Assessing cumulative evidence in genetic associations

Bethesda, December 2008

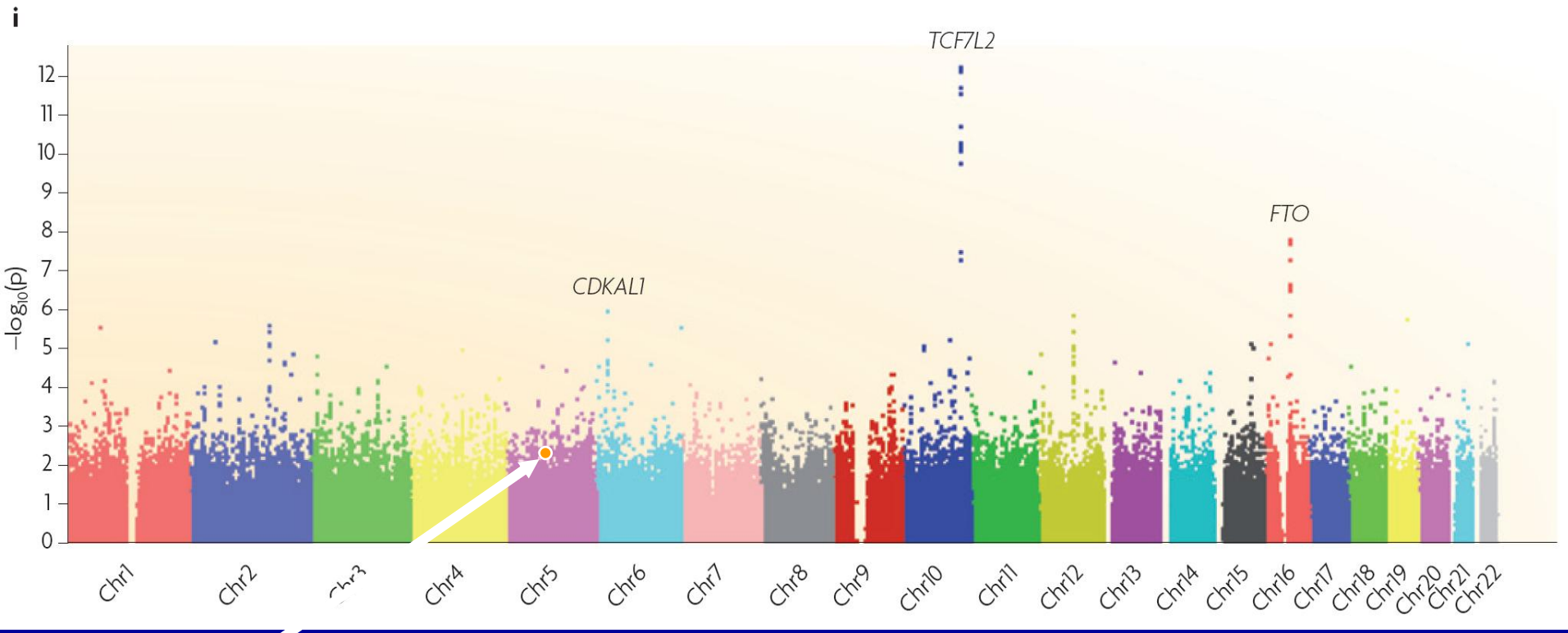
John P.A. Ioannidis

Professor and Chairman, Department of Hygiene and Epidemiology

University of Ioannina School of Medicine, Ioannina, Greece

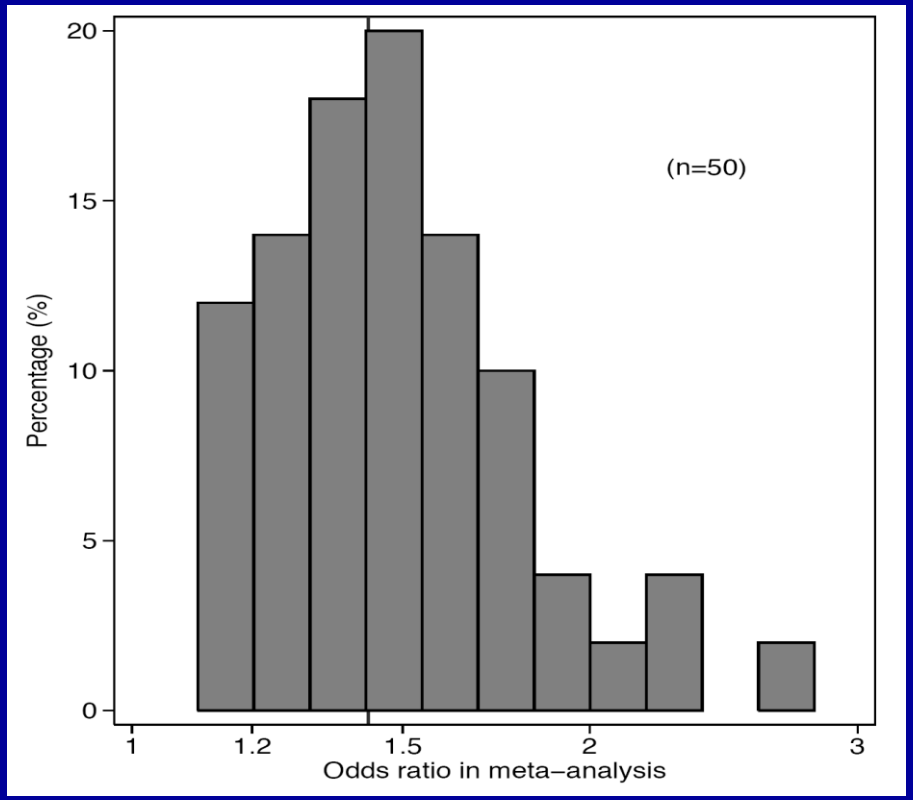
Professor of Medicine (adjunct) and Director, Genetics/Genomics Component, Tufts Clinical and Translational Science Institute and Center for Genetic Epidemiology and Modeling, Tufts Medical Center, Boston, USA

Agnostic search of skyscrapers: a single publication with 2 million analyses



Equivalent of 10,000 publications by a prestigious cohort of nutritional/lifestyle epidemiology

