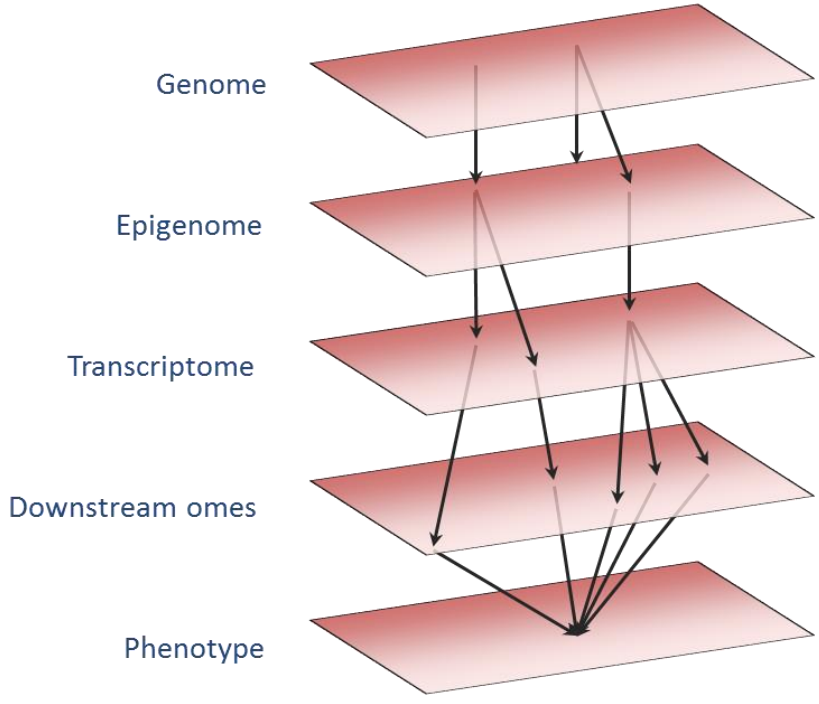
Analyzing multiple omics levels - Mendelian randomization

Bas Heijmans
Molecular Epidemiology
Dept. of Biomedical Data Sciences
Leiden University Medical Center
The Netherlands
bas.heijmans@lumc.nl

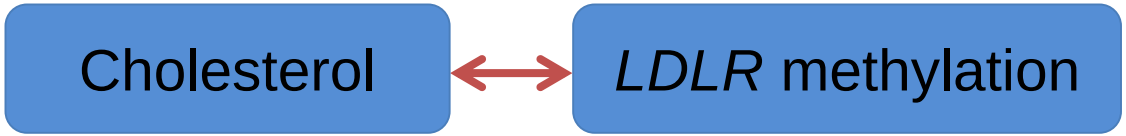
FOS course Molecular Data Science – 1 November 2019.

The single most important distinction in study designs

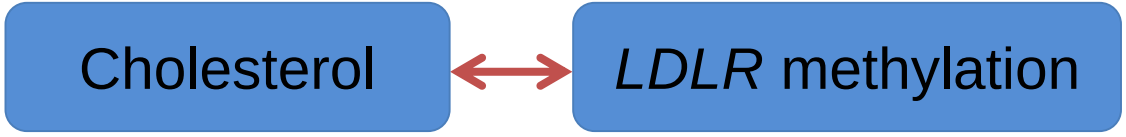

Genome
Epigenome
Transcriptome
Downstream omes
Phenotype

Epigenome-wide Association Study

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

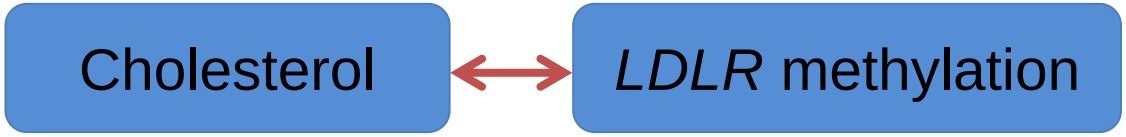


What's next?



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

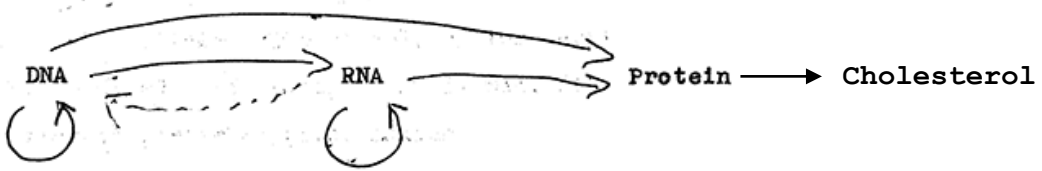
What's next?



