Human Genome Epidemiology: building the knowledge base for genetic variation and human health

John P.A. Ioannidis

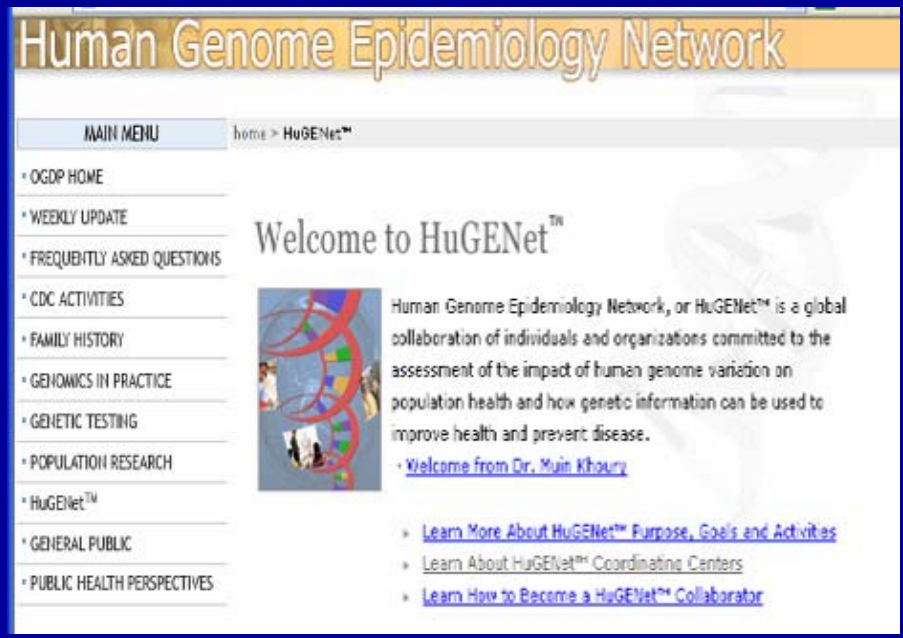
Department of Hygiene and Epidemiology

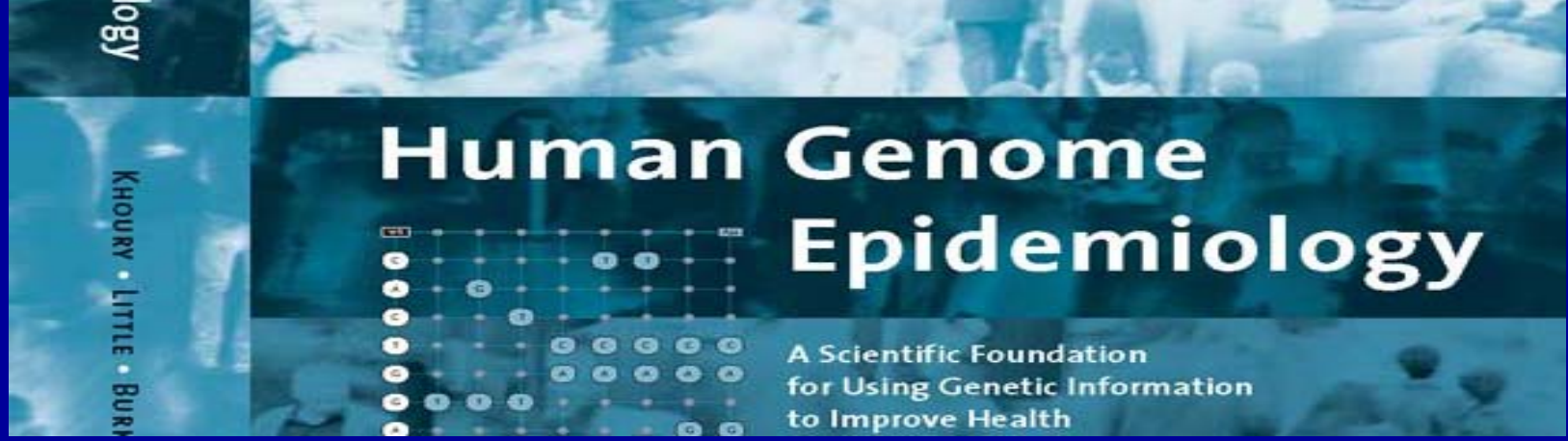
University of Ioannina School of Medicine, Ioannina, Greece

and Tufts University School of Medicine, Boston, USA

Human Genome Epidemiology Network (HuGENet)

- Global collaboration of individuals and organizations to assess population impact of genomics and how it can be used to improve health and prevent disease





"Systematic application of epidemiologic methods and approaches to assess the impact of human genetic variation on health and disease"

Khoury, Little and Burke, HuGE 2004

- Genotype prevalence
- Gene - disease association
- Gene - gene interactions
- Gene - environment interactions
- Assessment of Genetic tests

HuGE problem: 25,000 genes, their combinations and interactions with risk factors

From Genetics to Genomics

- Genetic Disorders
- **Mendelian Disorders**
- **Disease burden: 5%**
- **Mutations/One Gene**
- **High Disease Risk**
- **Environment +/-**
- **“Genetic Services”**

- Genetic Information
- **All Diseases**
- **Disease Burden: 95%**
- **Variants/MultiGenes**
- **Low Disease Risk**
- **Environment +++**
- **General Practice**

Human genome epidemiology: major challenges and evolving status

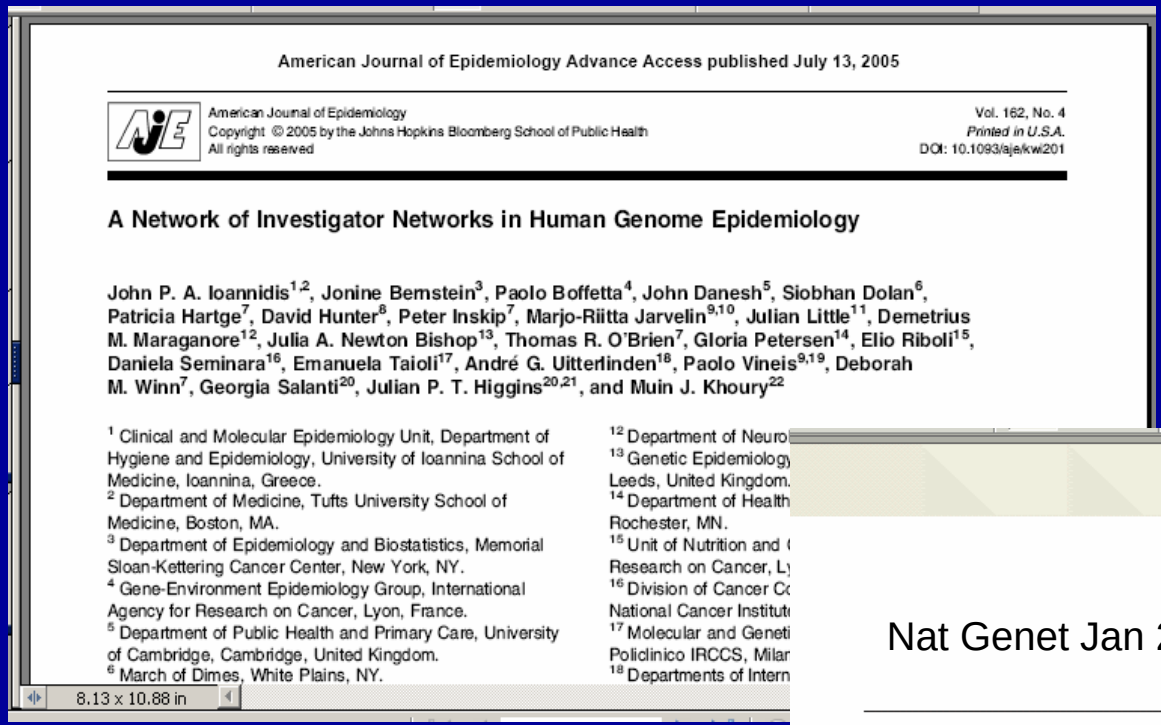
- Small sample sizes:
- Small effect sizes:
- Large number of biological factors:
- Interactions of genes:
- Questionable replication:
- Genuine variability across populations:
- Old-epidemiology problems - confounding (population stratification), misclassification
- Modifiable environment:

- *Can solve with consortia*
- *Have to live with it*
- *Better with current platforms*
- *Still difficult/impossible*
- *We are doing better (no?)*
- *Interesting to learn about*
- *Still with us, but design and reporting are hopefully improving*
- *Working on it*

Sample sizes

- Genetic epidemiology has evolved within a decade from a discipline of case series or case control studies of a few dozen participants to the accrual of large-scale teams and consortia of many teams including many thousands of participants

HuGENet "Network of Networks"



American Journal of Epidemiology Advance Access published July 13, 2005

American Journal of Epidemiology
Copyright © 2005 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved

Vol. 162, No. 4
Printed in U.S.A.
DOI: 10.1093/aje/kwi201

A Network of Investigator Networks in Human Genome Epidemiology

John P. A. Ioannidis[1,2], Jonine Bernstein[3], Paolo Boffetta[4], John Danesh[5], Siobhan Dolan[6], Patricia Hartge[7], David Hunter[8], Peter Inskip[7], Marjo-Riitta Jarvelin[9,10], Julian Little[11], Demetrius M. Maraganore[12], Julia A. Newton Bishop[13], Thomas R. O'Brien[7], Gloria Petersen[14], Elio Riboli[15], Daniela Seminara[16], Emanuela Taioli[17], André G. Uitterlinden[18], Paolo Vineis[9,19], Deborah M. Winn[7], Georgia Salanti[20], Julian P. T. Higgins[20,21], and Muin J. Khoury[22]

[1] Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.
[2] Department of Medicine, Tufts University School of Medicine, Boston, MA.
[3] Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY.
[4] Gene-Environment Epidemiology Group, International Agency for Research on Cancer, Lyon, France.
[5] Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.
[6] March of Dimes, White Plains, NY.
[12] Department of Neuro
[13] Genetic Epidemiology
Leeds, United Kingdom.
[14] Department of Health
Rochester, MN.
[15] Unit of Nutrition and
Research on Cancer, Ly
[16] Division of Cancer C
National Cancer Institut
[17] Molecular and Geneti
Policlinico IRCCS, Milar
[18] Departments of Intern

8.13 x 10.88 in

COMMENTARY

Nat Genet Jan 2006

A road map for efficient and reliable human genome epidemiology

John P A Ioannidis[1,2], Marta L Gwinn[3], Julian Little[4], Julian P T Higgins[5,6], Jonine L Bernstein[7], Paolo Boffetta[8], Melissa Bondy[9], Molly S Bray[10], Paul E Brenchley[11], Patricia A Buffler[12], Juan Pablo Casas[13], Anand Chokkalingam[12], John Danesh[14], George Davey Smith[15], Siobhan Dolan[16], Ross Duncan[17], Nelleke A Gruis[18], Patricia Hartge[19], Mia Hashibe[8], David Hunter[20], Marjo-Riitta Jarvelin[21,22], Beatrice Malmer[23], Teri Manolio[24], Demetrius M Maraganore[25], Julia A Newton-Bishop[26], Thomas R O'Brien[19], Gloria Petersen[27], Elio Riboli[8], Georgia Salanti[1,5], Daniela Seminara[28], Liam Smeeth[13], Emanuela Taioli[29], Nic Timpson[15], Andre G Uitterlinden[30], Paolo Vineis[20,31], Nick Wareham[32], Deborah M Winn[28], Ron Zimmern[6], Muin J Khoury[3] & the Human Genome Epidemiology Network and the Network of Investigator Networks

Networks of investigators have begun sharing best practices, tools and methods for analysis of associations between genetic variation and common diseases. A Network of Investigator Networks has been set up to drive the process,

1 of 3

Some examples of consortia

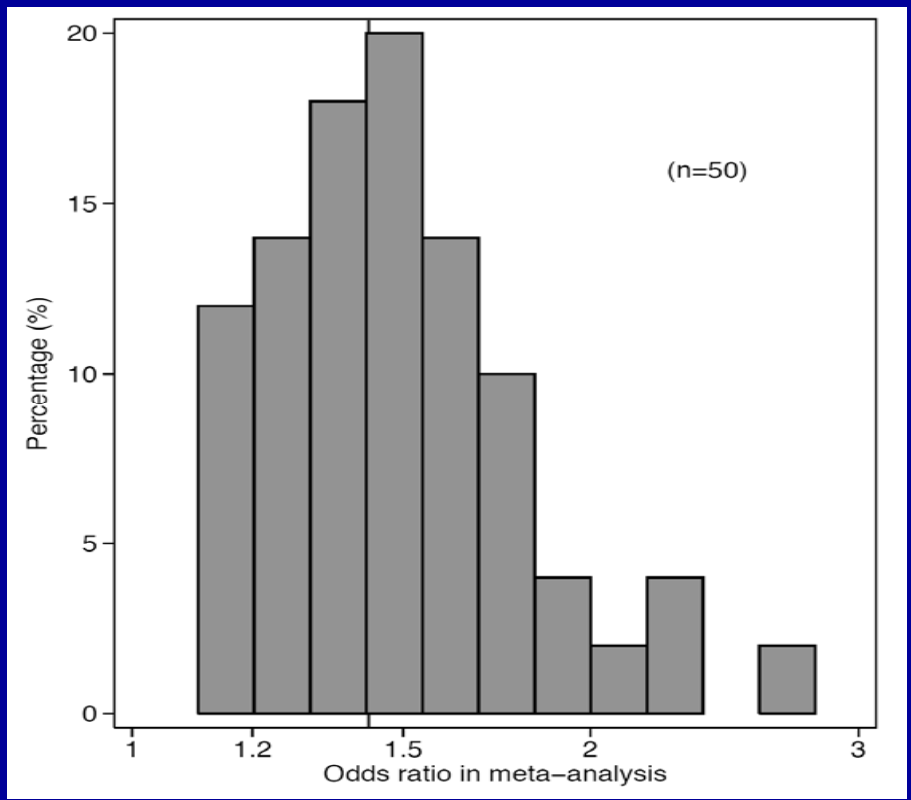
Disease	Consortium	Teams	Participants
Parkinson	GEO-PD	18	10,000
Osteoporosis	GEFOS	40	133,000
Osteoarthritis	TREAT-OA	20	30,000
Preterm birth	PREGENIA	10	20,000
Lymphoma	INTERLYMPH	15	20,000
Lung cancer	ILLCO	30	51,000
Head & Neck	INHANCE	13	28,000
Melanoma	GENOMEL	12	3,000
Pancreatic Ca	PACGENE	10	5,000

