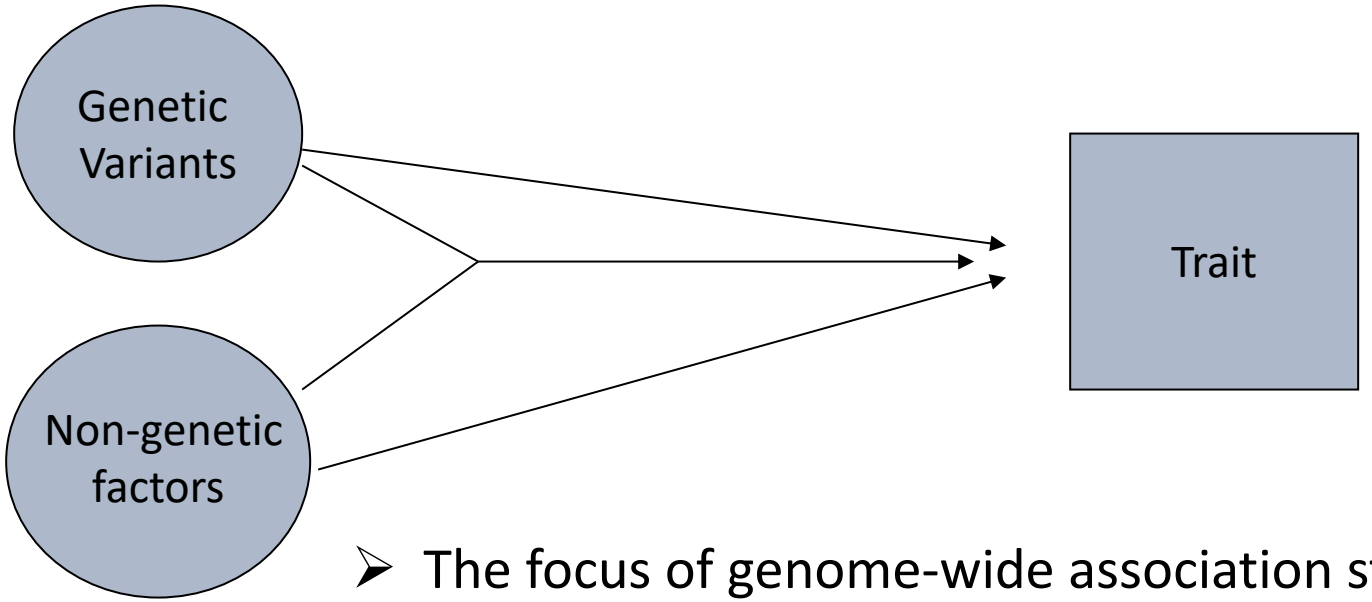
Gene x Environment Interactions

Ioanna Tzoulaki, PhD

@jotzou itzoulak@cc.uoi.gr

Complex Traits: Multifactorial Inheritance



- The focus of genome-wide association studies has typically been on complex genetic diseases.
- Characterized by:
  - Do not follow simple Mendelian mode of inheritance
  - Multiple susceptibility loci
  - Incomplete penetrance
  - Phenocopies, heterogeneity
  - Environmental risk factors
- Many of public health importance

Studies in Genetic Epidemiology

- Linkage analysis using families takes unbiased look at whole genome, but is underpowered for the size of genetic effects we expect to see for many complex genetic traits.

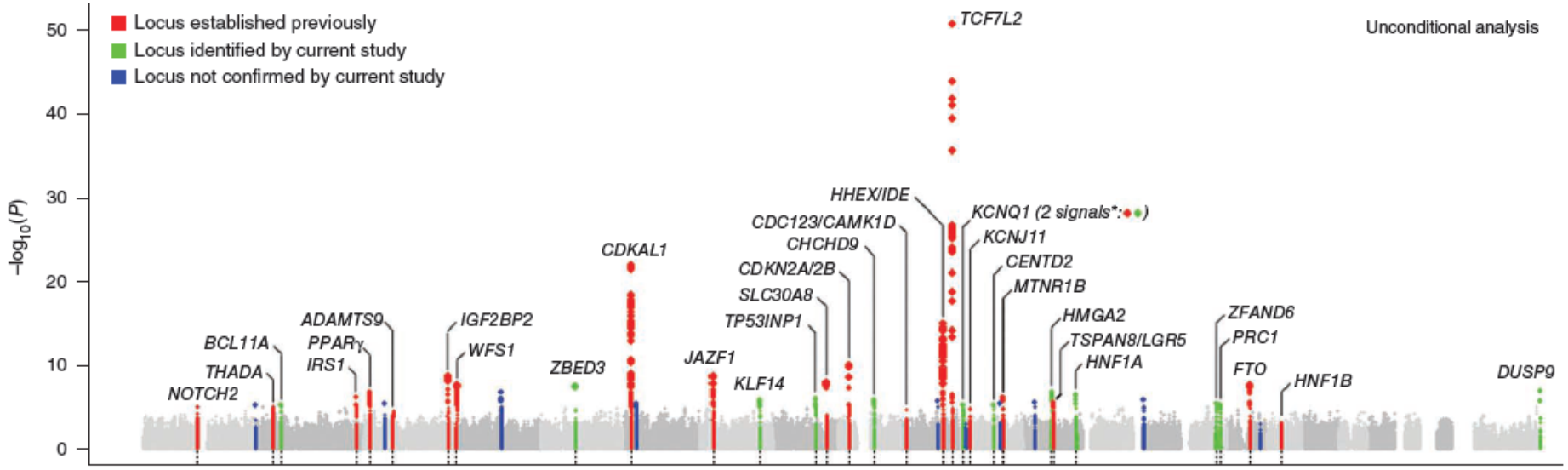
- Candidate-gene association studies have greater power to identify smaller genetic effects, but rely on *a priori* knowledge about disease etiology.

- Genome-wide association studies combine the genomic coverage of linkage analysis with the power of association studies to have much better chance of finding complex trait susceptibility variants.

    - Other advantages: agnostic search, large sample sizes, improved quality of genotyping, rigorous p-value thresholds, replication

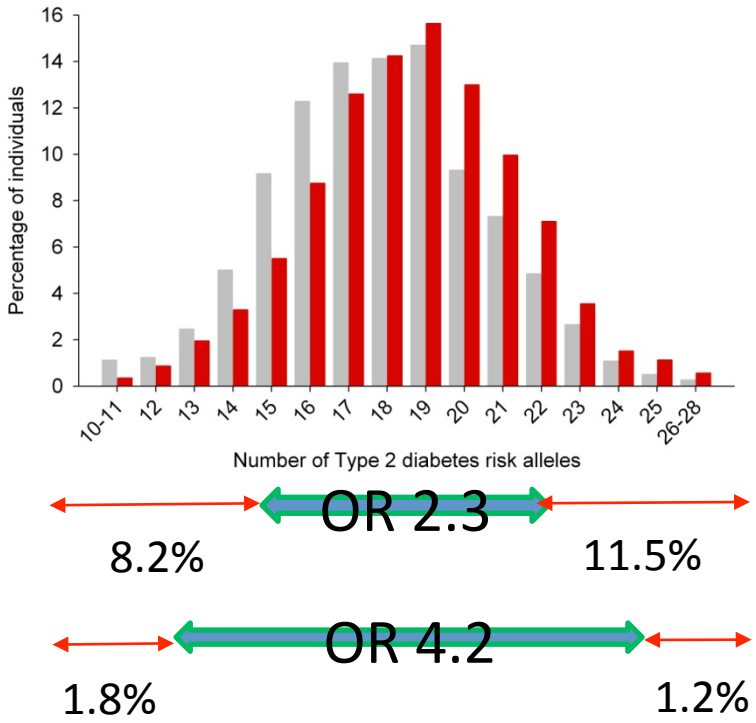
DIAGRAM+ meta-analysis

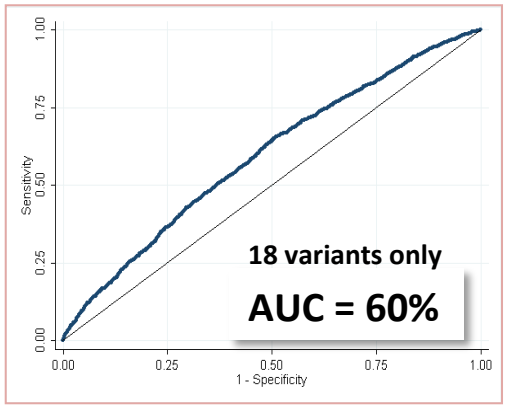
1 000 000 independent statistical tests
Statistical threshold:
P-value=0.05/1 000 000
P-value<0.00000005

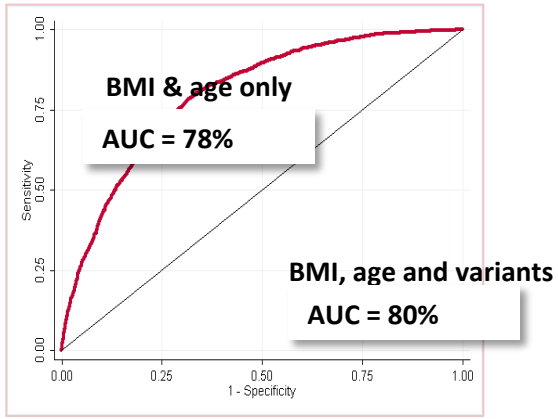


Voight et al, Nature Genetics, 2010

Prediction not (yet) possible







Even with 40 genetic variants prediction is poor

Individual effects are modest

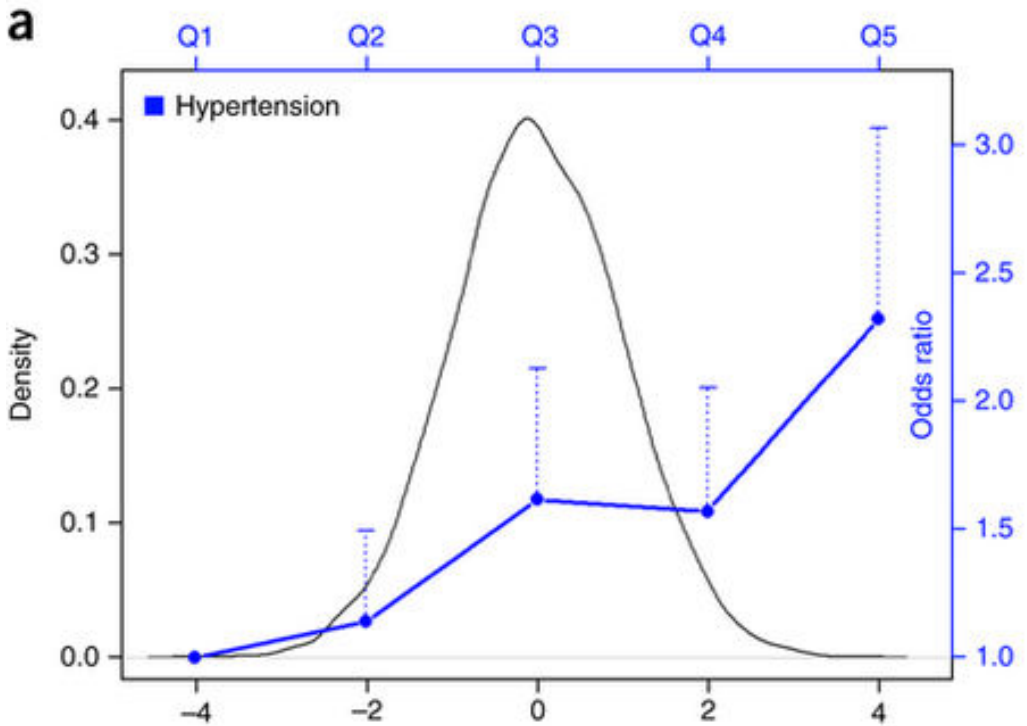
Only ~5-10% of genetic predisposition found

Weedon et al, PLOS, 2007
Lango et al, Diabetes 2008

nature genetics

Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk

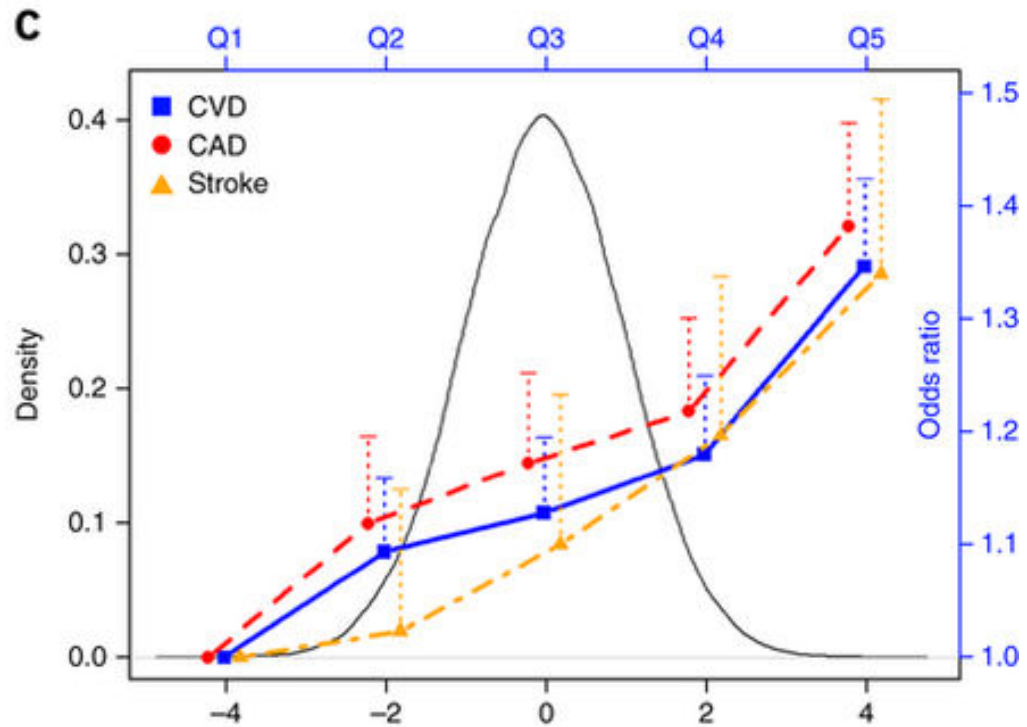
a



b

(mm Hg)	Q1	Q2	Q3	Q4	Q5
SBP	136.4 (17.2)	138.7 (16.5)	142.5 (18.8)	142.5 (19.4)	145.7 (19.7)
DBP	82.9 (9.9)	84.3 (10.3)	85.9 (11.1)	86.0 (10.9)	87.6 (11.9)
PP	53.4 (11.6)	54.4 (11.3)	56.7 (12.3)	56.5 (12.4)	58.2 (12.4)

c



d

(count)	Q1	Q2	Q3	Q4	Q5
CVD	2,462	2,689	2,759	2,874	3,229
CAD	1,783	1,996	2,078	2,154	2,417
Stroke	581	597	640	695	776

New England Journal of Medicine 2009

Genetic Risk Prediction — Are We There Yet?

