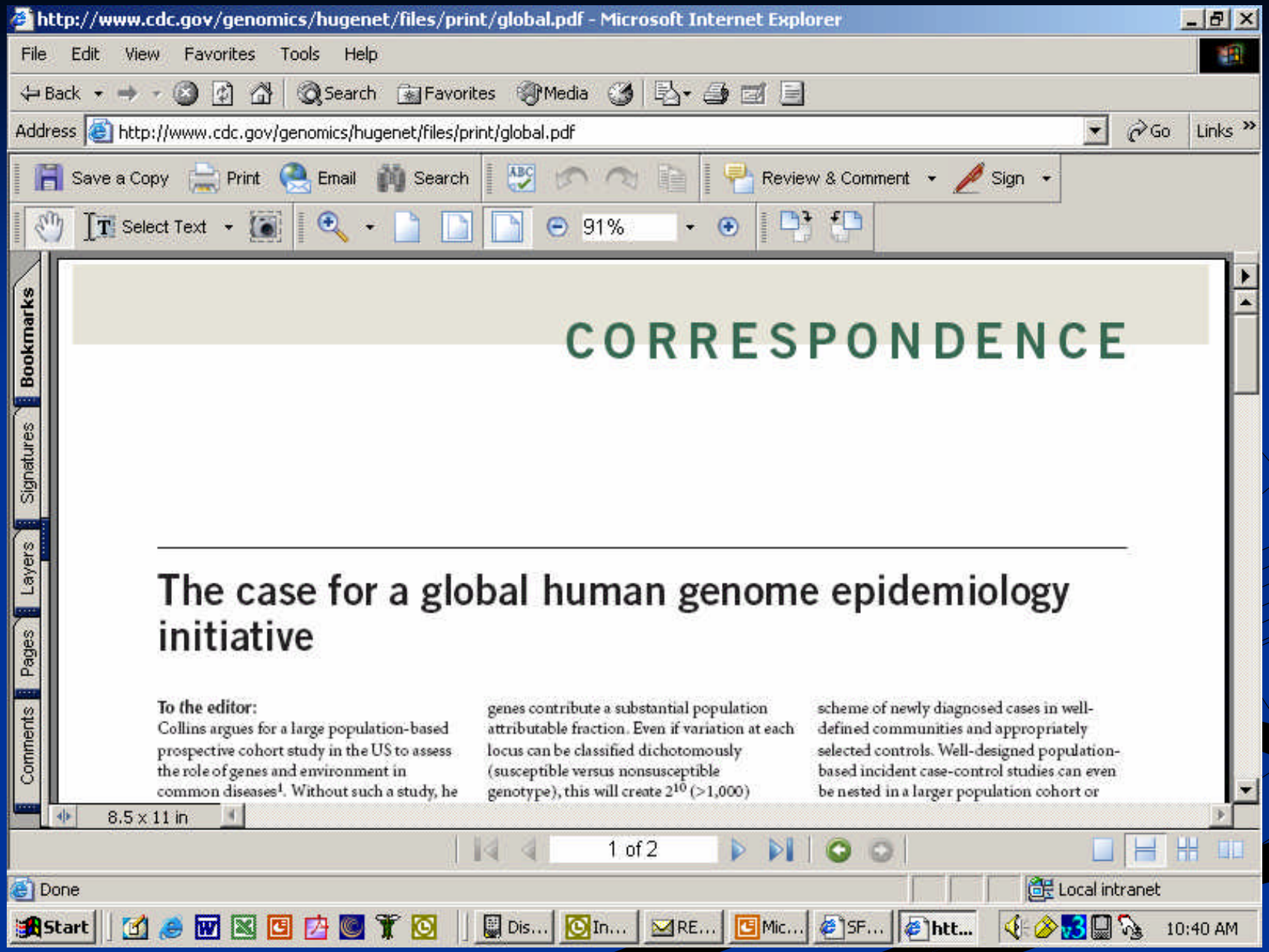
Building the Knowledge Base on Genes and Population Health: Need for Global Collaboration

Muin J. Khoury, MD, Ph.D.