Challenges in setting up consortia

- Assembling teams
- Overall project design
- Harmonization vs standardization
- Outcome definitions and ascertainment
- Risk factor definitions and ascertainment
- Gene selection and measurement of genotypes
- Other biological markers
- Integrating and understanding the environmental variables

Seminara et al. Epidemiology 2007

Genetic risks: quanta of small effects



GENE	Polymorphism	Fixed effects OR (95% CI)
---	rs9300039[a]	1.25 (1.15-1.37)
FTO	rs8050136	1.17 (1.12-1.22)
PPARG	rs1801282	1.14 (1.08-1.20)
CDKAL1	rs10946398[b]	1.12 (1.08-1.16)
SLC30A8	rs13266634	1.12 (1.07-1.16)
CDKN2B	rs564398	1.12 (1.07-1.17)
HHEX	rs5015480-rs1111875	1.13 (1.08-1.17)
KCNJ11	rs5215[c]	1.14 (1.10-1.19)
IGF2BP2	rs4402960	1.14 (1.10-1.18)
CDKN2B	rs10811661	1.20 (1.14-1.25)
TCF7L2	rs7901695[d]	1.37 (1.31-1.43)

Ioannidis, Trikalinos, and Khoury, Am J Epidemiol 2006 and Zeggini et al. Science 2007

Large number of biological risk factors: Counting fish in the sea of gene-disease associations

Multiplier	**Parameter**
>10000000	Gene variants
>1000	Diseases
>10	Outcomes
>10	Subgroups
>10	Genetic contrasts
>10	Investigators
1 quadrillion	**Candidate analyses**

Ioannidis, Trends Mol Med 2003

How many variants are we after?

- Assuming at least 1000 diseases/phenotypes involved
- Estimating typically 20-100 variants for each disease (range 1 to 500)
- Allowing for some genetic-phenotypic overlap (e.g. common variants for many autoimmune diseases), probably we aim for approximately 20,000-50,000 variants in an encyclopedia of common genetic variants for common diseases/phenotypes
- We have covered about 1% so far

Interactions between genes: not a task for computers beyond the basics

12,000,000 interacting variants in all possible combinations means… 10^{2085} analyses

If so, “genome-wide” statistical significance should be claimed at $p=10^{-2087}$

Questionable replication: bias or genuine variability

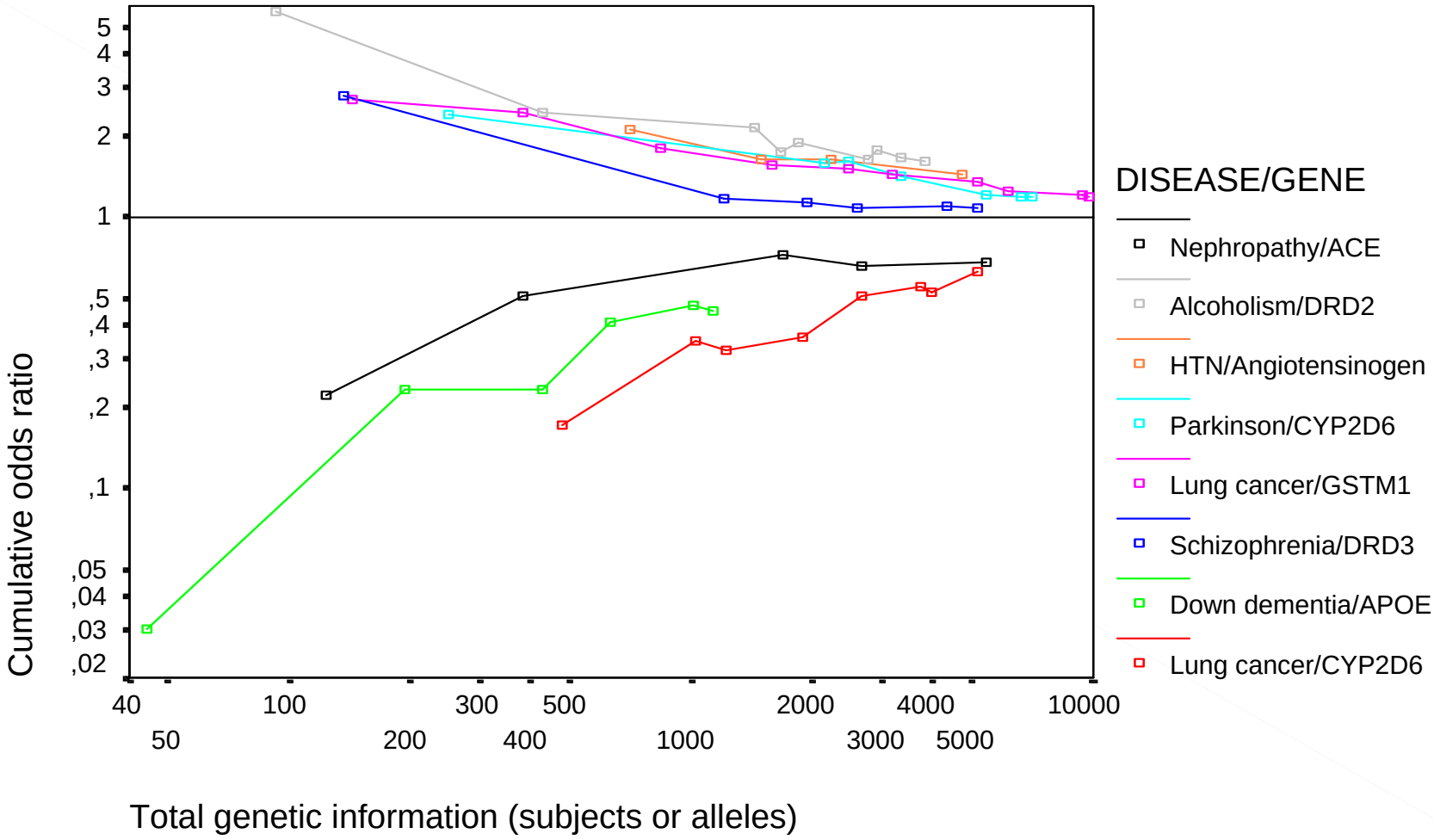
A research finding cannot reach credibility over 50% unless

$u < R$

i.e. bias must be less than the pre-study odds

Ioannidis, PLoS Med 2005

Non-replicated diminishing effects



Ioannidis et al, Nature Genetics 2001

Breast cancer meta-analyses of common variants on candidate genes

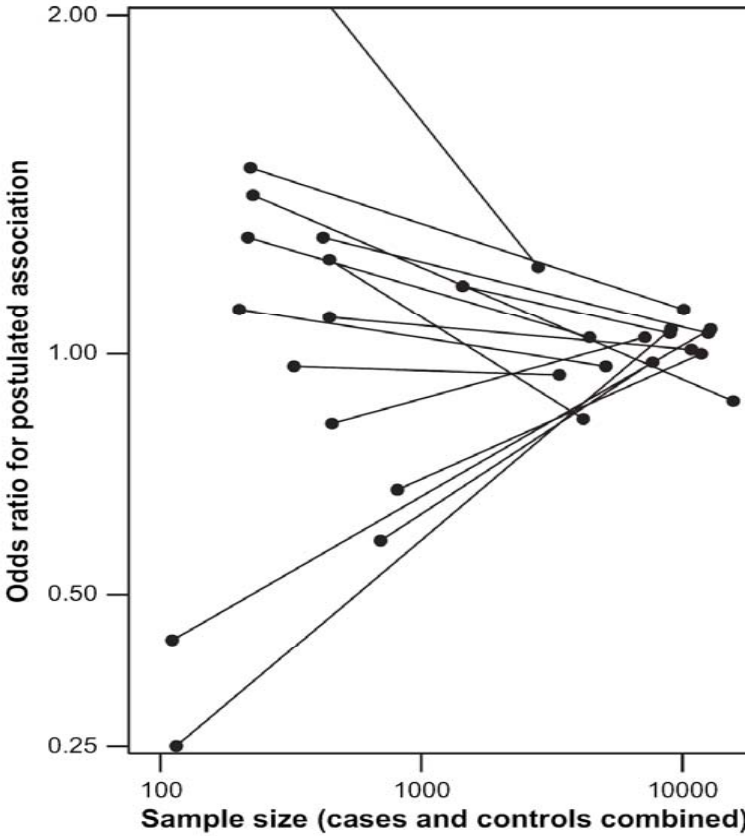
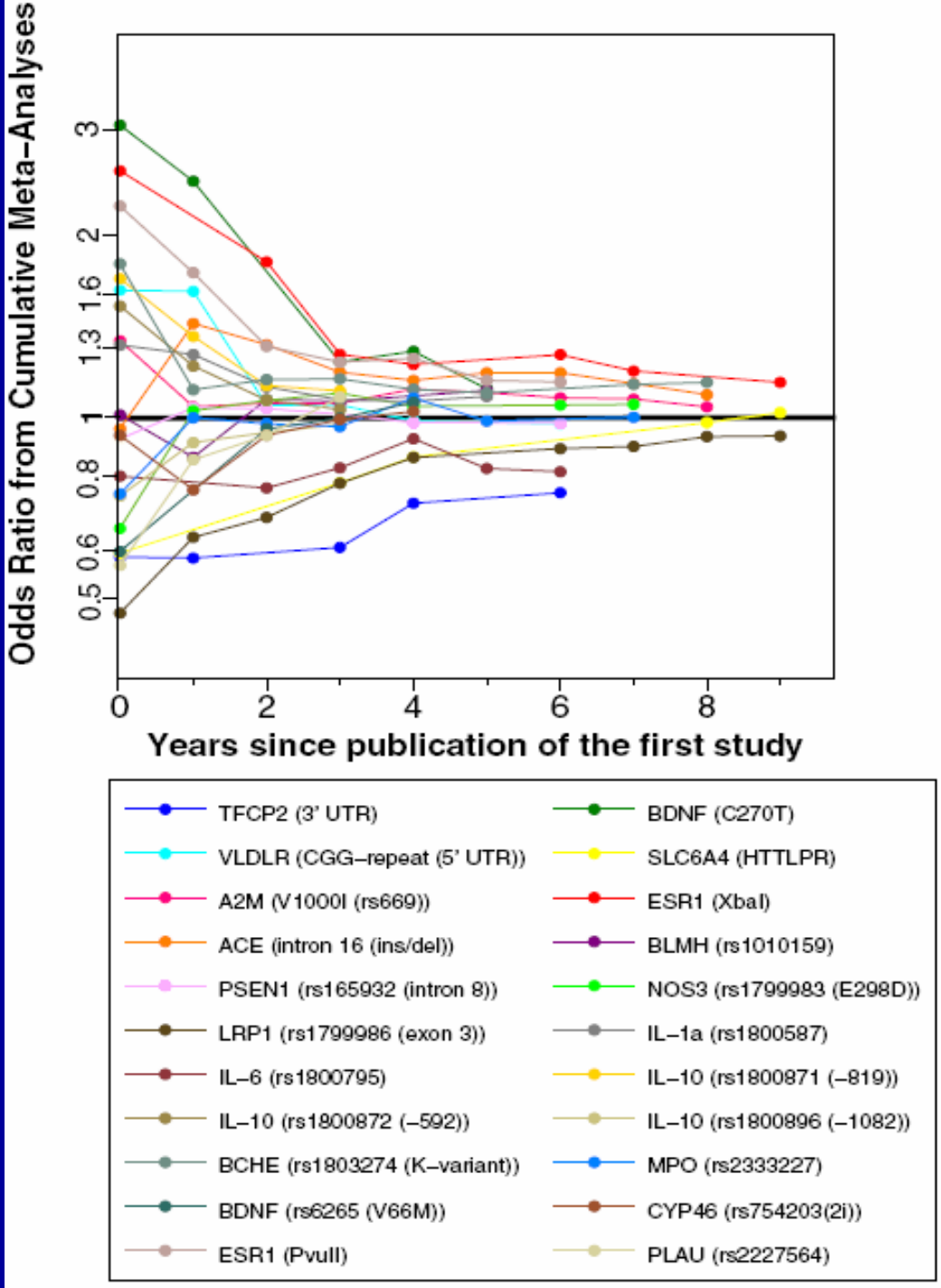
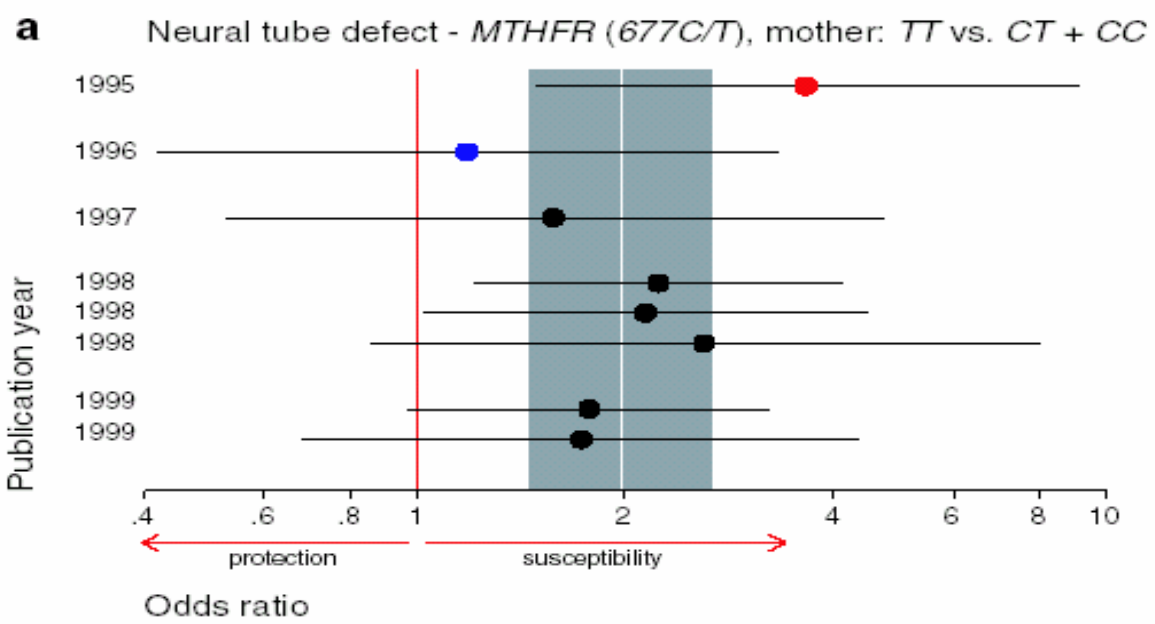


