

ENIGMA & Large Scale Imaging Association

Jason L. Stein

Geschwind Laboratory /

Imaging Genetics Center, Laboratory of Neuro Imaging
University of California, Los Angeles

steinja@ucla.edu

June 10, 2012

Organization for Human Brain Mapping
Introduction to Imaging Genetics Educational Course
Beijing, China



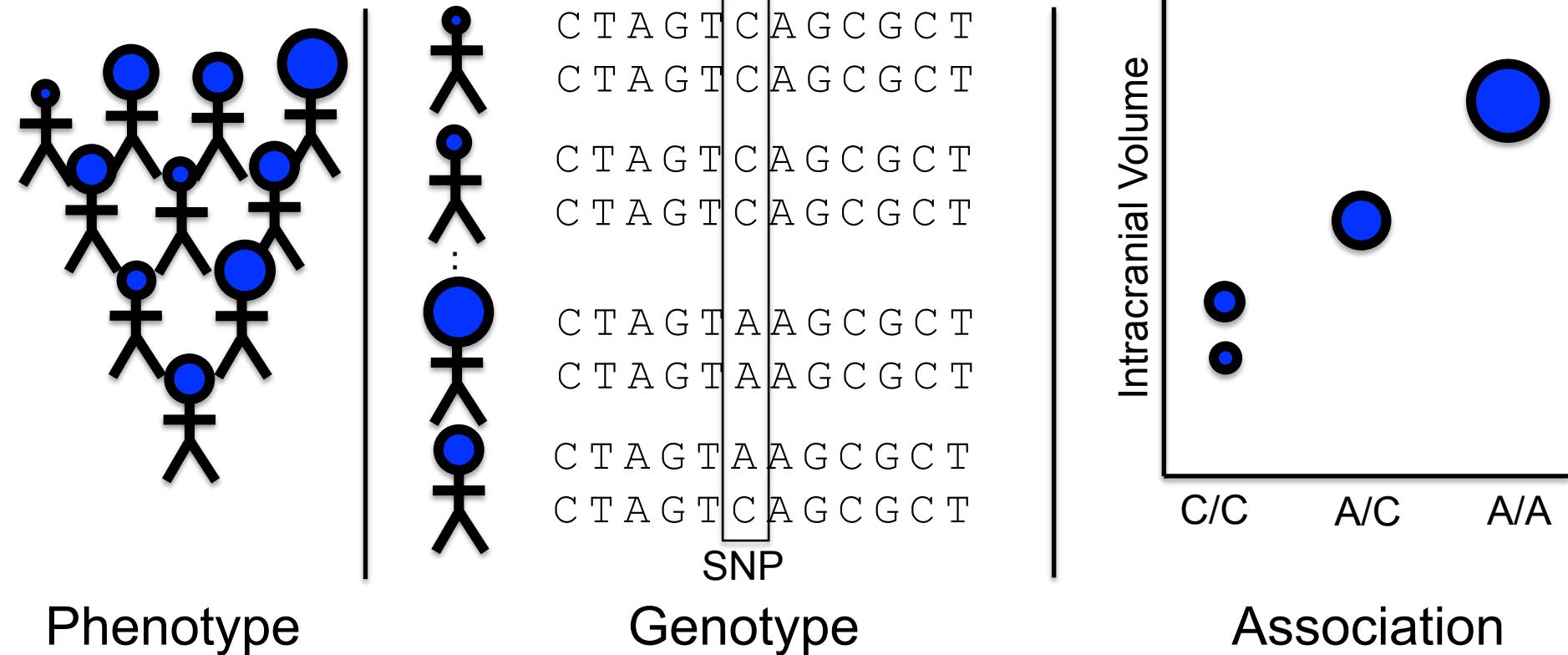
Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 5: Quality Checking
- Step 6: Meta-analysis
- Conclusions

Imaging Genomics Meta-Analysis

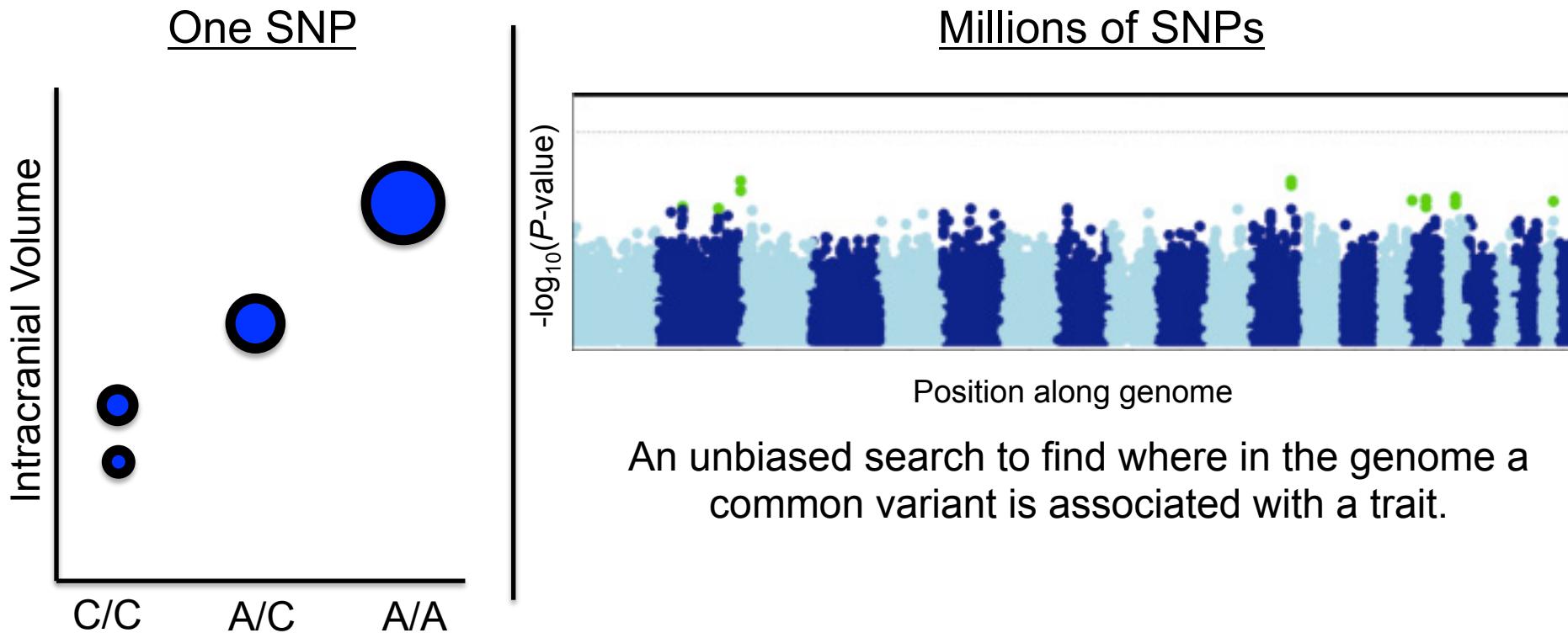
- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 5: Quality Checking
- Step 6: Meta-analysis
- Conclusions

Finding Genetic Variants Influencing a Quantitative Trait



Is there a relationship between genotype and phenotype?

Genome-wide association study



The Powerlessness of the Individual Researcher

When looking at genetic variants, there are (unfortunately!) small effects and many tests conducted. This means you need huge sample sizes to significantly detect variants.

However, both imaging and genotyping are extraordinarily expensive (compared to disease diagnosis and family history), so a consortium is needed.



E. Nigma

Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 6: Quality Checking
- Step 7: Meta-analysis
- Conclusions

Options for how to run a consortium

Mega-analysis

PSYCHIATRIC • GWAS • CONSORTIUM

- Raw phenotype and genotype data are uploaded for each subject to a central server

Meta-analysis



- Collaboration is necessary for significant results, but sharing data is difficult because of IRB and ownership issues.
- ENIGMA is a data free results filled network, we get uploads of statistical results – no raw data

Advantages of each

Mega-analysis

PSYCHIATRIC • GWAS • CONSORTIUM

- Ability to check every part of the analysis
- Moves as fast as you make it (no waiting)
- Greater analysis possibilities (polygenic score, structural equation modeling)

Meta-analysis

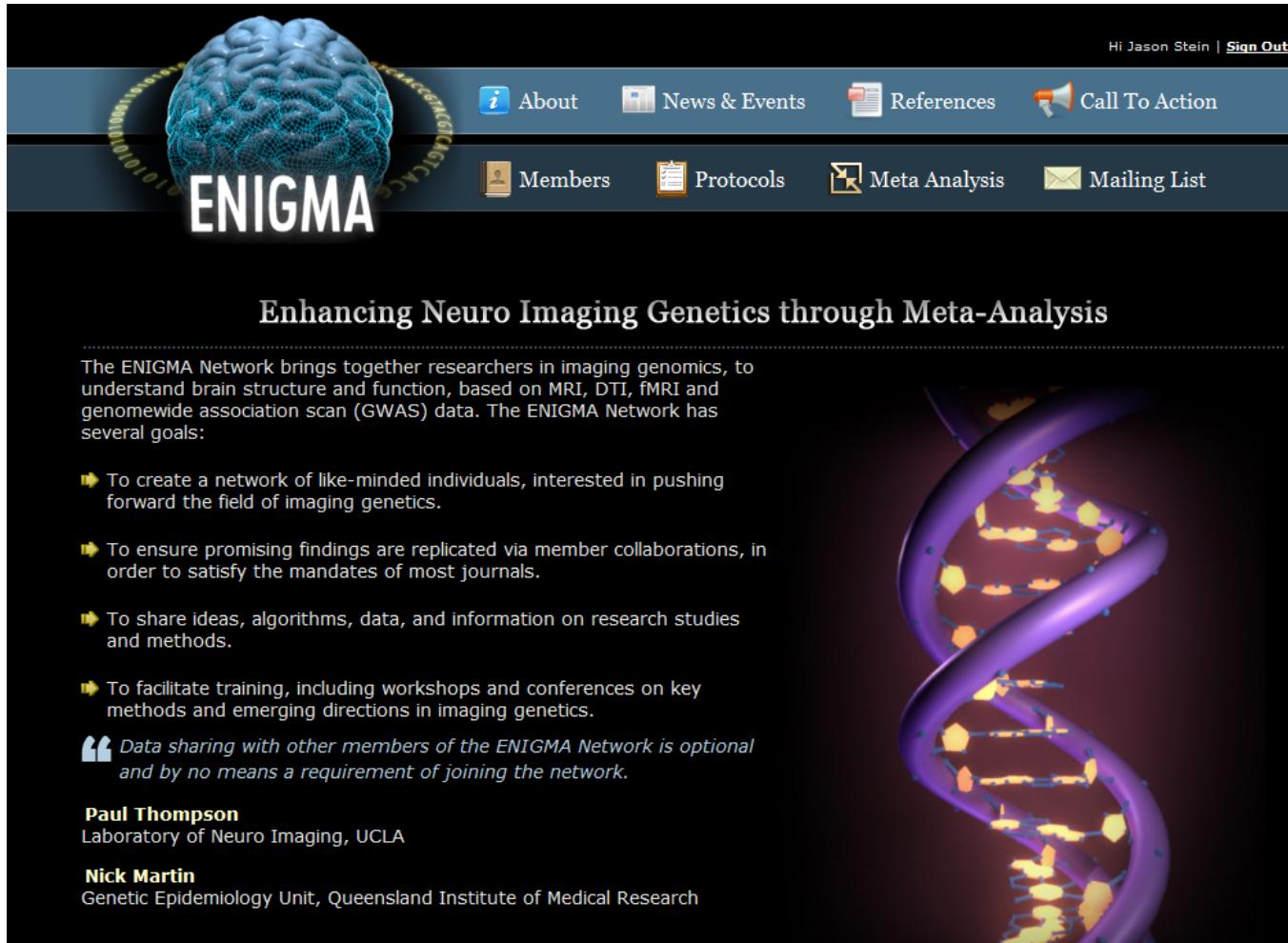
Collaborate



- (We think) more groups are willing to share results than data
- Distributed approach

Replication through collaboration

<http://enigma.loni.ucla.edu>

A screenshot of the ENIGMA Network website. The header features a 3D brain model and the word "ENIGMA". Navigation links include About, News & Events, References, Call To Action, Members, Protocols, Meta Analysis, and Mailing List. A sub-header reads "Enhancing Neuro Imaging Genetics through Meta-Analysis". Below this, a text block states: "The ENIGMA Network brings together researchers in imaging genomics, to understand brain structure and function, based on MRI, DTI, fMRI and genomewide association scan (GWAS) data. The ENIGMA Network has several goals: To create a network of like-minded individuals, interested in pushing forward the field of imaging genetics. To ensure promising findings are replicated via member collaborations, in order to satisfy the mandates of most journals. To share ideas, algorithms, data, and information on research studies and methods. To facilitate training, including workshops and conferences on key methods and emerging directions in imaging genetics." A quote from Paul Thompson follows: "Data sharing with other members of the ENIGMA Network is optional and by no means a requirement of joining the network." The footer lists Paul Thompson (Laboratory of Neuro Imaging, UCLA) and Nick Martin (Genetic Epidemiology Unit, Queensland Institute of Medical Research).

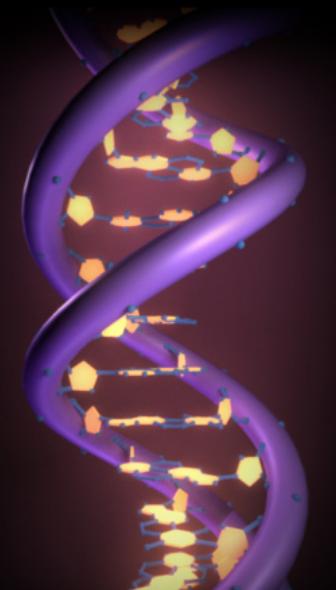
The ENIGMA Network brings together researchers in imaging genomics, to understand brain structure and function, based on MRI, DTI, fMRI and genomewide association scan (GWAS) data. The ENIGMA Network has several goals:

- ▶ To create a network of like-minded individuals, interested in pushing forward the field of imaging genetics.
- ▶ To ensure promising findings are replicated via member collaborations, in order to satisfy the mandates of most journals.
- ▶ To share ideas, algorithms, data, and information on research studies and methods.
- ▶ To facilitate training, including workshops and conferences on key methods and emerging directions in imaging genetics.

“Data sharing with other members of the ENIGMA Network is optional and by no means a requirement of joining the network.”

Paul Thompson
Laboratory of Neuro Imaging, UCLA

Nick Martin
Genetic Epidemiology Unit, Queensland Institute of Medical Research



> 200 scientist members from 12 countries all over the world!

Imaging Segmentation, Imputation, and Association protocols provided at our website.

Project Description

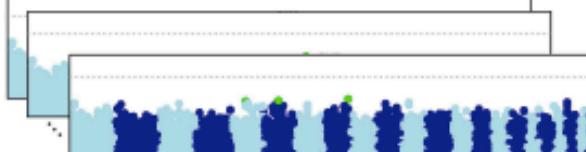
Completed at the level of the individual site

Image Acquisition

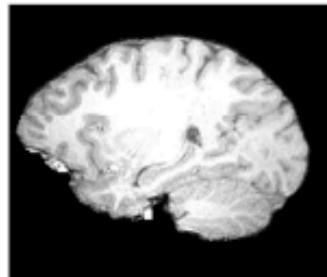
Image Processing

1

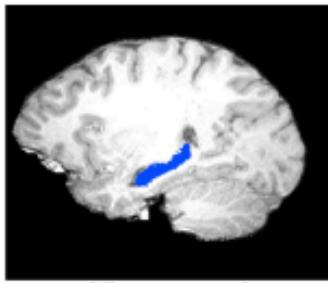
All Subjects



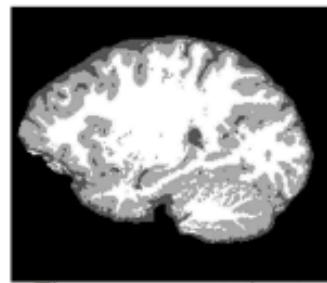
Imaging Protocols



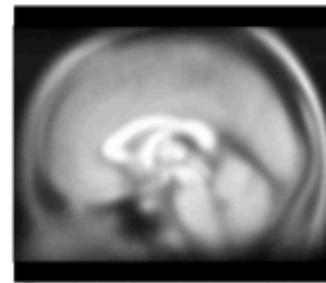
T_1 -weighted structural



Hippocampal segmentation



Tissue segmentation to calculate brain volume (white matter + grey matter)

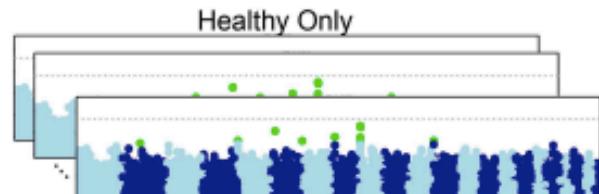


Registration to template to calculate ICV

Genetics Protocols

GGG GAA GGG T TGG GAA AAC C CCG GGG G GCG GGG GT T
TT GGG GAA AAC C CAG AAG AAC TCA CCT ATC G GCG GGG GTC
T GCG GCG TCT TTT TAC TTA AAC CCT G GCG GGG GTC
RATT T TCG GGG G GCG CAG AAG AAC TCA CCT ATC G GCG GGG GTC
TAT TTG TCT TTT CCT G GCG GGG G GCG GGG GTC
CCG GGT TGG GAG GAC CCT TGA GAG GGG G GCG GGG GTC
GCC GGT TGG GAG GAC CCT TGA GAG GGG G GCG GGG GTC
RAG RAG RAG ATT TTG TCT GAA TCC CCT G GCG GGG GTC
GAG GAG GAG CCG ATG CG G GCG GGG G GCG GGG GTC
GAG GGT TTT TCC CCG G GCG GGG G GCG GGG GTC
GTT TCT TCC CCG G GCG GGG G GCG GGG GTC
TTT TGG AGT GAG TGG GAG GGG G GCG GGG GTC
CAG GGT CT GCG GGG G GCG GGG G GCG GGG GTC
TTT TGG GAG GAG GGG G GCG GGG G GCG GGG GTC
GGG GGG GGG G GCG T GAG GGG G GCG GGG GTC
GAG GGT TCT GGG G GCG GGG G GCG GGG GTC
TAC GAG GCG G GCG G GCG GGG G GCG GGG GTC
GTT TGG AGT GAG TGG GAG GGG G GCG GGG GTC
CCT CCG GGT GAG TGG GAG GGG G GCG GGG GTC

