

Mendelian Randomization Study and Power

Sander W. van der Laan | s.w.vanderlaan-2@umcutrecht.nl | @swvanderlaan

Cystatin C and cardiovascular disease: an example Mendelian randomization study

1. What you should know before we start
2. So why do a Mendelian Randomization?
3. What is cystatin C and why are we interested anyway?
4. Who was a friar and what does he have to do with cystatin C?
5. What is the chicken and what is the egg?





UMC Utrecht

3 THINGS YOU SHOULD KNOW BEFORE WE START



We are all equally unique

~100 million genetic variations are known to date

among ~3 billion base pairs in the human genome



These genetic variants, i.e. single-nucleotide polymorphisms (SNPs), are correlated...

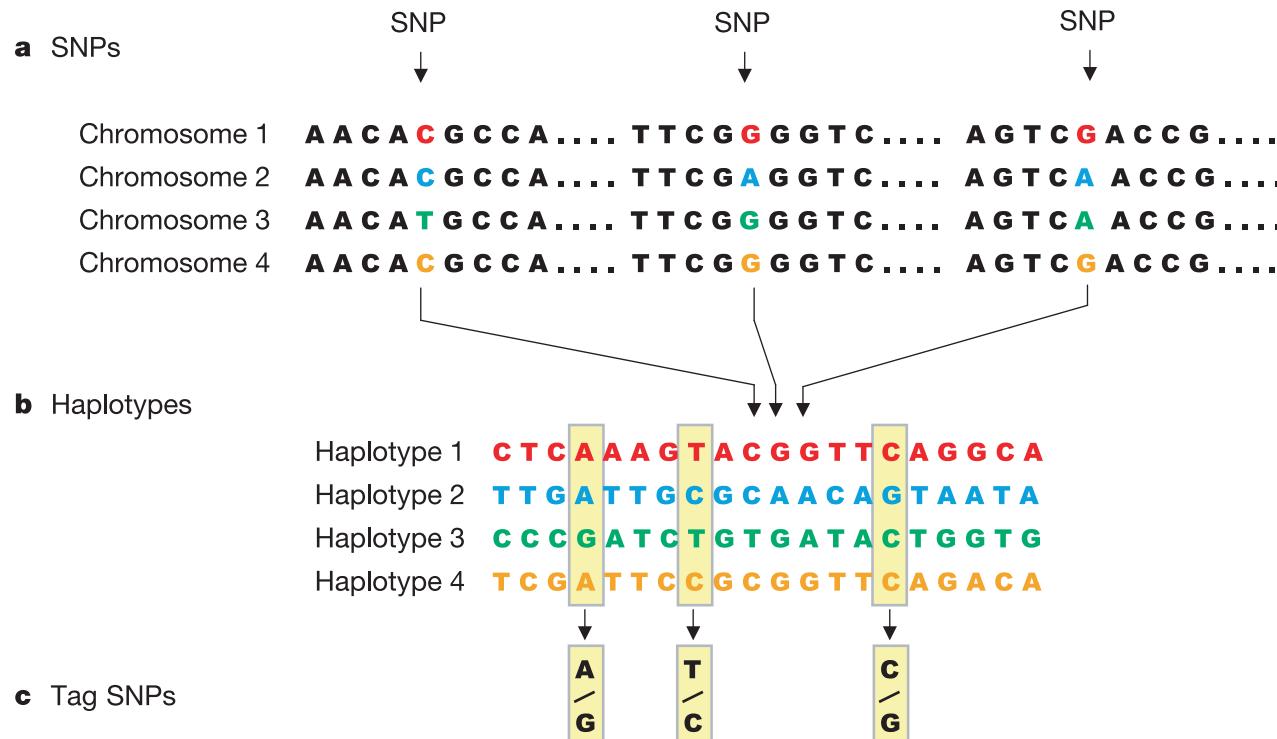


The International HapMap Project. *Nature*; 426:789-796; 2003

Pe'er I. *Genet Epidemiol*; 32(4):381-385; 2008

Dudbridge F. *Genet Epidemiol*; 32(3):227-234; 2008

...and form *haplotypes*...



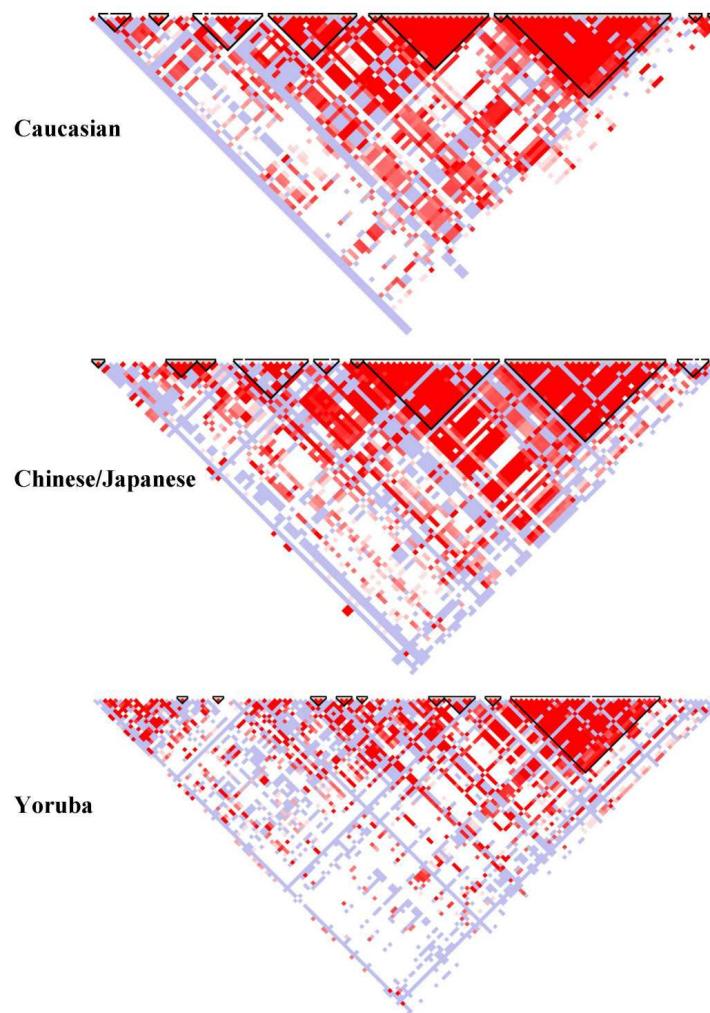
~ 1 million uncorrelated variants – thus $p < 5 \times 10^{-8}$

The International HapMap Project. *Nature*; 426:789-796; 2003

Pe'er I. *Genet Epidemiol*; 32(4):381-385; 2008

Dudbridge F. *Genet Epidemiol*; 32(3):227-234; 2008

...that differ between ancestral populations...

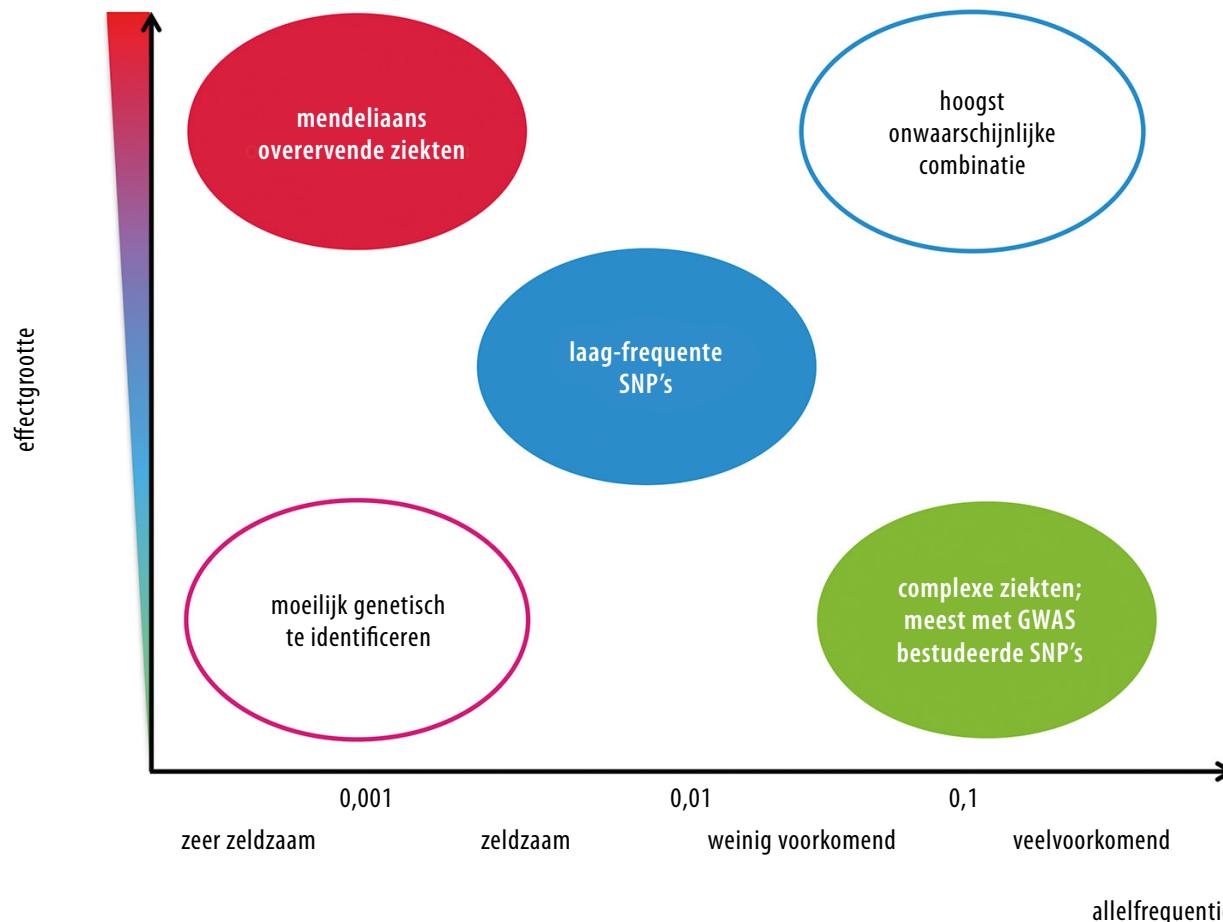


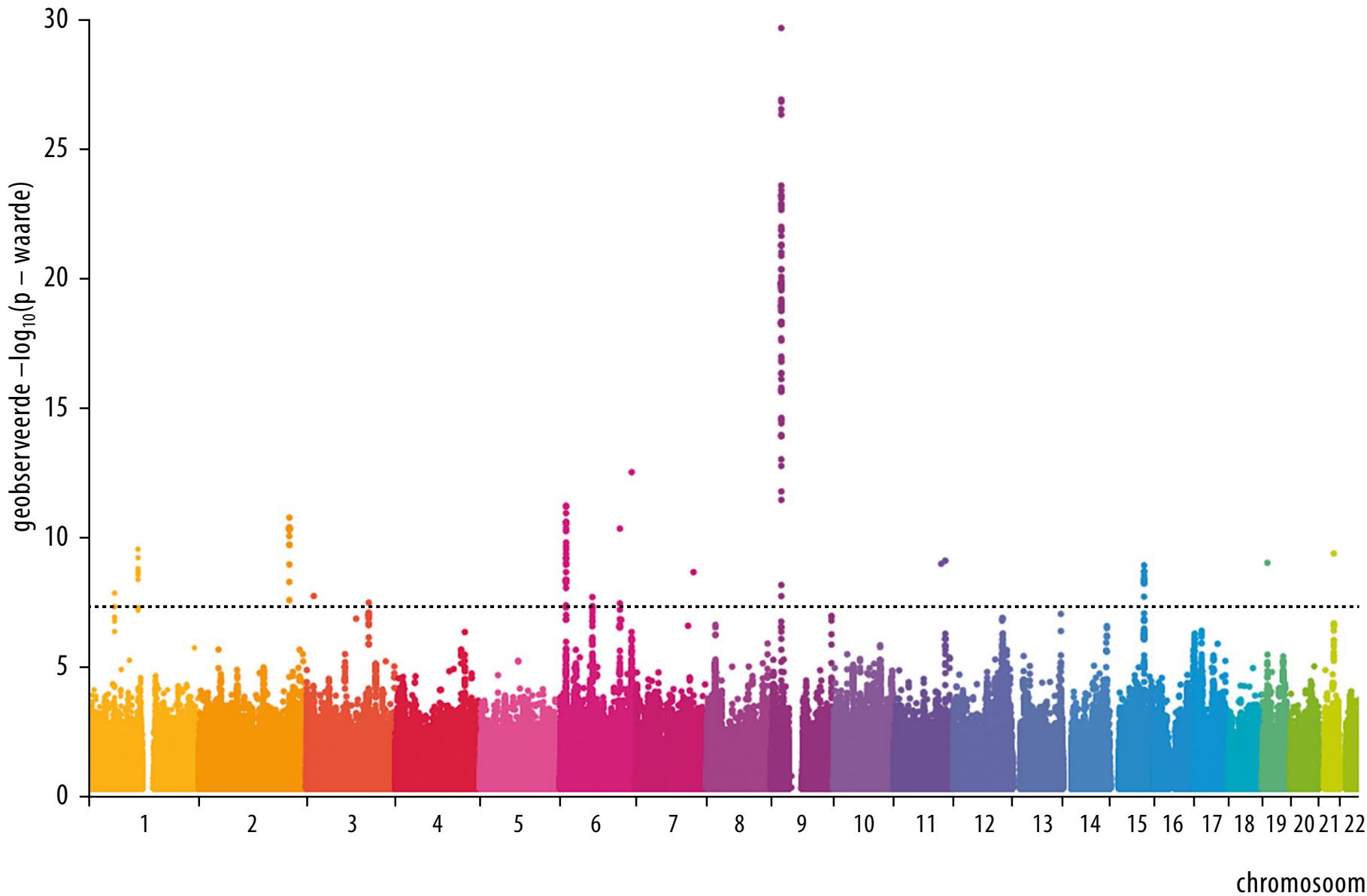
The International HapMap Project. *Nature*; 426:789-796; 2003

Pe'er I. *Genet Epidemiol*; 32(4):381-385; 2008

Dudbridge F. *Genet Epidemiol*; 32(3):227-234; 2008

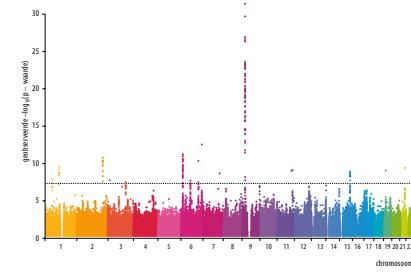
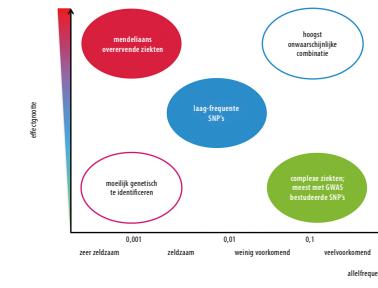
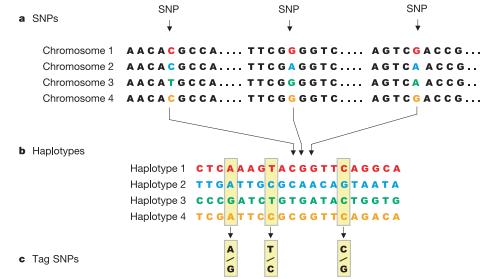
Common diseases and traits are complex and polygenic by nature many SNPs are involved with small effects





So now you know 3 things

1. We are all equally unique: ~100 million genetic variations are known to date, among ~3 billion base pairs in the human genome
2. These genetic variants, i.e. single-nucleotide polymorphisms (SNPs), are correlated and form *haplotypes* that differ between ancestral populations
3. Common diseases and traits are complex and polygenic by nature: many SNPs are involved with small effects





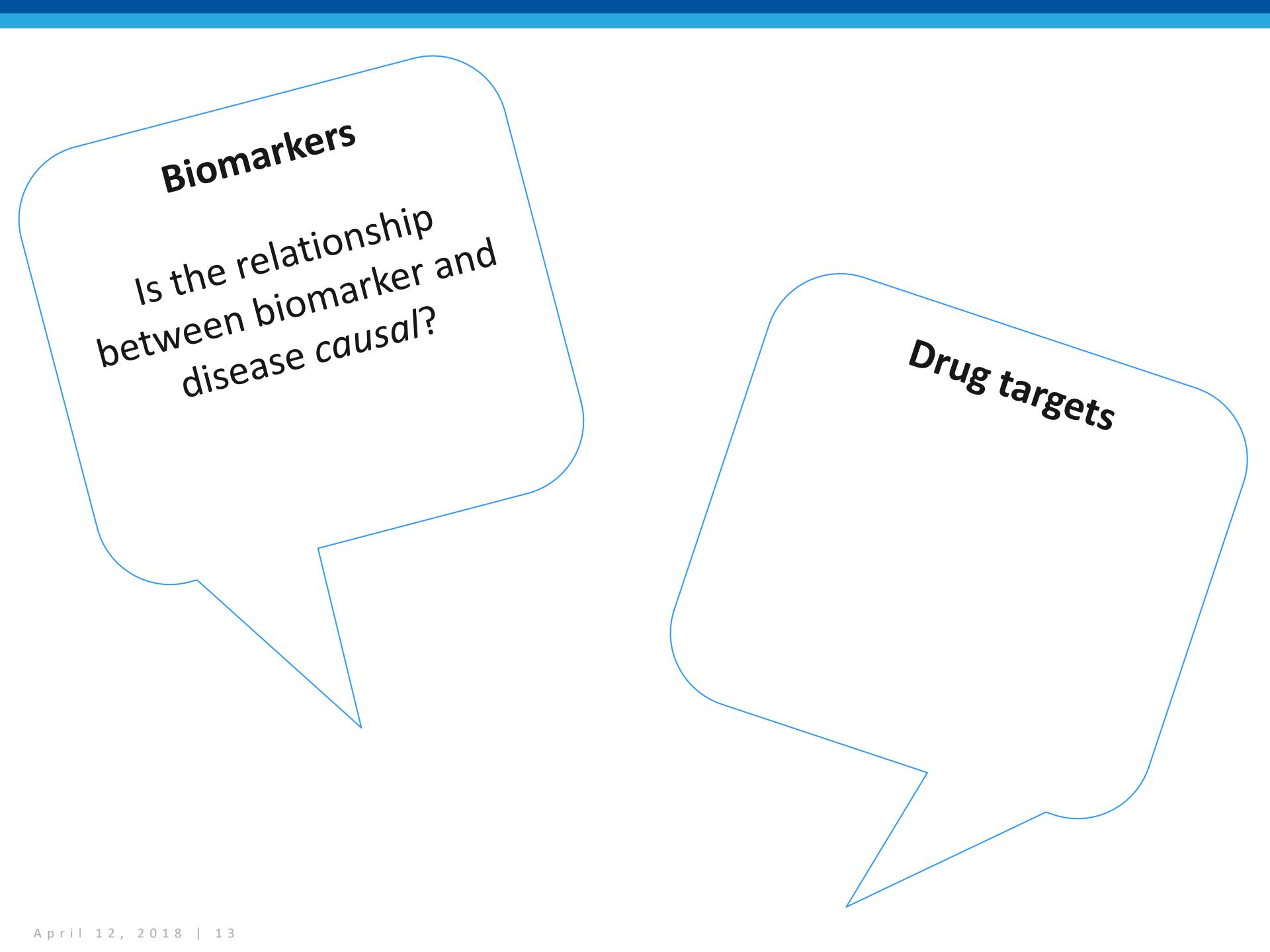
UMC Utrecht

WHY DO A MENDELIAN RANDOMIZATION STUDY?



Biomarkers

Drug targets



Biomarkers

Is the relationship
between biomarker and
disease causal?

Drug targets

Biomarkers

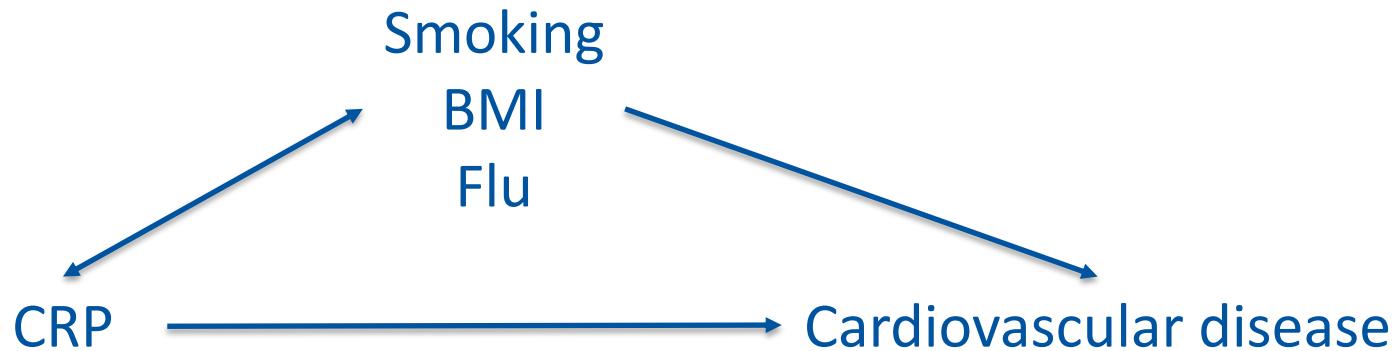
Is the relationship
between biomarker and
disease causal?

Drug targets

What are the
consequences for
biomarkers or risk
factors when you
specifically modulate a
target?

