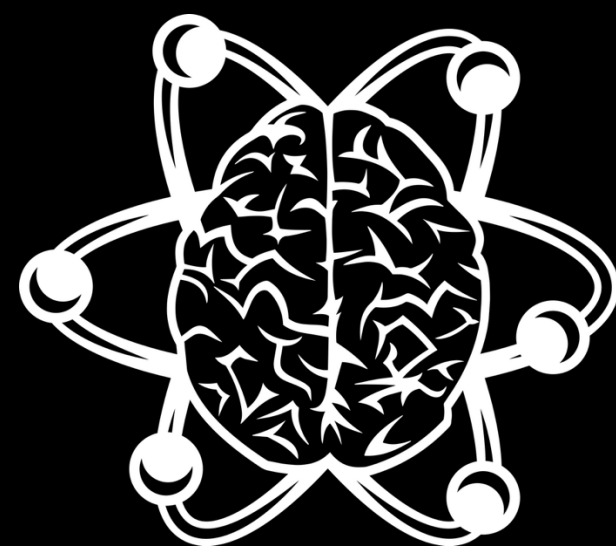
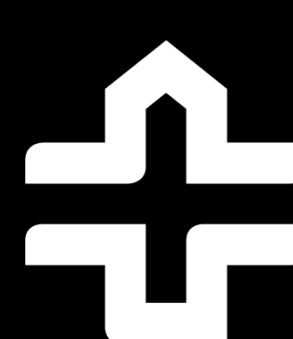




NEUROINFORMATICS AND LARGE- SCALE ANALYSIS

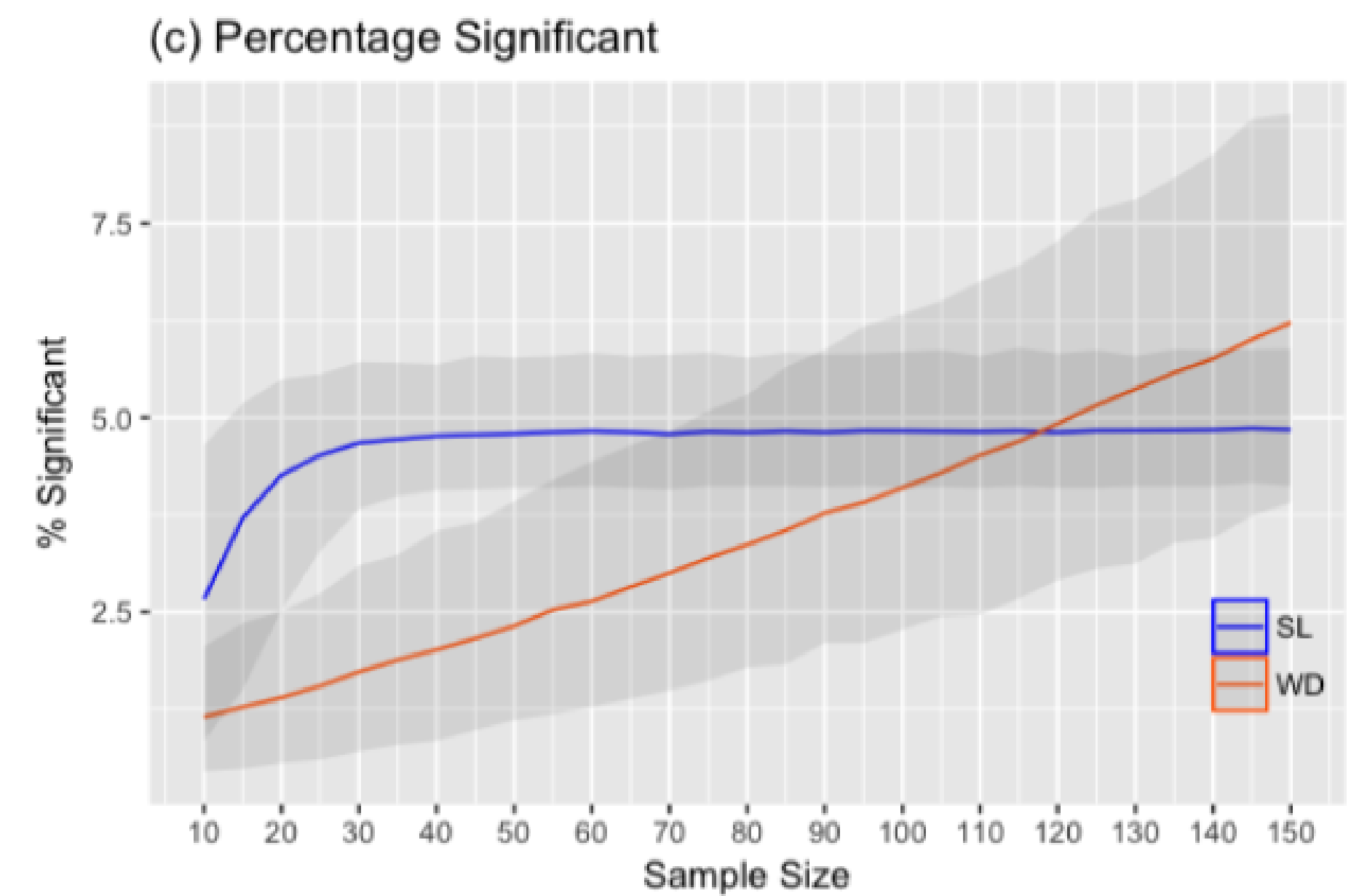
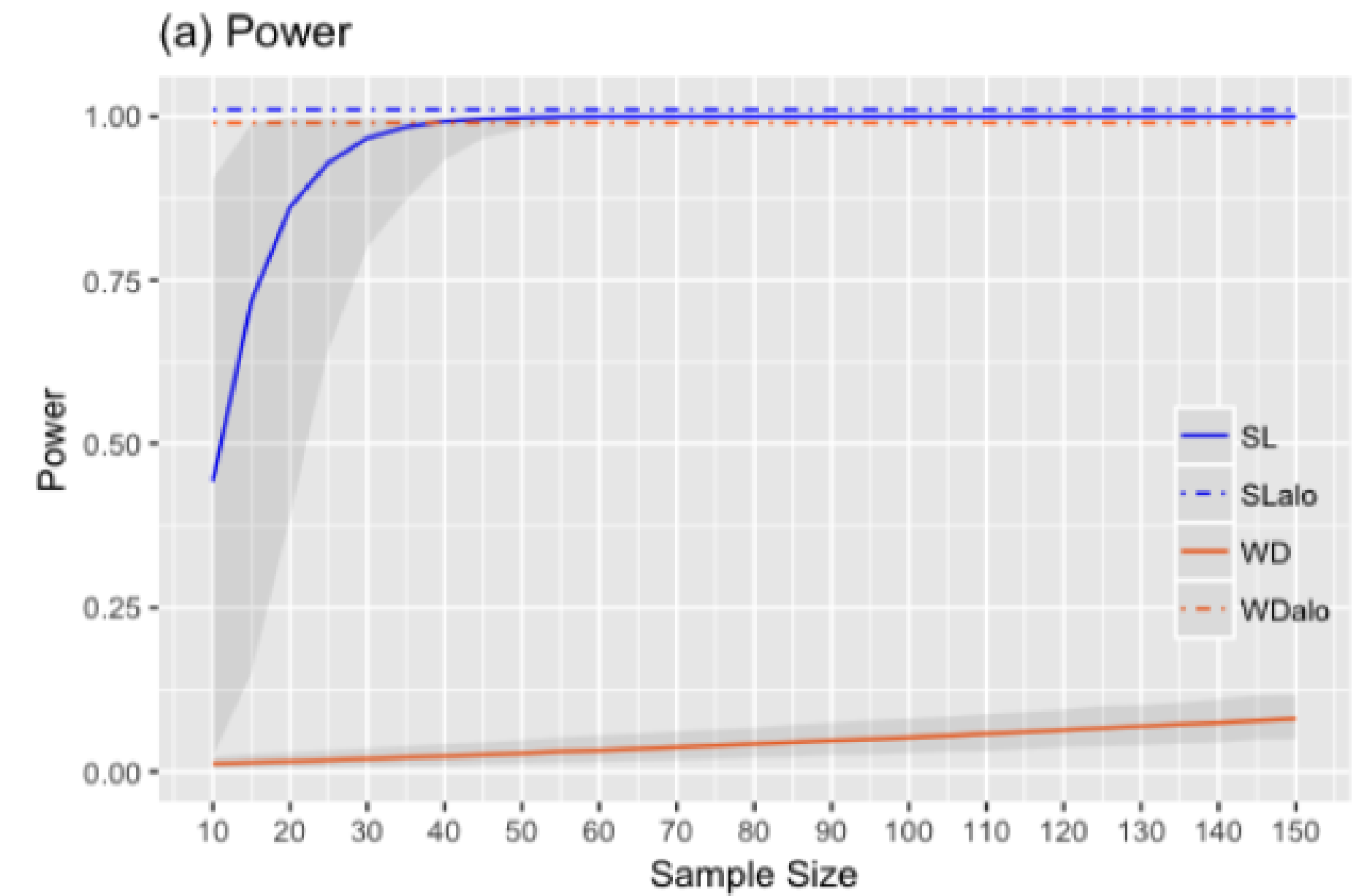
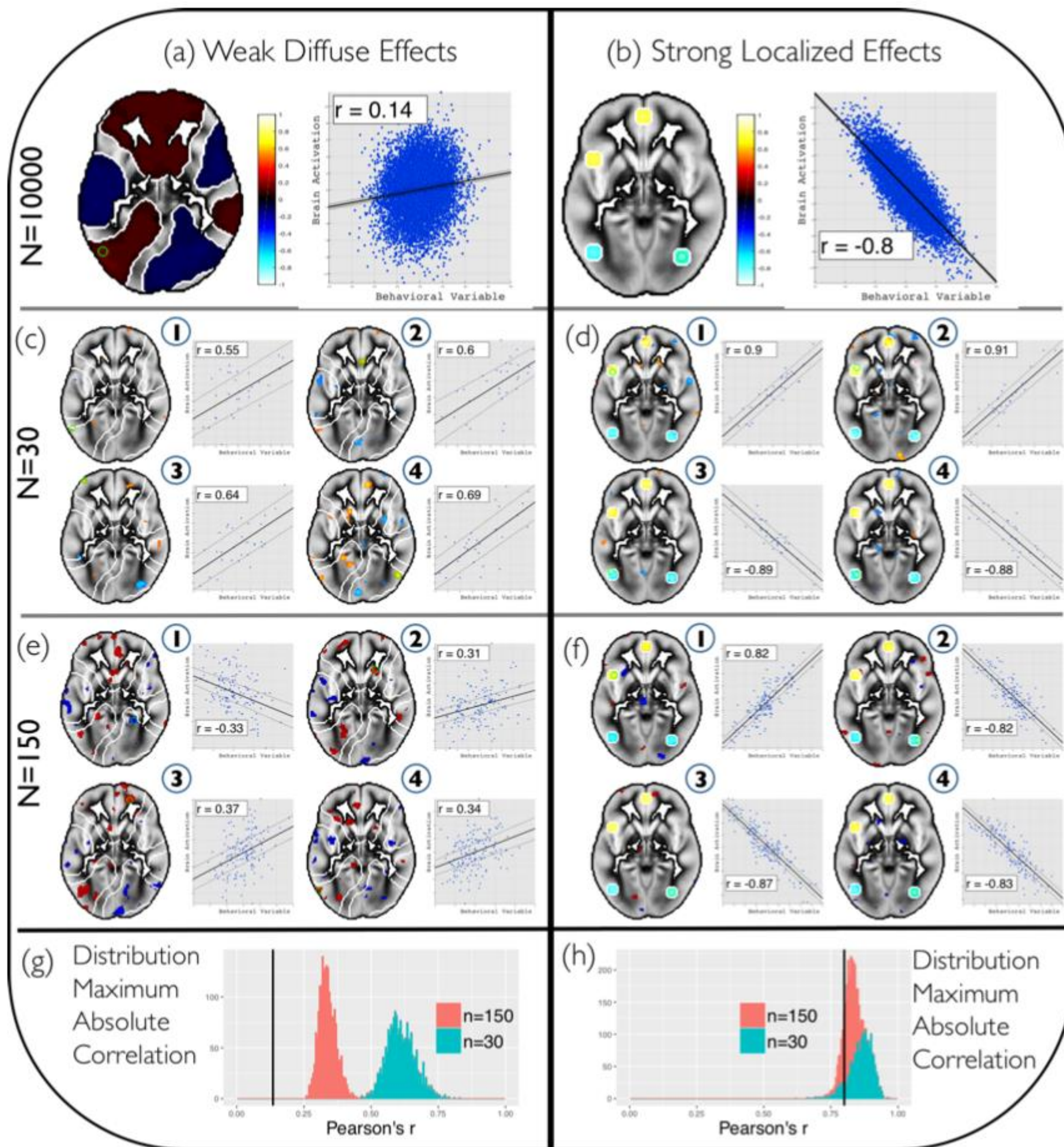
Turku PET Centre Brain Imaging Course 2025

Lauri Nummenmaa, Turku PET Centre



More is more

- Imaging is often financially prohibitive (>5000€ / PET scan) thus sample sizes are compromised
- Potential harm to subject needs to be minimised —> sample sizes kept at minimum
- Underpowered studies can result in experimental failure even when experimental design and measurements are otherwise sufficient
- Poor statistical power increases Type 1 & 2 error rates and lead to poor replicability.



Basic problems

Data storage

- Where are my **data**?
- What if my lab members **leave**?
- How can i **re-access** my data?

Data processing

- What are my postdocs **doing** all day long?
- Are you sure the files are **good**?

Data analysis

- Is everything done **lege artis**?
- How can we **reproduce** our results?

