

*Knowledge Integration in
Public Health Genomics:
Evaluation of the Epidemiologic Evidence*

Muin J. Khoury, MD, Ph.D.

CDC National Office of Public Health Genomics



SAFER • HEALTHIER • PEOPLE™



Society

Communication and Stakeholder Engagement

Population Sciences

Genome-based Science and Technology

Humanities and Social Sciences

Last month we explored the role of behavioral and social sciences in the translation of genomic discoveries into population health benefit

Education and Training

Research

ent
n

Society

Communication and Stakeholder Engagement

Population Sciences

Genome-based Science and Technology

Humanities and Social Sciences

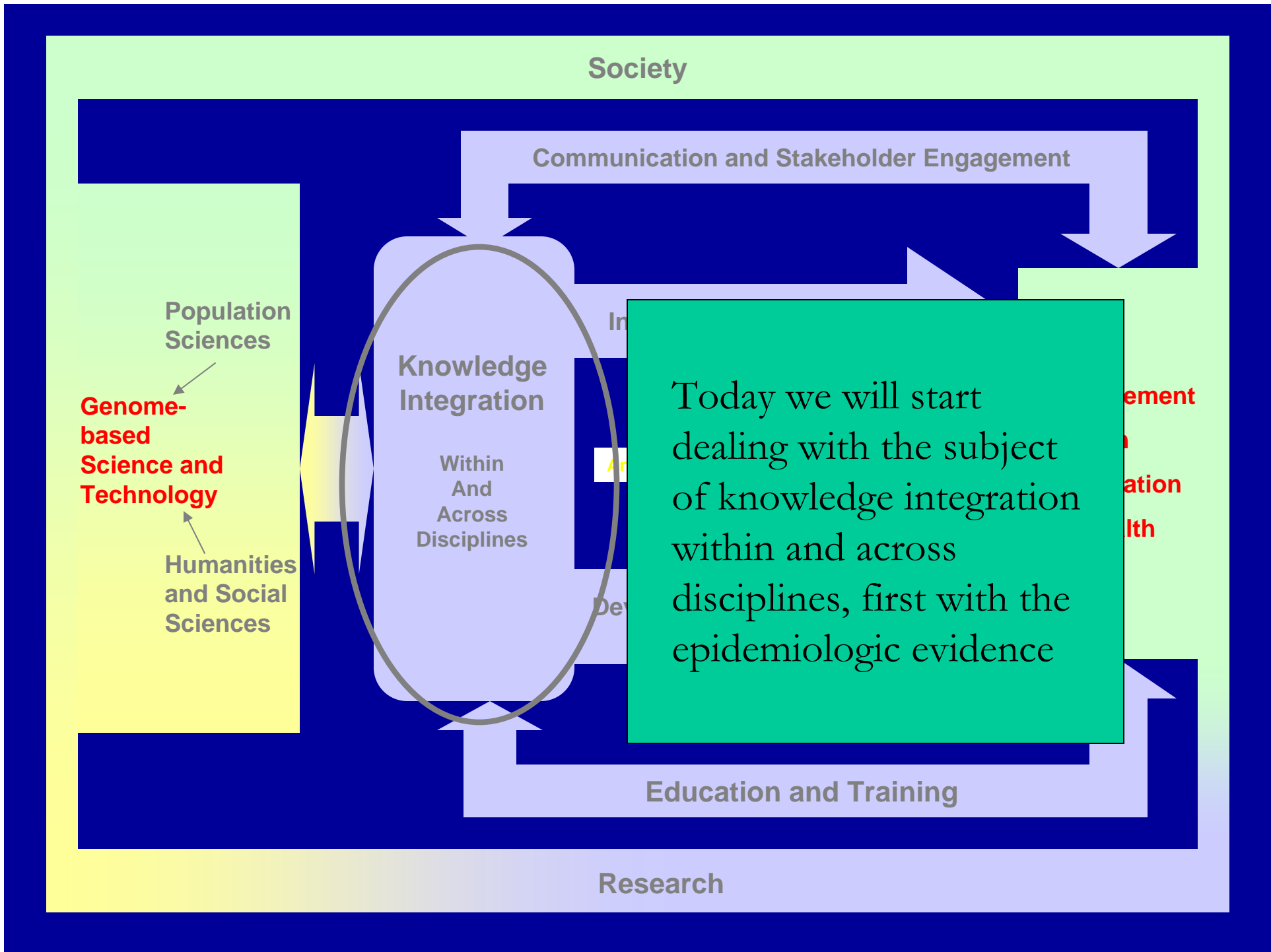
Knowledge Integration

Within And Across Disciplines

Today we will start dealing with the subject of knowledge integration within and across disciplines, first with the epidemiologic evidence

Education and Training

Research



Knowledge Integration

The activity that we call knowledge integration is the driving force or 'engine house' of the enterprise

It is the process of selecting, storing, collating, analysing, integrating and disseminating information both within and across disciplines for the benefit of population health and includes methodological development

It is the means by which information is transformed into knowledge

Interdisciplinarity is a key feature

Outline

- Introduction-Why Integrate?
- HuGENet Road Map
- Literature Scanning, Reporting, Synthesis and Network Collaboration
- Developing the Knowledge Base and Causal Inference

Why Integrate?

ORIGINAL CONTRIBUTION

JAMA April 11, 2007

Nonvalidation of Reported Genetic Risk Factors for Acute Coronary Syndrome in a Large-Scale Replication Study

Thomas M. Morgan, MD

Harlan M. Krumholz, MD, MS

Richard P. Lifton, MD, PhD

John A. Spertus, MD, MPH

COMPELLING EVIDENCE FROM twin and epidemiological studies suggests a genetic basis for atherosclerotic heart disease and acute coronary syn-

Context Given the numerous, yet inconsistent, reports of genetic variants being associated with acute coronary syndromes (ACS), there is a need for comprehensive validation of ACS susceptibility genotypes.

Objective To perform an extensive validation of putative genetic risk factors for ACS.

Design, Setting, and Participants Through a systematic literature search of articles published before March 10, 2005, we identified genetic variants previously reported as significant susceptibility factors for atherosclerosis or ACS. Restricting our analysis to white patients to reduce confounding from racial admixture, we identified 811 patients who presented from March 2001 through June 2003 with ACS at 2 Kansas City, Mo, university-affiliated hospitals. During 2005-2006, we genotyped the 811 patients along with 650 age- and sex-matched controls for 95 variants in 70 genes and attempted to replicate

Why Integrate?

Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when

factors that influence this problem and some corollaries thereof.

Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship

Why Integrate?

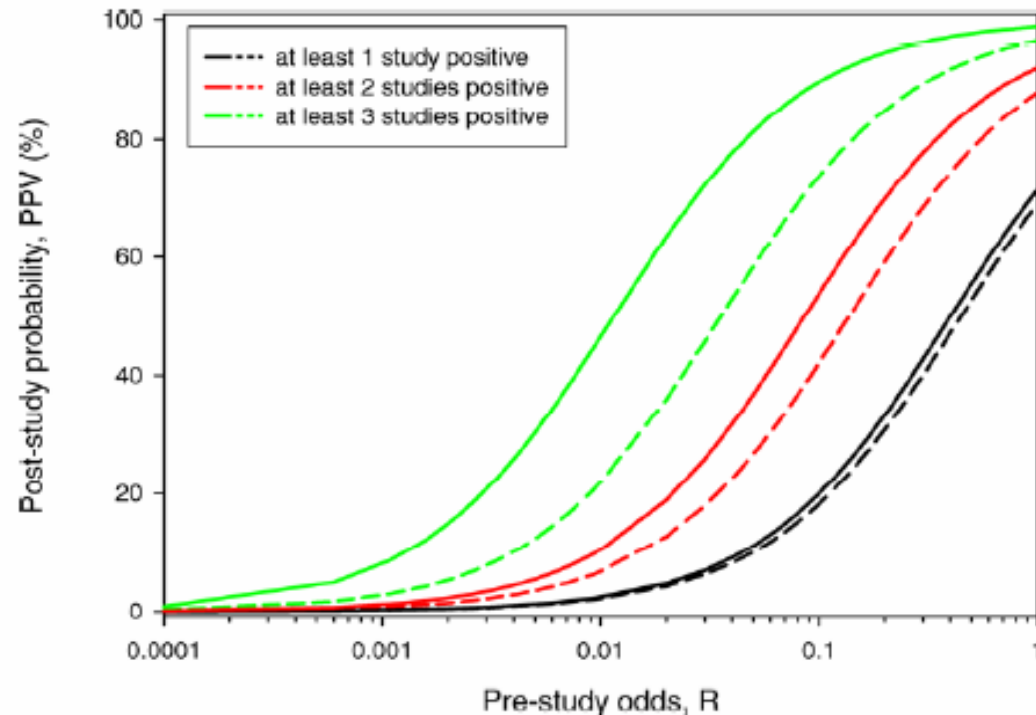
Essay

Most Published Research Findings Are False— But a Little Replication Goes a Long Way

Ramal Moonesinghe*, Muin J. Khoury, A. Cecile J. W. Janssens

We know there is a lot of lack of replication in research findings, most notably in the field of genetic associations [1–3]. For example, a survey of 600 positive associations between gene variants and common diseases showed that out of 166 reported associations studied three or more times, only six were replicated consistently [4]. Lack of replication results from a number of factors such as publication bias, selection bias, Type I errors, population stratification (the mixture of individuals from heterogeneous genetic backgrounds), and lack of statistical power [5].

In a recent article in *PLoS Medicine*, John Ioannidis quantified the theoretical basis for lack of replication by deriving the positive predictive value (PPV) of the truth of a research finding on the basis of a combination of factors. He showed elegantly that

