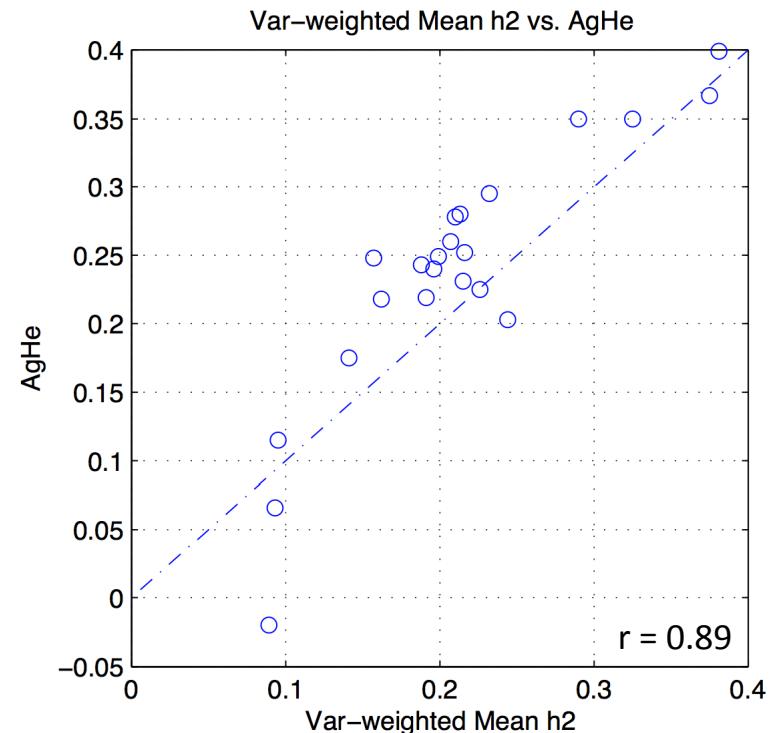
# HCP Phenotype Ranking: Estimates

- Good monotonic relationship
  - Tighter for  $\tilde{h}^2$  (variance-weighted mean  $h^2$ )

