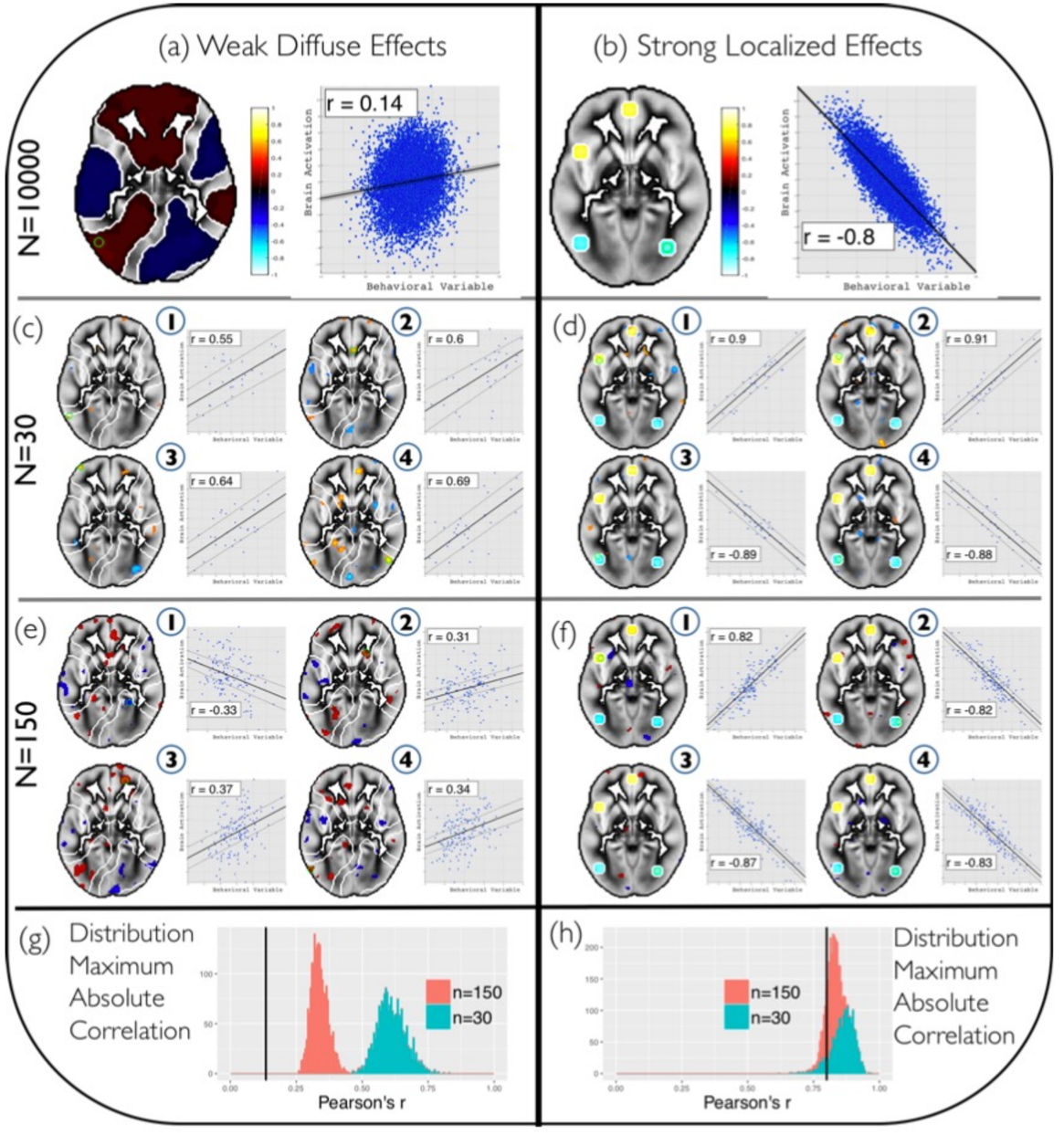
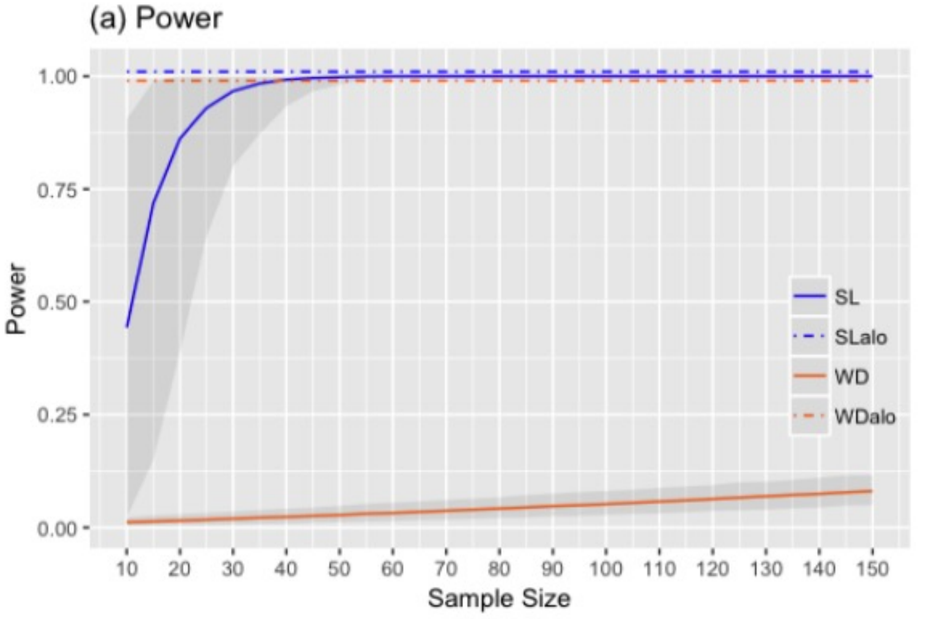
Neuroinformatics and large-scale analysis

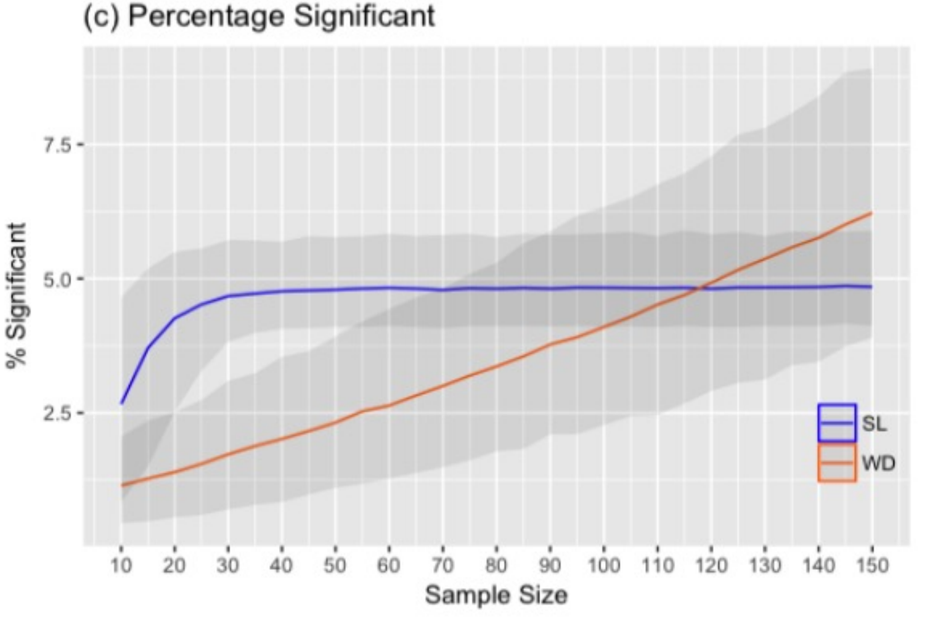
Lauri Nummenmaa
Turku PET Centre / TYKS

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.







Cremers et al (2017)

