INTRODUCTION TO EPIGENOME-WIDE ASSOCIATION STUDIES (EWAS)

4. META-ANALYSIS OF EPIGENOME-WIDE ASSOCAITION STUDIES (EWAS) (THEORY)

EPIGENOME-WIDE ASSOCIATION STUDY (EWAS)

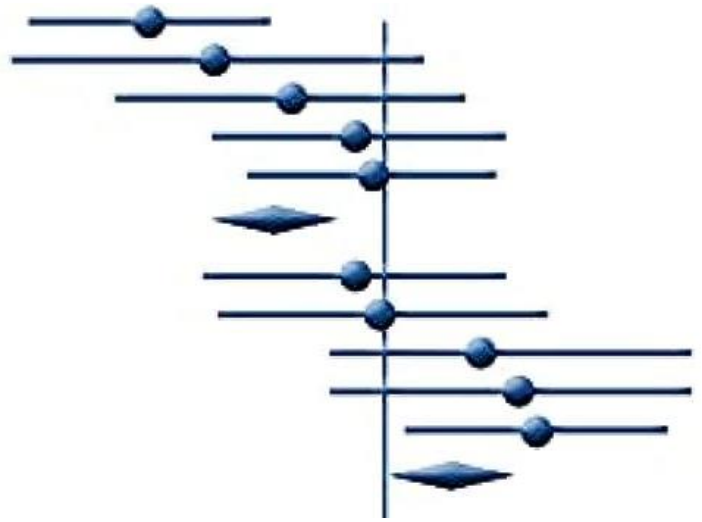
Workflow

1. Scientific question
2. Study population
3. Biological sample
4. DNA methylation data acquisition
5. Quality control of DNA methylation data
6. Epigenome-wide association study (EWAS)
7. Meta-EWAS or replication / validation
8. Biological interpretation

EPIGENOME-WIDE ASSOCIATION STUDY (EWAS)

Workflow

1. Scientific question
2. Study population
3. Biological sample
4. DNA methylation data acquisition
5. Quality control of DNA methylation data
6. Epigenome-wide association study (EWAS)
7. **Meta-EWAS or replication / validation**
8. Biological interpretation



7. META-EWAS

Robustness of results

Statistical power (methyl=smoking):

- Sample size
- Percentage of exposed (SD of continuous variable)
- Effect size on DNA methylation
- Type I error (and multiple-testing)

Statistical power calculator:

pwrEWAS for case/control studies (https://github.com/stefangraw/pwrEWAS)

Solutions:

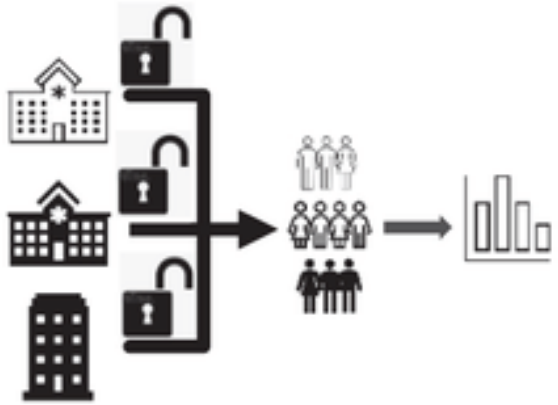
Validation with another method in same samples

Replication in an independent sample

Increase statistical power -> increase sample size -> combine data from different studies

7. META-EWAS

Pooled analysis



- NOT privacy-protected
- NOT communication-efficient
- Heterogeneity-aware
- Highly accurate

Meta-analysis

- Privacy-protected
- Communication-efficient
- Heterogeneity-aware
- Not highly accurate