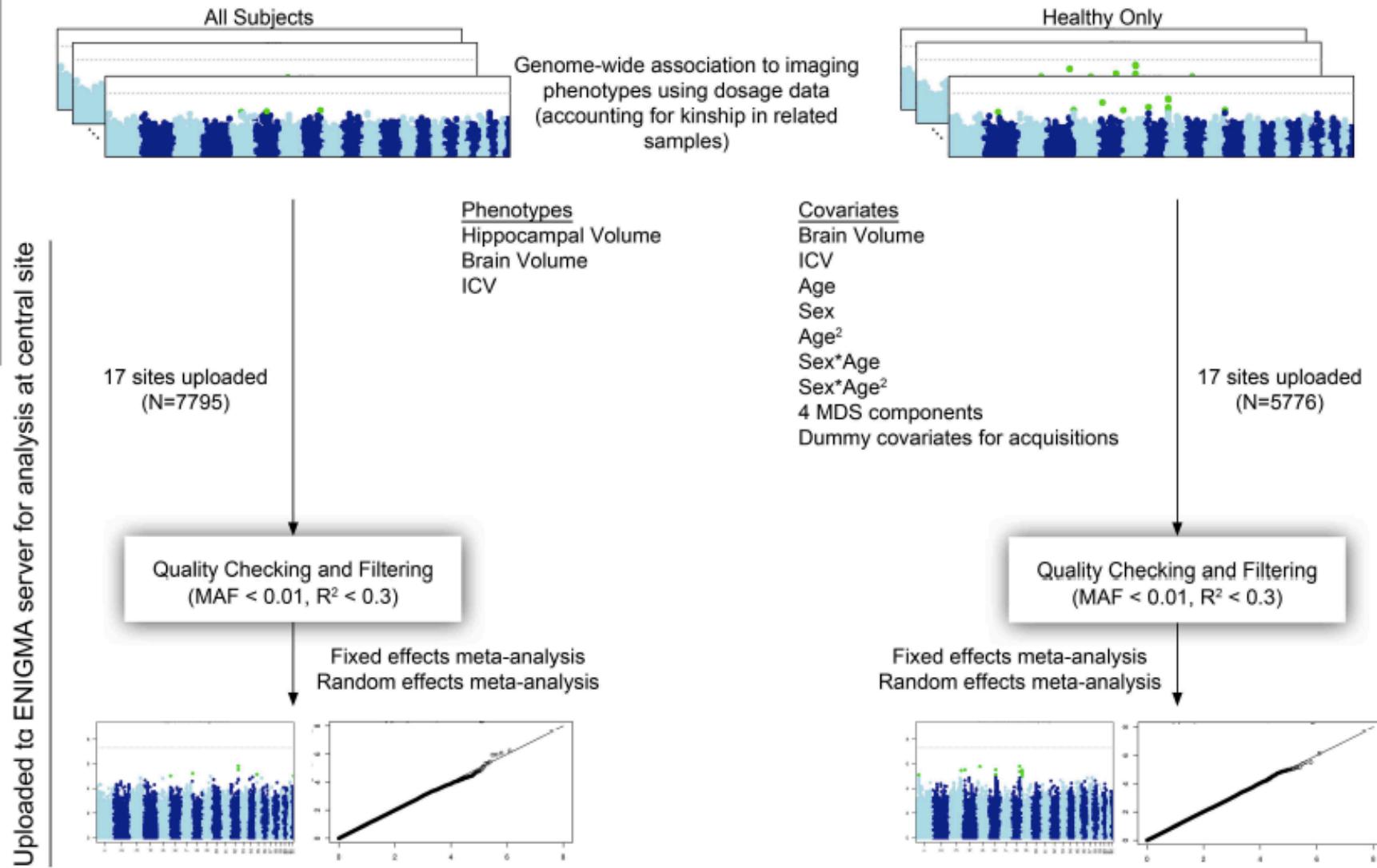
Genome-wide genotypes imputed to
1,387,466 autosomal single nucleotide
polymorphisms (SNPs) based on
HapMap III reference panels



Genome-wide association to imaging phenotypes using dosage data (accounting for kinship in related samples)

Project Description

Complete



Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 6: Quality Checking
- Step 7: Meta-analysis
- Conclusions

ENIGMA Upload Page

Having trouble? Please see the [FAQ](#) or [email Enigma](#).

Contact Information

Group/Project Name
Contact Person
Contact Email

Sample Description

Age Mean (all subjects)
Age Standard Deviation (all subjects)
Age Range (all subjects)
Number of males/number of females (all subjects)
Percentage of each ethnicity (all subjects)
Age Mean (healthy subjects only)
Age Standard Deviation (healthy subjects only)
Age Range (healthy subjects only)
Number of males/number of females (healthy subjects only)
Percentage of each ethnicity (healthy subjects only)
Scanner(s) used

Detailed scanning acquisition paragraph
(Including resolution, acquisition direction, MRI pulse sequence, and other standard MRI protocol summaries)

T1-weighted images of the brain were acquired with an inversion recovery rapid gradient echo sequence on a 4 Tesla MRI scanner (Bruker Medspec; with acquisition parameters: TI/TR/TE = 700/1500/3.35 ms; flip angle = 8 degrees; slice thickness = 0.9 mm, 256 x 256 x 256).

Genotyping chip used

Goal is to get as much information at the beginning as you possibly can

ENIGMA Automated Submission System

HapMap Count Post Filter 210,737

Total Uploaded SNPs 1,387,488

RSQR Failed 170,058

P Val Failed 0

Freq Failed 40,701

Total SNPs For Analysis 1,176,729

Stage 1

File Sample:

1	rs3131967	T	C	0.110	HippoVol	-123.792	164.682	2.509	0.123	0.4522	0.1099
0.0392											
1	rs1048488	C	T	0.160	HippoVol	-118.594	138.708	3.160	0.159	0.3926	0.1598
0.0449											
1	rs12562034	G	A	0.898	HippoVol	-13.048	36.165	0.026	0.028	0.7183	0.8979
0.9862											
1	rs12124819	A	G	0.788	HippoVol	5.057	99.895	0.007	0.001	0.9596	0.7885
0.0740											
1	rs4040617	G	A	0.114	HippoVol	-122.766	158.387	2.538	0.130	0.4383	0.1135
0.0433											
1	rs2905036	T	Z	1.000	HippoVol	-0.000	100000.000	0.000	0.000	1	1.0000

Stage 2

File Sample:

SNP	AL1	AL2	EFFECT	FREQ1	FREQ2	H2	LOD	PVALUE	RSQ	SE	TRAIT,,ML_RSQR	
rs11778460	C	G	82.726	0.912	0.920	0.946	0.03685	0.9911	39.631	HippoVol	0.9911	
rs333277	A	G	-13.554	0.629	0.6294	0.072	0.079	0.5465	0.9834	22.477	HippoVol	0.9834
rs10806671	T	C	22.571	0.410	0.4102	0.206	0.218	0.3163	0.9956	22.523	HippoVol	0.9956
rs6079035	T	G	4.090	0.650	0.6496	0.006	0.006	0.8638	0.9977	23.837	HippoVol	0.9977
rs7173425	C	T	18.975	0.917	0.9172	0.046	0.048	0.6395	0.9934	40.506	HippoVol	0.9934
rs2239669	G	A	11.094	0.687	0.6867	0.044	0.042	0.6588	0.9346	25.125	HippoVol	0.9346
rs9901367	T	G	39.442	0.4831	0.4831	0.650	0.736	0.06563	0.9801	21.425	HippoVol	0.9801
rs734883	C	T	31.274	0.989	0.9886	0.018	0.015	0.7902	0.8344	117.577	HippoVol	0.8344
rs10459518	G	A	21.100	0.834	0.8337	0.103	0.118	0.4618	0.9986	28.67	HippoVol	0.9986
rs6961069	T	C	-3.641	0.409	0.4090	0.005	0.006	0.8716	0.9984	22.531	HippoVol	0.9984

Stage 3

File Sample:

SNP	EFFECT	ALLELE	SE	NON_EFFECT	ALLELE	BETA	FREQN	P_VAL
rs11778460	C	39.631	G	82.726	0.912	579	0.03685	
rs333277	A	22.477	G	-13.554	0.629	579	0.5465	
rs10806671	T	22.523	C	22.571	0.410	579	0.3163	
rs6079035	T	23.837	G	4.090	0.650	579	0.8638	
rs7173425	C	40.506	T	18.975	0.917	579	0.6395	
rs2239669	G	25.125	A	11.094	0.687	579	0.6588	
rs9901367	T	21.425	G	39.442	0.483	579	0.06563	
rs734883	C	117.577	T	31.274	0.989	579	0.7902	
rs10459518	G	28.67	A	21.100	0.834	579	0.4618	
rs6961069	T	22.531	C	-3.641	0.409	579	0.8716	

Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 5: Quality Checking
- Step 6: Meta-analysis
- Conclusions

Filtering by Quality of Imputation

- **RSQR_HAT**: Estimates the squared correlation between imputed and true genotypes.
- Typically, a cut-off of **0.30** will flag most of the poorly imputed SNPs but only a small number (<1%) of well imputed SNPs

The true genotype is generally unknown so how do you get this?

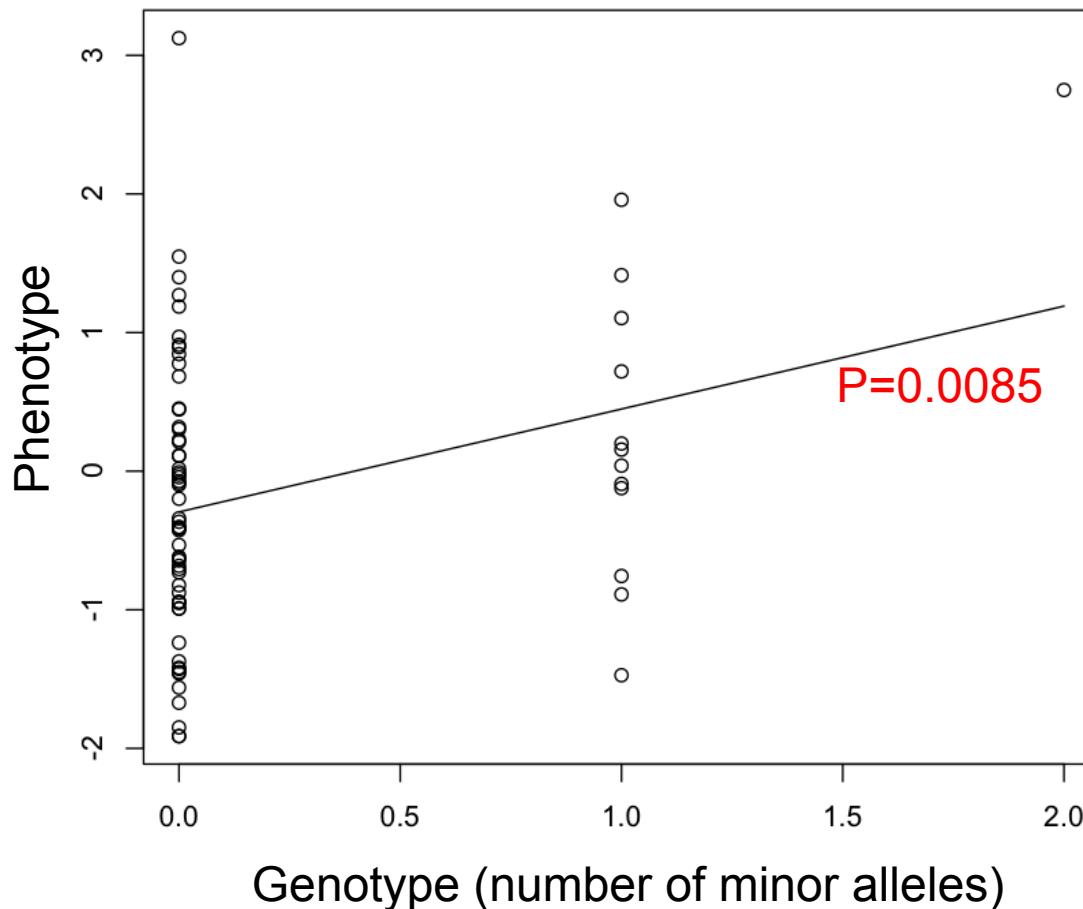
Take the ratio of the empirically observed variance of the allele dosage (from imputed results) to the expected binomial variance of HWE calculated by allele frequency from HapMap

How to filter based on imputation quality using mach2qtl output

TRAIT	INFORMATION FROM .info FILE				QTL ASSOCIATION ADDITIVE				
	MARKER	ALLELES	FREQ1	RSQR	EFFECT2	STDERR	CHISQ	PVALUE	N
HippoVol	rs9755941	T,C	.5160	.9164	3.271	2.147	2.3203	0.1277	743
HippoVol	rs13060385	G,A	.4851	.9091	-0.904	2.168	0.1740	0.6766	743
HippoVol	rs13072188	A,C	.6026	.9977	0.949	2.120	0.2005	0.6543	743
HippoVol	rs9713920	T,C	.8462	.9450	-0.387	2.951	0.0172	0.8958	743
HippoVol	rs9683305	T,C	.2971	.9471	-2.555	2.360	1.1724	0.2789	743
HippoVol	rs9681213	A,G	.3315	.9980	-1.302	2.211	0.3468	0.5559	743
HippoVol	rs11928872	A,-	1.00	.0000	-0.000	10000000000000000000	0.0000	1	743
HippoVol	rs13089679	C,T	.3101	.9656	-3.206	2.311	1.9243	0.1654	743
HippoVol	rs9682794	T,C	.7124	.9613	1.643	2.310	0.5059	0.4769	743
HippoVol	rs12637617	T,C	.8450	.9503	-0.369	2.936	0.0158	0.9	743
HippoVol	rs17075464	A,G	.9991	.5082	-42.070	47.885	0.7719	0.3796	743
HippoVol	rs9284831	A,G	.7447	.9551	-3.559	2.401	2.1971	0.1383	743
HippoVol	rs6790898	T,A	.9988	.3630	-30.931	50.891	0.3694	0.5433	743
HippoVol	rs1516320	C,T	.2693	.9644	-3.083	2.426	1.6147	0.2038	743
HippoVol	rs1548188	G,A	.7865	.9602	-3.669	2.512	2.1342	0.144	743
HippoVol	rs1400174	A,G	.7148	.9677	1.535	2.306	0.4433	0.5055	743
HippoVol	rs1516321	A,G	.7867	.9935	1.065	2.560	0.1730	0.6775	743
HippoVol	rs41342346	C,T	.9633	.9525	-6.101	5.813	1.1018	0.2939	743
HippoVol	rs1949341	C,A	.8876	.9614	-1.300	3.394	0.1467	0.7017	743
HippoVol	rs13088719	G,T	.8518	.9597	-2.560	3.047	0.7060	0.4008	743

```
% awk '{if (NR>32 && $5>=0.3) print  
$1,$2,$3,$4,$5,$6,$7,$8,$9,$10}' chr3-mach2qtl.out > chr3-  
rsqrfilt.out
```

Filter by low Minor Allele Frequency (MAF)



Associations with low MAF can be driven by outliers in the data. It is common to exclude SNPs with MAF < 0.01.

How to filter based on frequency using mach2qtl output

TRAIT	INFORMATION FROM .info FILE				QTL ASSOCIATION ADDITIVE				
	MARKER	ALLELES	FREQ1	RSQR	EFFECT2	STDERR	CHISQ	PVALUE	N
HippoVol	rs9755941	T,C	.5160	.9164	3.271	2.147	2.3203	0.1277	743
HippoVol	rs13060385	G,A	.4851	.9091	-0.904	2.168	0.1740	0.6766	743
HippoVol	rs13072188	A,C	.6026	.9977	0.949	2.120	0.2005	0.6543	743
HippoVol	rs9713920	T,C	.8462	.9450	-0.387	2.951	0.0172	0.8958	743
HippoVol	rs9683305	T,C	.2971	.9471	-2.555	2.360	1.1724	0.2789	743
HippoVol	rs9681213	A,G	.3315	.9980	-1.302	2.211	0.3468	0.5559	743
HippoVol	rs11928872	A,-	1.00	.0000	-0.000	10000000000000000000	0.0000	1	743
HippoVol	rs13089679	C,T	.3101	.9656	-3.206	2.311	1.9243	0.1654	743
HippoVol	rs9682794	T,C	.7124	.9613	1.643	2.310	0.5059	0.4769	743
HippoVol	rs12637617	T,C	.8450	.9503	-0.369	2.936	0.0158	0.9	743
HippoVol	rs17075464	A,G	.9991	.5082	-42.070	47.885	0.7719	0.3796	743
HippoVol	rs9284831	A,G	.7447	.9551	-3.559	2.401	2.1971	0.1383	743
HippoVol	rs6790898	T,A	.9988	.3630	-30.931	50.891	0.3694	0.5433	743
HippoVol	rs1516320	C,T	.2693	.9644	-3.083	2.426	1.6147	0.2038	743
HippoVol	rs1548188	G,A	.7865	.9602	-3.669	2.512	2.1342	0.144	743
HippoVol	rs1400174	A,G	.7148	.9677	1.535	2.306	0.4433	0.5055	743
HippoVol	rs1516321	A,G	.7867	.9935	1.065	2.560	0.1730	0.6775	743
HippoVol	rs41342346	C,T	.9633	.9525	-6.101	5.813	1.1018	0.2939	743
HippoVol	rs1949341	C,A	.8876	.9614	-1.300	3.394	0.1467	0.7017	743
HippoVol	rs13088719	G,T	.8518	.9597	-2.560	3.047	0.7060	0.4008	743

--More--(0%)

```
awk '{if (NR>32 && ($4 >= 0.01 && $4 <= 0.99) ) print  
$1,$2,$3,$4,$5,$6,$7,$8,$9,$10}' chr3-mach2qtl.out > chr3-  
freqfilt.out
```

Filtering P-values

P-value = NA
P-value = -1

	Genotype Dosage	Scanner 2
1	0.01	1 0 0
1	0.99	1 0 0
1	1.98	0 1 0
1	1.00	0 1 0
1	1.02	0 0 1

