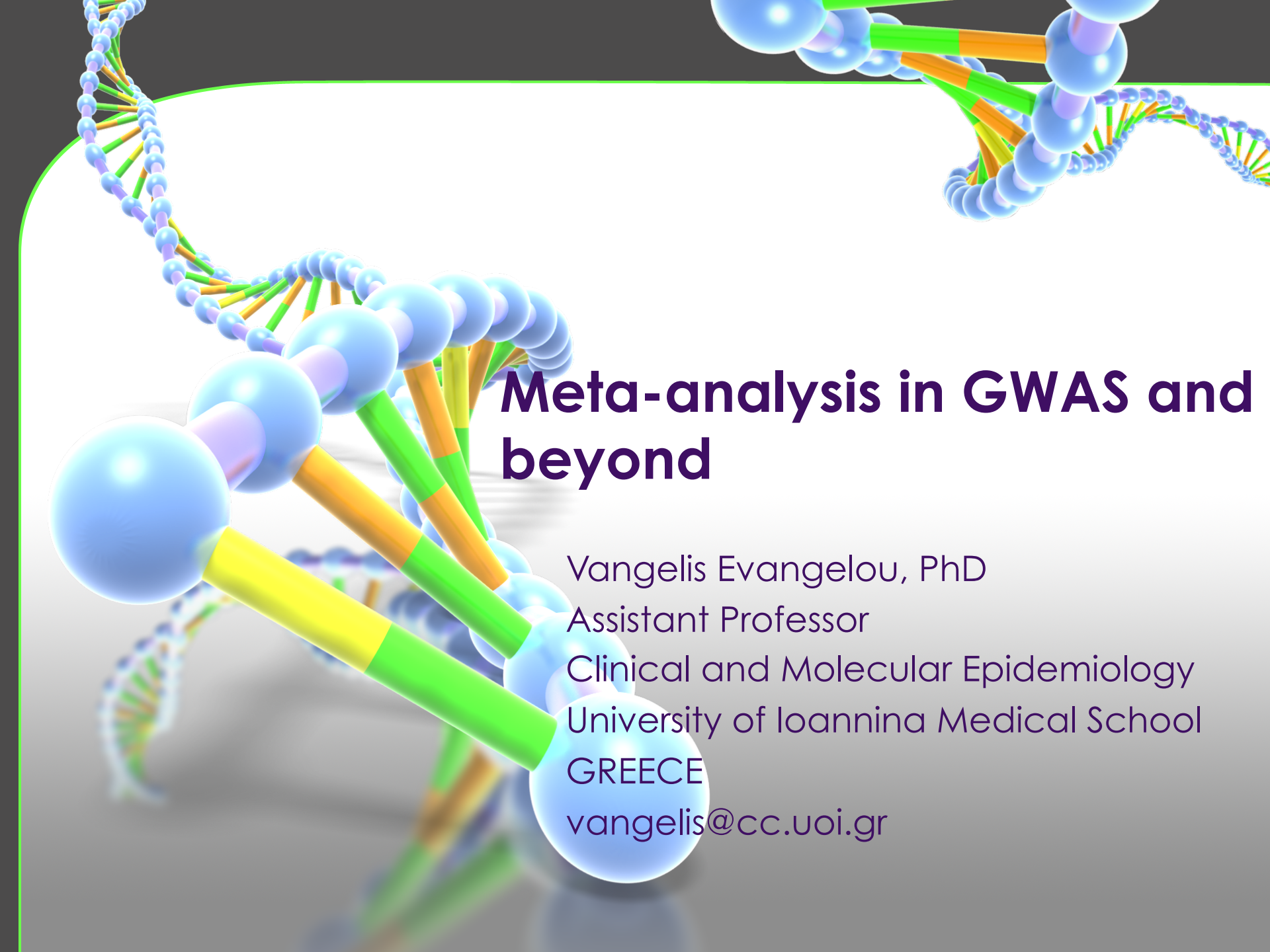


Meta-analysis in GWAS and beyond

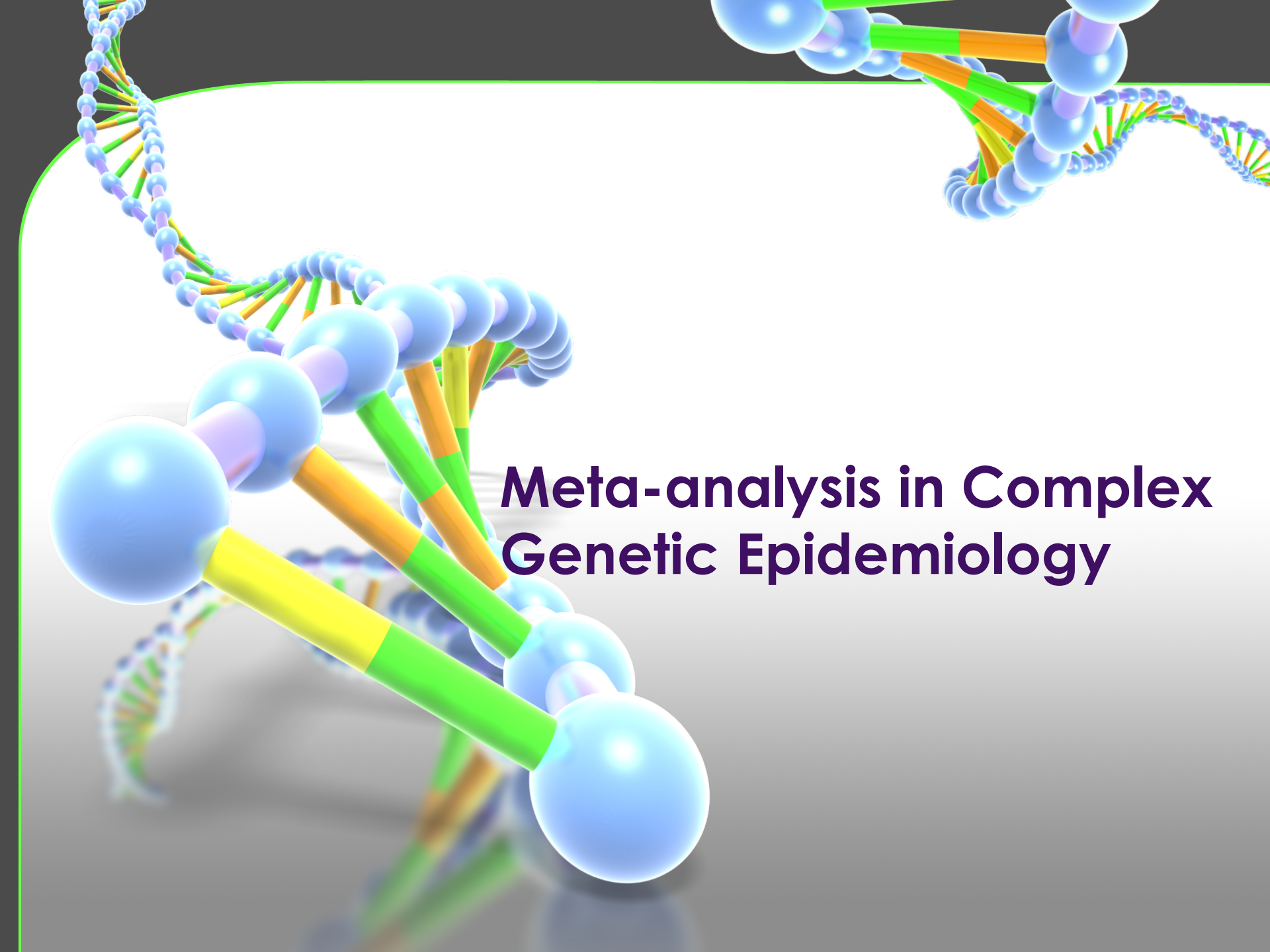
Vangelis Evangelou, PhD
Assistant Professor
Clinical and Molecular Epidemiology
University of Ioannina Medical School
GREECE
vangelis@cc.uoi.gr

GENOME-WIDE ASSOCIATION STUDIES

Meta-analysis methods for genome-wide association studies and beyond

Evangelos Evangelou[1] and John P. A. Ioannidis[2,3]

Abstract | Meta-analysis of genome-wide association studies (GWASs) has become a popular method for discovering genetic risk variants. Here, we overview both widely applied and newer statistical methods for GWAS meta-analysis, including issues of interpretation and assessment of sources of heterogeneity. We also discuss extensions of these meta-analysis methods to complex data. Where possible, we provide guidelines for researchers who are planning to use these methods. Furthermore, we address special issues that may arise for meta-analysis of sequencing data and rare variants. Finally, we discuss challenges and solutions surrounding the goals of making meta-analysis data publicly available and building powerful consortia.