Statistical synthesis of information from multiple independent studies

Luo et al Nat Comm 2022

7. META-EWAS

Steps

1. Sources

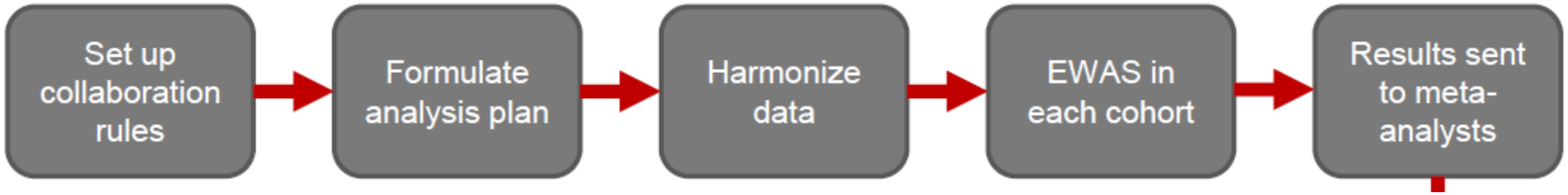
1.1. Individual-level data

- Directly from other studies

  -> Pregnancy and Childhood Epigenetics Consortium – PACE

  https://www.niehs.nih.gov/research/atniehs/labs/epi/pi/genetics/pace/index.cfm

- Online repositories of individual level data



PACE

Pregnancy And Childhood Epigenetics

7. META-EWAS

Steps

1. Sources

1.1. Individual-level data

- Directly from other studies

  -> Pregnancy and Childhood Epigenetics Consortium – PACE

  https://www.niehs.nih.gov/research/atniehs/labs/epi/pi/genetics/pace/index.cfm
- Online repositories of individual level data

1.2. Summarized results -> check units and covariates!!!

- Online repositories of summary statistics
- Papers

7. META-EWAS

Steps

1. Sources

1.1. Individual-level data

- Directly from other studies

  -> Pregnancy and Childhood Epigenetics Consortium – PACE

  https://www.niehs.nih.gov/research/atniehs/labs/epi/pi/genetics/pace/index.cfm

- Online repositories of individual level data

1.2. Summarized results

- Online repositories of summary statistics
- Papers

2. Transfer summarized results (CpG, effect, SE, pvalue)

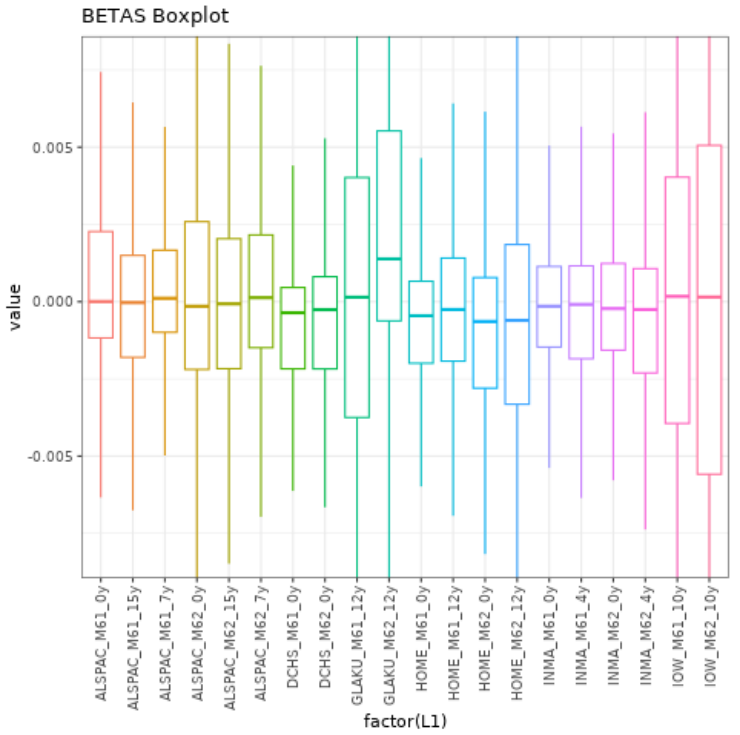
3. Quality control of the results of each study