Implicit Intercept Scanner 1 Scanner 3

Incorrect: Perfect collinearity with intercept

	Genotype Dosage	Scanner 2
1	0.01	1 0
1	0.99	1 0
1	1.98	0 1
1	1.00	0 1
1	1.02	0 0

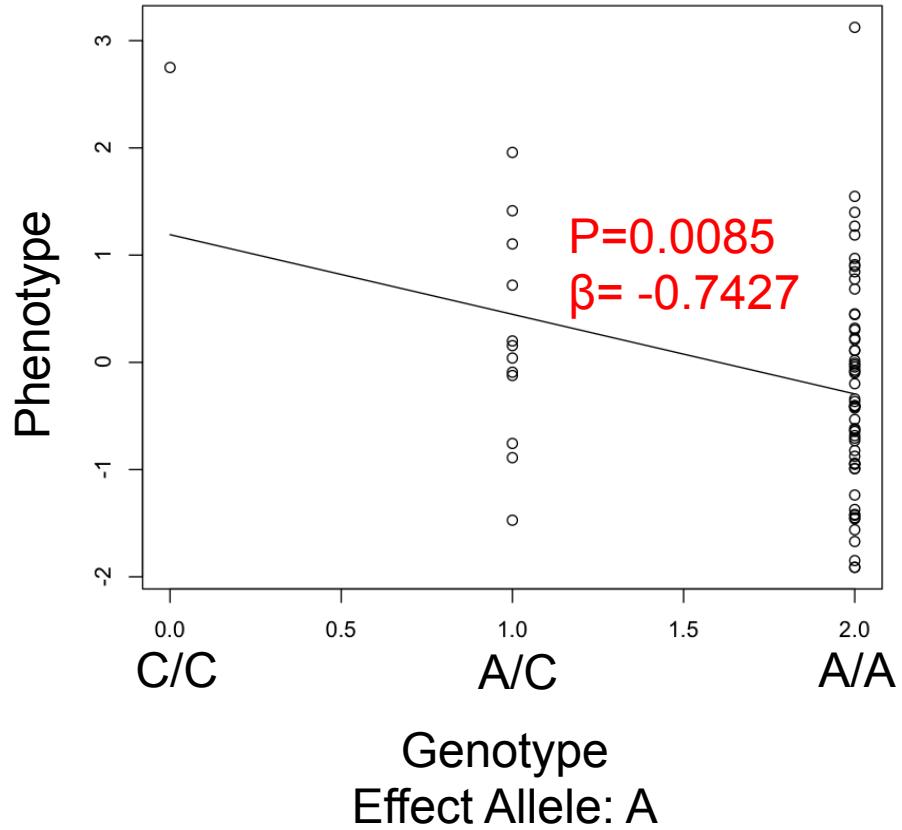
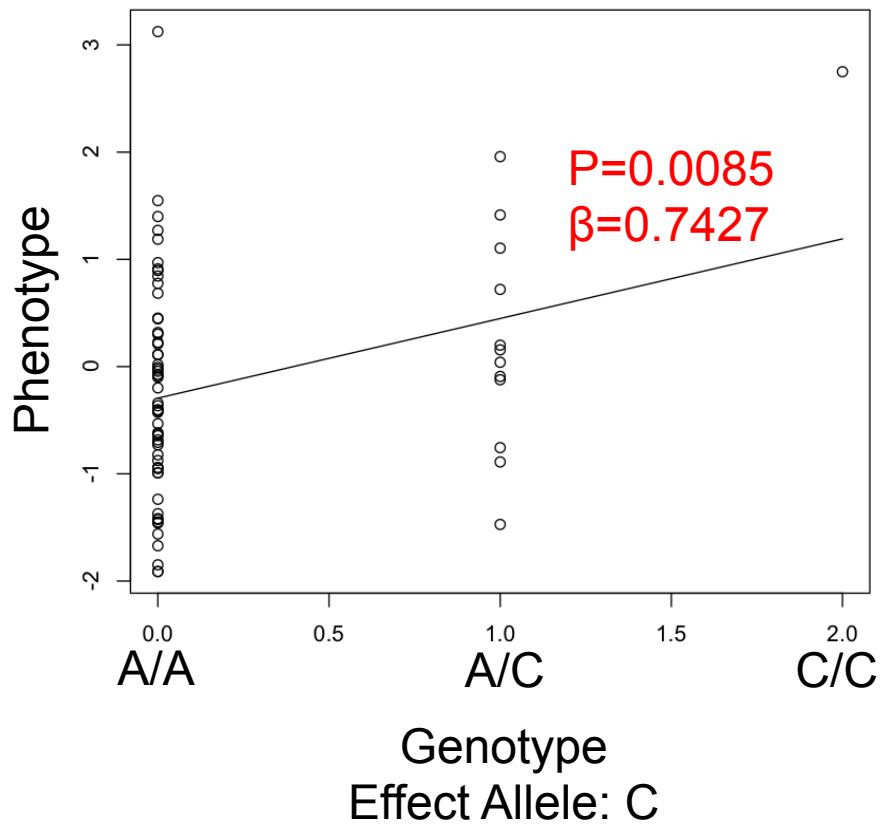
Implicit Intercept Scanner 1

Correct: Remove one dummy covariate

Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 5: Quality Checking
- Step 6: Meta-analysis
- Conclusions

What is the Effect Allele?



The effect allele is crucial in meta-analysis. Depending on how the association is coded, you can get the same P-value, but different direction of effect in meta-analysis.

Assignment of Effect Allele

Program	Effect Allele
mach2qtl	A2
Plink	A1
Merlin-Offline	A1
ProbABEL	A1
snptest	allele_B
Quicktest	alleleB
Solar	A1

Effect allele varies by association and imputation program. It is not always an easy or readily apparent thing to figure out either so you have to delve into the manuals!

Frequency Allele



TRAIT		INFORMATION FROM .info FILE			QTL ASSOCIATION ADDITIVE					
		MARKER	ALLELES	FREQ1	RSQR	EFFECT2	STDERR	CHISQ	PVALUE	N
HippoVol		rs9755941	T,C	.5160	.9164	3.271	2.147	2.3203	0.1277	743
HippoVol		rs13060385	G,A	.4851	.9091	-0.904	2.168	0.1740	0.6766	743
HippoVol		rs13072188	A,C	.6026	.9977	0.949	2.120	0.2005	0.6543	743

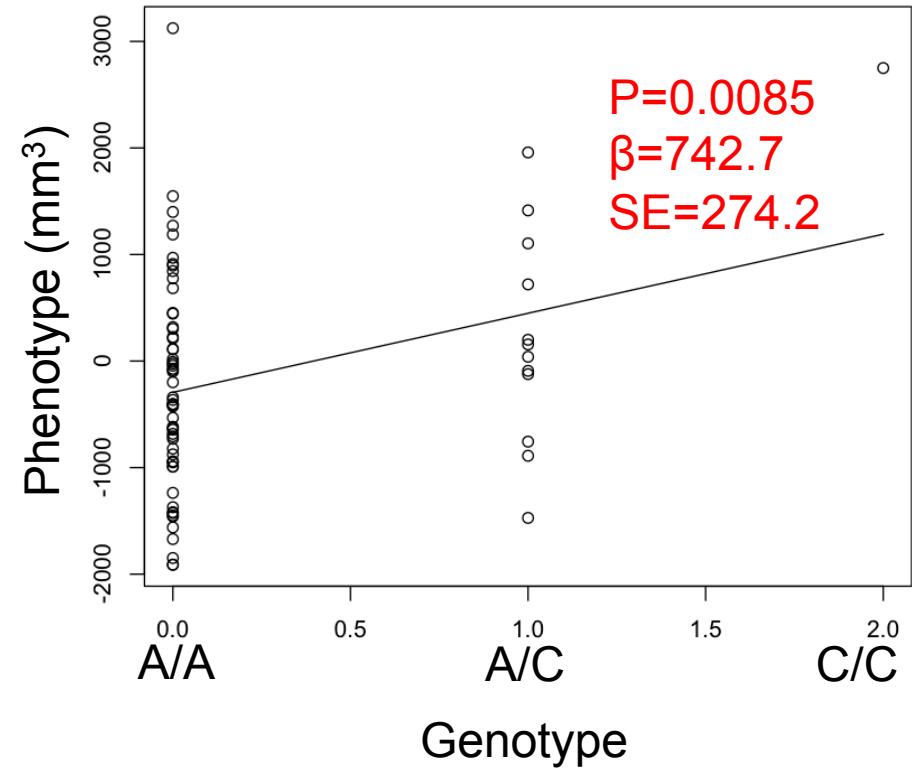
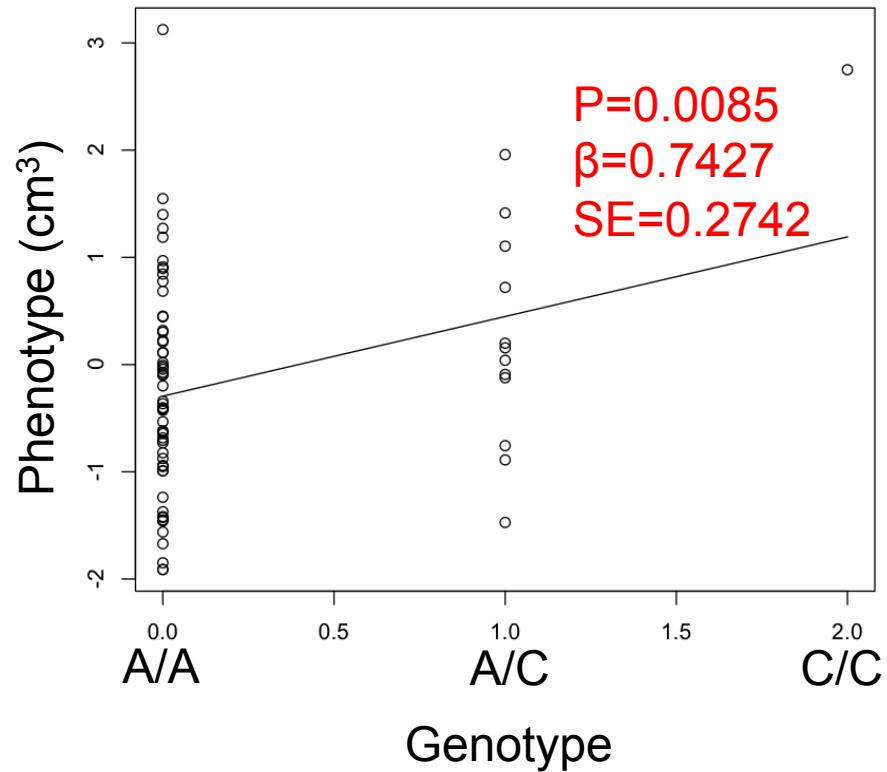
Effect allele is not the same as the frequency allele in mach2qtl
A simple 1-FREQ1 gives you the frequency of the effect allele

Assignment of Frequency Allele

Program	Effect Allele	Frequency Allele
mach2qtl	A2	A1
Plink	A1	A1
Merlin-Offline	A1	A1
ProbABEL	A1	A1
snptest	allele_B	Genotype counts given
Quicktest	alleleB	Genotype counts given
Solar	A1	A1

Frequency allele varies by association and imputation program. It is not always an easy or readily apparent thing to figure out either so you have to delve into the manuals!

Correct scale for beta / SE



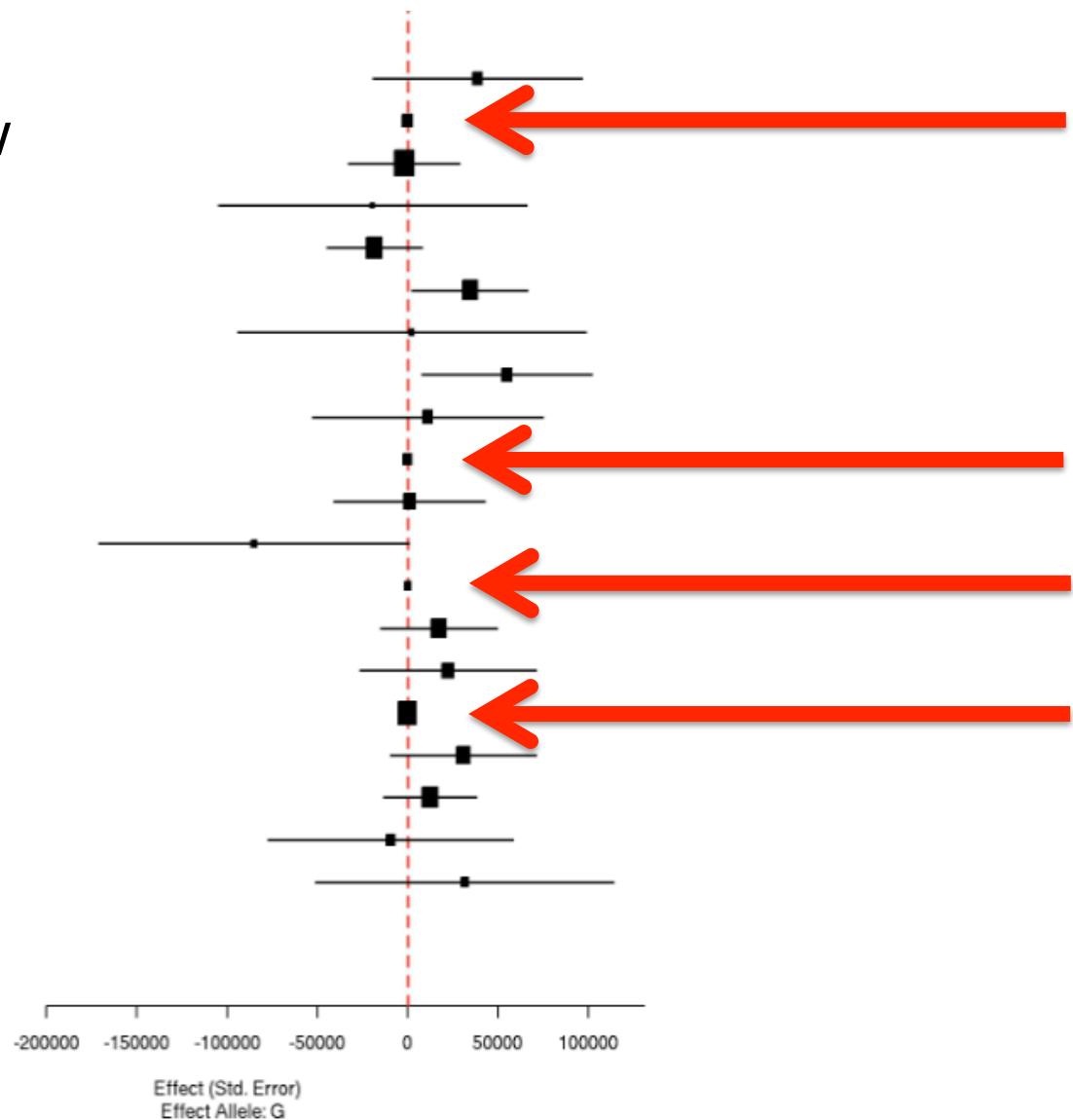
Units will affect beta values and standard errors in regression so need to be very careful to know what units analyses were run in. You can use simple scaling factors if you know the conversion:

$$\beta_{\text{mm}} = \beta_{\text{cm}} * 1000$$

$$SE_{\text{mm}} = SE_{\text{cm}} * 1000$$

When Scaling Goes Wrong

- Forest Plots can show you something is off in units relative to other groups
- Assuming similar effect sizes and standard errors between groups which should be a safe assumption



A side note on family based samples

Calculation of association statistics in family based samples is generally done through a **minimization algorithm** (see SOLAR or merlin-offline) which **generally fails with large phenotype values.**

You can use this Beta/SE trick in your favor in this case!

Divide the phenotype values by a constant prior to association, then multiply the Beta/SE values by the constant after association so meta-analysis is in the same scale.