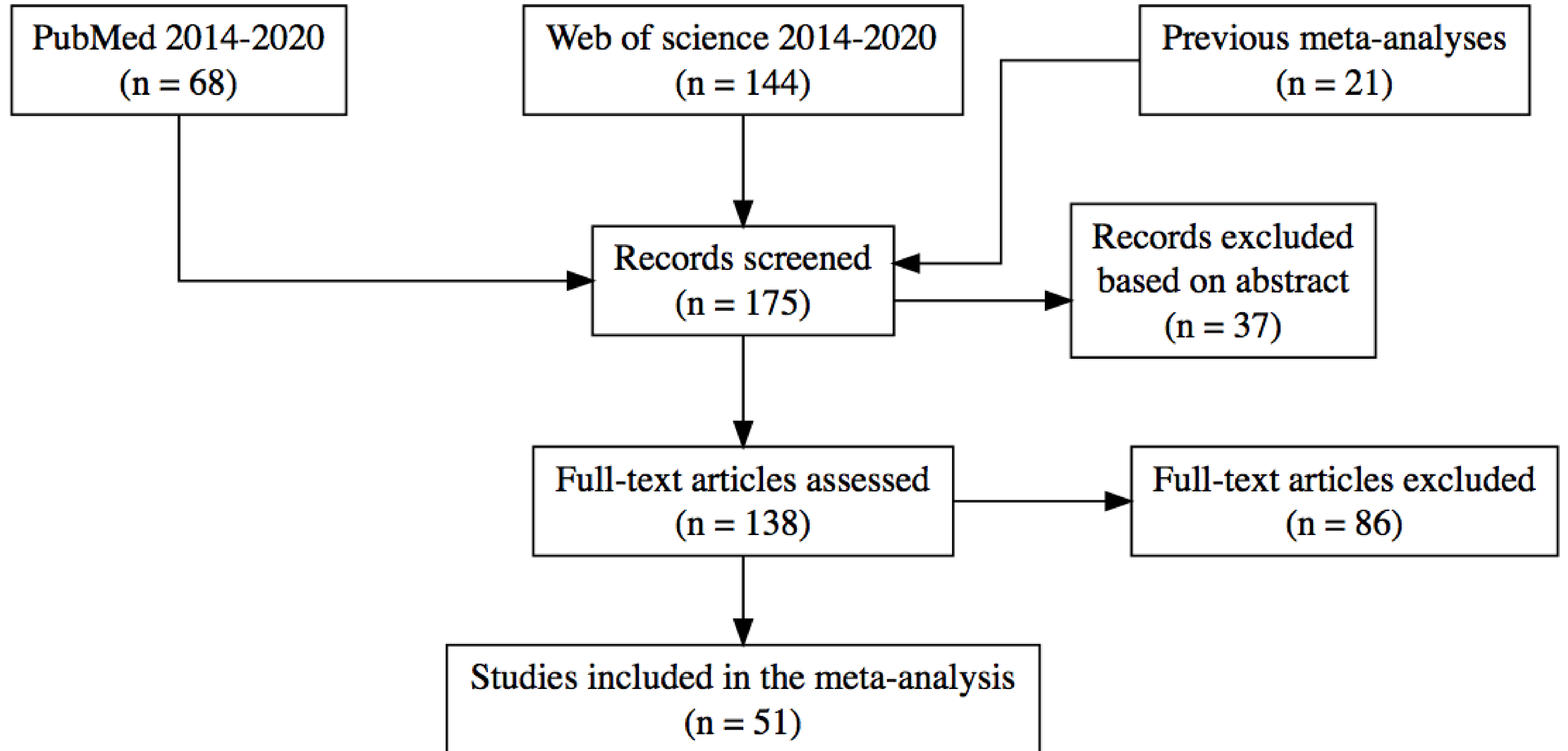
Data synthesis

- How can we **combine** data?
- **What** can we combine?

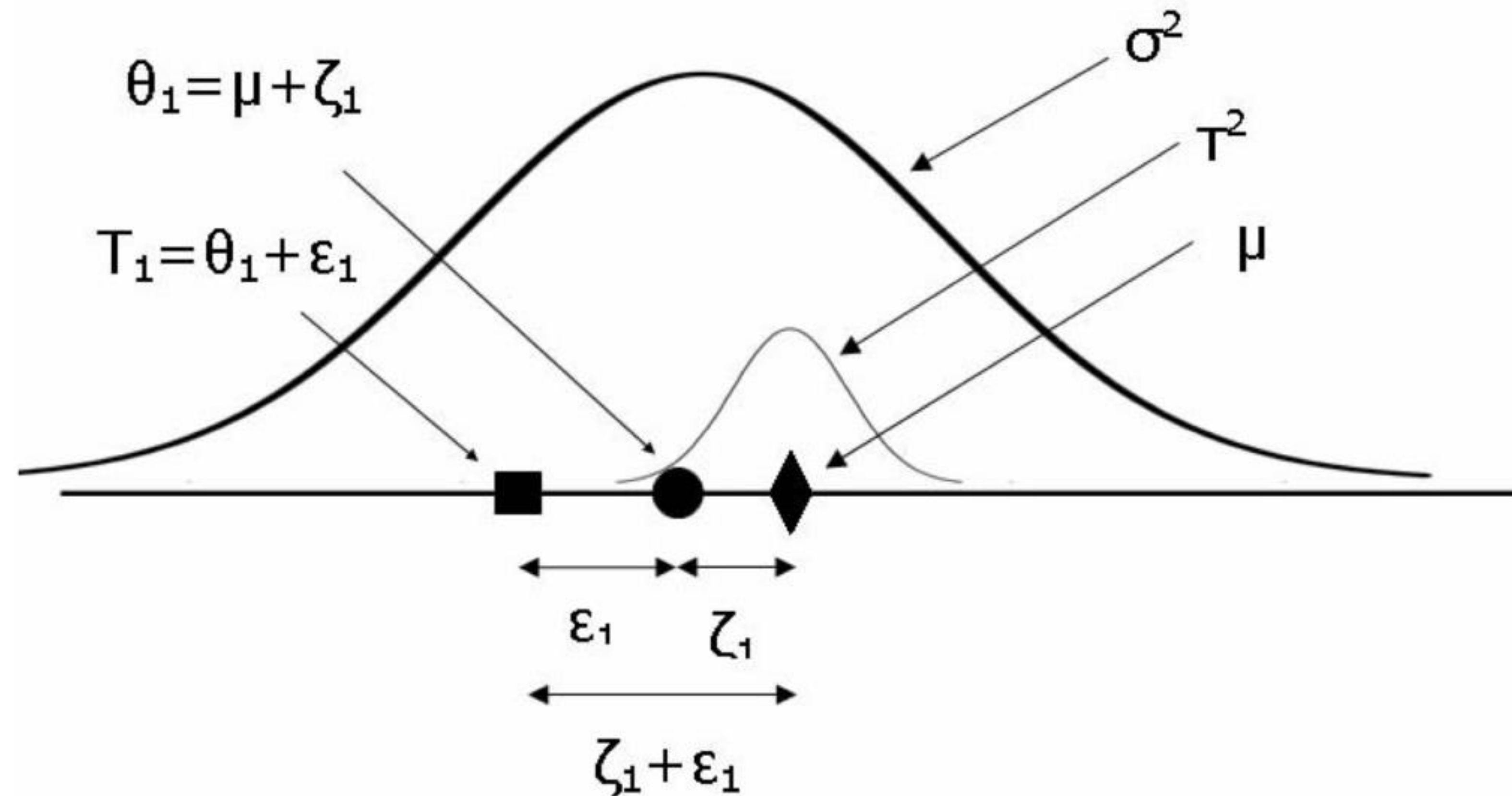
Solution 1: Meta-analysis

- Meta-analysis: Pooling standardised effect sizes to estimate population effect location and distribution
- For neuroscience, three main approaches
 - ROI level data and classic univariate meta-analysis
 - Coordinate-based data and volumetric meta-analysis
 - Combination of statistical maps from original studies





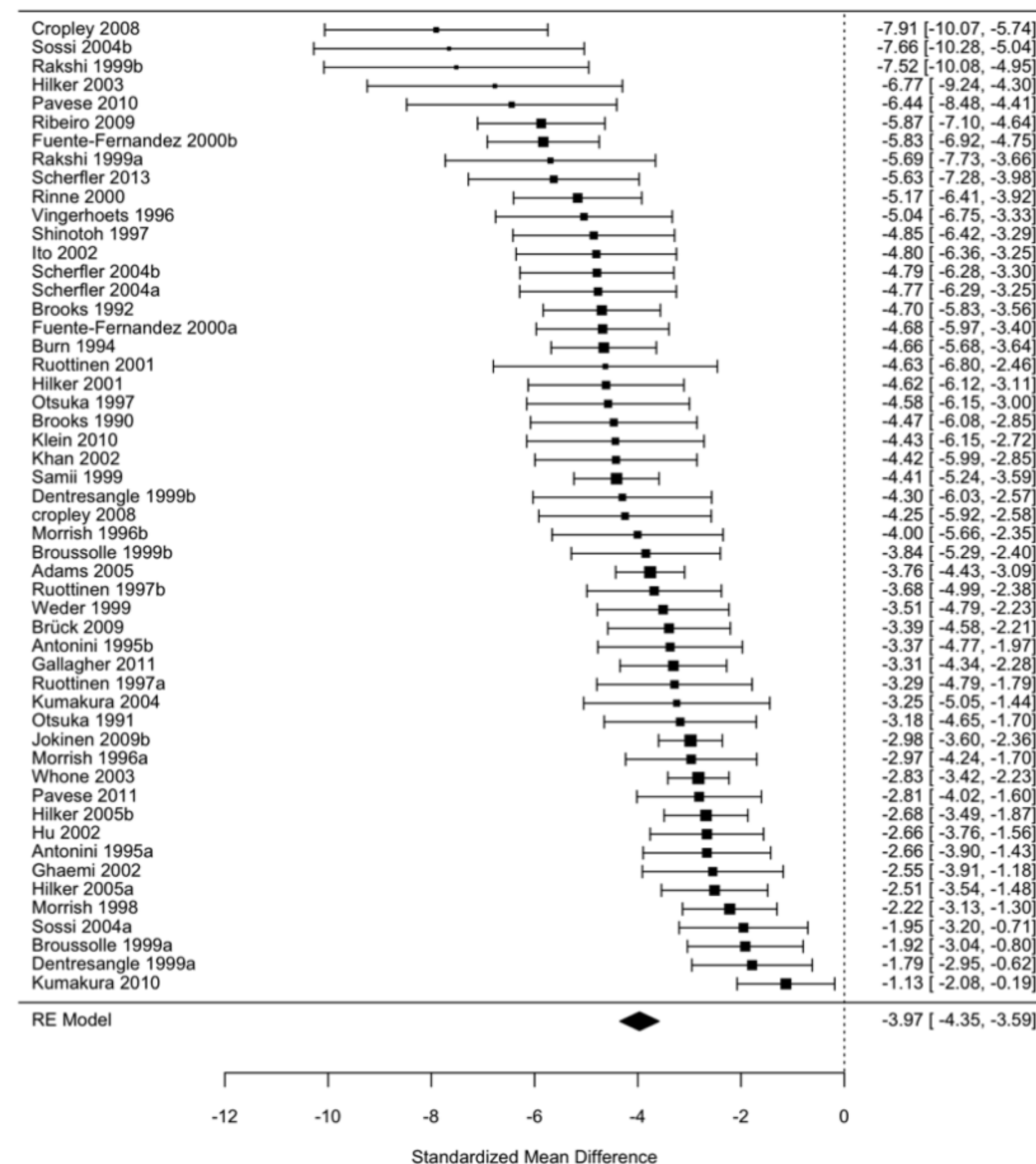
Classic mixed effects meta-analysis



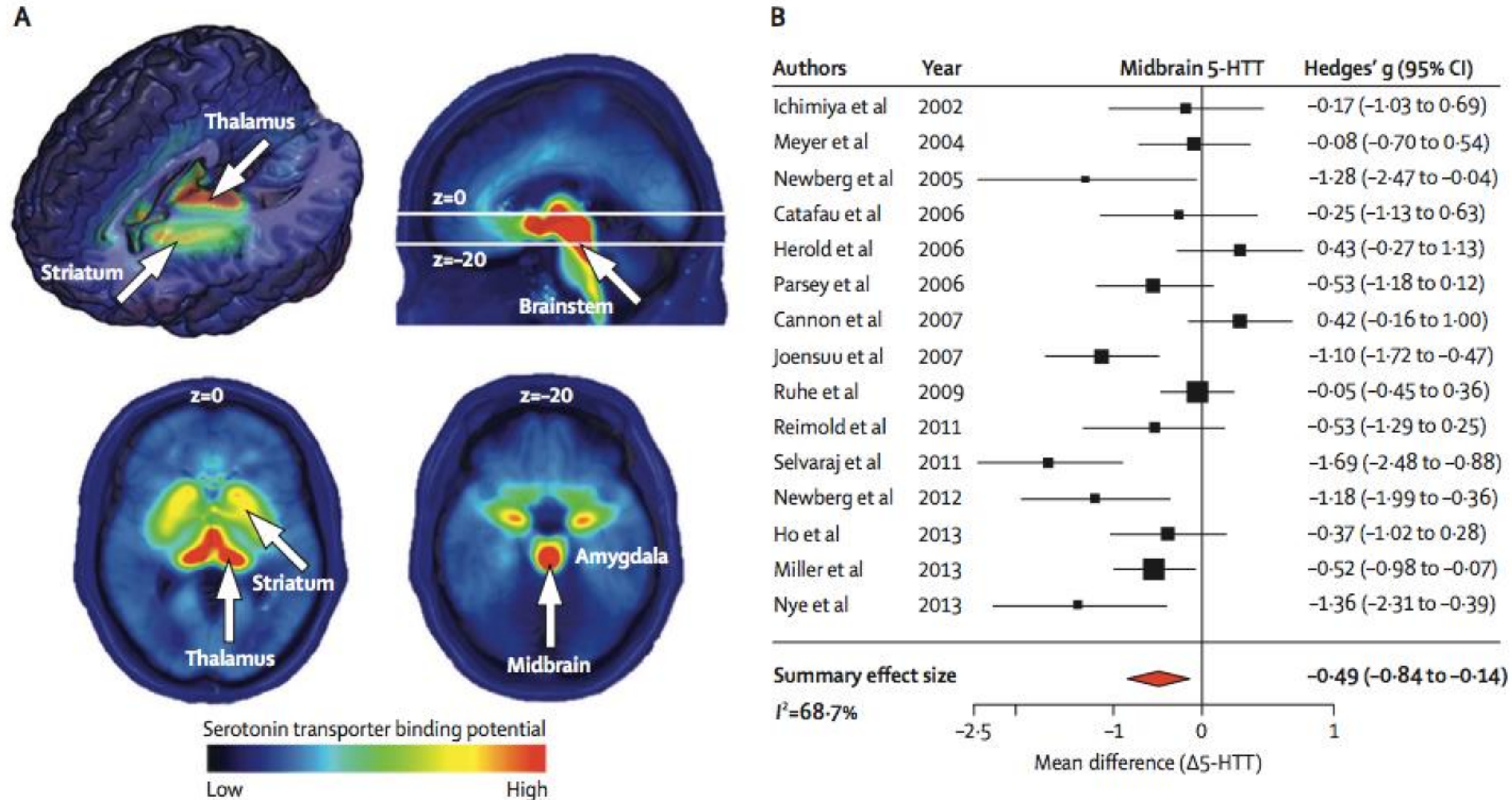
The observed effect T_1 is sampled from a distribution with true effect θ_1 , and variance σ^2 .
This true effect θ_1 , in turn, is sampled from a distribution with mean μ and variance τ^2 .