AgHe vs.  $\overline{h^2}$



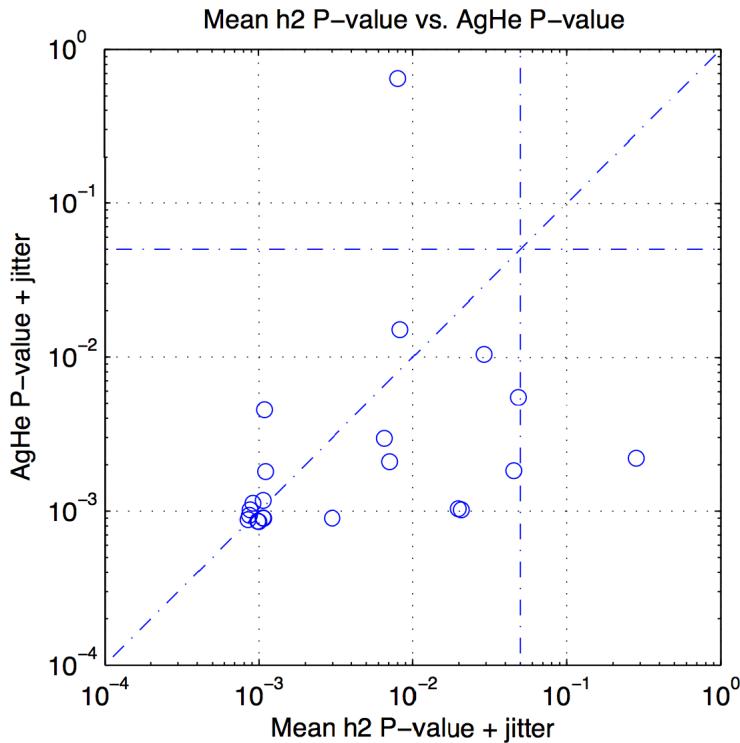
AgHe vs.  $\tilde{h}^2$



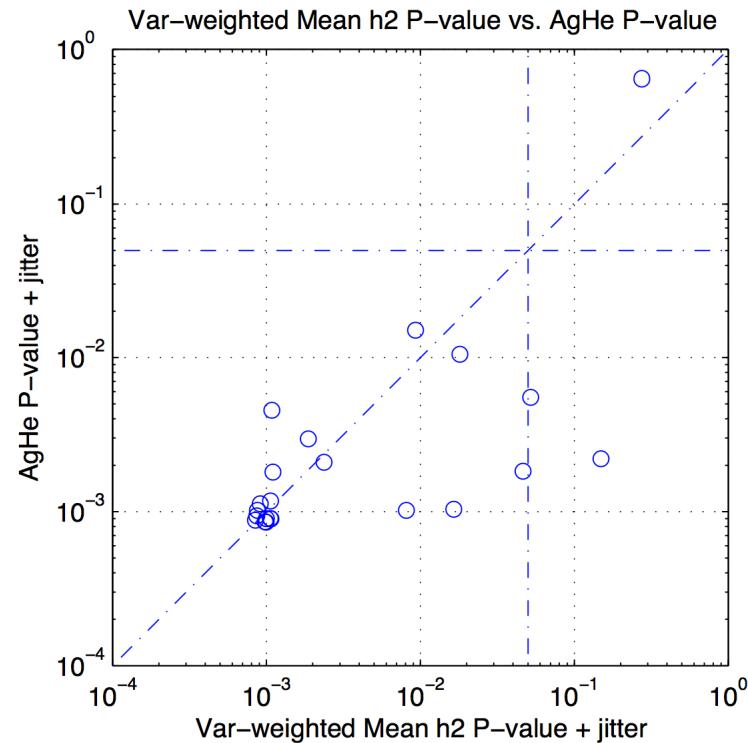
# HCP Phenotype Ranking: P-values

- Good agreement for strong significance
  - AgHe more optimistic... possibly due to  $(1 - \widetilde{\rho}^P)^{-1}$