"Strength of association" Difficult to assess 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

DIAGRAM results: meta-GWA

				Stage 1 (DGI, FUSION, WTCCC)		All data		
Chr	risk allele frequency	n samples for 80% power	nearest gene(s)	OR (95%CI)	*P* value	n_{eff}	OR (95%CI)	*P* value
7	0.501	10,610	*JAZF1*	1.14 (1.07-1.20)	1.5E-04	59,617	1.10 (1.07-1.13)	5.0E-14
10	0.183	9,334	*CDC123/CAMK1D*	1.15 (1.06-1.24)	4.2E-04	62,366	1.11 (1.07-1.14)	1.2E-10
12	0.269	23,206	*TSPAN8/LGR5*	1.18 (1.10-1.26)	1.8E-05	62,301	1.09 (1.06-1.12)	1.1E-09
2	0.902	9,624	*THADA*	1.25 (1.12-1.40)	1.8E-04	60,832	1.15 (1.10-1.20)	1.1E-09
3	0.761	9,748	*ADAMTS9*	1.13 (1.06-1.22)	5.4E-04	62,387	1.09 (1.06-1.12)	1.2E-08
1	0.106	21,568	*NOTCH2*	1.30 (1.17-1.43)	1.1E-04	58,667	1.13 (1.08-1.17)	4.1E-08
12	0.733	17,808	*DCD*	1.15 (1.08-1.23)	3.2E-05	62,301	1.08 (1.05-1.11)	1.8E-07
3	0.927	16,370	*SYN2/PPARG*	1.33 (1.18-1.50)	1.0E-05	59,682	1.15 (1.10-1.21)	2.0E-07
1	0.107	17,428	*ADAM30*	1.14 (1.05-1.25)	1.4E-03	60,048	1.10 (1.06-1.15)	4.0E-07
6	0.282	16,696	*VEGFA*	1.13 (1.06-1.21)	5.4E-05	63,537	1.06 1.04-1.09)	4.0E-06
2	0.724	13,502	*BCL11A*	1.17 (1.10-1.26)	3.4E-05	59,682	1.05 (1.03-1.08)	1.0E-04

Zeggini et al. Nat Genet 2008

Int. J. Epidemiol. Advance Access published September 26, 2007

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2007;1–11
doi:10.1093/ije/dym159

Assessment of cumulative evidence on genetic associations: interim guidelines

John P A Ioannidis,[1–3]* Paolo Boffetta,[4] Julian Little,[5] Thomas R O'Brien,[6] Andre G Uitterlinden,[7] Paolo Vineis,[8] David J Balding,[8] Anand Chokkalingam,[9] Siobhan M Dolan,[10] W Dana Flanders,[11] Julian P T Higgins,[12] Mark I McCarthy,[13,14] David H McDermott,[15] Grier P Page,[16] Timothy R Rebbeck,[17] Daniela Seminara[18] and Muin J Khoury[19]

Accepted 9 July 2007

Established guidelines for causal inference in epidemiological studies may be inappropriate for genetic associations. A consensus process was used to develop guidance criteria for assessing cumulative epidemiologic evidence in genetic associations. A proposed semi-quantitative index assigns three levels for the amount of evidence, extent of replication, and protection from bias, and also generates a composite assessment of 'strong', 'moderate' or 'weak' epidemiological credibility. In addition, we discuss how additional input and guidance can be derived from biological data. Future empirical research and consensus development are needed to develop an integrated model for combining epidemiological and biological evidence in the rapidly evolving field of investigation of genetic factors.

Keywords Epidemiologic methods, genetics, genomics, causality, evidence

Grading the evidence: the Venice criteria (IJE, 2008)

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

The three criteria

Table 1 Considerations for epidemiologic credibility in the assessment of cumulative evidence on genetic associations

Criteria	Categories	Proposed operationalization
Amount of evidence	A: Large-scale evidence B: Moderate amount of evidence C: Little evidence	Thresholds may be defined based on sample size, power or false-discovery rate considerations. The frequency of the genetic variant of interest should be accounted for. As a simple rule, we suggest that category A requires over 1000 subjects (total number of cases and controls assuming 1:1 ratio) evaluated in the least common genetic group of interest; B corresponds to 100–1000 subjects evaluated in this group and C corresponds to <100 subjects evaluated in this group (see 'Discussion' section in the text and Table 2 for further elaboration).[a]
Replication	A: Extensive replication including at least one well-conducted meta-analysis with little between-study inconsistency B: Well-conducted meta-analysis with some methodological limitations or moderate between-study inconsistency C: No association; no independent replication; failed replication; scattered studies; flawed meta-analysis or large inconsistency	Between-study inconsistency entails statistical considerations (e.g. defined by metrics such as I^2, where values of 50% and above are considered large and values of 25–50% are considered moderate inconsistency) and also epidemiological considerations for the similarity/standardization or at least harmonization of phenotyping, genotyping and analytical models across studies. See 'Discussion' section in the text for the threshold (statistical or others) required for claiming replication under different circumstances (e.g. with or without including the discovery data in situations with massive testing of polymorphisms).
Protection from bias	A: Bias, if at all present, could affect the magnitude but probably not the presence of the association B: No obvious bias that may affect the presence of the association but there is considerable missing information on the generation of evidence C: Considerable potential for or demonstrable bias that can affect even the presence or absence of the association	A prerequisite for A is that the bias due to phenotype measurement, genotype measurement, confounding (population stratification) and selective reporting (for meta-analyses) can be appraised as not being high (as shown in detail in Table 3) plus there is no other demonstrable bias in any other aspect of the design, analysis or accumulation of the evidence that could invalidate the presence of the proposed association. In category B, although no strong biases are visible, there is no such assurance that major sources of bias have been minimized or accounted for because information is missing on how phenotyping, genotyping and confounding have been handled. Given that occult bias can never be ruled out completely, note that even in category A, we use the qualifier 'probably'.

[a]For example, if the association pertains to the presence of homozygosity for a common variant and if the frequency of homozygosity is 3%, then category A amount of evidence requires over 30 000 subjects and category B between 3000 and 30 000.