CDC Office of Genomics and Disease Prevention



Khoury MJ, Nat Genet 2004;36:1027-8

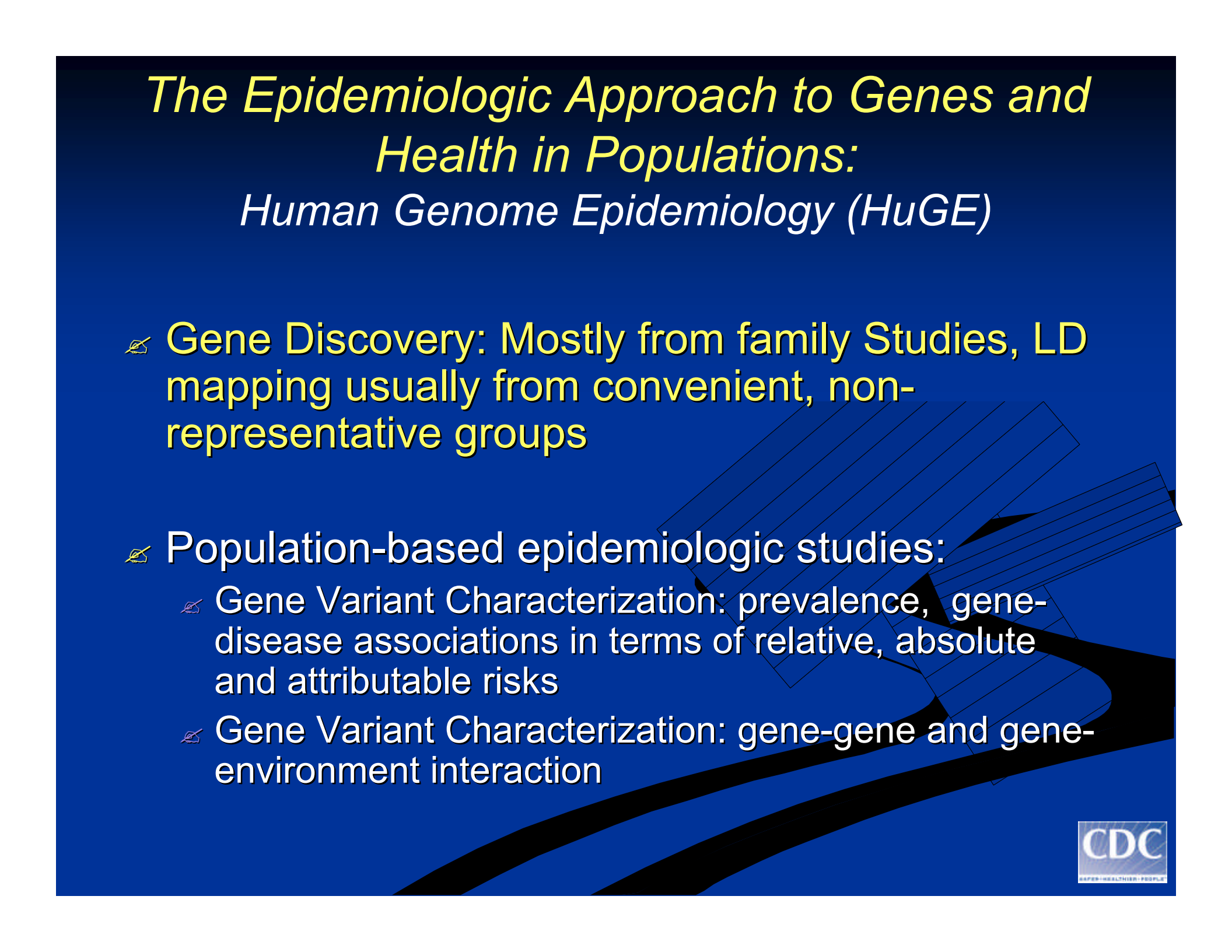
Outline: Three Messages

- Global Collaboration in Biobanks/Population-based Cohort Studies
  - One cohort study in one country is not enough!
- Integration of all Human Genome Epidemiologic Data (cohort, case-control, other)
  - There is more than one way to get there!
- Development of Evidence-based Processes that use Human Genome Epidemiologic Data
  - Direct link from epidemiology to practice!