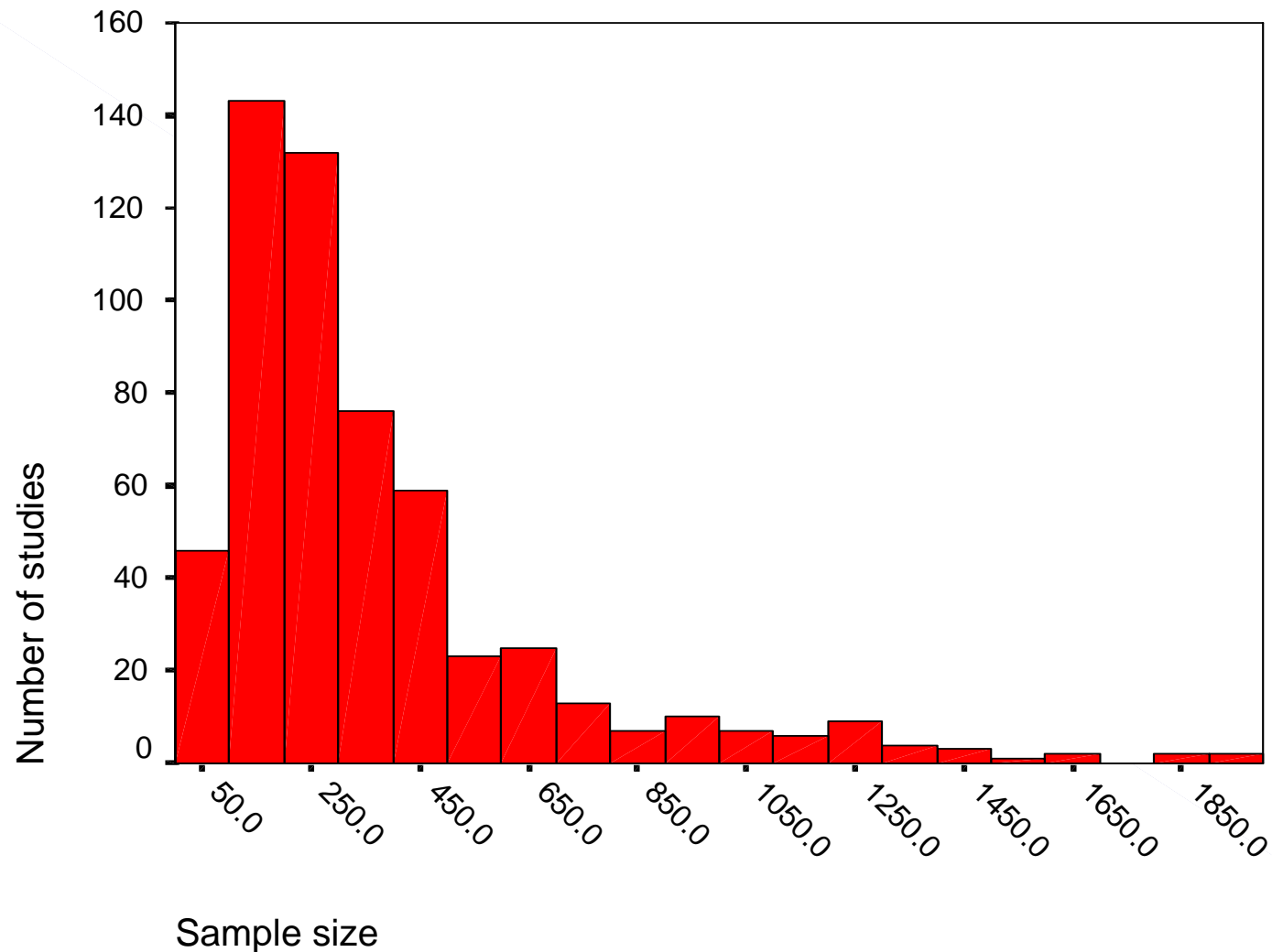


doi:10.1371/journal.pmed.0040028.g001

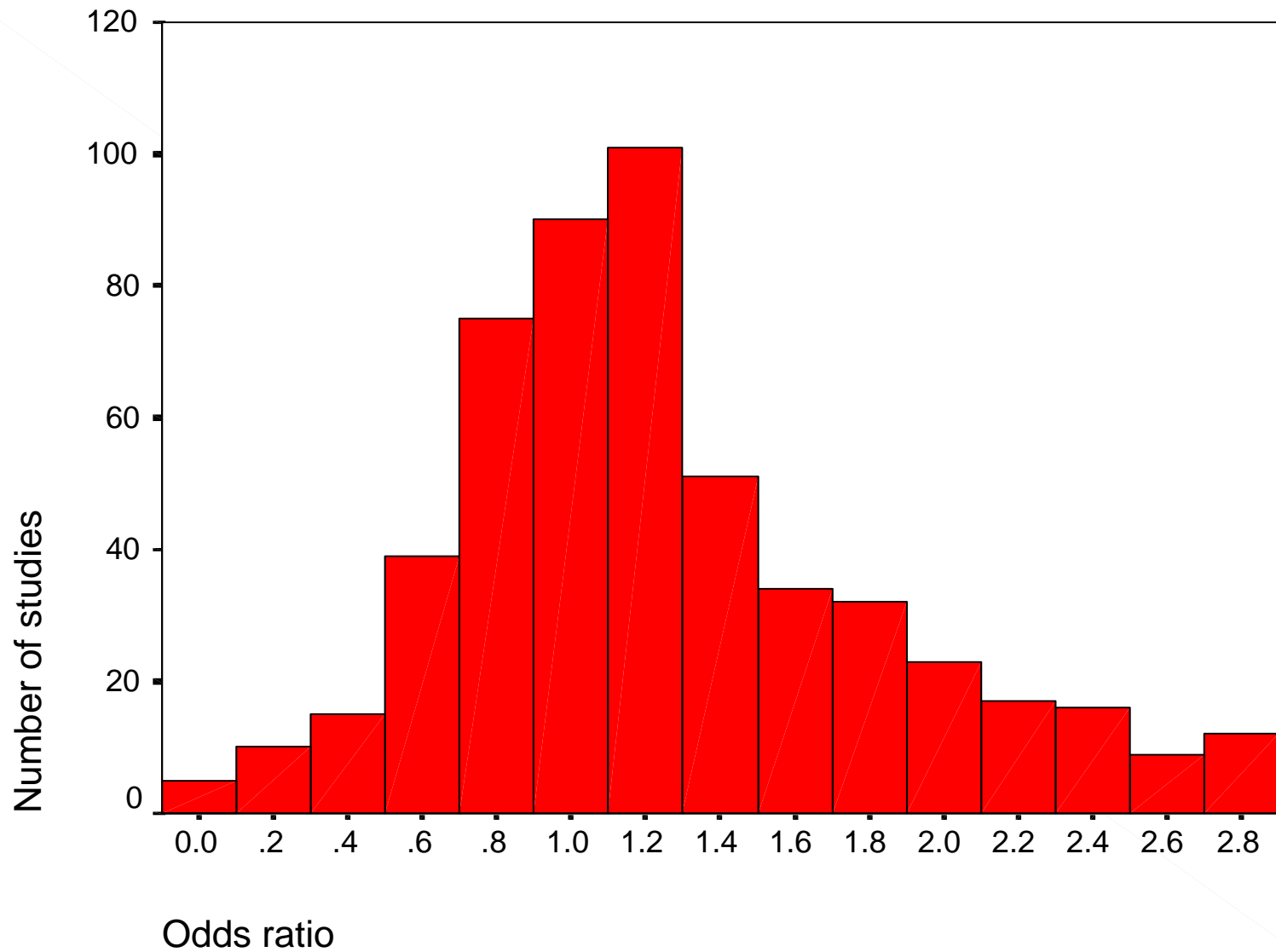
Why Integrate?

- Unmanageable amounts of emerging information
- Small sample size of individual studies
- Small effect size of individual studies
- Assess replication of associations
- Assess and explain heterogeneity-interactions
- Build the knowledge base: ‘what we know and what we don’t know’
- Produce information needed to calculate risks for use in clinical medicine and public health

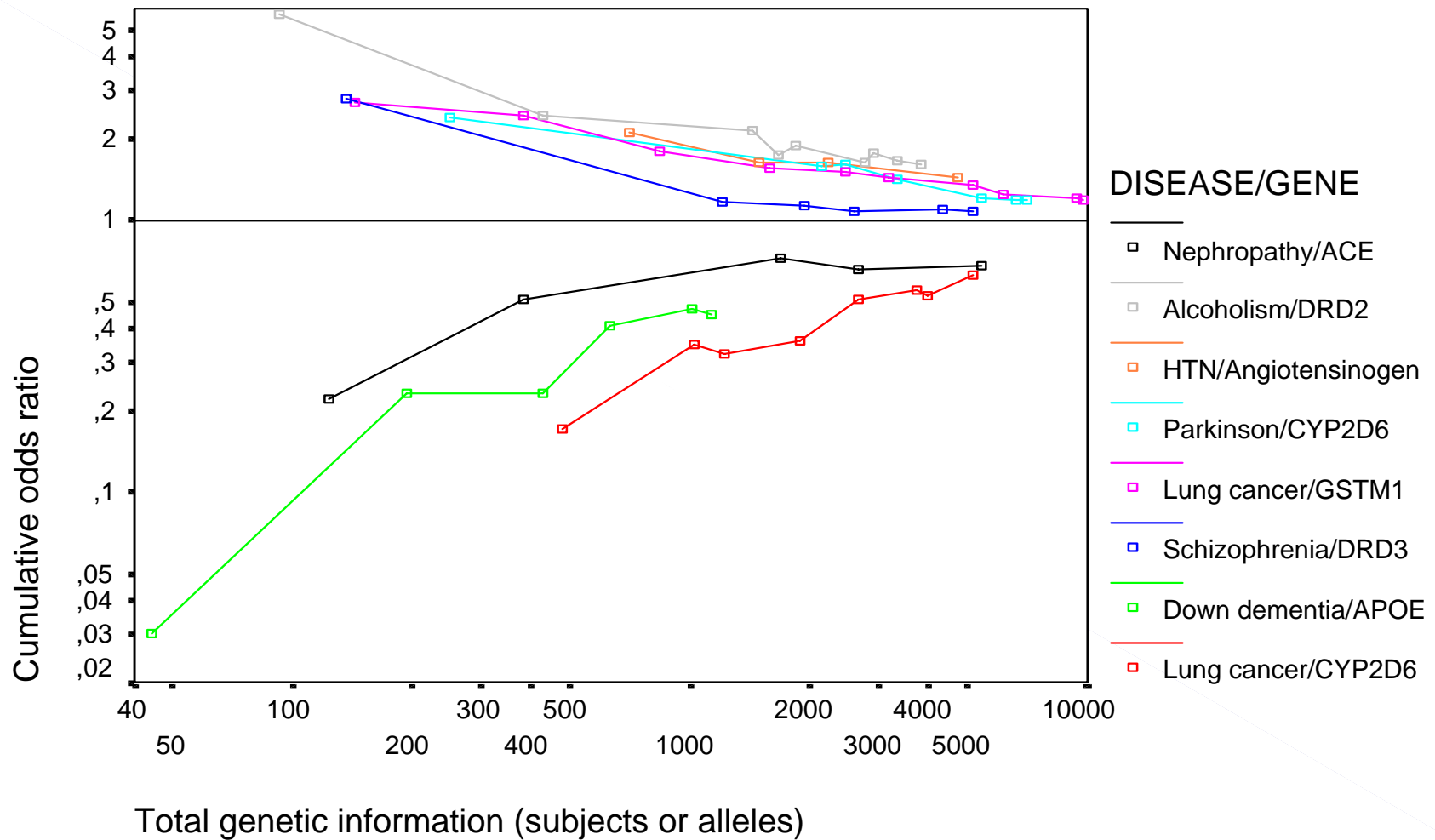
Small Sample Sizes of Individual Studies



Small Effect Sizes in Individual Studies



Evolving Genetic Associations: Effects that Diminish Over Time



Ioannidis et al, Nature Genetics 2001

Outline

- Introduction-Why Integrate?
- **HuGENet Road Map**
- Literature Scanning, Reporting, Synthesis and Network Collaboration
- Developing the Knowledge Base and Causal Inference

Options for Integration of Information

- Single, all-absorbing mega-studies (e.g. proposed US cohort on genes and environment)
- Meta-analyses of group data
- Meta-analyses of individual participant data (pooled analysis)

- All of these designs are unlikely to be successful unless they allow for evolving (often rapidly evolving) evidence

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
 - 4 coordinating centers
 - Dozens of networks
 - Hundreds of collaborators
 - 10 collaborating journals

Address <http://www.cdc.gov/genomics/hugenet/default.htm> Go Links

CDC National Office of Public Health Genomics

EVENTS TRAINING FUNDING LINKS SEARCH GENOMICS

Human Genome Epidemiology Network

MAIN MENU home > HuGENet™

- * NOPHG Home
- * Weekly Update
- * Frequently Asked Questions
- * CDC Activities
- * Family History
- * Genomics in Practice
- * Genetic Testing
- * Population Research
- * HuGENet™
- * General Public
- * Public Health Perspectives

Welcome to **HuGENet™**

Human Genome Epidemiology Network, or HuGENet™ is a global collaboration of individuals and organizations committed to the assessment of the impact of human genome variation on population health and how genetic information can be used to improve health and prevent disease.

