Genetic Epidemiology at the intersection between function and disease

Florian Kronenberg

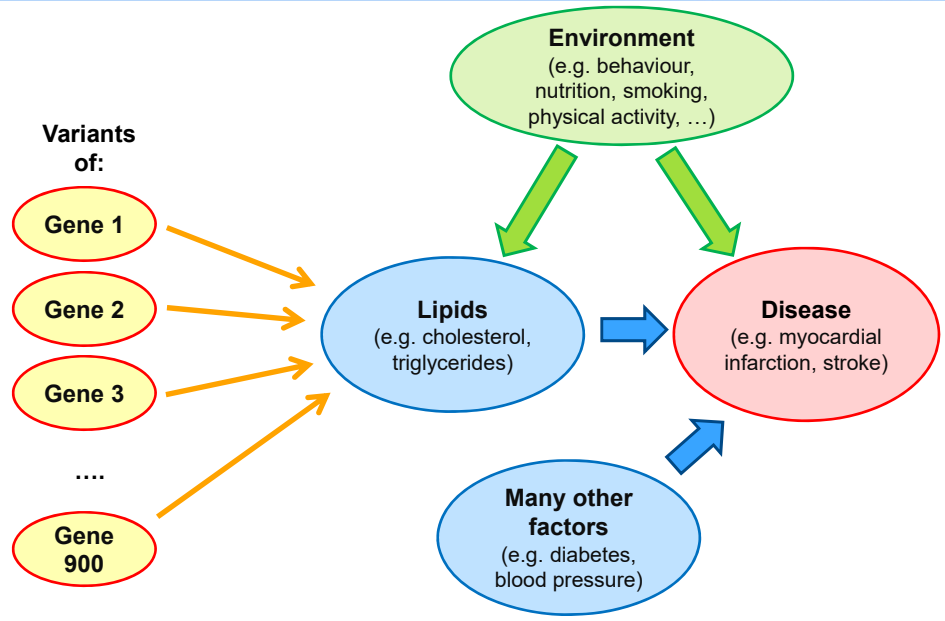
Institute of Genetic Epidemiology, Medical University of Innsbruck

Overview

1. **Background**
2. Association studies
3. Genomewide association studies (GWAS)

GENEPI INNSBRUCK

How is health and disease determined?



GENEPI INNSBRUCK

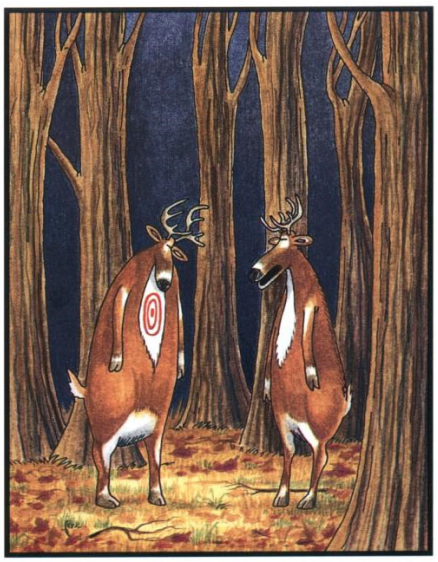
Why are we interested in "new" genes?

By Victor A. McKusick, M.D., Baltimore, Maryland

Ann. Int. Med. 49:556-567, 1958

Study of genetic factors is important:

(1) because potentially it will permit **recognition of genetic susceptibles**, for more effective application of preventive measures,



"Bummer of a birthmark, Hal."

GENEPI INNSBRUCK

Why are we interested in "new" genes?

By Victor A. McKusick, M.D., Baltimore, Maryland

Ann. Int. Med. 49:556-567, 1958

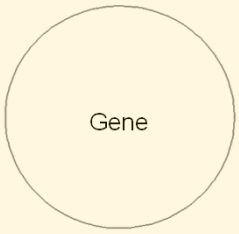
Study of genetic factors is important:

(2) because from our **understanding of the mechanism** whereby the gene or genes operate in these disorders can come preventive or therapeutic measures for breaking the chain leading to disease.

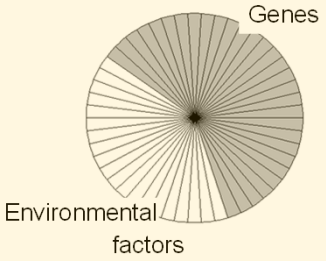
"This could be the discovery of the century. Depending, of course, on how far down it goes."

GENEPI INNSBRUCK

Monogenic and complex diseases



Monogenic diseases:
e.g. Morbus Huntington



Complex diseases:

e.g. Diabetes, myocardial infaction, overweight, cancer, …

Environmental factors are e.g. smoking, physical activity, nutrition, education, sun exposition, ….

GENEPI INNSBRUCK

Knockout versus small changes by polymorphisms

Knock-out

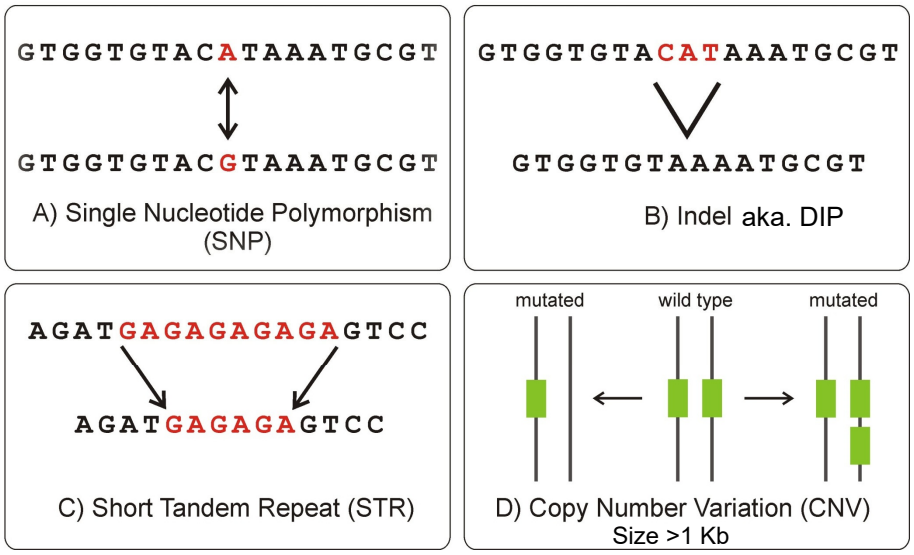
- Pronounced effects
- Animals: great models but not necessarily to extrapolate to humans
- Humans: often very rare cases

Polymorphism

- Small effects
- Usually investigated in humans
- Real in vivo conditions
- Thousands of people can be studied easily
- Sample sizes of thousands are required

GENEPI INNSBRUCK

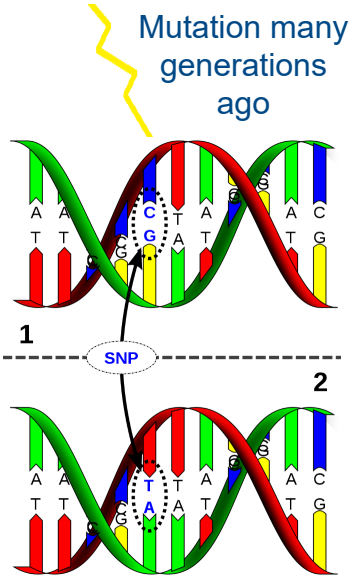
Genetic variability



GENEPI INNSBRUCK

Single Nucleotide Polymorphisms (SNPs)

- Variations of single base pairs (bp) in the DNA sequence
- Heritable and stable.
- Account for 90% of the genetic variability
- Every 300 – 1000 bp
- At least 3 – 4 million SNPs per individual
- 10,000 – 11,000 non-synonymous SNPs per individual
- 700 million SNPs are described in databases