The Epidemiologic Approach to Genes and Health in Populations:
Human Genome Epidemiology (HuGE)

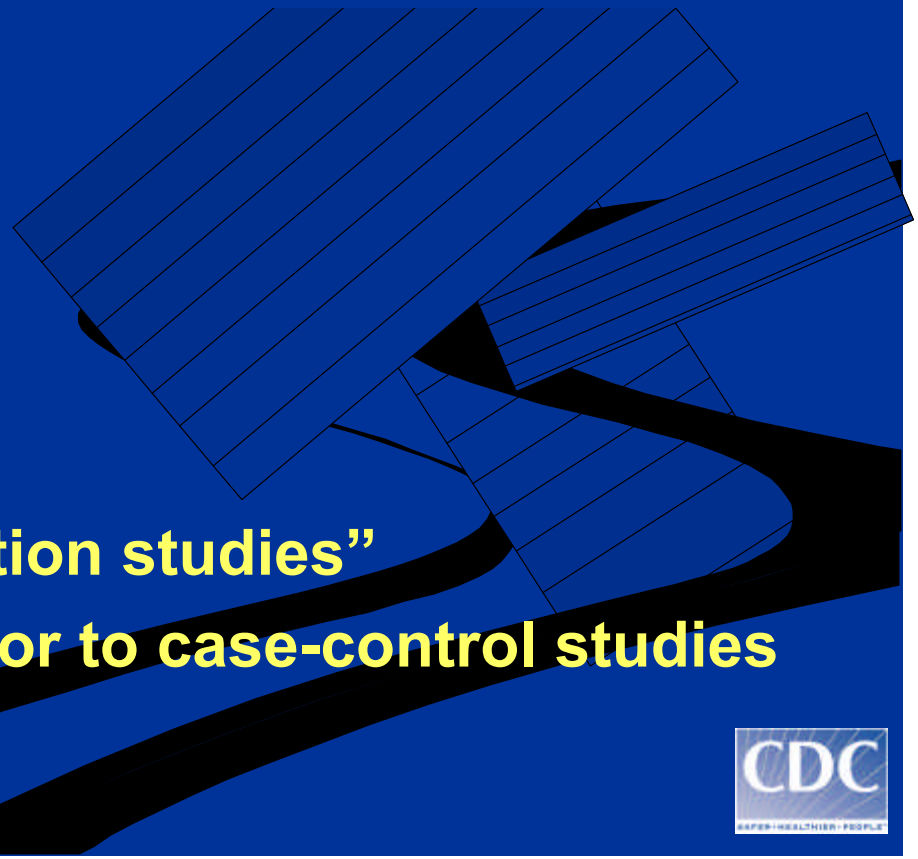
- Gene Discovery: Mostly from family Studies, LD mapping usually from convenient, non-representative groups

- Population-based epidemiologic studies:
  - Gene Variant Characterization: prevalence, gene-disease associations in terms of relative, absolute and attributable risks
  - Gene Variant Characterization: gene-gene and gene-environment interaction