P-values: AgHe vs.  $\overline{h^2}$



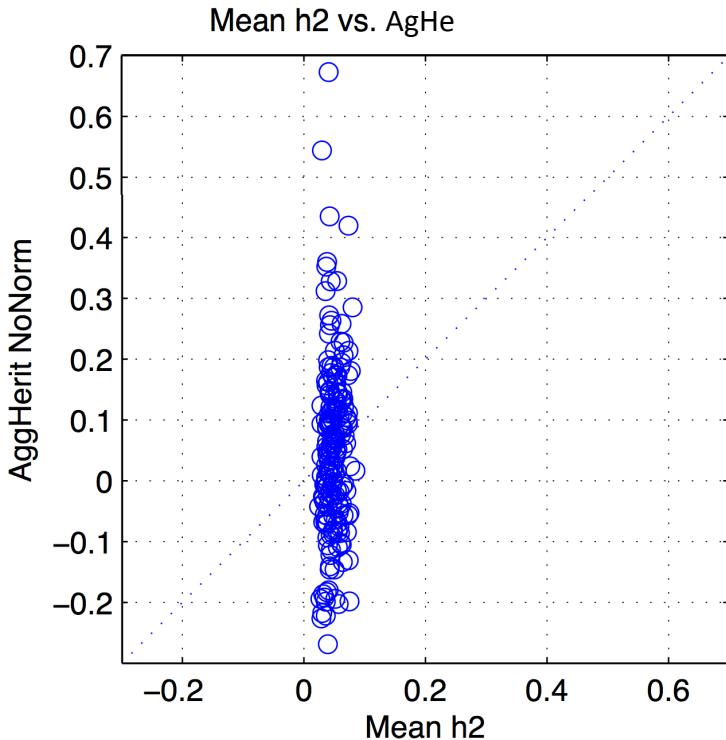
Pvalues: AgHe vs.  $\widetilde{h^2}$



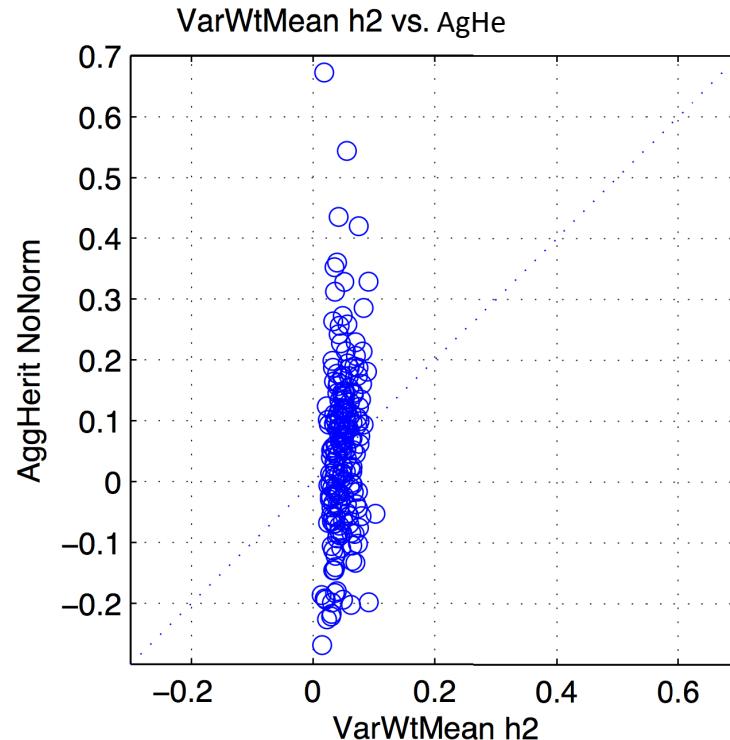
# Non-High-Dimensional Phenotypes: Not so good

- Ranking of each row of a 200-dimensional ICA netmat
  - 200 phenotypes, each with 199 elements:  
Connection strength to each other node
- AgHe not so biased, but huge variance

AgHe vs.  $\overline{h^2}$



AgHe vs.  $\widetilde{h^2}$



# Aggregate Heritability Conclusions

- Cheap and cheerful approach to heritability
- Gives a (biased) estimate of variance-weighted mean heritability
- But *valid* test of  $H_0$ : All  $h^2=0$
- Easy to implement
- Potentially useful for phenotype ranking

$$E(r_{MZ} - r_{DZ}) \approx \frac{1}{2} \frac{\widetilde{h^2} - \widetilde{ERV}}{1 - \widetilde{\rho^P}}$$

$$\widetilde{h^2} = \frac{1}{J} \sum_j \left( \frac{\sigma_j^2}{\overline{\sigma^2}} \right) h_j^2$$

# Heritability Application

- HCP Designed as N=1,200 study
  - Good estimates of heritability & genetic effects really need that N!
  - But still interesting look so far
- We considered heritability of all the phenotypes in the tabular spreadsheets in HCP500

# Examining Heritability in Behavioral Phenotypes

- 500 Subject release
- 392 phenotypes considered

Anatomy	Behavior	Behavior-IQ	Biomarker
179	182	17	14

## Behavior-IQ

CardSort_AgeAdj	CardSort_Unadj	Flanker_AgeAdj	Flanker_Unadj
ListSort_AgeAdj	ListSort_Unadj	PicSeq_AgeAdj	PicSeq_Unadj
PicVocab_AgeAdj	PicVocab_Unadj	PMAT24_A_CR	PMAT24_A_RTCR
PMAT24_A_SI	ProcSpeed_AgeAdj	ProcSpeed_Unadj	ReadEng_AgeAdj
ReadEng_Unadj			

## Biomarker

BMI	BPDiastolic	BPSystolic	Endurance_AgeAdj
Endurance_Unadj	GaitSpeed_Comp	HbA1C	Hematocrit_1
Hematocrit_2	Noise_Comp	Strength_AgeAdj	Strength_Unadj
ThyroidHormone	Weight		