• Message from Dr. Muin Khoury

- ▶ [Learn More About HuGENet™ Purpose, Goals and Activities](#)
- ▶ [Learn About HuGENet™ Coordinating Centers](#)
- ▶ [Learn How to Become a HuGENet™ Collaborator](#)

Human Genome Epidemiology Network (HuGENet)



www.cdc.gov/genomics/hugenet

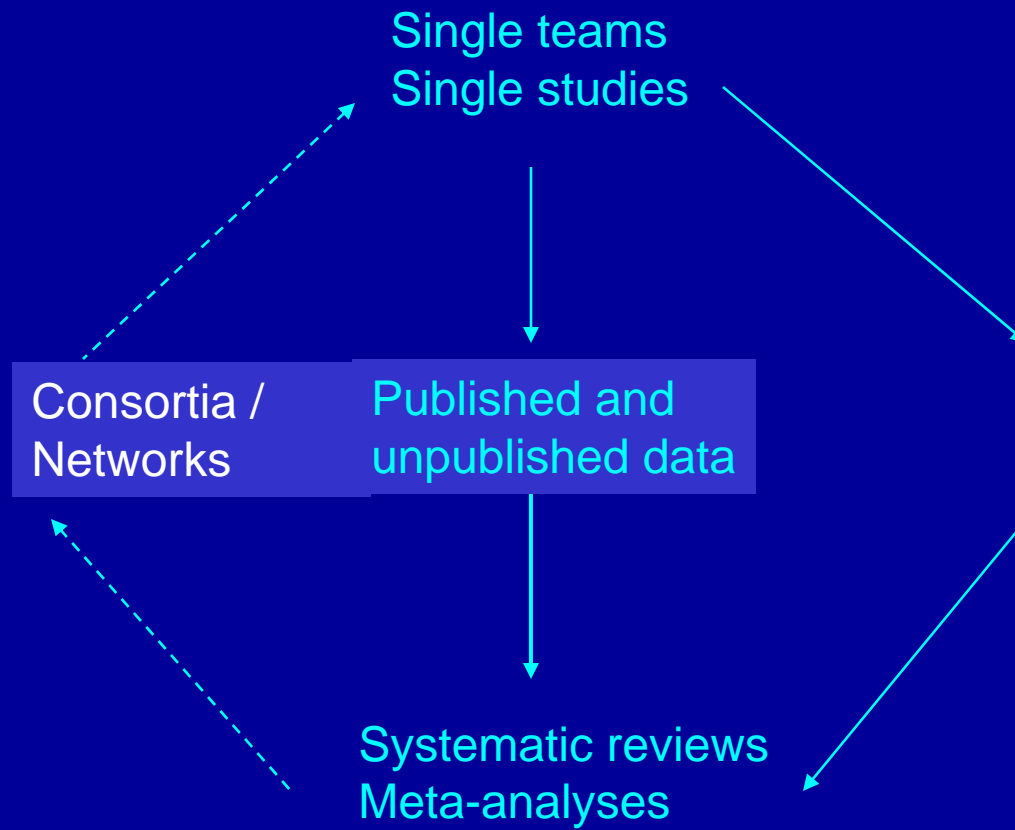
<http://www.hugenet.ca> HuGENet Canada

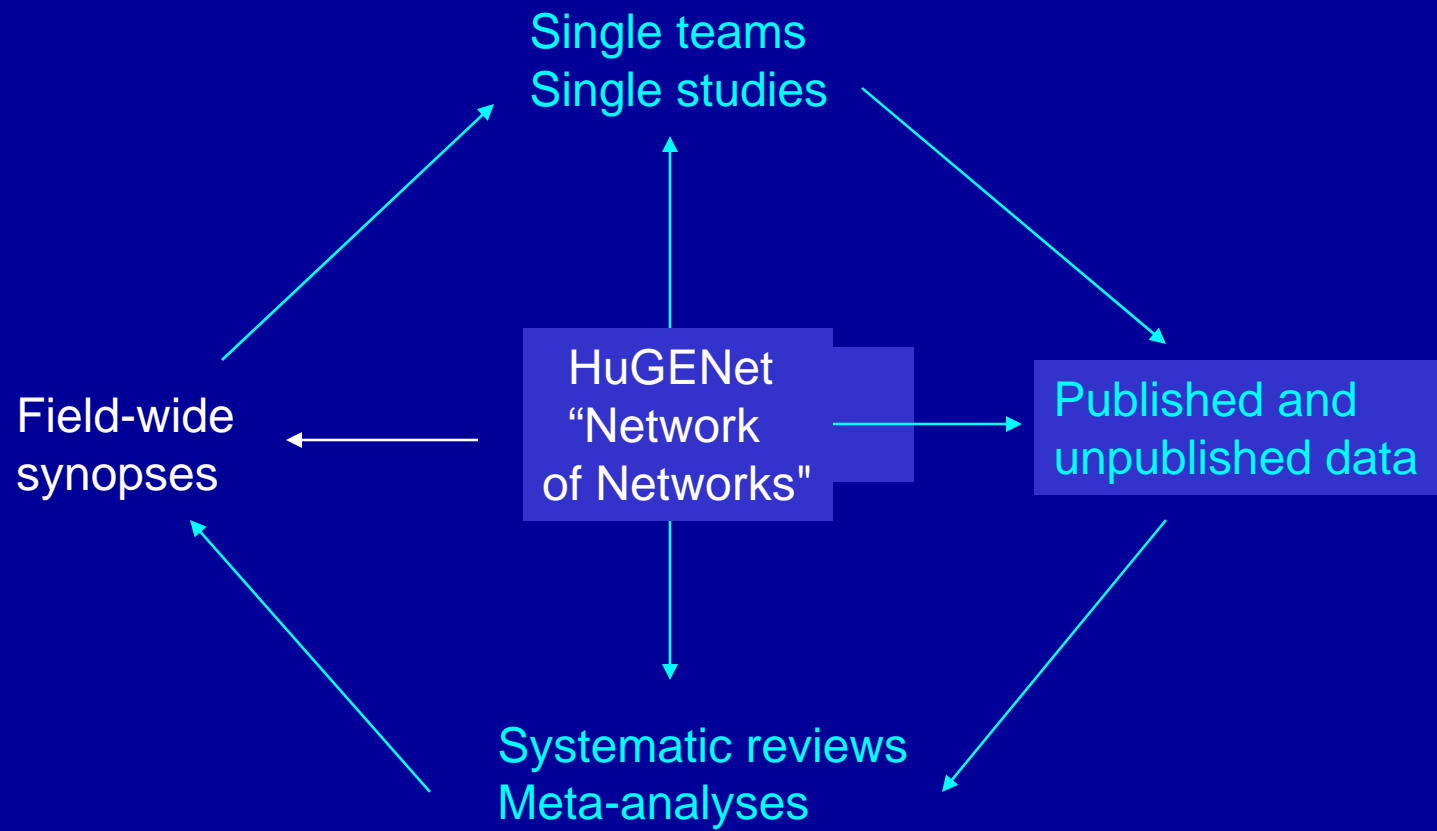
<http://www.hugenet.org.uk> UK HuGENet Coordinating Centre

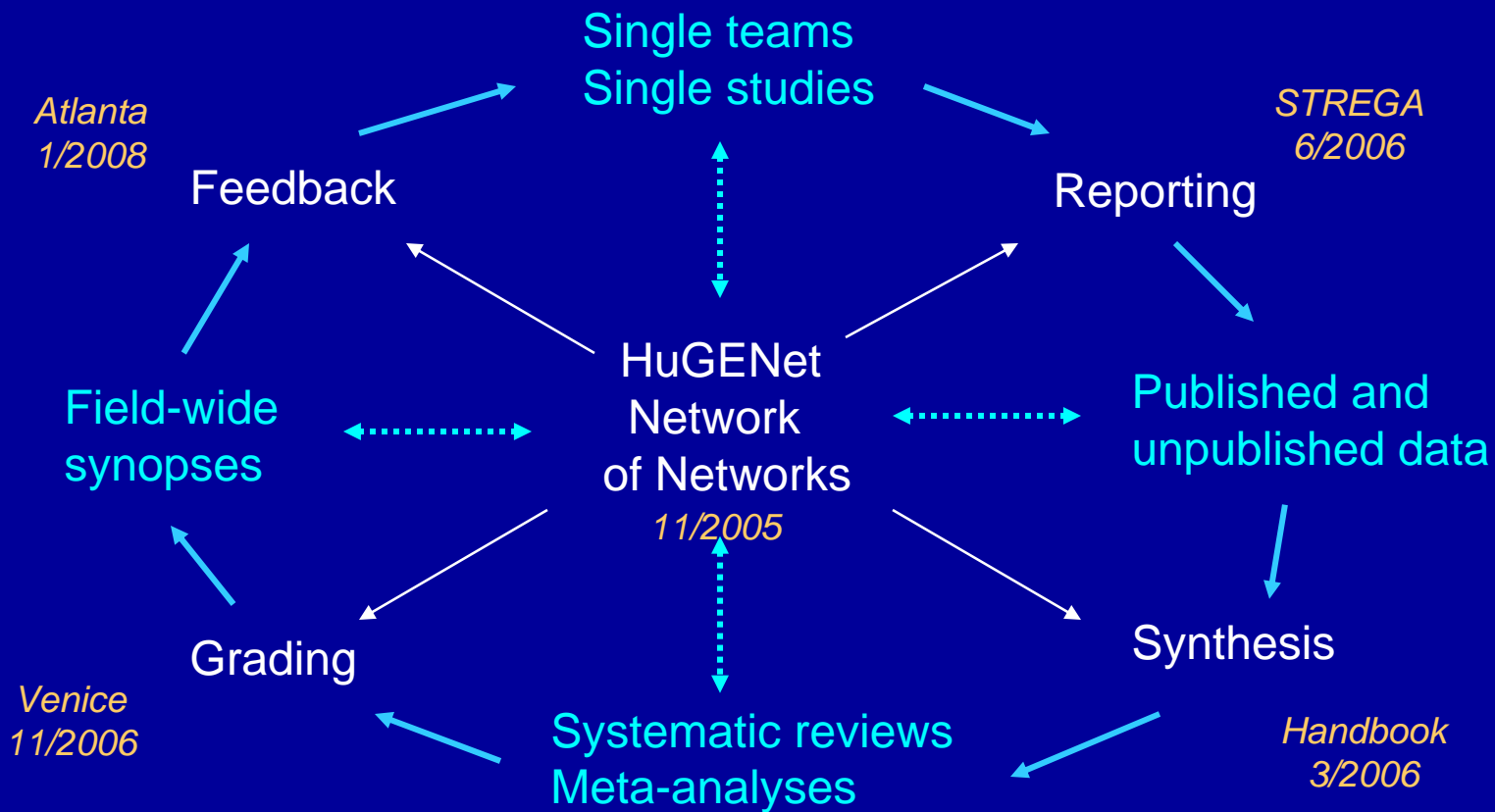
<http://www.dhe.med.uoi.gr/hugenet.htm>

Department of Hygiene and Epidemiology,
University of Ioannina School of Medicine