HapMap3 Specific Troubles

TRAIT	INFORMATION FROM .info FILE				QTL ASSOCIATION ADDITIVE					
	MARKER	ALLELES	FREQ1	RSQR	EFFECT2	STDERR	CHISQ	PVALUE	N	
HippoVol	AFFX-SNP_7351650__rs7539366G,C			.5220	.9396	101.171	61.240	2.7293	0.09853	48
HippoVol	AFFX-SNP_8199384__rs9430274A,G			.8558	.9980	39.402	83.564	0.2223	0.6373	48
HippoVol	AFFX-SNP_5748019__rs9660152C,T			.5238	1.00	-44.034	58.982	0.5574	0.4553	48
HippoVol	AFFX-SNP_12165450__rs1361396T,C			.7189	.7472	132.774	86.360	2.3637	0.1242	48
HippoVol	AFFX-SNP_11712252__rs6673996T,A			.9051	.9680	5.982	118.553	0.0025	0.9598	48
HippoVol	AFFX-SNP_10563010__rs12049021G,A			.7910	.9606	43.277	85.575	0.2558	0.6131	48
HippoVol	AFFX-SNP_9790694__rs1984490A,G			.7725	.9795	11.969	63.384	0.0357	0.8502	48
HippoVol	AFFX-SNP_10243548__rs3118040A,G			.5177	.9645	41.589	59.065	0.4958	0.4814	48
HippoVol	AFFX-SNP_9852944__rs2764670C,T			.8948	1.00	-286.16	123.251	5.3907	0.02024	48
HippoVol	AFFX-SNP_10105880__rs3737108A,T			.5440	.9576	92.364	63.150	2.1393	0.1436	48
HippoVol	AFFX-SNP_8016994__rs10802131A,T			.8855	1.00	-336.00	122.869	7.4782	0.006245	48
HippoVol	AFFX-SNP_11382670__rs1322374A,G			.7042	.9065	46.225	77.967	0.3515	0.5533	48
HippoVol	AFFX-SNP_10675579__rs284082C,T			.5906	.9979	16.443	62.788	0.0686	0.7934	48
HippoVol	AFFX-SNP_10857317__rs4950741C,T			.7809	1.00	-134.39	80.923	2.7581	0.09676	48
HippoVol	AFFX-SNP_220446__rs1553622A,G			.8306	.9673	-75.225	79.094	0.9046	0.3416	48
HippoVol	AFFX-SNP_8981529__rs6663947A,C			.5779	.9926	-1.631	61.551	0.0007	0.9789	48
HippoVol	AFFX-SNP_7236864__rs416768C,G			.9705	.9588	-246.36	180.484	1.8632	0.1723	48

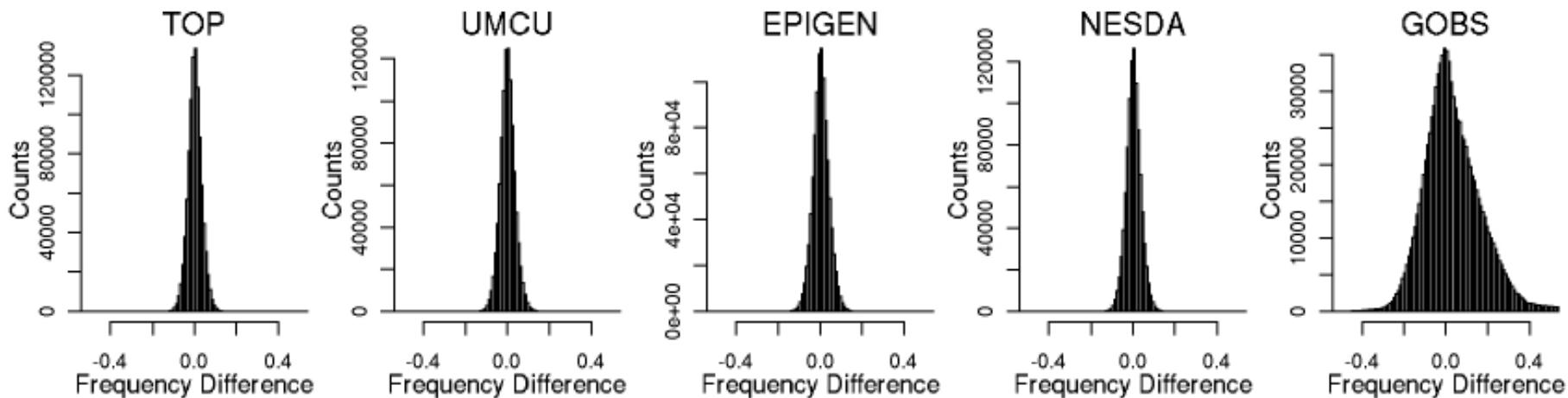
SNPs with prefix of AFFX have no white space between alleles and AFFX in mach2qtl (messes up parsing)

```
sed 's/A,/ A,/g; s/C,/ C,/g; s/T,/ T,/g; s/G,/ G,/g' chr3-mach2qtl.out > chr3-AFFXfix.out
```

Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 5: Quality Checking
- Step 6: Meta-analysis
- Conclusions

Frequency Differences Histograms Relative to CEU

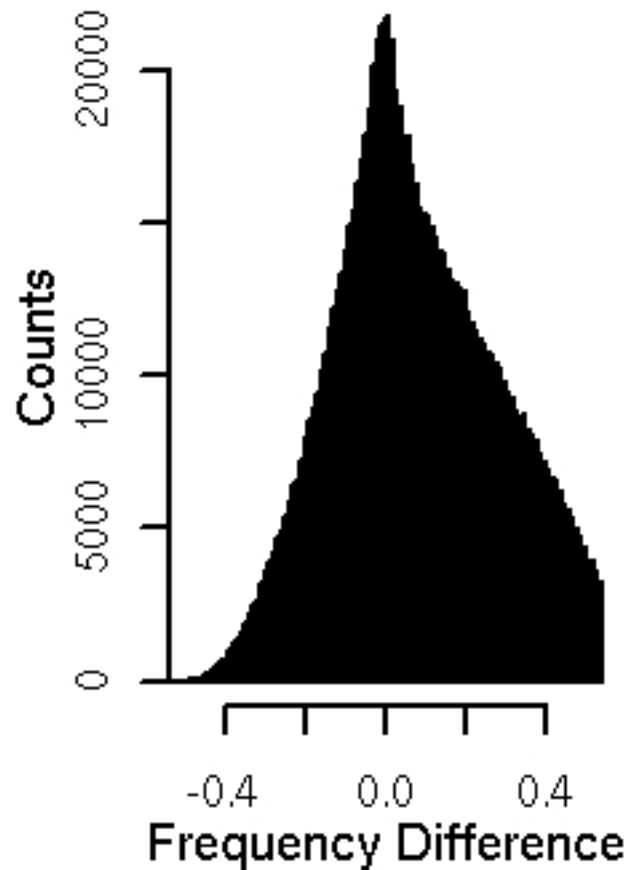


Histograms centered around zero with little variance imply that imputation and genotyping were done well (not different from the reference HapMap sample).

Also will tell you if frequency is assigned to the correct allele!

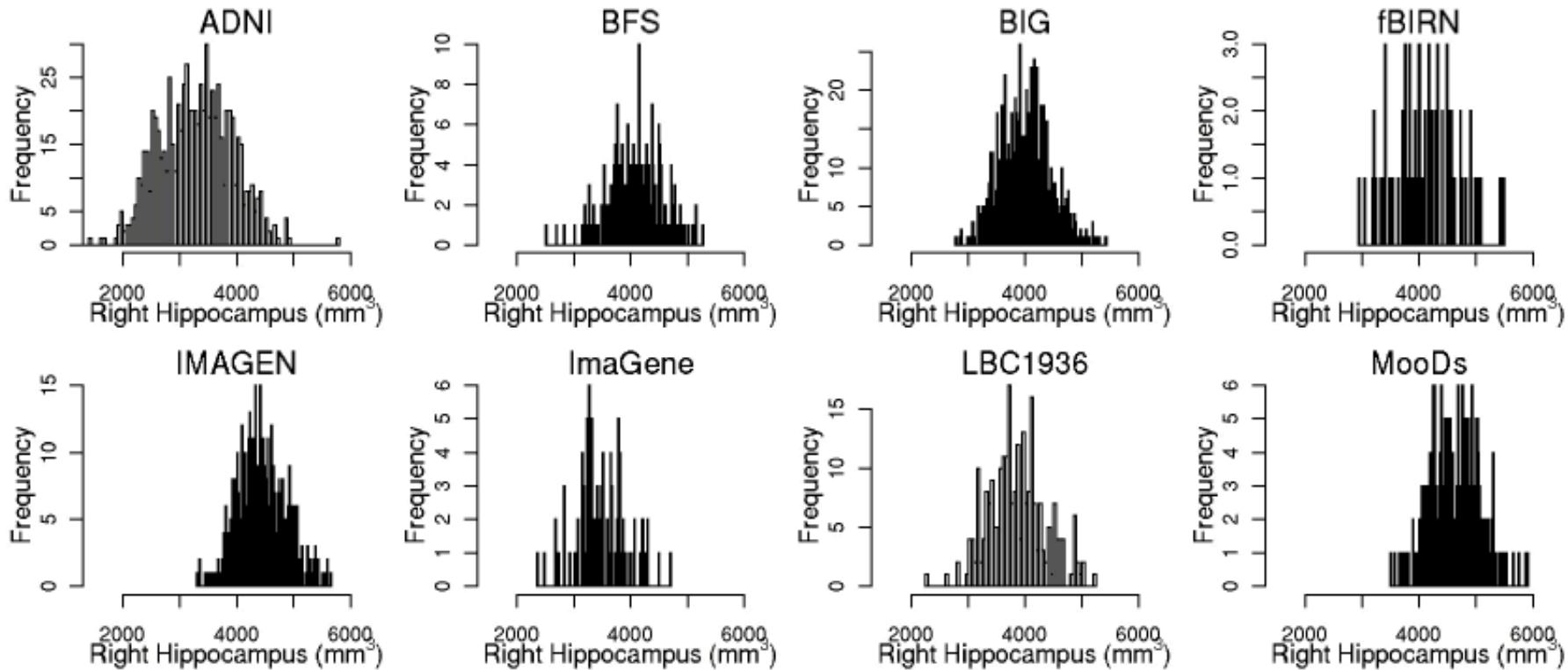
Here you can see that all samples are centered around zero which is evidence of good imputation and genotyping. Samples of different ethnicity (GOBS) have greater variance.

When Frequency Difference Plots go Wrong



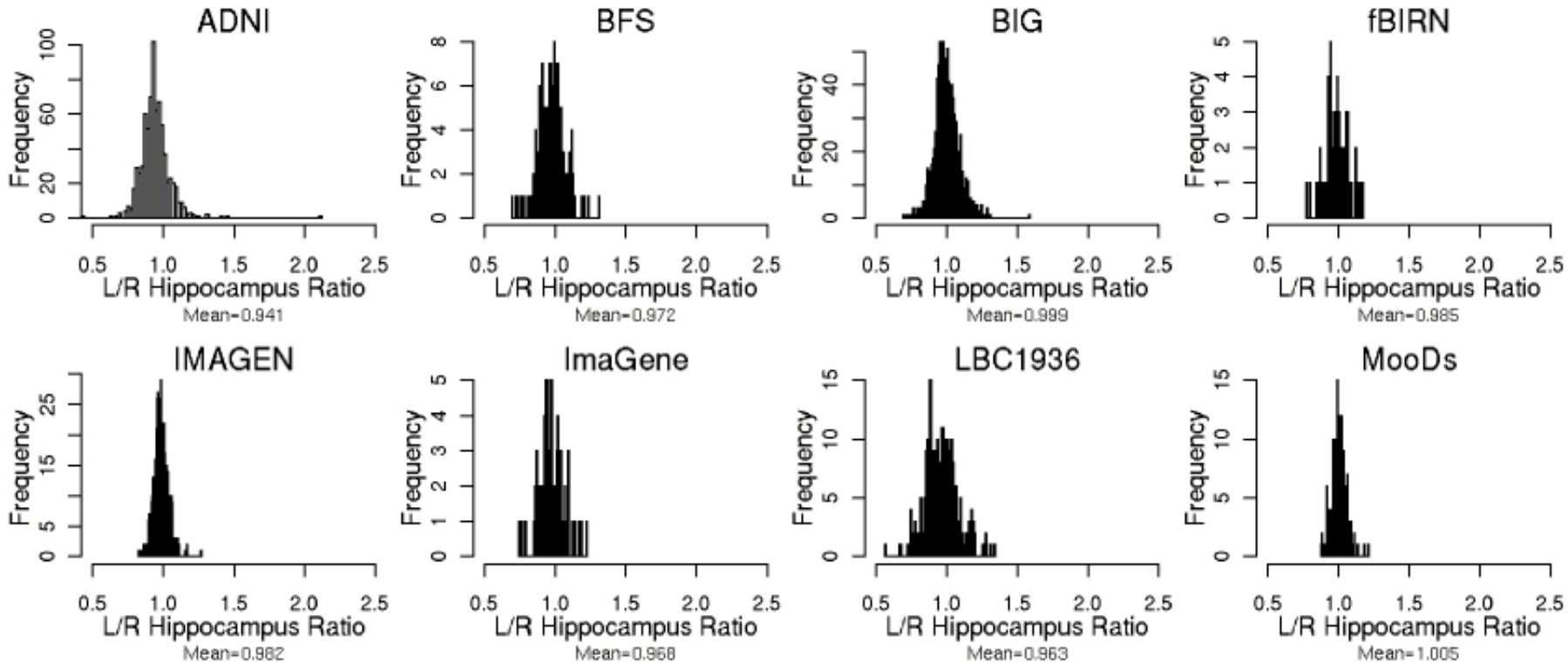
Generally indicative of poor imputation or poor genotyping

Right Hippocampus Phenotype



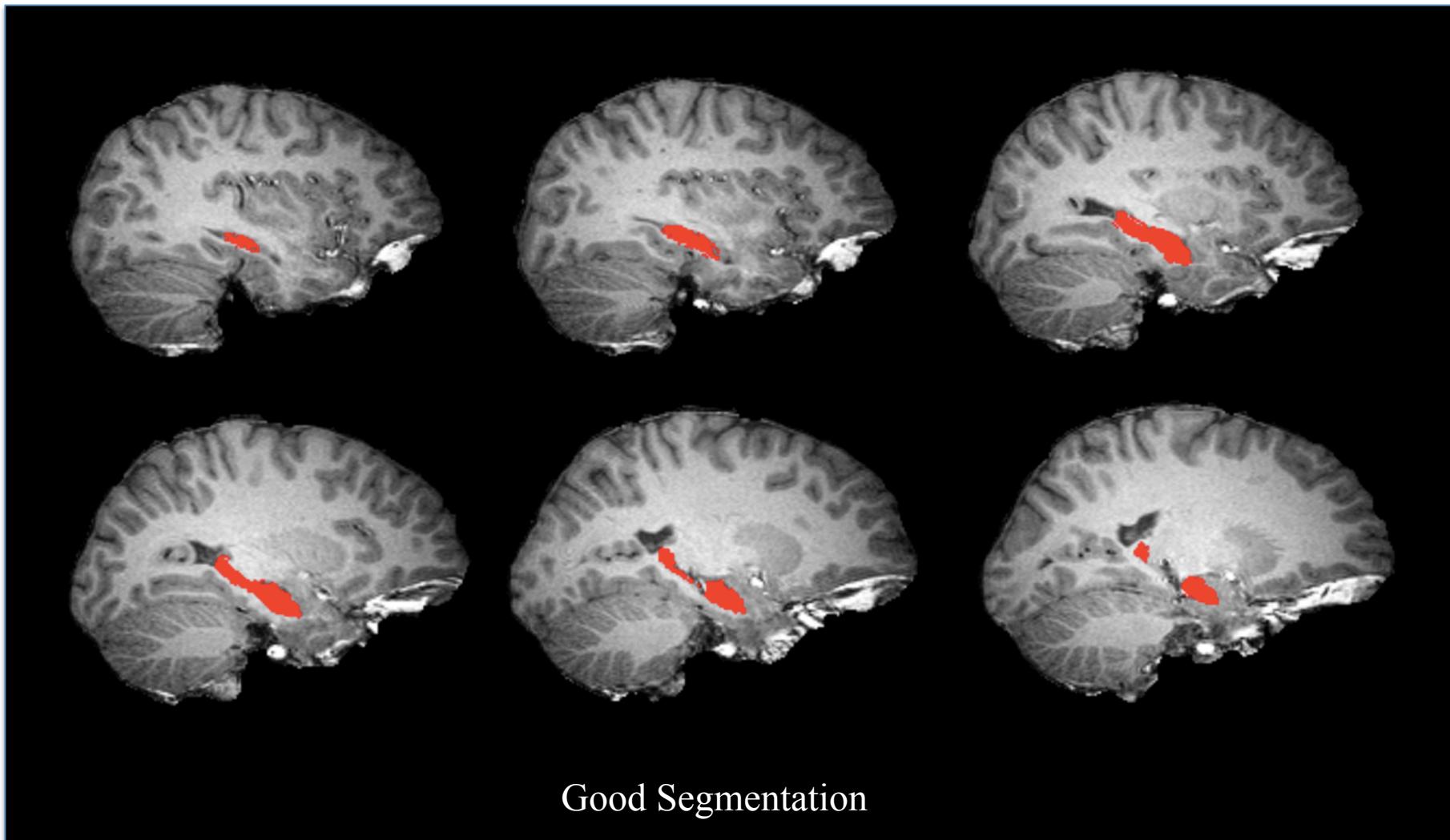
Histograms are a great way to visualize your data and a way to identify poorly segmented outliers.

L/R Hippocampus Ratio



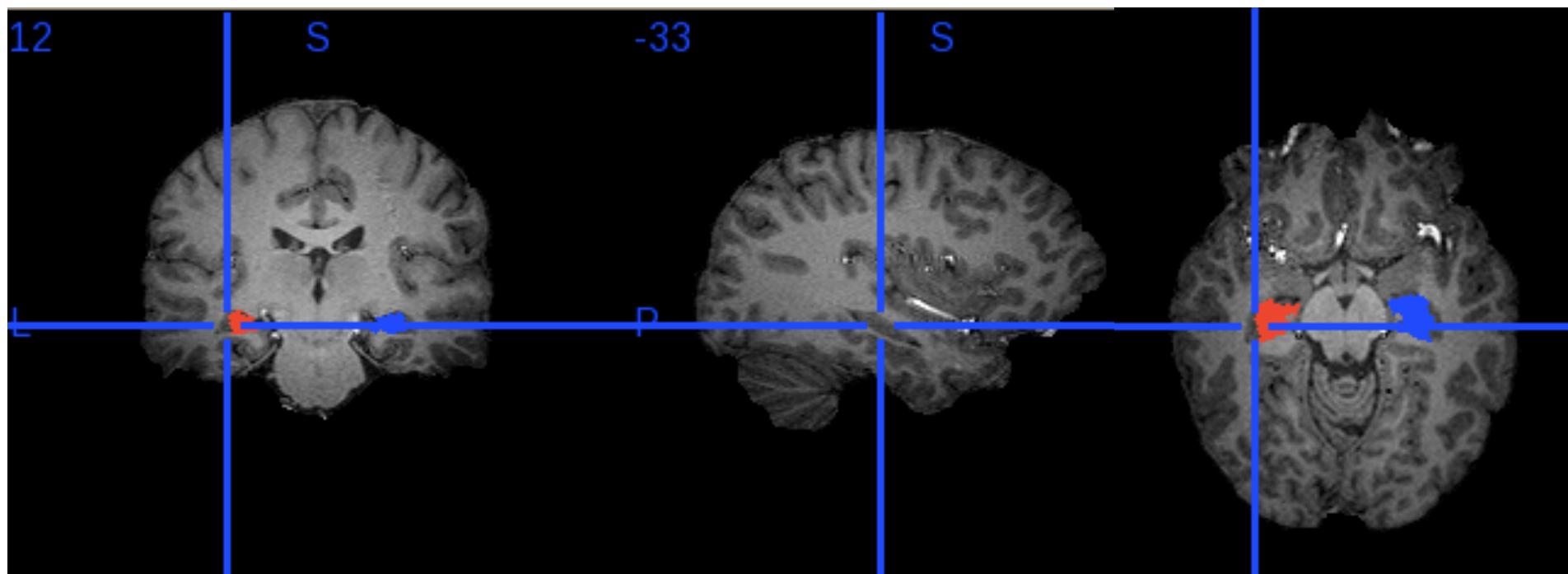
Neurologically normal adults (Weis et al., 1989; Watson et al., 1992; Jack et al., 2000) and children (Giedd et al., 1996; Pfluger et al., 1999; Utsunomiya et al., 1999) have significantly larger right hippocampi. Therefore the mean L/R ratio should generally be less than 1.

