Homogenizing data using ENIGMA-MEGA analysis

Peter Kochunov

University of Maryland, School of Medicine

Introduction

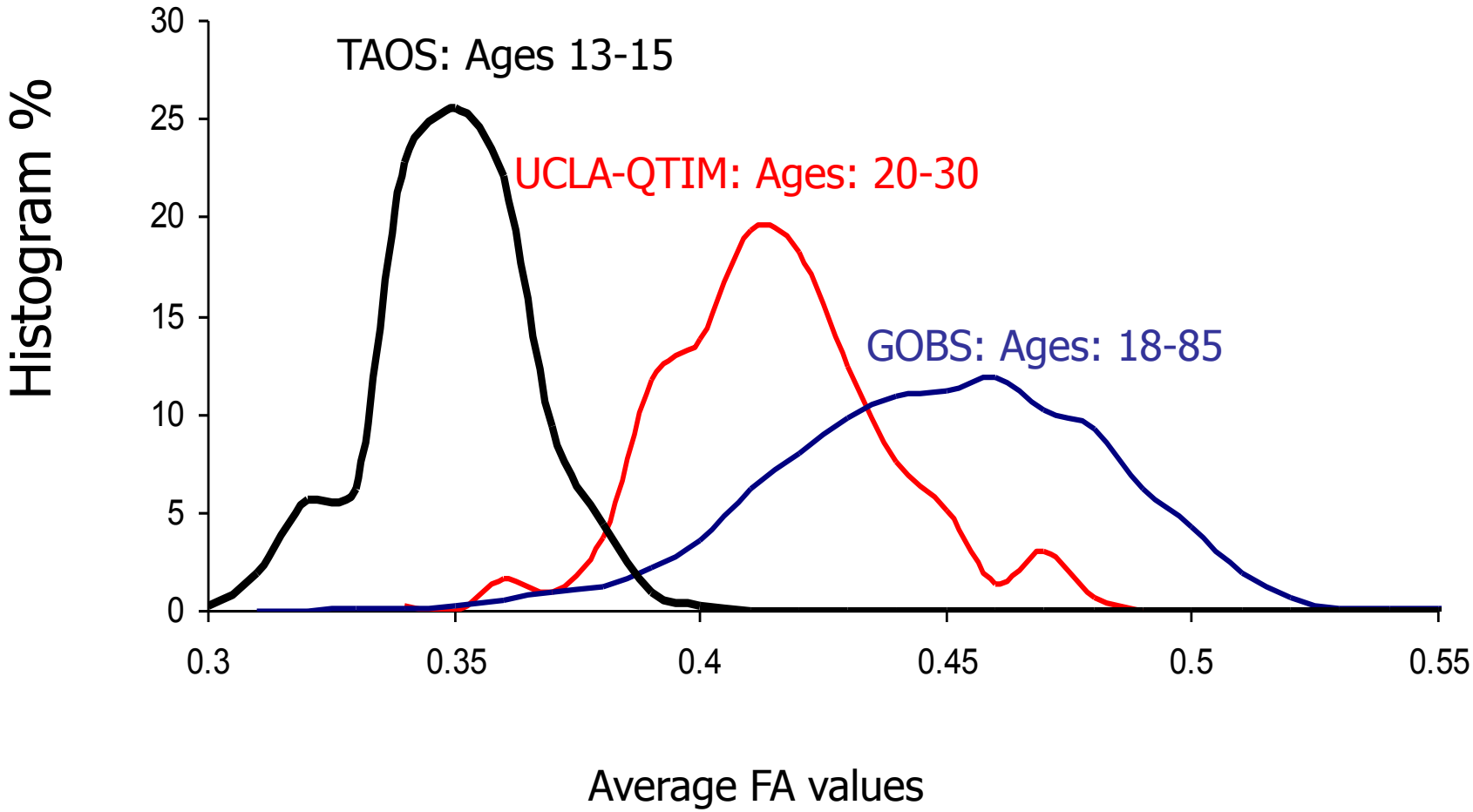
- What is mega-analysis
- MEGA-Analysis algorithm developed by ENIGMA
- Examples:
  - MEGA analysis of additive genetic effects
  - MEGA analysis of SCZ effects on white matter integrity

Mega-analysis

- Combining of “raw” data from multiple studies
- Pros:
  - Additive increase in degrees of freedom
  - Simplified analysis structure
  - Uniformly weighting by subjects
    - Removes weighting uncertainty in Meta analysis of genetic data
  - Ideal for familial genetics studies where analysis is performed per family

