Machine Learning for Healthcare
6.7930 [6.871], HST.956

Lecture 22: Genetics Part 2
Mendelian Randomization, eQTLs, patient subtypes, multi-modal integration

Prof. Manolis Kellis

Slides credit:
David Evans
Manny Rivas
Sek Kathiresan
Yosuke Tanigawa

Mendelian randomization

A method for using measured variation in genes of known function to examine the causal effect of a modifiable exposure on disease in observational studies.

The design has a powerful control for reverse causation and confounding, which often impede or mislead epidemiological studies.

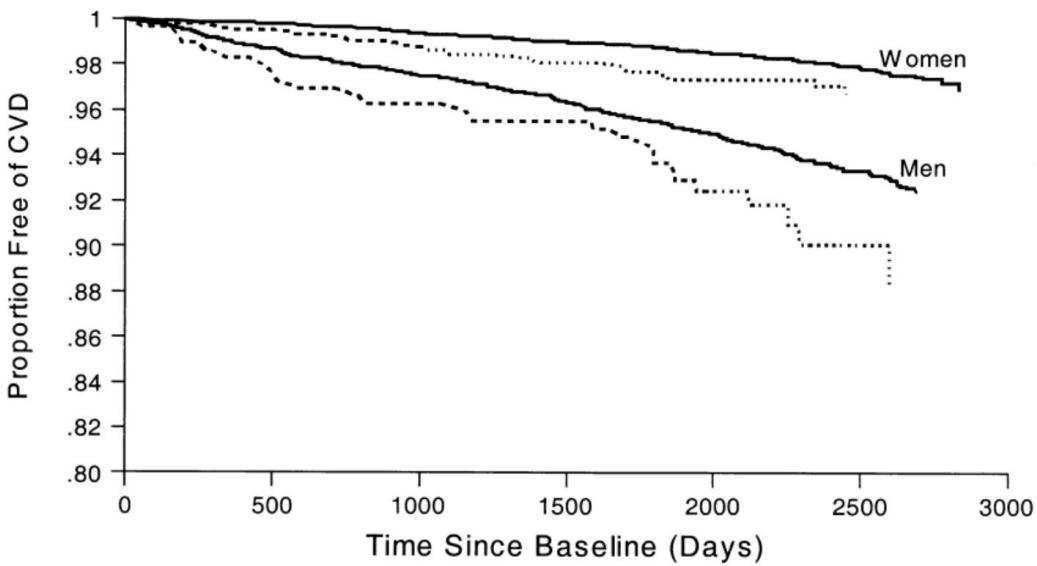
Slides from Sek Kathiresan

Recent history of CVD RCTs

1. Biomarker X is associated with Disease Y
2. Hypothesis: treatment to lower X will risk reduce risk for Y
3. Phase 3 randomized control trial to test hypothesis above

Slides from Sek Kathiresan

Example #1: Anemia and CVD



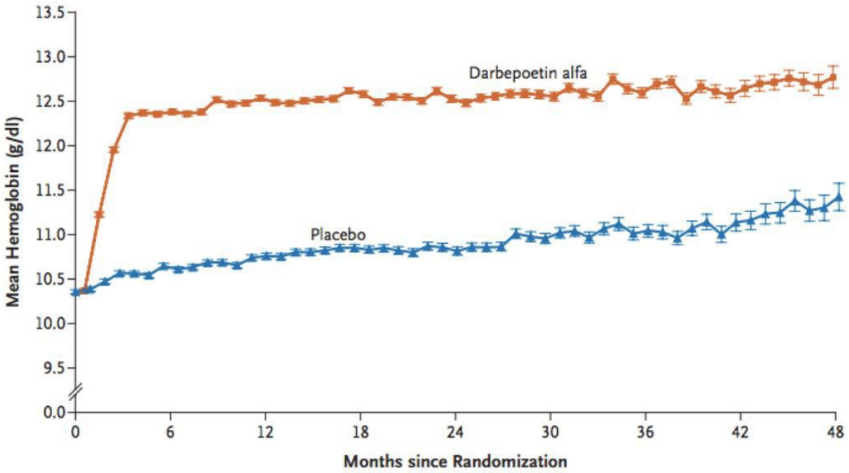
Sarnak, *J Am Coll Cardiol* 2002

Slides from Sek Kathiresan

Slides from Sek Kathiresan

Erythropoiesis stimulating agents (ESA) increases hemoglobin

TREAT trial: Treatment with ESA improved hemoglobin

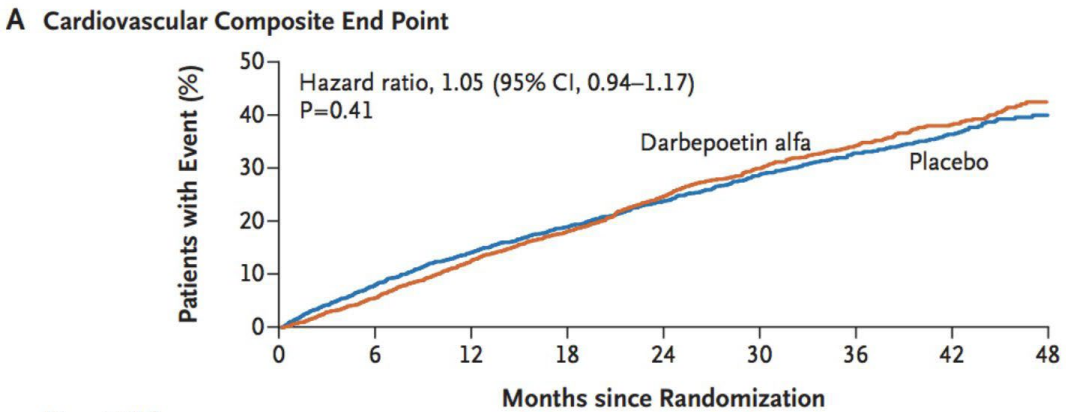


No. of Patients									
Darbepoetin alfa	2004	1768	1503	1300	946	635	404	253	97
Placebo	2019	1742	1460	1221	887	620	356	216	79

Pfeffer, *N Engl J Med* 2010

Slides from Sek Kathiresan

… but failed to reduce CVD risk

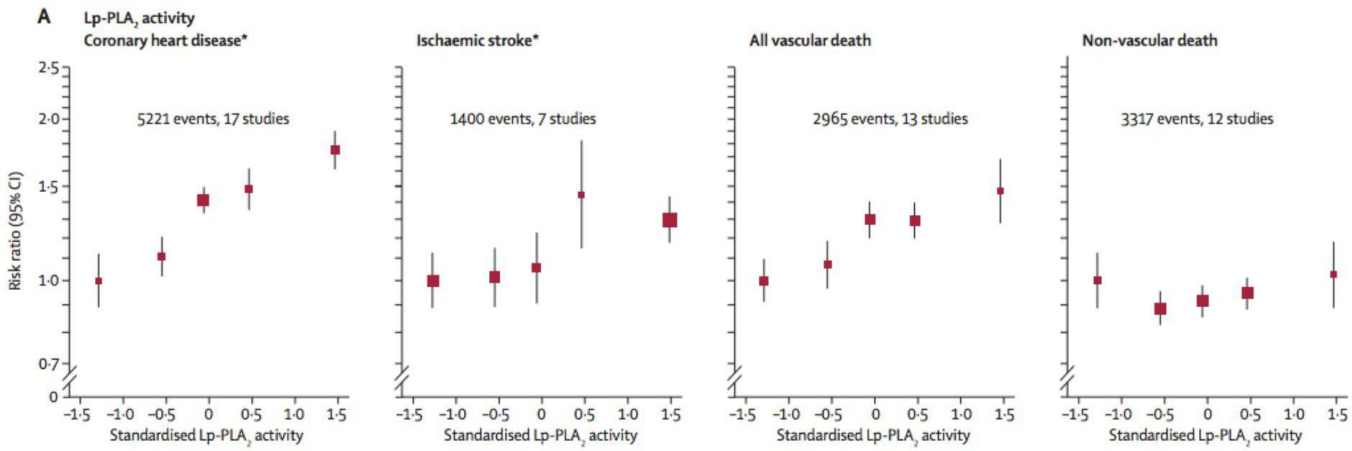


No. at Risk									
Darbepoetin alfa	2012	1882	1717	1515	1180	817	551	318	130
Placebo	2026	1836	1687	1487	1178	834	529	319	122

Pfeffer, *N Engl J Med* 2010

Slides from Sek Kathiresan

Example 2: Lipoprotein-associated phospholipase A2 & CHD



The Lp-PLA2 Studies Collaboration, *Lancet* 2010

Slides from Sek Kathiresan

Oral inhibitor of Lp-PLA2 - darapladib - inhibits enzymatic activity

Slides from Sek Kathiresan

Heartwire

Darapladib Fails in Large Phase 3 Study

Michael O'Riordan

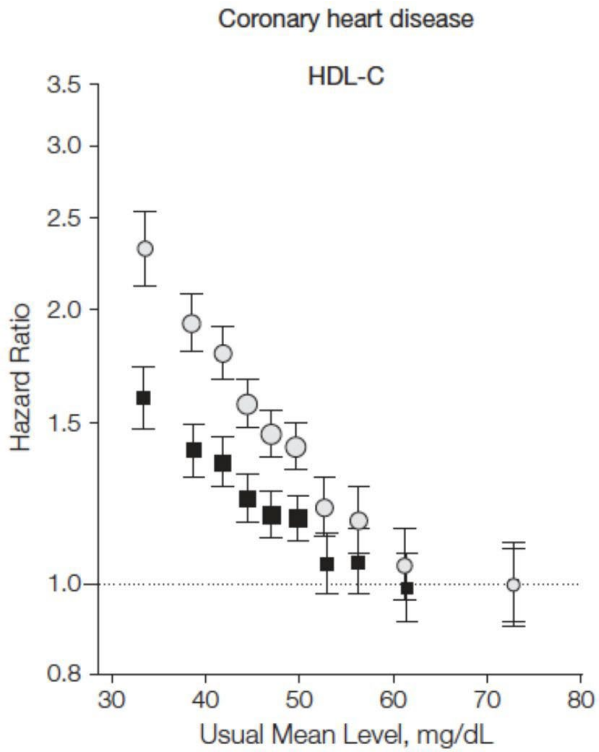
November 12, 2013

LONDON, UK – A large phase 3 study testing an inhibitor of the lipoprotein-associated A2 (Lp-PLA2) enzyme has failed to lower the risk of cardiovascular events in coronary heart disease patients who received the drug on top of statin therapy[1].

The Lp-PLA2 inhibitor **darapladib** (GlaxoSmithKline, London, UK) was tested in more than 15 000 patients in the **Stabilization of Atherosclerotic Plaque by Initiation of Darapladib Therapy** (STABILITY) study.

The trial ran to completion, but GlaxoSmithKline announced the top-line results today, stating the drug failed to provide a significant reduction in the risk of cardiovascular death, nonfatal MI, or nonfatal stroke when compared with patients treated with placebo.

Example 3: HDL cholesterol and CHD



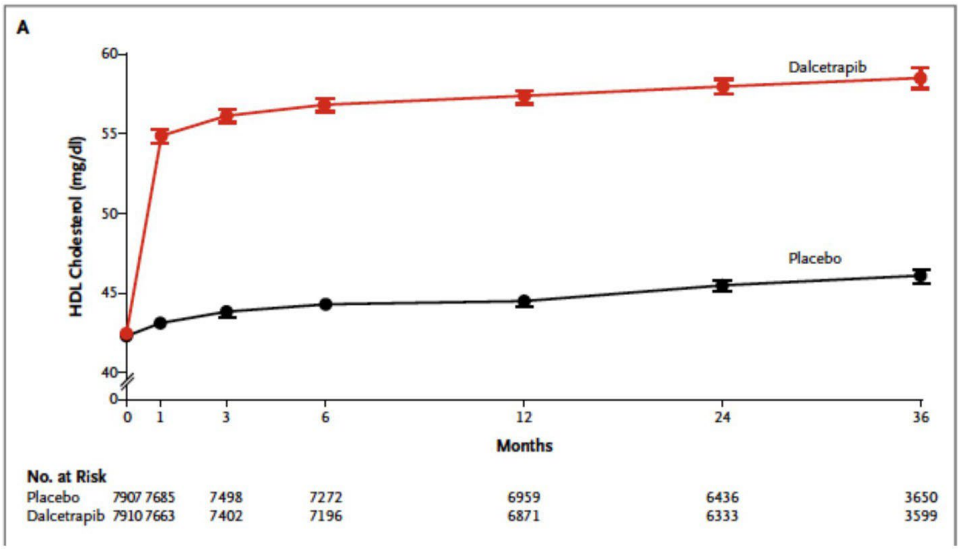
Emerging Risk Factors Collaboration, *JAMA* 2009

Slides from Sek Kathiresan

Dalcetrapib increases HDL cholesterol by 30%

Slides from Sek Kathiresan

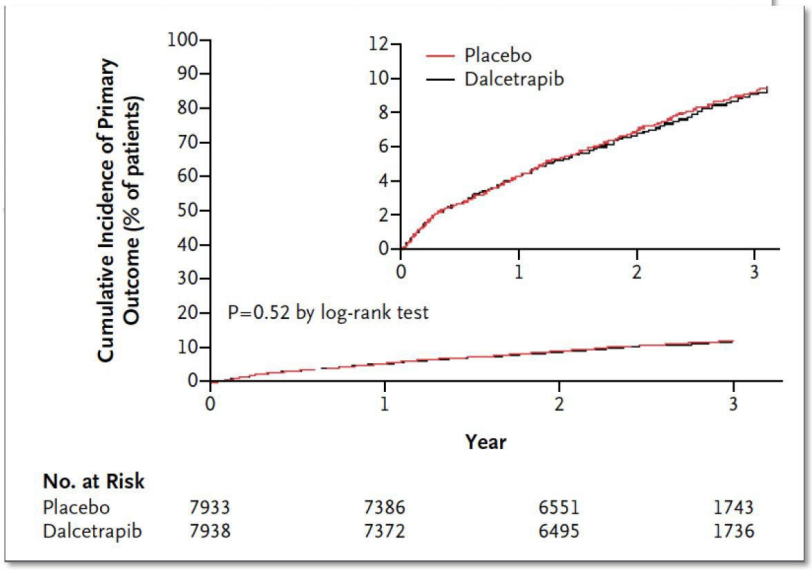
dal-OUTCOMES: treatment with dalcetrapib increased in HDL cholesterol



Schwartz, *N Engl J Med* 2012

Slides from Sek Kathiresan

...but failed to reduced CVD risk



Schwartz,
N Engl J Med 2012

16,000-person randomized controlled trial

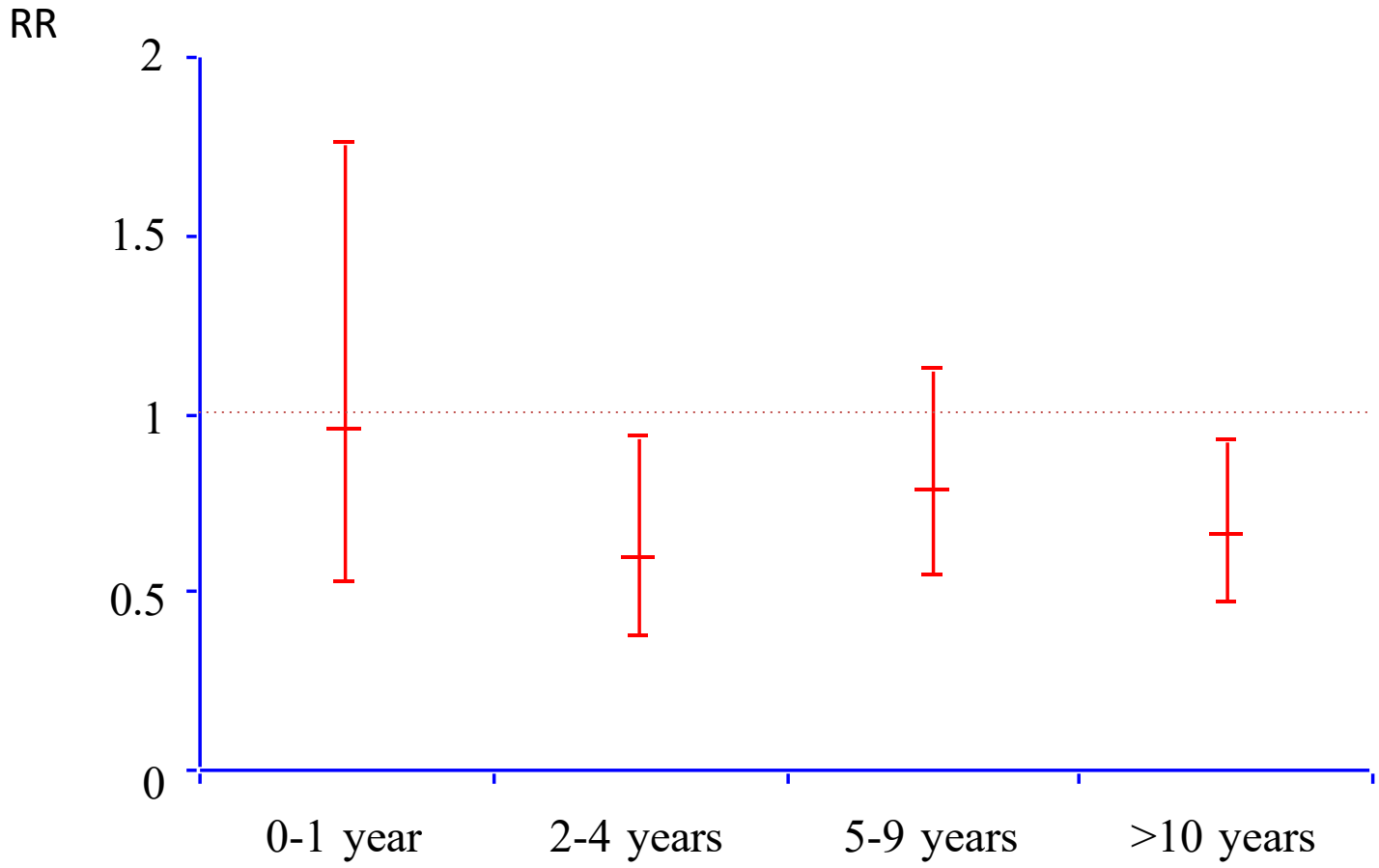
Mendelian Randomization

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

Problems with inferring causality in observational studies

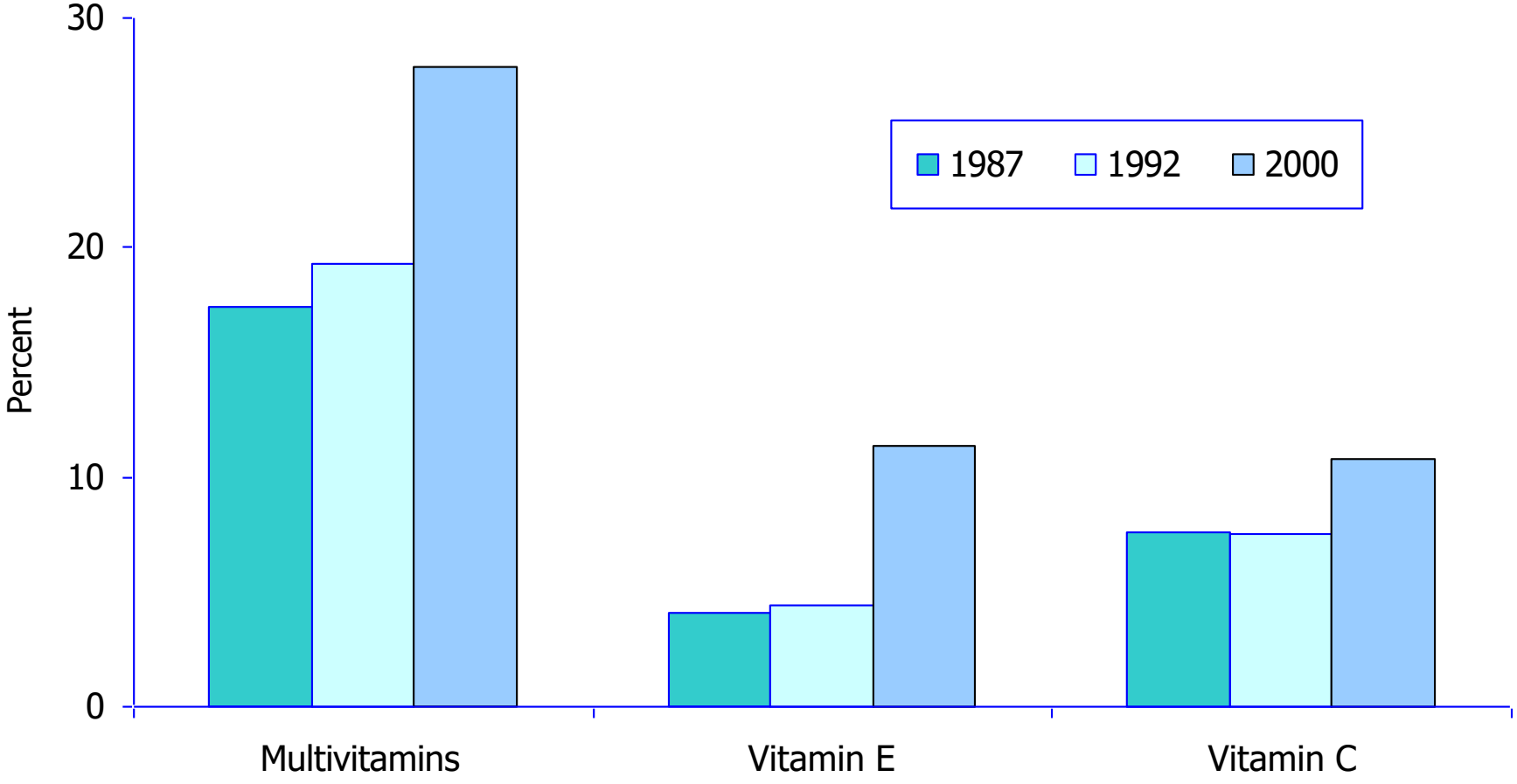
The Problem with Inferring Causality in Observational Studies

CHD risk according to duration of current Vitamin E supplement use compared to no use



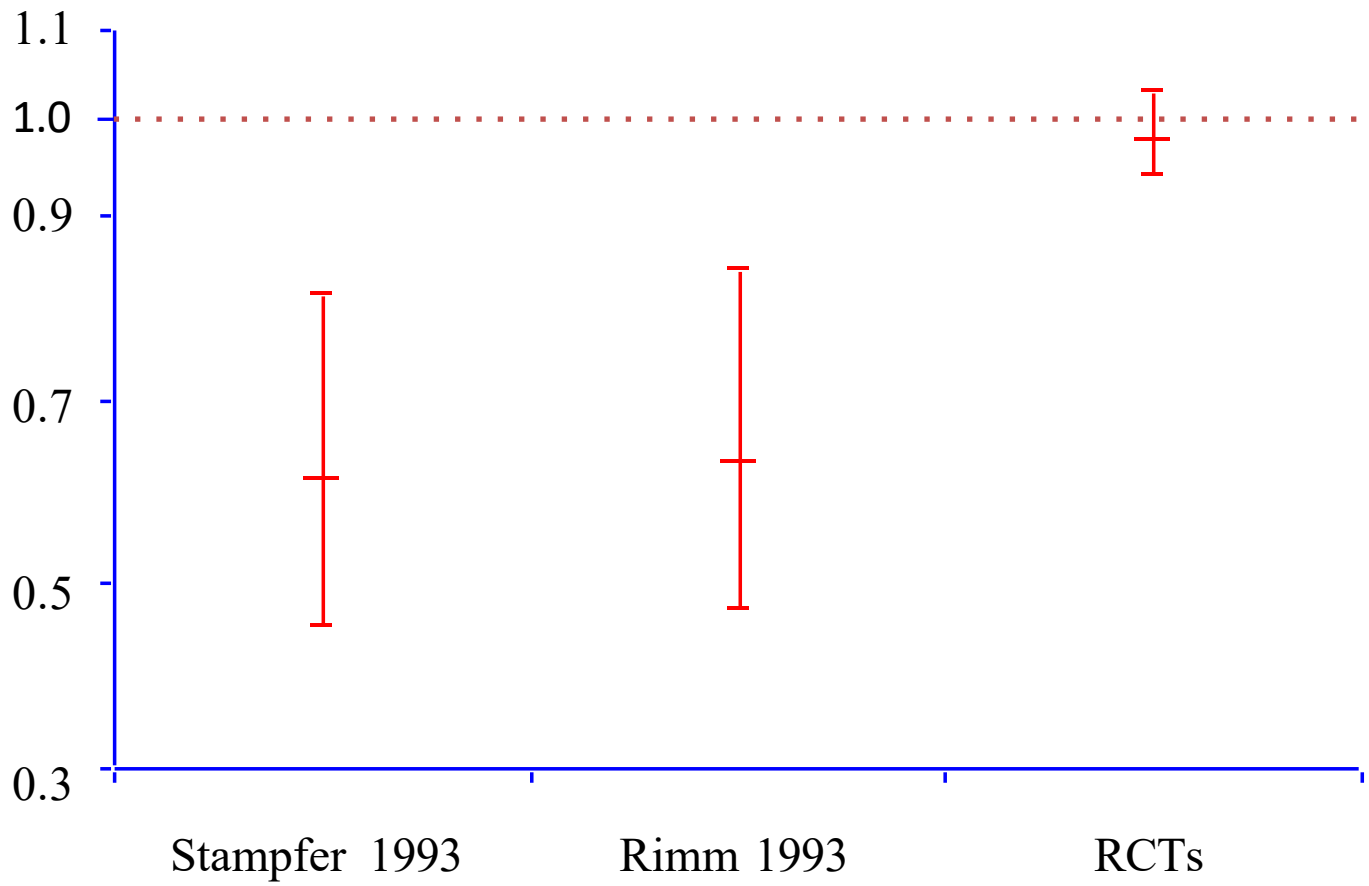
Rimm et al NEJM 1993; 328: 1450-6

Use of vitamin supplements by US adults, 1987-2000



Source: Millen AE, Journal of American Dietetic Assoc 2004;104:942-950

Vitamin E supplement use and risk of Coronary Heart Disease



Stampfer et al NEJM 1993; 328: 144-9; Rimm et al NEJM 1993; 328: 1450-6; Eidelman et al Arch Intern Med 2004; 164:1552-6

MANY OTHER EXAMPLES

VITAMIN C, VITAMIN A, HRT,
MANY DRUG TARGETS…….