- Published literature scan
- Systematic reviews
- Strengthened reporting
- Network collaboration







Commentary, *Nature Genetics* 38, 3 - 5 (2006)
 A road map for efficient and reliable human genome epidemiology

Examples of Network HuGE Study Platforms

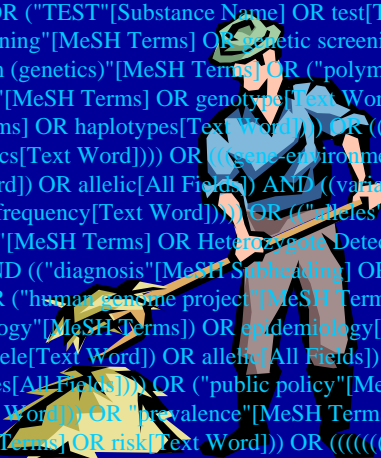
• Disease	Consortium	Teams	Subjects
• Parkinson	GEO-PD	18	10,000
• Osteoporosis	GENOMOS	10	30,000
• Preterm birth	PREGENIA	10	20,000
• Lymphoma	INTERLYMPH	15	20,000
• Lung cancer	ILLCO	30	51,000
• Head & Neck	INHANCE	13	28,000
• Melanoma	GENOMEL	12	3,000
• Pancreatic Ca	PACGENE	10	5,000

From Ioannidis J et al. *AJE* 2005;162:304

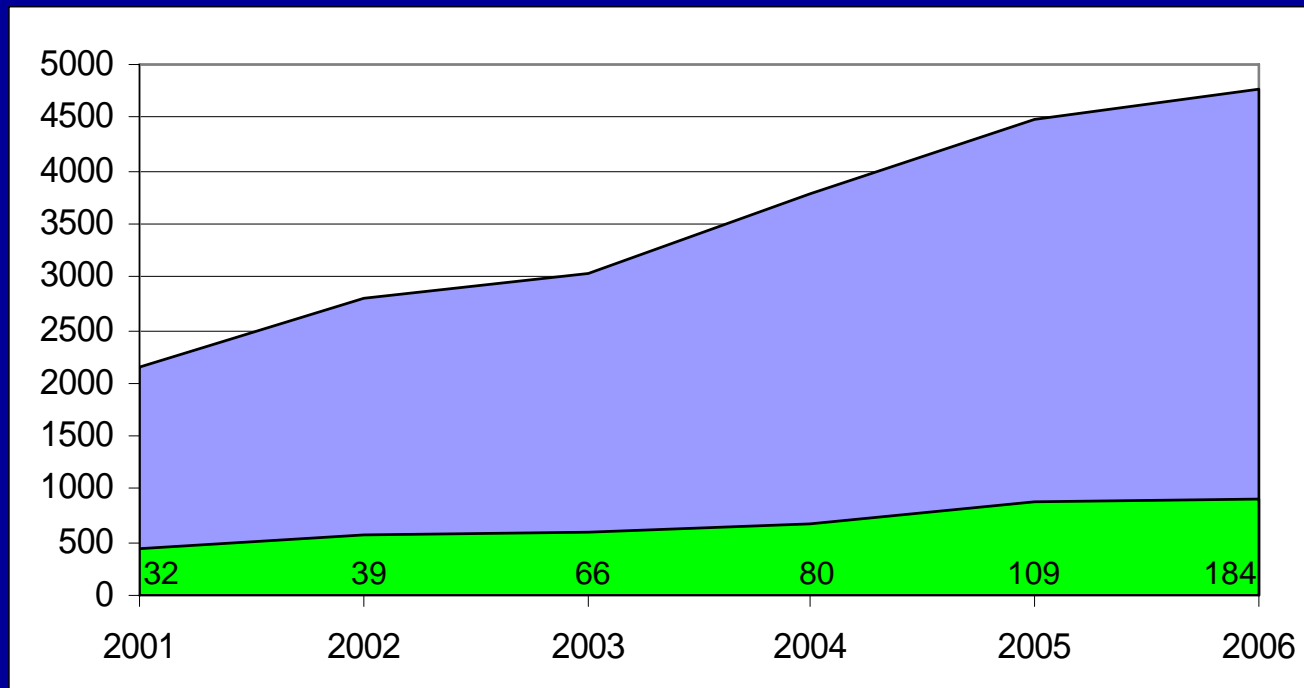
Outline

- Introduction-Why Integrate?
- HuGENet Road Map
- Literature Scanning, Reporting, Synthesis and Network Collaboration
- Developing the Knowledge Base and Causal Inference

(((((((((((((genetic[All Fields] AND ("disease"[MeSH Terms] OR disease[Text Word]) OR disease susceptibility[MeSH Terms] OR predisposition[Text Word]) OR (hereditary diseases"[MeSH Terms] OR genetic disorder[Text Word]) OR (genetic counseling"[MeSH Terms] OR counseling[Text Word])) OR ("disease Fields] AND ("MeSH [Substance Name] OR test[Text Word])) OR ("genetic[MeSH Terms] OR gene[Text Word] OR "genes"[MeSH Terms] OR genes[Text screening"[MeSH Terms] OR genetic screening[Text Word]) OR (genetic[All Fields] MeSH Terms] OR gene[Text Word] AND ("mutation"[MeSH Terms] OR AND ("risk"[MeSH Terms] OR risk[Text Word])) OR ("polymorphism[Text Word] AND ("genes"[MeSH Terms] OR gene[Text Word])) OR ("mutation"[MeSH (genetics)"[MeSH Terms] OR genetic polymorphism[Text Word]) OR genes"[MeSH Terms] OR gene[Text Word])) OR ("hereditary diseases"[MeSH Terms] OR ("genotype"[MeSH Terms] OR genotype[Text Word]) OR ("genome"[MeSH Terms] OR ("TEST"[Substance Name] OR test[Text Word]) OR ("research OR genome[Text Word]) OR (gene-environment interaction[All Fields] OR genetic screening"[MeSH Terms] OR genetic screening[Text Word]) OR environment[All Fields]) OR (genetic[All Fields] OR ("genes"[MeSH Terms] OR morphism (genetics)"[MeSH Terms] OR ("polymorphism (genetics)"[MeSH gene[Text Word])) AND variant[All Fields]) AND genotype"[MeSH Terms] OR ("genotype"[MeSH Terms] OR genotype[Text Word]) OR genotyping[All Fields]) OR (((("epidemiology"[Subheading] OR epidemiology"[MeSH Terms]) OR ("haplotypes"[MeSH Terms] OR haplotypes[Text Word]) OR ("genome"[MeSH Terms] OR epidemiology[Text Word] OR ("public health"[MeSH Terms] OR public health[Text Word]) OR (gene-environment) OR (gene AND Word)) OR ("alleles"[MeSH Terms] OR allele[Text Word]) AND (variant[All Fields] OR variants[All Fields]) OR ("epidemiology"[Subheading] OR epidemiology"[MeSH Terms]) OR frequency[Text Word]) OR ("alleles"[MeSH Terms] OR allele[Text Word]) OR ("public policy"[MeSH Terms] OR policy[Text Word]) OR zygote detection"[MeSH Terms] OR Heterozygote Detection[Text Word]) OR ("education"[Subheading] OR education"[MeSH Terms]) OR education[Text Word]) AND ("diagnosis"[MeSH Subheading] OR "mass screening"[MeSH OR ("epidemiology"[Subheading] OR prevalence"[MeSH Terms]) OR ic[All Fields] OR ("human genome project"[MeSH Terms] OR human genome epidemiology"[MeSH Terms]) OR ("Prevalence"[MeSH Terms]) OR prevalence[Text Word] OR epidemiology[Text Word] OR ("public Word]) OR ("prevention and control"[Subheading] OR prevention[Text Word]) OR OR allele[Text Word] OR allele[All Fields] AND ("epidemiology"[MeSH ("risk"[MeSH Terms] OR risk[Text Word]) OR ("population"[MeSH Terms] OR frequencies[All Fields]) OR ("public policy"[MeSH Terms] OR policy[Text population[Text Word]) AND study[All Fields])) NOT ("animals"[MeSH Terms] OR [Text Word]) OR prevalence"[MeSH Terms] OR prevalence[Text Word] animal"[MeSH Terms]) OR ("prevention and control"[Subheading] OR prevention[Text Word]) OR ("risk"[MeSH Terms] OR risk[Text Word]) OR (((((population[Text Word] OR (a number of) OR genetic[All Fields] OR comparative[All Fields] OR prospective[All Fields] OR cohort[All Fields] OR cross-section[All Fields] OR cross-sectional[All Fields] OR case-control[All Fields] AND (studies OR study[All Fields])) OR (clinical trial[All Fields] OR randomized controlled trial[All Fields]) OR ("drug interactions"[MeSH Terms] OR interactions[Text Word]) OR ("interpersonal relations"[MeSH Terms] OR "drug interactions"[MeSH Terms] OR interaction[Text Word]) OR ("questionnaires"[MeSH Terms] OR questionnaire[Text Word]) OR ("sensitivity and specificity"[MeSH Terms] OR sensitivity[Text Word] OR ("sensitivity and specificity"[MeSH Terms] OR specificity[Text Word])) OR (((case[All Fields] OR cases[All Fields]) OR ("patients"[MeSH Terms] OR patients[Text Word]) OR (study[All Fields] AND group[All Fields])) OR (((("prevention and control"[MeSH Subheading] OR control[Text Word]) OR controls[All Fields] OR (healthy[All Fields] AND subjects[All Fields])) OR ("child"[MeSH Terms] OR children[Text Word]) OR ("adult"[MeSH Terms] OR adults[Text Word]) OR individuals[All Fields])) OR (((("association"[MeSH Terms] OR association[Text Word]) OR ("association"[MeSH Terms] OR associations[Text Word]) OR ("disease"[MeSH Terms] OR disease[Text Word]) AND ("genes"[MeSH Terms] OR gene[Text Word] OR ("genes"[MeSH Terms] OR genes[Text Word])) OR oversight[All Fields] OR ("genotype"[MeSH Terms] OR genotype[All Fields] OR allele[All Fields] AND distribution[Text Word]) OR (((("genotype"[MeSH Terms] OR genotype[Text Word] AND ("phenotype"[MeSH Terms] OR phenotype[Text Word]) OR genotype-phenotype[All Fields] AND correlation[All Fields]) OR ((positive OR negative) AND predictive value)) OR (odds ratio) OR ("ethics"[MeSH Terms] OR ethics[Text Word]) OR ethical[All Fields])))) AND "2004/7/7 8.00"[MHDA]:"2004/7/14 8.00"[MHDA])