Slice by Slice QC of Hippocampal Segmentations in Sagittal View

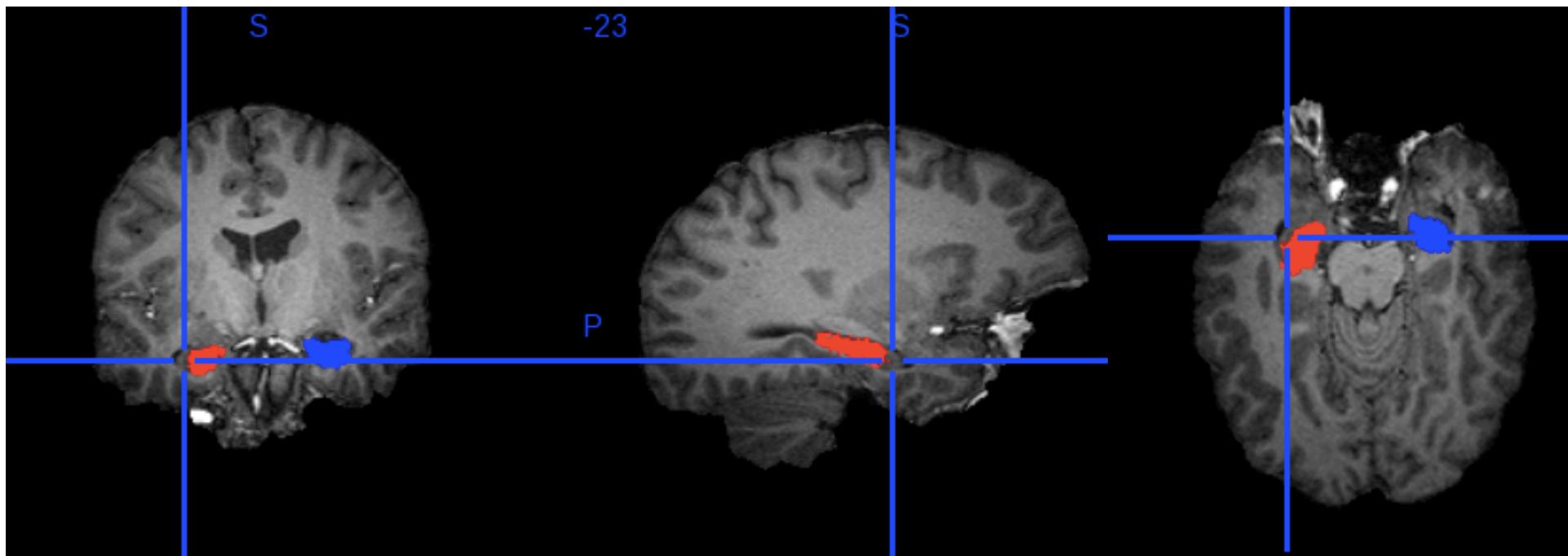


Good Segmentation

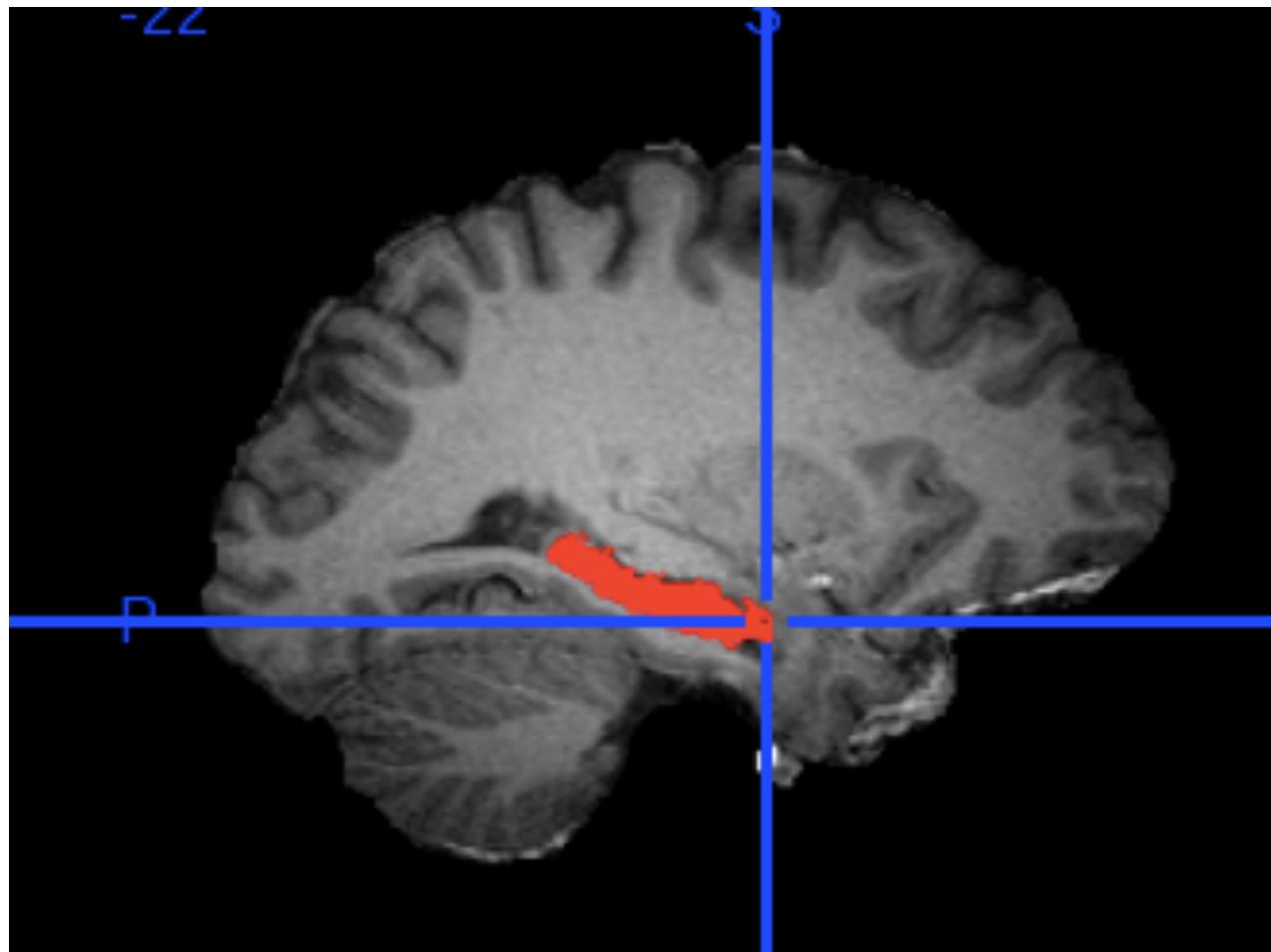
Common Errors: Missing Large Lateral Sections of Hippocampus



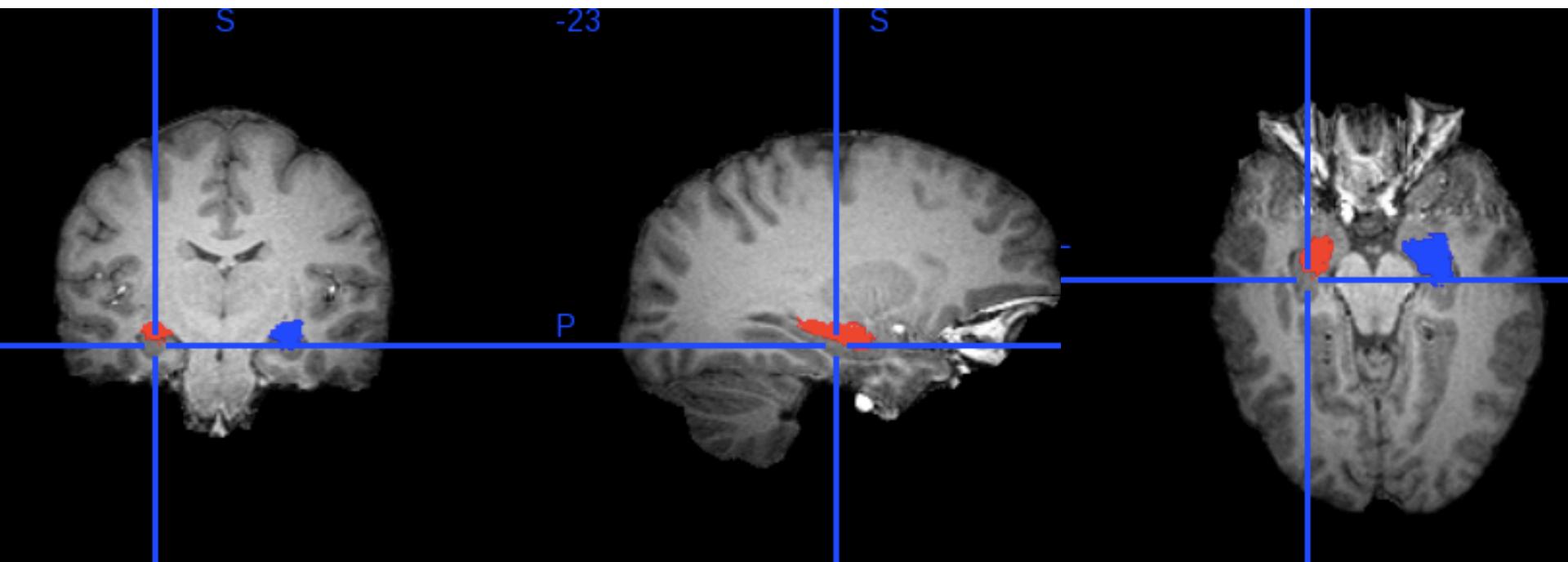
Common Errors: Missing Anterior Portions of Hippocampus



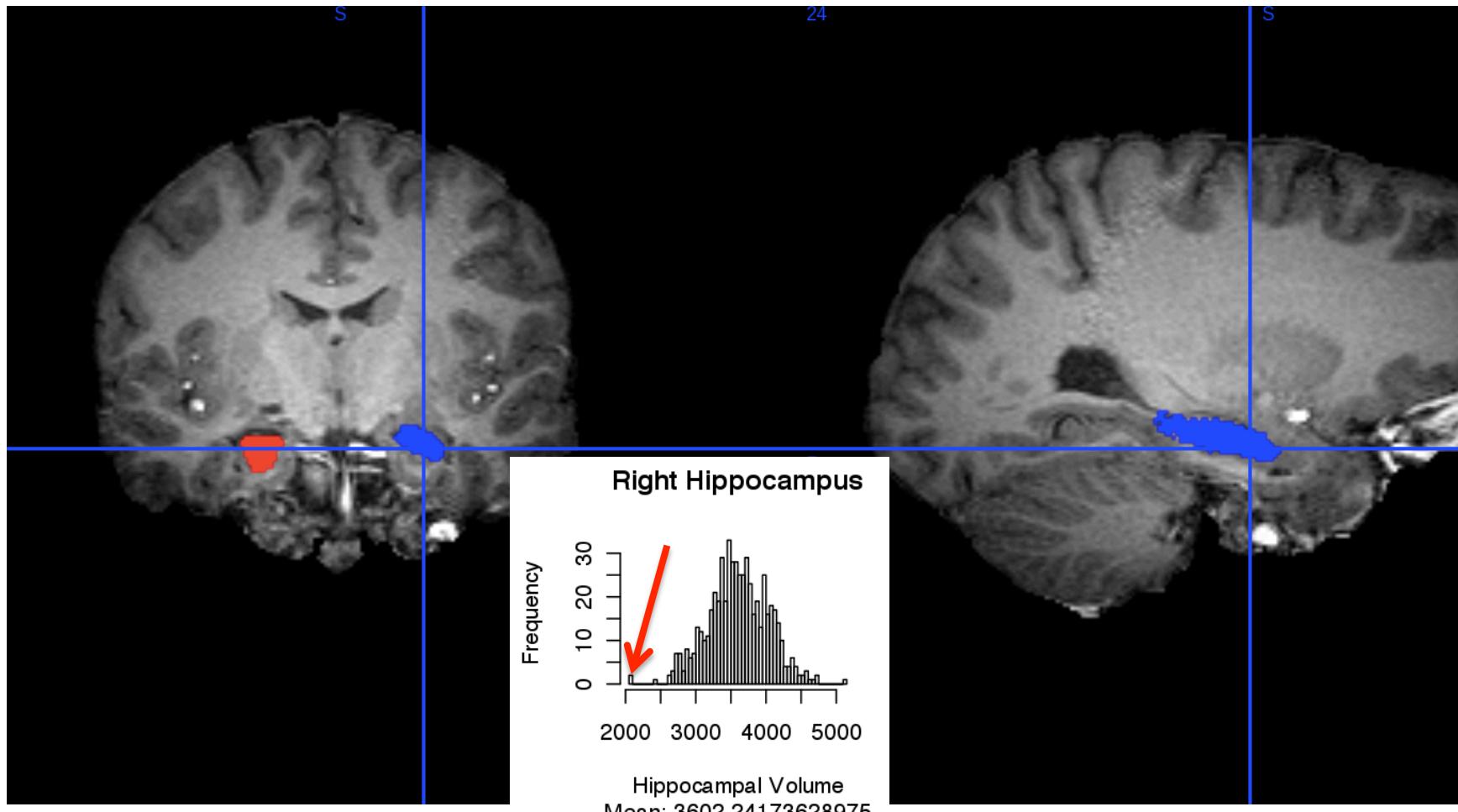
Common Errors: Extension into amygdala



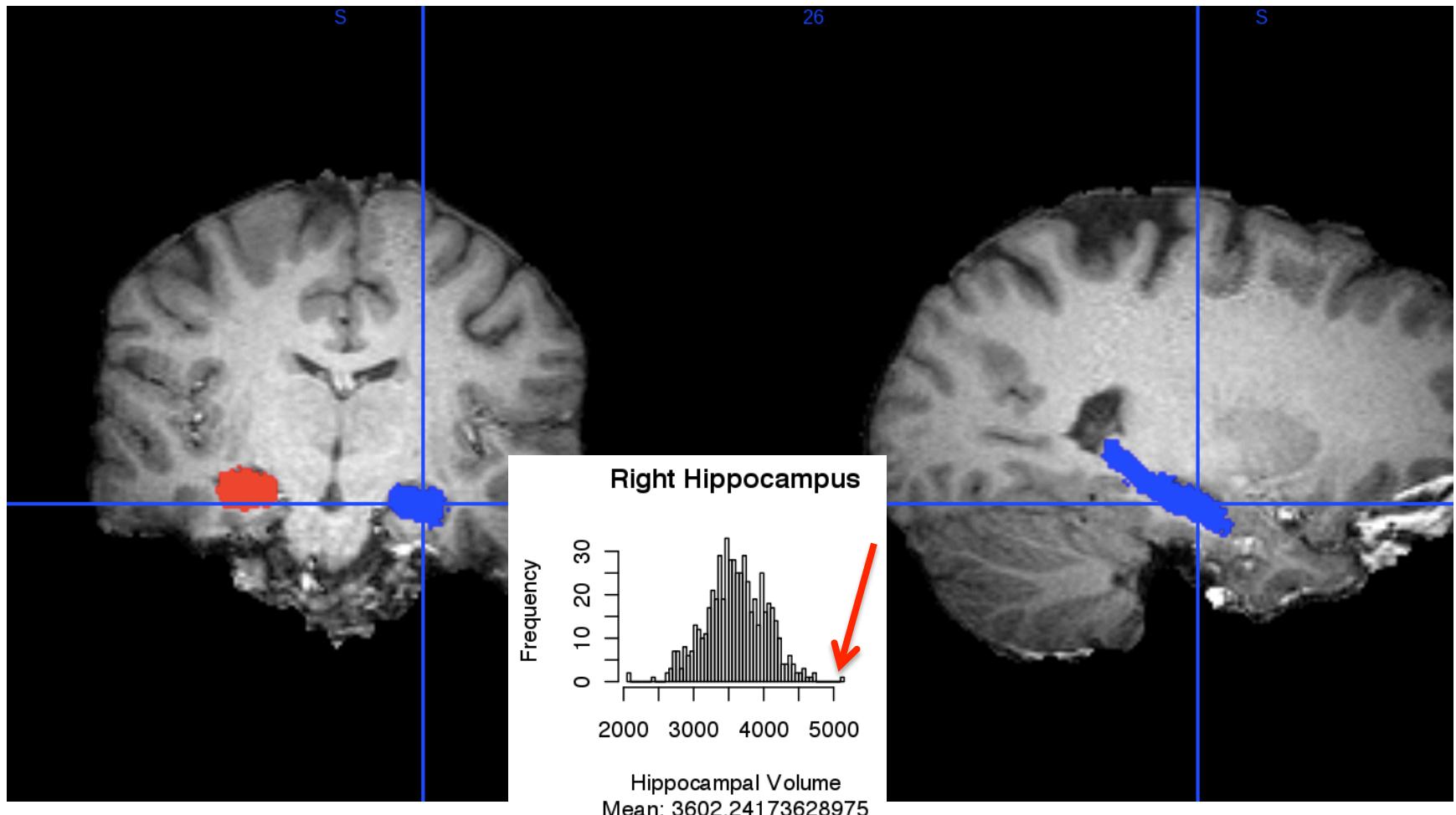
Less Common Error: Segmentation Completely Off