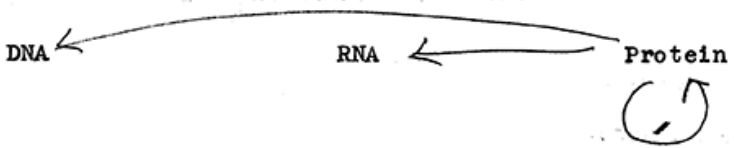
Alternative for experiment

Cholesterol ↔ *LDLR* methylation

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have



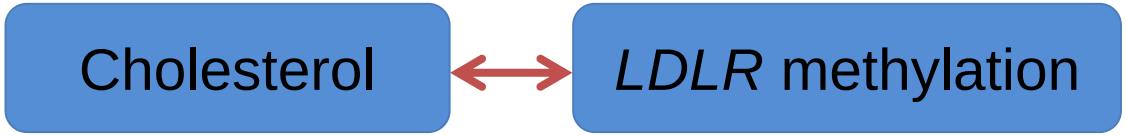
but never



where the arrows show the transfer of information.

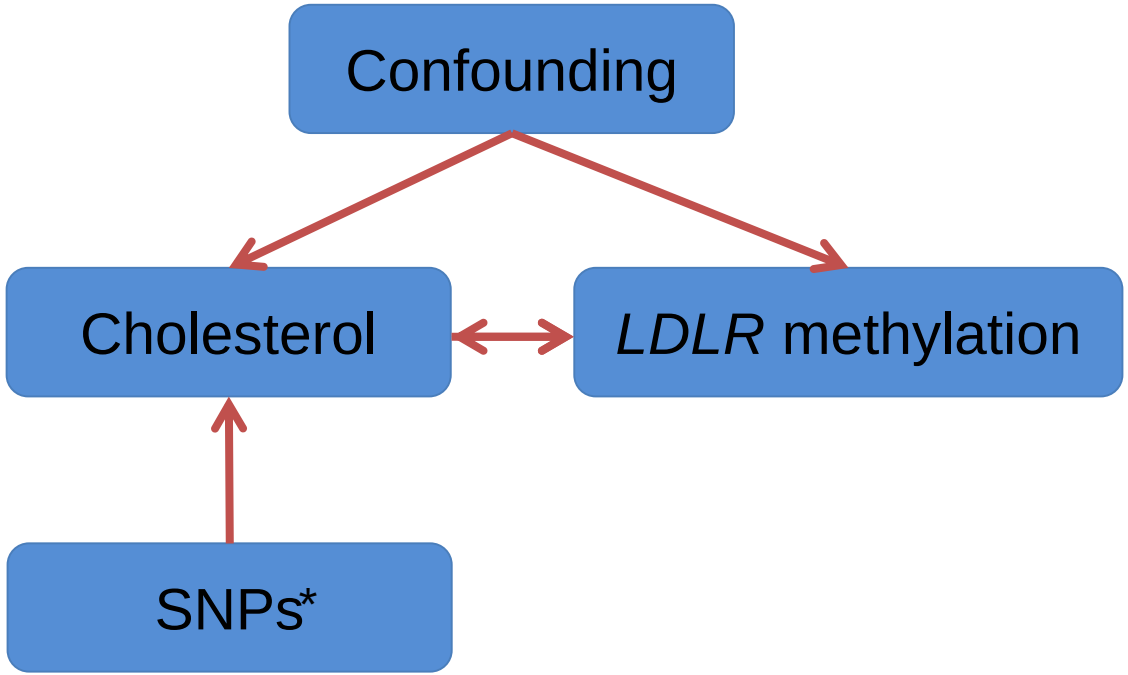
Crick, 1956

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



**Which we know from GWASs*

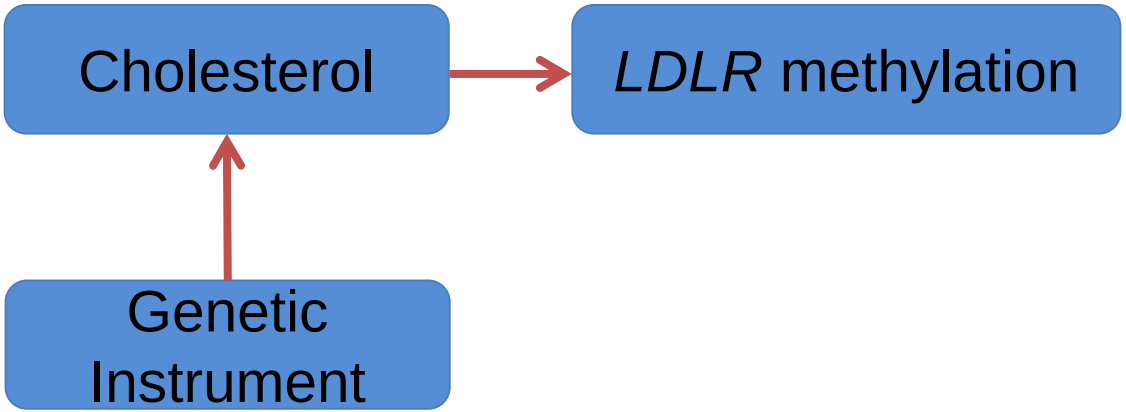
Mendelian randomization

1. Identify (sufficiently strong) genetic instrument.
   Here: SNP associated with cholesterol from GWAS (see last week's 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-stage least squares model

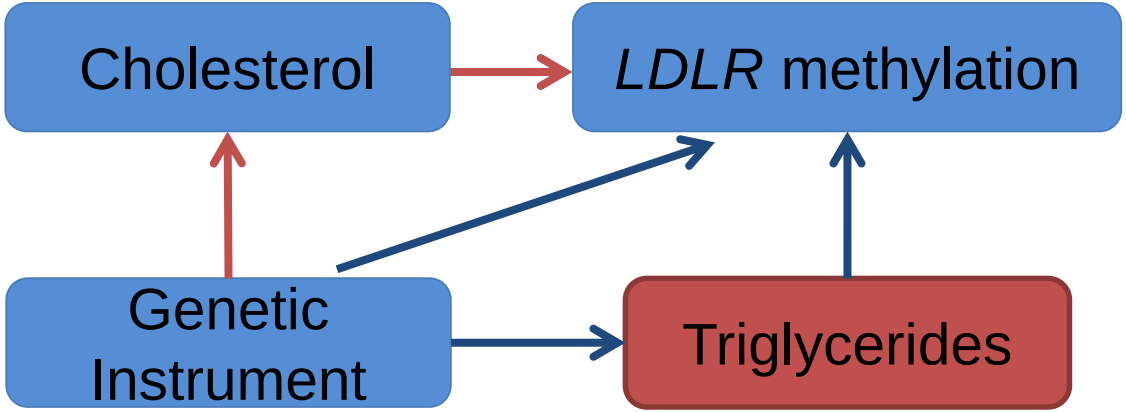
1. Pred(Chol) = γ_0 + γ_1 x genotype + γ_2 x age + γ_3 x batch + …