HuGE Published Literature 2001-2006



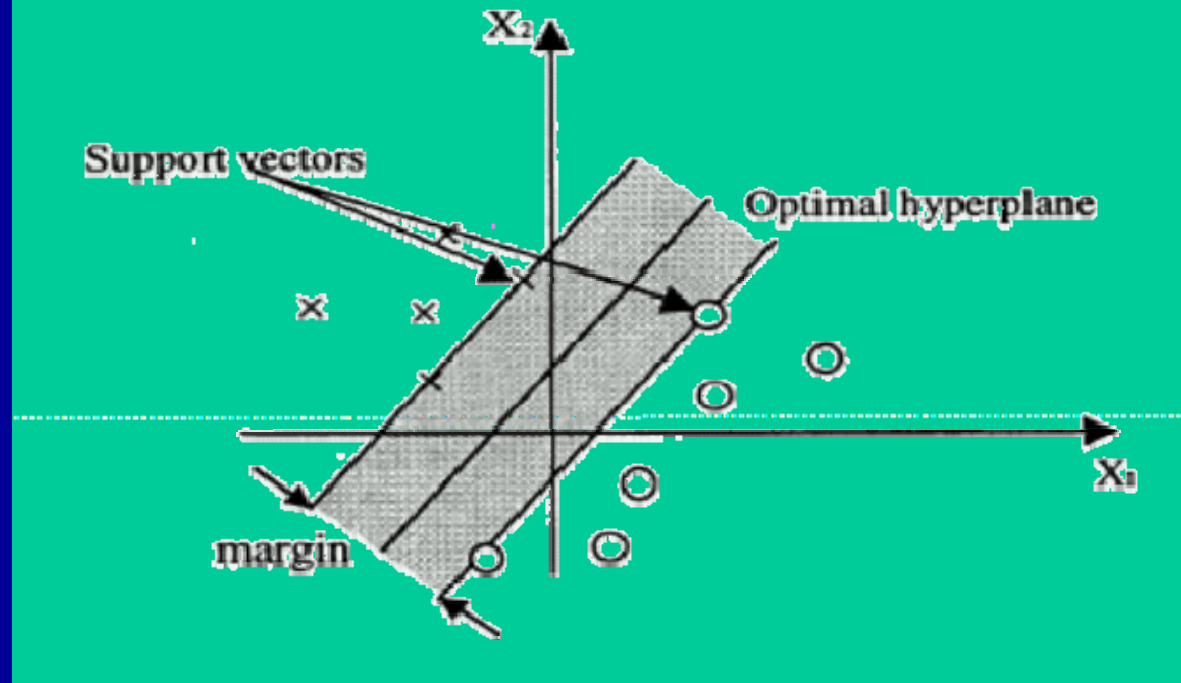
- Gene-disease association
- Gene-environment interaction
- 32 Meta-analysis/HuGE review



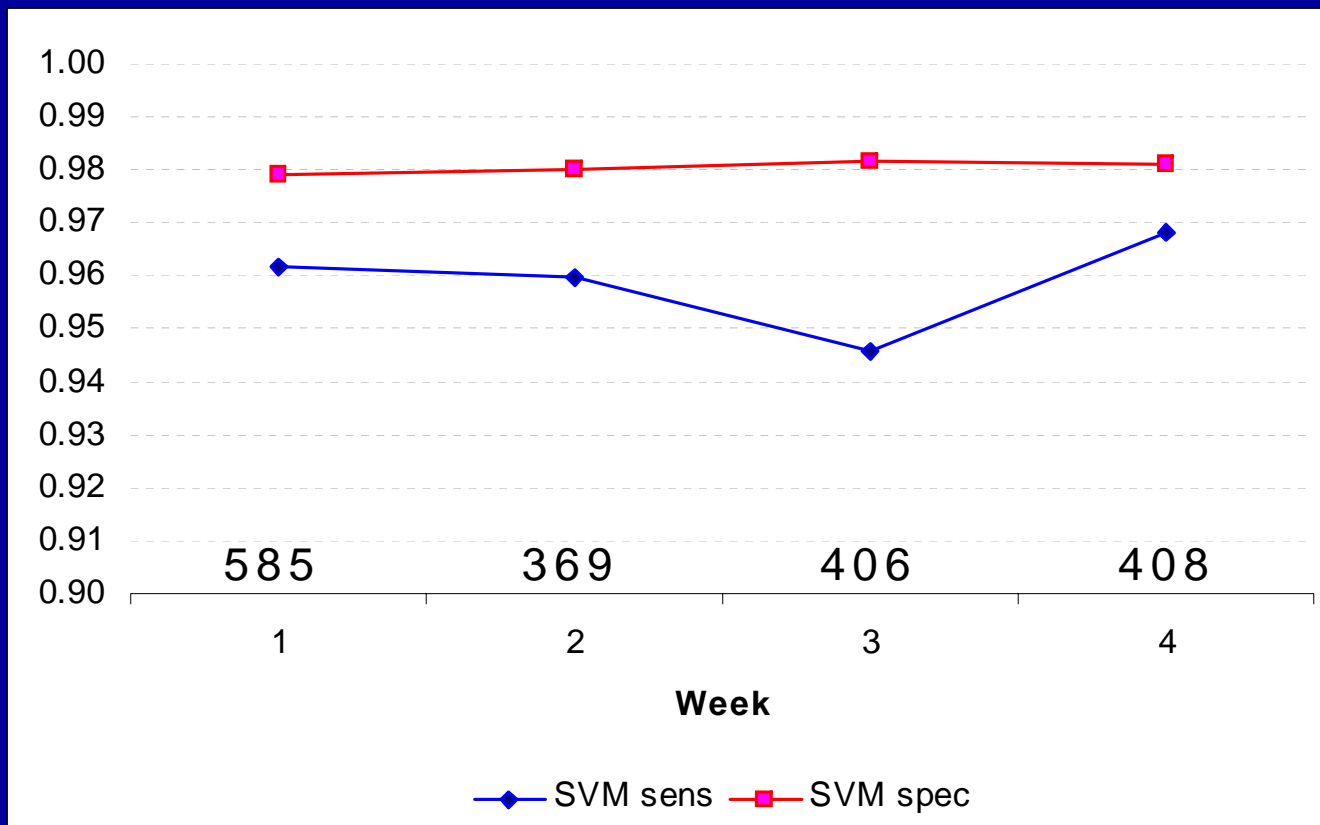
HuGE



Classification Problem: Man vs. Machine



Support Vector Machine = SVM



HuGE published literature: SVM model vs human method
Feb-Mar, 2007

Hand searching was reduced by 90%



SAFER • HEALTHIER • PEOPLE™

CDC Home

Search

Health Topics A-Z

National Office of
Public Health Genomics

S M T W
1 2 3 4
8 9 10 11
15 16 17 18
22 23 24 25
29 30 31 1

EVENTS

TRAINING

FUNDING

LINKS

SEARCH GENOMICS

HuGE Navigator

HuGE Navigator is a knowledge base that provides integrated information and knowledge in human genome epidemiology (HuGE).

Currently, HuGE Navigator consists of the following applications:

- [HuGEPedia](#) : an online Human Genome Epidemiology encyclopedia for human diseases.
- [GeneSelectAssist](#) : a search engine for finding possible candidate genes based on the NCBI Entrez Gene, PubMed and HuGE Pub Lit databases.
- [HuGE Literature Finder](#) : a search engine for finding PubMed articles related to human genome epidemiology.
- [HuGE Investigator Browser](#) : a search engine for finding investigators or potential collaborators in a particular HuGE field.
- [US Genome Variation Database](#) : a collection of genotype prevalence data from CDC NHANES genotyping project.
- [HuGE Reality Checker](#) : a calculator for evaluating the predictive value of genetic markers.



HuGE Investigator Browser

- search HuGE investigators -

[HuGE Navigator](#) | [Home](#) | [About](#) | [Search Instructions](#) | [FAQs](#) | [Contact Us](#)

Search for First & Last Authors All Authors

- Enter search terms into the text box.
- Search terms can include disease, exposure, gene, author, journal, etc.
- Author options determine the authorship status of investigators of interest.
- Simple Boolean operators are allowed (such as AND or OR).
- Use the Search dropdown list to switch to other HuGE Navigator applications.

HuGE Investigator Browser is a search engine for finding investigators or collaborators in human genome epidemiology based on study interests such disease/condition, environmental risk factors, or gene. Investigator profile information is extracted using an accessory utility that automatically parses the affiliation data provided by PubMed. .



HuGE Investigator Browser

- search HuGE investigators -

[HuGE Navigator](#) | [Home](#) | [About](#) | [Search Instructions](#) | [FAQs](#) | [Contact Us](#)

Search for First & Last Authors All Authors

- Enter search terms into the text box.
- Search terms can include disease, exposure, gene, author, journal, etc.
- Author options determine the authorship status of investigators of interest.
- Simple Boolean operators are allowed (such as AND or OR).
- Use the Search dropdown list to switch to other HuGE Navigator applications.

HuGE Investigator Browser is a search engine for finding investigators or collaborators in human genome epidemiology based on study interests such disease/condition, environmental risk factors, or gene. Investigator profile information is extracted using an accessory utility that automatically parses the affiliation data provided by PubMed. .



HuGE Investigator Browser

- search HuGE investigators -

[HuGE Navigator](#) | [Home](#) | [About](#) | [Search Instructions](#) | [FAQs](#) | [Contact Us](#)

Search for First & Last Authors All Authors

Search Criteria: osteoporosis[Query]

Filtered By

[Click to re-sort the table based on]

Investigator Name
Deng HW
Ralston SH
Uitterlinden AG
Emi M
Yamada Y
Shimokata H
Chen HY
Langdahl BL
T...

Investigator Information - Microsoft Internet Explorer

Investigator Information

Investigator Name: Deng HW

Query/Domain: osteoporosis

Publications:

All HuGE	32
Query Specific HuGE (all author)	16
Query Specific HuGE (first/last author)	16
All PubMed (query specific)	Link

Possible Affiliation Information:

Institute	Creighton University
Country	United States
Email	
Address	Osteoporosis Research Center, Creighton University, Omaha, Nebraska 68131, USA



HuGE Investigator Browser

- search HuGE investigators -

[HuGE Navigator](#) | [Home](#) | [About](#) | [Search Instructions](#) | [FAQs](#) | [Contact Us](#)

Search for First & Last Authors All Authors

Search Criteria: osteoporosis[Query]

Filtered By

464 investigators were found.

[Click to re-sort the table based on either investigator names (left icon) or the publication count (right icon)]

	Investigator Name	Number of Publications (F/L)	
	Deng HW	16	
	Ralston SH	12	
	Uitterlinden AG	11	
	Emi M	10	
	Yamada Y	9	
	Shimokata H	8	
	Chen HY	7	
	Langdahl BL	6	
	T...	...	