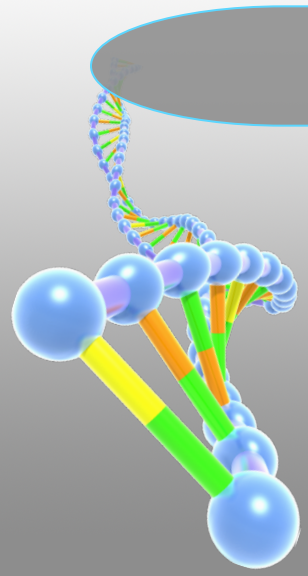


Meta-analysis in Complex Genetic Epidemiology

•Overview of the lecture

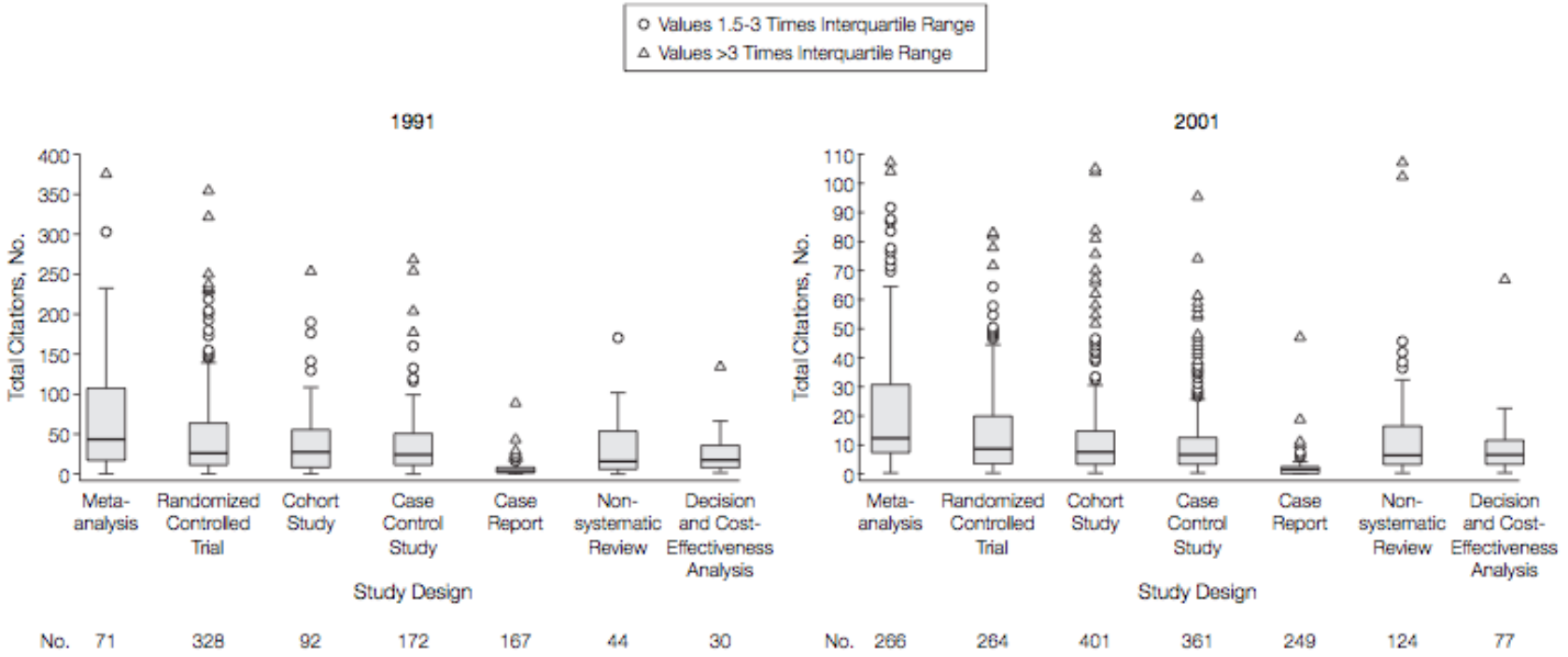
A second line of text could go here

- Meta-analysis
- Genetic epidemiology
- Meta-analysis' applications in genetics
- Future efforts



•What is the most cited design?

Figure 2. Box Plots for Total Citation Counts for Various Study Designs for Articles Published in 1991 and 2001



Box length and error bars represent the interquartile range and 1.5 times the interquartile range, respectively. Outliers (high values extending beyond 1.5 times and up to 3 times the interquartile range) are shown by circles and extremes (high values extending beyond 3 times the interquartile range) are shown by triangles. There were 25 articles with citation counts outside the depicted range (3 randomized controlled trials and 1 review in 1991; 7 meta-analyses, 10 randomized controlled trials, 3 cohort studies, 1 review in 2001). The thick lines in the boxes represent medians.

•Features of "Evidence-Based X"

- Systematic approach to information
- Information counts more than opinion
- Careful attention to study design and biases
- Emphasis on synthesis of data from diverse studies on each question of interest

•Two different approaches

A second line of text could go here

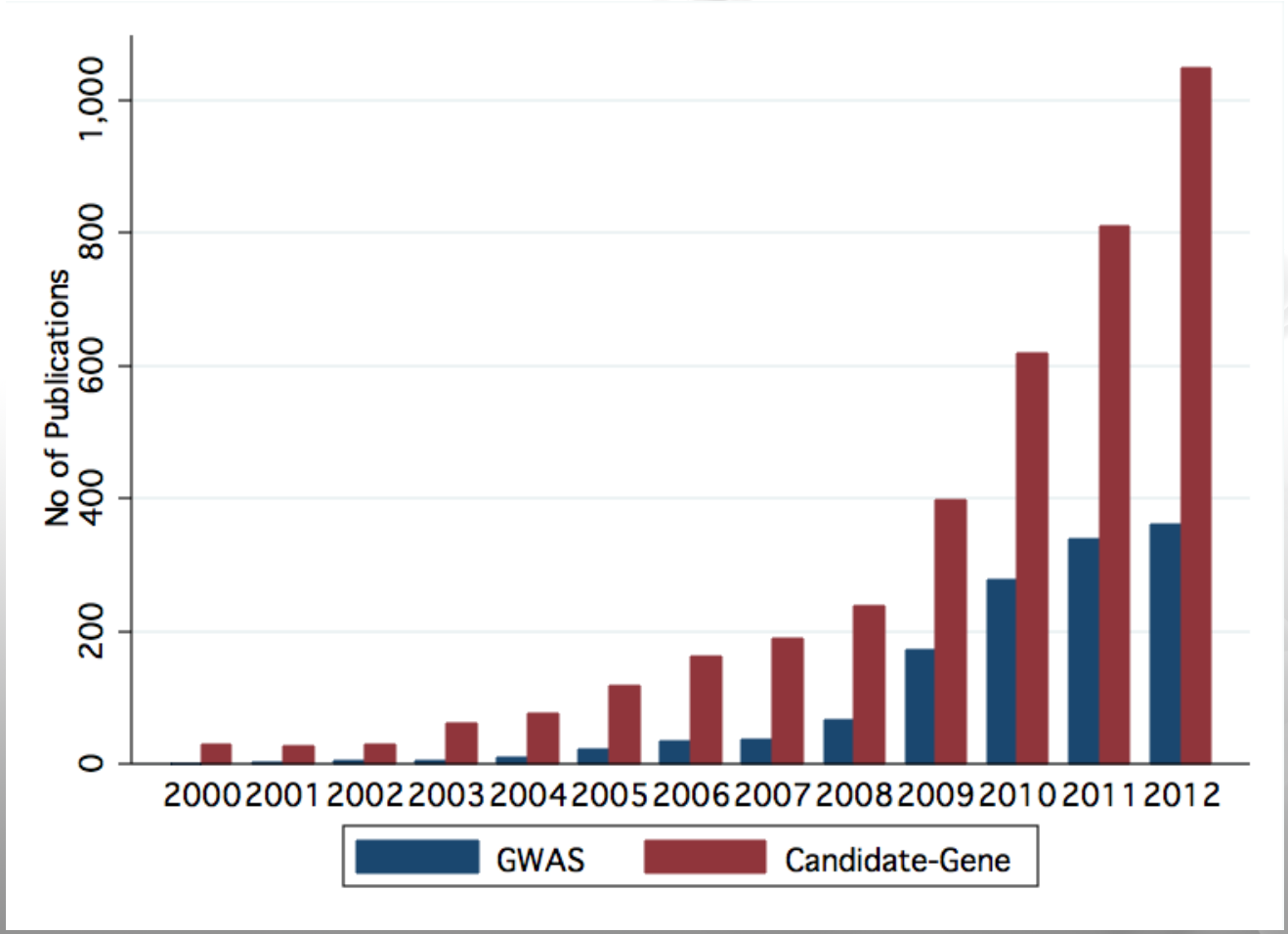
GENOME-WIDE SCANS

- Uses high-throughput genotyping technologies to essay hundred of thousand of SNPs
- Hypothesis free-agnostic approach
- Millions of associations tested simultaneously
- Adjust for multiple comparison
- Replication of most significant findings

CANDIDATE GENES

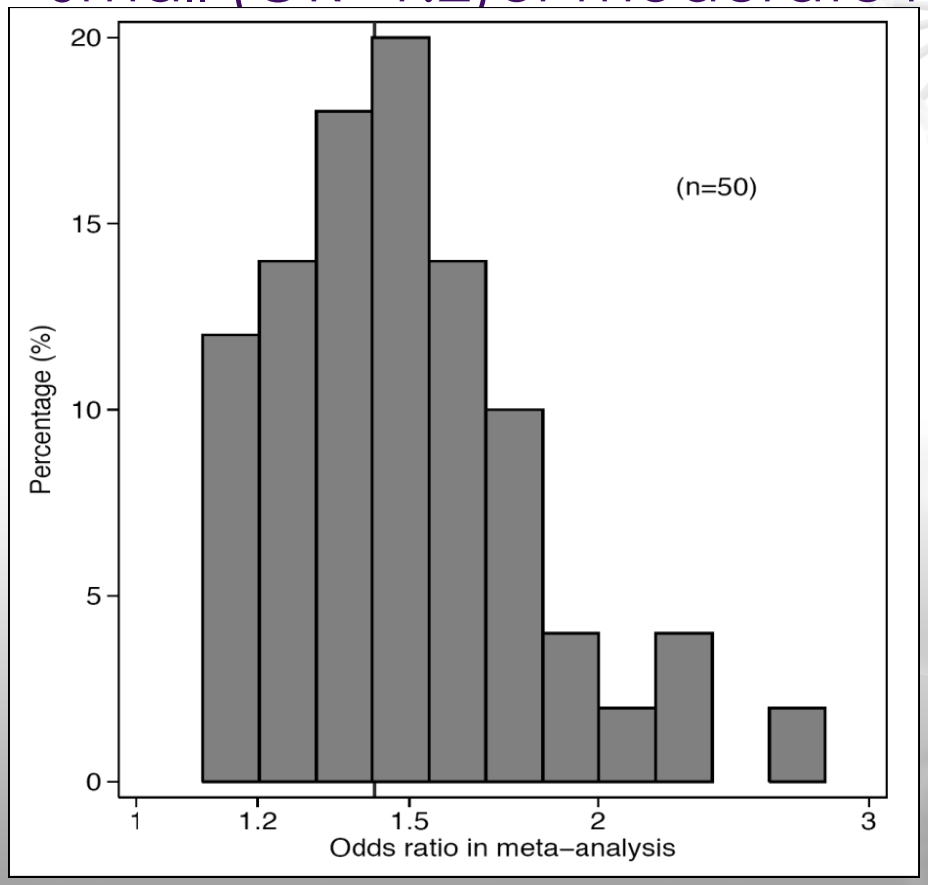
- Research based on previous hypothesis
- Biological-functional background
- Ad hoc analysis of published results
- Replication