2. DNAm = β_0 + β_1 x pred(Chol) + β_2 x age + β_3 x batch + …



Beware of assumptions

Pleiotropy

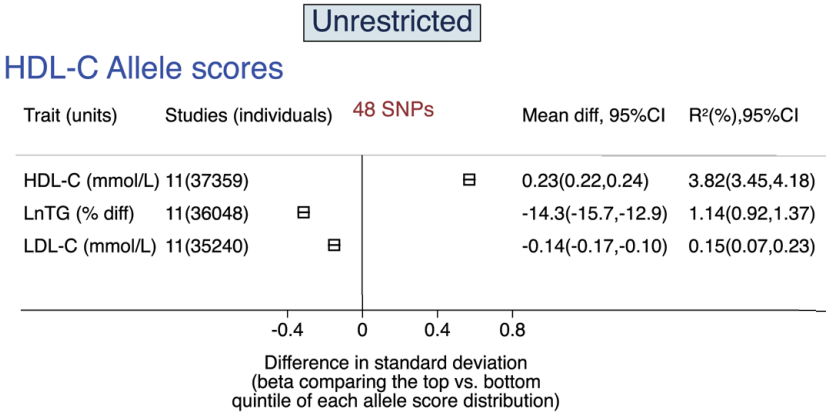


Example: good and bad cholesterol

HDL LDL

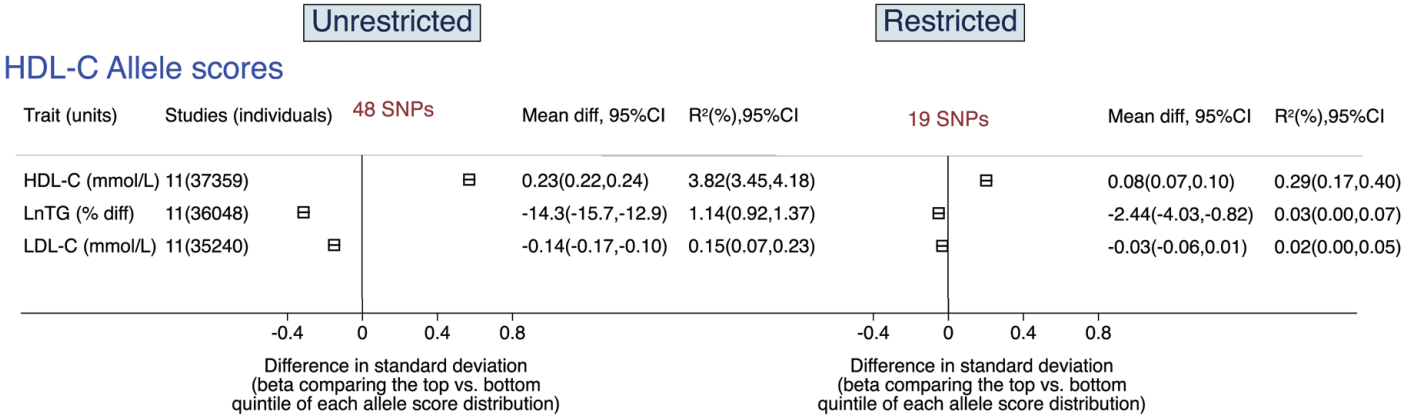
- What is the evidence for being good and bad?
- Do you know medication targeting good or bad cholesterol?