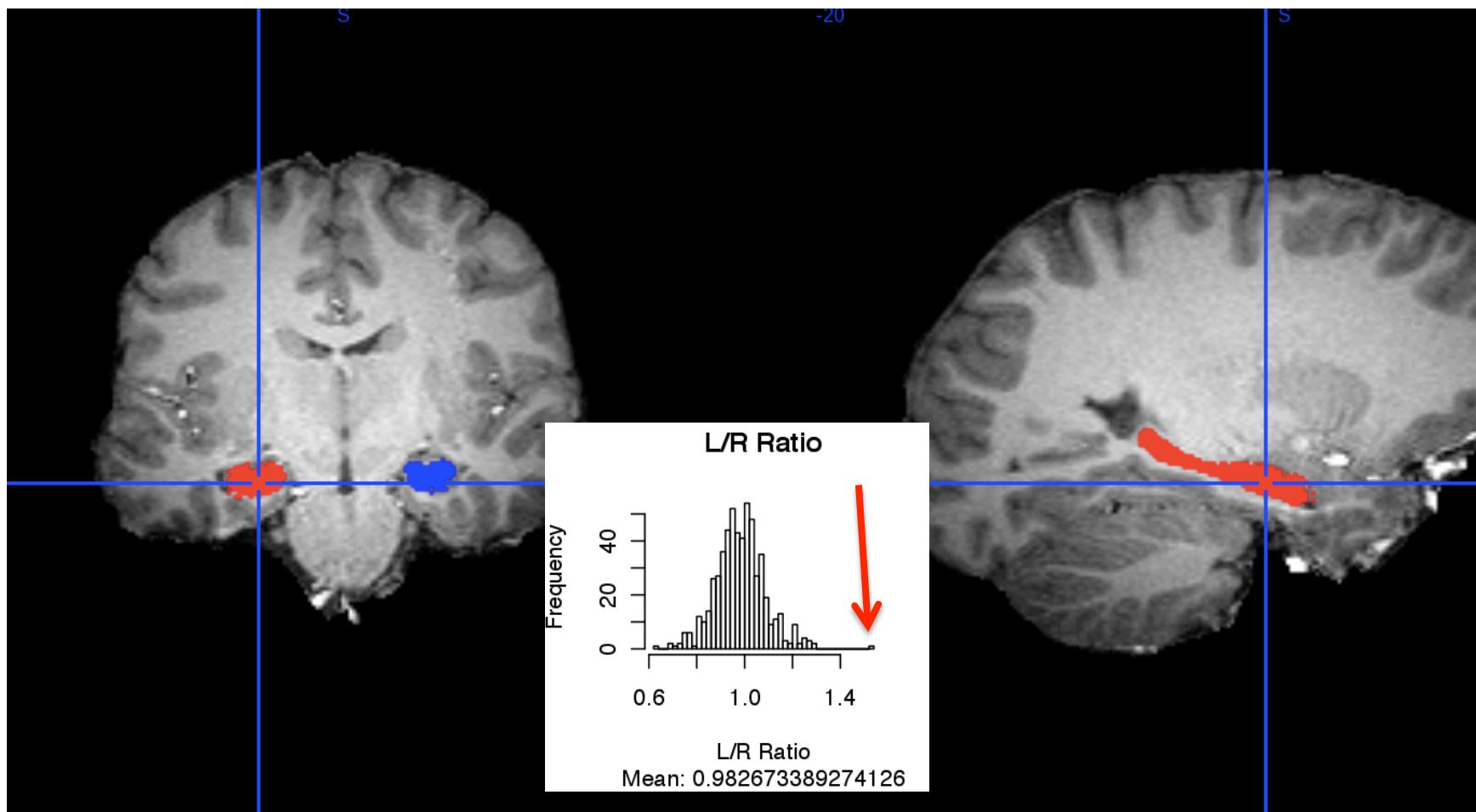
Subjects with Well Delineated but Small Hippocampi



Subjects with Good but Big Hippocampi

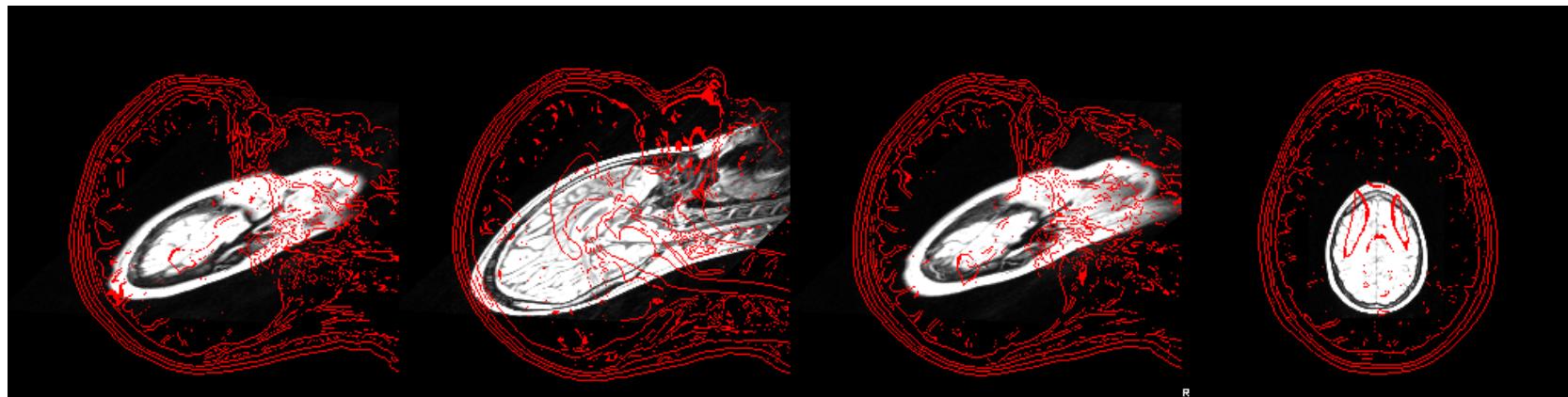


Subject with good segmentations large L/R Ratio



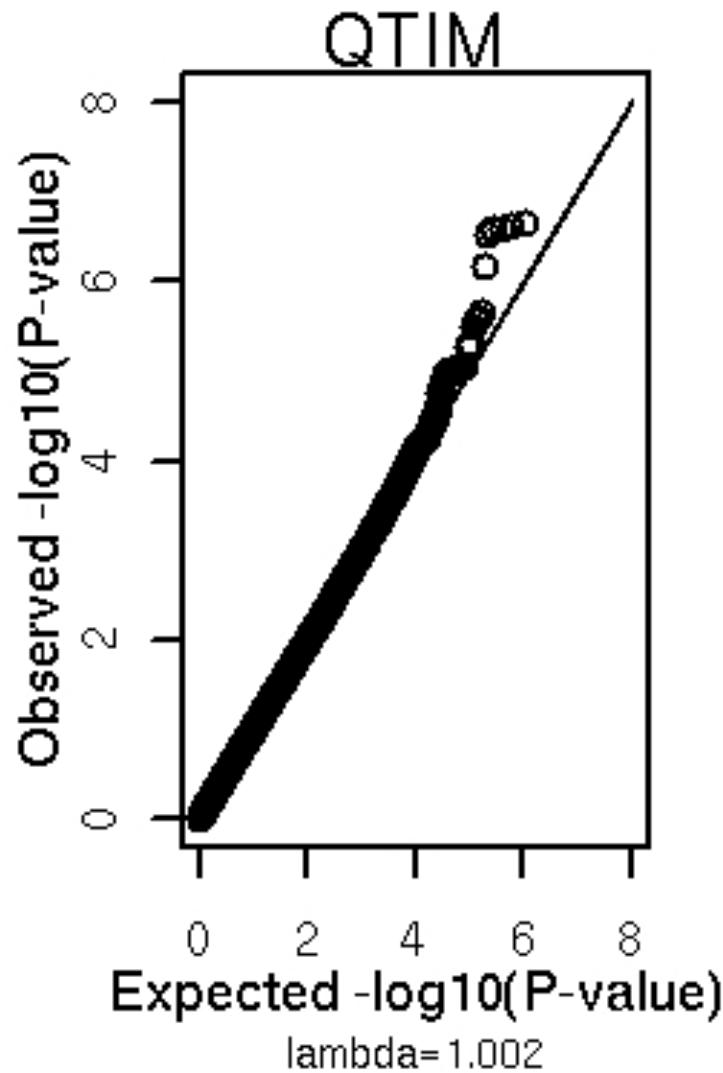
$$L/R \text{ Ratio} = 1.525996$$

eTIV errors from poor registration

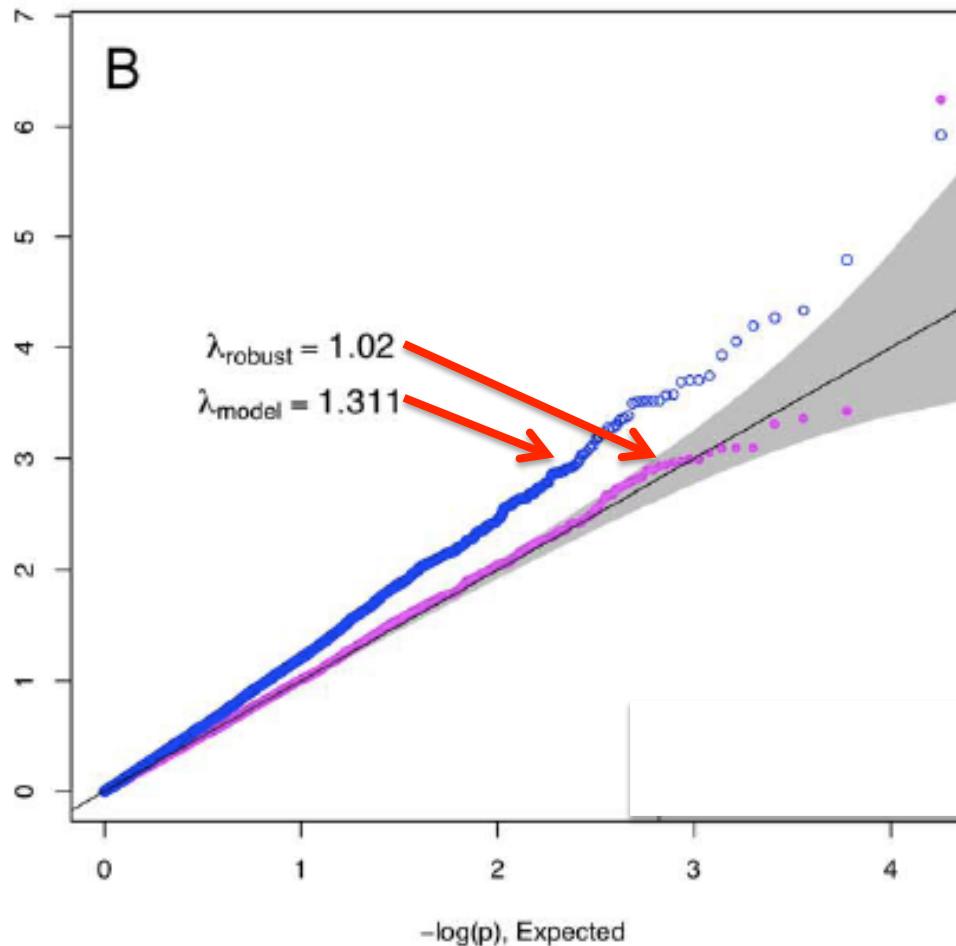


Poor registration gives poor estimates of the scaling used to determine estimated total intracranial volume

Individual Site QQ Plots

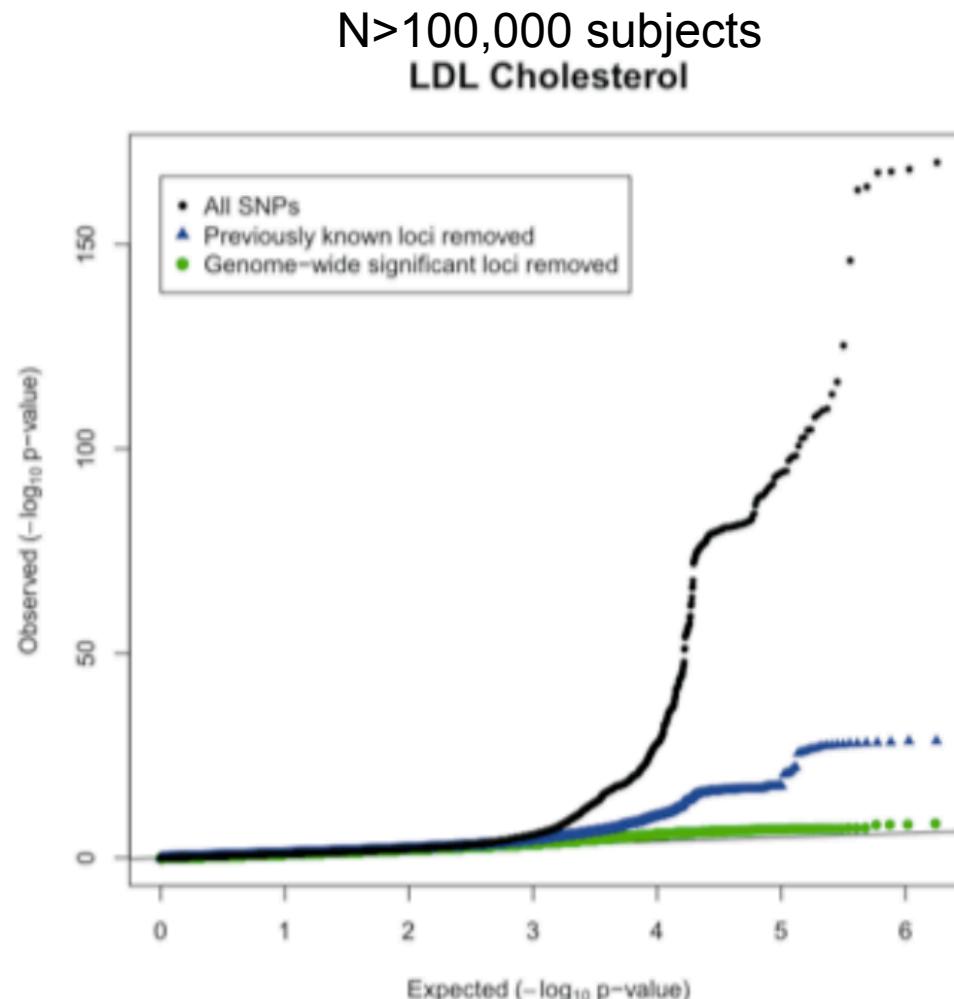


When QQ Plots Go Wrong



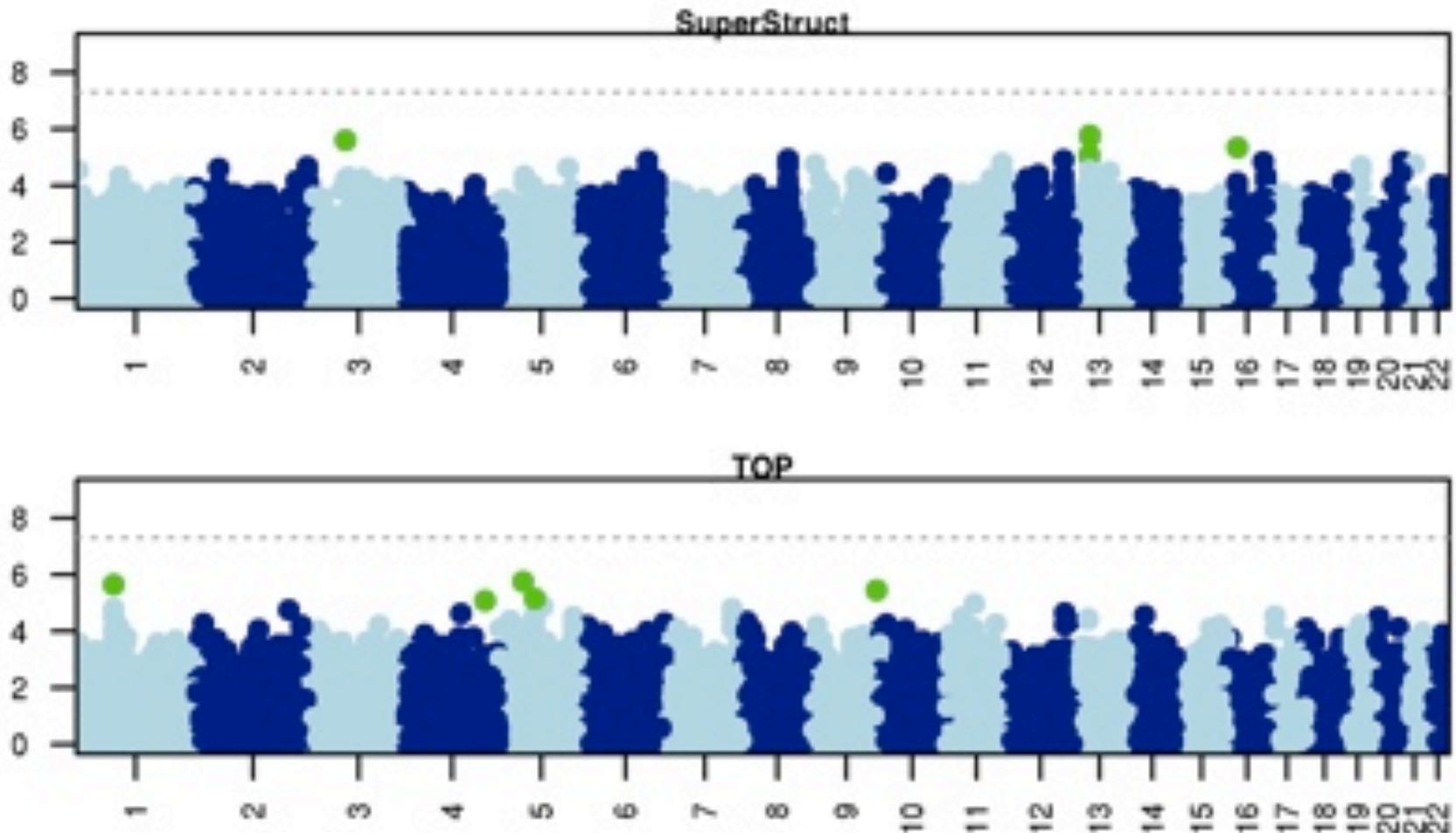
Can show evidence of unaccounted for population stratification, cryptic relatedness, or just that your data does not follow expected distributions

When QQ Plots Go Really Well

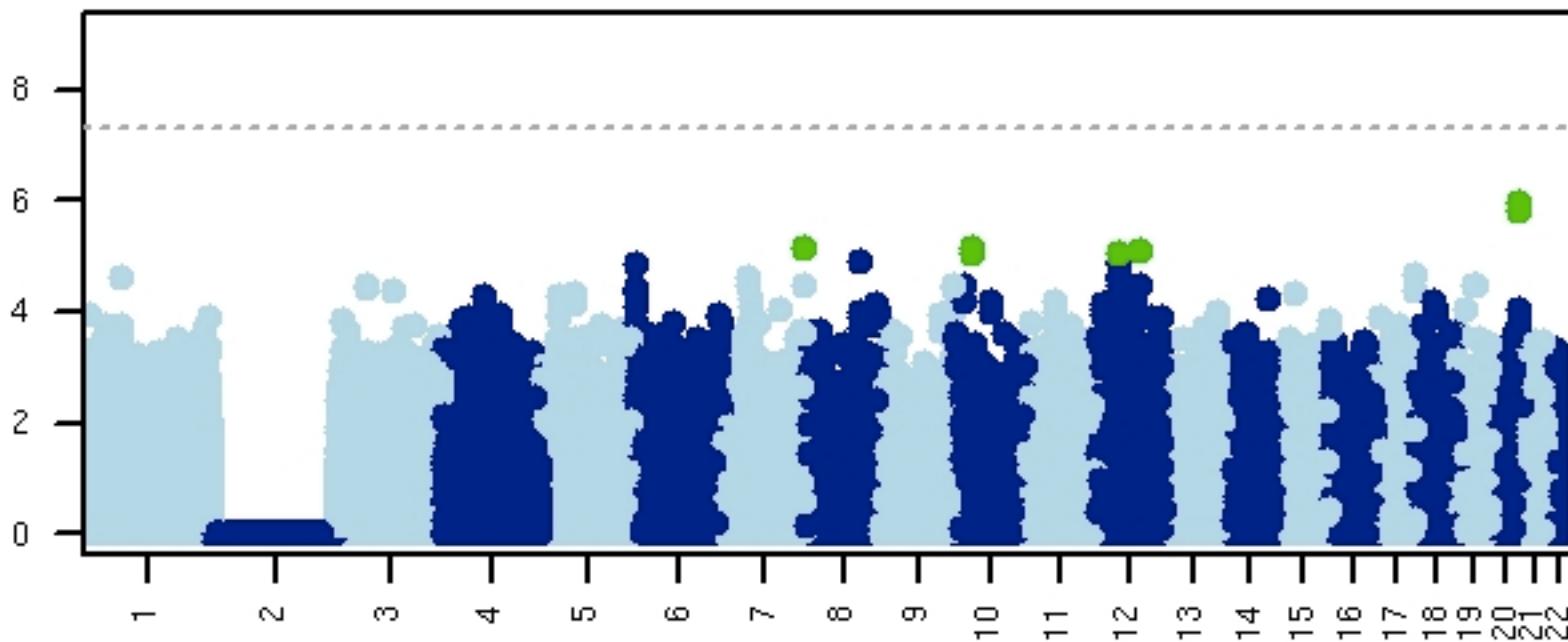


The observed distribution only deviates from the expected at low P-values. Would not expect something like this without huge effect sizes or huge sample sizes.