•Meta-analysis publications



•Magnitude of genetic risks

- Small (OR<1.2)or moderate risks (1.2<OR<1.4)



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)

In the GWA era the effect sizes are even smaller

•Why meta-analysis

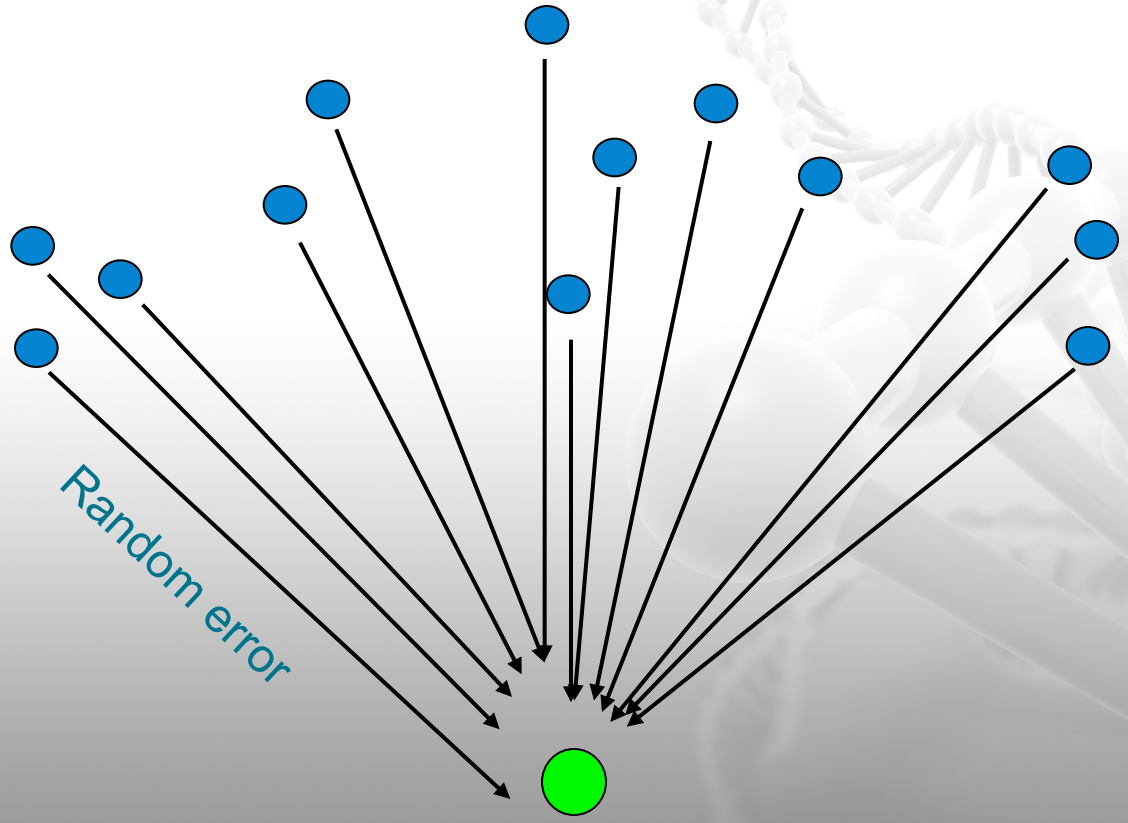
- To improve power
- To assess heterogeneity
- To explain heterogeneity
- To detect and/or exclude bias

•MODELS FOR DATA SYNTHESIS

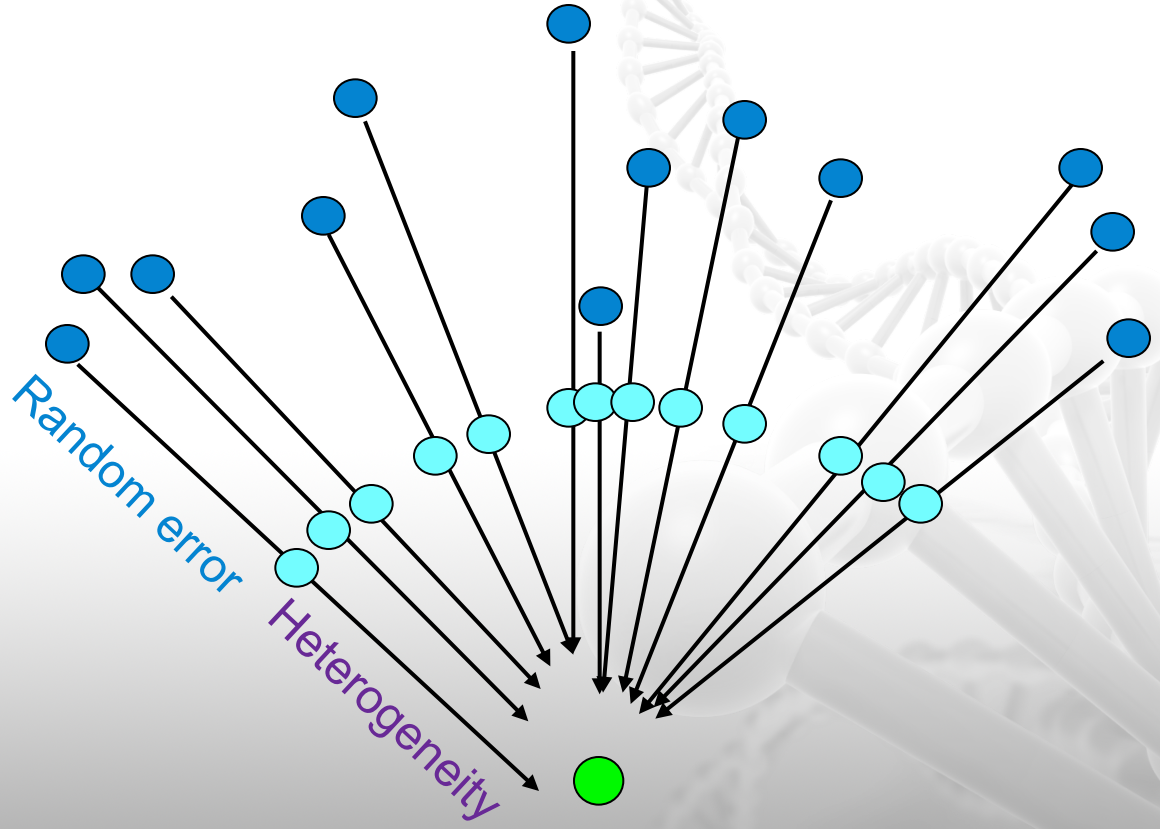
Popular methods for data synthesis

- Fixed effects; all studies have a common true effect; difference are due to chance; we try to estimate the common effect and uncertainty thereof
- Random effects; all studies have variable effects that are derived from a population of effects; we try to estimate the mean and the uncertainty of the mean effect
- P-values; most common method is Fisher's approach

•Fixed effects



•Random effects



•Types of meta-analyses

- Meta-analysis of effect sizes→ summary effect size, p-value
- Meta-analysis of z→summary p-value
- Meta-analysis of p-values →summary p-value

•Other extensions

- Bayesian approaches
  - Need to specify priors for other parameters e.g. for between study variance
  - Typically this introduces additional uncertainty
- Multivariate approaches
  - Correlated phenotypes; correlated variants
- Alternative random effects
  - Power boost in presence of heterogeneity
- Cross-phenotype meta-analysis
  - Statistical metrics for analysis of phenotypes that may share a common genetic background