Search Criteria: osteoporosis[Query]>>Deng HW [First/Last Author]

[\[Query Detail\]](#)

Filtered By [?](#) [?](#) [?](#) [?](#) [?](#) [?](#) [?](#)

Articles 1 - 16 of 16

[PubMed It](#)

Display on of 1

1. [Tests of linkage and association of PTH/PTHrP receptor type 1 gene with bone mineral density and height in Caucasians.](#) [\[Detail\]](#)
Journal of bone and mineral metabolism. 2006 24 (1): 36-41.
Zhang YY, Liu PY, Lu Y, Xiao P, Liu YJ, Long JR, Shen H, Zhao LJ, Elze L, Recker RR, Deng HW
2. [Is a gene important for bone resorption a candidate for obesity? An association and linkage study on the RANK \(receptor activator of nuclear factor-kappaB\) gene in a large Caucasian sample.](#) [\[Detail\]](#)
Human genetics 2006 Nov 120 (4): 561-70.
Zhao LJ, Guo YF, Xiong DH, Xiao P, Recker RR, Deng HW
3. [Robust and comprehensive analysis of 20 osteoporosis candidate genes by very high-density single-nucleotide polymorphism screen among 405 white nuclear families identified significant association and gene-gene interaction.](#) [\[Detail\]](#)
Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research 2006 Nov 21 (11): 1678-95.
Xiong DH, Shen H, Zhao LJ, Xiao P, Yang TL, Guo Y, Wang W, Guo YF, Liu YJ, Recker RR, Deng HW
4. [The human calcium-sensing receptor and interleukin-6 genes are associated with bone mineral density in Chinese.](#) [\[Detail\]](#)
Yi chuan xue bao = Acta genetica Sinica. 2006 Oct 33 (10): 870-80.
Wang YB, Guo JJ, Liu YJ, Deng FY, Jiang DK, Deng HW

STrengthening REporting of Genetic Associations (STREGA)

- STROBE:
International collaborative initiative for STrengthening the Reporting of OBservational studies in Epidemiology



American Journal of Epidemiology
Copyright © 2002 by the Johns Hopkins Bloomberg School of Public Health
All rights reserved

Vol. 156, No. 4
Printed in U.S.A.
DOI: 10.1093/aje/kwf054

Reporting, Appraising, and Integrating Data on Genotype Prevalence and Gene-Disease Associations

Julian Little¹, Linda Bradley², Molly S. Bray³, Mindy Clyne⁴, Janice Dorman⁵, Darrell L. Ellsworth⁶, James Hanson⁷, Muin Khoury⁴, Joseph Lau⁸, Thomas R. O'Brien⁷, Nat Rothman⁷, Donna Stroup⁹, Emanuela Taioli¹⁰, Duncan Thomas¹¹, Harri Vainio¹², Sholom Wacholder⁷, and Clarice Weinberg¹³

¹ Epidemiology Group, Department of Medicine and Therapeutics, University of Aberdeen, Aberdeen, Scotland.

² Sunrise Medical Laboratory, Hauppauge, NY.

³ University of Texas Health Science Center at Houston, Houston, TX.

⁴ Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, GA.

⁵ Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.

STROBE + HuGE = STREGA

STrengthening REporting of Genetic Associations (STREGA)

- STROBE:
International collaborative initiative for STrengthening the REporting of OBservational studies in Epidemiology

STROBE statement: Checklist of essential items Version 2 (April 2005)

	<i>Item #</i>	<i>Cohort</i>	<i>Case-control</i>	<i>Cross-sectional</i>
<i>TITLE & ABSTRACT</i>	1	(a) Identify the article as a cohort study in the title or the abstract.	(a) Identify the article as a case-control study in the title or the abstract.	(a) Identify the article as a cross-sectional study in the title or the abstract.
		(b) The abstract should be a highly informative structured summary of the article, taking account of all issues in the checklist below.		
<i>INTRODUCTION</i>				
<i>Background / Rationale</i>	2	Explain scientific background and rationale for the study.		
<i>Objectives</i>	3	State specific objectives and hypotheses.		
<i>METHODS</i>				
<i>Study design</i>	4	Present key elements of study design. State purpose of original study, if article is one of several from an ongoing study.		
<i>Setting</i>	5	Describe setting, locations and dates defining periods of data collection.		

STROBE + HuGE = STREGA

Reporting Characteristics of 315 HuGE Articles 2001-2003: General

Reporting characteristic	Count	%
Number of study participants		
<100	49	15.6
100-499	190	60.3
500-999	47	14.9
>= 1000	29	9.2
Reported the available power of the study		
No	275	87.3
Yes	40	12.7
Reported that multiple study participant or case or control groups were used		
No	235	74.6
Yes	80	25.4
Provided any information on the origin of the study participants		
No	38	12.1
Yes	277	87.9
Provided any information on the enrollment criteria for the study participants		
No	8	2.5
Yes	307	97.5

Yesupriya et al.

Reporting Characteristics of 315 HuGE Articles 2001-2003: Genotyping

Reporting characteristic	Count	%
Length in pages dedicated to describing genetic testing method		
0	4	1.3
0.01-0.24	184	58.4
0.25-0.49	88	27.9
>0.5	39	12.4
Reference to the genotyping method of another study		
No	62	19.7
Yes	253	80.3
Reported that the genotyping results were validated by using duplicate samples		
No	293	93.0
Yes	22	7.0
Reported that the genotyping results were validated by using a different method		
No	284	90.2
Yes	31	9.8
Reported that the evaluation of the genetic test was blind to the outcomes or phenotypes		
Blind	35	11.1
Unclear	280	88.9
Reported that the evaluation of the outcomes or phenotypes was blind to the genetic test		
Blind	12	3.8
Unclear	303	96.2

Yesupriya et al.

Reporting Characteristics of 315 HuGE Articles 2001-2003: Subject Selection

Reporting characteristic	Count	%
Explicitly stated that all study participants were drawn from the same ethnic population		
Unclear	130	41.3
Stated	185	58.7
Analysis conducted by using different ethnic groups		
No	285	90.5
Yes	30	9.5
If different ethnic groups were included, how was ethnicity treated in the analysis (n=30)		
Stratified by or adjusted for ethnic groups	23	76.7
Pooled ethnic groups together	2	6.7
Unclear	5	16.6
Reported that unlinked genetic markers were used to identify population stratification		
No	313	99.4
Yes	2	0.6
Reported that cases and controls were drawn from the same population in regards to geography (n=227)		
No	79	34.8
Yes	148	65.2
Reported that cases and controls were drawn from the same population in regards to the clinical population (n=227)		
No	180	79.3
Yes	47	20.7

Yesupriya et al.

Reporting Characteristics of 315 HuGE Articles 2001-2003: Analysis

Reported that all genetic variants were examined for Hardy-Weinberg equilibrium		
No	164	52.1
Yes	151	47.9
If Hardy-Weinberg equilibrium was reported, did any polymorphism reportedly fail Hardy-Weinberg equilibrium (n=151)		
No	141	93.4
Yes	10	6.6
Sufficient data reported on all genetic variants of interest for all outcomes		
No	41	13.0
Yes	274	87.0
Reported that analyses were conducted by using allele-based genetic contrasts		
No	143	45.4
Yes	172	54.6
Reported that analyses were conducted by using genotype-based genetic contrasts		
No	45	14.3
Yes	270	85.7
If the analyses were conducted by using genotypes, were selected contrasts or all possible contrasts assessed (n=270)		
All possible	214	79.3
Selected	56	20.7
Justifications given for the selection of specific genetic contrasts (n=56)		
No	33	58.9
Yes	23	41.1

Yesupriya et al.

home > HuGENet™ > reviews

HuGENet™

HuGE Reviews

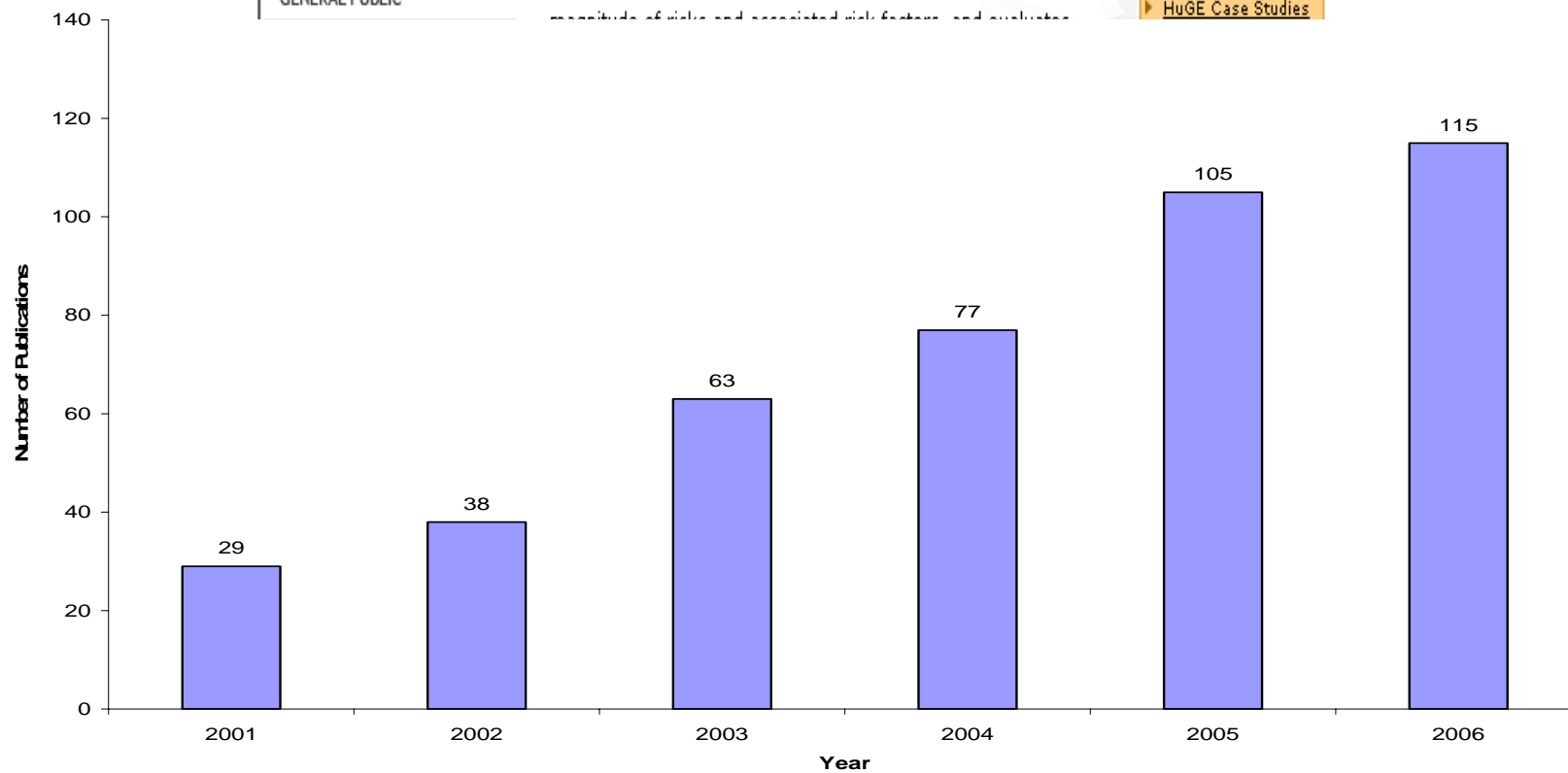
A HuGE Review identifies human genetic variations at one or more loci, and describes what is known about the frequency of these variants in different populations, identifies diseases that these variants are associated with and summarizes the magnitude of risks and associated risk factors, and evaluates

HuGENet™ Menu

- HuGENet™ Home
- What's New
- HuGE Funding
- GDP Info
- Genotype Prev. Db
- HuGE Book
- HuGE Case Studies