Classical meta-analysis

- Easy to perform even with limited statistical knowledge
- Most required functions available in R package **metafor**:
- Effect size calculation / conversion
- Model estimation
- Plotting



Approach 1: Regional analysis



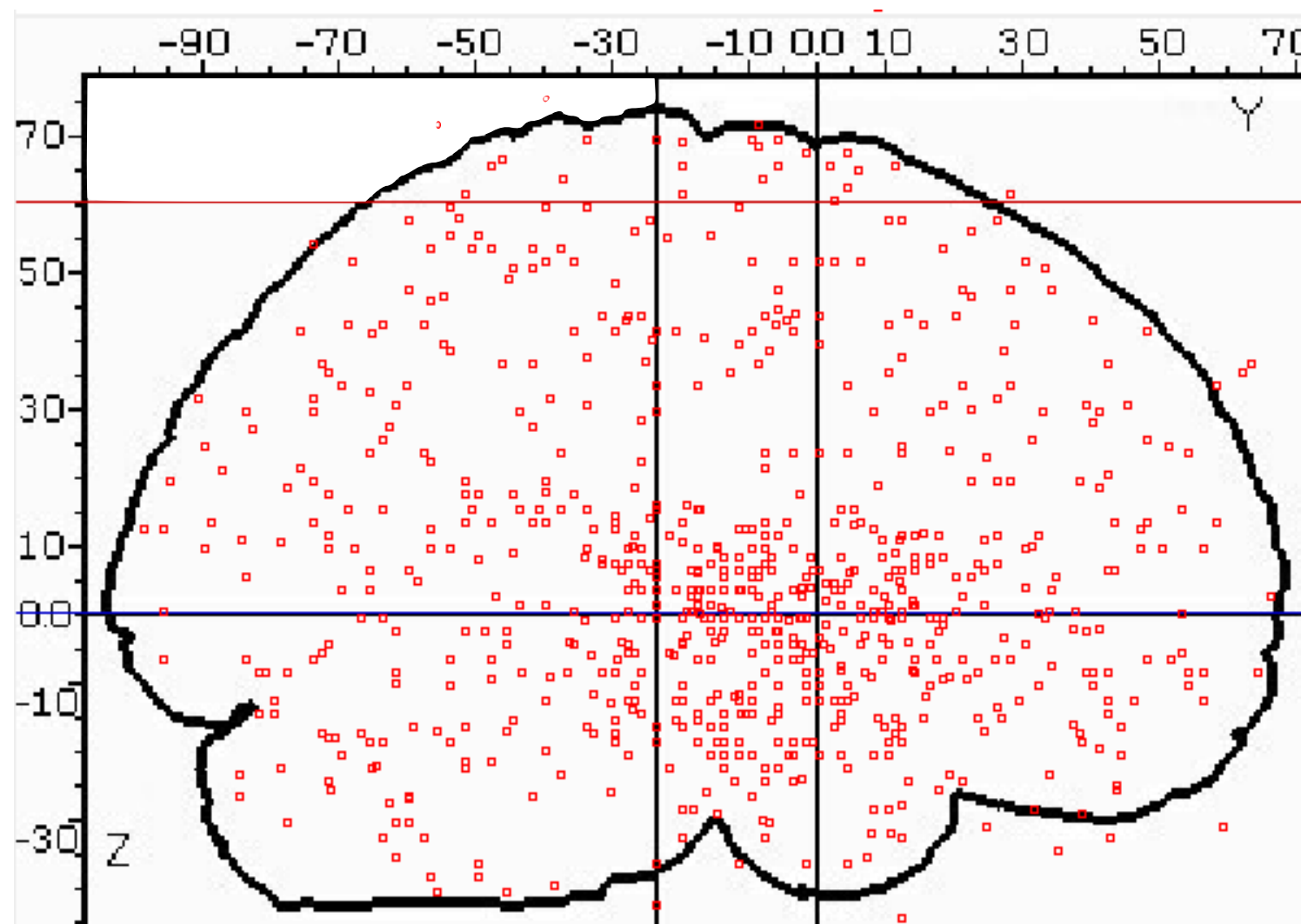
Spies et al (2015)

Regional meta-analysis: Pros and cons

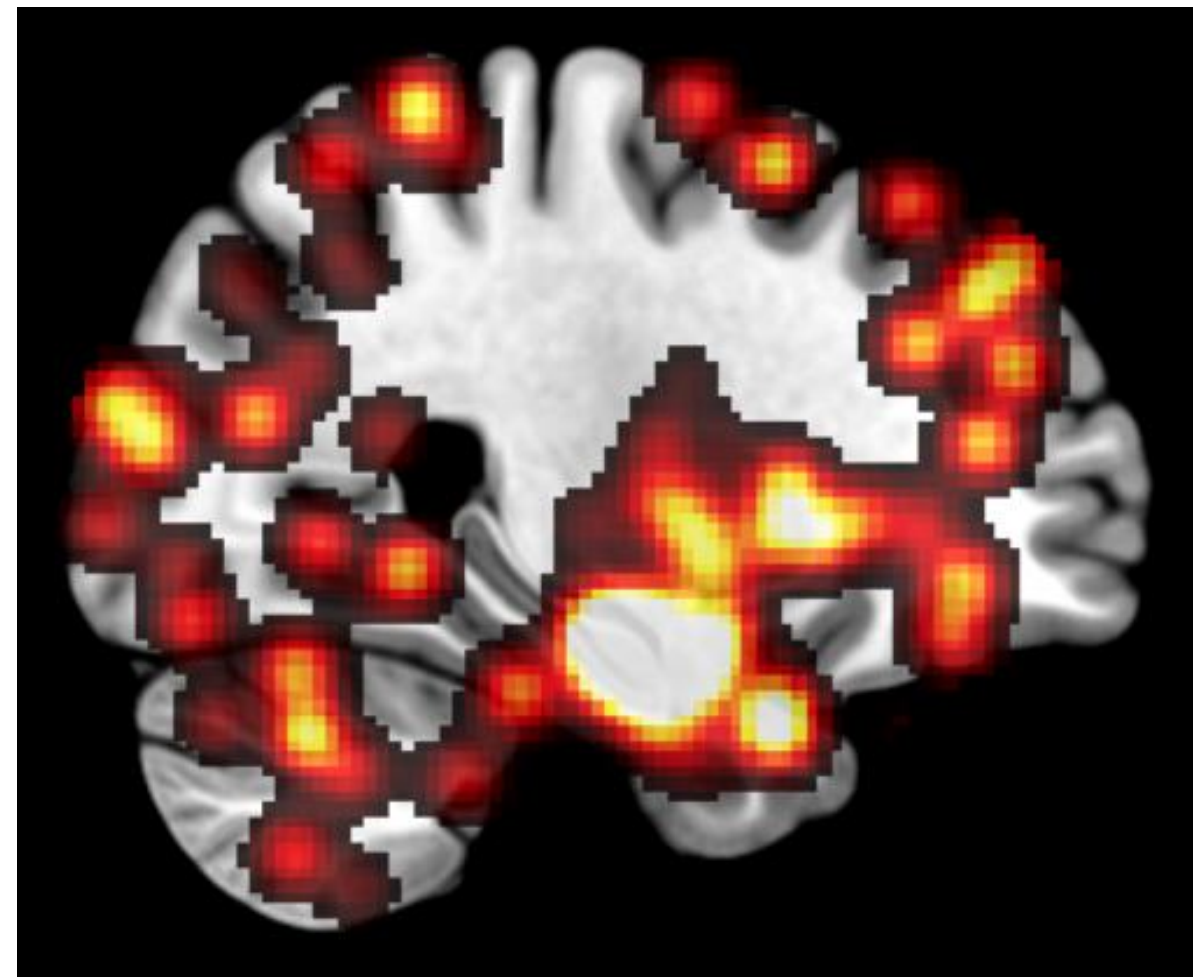
- Easy to analyse and interpret
- Data comparable in statistical terms
- No need to worry about normalization etc.
- Laborious
- Anatomical nomenclature not consistent
- Misses effects outside chosen ROIs