WHAT'S THE EXPLANATION?

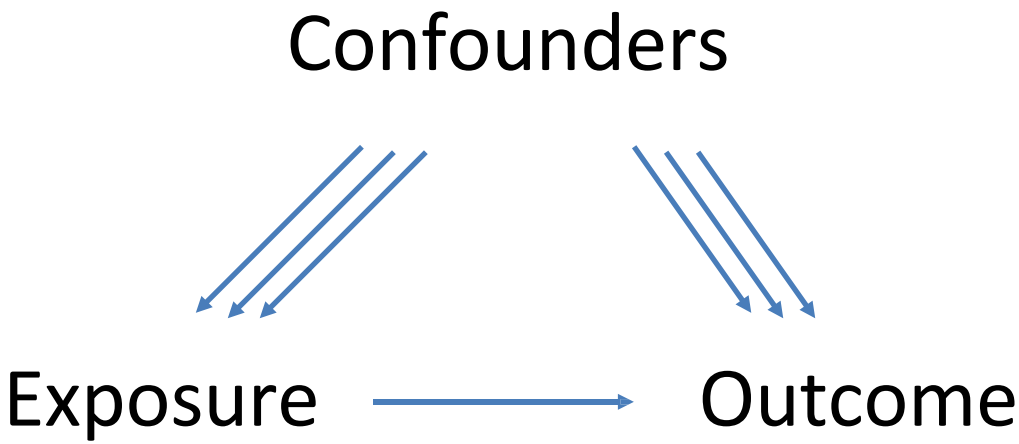
Vitamin E levels and confounding risk factors:

Childhood SES	↓
Manual social class	↓
No car access	↓
State pension only	↓
Smoker	↓
Obese	↓
Daily alcohol	↑
Exercise	↑
Low fat diet	↑
Height	↑
Leg length	↑

Women's Heart and Health Study
Lawlor et al, Lancet 2004

Confounding

Smoking, diet, alcohol, socioeconomic position….

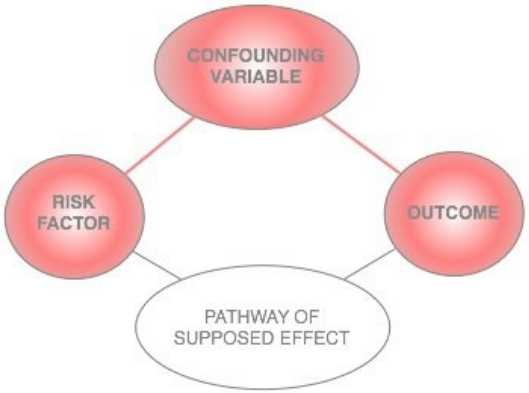


Vitamin E

Heart disease

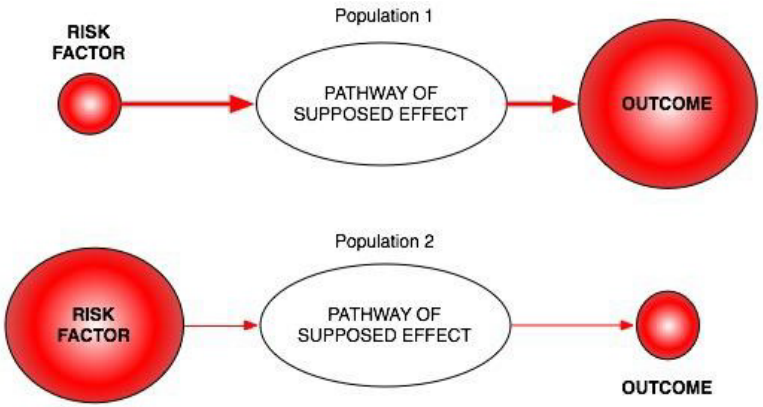
Classic limitations to "observational" science

- **Confounding**



- **Reverse Causation**

- Bias

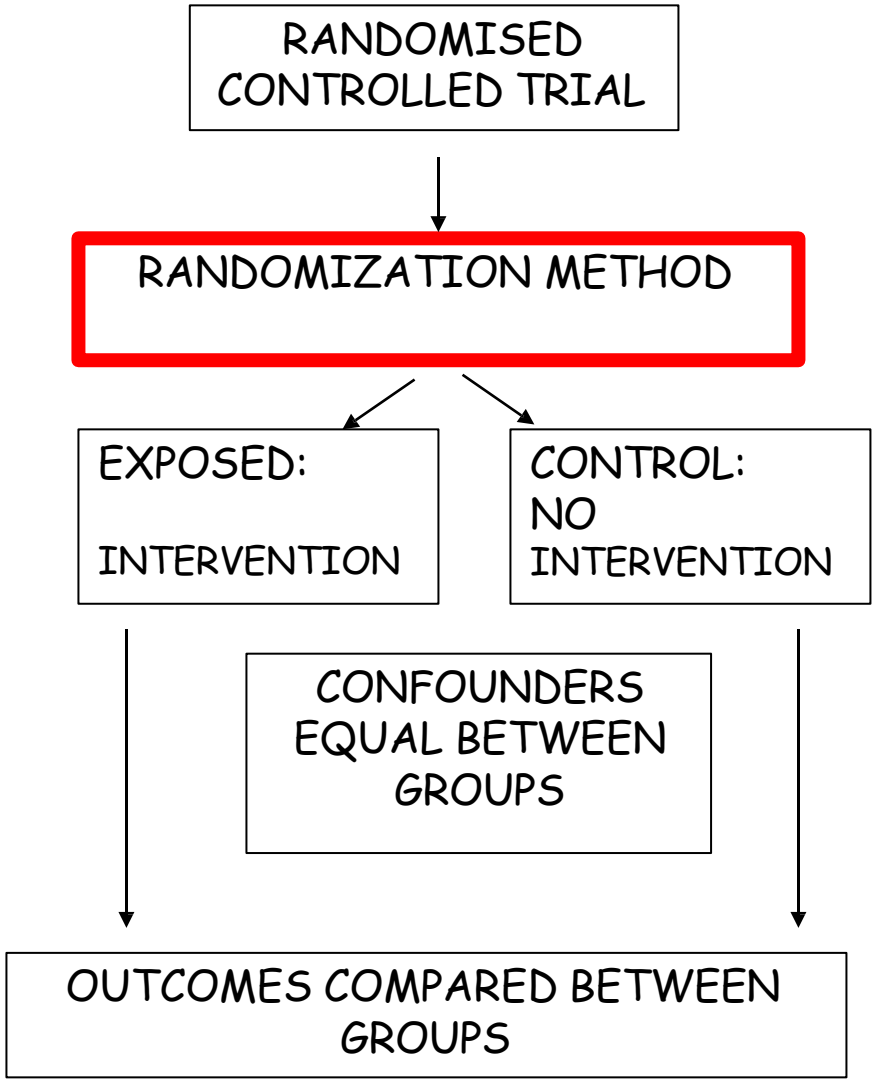


Mendelian Randomization

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

RCTs: the Gold Standard in Inferring Causality

Randomization makes causal inference possible



The Need for Observational Studies

- **Randomized Controlled Trials (RCTs):**
  - Not always ethical or practically feasible eg anything toxic
  - Expensive, requires experimentation in humans
  - Impractical for long follow up times
  - Should only be conducted on interventions that show very strong observational evidence in humans

- **Observational studies:**
  - Association between environmental exposures and disease measured in observational designs (non-experimental)
    eg case-control studies or cohort studies
  - Reliably assigning causality in these types of studies is ***very limited***

The Wide Applicability of MR

- **Traditional Observational Epidemiological Studies**
- **Behavior Genetics and the Social Sciences**
- **Molecular Studies**
- **Pharmacogenomics**

Mendelian Randomization

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

How does Mendelian randomization work?

What does MR do?

- **Assess causal relationship between two variables**
- **Estimate magnitude of causal effect**

How does it do this?

By harnessing Mendel's laws of inheritance

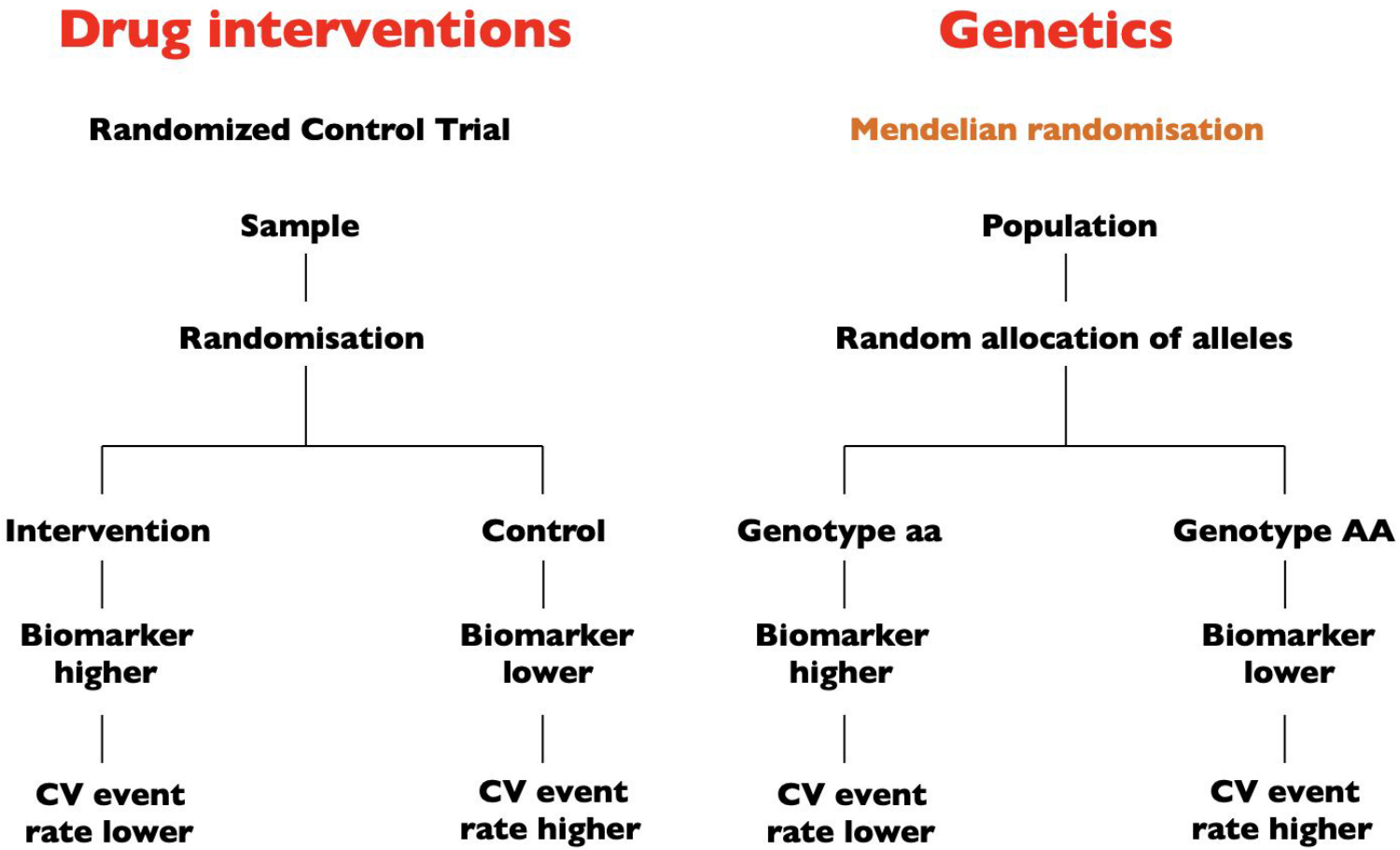
Mendel's Laws of Inheritance

Mendel in 1862

1. **Segregation:** alleles separate at meiosis and a randomly selected allele is transmitted to offspring

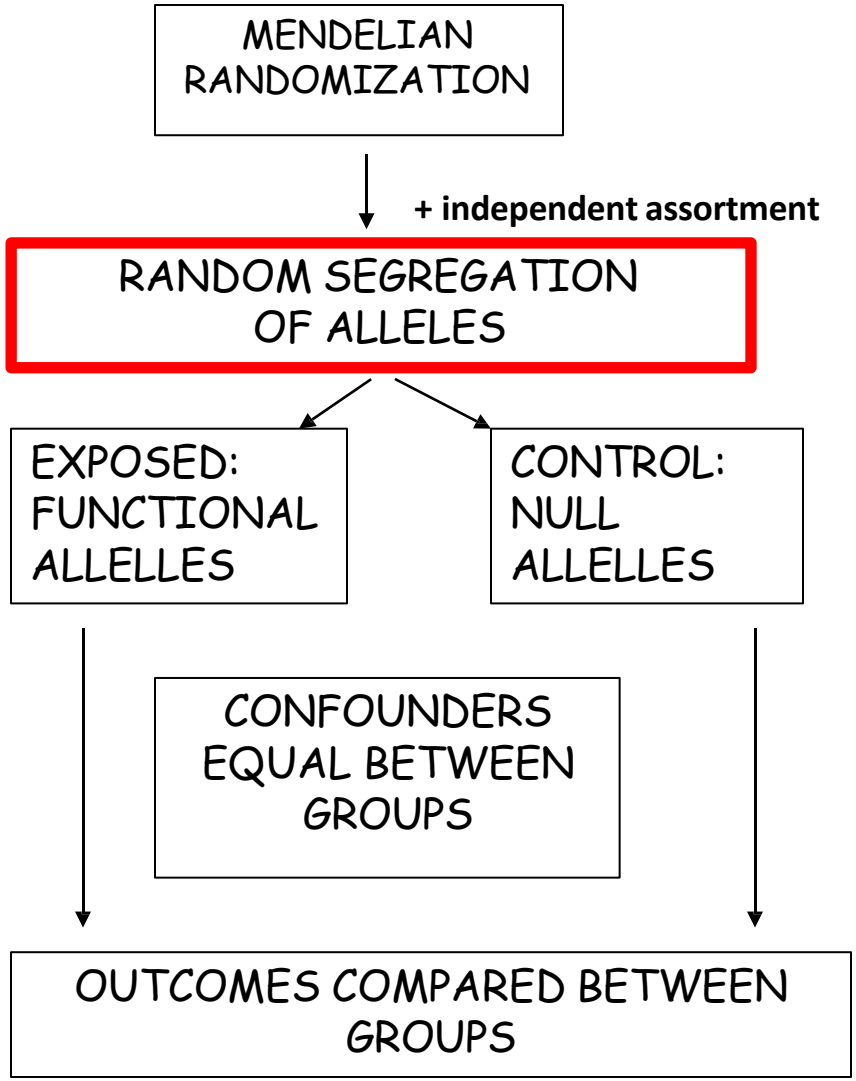
2. **Independent assortment:** alleles for separate traits are transmitted independently of one another

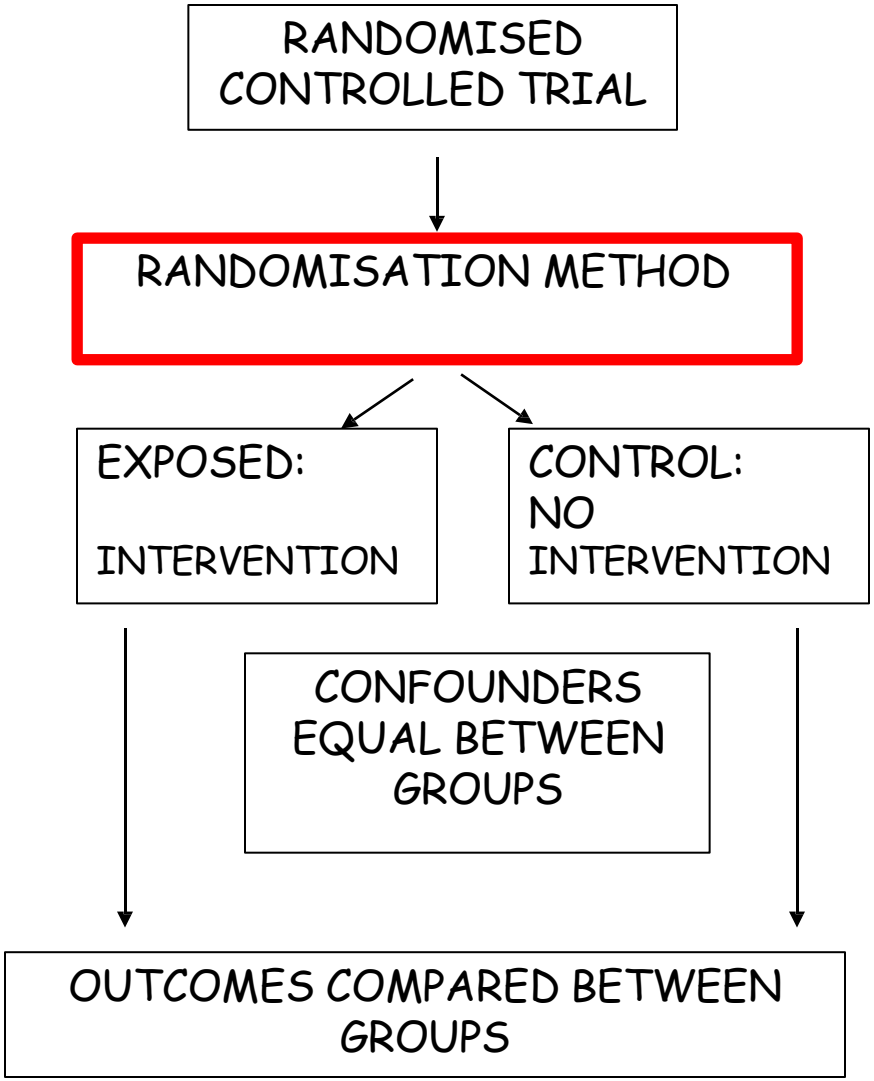
Treat genetics as randomized assignment variable



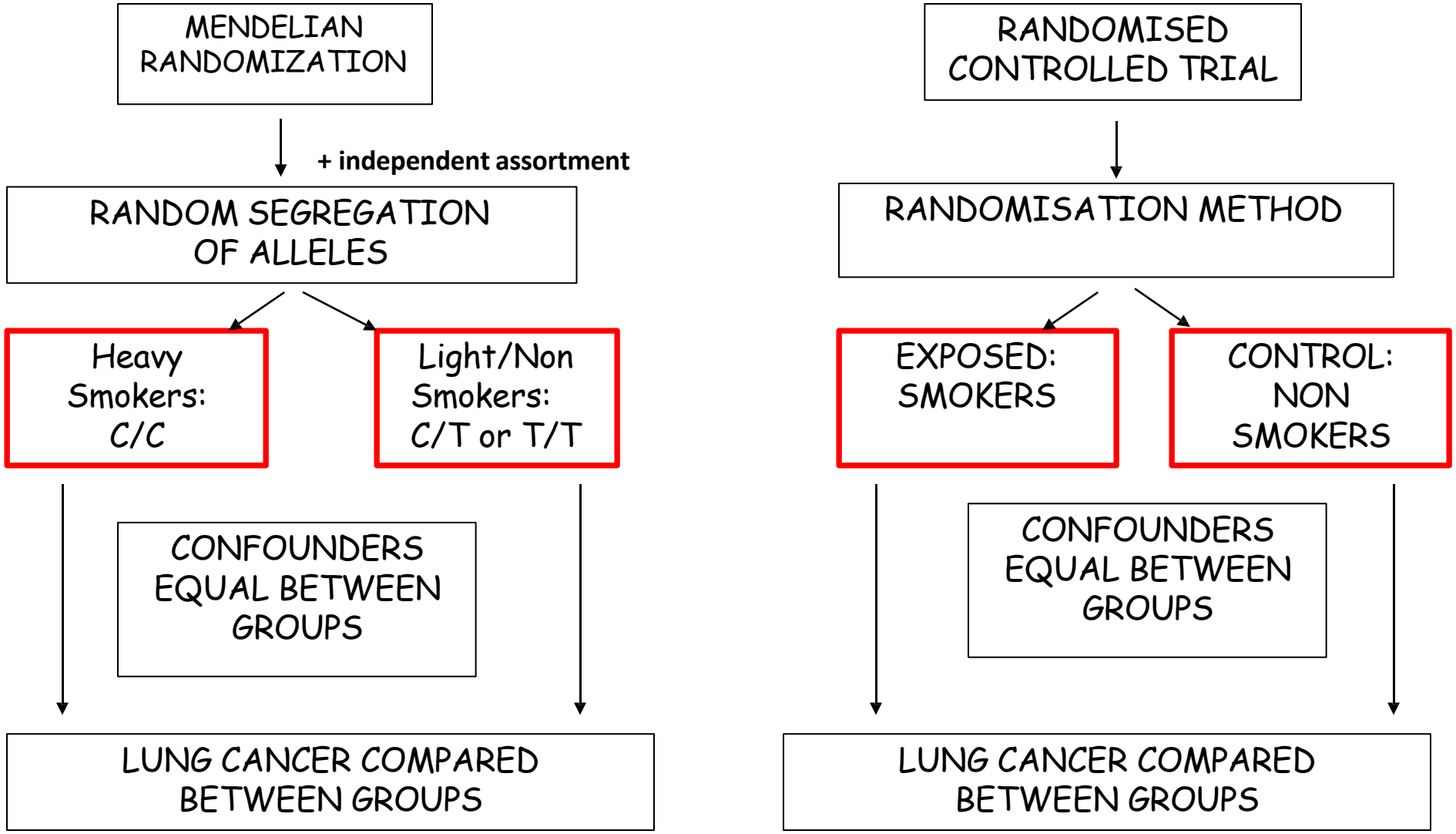
Slide courtesy of John Danesh
Hingorani et al, *Lancet* 2005