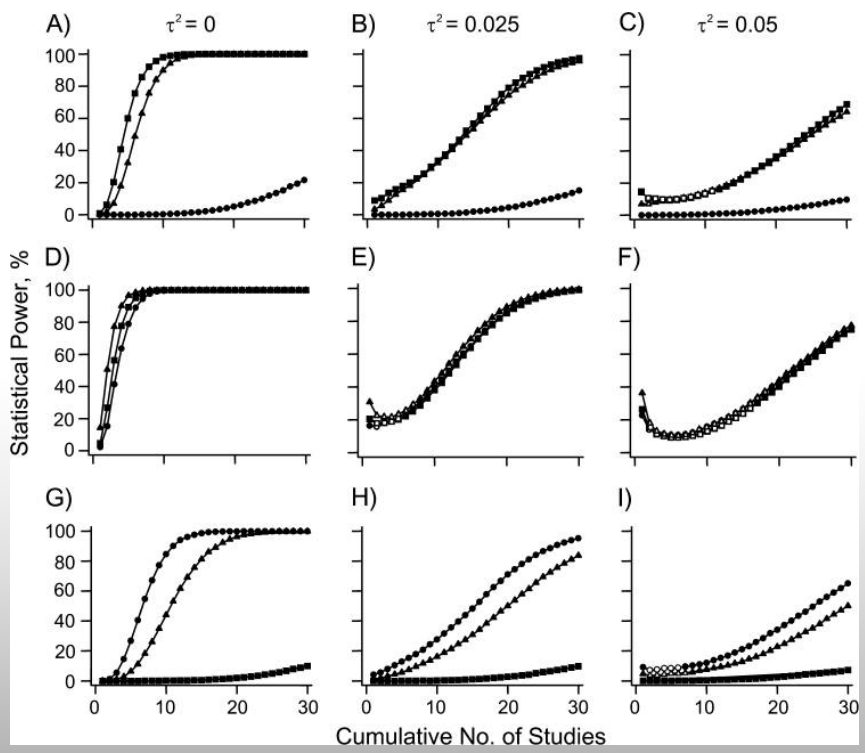
Method	Description	Advantages	Disadvantages	Main software used
P value meta-analysis	Simplest meta-analytical approach	Allows meta-analysis when effects are not available	Direction of effect is not always available; inability to provide effect sizes; difficulties in interpretation	METAL, GWAMA, R packages
Fixed effects	Synthesis of effect sizes. Between-study variance is assumed to be zero	Effects readily available through specialized software	Results may be biased if a large amount of heterogeneity exists	METAL, GWAMA, R packages
Random effects	Synthesis of effect sizes. Assumes that the individual studies estimate different effects	Generalizability of results	Power deserts in discovery efforts; may yield spuriously large summary effect estimates when there are selection biases	GWAMA, R packages
Bayesian approach	Incorporates prior assessment of the genetic effects	Most direct method for interpretation of results as posterior probabilities given the observed data	Methodologically challenging; GWAS-tailored routine software not available; subjective prior information used	R packages
Multivariate approaches	Incorporates the possible correlation between outcomes or genetic variants	Increased power can identify variants that conventional meta-analysis do not reveal using the same data sets	Computationally intensive; software not available for all analyses; some may require individual-level data	GCTA for multi-locus approaches
Other extensions	A set of different approaches that allows for the identification of multiple variants across different diseases	Summary results of previous meta-analyses can be used	May need additional exploratory analyses for the identification of variants; prone to systematic biases	Software developed by the authors of the proposed methodologies

•Fixed or random effects?

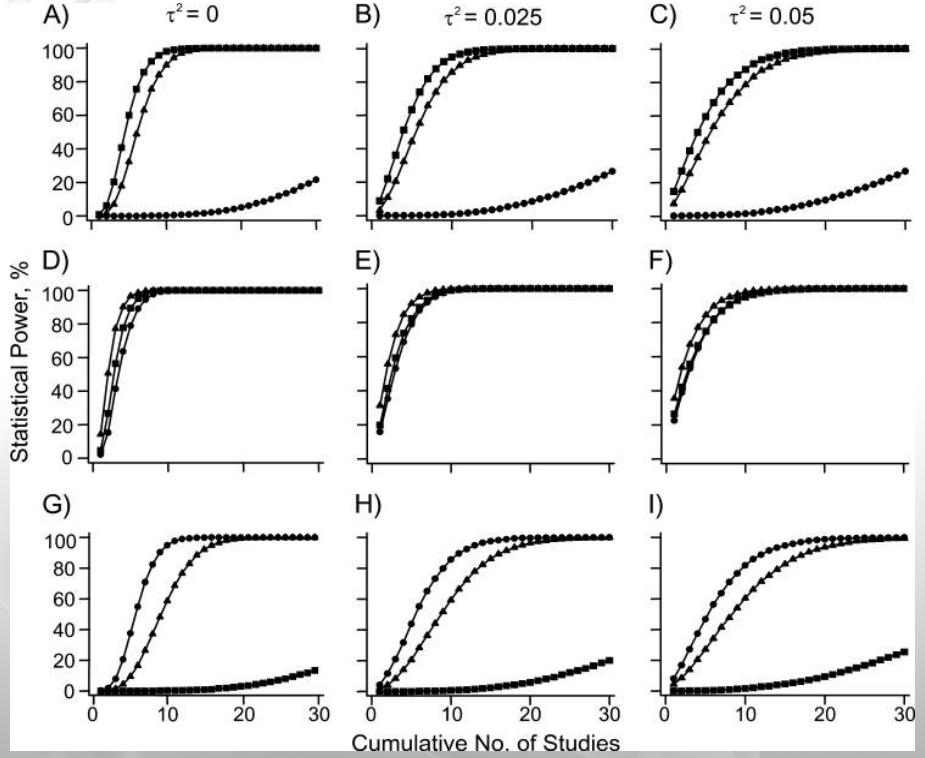
- Fixed effects are preferable for initial screenings and discovery efforts
- Random effects are preferable for generazibility applications

•Fixed or random effects?

RANDOM EFFECTS



FIXED EFFECTS



Pereira TV et al., AJE, 2009

- INFERENCE

•Inferential tools

- P-values
  - <5x10-8
  - <2x10-8 in Africans
  - ?? NGS data
- Q-values
- Bayes Factor
  - Posterior OR=BF x Prior OR
  - Differences can be observed for low MAF compared to P-values

- HETEROGENEITY

•Why meta-analysis

- To improve power
- To assess heterogeneity
- To explain heterogeneity
- To detect and/or exclude bias

•Heterogeneity

- Q statistic (chi-square based, underpowered with few studies; overpowered with over 30 studies, considered significant at the $p<0.1$ level
- Between study variance (tau square)
- I^2 statistic (Q-df/Q) (independent of the number of studies).
  - The percentage of total variation across studies that is explained beyond chance
  - 0-25% low
  - 25-50% modest
  - 50-75% large
  - >75% very large

•Sources of heterogeneity

- Poor/Differential phenotyping
- Poor genotyping
- Different SNP platforms (imputed vs genotyped SNPs)
- Unaccounted population stratification
- Genuine differences in the pertinent LD blocks across population
- Differences in environment exposures across populations

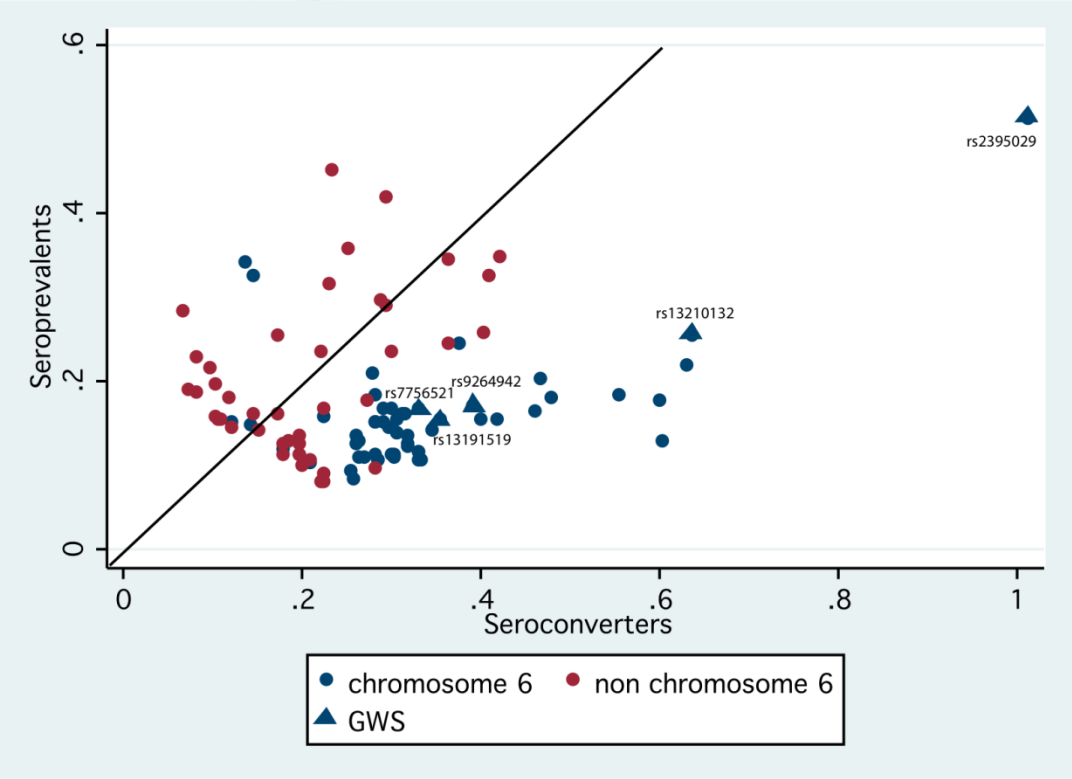
•Phenotype-based heterogeneity

HIV GWAs

Differences in the phenotype definition may affect the magnitude of the effect size

The effect estimates (of the top hits) were 0.09 log10 VL larger in seroconverters compared to seroprevalent subjects

Evangelou et al. AJE



•Ancestral-based heterogeneity

nature genetics

'Racial' differences in genetic effects for complex diseases

John P A Ioannidis[1–3], Evangelia E Ntzani[1] & Thomas A Trikalinos[1,3]