C/T polymorphism

GENEPI INNSBRUCK

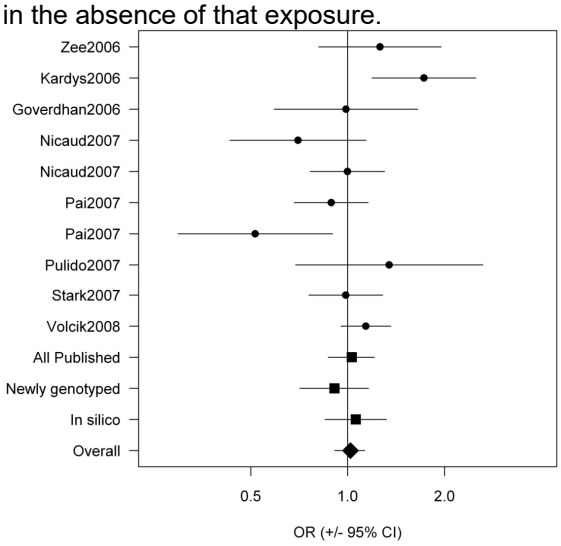
Single Nucleotide Polymorphism (SNP)

- **Coding SNPs within a gene**
  - synonymous exchanges: without influence on protein
  - non-synonymous exchanges: resulting in an AA exchange
- **SNPs within the regulatory regions:**
  - when and why a gene will be switched on or off
  - effect on quantity of protein production
- **SNPs within the untranslated regions**
  - with influence on mRNA stability
- **SNPs in intergenic regions**
  - functional consequences have to be evaluated

GENEPI INNSBRUCK

Some basics from epidemiology

- **Odds ratio**
  - Represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure.
  - Values between 0 and infinite (∞)
  - 1.00 = same odds
  - 1.50 = 50% higher odds
  - 2.00 = 100% higher odds
  - 0.50 = 50% lower odds
- **95% confidence interval (CI)**
- **Hazard ratio**
  - In case of prospective studies
- **Meta-analysis**
  - Combining data from more than one study



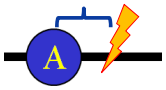
GENEPI INNSBRUCK

Overview

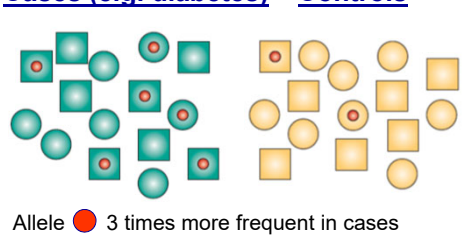
1. Background
2. **Association studies**
3. Genomewide association studies (GWAS)

GENEPI INNSBRUCK

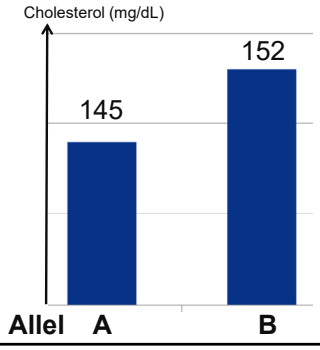
Principle of association studies



- **Qualitative analysis:**
  Preferential association of an allele with a disease status



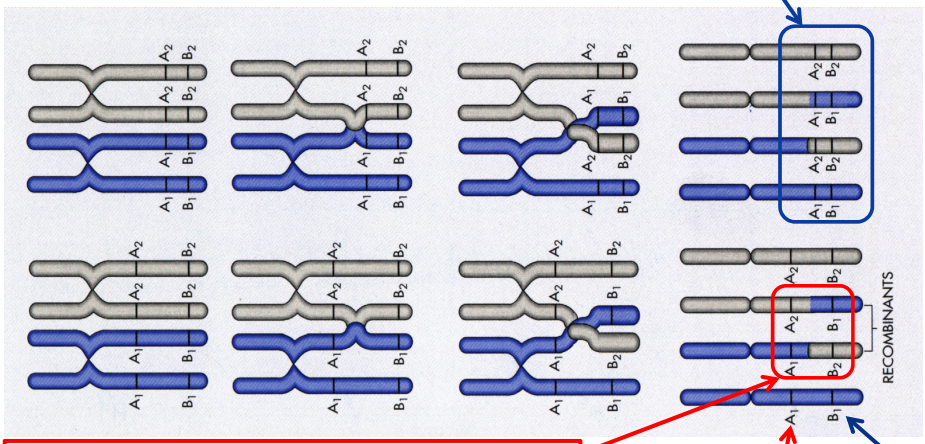
- **Quantitative analysis:**
  Carriers of a various alleles differ in the mean values of the investigated parameter (e.g. cholesterol level)



GENEPI INNSBRUCK

Remember: crossover and recombination during meiosis I

No recombination of alleles of the gene loci A and B since crossover outside the region between A and B



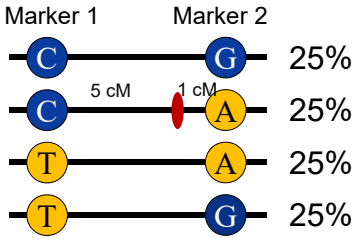
Recombination of alleles in the two gametes due to crossover

"Wanted" disease locus

Genotyped marker locus

Jorde, Carey, Bamshad, White: Medical Genetics

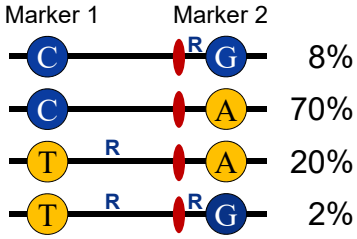
Basis of association: Linkage disequilibrium



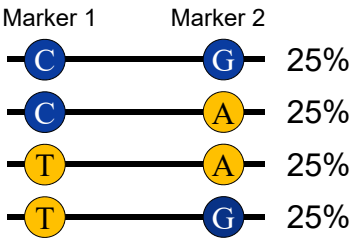
	Patients	Controls
Allele 2A	**90%**	**50%**
Allele 1C	78%	50%

After many generations

Patients



Controls



GENEPI INNSBRUCK

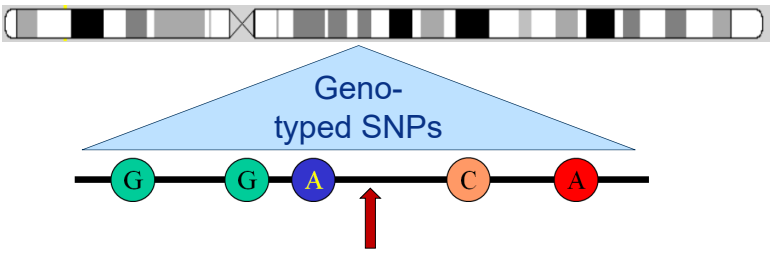
R … Recombination

What does a significant genetic association mean?

- **Direct association**
  - The investigated genetic variant is indeed the causal disease-causing variant
  - This was rarely the case in earlier times; improves nowadays by the dense map of markers we can investigate
  - Optimum pocedure: functional characterisation goes hand in hand
- **Indirect association**
  - The investigated genetic variant is in linkage disequillibrium with the causal variant

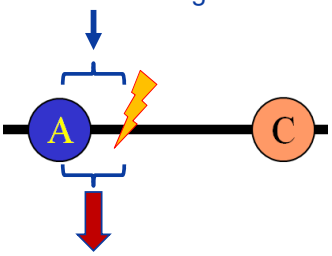
GENEPI INNSBRUCK

Indirect association



Disease-causing variant we are searching

Due to the small distance there are rarely crossovers and recombinations during meiosis



Allele A can be observed more frequently in the diseased group

ACTAGAGCTACTACGAGGGACTAC...TACGAGCATCGACTA...GAGG
TAGAGCTATA...TTCT AGGCTA...CTACGATCGATC...ACGTAG...

High correlation

The genotyped SNP is a marker of the „non-genotyped“ disease-causing variant we are searching

GENEPI INNSBRUCK

What does a significant genetic association mean?