Mendelian randomization and RCTs

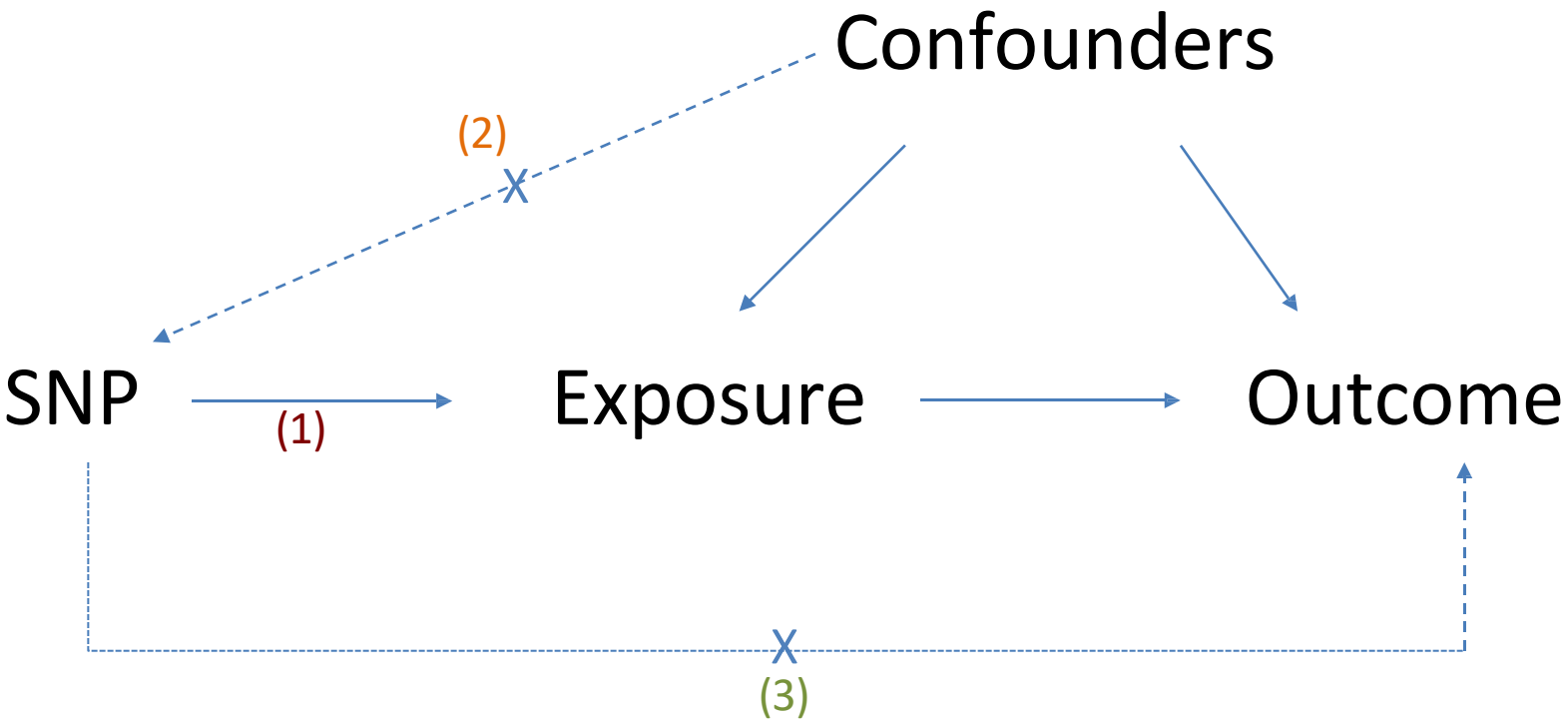




Mendelian randomization: Smoking and Lung Cancer



Mendelian Randomization: 3 Core Assumptions



(1) SNP is associated with the exposure

(2) SNP is NOT associated with confounding variables

(3) SNP ONLY associated with outcome through the exposure

Why are genetic associations special?

- Robustness to confounding due to Mendel's laws:
  - Law of segregation: inheritance of an allele is random and independent of environment etc
  - Law of independent assortment: genes for different traits segregate independently (assuming not in LD)

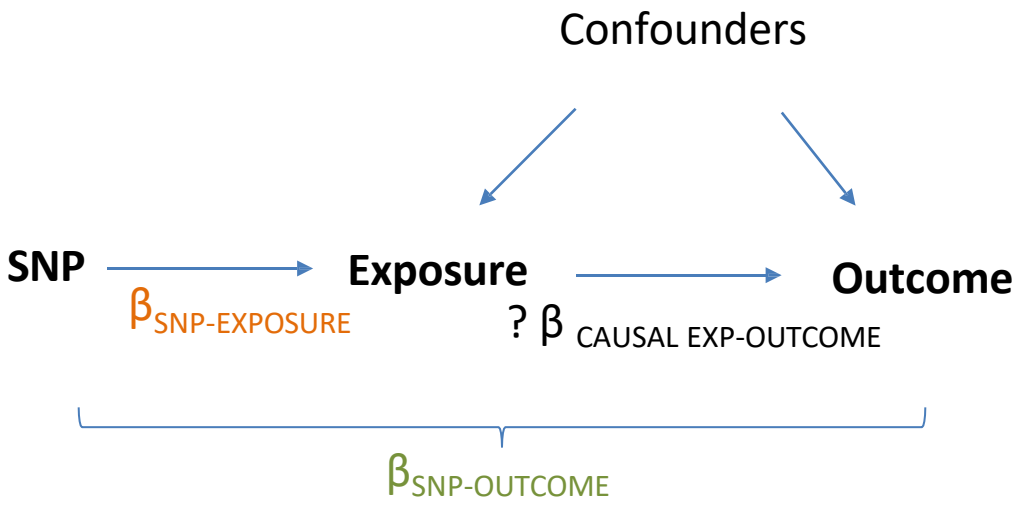
- The direction of causality is known – always from SNP to trait
- Genetic variants are **potentially** very good instrumental variables
- Using genetic variants as IVs is a special case of IV analysis, known as Mendelian randomization

Mendelian Randomization

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

Calculating causal effect estimates

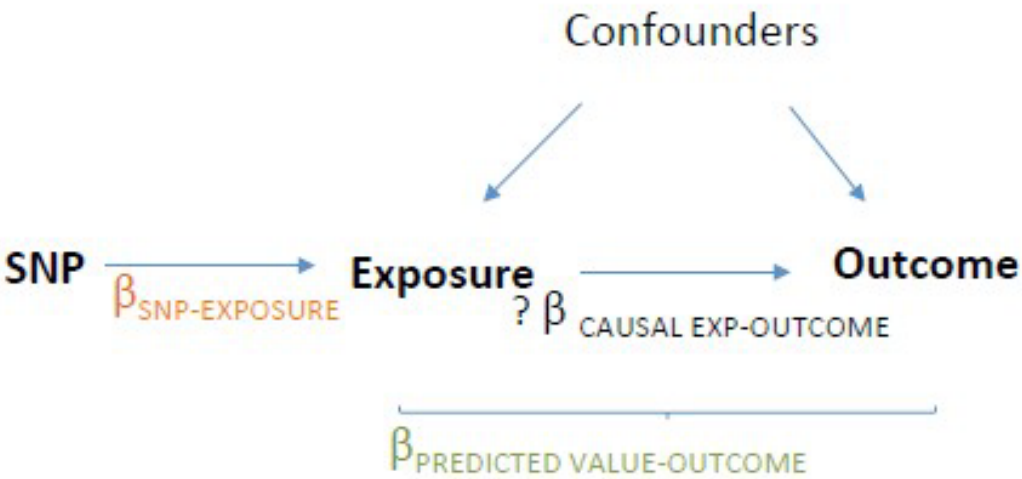
Calculating Causal Effect Estimates

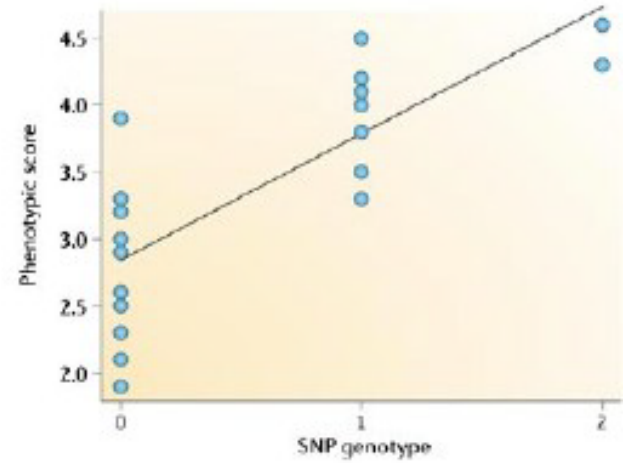


After SNP identified robustly associated with exposure of interest:

- Wald Estimator
- Two-stage least-squares (TSLS) regression

Calculating Causal Effect Estimates





Two-stage Least Squares (2SLS):

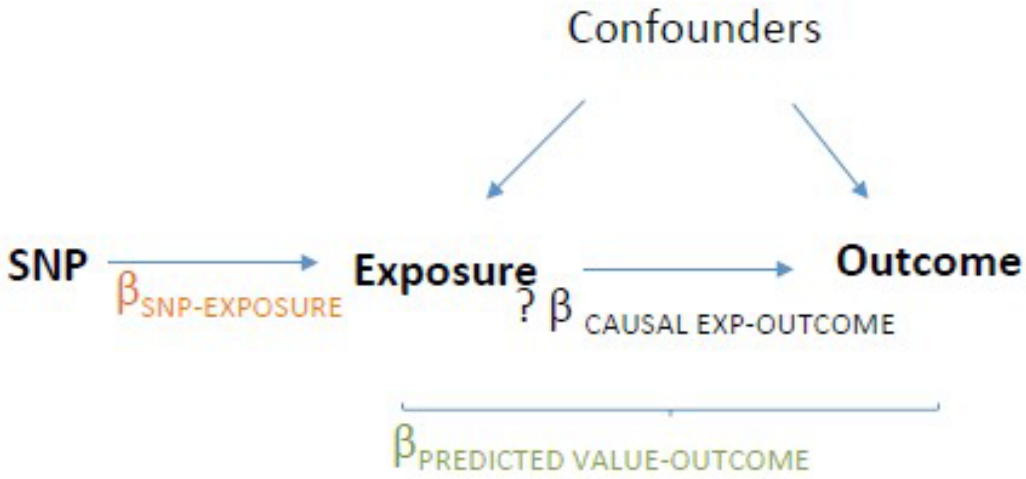
(1) Regress exposure on SNP & obtain predicted values

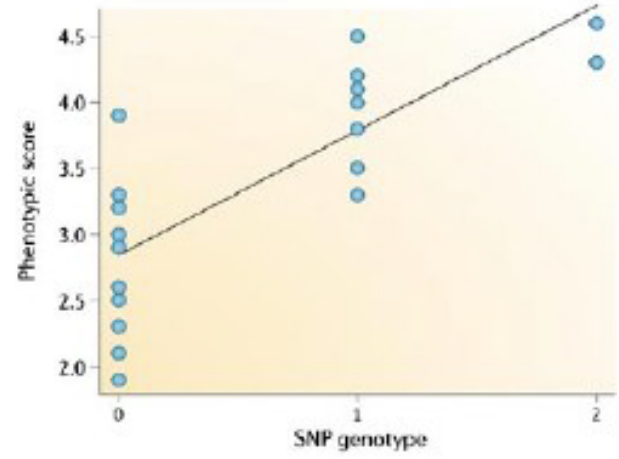
(2) Regress outcome on **predicted** exposure (from 1st stage regression)

(3) Adjust standard errors

*Needs to be done in the one sample ("Single sample MR")

Calculating Causal Effect Estimates





Two-stage Least Squares (2SLS):

(1) Regress exposure on SNP & obtain predicted values

(2) Regress outcome on **predicted** exposure (from 1st stage regression)

(3) Adjust standard errors

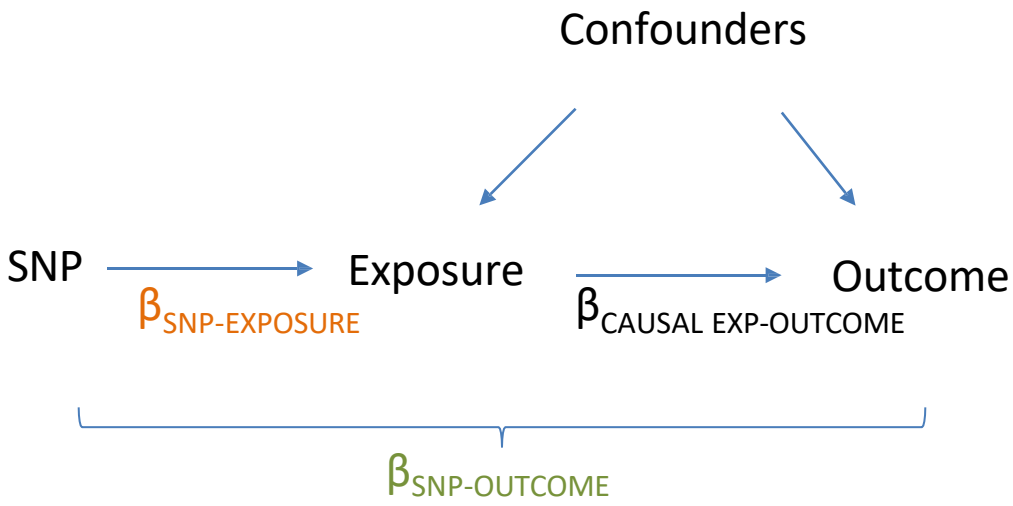
This gives you: difference in outcome per unit change in (genetically-predicted) exposure

Genetically determined exposure → "randomized" → can ascribe causality

(if assumptions are met)

*Needs to be done in the one sample ("Single sample MR")

Calculating Causal Effect Estimates



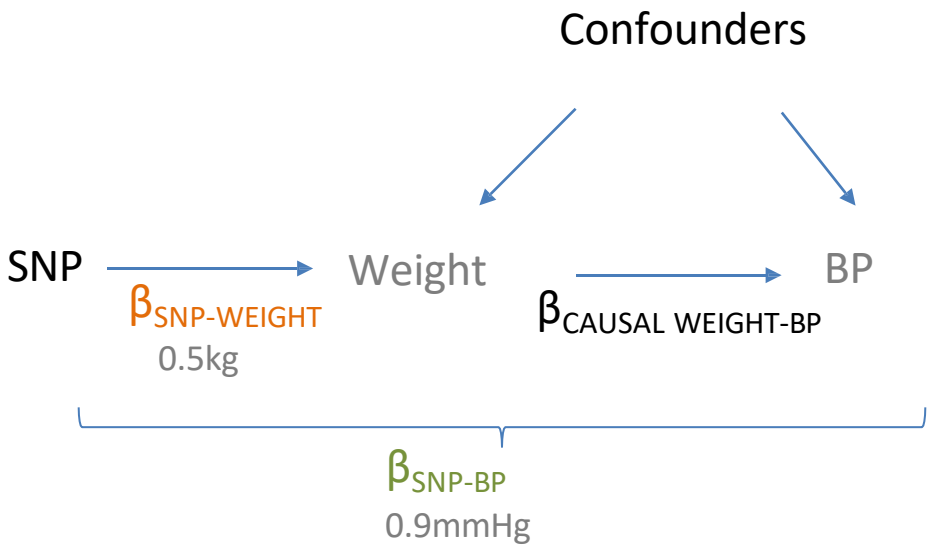
Causal effect by Wald Estimator* :

$$\frac{\hat{\beta}_{SNP\text{-}OUTCOME}}{\hat{\beta}_{SNP\text{-}EXPOSURE}}$$

$$\beta_{SNP\text{-}OUTCOME} = \beta_{CAUSAL\ EXP\text{-}OUTCOME} \times \beta_{SNP\text{-}EXPOSURE}$$

*Can be used in different samples ("Two sample MR")

Calculating Causal Effect Estimates



Causal effect by Wald Estimator* :

$$\frac{\hat{\beta}_{SNP\text{-}OUTCOME}}{\hat{\beta}_{SNP\text{-}EXPOSURE}}$$

= change in outcome per unit change in exposure

BP and weight:

$$\frac{0.9\ \text{mmHg/allele}}{0.5\ \text{kg/allele}}$$

=1.8 mmHg/kg

*Can be used in different samples ("Two sample MR")

MR can also be performed using just the results from GWAS

- Also known as two-sample MR, SMR, or MR with summary data etc
- Advantages:
  - The data is readily available, non-disclosive, free, open source
  - The exposure and outcome might not be measured in the same sample
  - The sample size of the outcome variable, key to statistical power, is not limited by requiring overlapping measures of the exposure
- Disadvantages:
  - Some extensions of MR not possible, e.g. non-linear MR, use of GxE for negative controls, various sensitivity analyses

Mendelian Randomization

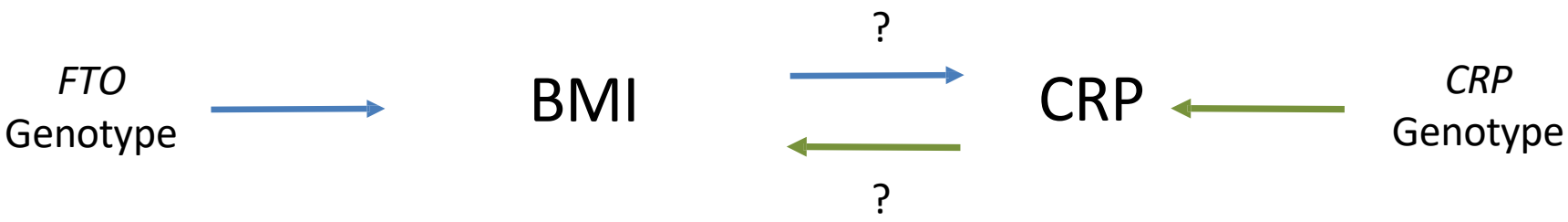
- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

An Example using Mendelian randomization

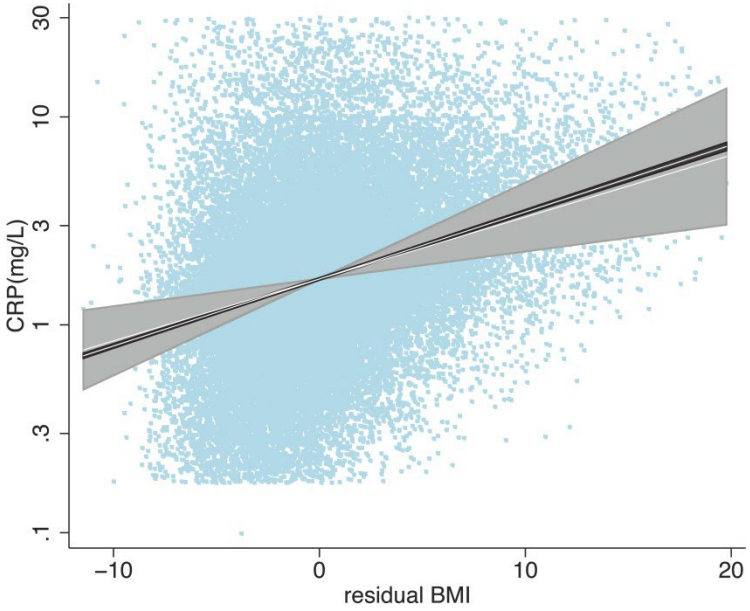
MR Example using CRP

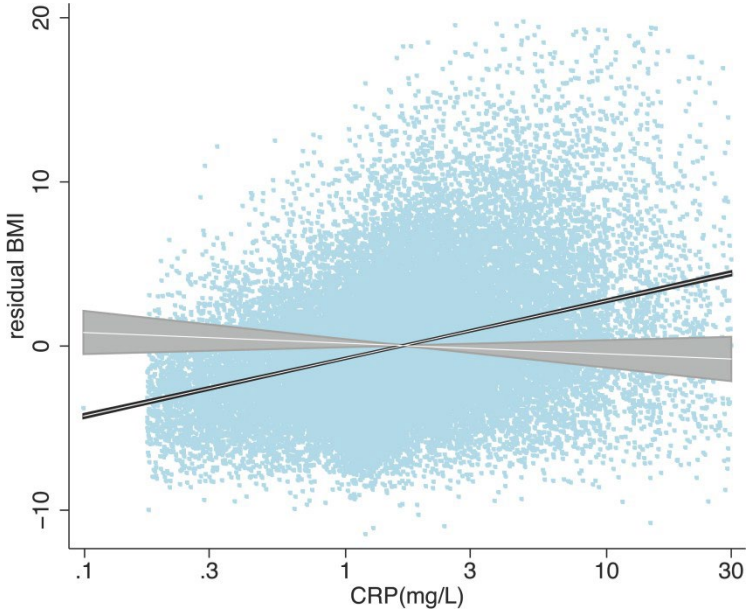
- C-Reactive Protein (CRP) is a biomarker of inflammation
- It is associated with BMI, metabolic syndrome, CHD and a number of other diseases
- It is unclear whether these observational relationships are causal or due to confounding or reverse causality
- This question is important from the perspective of intervention and drug development

“Bi-directional Mendelian Randomization”: Testing causality and reverse causation



	Effect estimates				
Outcome / explanatory variable	**Observational**	**Instrumental variable**	P_{IV}	P_{diff}	F_{first}
CRP/BMI	1.075 (1.073, 1.077)	1.06 (1.02, 1.11)	0.002	0.6	50.2
BMI/CRP	1.58 (1.53, 1.62)	-0.30 (-0.78, 0.18)	0.2	<0.00001	78.3





Mendelian Randomization

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

Limitations to Mendelian randomization

Limitations to Mendelian Randomization

1 Violations of assumptions

2 Population stratification

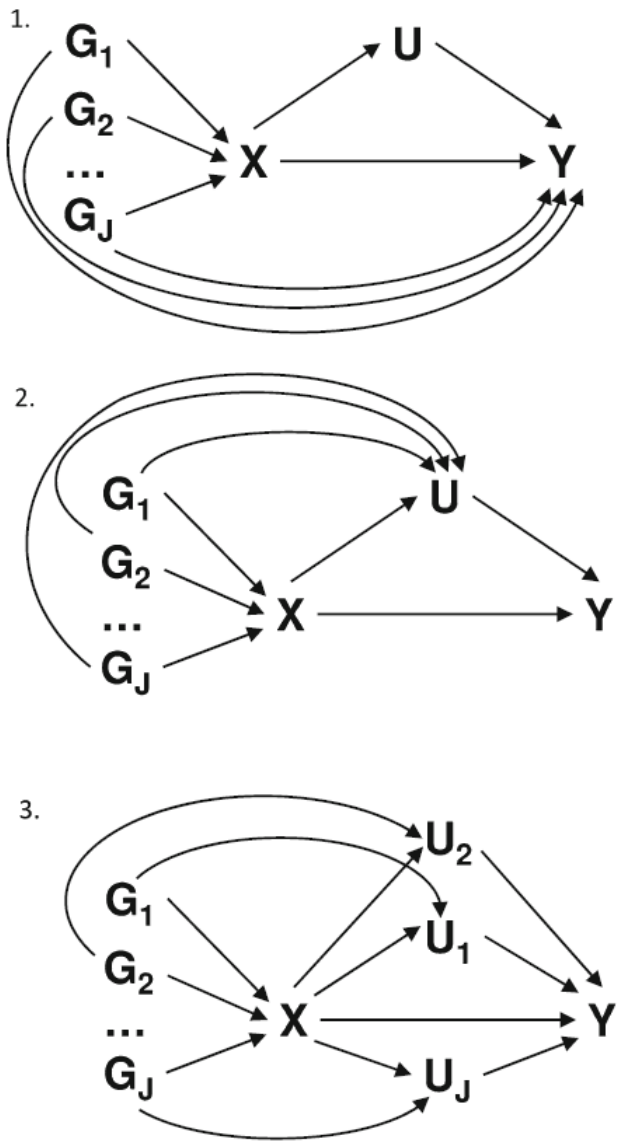
3 Canalisation ("Developmental compensation")

4 The existence of instruments

5 Power and "weak instrument bias"

6 Pleiotropy

Assumption: INstrument Strength Independent of Direct Effect (InSIDE)



<u>Top: okay.</u> pleiotropic effects act directly on the outcome (InSIDE satisfied)

Middle: pleiotropic effects act on the outcome via single confounder
(InSIDE violated)

Bottom: pleiotropic effects act on the outcome via different confounders
(InSIDE still violated).