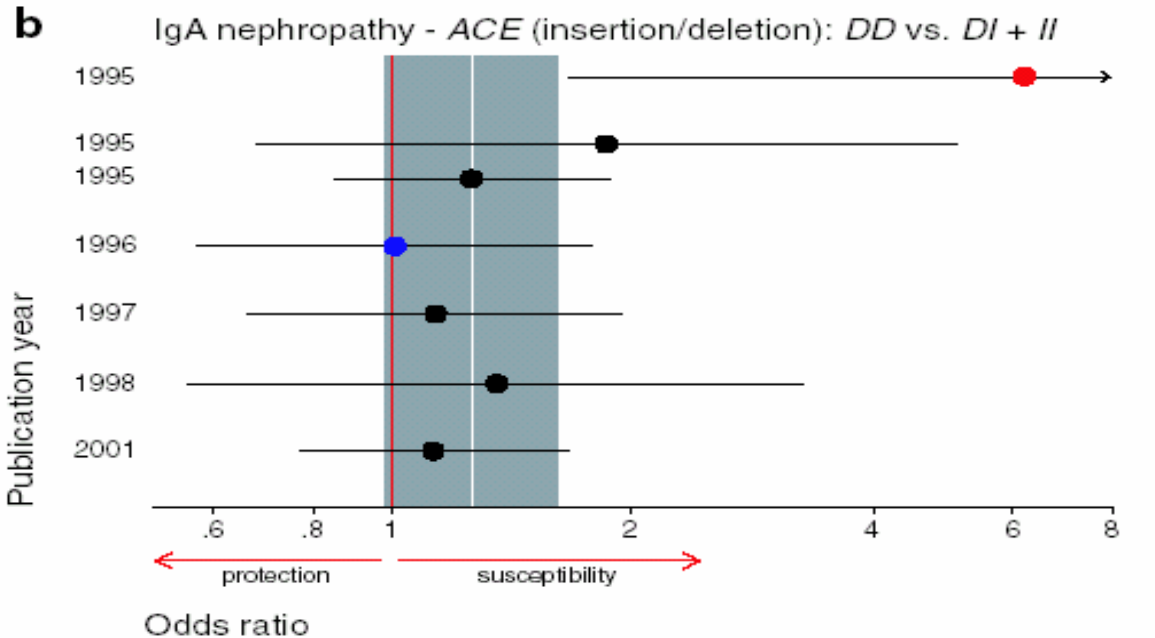
Fig. 1. Meta-analyses of breast cancer candidate gene variants and regression to the null. The 16 hypothesized gene–disease associations listed in Table 1 are shown. For each association, a line connects the results of the first published study (odds ratio [OR] as a function of sample size) with the summary results

Ioannidis, JNCI 2006 and Pharoah et al. JNCI 2006



Heterogeneous meta-analyses with excess of statistically significant single studies in Alzheimer's disease genetics: genuine heterogeneity or bias?





Succession of early extremes: the Proteus phenomenon

Ioannidis and Trikalinos, J Clin Epidemiol 2005

Proteus phenomenon in the GWA era:

13 SNPs proposed for Parkinson's disease in 2-stage GWA

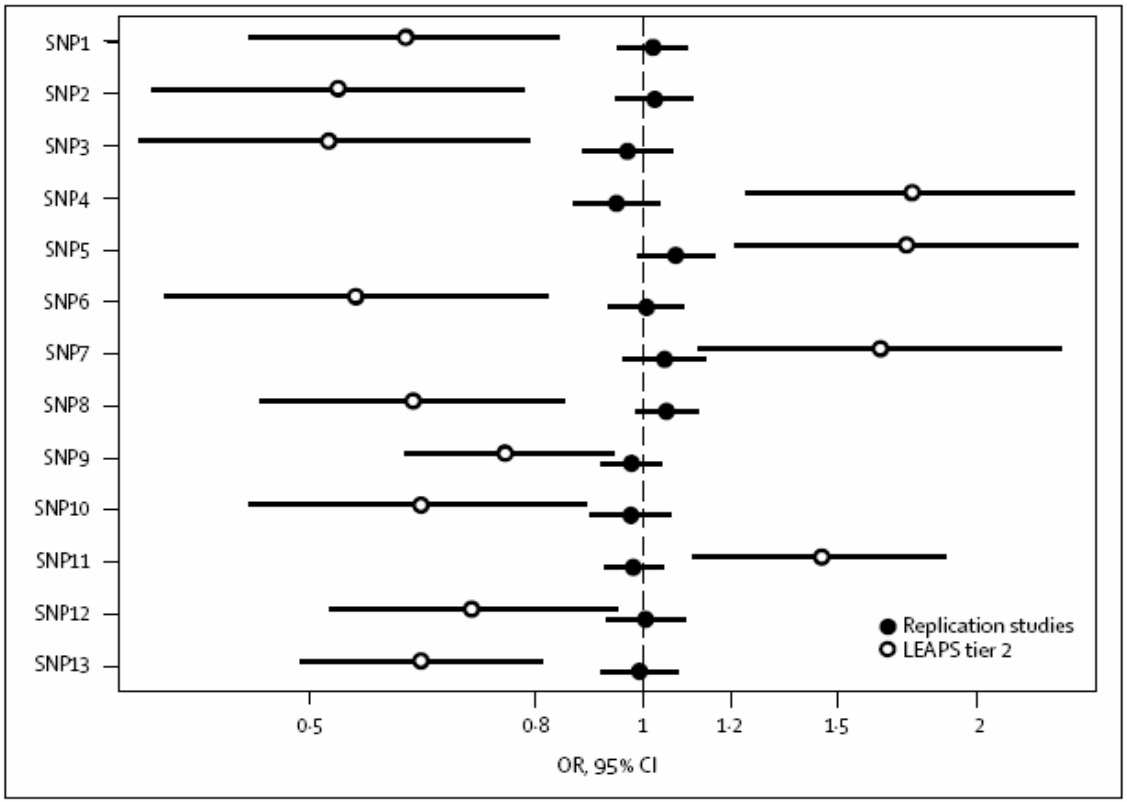


Figure 2: **Tier 2 results from the whole genome-association versus meta-analysis of all replication data**
Point estimates and 95% CIs are shown (random-effects calculations for the meta-analysis). Results are shown for each of the 13 SNPs.

Elbaz et al, Lancet Neurology 2006

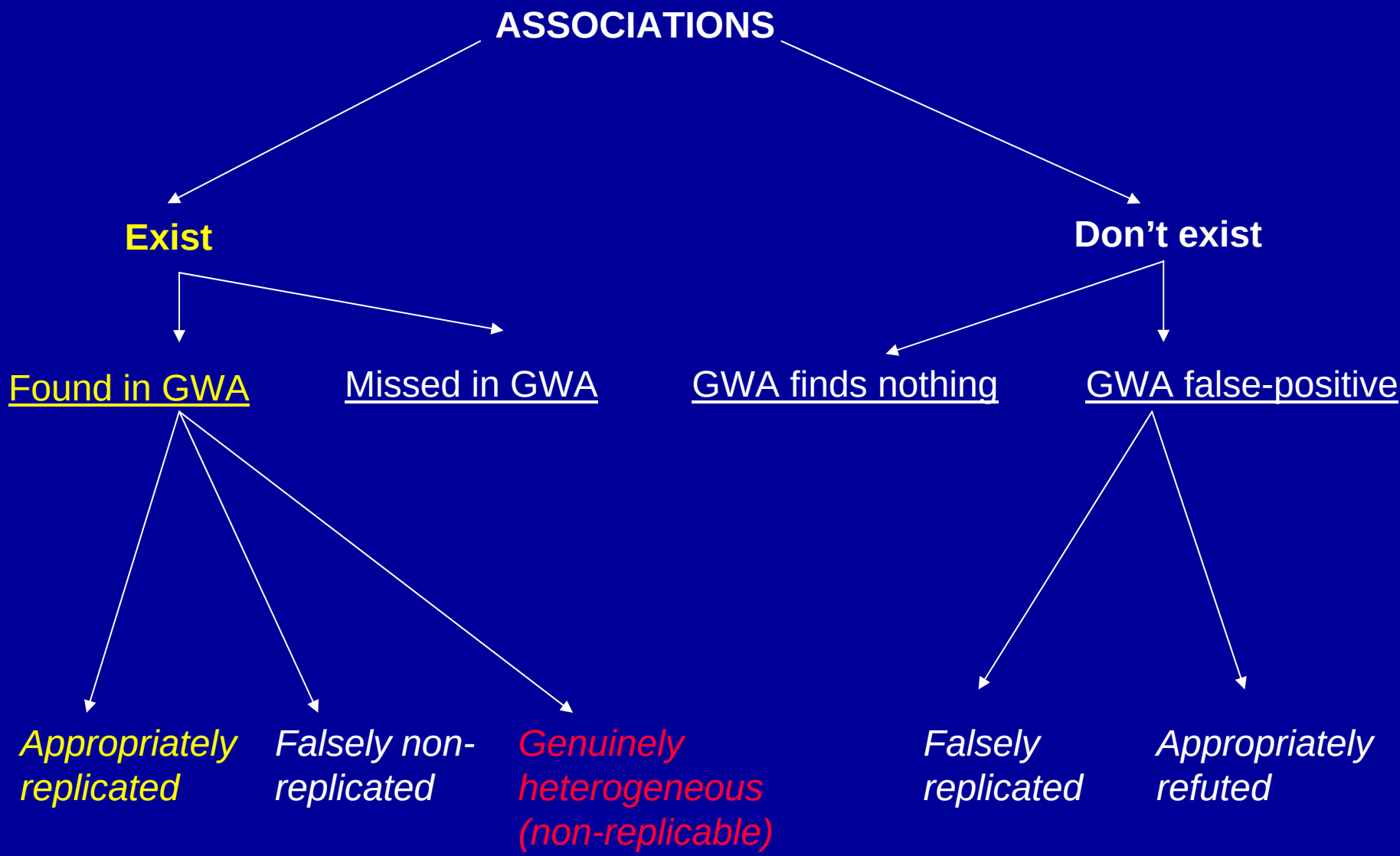
GWA, early replication, and late replication

Definition of replication, non-replication and inconsistency based on meta-analysis considerations

MA including all data		*Without early replication data*		*Status of evidence*
Effect	*Heterogeneity*	*Effect*	*Heterogeneity*	
Yes	No	Yes	No	Replication
Yes	Yes	Yes	No	Replication with winner's curse
Yes	Yes	Yes	Yes	Inconsistency
Yes	No	No	No	Non-replication
Yes	Yes	No	No	Non-replication
Yes	Yes	No	Yes	Non-replication or inconsistency
No	No	No	No	Non-replication
No	Yes	No	No	Non-replication with winner's curse
No	Yes	No	Yes	Non-replication or inconsistency

Ioannidis, Human Heredity 2007

Associations: existing or not, found or not



Potential reasons for genuinely inconsistent findings

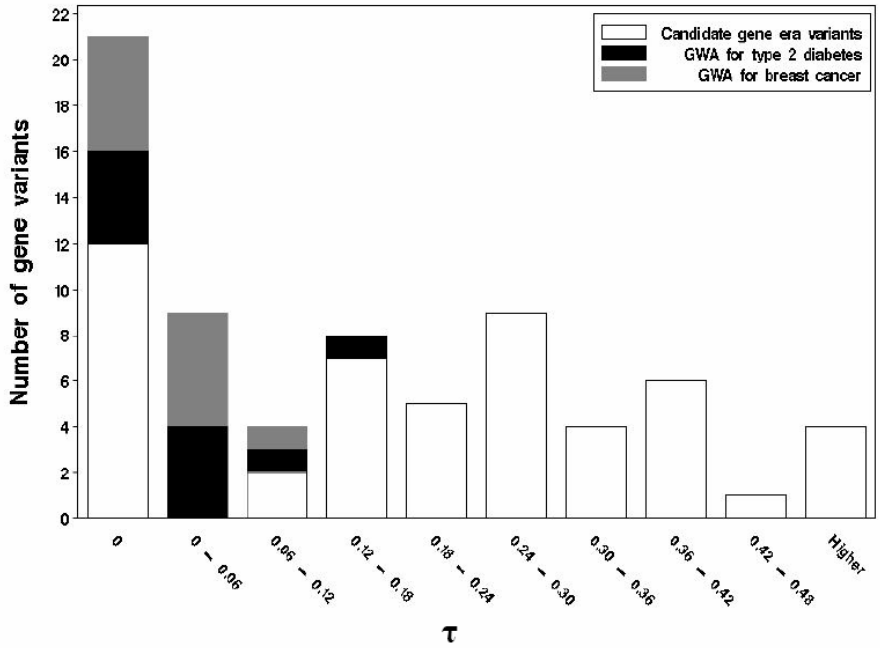
TagSNP with variable linkage disequilibrium across populations