- **Direct association**
  - The investigated genetic variant is indeed the causal disease-causing variant
  - This was rarely the case in earlier times; improves nowadays by the dense map of markers we can investigate
  - Optimum pocedure: functional characterisation goes hand in hand
- **Indirect association**
  - The investigated genetic variant is in kinkage disequillibrium with the causal variant
- **False-positive finding (spurious association)**
  - Random finding (sample size!)
  - Confounding: e.g. population stratification
  - Often observed in small studies without replication

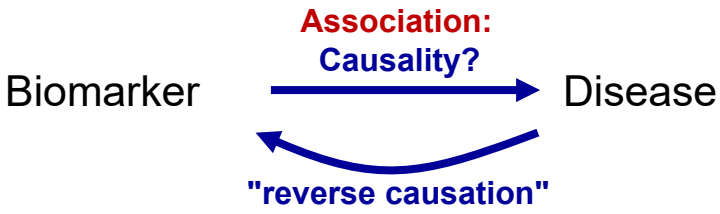
→ **replicate, replicate, replicate!**

GENEPI INNSBRUCK

Mendelian Randomization study

GENEPI INNSBRUCK

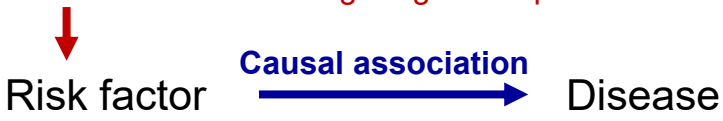
Biomarker for diseases: causality or consequence?



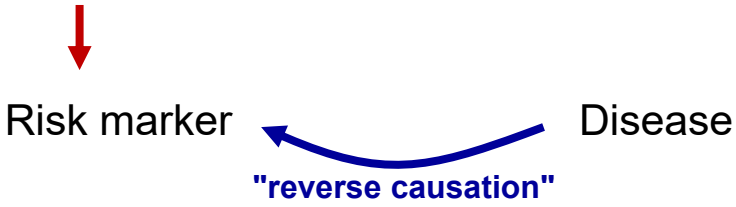
GENEPI INNSBRUCK

Biomarker for diseases: causality or consequence?

- Intervention including drug development



- No direct reason for intervention or drug development
- Useful for diagnostic purposes?



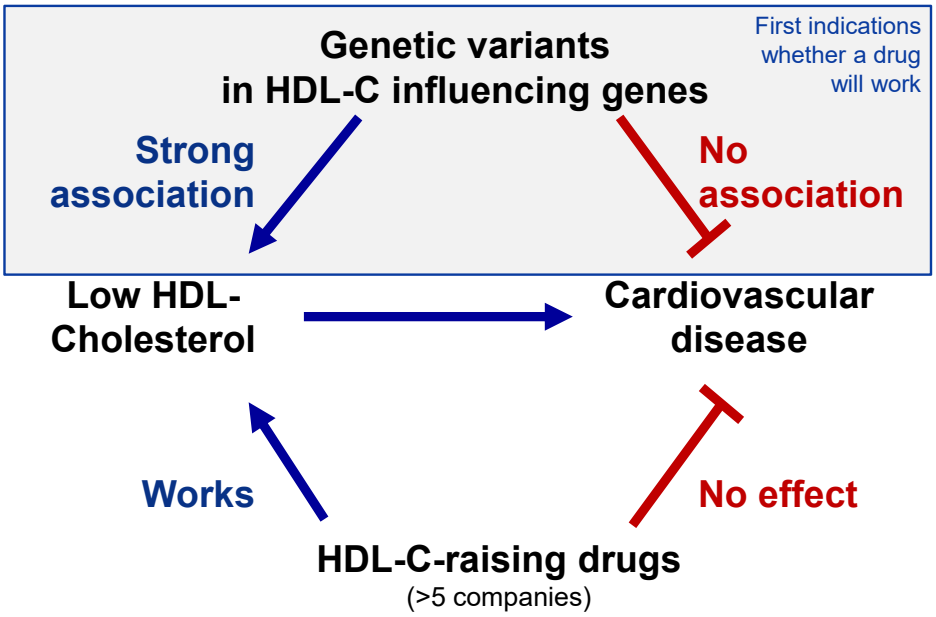
GEWEPI INNSBRUCK

The big question for biomarkers

- You found an association with a disease
- **Risk factor or risk marker?**
- Classical epidemiological studies with prospective observation will last a long time and will not prove causality
- You have to decide now whether to go for drug development or not
- Worst case scenario: after 10-15 years of development the drug flops
- One reason might be that it is only a risk marker and not a risk factor.