Peter Kraft, Ph.D., and David J. Hunter, M.B., B.S., Sc.D., M.P.H.

A major goal of the Human Genome Project was to facilitate the identification of inherited genetic variants that increase or decrease the risk of comp diseases. The completion of t International HapMap Project a the development of new metho for genotyping individual D samples at 500,000 or more l have led to a wave of discover through genomewide associati studies. These analyses have id tified common genetic varia that are associated with the r of more than 40 diseases and l man phenotypes. Several com nies have begun offering dire to-consumer testing that uses

tests of genetic predisposition to important diseases would have major clinical, social, and economic ramifications. But the great ma-

est relative risks are almost certainly overrepresented in the first wave of findings from genome-wide association studies, since

Genetic Cardiovascular Risk Prediction

Will We Get There?

George Thanassoulis, MD; Ramachandran S. Vasan, MD

Circulation 2010

Major advances in genetics, including the sequencing of the human genome in 2001[1,2] and the publication of the HapMap in 2005,[3] have paved the way for a revolution in our understanding of the genetics of complex diseases, including cardiovascular disease (CVD). A results and failure to replicate put ciations, high-throughput technolo than 500 000 genetic markers k polymorphisms [SNPs]) and novel a virtual explosion of novel genet complex human diseases. In th advances have been remarkably many novel genetic associations (MI) and cardiovascular risk fac pressure, diabetes, and obesity. A studies has always been to prov biology of CVD. However, a high these discoveries has been to use usher in a new era of personalized genetic information into risk pre

these factors, a number of risk prediction algorithm scores have been developed, including the Framingham risk score, that provide an estimate of the 10-year risk (and recently, the 30-year risk) of CVD.[6–9] Generally speaking, the metrics

Clinical Utility of Genetic Variants for Cardiovascular Risk Prediction

A Futile Exercise or Insufficient Data?

Emanuele Di Angelantonio, MD, MSc, PhD; Adam S. Butterworth, MSc, PhD

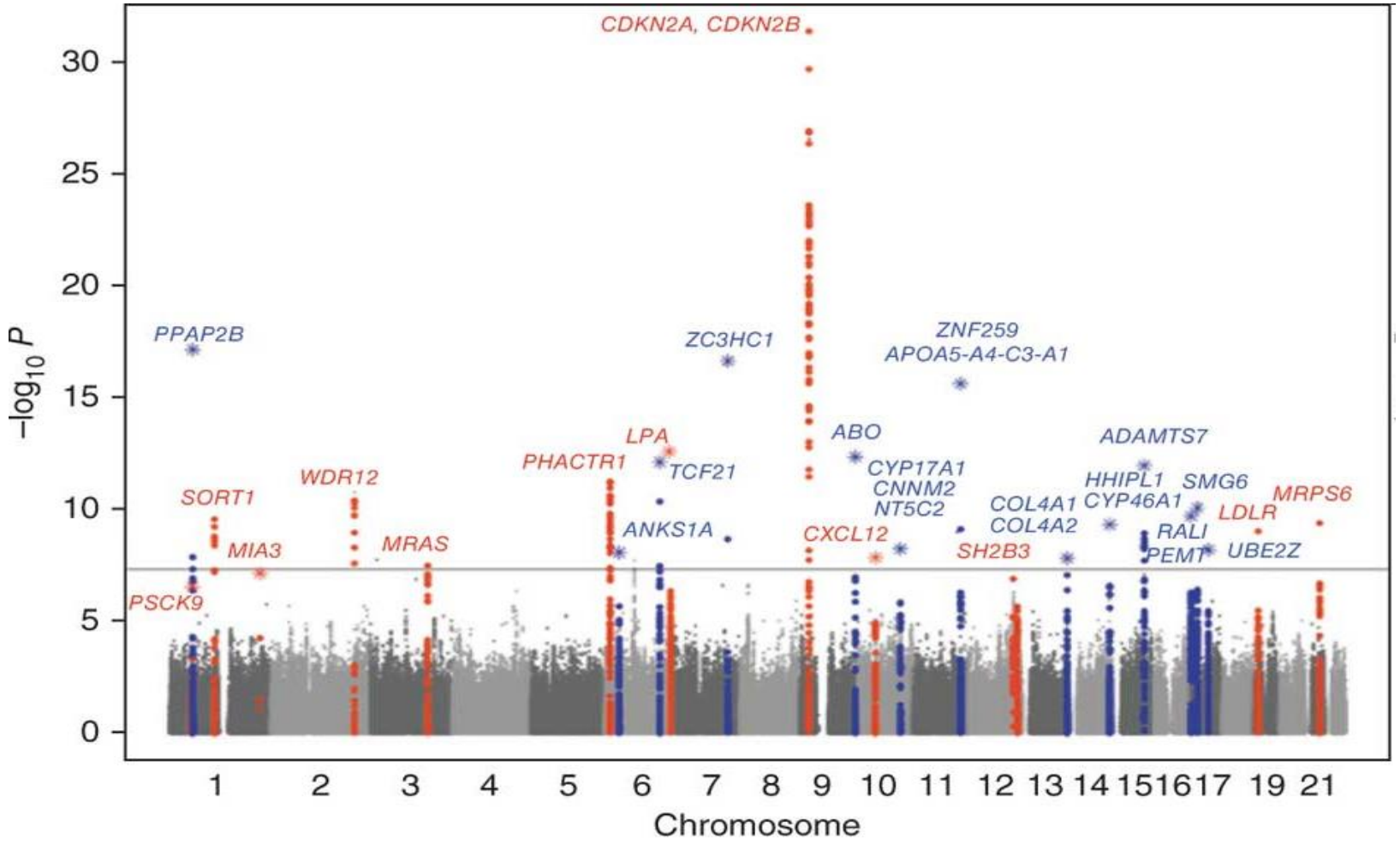
Estimation of an individual's cardiovascular disease (CVD) risk usually involves measurement of risk factors correlated with risk of CVD to identify people who may especially benefit from preventive action, such as lifestyle advice or pharmacologic agents.[1] Since the Framingham Risk Score was first developed, several other risk-prediction algorithms have been proposed, each involving a core set of the same established risk factors (ie, age, sex, smoking, blood pressure, and total cholesterol), but differing in their inclusion of various other characteristics (eg, ethnicity or presence of diabetes mellitus).[2] The challenge in recent years has been to improve existing CVD risk-prediction models by including additional information to the traditional risk factors generally included in risk scores. Several additional soluble biochemical factors have been advocated for inclusion, but contradictory evidenc been reported on the incremental predictive gain afford these markers, and there is divergence of expert opinion

Until a few years ago, genetic epidemiologic studies of CVD were predominantly candidate gene studies involving focused investigation of relatively few genetic variants based on plausible biological hypotheses. Many of these studies had anticipated identification of variants that are common in populations with moderate-to-large effects on disease risk. However, the combination of the low prior odds of the variants selected for study, inadequate power (ie, small sample size), and overliberal declarations of significance, resulted in the reporting of many seemingly positive findings that remain unreplicated or directly refuted.[7] In recent years, genome-wide association studies (GWAS) have demonstrated that so-called hypothesis-free global-testing methods can advance discovery and understanding of genetic variants in relation to chronic

Circ Cardiovasc Genet. 2012

GWAS on Coronary Artery Disease

(The CARDIoGRAM Consortium: 22,500 cases, 65,000 controls – 23 loci)

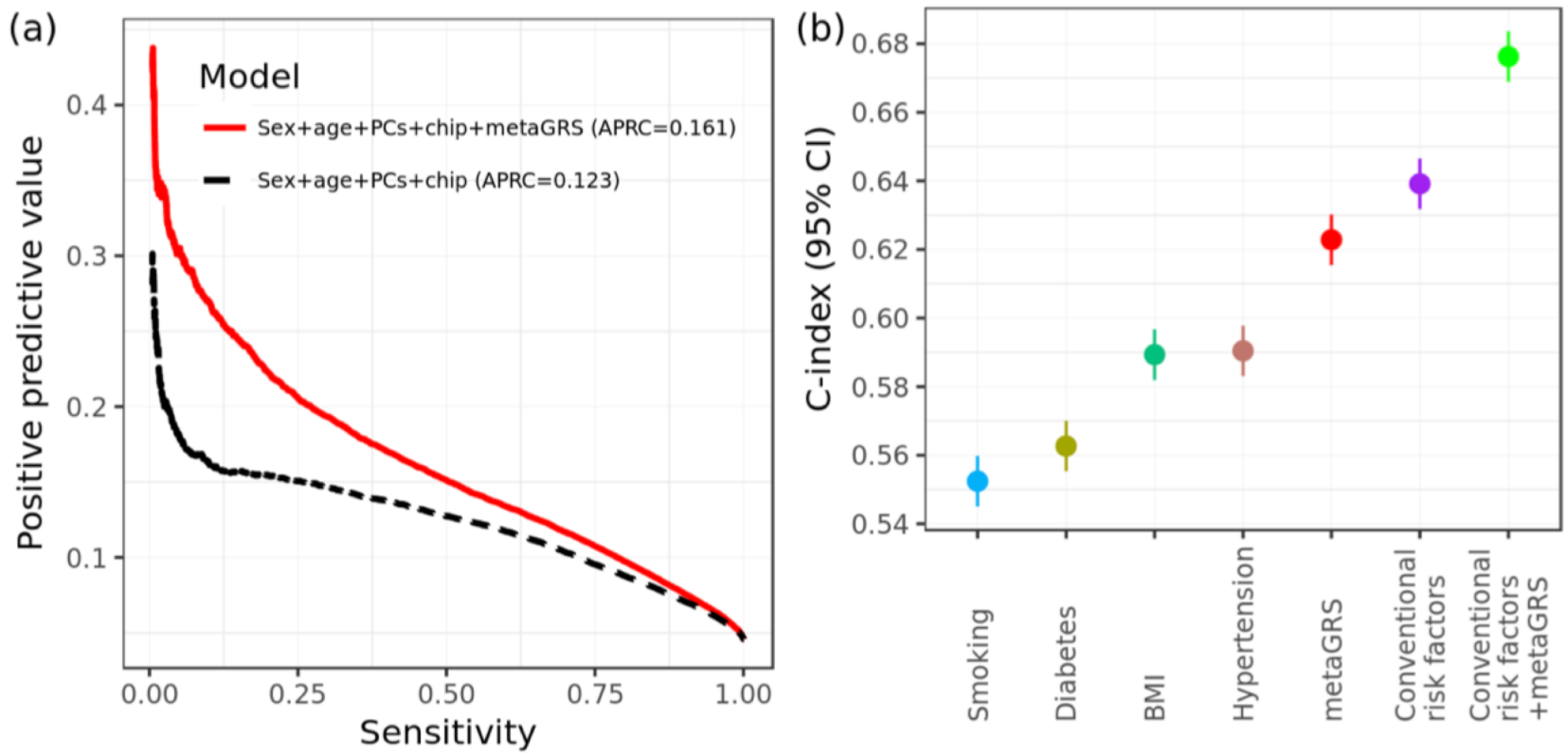


Schunkert H, et al. Nat Genet. 2011

Genomic risk prediction of coronary artery disease in nearly 500,000 adults: implications for early screening and primary prevention

Michael Inouye[1,2,3,#,*], Gad Abraham[1,2,3,#,*], Christopher P. Nelson[4], Angela M. Wood[2], Michael J. Sweeting[2], Frank Dudbridge[2,5], Florence Y. Lai[4], Stephen Kaptoge[2,6], Marta Brozynska[1,2,3], Tingting Wang[1], Shu Ye[4], Thomas R Webb[4], Martin K. Rutter[7,8], Ioanna Tzoulaki[9,10], Riyaz S. Patel[11,12], Ruth J.F. Loos[13], Bernard Keavney[14,15], Harry Hemingway[16], John Thompson[5], Hugh Watkins[17,18], Panos Deloukas[19], Emanuele Di Angelantonio[2,6], Adam S. Butterworth[2,6], John Danesh[2,6,20], Nilesh J. Samani[4,#,*] for The UK Biobank CardioMetabolic Consortium CHD Working Group

