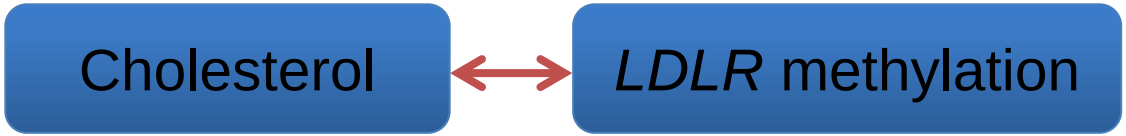
Analyzing multiple omics levels

Bas Heijmans
Molecular Epidemiology
Leiden University Medical Center
The Netherlands
bas.heijmans@lumc.nl

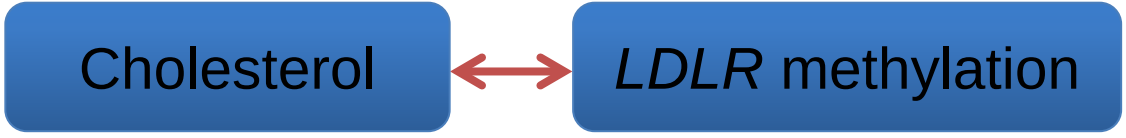
FOS course ‘Molecular Epidemiology in Ageing Research – 19 June 2017.

Epigenome-wide Association Study

1. Data: methylation at 450 thousand CpGs + lipids levels in 2000 individuals
2. Test per CpG: DNAm ~ cholesterol + gender + age + cell counts + batches

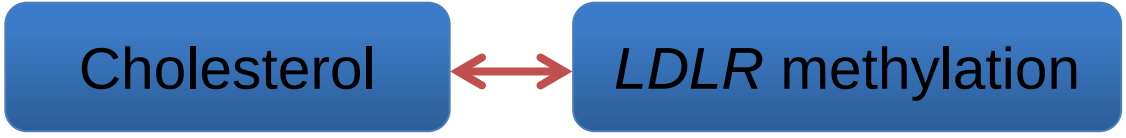


What's next?



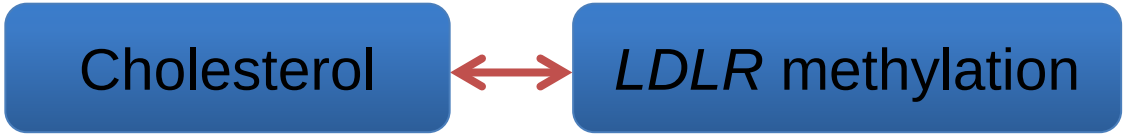
- Can we make conclusions stronger?
- What are the main limitations in observational epidemiology?

What's next?



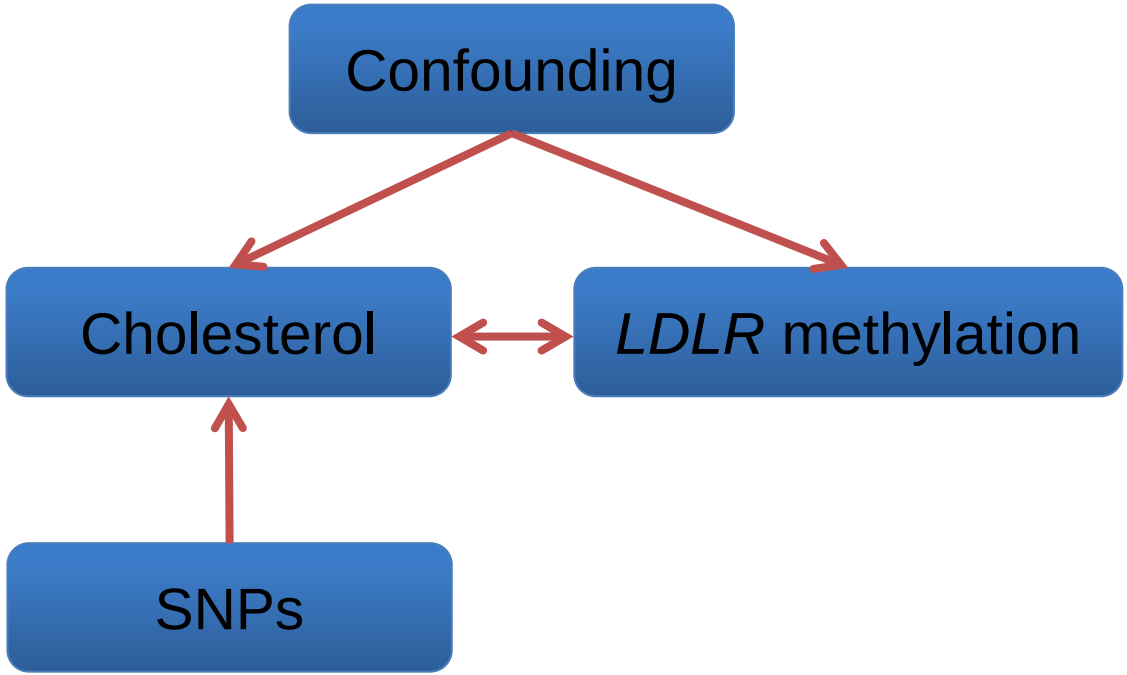
- Causality
- Confounding

Alternative for experiment



- An experiment of nature using genetic variation as causal anchor
- ‘Mendelian randomization’: a natural trial with exposures to genetic variations randomized according to Mendel’s law and with the exposed blinded towards exposure.
- Uses genetic variant(s) as ‘instrumental variable’ instead of measured variable itself.