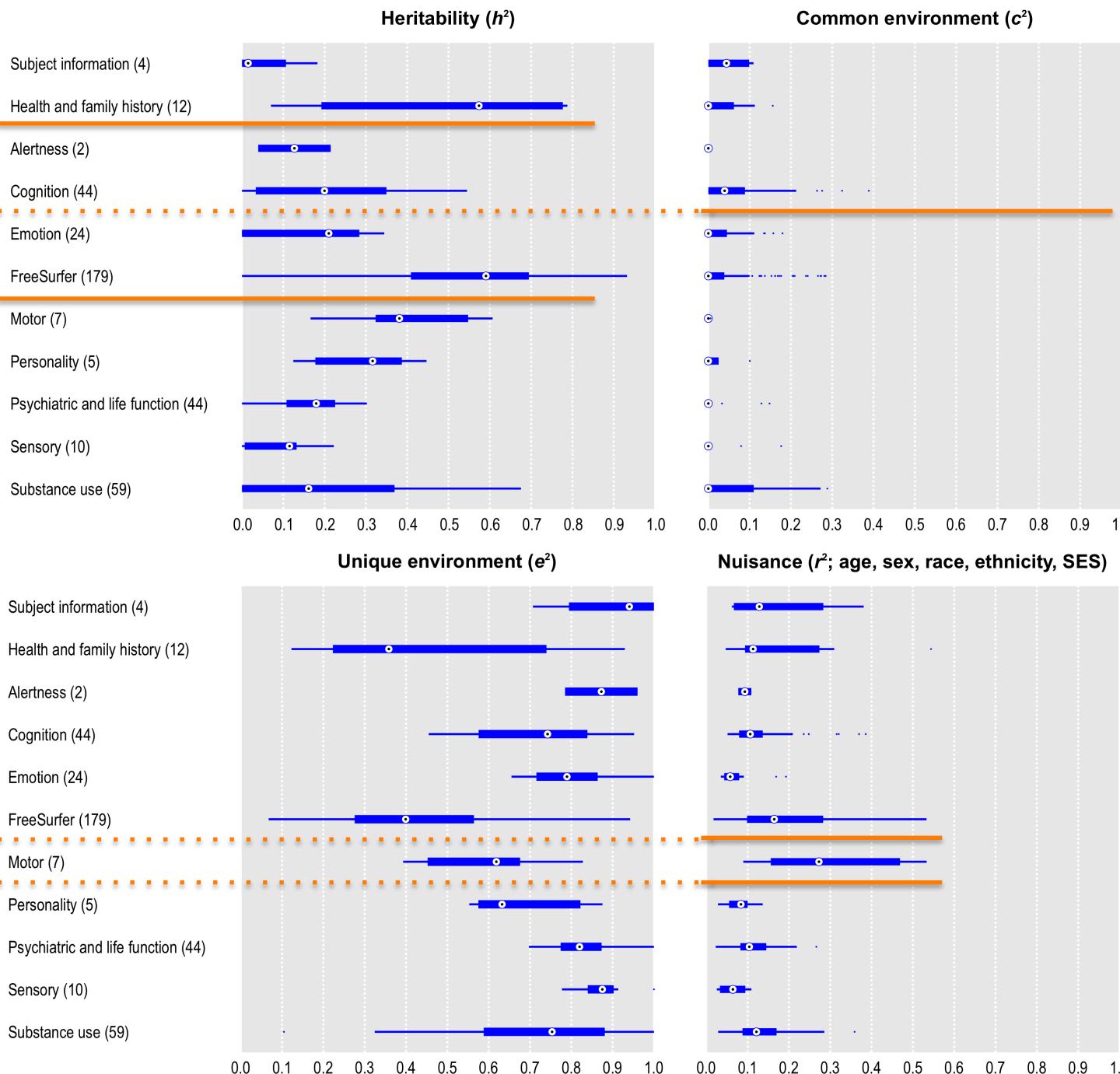
# Heritability Analysis

- SOLAR
  - Nuisance variables regressed out
    - age, sex, age  $\times$  sex, age<sup>2</sup>, and age<sup>2</sup>  $\times$  sex
    - asian, asian  $\times$  sex, black, black  $\times$  sex, white, white  $\times$  sex, hispanic & hispanic  $\times$  sex
- ACE model
  - Common environmental effect also modelled
  - Defined as “Household” effect, by individuals sharing 2 parents

$h^2$   $c^2$

$e^2$  "R<sup>2</sup>"

- Health, brain structure most heritable
- Cognition stands out with relatively high  $c^2$
- Motor & FS large R<sup>2</sup>