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

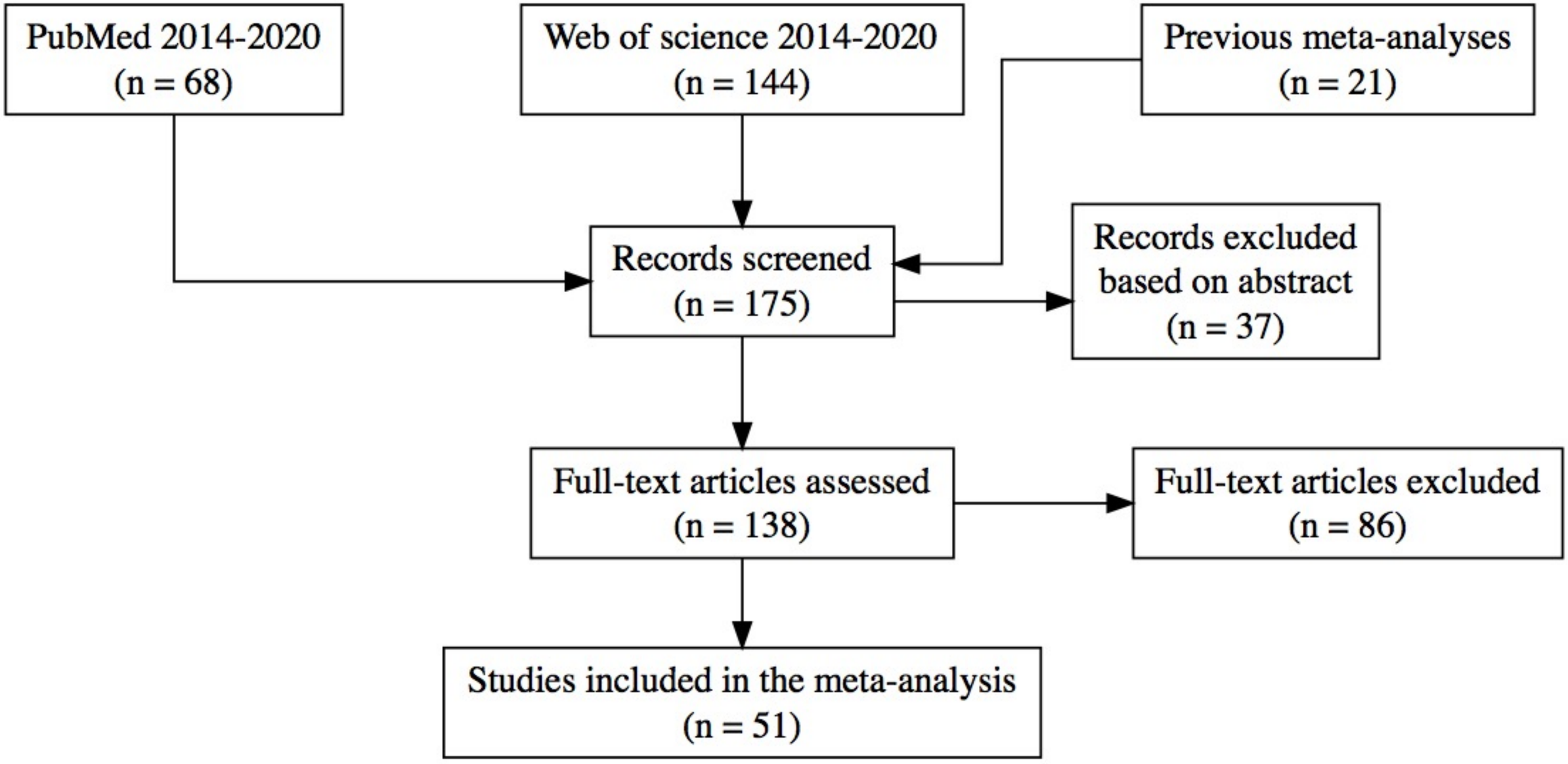
- Is everything done **lege artis**?
- How can we **reproduce** our results?

Data synthesis

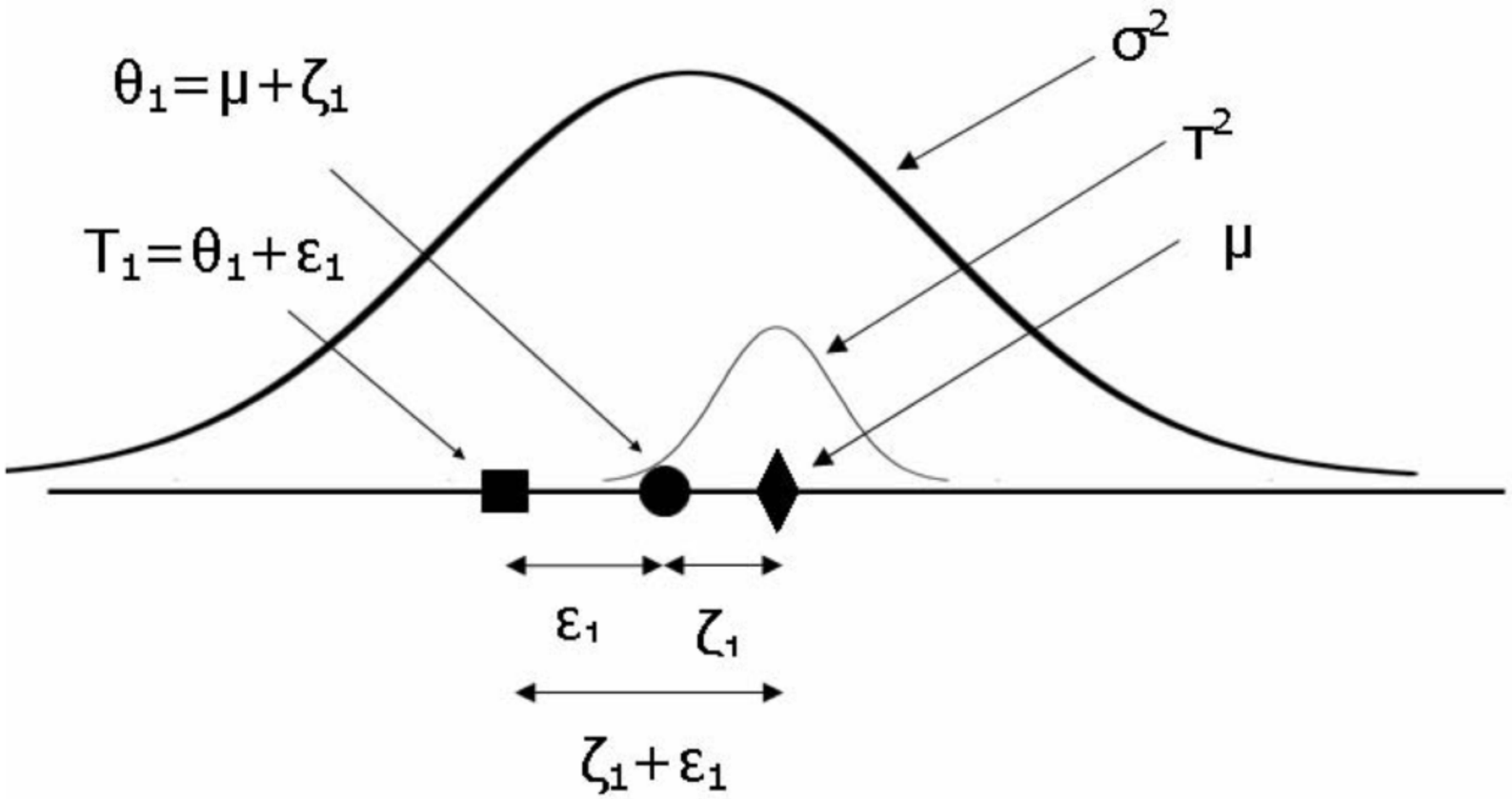
- How can we **combine** data?
- **What** can we combine?

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

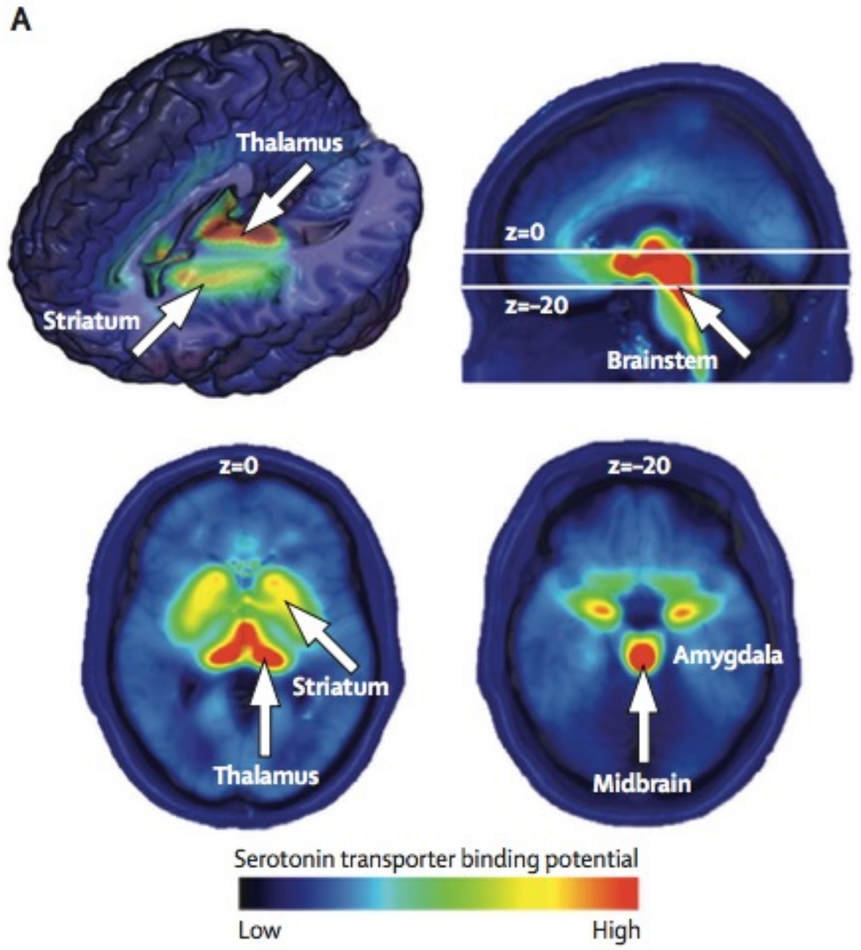

PubMed 2014-2020
(n = 68)
Web of science 2014-2020
(n = 144)
Previous meta-analyses
(n = 21)
Records screened
(n = 175)
Records excluded
based on abstract
(n = 37)
Full-text articles assessed
(n = 138)
Full-text articles excluded
(n = 86)
Studies included in the meta-analysis
(n = 51)