MR – validation PolyGenic Scores



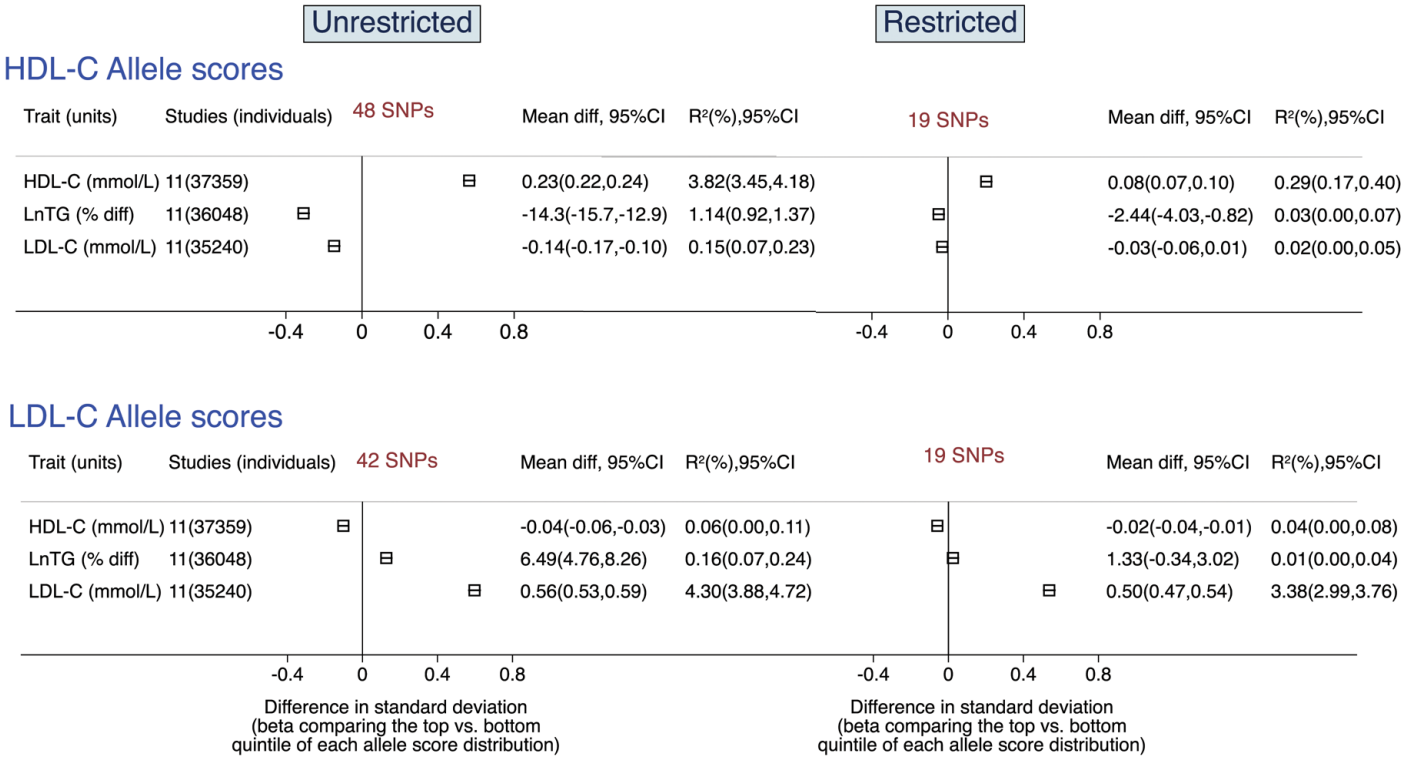
Holmes et al. *Eur Heart J* 2015

MR – validation PolyGenic Scores



Holmes et al. *Eur Heart J* 2015

MR – validation PolyGenic Scores



1. Pred(Lipid) = γ_0 + γ_1 x PGS + ...

Holmes et al. *Eur Heart J* 2015

PolyGenic Scores (PGSs)