Approach 2: Peak-based analysis

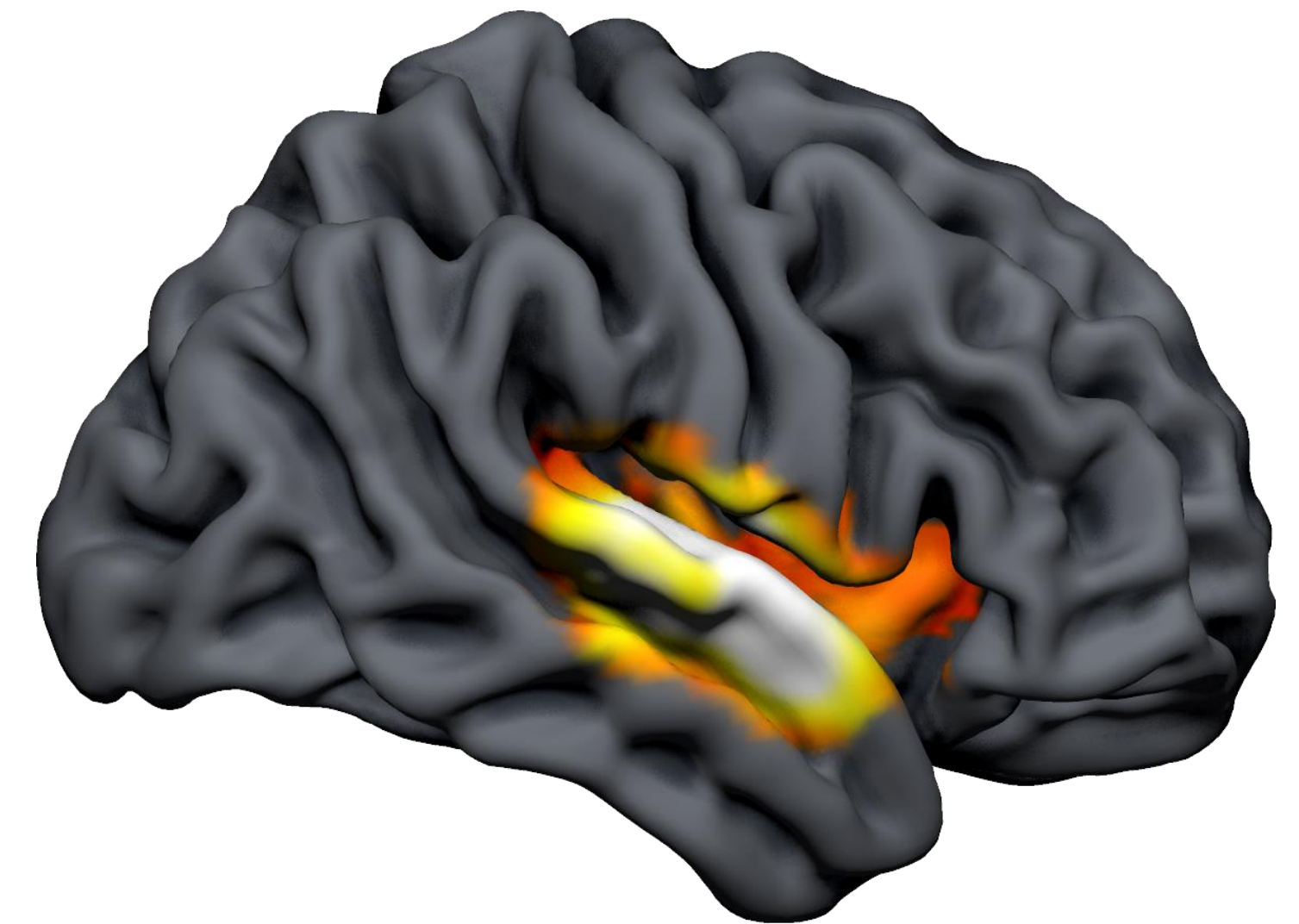
Individual foci



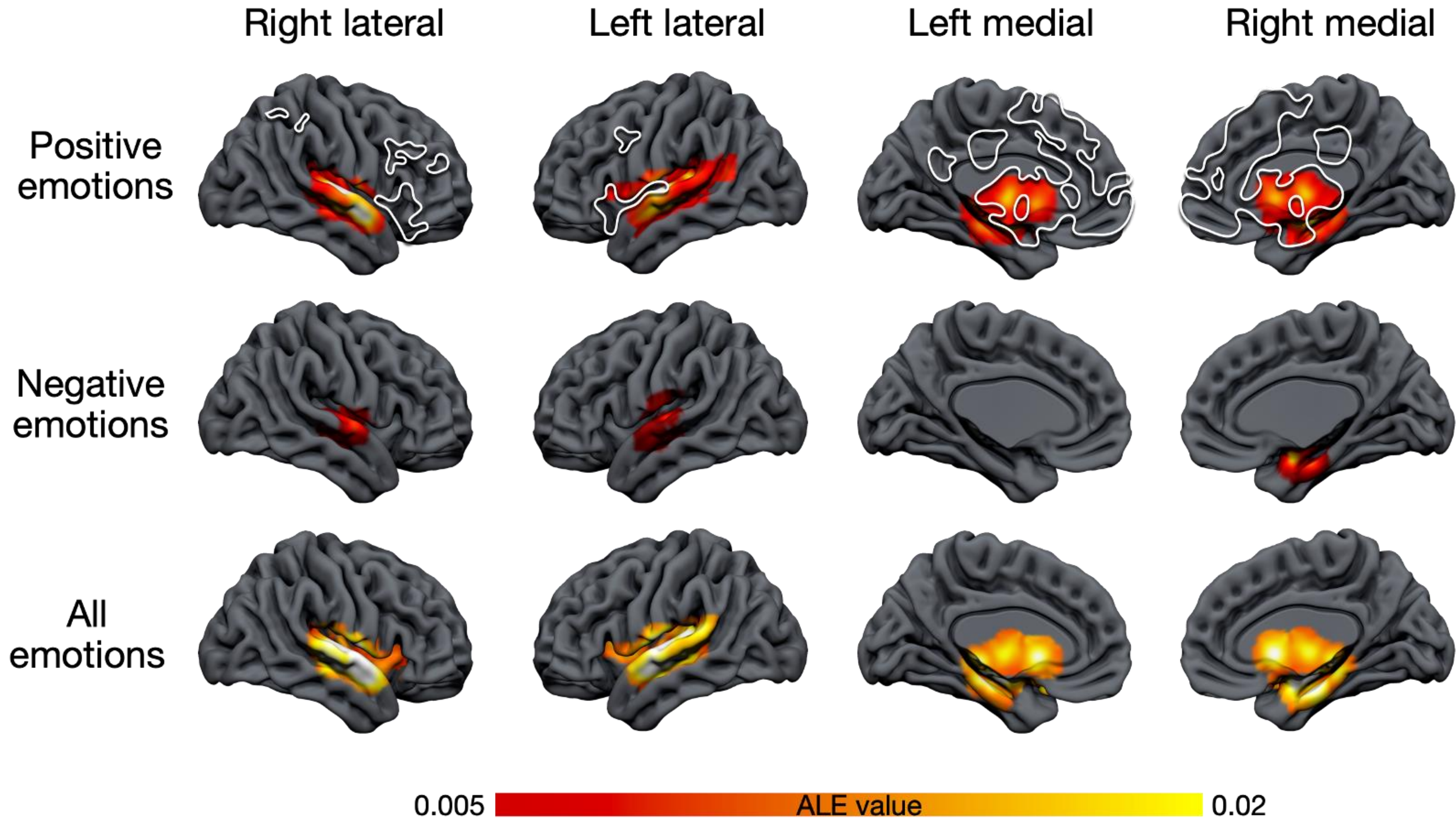
Permutations



Thresholding



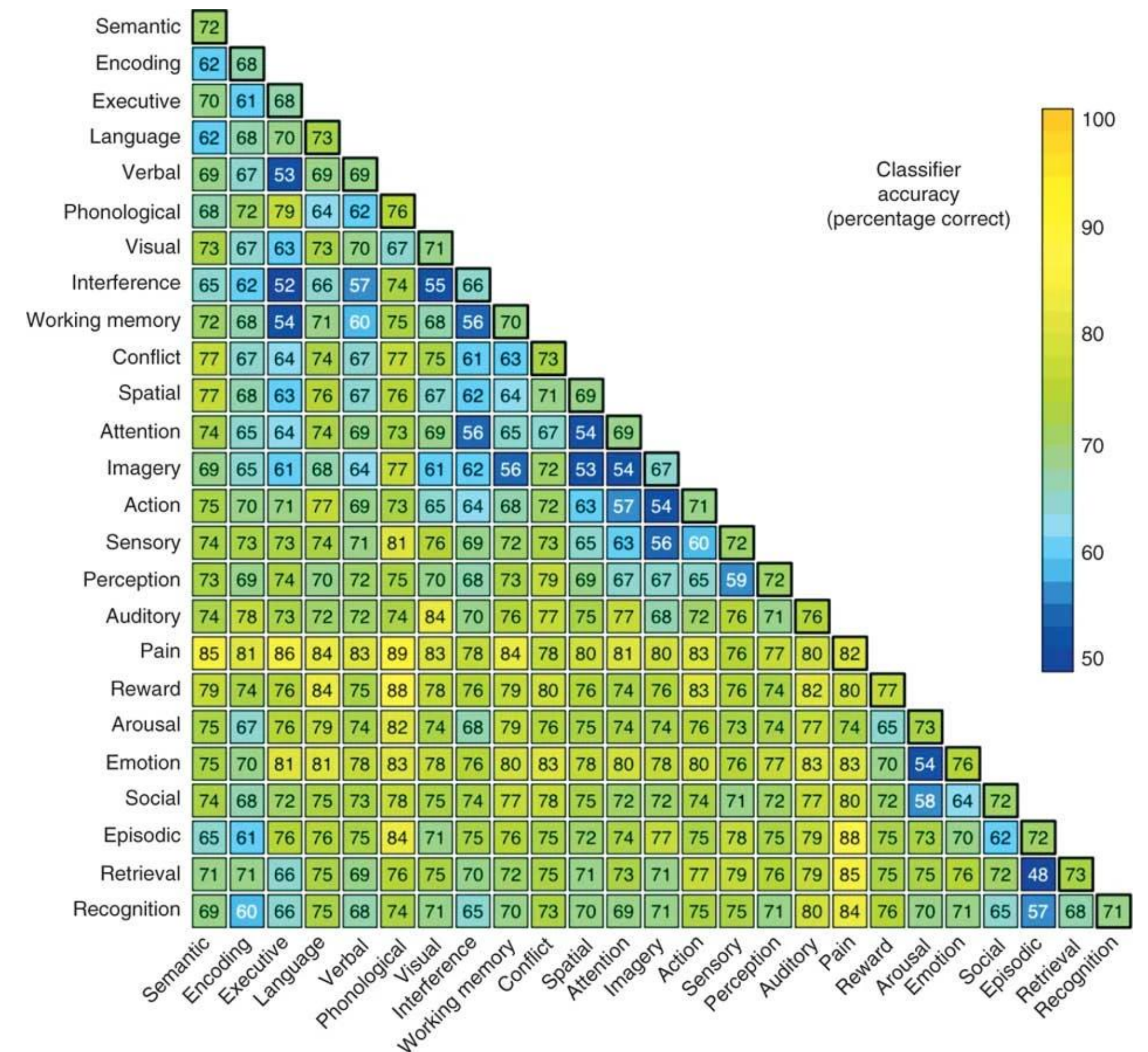
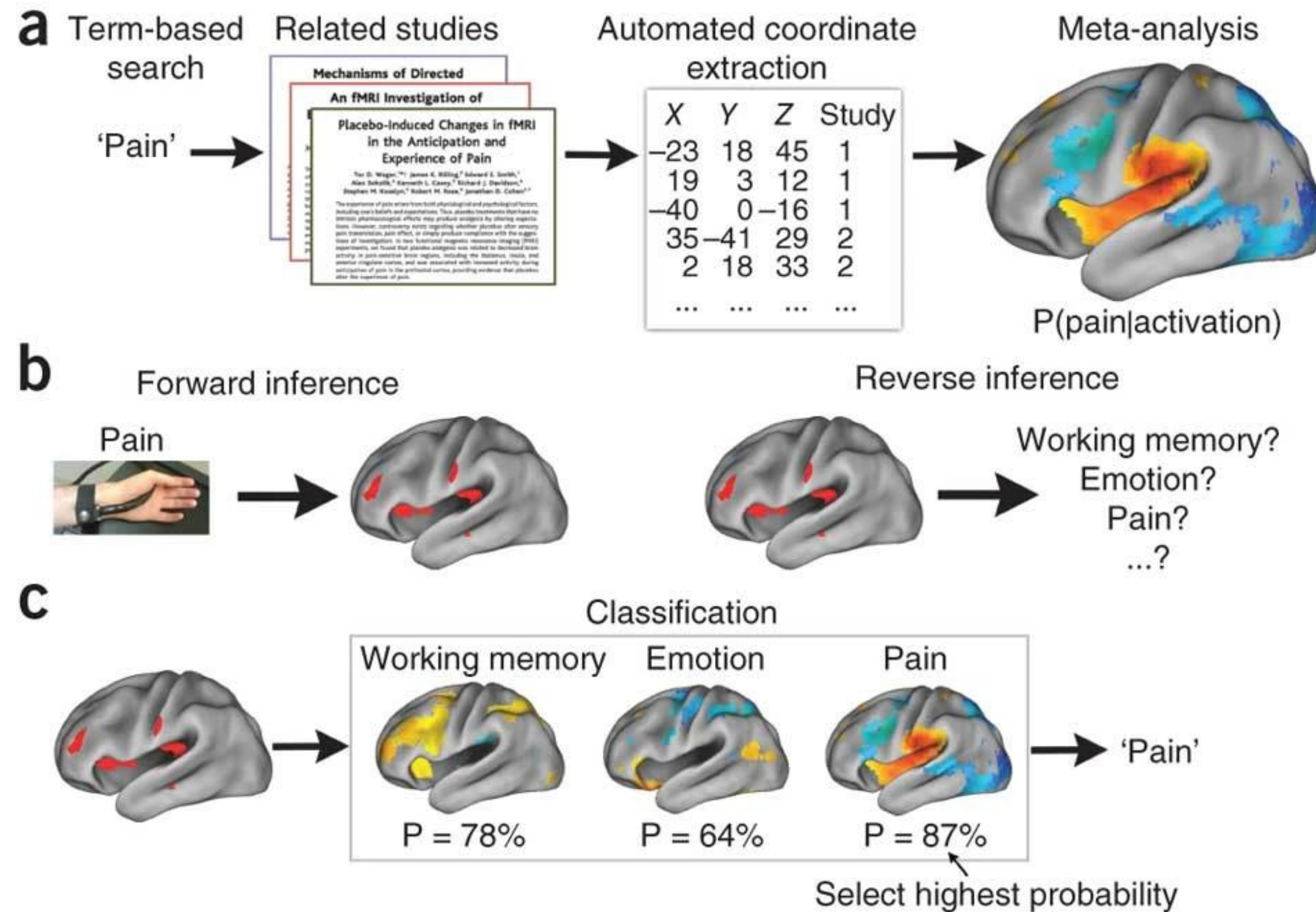
Convergence of activation locations at given threshold



ALE Pros and cons

- Relatively easy to analyze and interpret
- Full-volume analysis
- No need to worry about normalization
- Effect direction can be accounted for
- Effect sizes scaled only by sample size
- Requires coordinate-levels data
- Data modelled per peaks —> cluster size not taken into consideration

Approach 3: Automated data mining

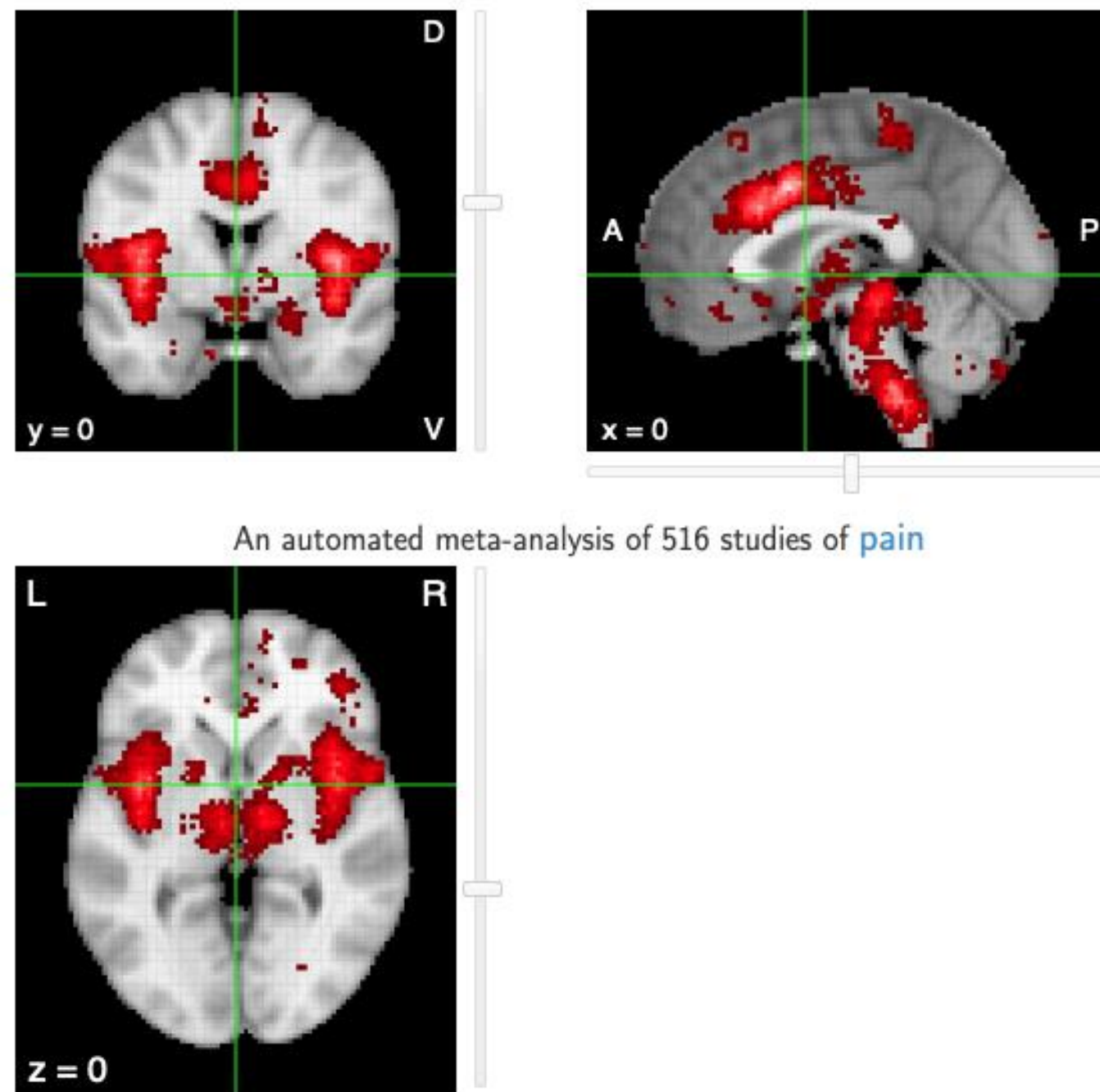


NeuroSynth (Yarkoni et al, 2011)

neurosynth.org

Neurosynth is a platform for large-scale, automated synthesis of functional magnetic resonance imaging (fMRI) data.

It takes thousands of published articles reporting the results of fMRI studies, chews on them for a bit, and then spits out images that look like this:



- Keyword-based search
- Precompiled meta-analytic maps based on automatically parsed data
- Result maps can be downloaded as 3D nifti files for further analysis
- Custom analyses and queries possible

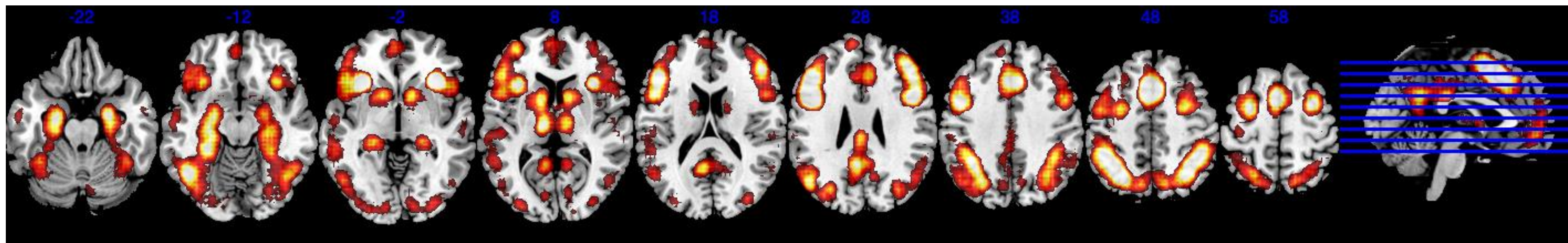
Database Status

507891 activations reported in [14371 studies](#)

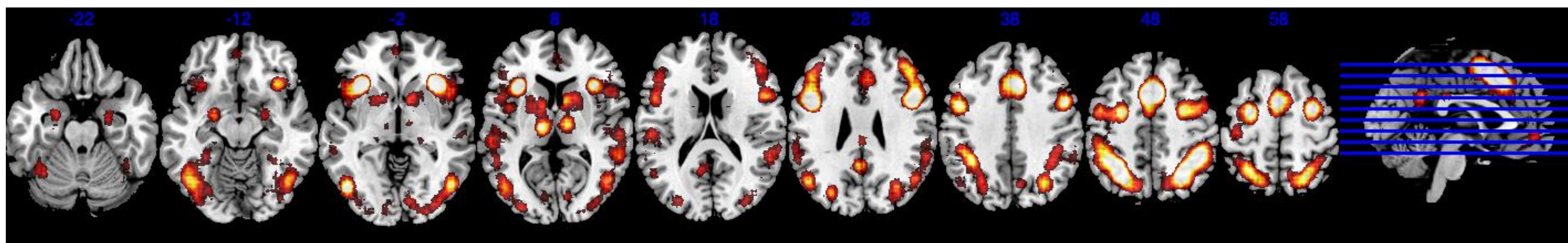
Interactive, downloadable meta-analyses of [1334 terms](#)

Functional connectivity and coactivation maps for over [150,000 brain locations](#)