Options of amount of evidence

- Simple operational: sample size of the least common genetic group among those compared (it could reflect participants or alleles, depending on the model)
- Power
- False-discovery rate
- Bayesian credibility

Replication: have we had enough?

- Data from the National Human Genome Research Institute (NHGRI) GWA studies catalog as of October 14, 2008

- 233 discovered associations for binary outcome phenotypes with $p<10^{-5}$
- Only 142 have a p-value $<10^{-7}$
- Only 87 (39%) have a p-value$<10^{-10}$.

- Most GWAS-discovered loci need further exact replication with more large-scale evidence, before they can be considered sufficiently reliable even as simple markers.

Consistency of replication: why do almost all GWAS use a meta-analysis method developed in 1932 and largely abandoned in the 1970s?

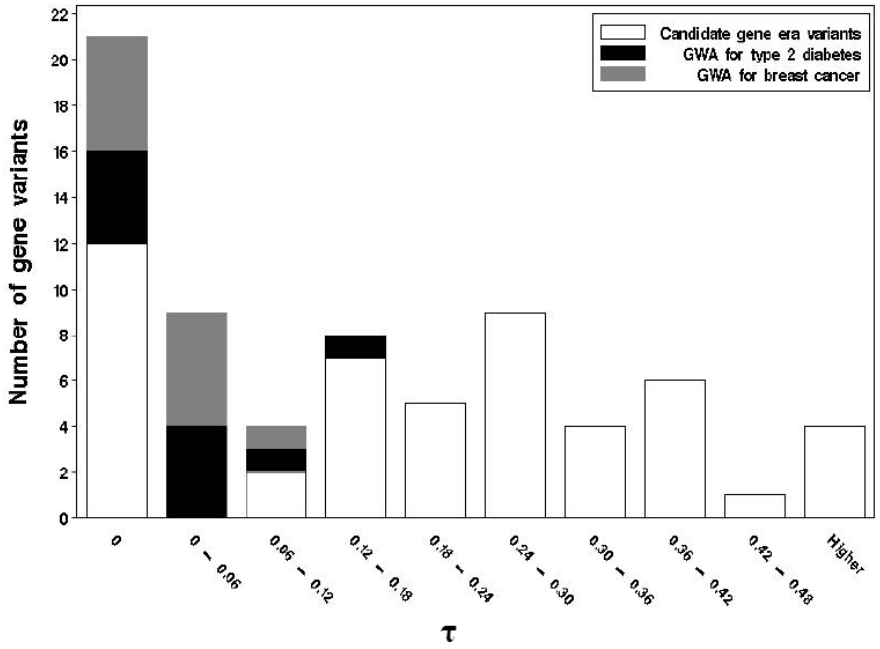
A: Extensive replication including at least one well-conducted meta-analysis with little between-study inconsistency

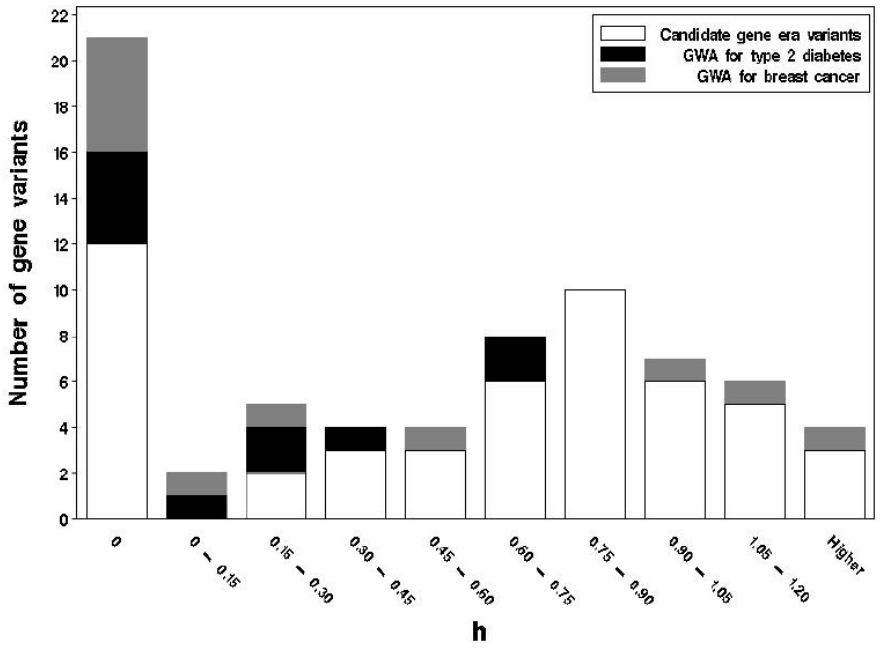
B: Well-conducted meta-analysis with some methodological limitations or moderate between-study inconsistency

C: No association; no independent replication; failed replication; scattered studies; flawed meta-analysis or large inconsistency

Between-study inconsistency entails statistical considerations (e.g. defined by metrics such as I^2, where values of 50% and above are considered large and values of 25–50% are considered moderate inconsistency) and also epidemiological considerations for the similarity/standardization or at least harmonization of phenotyping, genotyping and analytical models across studies.

See 'Discussion' section in the text for the threshold (statistical or others) required for claiming replication under different circumstances (e.g. with or without including the discovery data in situations with massive testing of polymorphisms).