Mega-Analysis: Cons

- Data may have a site-specific bias

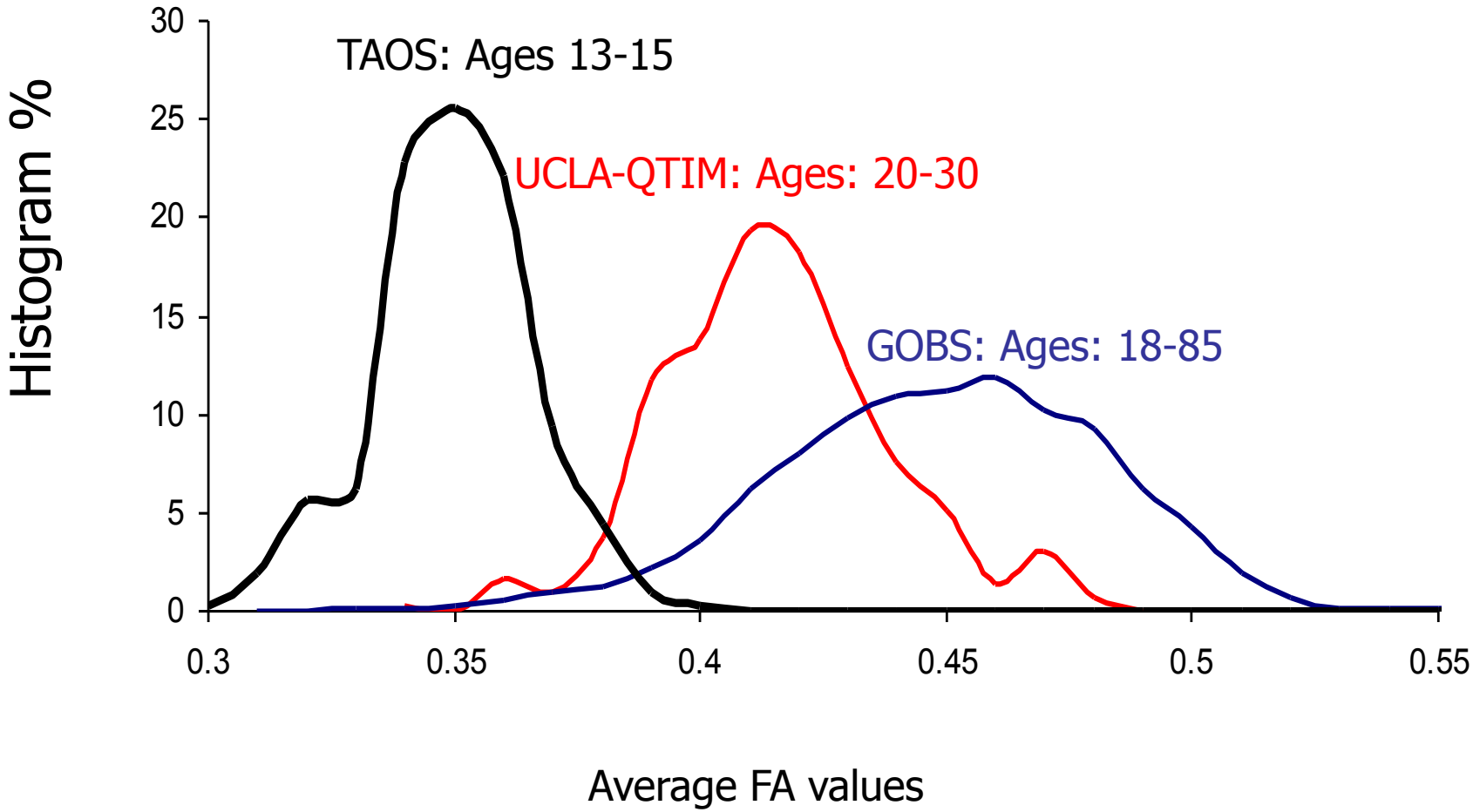


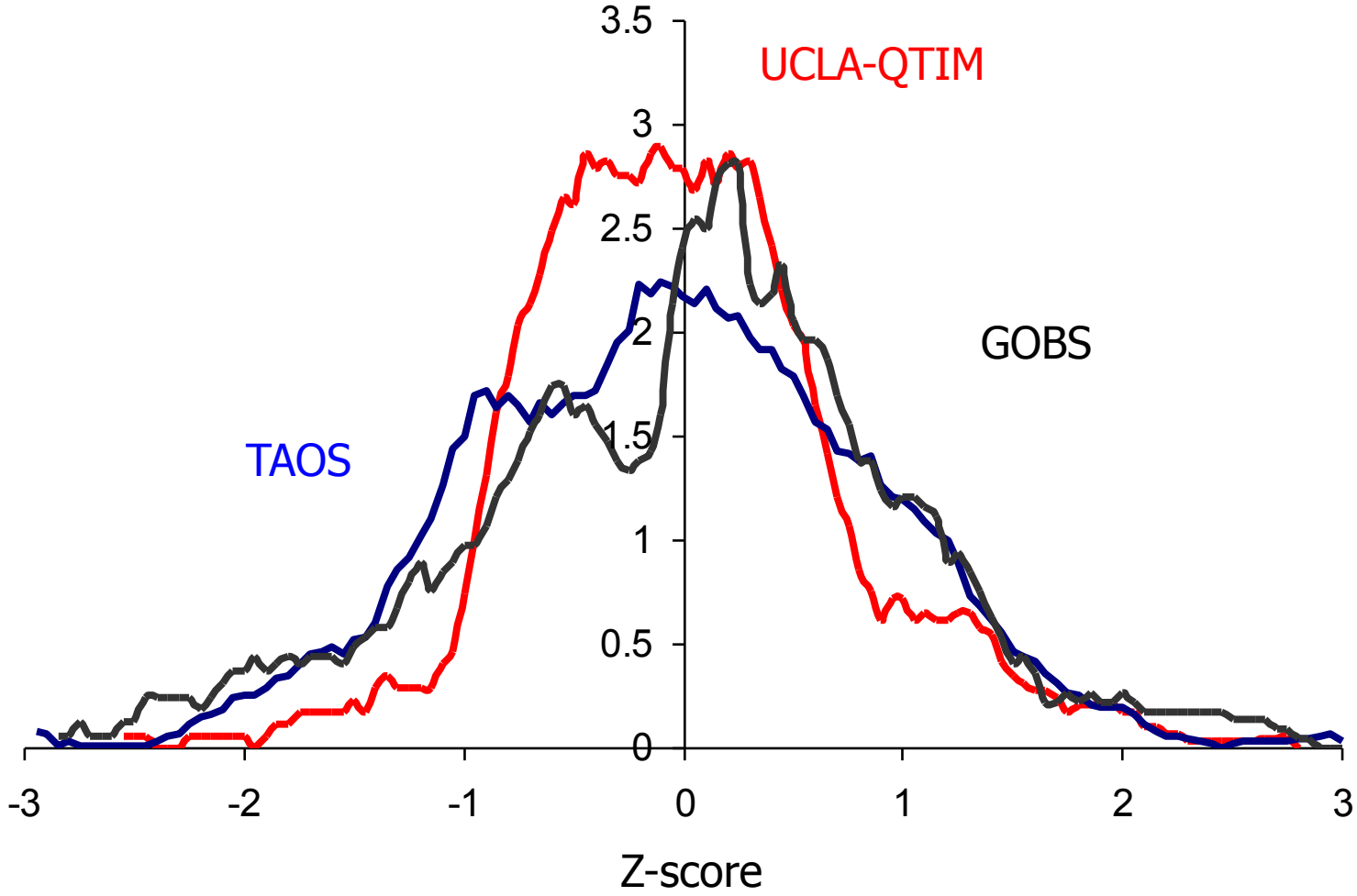
Mega-Analysis Algorithm

- Developed by Neda and Me
- Coded in SOLAR-Eclipse
- Tried in the following analyses
  - Additive Genetic Analysis (Heritability)
  - Additive Genetic Analysis (Genetic correlation)
  - Association Analyses (GWA)
  - Disorder effect analyses
    - Effects of Schizophrenia on white matter