Observational studies are inherently problematic

- Many complex diseases (and traits) are by **confounders**



- Reverse causation further complicates the interpretation





UMC Utrecht

THE CONCEPT OF MENDELIAN RANDOMIZATION



APOLIPOPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR,—It is unclear whether the relation between low serum cholesterol levels and cancer¹ is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study cohorts with widely different diets and corresponding differences in mean cholesterol levels experienced similar mean cancer rates.^{2,3} On the other hand, within each region cancer incidence was higher in men with a serum cholesterol in the lowest part of the cholesterol distribution for that country.³ Thus, naturally low cholesterol levels are sometimes associated with increased cancer risk.^{1,3}

Differences in the aminoacid sequence of apolipoprotein E (apo E) are major determinants of differences in plasma cholesterol levels within a population. Apo E has a key role in the clearance of cholesterol from plasma.⁴ The synthesis of apo E is under the control of three independent alleles, located at a single gene locus, coding for the major isoforms E-2, E-3, and E-4 with respective population frequencies of about 8, 77, and 15%.⁵ The homozygous E-3/E-3 is the most common phenotype encountered and E-2/E-2 is the least common. From apo E-2 to apo E-3, one cysteine residue is replaced by arginine, and from apo E-3 to apo E-4 another cysteine residue is replaced. As a result the avidity of apo E containing lipoproteins for lipoprotein receptors increases from apo E-2 to apo E-3 to apo E-4. In several populations,⁶⁻⁸ including the Finns and the Japanese (Dr G. Utermann, personal communication), the gradient in serum cholesterol levels in the population is associated with a gradient in apo E phenotype, E-2 being associated with lower serum low-density lipoprotein and total cholesterol levels than E-3 and E-4. Thus, if a naturally low cholesterol favours tumour

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

Unlike most other indices of lipid metabolism, apolipoprotein aminoacid sequences are not disturbed by disease, and the apo E phenotype found in a patient will have been present since birth. A comparison of apo E phenotypes in cancer patients with those in matched controls might thus shed light on the relation between low cholesterol and cancer. If it is causal then the E-2 allele should be more common among patients and E-3 and E-4 more common among controls. On the other hand, equal distribution of apo E phenotypes among cases and controls would suggest that the association between low cholesterol and cancer is spurious. Measurement of apo E phenotype by isoelectric focusing of plasma is a routine determination in lipid laboratories; epidemiologists interested in cholesterol and cancer should include it in their studies.

Department of Human Nutrition,
Agricultural University,
6703 BC Wageningen, Netherlands

MARTIJN B. KATAN

1. McMichael AJ, Jensen OM, Parkin DM, Zaridze DG. Dietary and endogenous cholesterol and human cancer. *Epidemiol Rev* 1984; **6**: 192-216.
2. Keys A, Aravanis C, Blackburn H, et al. Serum cholesterol and cancer mortality in the seven countries study. *Am J Epidemiol* 1985; **121**: 870-83.
3. Katan MB. Effects of cholesterol-lowering diets on the risk for cancer and other non-cardiovascular diseases. In: Nestel PJ, et al, eds. Atherosclerosis VII: Proceedings of the Seventh International Atherosclerosis Symposium. Amsterdam: Elsevier, 1986.
4. Brown MS, Kovanen PT, Goldstein JL. Regulation of plasma cholesterol by lipoprotein receptors. *Science* 1981; **212**: 628-35.
5. Utermann G, Steinmetz A, Weber W. Genetic control of human apolipoprotein E polymorphism: comparison of one- and two-dimensional techniques of isoprotein analysis. *Hum Genet* 1982; **60**: 344-51.
6. Utermann G, Kindermann I, Kaffarnik H, Steinmetz A. Apolipoprotein E phenotypes and hyperlipidemia. *Hum Genet* 1984; **65**: 232-36.
7. Robertson FW, Cumming AM. Effects of Apoprotein E Polymorphism on Serum Lipoprotein Concentration. *Arteriosclerosis* 1985; **5**: 283-92.
8. Utermann G. Genetic polymorphism of apolipoprotein E: impact on plasma lipoprotein metabolism. In: Crepaldi G, et al, eds. Diabetes, obesity and hyperlipidemias III. Amsterdam: Elsevier, 1985: 1-28.



APOLIPOPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR.—It is unclear whether the relation between low serum cholesterol levels and cancer¹ is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study cohorts with widely different diets and corresponding differences in mean cholesterol levels experienced similar mean cancer rates.^{2,3} On the other hand, within each region cancer incidence is higher among patients with naturally low cholesterol levels than among those with higher cholesterol levels.⁴ Thus, if a naturally low cholesterol favours tumour growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

Differences in the aminoacid sequence of apolipoprotein E (apo E) are major determinants of differences in plasma cholesterol levels within a population. Apo E has a key role in the clearance of cholesterol from

control of three genes coding for the protein. The apo E-2 allele is the most common in European populations, while the apo E-3 allele is the least common. From apo E-2 to apo E-3, one cysteine residue is replaced by arginine, and from apo E-3 to apo E-4 another cysteine residue is replaced. As a result the avidity of apo E containing lipoproteins for lipoprotein receptors increases from apo E-2 to apo E-3 to apo E-4. In several populations,⁶⁻⁸ including the Finns and the Japanese (Dr G. Utermann, personal communication), the gradient in serum cholesterol levels in the population is associated with a gradient in apo E phenotype, E-2 being associated with lower serum low-density lipoprotein and total cholesterol levels than E-3 and E-4. Thus, if a naturally low cholesterol favours tumour

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

Unlike most other indices of lipid metabolism, apolipoprotein E aminoacid sequences are not disturbed by disease, and the apo E phenotype found in a patient will have been present since birth. A comparison of apo E phenotypes in cancer patients with those in matched controls might thus shed light on the relation between low cholesterol and cancer. If it is causal then the E-2 allele should be more common among patients and E-3 and E-4 more common among controls. On the other hand, equal distribution of apo E phenotypes among cases and controls would suggest that the association between low cholesterol and cancer is spurious.

If focusing of plasma lipoprotein metabolism on the apo E genes; epidemiologists interested in cholesterol and cancer should include it in their studies.

Department of Human Nutrition,
Agricultural University,
6703 BC Wageningen, Netherlands

MARTIJN B. KATAN

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

cardiovascular diseases. In: Nestel PJ, et al, eds. *Atherosclerosis VII: Proceedings of the Seventh International Atherosclerosis Symposium*. Amsterdam: Elsevier, 1986.

4. Brown MS, Kovanen PT, Goldstein JL. Regulation of plasma cholesterol by lipoprotein receptors. *Science* 1981; **212**: 628-35.
5. Utermann G, Steinmetz A, Weber W. Genetic control of human apolipoprotein E polymorphism: comparison of one- and two-dimensional techniques of isoprotein analysis. *Hum Genet* 1982; **60**: 344-51.
6. Utermann G, Kindermann I, Kaffarnik H, Steinmetz A. Apolipoprotein E phenotypes and hyperlipidemia. *Hum Genet* 1984; **65**: 232-36.
7. Robertson FW, Cumming AM. Effects of Apoprotein E Polymorphism on Serum Lipoprotein Concentration. *Arteriosclerosis* 1985; **5**: 283-92.
8. Utermann G. Genetic polymorphism of apolipoprotein E: impact on plasma lipoprotein metabolism. In: Crepaldi G, et al, eds. *Diabetes, obesity and hyperlipidemias III*. Amsterdam: Elsevier, 1985: 1-28.





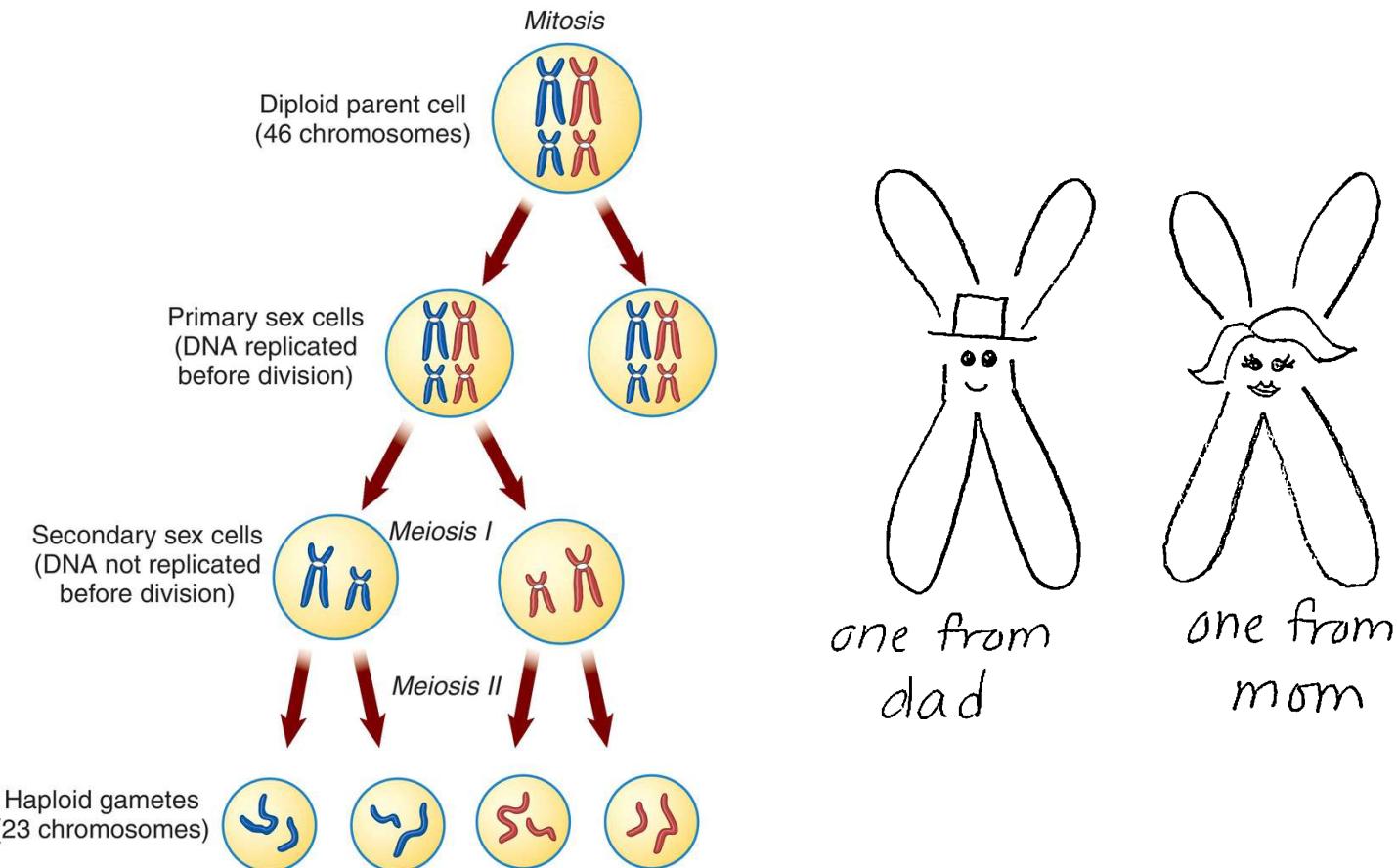
Gregor Johann Mendel - 20 July 1822 – 6 January 1884

Can you name the Laws of Mendel?

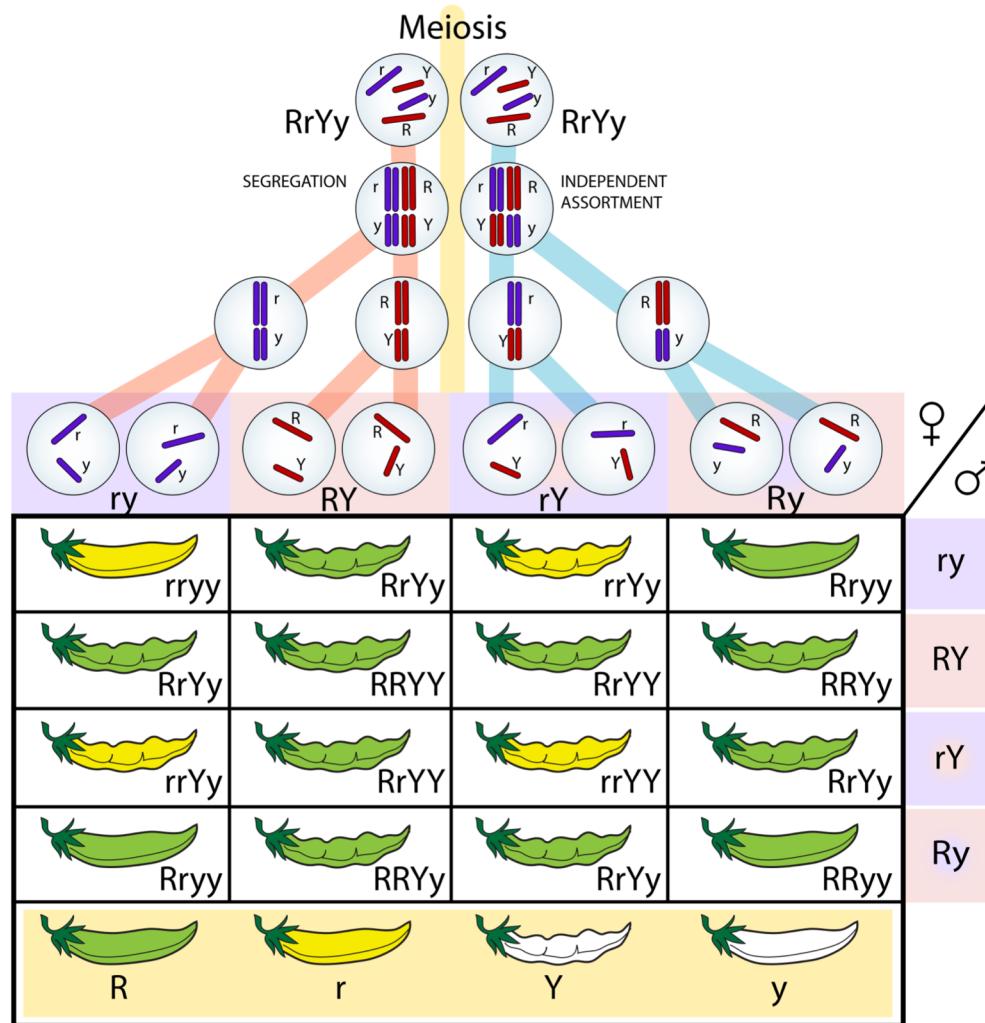


#1: Law of Segregation

- Process where a single cell divides **2 times** to produce **4 cells** containing **half** of the original amount of genetic information



#2: Law of Independent Assortment



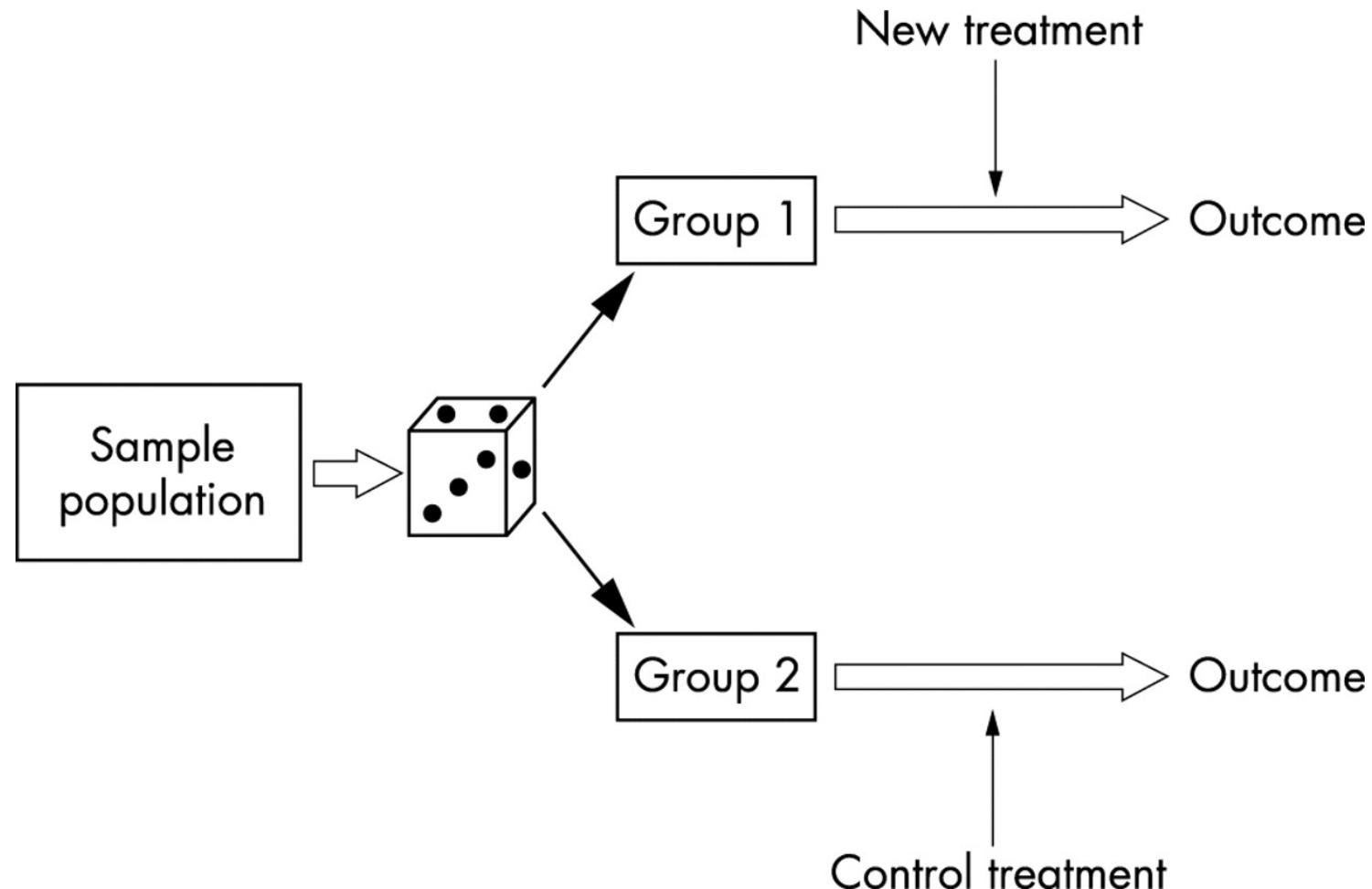
Can you name the Laws of Mendel?

TABLE 1: MENDEL'S LAWS OF INHERITANCE.

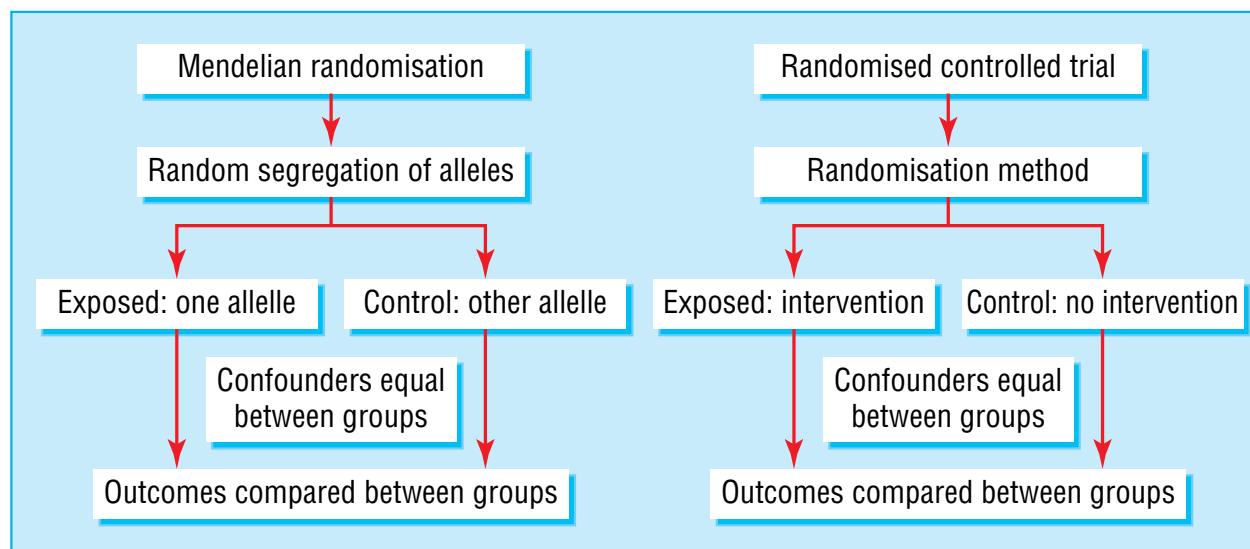
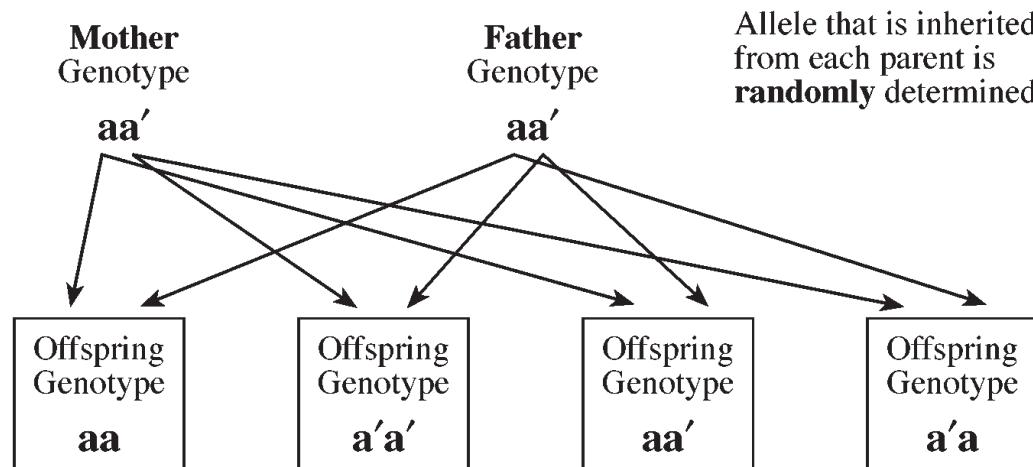
Law of Segregation	During gametogenesis alleles (copy) of each gene will segregate such that each gamete will carry only one of these alleles.
Law of Independent Assortment	Alleles associated with traits can segregate independently from parent to offspring.
Law of Dominance	Alleles can be dominant or recessive. Organisms carrying at least one dominant allele will show the effect associated with that allele.