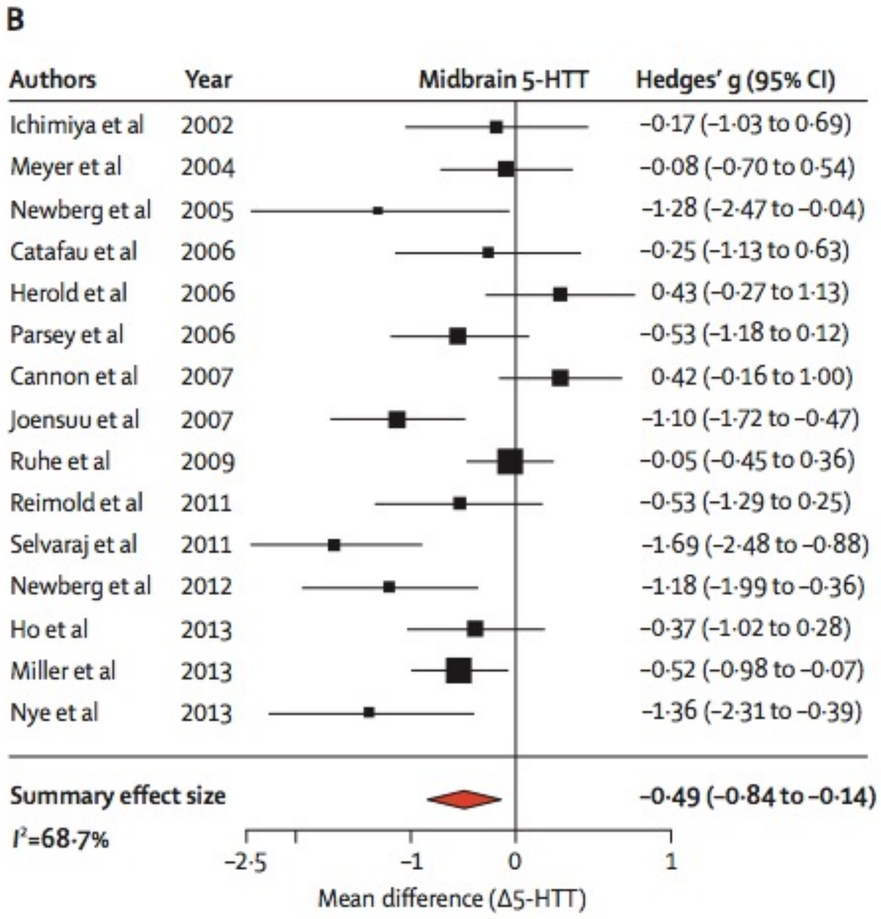
Classic mixed effects meta-analysis



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.

Approach 1: Regional analysis





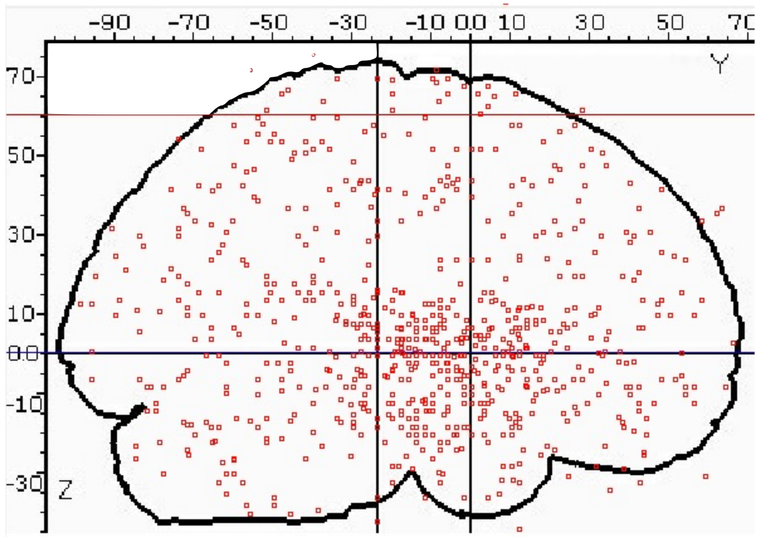
Spies et al (2015)

Regional analysis: Pros and cons

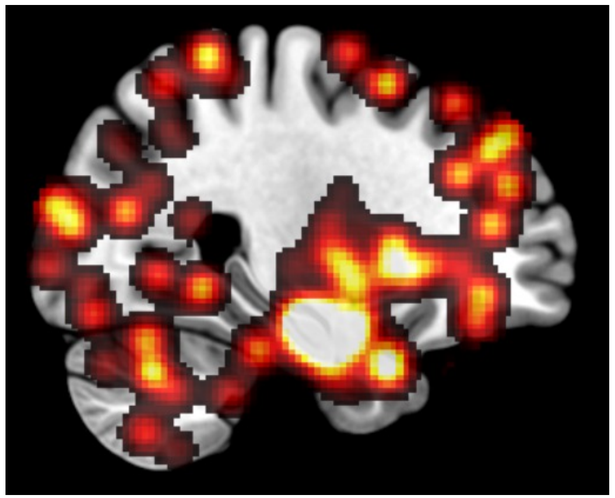
- Easy to analyze and interpret
- Data comparable in statistical terms
- No need to worry about normalization etc.
- Laborious
- Anatomical nomenclature not consisten
- May miss effects outside chosen ROIs

Approach 2: Peak-based analysis

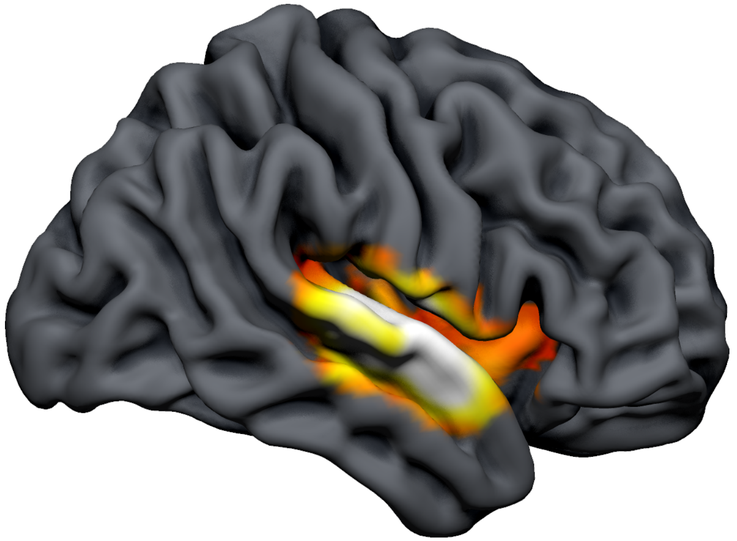
Individual foci



Permutation

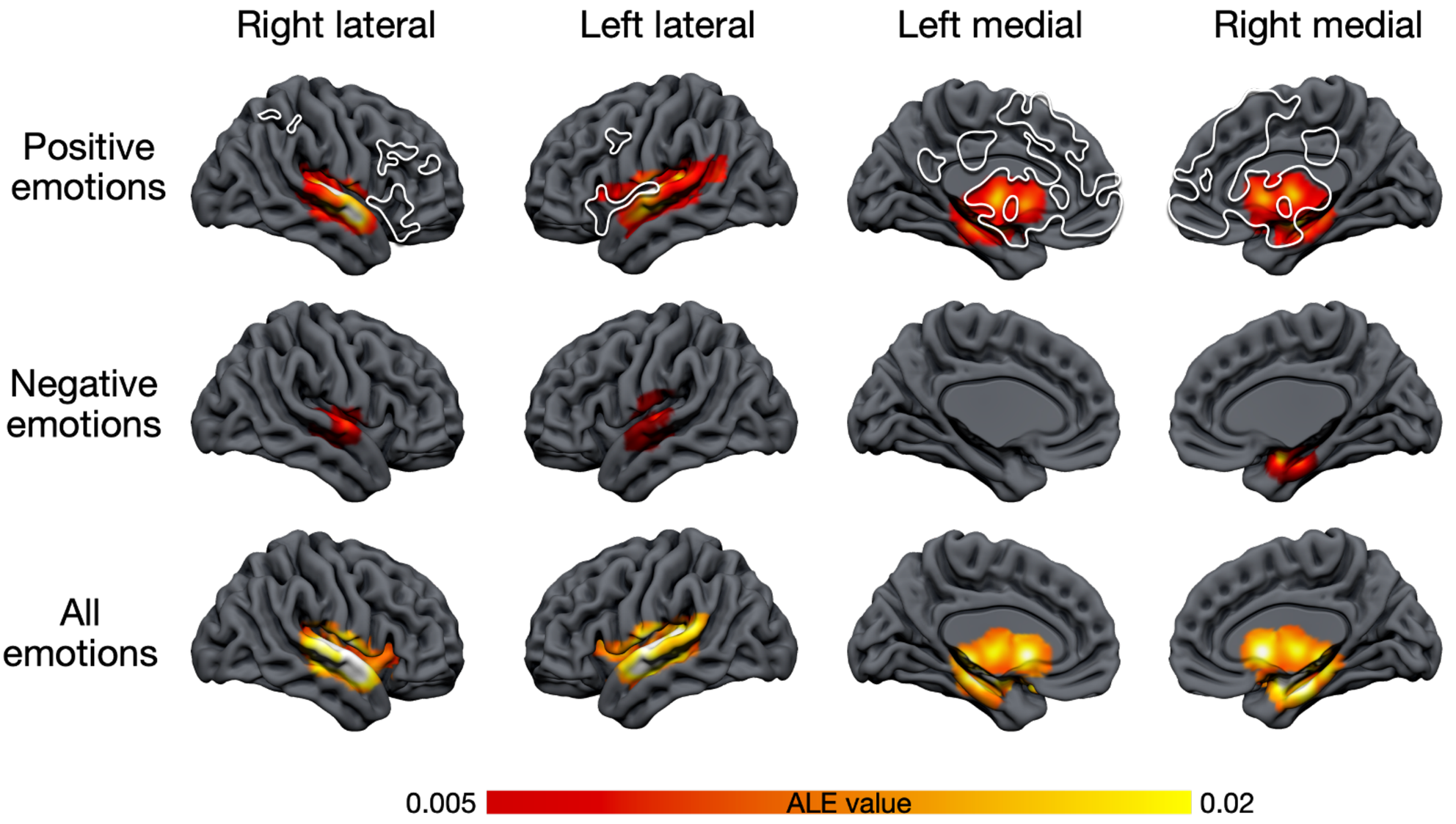


Thresholding



Convergence of activation locations at given threshold

Activation Likelihood Estimation (Eickhoff et al 2015)