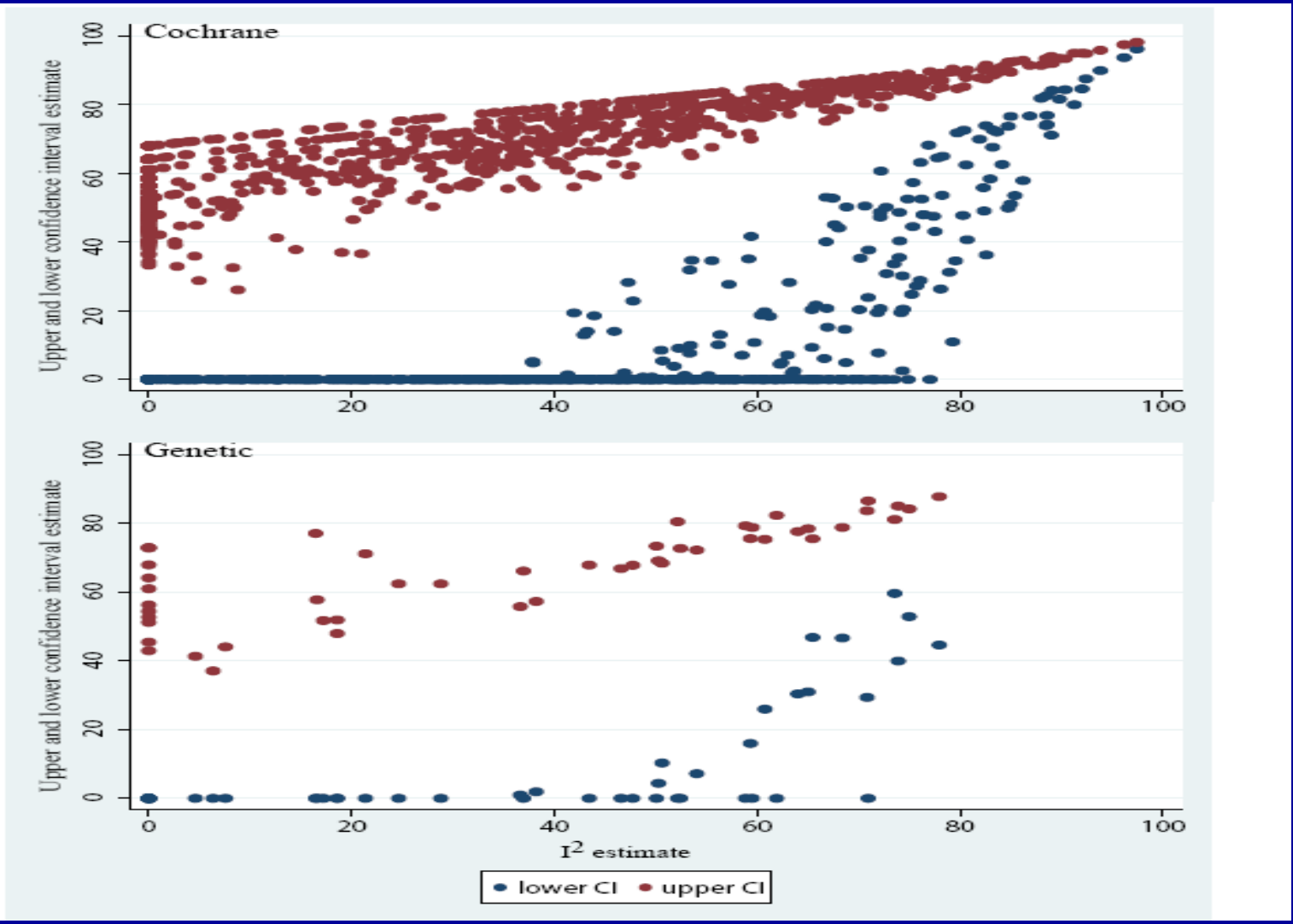




Heterogeneity in candidate gene era and GWA era

Mooneshinghe, et al. PNAS 2008

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



Ioannidis, Patsopoulos, Evangelou, BMJ 2007

Protection from bias

A: Bias, if at all present, could affect the magnitude but probably not the presence of the association

B: No obvious bias that may affect the presence of the association but there is considerable missing information on the generation of evidence

C: Considerable potential for or demonstrable bias that can affect even the presence or absence of the association

A prerequisite for A is that the bias due to phenotype measurement, genotype measurement, confounding (population stratification) and selective reporting (for meta-analyses) can be appraised as not being high (as shown in detail in Table 3) plus there is no other demonstrable bias in any other aspect of the design, analysis or accumulation of the evidence that could invalidate the presence of the proposed association. In category B, although no strong biases are visible, there is no such assurance that major sources of bias have been minimized or accounted for because information is missing on how phenotyping, genotyping and confounding have been handled. Given that occult bias can never be ruled out completely, note that even in category A, we use the qualifier 'probably'.

Table 3 Typical biases and their typical impact on associations depending on the status of the evidence

Biases	Status of the evidence	Likelihood of bias to invalidate an observed association		
		Small OR <1.15	Typical OR 1.15–1.8	Large OR >1.8
Bias in phenotype definition	Not reported what was done	Unknown	Unknown	Unknown
	Unclear phenotype definitions	Possible/High	Possible/High	Possible/High
	Clear widely agreed definitions of phenotypes	Low/None	Low/None	Low/None
	Efforts for retrospective harmonization	Possible/High	Low	Low/None
	Prospective standardization of phenotypes	Low/None	Low/None	Low/None
Bias in genotyping	Not reported what was done	Unknown	Unknown	Unknown
	No quality control checks	Possible/High	Low	Low
	Appropriate quality control checks	Low	Low	Low/None
Population stratification	Not reported what was done	Unknown	Unknown	Unknown
	Nothing done[a]	Possible/High	Possible/High	Possible/High
	Same descent group[b]	Possible/High	Low	Low/None
	Adjustment for reported descent	Possible/High	Low	Low/None
	Family-based design	Low/None	Low/None	Low/None
	Genomic control, PCA or similar method	Low/None	Low/None	Low/None
Selective reporting biases	Meta-analysis of published data	Possible/High	Possible	Possible
	Retrospective efforts to include unpublished data	Possible/High	Possible	Possible
	Meta-analysis within consortium	Low/None	Low/None	Low/None

A research finding cannot reach credibility over 50% unless

$u < R$

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

Ioannidis. PLoS Med 2005

Bias checks for retrospective meta-analysis

"Automated checks"

- Effect size <1.15-fold from the null effect
- Association lost with exclusion of first study
- Association lost with exclusion of HWE-violating studies or with adjustment for HWE
- Evidence for small-study effect in an asymmetry regression test with proper type I error (e.g. Harbord, Stat Med)
- Evidence for excess of single studies with formally statistically significant results (Ioannidis and Trikalinos, Clinical Trials)

"Consider whether they are problems"

- Unclear/misclassified phenotypes with possible differential misclassification against genotyping
- Differential misclassification of genotyping against phenotypes
- Major concerns for population stratification (need to justify for affecting OR>1.15-fold, not invoked to-date)
- Any other reason (case-by-case basis) that would destroy the association