Hum Genet (2012) 131:1057–1071
DOI 10.1007/s00439-011-1124-4

ORIGINAL INVESTIGATION

Consistency of genome-wide associations across major ancestral groups

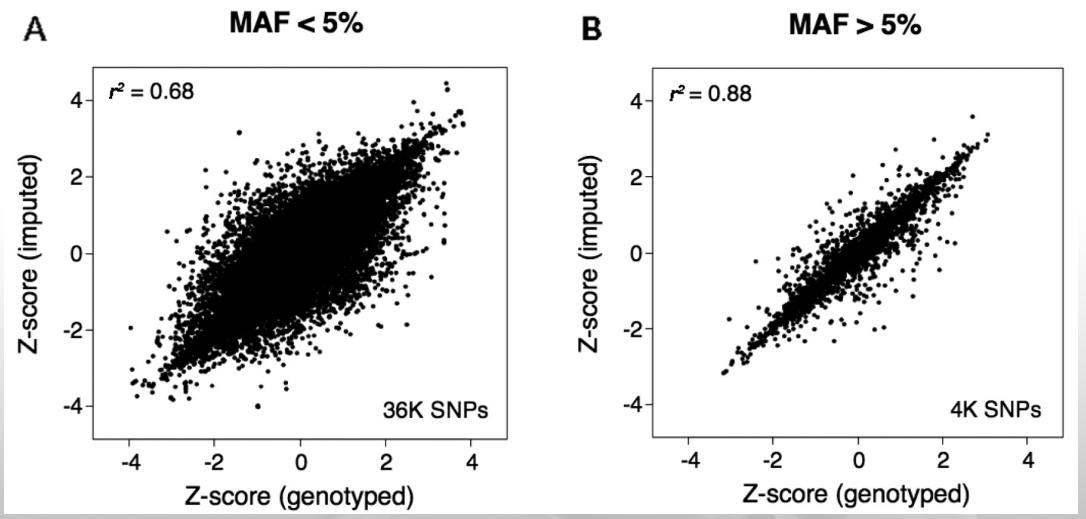
Evangelia E. Ntzani · George Liberopoulos ·
Teri A. Manolio · John P. A. Ioannidis

Trans-ethnic meta-analysis: Takes into account the similarility of allelic effect between related populations while allowing for heterogeneity between more diverse ethnic groups

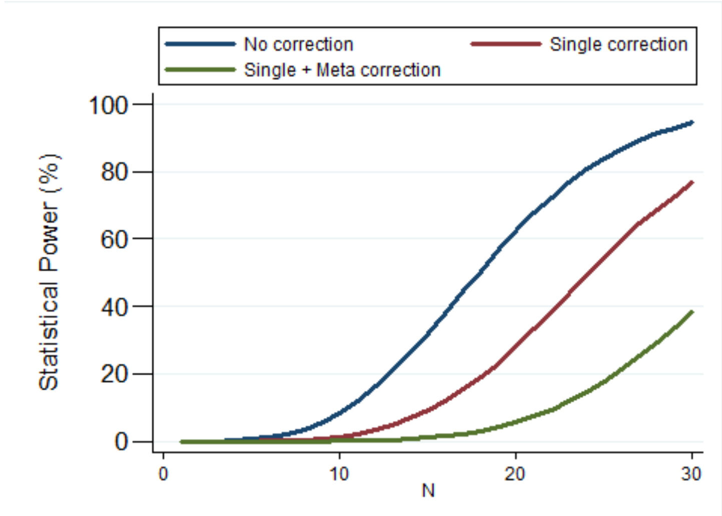
•Incorporating Imputation uncertainty

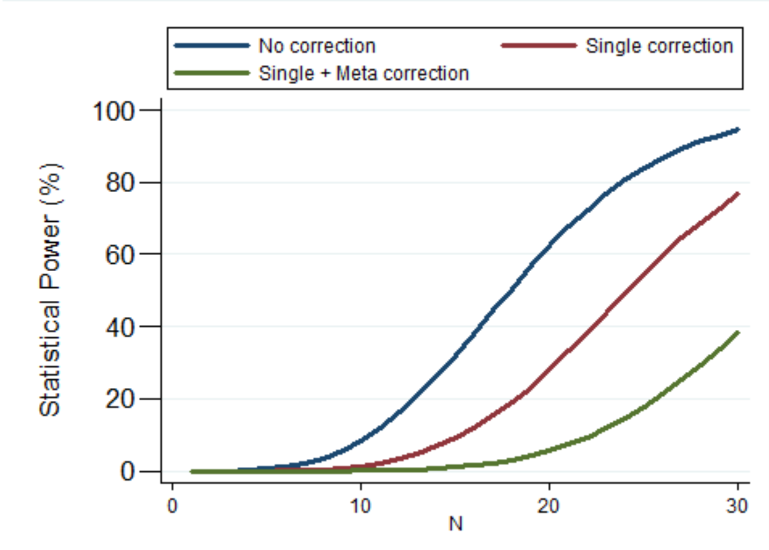
HapMap is biased towards common variants

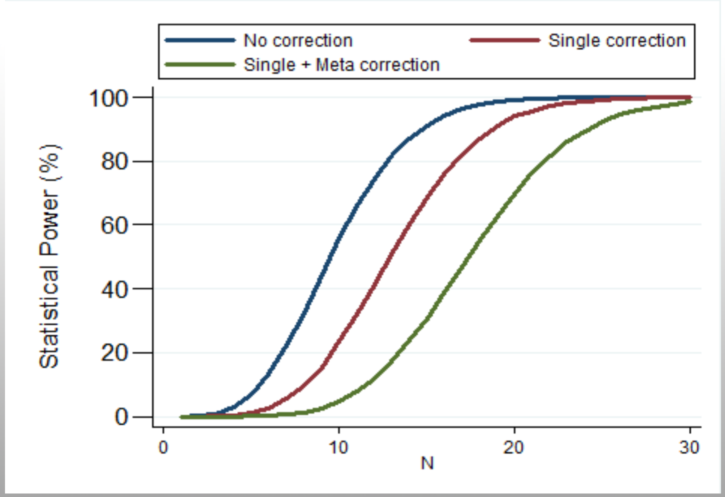
No single study should be able to disproportionally contribute in the meta-analysis

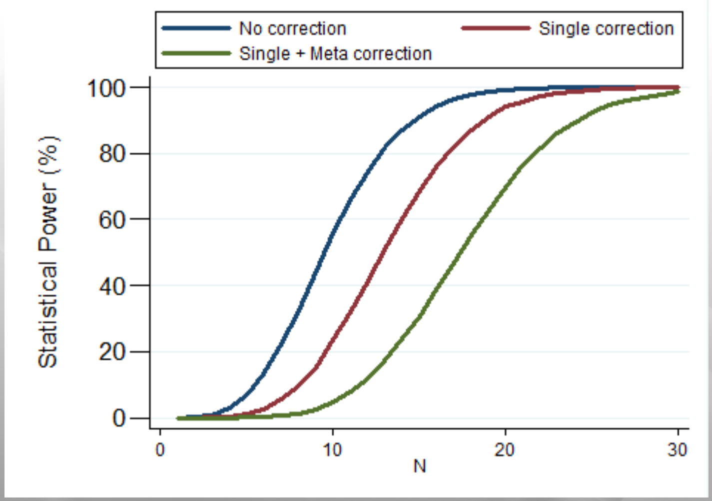


•Taking into account population stratification









•Measurement errors

- Insights from a collaborative effort

- 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

•Heterogeneity can be informative

GENE	Polymorphism	Q (df)[a] [p]	I^2 (95% CI)	Random effects OR (95% CI)	Fixed effects OR (95% CI)	Random effects p-value	Fixed effects p-value
—	rs9300039	8.38 (3) [0.039]	64% (0–86)	1.29 (1.11–1.50)	1.26 (1.15–1.37)	0.001	2.8×10^{-8}
FTO	rs8050136	12.98 (4) [0.011]	69% (0–86)	1.15 (1.06–1.25)	1.17 (1.12–1.23)	0.001	2.5×10^{-12}
PPARG	rs1801282	6.93 (4) [0.14]	42% (0–76)	1.14 (1.06–1.23)	1.13 (1.08–1.20)	0.0007	3.4×10^{-6}
CDKAL1	rs10946398	8.76 (5) [0.12]	43% (0–76)	1.13 (1.07–1.18)	1.12 (1.08–1.15)	1.2×10^{-6}	1.9×10^{-10}
SLC30A8	rs13266634	3.17 (5) [0.67]	0 (0–61)	1.13 (1.08–1.17)	1.13 (1.08–1.17)	4.1×10^{-9}	4.1×10^{-9}
CDKN2B	rs564398	3.62 (4) [0.46]	0% (0–64)	1.11 (1.06–1.15)	1.11 (1.06–1.15)	5.8×10^{-7}	5.8×10^{-7}
HHEX	rs5015480–rs1111875	6.20 (5) [0.29]	19% (0–68)	1.13 (1.08–1.17)	1.12 (1.08–1.17)	2.2×10^{-8}	3.2×10^{-10}
KCNJ11	rs5215	3.50 (4) [0.48]	0% (0–64)	1.14 (1.09–1.18)	1.14 (1.09–1.18)	9×10^{-11}	9×10^{-11}
IGF2BP2	rs4402960	7.08 (5) [0.21]	29% (0–71)	1.15 (1.10–1.20)	1.15 (1.11–1.19)	2.9×10^{-10}	1.1×10^{-15}
CDKN2B	rs10811661	4.15 (5) [0.53]	0% (0–61)	1.20 (1.15–1.25)	1.20 (1.15–1.25)	2.7×10^{-15}	2.7×10^{-15}
TCF7L2	rs7901695	1.31 (4) [0.86]	0% (0–64)	1.37 (1.32–1.43)	1.37 (1.32–1.43)	1.0×10^{-48}	1.0×10^{-48}

CI: confidence interval; OR: odds ratio

[a]df = degrees of freedom; not all markers were tested by all 3 investigations in their replication efforts, thus even with splitting the discovery and replication phases, there are fewer than 6 datasets (df = 5) for some variants.

doi:10.1371/journal.pone.0000841.t002

•Absence of heterogeneity

- Absence of statistical heterogeneity does not mean absence of clinical and biological heterogeneity
- Means almost always tremendous heterogeneity for single patients
- A single true effect at the relative scale can mean enormous differences in absolute risk differences for single patients

•Why meta-analysis

- To improve power
- To assess heterogeneity
- To explain heterogeneity
- To detect and/or exclude bias

•Small study effect

- Begg and Mazumbar test
- Egger test
- Modified regression test (Harbord test)

•Early vs late studies

- Winner's curse phenomenon
  - Early studies suggest stronger effects
  - The magnitude of the winner's curse in inversely related to the power of the study
  - Analytical methods for estimating the amount of the inflation (Zollner S et al, Am J Hum Genet) <<<evaluation of the association in additional datasets