Alleles are randomly assigned: a “natural” treatment in a “natural” randomized trial



Alleles are randomly assigned: a “natural” treatment in a “natural” randomized trial



Mendelian Randomization Studies: why

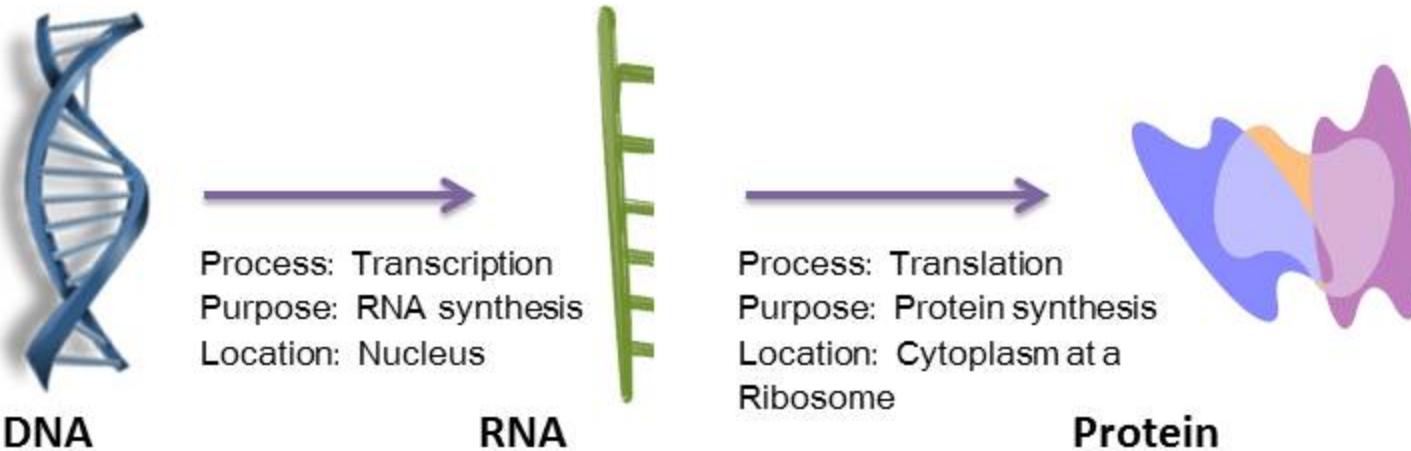
- Alleles are randomly, and independent assorted
- Eliminates confounding
- Useful and cheap – especially when there are...
 - ...no drugs,
 - ...the trial is not feasible...
 - ...or not ethical



Do you know what the Central Dogma and how does it relates to Mendelian Randomization?

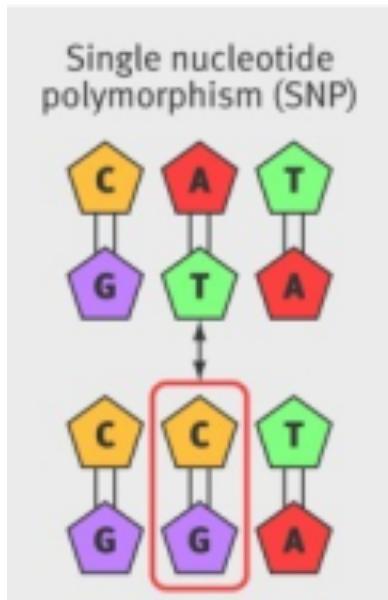


The Central Dogma



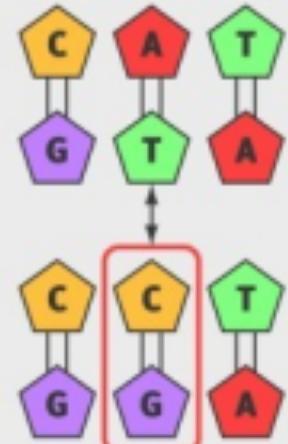
DNA contains the original codes for making the proteins that living cells need. mRNA is a copy of a gene located on the DNA molecule. mRNA will leave the nucleus of the cell and the ribosome will read its coding sequences and put the appropriate amino acids together.

Many types of genetic variants

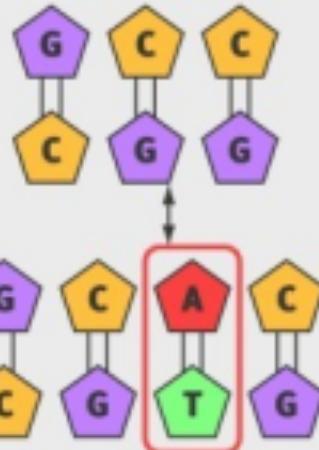


Many types of genetic variants

Single nucleotide polymorphism (SNP)

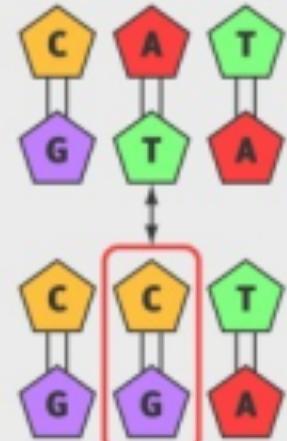


Insertion and deletion polymorphism (indel)

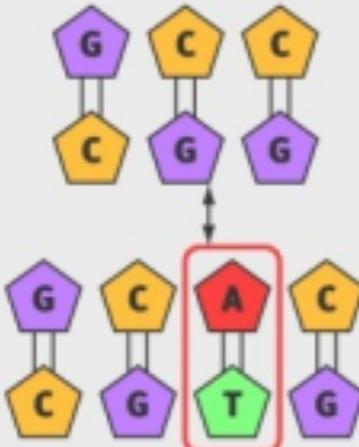


Many types of genetic variants

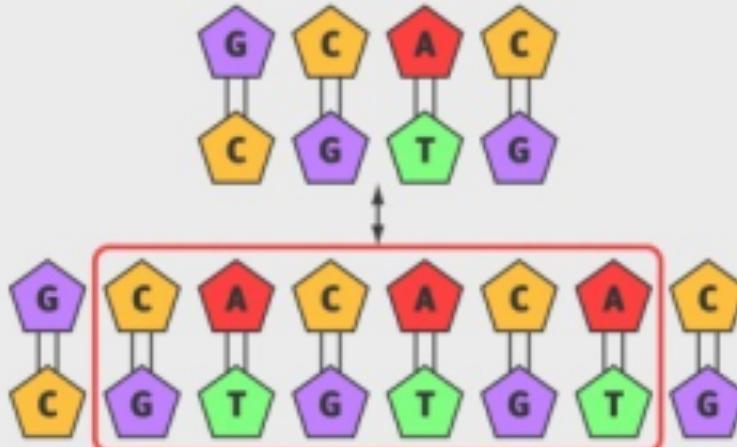
Single nucleotide polymorphism (SNP)



Insertion and deletion polymorphism (indel)

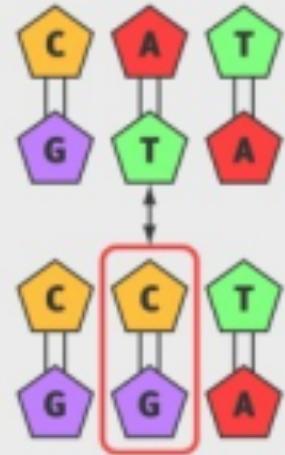


Nucleotide repeat polymorphism

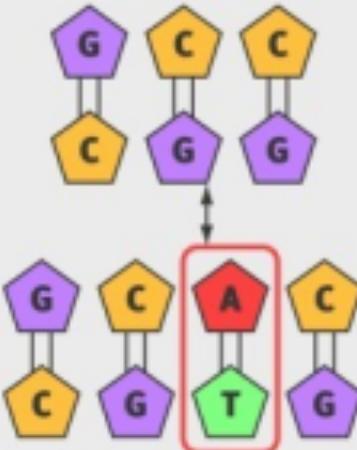


Many types of genetic variants

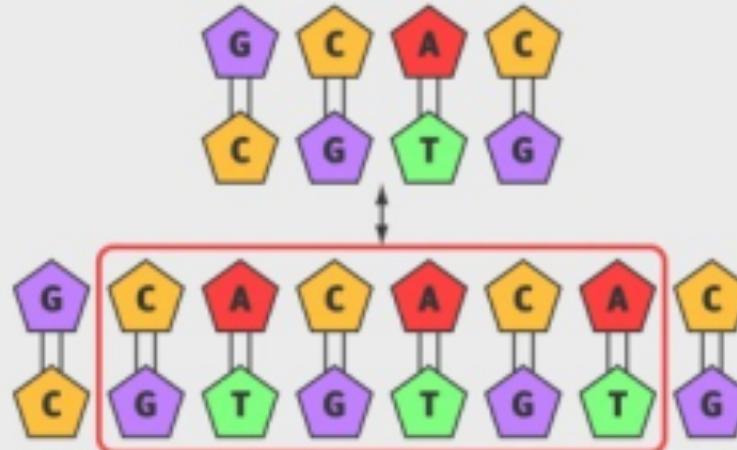
Single nucleotide polymorphism (SNP)



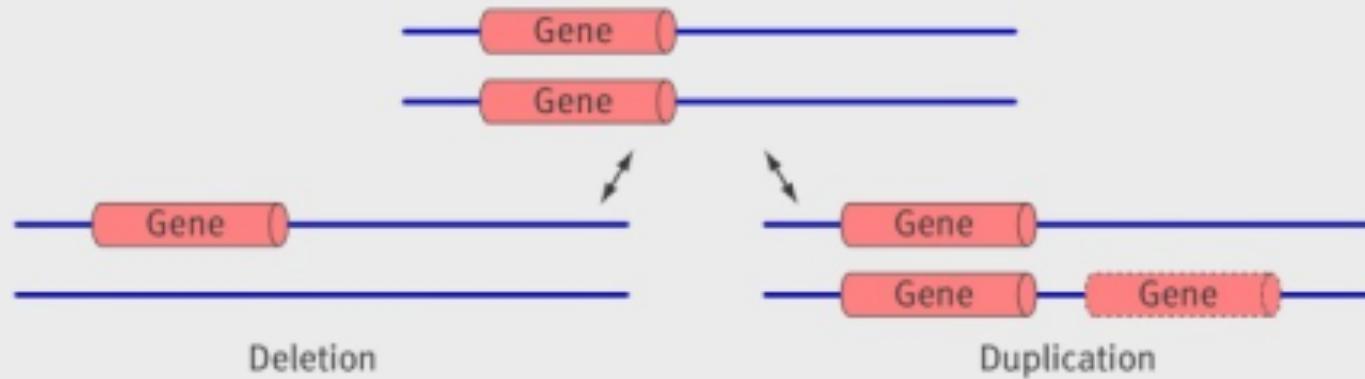
Insertion and deletion polymorphism (indel)



Nucleotide repeat polymorphism



Copy number variation



Genetic variants influence proteins

- GWAS of biomarkers identified many associated loci
- Genetic variant should be...
 - ...robustly associated to the biomarker (a strong “instrument”)
 - ...independent of any potential confounder
 - ...indirectly associated to the ‘endpoint’, *i.e.* via the putative causal biomarker
 - ...*common* enough in the general population (**why?**)



The concept of Mendelian Randomization

APOLIPOPROTEIN E ISOFORMS, SERUM CHOLESTEROL, AND CANCER

SIR,—It is unclear whether the relation between low serum cholesterol levels and cancer¹ is causal. In many studies occult tumour may have depressed cholesterol levels though in others the relation was found when serum cholesterol had been measured many years before the cancer was diagnosed. The relation is probably not explained by diet, because in the Seven Countries Study² men with different diets had corresponding differences in mean cholesterol levels experienced similar mean cancer rates.^{3,4} On the other hand, within each region cancer incidence was higher in men with a serum cholesterol in the lowest part of the cholesterol distribution for that country.⁵ Thus, naturally low cholesterol levels are sometimes associated with increased cancer risk.^{1,3}

Differences in the aminoacid sequence of apolipoprotein E (apo E) are major determinants of differences in plasma cholesterol levels within a population. Apo E has a key role in the clearance of cholesterol from plasma.⁶ The synthesis of apo E is under the control of several independent loci, each containing genes coding for the three major isoforms, E-2, E-3, and E-4, with respective population frequencies of about 8, 77, and 15%.⁵ The homozygous E-3/E-3 is the most common phenotype encountered and E-2/E-2 is the least common. From apo E-2 to apo E-3, one cysteine residue is replaced by arginine, and from apo E-3 to apo E-4 another cysteine residue is replaced. As a result the avidity of apo E containing lipoproteins for lipoprotein receptors increases from apo E-2 to apo E-3 to apo E-4. In several populations,^{6,8} including the Finns and the Japanese (Dr G. Utermann, personal communication), the gradient in serum cholesterol within the population is associated with a gradient in apo E phenotype. E-2 being associated with lower serum low-density lipoprotein and total cholesterol levels than E-3 and E-4. Thus, if a naturally low cholesterol favours tumour

growth, then subjects with the E-2/E-2 or E-2/E-3 phenotype should have an increased risk of cancer.

Unlike most other indices of lipid metabolism, apolipoprotein aminoacid sequences are not disturbed by disease, and the apo E phenotype found in a patient will have been present since birth. A comparison of apo E phenotypes in cancer patients with those in matched controls might thus shed light on the relation between low cholesterol and cancer. If it is causal then the E-2 allele should be more common among patients and E-3 and E-4 more common among controls. On the other hand, equal distribution of apo E phenotypes among cases and controls would suggest that the association between low cholesterol and cancer is spurious. Measurement of apo E phenotype by isoelectric focusing of plasma is a routine determination in lipid laboratories; epidemiologists interested in cholesterol and cancer should include it in their studies.

Department of Human Nutrition,
Agricultural University,
6703 BC Wageningen, Netherlands

MARTIJN B. KATAN

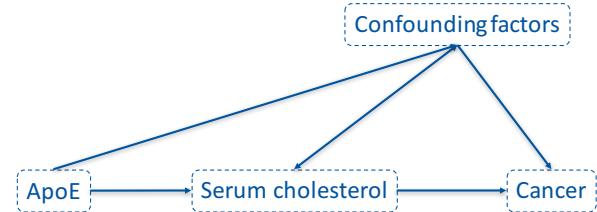
1. McMichael AJ, Jensen OM, Parkin DM, Zaridze DG. Dietary and endogenous cholesterol and human cancer. *Epidemiol Rev* 1984; **6**: 193–216.
2. Keys A, Aravanis C, Blackburn H, et al. Serum cholesterol and cancer mortality in the Seven Countries Study. *N Engl J Med* 1985; **312**: 870–83.
3. Katan MB. Effects of cholesterol-lowering diets on the risk for cancer and other non-cardiovascular diseases. In: Nestel PJ, et al., eds. *Atherosclerosis VII: Proceedings of the Second International Conference*. Amsterdam: Elsevier, 1986.
4. Brown MS, Kovacs PT, Goldstein JL. Regulation of plasma cholesterol by lipoprotein receptors. *Science* 1981; **212**: 628–35.
5. Utermann G. Human apolipoprotein E polymorphism: comparison of one- and two-dimensional techniques of isoprotein analysis. *Hum Genet* 1982; **60**: 344–51.
6. Utermann G, Kindermann M, Kahlffer S, Steinonen A. Apolipoprotein E phenotypes and hyperlipidemia. *New Eng J Med* 1984; **301**: 232–36.
7. Robertson FW, Cumming AM. Effects of Apolipoprotein E Polymorphism on Serum Lipoprotein Concentrations. *Atherosclerosis* 1985; **51**: 285–92.
8. Utermann G. Human apolipoprotein E polymorphism: its impact on plasma lipoprotein metabolism. In: Crepaldi G, et al., eds. *Diabetes, obesity and hyperlipidemias III*. Amsterdam: Elsevier, 1985: 1–28.

Confounding factors



The concept of Mendelian Randomization

- Use genetic variants *independent of any confounder*
- Include *prospective studies* to reduce reverse causation
- Include *deeply phenotyped studies* to correct for potential confounders in the observational analysis





Cystatin C and Cardiovascular Disease

AN EXAMPLE MR STUDY



UMC Utrecht





UMC Utrecht

Cystatin C and Cardiovascular Disease

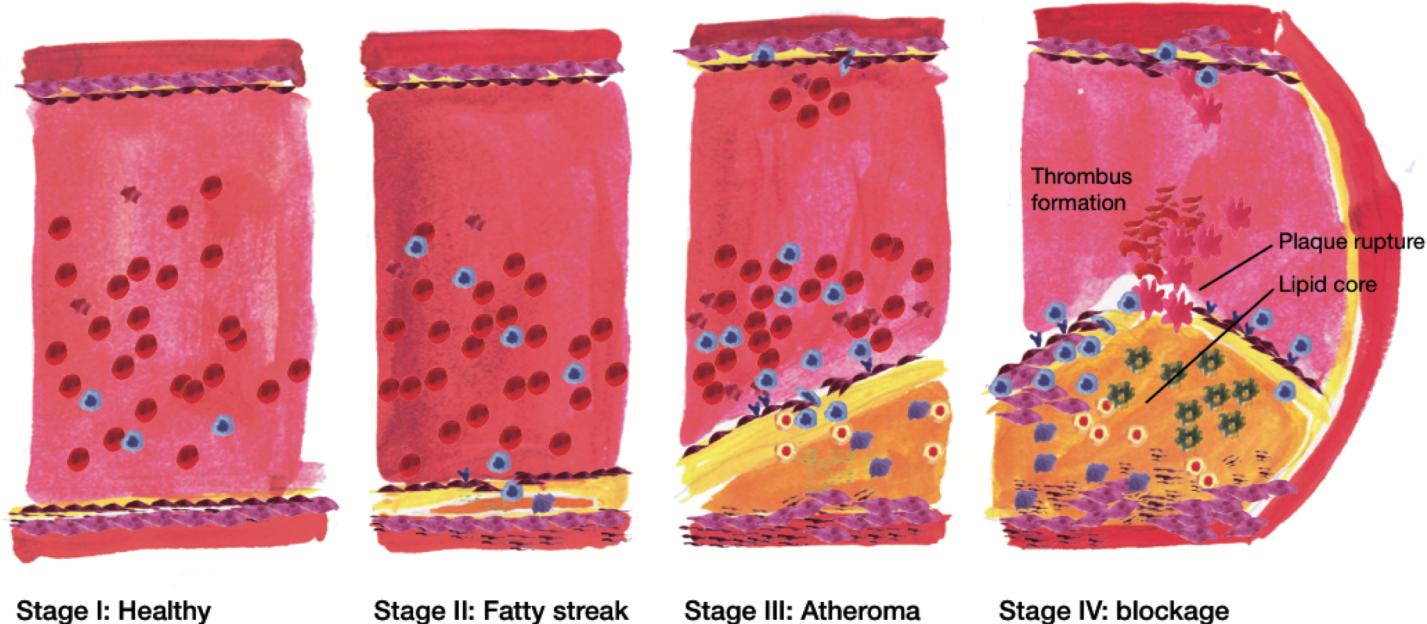
BACKGROUND



Atherosclerosis is the primary underlying cause of cardiovascular disease

Legend

- Monocyte
- Macrophage
- Foam cell
- Apoptotic foam cell
- Red blood cell
- Platelet
- Activated platelet
- Smooth muscle cell
- Endothelial cell
- Extracellular matrix
- Matrix metalloproteases
- Receptor



CVD: myocardial infarction or cerebrovascular accident (a.k.a. stroke)

A

Gemiddeld "klopt" het hart 60 maal, daarvoor is veel zuurstof nodig, die het hart krijgt via de kransslagaders.



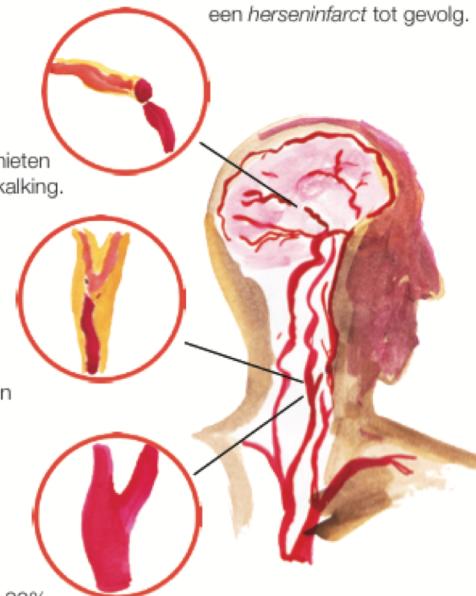
B

De losgeschoten thrombus kan verderop vast komen te zitten en zo een bloedvat afsluiten met een *herseninfarct* tot gevolg.

Een *thrombus* kan losbreken van de plek met aderverkalking.