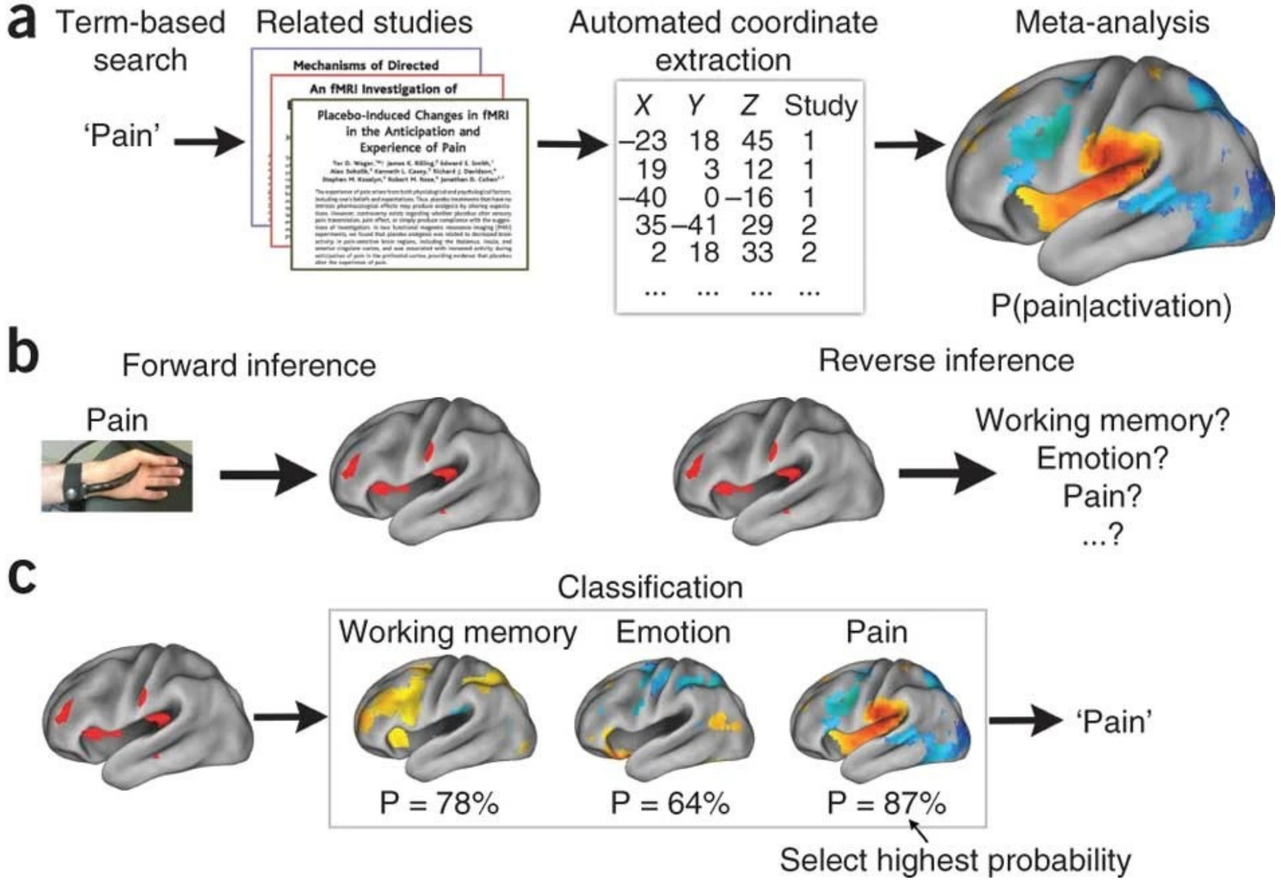


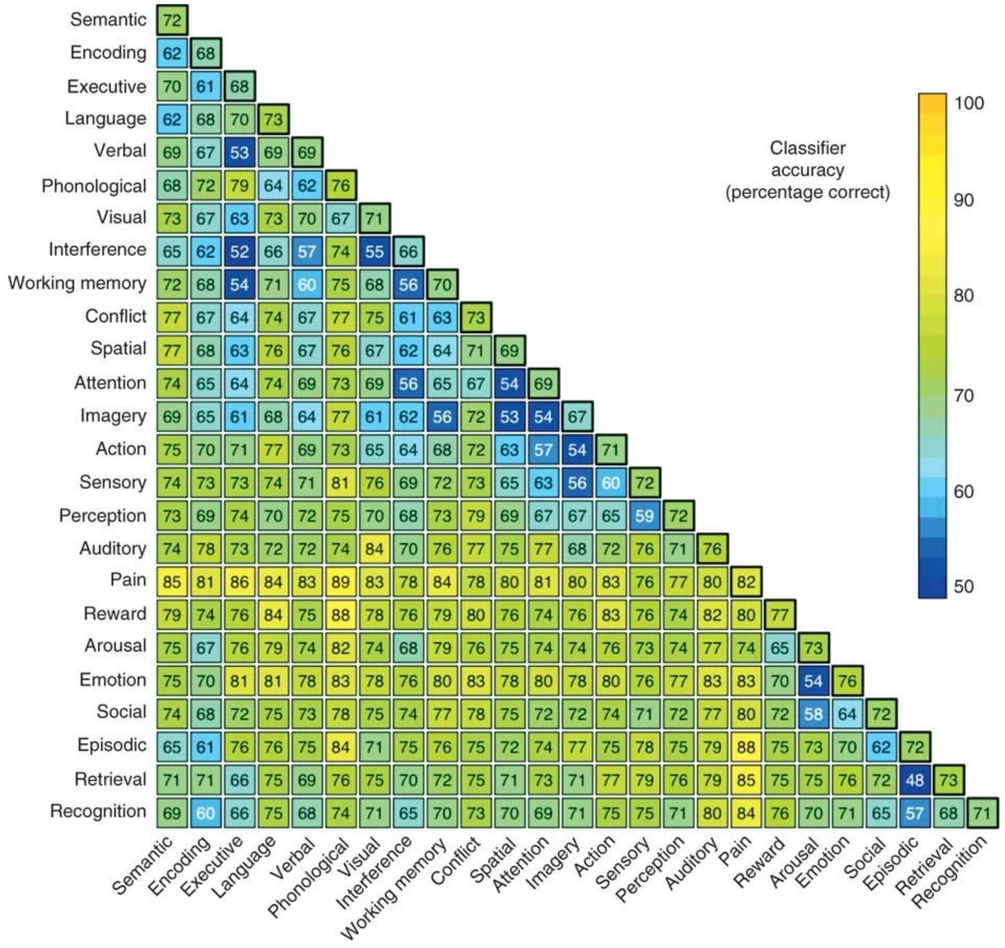
Nummenmaa, Putkinen & Sams (COiBS 2021)