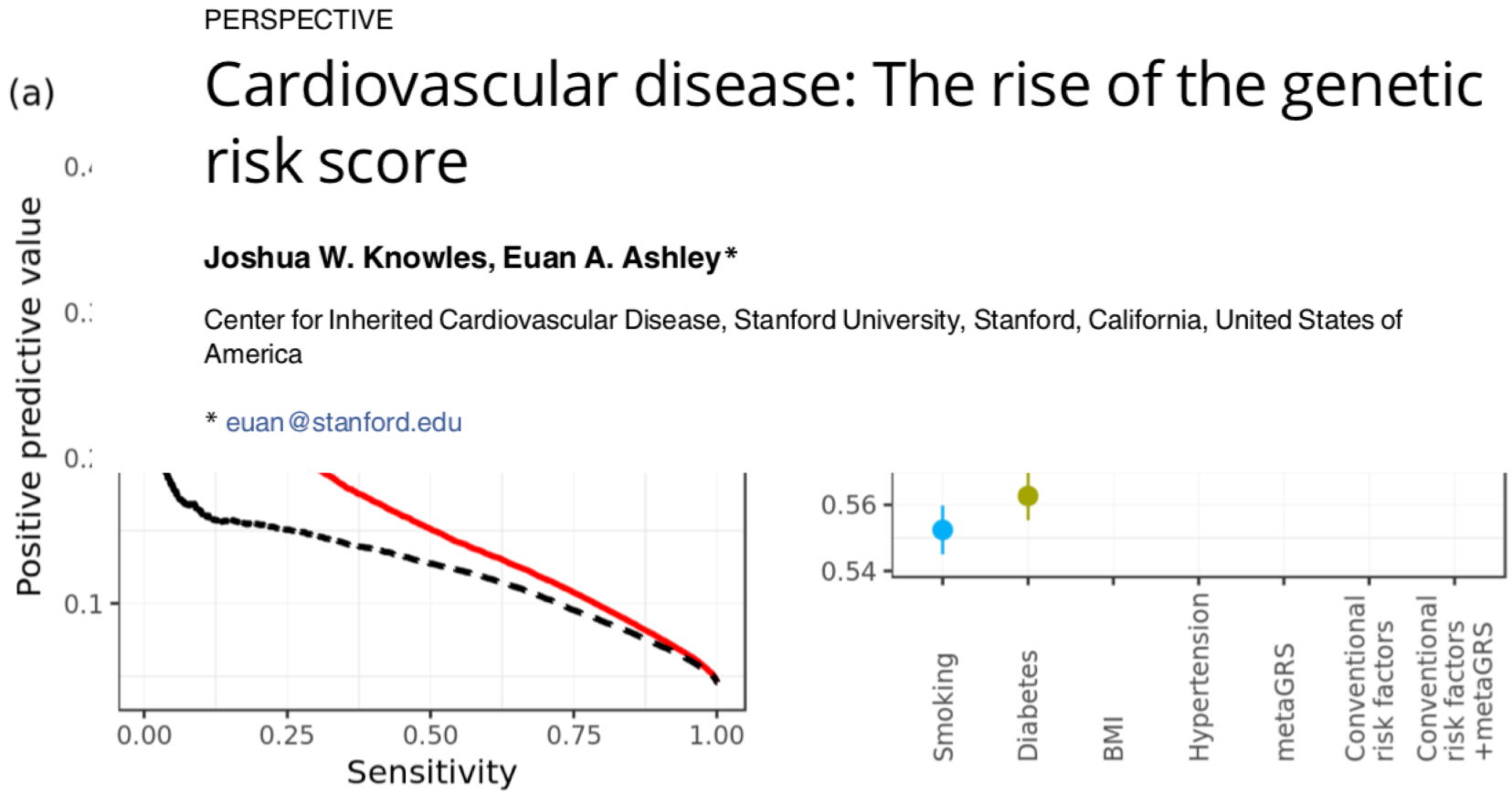
Figure 2: Predictive measures of CAD using the metaGRS and conventional risk factors



Genomic risk prediction of coronary artery disease in nearly 500,000 adults: implications for early screening and primary prevention

Michael Inouye[1,2,3,#,*], Gad Abraham[1,2,3,#,*], Christopher P. Nelson[4], Angela M. Wood[2], Michael J. Sweeting[2], Frank Dudbridge[2,5], Florence Y. Lai[4], Stephen Kaptoge[2,6], Marta Brozynska[1,2,3], Tingting Wang[1], Shu Ye[4], Thomas R Webb[4], Martin K. Rutter[7,8], Ioanna Tzoulaki[9,10], Riyaz S. Patel[11,12], Ruth J.F. Loos[13], Bernard Keavney[14,15], Harry Hemingway[16], John Thompson[5], Hugh Watkins[17,18], Panos Deloukas[19], Emanuele Di Angelantonio[2,6], Adam S. Butterworth[2,6], John Danesh[2,6,20], Nilesh J. Samani[4,#,*] for The UK Biobank CardioMetabolic Consortium CHD Working Group

Figur



Missing Heritability?

The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Missing Heritability?

EDITORIAL

Missing Heritability and GWAS Utility

Clifton Bogardus*

doi:10.1038/oby.2008.613

Vol 461|8 October 2009|doi:10.1038/nature08494 nature

REVIEWS

Finding the missing heritability of complex diseases

Teri A. Manolio[1], Francis S. Collins[2], Nancy J. Cox[3], David B. Goldstein[4], Lucia A. Hindorff[5], David J. Hunter[6], Mark I. McCarthy[7], Erin M. Ramos[5], Lon R. Cardon[8], Aravinda Chakravarti[9], Judy H. Cho[10], Alan E. Guttmacher[1], Augustine Kong[11], Leonid Kruglyak[12], Elaine Mardis[13], Charles N. Rotimi[14], Montgomery Slatkin[15], David Valle[9], Alice S. Whittemore[16], Michael Boehnke[17], Andrew G. Clark[18], Evan E. Eichler[19], Greg Gibson[20], Jonathan L. Haines[21], Trudy F. C. Mackay[22], Steven A. McCarroll[23] & Peter M. Visscher[24]

The case o

When scientists opened up t
common traits and diseases.
six places where the missing loot could be stashed away.

Reasons for missing heritability

- “Common disease, common variant” is incorrect – study rarer variants
- Calculation of heritability effects is wrong?
- Not enough common variants of small effect detected
- Structural or other genomic variants more important
- Difficult to analyse gene-gene/gene-environment interactions and in general high-dimensional and systems biology data (i.e., combination of genomic, transcriptomic, proteomic, metabolomic data)

Ways forward…

- Further genetic discovery (denser genotyping)
- Better characterization of validated genes
- Use genes for causal inference (Mendelian randomization)
- Whole genome sequencing
- Systems biology approaches
- Development of clinically useful risk prediction models
- Other translation

Outline

- Gene-Environment Interaction
  - Conceptual Overview
  - Rationale
  - Challenges
  - Study designs
  - Established Examples

- Mendelian Randomization
  - Conceptual Overview
  - Assumptions
  - Effect estimation
  - Examples
  - Limitations and Current Advances

Definitions of gene-environment interaction

- *"Variation in the measure of effect of an environmental risk factor on an outcome according to genotype"*

- *"Joint effect of one or more genes with one or more environmental factors that cannot be readily explained by their separate marginal effects"*

- Examples: Individuals with different genotypes could differ in terms of:
  - Susceptibility to the health effects of exposures such as diet, smoking, drinking, sedentary lifestyle, etc.
  - Responses to life events such as trauma
  - Responses to medications (pharmacogenomics)

Thomas D. Nat Rev Gen 2010;259-272

Types of gene-environment interaction (I)

Model	Interpretation
No interaction	The **same effect** of the exposure on the outcome in individuals with different genotypes
Statistical interaction	A departure from a pure main effects model observed in one or a few studies
Positive interaction or synergism	**Greater effect** of the exposure on the outcome in individuals with a genotype of interest than in individuals with other genotypes
Negative interaction or antagonism	**Smaller effect** of the exposure on the outcome in individuals with a genotype of interest than in individuals with other genotypes
Multiplicative interaction	Interaction observed in multiplicative/relative measures of effect (e.g., OR, RR, HR, etc.)
Additive interaction	Interaction observed in additive/absolute measures of effect (e.g., RD, etc.)

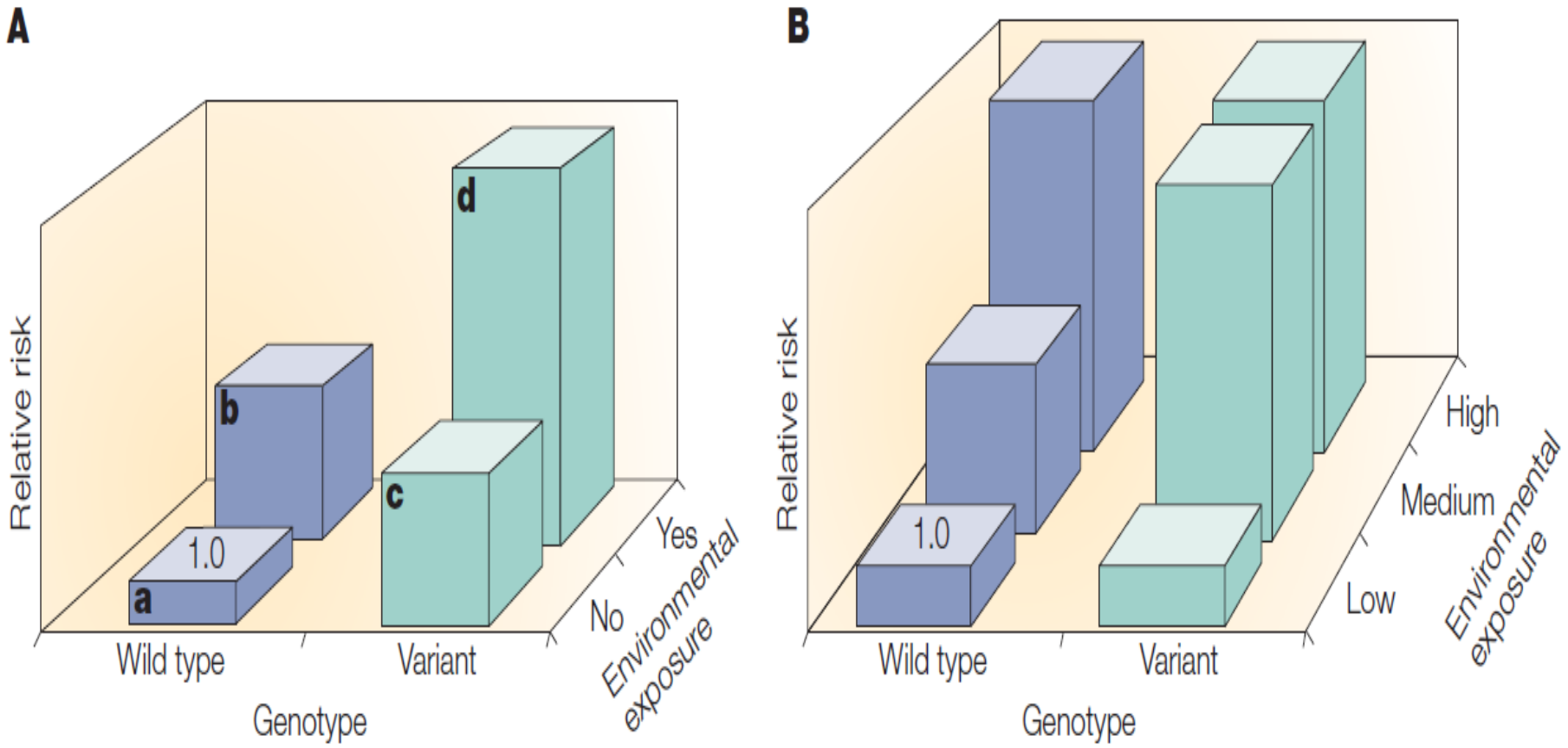
Types of gene-environment interaction (II)

Model	Interpretation
Quantitative interaction	Interaction in which the effects of the exposure on the outcome go in the same direction for different genotypes, but differ in magnitude
Qualitative interaction	Interaction in which the effects of the exposure on the outcome go in opposite directions (e.g., deleterious in carriers and protective in non-carriers) for different genotypes
Biological or causal interaction	An interaction that is present in nature (and is supported by the totality of the evidence)

Uses of gene-environment interaction

- Understanding biological mechanisms and pathways
  - Tobacco smoking – *NAT2* – bladder cancer
- Understanding heterogeneity in results across studies
- Identifying novel genes acting only through interactions
  - Could explain missing heritability (e.g., genetic susceptibility to air pollution in childhood asthma)
- Predicting individual risk of disease or prognosis
  - Optimal mammographic screening interval for *BRCA1* or *BRCA2* mutation carriers
  - Folate supplementation for colorectal cancer risk could depend on *MTHFR*
- Choosing the best treatment for an individual based on genetic predisposition
  - Statins – *SLCO1B1* - cardiomyopathy