# Some Cognitive Measures Have Surprisingly Low $h^2$

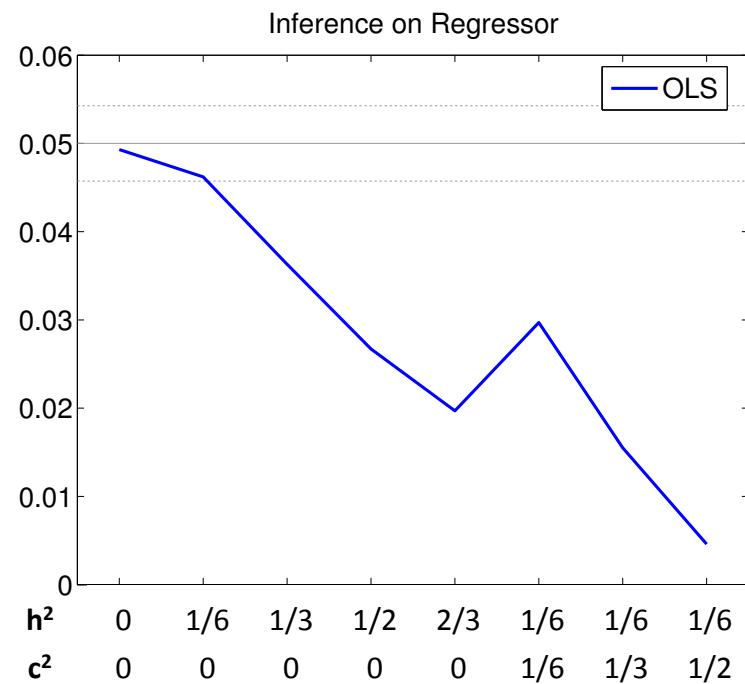
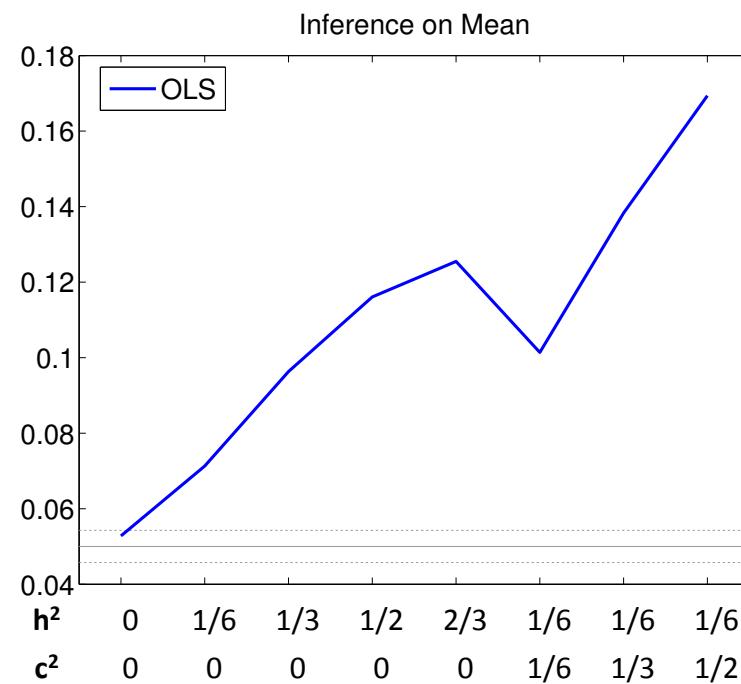
- IQ measures usually cited around 40%
  - But deep literature on this
- Here some as low as 0%
- Indicates substantial environmental effects

Trait	$h^2$	$h^2 p$	$h^2 p_{FDR}$	$c^2$	$c^2 p$	$c^2 p_{FDR}$	$r^2$	
Crystallized	CardSort_AgeAdj (F)	0.0848	3.82E-01	4.58E-01	0.2318	6.54E-02	1.31E-01	0.0945
	Flanker_AgeAdj (F)	0.1648	3.67E-02	1.32E-01	0.0000	-	-	0.0972
	ListSort_AgeAdj (F)	0.5502	4.32E-08	3.89E-07	0.0000	-	-	0.1207
	PicSeq_AgeAdj (F)	0.3827	4.77E-02	1.43E-01	0.0513	3.57E-01	5.84E-01	0.1367
	ProcSpeed_AgeAdj (F)	0.0000	5.00E-01	5.00E-01	0.2715	5.23E-02	1.31E-01	0.1180
	PicVocab_AgeAdj (C)	0.1787	2.00E-01	3.00E-01	0.3149	6.52E-03	3.91E-02	0.3699
	ReadEng_AgeAdj (C)	0.2682	1.12E-01	2.02E-01	0.2042	5.70E-02	1.31E-01	0.3144
	PMAT24_A_CR	0.0000	5.00E-01	5.00E-01	0.1767	4.47E-02	1.31E-01	0.2486
	PMAT24_A_RTCR	0.0000	5.00E-01	5.00E-01	0.0930	1.49E-01	2.68E-01	0.0994
	Fluid	0.2454	1.77E-01	2.90E-01	0.2087	6.57E-02	1.31E-01	0.2026
Prog. Mat.	Crystallised	0.2806	6.99E-02	1.43E-01	0.3196	2.91E-03	2.62E-02	0.3828
	Fluid PC1	0.1522	2.88E-01	3.99E-01	0.2401	4.68E-02	1.31E-01	0.1770
	Fluid PC2	0.4956	2.40E-08	3.89E-07	0.0000	-	-	0.0576
	Fluid PC3	0.3101	4.22E-04	2.53E-03	0.0000	-	-	0.0803
	Fluid PC4	0.1546	5.79E-02	1.43E-01	0.0000	-	-	0.0523
	Fluid PC5	0.1412	3.16E-01	4.06E-01	0.0368	4.13E-01	6.20E-01	0.0319
	Crystallised PC1	0.2785	7.15E-02	1.43E-01	0.3204	2.89E-03	2.62E-02	0.3831
	Crystallised PC2	0.2031	1.11E-02	5.00E-02	0.0000	-	-	0.0382