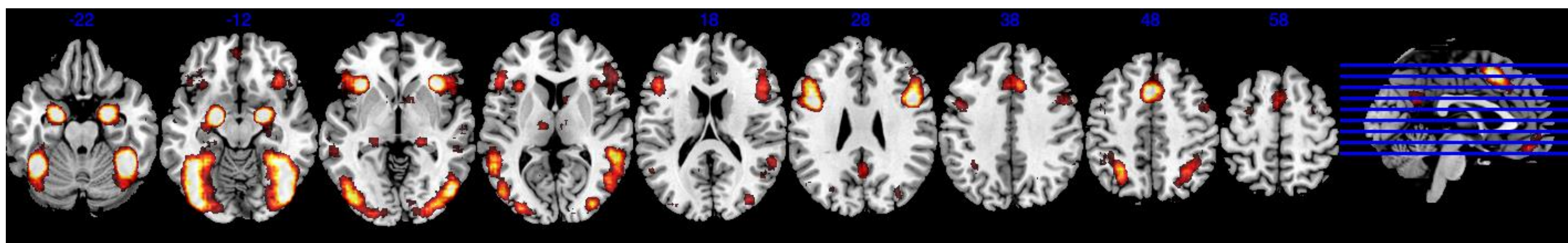
Memory



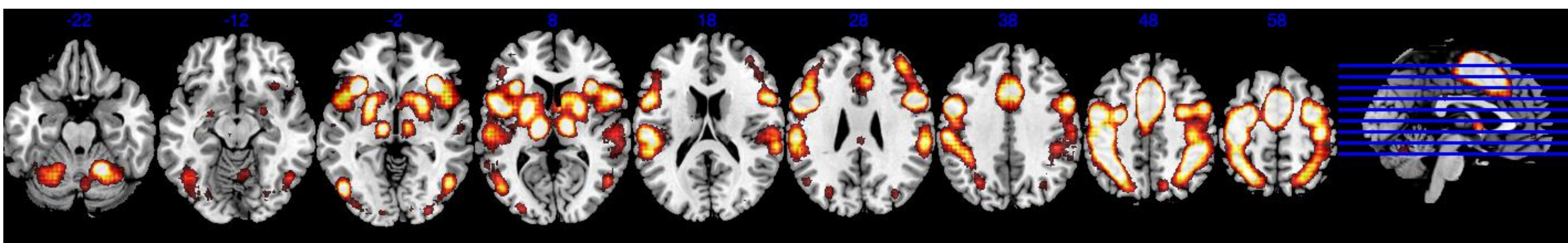
Attention



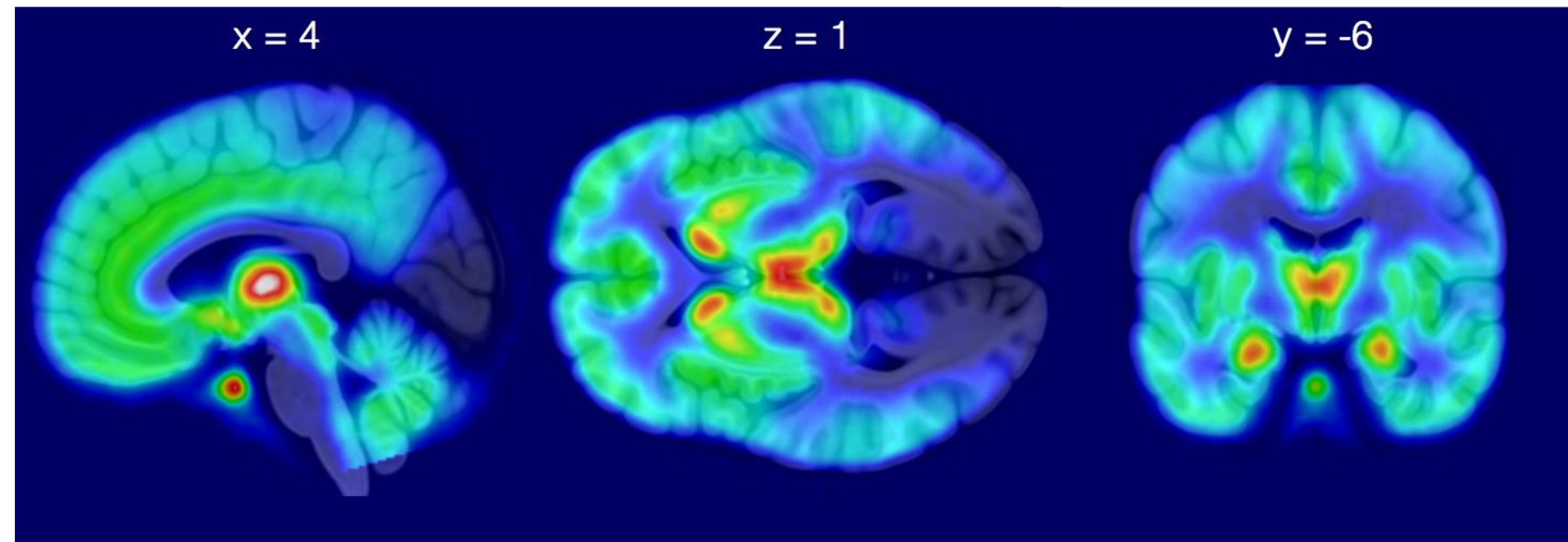
Face



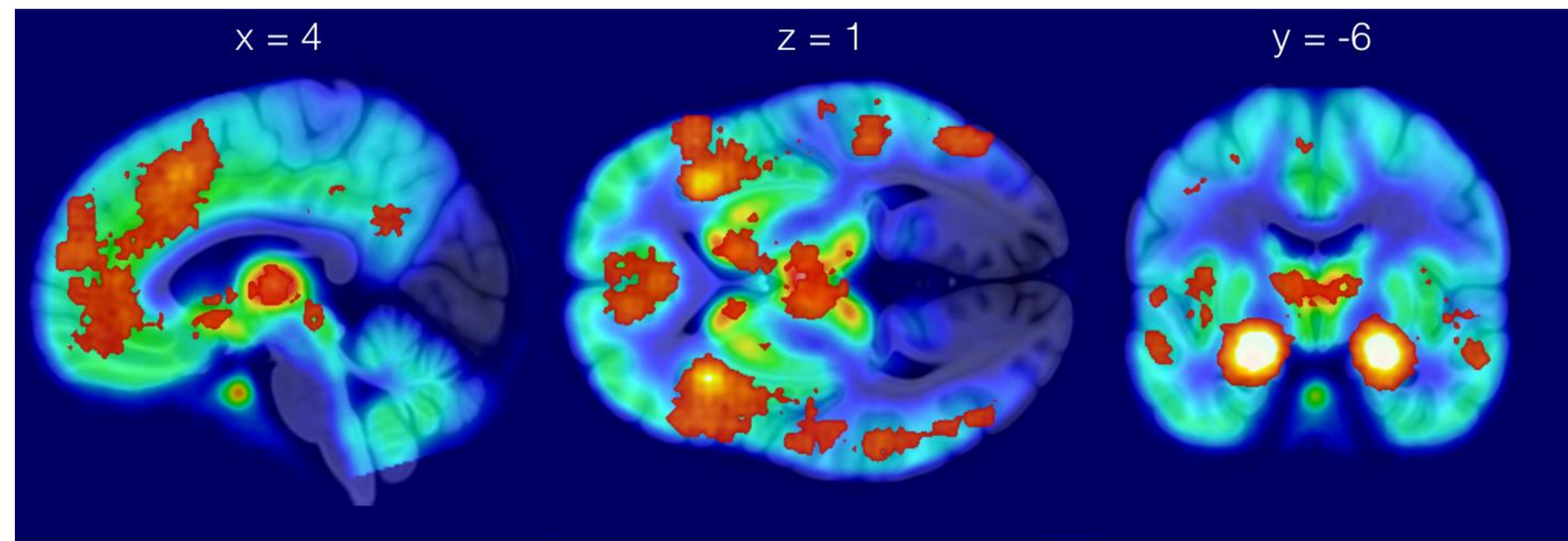
Motor



A) Distribution of μ -opioid receptors in the brain as measured with [^{11}C]carfentanil PET



B) Overlap between human emotion circuit and the μ -opioid receptor system



0 BP_{ND} 3.5 FDR Z-score 20

$$r_{\text{all}} = 0.38$$
$$r_{\text{pleasure}} = 0.44$$

Neurosynth: Pros and cons

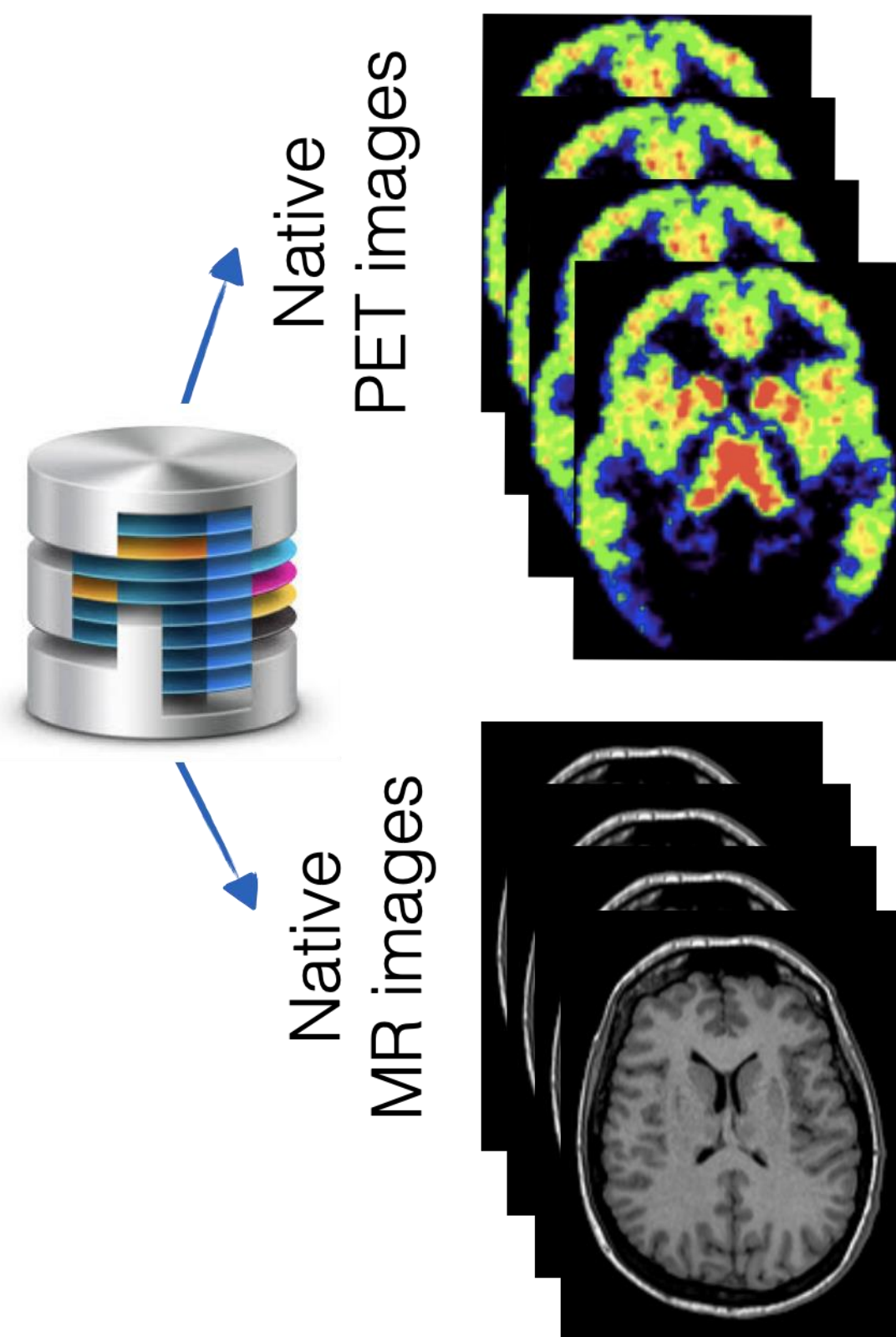
- Very easy to analyze and interpret
- Data readily available, allows custom analyses
- Full-volume analysis
- If large n, correlates well with ALE on manual extracted peaks
- Quality contingent on the parser & reporting in studies
- Currently distinguishing activation / condition direction difficult

Solution 2: Large-scale synthesis of old datasets

- Between-study variability and reliance on statistical estimates (rather than raw data) lower the power of meta-analysis
- Existing data are often available and cheap to use given permissions can be reanalysed (in Finland with “Rekisteritutkimus”)
- Data however have to be extracted, reprocessed and the metadata needs to be extracted