- Proteus phenomenon
  - First study gives strongest effect ever observed soon followed by a study the least strong effect over observed

•Meta-analysis stages

- Formulation of the research question
- Identification of the eligible studies
- Synthesis of the available evidence
- Assessing and addressing potential biases
- Interpreting the results

•Two different approaches

A second line of text could go here

GENOME-WIDE SCANS

- Uses high-throughput genotyping technologies to essay hundred of thousand of SNPs
- Hypothesis free-agnostic approach
- Millions of associations tested simultaneously
- Adjust for multiple comparison
- GW significance: $5x10^{-8}$
- Replication of most significant findings

CANDIDATE GENES

- Research based on previous
- Biological-functional background
- Ad hoc analysis of published results
- Replication

UCHL1 Is a Parkinson's Disease Susceptibility Gene

Ann Neurol

UCHL-1 Is Not a Parkinson's Disease Susceptibility Gene

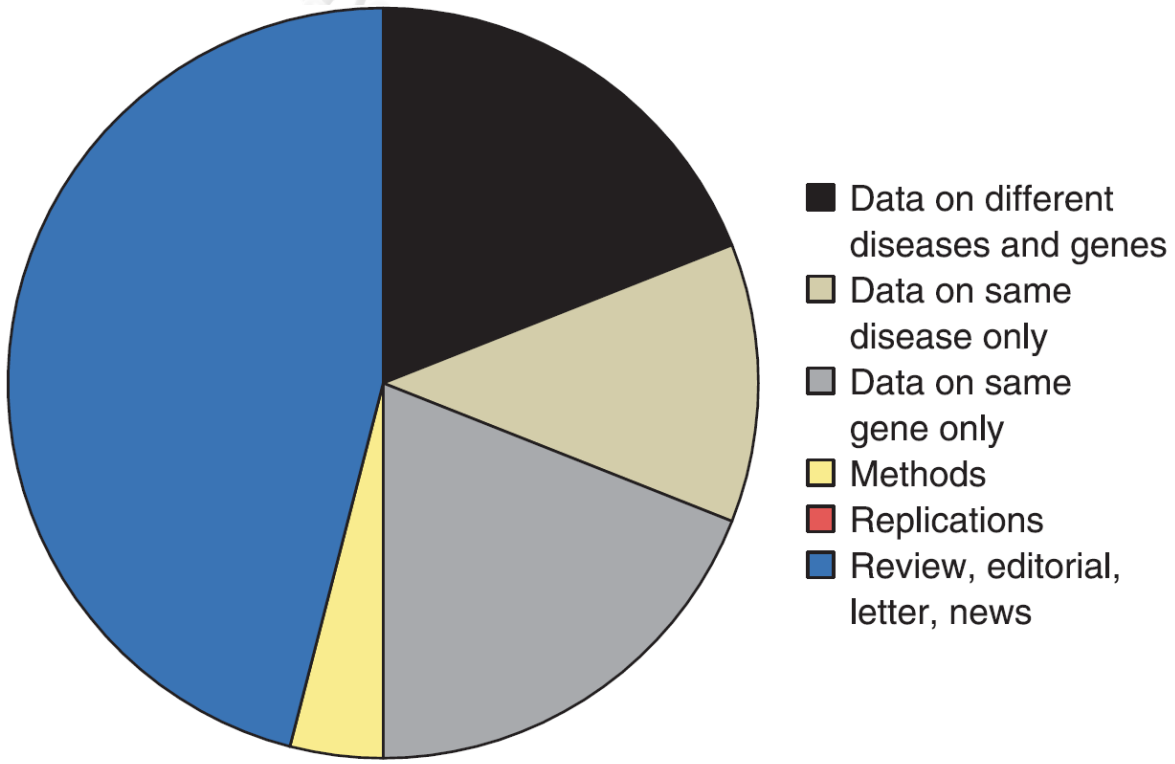
•Evolving evidence to replication

Early genetic Epidemiology

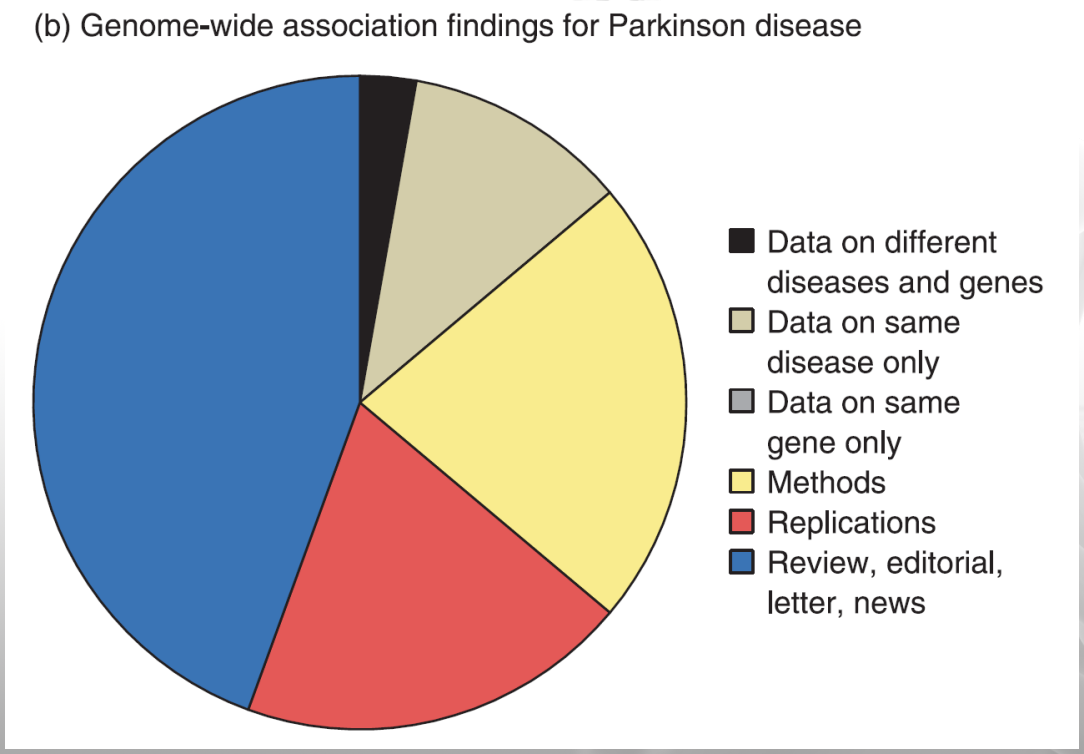
Nature 1994

TNFA associates with cerebral malaria

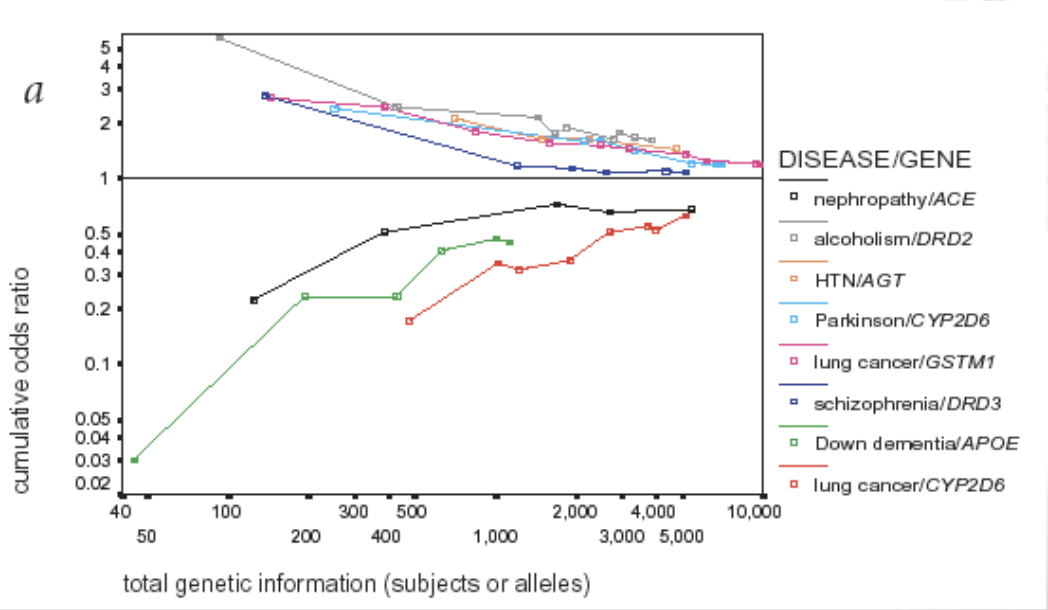
>1000 citations to-date



•Shifting attention to replication



•Non replicated-diminishing effects



Ioannidis JP et al. Nat Genet, 2001

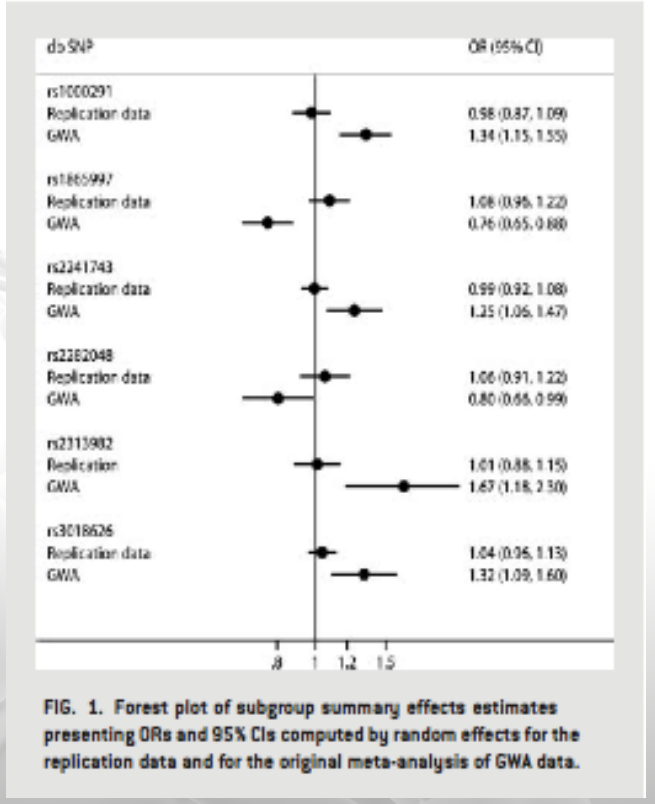


FIG. 1. Forest plot of subgroup summary effects estimates presenting ORs and 95% CIs computed by random effects for the replication data and for the original meta-analysis of GWA data.