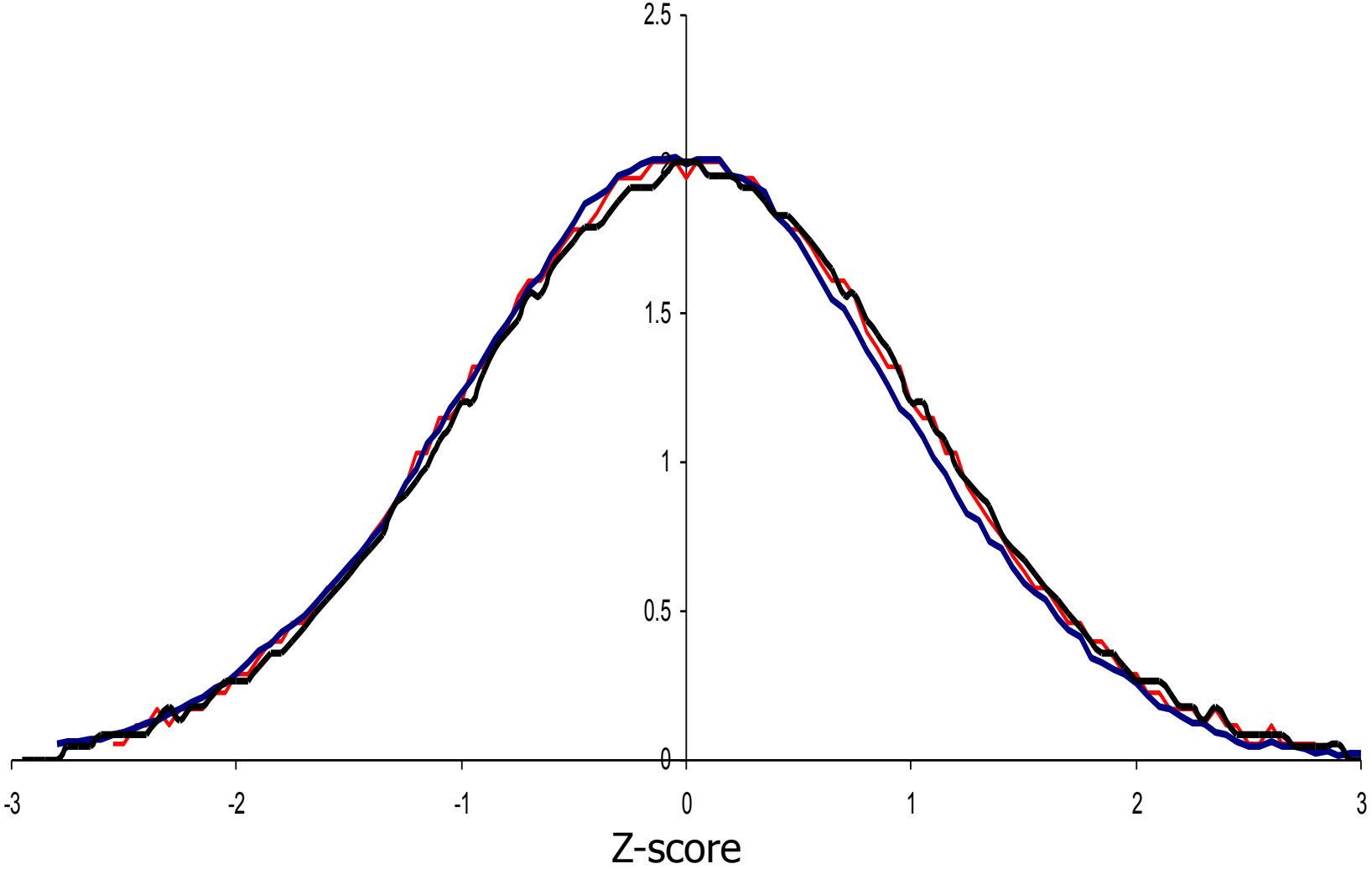
Step 1. Regression of nuisance covariates is performed per site

Remove effects of the covariates that don't act as "contrast" to make data "equivalent" per site





Step 2 ➔ Inverse normalization



Step 3 ➔ Testing for Stratification
ANOVA of heritability estimates

Test of homogeneity of the effect per group

Measure h^2 per sample. Perform ANOVA

UCLA= 0.56±0.25; p=.0001

TAOS=0.49±0.23; p=0.04

GOBS= 0.45±0.07; $p=10^{-8}$

No difference among groups

We can combine them into a single pedigree with the weight assigned based on the relativeness and the pedigree strength .

Significance of additive effects: Mega vs. Meta

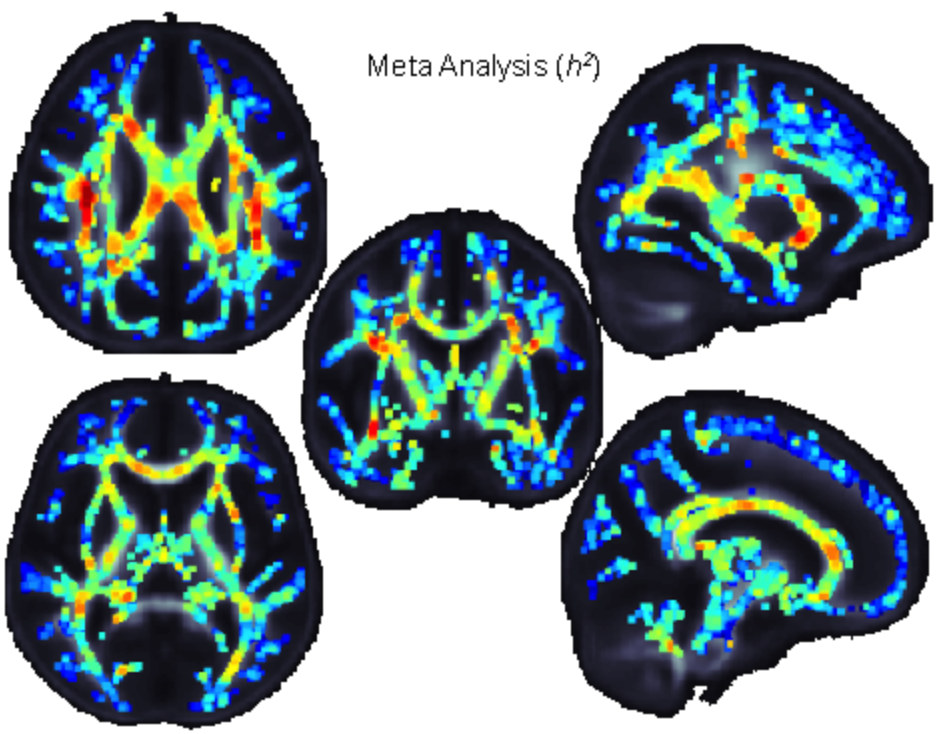
- Mega Analysis: lowest SE and highest significance
  - $h^2=0.47\pm0.02$; $p=10^{-16}$
- Meta Analysis StdErr-Weighted
  - $h^2=0.48 \pm0.09$; $p=0.004$
- Meta Analysis N-Weighted
  - $h^2=0.44 \pm0.03$; $p=10^{-6}$