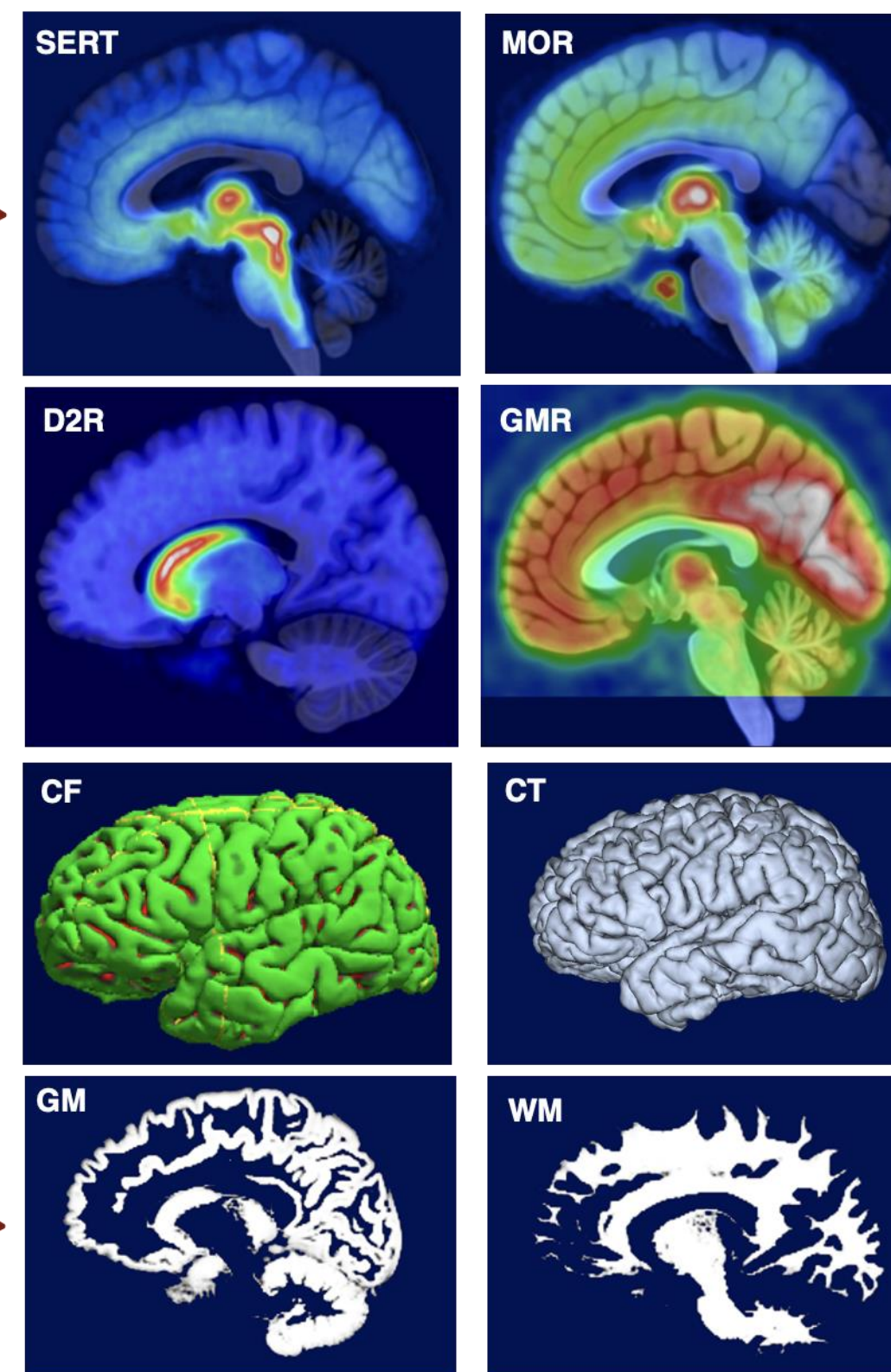
Integrated approach at PET Centre

Hospital PACS

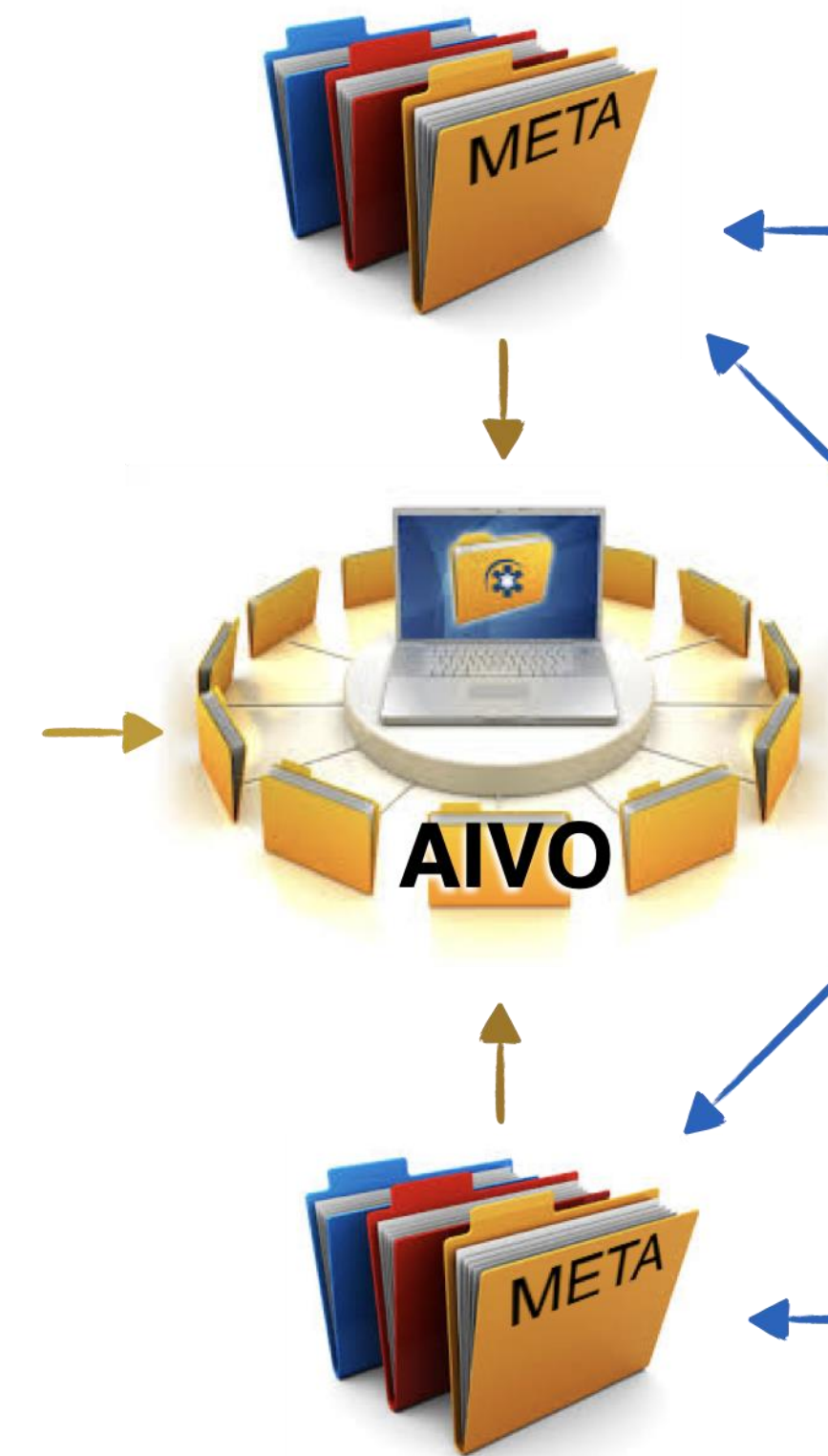


AUTOMATED KINETIC MODELLING, SEGMENTATION AND MESH GENERATION

Preprocessed **BRAIN** data



AIVO database



EXTERNAL register sources

SOMATIC WELL BEING

1. ICD codes for diagnoses
2. Laboratory results
3. Frequency of hospitalisation & sickness leaves

PSYCHOLOGICAL FUNCTIONING

1. Psychopathology (ICD codes)
2. Personality structure
3. Disorders of cognition (ICD codes)
4. Fluid intelligence and school achievement

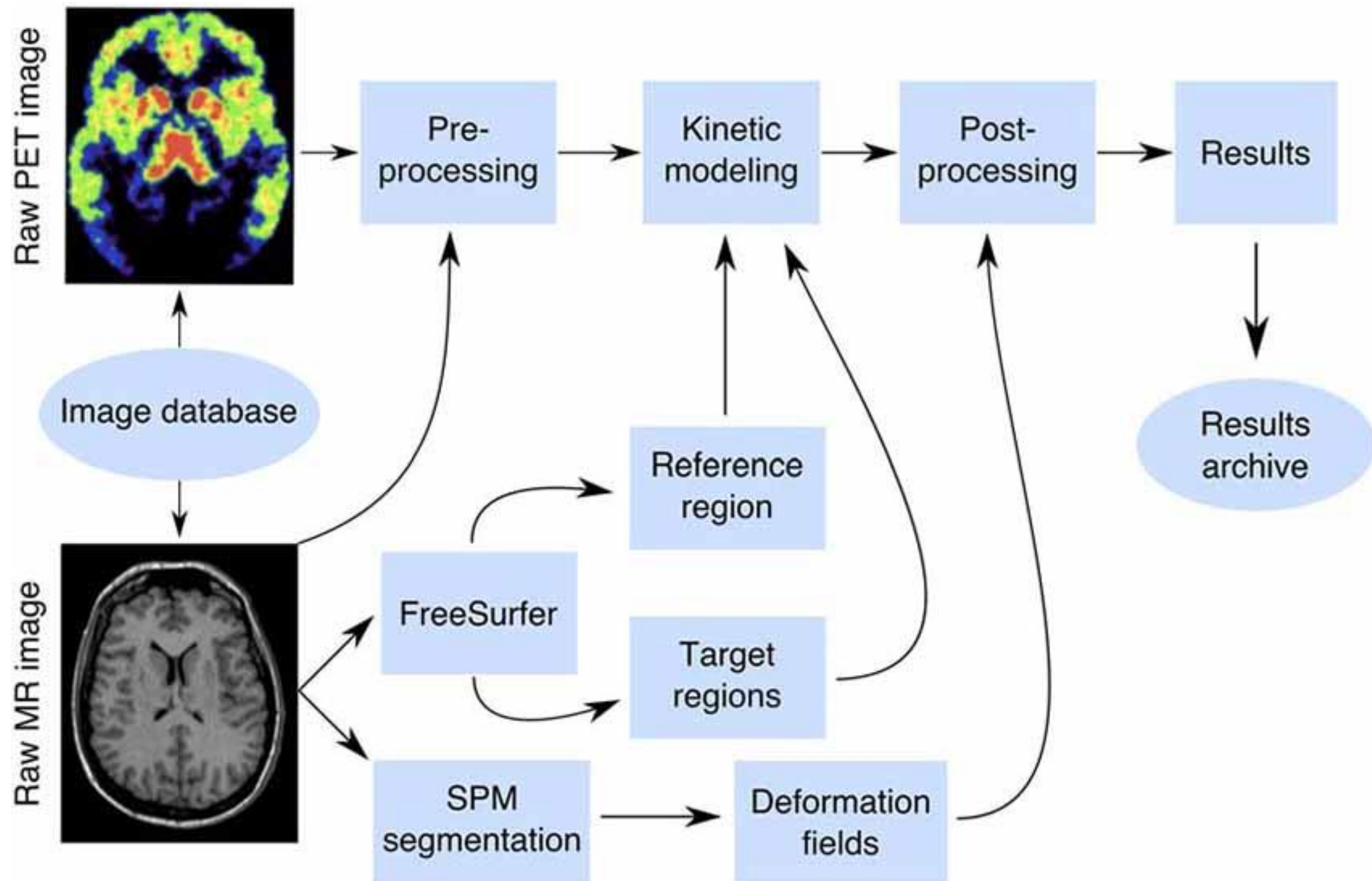
SOCIOECONOMIC WELL BEING

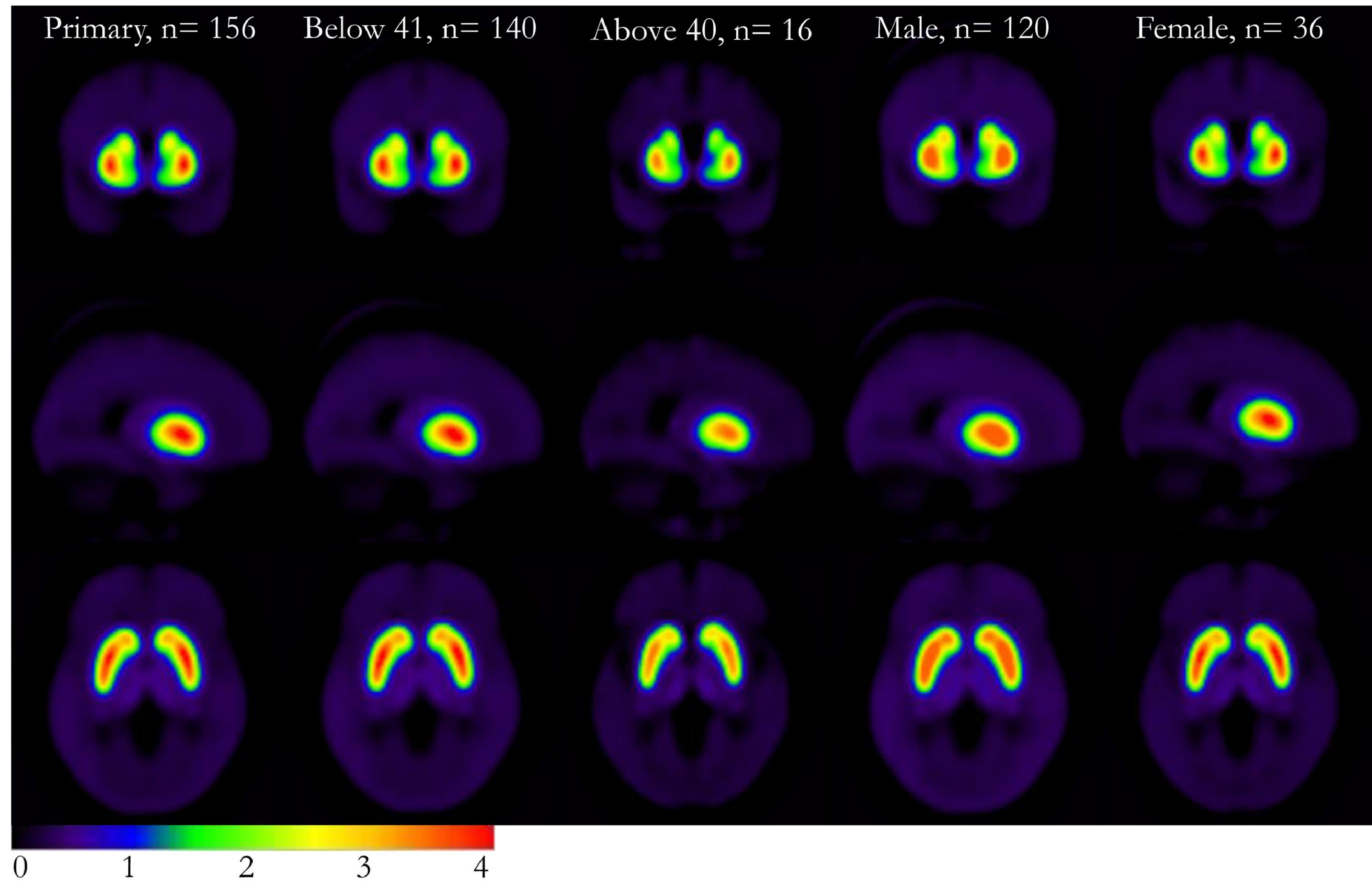
1. Earning and income transfers
2. Education and social status
3. Labour market attachment
4. Developmental socioeconomic stressors

SOCIAL ATTACHMENT BEHAVIOUR

1. Marriage and cohabitation
2. Family establishment
3. Reproduction and family size

Automated and supervised quality control





Malen et al (2022 Neuroimage)



NEUROVAULT

A public repository of unthresholded statistical maps, parcellations, and atlases of the brain.

What is it?

A place where researchers can publicly store and share unthresholded statistical maps, parcellations, and atlases produced by MRI and PET studies.

Why use it?

- Interactive visualization
- A permanent URL
- Publicly shareable
- Improves meta-analyses

Supported by



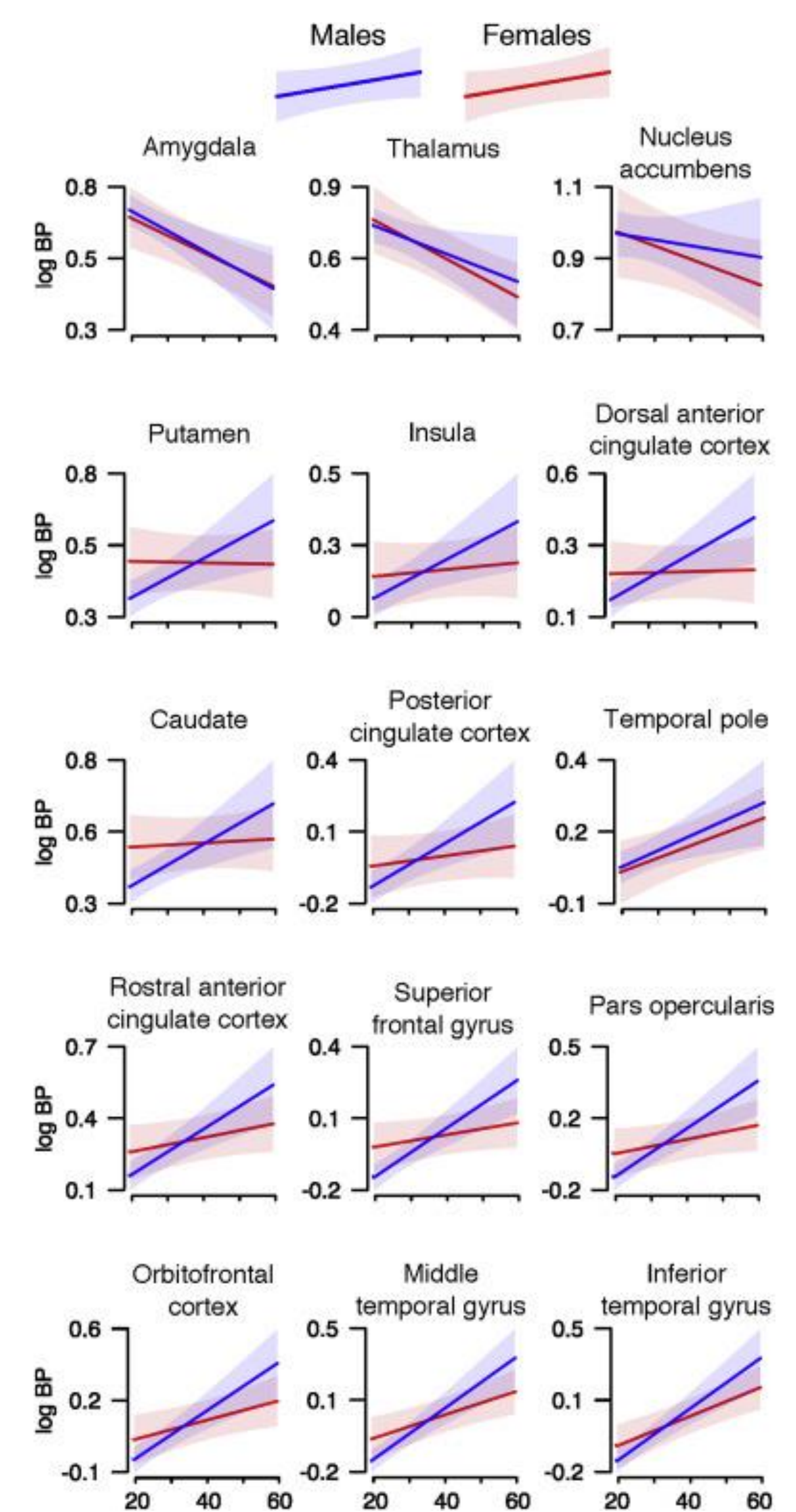
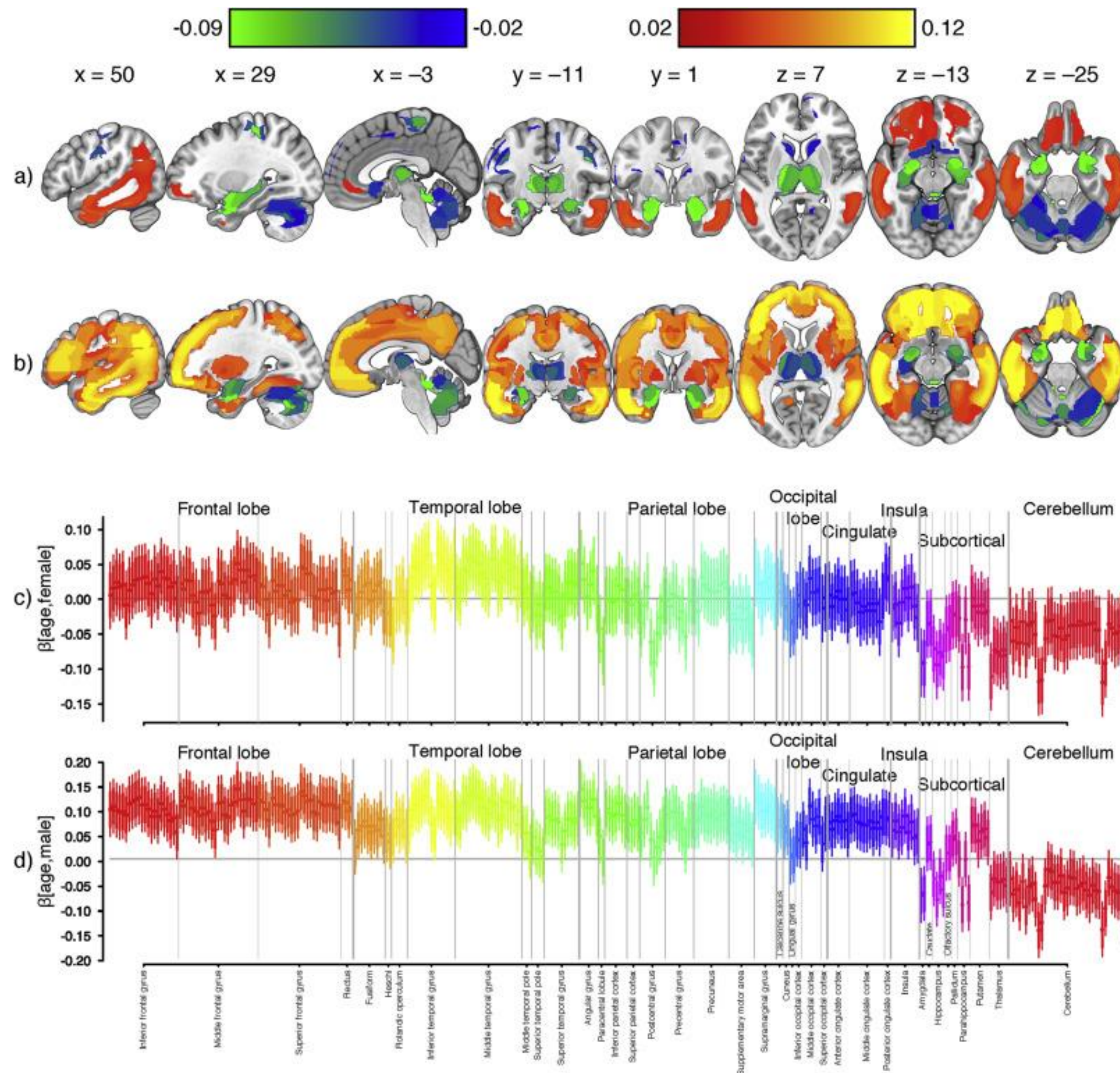
Stanford
University