MAIN MENU

- OGDP HOME
- WEEKLY UPDATE
- FREQUENTLY ASKED QUESTIONS
- CDC ACTIVITIES
- FAMILY HISTORY
- GENOMICS IN PRACTICE
- GENETIC TESTING
- POPULATION RESEARCH
- HuGENet™
- GENERAL PUBLIC



The HuGENet™ HuGE Review Handbook, version 1.0

Editors:

Julian Little⁶
Julian Higgins^{2,3}

Writing group:

Molly Bray¹
Julian Higgins^{2,3}
John Ioannidis⁴
Muin Khoury⁵
Julian Little⁶
Teri Manolio⁷
Liam Smeeth⁸
Jonathan Sterne⁹

on behalf of the HuGE Net Systematic Review and Meta-Analysis Working Group

American Journal of Epidemiology
Genetics in Medicine
Epidemiologic Reviews
Pediatric Perinatal Epidemiology
International Journal Epidemiology

American J Obs Gyn
Gastroenterology
CEBP
Teratology/Birth Def
Epidemiology

Advantages of MIPD

Ioannidis et al, Am J Epidemiol 2002

MIPD* versus MPL*

Advantages

Data

- More information
- Inclusion of extended databases from published studies
- Inclusion of data from unpublished studies

Better standardization of information

- Categorization of eligible participants

Outcomes

- Definition of follow-up period and censoring criteria

Analysis

Better time-to-event analyses

- Standardized statistical models
- Evaluation of time-dependency

Better adjusted/multivariate analyses

Consistent treatment of loci in linkage disequilibrium

Evaluation of dose-response effects for multiple genes or double doses of a single gene

Evaluation of subgroup effects, including racial heterogeneity

Interpretation

- Assessment of heterogeneity
- Assessment of sampling bias in specific studies

Other

- Establishment of international networks of collaborating investigators

Disadvantages of MIPD

Data

Data may not be made available from all published studies

Interpretation

Potential post hoc conflicts with collaborators regarding findings

Resources

Substantial effort and infrastructure required to:

Develop and administer a standardized protocol

Collect, manage, and analyze data

Communicate with collaborators

Stages in Integrating Evidence

- Formulating the problem
- Identification of studies and publication bias
- Critical appraisal of studies
- Abstraction of data
- Synthesis

Critical Appraisal

- Independent reviewers
- Inclusion/exclusion criteria
- Sequential or multiple publications of analyses of same or overlapping data sets
- Assessment of study quality

Synthesis of the Evidence

- Volume of evidence
- Evidence tables
 - Publication details
 - Study type
 - Factors relating to study quality
 - Measure of association, with indication of its precision

Summary of Studies of Colorectal Cancer & *GSTT1*

Area of study; Recruitment period	Cases Type	N	Controls Type	N	% <i>GSTT1</i> null	RR (95% CI) for null vs other genotypes	Adjustment	Subgroup analysis reported	Exposure assessment	Reference
Australia, Queensland; period not stated	Patients with <i>colorectal adenocarcinoma</i>	125	Unselected subjects (n=94; source not stated) and geriatric (n=54) patients without cancer or a family history of cancer	94 54	19% 9%	0.7 (0.3-1.4) 1.5 (0.6-4.3)	None None	Position of tumour; age	None	Chenevix-Trench et al., 1995 (78)
UK, North Staffordshire Cases & controls 1990-94	Unrelated English "Caucasian" patients with <i>colorectal cancer</i> recruited from 1 hospital	211	Hospitalised English "Caucasian" subjects without malignancy or inflammatory pathologies; recruited in same hospital as cases	509	18%	1.9 (1.3-2.7)	None	Position of tumour	None	Deakin et al., 1996 (49)
Japan, Kitakyusko City Cases 1991-95, controls 1993-95	Consecutive patients with <i>colorectal adenocarcinoma</i> diagnosed in 2 hospitals and 1 medical centre; 65% male; mean age 64.4 years	103	Subjects who had visited local medical centres for regular health check-ups; no gastrointestinal symptoms and no current or previous diagnosis of cancer; 57% male; mean age 61.9 years	126	44%	1.2 (0.7-2.0)	None	Position of tumour	Medical, residential, occupational and smoking history assessed by interview	Katoh et al., 1996 (14)
Australia, Adelaide; period not stated	White adults with sporadic <i>colorectal cancer</i> ; source not stated	219	White blood donors	200	19%	3.4 (2.1-5.4)	None	None	None	Butler et al., 1997 (77) (reported in abstract only)
USA (nested case- control study in Physicians Health Study (PHS)); cases 1982-96	Cases with <i>colorectal cancer</i> from those randomised in PHS; physicians excluded from randomisation if they had history of myocardial infarction, stroke, transient ischemic attack, cancer, renal or liver disease, peptic ulcer, gout	212	Sample of subjects not diagnosed with colorectal cancer in PHS (same exclusion criteria as listed for cases); matched on year of birth and smoking history	221	23%	0.8 (0.5-1.2)	BMI, physical activity, alcohol use	Position of tumour, age smoking	Smoking history, alcohol intake, diet, frequency of meat intake, physical activity, disease diagnoses	Certig et al., 1998 (58)
Singapore, period not stated	Chinese <i>colorectal carcinoma</i> patients recruited from a surgical department	300	Chinese patients obtained from clinical chemistry department with no history of neoplasms	183	Not stated ¹	-	-	Position of tumour, tumour histology	None	Lee et al., 1998 (23)

This paper was published with modifications in Am J Epidemiol 2000 May 1;151(9):862-877

PMID: 10791559; UI: 20250198

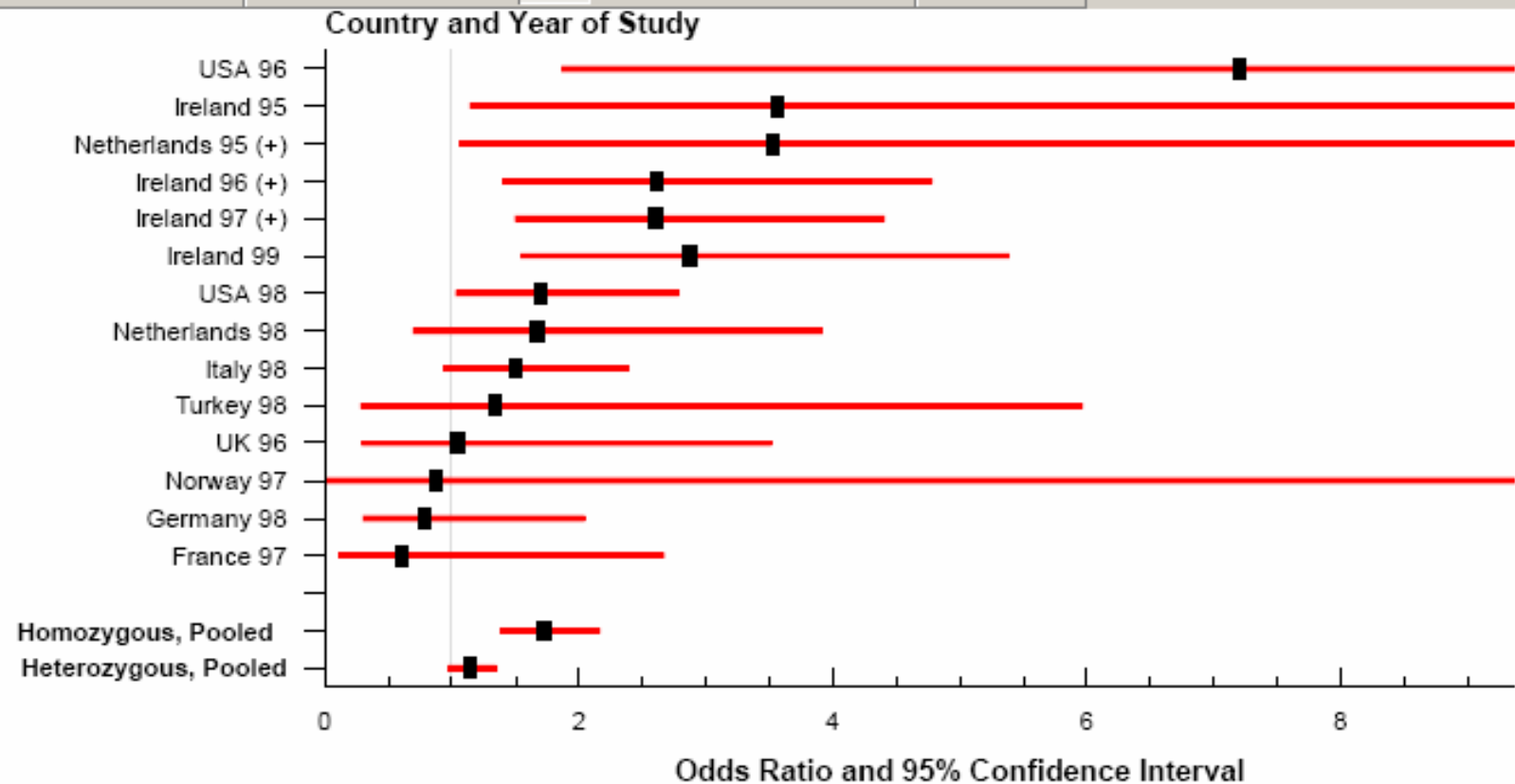
5, 10-Methylenetetrahydrofolate reductase (*MTHFR*) Gene Variants and Congenital Anomalies

Lorenzo D. Botto and Quanhe Yang

[Download print version](#)

Authors' affiliations:

Centers for Disease Control and Prevention, National Center for Environmental Health, Division of Birth





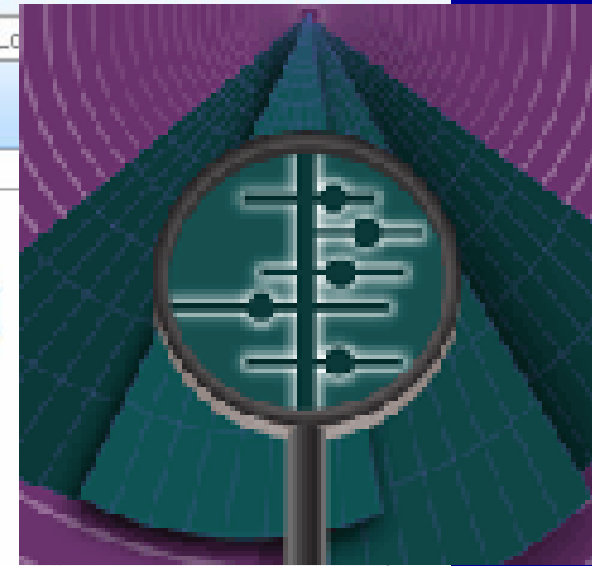
RESEARCH ARTICLE



Epidemiology and Reporting Characteristics of Systematic Reviews

David Moher^{1,2,3*}, Jennifer Tetzlaff¹, Andrea C. Tricco^{1,4}, Margaret Sampson¹, Douglas G. Altman⁵

¹ Chalmers Research Centre, Department of Paediatric Epidemiology and Community Institute of Population Science, Oxford, United Kingdom



EDITORIAL



Many Reviews Are Systematic but Some Are More Transparent and Completely Reported than Others

The *PLoS Medicine* Editors

Citation: The *PLoS Medicine* Editors (2007) Many Reviews Are Systematic but Some Are More Transparent

DOWNLOAD

- XML (16 KB)
- PDF - Large (82 KB)
- PDF - Small (54 KB)
- Citation

CONTRIBUTE

- Write a Response
- Feedback

Methodologic Issues in Meta-Analysis of Gene-Disease Associations

- 37 Meta analyses
- 22% (8) described search terms
- 51% (19) had no inclusion/exclusion criteria
- 76% (28) assessed heterogeneity
- 19% (7) checked for publication bias
- 24% (9) assessed Hardy-Weinberg equilibrium
- 22% (8) had biologic rationale for genetic model

Attia et al, 2003

What about Genome-Wide Association Studies?