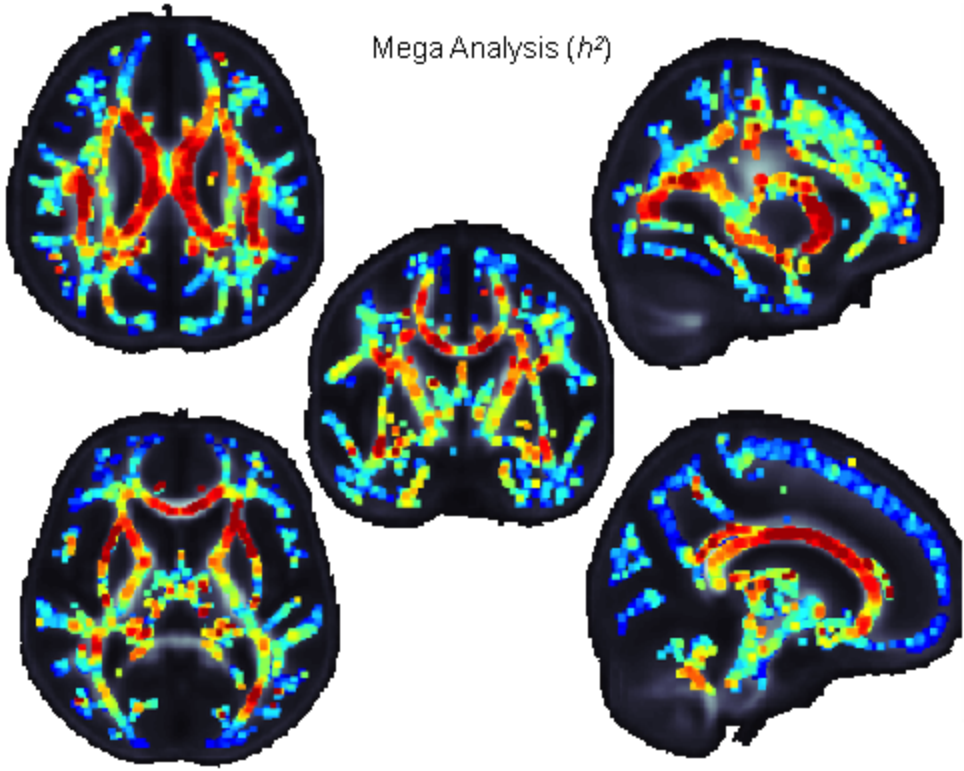
Greatly influenced by the small samples

Difficult to justify given that subjects don't contribute equally

Similar trends in voxel-wise data

P-values for heritability estimates (-log10)



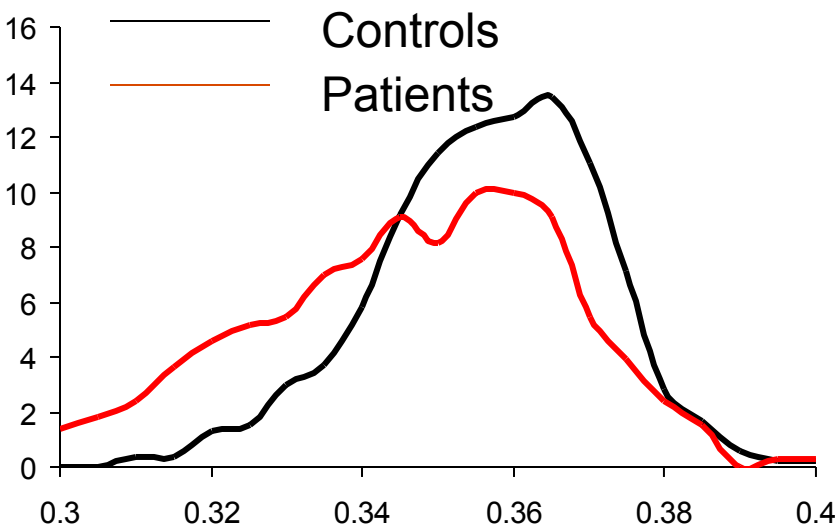


Effects of SCZ on white matter integrity

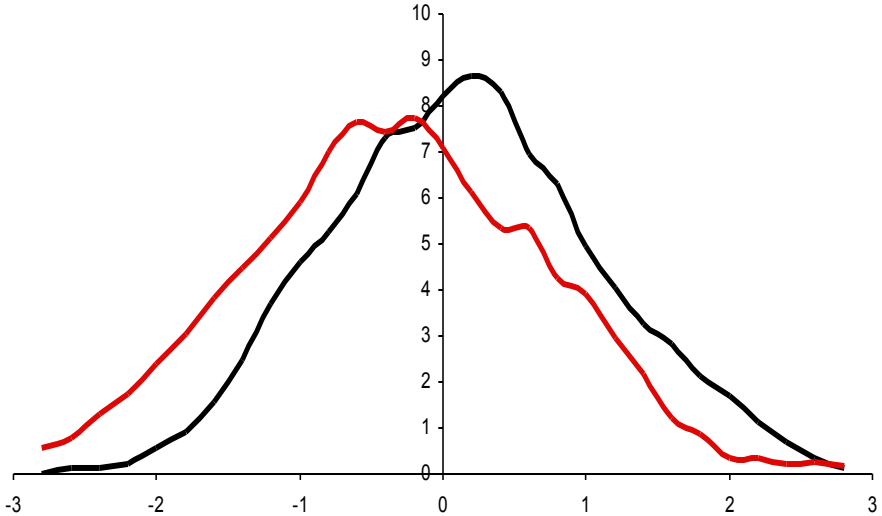
- Apply mega-analysis to study effects of disorder on FA values
- Use three samples collected on three scanners
- Some cross-over of subjects to directly study effects of data transform

Effects per Site: Site 1 (N=350)