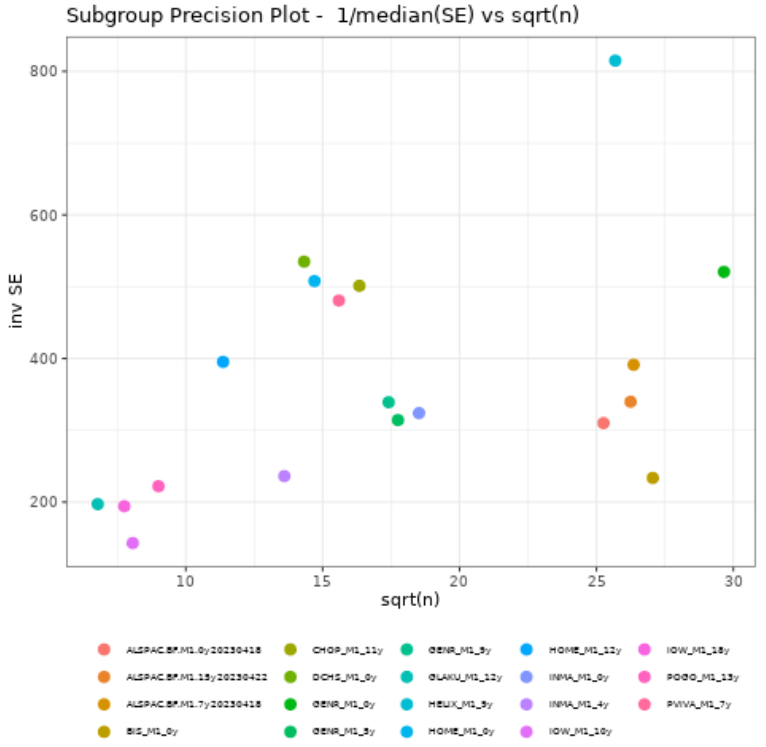
4. Meta-analysis

7. META-EWAS

Quality control of results of each study

- QQ plot and lambda inflation factor
- Check consistency of results (effects, SE, p-values)
- Check filtering of CpGs
- Precision-plot





7. META-EWAS

Statistical tests

- Meta-analysis of p-values
- Inverse-variance weighted meta-analysis (IVW)
  - More precise studies (usually larger ones) will have more weight in the meta-analysis
  - Types:
    - Fixed effect IVW MA: assumes one true effect size share by all studies (diff. are sampling error)
    - Random effect IVW MA: allows true effect sizes to vary from study to study (heterogeneity)

      –> less power
  - Estimates of heterogeneity: I^2, p-value het
    - Low-moderate (I^2<50)
    - Moderate-high (I^2>50)

Tools

- GWAMA (fixed and random-effects, https://genomics.ut.ee/en/tools)
- METAL (fixed-effects IVW, https://genome.sph.umich.edu/wiki/METAL_Documentation)
- metafor R package