ALE Pros and cons

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

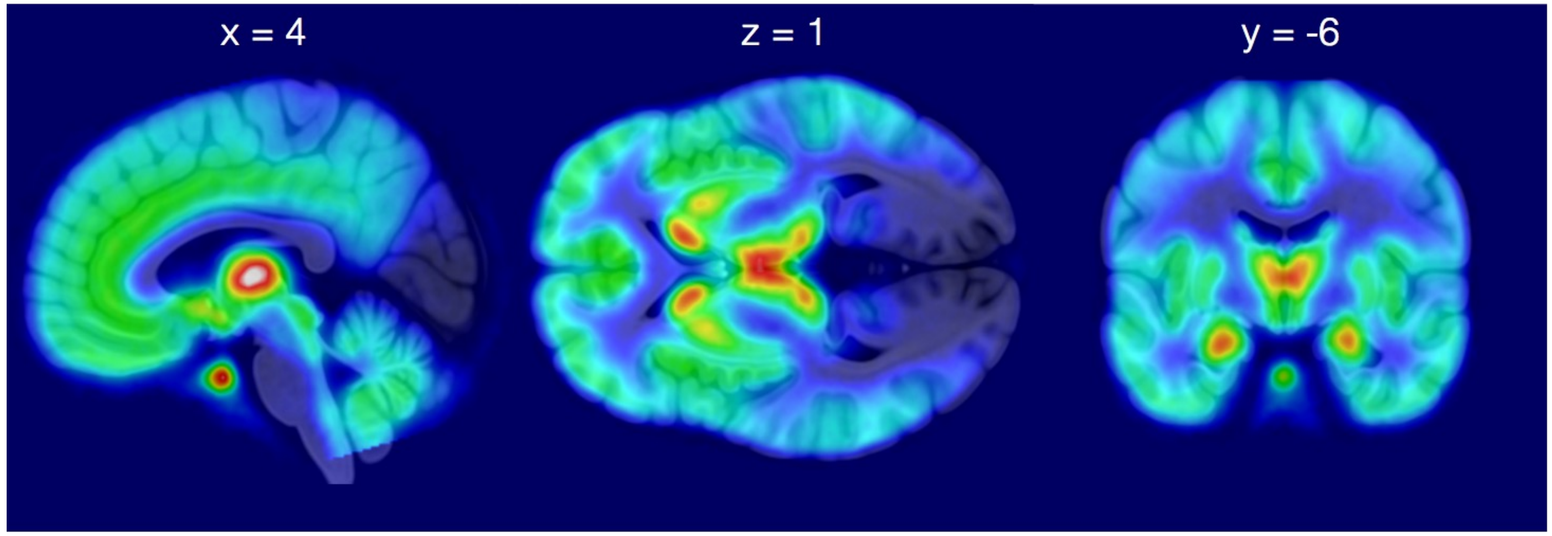
Approach 3: Automated data mining



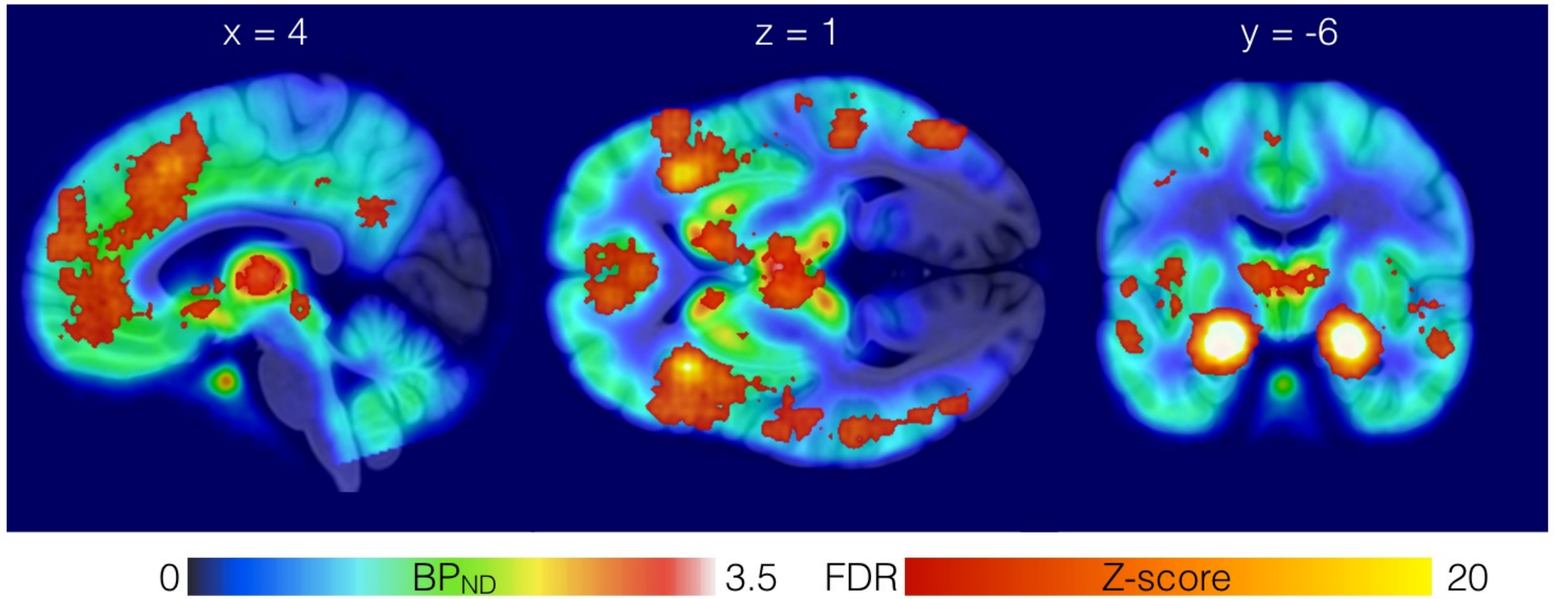


NeuroSynth (Yarkoni et al, 2011)

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



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



$r_{all} = 0.38$

$r_{pleasure} = 0.44$

Nummenmaa & Tuominen (2018 Br J Pharmac); Kantonen et al (2020 NeuroImage)

Neurosynth: Pros and cons

- Very easy to analyze and interpret
- Data readily available, allows custom analyses
- Full-volume analysis
- 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 is often available and cheap to use given permissions can be reanalysed
- Data however have to be extracted, reprocessed and the metadata needs to be extracted