# Dangers of *Totally Ignoring Heritability*

- 120 subjects, 30 families of 4
  - 15 MZ families, 15 DZ families
- Varying  $h^2$  &  $c^2$  values
- Simple linear regression
  - Inference on Mean & Regressor (uniform random)

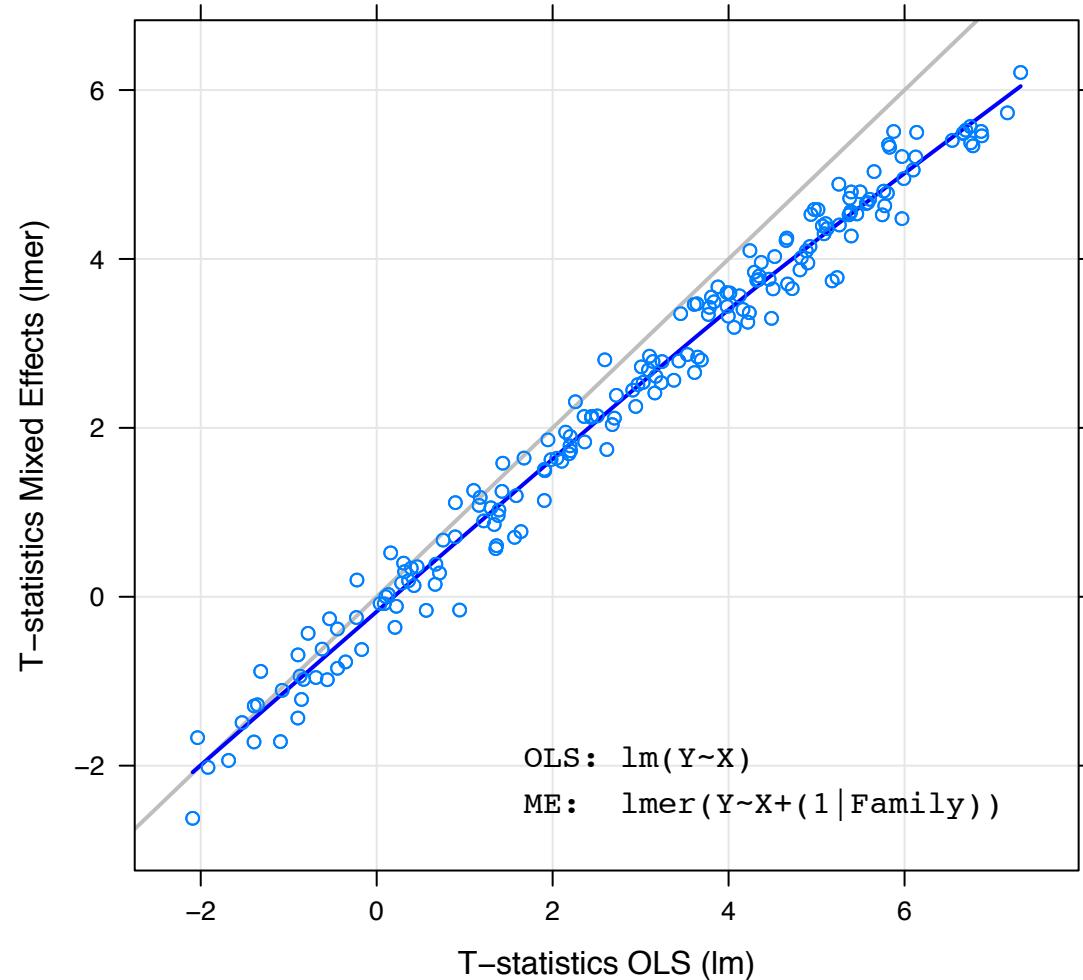
$h^2$	0	1/6	1/3	1/2	2/3	1/6	1/6	1/6
$c^2$	0	0	0	0	0	1/6	1/3	1/2