Bias checks for a prospective consortium analysis

- Magnitude of effect size, small-study effects, excess of studies with significant findings are not an issue here, provided there is no selective reporting (basic trust)
- The other considerations still need to be raised

Calibration of credibility with spike and smear prior

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

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

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

Ioannidis, Am J Epidemiol 2008 and Am J Med Genet 2008

Calibration of credibility for various proposed GWA associations

Table 2. Credibility estimates for proposed genome wide associations

Gene	Variant	With prior credibility C_0 =0.0001			With prior credibility C_0 =0.00001			With prior credibility C_0 =0.000001		
		θ_A =0.049	θ_A =0.262	θ_A =0.588	θ_A =0.049	θ_A =0.262	θ_A =0.588	θ_A =0.049	θ_A =0.262	θ_A =0.588
Periodic limb movements in sleep										
BTBD9	rs3923809	0.948	1.000	1.000	0.645	1.000	1.000	0.154	1.000	1.000
Type 2 diabetes mellitus										
---	rs9300039	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FTO	rs8050136	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PPARG	rs1801282	0.005	0.007	0.004	0.001	0.001	0.000	0.000	0.000	0.000
CDKAL1	rs10946398	0.325	0.252	0.136	0.046	0.033	0.015	0.005	0.003	0.002
SLC30A8	rs13266634	0.154	0.124	0.062	0.018	0.014	0.007	0.002	0.001	0.001
CDKN2B	rs564398	0.886	0.887	0.788	0.437	0.440	0.270	0.072	0.073	0.036
HHEX	rs5015480-rs1111875	0.999	0.999	0.998	0.990	0.992	0.984	0.911	0.927	0.857
KCNJ11	rs5215	1.000	1.000	1.000	0.999	0.999	0.998	0.989	0.993	0.985
IGF2BP2	rs4402960	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.999	0.998
CDKN2B	rs10811661	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TCF7L2	rs7901695	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Parkinson's disease										
SEMA5A	rs7702187	0.001	0.179	0.222	0.000	0.021	0.028	0.000	0.002	0.003
---	rs10200894	0.000	0.084	0.125	0.000	0.009	0.014	0.000	0.001	0.001
---	rs2313982	0.000	0.062	0.123	0.000	0.007	0.014	0.000	0.001	0.001
---	rs17329669	0.000	0.079	0.094	0.000	0.008	0.010	0.000	0.001	0.001
---	rs7723605	0.000	0.054	0.071	0.000	0.006	0.008	0.000	0.001	0.001
---	ss46548856	0.000	0.042	0.068	0.000	0.004	0.007	0.000	0.000	0.001
GALNT3	rs16851009	0.000	0.040	0.060	0.000	0.004	0.006	0.000	0.000	0.001
PRDM2	rs2245218	0.000	0.046	0.051	0.000	0.005	0.005	0.000	0.000	0.001
PASD1	rs7878232	0.001	0.040	0.027	0.000	0.004	0.003	0.000	0.000	0.000
---	rs1509269	0.000	0.024	0.029	0.000	0.003	0.003	0.000	0.000	0.000
---	rs11737074	0.000	0.019	0.016	0.000	0.002	0.002	0.000	0.000	0.000

The value 0.000 corresponds to estimated credibility <0.001

Subjecting traditional epidemiology to the same rules?