- Find a GWAS publication with SNPs and effect sizes.
- Have genotypes for your own study.
- Count the number of minor alleles an individual has.
- Multiply this number by the effect size of the allele.
- The result is the genetically predicted level for an individual.

Constructing PGS

Minor allele effect size

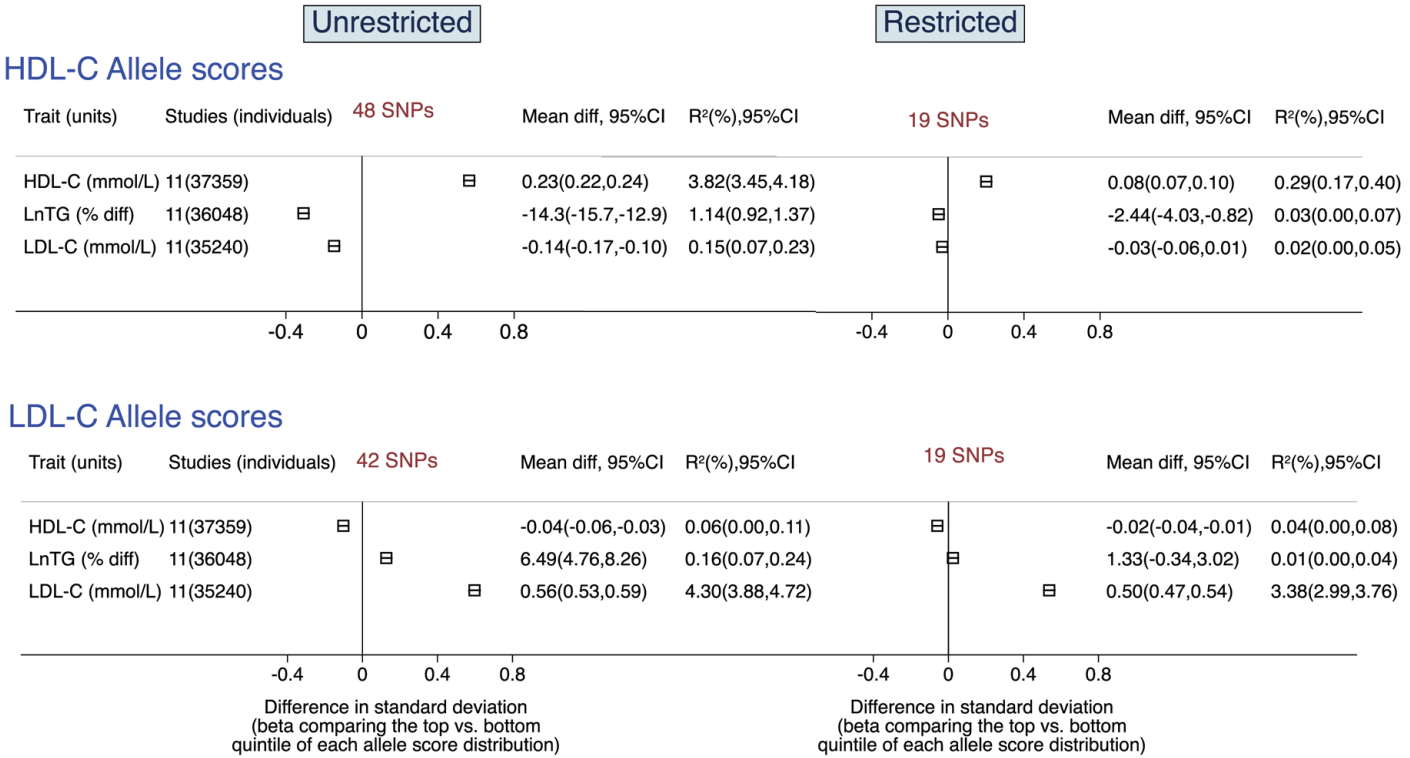
Nearest gene	MarkerName	Chr	hg19 Position (Mb)	Primary trait, Secondary trait(s)	MAF	Alleles minor/ major	Effect	Joint N (in 1000s)	Joint *P*-value
Loci Primarily Associated with HDL Cholesterol									
LCAT	rs16942887	16	67.93	HDL	.14	A/G	.083	186	$8x10^{-54}$
CMIP	rs2925979	16	81.53	HDL	.31	T/C	-.035	186	$1x10^{-19}$
STARD3	rs11869286	17	37.81	HDL	.35	G/C	-.032	178	$3x10^{-17}$

Let genotypes for individual 'Harry' be:
GG (*LCAT*), TC (*CMIP*), GG (*STARD3*)

- What is his PGS (or genetically predicted HDL level)?