- Arrows from the genetic variants to the risk factor may not be present for all variants
- some variants may affect the confounder directly and not the risk factor.

<u>Notation:</u>

G1, G2, . . ., GJ , genetic variants
X, risk factor
Y, outcome
U, confounder.
Curved arrows: Pleiotropic effects

Power and Weak Instruments

- Power:
  - Genetic variants explain very small amounts of phenotypic variance in a given trait
  - VERY large sample sizes are generally required

- Weak instruments:
  - Genetic variants that are weak proxies for the exposure
  - Results in biased causal estimates from MR

- Different impact of the bias from weak instruments:
  - **Single Sample MR:** to the confounded estimate
  - **Two-Sample MR:** to the null

Using Multiple Genetic Variants as Instruments

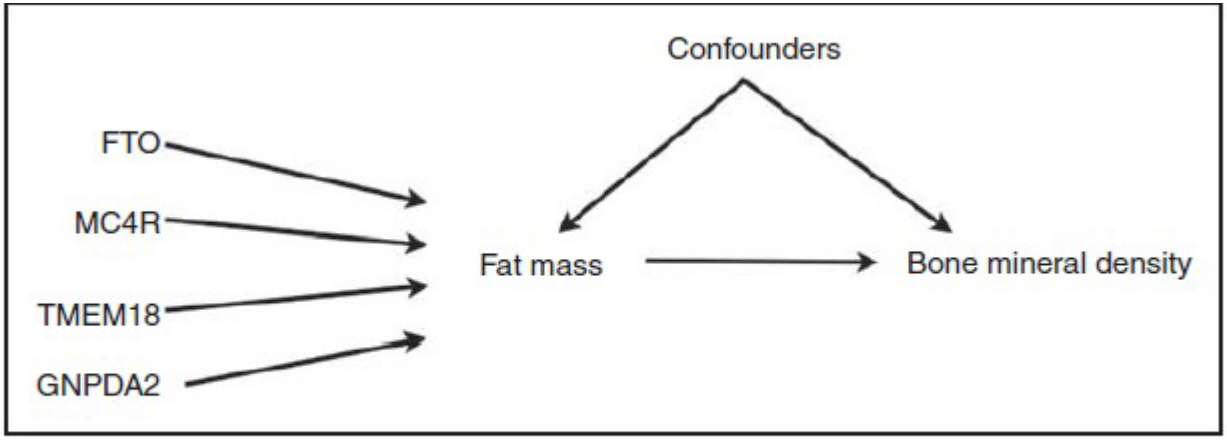


Figure 1. DAG for a Mendelian randomisation analysis using four genetic variants as instrumental variables for the effect of fat mass on bone mineral density.

Palmer et al (2011) Stat Method Res

- Allelic scores
- Testing multiple variants individually
- Meta-analyse individual SNPs

Calculating Power in Mendelian Randomization Studies

http://cnsgenomics.com/shiny/mRnd/ | mRnd: Power calculations f...

File Edit View Favorites Tools Help

mRnd: Power calculations for Mendelian Randomization

Input

Calculate:

- (●) Power
- ( ) Sample size

Provide:

Sample size

1000

α

0.05

Type-I error rate

β_{yx}

0

Continuous outcome | Binary outcome | Binary outcome derivations | Citation | About

Two-stage least squares

Power	0.05	
NCP	0.00	Non-Centrality-Parameter
F-statistic	11.10	The strength of the instrument

Power or sample size calculations for two-stage least squares Mendelian Randomization studies using a genetic instrument Z (a SNP or allele score), a continuous exposure variable X (e.g. body mass index [BMI, $\frac{kg}{m^2}$]) and a continuous outcome variable Y (e.g. blood pressure [mmHg]).

YZ association

Power	0.05	
NCP	0.00	Non-Centrality-Parameter

Power or sample size calculations for the regression association of a genetic instrument Z (e.g. a BMI SNP), with a continuous outcome variable Y (blood pressure).

EN 2:26 PM 6/02/2017

Limitations to Mendelian Randomization

1 Population stratification

2 Canalisation (“Developmental compensation”)

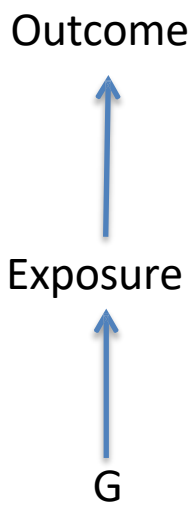
3 The existence of instruments

4 Power (also “weak instrument bias”)

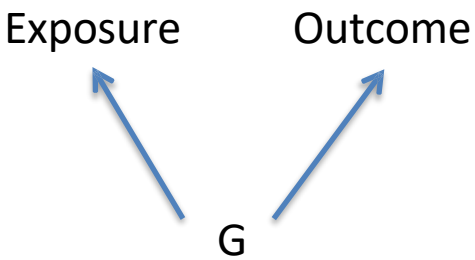
5 Pleiotropy

Pleiotropy

- Genetic variant influences more than one trait
- Horizontal vs Vertical pleiotropy



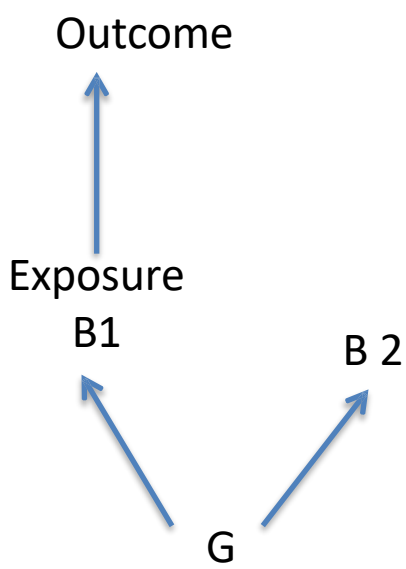
Vertical Pleiotropy

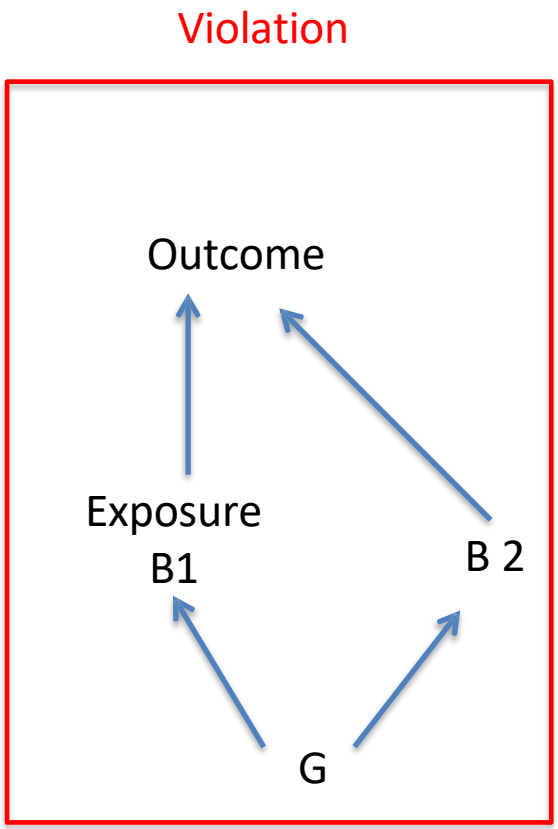


Horizontal Pleiotropy

Pleiotropy

- Genetic variant influences more than one trait
- Pleiotropy only violates MR's assumptions if it involves a pathway outside that of the exposure and is a pathway that <u>affects your outcome</u>



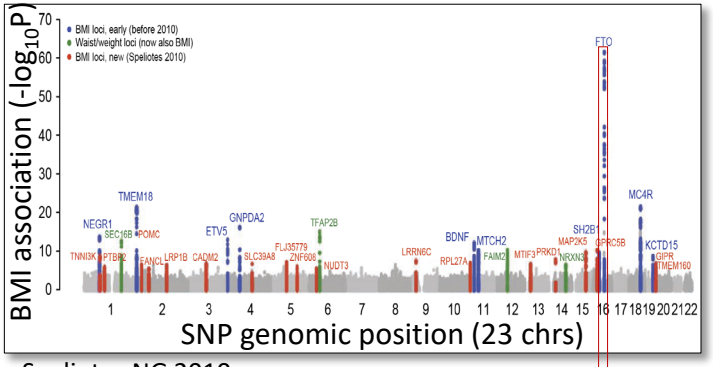


Molecular QTL mapping and causal inference for gene-regulatory mechanisms

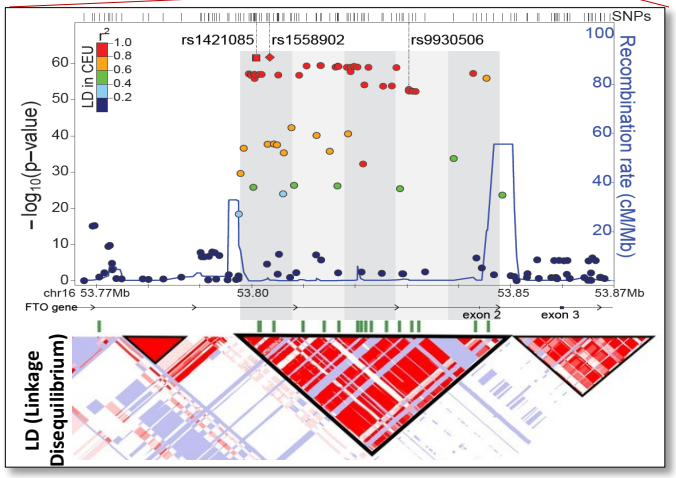
- **Concept of molecular QTL mapping**
- Basic methods for eQTL discovery
- Molecular QTL mapping in single-cell genomics
- Mediation analysis to understand mechanisms
- Causality inference: A battle against confounding variables

Genomic medicine: challenge and promises

GWAS Manhattan Plot: simple χ^2 statistical test



Speliotes NG 2010



Dina NG 2007, Frayling Science 2007, Claussnitzer NEJM 2015

The promise of genetics

- Path to causality
- Disease mechanism
- New target genes
- New therapeutics
- Personalized medicine

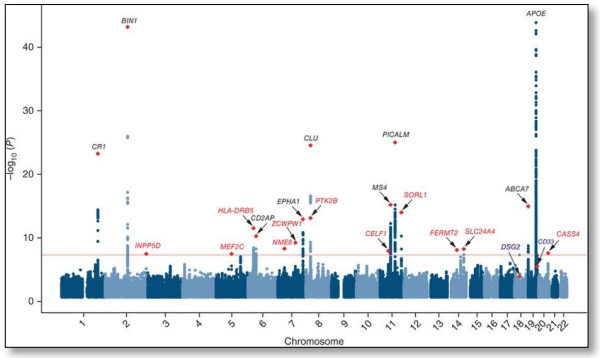
The challenge of mechanism

- **90+% disease hits non-coding**
- Target gene not known
- Causal variant not known
- Cell type of action not known
- Relevant pathways not known
- Mechanism not known

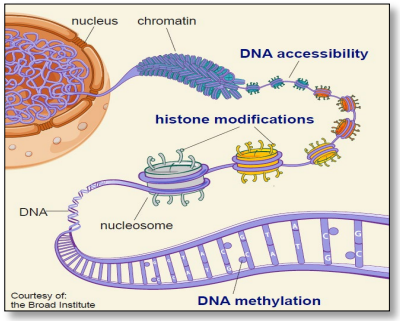
Ward NBT'12

Claussnitzer NEJM'15

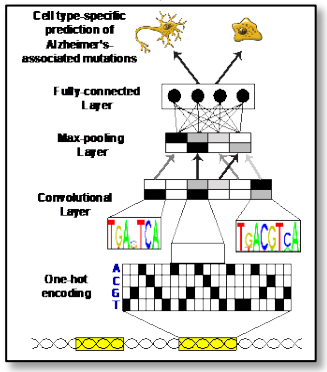
Dissect mechanisms of disease-associated regions



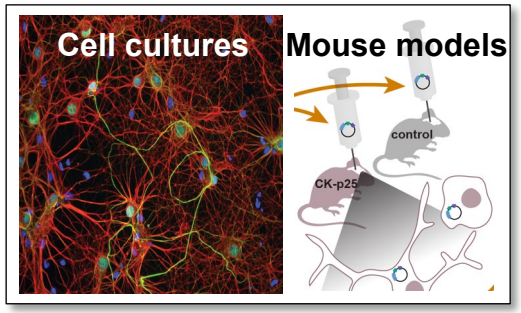
1. Disease genetics reveals common + rare variants/regions



2. Profile RNA + Epigenome in healthy + disease samples



3. Integrate data to predict driver genes, regions, cell types



4. Validate predictions in human cells + mouse models

5. Disseminate results

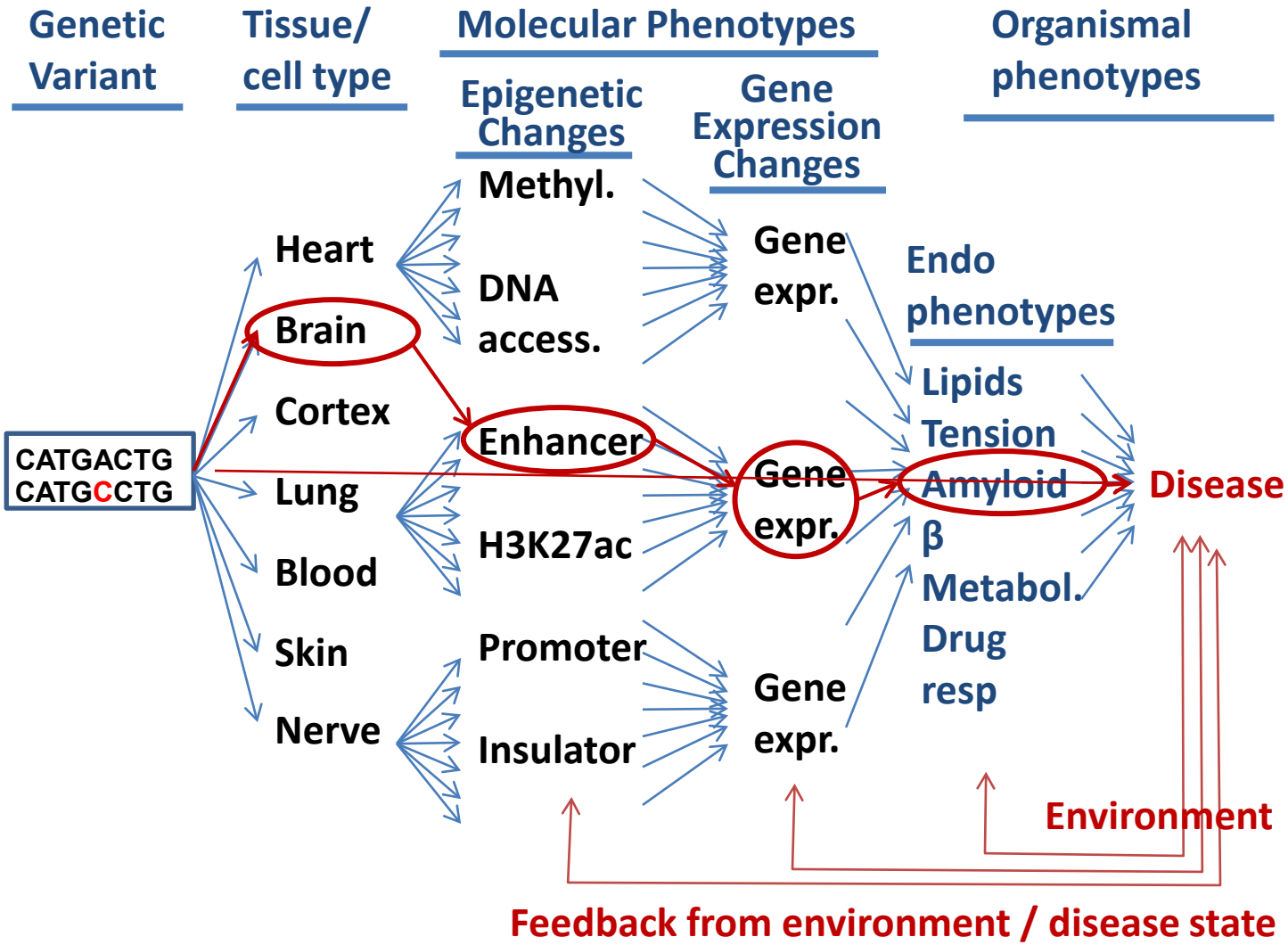
Roadmap Nature 15

Boix EpiMap Nature 21

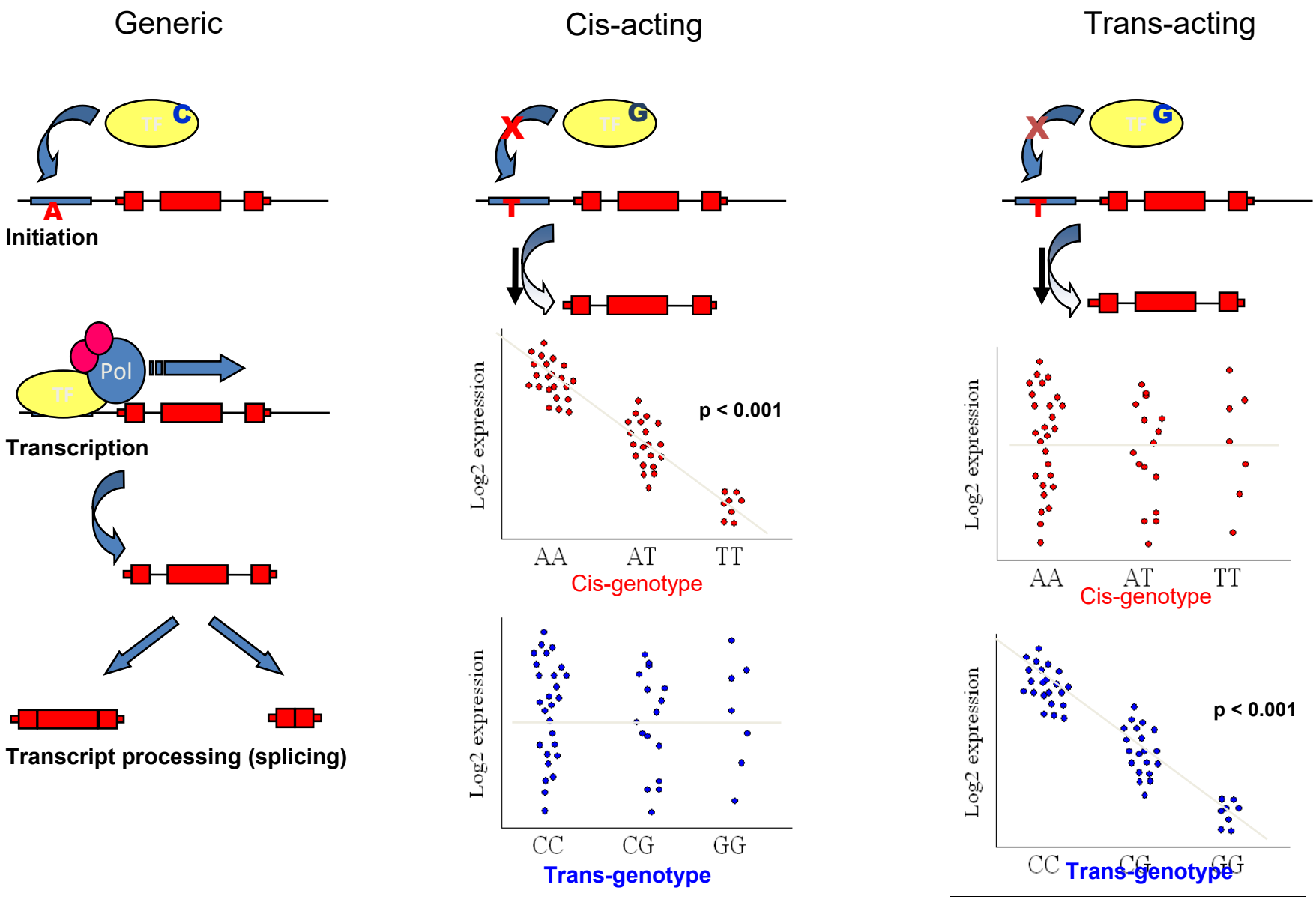
Claussnitzer NEJM'15

Blanchard, Nature, 2022

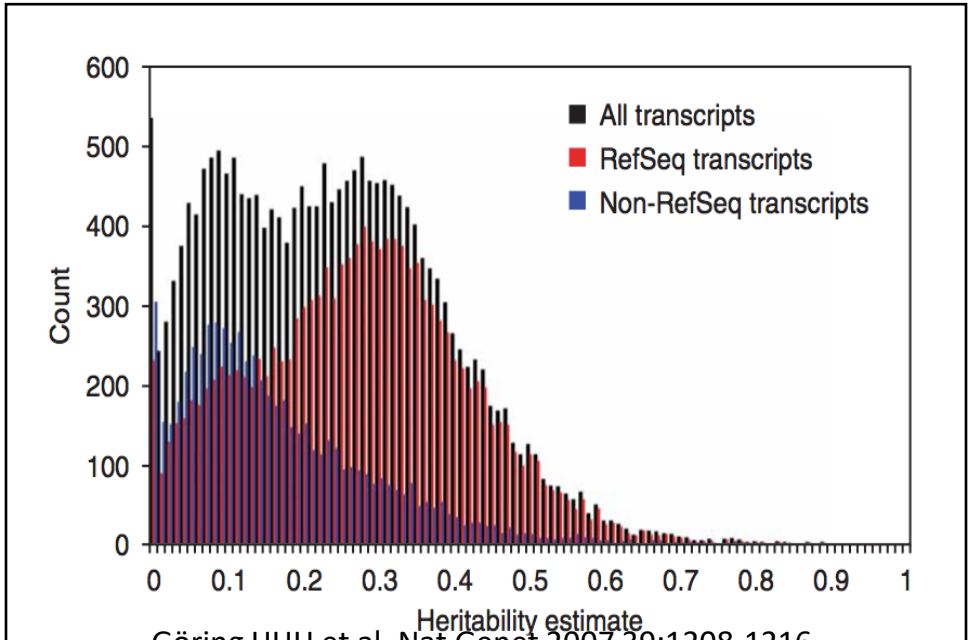
Park NBT 15


Genetic Variant
Tissue/ cell type
Molecular Phenotypes
Epigenetic Changes
Gene Expression Changes
Organismal phenotypes
CATGACTG
CATGCCTG
Heart
Brain
Cortex
Lung
Blood
Skin
Nerve
Methyl.
DNA access.
Enhancer
H3K27ac
Promoter
Insulator
Gene expr.
Gene expr.
Gene expr.
Endo phenotypes
Lipids
Tension
Amyloid β
Metabol.
Drug resp
Disease
Environment
Feedback from environment / disease state