GEWEPI INNSBRUCK

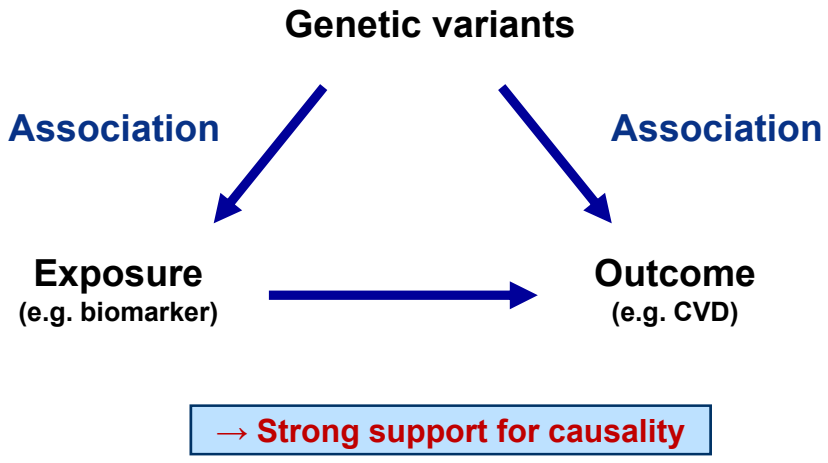
Example of a flop: CETP inhibitors



GENEPI INNSBRUCK

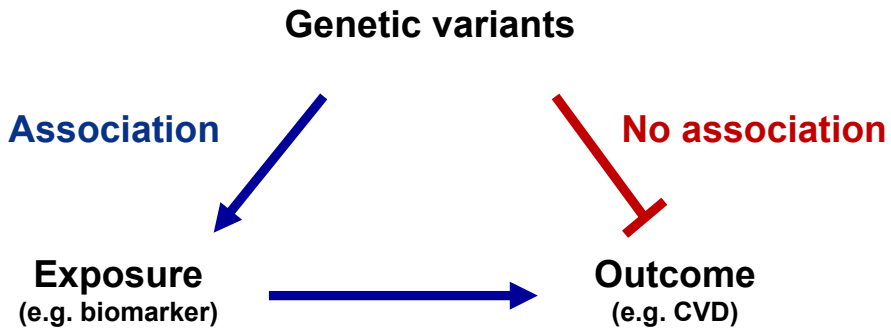
Kronenberg: Kidney Int. 89:747-9, 2016

Mendelian randomization approach



GENEPI INNSBRUCK

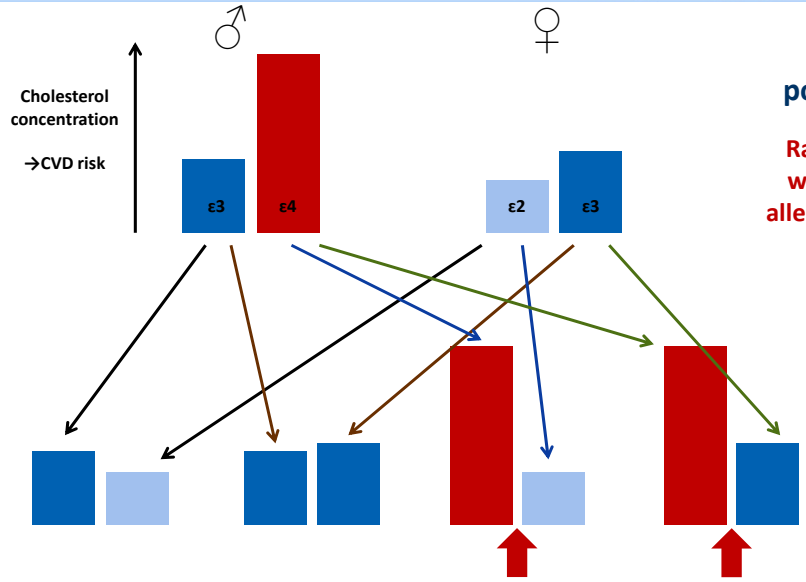
Mendelian randomization approach



→ Causality is unlikely if the study is sufficiently powered

GENEPI INNSBRUCK

Mendelian randomization at the time of conception



ApoE polymorphism

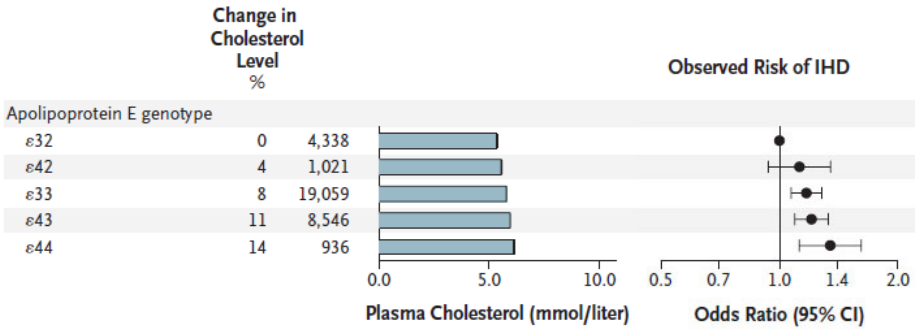
Random decision which of the two alleles is transmitted to the child

Increased risk for CVD due to lifelong exposure to higher cholesterol concentrations

GENEPI INNSBRUCK

Adapted from Cardiovasc.Drugs Ther. 30:87-100, 2016

ApoE, cholesterol and risk for CVD



Zacho et al.: NEJM 359:1897-1908, 2008

Example: Lipoprotein(a)

Lp(a)

Experience of a young widow

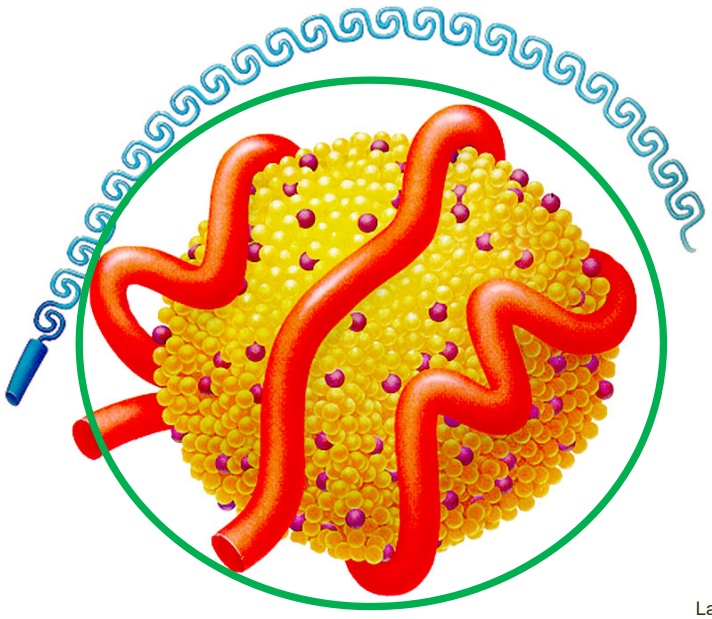
- Husband 39 years of age
- Loses consciousness, cardiac arrest, revival not successful

- No classical risk factors
- Healthy lifestyle, physically active
- Health checkup on a yearly basis

- Autopsy: most severe heart disease
- **Very high Lp(a) concentrations**

GENEPI INNSBRUCK

Lp(a) – the mysterious brother of LDL



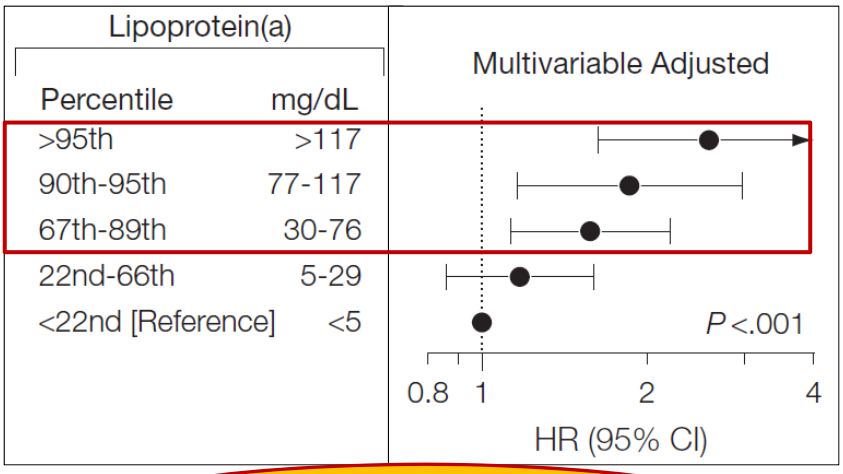
GENEPI INNSBRUCK

Lawn; Sci.Am. 266:54-60, 1992

Lp(a) and risk for myocardial infarction

Results from the Copenhagen City Heart Study

A third of the entire population!



Ist this association causal?

GENEPI INNSBRUCK

Kamstrup et al.: JAMA 301:2331-9, 2009

Lp(a) and CHD: truth or consequence?

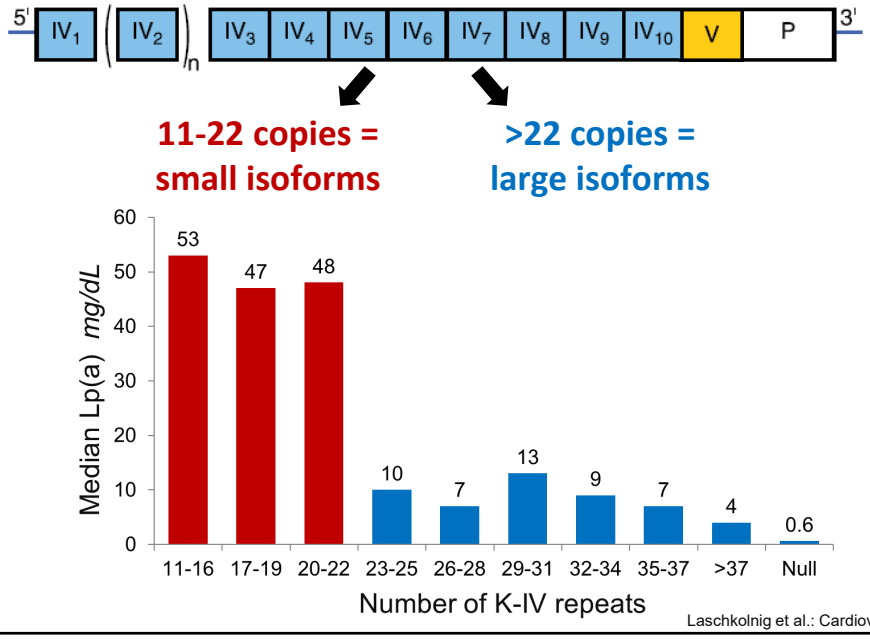
Genetic instrument
(=genetic variants)?

High Lp(a) → **causal association?** → CHD

"reverse causation"?