Deloukas et al. *Nat Genet* 2013

MR – validation PolyGenic Scores

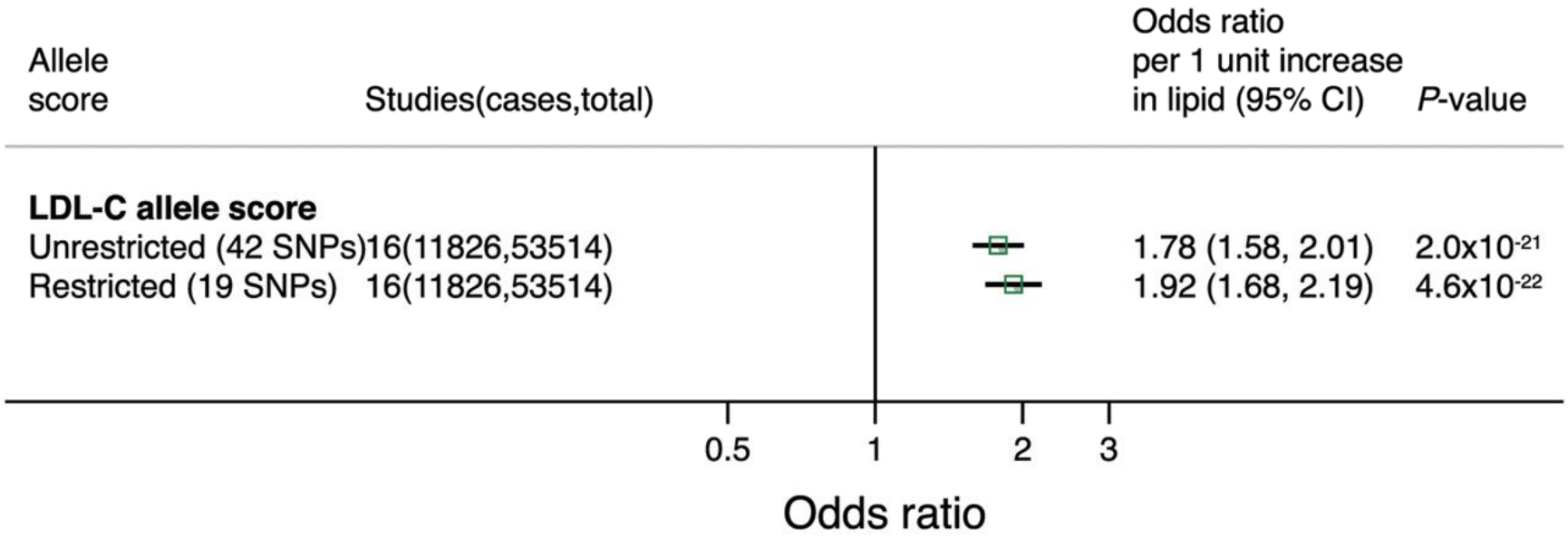


1. Pred(Lipid) = γ_0 + γ_1 x PGS + ...

Holmes et al. *Eur Heart J* 2015

MR – causal inference

CHD (incident/prevalent)