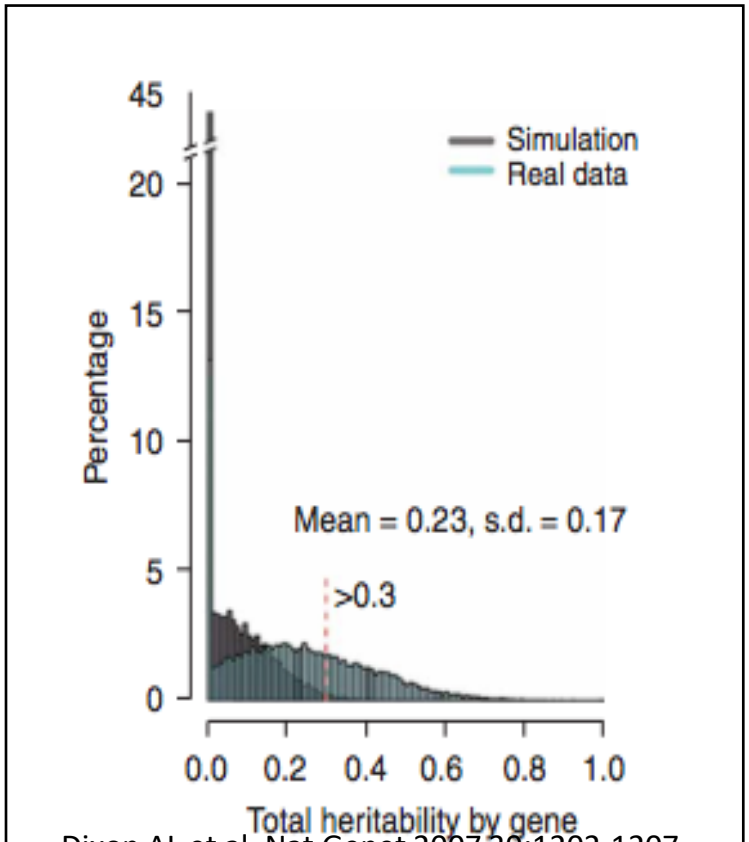
eQTL mapping: a population genetic approach for regulatory variant identification



Gene expression is a heritable trait

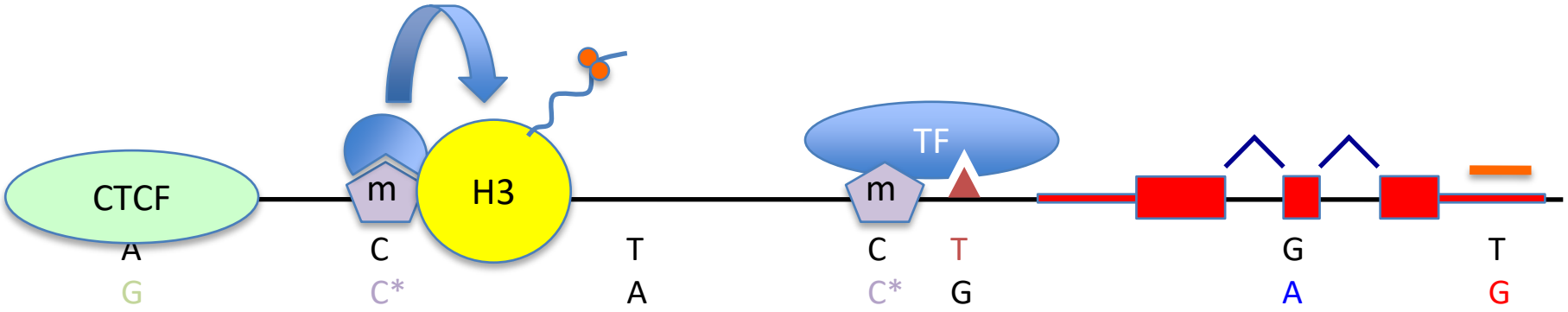


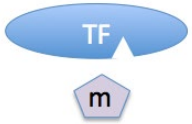
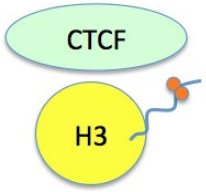
Göring HHH et al. Nat Genet 2007 39:1208-1216.



Dixon AL et al. Nat Genet 2007 39:1202-1207.

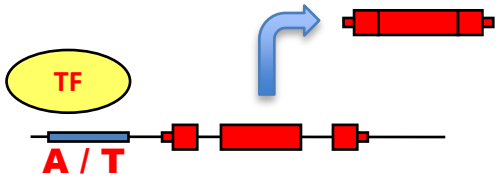
Types of regulatory variants



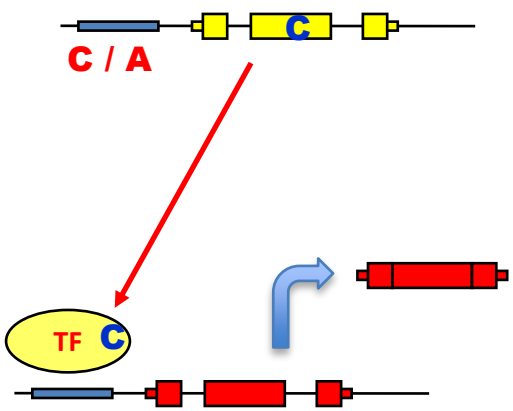
	Altered TF Binding
	Histone Modifications
	Altered splicing
	Altered miRNA silencing

Cis vs. Trans elements

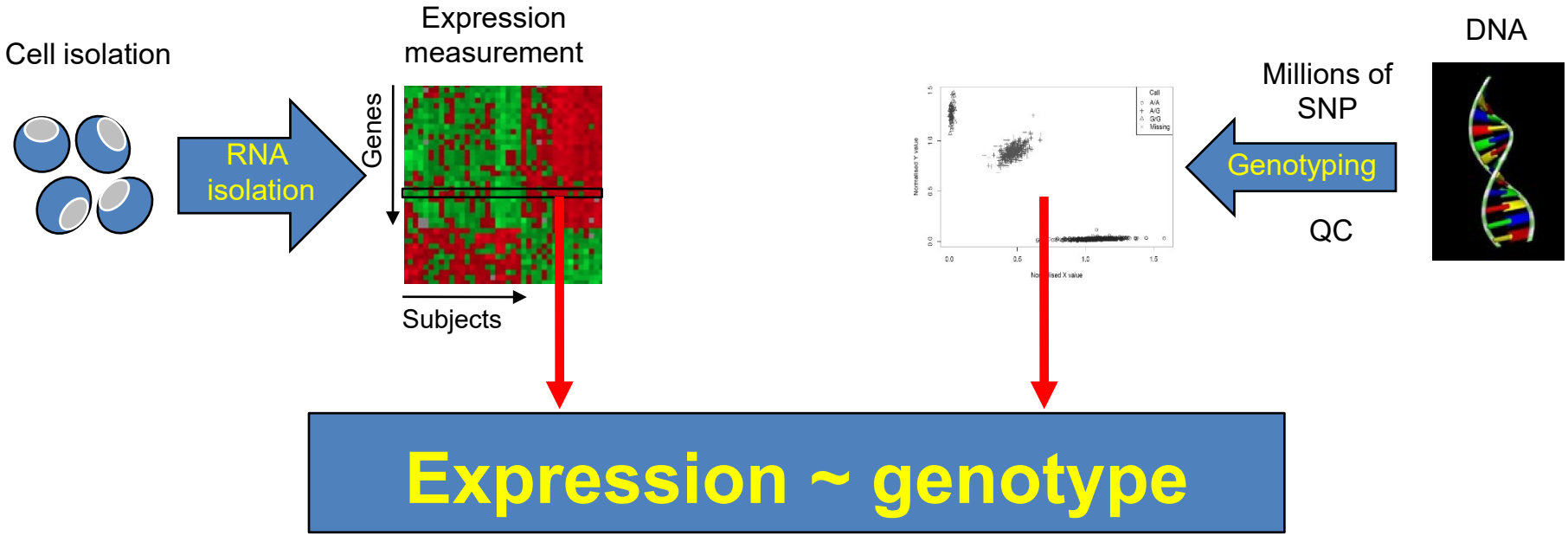
- *cis*-eQTL: variant resides in close proximity to target gene location
  - Multiple mechanisms implicated
    - Promoter
    - Splicing
    - Methylation
    - Chromatin modification



- *trans*-eQTL: variant resides very distant to the target
  - Alternative chromosome
  - Same chromosome, but far away
  - Mechanisms less clear

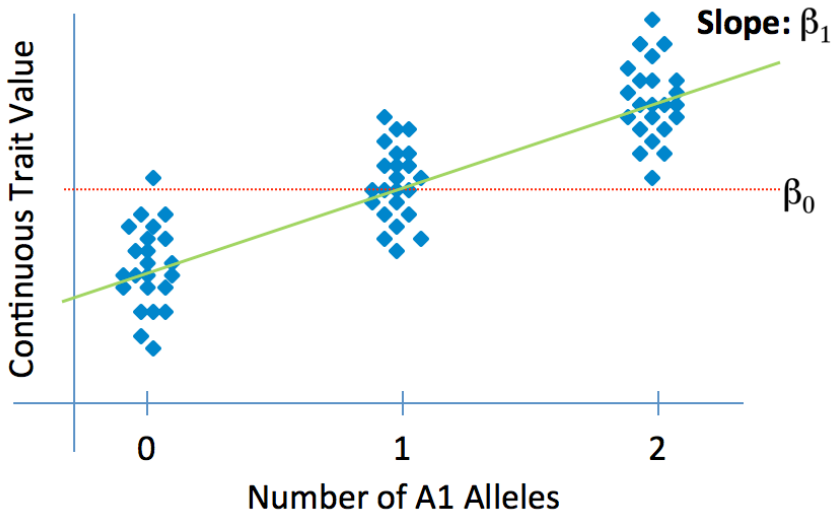


The nuts and bolts of an eQTL study



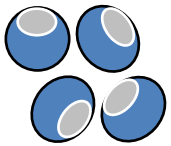
Linear Regression Equation

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$



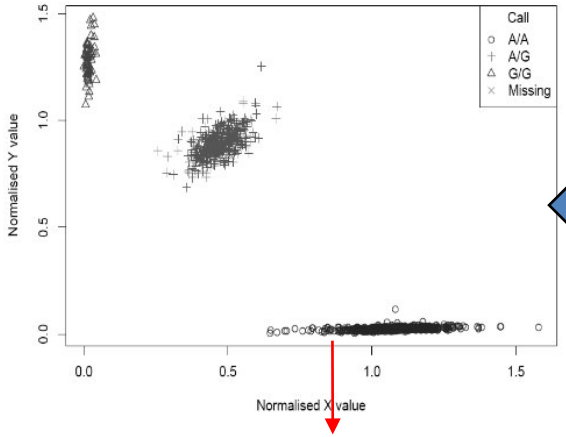
The nuts and bolts of an eQTL study

Cell isolation



RNA isolation

Expression measurement

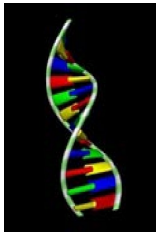


Genotyping

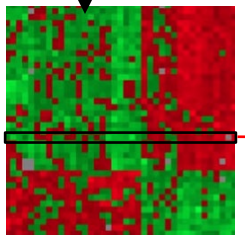
Millions of SNP

QC

DNA



Filter transcripts

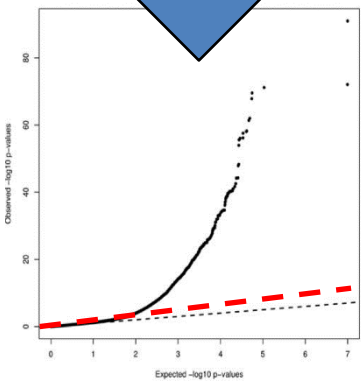


Genes

Subjects

Linear Regression
Expression = genotype + covariates

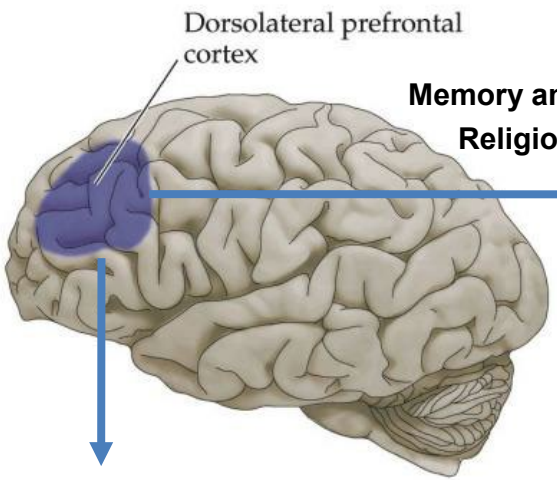
Age, gender
Pop stratification
Technical Covs



Determine significance threshold

Annotation
Visualization
Interpretation

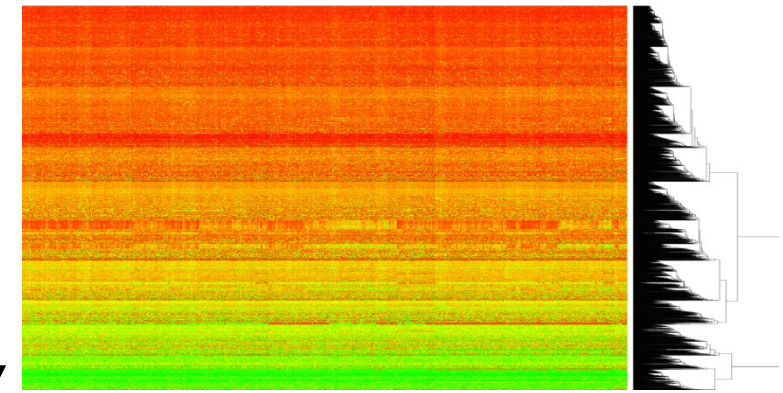
Methylation in 750 Alzheimer patients/controls

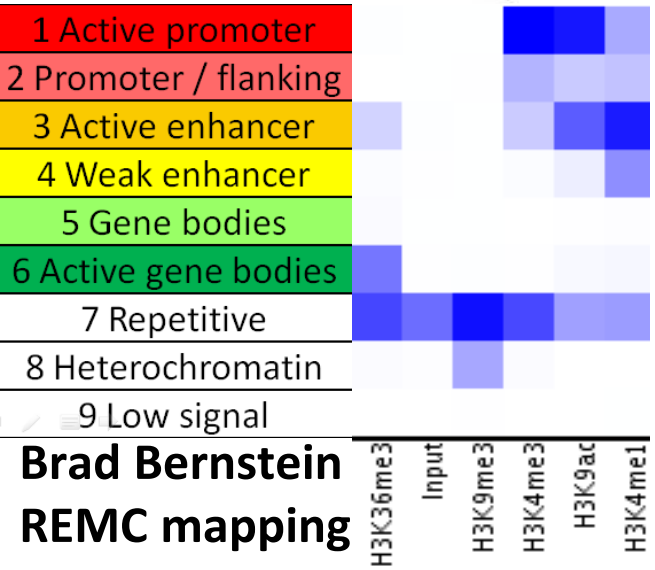


Memory and Aging Project
Religious Order Study

750 individuals (~50% w/AD)

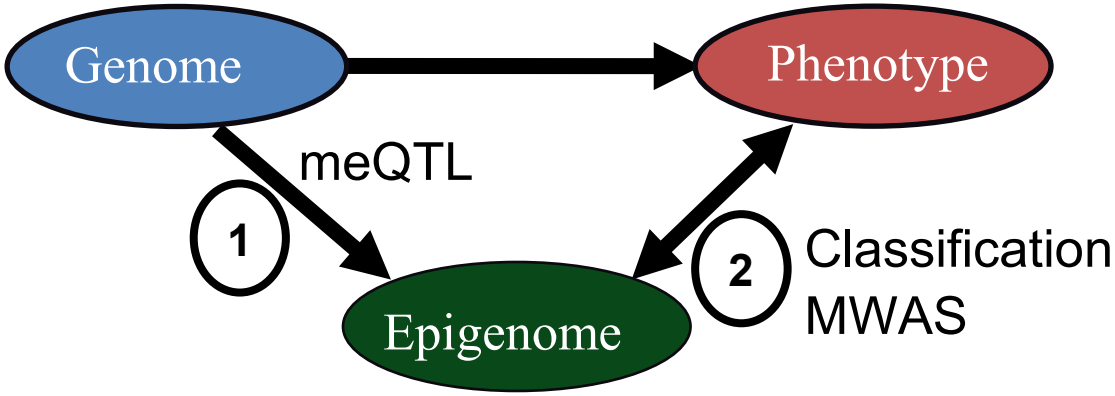
486,000 methylation probes





Brad Bernstein
REMC mapping

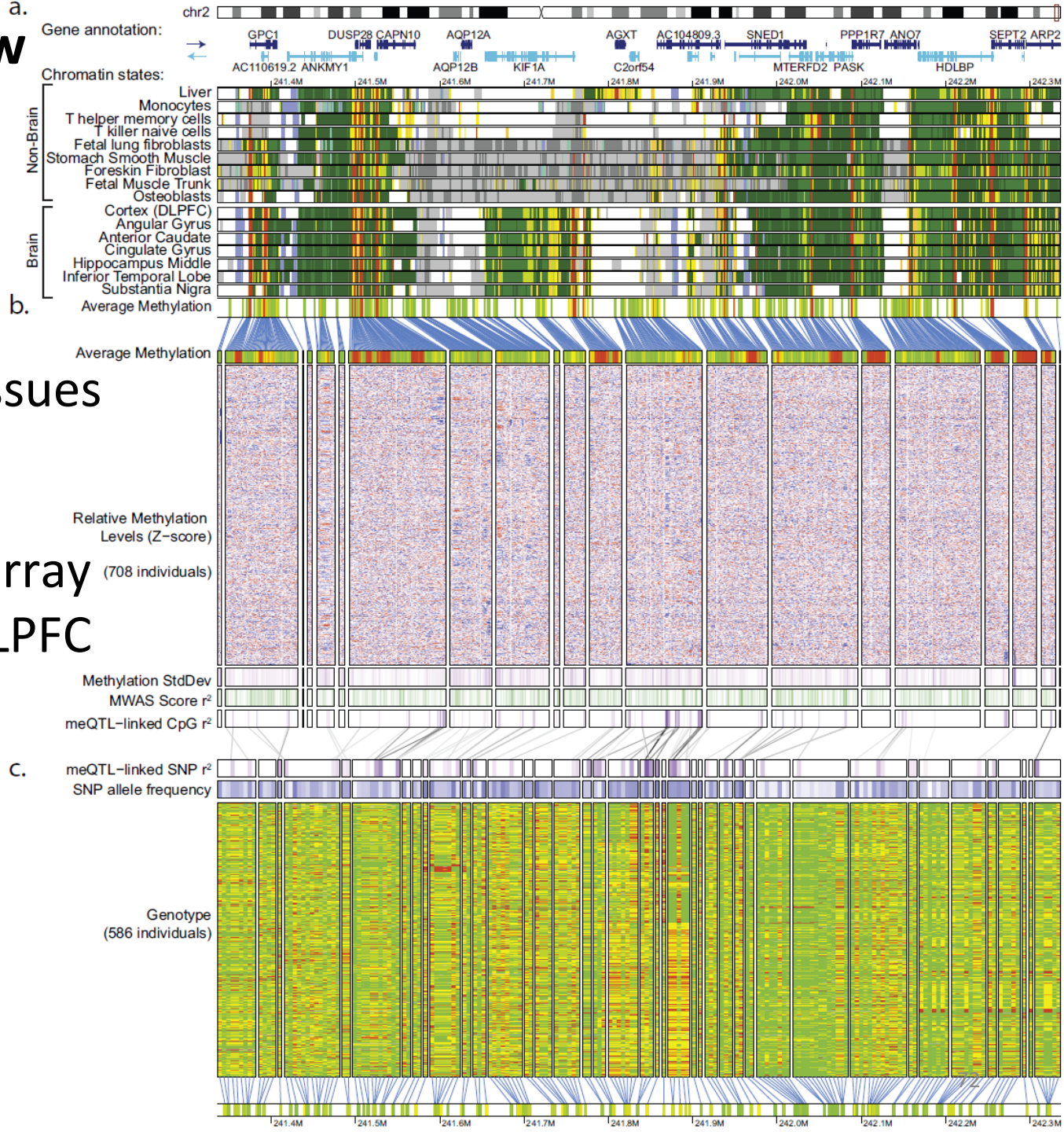
Philip deJager, Epigenomics Roadmap



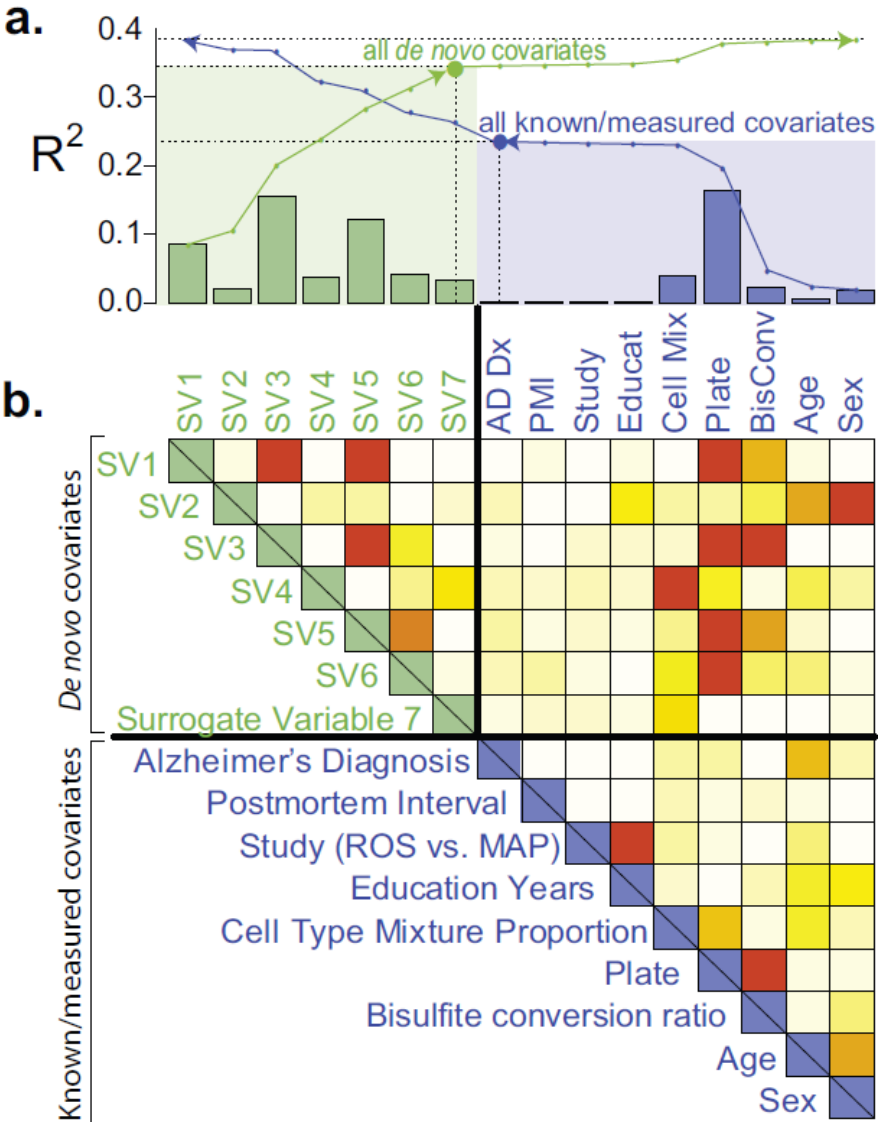
- Patients followed for 10+ years with cognitive evaluations
- Brain samples donated post-mortem methylation/genotype
- Seek predictive features: SNPs, QTLs, mQTLs, regulation