[Get started and upload an image!](#)

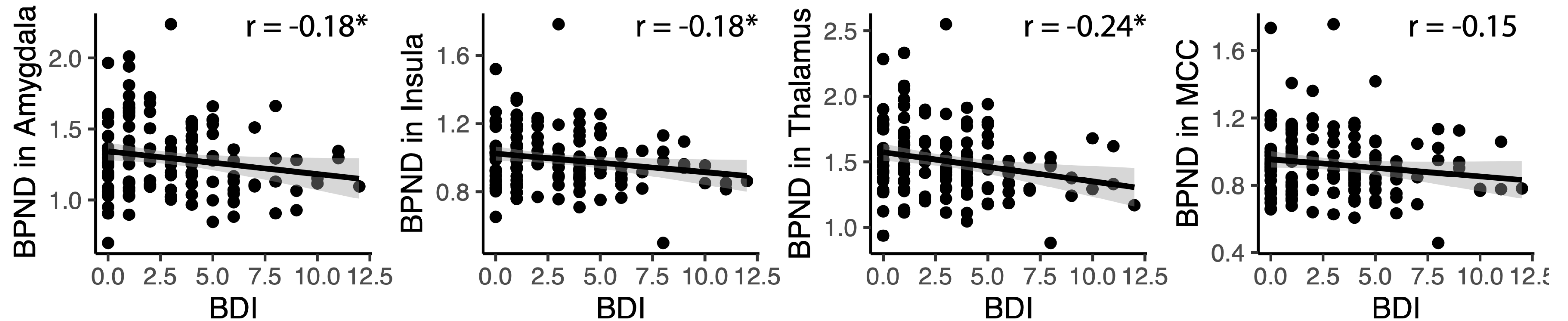
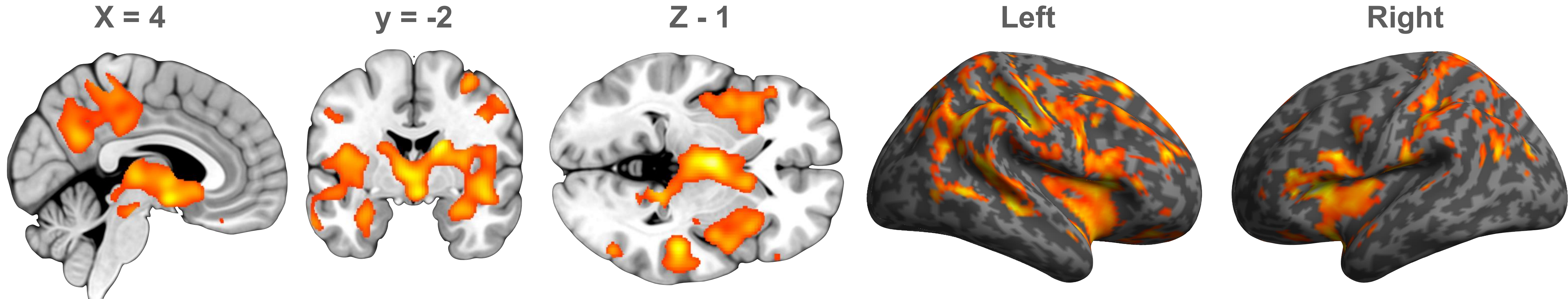
Recently added collections of images from published papers

Name	Number of images
Genetic, cellular, and connectomic characterization of the brain regions commonly plagued by glioma	1
Principal component analysis reveals multiple consistent responses to naturalistic stimuli in children and adults	22
Attention- versus significance-driven memory formation: Taxonomy, neural substrates, and meta-analyses	3
A dynamic gradient architecture generates brain activity states	12
Towards the Interpretability of Deep Learning Models for Multi-modal Neuroimaging: Finding Structural Changes of the Ageing Brain	27
Whole-brain functional correlates of memory formation in mesial temporal lobe epilepsy	48
HIV infection is linked with reduced error-related default mode network suppression and poorer medication management abilities	6
What Executive Function Network is that? An Image-Based Meta-Analysis of Network Labels	8
Reinforcement learning with associative or discriminative generalization across states and actions:	49
Residential green space and air pollution are associated with brain activation in a social-stress paradigm	17



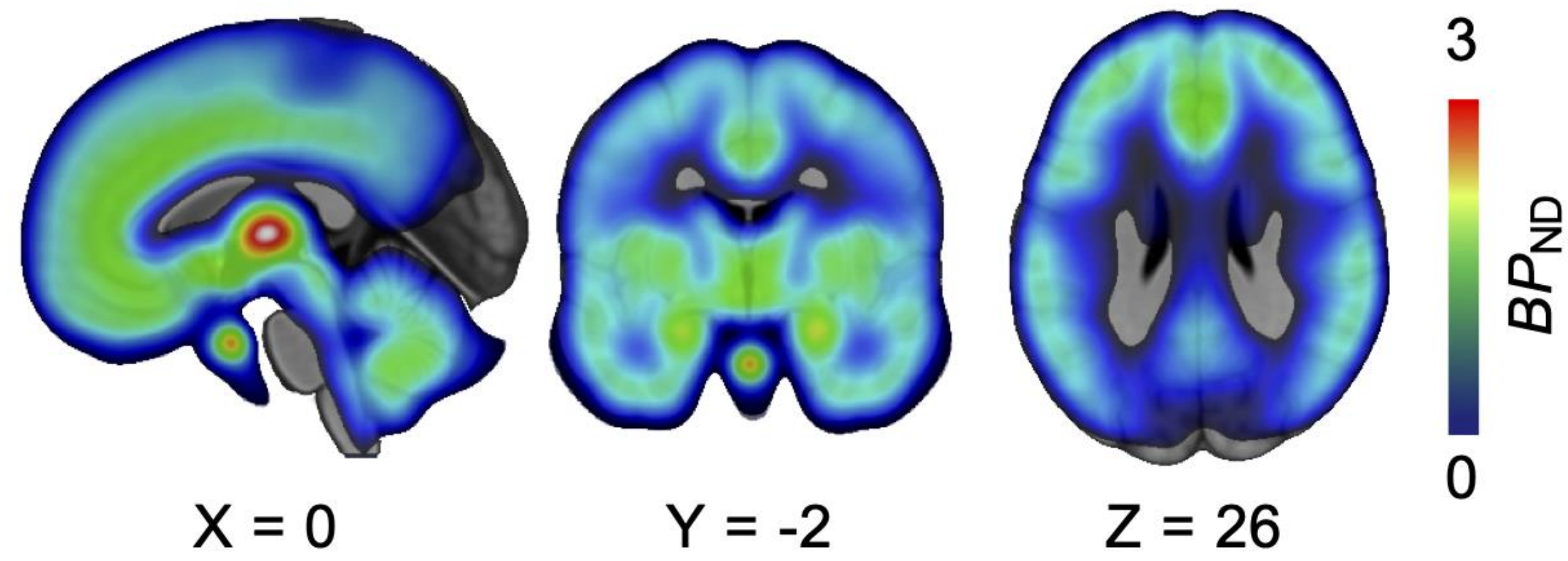
Kantonen et al (2020 NeuroImage)