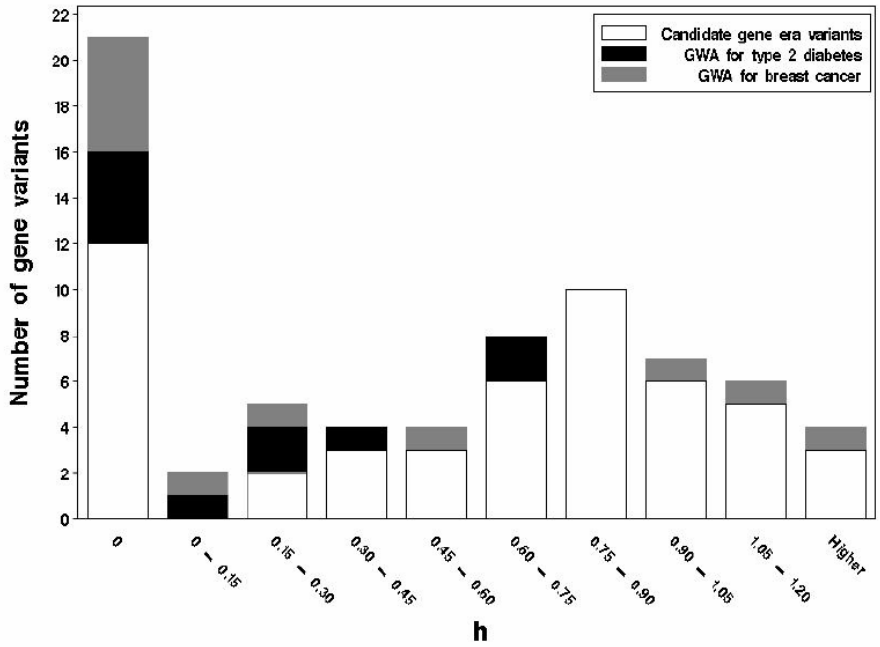
Individual- and population-specific genetic effects

- Independent of other genetic variants and environmental exposures
- Due to epistasis (gene-gene interactions)
- Due to gene-environment interactions

Exchangeable genetic variants and multi-gene signatures thereof

- Functional pathways
- Gene ontology
- Other known or unknown common denominator for genes

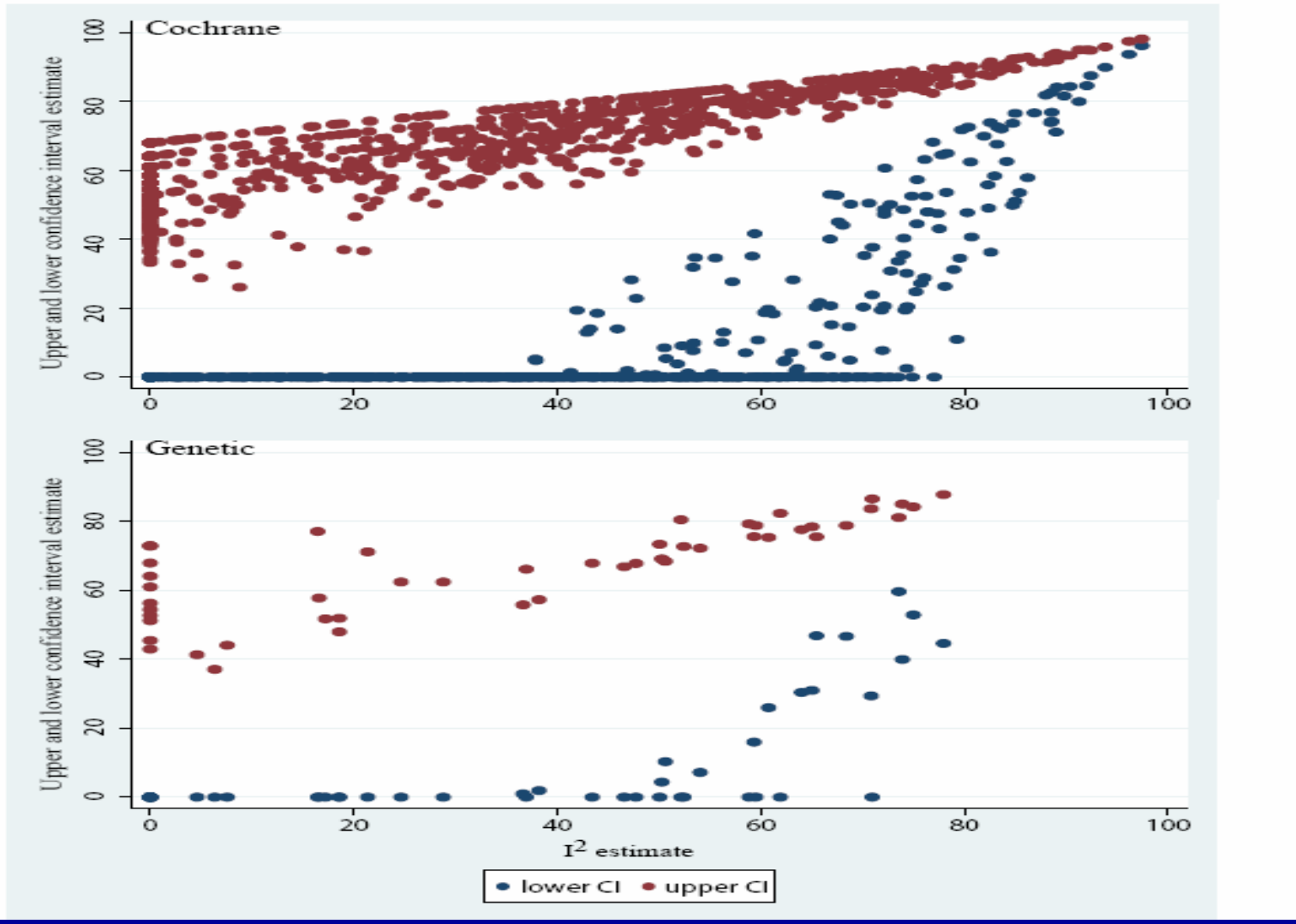




Heterogeneity in candidate gene era and GWA era

Mooneshinghe, et al. PNAS 2008

Uncertainty of I^2 estimates of heterogeneity in meta-analyses



Ioannidis, Patsopoulos, Evangelou, BMJ 2007

OPEN ACCESS Freely available online

PLoS one

Heterogeneity in Meta-Analyses of Genome-Wide Association Investigations

John P. A. Ioannidis[1,2,3]*, Nikolaos A. Patsopoulos[1], Evangelos Evangelou[1]

1 Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, 2 Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece, 3 Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, United States of America

Background. Meta-analysis is the systematic and quantitative synthesis of effect sizes and the exploration of their diversity across different studies. Meta-analyses are increasingly applied to synthesize data from genome-wide association (GWA) studies and from other teams that try to replicate the genetic variants that emerge from such investigations. Between-study heterogeneity is important to document and may point to interesting leads. ***Methodology/Principal Findings.*** To exemplify these issues, we used data from three GWA studies on type 2 diabetes and their replication efforts where meta-analyses of all data using fixed effects methods (not incorporating between-study heterogeneity) have already been published. We considered 11 polymorphisms that at least one of the three teams has suggested as susceptibility loci for type 2 diabetes. The I^2 inconsistency metric (measuring the amount of heterogeneity not due to chance) was different from 0 (no detectable heterogeneity) for 6 of the 11 genetic variants; inconsistency was moderate to very large ($I^2 = 32–77\%$) for 5 of them. For these 5 polymorphisms, random effects calculations incorporating between-study heterogeneity revealed more conservative p-values for the summary effects compared with the fixed effects calculations. These 5 associations were perused in detail to highlight potential explanations for between-study heterogeneity. These include identification of a marker for a correlated phenotype (e.g. *FTO* rs8050136 being associated with type 2 diabetes through its effect on obesity); differential linkage disequilibrium across studies of the identified genetic markers with the respective culprit polymorphisms (e.g., possibly the case for *CDKAL1* polymorphisms or for rs9300039 and markers in linkage disequilibrium, as shown by additional studies); and potential bias. Results were largely similar, when we treated the discovery and replication data from each GWA investigation as separate studies. ***Significance.*** Between-study heterogeneity is useful to document in the synthesis of data from GWA investigations and can offer valuable insights for further clarification of gene-disease associations.

Citation: Ioannidis JPA, Patsopoulos NA, Evangelou E (2007) Heterogeneity in Meta-Analyses of Genome-Wide Association Investigations. PLoS ONE 2(9): e841. doi:10.1371/journal.pone.0000841

Table 1. Between-study heterogeneity and random versus fixed effects calculations for polymorphisms that were considered "confirmed"

GENE	Polymorphism	Q (p)	I^2 (95% CI)	Random effects OR (95% CI)	Fixed effects OR (95% CI)	Random effects p-value	Fixed effects p-value
—	rs9300039[a]	7.98 (0.019)	75% (0–90)	1.25 (1.04–1.50)	1.25 (1.15–1.37)	0.015	4.3×10-7
FTO	rs8050136	8.62 (0.013)	77% (0–91)	1.13 (1.02–1.25)	1.17 (1.12–1.22)	0.015	1.3×10^{-12}
PPARG	rs1801282	3.80 (0.15)	47% (0–84)	1.16 (1.07–1.25)	1.14 (1.08–1.20)	0.0003	1.7×10^{-6}
CDKAL1	rs10946398[b]	3.73 (0.16)	46% (0–84)	1.12 (1.07–1.17)	1.12 (1.08–1.16)	3.2×10^{-6}	4.1×10^{-11}
SLC30A8	rs13266634	2.92 (0.23)	32% (0–81)	1.12 (1.07–1.18)	1.12 (1.07–1.16)	8.7×10^{-6}	5.3×10^{-8}
CDKN2B	rs564398	1.48 (0.48)	0% (0–73)	1.12 (1.07–1.17)	1.12 (1.07–1.17)	1.2×10^{-7}	1.2×10^{-7}
HHEX	rs5015480–rs1111875	0.45 (0.80)	0% (0–73)	1.13 (1.08–1.17)	1.13 (1.08–1.17)	5.7×10^{-10}	5.7×10^{-10}
KCNJ11	rs5215[c]	0.56 (0.76)	0% (0–73)	1.14 (1.10–1.19)	1.14 (1.10–1.19)	5×10^{-11}	5×10^{-11}
IGF2BP2	rs4402960	2.65 (0.27)	25% (0–79)	1.15 (1.10–1.19)	1.14 (1.10–1.18)	6.5×10^{-12}	8.6×10^{-16}
CDKN2B	rs10811661	0.03 (0.99)	0% (0–73)	1.20 (1.14–1.25)	1.20 (1.14–1.25)	7.8×10^{-15}	7.8×10^{-15}
TCF7L2	rs7901695[d]	0.24 (0.89)	0% (0–73)	1.37 (1.31–1.43)	1.37 (1.31–1.43)	1.0×10^{-48}	1.0×10^{-48}

Additive models are presented, as in the main analyses of the original papers. Fixed effects calculations are Mantel-Haenszel estimates as in the original papers. Random effects calculations use the DerSimonian and Laird estimators for the between-study variance.