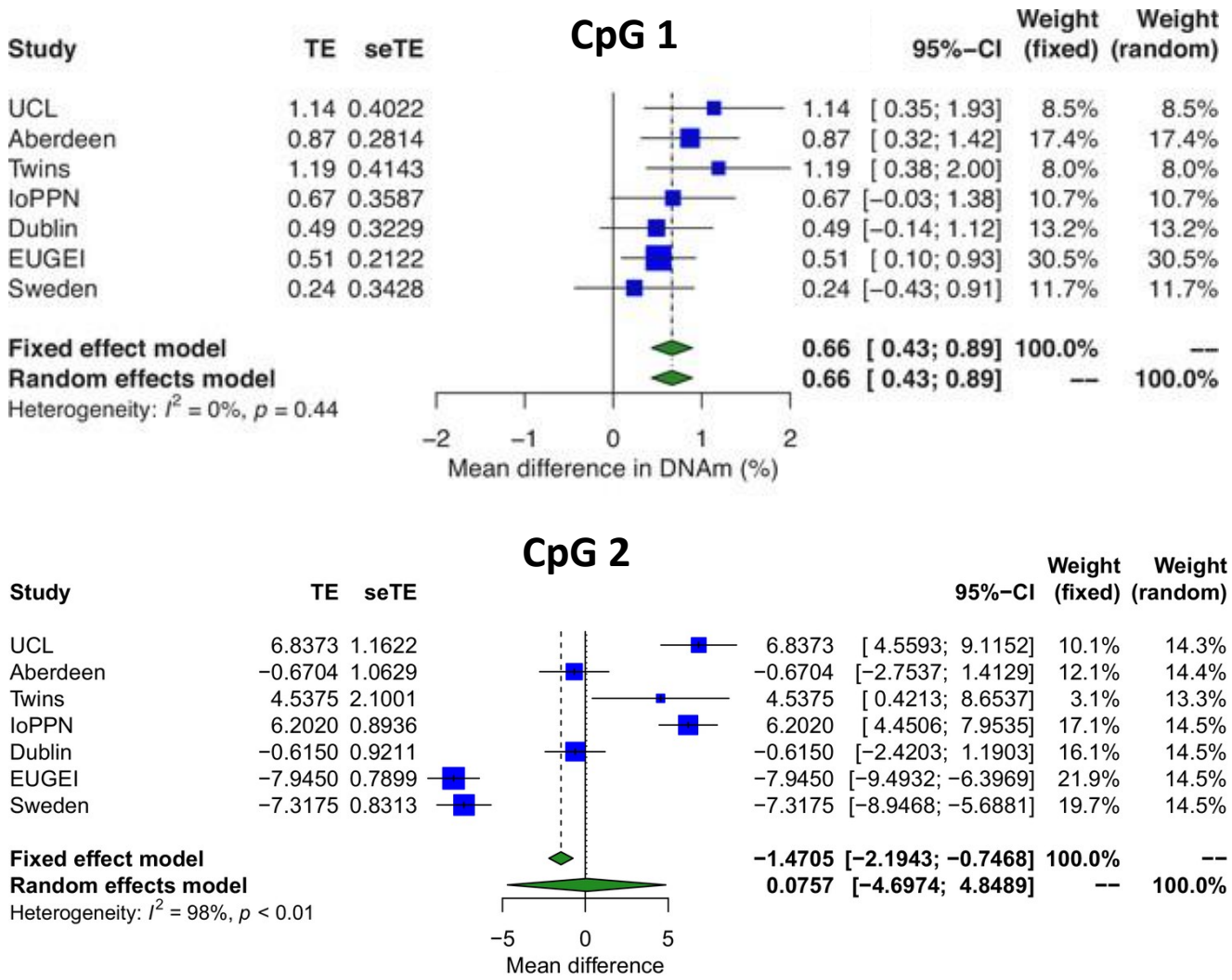
Rice, Higgins & Lumley, 2018

7. META-EWAS

Visualization of results

- Table with results
- Manhattan plot
- Volcano plot
- Scatter plots
- Comet plot
- Forest plot
- Leave-one-out MA plot

Forest plots



INTRODUCTION TO EPIGENOME-WIDE ASSOCIATION STUDIES (EWAS)

4. META-ANALYSIS OF EPIGENOME-WIDE ASSOCAITION STUDIES (EWAS) (PRACTICAL SESSION)

META-EWAS OF CURRENT AND FORMER SMOKING

Data: Cohort 1 (N = 294), Cohort 2 (N=273), Cohort 3 (N=260)

- Array: 450K
- Tissue: blood
- Ancestry: White European
- Sex: males and females
- Smoking: never, former, current
- Age: yes
- Array batch: yes
- Cells: yes

Input (for each cohort and current and former smoking): results dataframe (adj and sva)

Output (for current and former / each cohort): QC boxplot, QC precision plot, meta-results dataframe, meta-results FDR sig dataframe, lambda, QQ plot, Volcano plot, forestplot

Tool: metafor R package

Questions:

- What do we expect a precision plot to show? Does our precision plot follow the theory?
- How many FDR CpGs are associated with current smoking? and how many with former smoking?
- Which is the CpG with the lowest p-value for current smoking? Does it show increased or decreased methylation? Does it show heterogeneity across studies? Is the effect size of this CpG similar between current and never smokers and between former and never smokers?