Net als in het hart kan door risicofactoren *aderverkalking* ontstaan in bijvoorbeeld de halsslagader.

Het brein heeft ongeveer 20% van de totale energiebehoefte van het lichaam nodig per dag. Dit krijgt het via het bloed dat normaal door de gezonde linker en rechter *halsslagader* en de *wervelvliegaders* stroomt.



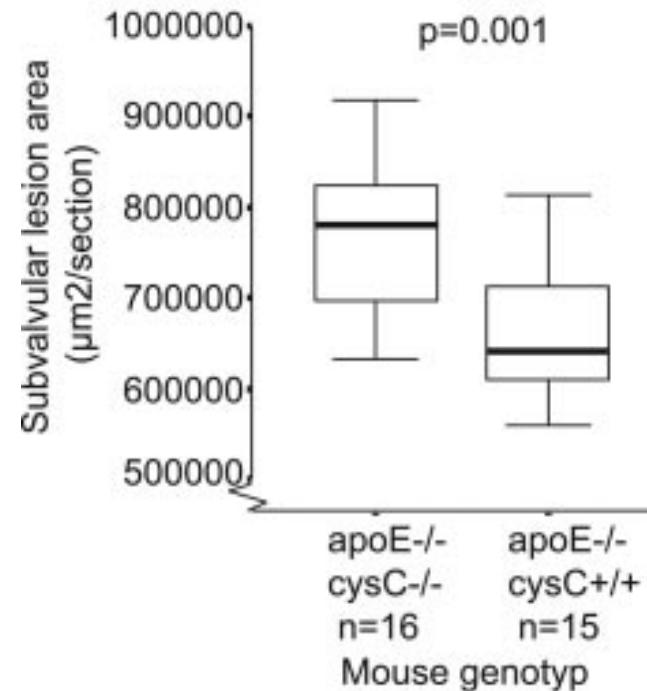
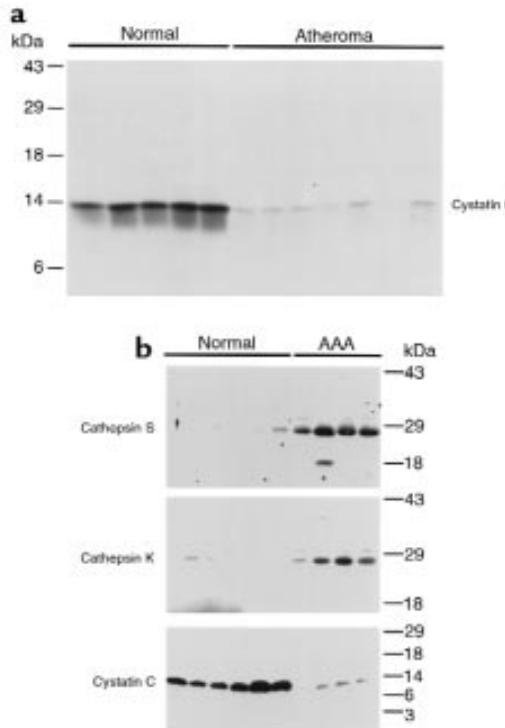
Cystatin C (CystC)

- Encoded by *CST3* on 20p11.21
- “Protease inhibitor”
 - It inhibits the activity of cathepsins (cysteine proteases) in vascular smooth muscle cells
 - Prevention of plaque formation by binding amyloid-beta precursor protein (APP) in Alzheimer’s disease



A role in human and murine atherosclerosis

- Reduction of CystC in abdominal aortic aneurysm (in smooth muscle cells)
- Lack of CystC in ApoE^{-/-} mice promotes atherosclerosis



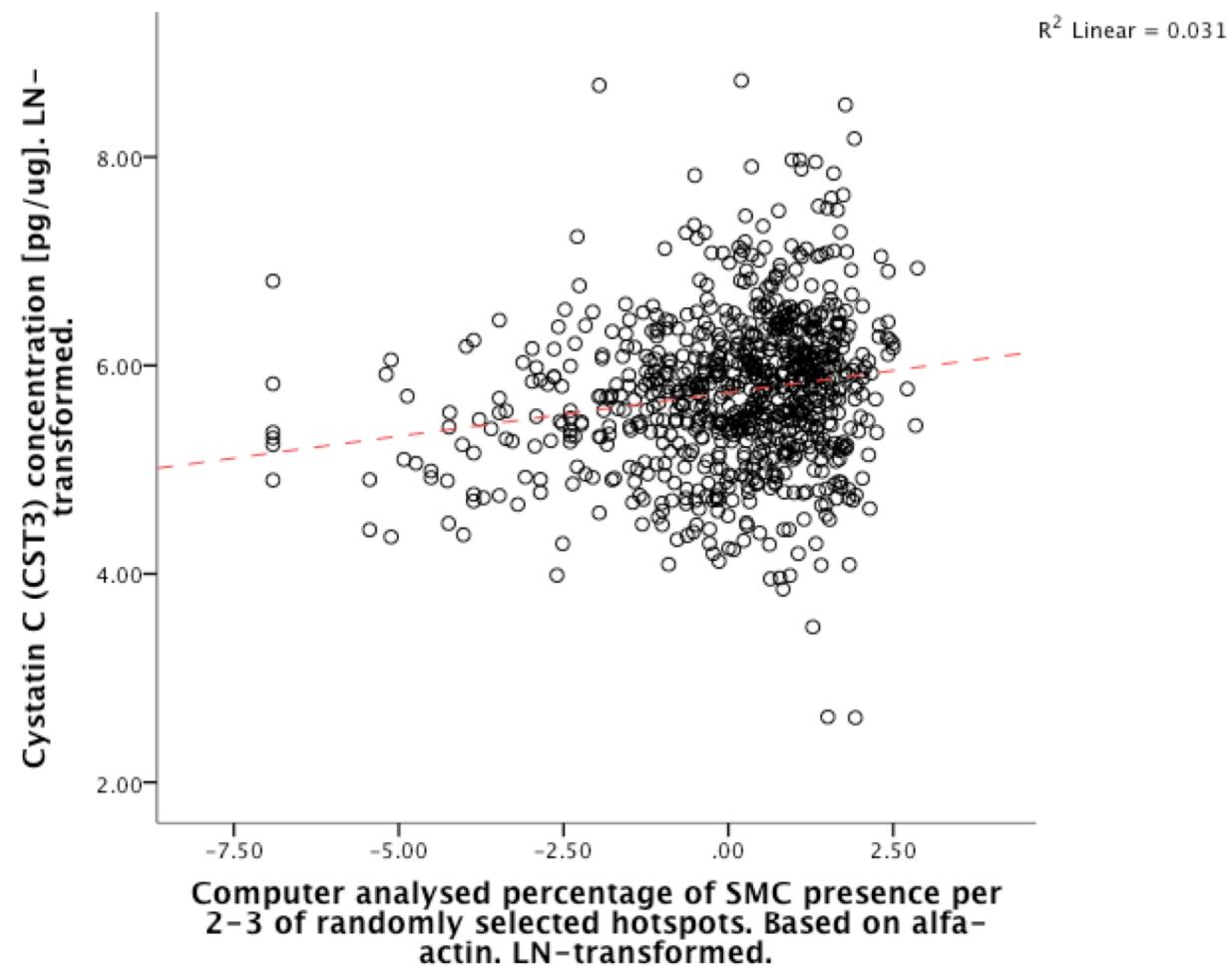
Shi *et al.* JCI 1999; Bengtsson *et al.* ATVB 2005



In carotid plaques CystC positively correlates with the percentage of smooth muscle cells

Rho = 0.185 (R = 0.176)

p = 1.10x10⁻⁷



CystC in atherosclerosis

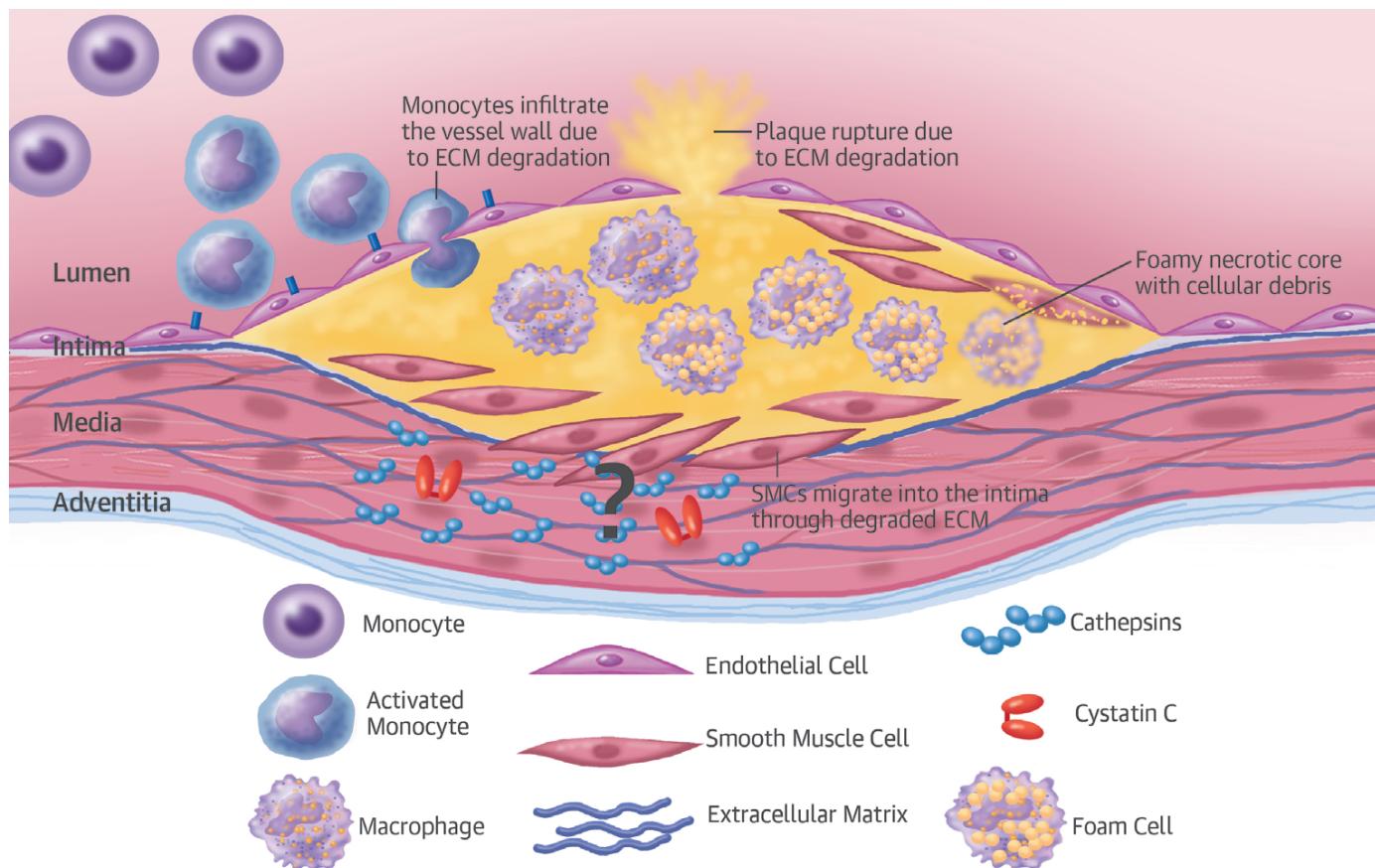


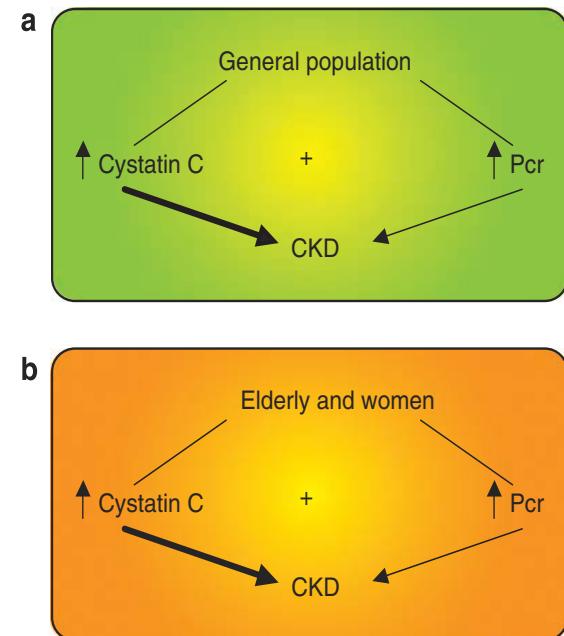
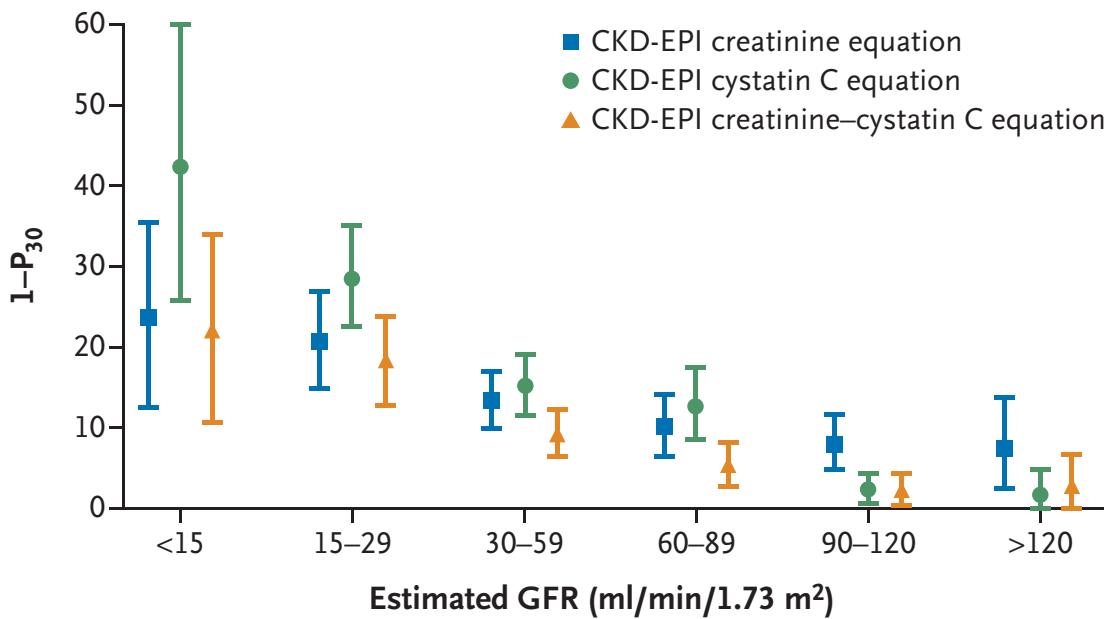
FIGURE 1 PRESUMED MECHANISM OF CYSTATIN C IN PLAQUES.

In vivo and *in vitro* animal and human studies have shown elevated levels of cathepsins and lower levels of cystatin C – a potent cathepsin inhibitor – in atherosclerotic tissue. Cathepsins are thought to degrade the extracellular matrix ECM, thus facilitating the migration of smooth muscle cells to the plaque core and promoting the destabilization.



CystC as a biomarker for CKD

- One of the best naturally occurring biomarker for human estimated glomerular filtration rate (eGFR)
- It is a better biomarker than creatinine when eGFR>90 (mL/min/1.73m³)



Fried *et al.* Kid Intern 2009; Inker *et al.* NEJM 2012

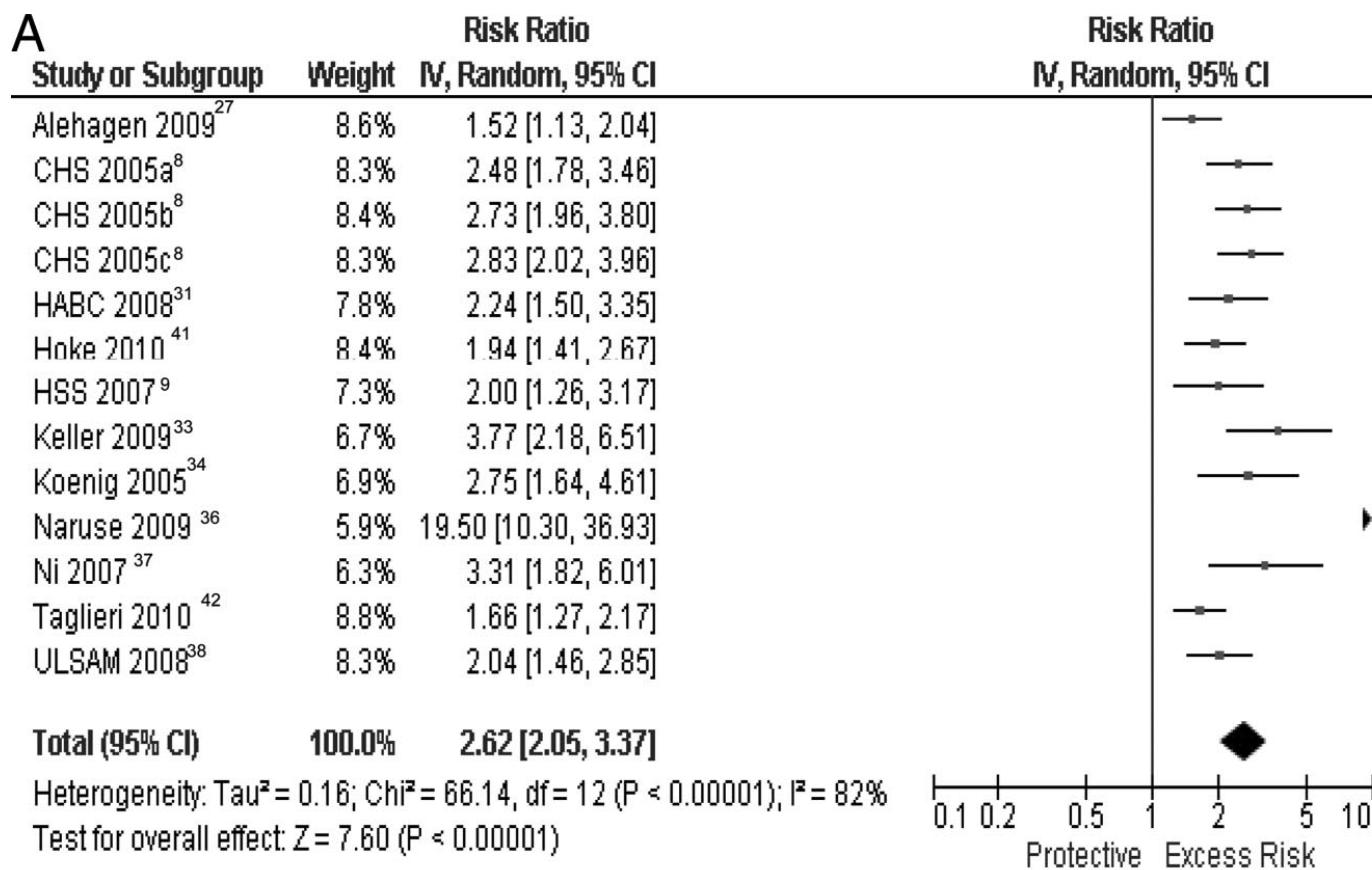
Epidemiological evidence

- In elderly without chronic kidney disease (CKD), CystC is prognostic of death, cardiovascular disease (CVD) and CKD
- Asymptomatic carotid atherosclerosis patients, CystC prognostic for CVD
- CystC associated with lesion size in acute stroke



Literature meta-analyzed

- CystC is associated with a higher risk of CVD, stroke and mortality in high-risk populations



Clinical applications

- 2 observational studies and 5 clinical trials focused on
 - Biomarker for diagnosis and prognosis
 - Marker of efficacy (treatment)
 - At least one drug in development: “RO4905417/INCLACUMAB” to reduce damage during PCI (Roche)
- Patents
 - Cavadis B.V.: exosome derived biomarker for adverse cardiovascular events (several patents [pending])
 - Biomarker for peripheral artery disease (2469279)
 - CystC as a “drug” for the treatment of atherosclerosis (WO/2001/030370)
 - Roche: selection of therapy in heart failure (WO/2014/086833)
 - Siemens: biomarker for pulmonary hypertension (WO/2014/052803)



Hypothesis...

- There is a large body of evidence from human and murine experiments and specimens that CystC is involved in atherosclerosis
- Epidemiological studies show that CystC is associated with increased risk CVD
- Plaque CystC and the percentage of SMCs in plaques are correlated

Is CystC causal to cardiovascular disease?