CI: confidence interval; OR: odds ratio

[a]multi-marker tag in DGI and rs1514823 in the UK study

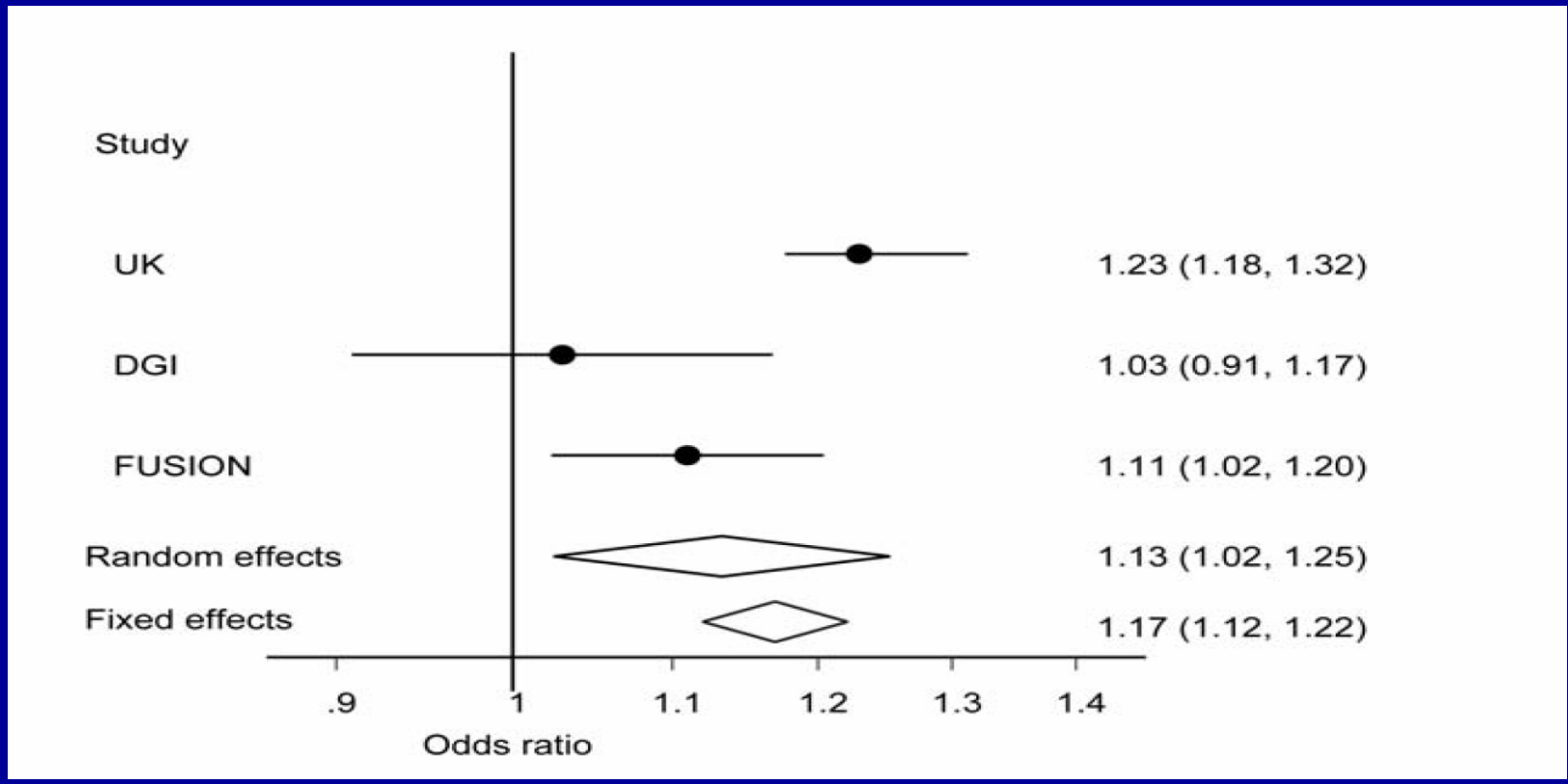
[b]rs7754840 in FUSION

[c]rs5219 in FUSION and DGI

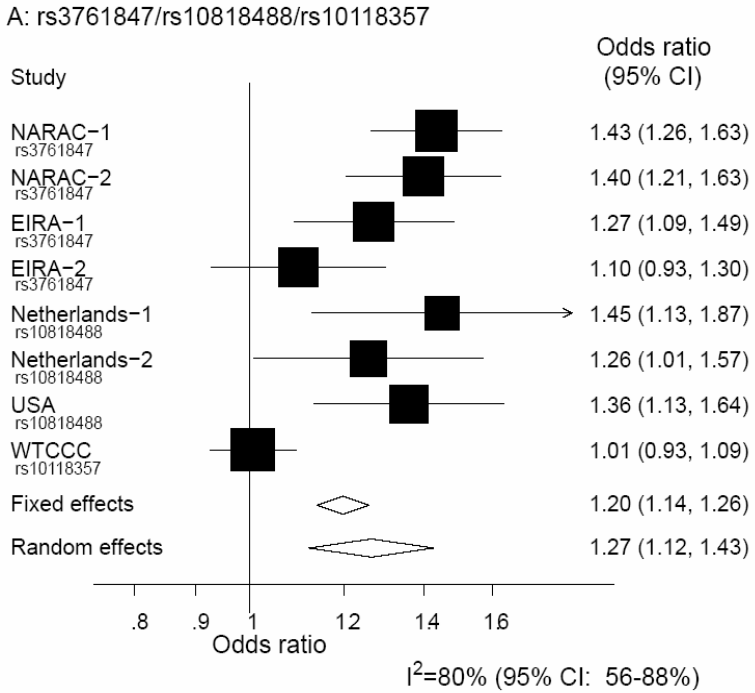
[d]rs7903146 in FUSION and DGI

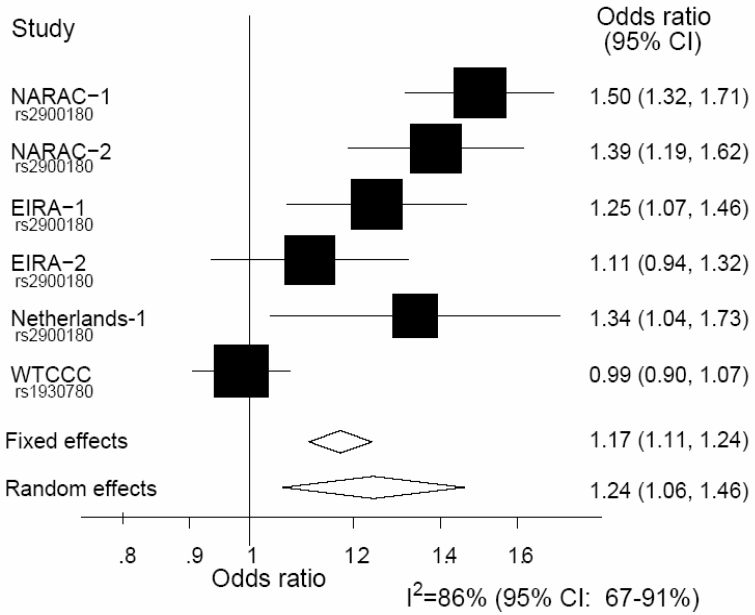
doi:10.1371/journal.pone.0000841.t001

An inconsistent association mirroring a different association: *FTO*, type 2 diabetes, and obesity



Ioannidis, Patsopoulos, Evangelou, PLoS ONE 2007





An inconsistency for rheumatoid arthritis: bias, LD or we still don't know what disease we are after?

Inconsistency and non-replicability threshold

- Inconsistency may be due to either bias or genuine between-study heterogeneity
- Beyond a given threshold of inconsistency, no matter how large studies we conduct, we may never have enough power to replicate an association (non-replicability threshold)
- This means that we need to decrease bias to a minimum so that we have to face only the genuine heterogeneity
- The main question is shifting from whether chance can create an association to whether bias of whatever kind can create an association of the observed magnitude

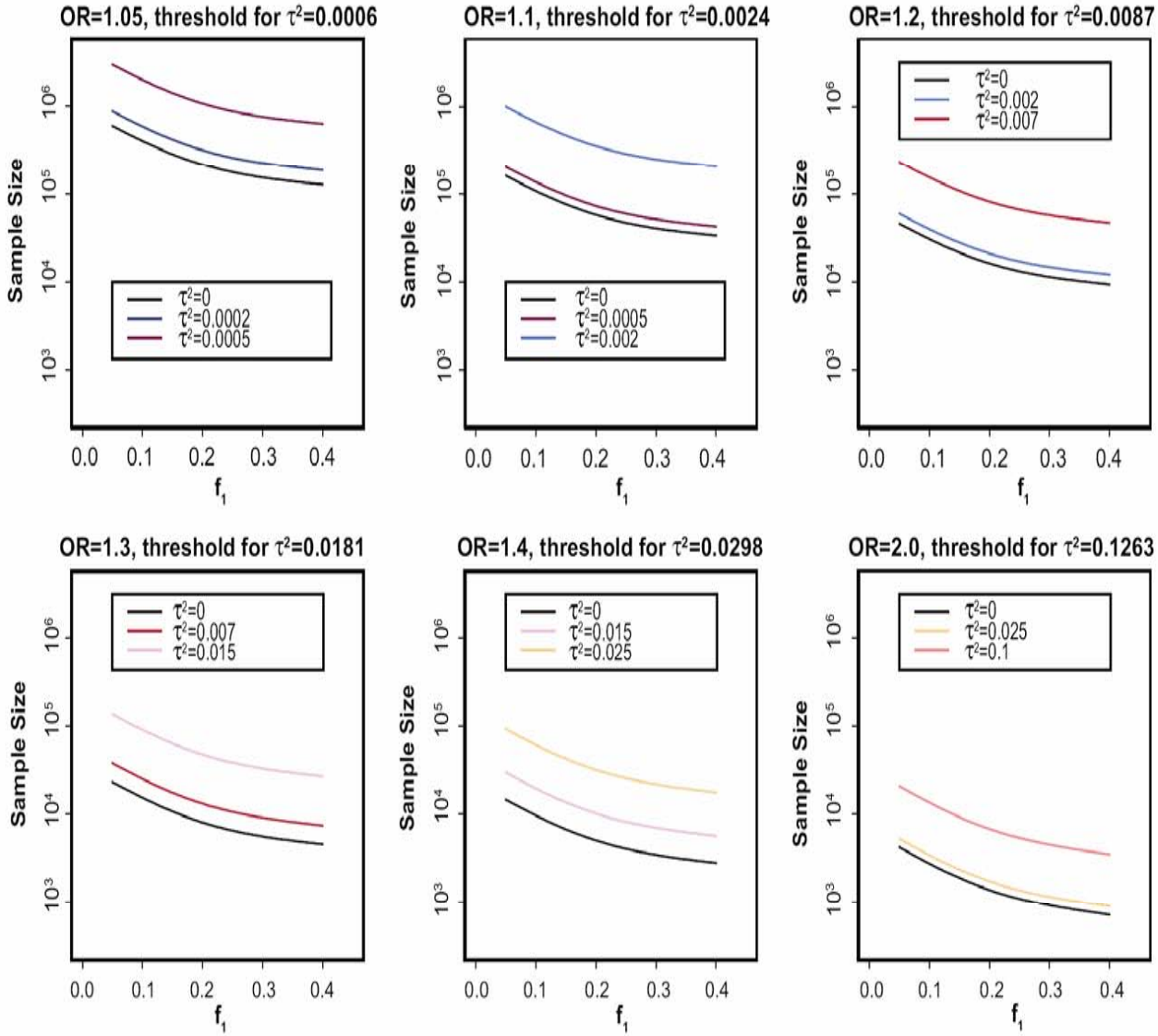


Fig. 2. Mean sample sizes required to detect odds ratios of 1.05, 1.1, 1.2, 1.3, 1.4, and 2.0 with power 80% at $\alpha = 0.0000001$ as a function of genotype frequency f_1 for a metaanalysis of 10 equally large studies.

Mooneshinghe, Khoury, Liu and Ioannidis PNAS 2008

American Journal of Epidemiology

Vol. 166, No. 8
DOI: 10.1093/aje/kwm248
Advance Access publication September 5, 2007

Editorial

Turning the Pump Handle: Evolving Methods for Integrating the Evidence on Gene-Disease Association

Julian P. T. Higgins[1], Julian Little[2], John P. A. Ioannidis[3,5], Molly S. Bray[4], Teri A. Manolio[6], Liam Smeeth[7], Jonathan A. Sterne[8], Betsy Anagnostelis[9], Adam S. Butterworth[10], John Danesh[10], Carol Dezateux[11], John E. Gallacher[12], Marta Gwinn[13], Sarah J. Lewis[8], Cosetta Minelli[14], Paul D. Pharoah[15], Georgia Salanti[3], Simon Sanderson[10], Lesley A. Smith[16], Emanuela Taioli[17], John R. Thompson[18], Simon G. Thompson[1], Neil Walker[19], Ron L. Zimmern[20], and Muin J. Khoury[13]

[1] MRC Biostatistics Unit, Cambridge, United Kingdom.
[2] Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.
[3] Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.
[4] Center for Human Genetics, Institute of Molecular Medicine and School of Public Health, University of Texas, Houston, TX.
[5] Department of Medicine, Tufts University School of Medicine, Boston, MA.
[6] National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.
[7] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom.
[8] Department of Social Medicine, University of Bristol, Bristol, United Kingdom.
[9] Royal Free Hospital Medical Library, University College London, London, United Kingdom.
[10] Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.
[11] Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, London, United Kingdom.
[12] Department of Epidemiology, Cardiff University, Cardiff, Wales, United Kingdom.
[13] National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA.
[14] National Heart and Lung Institute, Imperial College, London, United Kingdom.
[15] Cancer Research UK Human Cancer Genetics Group, Department of Oncology, University of Cambridge, Cambridge, United Kingdom.
[16] School of Health and Social Care, Oxford Brookes University, Oxford, United Kingdom.
[17] University of Pittsburgh Medical Center, Pittsburgh, PA.
[18] Department of Health Sciences, University of Leicester, Leicester, United Kingdom.
[19] Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom.
[20] PHG Foundation, Cambridge, United Kingdom.