# Heritability as Nuisance: Empirical Evaluation

- 523 subjects
- 189 Freesurfer traits
- Compare
  - Naïve OLS
  - Mixed Effects, random Family effect
- OLS inflates significance

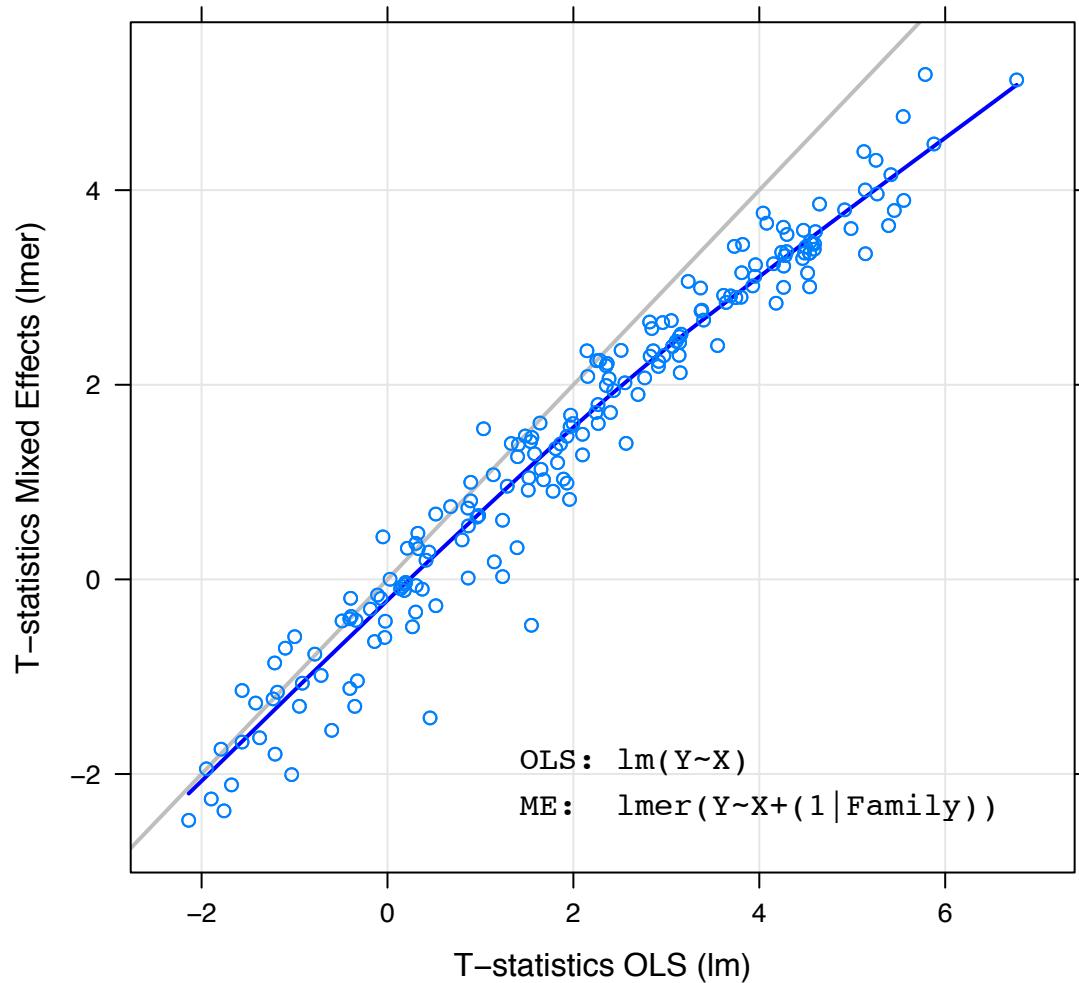
OLS vs Mixed Effects with Family Effect  
189 FS Traits on 'ReadEng\_AgeAdj'



# Heritability as Nuisance: Empirical Evaluation

- 523 subjects
- 189 Freesurfer traits
- Compare
  - Naïve OLS
  - Mixed Effects, random Family effect
- OLS inflates significance