Integrated approach at PET Centre

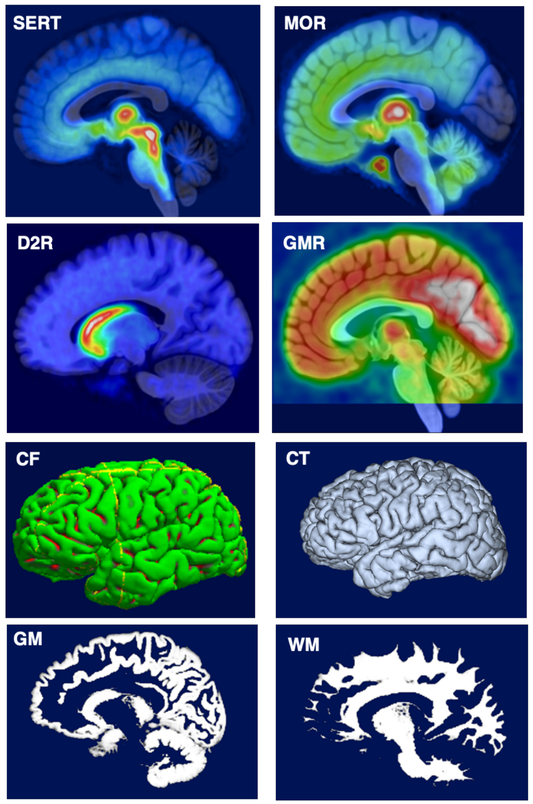
Hospital **PACS**

Native PET images

Native MR images

AUTOMATED KINETIC MODELLING, SEGMENTATION AND MESH GENERTATION

Preprocessed **BRAIN** data



Automated and supervised quality control

AIVO database

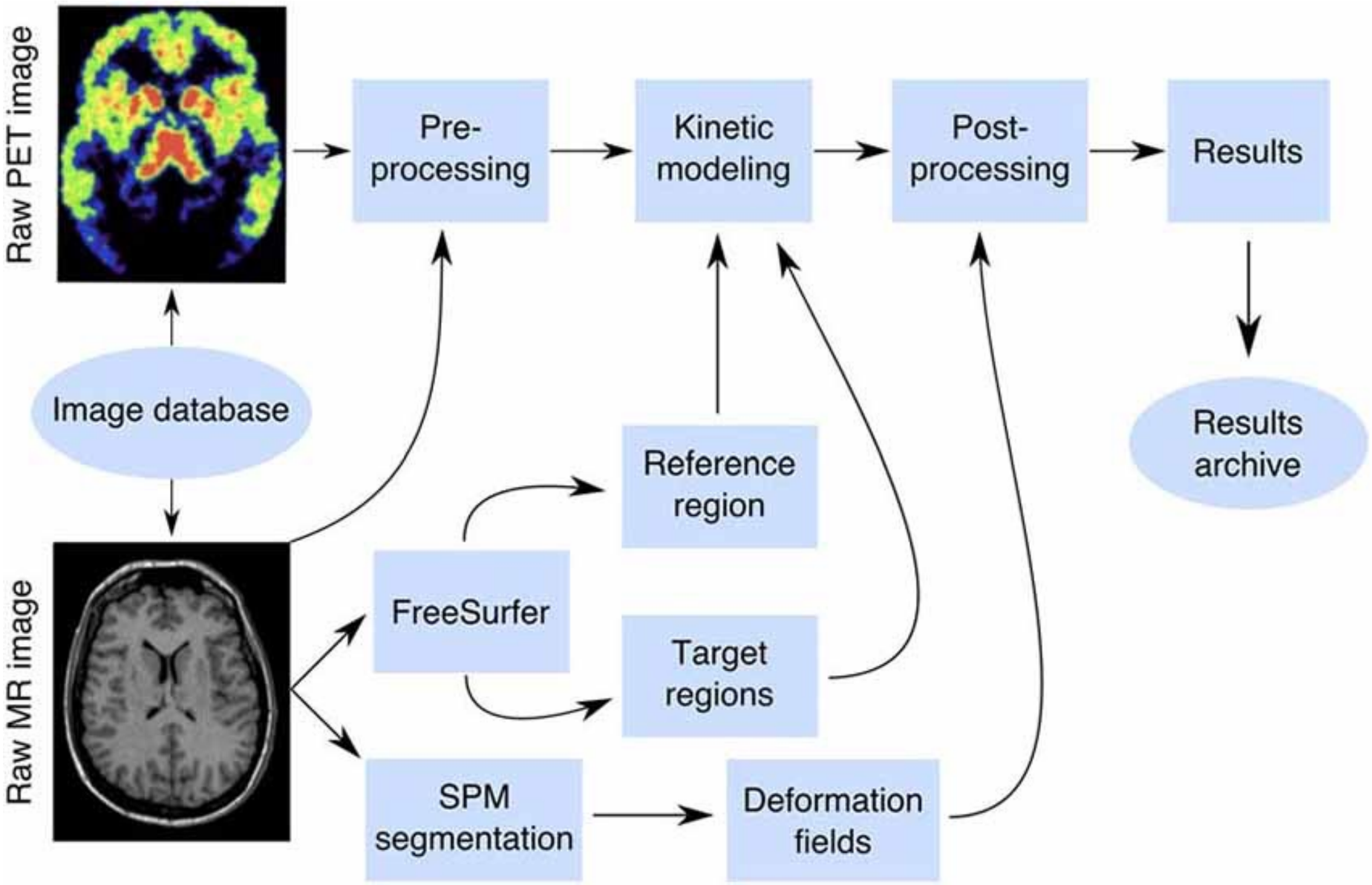
EXTERNAL register sources

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

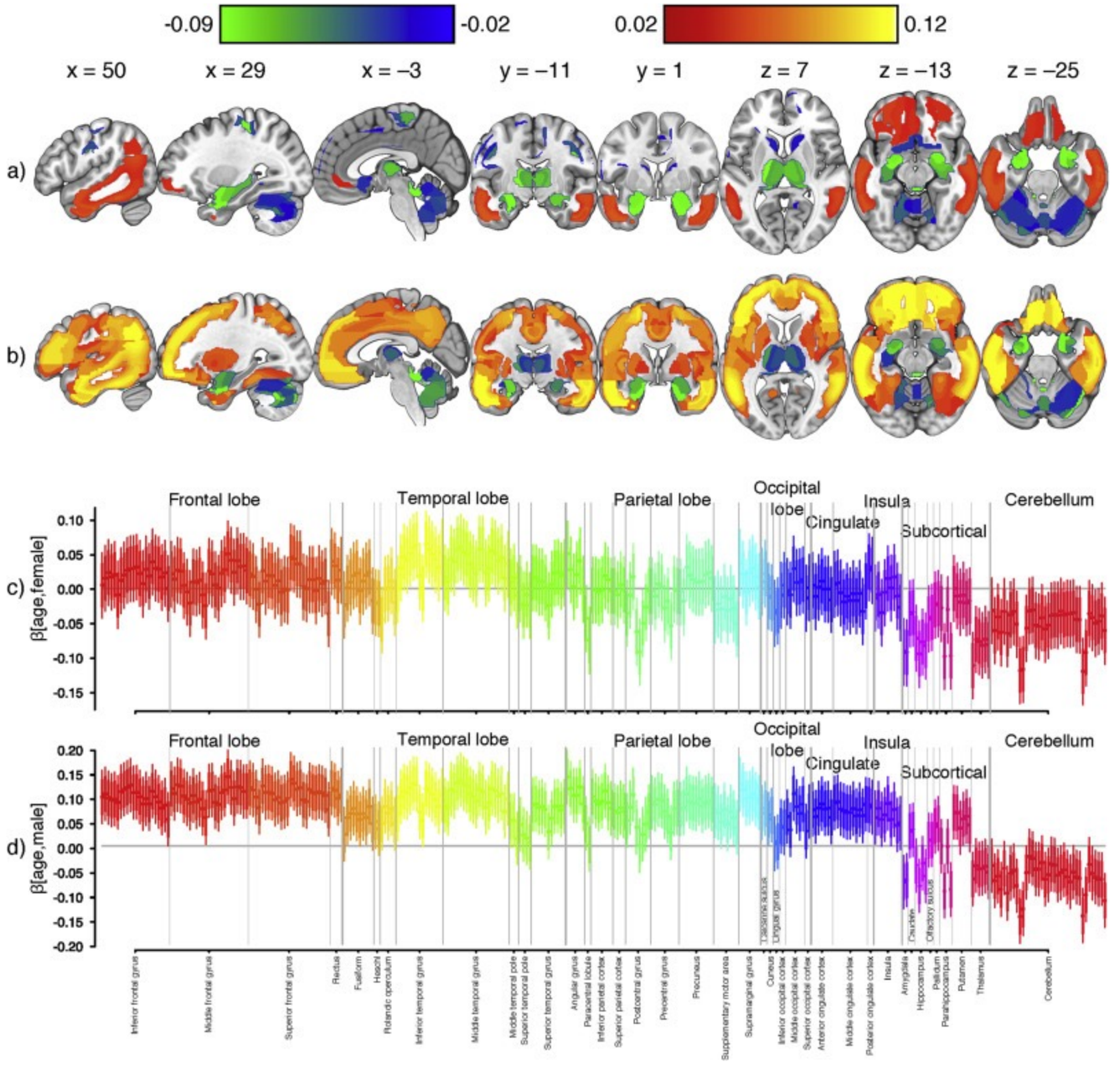
PSYCHOLOGICAL FUCNTIONING
1. Psychopathology (ICD codes)
2. Personality structure
3. Disorders of cognition (ICD codes)
4. Fluid intelligence and school achievement

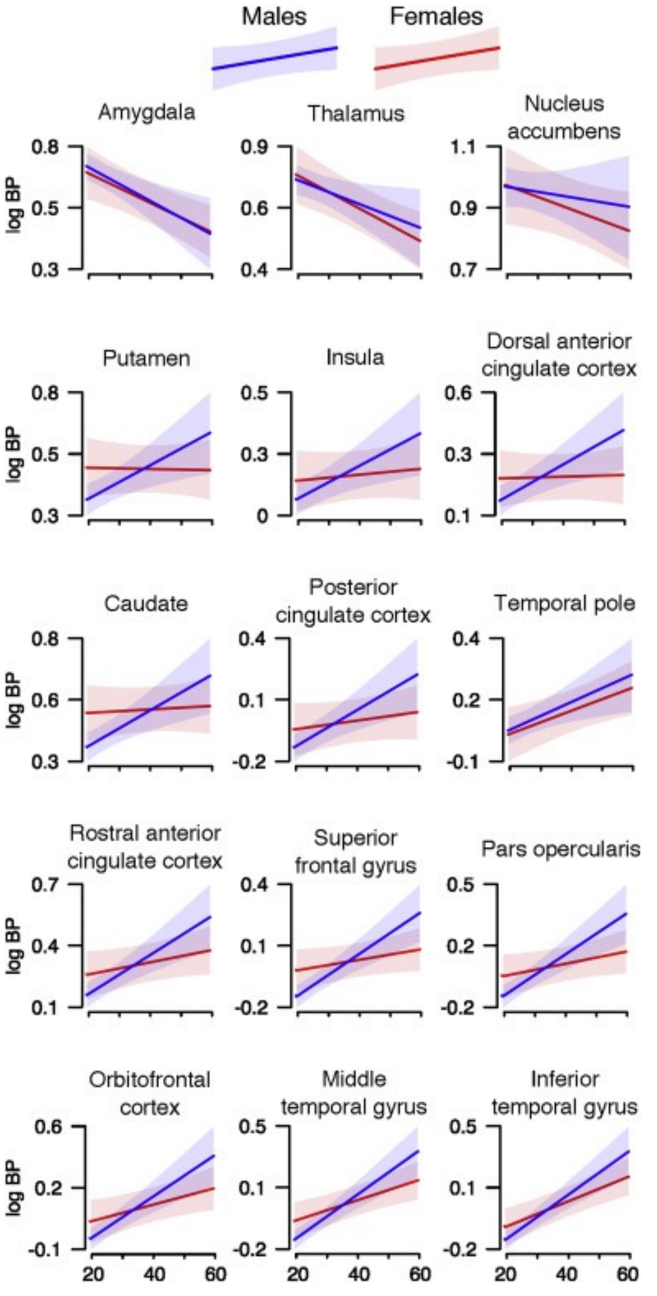
SOCIOECONOMIC WELL BEING
1. Earning and income transfers
2. Education and social status
3. Labour market attachment
4. Developmental socioeconomic stressors