National Cancer Institute
U.S. National Institutes of Health | www.cancer.gov

CGEMS
Cancer Genetic Markers of Susceptibility

CORE GENOTYPING FACILITY

Welcome | Browse Data | Bulk Data Download | Support | Feedback | Login/Registration

Address: <http://www.wtccc.org.uk/>

FOUNDATION FOR THE NATIONAL INSTITUTES OF HEALTH

HOME | ABOUT US | PROGRAMS | PARTNERS & DONORS | NEWS & EVENTS | MAKE A CONTRIBUTION | CONTACT US

GAIN Program

GAIN HOME PAGE
▶ Partnerships
▶ Overview

GENETIC ASSOCIATION INFORMATION NETWORK (GAIN)

Genetic Association Information Network (GAIN) is a public-private partnership of the National Institutes of Health, Inc. (NIH), which will include corporations, advocacy groups, concerned individuals, and the National Institutes of Health. This initiative will take the next step in the search to understand the genetic risk for complex diseases. Through a series of whole genome scans, using samples from existing case-control studies of common diseases, the goal is to identify genetic pathways that make us more susceptible to disease and thereby facilitate discovery of new molecular targets for prevention, diagnosis, and treatment.

WTCCC
Home

Links
Overview
Participants
Press Release [28 09 2005]
Data release and access policy
Guidelines for data use
Analysis data format
Simple data format
Data access

The Wellcome Trust Case Control Consortium

The Wellcome Trust Case Control Consortium (WTCCC) is a collaboration of 24 leading human geneticists, who will analyse thousands of DNA samples from patients suffering with different diseases to identify common genetic variations for each condition. It is hoped that by identifying these genetic signposts, researchers will be able to understand which people are most at risk, and also produce more effective treatments.

The WTCCC will search for the genetic signposts for tuberculosis, coronary heart disease, type 1 diabetes, type 2 diabetes, rheumatoid arthritis, Crohn's disease and ulcerative colitis, bipolar

Outline

- Introduction-Why integrate?
- The HuGE Net movement
- Collaboration across epidemiologic platforms
- Strengthening the reporting of genetic associations (STREGA)
- Integration across studies-HuGE Reviews
- Developing the knowledge base and causal inference

“Guidelines” for Causal Inference

Consistency

Strength

Dose-response

Biological plausibility

Specificity

Temporality

Experimentation

Coherence

Analogy

(Hill, 1965; US Surgeon General’s Committee, 1964)

The Legend of Biologic Plausibility

- In 2002, studies were published addressing the relationship of the APOE polymorphism with Alzheimer's disease; colorectal cancer; fatty liver; atherosclerosis; hyperlipidemia; acute ischemic stroke; spina bifida; coronary artery disease; normal tension glaucoma; hypertension; Parkinson's disease, diabetic nephropathy; pre-eclampsia; hepatitic C-related liver disease; cerebrovascular disease; coronary artery disease post-renal transplantation; non-specified cognitive impairment; childhood nephrotic syndrome; spontaneous abortion; multiple sclerosis; alcohol withdrawal; cognitive dysfunction after coronary artery surgery; alcoholic chronic pancreatitis; alcoholic cirrhosis; macular toxicity from chloroquine; macular edema; aortic valve stenosis; vascular dementia; type II diabetes mellitus; and migraine.

– Source. J Ioannidis

Assessment of Cumulative Evidence on Genetic Associations: Interim Guidelines

John P. A. Ioannidis¹⁻³, Paolo Boffetta⁴, Julian Little⁵, Thomas R. O'Brien⁶, Andre G. Uitterlinden⁷, Paolo Vineis⁸, David J. Balding⁸, [Anand Chokkalingam](#)⁹, Siobhan Dolan¹⁰, W. Dana Flanders¹¹, Julian P. T. Higgins¹², Mark I. McCarthy^{13,14}, David H. McDermott¹⁵, Grier P. Page¹⁶, Timothy R. Rebbeck¹⁷, Daniela Seminara¹⁸, Muin J. Khoury¹⁹

1 Clinical and Molecular Epidemiology Unit, University of [Ioannina School of Medicine](#), [Ioannina 45110](#), Greece

2 Biomedical Research Institute, Foundation for Research and Technology – Hellas, [Ioannina 45110](#), Greece

3 Department of Medicine, Tufts University School of Medicine, Boston MA 02111, USA

4 [International Agency for Research on Cancer](#), Lyon 69008, France

5 Canada Research Chair in Human Genome Epidemiology, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario K1H 8M5, Canada

6 [Division of Cancer Epidemiology and Genetics](#), National Cancer Institute, NIH, Rockville MD 02982, USA

7 Departments of Internal Medicine and Epidemiology & Biostatistics, Erasmus MC, Rotterdam 3000 CA, [The](#)

Assessing Cumulative Evidence on Genetic Associations (Venice Guidelines)

- Epidemiologic Credibility
- Biology
- Clinical/Public Health Relevance

Amount of evidence	A Large-scale evidence	
	B Moderate amount of evidence	
	C Little evidence	
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	
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 and accumulation of evidence	
	C Clear presence of bias that can affect even the presence or not of the association	

Assessing Epidemiologic Credibility of Cumulative Evidence on Genetic Associations- Venice Guidelines

First letter = amount

Second letter = replication

Third letter = protection from bias

AAA	ABA	ACA
AAB	ABB	ACB
AAC	ABC	ACC

BAA	BBA	BCA
BAB	BBB	BCB
BAC	BBC	BCC

CAA	CBA	CCA
CAB	CBB	CCB
CAC	CBC	CCC



Strong evidence



Moderate evidence



Weak evidence

Table 2: Considerations for assessment of clinical and public health relevance and importance of genetic associations

Magnitude of effect

Effect size

Frequency of genetic variant in population

Clinical and public health importance

Type of phenotype: biological, endophenotype, hard clinical outcome

Disease burden; incidence, severity, and mortality

Interaction with identified modifiable environmental exposures

Potential to prevent disease through intervention (e.g. through Mendelian randomization insights)

An Online Encyclopedia for Genome Variation and Health?

EDITORIAL

Nature Genetics 38, 1 (2006)
doi:10.1038/ng0106-1

Embracing risk

In response to requests from researchers for a way to publish and credit well-executed genetic association studies regardless of the outcome, we offer an experimental solution: the journal will now consider for publication—in Analysis format—annual synopses of all adequate association studies on a particular disease or phenotype. The synopsis may be written by a consortium of the authors of unpublished but publicly deposited studies, and it is hoped that the referees of the synopses will publish their comments as a counterpoint.

Much remains to be done. First, researchers need to decide on minimally acceptable criteria

Models for Online Encyclopedia 1

Address  <http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=219700>

 [Go](#) [Links >>](#)



MIM #219700

- Description
- Clinical Features
- Inheritance
- Cytogenetics
- Mapping
- Molecular Genetics
- Heterogeneity
- Pathogenesis
- Diagnosis
- Clinical Management
- Population Genetics
- Evolution
- Genetic Variability
- Animal Model
- History
- See Also
- References
- Contributors
- Creation Date
- Edit History

XXXXXXXXXX **OMIM** XXXXXXXXXXXX  Johns Hopkins University
Online Mendelian Inheritance in Man

All Databases PubMed Nucleotide Protein Genome Structure PMC Taxonomy **OMIM**

Search for

Display Show Send to

All: 1 

[#219700](#) [GeneTests, Links](#)

CYSTIC FIBROSIS; CF

Alternative titles; symbols

MUCOVISCIDOSIS

Gene map locus [7q31.2](#)

TEXT

Models for Online Encyclopedia 2

Address  <http://www.genetests.org/servlet/access?db=geneclinics&site=gt&id=8888891&key=eeT7kUdx42arK&gry=&fcn=y&fw=cMYX&>  Go  Links

[Home Page](#)

[About GeneTests](#)

 [GeneReviews](#)

[Laboratory Directory](#)

[Clinic Directory](#)

[Educational Materials](#)

[\[Printable Copy\]](#)

CFTR-Related Disorders

[Summary](#)
[Diagnosis](#)
[Clinical Description](#)
 [Prevalence](#)
[Differential Diagnosis](#)
[Management](#)
[Genetic Counseling](#)
[Molecular Genetics](#)
[Resources](#)
[References](#)
[Author Information](#)
[Top of Page](#)

[Disable Glossary](#)
(Returns to top)

CFTR-Related Disorders

[Includes: Cystic Fibrosis (CF, Mucoviscidosis) and Congenital Bilateral Absence of the Vas Deferens (CBAVD)]

Authors: Samuel M Moskowitz, MD
Ronald L Gibson, MD, PhD
Darci L Sternes, MS, CGC
Edith Cheng, MS, MD
Garry R Cutting, MD

[About the Authors / Author History](#)





Initial Posting:
26 March 2001

Last Revision:
24 August 2005

Click on [defined terms](#); definition displays here.

Models for Online Encyclopedia 3

Address  http://en.wikipedia.org/wiki/Main_Page

  Go  Links 

 [Sign in / create account](#)



WIKIPEDIA
The Free Encyclopedia

navigation

- [Main Page](#)
- [Community Portal](#)
- [Featured articles](#)
- [Current events](#)
- [Recent changes](#)
- [Random article](#)
- [Help](#)
- [Contact Wikipedia](#)
- [Donations](#)

search

[main page](#)

[discussion](#)

[view source](#)

[history](#)

Your continued donations keep Wikipedia running!

Welcome to **Wikipedia**,
the free encyclopedia that *anyone* can edit.
1,461,789 articles in English

- [Arts](#)
- [History](#)
- [Society](#)
- [Biography](#)
- [Mathematics](#)
- [Technology](#)
- [Geography](#)
- [Science](#)
- [All portals](#)

[Overview](#) · [Searching](#) · [Editing](#) · [Questions](#) · [Help](#)

[Categories](#) · [Featured content](#) · [A–Z index](#)

Today's featured article



Kochi is a city in the [Indian](#) state of [Kerala](#), and one of the principal [seaports](#) of the country. Kochi is located in the district of [Ernakulam](#), about 220 km north of the state capital

[Thiruvananthapuram](#). The city has an estimated population of 650,000, with an [extended metropolitan population](#) of over 1.6 million, making it the largest [urban agglomeration](#) and the second largest city in

In the news

- The [World Confederation of Labour](#) and the [International Confederation of Free Trade Unions](#) merge to form the **International Trade Union Confederation**, representing 166 million members.
- [Hassan Nasrallah](#), the [Secretary General](#) of [Hezbollah](#), says negotiations



Outline

- Introduction-Why Integrate?
- HuGENet Road Map
- Literature Scanning, Reporting, Synthesis and Network Collaboration
- Developing the Knowledge Base and Causal Inference