Models of gene-environment interaction



Hunter DJ. Nat Rev Gen 2005;287-298

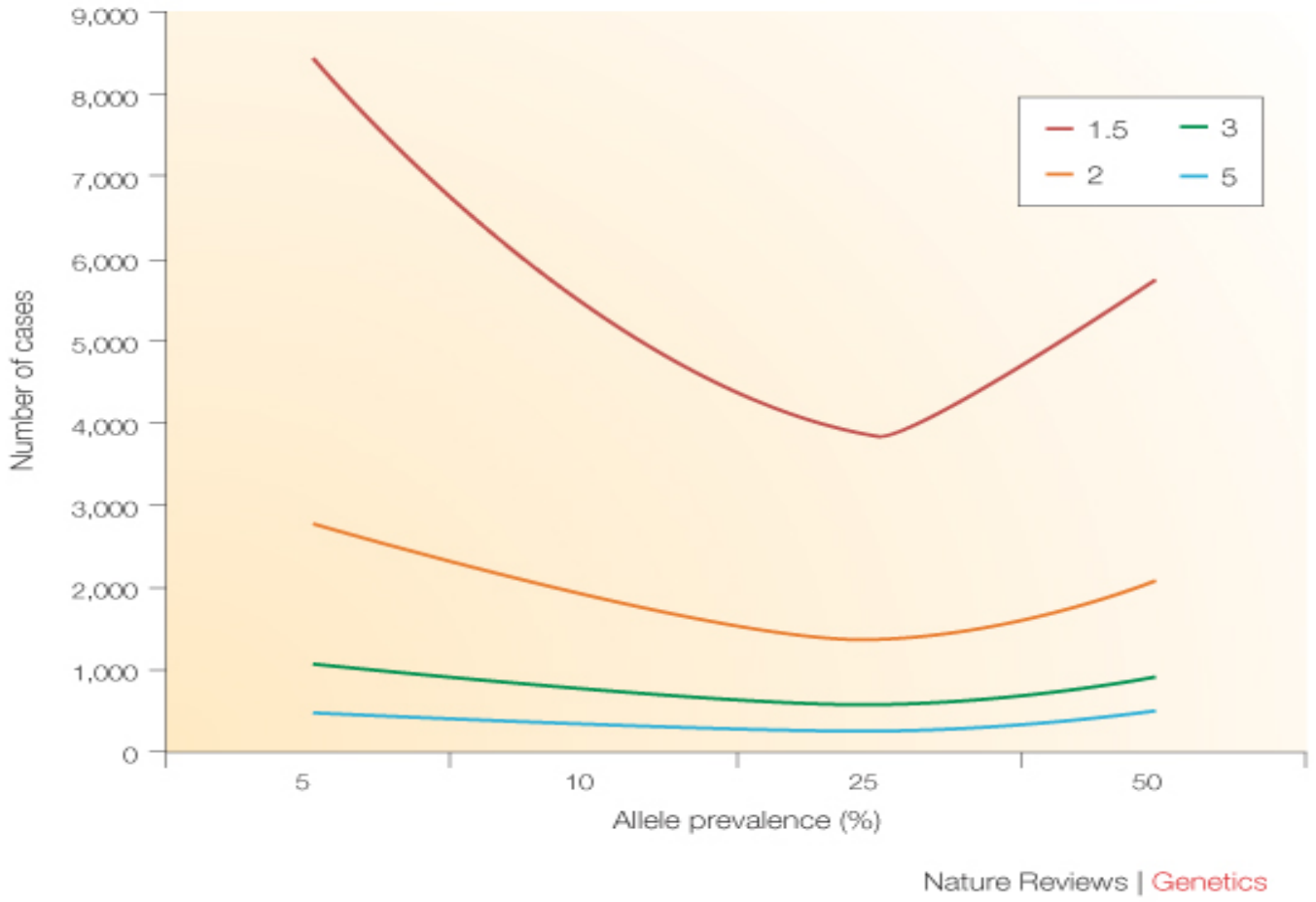
Challenges of gene-environment interaction

- Exposure assessment
  - Multidimensional, time-varying exposures
  - Interactions will be biased only if measurement errors are differentially related to both exposure and genotype

- Sample size and power
  - Sample size requirements can be enormous
  - Interactions require samples ~four times larger than are needed to find genetic main effects
  - Some of the poor replication ability of GxE interactions are due to underpowered studies

- Heterogeneity and replication

Example of Sample Size Issue for detecting ONE interaction for a dichotomous trait and a 10% exposure prevalence



Hunter DJ. Nat Rev Gen 2005;287-298

Study designs for gene-environment interaction (I)

Design	Approach	Advantages	Disadvantages	Settings	Examples
Basic epidemiologic designs					
Cohort	Comparison of incidence of new cases across groups defined by E and G	Freedom from most biases; clear temporal sequence of cause and effect	Large cohorts and/or long follow-up needed to obtain sufficient numbers of cases; possible biased losses to follow-up; changes in exposure may require recurring observation	Common Ds or multiple end points; commonly used in biobanks	*ITGB3* × fibrinogen in platelet aggregation in Framingham cohort[154]
Case–control	Comparison of prevalence of E and G between cases and controls	Modest sample sizes needed for rare Ds; can individually match on confounders	Recall bias for E; selection bias, particularly for control group	Rare Ds with common E and G risk factors	*CYP1A2*, *NAT2*, smoking and red meat in colorectal cancer[57]
Case-only	Test of G–E association among cases, assuming G–E independence in the source population	Greater power than case–control or cohort	Bias if G–E assumption is incorrect	G×E studies in which G–E independence can be assumed	Radiotherapy × DNA repair genes in second breast cancers[32]
Randomized trial	Cohort study with random assignment of E across individuals	Experimental control of confounders	Prevention trials for D incidence can require very large sample sizes	Experimental confirmation for chronic effects	Albuteral and *B2AR* in asthmatics[126]

Thomas D. Nat Rev Gen 2010;259-272

Study designs for gene-environment interaction (II)

Design	Approach	Advantages	Disadvantages	Settings	Examples
Hybrid designs					
Nested case–control	Selection of matched controls for each case from cohort members who are still D-free	The freedom from bias of a cohort design combined with the efficiency of a case–control design; simple analysis	Each case group requires a separate control series	Studies within cohorts requiring additional data collection	Antioxidants × MPO in breast cancer[155]
Case–cohort	Unmatched comparison of cases from a cohort with a random sample of the cohort	Same advantages as nested case–control; the same control group can be used for multiple case series	Complex analysis	Studies within cohorts with stored baseline biospecimens	*APOE* and smoking for CHD in Framingham offspring cohort[156]
Two-phase case–control	Stratified sampling on D, E and G for additional measurements (for example, biomarkers)	High statistical efficiency for subsample measurements	Complex analysis	Substudies for which outcome and predictor data are already available	*GST* genes and tobacco smoking in CHD[47]
Counter-matching	Matched selection of controls who are discordant for a surrogate for E	Permits individual matching; highly efficient for E main effect and G×E interactions	Complex control selection	Substudies in which a matched design is needed	Radiotherapy × DNA repair genes in second breast cancers[49]
Joint case-only and case–control	Bayesian compromise between case-only and case–control comparisons	Power advantage of case-only combined with robustness of case–control	Some bias when G–E association is moderate	G×E studies for which G–E independence is uncertain	*GSM1*, *NAT2*, smoking and diet in colorectal cancer[34]
Family-based designs					
Case–sibling (or –cousin)	Case–control comparison of E and G using unaffected relatives as controls	More powerful than case–control for G×E; immune to population stratification bias	Discordant sibships difficult to enroll; overmatching for G main effects	Populations with potential substructure	*GSTM1* × air pollution in childhood asthma[17]
Case–parent triad	Comparison of Gs for cases with Gs that could have been inherited from parents, stratified by case's E	More powerful than case–control for G×E; immune to population stratification bias for G main effects	Difficult to enroll complete triads; possible bias in G×E if G and E are associated within parental mating types	Substructured populations, particularly for Ds of childhood	*TGFA* × maternal smoking, alcohol and vitamins in cleft palate[157]
Twin studies	Comparison of D concordance between MZ and DZ pairs in different environments	No genetic data required; can be extended to include half-siblings, twins reared together or apart, or to compare discordant pairs on measured G and E	Used mainly to identify interactions with unmeasured genes; assumption of similar E between MZ and DZ pairs	Exploratory studies of potential for G×E before specific genes have been identified	Concordance of insulin levels in relation to non-genetic variation in obesity[158]

Study designs for gene-environment interaction (III)

Design	Approach	Advantages	Disadvantages	Settings	Examples
GWA designs					
Two-stage genotyping	Use of high-density panel on part of a case–control sample to select a subset of SNPs with suggestive Gs or G×E interaction for testing; the SNPs are tested using a custom panel in an independent sample, with joint analysis of both samples	Highly cost efficient	Only part of sample has GWA genotypes	GWA studies for which complete SNP data on all subjects is not needed	None identified
Two-step interaction analysis	Preliminary filtering of a GWA scan for G–E association in combined case–control sample, followed by G×E testing of a selected subset	Much more powerful for G×E or G×G interactions than a single-step analysis	Can miss some interactions	GWA studies with complete SNP data and focus on G×E	G × *in utero* tobacco in childhood asthma

Thomas D. Nat Rev Gen 2010;259-272

GxE Interaction: Testing for Additive/Multiplicative Effects

Stratum	Cases	Controls
Gene (G+), Environment (E+)	a	b
Gene (G+), No Environment (E-)	c	d
No Gene (G-), Environment (E+)	e	f
No Gene (G-), No Environment (E-)	g	h

Identifying GxE Interaction

Strata	Cases	Controls
G+E+	a	b
G+E-	c	d
G-E+	e	f
G-E-	g	h

Odds Ratio (OR)

ah / bg

ch / dg

eh / fg

1 (Ref)

GxE Interaction:
4 groups defined by genotype and exposure

COHORT STUDY:	**G+ E+**	**G+ E-**	**G- E+**	**G- E-**
Affected	a	b	e	f
Unaffected	c	d	g	h
Risk	a/(a+c)	b/(b+d)	e/(e+g)	f/(f+h)
Relative risk	$RR_{G+} = \frac{a/(a+c)}{b/(b+d)}$		$RR_{G-} = \frac{e/(e+g)}{f/(f+h)}$	
Risk difference	$RD_{G+} = a/(a+c) - b/(b+d)$		$RD_{G-} = e/(e+g) - f/(f+h)$	

Test for interaction: Is the effect of the exposure the same in people with and without the high-risk genotype?

Multiplicative scale: No interaction implies $RR_{G+} = RR_{G-}$
Additive scale: No interaction implies $RD_{G+} = RD_{G-}$

Example: Factor V Leiden Mutations, Oral Contraceptive Use, and Venous Thrombosis

Strata	Cases	Controls	OR
G+E+	25	2	34.7
G+E-	10	4	6.9
G-E+	84	63	3.7
G-E-	36	100	Reference
Total	155	169	

Vanderbroucke et al., The Lancet 1994

Factor V Leiden Mutations, Oral Contraceptive Use, and Venous Thrombosis

Evidence for Interaction?

Strata OR

G+E+ 34.7

G+E- 6.9

G-E+ 3.7

G-E- Ref

$OR_{Interaction}$ =

34.7 / 6.9 x 3.7 = 1.4

Risk of thrombosis in women using OCs is much greater among those with Factor V Leiden Mutations than those without

Examples of gene-environment interactions

Table 2 | **Selected examples of gene–environment interactions observed in at least two studies**

Gene symbol	Variant(s)	Environmental exposure	Outcome and nature of interaction
Genes for skin pigmentation (for example, *MC1R*)	Variants for fair skin colour	Sunlight or ultraviolet light B	Risk of skin cancer is higher in people with fair skin colour that are exposed to higher amounts of sunlight
CCR5	Δ-32 deletion	HIV	Carriers of the receptor deletion have lower rates of HIV infection and disease progression
MTHFR	Ala222Val polymorphism	Folic acid intake	Homozygotes for the low activity Ala222Val variant are at different risk of colorectal cancer and adenomas if nutritional folate status is low
NAT2	Rapid versus slow acetylator SNPs	Heterocyclic amines in cooked meat	Red meat intake is more strongly associated with colorectal cancer among rapid acetylators
F5	Leiden prothrombotic variant	Hormone replacement	Venous thromboembolism risk is increased in factor V Leiden carriers who take exogenous steroid hormones
UGT1A6	Slow-metabolism SNPs	Aspirin	Increased benefit of prophylactic aspirin use in carriers of the slow metabolism variants
APOE	*E4* allele	Cholesterol intake	Exaggerated changes in serum cholesterol in response to dietary cholesterol changes in *APOE4* carriers
ADH1C	γ-2 alleles	Alcohol intake	Inverse association between ethanol intake and myocardial infarction; risk is stronger in carriers of slow-oxidizing γ-2 alleles
PPARG2	Pro12Ala	Dietary fat intake	Stronger relation between dietary fat intake and obesity in carriers of the Pro12Ala allele
HLA-DPB1	Glu69	Occupational beryllium	Exposed workers who are carriers of the Glu69 allele are more likely to develop chronic beryllium lung disease
TPMT	Ala154Thr and Tyr240Cys	Thiopurine drugs	Homozygotes for the low-activity alleles of *TPMT* are likely to experience severe toxicity when exposed to thiopurine drugs
ADRB2	Arg16Gly	Asthma drugs	Arg16Gly homozygotes have a greater response in the airway to albuterol

Hunter DJ. Nat Rev Gen 2005;287-298

Gene gene interaction

- Gene-gene (also known as epistasis)
- Gene-gene-environment interactions
- Regression-based analyses

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \mathrm{SNP1} + \beta_2 \mathrm{SNP2} + \beta_3 \mathrm{SNP1} \times \mathrm{SNP2}$$

- Pairwise gene-gene interactions → too many tests
  - data reduction approaches
    - LD prunning
  - hypothesis-driven approach on biological approach
  - prior statistical knowledge

Multiple testing correction

- Hundreds of thousands or millions of variants are considered
- Multiple environmental factors
- 500,000 SNPs → 2.5 x 10^{11} tests
- Bonferroni correction overly conservative
- False Discovery Rate
- Permutation testing → computationally intensive

Power of interaction analysis

	Number of cases and controls					
Interaction type	5000	10,000	20,000	30,000	40,000	50,000
G-G	0.000	0.003	0.164	0.654	0.940	0.995
G-E	0.024	0.032	0.317	0.717	0.928	0.988