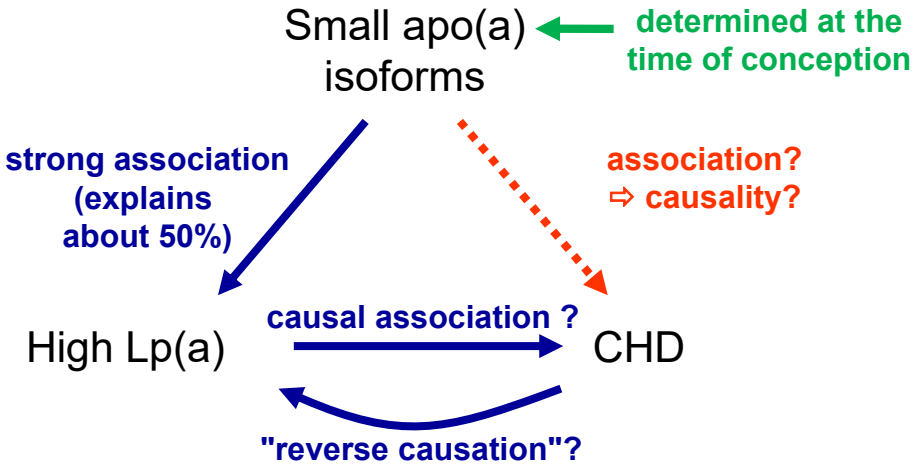
GENEPI INNSBRUCK

Apolipoprotein(a) - Mr 300-800 kDa



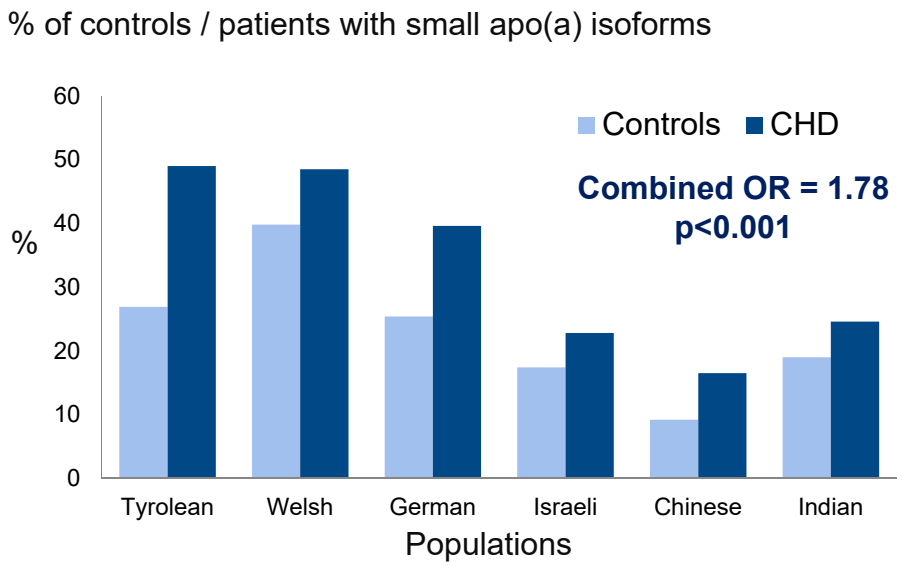
Laschkolnig et al.: Cardiovasc Research 103: 28-36, 2014

Lp(a) and CHD: Mendelian randomization



Do carriers of small apo(a) isoforms more often have CHD?

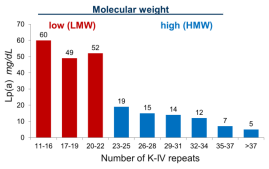
Apo(a) isoforms and risk for CHD



GENEPI INNSBRUCK

Sandholzer et al.: Arterioscler Thromb 12: 1214-26, 1992

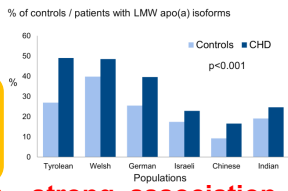
Lp(a) concentrations, apo(a) isoforms and CVD



strong association (explains about 50%)

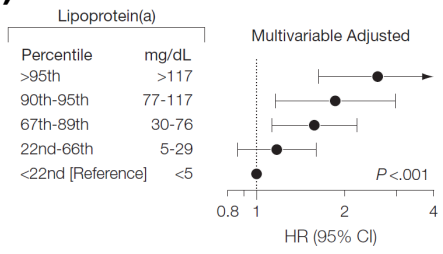
Small apo(a) isoforms

25-35% of the population



strong association

High Lp(a) — causal association → **CVD**



GENEPI INNSBRUCK

Cardiovasc.Drugs Ther.: 30:87-100, 2016

ESC European Society of Cardiology
European Heart Journal (2022) 43, 3925–3946
https://doi.org/10.1093/eurheartj/ehac361

SPECIAL ARTICLE
Miscellaneous

Lipoprotein(a) in atherosclerotic cardiovascular disease and aortic stenosis: a European Atherosclerosis Society consensus statement

Genetic variants associated with increased Lp(a) concentrations …

… are associated with increased cardiovascular risk

LPA genetic score Number of Lp(a)-increasing variants *	No. participants	No. events	HR (95%CI)
0 (median Lp(a) 14 nmol/L)	358,464	20,610	1.00 [reference]
1 (median Lp(a) 146 nmol/L)	77,655	6,167	1.47 [1.43, 1.52]
2 (median Lp(a) 262 nmol/L)	4,249	410	1.89 [1.70, 2.10]

*** variants of rs10455872 and rs3798220 described by R. Clarke**

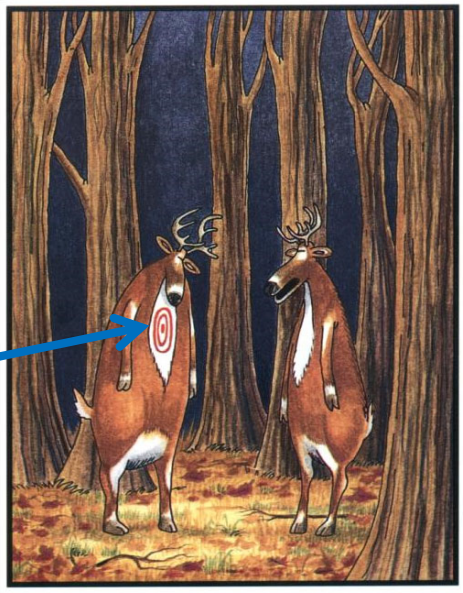
EAS

Data are provided by Ference, Catapano et al. using data from the UK Biobank
Kronenberg, Mora, Stroes et al.: Eur. Heart J. 43:3925-46, 2022

Lp(a) concentrations and apo(a) isoform as a risk factor

Having a small apo(a) isoform doubles the odds for CVD in 25-35% of the population

→ strongest genetically determined risk factor for CVD

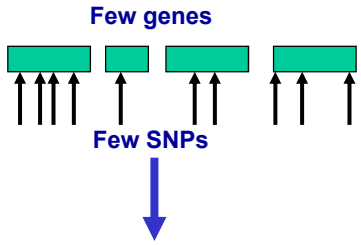


"Bummer of a birthmark, Hal."

GENEPI INNSBRUCK

Candidate gene approach vs. GWAS

Candidate gene approach



- Hypothesis-driven
- Biochemical or physiological *a priori* knowledge
- Few genes identified

GENEPI INNSBRUCK

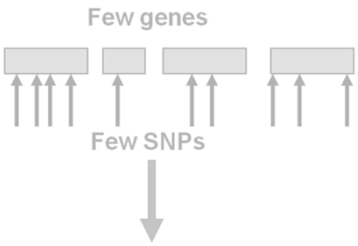
Overview

1. Background
2. Association studies
3. **Genomewide association studies (GWAS)**

GENEPI INNSBRUCK

Candidate gene approach vs. GWAS

Candidate gene approach



Association with phenotype

- Hypothesis-driven
- Biochemical or physiological *a priori* knowledge
- Few genes identified

Genomwide association study GWAS

23 chromosomes

0.5 – 10 million SNPs

Association with phenotype

- **"Hypothesis-free"** (unbiased)
- No *a priori* knowledge
- New pathways
- Small effects detectable
- **Very large sample sizes required**