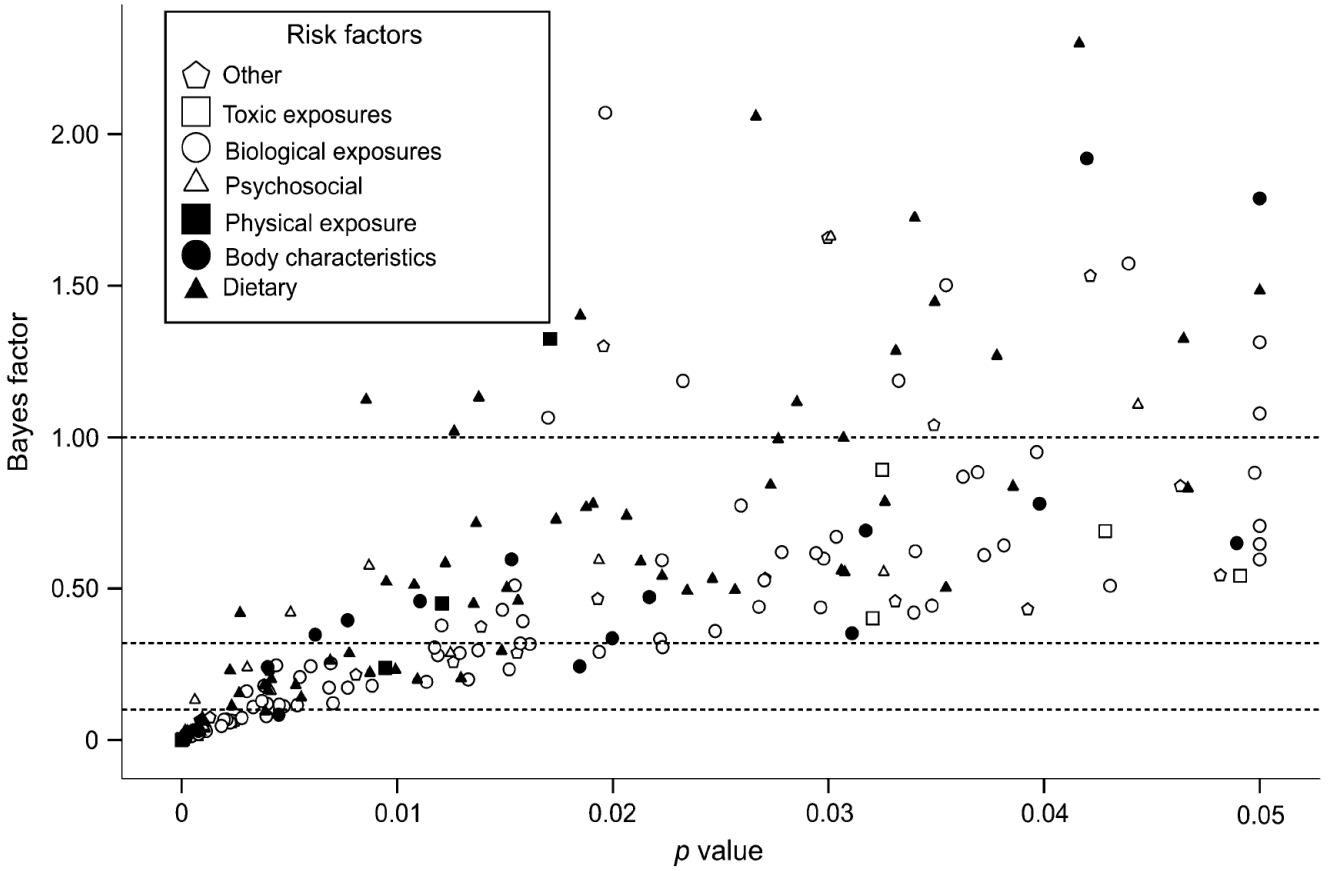


FIGURE 2. Estimated Bayes factors for 272 epidemiologic studies with formally statistically significant results. The Bayes factor is plotted against the observed *p* value in each study. Shown are calculations assuming θ_A of 1.50 (relative risk = 4.48). The dashed lines correspond to threshold values (1.00, 0.32, 0.10) separating different Bayes factor categories.

A step further: 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.

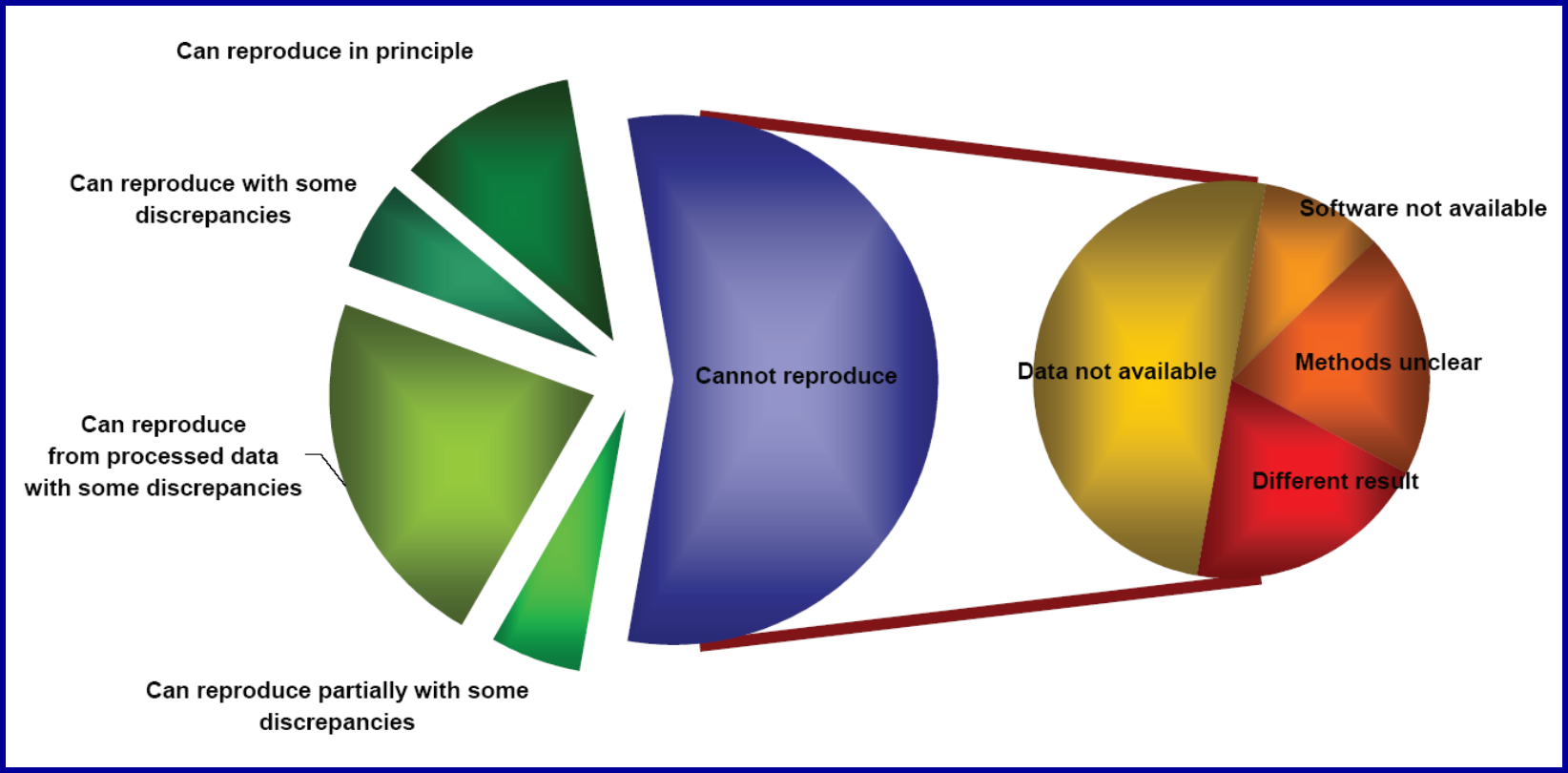
Results We appraised 432 sex-difference claims in 77 eligible articles. Authors stated that sex comparisons were decided a priori for 286 claims (66.2%), while the entire sample size was used in 210 (48.6%) claims. Appropriate documentation of gene-sex interaction was recorded in 55 claims (12.7%); documentation 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

Publicly available data
Replication>Reproducibility>Repeatability



Ioannidis, Allison, Ball, et al. Nature Genetics (in press)

“Conglomerate” evidence

Various combinations of

* scattered studies
* retrospective meta-analyses
* scattered single GWAS
* prospective consortia analyses, including multiple GWAS
* more scattered studies

in various time sequence: consider the highest level of evidence? Or all the evidence?