Examples of ‘established’ GxE Interactions

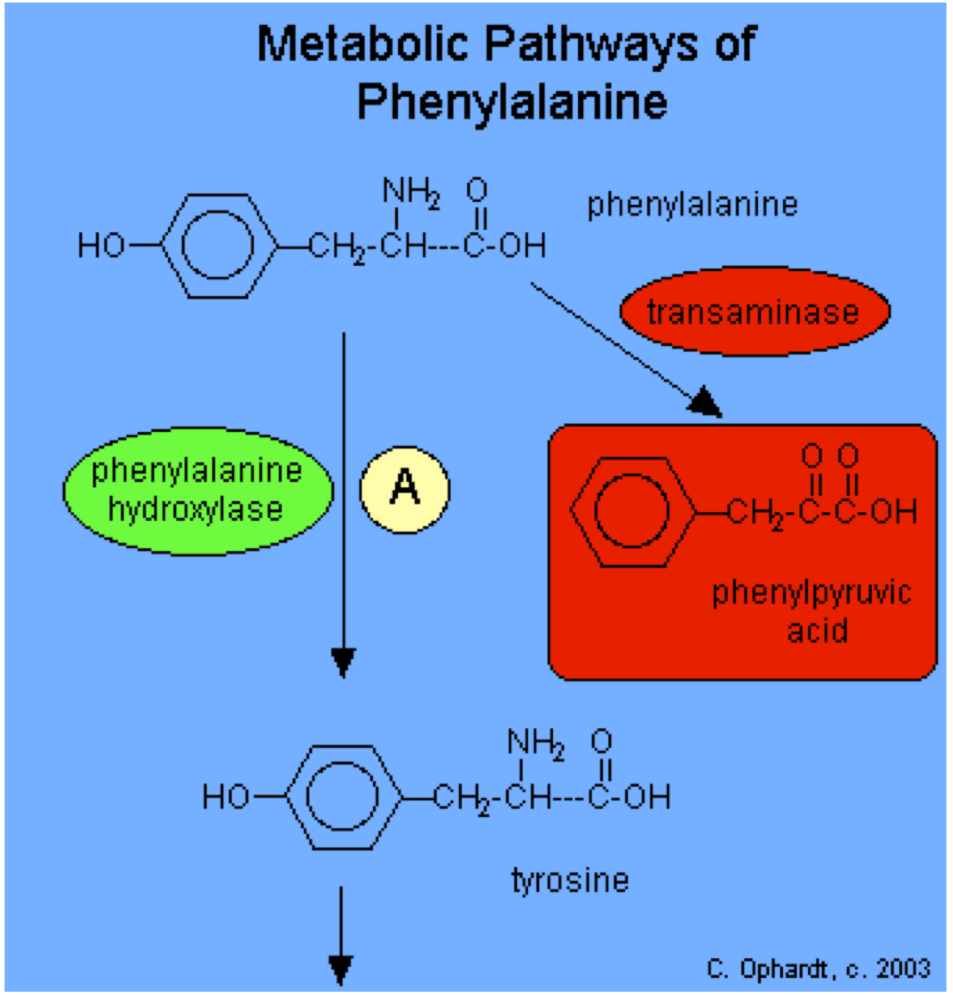
- Have any GXE Interactions been identified with certainty?
  - Few Established Examples to Date
    - Phenylketonuria
    - Lactose Intolerance
    - Smoking, NAT2 and Bladder Cancer
    - Coffee, GRIN2A, and Parkinson’s Disease?

Phenylketonuria

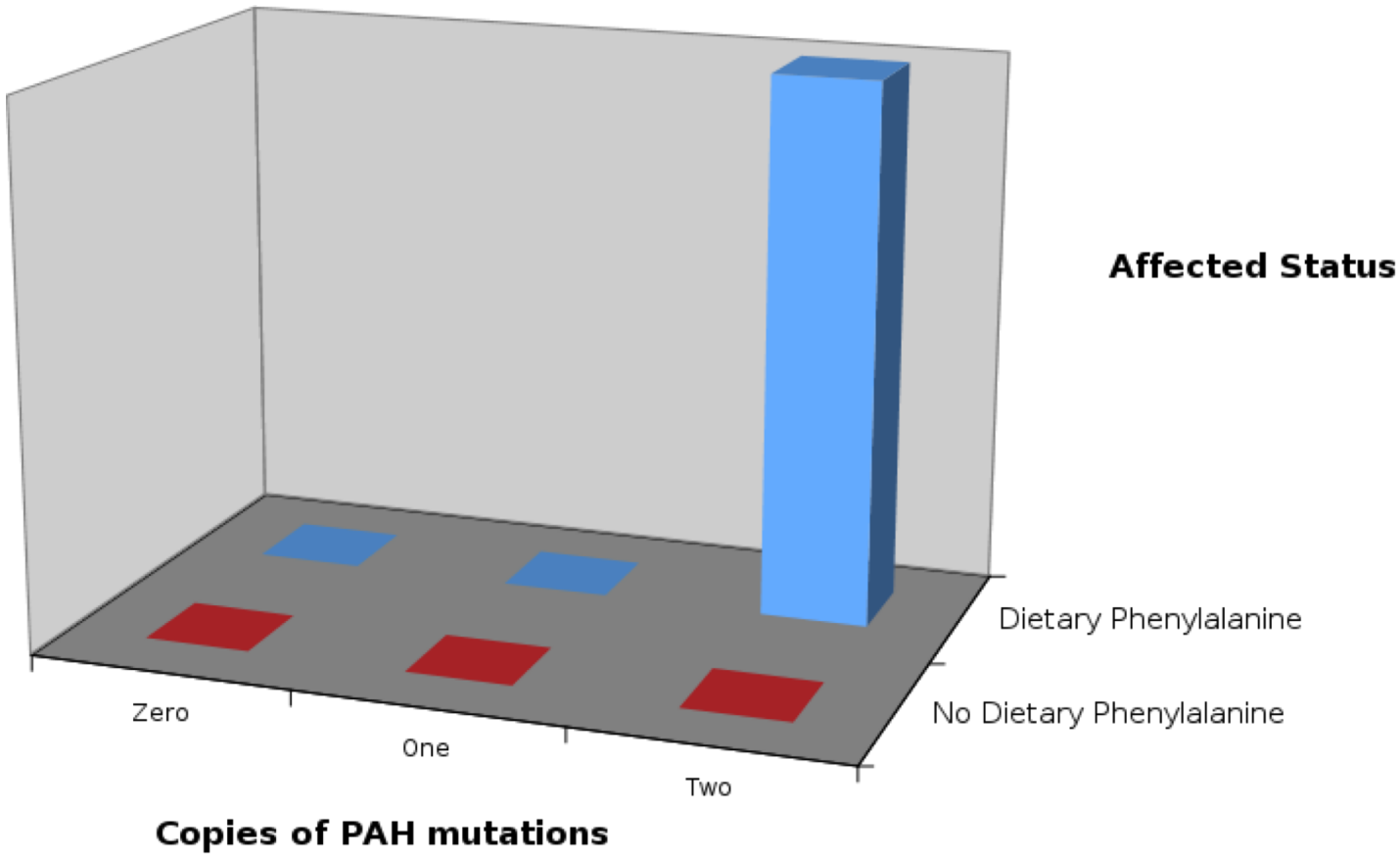
- Mental retardation and seizures
- 1/15,000 live births
  - 1/100,000 in Finland
  - 1/2,600 in Turkey

- Mutations in Phenylalanine Hydroxylase (PAH) (G)
- Dietary Phenylalanine (E)

- Both are necessary
- Neither is sufficient for disease

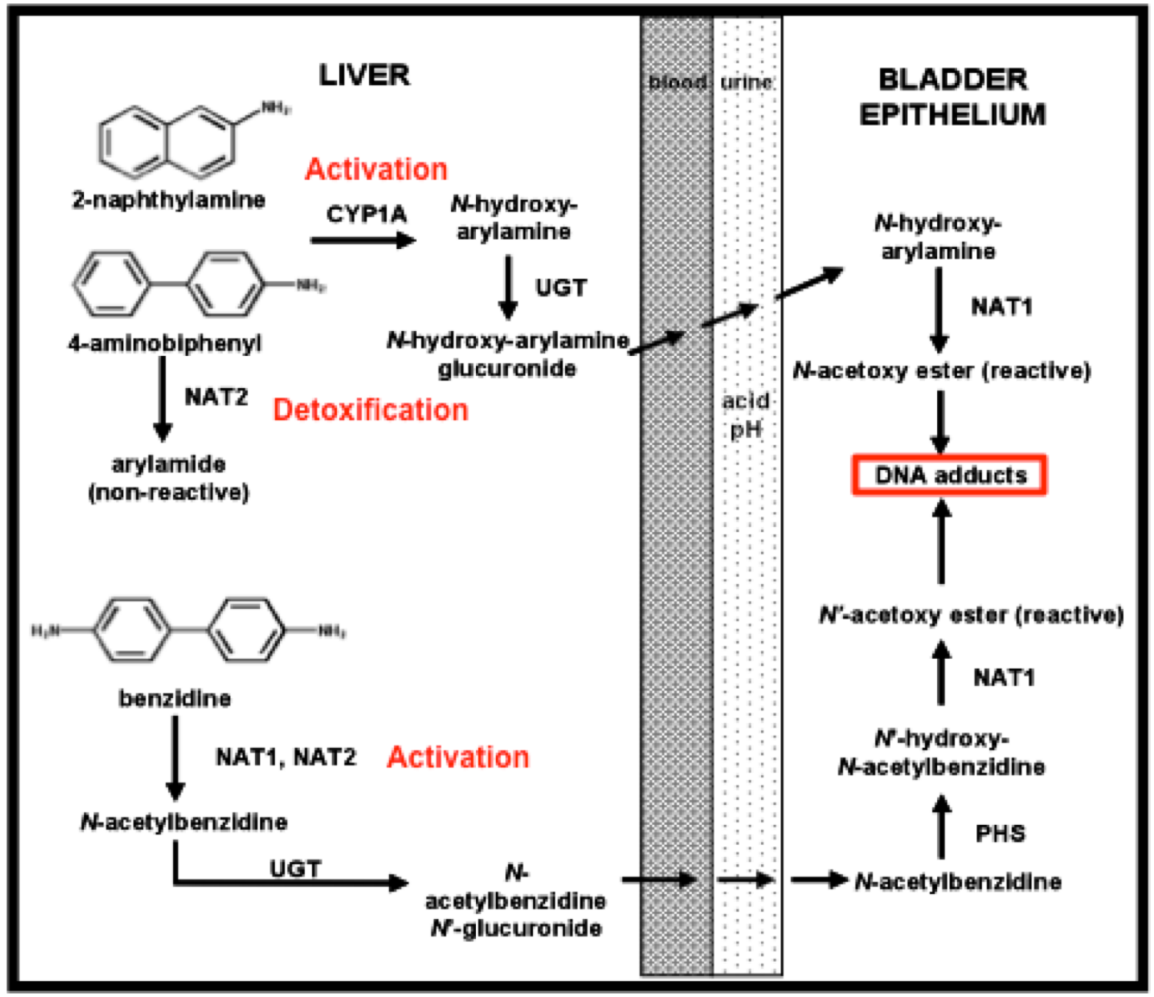


Phenylketonuria: Example of Gene-Nutrition Interaction



Common variation in metabolizing genes could modify the effects of arylamine exposure

Metabolism of aromatic amines and bladder carcinogenesis



Strong in vitro and in vivo evidence that interaction exists

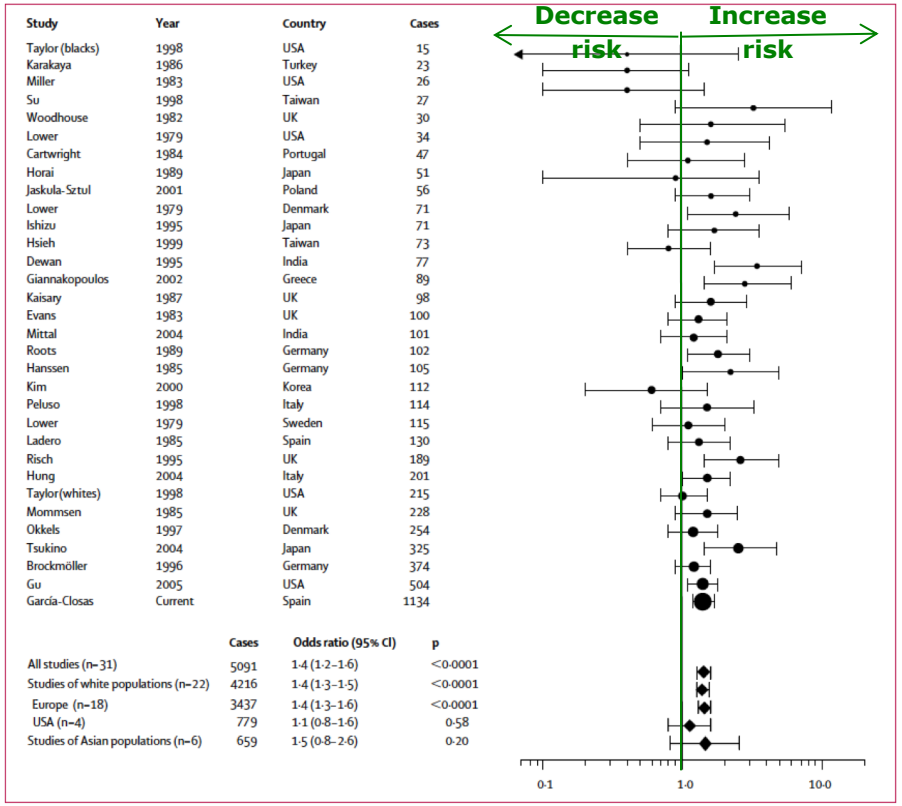
Lancet 2005; 366: 649–59

NAT2 slow acetylation, *GSTM1* null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses

Montserrat García-Closas, Núria Malats, Debra Silverman, Mustafa Dosemeci, Manolis Kogevinas, David W Hein, Adonina Tardón, Consol Serra, Alfredo Carrato, Reina García-Closas, Josep Lloreta, Gemma Castaño-Vinyals, Meredith Yeager, Robert Welch, Stephen Chanock, Nilanjan Chatterjee, Sholom Wacholder, Claudine Samanic, Montserrat Torà, Francisco Fernández, Francisco X Real, Nathaniel Rothman

NAT2 slow acetylation increases bladder cancer risk by 40%

OR=1.4 95%CI (1.2-1.6)



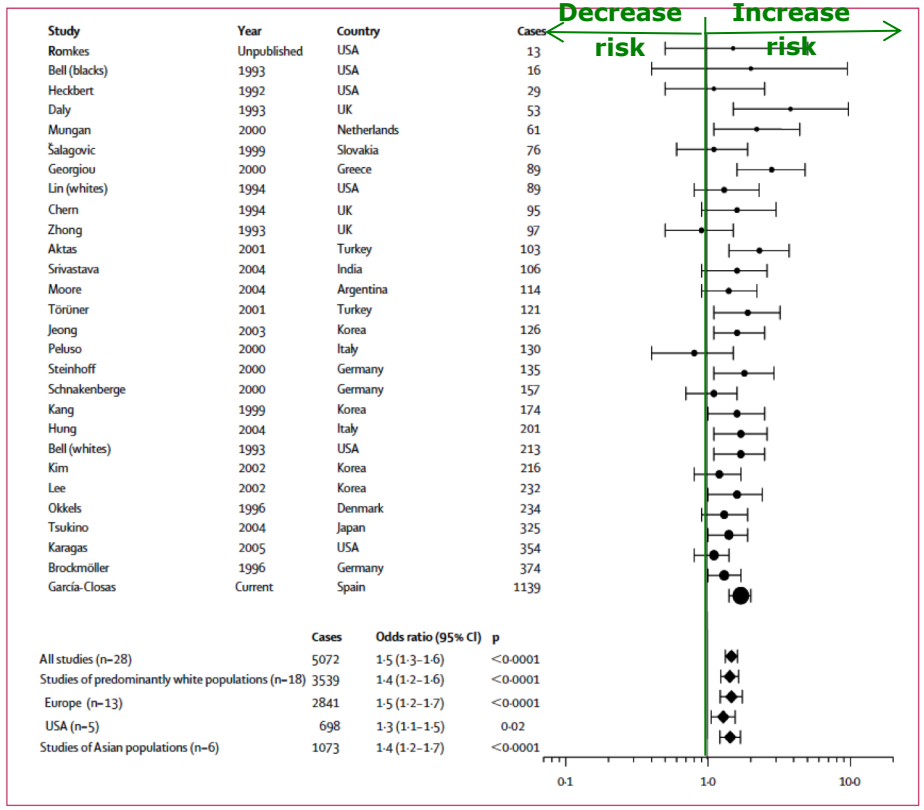
Study	Year	Country	Cases
Taylor (blacks)	1998	USA	15
Karakaya	1986	Turkey	23
Miller	1983	USA	26
Su	1998	Taiwan	27
Woodhouse	1982	UK	30
Lower	1979	USA	34
Cartwright	1984	Portugal	47
Horai	1989	Japan	51
Jaskula-Sztul	2001	Poland	56
Lower	1979	Denmark	71
Ishizu	1995	Japan	71
Hsieh	1999	Taiwan	73
Dewan	1995	India	77
Giannakopoulos	2002	Greece	89
Kaisary	1987	UK	98
Evans	1983	UK	100
Mittal	2004	India	101
Roots	1989	Germany	102
Hanssen	1985	Germany	105
Kim	2000	Korea	112
Peluso	1998	Italy	114
Lower	1979	Sweden	115
Ladero	1985	Spain	130
Risch	1995	UK	189
Hung	2004	Italy	201
Taylor(whites)	1998	USA	215
Mommsen	1985	UK	228
Okkels	1997	Denmark	254
Tsukino	2004	Japan	325
Brockmöller	1996	Germany	374
Gu	2005	USA	504
García-Closas	Current	Spain	1134

	Cases	Odds ratio (95% CI)	p
All studies (n=31)	5091	1·4 (1·2–1·6)	<0·0001
Studies of white populations (n=22)	4216	1·4 (1·3–1·5)	<0·0001
Europe (n=18)	3437	1·4 (1·3–1·6)	<0·0001
USA (n=4)	779	1·1 (0·8–1·6)	0·58
Studies of Asian populations (n=6)	659	1·5 (0·8–2·6)	0·20

Figure 2: **Meta-analysis of studies of *NAT2* slow-acetylation genotype and bladder-cancer risk**
Numbers of cases are individuals with *NAT2* information.

GSTM1 deletion increases bladder cancer risk by 50%

OR=1.5 95%CI (1.3-1.6)



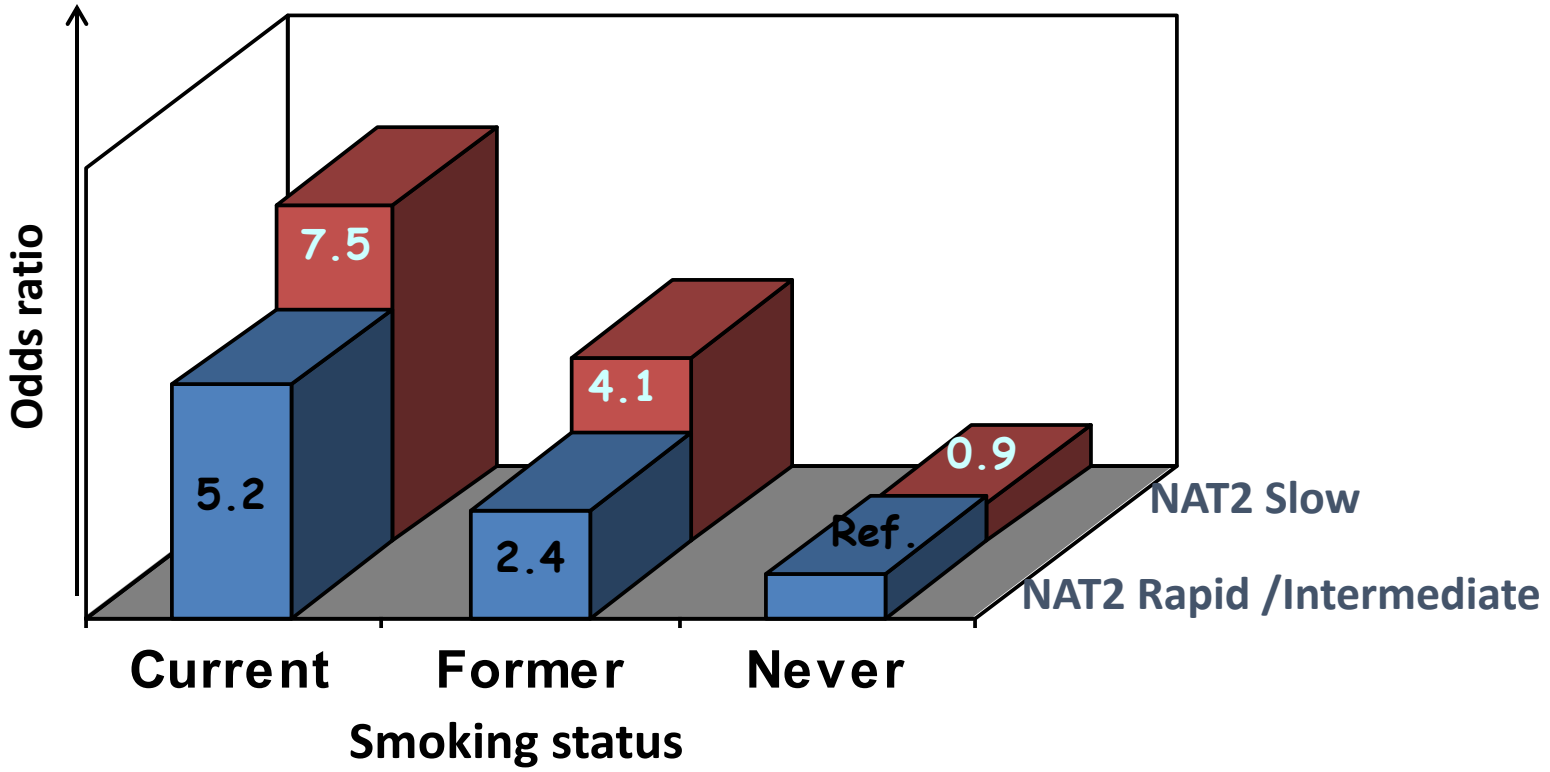
Study	Year	Country	Cases
Romkes	Unpublished	USA	13
Bell (blacks)	1993	USA	16
Heckbert	1992	USA	29
Daly	1993	UK	53
Mungan	2000	Netherlands	61
Šalagovic	1999	Slovakia	76
Georgiou	2000	Greece	89
Lin (whites)	1994	USA	89
Chern	1994	UK	95
Zhong	1993	UK	97
Aktas	2001	Turkey	103
Srivastava	2004	India	106
Moore	2004	Argentina	114
Törüner	2001	Turkey	121
Jeong	2003	Korea	126
Peluso	2000	Italy	130
Steinhoff	2000	Germany	135
Schnakenberge	2000	Germany	157
Kang	1999	Korea	174
Hung	2004	Italy	201
Bell (whites)	1993	USA	213
Kim	2002	Korea	216
Lee	2002	Korea	232
Okkels	1996	Denmark	234
Tsukino	2004	Japan	325
Karagas	2005	USA	354
Brockmöller	1996	Germany	374
García-Closas	Current	Spain	1139

	Cases	Odds ratio (95% CI)	p
All studies (n=28)	5072	1·5 (1·3–1·6)	<0·0001
Studies of predominantly white populations (n=18)	3539	1·4 (1·2–1·6)	<0·0001
Europe (n=13)	2841	1·5 (1·2–1·7)	<0·0001
USA (n=5)	698	1·3 (1·1–1·5)	0·02
Studies of Asian populations (n=6)	1073	1·4 (1·2–1·7)	<0·0001

Figure 4: **Meta-analysis of studies of *GSTM1* null genotype and bladder-cancer risk**
Number of cases for studies in Engel et al[8] are based on table 1 of that paper.

NAT2 slow acetylators are at higher risk of developing bladder cancer from smoking

Joint effect of smoking and NAT2 acetylation on bladder cancer risk: Spanish Bladder Cancer Study



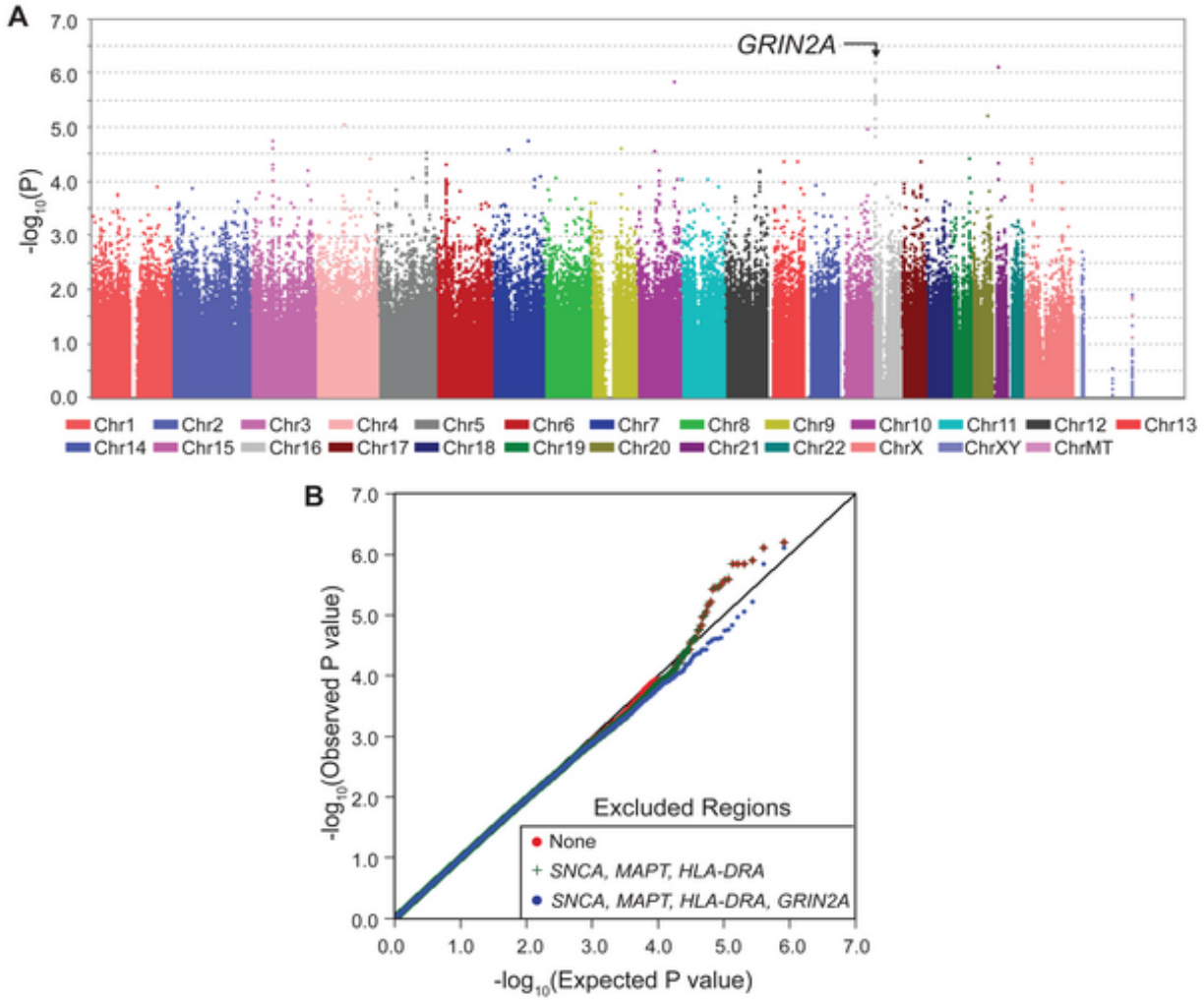
Garcia-Closas M et al. Lancet, 2005; 366: 649-658.