UMC Utrecht

Cystatin C and Cardiovascular Disease

FULFILLING ASSUMPTIONS





UMC Utrecht

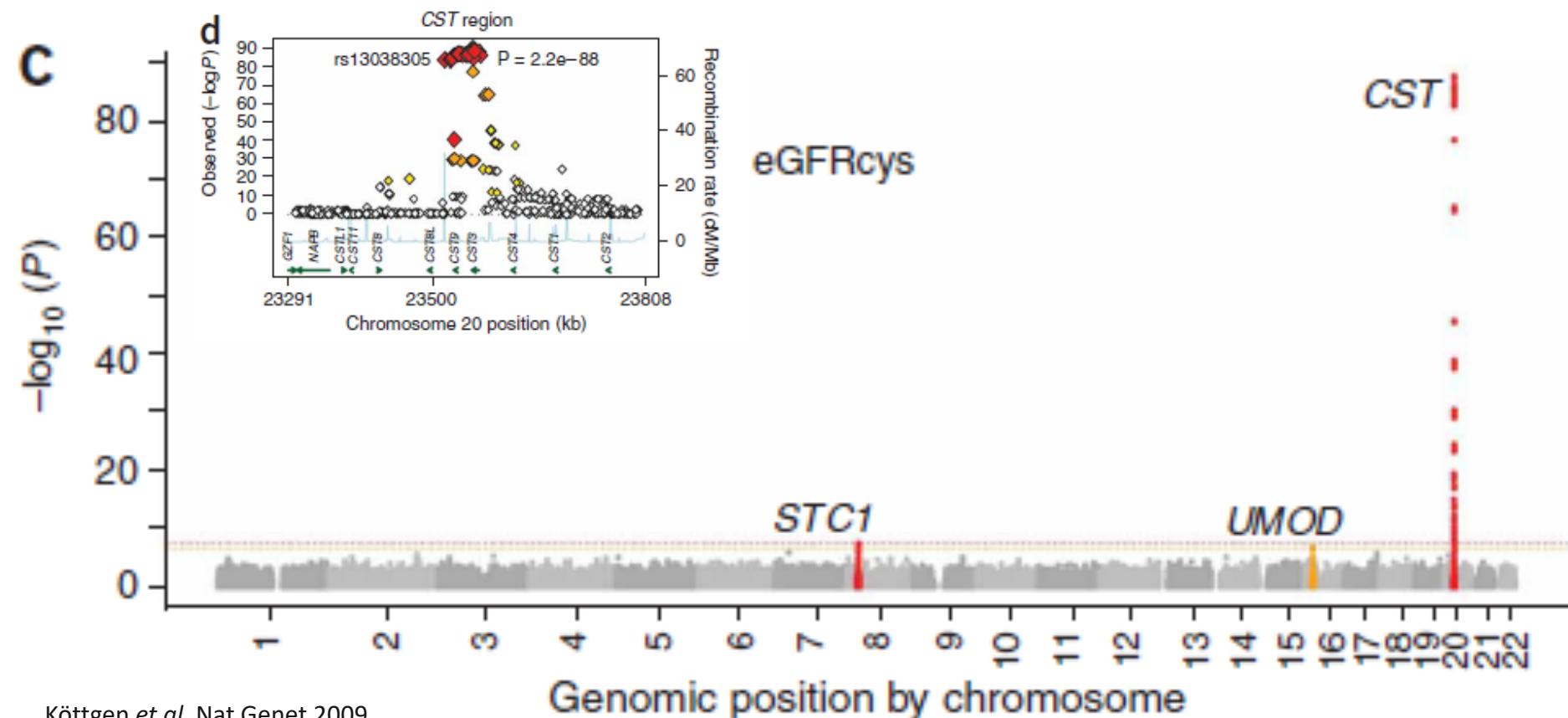
Fulfilling assumptions

SNP SELECTION



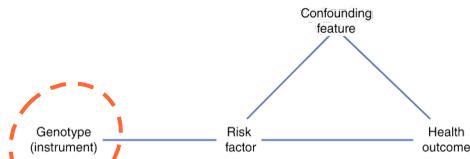
GWAS: locus with *CST3* pops up (naturally)

- Four Genome–Wide Association Study Meta-Analyses of eGFR and CKD revealed a region containing *CST3* (Cystatin C) as significantly associated
 - Identification of three loci associated with eGFRcys aka CystC expression: *STC1*, *UMOD*, *CST3*



Köttgen et al. Nat Genet 2009

SNP selection



- One GWAS on CystC, rs1158167 in Framingham Heart Study
- Three (meta-analyses of) GWAS on eGFR_{CystC}
 - rs911119, rs13038305
 - $eGFR_{CystC} = 76.7 \times (\text{serum CystC})^{-1.19}$
 - Log linear relation between serum CystC and eGFR

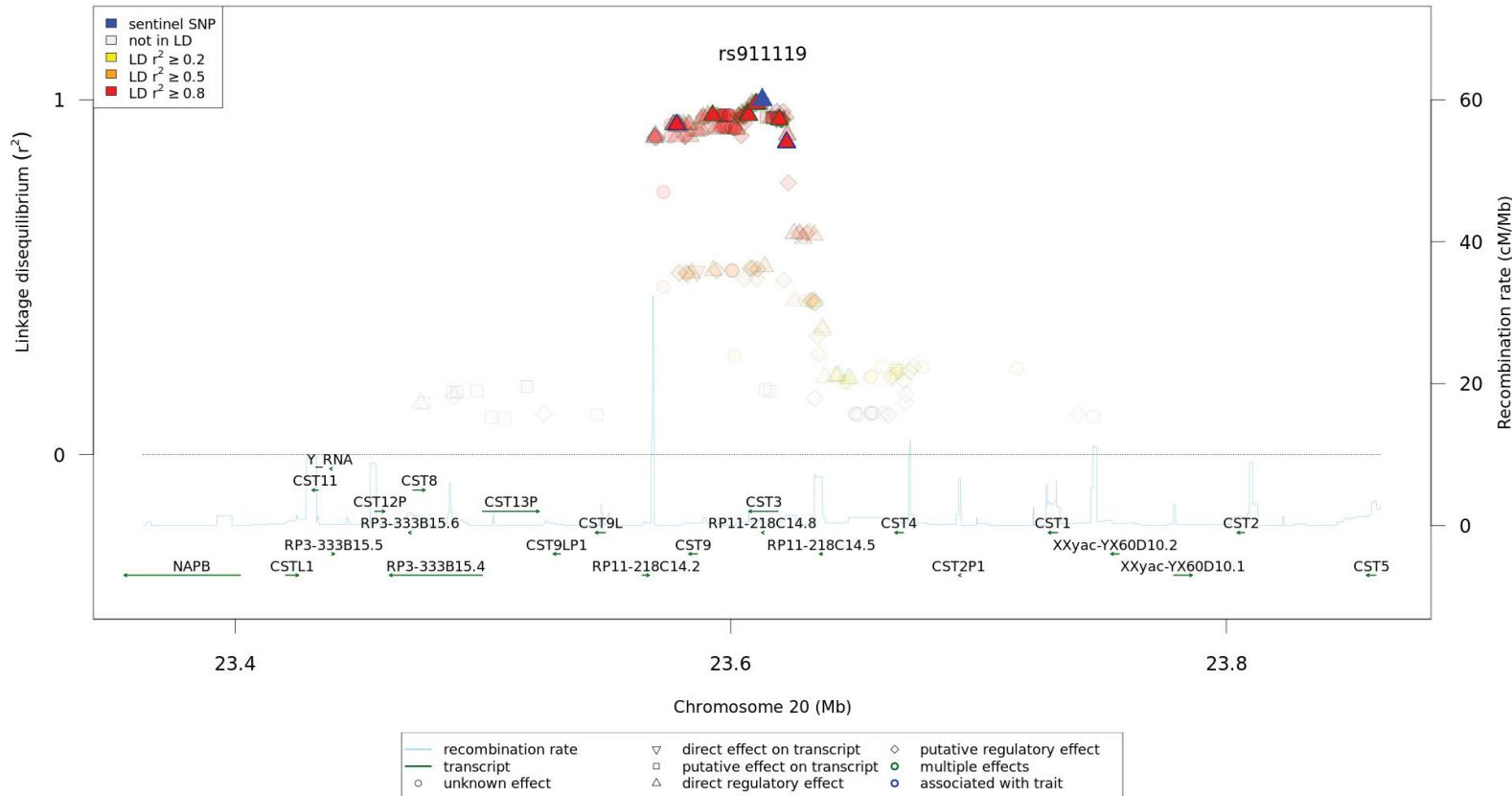
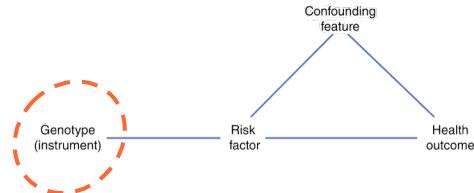
SUPPLEMENTAL TABLE 4: THE SELECTED SNP (RS911119, IN BOLD) AND ITS PROXIES USED IN THIS STUDY.

PROXY	DISTANCE	R ²	CHR	POSITION	MINOR	MAJOR	MAF	VARIANT TYPE	GENE
rs1158167	-34,548	0.913	20	23,526,189	G	A	0.240	downstream	-
rs17751897	-20,023	0.955	20	23,540,714	C	T	0.237	downstream	-
rs12625716	-5,892	0.955	20	23,554,845	A	G	0.236	downstream	-
rs6048952	-5,480	0.955	20	23,555,257	G	A	0.237	downstream	-
rs13038305	-2,475	1.000	20	23,558,262	T	C	0.233	intronic	CST3
rs911119	-	-	20	23,560,737	C	T	0.233	intronic	CST3
rs3827143	6,880	1.000	20	23,567,617	G	A	0.225	upstream	-
rs6114208	8,997	1.000	20	23,569,734	G	C	0.233	upstream	-
rs2424582	9,813	0.954	20	23,570,550	G	A	0.219	upstream	-

Distance of the SNP (in base pairs) from rs911119, linkage disequilibrium correlation (r^2), alleles, minor allele frequencies (MAF), and chromosomal positions are from 1000G pilot data (based on a European ancestral population [CEU]) available in SNAP and based on genome build 37³⁸⁴. The Variant type was determined using 1000G data from SNiPA¹¹⁴.



SNP selection

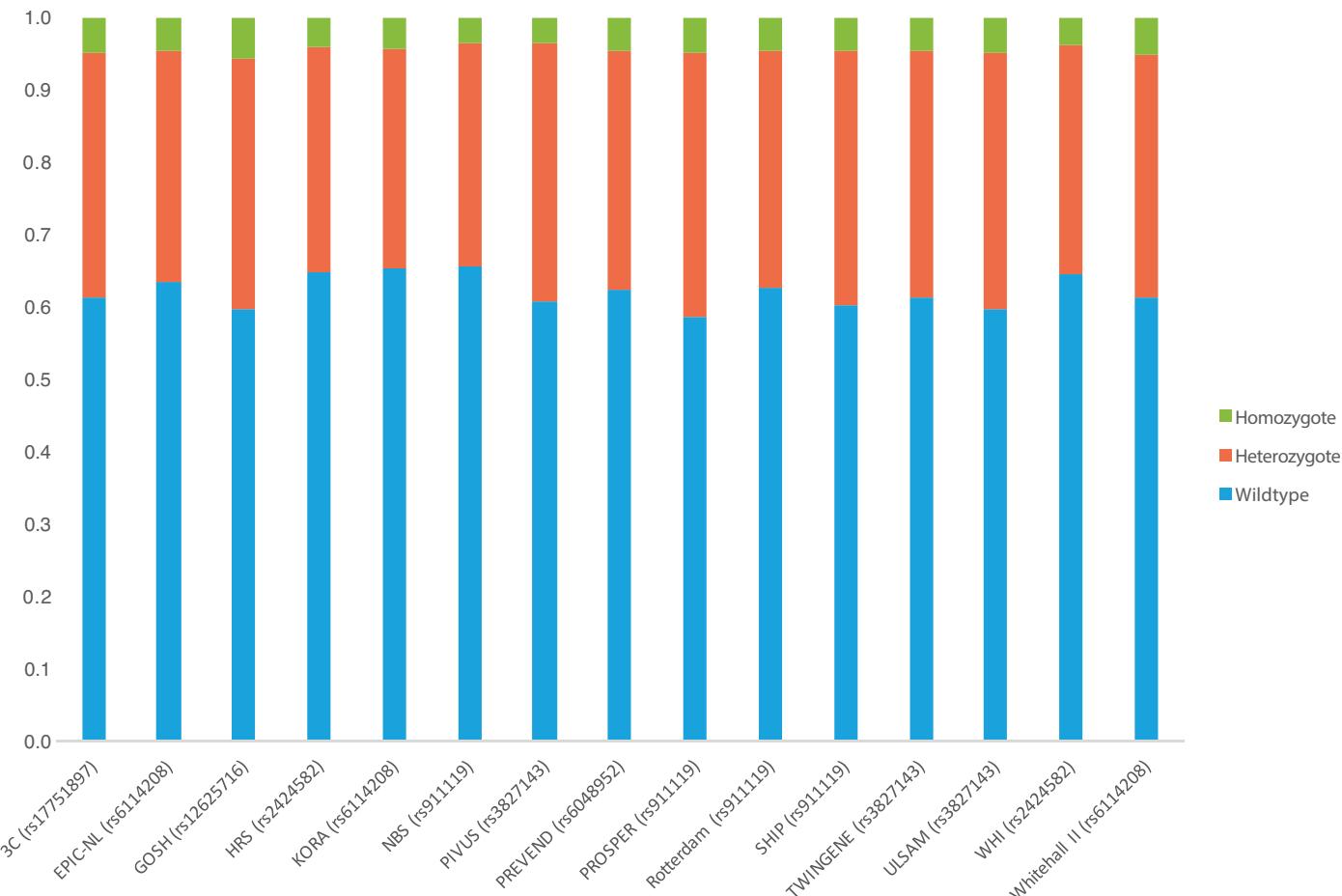
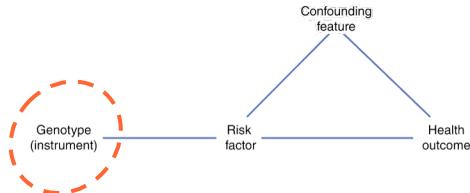


SUPPLEMENTAL FIGURE 1: OUR SENTINEL SNP (RS911119 IN BOLD) RELATIVE TO ITS GENOMIC LOCATION.

Each dot represents a SNP and is colored based on its linkage disequilibrium correlation (r^2) relative to rs911119 (y-axis). Chromosomal positions (x-axis) are from 1000G (phase 3, version 5, EUR population) available in SNiPA and based on genome build 37¹⁴. Functional SNP annotations are based on Ensembl77⁸⁵.



SNP selection



SUPPLEMENTAL FIGURE 2: GENOTYPE FREQUENCIES OF THE SNPs USED BY COHORTS IN THIS STUDY.

There was no SNP data available for the Tromsø Study.





UMC Utrecht

Fulfilling assumptions

STUDIES & CONFOUNDING



Studies included

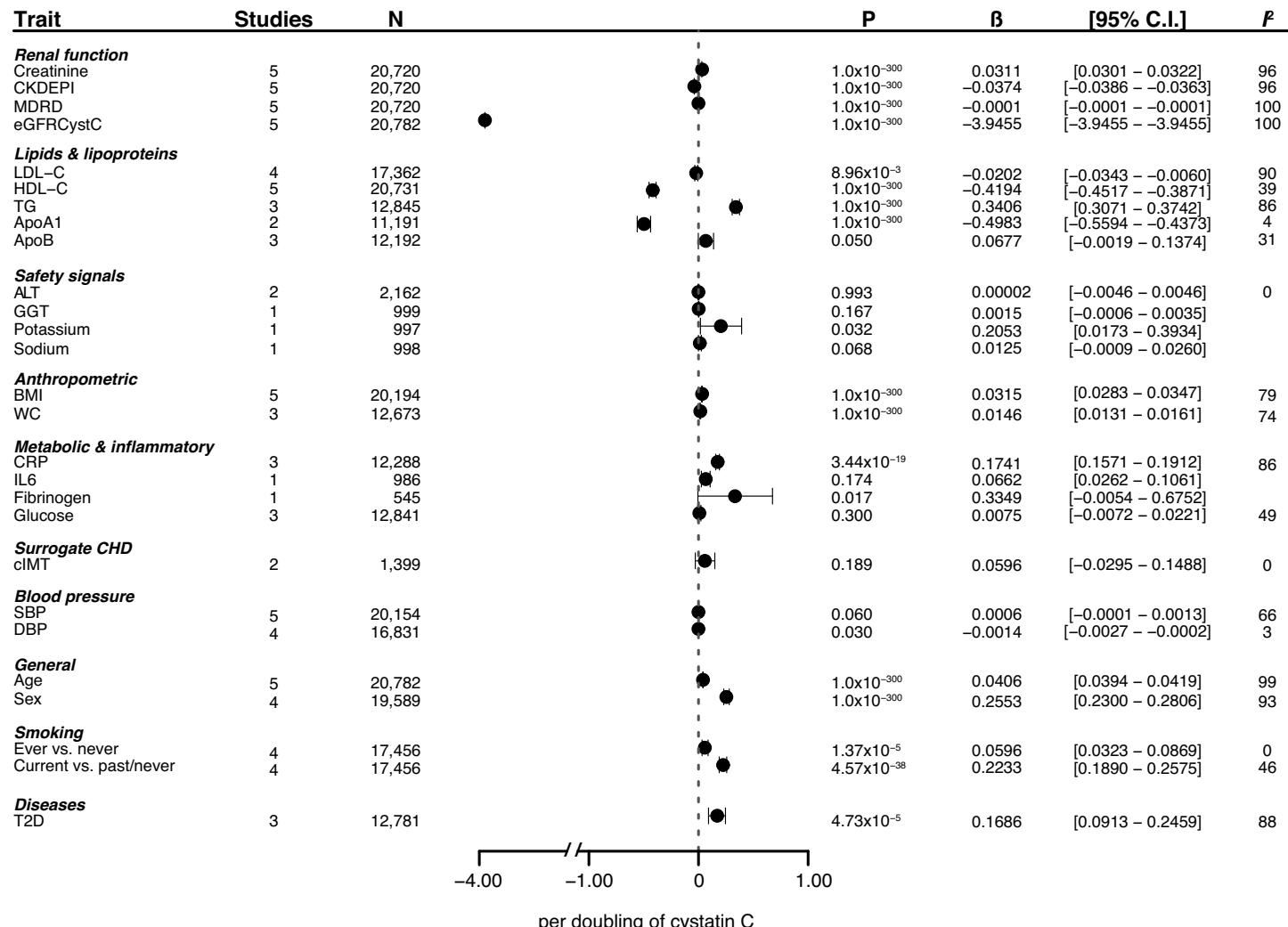
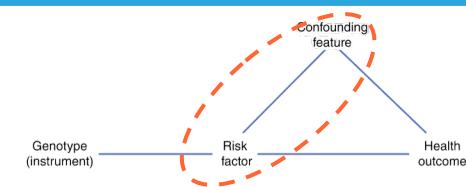
TABLE 1: CHARACTERISTICS OF THE 16 PROSPECTIVE COHORTS WITH GENOTYPE AND MEASURES OF CYSTATIN C.

STUDY	N _{TOTAL}	N _{SNP}	N _{CYSTC} ***	N _{CVD}	N _{CHD}	N _{IS}	N _{HF}	N _{MI}	MEN (%)	AGE (S.D.)	CYSTC (IQR)
3C	6,440	6,435	1,244	1,717	1,235	459	439	486	39.19%	74.30 (5.52)	0.92 (0.24)
EPIC-NL	6,265	5,192	-	1,967	1,430	537	-	1,430	22.39%	53.80 (10.23)	-
GOSH	1,478	1,479	-	493	111	235	233	-	42.08%	51.08 (11.86)	-
HRS	7,844	5,585	5,777	-	-	-	-	-	-	-	0.64 (0.34)
KORA	4,856	1,867	4,676	540	341	255	-	341	49.53%	49.75 (14.11)	0.80 (0.21)
NBS	1,819	1,297	-	66	-	66	-	170	49.48%	61.05 (10.26)	-
PIVUS	1,016	949	1,004	255	175	71	75	105	49.90%	70.20 (0.17)	0.90 (0.19)
PREVEND	3,245	3,245	3,245	236	190	58	-	-	50.26%	49.42 (12.25)	0.87 (0.17)
PROSPER*	5,244	5,150	-	2,561	2,034	779	211	762	48.13%	75.34 (3.35)	-
ROTTERDAM	7,983	5,974	3,906	3,579	1,934	1,328	1,625	1,176	38.90%	73.06 (7.49)	1.11 (0.28)
SHIP	3,224	3,224	3,212	114	19	87	-	134	48.08%	54.46 (15.26)	0.88 (0.30)
TROMSØ	6,129	-	6,129	1,251	-	494	-	881	47.59%	60.59 (10.25)	0.86 (0.18)
TWINGENE**	6,902	6,902	6,740	932	610	287	206	-	47.23%	64.83 (8.26)	1.02 (0.30)
ULSAM	1,221	1,107	1,193	503	285	175	220	-	100.00%	71.00 (0.64)	1.25 (0.27)
WHI	7,854	7,844	-	4,831	2,934	2,115	-	2,934	0.00%	67.97 (6.58)	-
WHITEHALL II	4,961	5,011	-	349	254	111	-	254	74.58%	49.19 (5.99)	-
OVERALL	76,481	61,261	37,126	19,394	11,552	7,057	3,009	8,673	-	-	-