Alternative for experiment



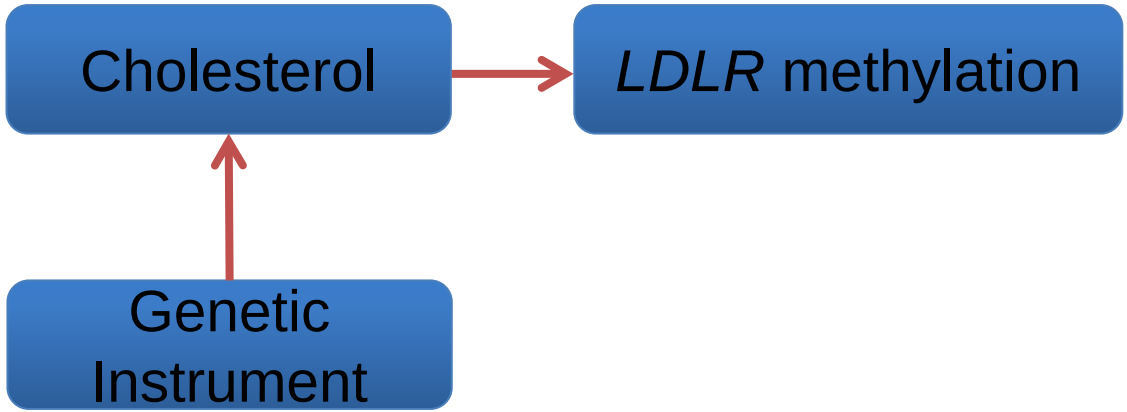
Mendelian randomization

1. Identify (sufficiently strong) genetic instrument.
   Here: SNP associated with cholesterol from GWAS (see wednesday practical)
2. Predict cholesterol level for every individual on basis of one's genotype.
   Will explain only small proportion of variation
3. Test whether predicted (genetic) level is associated with methylation.
   No confounding (unbiased effect estimate) & only one possible direction of causality

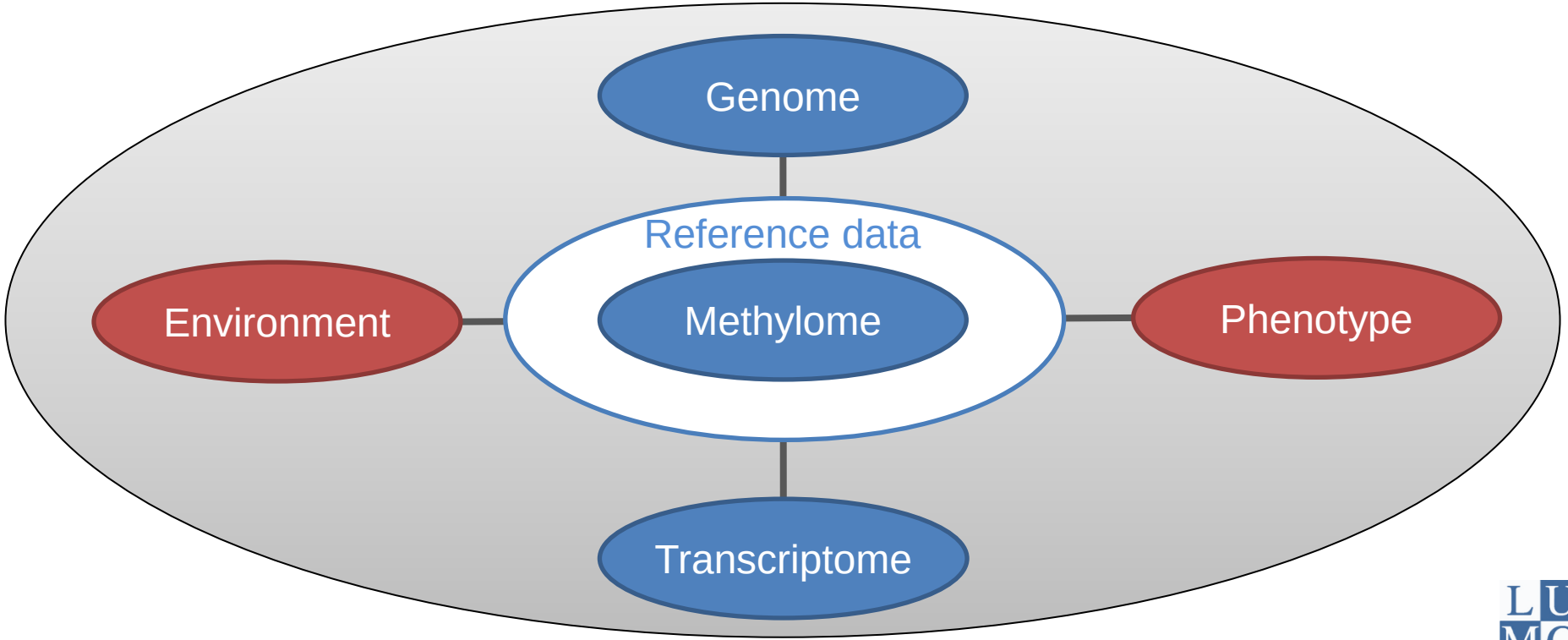
Two-step least squares model

1. Pred(Chol) = $\gamma_0 + \gamma_1$ x genotype + γ_2 x age + γ_3 x batch + ε_i (i=1,…, n)

2. DNAm = $\beta_0 + \beta_1$ x pred(Chol) + β_2 x age + β_3 x batch + ε_i (i=1,…, n)



Integrative Genomics



Population Genomics

The human as ‘model organism’:
Exploiting natural variation in large-scale population studies

- Genome biology
- Disease mechanisms
- Biomarkers