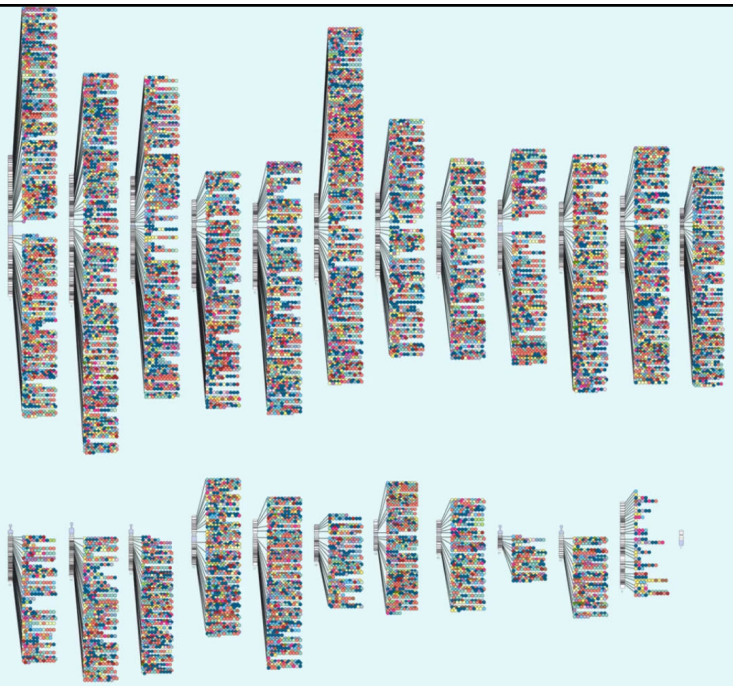
GENEPI INNSBRUCK

2018 Apr

Associations: 69,885

Studies: 5,152

Papers: 3,378



www.ebi.ac.uk/gwas

Gain in detected genes by GWAS

Examples for metabolic traits

Disease	before 2007	2007 onward
Type 2 DM	3	50
Body mass index	1	30
Glucose or insulin	1	15
Fat distribution	0	20
Lipids	16	95
Total	**21**	**202**

7 examples of autoimmune diseases

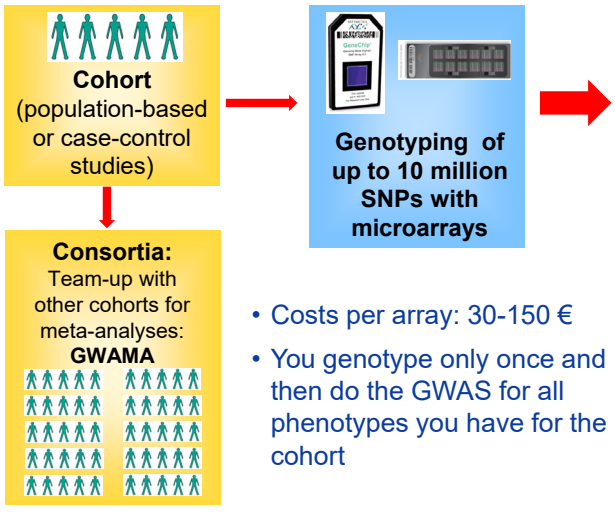
Disease	before 2007	2007 onward
Ankylosis spondylitis	1	13
Rheumatoid arthritis	3	30
Systemic lupus eryth.	3	31
Type 1 DM	4	40
Multiple sclerosis	1	51
Crohn's disease	4	67
Ulcerative colitis	3	44
Total	**19**	**277**

Since 2012 the number of known genes has further increased by 5- to 10-fold

GENEPI INNSBRUCK

Visscher et al.: Am.J.Hum.Genet. 90:7-24, 2012 (updated)

Design and cost-performance ratio



- Costs per array: 30-150 €
- You genotype only once and then do the GWAS for all phenotypes you have for the cohort

Association with phenotypes (e.g.)

- BMI
- Waist
- Blood pressure
- QT interval
- Smoking

Lab values

- Lipids
- Kidney function
- CRP
- Hemoglobin
- …

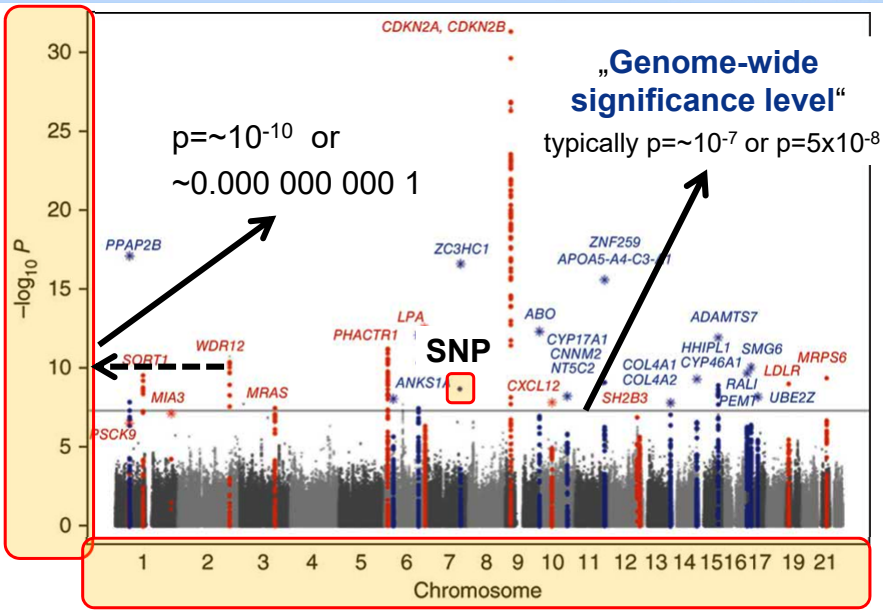
Diseases

- CAD
- Stroke
- Ankle-brachial-index
- Cancer types

Whatever has a genetic component and is measured

GENEPI INNSBRUCK

The Manhattan Plot



GENEPI INNSBRUCK

Schunkert et al.: Nature Genetics 43:333-338, 2011

Genome-wide association studies (GWAS)

- **Study design**
- **Examples:**
  - Lipids
  - Type 2 diabetes mellitus
  - Blood pressure
  - Kidney function
  - Addiction (smoking quantity)
  - BMI

GENEPI INNSBRUCK

GWAS: Lipids

- **Consortium:**
  - Established during fall 2005
  - Together with Helmholtz-Zentrum München
  - 11 members studying various phenotypes
- **Innsbruck Group:**
  - Lipid metabolism
  - HDL-C as a starting point
  - Quantitative trait considered more powerful
- **Population and Genotyping:**
  - 1644 population-based subjects from KORA
  - Affymetrix 500K SNP chip