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



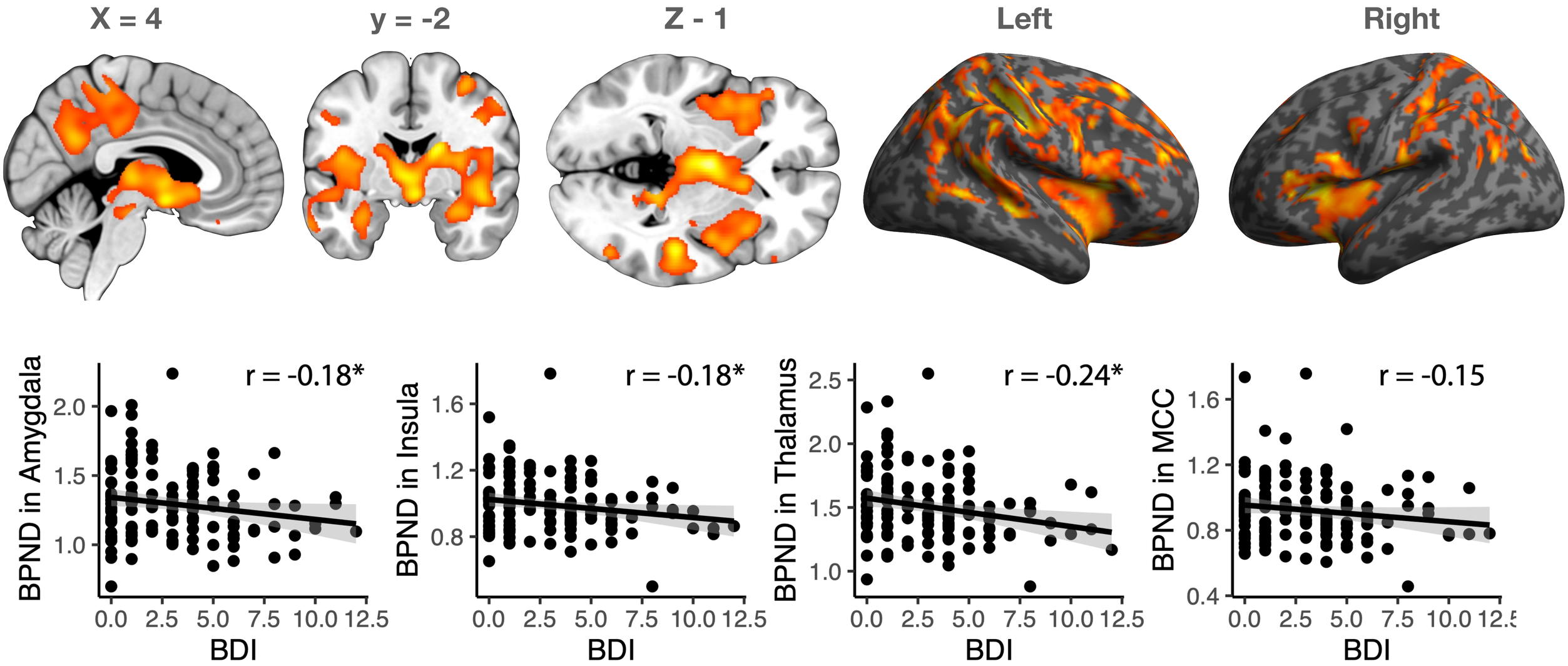
Karjalainen et al (2020)



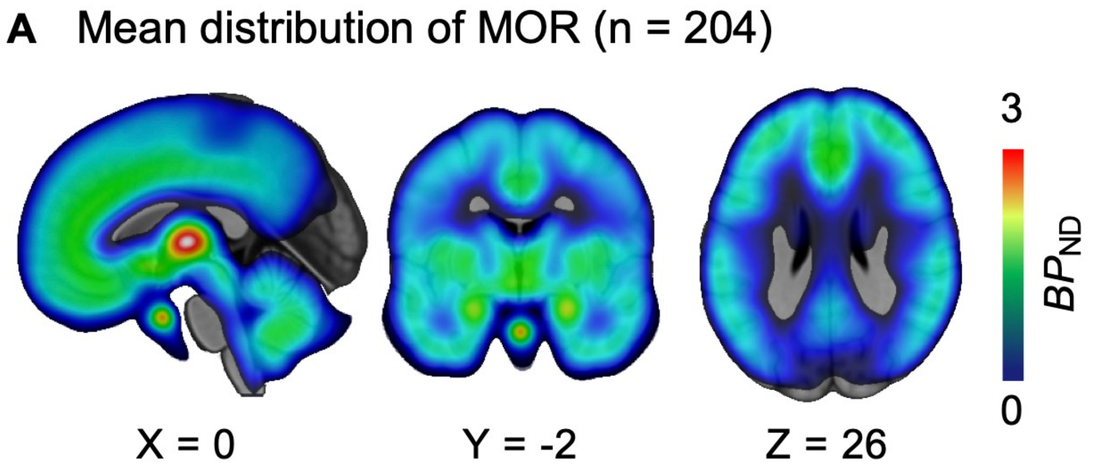


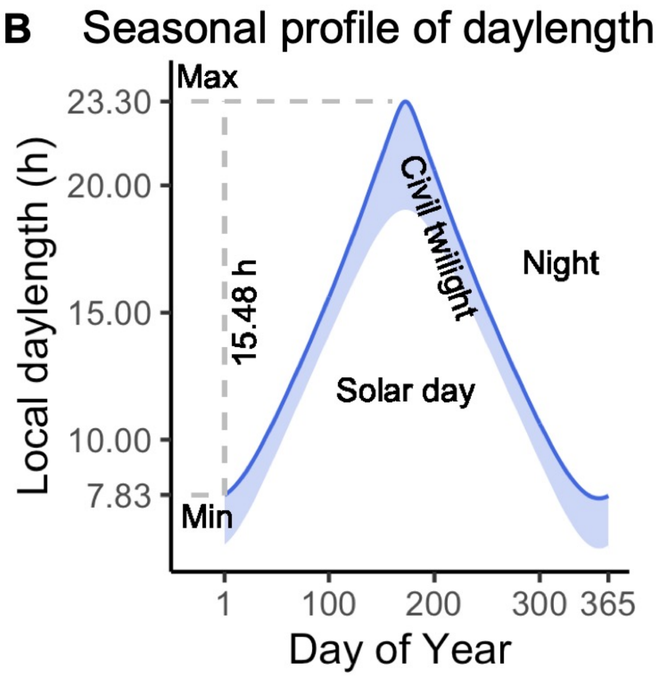
Kantonen et al (2020 NeuroImage)

Lowered mu-opioid receptor availability in subclinical depression and anxiety

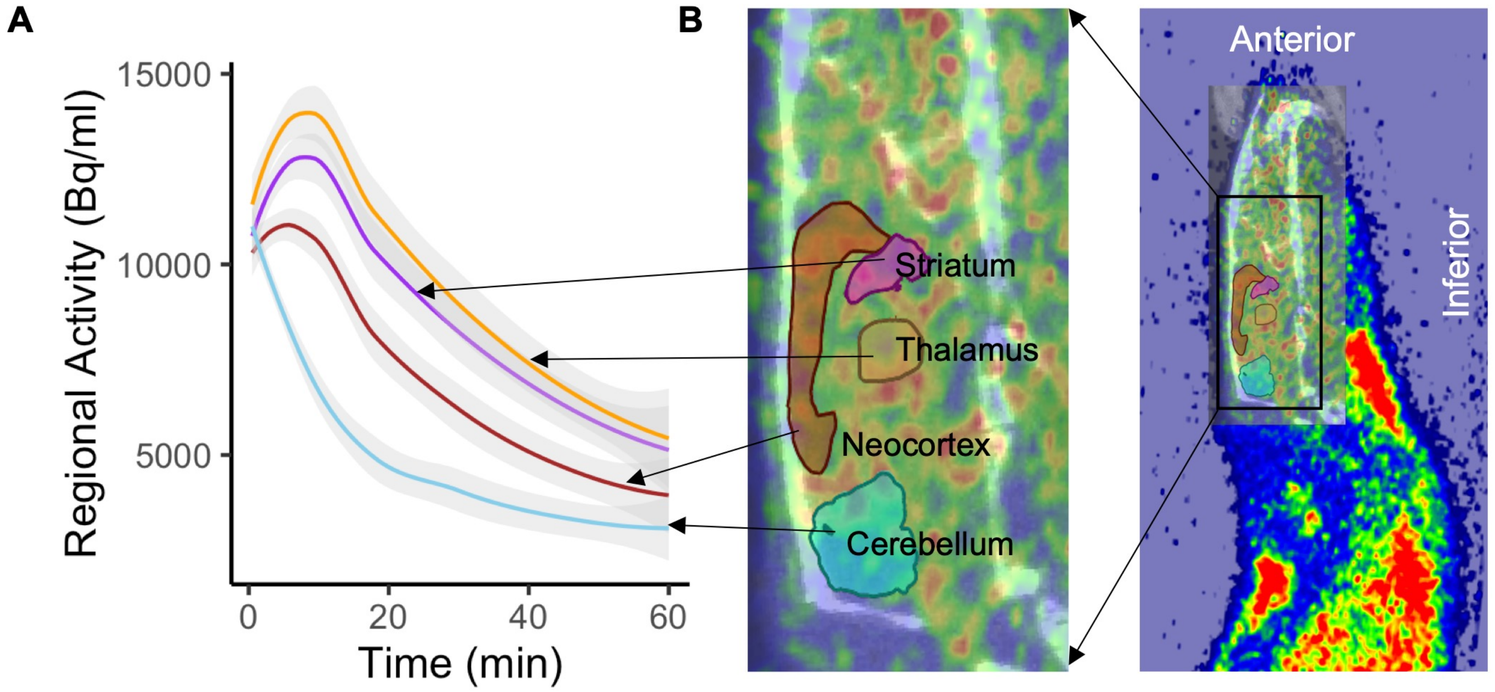


Nummenmaa et al (2020 Neuropsychopharmacology)





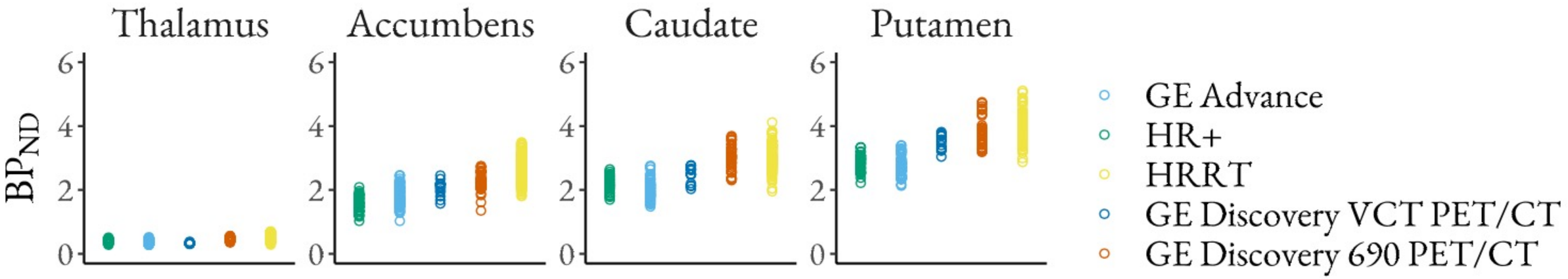
Sun et al (2021 J Neurosci)