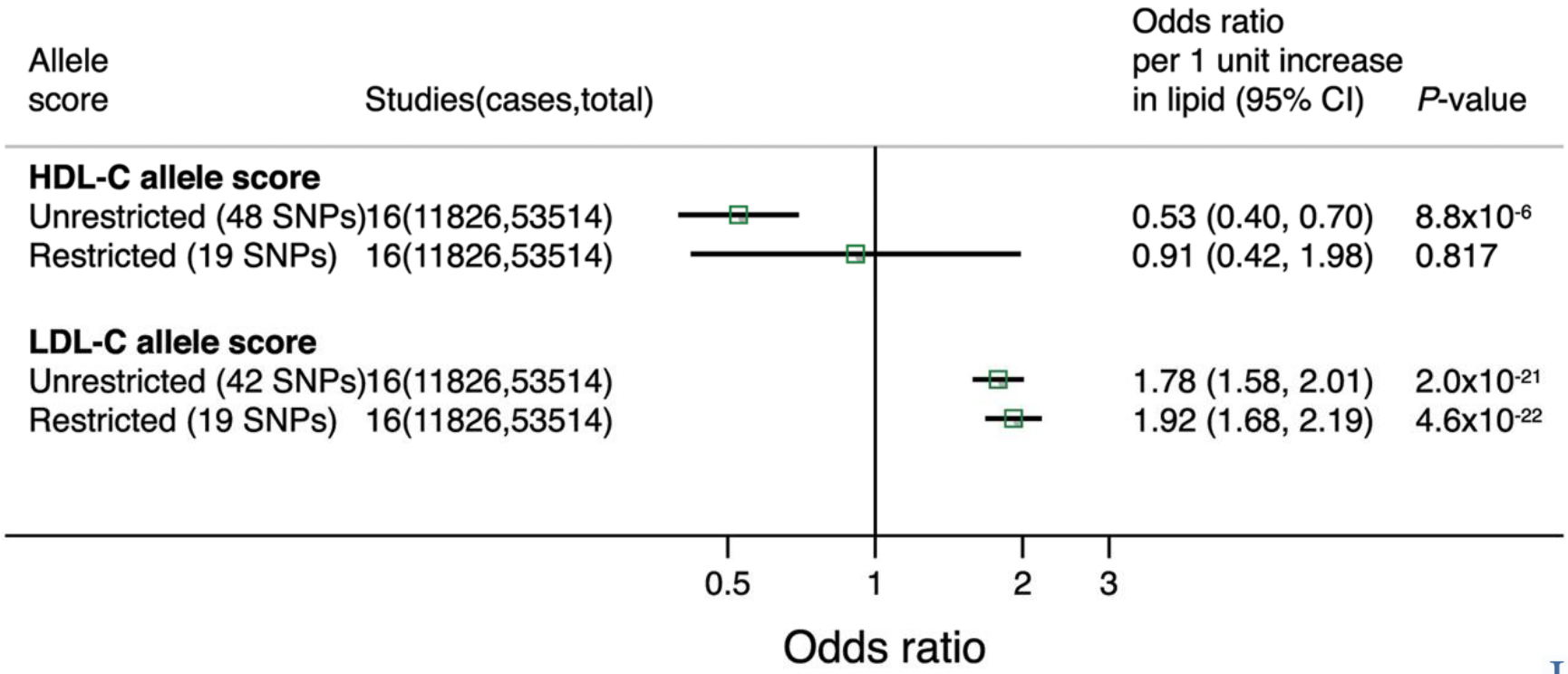


2. CHD = β_0 + β_1 x pred(Lipid) + ….

Holmes et al. *Eur Heart J* 2015

MR – causal inference

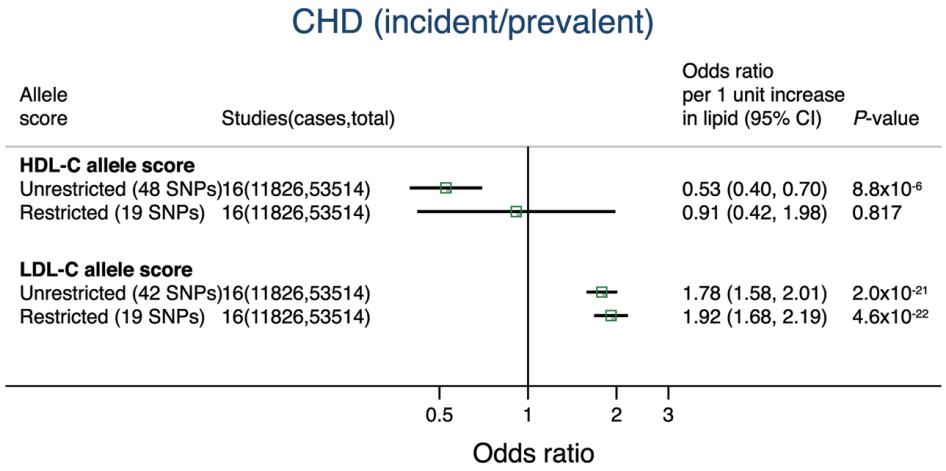
CHD (incident/prevalent)