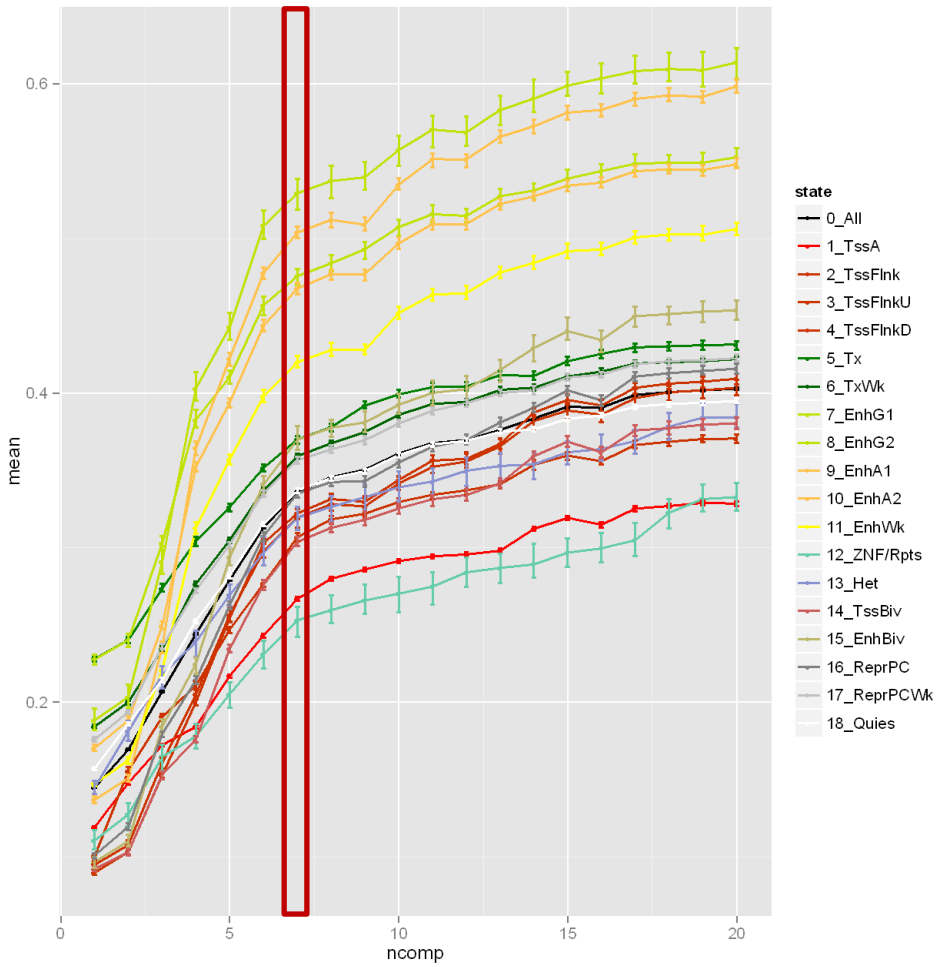
Dataset overview

- Chromatin state
  - 18 states
  - 6 marks
  - DLPFC
  - Joint w/ 127 tissues
- Methylation level
  - 450k Illumina array
  - Brain Cortex DLPFC
  - 708 individuals
- Genotype
  - 620k SNPs
  - 586 individuals
  - Blood



Pre-processing and covariate elimination





- Eliminate 7 *de novo* co-variates, and 8 known co-variates
- Correlate with Plate, Cell Mixture, Conversion, Sex, age

Most methylation probes are high or low, with little variability

a. Chromatin state definitions

b. Distribution of CpG **avg** methylation levels (in Illumina 450k array)

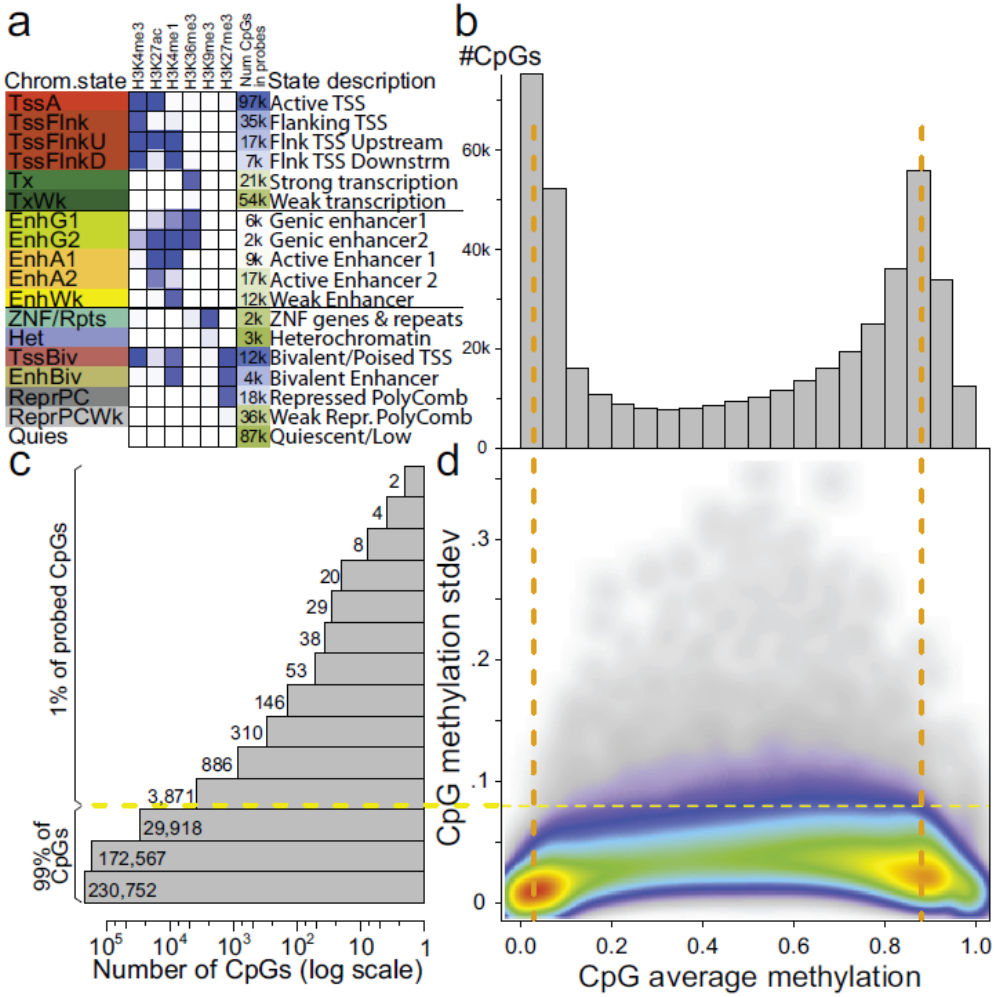
– Average methylation across 708 individuals

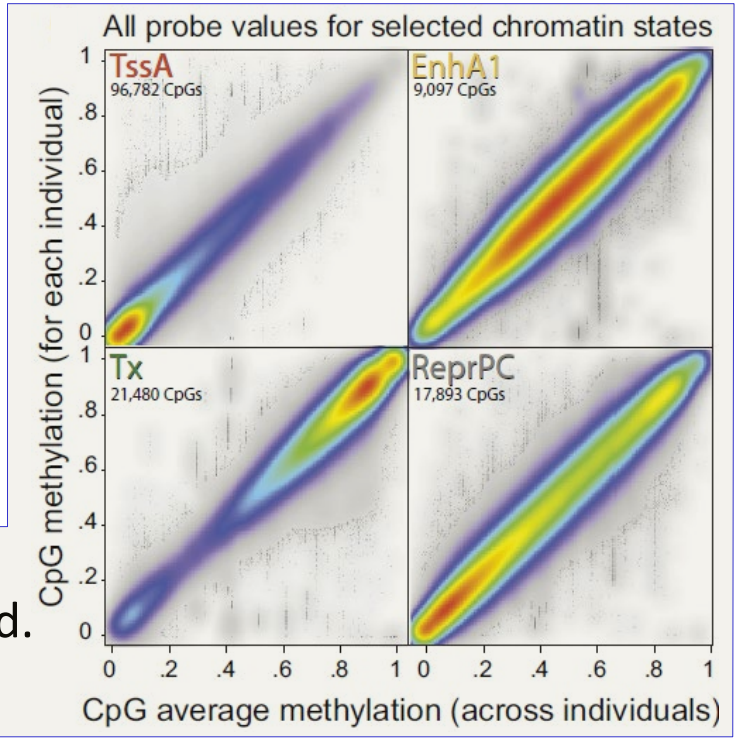
c. Distribution of CpG methylation **variance** across individuals

– Log: **Very few probes** show high variance

d. 2D distribution: average vs. variance

– Highest variance intermediate-methylation

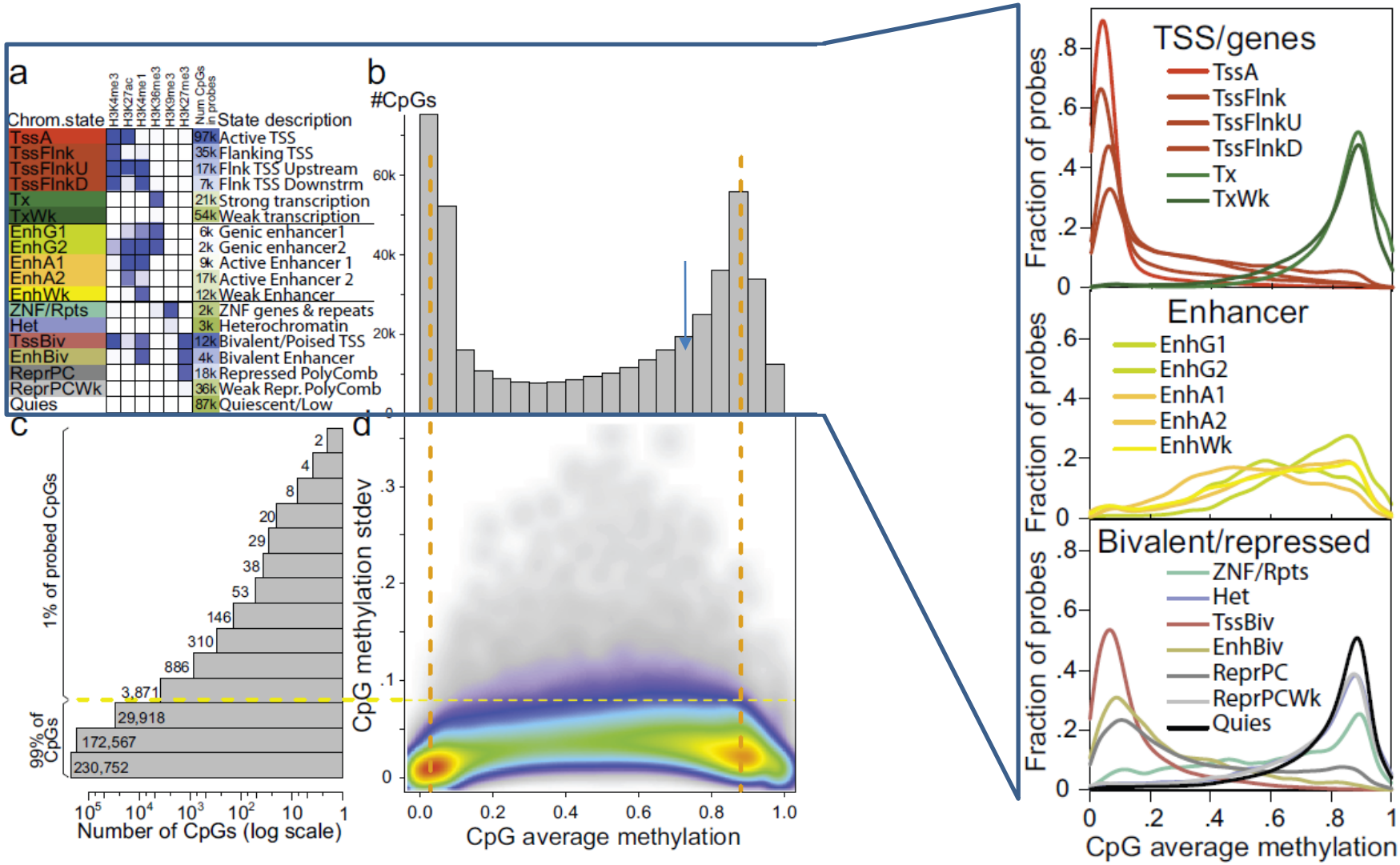




- However: Intermediate methylation is not just an artifact of averaging bimodal levels between individ.

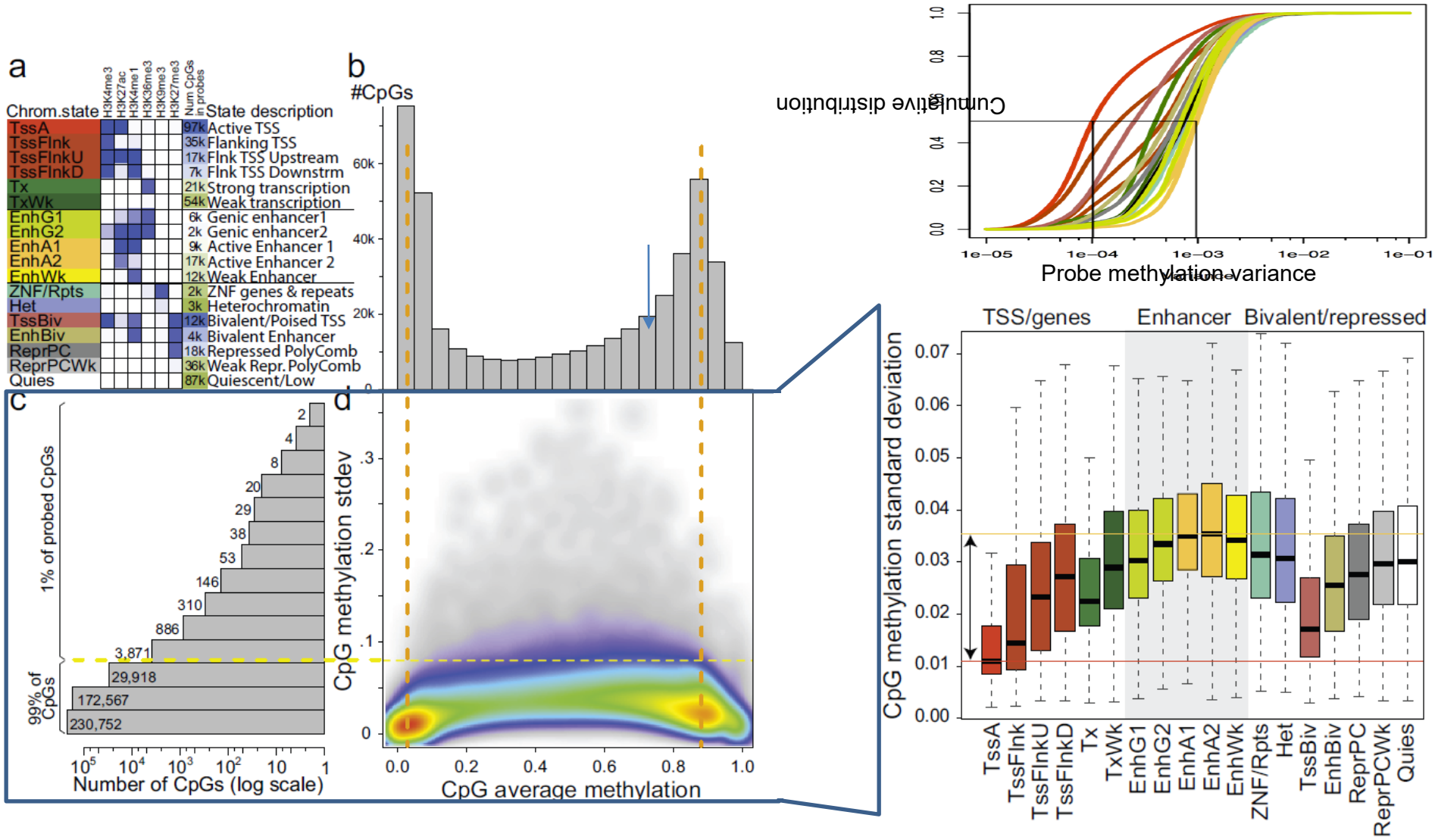
➢ Intermediate methylation is truly intermediate

Enhancer regions show intermediate methylation



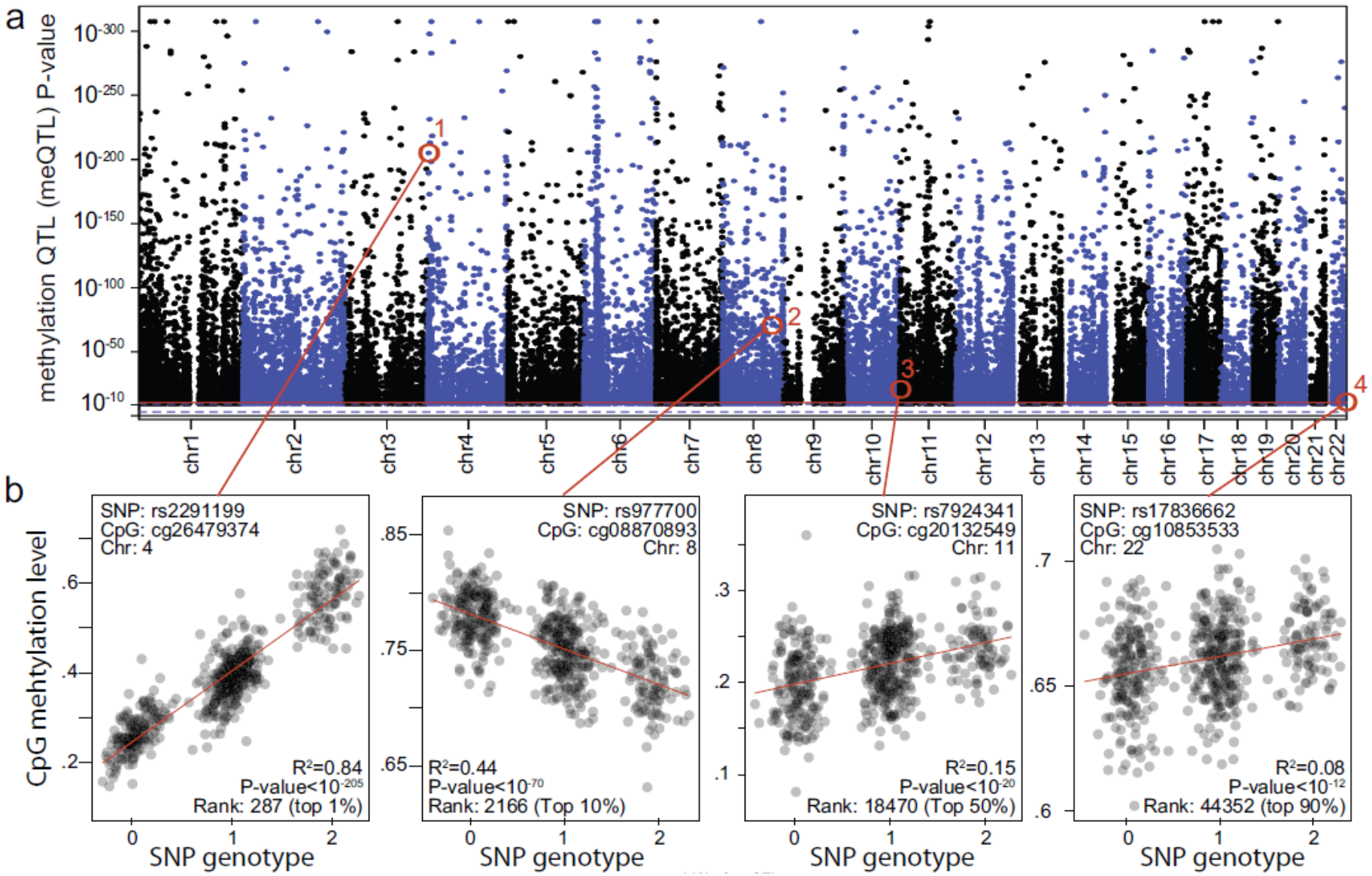
- Enhancer states: Intermediate (EnhG1/G1/A1/A2/Wk)
- Active states: Promoters: low. Tx: high.
- Repressed states: TssBiv/EnhBiv/ReprPC: low. Quies/ReprPCWk: high