raw significance $p=2*10^{-6}$

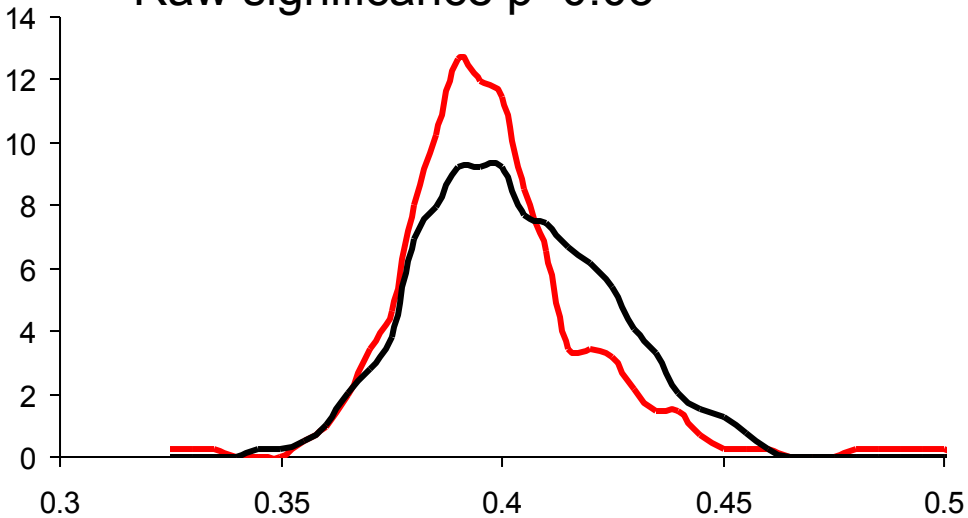


Transformed significance $p=10^{-6}$

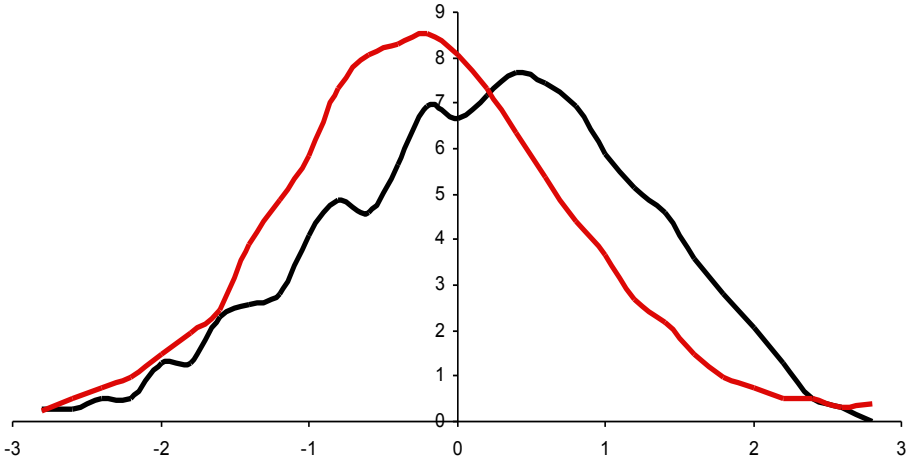


Effects per Site: Site 2 (N=220)

Raw significance p=0.03

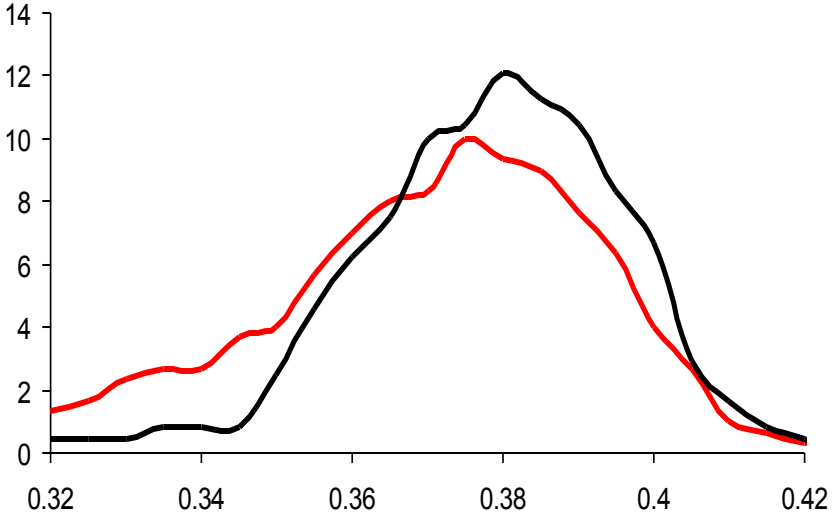


Transformed significance p=0.01

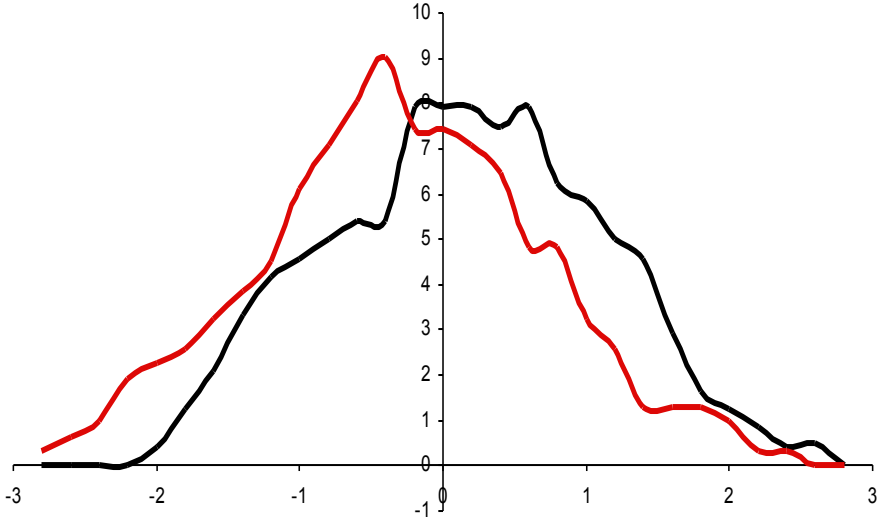


Effects per Site: Site 3 (N=120)