Coffee, *GRIN2A* and Parkinson's Disease?

- Coffee shown to be inversely associated with PD in observational studies (though not all benefit equally)
- Conducted GWAS (>800,000 SNPs; agnostic)
- 1,458 persons with PD and 931 without PD from the NeuroGenetics Research Consortium (NGRC),
- *GRIN2A* as a novel PD modifier gene. *GRIN2A* encodes a subunit of the NMDA-glutamate-receptor which is well known for regulating excitatory neurotransmission in the brain and for controlling movement and behavior.
- Proof of concept that inclusion of environmental factors can help identify genes that are missed in GWAS.

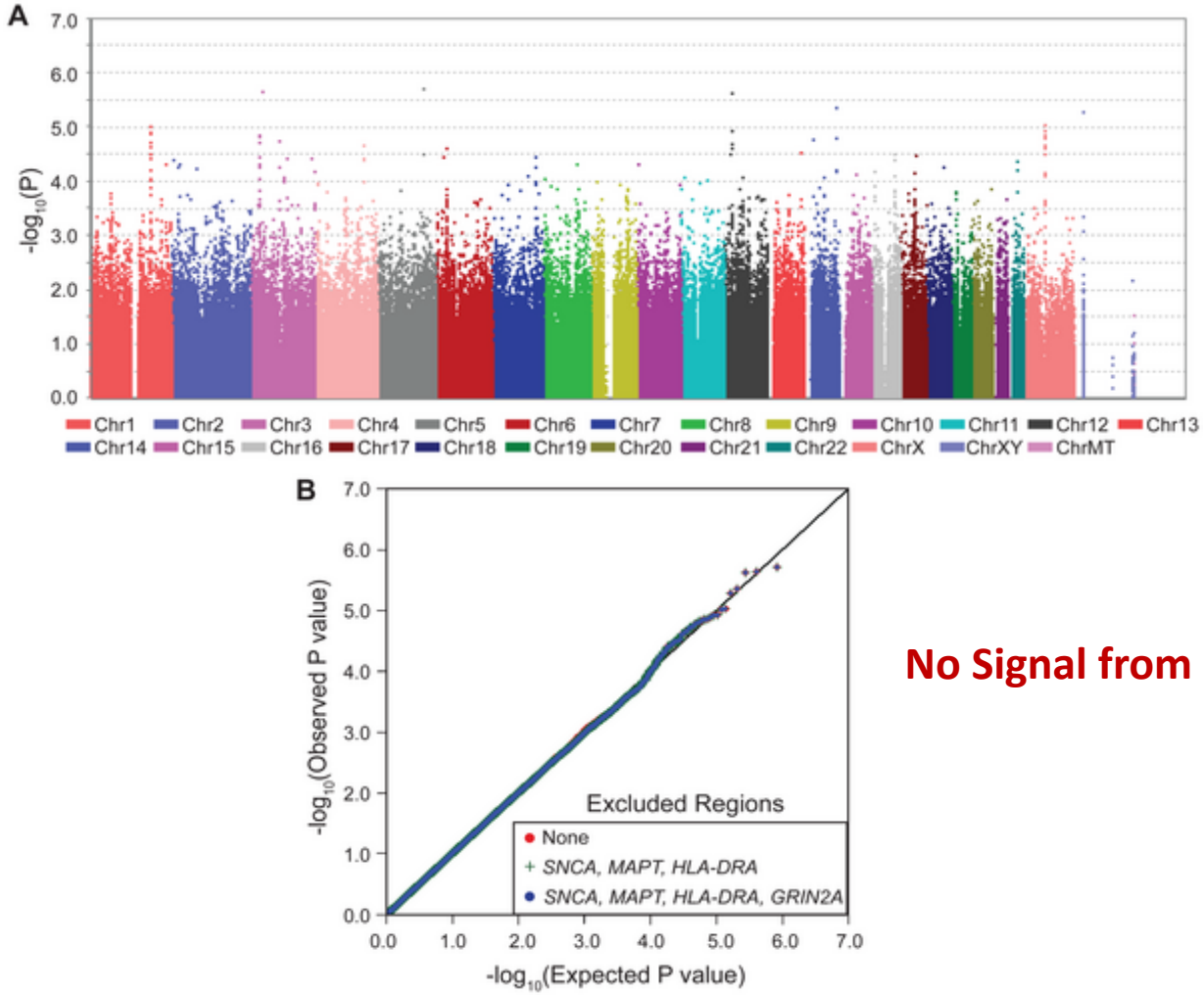
Hamza TH, et al. (2011) Genome-Wide Gene-Environment Study Identifies Glutamate Receptor Gene GRIN2A as a Parkinson's Disease Modifier Gene via Interaction with Coffee. PLoS Genet 7(8): e1002237.

- **GWAS in heavy coffee-drinkers.**



Hamza TH, Chen H, Hill-Burns EM, Rhodes SL, et al. (2011) Genome-Wide Gene-Environment Study Identifies Glutamate Receptor Gene GRIN2A as a Parkinson's Disease Modifier Gene via Interaction with Coffee. PLoS Genet 7(8): e1002237. doi:10.1371/journal.pgen.1002237
http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1002237

- **GWAS in light coffee-drinkers**



No Signal from GRIN2A

Hamza TH, Chen H, Hill-Burns EM, Rhodes SL, et al. (2011) Genome-Wide Gene-Environment Study Identifies Glutamate Receptor Gene GRIN2A as a Parkinson's Disease Modifier Gene via Interaction with Coffee. PLoS Genet 7(8): e1002237. doi:10.1371/journal.pgen.1002237
http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1002237

BPC3: GxE interaction studies for prostate cancer

American Journal of Epidemiology

DOI: 10.1093/aje/kwr423

Original Contribution

Interactions Between Genome-wide Significant Genetic Variants and Circulating Concentrations of Insulin-like Growth Factor 1, Sex Hormones, and Binding Proteins in Relation to Prostate Cancer Risk in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium

Konstantinos K. Tsilidis*, Ruth C. Travis, Paul N. Appleby, Naomi E. Allen, Sara Lindstrom, Fredrick R. Schumacher, David Cox, Ann W. Hsing, Jing Ma, Gianluca Severi, Demetrius Albanes, Jarmo Virtamo, Heiner Boeing, H. Bas Bueno-de-Mesquita, Mattias Johansson, J. Ramón Quirós, Elio Riboli, Afshan Siddiq, Anne Tjønneland, Dimitrios Trichopoulos, Rosario Tumino, J. Michael Gaziano, Edward Giovannucci, David J. Hunter, Peter Kraft, Meir J. Stampfer, Graham G. Giles, Gerald L. Andriole, Sonja I. Berndt, Stephen J. Chanock, Richard B. Hayes, and Timothy J. Key

* Correspondence to Dr. Konstantinos K. Tsilidis, Department of Hygiene and Epidemiology, School of Medicine, University of Ioannina, Ioannina 45110, Greece (e-mail: kostas.tsilidis@ceu.ox.ac.uk).

Initially submitted July 13, 2011; accepted for publication October 27, 2011.

Genome-wide association studies (GWAS) have identified many single nucleotide polymorphisms (SNPs) associated with prostate cancer risk. There is limited information on the mechanistic basis of these associations, particularly about whether they interact with circulating concentrations of growth factors and sex hormones, which may be important in prostate cancer etiology. Using conditional logistic regression, the authors compared per-allele odds ratios for prostate cancer for 39 GWAS-identified SNPs across thirds (tertile groups) of circulating concentrations of insulin-like growth factor 1 (IGF-1), insulin-like growth factor binding protein 3 (IGFBP-3), testosterone, androstenedione, androstanediol glucuronide, estradiol, and sex hormone-binding globulin (SHBG) for 3,043 cases and 3,478 controls in the Breast and Prostate Cancer Cohort Consortium. After allowing for multiple testing, none of the SNPs examined were significantly associated with growth factor or hormone concentrations, and the SNP-prostate cancer associations did not differ by these concentrations, although 4 interactions were marginally significant (*MSMB*-rs10993994 with androstenedione (uncorrected $P = 0.008$); *CTBP2*-rs4962416 with IGFBP-3 (uncorrected $P = 0.003$); 11q13.2-rs12418451 with IGF-1 (uncorrected $P = 0.006$); and 11q13.2-rs10896449 with SHBG (uncorrected $P = 0.005$)). The authors found no strong evidence that associations between GWAS-identified SNPs and prostate cancer are modified by circulating concentrations of IGF-1, sex hormones, or their major binding proteins.

BPC3: GxE interaction studies for prostate cancer

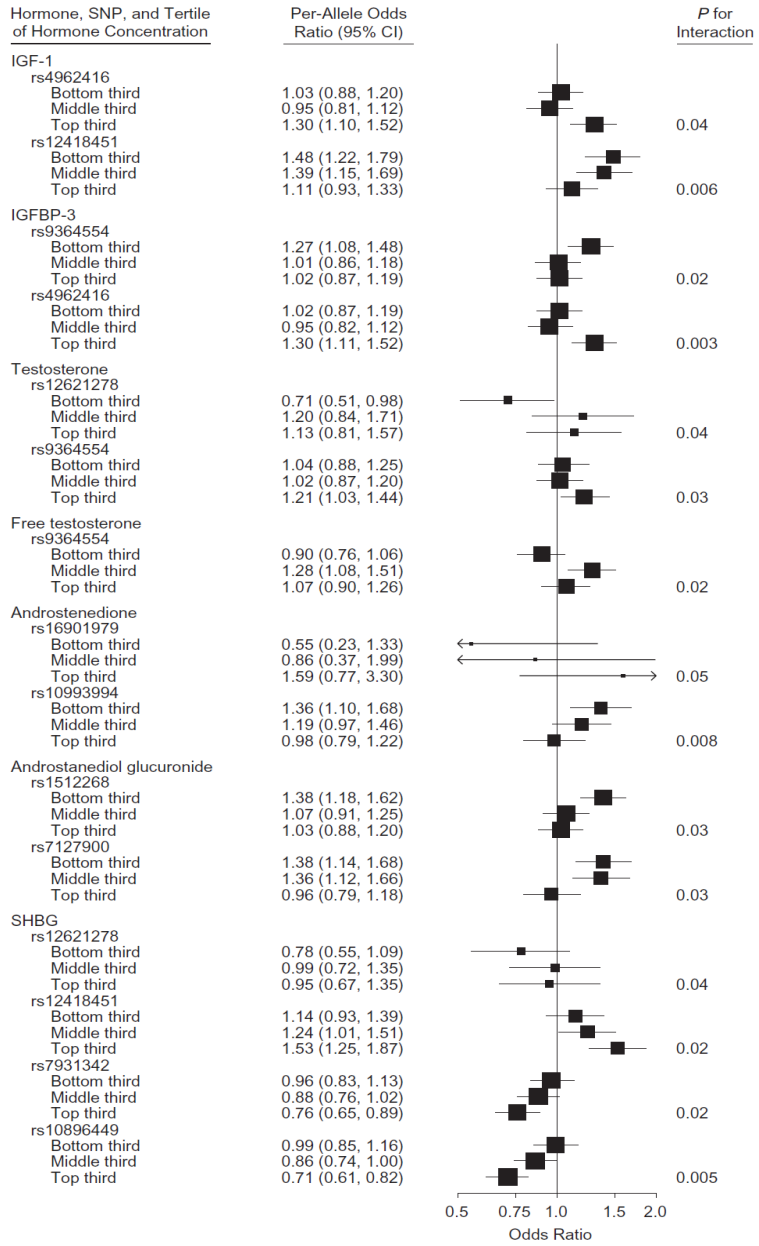


Figure 1. Per-allele associations between single nucleotide polymorphisms (SNPs) identified in genome-wide association studies and risk of prostate cancer, according to circulating concentrations of insulin-like growth factor and steroid sex hormones, for the 15 nominally significant interactions in the Breast and Prostate Cancer Cohort Consortium. Results were obtained from a conditional logistic regression model using cohort-specific thirds of the hormone concentrations (see Web Table 1), matched for age at blood draw, cohort, and country (within the European Prospective Investigation into Cancer and Nutrition), and adjusted for age at blood draw (years; continuous) and body mass index. The *P* values for interaction were calculated using 1-df likelihood ratio tests based on per-allele odds ratios and a continuous hormone variable. Conventional *P* values are shown; all *P* values were nonsignificant after allowance for multiple testing. Bars, 95% confidence interval (CI). (IGF-1, insulin-like growth factor 1; IGFBP-3, insulin-like growth factor binding protein 3; SHBG, sex hormone-binding globulin).

GxE Interaction: Epigenetics

- The study of reversible heritable changes in gene function that occur without a change in the sequence of nuclear DNA (focus of today's talk has been non-reversible heritable changes….)