Evangelou E et al, Am J med Genet B, 2010

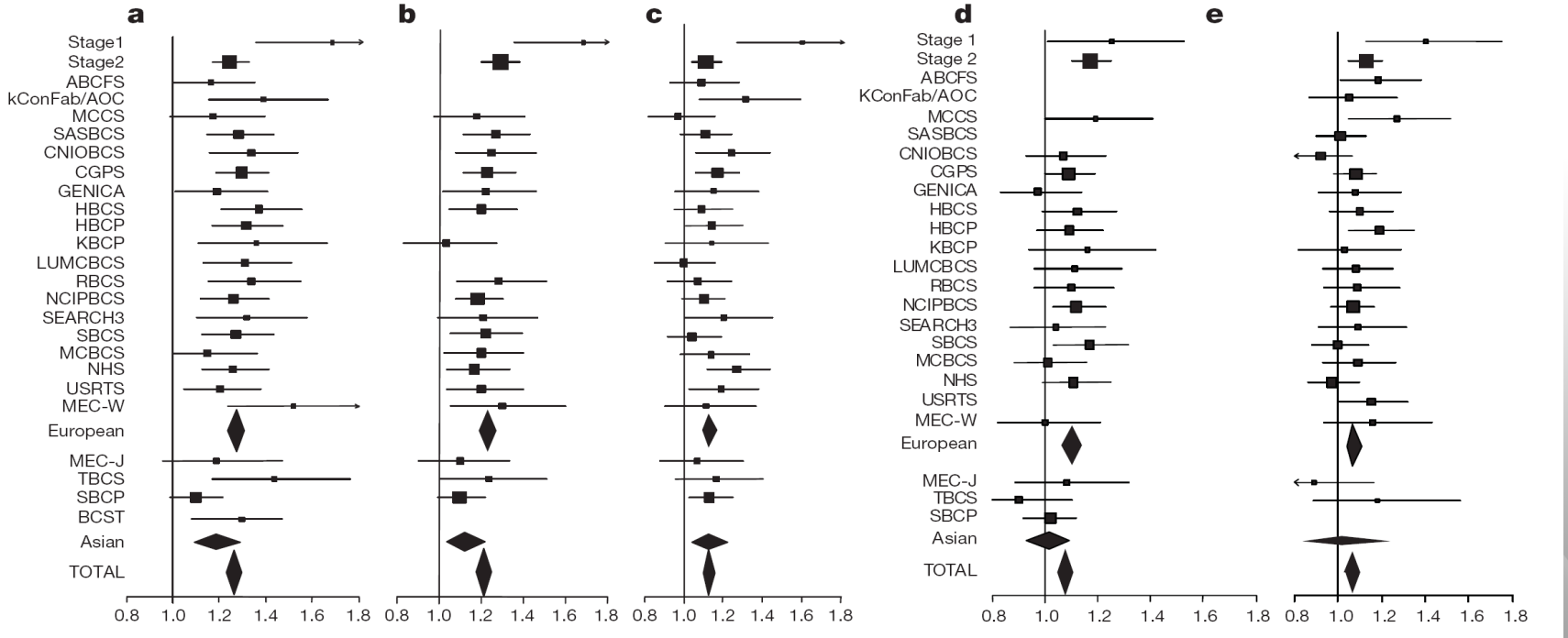
•Ultrafast replication as a sine qua non

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],



Easton et al, Nature 2007

•The Human Genome Epidemiology Network

- Global collaboration of individuals and organizations interested in accelerating the develpment of the knowledge base on genetic variation and common diseases
- HuGE reviews→ reviews and meta-analyses published on human genome epidemiology topics
- www.hugenavigator.org

•Multiple meta-analysis

- Umbrella reviews
- Field synopsis
  - Systematic, comprehensive synopsis of genetic association studies in certain diseases assembled to date
  - Criteria suggested by HugeNet Roadmap
  - Quantitative methods to derive summary effect estimates by means of meta-analyses on all variants with genotype data available
  - Systematic assessment of sources of bias

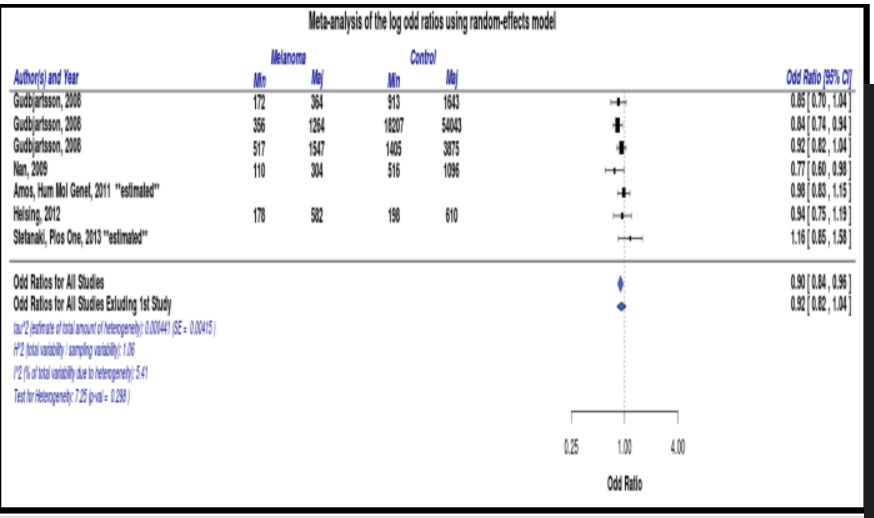
•Continuously updated databases

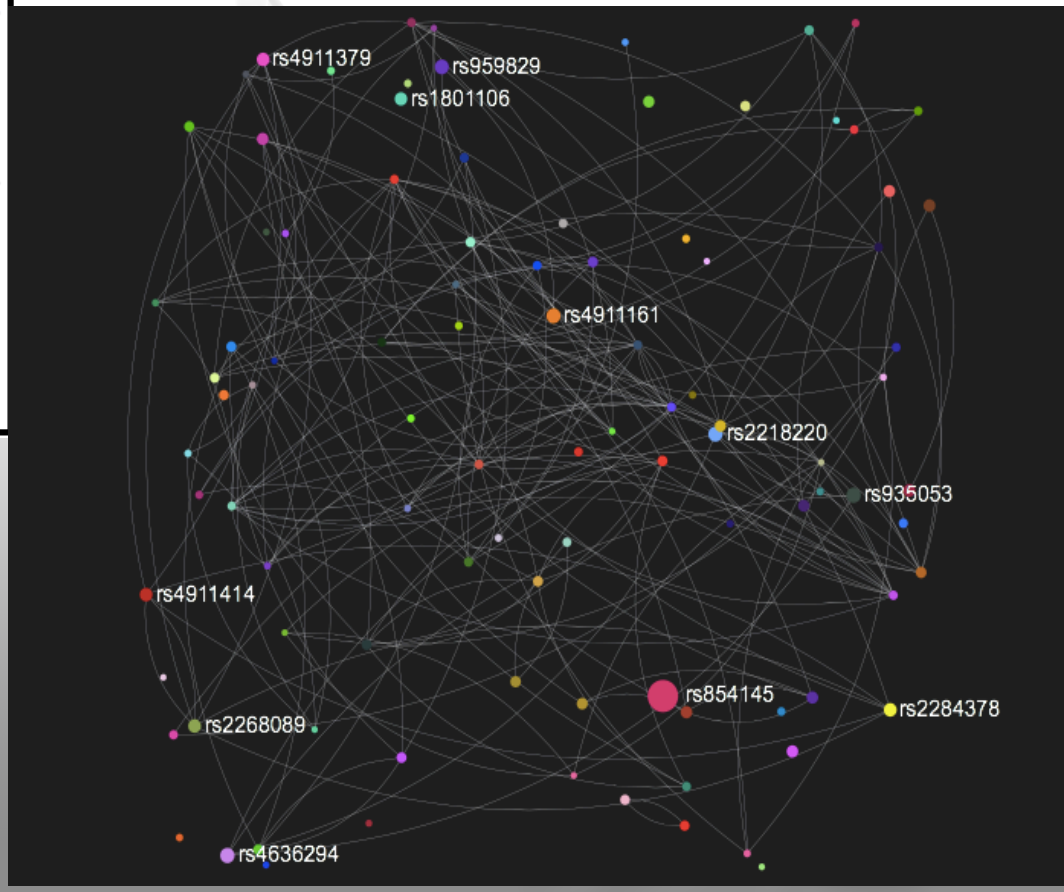
- SzGene
- AlzGene
- MelGene
- CUMAGAS

•SzGene

- 1179 publications of common genetic variants and schizophrenia

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





Chatzinasiou F et al. 2011, JNCI
Antonopoulou K, et al. 2014, JID
Athanasiadis E, et al. 2014, Database

•Candidate genes and GWAs

ARTICLE | Annals of Internal Medicine

Collaborative Meta-analysis: Associations of 150 Candidate Genes With Osteoporosis and Osteoporotic Fracture