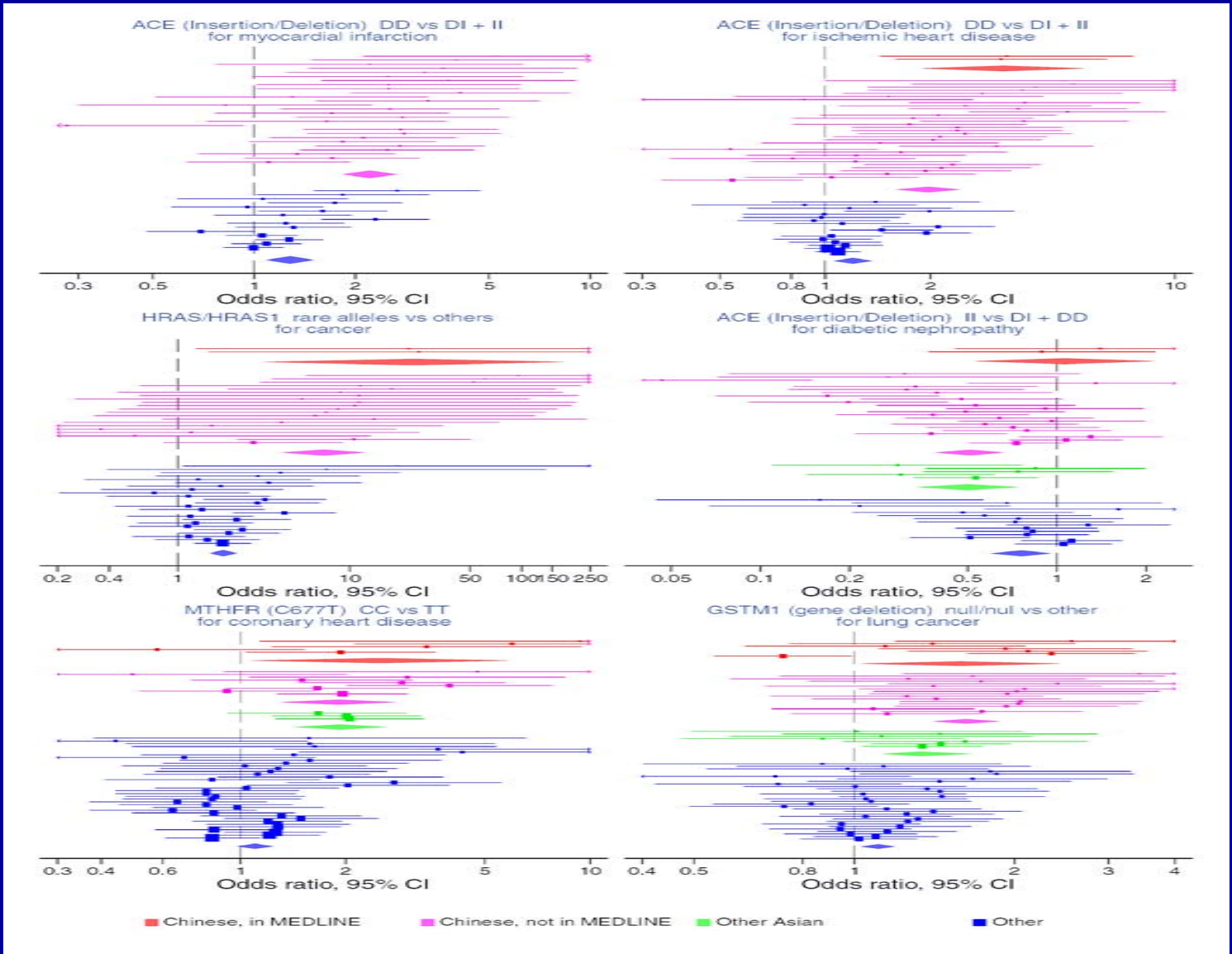
Received for publication July 2, 2007; accepted for publication August 1, 2007.

Measurement error: insight from a collaborative analysis

- Of 18 teams of investigators participating in the collaborative analysis of alpha-synuclein REP-I variation and Parkinson's disease risk, we found that 7 had to be excluded from the main analyses because of laboratory error exceeding 10% and/or overt violation of HWE in the controls
- Two other teams who had published an inverse association apparently had miscoded the alleles in their databases.

Maraganore et al, JAMA 2006

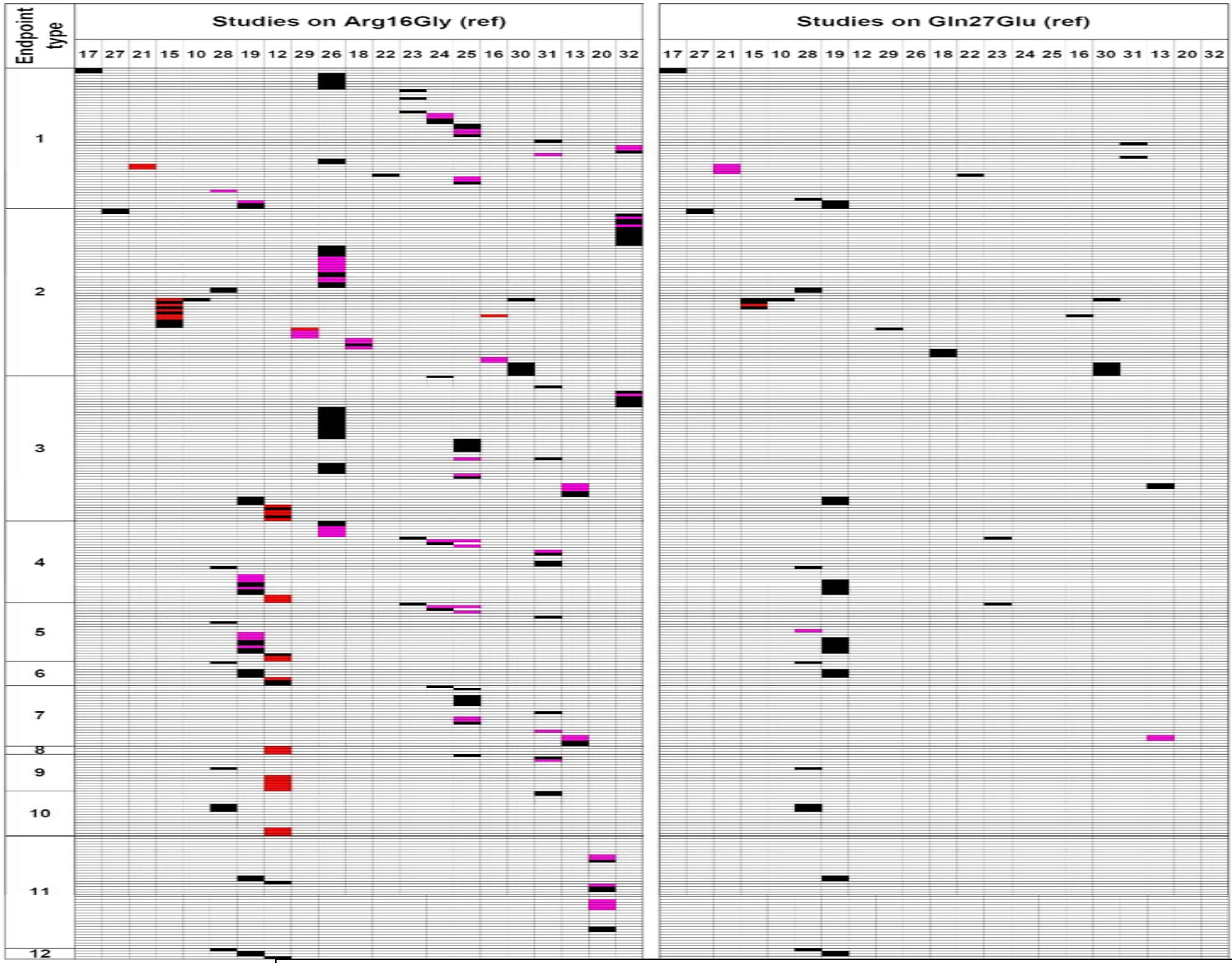
Language bias and global science



Pan et al. PLoS Med 2005

Defining and harmonizing multifarious phenotypes: the Lernean Hydra bias

-Hercules, I think we have a serious multiplicity problem!!!



Contopoulos-Ioannidis et al. Pharmacogenetics and Genomics 2006

Talking about sex and other interesting subgroups

Claims of Sex Differences

An Empirical Assessment in Genetic Associations

Nikolaos A. Patsopoulos, MD
Athina Tatsioni, MD
John P. A. Ioannidis, MD

SEX IS A FACTOR THAT HAS BEEN invoked extensively in the past as a modulator of effects in clinical research. However, empirical data from randomized trials suggest that many claimed subgroup differences based on sex have been spurious and led to serious misconceptions.[1] For example, aspirin was believed to be ineffective in secondary prevention of stroke in women for more than 10 years based on an underpowered subgroup analysis.[2]

In the human genome era, for many common diseases, published research has often considered that some common gene variants may have different effects in men vs women. Many diseases or traits with strong genetic backgrounds have different prevalence in the 2 sexes. For example, autoimmune diseases, endocrinopathies, and longevity are more common in women, while coronary artery disease, ischemic stroke, and high cholesterol levels are more common in men.[3] These observations do not necessarily mean that a specific gene variant should also have a different effect in men vs women. For most phenotypes, many common gene variants are likely to be responsible for determining susceptibility to disease.[4] Among autosomal variants, only some of them, if any, may interact with sex

Context Many studies try to probe for differences in risks between men and women, and this is a major challenge in the expanding literature of associations between genetic variants and common diseases or traits.

Objective To evaluate whether prominently claimed sex differences for genetic effects have sufficient internal and external validity.

Data Sources We searched PubMed through July 6, 2007, for genetic association studies claiming sex-related differences in the articles' titles. Titles and abstracts and, if necessary, the full text of the article were assessed for eligibility.

Study Selection Two hundred fifteen articles were retrieved by the search. We considered eligible all retrieved association studies that claimed different genetic effects across sexes of 1 or more gene variants for any human disease or phenotype. We considered both biallelic and multiallelic markers (including haplotypes) and both binary and continuous phenotypes and traits. We excluded non–English-language studies; studies evaluating only 1 sex; studies in which sex was treated only as an independent predictor of disease; studies that did not address any association of the investigated genetic variant with a disease or trait; studies not involving humans; and studies in which the authors did not claim any sex difference.

Data Extraction Two evaluators independently extracted data with a third evaluator arbitrating their discrepancies. Data evaluation included whether analyses were stated to have been specified a priori; whether sex effects were evaluated in the whole study or subgroups thereof; and whether the claims were appropriately documented, insufficiently documented, or spurious. For appropriately and insufficiently documented claims we performed the calculations for gene-sex interaction whenever raw data were available. Finally, we compared the sex-difference claims with the best internal validity against the results of other studies addressing the same interaction.

Results We appraised 432 sex-difference claims in 77 eligible articles. Authors stated that sex comparisons were decided a priori for 286 claims (66.2%), while the entire sample size was used in 210 (48.6%) claims. Appropriate documentation of gene-sex interaction was recorded in 55 claims (12.7%); documentation was insufficient for 303 claims and spurious for the other 74. Data for reanalysis of claims were available for 188 comparisons. Of these, 83 (44.1%) were nominally statistically significant at a $P=.05$ threshold, and more than half of them (n=44) had modest P values between .01 and .05. Of 60 claims with seemingly the best internal validity, only 1 was consistently replicated in at least 2 other studies.

Conclusion In this sample of highly prominent claims of sex-related differences in genetic associations, most claims were insufficiently documented or spurious, and claims with documented good internal and external validity were uncommon.

JAMA. 2007;298(8):880-893 www.jama.com

Patsopoulos et al. JAMA 2007

Calibration of credibility

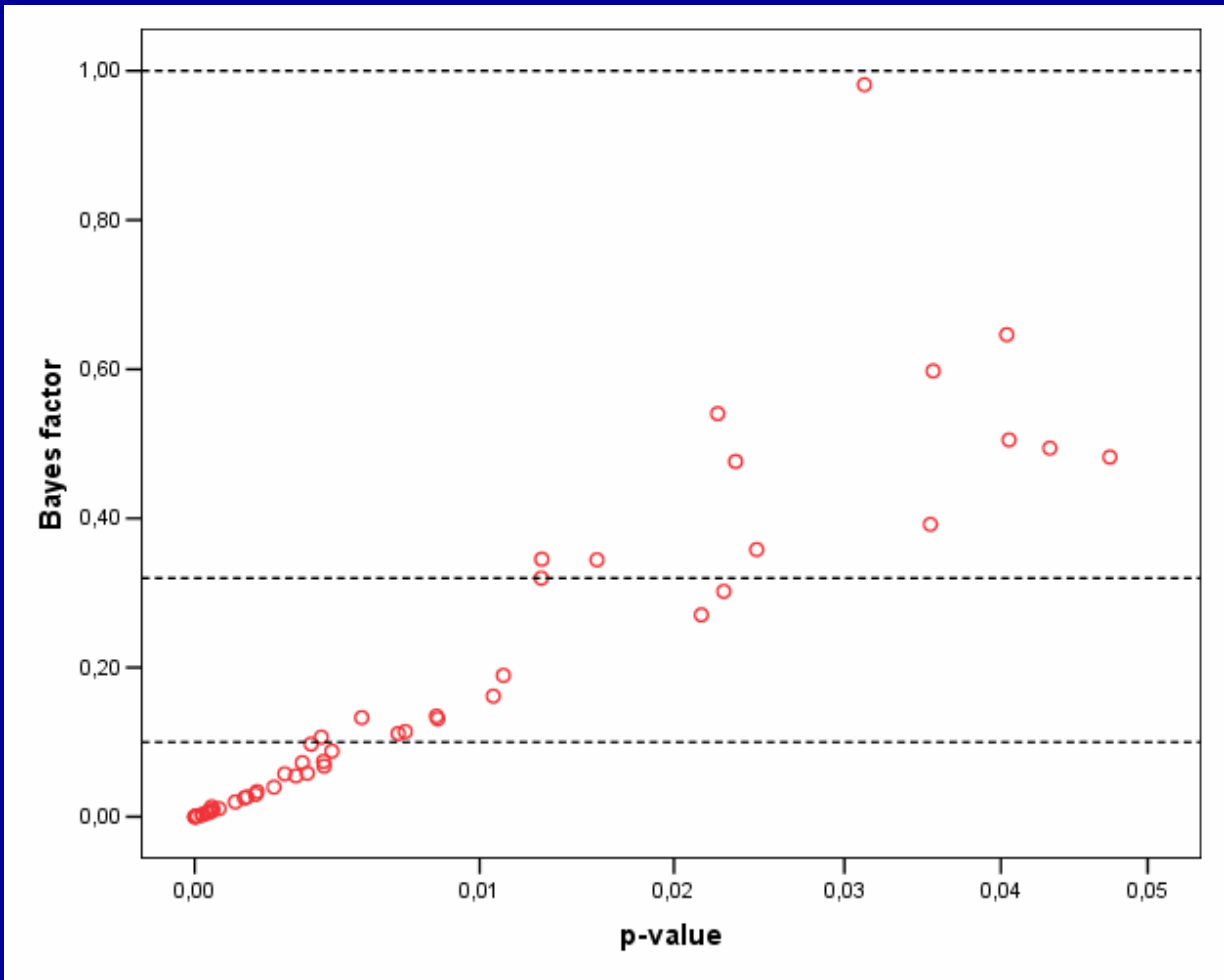
$$B = \sqrt{(1+(m/n_0)}\exp[(-z_m^2)/(2(1+(n_0/m)]$$

$$n_0 = 2\sigma^2/(\pi\theta_A^2) = 2m\,\mathrm{var}(\theta)/(\pi\theta_A^2)$$

$$n_0/m = 2\mathrm{var}(\theta)/(\pi\theta_A^2)$$

Ioannidis, Am J Epidemiol (in press) and Am J Med Genet (in press)