- Gene-regulatory information that is not expressed in DNA sequences but that is transmitted from one generation (of cells or organisms) to the next (e.g. as methylation changes to DNA structure)

- Strongly influenced by environmental exposures such as diet (in utero nutrition etc)

- Likely to influence GxE interactions….future studies may incorporate epigenetic data into GxE estimations….and beyond.

Circulation

ORIGINAL RESEARCH ARTICLE

Genetic Predisposition to High Blood Pressure and Lifestyle Factors

Associations With Midlife Blood Pressure Levels and Cardiovascular Events

Editorial, see p 662

BACKGROUND: High blood pressure (BP) is a major risk factor for cardiovascular diseases (CVDs), the leading cause of mortality worldwide. Both heritable and lifestyle risk factors contribute to elevated BP levels. We aimed to investigate the extent to which lifestyle factors could offset the effect of an adverse BP genetic profile and its effect on CVD risk.

METHODS: We constructed a genetic risk score for high BP by using

Raha Pazoki, MD, PhD
Abbas Dehghan, MD, PhD
Evangelos Evangelou, PhD
Helen Warren, PhD
He Gao, PhD
Mark Caulfield, MD, PhD
Paul Elliott, MD, PhD
Ioanna Tzoulaki, PhD

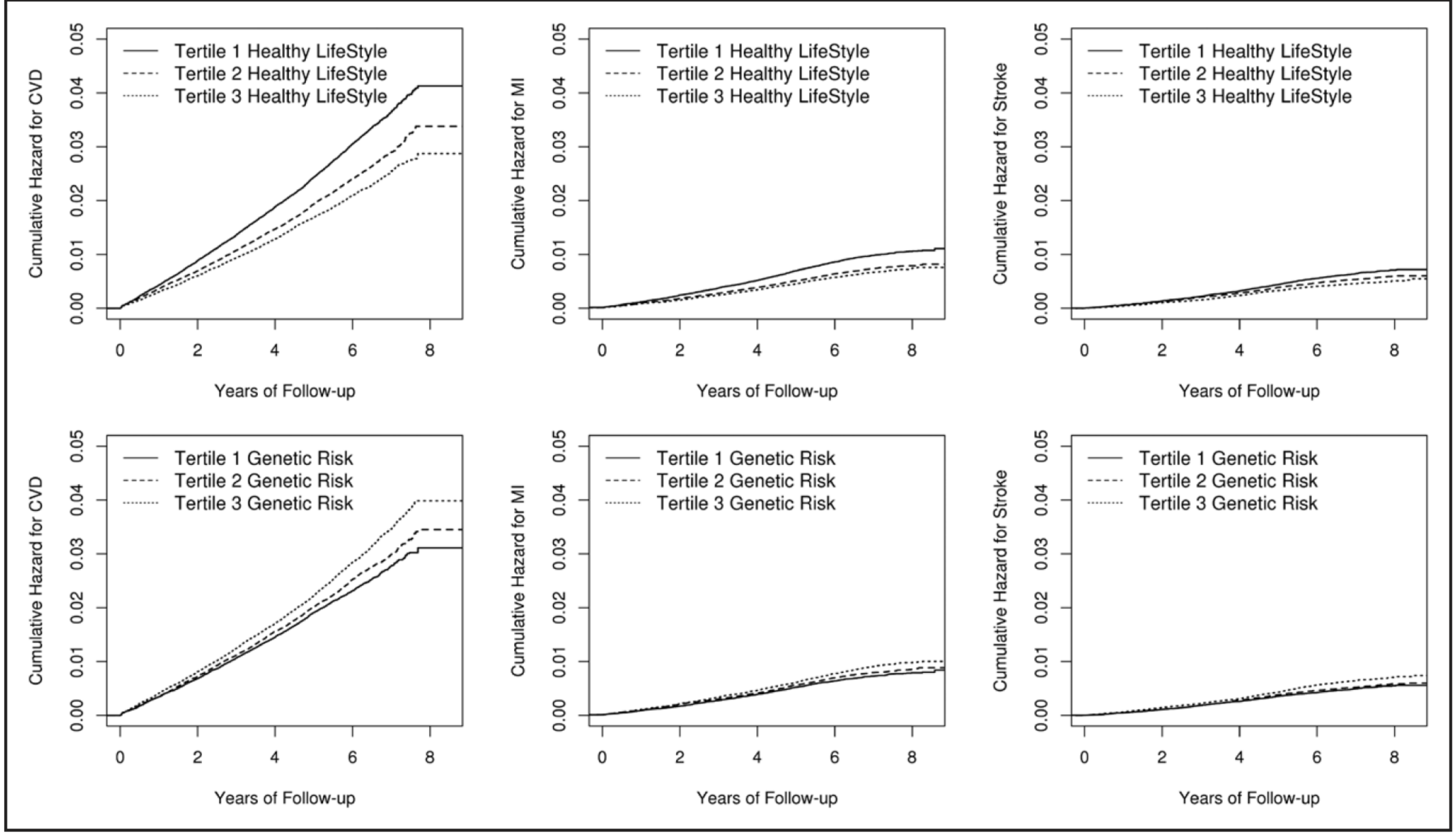
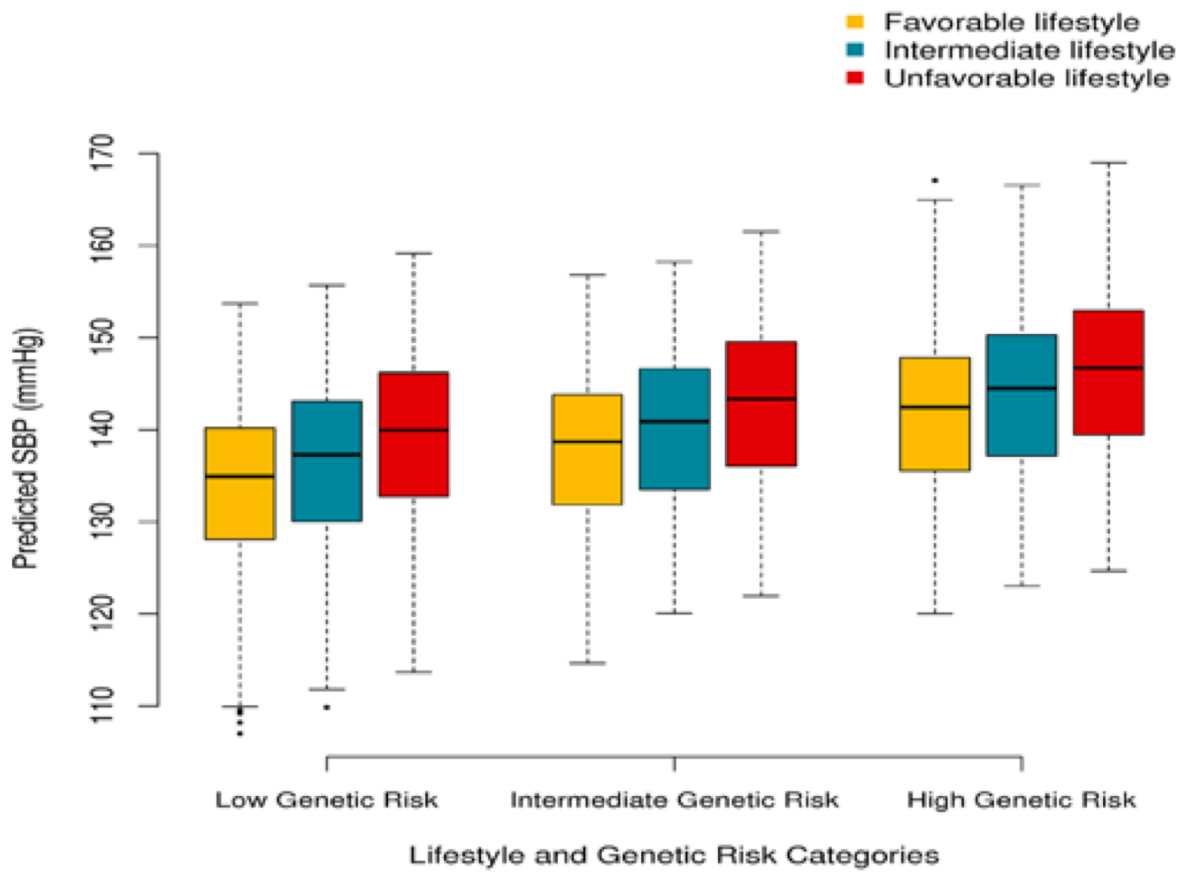


Figure 1. Cumulative hazard rates according to genetic and lifestyle risk tertiles in the UK Biobank study.
The graphs compare different tertiles of genetic risk and lifestyle risk for hazard of CVD (left-hand graphs), myocardial infarction (middle graphs), and stroke (right-hand graphs) (see Table II in online-only Data Supplement for definition of CVD). Cox regression models were adjusted for age and sex. CVD indicates cardiovascular disease; and MI, myocardial infarction.

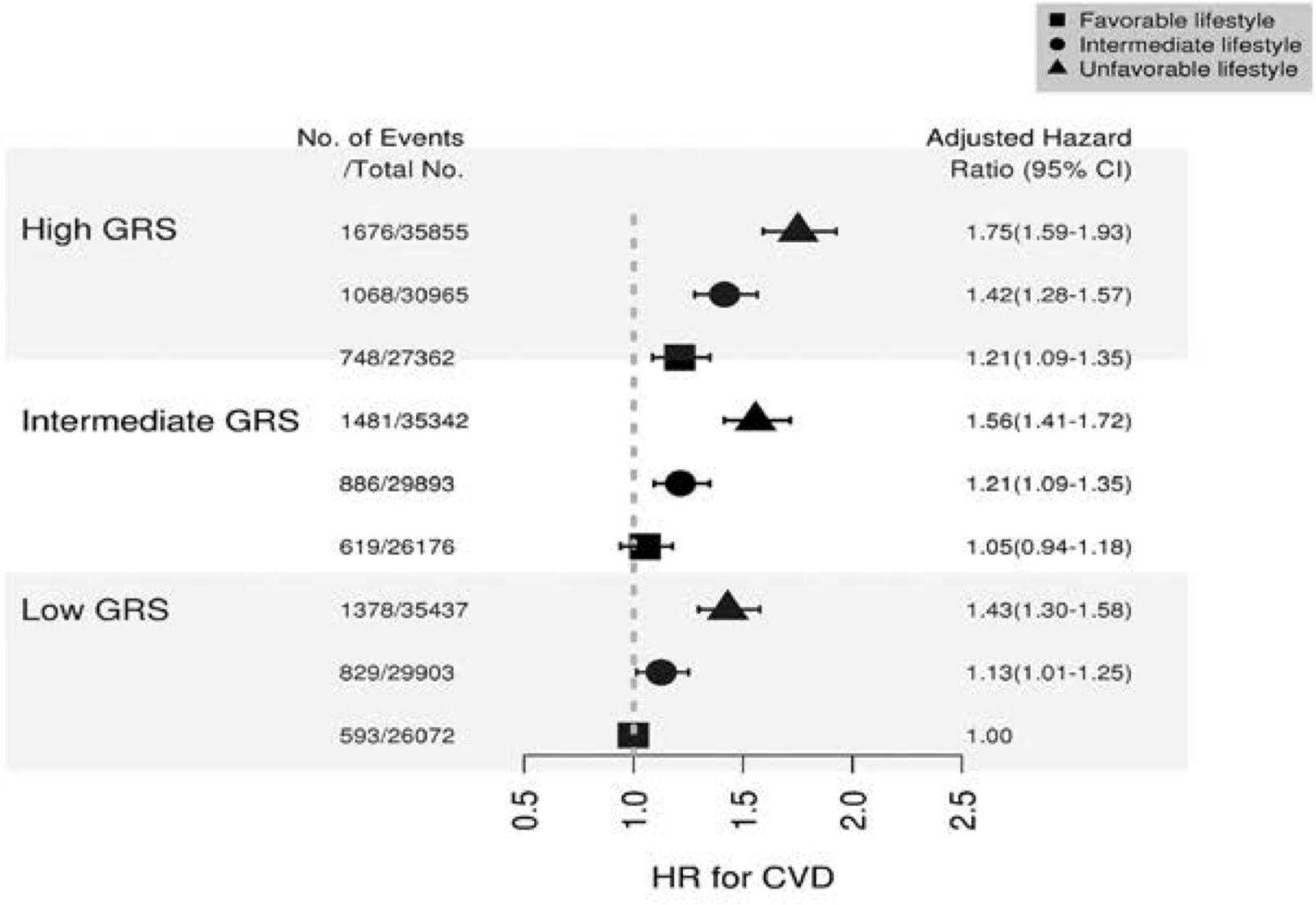
Genetic Predisposition to High Blood Pressure and Lifestyle Factors

Associations With Midlife Blood Pressure Levels and Cardiovascular Events



Genetic Predisposition to High Blood Pressure and Lifestyle Factors

Associations With Midlife Blood Pressure Levels and Cardiovascular Events



Reading list

- Thomas D. Gene—Environment-wide association studies: emerging approaches. Nat Rev Genet 2010;11(4):259-272.
- Population health aspects of genetic epidemiology: genomic profiling, personalised medicine and Mendelian randomization. In: Palmer LJ, Burton P, Davey Smith G. An Introduction to Genetic Epidemiology. Health & Society Series, 2011. pp. 175-217.
- Chapter 7; Genetic Epidemiology
- Gusareva et al. Practical aspects of genome-wide association interaction analysis. Hum Genet (2014) 133:1343–1358
- Torkamani et al. The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics* (2018) doi:10.1038/s41576-018-0018-x