Individual Site Manhattans

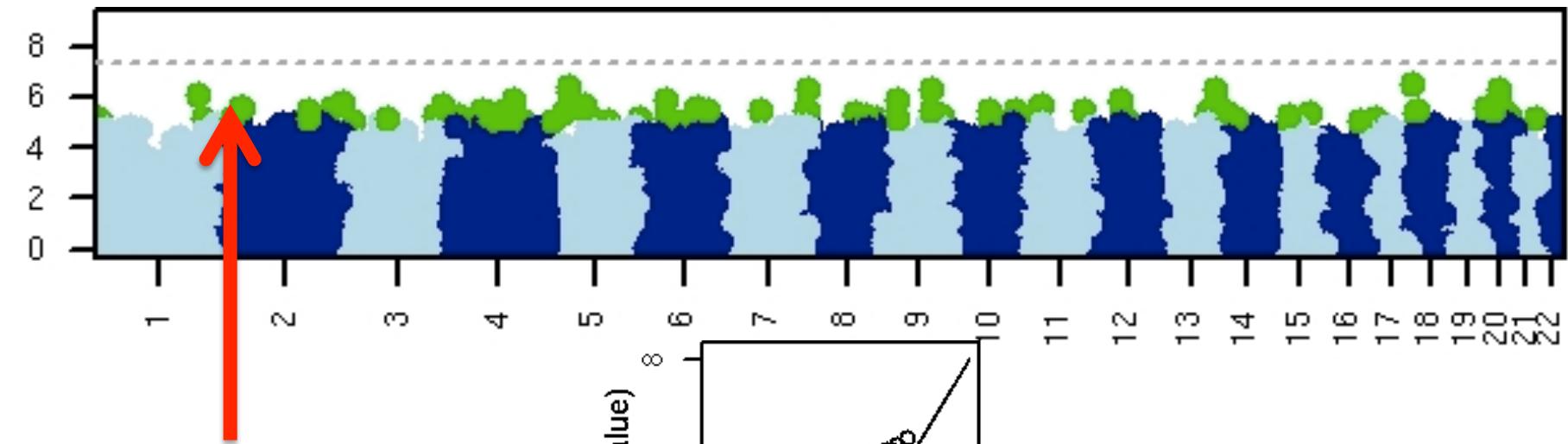


When Manhattan Plots Go Wrong

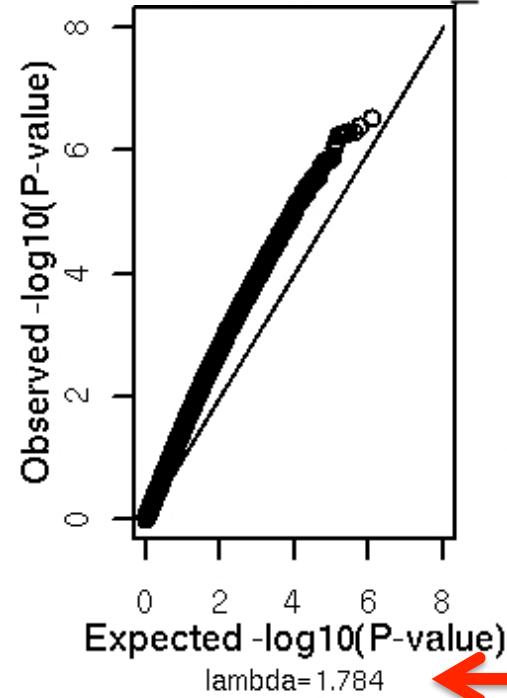


Can be evidence that imputation failed on one chromosome or that somebody just typed something incorrectly when running association

Population Stratification

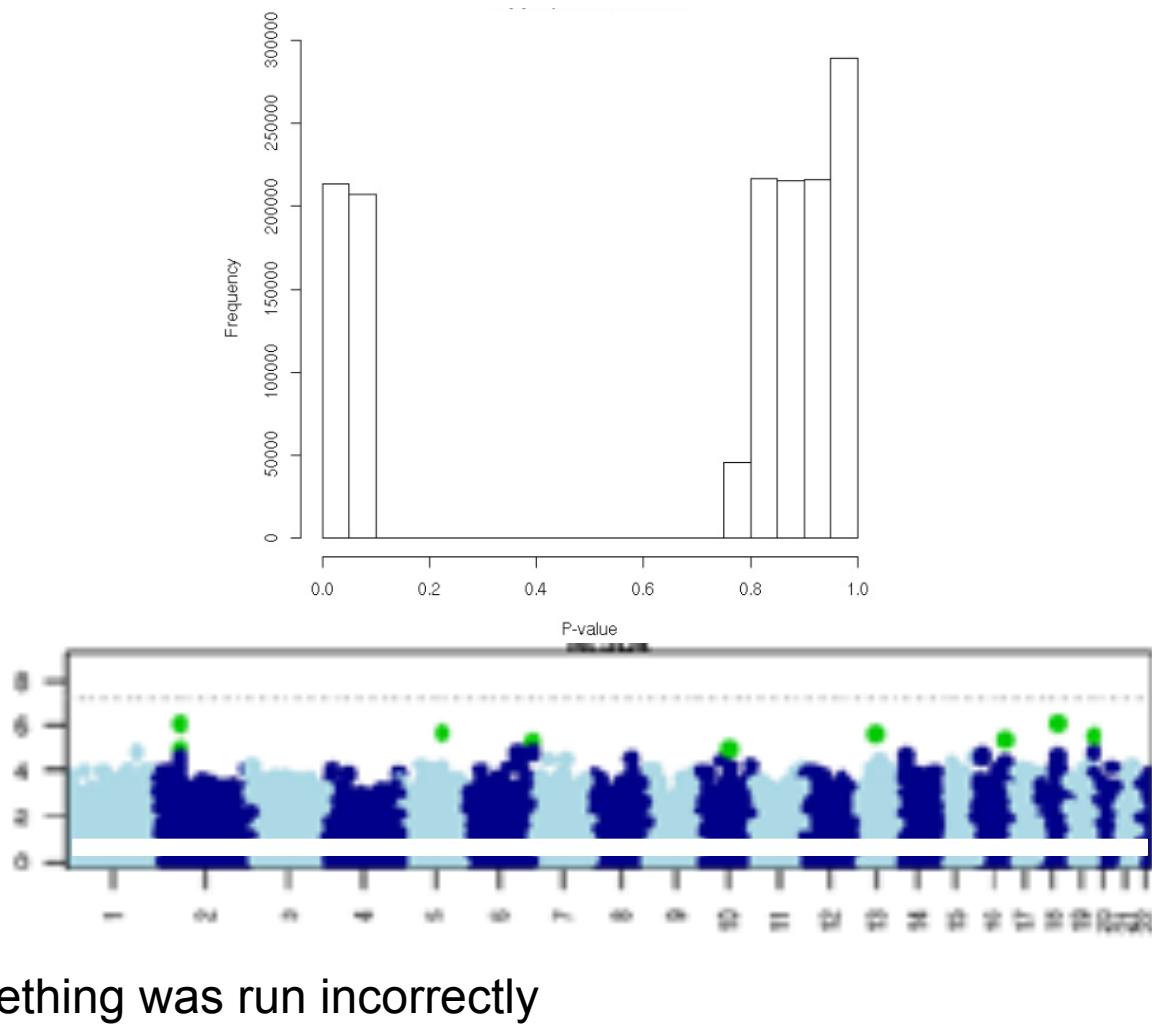
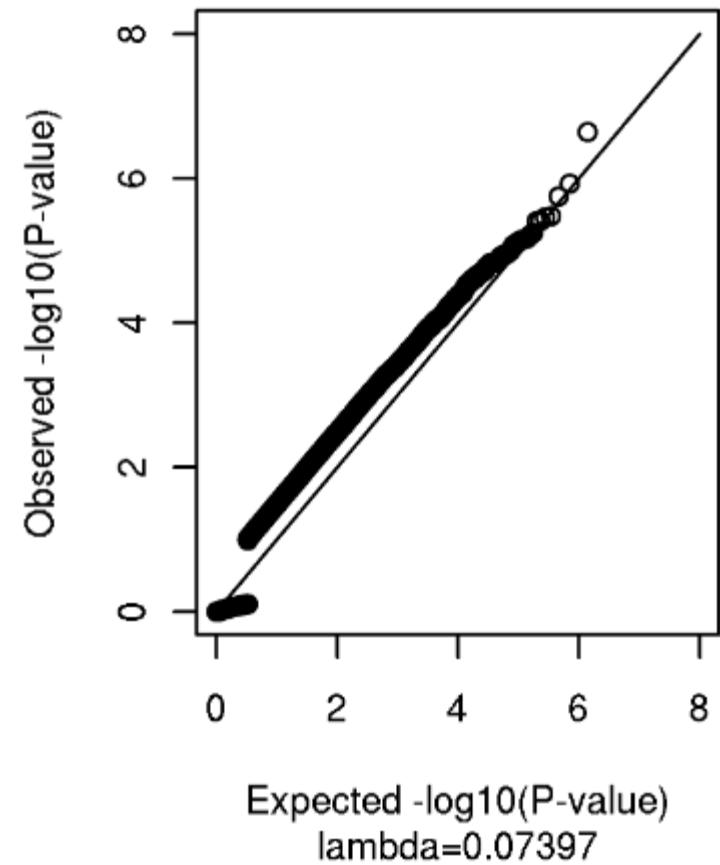


Instead of a few hits,
most everything is
 $P < 1e-5$



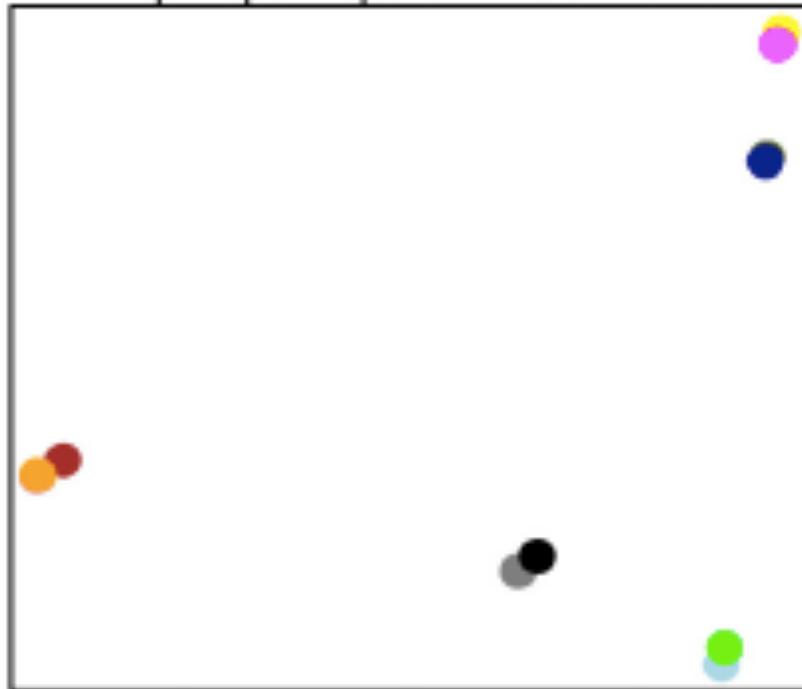
Lambda should be
near 1

Other Interesting Data

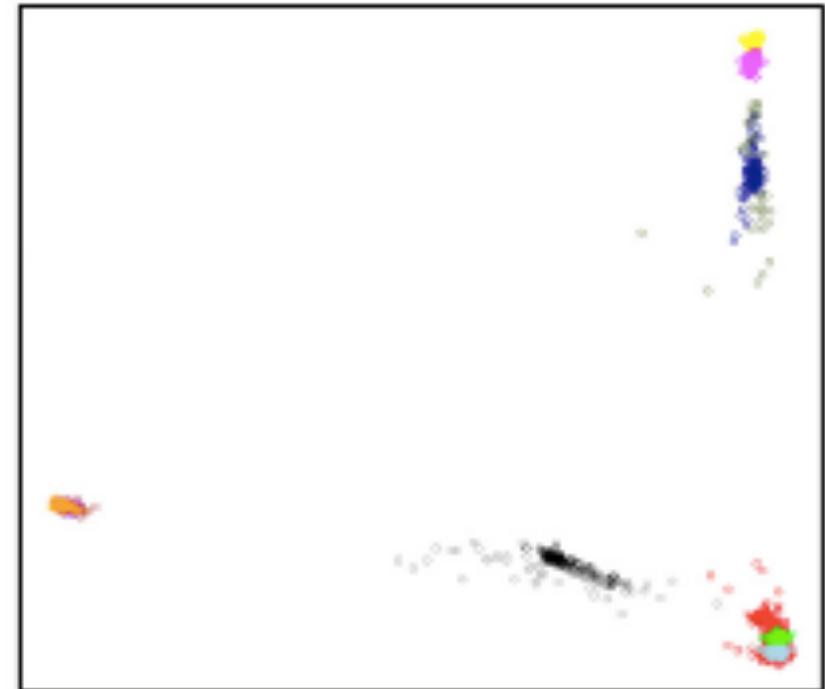


MDS Plots

HapMap3 Population Centroids



ADNI



Can show you which HapMap population best describes your sample

Good for finding and removing ancestry outliers

Imaging Genomics Meta-Analysis

- The need for large scale meta-analysis
- Step 1: Set up the Consortium
- Step 2: Receiving the data
- Step 3: Filtering data
- Step 4: Processing Data
- Step 5: Quality Checking
- Step 6: Meta-analysis
- Conclusions

Fixed Effects MA Description

Fixed effects assume that the genetic effects are the same across the combined investigations and all differences are due to chance

$$z_{\text{meta}} = \sum_i z_i \times w_i$$

$$w_i = \sqrt{\frac{N_i}{N_{\text{total}}}}$$

P-value based meta-analysis
(Weighted sum of Z-scores)

the units of the beta coefficients and standard errors **need not be the same** across studies

Fixed Effects MA Description

$$\langle \beta \rangle = \frac{\sum_i [\beta_i / (\text{SE}_i)^2]}{\sum_i [1 / (\text{SE}_i)^2]}.$$

$$\langle \text{SE} \rangle = \sqrt{\frac{1}{\sum_i [1 / (\text{SE}_i)^2]}}$$

$$z_{\text{meta}} = \frac{\langle \beta \rangle}{\langle \text{SE} \rangle}$$

pooled inverse variance weighted

the units of the beta coefficients and standard errors **must be** the same across studies

How to run METAL: Stage 1

http://genome.sph.umich.edu/wiki/METAL_Documentation

SNP	NON_EFFECT_ALLELE	EFFECT_ALLELE	BETA	FREQ	N	P_VAL	SE
rs9586302	C	A	24.297	0.0501	550	0.5288	38.575
rs354417	T	G	8.233	0.815	550	0.7102	22.154
rs1927207	T	C	10.124	0.0827	550	0.7388	30.361
rs9582391	A	C	-22.281	0.2143	550	0.2723	20.296
rs507529	A	G	-26.772	0.4256	550	0.1227	17.343
rs7490444	T	C	90.795	0.1035	550	0.005564	32.75
rs12584241	T	C	-54.848	0.0876	550	0.06505	29.729

/home/enigma/site1.tbl

Make sure all data are in the same format as white space delimited text files. Pay special attention to making sure effect_allele is correct!