Calibration of credibility: genetic meta-analyses



Evolving credibility in genetic meta-analyses

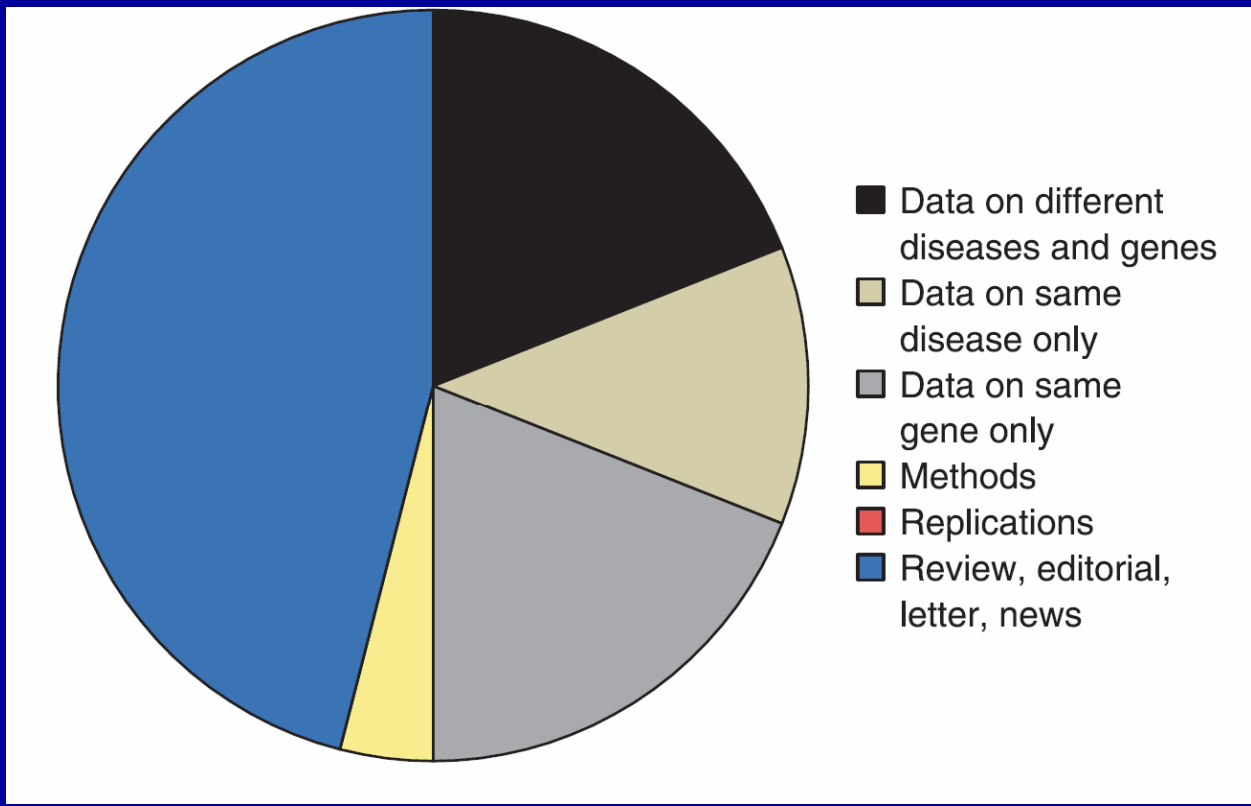
Earlier M-A (author and year)	Gene (variant); Contrast	Disease	OR (95% CI) in M-A	OR (95% CI) M-A2	M-A2 (author and year)	Differences	Bayes
No substantial support							
Boekholdt 2001	*FGB*/FGB promoter (455G/A); AA vs GG	MI	1.46 (1.00, 2.13)	1.12 (0.90, 1.41)	Smith 2005	Allele/wider	0.48/NP
Maraganore 2004	UCH-L1 (S18Y); S/S vs. other	Parkinson	1.20 (1.02, 1.40)	0.96 (0.86, 1.08)	Healy 2006	None/None	0.48/NP
Kosmas 2004	MTHFR (677C/T); TT vs. other	Preeclampsia	1.21 (1.01, 1.45)	1.01 (0.79, 1.29)	Lin 2005	None/None	0.60/NP
Burzotta 2004	F2 (20210G/A); other vs. GG	MI	1.32 (1.01, 1.72)	1.25 (1.05, 1.50)	Ye 2006	Allelle	0.51/0.28
Jonsson 2003	DRD3 (Ser9Gly) SerSer vs. other	Schizophrenia	1.10 (1.01, 1.21)	1.05 (0.97, 1.13)	Jonsson 2004	None/None	0.98/NP
Combarros 2003	IL1A (-889); 2/2 vs. Other	Alzheimer	2.35 (1.03, 5.37)	1.08 (0.98, 1.18)	Bertram 2007	Allele/wider	0.49/NP

There is certainly great news

- The replication process is accelerating

Early genetic epi: forlorn replication in search for complexity

- *Nature* 1994
- TNFA associates with cerebral malaria
- >800 citations to-date



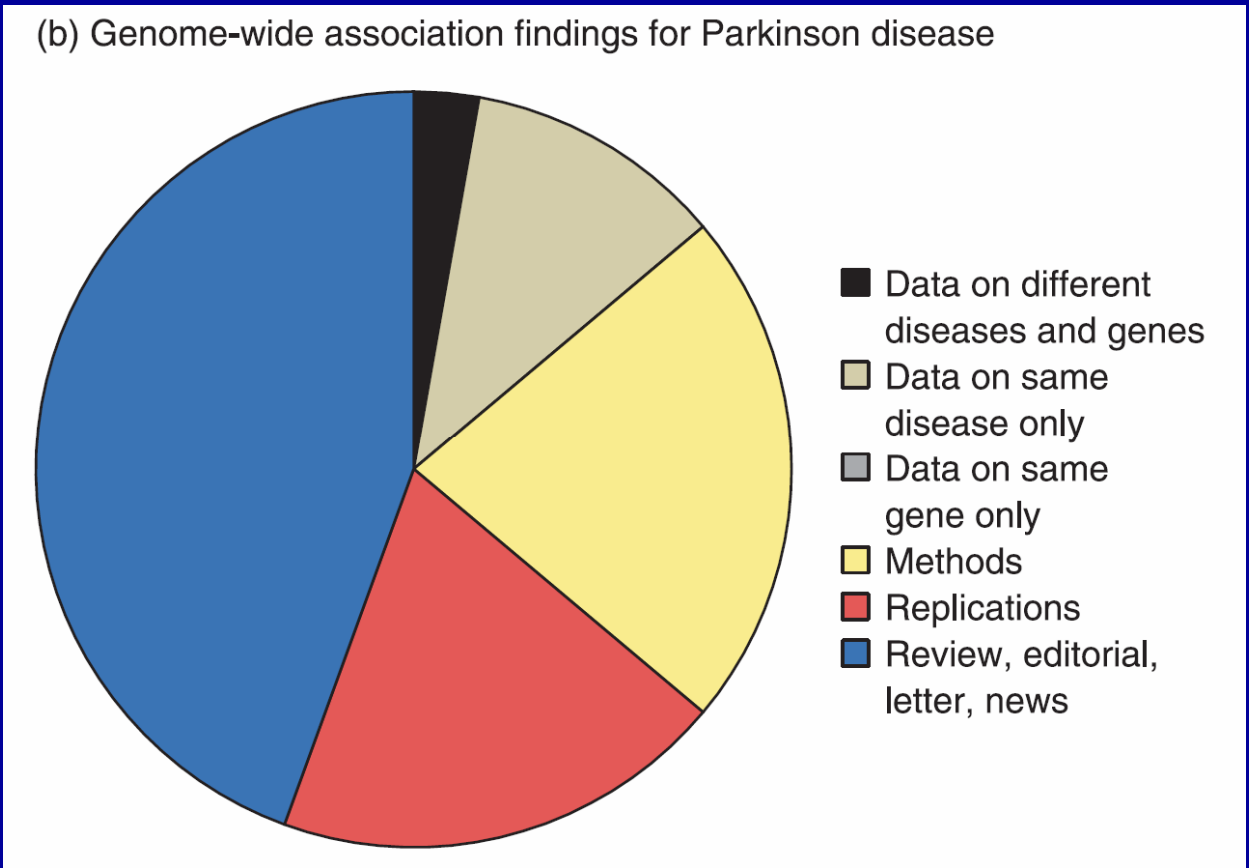
Pie chart analysis of the first 100 citations to the Nature paper

Ioannidis Eur J Clin Invest 2007

Discovery claims are a rapidly spreading infectious disease

- Within the first 100 citations to the Nature paper, 19 probed associations of *TNFA* genetic variability with various other conditions and phenotypes with 12 of these 19 studies proposing significant associations.
- In all 800 citations, more than 100 new associations were proposed.
- The proposing team subsequently also published on a different *TNFA* polymorphism that would modulate malarial outcomes, and also claimed that different alleles conferred susceptibility to severe anemia from malaria vs. cerebral malaria.
- Independent teams recently found no association with the original proposed polymorphism with either cerebral malaria or severe anemia – in much larger studies.
- What was probably a false-positive finding, not only got entrenched in the literature, but it also lent citation support for probably over 100 other proposed associations, many/most of which are likely to be also spurious.

Shifting attention to replication



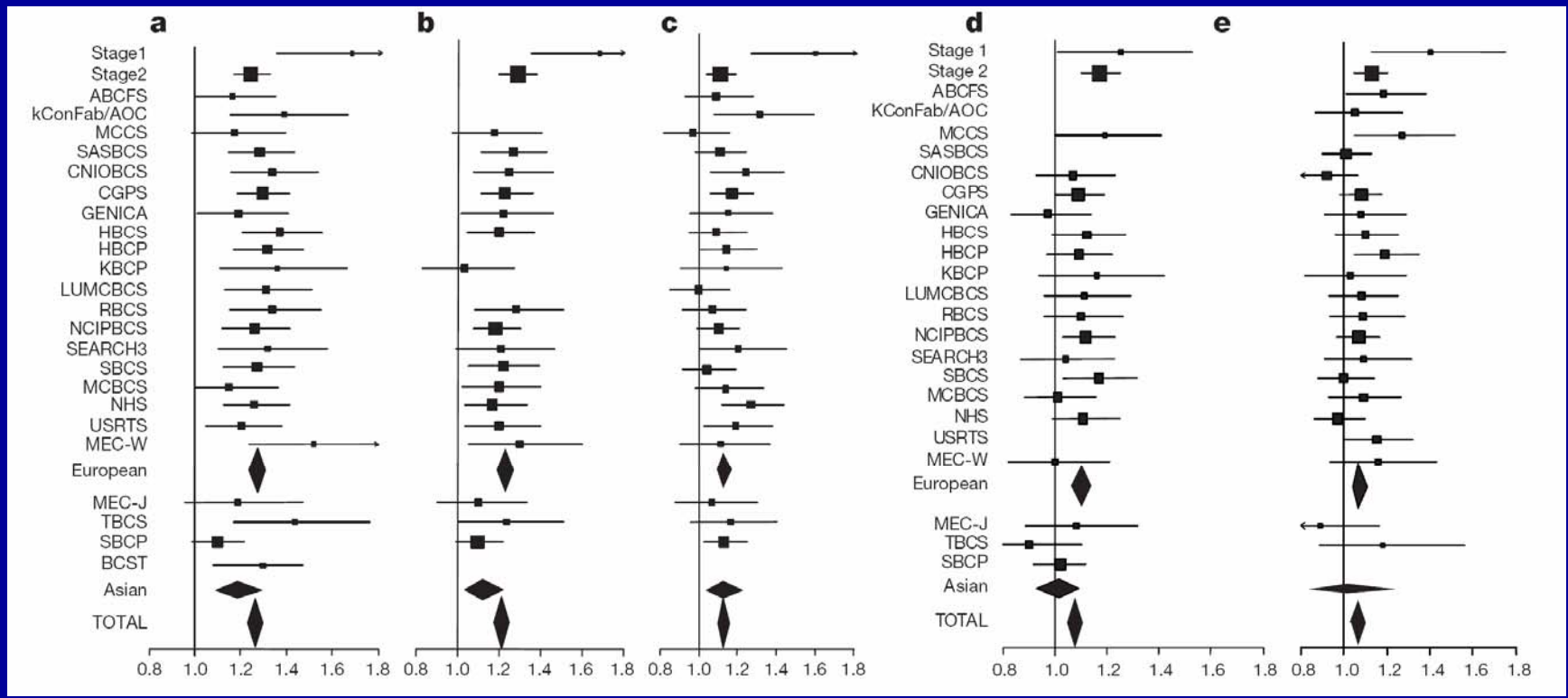
Ultrafast replication as a sine qua non

doi:10.1038/nature05887 nature

ARTICLES

Genome-wide association study identifies novel breast cancer susceptibility loci

Douglas F. Easton[1], Karen A. Pooley[2], Alison M. Dunning[2], Paul D. P. Pharoah[2], Deborah Thompson[1], Dennis G. Ballinger[3], Jeffery P. Struewing[4], Jonathan Morrison[2], Helen Field[2], Robert Luben[5], Nicholas Wareham[5], Shahana Ahmed[2], Catherine S. Healey[2], Richard Bowman[6], the SEARCH collaborators[2*], Kerstin B. Meyer[7],