Iris Heid


Stefan Coassin

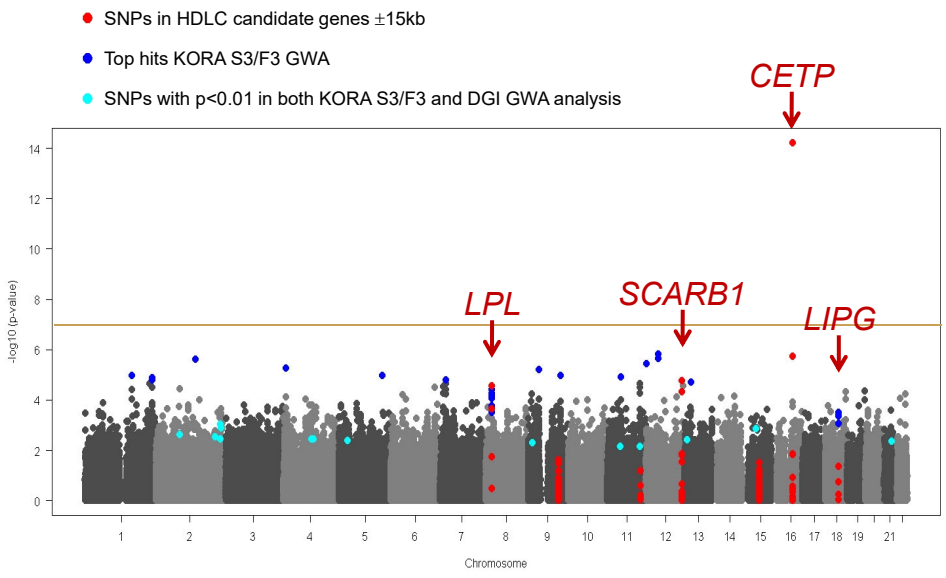

Eva Boes


Barbara Kollerits


Anita Brandstätter


Claudia Lamina

GWAS: HDL cholesterol



Heid et al.: Circ. Cardiovasc. Genet. 1:10-20, 2008

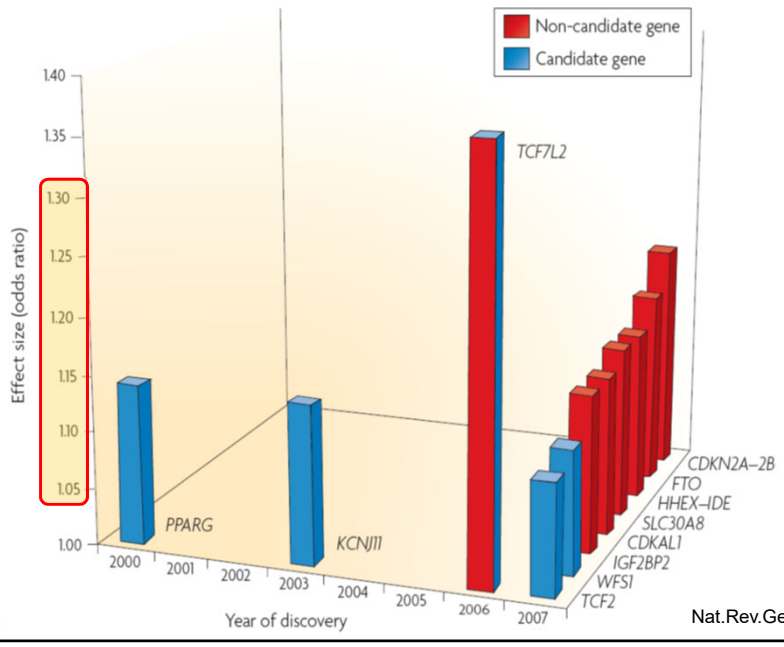
GWAS on Lipids (TC, HDLC, LDLC, TG): next steps

- **First own GWAS** *(Circ. Cardiovasc. Genet. 2008)*
  1644 probands: found nothing new
- **Engage Consortium:** *(Nature Genetics 2009)*
  - 22,000 probands: **22 genes found associated**
- **Global Lipids Genetics Consortium:** *(Nature 2010)*
  - >100,000 probands: **95 genes found associated**
- **Global Lipids Genetics Consortium:** *(Nature Genetics 2013)*
  - >188,000 probands: **roughly 155 genes found associated**
- **Global Lipids Genetics Consortium:** *(Nature 2021)*
  - 1,65 million probands: **>900 Gene**

Functional characterisation for most of the genes has to be done

GENEPI INNSBRUCK

GWAS: Type 2 diabetes mellitus



2017:

≈ 86 genes

2019:

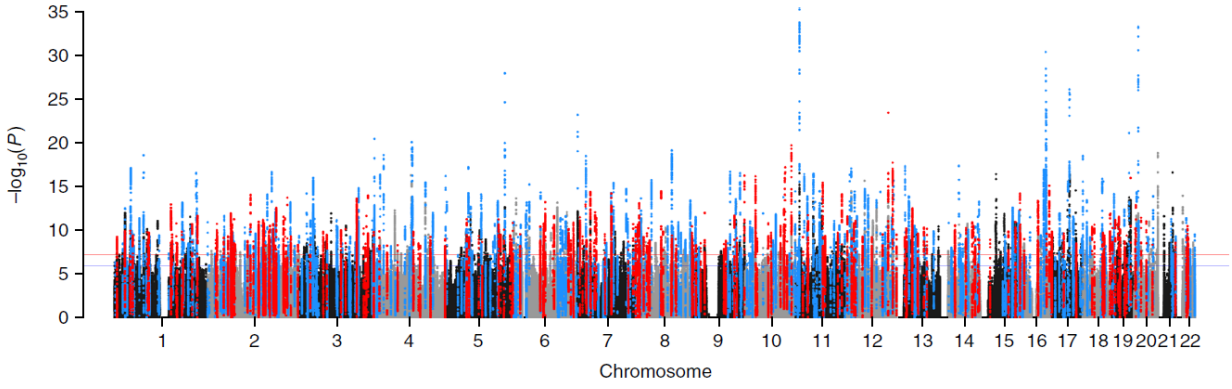
≈ 250 genes

T. Frayling:
Nat.Rev.Genet. 8:657-62, 2007

GENEPI INNSBRUCK

GWAS and blood pressure traits

- Phenotypes: systolic and diastolic BP, pulse pressure
- > 1 million study participants
- 901 genetic loci in total (535 novel)

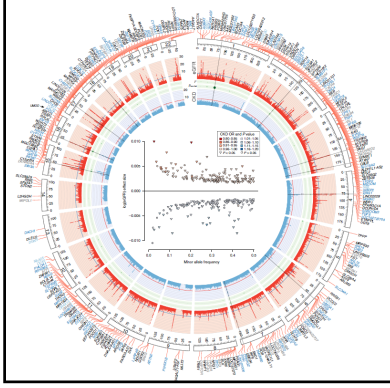


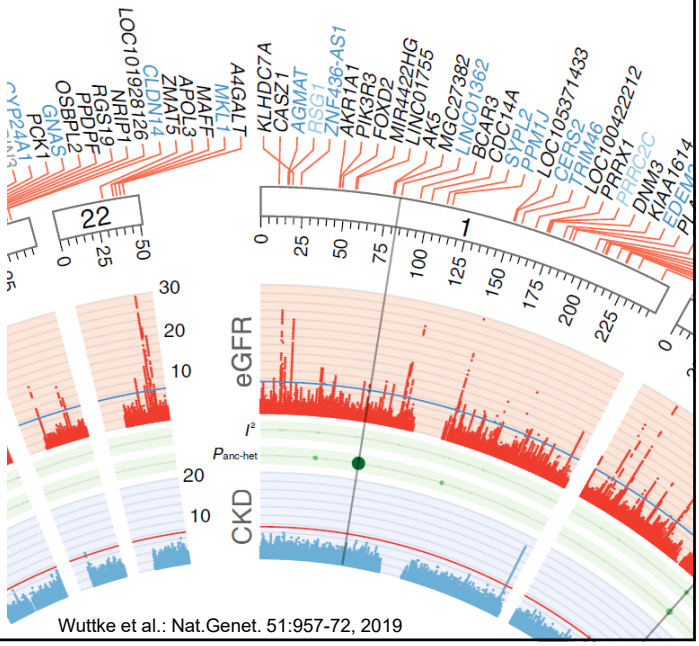
GENEPI INNSBRUCK

Evangelou et al.: Nature Genet. 50:1412-25, 2018

GWAS on kidney function

- **CKDGen consortium**
  - 1,046,070 persons
  - 264 associated loci (166 new)
  - Circos plot
  - comprehensive priority list of molecular targets for translational research

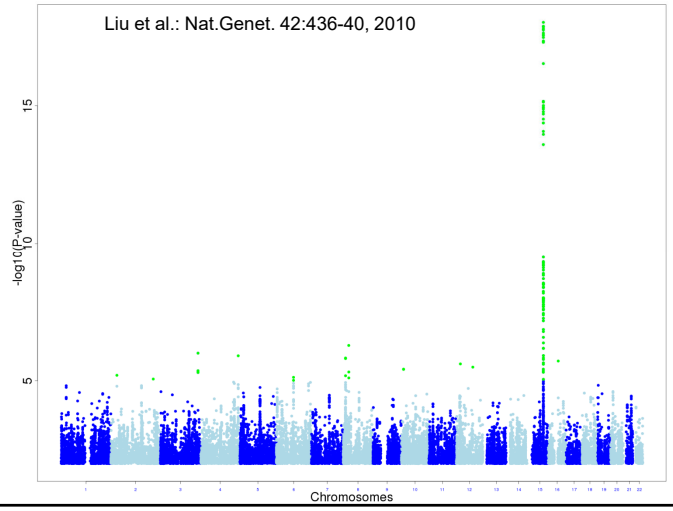




Wuttke et al.: Nat.Genet. 51:957-72, 2019

GWAS on smoking (quantity)