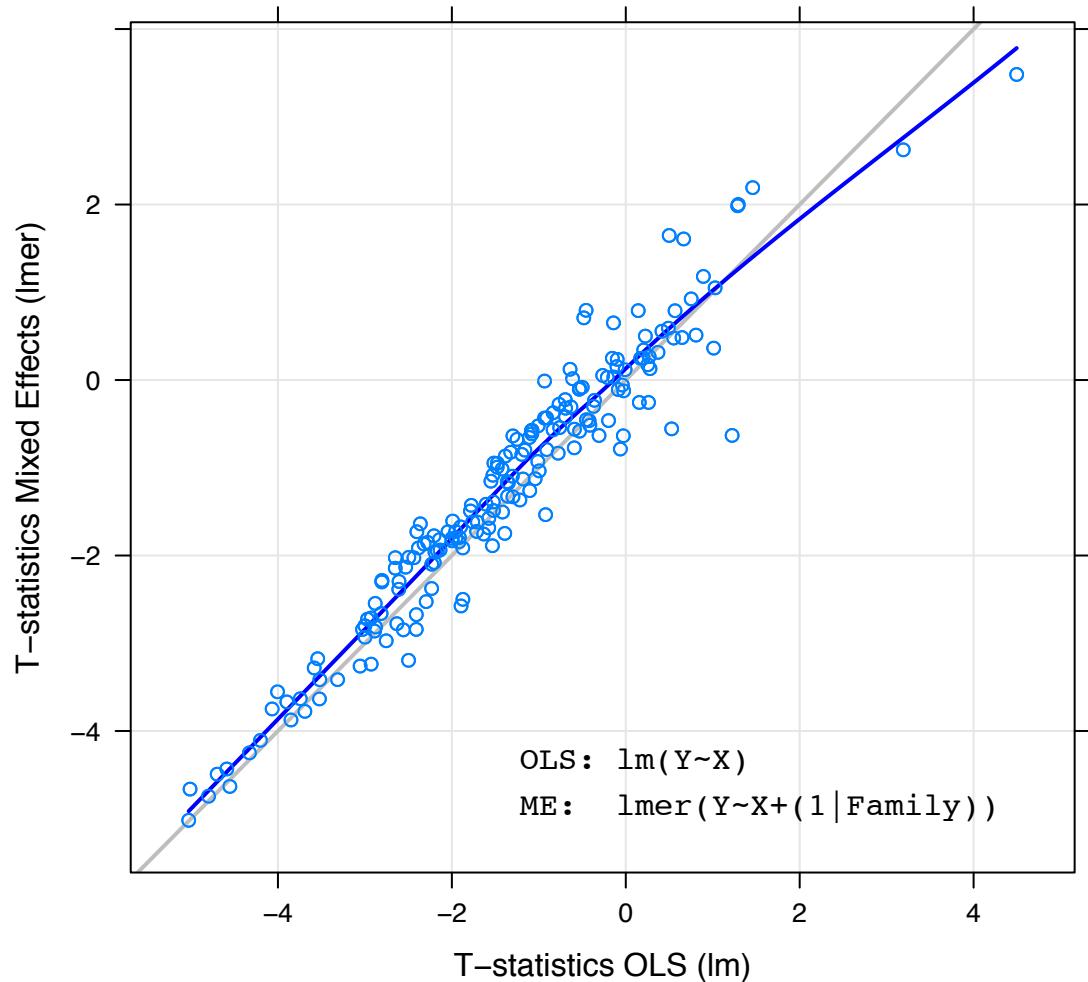
OLS vs Mixed Effects with Family Effect  
189 FS Traits on 'PMAT24\_A\_CR'



# Heritability as Nuisance: Empirical Evaluation

OLS vs Mixed Effects with Family Effect  
189 FS Traits on 'Age\_in\_Yrs'

- 523 subjects
- 189 Freesurfer traits
- Compare
  - Naïve OLS
  - Mixed Effects, random Family effect
- Not so different in null



# Ignoring Heritability

- Available tools
  - SOLAR & SOLAR-Eclipse
    - Will make inference on regressors accounting for family structure
  - OpenMx
    - No inference on regressors (e.g. IQ)
  - Mixed effects, repeated measures code (e.g. R's lmer)
    - Just treat family as a “cluster” or random categorical variable
    - Ignores strong MZ correlation (but is a rare relationship)
      - E.g. in my sim, of 180 kinships, only 15 different due to MZ
  - PALM
    - Permutation inference accounting for family structure  
<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/PALM>

# Conclusions

- Family Structure in HCP data...
- Can be a treasure trove
  - Interesting heritability analyses
  - Will support powerful genetic association analyses
- Can be a nuisance
  - Treating families as a random effect mostly
  - Need to use tools like PALM / SOLAR to account family structure