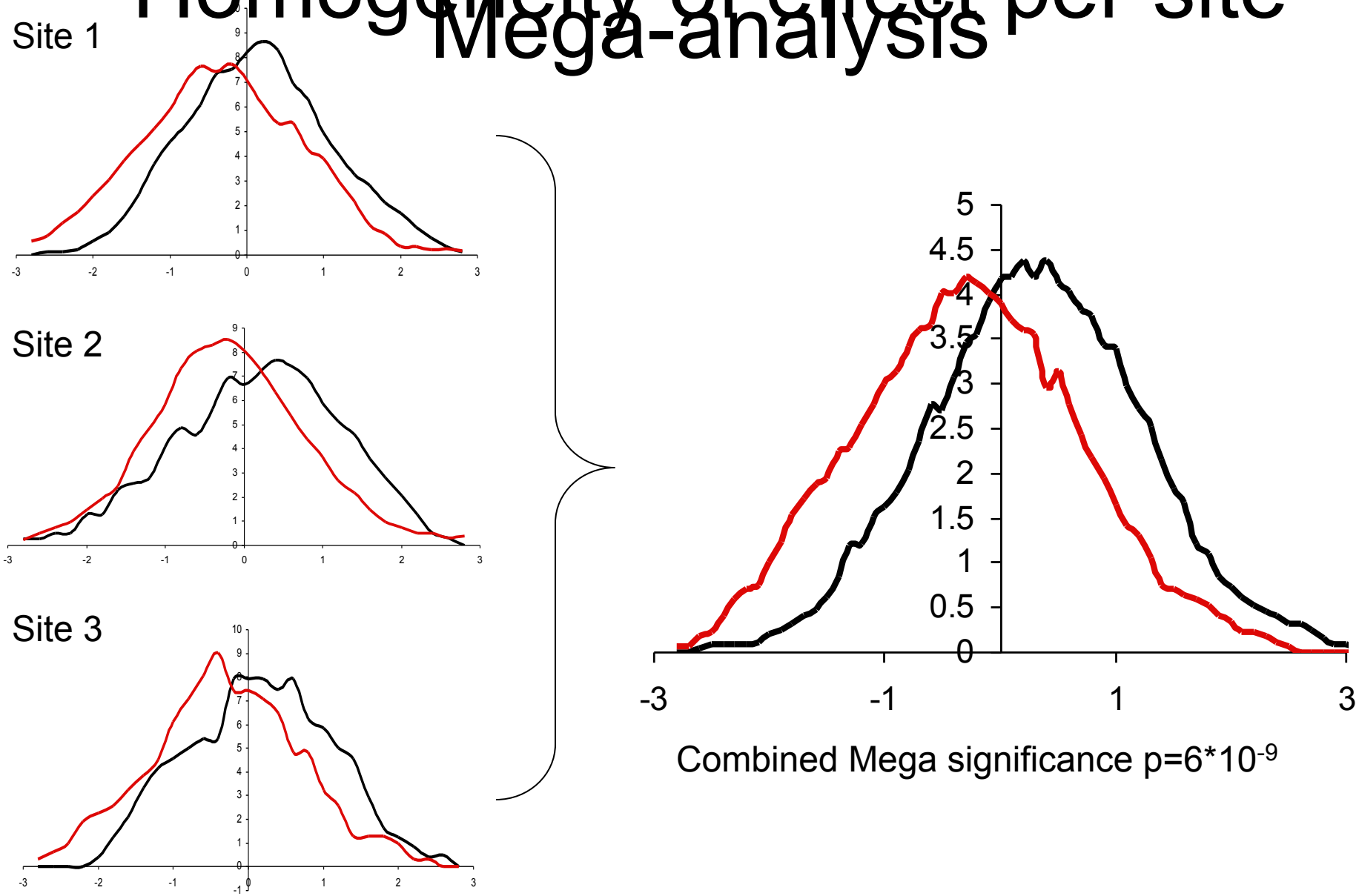
raw significance p=0.03



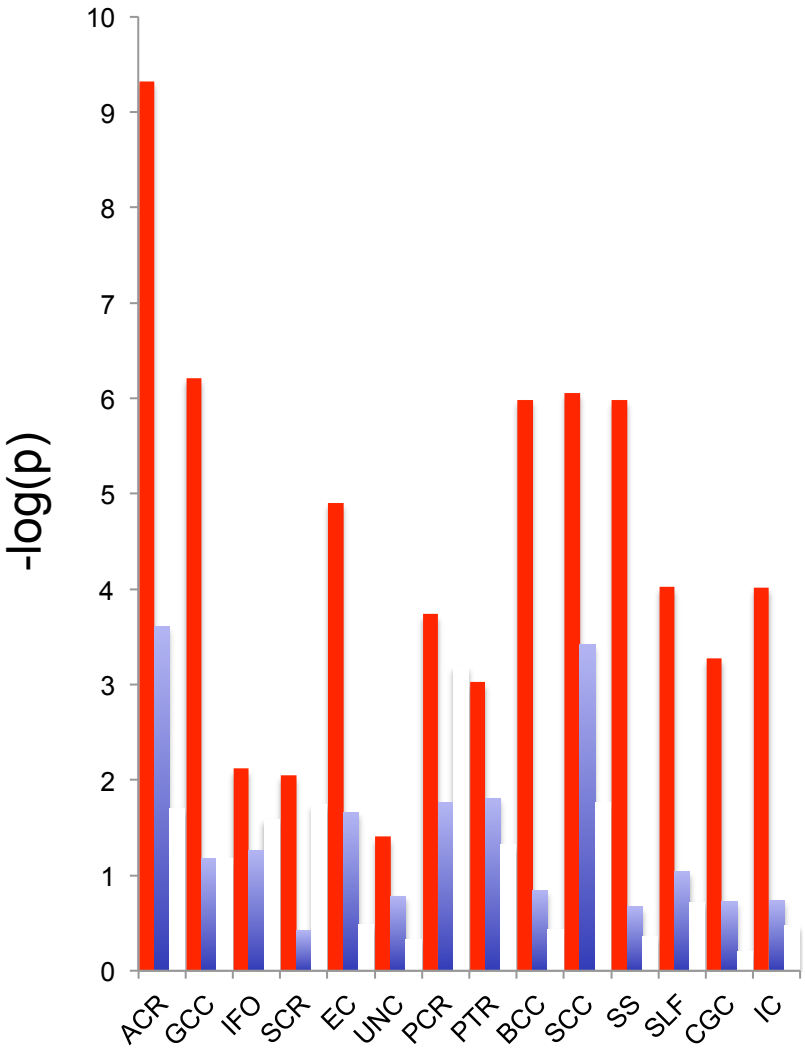
Transformed significance p=0.03

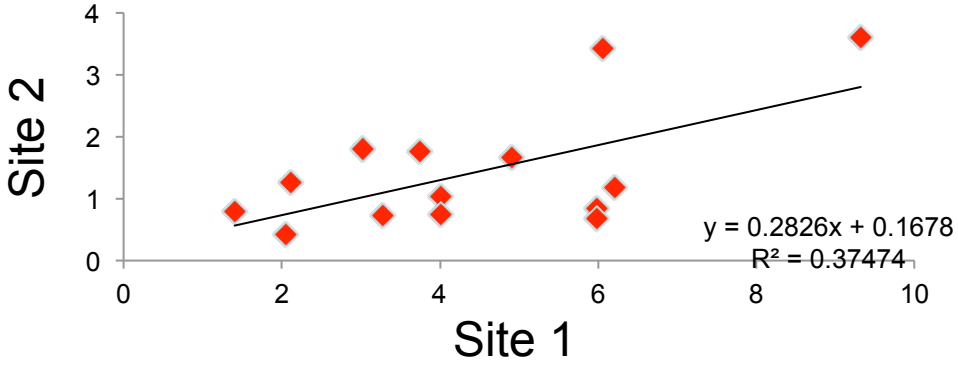


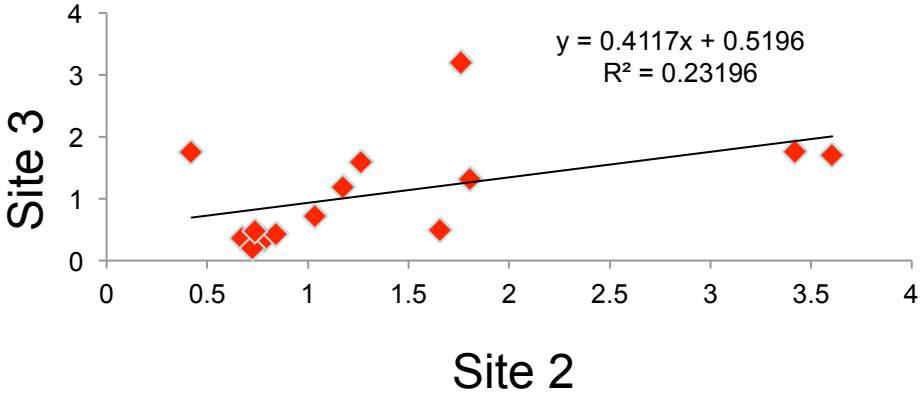
Homogeneity of effect per site
Mega-analysis