Lowered mu-opioid receptor availability in subclinical depression and anxiety

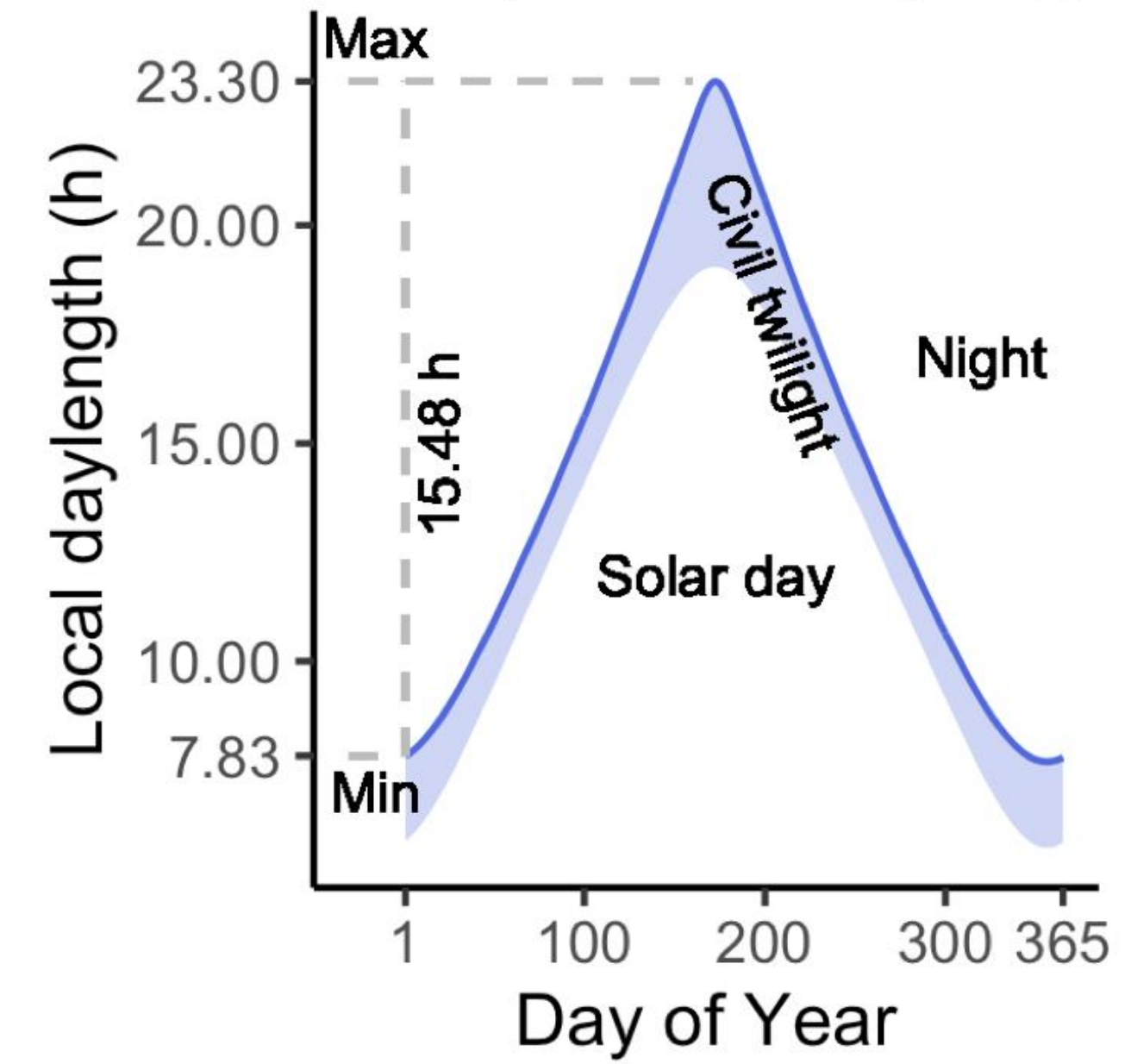


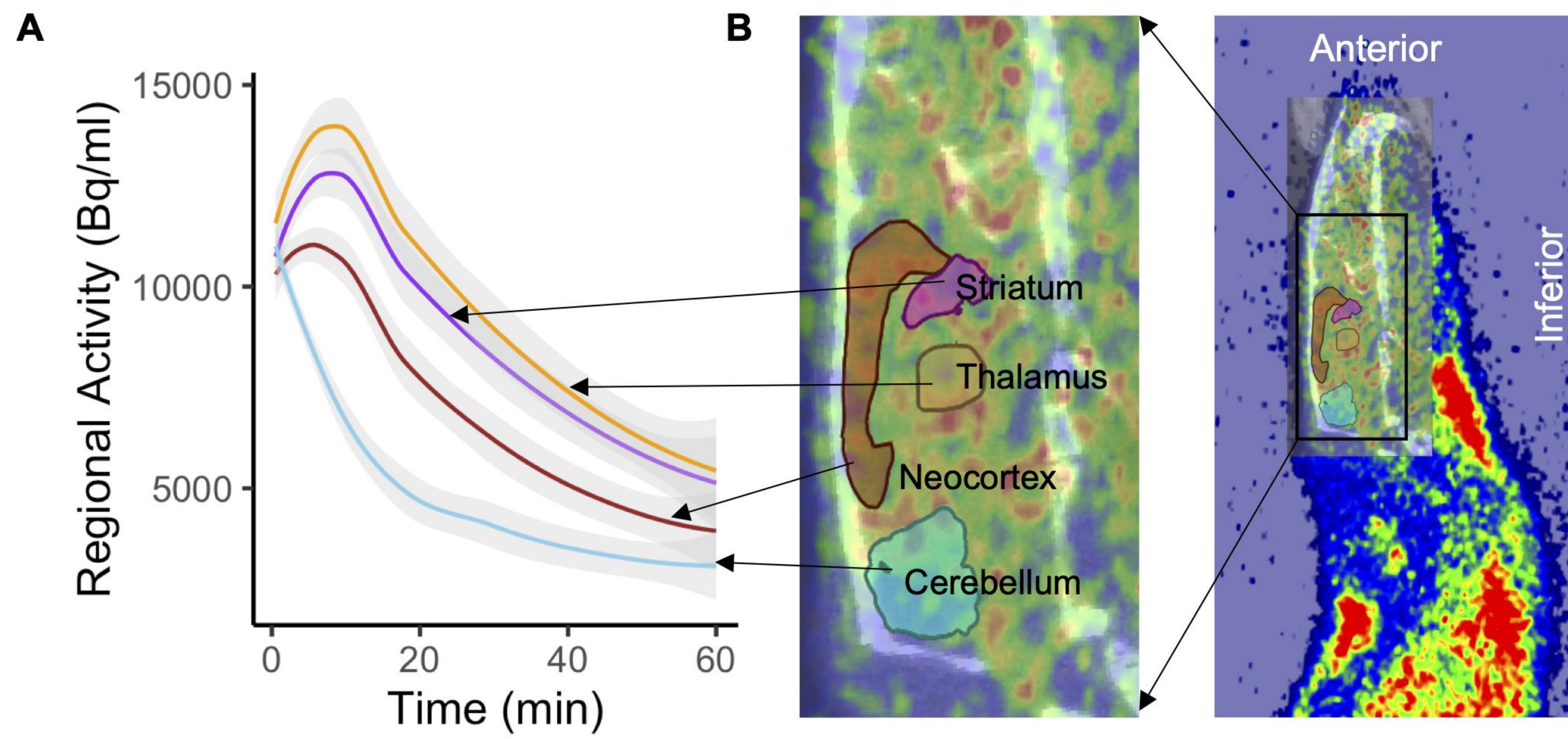
FDR T-score 4

A Mean distribution of MOR (n = 204)



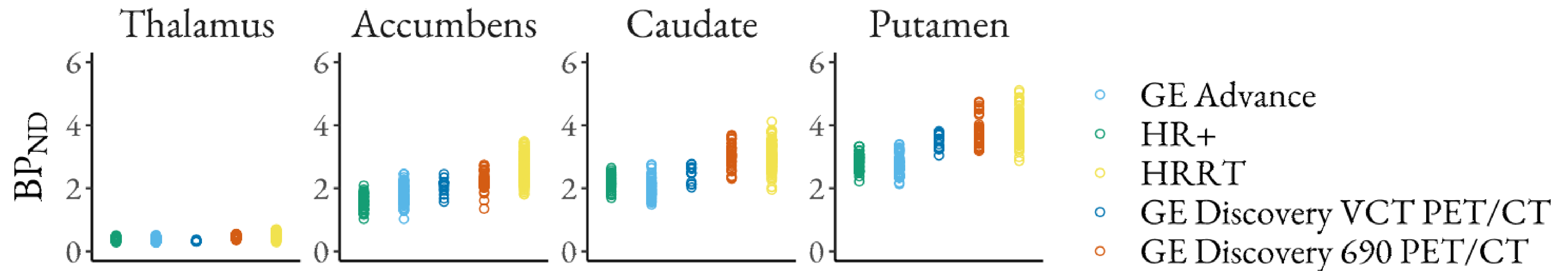
B Seasonal profile of daylength



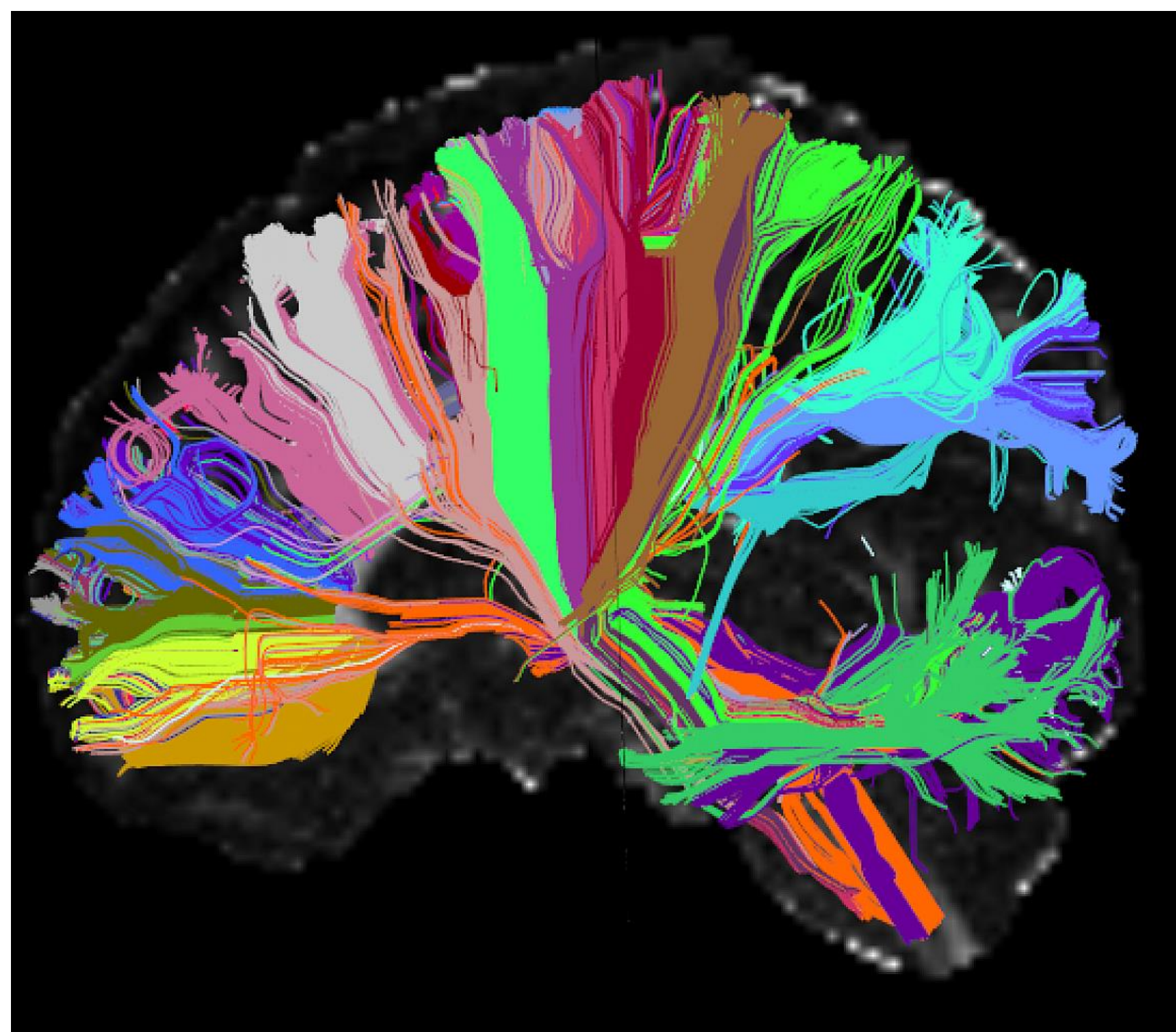


Common problems with data integration

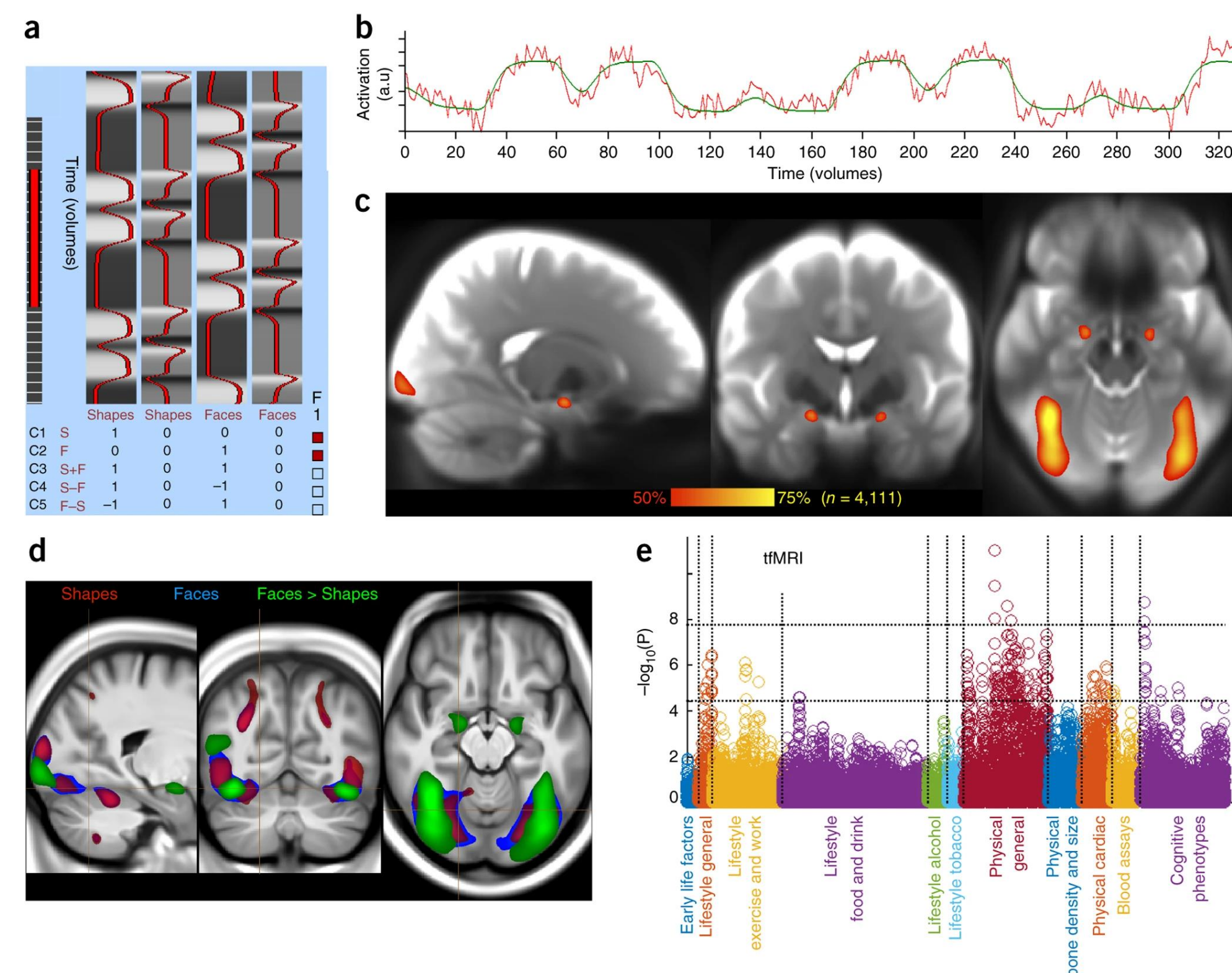
- Variable imaging equipment
- Standardization of data acquisition
- Metadata description
- Processing pipelines
- Comparability of conditions
- Specificity of effects



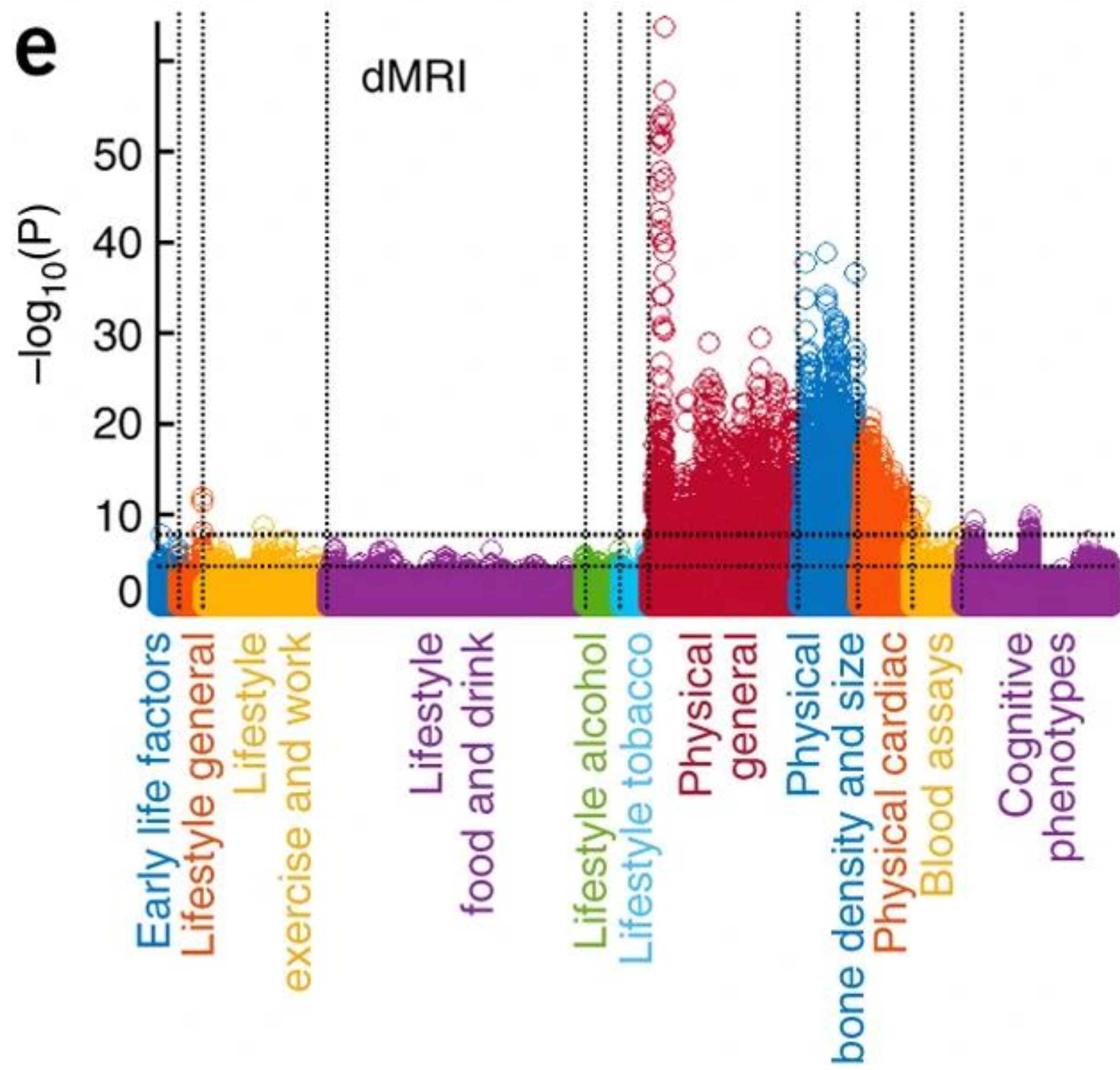
Solution 3: More is more in the first place



Human Connectome Project



UK Biobank



Comparison of the approaches

	Specificity	Price	Computational demands
Meta-analysis	Low-medium	Low	Low
Retrospective reuse	Medium	Medium-high	Moderate
Dedicated large-scale study	High	High-stratospheric	High