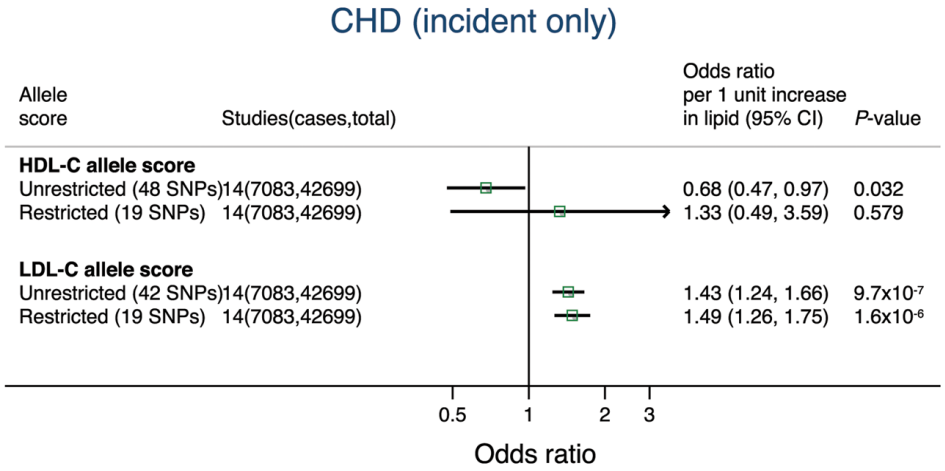


2. CHD = β_0 + β_1 x pred(Lipid) + ….

Holmes et al. *Eur Heart J* 2015

Mendelian randomization



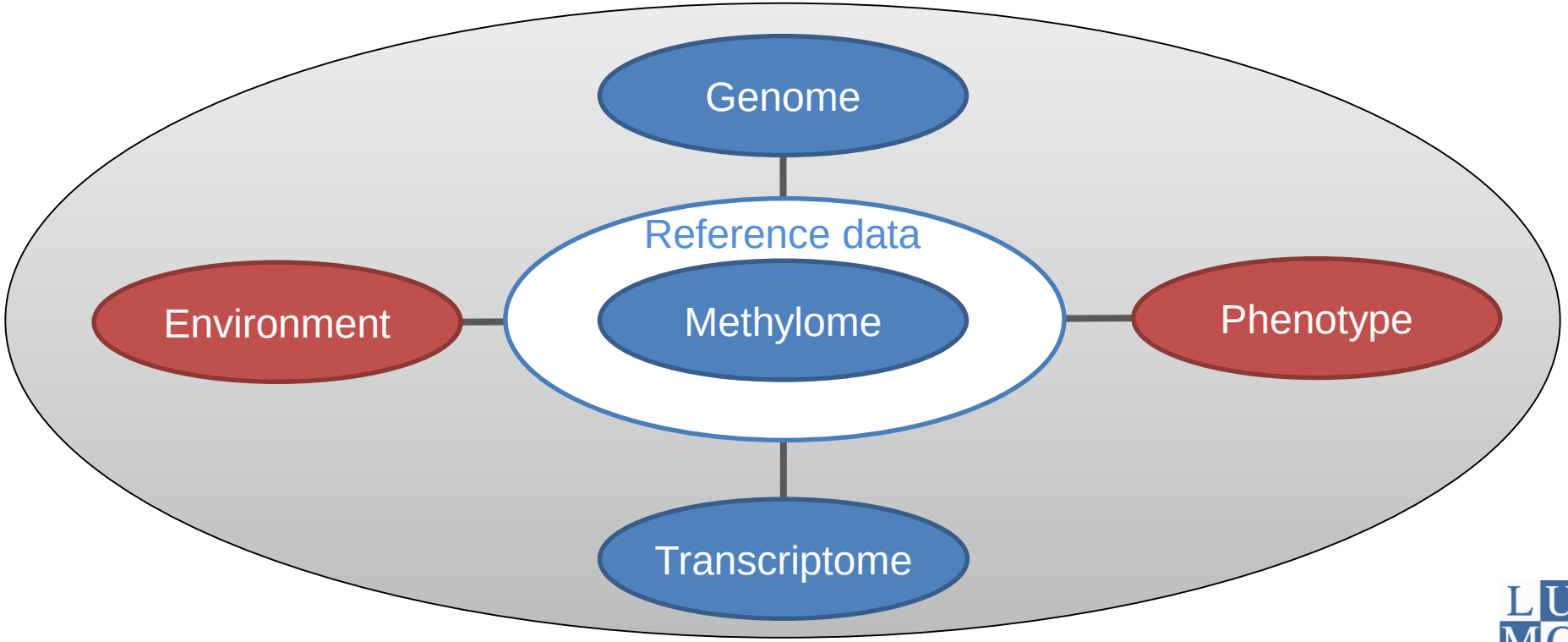


Holmes et al. *Eur Heart J* 2015

HDL and LDL in human studies

- Observational: cross-sectional & prospective case/control studies
- Experimental: drugs in clinical trial
- ‘Natural experiment’: Mendelian randomization

Integrative Genomics



Molecular Data Science in populations

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

- Genome biology
- Disease mechanisms
- Biomarkers