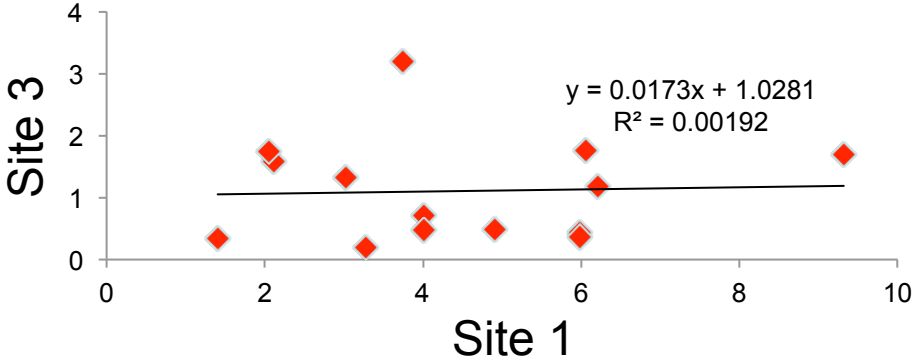


Regional Specificity?









Mega-analysis results of regional effects

Greatest impact with Schizophrenia

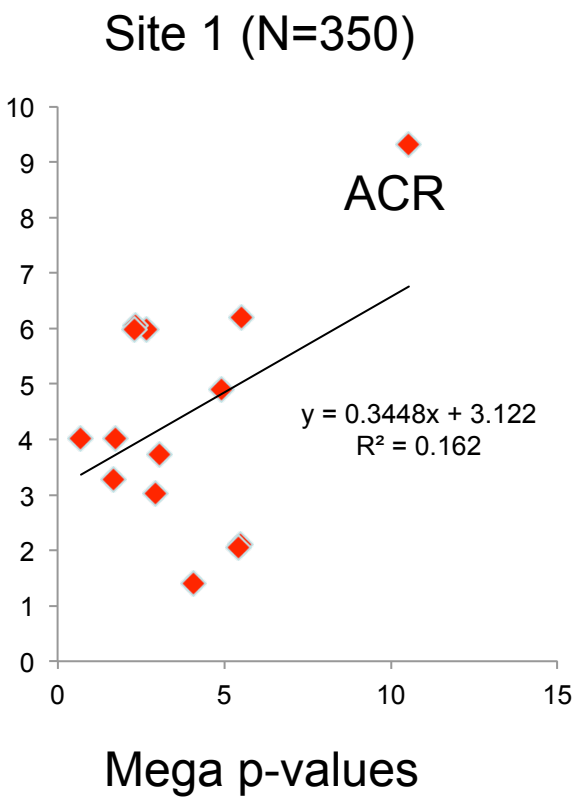
- **Anterior corona radiata $p<10^{-11}$**
- **Genu of Corpus Calosum $p<10^{-6}$**
- **Inferior Frontal Occipital $p<10^{-5}$**
- **Superior Corona radiata $p<10^{-5}$**

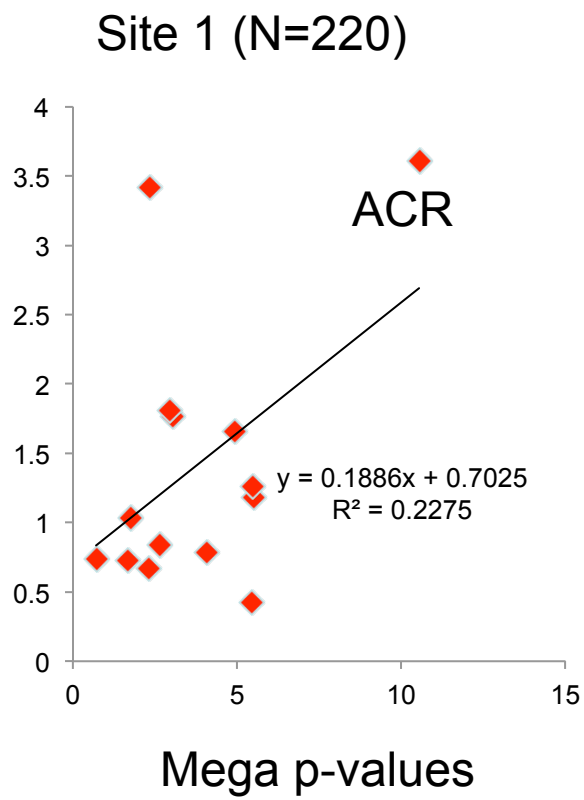
Mega-analysis results of regional effects

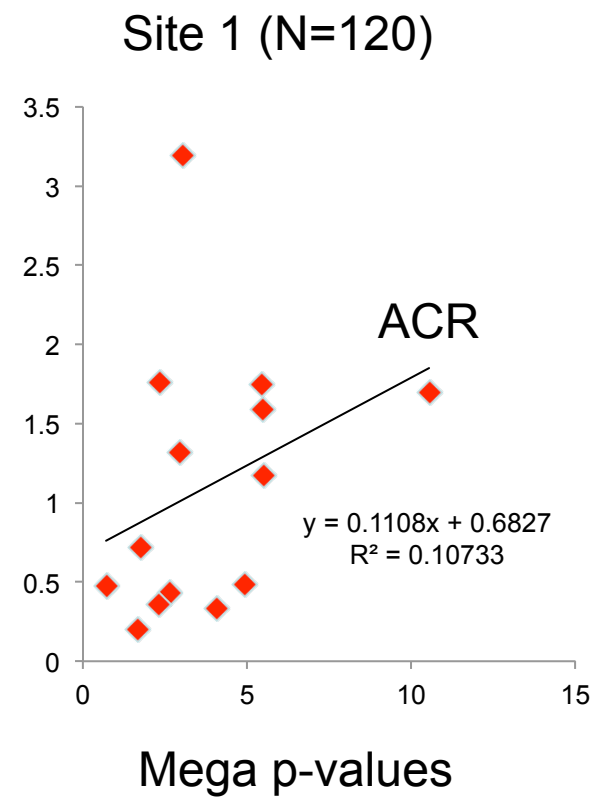
Least impact with Schizophrenia

- Cortico-Spinal Tract p=0.2
- Superior-Frontal Occipital 0.05
- Uncinate fasiculos p=0.02