Summarizing and grading the evidence in its totality

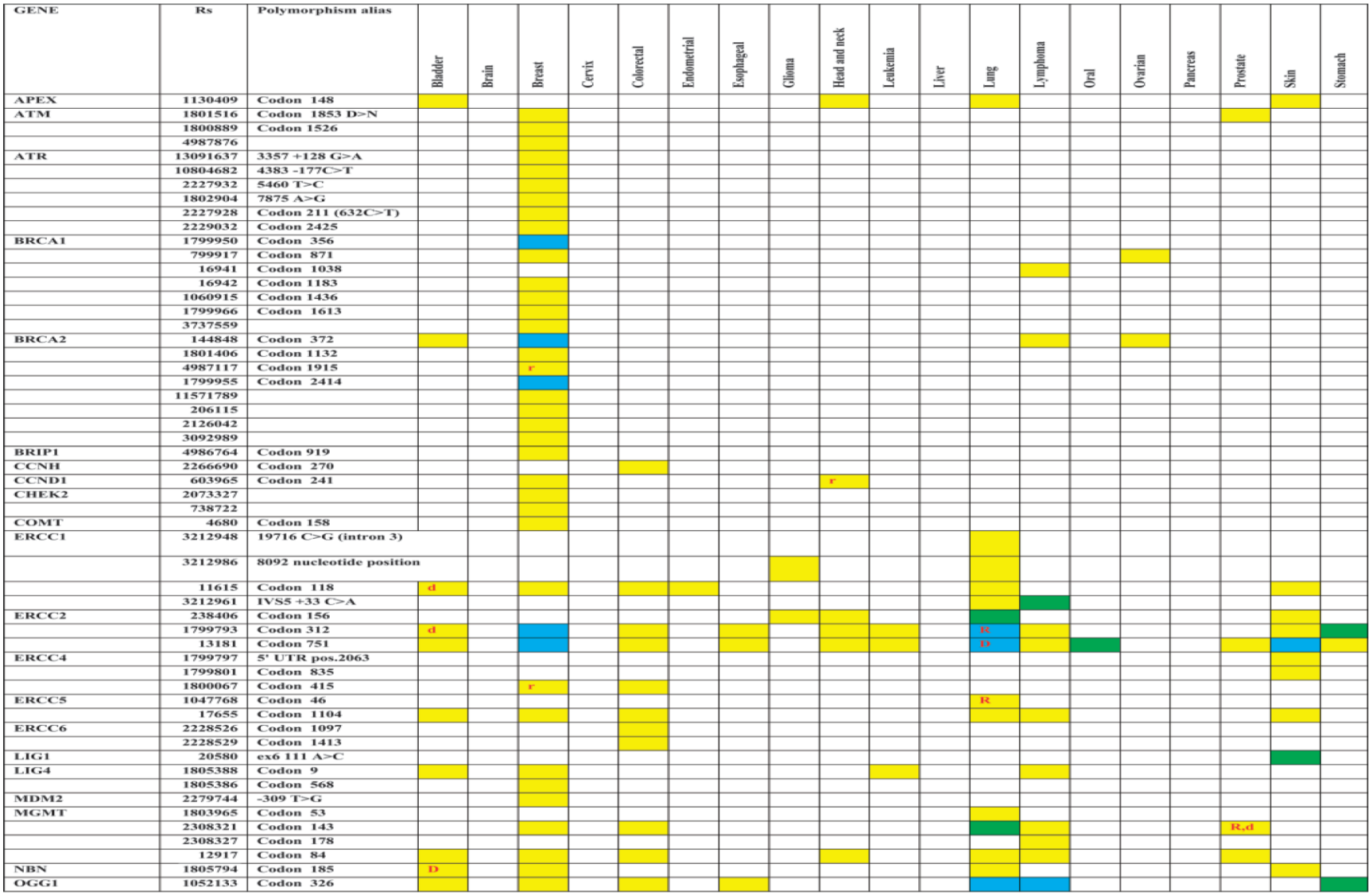
- Field synopses, including all data from candidate and agnostic studies in a specific field