J. Brent Richards, MD, MSc; Fotini K. Kavvoura, MD, PhD; Fernando Rivadeneira, MD, PhD; Unnur Styrkársdóttir, PhD; Karol Estrada, MSc; Bjarni V. Halldórsson, PhD; Yi-Hsiang Hsu, MD, ScD; M. Carola Zillikens, MD; Scott G. Wilson, PhD; Benjamin H. Mullin, BSc; Najaf Amin, MSc; Yurii S. Aulchenko, PhD; L. Adrienne Cupples, PhD; Panagiotis Deloukas, PhD; Serkalem Demissie, PhD; Albert Hofman, MD, PhD; Augustine Kong, PhD; David Karasik, PhD; Joyce B. van Meurs, PhD; Ben A. Oostra, PhD; Huibert A.P. Pols, MD, PhD; Gunnar Sigurdsson, MD, PhD; Unnur Thorsteinsdottir, PhD; Nicole Soranzo, PhD; Frances M.K. Williams, MD, PhD; Yanhua Zhou, MSc; Stuart H. Ralston, MD; Gudmar Thorleifsson, PhD; Cornelia M. van Duijn, PhD; Douglas P. Kiel, MD, MPH; Kari Stefansson, MD, PhD; André G. Uitterlinden, PhD; John P.A. Ioannidis, MD, PhD; and Tim D. Spector, MD, MSc, for the GEFOS (Genetic Factors for Osteoporosis) Consortium

•Grading the evidence

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

CAA	CBA	CCA
CAB	CBB	CCB
CAC	CBC	CCC

Strong evidence
Moderate evidence
Weak evidence

•GWAs and meta-analysis

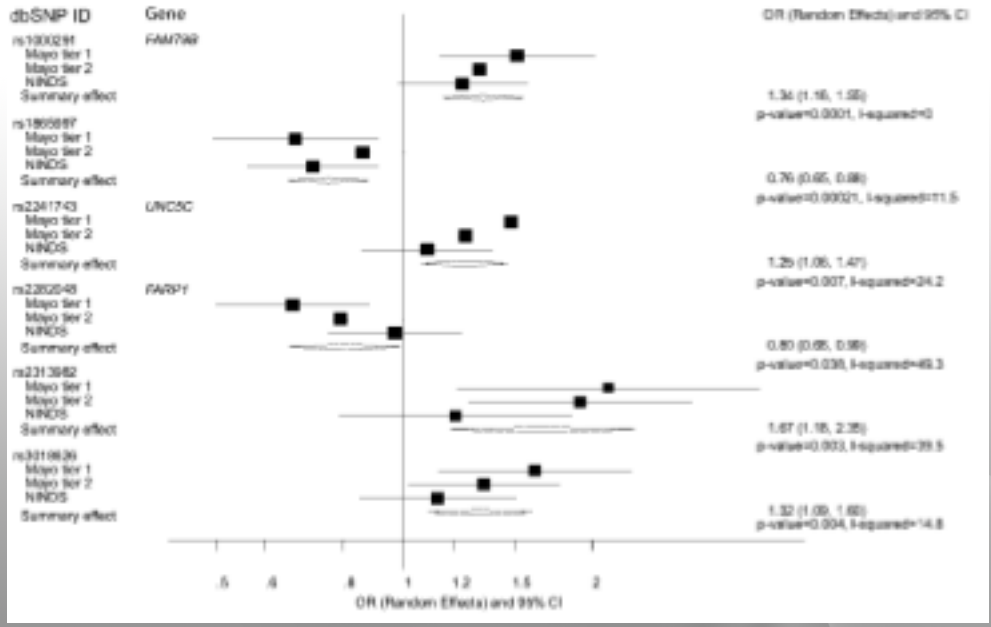
- Collaboration basics

OPEN ACCESS Freely available online

PLoS one

Meta-Analysis in Genome-Wide Association Datasets: Strategies and Application in Parkinson Disease

Evangelos Evangelou[1], Demetrius M. Maraganore[2], John P. A. Ioannidis[1,3,4]*



•GWAs and meta-analysis

- Collaboration basics

ARTICLE

doi:10.1038/nature14132

New genetic loci link adipose and insulin biology to body fat distribution

Collaborators (2208)

Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, Henneman P, Heid IM, Kizer JR, Lyytikainen LP, Fuchsberger C, Tanaka T, Morris AP, Small K, Isaacs A, Beekman M, Coassin S, Lohman K, Qi L, Kanoni S, Pankow JS, Uh HW, Wu Y, Bidulescu A, Rasmussen-Torvik LJ, Greenwood CM, Ladouceur M, Grimsby J, Manning AK, Liu CT, Kooner J, Mooser VE, Vollenweider P, Kapur KA, Chambers J, Wareham NJ, Langenberg C, Frants R, Willemsvan-vanDijk K, Oostra BA, Willems SM, Lamina C, Winkler T, Psaty BM, Tracy RP, Brody J, Chen I, Viikari J, Kähönen M, Pramstaller PP, Evans DM, St Pourcain B, Sattar N, Wood A, Bandinelli S, Carlson OD, Egan JM, Böhringer S, van Heemst D, Kedenko L, Kristiansson K, Nuotio ML, Loo BM, Harris T, Garcia M, Kanaya A, Haun M, Klopp N, Wichmann HE, Deloukas P, Katsareli E, Couper DJ, Duncan BB, Kloppenburg M, Adair LS, Borja JB, Wilson JG, Musani S, Guo X, Johnson T, Semple R, Teslovich TM, Allison MA, Redline S, Buxbaum SG, Mohlke KL, Meulenbelt I, Ballantyne CM, Dedoussis GV, Hu FB, Liu Y, Paulweber B, Spector TD, Slagboom P, Ferrucci L, Jula A, Perola M, Raitakari O, Florez JC, Salomaa V, Eriksson JG, Frayling TM, Hicks AA, Lehtimäki T, Smith GD, Siscovick DS, Kronenberg F, van Duijn C, Loos RJ, Waterworth DM, Meigs JB, Dupuis J, Richards JB, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, Ingelsson E, Saleheen D, Erdmann J, Goldstein BA, Stirrups K, König IR, Cazier JB, Johansson Å, Hall AS, Lee JY, Willer CJ, Chambers JC, Esko T, Folkersen L, Goel A, Grundberg E, Havulinna AS, Ho WK, Hopewell JC, Eriksson N, Kleber ME, Kristiansson K, Lundmark P, Lyytikäinen LP, Rafelt S, Shungin D, Strawbridge RJ, Thorleifsson G, Tikkanen E, Van Zuydam N, Voight BF, Waite LL, Zhang W, Ziegler A, Absher D, Altshuler D, Balmforth AJ, Barroso I, Braund PS, Burgdorf C, Claudi-Boehm S, Cox D, Dimitriou M, Do R, Doney AS, El Mokhtari N, Eriksson P, Fischer K, Fontanillas P, Franco-Cereceda A, Gigante B, Groop L, Gustafsson S, Hager J, Hallmans G, Han BG, Hunt SE, Kang HM, Illig T, Kessler T, Knowles JW, Kolovou G, Kuusisto J, Langenberg C, Langford C, Leander K, Lokki ML, Lundmark A, McCarthy MI, Meisinger C, Melander O, Mihailov E, Maouche S, Morris AD, Müller-Nurasyid M, Nikus K, Peden JF, Rayner NW, Rasheed A, Rosinger S, Rubin D, Rumpf MP, Schäfer A, Sivananthan M, Song C, Stewart AF, Tan ST, Thorgeirsson G, van der Schoot CE, Wagner PJ, Wells GA, Wild PS, Yang TP, Amouyel P, Arveiler D, Basart H, Boehnke M, Boerwinkle

•Meta-analysis of GWAs

- Harmonization of dataset

- Studies differ in design, sample collection, genotyping platform, association analysis methods

- Investigators should have made sensible agreements about phenotype definitions, necessary sample exclusions and appropriate covariate modeling

Minimization of spurious heterogeneity

•Have we reached our limits?

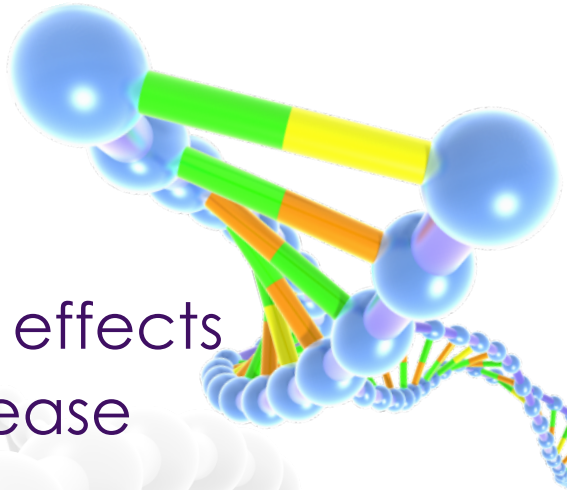
A Compendium of Genome-Wide Associations for Cancer: Critical Synopsis and Reappraisal

John P. A. Ioannidis, Peter Castaldi, Evangelos Evangelou

- GWAs is an effective tool in identifying signals with moderate effect
- Identifying risks with small effect and rare variants would require major new efforts
- iGOGs-74 new variants susceptible for different types of cancers using traditional methods but bringing all teams together

•Where are the lost variants

- Add a subtitle here



- Genetic risks are very small
- We are underpowered to detect such effects
- Gene-gene interaction- polygenic disease
- Rare variants?
- Lack of standardization of phenotypes
- Other environmental risk factors

- **Meta-analysis of sequencing data and rare variants**
  - Whole-genome and exome sequencing
  - Low event rates and zero counts
  - Merging rare variation
  - Genome-wide significance threshold for sequencing

•Merging rare variation

- Most of the methods provide a p-value or a test statistic
- Software has been developed for the synthesis of the available evidence
  - metaSKAT, rare-METAL