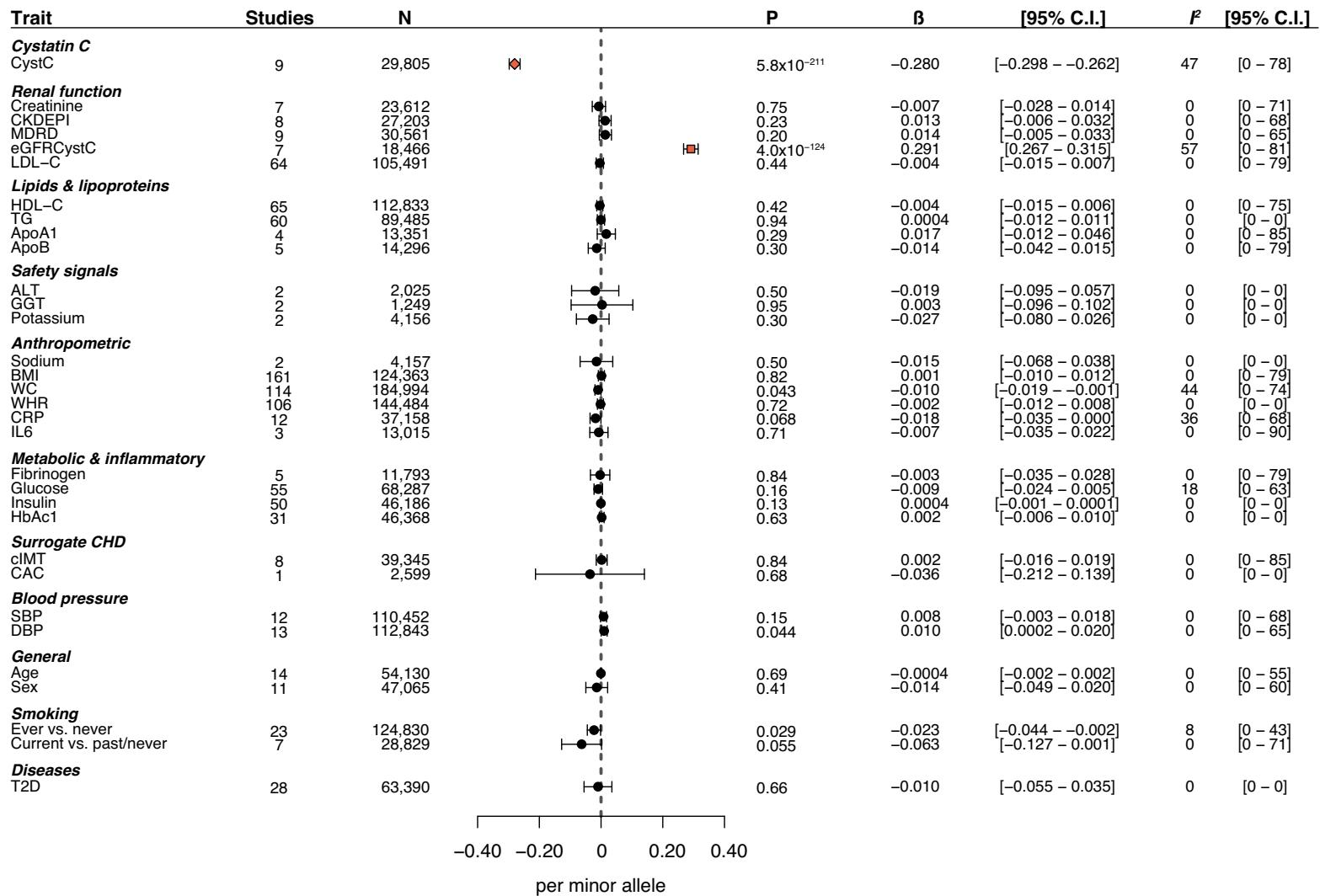
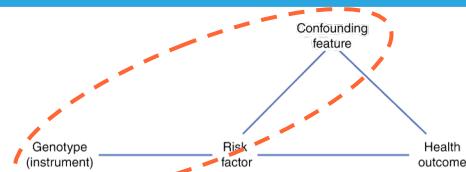
N_{total}: the total number of individuals in the study; *N_{SNP}*: the total number of individuals with genotype data; *N_{cystatin C}*: the total number of individuals with measurements of serum cystatin C; *N_{CVD}*: the total number of cardiovascular disease cases (as a composite of incident and prevalent fatal and non-fatal CHD, IS and HF); *N_{CHD}*: the total number of coronary heart disease incident and prevalent cases; *N_{IS}*: the total number of ischemic stroke incident and prevalent cases; *N_{HF}*: the total number of heart failure incident and prevalent cases; *N_{MI}*: the total number of myocardial infarction incident and prevalent cases; *Mean age* in years with standard deviation (s.d.); *Median Cystatin C* in mg/L with interquartile range (IQR). * PROSPER is a randomized clinical trial. **For the association of SNP with cystatin C concentrations 9,488 samples were available in TWINGENE (see **Material and Methods**). *** Genetic data were available in 29,805 of the 37,126 individuals that had values for cystatin C, which we used to associate rs911119 to circulating cystatin C. Note that for the genetic analysis of CVD, CHD, IS, and HF we excluded cohorts that contributed towards consortia.



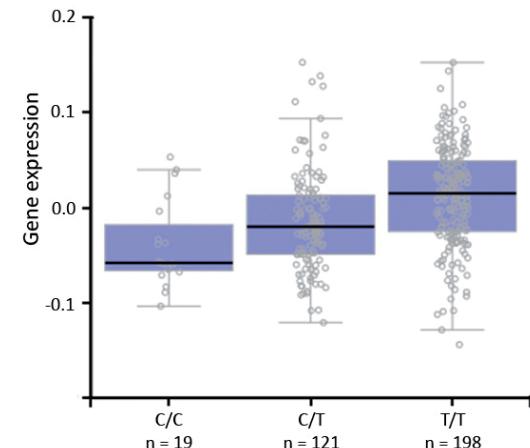
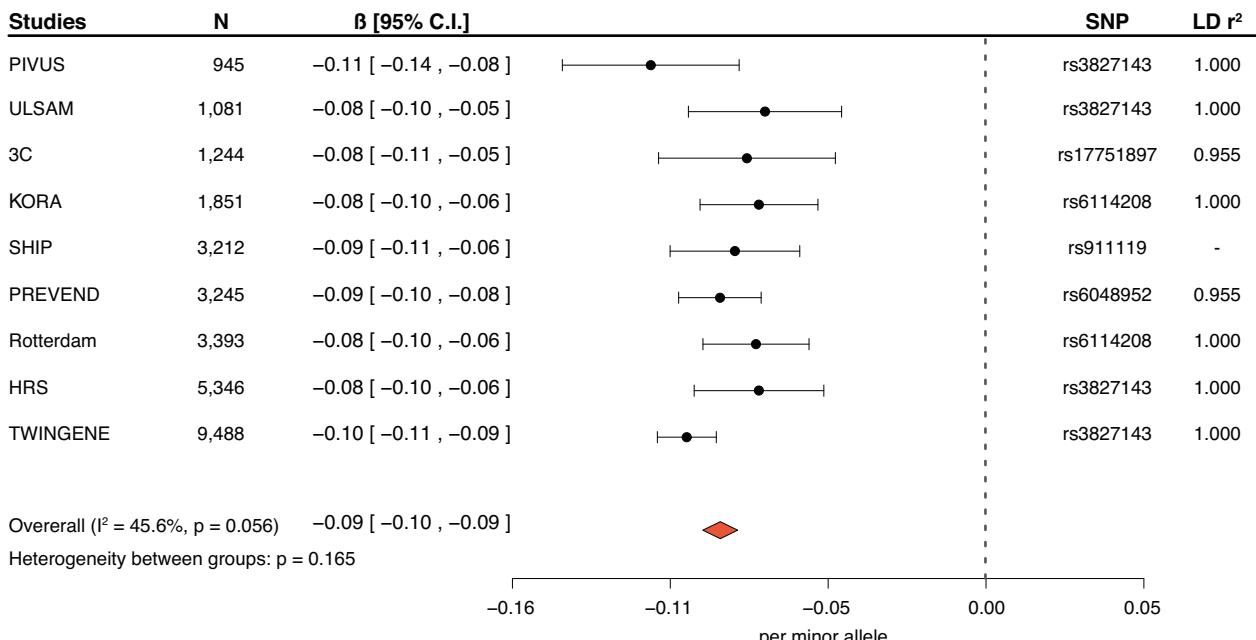
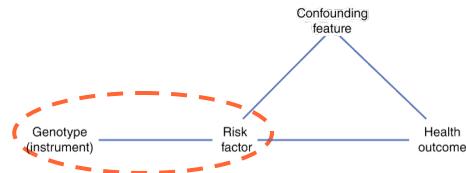
Serum cystatin C vs. risk factors



Cystatin C variant vs. risk factors



SNP vs. cystatin C



SUPPLEMENTAL FIGURE 4: ASSOCIATION OF GENETIC VARIANTS IN CST3 WITH CIRCULATING CYSTATIN C PER COHORT.

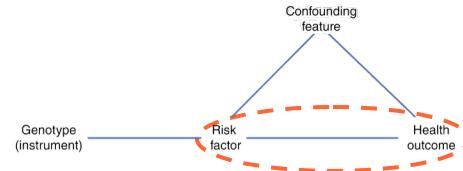
Effect sizes (β) are presented as log₂-transformed cystatin C concentrations in mg/L per minor allele. N indicates the total sample size. C.I.: confidence interval. SNP: the single-nucleotide polymorphism available for analysis. LD: the linkage disequilibrium r^2 relative to rs911119.

- Per minor allele there is 6.13% decrease in CystC [mg/L]
- $\beta = -0.09 [-0.10 - -0.09]$, $p = 5.95 \times 10^{-163}$, $N = 29,805$
- This explains $\approx 2.75\%$ of the phenotypic variation

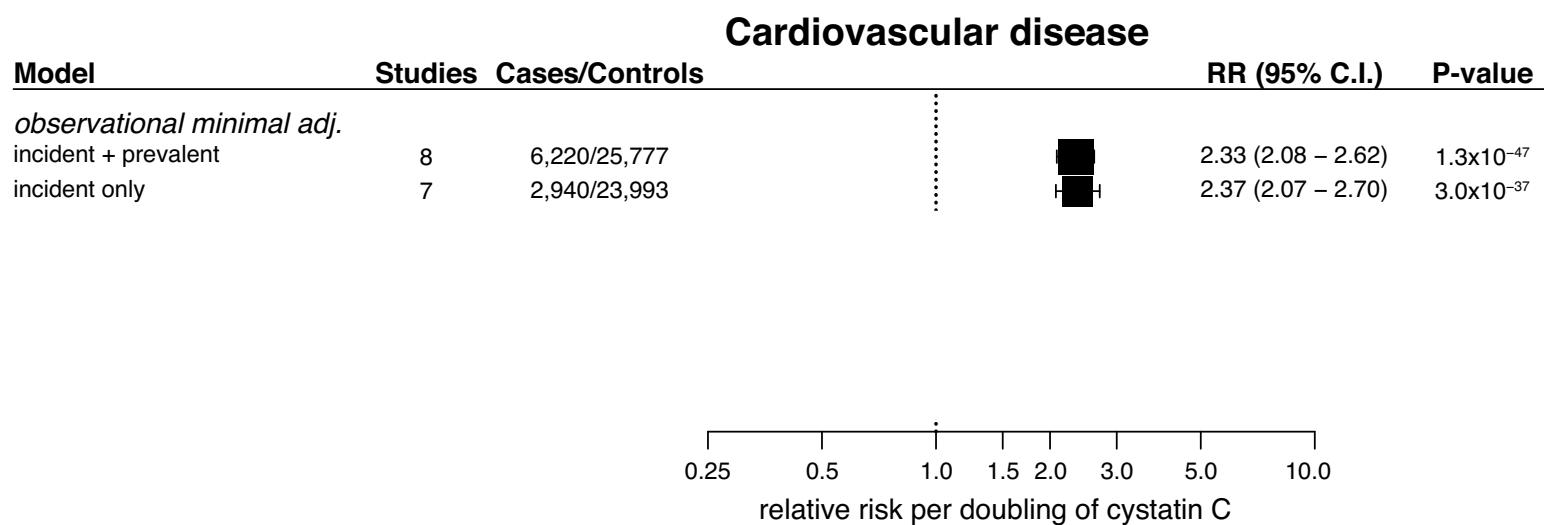
Van der Laan S.W. et al. JACC 2016



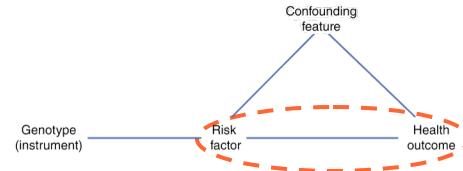
Observational analysis



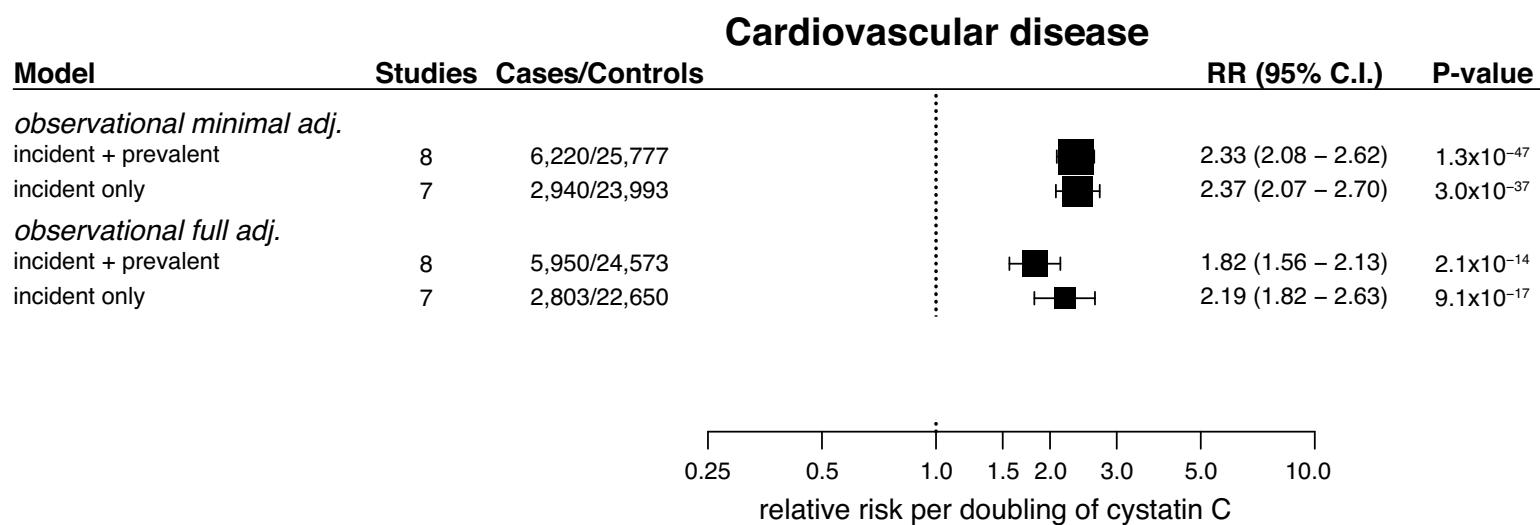
- Minimally adjusted model: age + sex



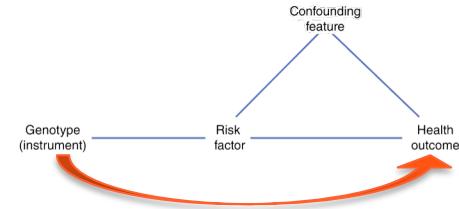
Observational analysis



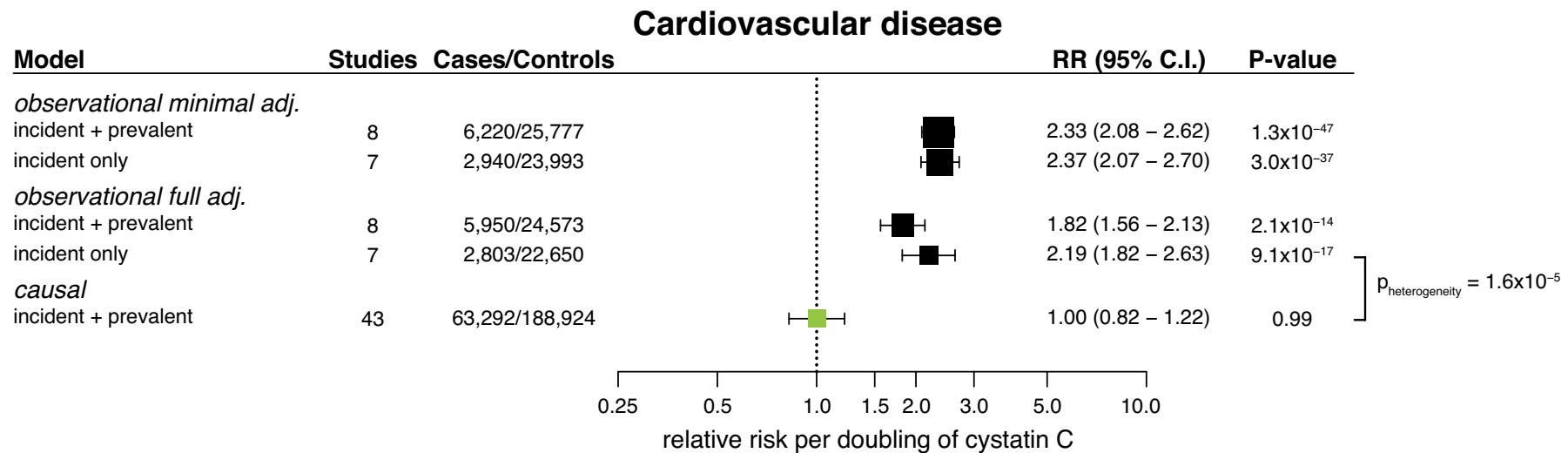
- Minimally adjusted model: age + sex
- Fully adjusted model: age + sex + HDL + systolic blood pressure + BMI + smoking status + CKDEPI



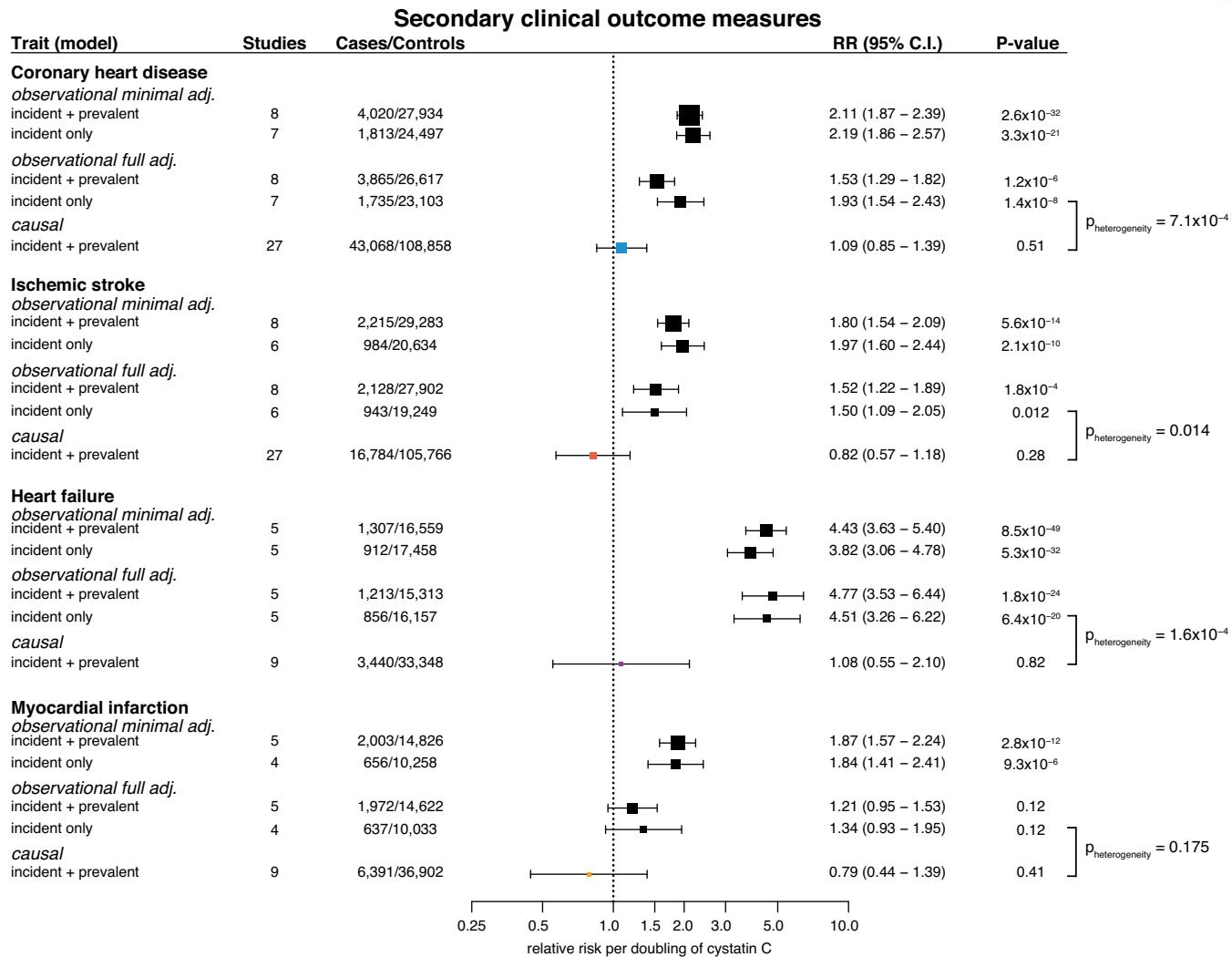
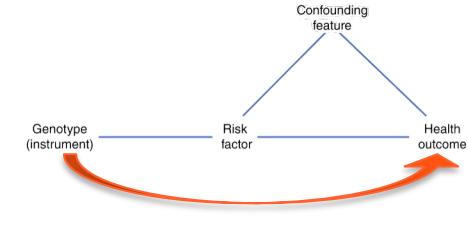
Mendelian randomization analysis



- Minimally adjusted model: age + sex
- Fully adjusted model: age + sex + HDL + systolic blood pressure + BMI + smoking status + CKDEPI
- Causal effect: *nada, nothing, niets, rien, nichts, ничего*
- Interaction analysis*: significant – models are different



Other outcomes





UMC Utrecht

Fulfilling assumptions

POWER?