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

Allen et al., Nat Genetics 2008

DNA repair genes: A thousand studies in one slide…



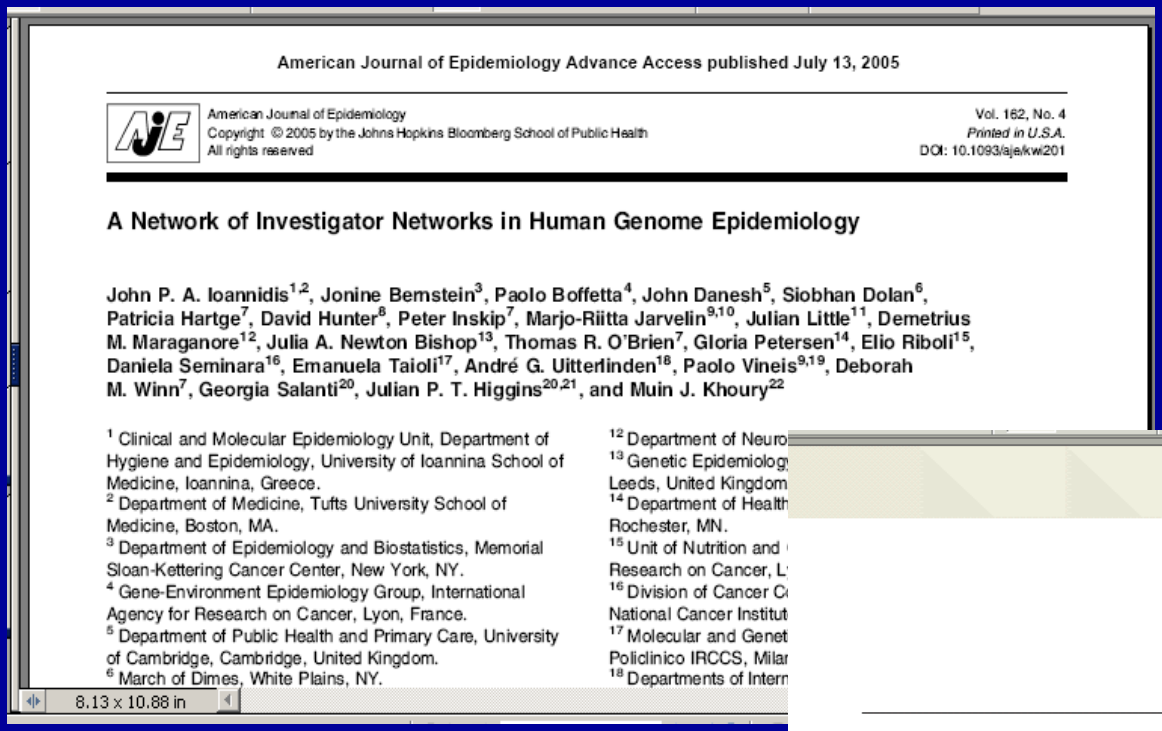
GENE	Rs	Polymorphism alias	Bladder	Brain	Breast	Cervix	Colorectal	Endometrial	Esophageal	Glioma	Head and neck	Leukemia	Liver	Lung	Lymphoma	Oral	Ovarian	Pancreas	Prostate	Skin	Stomach
APEX	1130409	Codon 148																			
ATM	1801516	Codon 1853 D>N																			
	1800889	Codon 1526																			
	4987876																				
ATR	13091637	3357 +128 G>A																			
	10804682	4383 -177C>T																			
	2227932	5460 T>C																			
	1802904	7875 A>G																			
	2227928	Codon 211 (632C>T)																			
	2229032	Codon 2425																			
BRCA1	1799950	Codon 356																			
	799917	Codon 871																			
	16941	Codon 1038																			
	16942	Codon 1183																			
	1060915	Codon 1436																			
	1799966	Codon 1613																			
	3737559																				
BRCA2	144848	Codon 372																			
	1801406	Codon 1132																			
	4987117	Codon 1915			r																
	1799955	Codon 2414																			
	11571789																				
	206115																				
	2126042																				
	3092989																				
BRIP1	4986764	Codon 919																			
CCNH	2266690	Codon 270																			
CCND1	603965	Codon 241									r										
CHEK2	2073327																				
	738722																				
COMT	4680	Codon 158																			
ERCC1	3212948	19716 C>G (intron 3)																			
	3212986	8092 nucleotide position																			
	11615	Codon 118	d																		
	3212961	IVS5 +33 C>A																			
ERCC2	238406	Codon 156																			
	1799793	Codon 312	d											R							
	13181	Codon 751												D							
ERCC4	1799797	5' UTR pos.2063																			
	1799801	Codon 835																			
	1800067	Codon 415			r																
ERCC5	1047768	Codon 46												R							
	17655	Codon 1104																			
ERCC6	2228526	Codon 1097																			
	2228529	Codon 1413																			
LIG1	20580	ex6 111 A>C																			
LIG4	1805388	Codon 9																			
	1805386	Codon 568																			
MDM2	2279744	-309 T>G																			
MGMT	1803965	Codon 53																			
	2308321	Codon 143																	R,d		
	2308327	Codon 178																			
	12917	Codon 84																			
NBN	1805794	Codon 185	D																		
OGG1	1052133	Codon 326																			

Vineis, Manuguerra, Kavvoura et al., JNCI in press 2008

well, maybe two slides…

PARP1	1136410	Codon 762																		
	1805410	IVS9 +104 A>G			R															
PPP1R13L	6966	2485 A>T (3 UTR)																		
	1970764	IVS1 4364 A>G																		
POLI	8305	Codon 706												D						
RAD23B	1805329	Codon 249																		
RAD51	1801321	172 G>T																		
	1801320	5` UTR																		
RAD52	11226	2259 C>T																		
RAG1	2227973	Codon 820																		
TP53	1042522	Codon 72				d								D						
	1614984	1474 bp 3STP C>T																		
	9894946	1846 bp 3STP T>C																		
	8079544	IVS1 -112 G>A																		
	1642785	IVS2 +38 C>G																		
	2909430	IVS4 -91 A>G																		
	12947788	IVS7 +72 T>C																		
	12951053	IVS7 +92 T>G																		
	1625895	Intron 6 (Msp I)																		
	17878362	PIN3																		
TP53BP1	1869258	-885 T>G																		
	2602141	Codon 1136																		
	560191	Codon 353																		
WRN	1346044	Codon 1367																		
XPA	1800975	23 G>A												R						
XPC	2228000	Codon 500																		
	2228001	Codon 939																		
		PAT Intron 9																		
XRCC1	1799782	Codon 194							r		R								d	d
	915927	Codon 206																		
	25489	Codon 280																		
	25487	Codon 399				r														
	3213245	-77 T>C												D						
XRCC2	3218536	Codon 188					d													
XRCC3	1799794	5' region pos.4541			d															
	861539	Codon 241			R															r
	1799796	IVS7 17893 A>G			r															
XRCC4	1805377	IVS7 -1 A>G	D																	
XRCC6	132788	Codon 593																		

Networks and Networks 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, L
[16] Division of Cancer C
National Cancer Institut
[17] Molecular and Genet
Policlinico IRCCS, Milar
[18] Departments of Intern

8.13 x 10.88 in

COMMENTARY

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

Convincing predictive ability and improvement in decision-making: it takes far more than just highly credible epidemiology, but is impossible without it

Table 1. Recent studies of predictive performance with common genetic variants, traditional risk factors, and both

Disease (ref)	Design, number of cases	Gene variants	AUC for gene variants	AUC for traditional risk factors	AUC including both
Cardiovascular events (5)	Cohort, n=238	9*	ND	0.80 (15 factors)	0.80
Prostate cancer (6)	Case-control, n=2893	5	ND	0.608 (age, region, family history)	0.633
Type 2 diabetes (7)	Case-control, n=2309	18	0.60	0.78 (age, gender, body mass index)	0.80
Progression of age-related macular degeneration (8)	Cohort, n=281	2	0.758	ND (smoking, body mass index)	0.768

Ioannidis J. Personalized genetic prediction: too limited, too expensive, too soon? Ann Intern Med Jan 20, 2009

Some concluding comments

- Assessment of the cumulative evidence on genetic associations focuses on amount of evidence, consistency of replication, and protection from bias
- Evidence is often uncertain and tenuous, and the uncertainty is often under-appreciated
- Evidence is likely to become more reliable when its integration is transparent and anticipated prospectively by all involved partners
- Discovery and integration should ideally proceed in parallel