Enhancers are most variable, promoters least



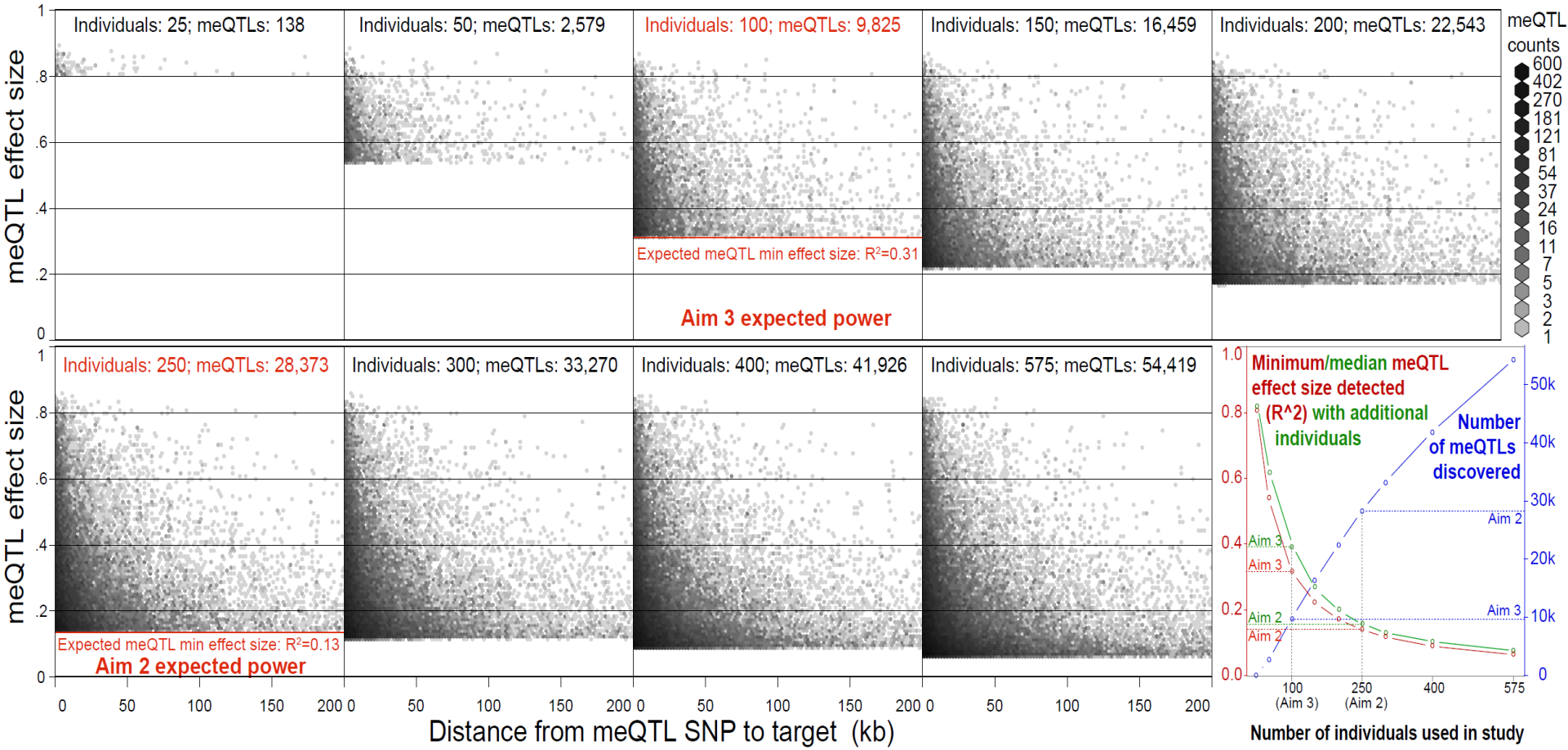
- Chromatin states vary 10-fold in methylation variance, 3-fold in stdev
- Active states: EnhA > EnhWk > EnhG > TxWk > TssFlnk >> TssA
- Repressed states: Quies > ReprPC > EnhBiv >> TssBiv

Discover 50,000 methylation QTLs after Bonferroni



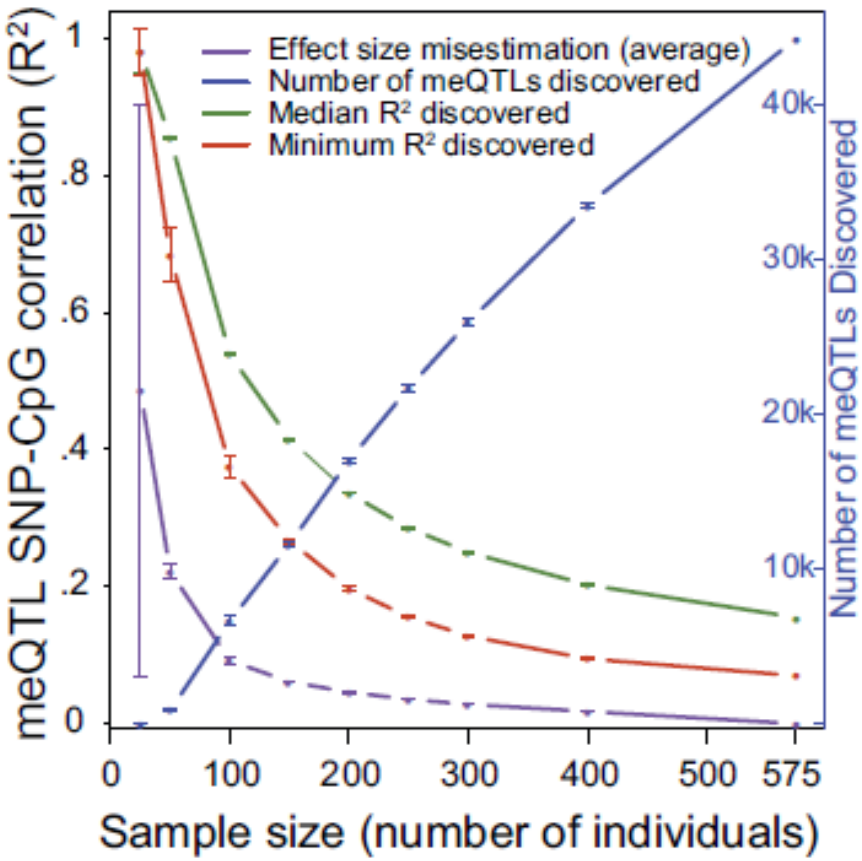
- Overlay meQTL discovery plot

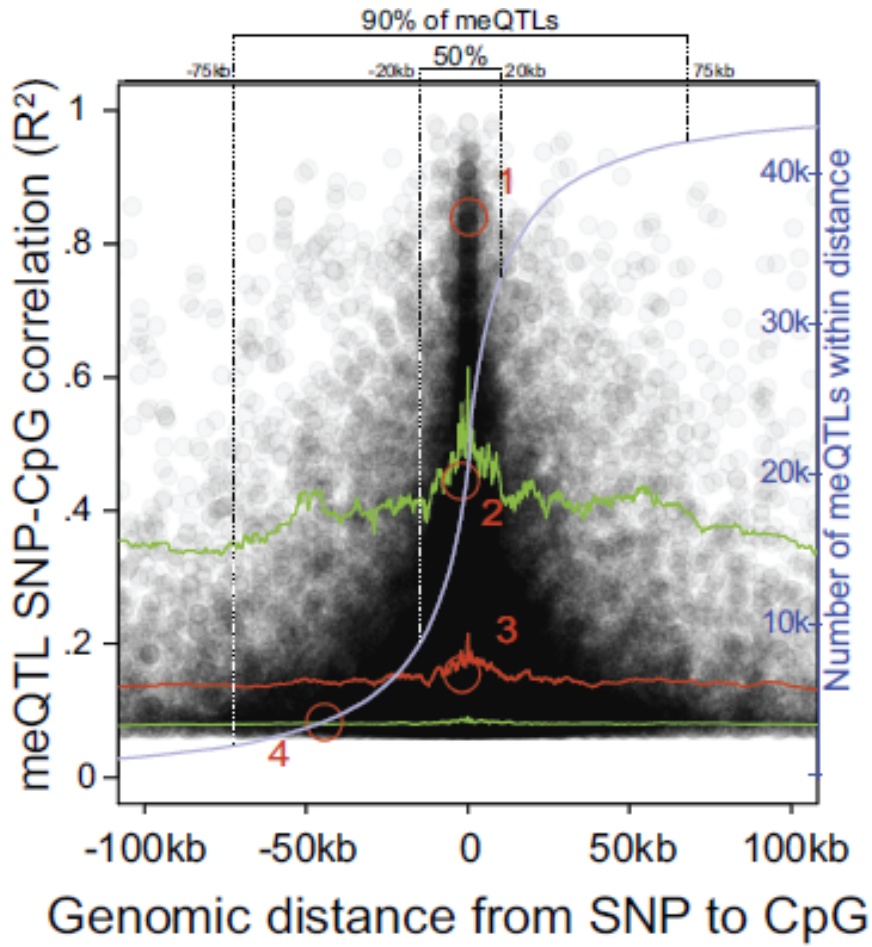
meQTL discovery vs. distance vs. cohort size



- Vary: (1) distance from CpG; (2) effect size; (3) cohort size
- Strongest effects within 20 kb of tested CpGs
- Expectation for 100, 150, 200 individuals (if searching a 1Mb region)

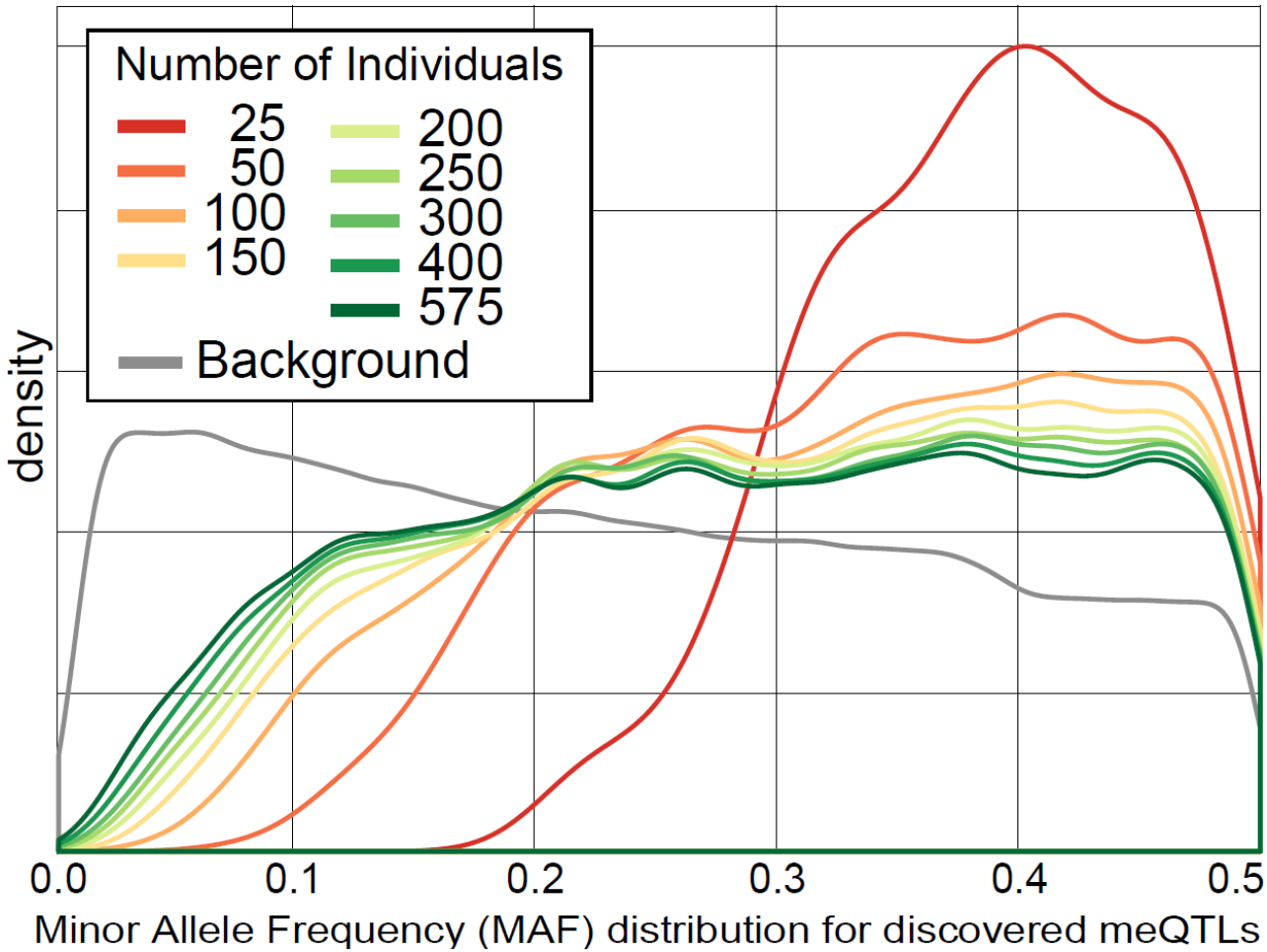
Selection of the number of individuals





- More individuals ➊ linearly more meQTLs, but smaller effect size
- Strongest effects concentrated within 20 kb of tested CpGs
  ➊ can be used to increase power for smaller sample sizes.

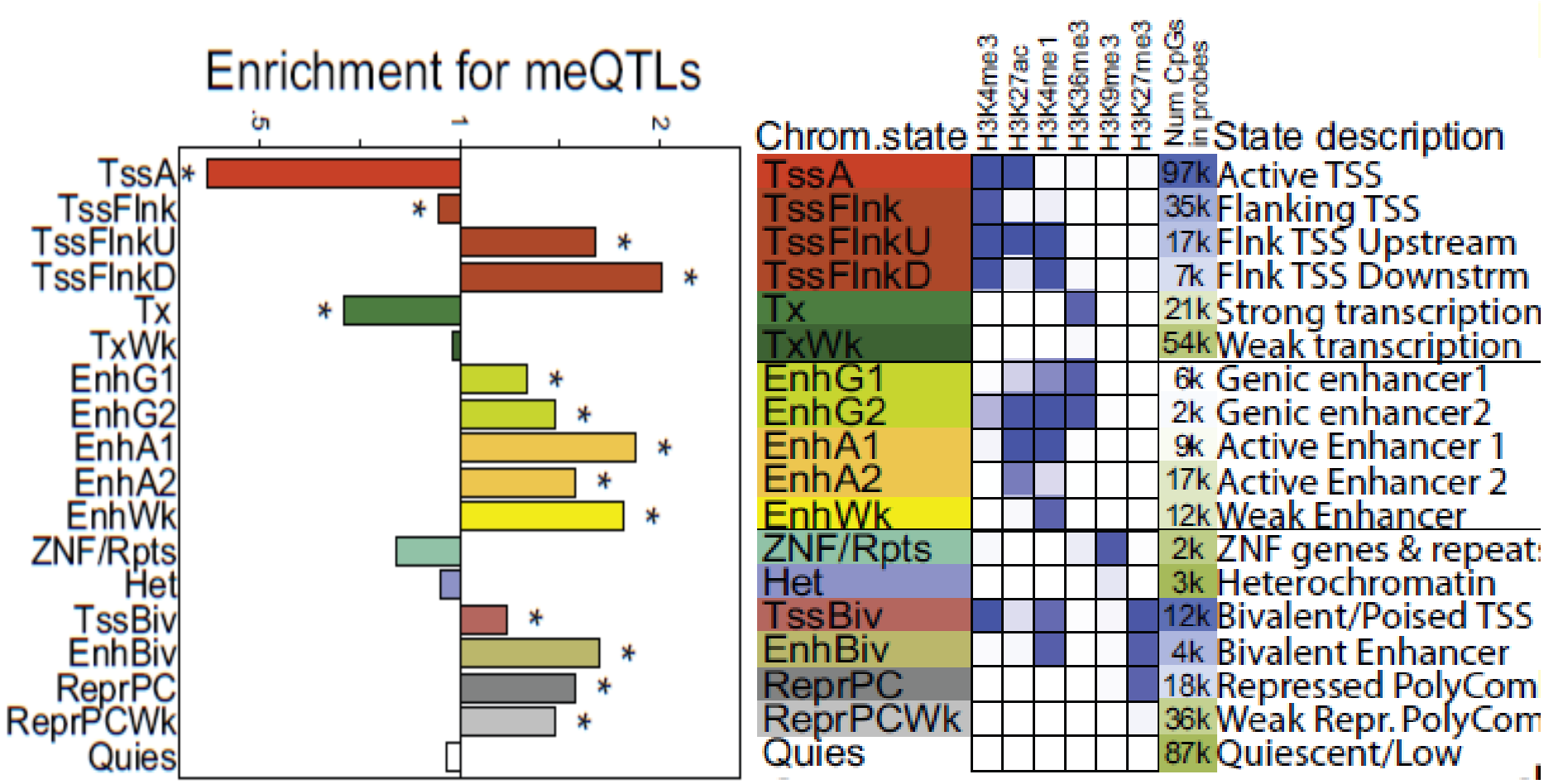
of individuals ⇔ MAF of meQTL SNPs



Minor Allele Frequency (MAF) of discovered meQTL SNPs. Discovery power is greater for high-MAF SNPs, resulting in skewed distributions. Thus, we expect the majority of meQTLs to have both alleles represented in samples of 20 individuals (40 chromosomes). For

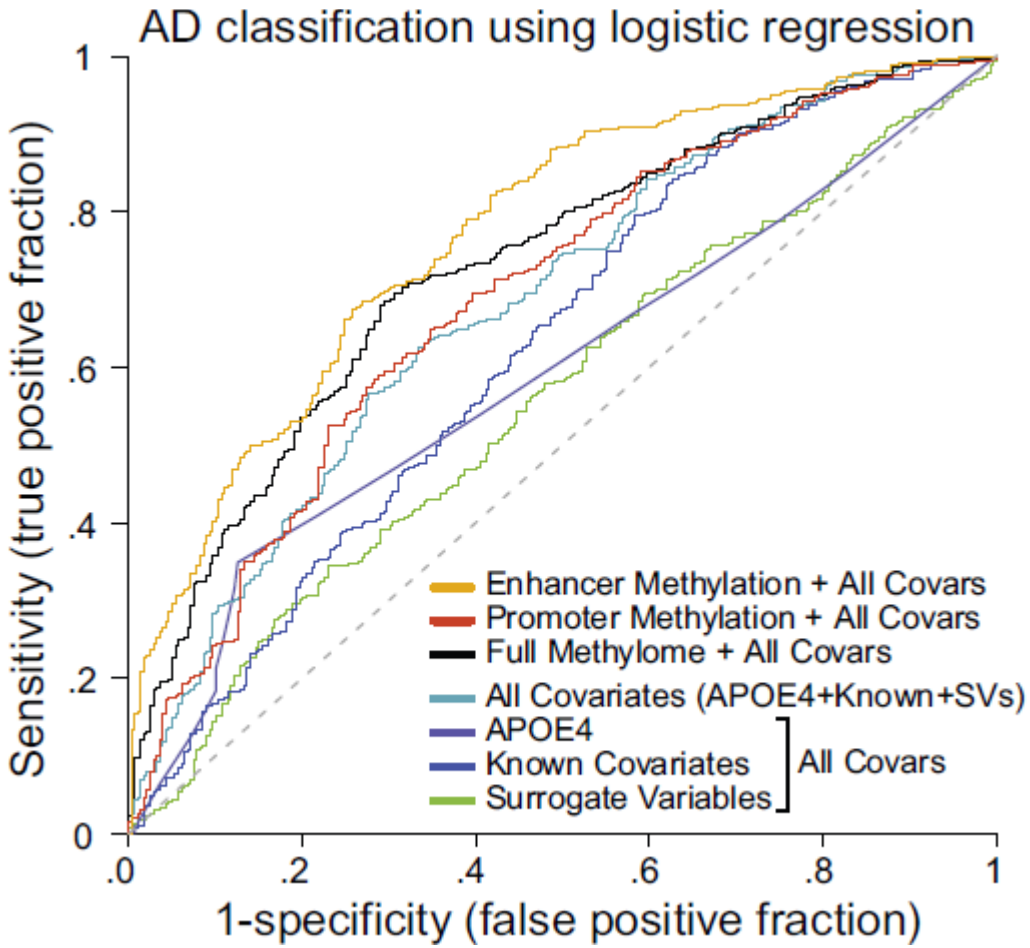
- Focusing on 100-150 individuals, MAF > 0.1, as expected
- Large number of SNPs never probed even with 600 indiv

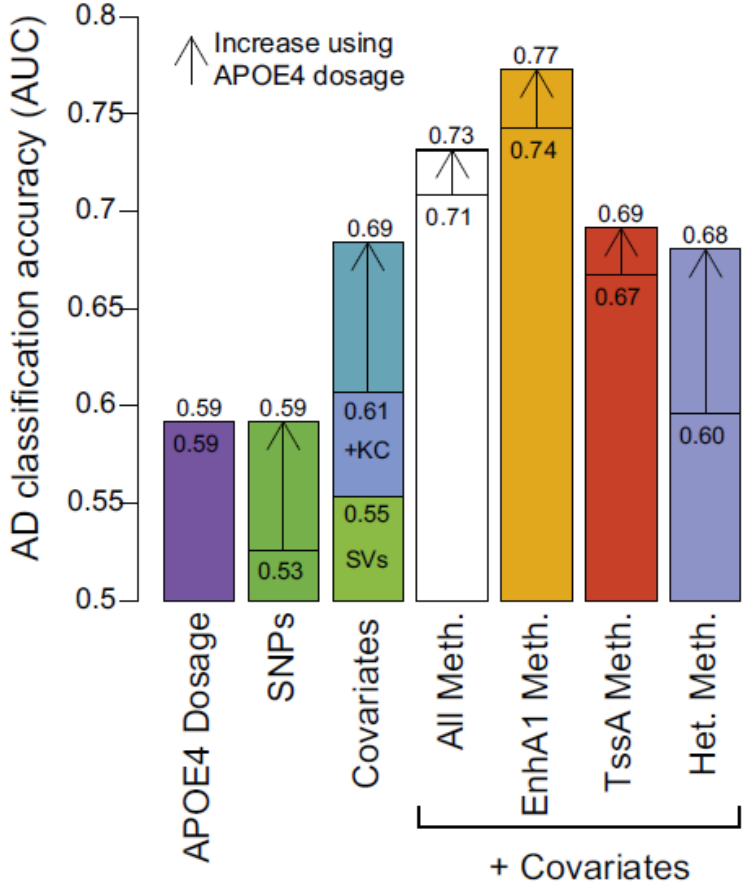
meQTL probes are enriched in enhancers + TssFlnk



- Prioritize EnhA, EnhWk, TssFlnk regions for meQTLs
- Profile variation in H3K27ac directly (ChIP-seq component)