Power

Power to detect association

$$E[\chi^2] \approx N\gamma^2 p(1-p)R^2$$

Effect size

Sample size

LD with causal variant

Allele frequency

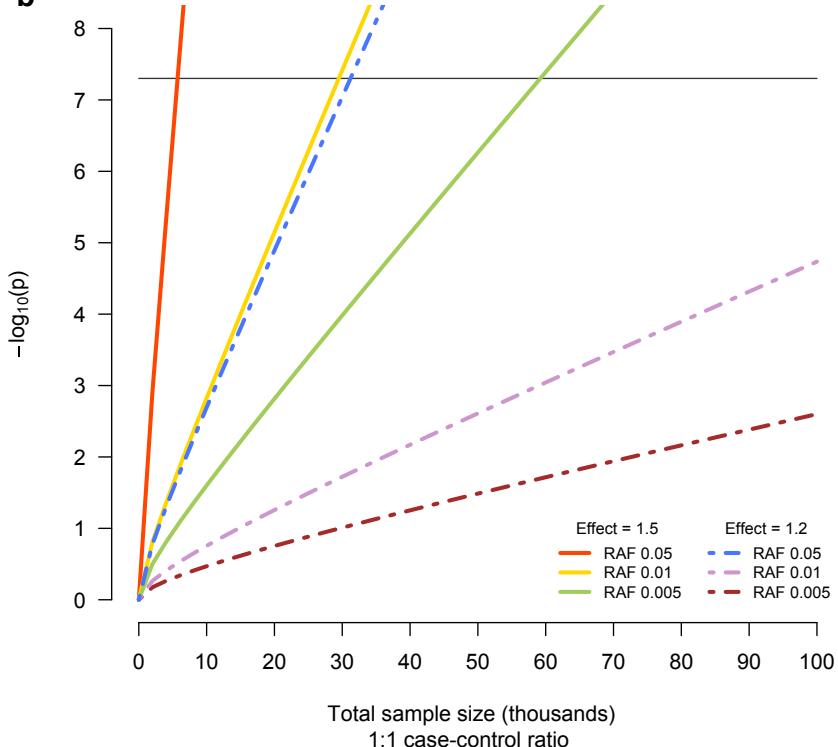


The power problem

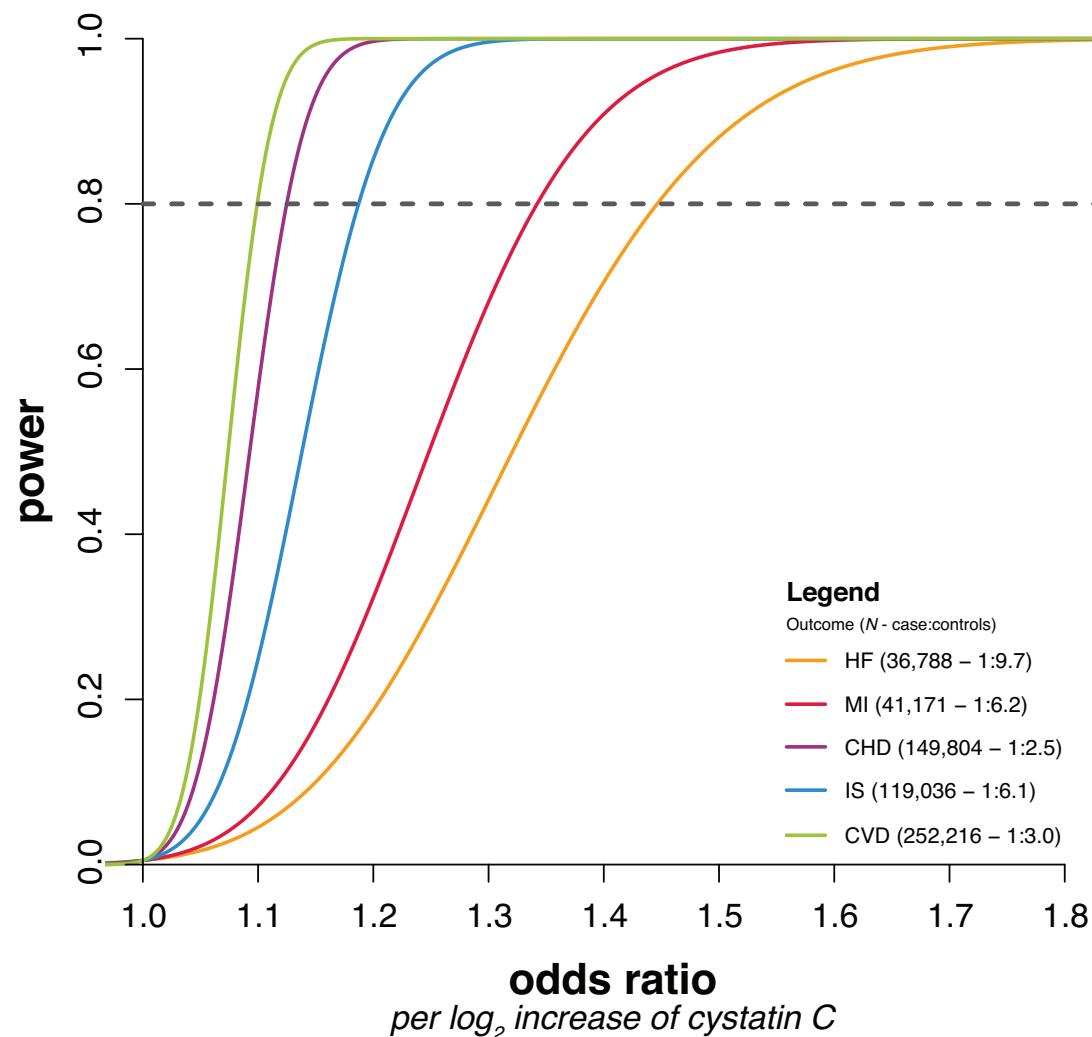
a

$$\begin{array}{c} \text{Power} \downarrow \\ E[X^2] \approx \text{Sample size} * \text{Effect size} * \text{Allele frequency} * \text{LD to causal variant} \\ N * \gamma^2 * p(1-p) * R^2 \end{array}$$

b



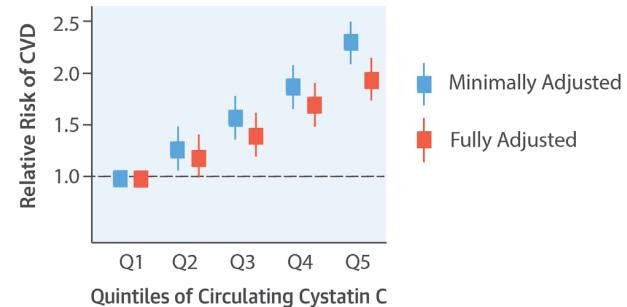
Did we have power?



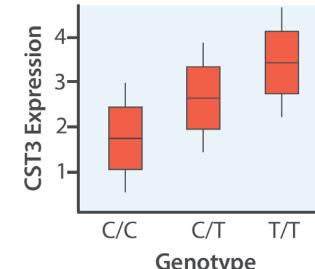
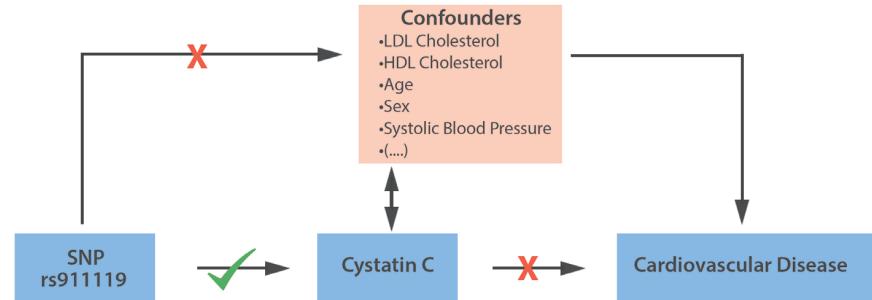
So actually you have learned more today...

- There is a correlation between the percentage of SMCs and CystC in plaques, and CystC levels appear to be lower in SMCs from AAAs
- Strong epidemiological evidence exists for a role of CystC in CVD
- Variants in CystC associated with serum CystC levels, eGFR_{Cystc}
- MR analysis: we find no evidence for a causal effect of CystC on CVD

A. Observational Epidemiology



B. Mendelian Randomization



Acknowledgements

Supervisors & Mentors

Gerard Pasterkamp, Paul de Bakker, **Folkert Asselbergs**, Hester den Ruijter & Dominique de Kleijn

Department Vascular Surgery

Frans Moll, Gert-Jan de Borst & Jean-Paul de Vries

Department of Pathology

Aryan Vink & Petra van der Kraak

Uppsala University, Sweden

Tove Fall, Johan Ärnlöv

Pennsylvania University, Philadelphia

Michael V. Holmes

Data Acquisition

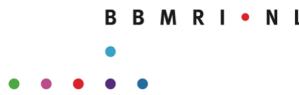
Arjan Schoneveld, Sander van de Weg, Loes Collé, Claudia Tersteeg, Evelyn Velema, Mirjam Smeets, Pleunie van den Borne, Sanne Willems & Kristy Vons

Data Analysis

Jessica van Setten, Kim Jie & Joyce Vrijenhoek, Vinicius Tragante do Ó

Financial support

ICIN (09.001), GENIUS (CVON2011-19), FP7 EU project
CVgenes@target HEALTH-F2-2013-601456, BBMRI CP2013-63,
Cavadis B.V.



Disclosure: this work is partly financed by Cavadis



A Mendelian Randomization study in practice

Sander W. van der Laan | s.w.vanderlaan-2@umcutrecht.nl | @swvanderlaan



Disclosure: this work is partly financed by Cavadis

CAVADIS
www.cavadis.com

What do we need?

- Can you think of the elements needed for a MR study?



What do we need?

- Can you think of the elements needed for a MR study?
 - Relevant question – duh
 - SNP → where, how?
 - Biomarker → observational analyses
 - Outcome → observational analyses
 - Studies → observational analyses, PubMed
 - Power → R script, online tools
 - Analysis plan
 - Time



Your question leads you to...

- Your biomarker & Outcome
- The studies → search PubMed

...and your SNP selection...

- PubMed
- GWAS Central
- GTEx Portal



Calculate Power

- There are a couple of sites:
- <https://sb452.shinyapps.io/power/>
- <https://cnsgenomics.shinyapps.io/mRnd/> -- a bit more extensive
- Assignment



Calculate Power: assignment

- PubMed: Elliott P et al. JAMA 2009
- Calculate power given their data using one of the aforementioned calculators



Perform a MR study...

- ...wait...what?
- Yes, you can: <http://www.mrbase.org/beta/>



Perform a MR study...: assignment

- Think of your favorite gene/biomarker...
- ...find an appropriate SNP...,
- ...and your favorite outcome measure (can also be continuous)...,
- ...and use the website to perform the MR

- But hold on: there are couple of other decisions you'll have to make... 😊



The screenshot shows the MRBase beta website. At the top, there's a sidebar with links: Welcome to MR Base, About, Acknowledgements, and Data access agreement. Below this is a main header with the MRBase logo (two red hexagons) and the text "MRBASE". A sub-header reads: "A platform for Mendelian randomisation using summary data from genome-wide association studies". A message at the top says: "To begin analysis please review the data access agreement and accept by logging in with your google account." A blue "Get started" button is present. On the right, there are two boxes: one for SNP-PHE with 3,994,161 entries and another for TRAITS with 340,161 entries. The bottom of the sidebar has "Current status" and "Beta phase release". The browser's address bar shows "www.mrbase.org/beta/".

This screenshot shows the "Data access agreement" page. It contains several sections of text with numbered headings (6, 7, 9, 10, 11) detailing terms and conditions. At the bottom, it says "Login button should appear below..." followed by a large blue "Accept the data access agreement via Google to get started" button. The browser's address bar shows "www.mrbase.org/beta/".

6. Identification of data subjects: You agree not to use, combine, manipulate or transform the GWAS Data in any way that would or might enable you to identify any living individual to which the GWAS Data relates, in breach of data protection laws anywhere in the world.

7. Disclaimers: We do not guarantee that (a) the Platform or any GWAS Data will always be available or uninterrupted; (b) the Platform or any GWAS Data will be accurate, complete, free from errors or omissions or secure or free from bugs or viruses; or (c) that the result of using the Platform or any GWAS Data will be accurate, adequate or fit for any particular purpose (more general information on the principles, assumptions and limitation of Mendelian randomization can be found on the papers recommended in the "About" tab). Where the Platform contains links to other sites or resources provided by third parties, these links are provided for your information only and you acknowledge that we have no control over the content of those sites or resources. All warranties, representations, conditions and all other terms of any kind whatsoever implied by statute or common law, to the fullest extent permitted by applicable law, excluding from these Terms.

8. Limitation of liability: You assume sole responsibility for the results obtained from the use of the Platform and any GWAS Data and for conclusions drawn from such use. The University shall not be liable for any damage or other loss whatsoever arising out of or in connection with your use of the Platform or any GWAS Data. The University shall not be liable in any circumstances whether in contract, tort (including for negligence), misrepresentation (whether innocent or negligent), restitution or otherwise for any special, indirect or consequential loss, costs, damages, charges or expenses however arising under these Terms including but not limited to loss of funding or loss of opportunity, goodwill or reputation. Nothing in these Terms excludes or limits any liability which cannot be excluded or limited by applicable law.

9. Suspension and termination: The University may, at its sole discretion, suspend, withdraw, discontinue or change all or any part of the Platform (in respect of any single user, group of users or all users of the Platform) without notice and whether or not arising from any breach of these Terms. The University will not be liable to you in such circumstances.

10. Changes to these Terms: The University may revise these Terms at any time by amending this page. Please check this page from time to time to take notice of any changes, as they will be legally binding on you.

11. Governing law and jurisdiction: These Terms and any claims or disputes arising out of or in connection with them (including non-contractual disputes) shall be governed by the laws of England and Wales whose courts shall have exclusive jurisdiction to settle the same.



Safari File Edit View History Bookmarks Window Help ⚡ 58% Tue 28 Mar 23:40

Huis Apple ID iCloud Reference WaldamarGroup Family Apple Cavadis deCODE Google Network News Shop Omica Science SNP MR

mrbase.org

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.

Welcome to MR Base

About Acknowledgements Data access agreement

Logged in as Sander W. van der Laan s.w.vanderlaan@gmail.com

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

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

GTEX eQTL catalog

These data were generated by the GTEX consortium and were originally obtained from <http://www.gtexportal.org/>. If you use these data please cite:

GTEX Consortium. Human genomics. The Genotype-Tissue Expression (GTEX) pilot analysis: multitissue gene regulation in humans. *Science*. 2015 May 8;348(6235):648-60

The SNPs present in this catalog are those flagged by GTEX as being the best SNP for each gene x tissue combination.

Gene name

SPP1

SPP1 CSPP1

Safari File Edit View History Bookmarks Window Help

mrbase.org

Huis Apple ID iCloud Reference WaldamarGroup Family Apple Cavadis deCODE Google Network News Shop Omica Science SNP MR

www.mrbase.org/beta/?state=RQcDoizmpMV8y74RSr7Q&code=4/IHbUXwfJixrp01P8q3kpawQ3KJfI9_vFhvIMGdIhw#

tissue SNP effect_allele se n
 gene_name other_allele

Row selection

Select All Deselect All

Search:

tissue	gene_name	SNP	beta	se	pval
Gastroesophageal Junction					
5 Esophagus Mucosa	SPP1	rs28449439	-0.436453	0.0856513	7.89851e-7
6 Esophagus Muscularis	SPP1	rs28628889	-0.434575	0.0705272	4.54794e-9
7 Heart Atrial Appendage	SPP1	rs12642655	-0.407315	0.0931665	0.0000260727
8 Heart Left Ventricle	SPP1	rs398092589	-0.308101	0.0721369	0.000034032
9 Nerve Tibial	SPP1	rs55717206	-0.461477	0.0651537	2.00418e-11
10 Pancreas	SPP1	rs2853749	-0.351531	0.0743794	0.00000669654
11 Spleen	SPP1	rs200563570	-0.622558	0.130326	0.0000100835
12 Stomach	SPP1	rs4585264	-0.307493	0.0699515	0.0000223815
13 Testis	SPP1	rs71594829	0.804118	0.0904835	7.16718e-15
14 Whole Blood	SPP1	rs6532024	-0.218013	0.0484261	0.0000096925

All All All All All All

Showing 1 to 14 of 14 entries

Safari File Edit View History Bookmarks Window Help mrbase.org Tue 28 Mar 23:41

mrbase.org

Select outcomes for analysis

The MR Base database houses a large collection of summary statistic data from hundreds of GWAS studies. In order to perform two sample MR, the SNPs that were selected for the exposures will be extracted from the outcomes that you select here.

Please select the outcomes that you want to test for being causally influenced by the exposures.

Studies available in MR base

Display columns

ID	Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Category	Year	Population	Priority	Sd	Sex	Subcategory	Unit
1	Adiponectin	Dastani Z	ADIPOGEN													
2	Body mass index	Locke AE	GIANT													
6	Coronary heart disease	Peden	C4D		15420	15062										
7	Coronary heart disease	Nikpay	CARDioGRAMplusC4D		60801	123504										

Show 10 entries Search:

Perform MR analysis

Trait	Note	First author	Consortium	Number of cases	Number of controls	Sample size	Number of variants	Year	Category	Subcategory		
1	Adiponectin	Dastani Z	ADIPOGen					39883	2675209	2012	Risk factor	Protein
2	Body mass index	Locke AE	GIANT					339224	2555511	2015	Risk factor	Anthropometric
6	Coronary heart disease	Peden	C4D		15420	15062	30482	540233	2011	Disease	Cardiovascular	
7	Coronary heart disease	Nikpay	CARDioGRAMplusC4D	60801	123504	184305	9455779	2015	Disease	Cardiovascular		
8	Coronary heart disease	Schunkert H	CARDioGRAM		22233	64762	86995	2420361	2011	Disease	Cardiovascular	
9	Coronary heart disease	Deloukas	CARDioGRAMplusC4D		63746	130681	194427	79129	2013	Disease	Cardiovascular	
10	Crohn's disease	Jostins L	IIBDGC		14763	15977	30740	13898	2012	Disease	Autoimmune / inflammatory	
11	Crohn's disease	Liu	IIBDGC		1690	3719	5409	154590	2015	Disease	Autoimmune / inflammatory	
12	Crohn's disease	Liu	IIBDGC		17897	33977	51874	124888	2015	Disease	Autoimmune / inflammatory	
13	Crohn's disease	Liu	IIBDGC		184	990	1174	154590	2015	Disease	Autoimmune / inflammatory	

Quick SNP lookup

Safari File Edit View History Bookmarks Window Help

mrbase.org

Huis Apple ID iCloud Reference WaldamarGroup Family Apple Cavadis deCODE Google Network News Shop Omica Science SNP MR

www.mrbase.org/beta/?state=RQcDoizmpMV8y74RSr7Q&code=4/IHbUXwfJixrpo1P8q3kpawQ3KJfI9_vFhvIMGdhw#

MR BASE

Welcome to MR Base

About Acknowledgements Data access agreement

Logged in as Sander W. van der Laan s.w.vanderlaan@gmail.com

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?

- Use proxies?

Minimum LD Rsq value

0.6 0.64 0.68 0.72 0.76 0.8 0.84 0.88 0.92 0.96 1

Allow palindromic SNPs?

- Allow palindromic SNPs?

MAF threshold for aligning palindromes

0.01 0.06 0.11 0.16 0.21 0.26 0.3 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:

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

Submit

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

⚡ Perform MR analysis

Safari File Edit View History Bookmarks Window Help

mrbase.org

www.mrbase.org/beta/?state=RQcDoizmpMV8y74RSr7Q&code=4/IHbUXwfJixrpo1P8q3kpawQ3KJfI9_vFhvIMGdhw#

Do we need to LD prune our SNP selection...?

Welcome to MR Base

How do we handle missing SNPs...?

Logged in as Sander W. van der Laan

What do we do with A/T and C/G SNPs...?

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?

Use proxies?

Minimum LD Rsq value

0.6 0.64 0.68 0.72 0.76 0.8 0.84 0.88 0.92 0.96 1

Allow palindromic SNPs?

MAF threshold for aligning palindromes

0.01 0.06 0.11 0.16 0.21 0.26 0.3 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:

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

Select a method for MR...

Perform MR analysis



Cystatin C and cardiovascular disease: a Mendelian randomization study

WHY WERE WE INTERESTED?