Grading the evidence: the Venice criteria (IJE, 2007)

AAA	ABA	ACA
AAB	ABB	ACB
AAC	ABC	ACC

First letter = amount
Second letter = replication
Third letter = protection from bias

BAA	BBA	BCA
BAB	BBB	BCB
BAC	BBC	BCC

Strong evidence
Moderate evidence
Weak evidence

CAA	CBA	CCA
CAB	CBB	CCB
CAC	CBC	CCC

Let us add the environment

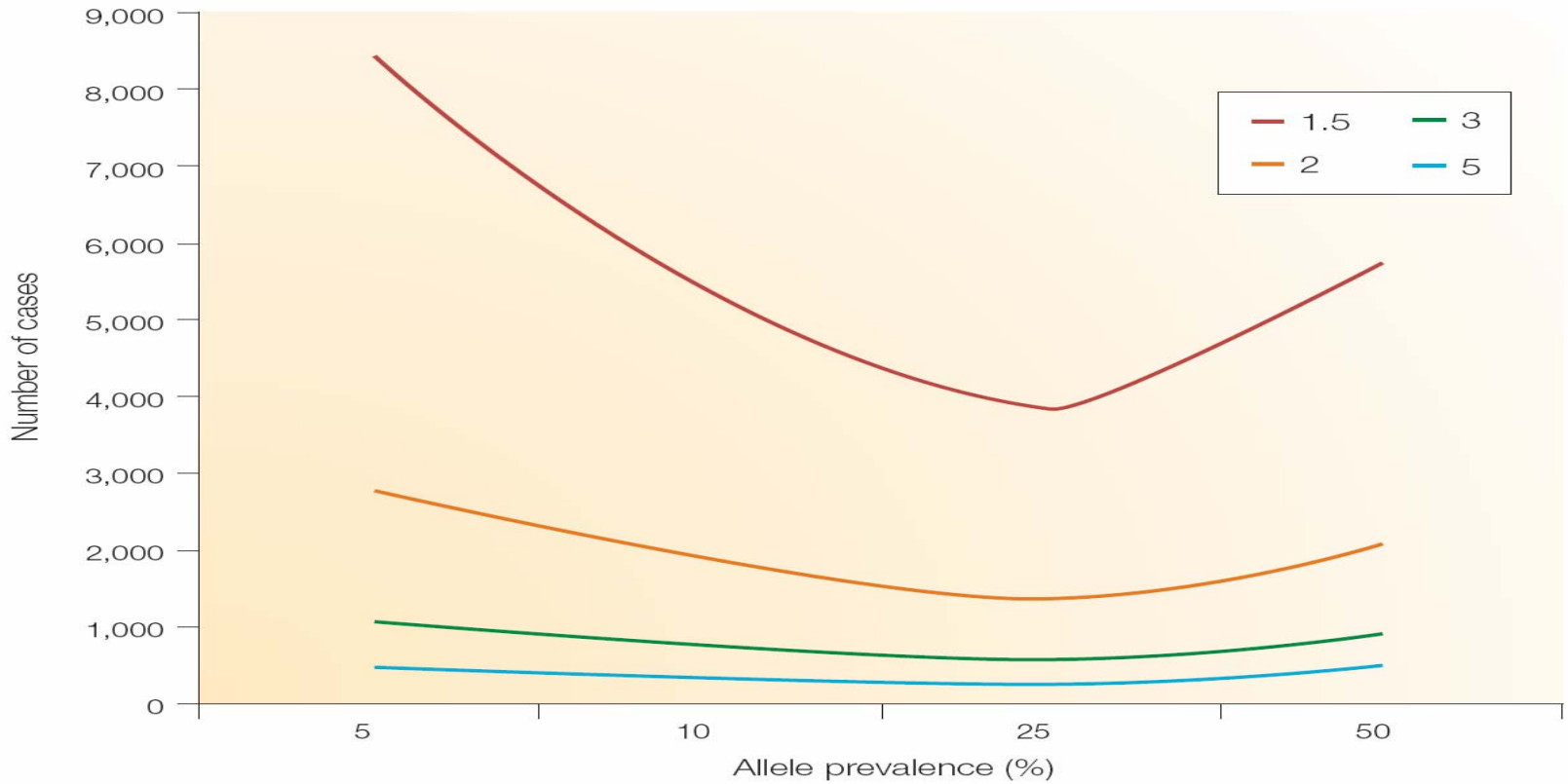


Figure 2 | **Number of cases needed to detect a range of multiplicative interactions, according to allele prevalence.** The model assumes the following: a dominant genetic model, a dichotomous exposure prevalence of 10%, a relative risk for a genotype of 1.5, a relative risk for exposure of 1.5 and a 1:1 case:control ratio. As the graph shows, thousands of cases and controls are needed to detect interactions with relative risks of 1.5 and 2. Calculations were carried out using Quanto Beta version 0.5 (REF. 13).

Hunter D. Nat Rev Genet 2005

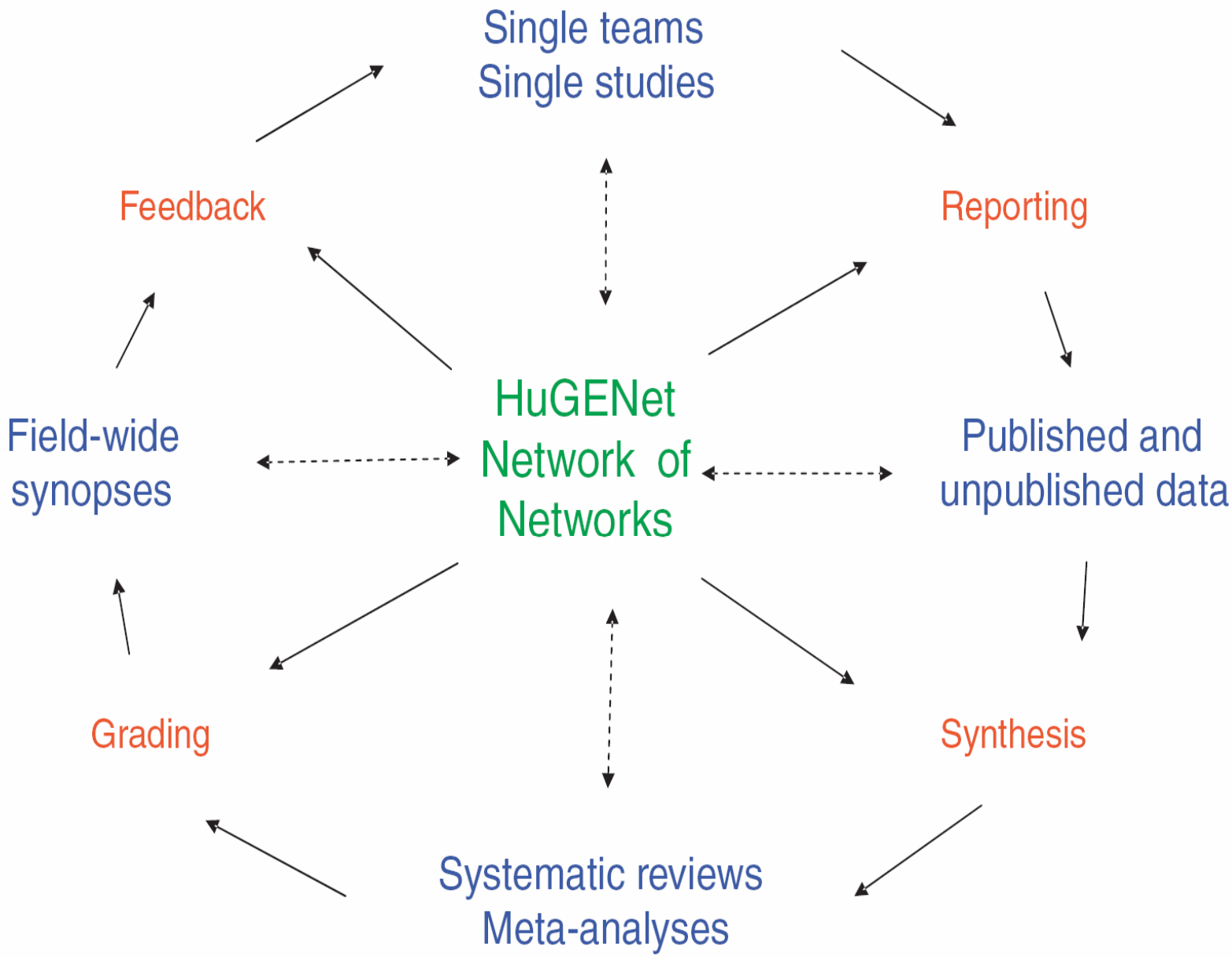


Figure 1 Framework for risk evaluation in genetic association studies.

SzGene synopsis: 1179 publications of common genetic variants and schizophrenia (including two GWA studies)

Gene	Polymorphism	Model	Cases vs. controls (# independent samples)	OR (95% CI)†	P-value	I^2	Grade
APOE	APOE (ε2/3/4) E4 vs. E3	E4 vs. E3, Caucasian[a]	1500 vs. 2702 (15)	1.16 (1.00-1.34)	0.043	0	B
COMT	rs165599	G vs. A, all ethnicities	2628 vs. 7340 (6)	1.11 (1.02-1.21)	0.019	25	C
COMT	rs737865	C vs. T, Caucasian[a]	1605 vs. 4021 (3)	1.13 (1.01-1.28)	0.039	34	C
DAO	rs4623951	C vs. T, all ethnicities	1509 vs. 1521 (4)	0.88 (0.79-0.98)	0.026	0	C
DRD1	rs4532 (DRD1_48A/G)	G vs. A, all ethnicities	725 vs. 1075 (5)	1.18 (1.01-1.38)	0.037	0	A
DRD2	rs1801028 (Ser311Cys)	G vs. C, Caucasian[b]	2299 vs. 3777 (15)	1.52 (1.09-2.12)	0.013	16	B
DRD2	rs6277 (Pro319Pro)	C vs. T, Caucasian[b]	473 vs. 896 (3)	1.45 (1.21-1.73)	<0.00004	15	C
DRD4	rs1800955 (521T/C)	C vs. T, all ethnicities	2002 vs. 1986 (6)	1.15 (1.05-1.26)	0.003	0	C
DRD4	120-bp TR	S vs. L, all ethnicities	1236 vs. 1199 (4)	0.81 (0.70-0.94)	0.005	7.	C
DTNBP1	rs1011313 (P1325)	T vs. C, Caucasian[a]	2696 vs. 2849 (8)	1.23 (1.07-1.40)	0.003	0	A
GABRB2	rs1816072	C vs. T, Caucasian[a]	1129 vs. 995 (4)	0.82 (0.72-0.93)	0.002	0	C
GABRB2	rs1816071	G vs. A, Caucasian[a]	1133 vs. 993 (4)	0.82 (0.72-0.93)	0.002	0	C
GABRB2	rs194072	C vs. T, Caucasian[a]	1137 vs. 991 (4)	0.83 (0.69-1.00)	0.048	7	B
GABRB2	rs6556547	T vs. G, Caucasian[a]	774 vs. 620 (3)	0.70 (0.52-0.95)	0.022	0	B
GRIN2B	rs7301328 (366G/C)	G vs. C, all ethnicities	903 vs. 810 (4)	1.16 (1.01-1.33)	0.034	27	C
GRIN2B	rs1019385 (200T/G)	G vs. T, all ethnicities	502 vs. 466 (4)	1.45 (1.14-1.85)	0.003	44	C
HP	Hp1/2	1 vs. 2, all ethnicities	1346 vs. 2018 (6)	0.88 (0.80-0.98)	0.016	0	C
IL1B	rs16944 (C511T)	T vs. C, Caucasian[b]	819 vs. 1302 (5)	0.78 (0.65-0.93)	0.006	26	C
MTHFR	rs1801133 (C677T)	T vs. C, all ethnicities	3327 vs. 4093 (14)	1.16 (1.05-1.30)	0.005	56	C
MTHFR	rs1801131 (A1298C)	C vs. A, Caucasian[b]	1211 vs. 1729 (5)	1.19 (1.07-1.34)	0.002	0	A
PLXNA2	rs752016	C vs. T, all ethnicities	1122 vs. 1211 (6)	0.82 (0.69-0.99)	0.037	33	C
SLC6A4	5-HTTVNTR	10 vs. 12, all ethnicities	2335 vs. 2688 (11)	0.86 (0.74-0.99)	0.036	50	C
TP53	rs1042522	C vs. G, all ethnicities	1418 vs. 1410 (5)	1.13 (1.01-1.26)	0.029	0	C
TPH1	rs1800532 (218A/C)	A vs. C, all ethnicities	829 vs. 1268 (5)	1.31 (1.15-1.51)	<0.00008	13	A

Human genome epidemiology

- Human genome epidemiology has made major progress in the last decade
- The pace of discovery and replication has accelerated a lot
- Methods and awareness of caveats has been heightened and solutions have been proposed for many of the problems of the early years
- The best is yet to come