Sun et al (2021 J Neurosci)

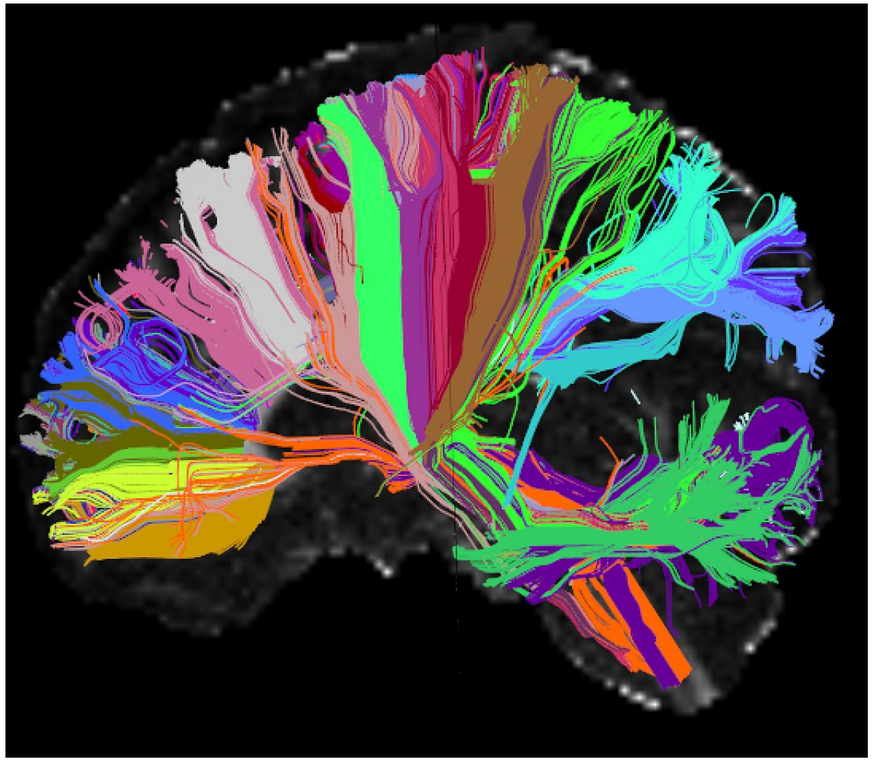
Common problems with data integration

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

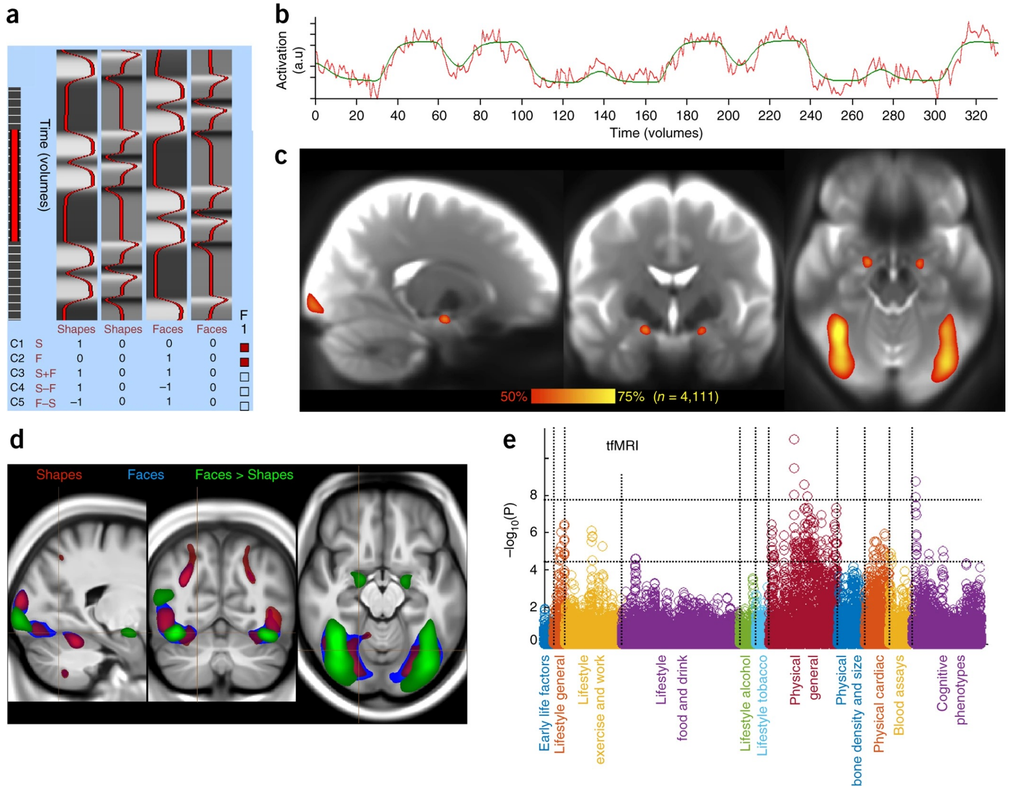


Malen et al (submitted)

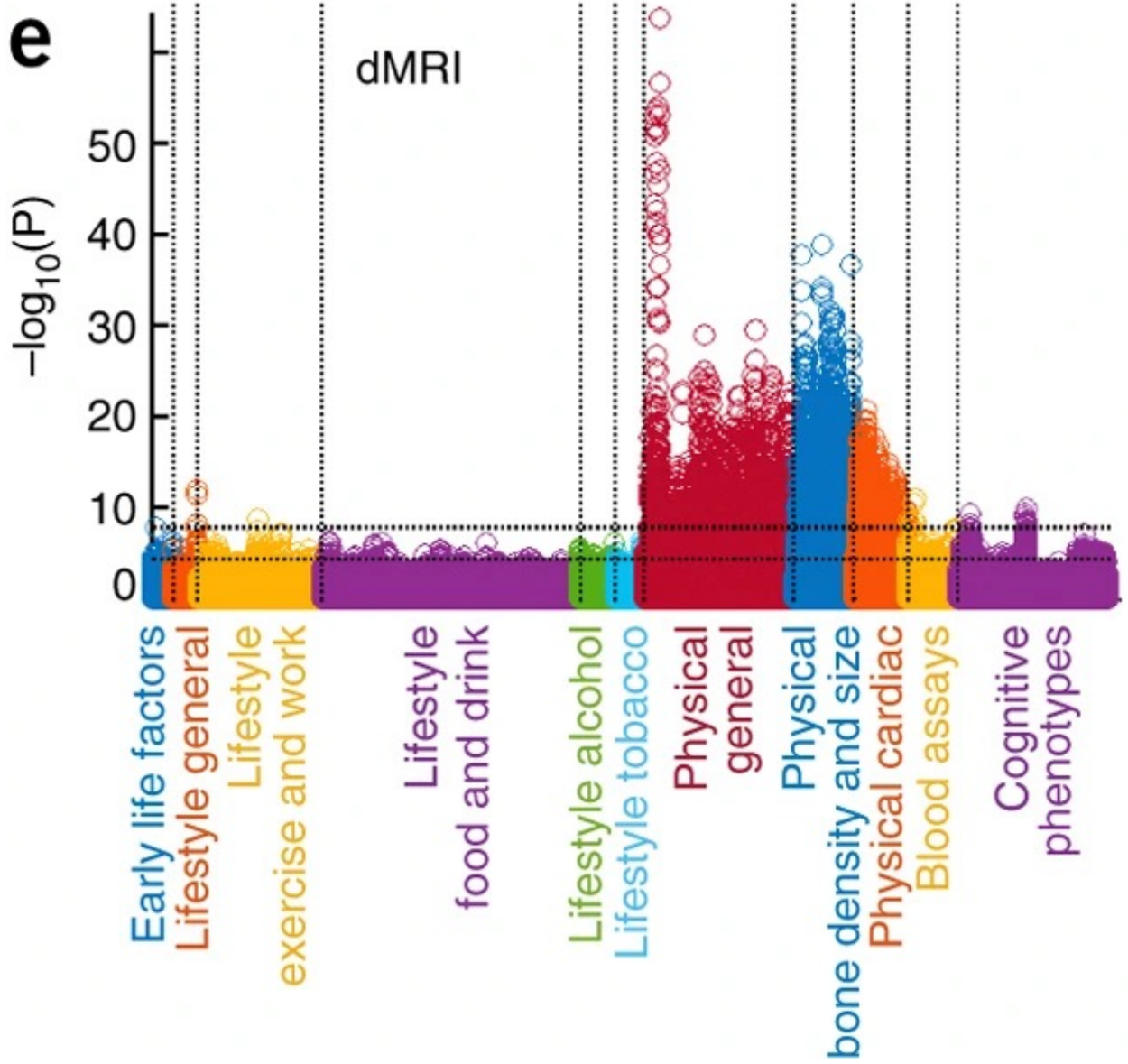
Solution 3: More is more in the first place



Human Connectome Project



UK Biobank



Miller et al (2016)

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


"That's all Folks!"