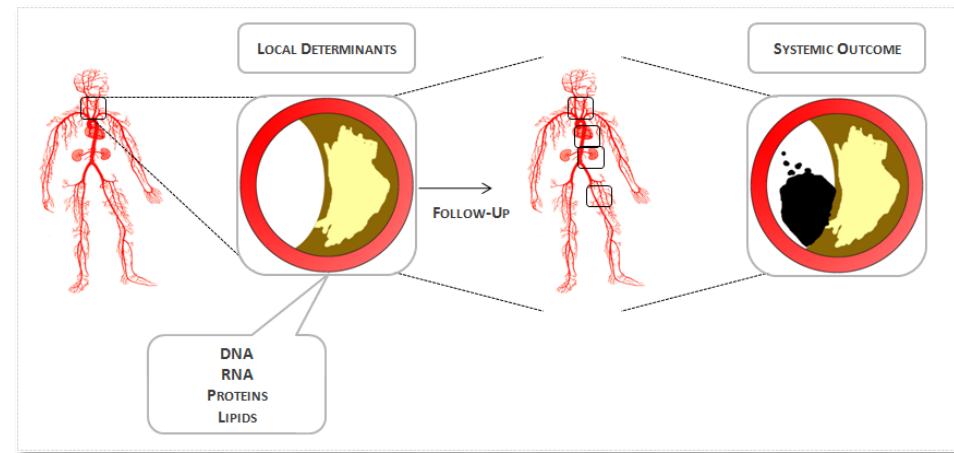
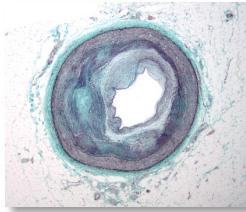


UMC Utrecht



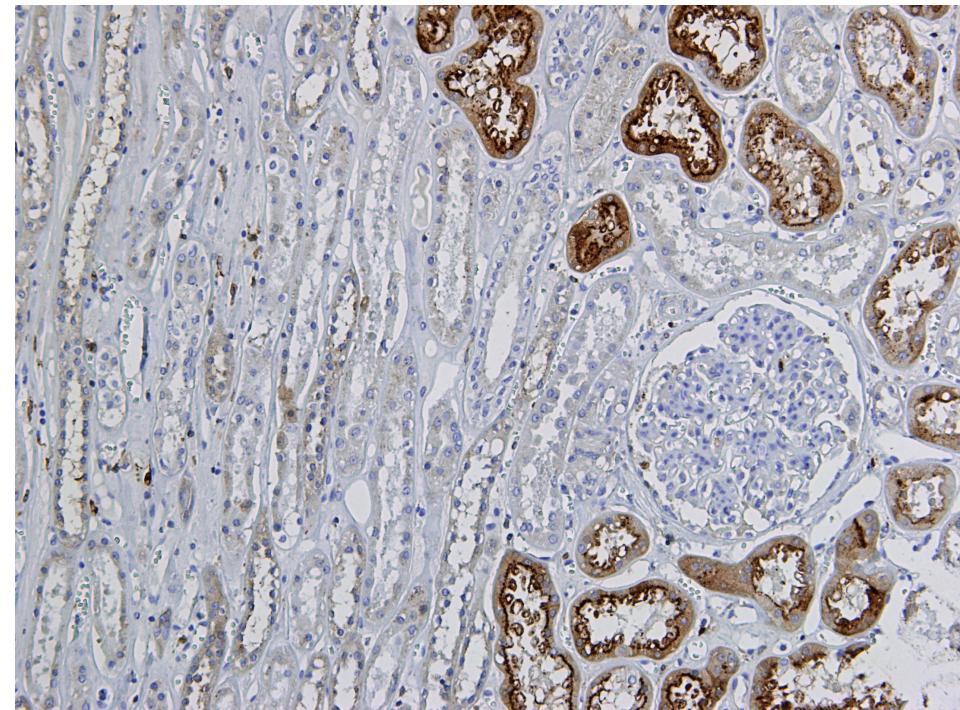
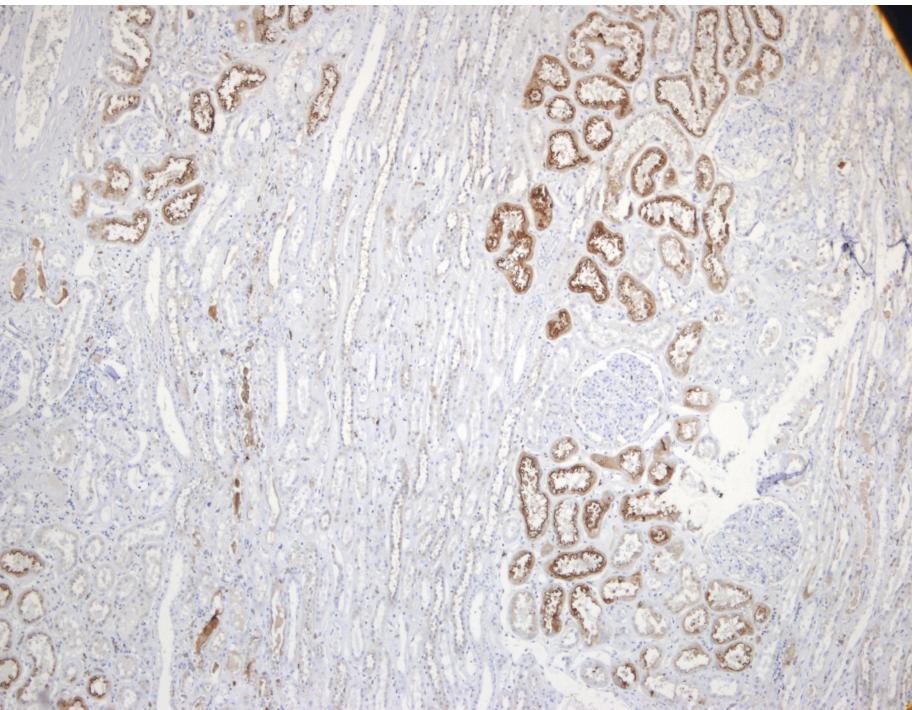
This was the start of our work...

- CystC expressed in smooth muscle cells inhibiting cathepsins
 - CystC is down regulated in AAA samples
 - ApoE^{-/-}–CystC^{-/-} mice have more atherosclerosis
-
- Is there an association of CystC associated variant with CystC and SMCs in the carotid plaques from the Athero-Express?



Histological staining for CystC

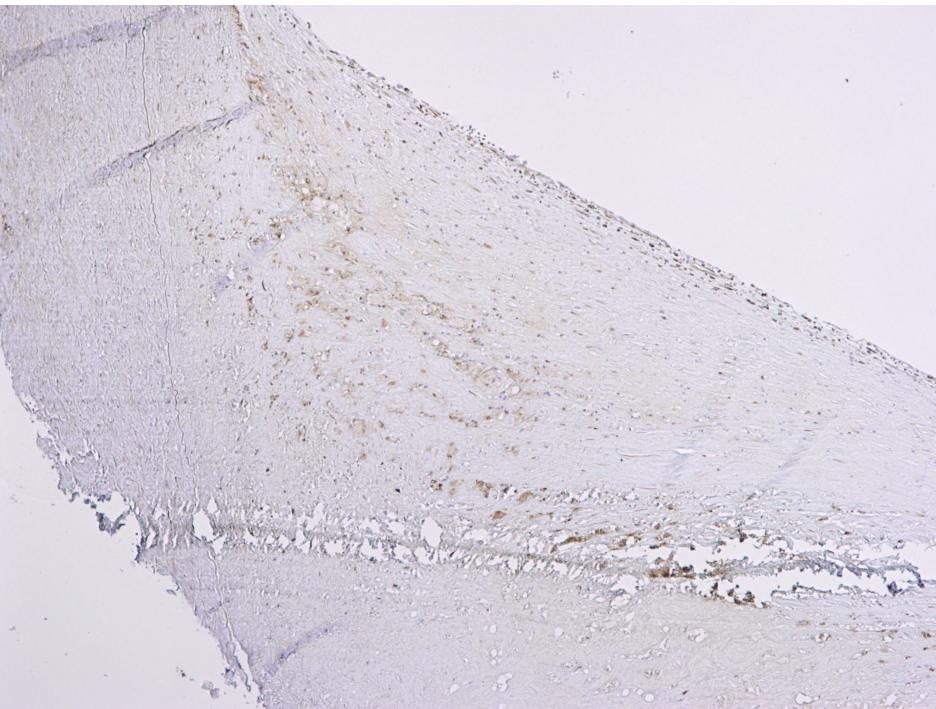
- First a positive control: the human kidney
- CystC present in proximal (?) tubuli
- Appears to be present in granules



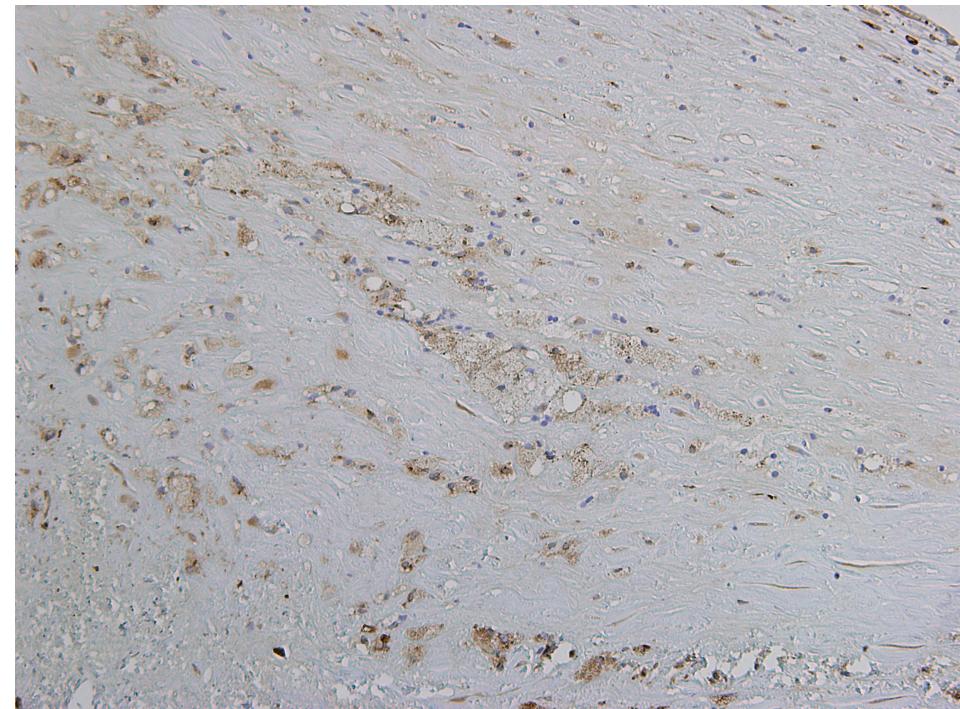
Van der Laan *et al.* unpublished

Carotid endarterectomy sample

- Shoulder region, clear staining for CystC
- Appears to be stronger in foamy cells, (foamy) smooth muscle cells

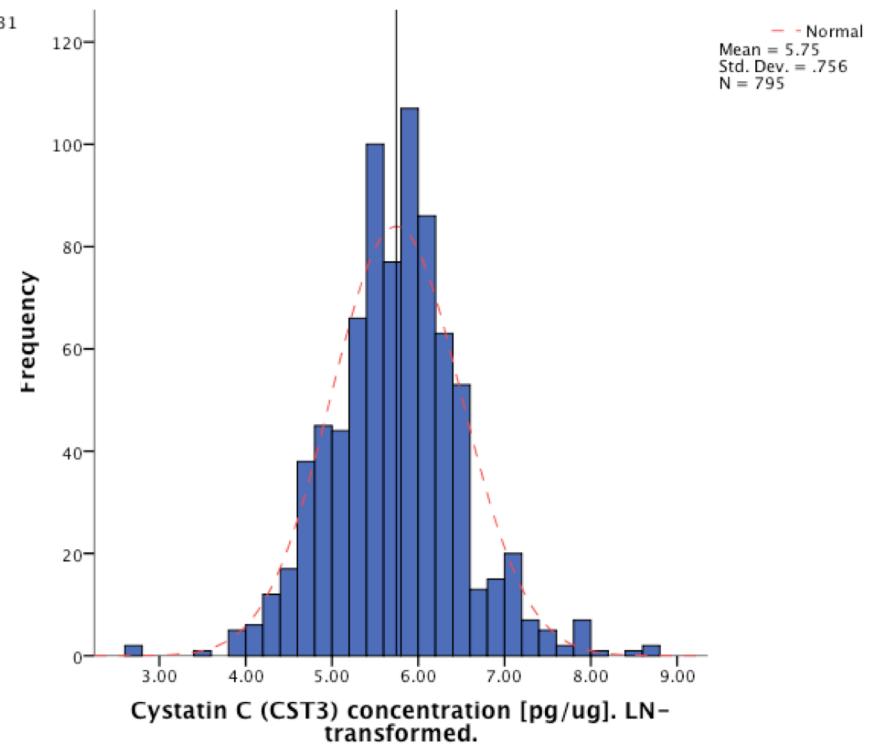
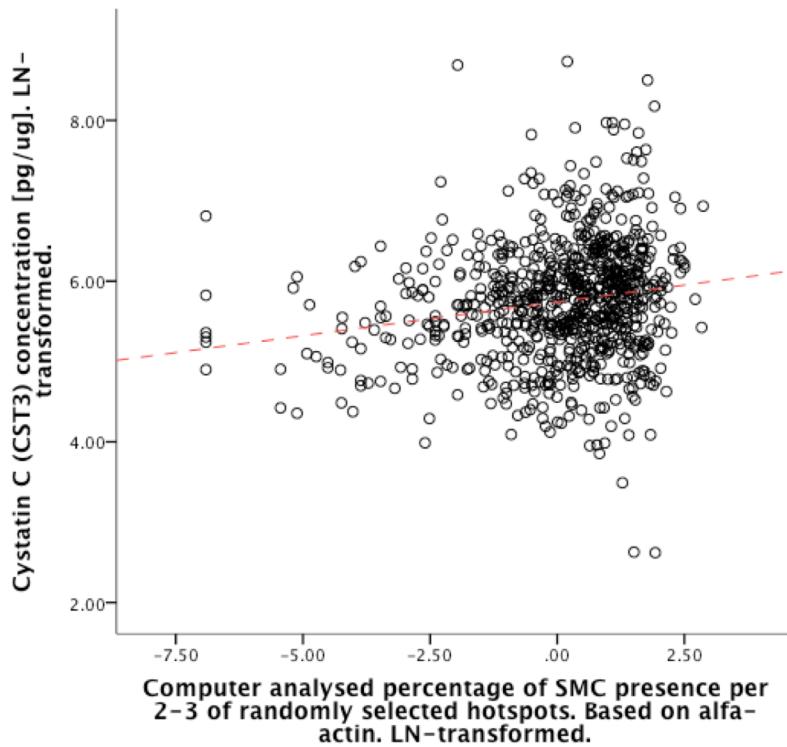
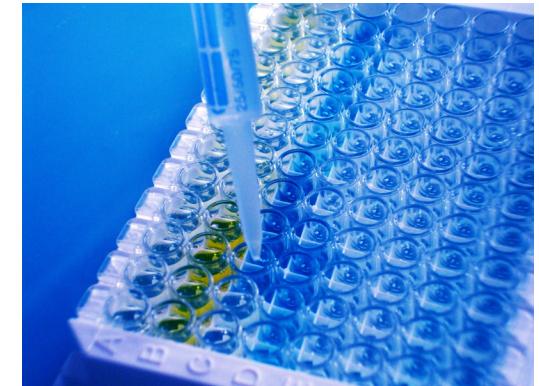


Van der Laan *et al.* unpublished



Luminex CystC plaque levels

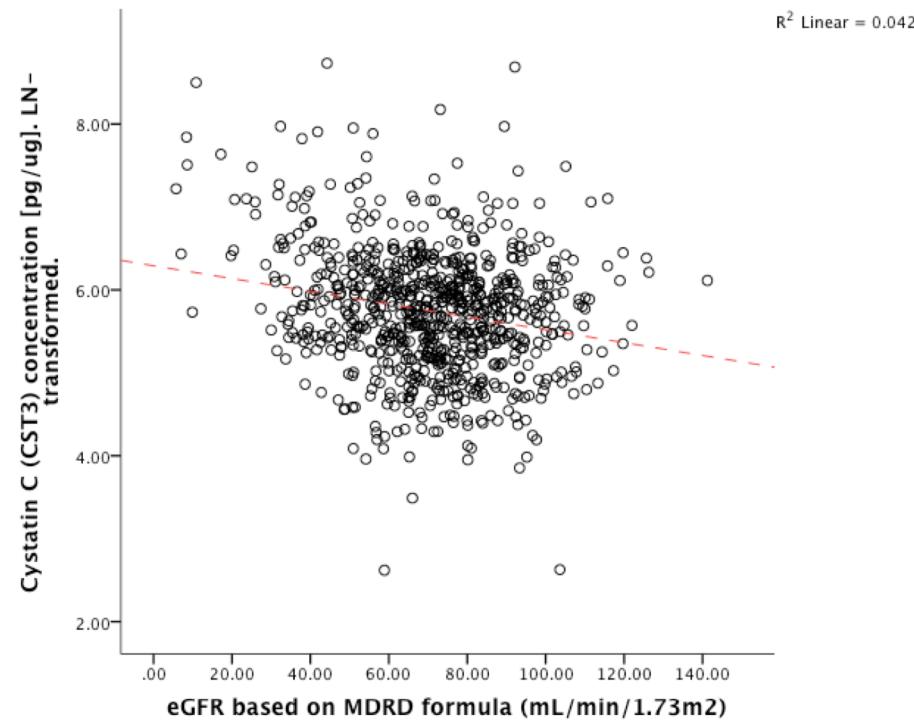
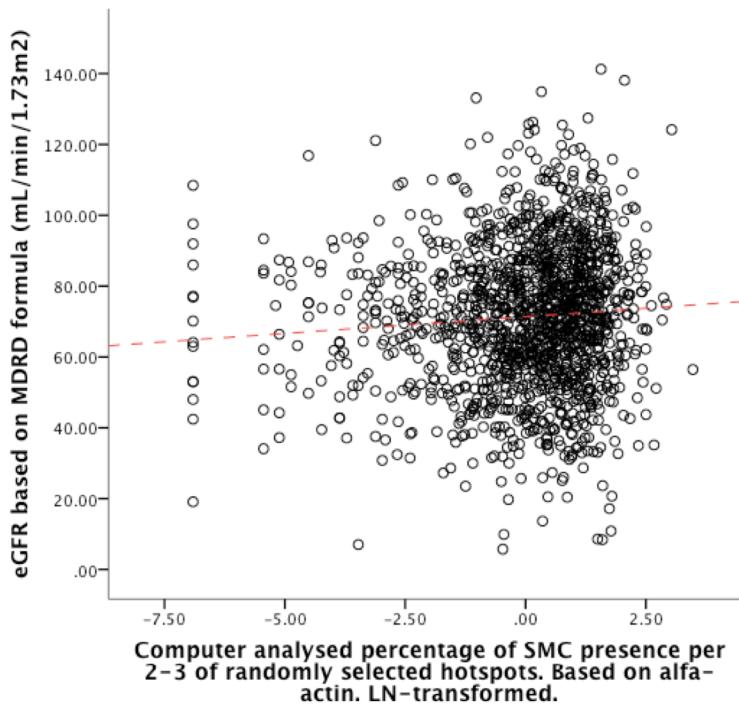
- N = 795
- Rho = 0.185 ($R^2 = 0.176$), p = 1.10×10^{-7}



Van der Laan *et al.* unpublished

Associations with eGFR

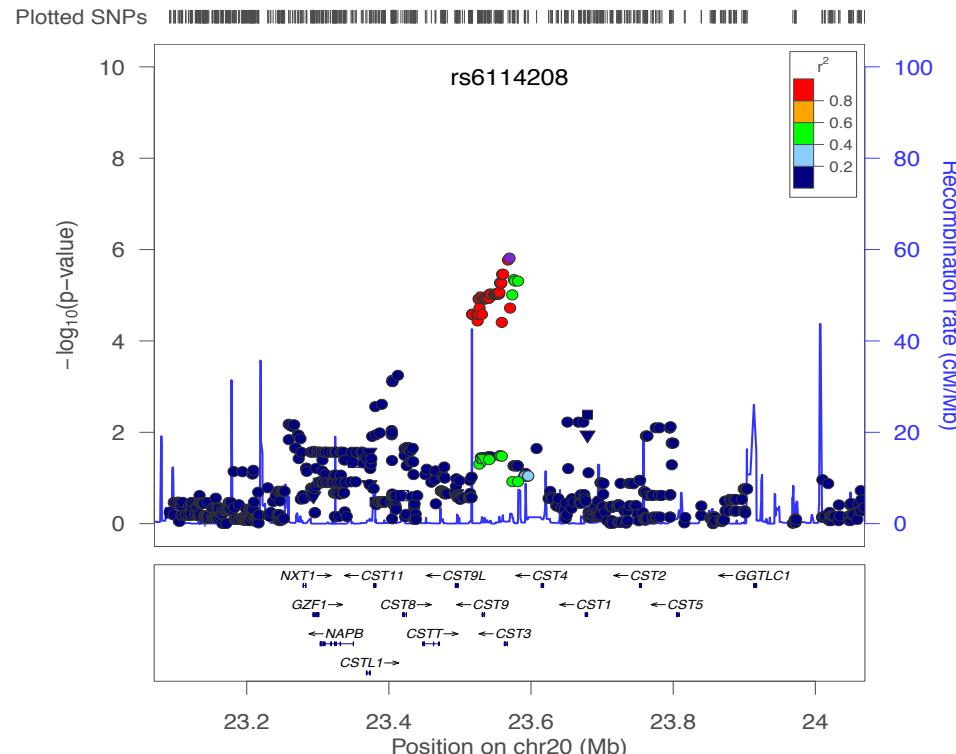
- SMC vs. eGFR: Rho = 0.084 ($R^2 = 0.076$), $p = 1.35 \times 10^{-3}$
- eGFR vs. plaque CST3: Rho = -0.161 ($R^2 = -0.206$), $p = 4.92 \times 10^{-3}$



Van der Laan *et al.* unpublished

CystC-variant associated with SMC in plaques

- In Athero-Express Genomics Study 1 ($N=571$) rs6114208 negatively correlated with SMCs in plaques, $p=1.53\times 10^{-6}$
- When analyzed with additional samples (Athero-Express Genomics Study 2, $N=868$), $p = 0.0083$, $\beta = -0.161$ (0.061), about 10.6% decrease in SMCs



Van der Laan *et al.* unpublished





UMC Utrecht

Fulfilling assumptions

STUDIES & CONFOUNDING