- Addiction research
- >40,000 persons
- Neuronal nicotinic acetylcholine receptor subunits
- Same region was found for lung cancer, COPD,lung function and PAD



GENEPI INNSBRUCK

Genes for body mass index and overweight

Own behavior (lifestyle)

Genetics

- 536 genetic loci detected
- Many of them play a role in the brain by
  - ⇨ **Regulation of appetite**
  - ⇨ Neuronal component of overweight
- Explain roughly 5% of BMI

GENEPI INNSBRUCK

Contribution of single genes to overweight

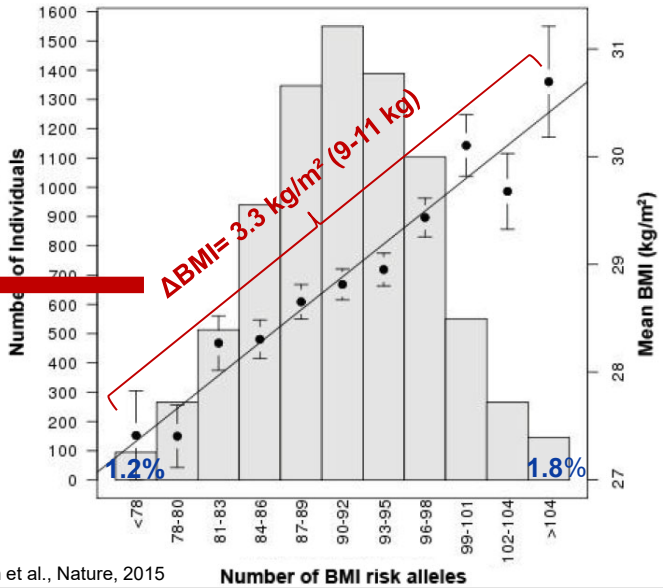
- **Very few with strong effects:** risk increase by 10 to 30% per allele
- **More with moderate effects:** risk increase by 3 to 10% per allele
- **Many more with tiny effects:** risk increase by 0.1 to 3% per allele

GENEPI INNSBRUCK

Development of SNP-Risk-Scores

GENEPI INNSBRUCK

Risk score of 97 SNPs for BMI: 0-194 risk variants



GENEPI INNSBRUCK

Shungin et al., Nature, 2015

Risk score for 65 SNPs for diabetes

Many risk variants (top quintile)

2.7-fold increased risk to develop a diabetes in the future

Few risk variants (bottom quintile)

GENEPI INNSBRUCK

Talmud et al., Diabetes 64:1830-40, 2015

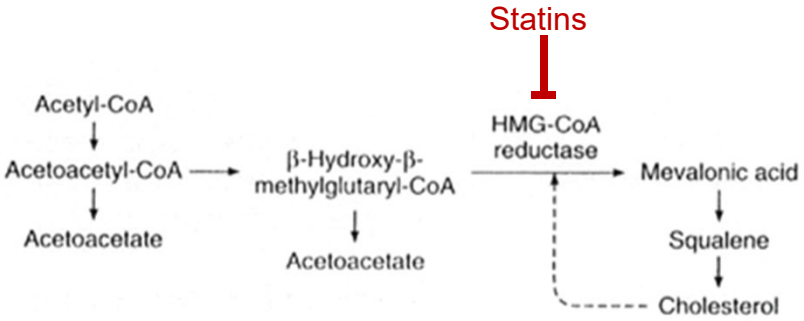
Where is the reward?

Can a single gene explaining less than 1% of the traits' variance still be useful for anything?

GENEPI INNSBRUCK

Statins: HMG-CoA-Reductase-Inhibitors

- **Mechanism of action**
  - Inhibition of HMG-CoA-Reductase: this enzyme catalyzes the conversion of HMG-CoA to mevalonic acid: an early and rate-limiting step in cholesterol biosynthesis.
  - Results in higher expression of LDL receptor which decreases LDL cholesterol



GENEPI INNSBRUCK

Polymorphisms in HMG-CoA-R gene region

- **GWAS results for HMG-CoA-reductase**
  - Very small effects
  - Were not detected in the first GWAS
  - This gene was only detected after investigation of at least 10.000 subjects
  - Single polymorphisms explain far less than 1% of the cholesterol concentrations within a population
  - Nevertheless, the most sucessful drug target for lipid metabolism
- **Other drug targets within the 157 lipid genes?**
  - CETP, ABCA1, PCSK9
  - Others?

GENEPI INNSBRUCK

Conclusions on GWAS

- An **hypothesis-free approach**
- Never before such a gain in gene-phenotypic information
- New genes for CAD, diabetes, cancer, kidney function…
- Odds ratios between 1.02 and 1.40
- To have the equipment is only the smallest step
- Very large studies of well phenotyped cohorts are necessary
- Works only within a very well constructed network between genetics, epidemiology, statistics, informatics, genomics
- Data sharing (a lot is already on the web)
- Non-coding SNPs and "gene deserts" can no longer be neglected
- A lot to learn about regulatory regions
- Functional characterization of "new" genes will need decades

GENEPI INNSBRUCK

Why are we searching these many genes?

Improvement of risk prediction (gene risk scores)

Identification of new drug targets

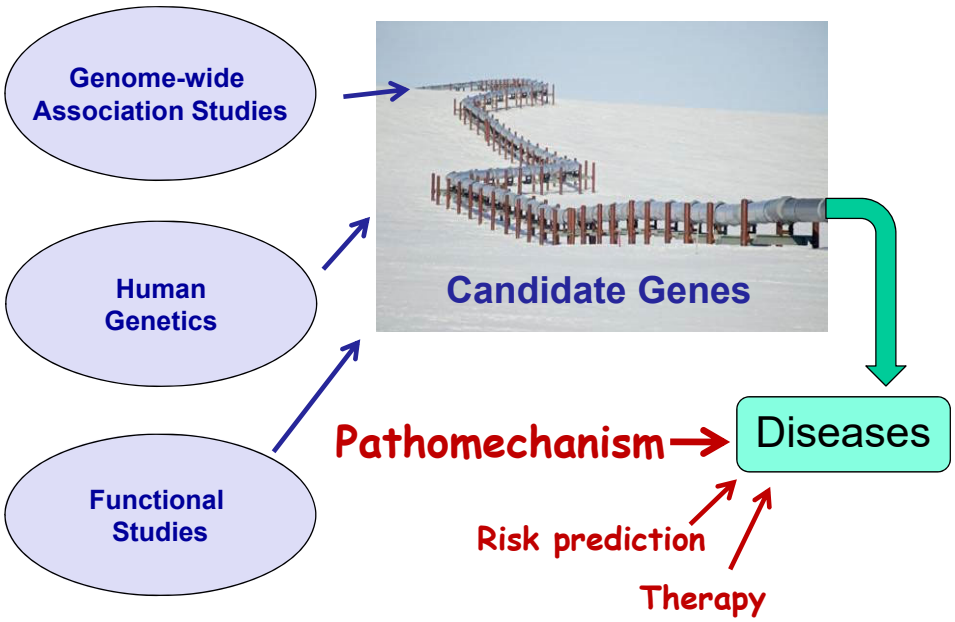
- PCSK9 increases LDL cholesterol: discovered by genetic studies
- PCSK9 inhibitors lower LDL cholesterol by 60%

Exclusion of drug targets

- CETP increases the "good" cholesterol
- Development of CETP inhibitors to increase HDL cholesterol
- Billions of investment without lowering of heart attacks
- Genetic studies would have predicted the failure of these drugs

GENEPI INNSBRUCK

Gene hunting: an interdisciplinary approach



GENEPI INNSBRUCK