Regional MEGA vs site



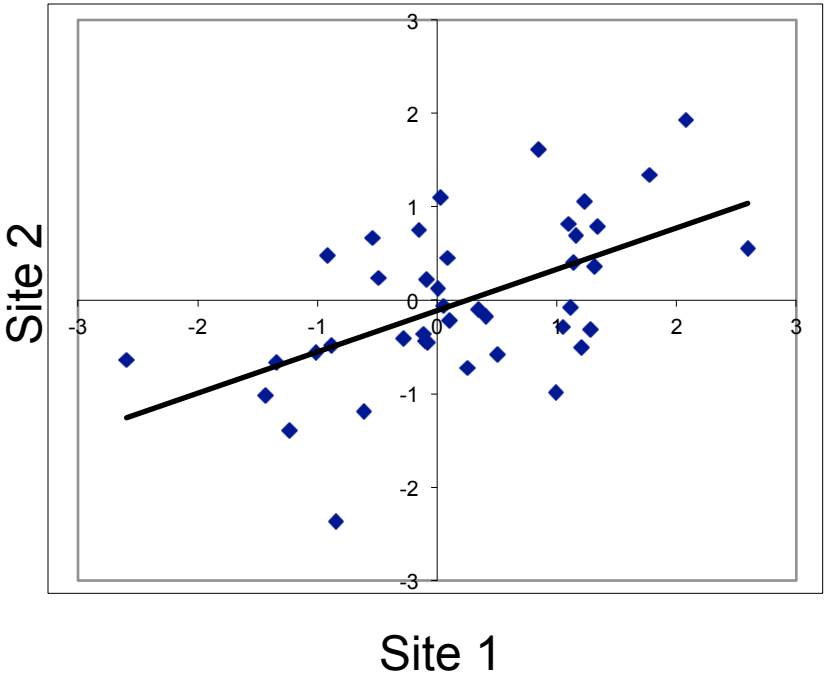




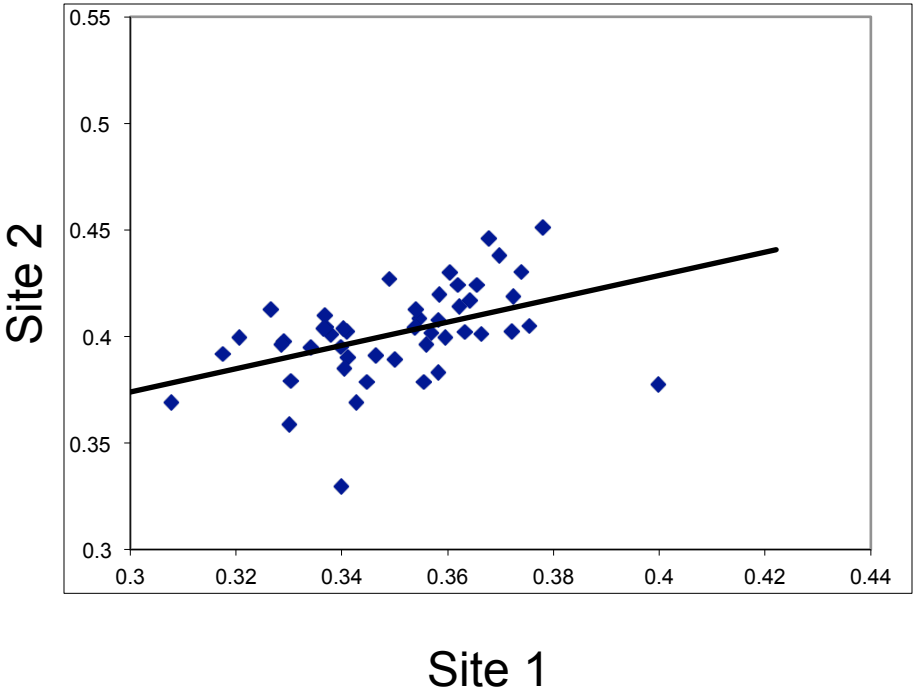
How does this work on individual subjects?

N=35 subjects were imaged at Site 1 and 2 in studies 5 years apart

R=0.55

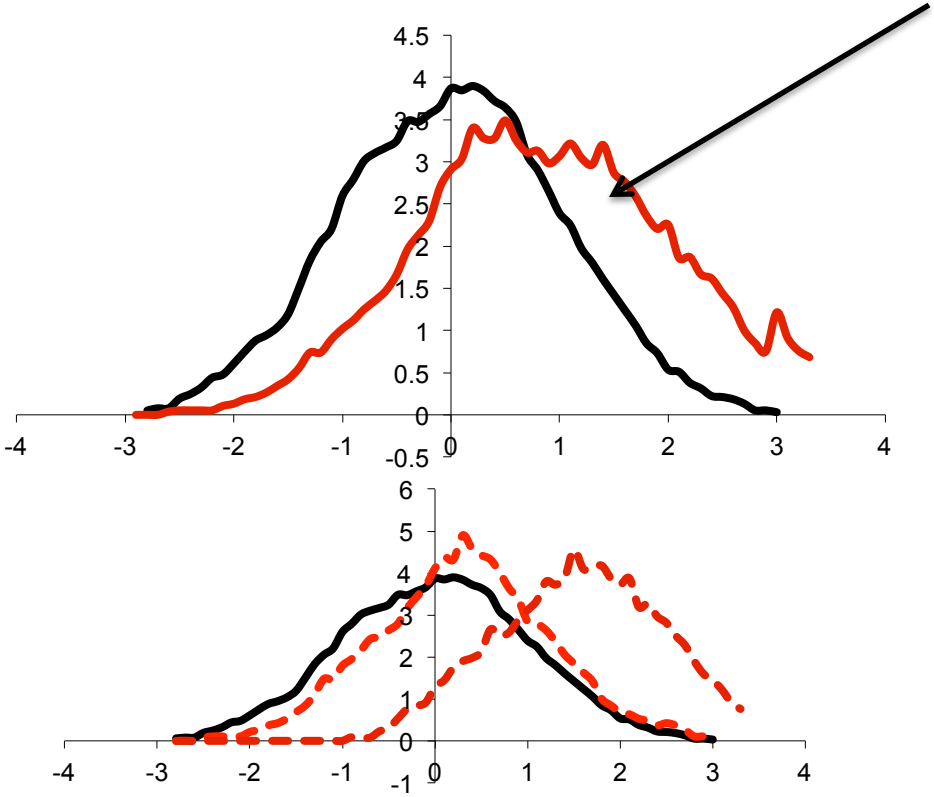


R=0.43



Limitations: Normality

- Data for patients and controls has to be transformable to “normal” state
  - Violated if patients have bi-modal distribution



Excessive Kurtosis

- Caused by bi-modal distribution of FLAIR lesions in patients
- Use inverse normal mapping parameters from the Controls
- Use bi-Gaussian fit to probabilistically separate patients

Acknowledgement

- ENIGMA Team
  - Paul Thompson, Neda Jahanshad, Siniad Kelly, Jessica Turner
- The PIs of the GOBS project: John Blangero and David Glahn
- NIH
  - R01s MH085646, R01DA027680, R01EB015611, MH078111, MH0708143 and MH083824
  - U54EB020403 and P50MH103222