How to run METAL: Stage 2

```
MARKERLABEL      SNP
EFFECTLABEL     BETA
WEIGHTLABEL      N
PVALUELABEL     P_VAL
MINWEIGHT        1000
SCHEME           SAMPLESIZE
AVERAGEFREQ      ON
MINMAXFREQ       ON
FREQLABEL        FREQ
GENOMICCONTROL   ON
OUTFILE          MetaOutput .tbl

ALLELELABELS     EFFECT_ALLELE NON_EFFECT_ALLELE

PROCESS /home/enigma/site1.tbl #Site 1
PROCESS /home/enigma/site2.tbl #Site 2
PROCESS /home/enigma/site3.tbl #Site 3
PROCESS /home/enigma/site4.tbl #Site 4

ANALYZE          HETEROGENEITY
QUIT
```

Minimum sample size needed to calculate meta-analysis

/home/enigma/allsites.metal

How to run METAL: Stage 2

```
MARKERLABEL      SNP
EFFECTLABEL     BETA
WEIGHTLABEL      N
PVALUELABEL     P_VAL
MINWEIGHT       1000
SCHEME          SAMPLESIZE ←
AVERAGEFREQ     ON
MINMAXFREQ      ON
FREQLABEL       FREQ
GENOMICCONTROL  ON
OUTFILE         MetaOutput .tbl

ALLELELABELS    EFFECT_ALLELE NON_EFFECT_ALLELE

PROCESS /home/enigma/site1.tbl #Site 1
PROCESS /home/enigma/site2.tbl #Site 2
PROCESS /home/enigma/site3.tbl #Site 3
PROCESS /home/enigma/site4.tbl #Site 4

ANALYZE          HETEROGENEITY
QUIT
```

Can specify P-value based meta-analysis or inverse standard error weighted meta-analysis here

/home/enigma/allsites.metal

How to run METAL: Stage 2

```
MARKERLABEL      SNP
EFFECTLABEL     BETA
WEIGHTLABEL      N
PVALUENAME       P_VAL
MINWEIGHT        1000
SCHEME           SAMPLESIZE
AVERAGEFREQ      ON
MINMAXFREQ       ON
FREQLABEL        FREQ
GENOMICCONTROL   ON ← Adjust the test statistics at the individual
OUTFILE          MetaOutput .tbl site level using the lambda factor
ALLELELABELS     EFFECT_ALLELE NON_EFFECT_ALLELE
PROCESS /home/enigma/site1.tbl #Site 1
PROCESS /home/enigma/site2.tbl #Site 2
PROCESS /home/enigma/site3.tbl #Site 3
PROCESS /home/enigma/site4.tbl #Site 4
ANALYZE          HETEROGENEITY
QUIT
/home/enigma/allsites.metal
```

How to run METAL: Stage 2

```
MARKERLABEL      SNP
EFFECTLABEL     BETA
WEIGHTLABEL      N
PVALUELABEL     P_VAL
MINWEIGHT       1000
SCHEME          SAMPLESIZE
AVERAGEFREQ     ON
MINMAXFREQ      ON
FREQLABEL       FREQ
GENOMICCONTROL  ON
OUTFILE         MetaOutput .tbl

ALLELELABELS    EFFECT_ALLELE NON_EFFECT_ALLELE

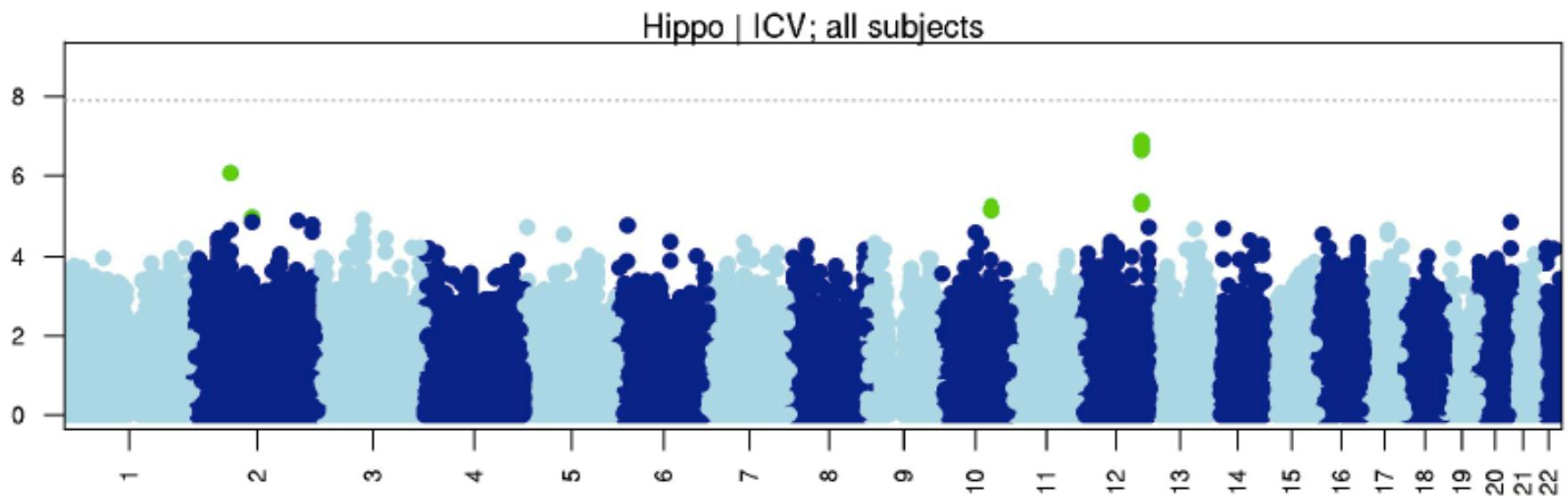
PROCESS /home/enigma/site1.tbl #Site 1
PROCESS /home/enigma/site2.tbl #Site 2
PROCESS /home/enigma/site3.tbl #Site 3
PROCESS /home/enigma/site4.tbl #Site 4

ANALYZE          HETEROGENEITY
QUIT
```

```
/home/enigma/allsites.metal
metal /home/enigma/allsites.metal > /home/enigma/
allsites.metal.log
```

← Specify the files to process

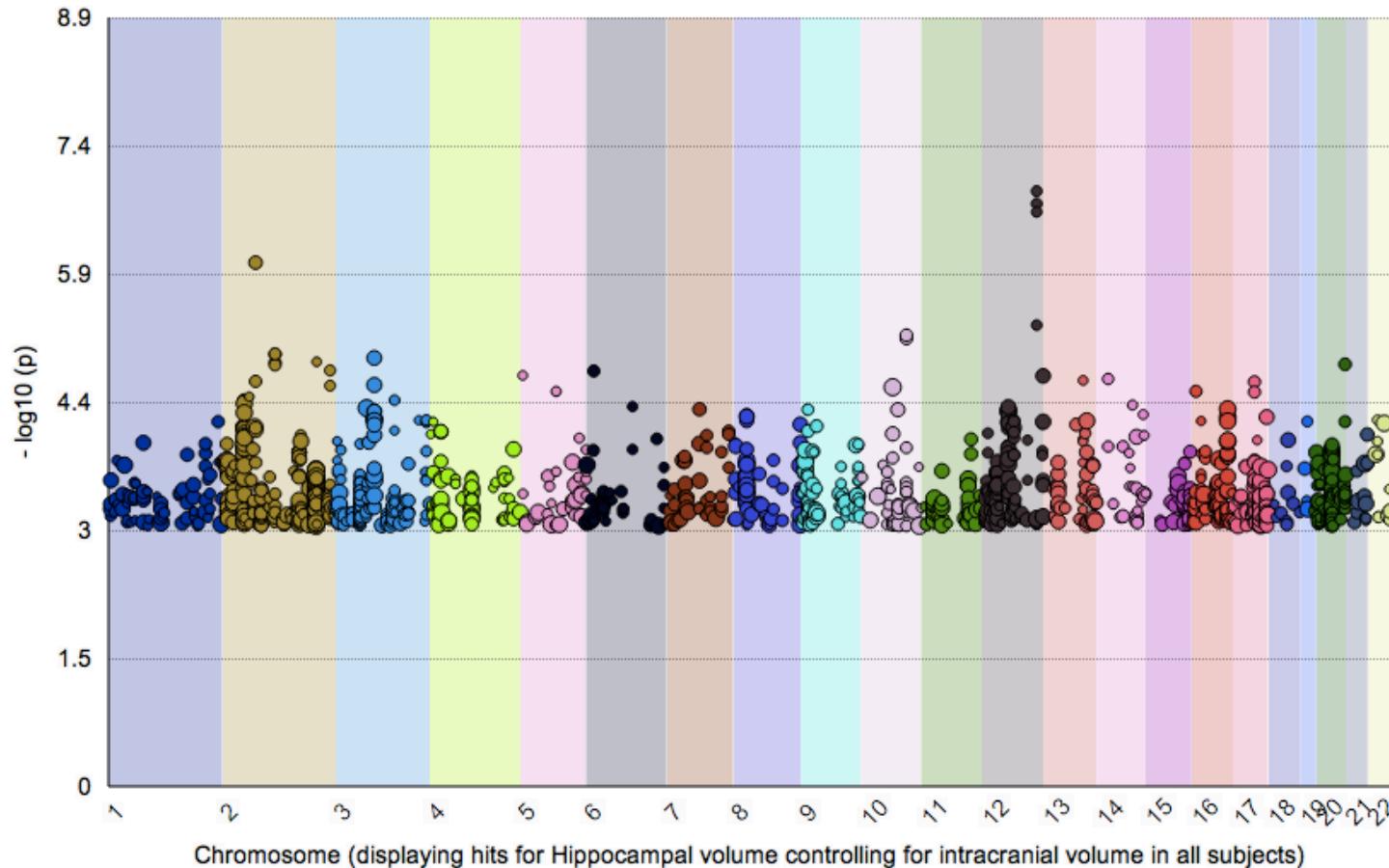
The Final Product



Each SNP P-value is the combined evidence from all contributing studies

Visualization of Results

Manhattan plot showing all SNPs with $p < 0.001$



For More Information

Human Molecular Genetics, 2008, Vol. 17, Review Issue 2 RI22–RI28
doi:10.1093/hmg/ddn288

Practical aspects of imputation-driven meta-analysis of genome-wide association studies

Paul I.W. de Bakker^{1,2,*}, Manuel A.R. Ferreira^{2,3,†}, Xiaoming Jia⁴, Benjamin M. Neale^{2,3}, Soumya Raychaudhuri^{2,3,5} and Benjamin F. Voight^{2,3}

LETTERS

nature
genetics

Identification of common variants associated with human hippocampal and intracranial volumes

Conclusions

- Consortia are important because **effect sizes are small**
- However, the more data you combine with different groups the more **quality checking and filtering** matters
- Clear demonstration of **Murphy's Law**
- Did I mention quality checking?

Acknowledgements

Jason L Stein^{1,127}, Sarah E Medland^{2-4,127}, Alejandro Arias Vasquez^{5-7,127}, Derrek P Hibar^{1,127}, Rudy E Senstad¹, Anderson M Winkler^{8,9}, Roberto Toro¹⁰⁻¹², Katja Appel^{13,14}, Richard Bartecik¹⁵, Ørjan Bergmann¹⁶, Manon Bernard¹⁷, Andrew A Brown^{16,18}, Dara M Cannon¹⁹, M Mallar Chakravarty²¹, Andrea Christoforou^{22,23}, Martin Domin²⁴, Oliver Grimm²⁵, Marisa Hollinshead^{26,27}, Avram J Holmes²⁶, Georg Homuth²⁸, Jouke-Jan Hottenga²⁹, Camilla Langan²⁰, Lorna M Lopez^{30,31}, Narelle K Hansell², Kristy S Hwang^{1,32}, Sunguen Kim^{33,34}, Gonzalo Laje³⁵, Phil H Lee^{36,37}, Xinmin Liu^{35,38}, Eva Loth³⁹, Anbarasu LourduSamy³⁹, Morten Mattingdal^{16,40}, Sebastian Mohnke⁴¹, Susana Muñoz Maniega^{30,42,43}, Kwangsik Nho^{33,44}, Allison C Nugent⁴⁵, Carol O'Brien^{46,47}, Martina Papmeyer⁴⁸, Benno Pütz⁴⁹, Adaikalavan Ramasamy⁵⁰, Jerod Rasmussen⁵¹, Mark Rijpkema^{7,52}, Shannon L Risacher³³, J Cooper Roddey⁵³, Emma J Rose^{46,47}, Mina Ryten⁵⁴, Li Shen^{33,34}, Emma Sprooten⁴⁸, Eric Strengman^{55,56}, Alexander Teumer²⁸, Daniah Trabzuni^{54,57}, Jessica Turner⁵⁸, Kristel van Eijk^{55,56}, Theo G M van Erp⁵¹, Marie-Jose van Tol⁵⁹⁻⁶¹, Katharina Wittfeld¹³, Christiane Wolf⁴⁹, Saskia Woudstra⁶², Andre Aleman⁶¹, Saud Alhusaini⁶³, Laura Almasy⁶⁴, Elisabeth B Binder⁴⁵, David G Brohawn³⁶, Rita M Cantor⁶⁵, Melanie A Carless⁶⁴, Aiden Corvin^{46,47}, Michael Czisch⁴⁹, Joanne E Curran⁶⁴, Gail Davies³¹, Marcio A A de Almeida⁶⁴, Norman Delanty^{63,66}, Chantal Depoort⁶⁷, Ravi Duggirala⁶⁴, Thomas D Dyer⁶⁴, Susanne Erk⁴¹, Jesen Fagerness³⁶, Peter T Fox⁶⁹, Nelson B Freimer⁶⁵, Michael Gill^{46,47}, Harald H H Göring⁶⁴, Donald J Hagler⁷⁰, David Hoehn⁴⁹, Florian Holsboer⁴⁹, Martine Hoogman^{5,7,71,72}, Norbert Hosten²⁴, Neda Jahanshad¹, Matthew P Johnson⁶⁴, Dalia Kasperaviciute⁷³, Jack W Kent Jr⁶⁴, Peter Kochunov^{69,74}, Jack L Lancaster⁶⁹, Stephen M Lawrie⁴⁸, David C Liewald³⁰, René Mandl¹⁵, Mar Matarin⁷³, Manuel Mattheisen⁷⁷, Eva Meissenzahl⁷⁸, Ingrid Melle^{16,79}, Eric K Moses⁶⁴, Thomas W Mühlleisen^{75,76}, Matthias Nauck⁸⁰, Markus M Nöthen^{75,76,81}, René L Olvera⁸², Massimo Pandolfi⁶⁷, G Bruce Pike⁸³, Ralf Puls²⁴, Ivar Reinvang^{84,85}, Miguel E Renteria^{2,86}, Marcella Rietschel²⁵, Joshua L Roffman³⁷, Natalie A Royle^{30,42,43}, Dan Rujescu⁷⁸, Jonathan Savitz^{45,87}, Hugo G Schnack¹⁵, Knut Schnell^{88,89}, Nina Seiferth⁴¹, Colin Smith⁹⁰, Vidar M Steen^{22,23}, Maria C Valdés Hernández^{30,42,43}, Martijn Van den Heuvel¹⁵, Nic J van der Wee^{59,60}, Neeltje E M Van Haren¹⁵, Joris A Veltman⁵, Henry Völzke⁹¹, Robert Walker⁹⁰, Lars T Westlye⁸⁴, Christopher D Whelan⁶³, Ingrid Agartz^{16,92}, Dorret I Boomsma²⁹, Gianpiero L Cavalleri⁶³, Anders M Dale^{53,70}, Srdjan Djurovic^{16,93}, Wayne C Drevets^{45,87}, Peter Hagoort^{7,52,72}, Jeremy Hall⁴⁸, Andreas Heinz⁴¹, Clifford R Jack Jr⁹⁴, Tatiana M Foroud^{34,95}, Stephanie Le Hellard^{22,23}, Fabio Macciardi⁵¹, Grant W Montgomery², Jean Baptiste Poline⁹⁶, David J Porteous^{30,97}, Sanjay M Sisodiya⁷³, John M Starr^{30,98}, Jessica Sussmann⁴⁸, Arthur W Toga¹, Dick J Veltman⁶², Henrik Walter^{41,89}, Michael W Weiner⁹⁹⁻¹⁰², the Alzheimer's Disease Neuroimaging Initiative (ADNI)¹⁰³, EPICEN Consortium¹⁰³, IMAGEN Consortium¹⁰³, Saguenay Youth Study Group (SYS)¹⁰³, Joshua C Bis¹⁰⁴, M Arfan Ikram¹⁰⁵⁻¹⁰⁷, Albert V Smith^{108,109}, Vilmundur Guðnason^{108,109}, Christophe Tzourio^{10,111}, Meike W Vernooij¹⁰⁵⁻¹⁰⁷, Lenore J Launer¹¹², Charles DeCarli^{113,114}, Sudha Seshadri^{115,116}, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium¹⁰³, Ole A Andreassen^{16,79}, Liana G Apostolova^{1,32}, Mark E Bastin^{30,42,43,117}, John Blangero⁶⁴, Han G Brunner⁵, Randy L Buckner^{26,27,37,68}, Sven Cichon^{75,76,118}, Giovanni Coppola^{32,119}, Greig I de Zubicaray⁸⁶, Ian J Deary^{30,31}, Gary Donohoe^{46,47}, Eco J C de Geus²⁹, Thomas Espeseth^{84,85,120}, Guillén Fernández^{7,52,71}, David C Glahn^{8,9}, Hans J Grab^{13,121}, John Hardy⁵⁴, Hilleke E Hulshoff Pol¹⁵, Mark Jenkins¹²², René S Kahn¹⁵, Colm McDonald²⁰, Andrew M McIntosh⁴⁸, Francis J McMahon³⁵, Katie L McMahon¹²³, Andreas Meyer-Lindenberg²⁵, Derek W Morris^{46,47}, Bertram Müller-Myhsok⁴⁹, Thomas E Nichols^{122,124}, Tomas Paus²¹, Zdenka Pausova¹⁷, Brenda W Penninx^{59,60,62,125}, Steven G Potkin⁵¹, Philipp G Sämann⁴⁹, Andrew J Saykin^{33,34,95}, Gunter Schumann³⁹, Jordan W Smoller^{36,37}, Joanna M Wardlaw^{30,42,43}, Nijholt Wijlaars⁵⁰, Niclasius C Märtti^{2,128}, Barbara Franke^{5-7,128}, Margaret J Wright^{2,128} & Paul M Thompson^{1,128} for the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) Consortium¹²⁶

207 co-authors



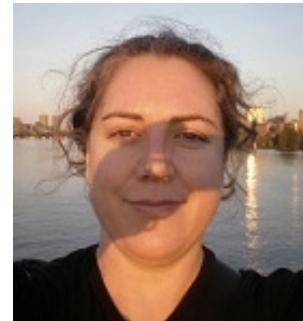
Paul Thompson
UCLA



Rudy Senstad
UCLA



Derrek Hibar
UCLA



Sarah Medland
Queensland Institute of
Medical Research,
Australia



Alejandro Arias Vasquez
Radboud University
Nijmegen, Nijmegen, The
Netherlands