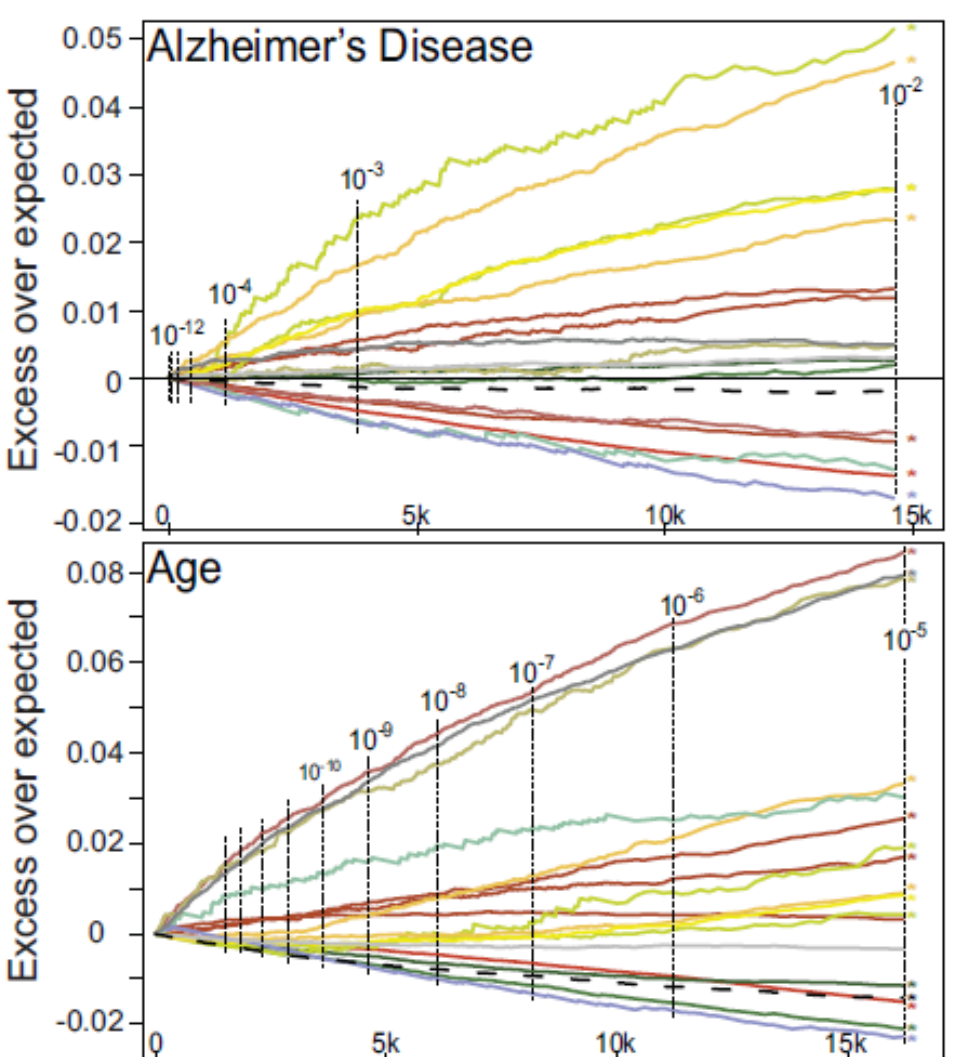
Enhancer variation correlated with AD diagnosis

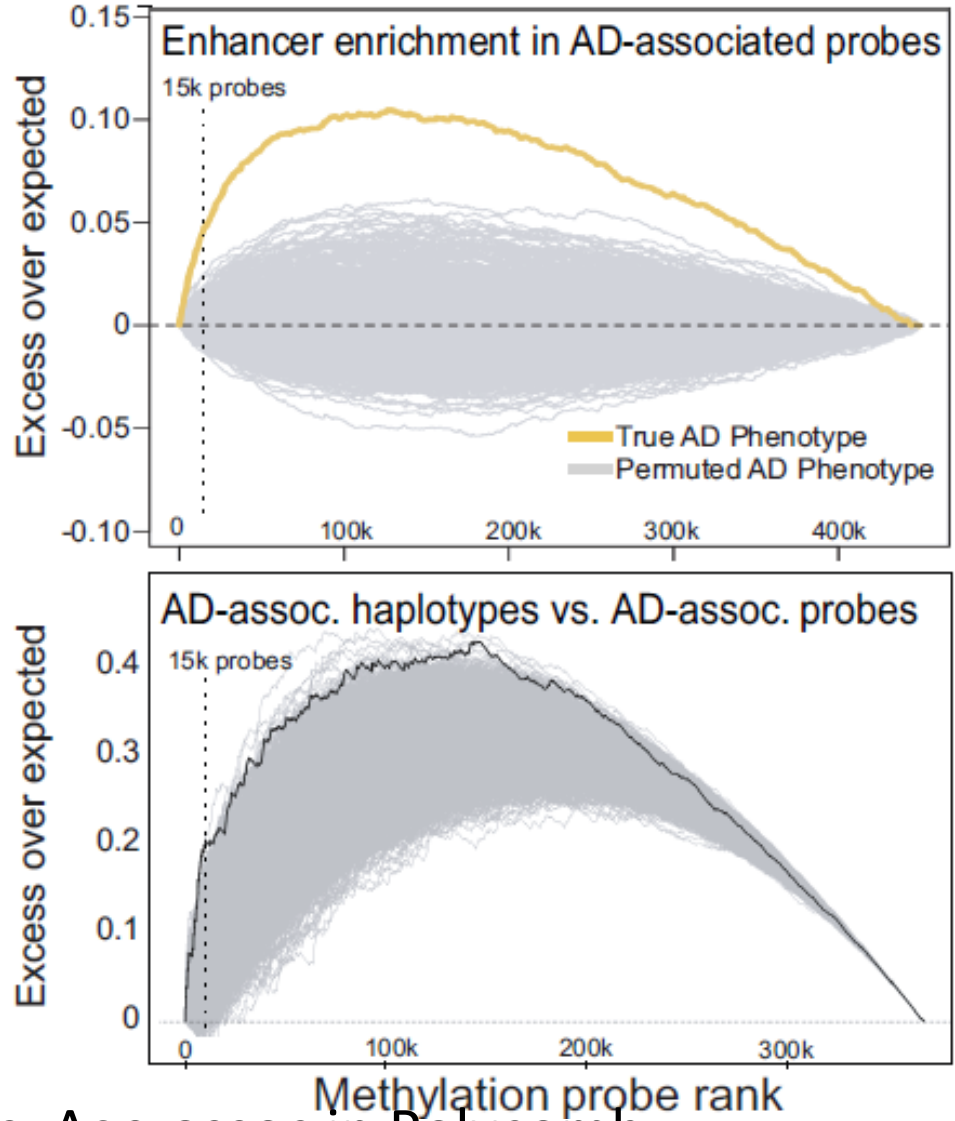




- Enhancer variation is actually biologically meaningful (not just an artifact of meaningless variation)
- Enhancers > all methylation > Promoters > APOE4 >> SNPs

Functional enrichments persist across 1000 probes





- AD-associated probes in enhancers. Age-assoc in Polycomb
- 10,000 phenotype permutations ❾ Statistical significance
- AD top 1k GWAS enrichment persists across 100k+ probes

Imputed MWAS: increased power, genetic component

GWAS:	G	→			D	N=74k
meQTL:	G	→	M			N=800
MWAS:			M	↔	D	N=800
iMWAS:	G	→	iM	→	D	N=74k

Learn G→D directly (complex phenotype)

Learn G→M (simpler phenotype)

M□D (no causality)

Apply G→M to get iM
iM→D (causality)

<u>Key Idea:</u>

- Learn G→M model (ROSMAP n=800) Fewer indiv. Simpler phenotype
- Impute methylation iM for GWAS cohort (n=74k)
- iMWAS between <u>genotype-driven</u> M and AD phenotype (n=47k)

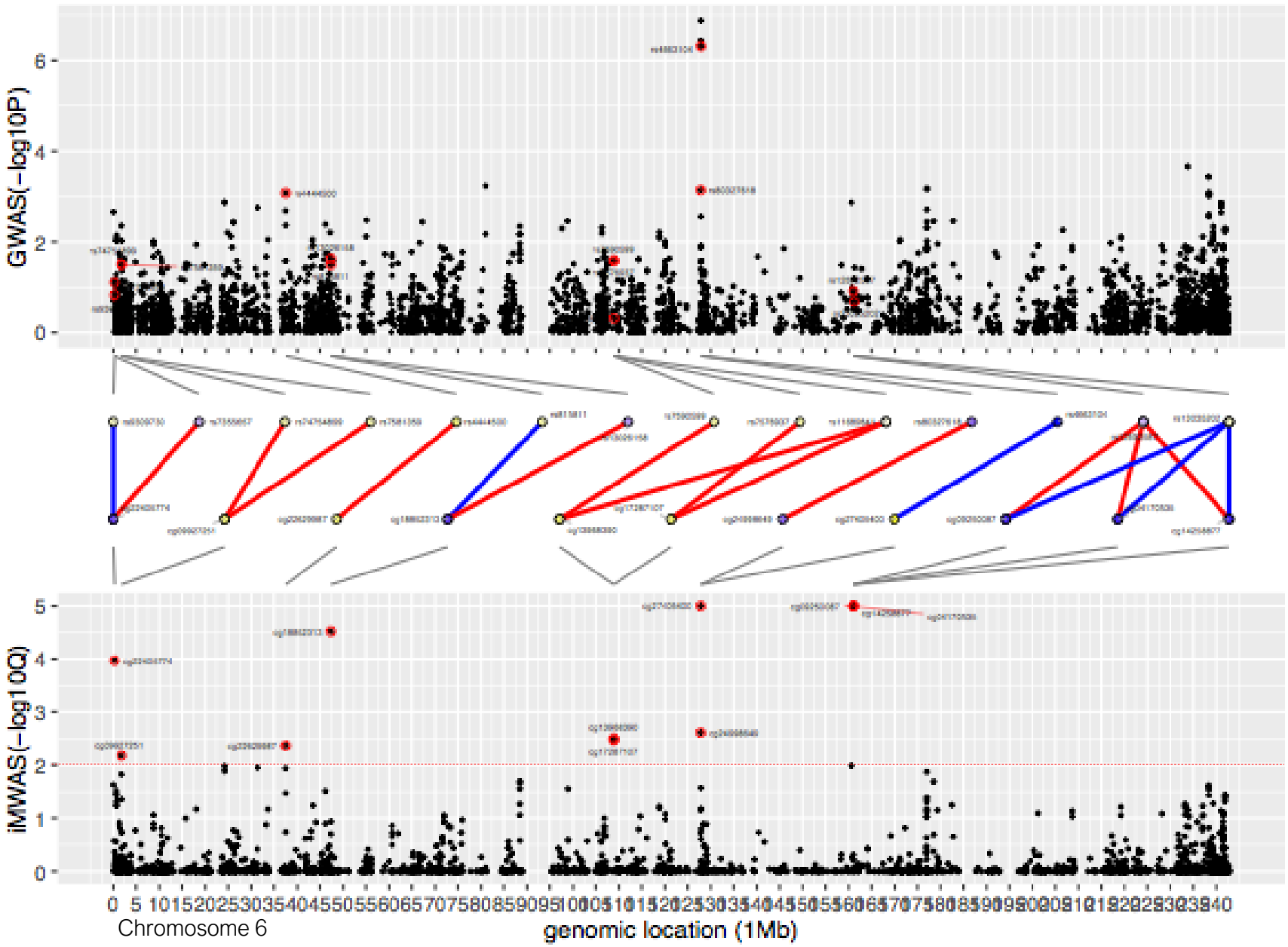
<u>Advantage:</u>

- Much larger GWAS cohorts (>>MWAS): increased power
- Genetic component of methyl. variation

<u>Logistical challenge:</u>

- Summary stats, not full genotypes ❾inear model, impute stats direct

iMWAS results: new loci, multiple contributing SNPs



Chromosome 6

Mendelian Randomization

- Problems with observational data
- Randomized controlled trials
- Mendelian Randomization (MR):
  - How it works
  - Core assumptions
  - Calculating causal effect estimates
- MR example
- Limitations of MR

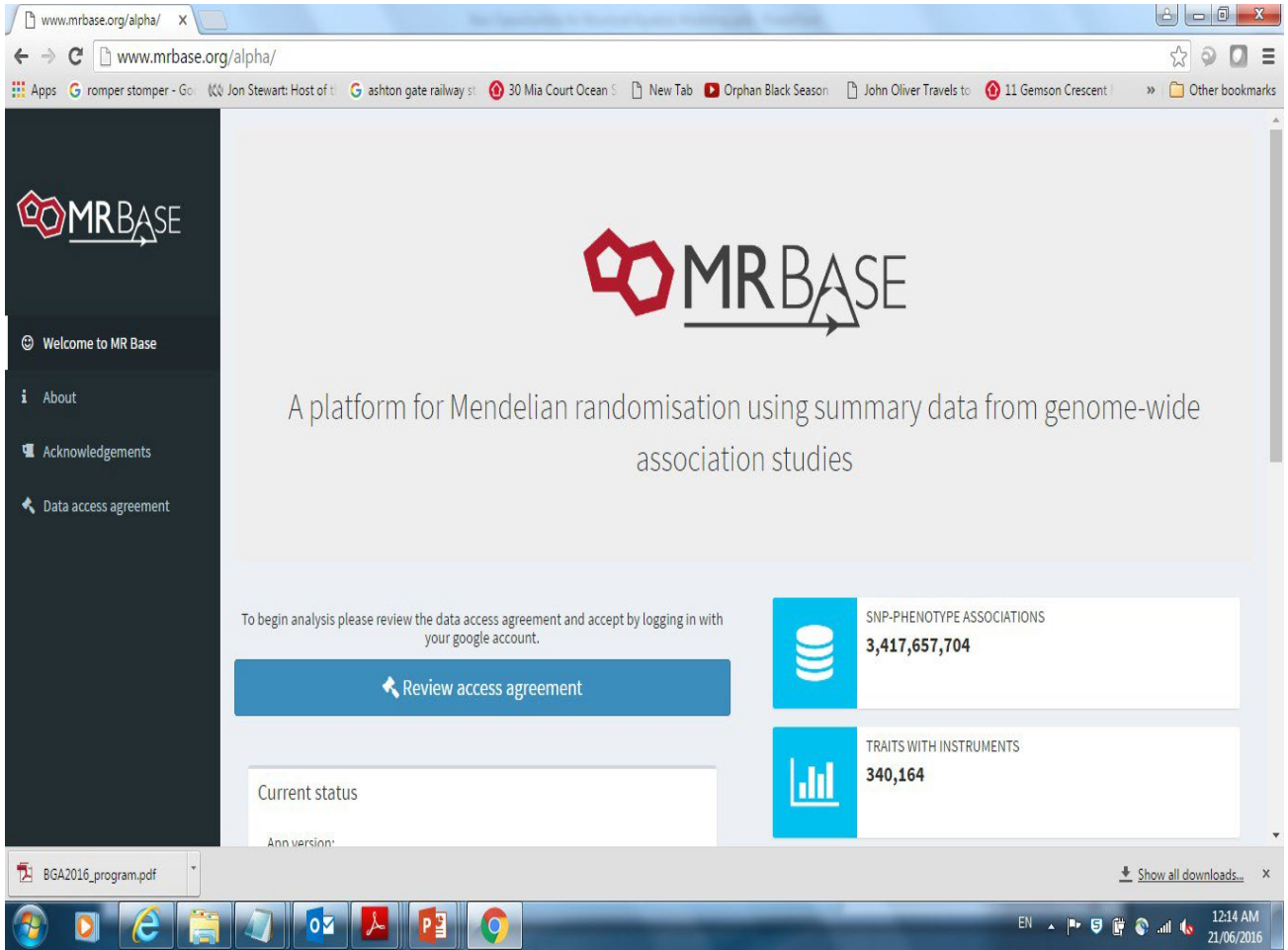
MR Base

Jie “Chris” Zheng

http://www.mrbase.org/

Gib Hemani

Phil Haycock



SingleSampleExtensionV | Inbox (688) - epxde@bri | mRnd: Power calculation | www.mrbase.org/beta/?s

www.mrbase.org/beta/?state=xeFWkjTVJMWm6uuhL8nr&code=4/HxGD3SyDFKXUWcqz97amNkiK08ztsdmGN8-14I2SFd8#

MRBASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

Logged in as
David Evans
epxde@bristol.ac.uk

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
- Quick SNP lookup

Choosing instruments for the exposure

To use two sample MR to estimate the causal effect of an exposure on an outcome, the first step is to identify SNPs that are robustly associated with the exposure. These summary statistics for these SNPs can be taken from a sample from which there is no data on the outcome.

Please provide instruments by choosing from one of the data sources below, or by uploading your own data. You can choose multiple exposures to be analysed, and multiple instruments per exposure.

Choose instruments

Select exposure source

- (●) Manual file upload
- ( ) NHGRI-EBI GWAS catalog
- ( ) MR Base GWAS catalog
- ( ) Gene expression QTLs
- ( ) Protein level QTLs
- ( ) Metabolite level QTLs
- ( ) Methylation level QTLs

Manual file upload

The file must be a plain text file.

To do simple SNP look ups it must have at least one column with the header **SNP** .

To do an MR analysis it must have the following column headers:

- **SNP** - rs IDs of the instruments for the exposure
- **beta** - effect sizes for each SNP
- **se** - standard errors
- **effect_allele** - Effect allele

It's useful to have these columns too:

- **other_allele** - Other allele
- **eaf** - Effect allele frequency

You can see an example file here: telomere_length.txt

Upload plain text file *Preview of uploaded table*

Browse No file s

ben.jpg | underconstruction.png | abstract_v2.docx | APP1137714_Asse....pdf | APP1137714_Sum....pdf | Show all

EN 2:44 PM 6/02/2017

Inbox (572) - ep | www.mrbase.or

www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#

MRBASE

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

Logged in as
David Evans
epxde@bristol.ac.uk

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
- Quick SNP lookup

LD clumping

Most two sample MR methods require that the instruments do not have LD between them.

Linkage disequilibrium

- (●) Do not check for LD between SNPs
- ( ) Use clumping to prune SNPs for LD

LD proxies

If a particular exposure SNP is not present in an outcome dataset, should proxy SNPs be used instead through LD tagging?

- [x] Use proxies?

Minimum LD Rsq value

0.6 — 0.8 — 1

0.6 0.64 0.68 0.72 0.76 0.8 0.84 0.88 0.92 0.96 1

- [x] Allow palindromic SNPs?

MAF threshold for aligning palindromes

0.01 — 0.3 — 0.49

0.01 0.06 0.11 0.16 0.21 0.26 0.31 0.36 0.41 0.46 0.49

Select methods for analysis

Many methods exist for performing two sample MR. Different methods have sensitivities to different potential issues, accommodate different scenarios, and vary in their statistical efficiency.

Choose which methods to use:

- [x] Wald ratio
- [ ] Fixed effects meta analysis (simple SE)
- [ ] Fixed effects meta analysis (delta method)
- [ ] Random effects meta analysis (delta method)
- [ ] Maximum likelihood
- [x] MR Egger
- [ ] MR Egger (bootstrap)
- [x] Weighted median
- [ ] Penalised weighted median
- [x] Inverse variance weighted

Submit

Once you have selected exposures, outcomes, and analysis options you are ready to perform the analysis.

Perform MR analysis

Watanabe2017.pdf | Lawloretal2017.pdf | Lawloretal2017.pdf Failed - Network error | Show all

EN 12:59 AM 11/03/2017

Inbox (572) - epxde@bri | www.mrbase.org/beta/?s

www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#

Run MR

MR Results

Quick SNP lookup

5e-08

Perform clumping

Display columns

- ID
- Trait
- Note
- First author
- Consortium
- Number of cases
- Number of controls
- Sample size
- Number of variants
- Year
- PubmedID
- Access
- Category
- Population
- Priority
- Sd
- Sex
- Subcategory
- Unit

Search: ldl

	ID	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	PubmedID	Access	Category	Po
300	300	LDL cholesterol		Willer CJ	GLGC			173082	2437752	2013	24097068	public	Risk factor	
781	781	LDL cholesterol	Metabo-chip	Willer CJ	GLGC			83198	120251	2013	24097068	public	Risk factor	
880	880	Total cholesterol in large LDL	L.LDL.C	Kettunen				21552	11871461	2016	27005778	public	Metabolites	
881	881	Cholesterol esters in large VLDL	L.LDL.CE	Kettunen				19273	11820655	2016	27005778	public	Metabolites	

Watanabe2017.pdf | Lawloretal2017.pdf | Lawloretal2017.pdf Failed - Network error | Show all

EN 1:10 AM 11/03/2017

FW: Decision on Nature | www.mrbase.org/beta/?s

www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#

- About
- Acknowledgements
- Data access agreement

Logged in as
David Evans
epxde@bristol.ac.uk

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
  - MR Results
- Quick SNP lookup

Display columns

- [ ] ID
- [x] Trait
- [x] Note
- [x] First author
- [x] Consortium
- [x] Number of cases
- [x] Number of controls
- [x] Sample size
- [x] Number of variants
- [x] Year
- [ ] PubmedID
- [ ] Access
- [ ] Category
- [ ] Population
- [ ] Priority
- [ ] Sd
- [ ] Sex
- [x] Subcategory
- [ ] Unit

Show 10 entries

Search: coro

	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	Subcategory
6	Coronary heart disease		Peden	C4D	15420	15062	30482	540233	2011	Cardiovascular
7	Coronary heart disease		Nikpay	CARDIoGRAMplusC4D	60801	123504	184305	9455779	2015	Cardiovascular
8	Coronary heart disease		Schunkert H	CARDIoGRAM	22233	64762	86995	2420361	2011	Cardiovascular
9	Coronary heart disease		Deloukas	CARDIoGRAMplusC4D	63746	130681	194427	79129	2013	Cardiovascular
	All	All	All	All	All	All	All	All	All	All

Showing 1 to 4 of 4 entries (filtered from 1,033 total entries)

Previous 1 Next

Watanabe2017.pdf | Lawloretal2017.pdf | Lawloretal2017.pdf Failed - Network error | Show all

EN 1:11 AM 11/03/2017

FW: Decision on Nature | www.mrbase.org/beta/?s

www.mrbase.org/beta/?state=3DNYEn5uYpGOFV1AR1x8&code=4/c_PBiWWgLEId0E5VXsQP0vhJ0ExNeHxmjpZeUz4u4vI#

- Welcome to MR Base
- About
- Acknowledgements
- Data access agreement

Logged in as
David Evans
epxde@bristol.ac.uk

- Perform MR analysis
  - Choose exposures
  - Choose outcomes
  - Run MR
  - MR Results
- Quick SNP lookup

Linkage disequilibrium

- (●) Do not check for LD between SNPs
- ( ) Use clumping to prune SNPs for LD

LD proxies

If a particular exposure SNP is not present in an outcome dataset, should proxy SNPs be used instead through LD tagging?

- [ ] Use proxies?

Allele harmonisation

An important step in two sample MR is making sure that the effects of the SNPs on the exposure correspond to the same allele as their effects on the outcome. This is potentially difficult with palindromic SNPs.

Handling reference alleles

- (●) All effect alleles are definitely on the positive strand
- ( ) Attempt to align strands for palindromic SNPs
- ( ) Exclude palindromic SNPs

potential issues, accommodate different scenarios, and vary in their statistical efficiency.

Choose which methods to use:

- [x] Wald ratio
- [ ] Fixed effects meta analysis (simple SE)
- [ ] Fixed effects meta analysis (delta method)
- [ ] Random effects meta analysis (delta method)
- [ ] Maximum likelihood
- [x] MR Egger
- [ ] MR Egger (bootstrap)
- [x] Weighted median
- [ ] Penalised weighted median
- [x] Inverse variance weighted

Perform MR analysis

Watanabe2017.pdf | Lawloretal2017.pdf | Lawloretal2017.pdf Failed - Network error | Show all

EN 1:12 AM 11/03/2017

Useful References

- Brion et al (2013). Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol*, 42(5), 1497-501.
- Davey-Smith & Hemani (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*, 23(1), R89-98.
- Davey-Smith & Ebrahim (2003). "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? IJE, 32, 1-22.
- Davies et al (2018). Reading Mendelian randomization studies: a guide, glossary, and checklist for clinicians. BMJ, Jul 12, 362:k601.
- Evans & Davey-Smith (2015). Mendelian randomization: New applications in the coming age of hypothesis free causality. Annu Rev Genomics Hum Genet, 16, 327-50.
- Hemani et al. (2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife, May 30, 7, e34408.
- Zheng et al. (2017). Recent developments in Mendelian randomization studies. Curr Epidemiol Rep, 4(4), 330-345.