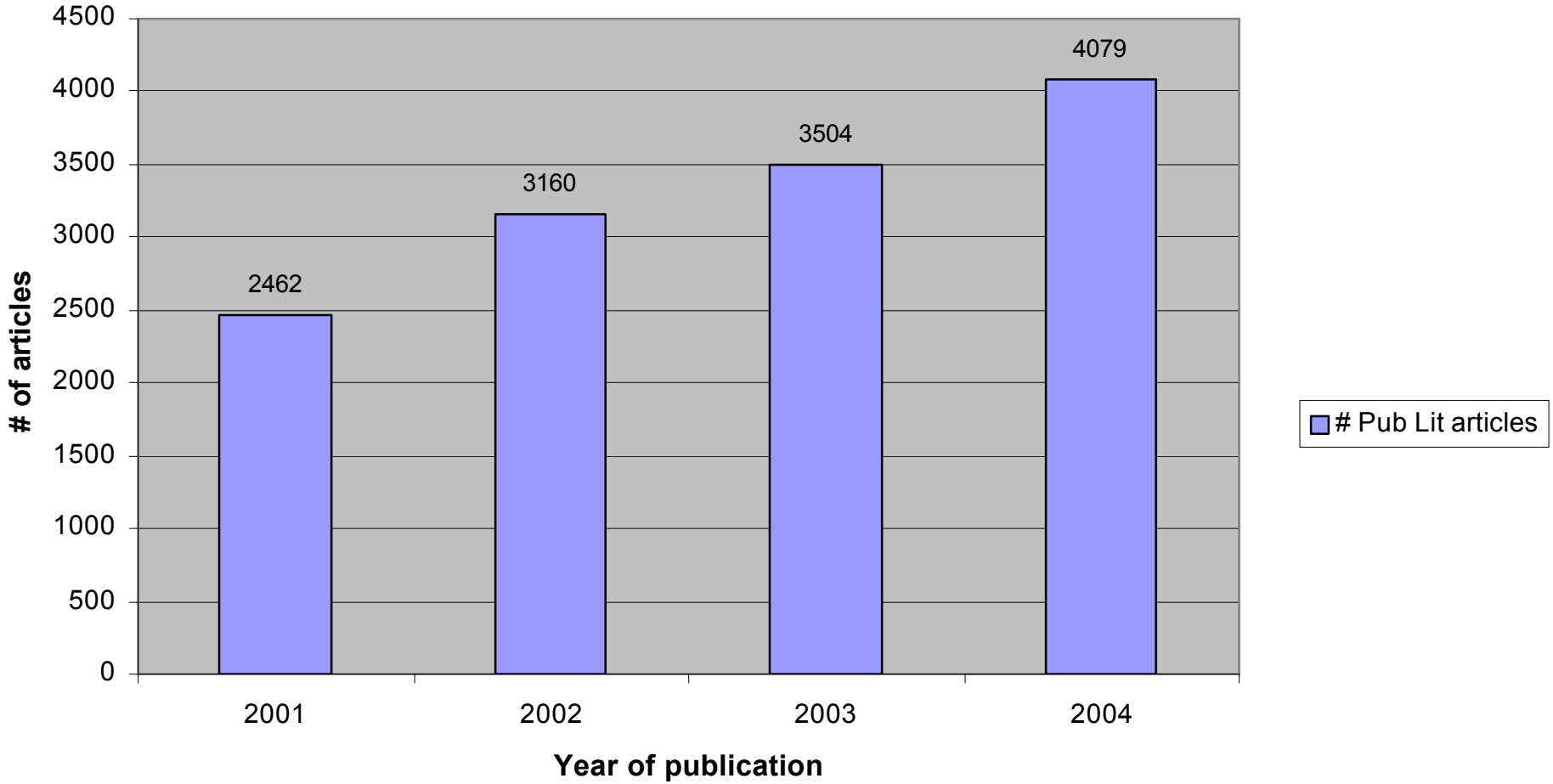
Types of Epidemiologic Studies for Gene Characterization in Populations

- Cross sectional studies
- Case-control studies
- Cohort studies
- Population biobanks



- **Myths:**
  - **Stigma around "association studies"**
  - **Cohort inherently superior to case-control studies**

No. of articles in Huge Published Literature db, 2001-2004*



(*As of Jan 14, 2005. Count excludes review articles, meta & pooled analyses.)

Number of Published HuGE Papers 2001-2004*

Year	Prevalence	Associations	Interactions
2001	308	2141	436
2002	349	2799	569
2003	323	3010	598
2004	368	3486	604

- **MOST ASSOCIATION STUDIES ARE CASE-CONTROL STUDIES**

NHANES III DNA BANK
Prevalence of Genes of Public Health Significance

Background

NHANES III DNA Bank

- National Health and Nutrition Examination Survey (NHANES) is a nationally representative survey
- Detailed interviews, clinical, laboratory and radiologic examinations are conducted
- Phenotypic data, such as serostatus for many infectious exposures, blood count, chemistries, etc. were collected
- During second phase NHANES III (1991-1994), white blood cells were frozen and cell lines were immortalized with EBV
- NHANES III DNA bank is located at NCEH,CDC, with specimens available from over 7000 participants
- In 2002, NCHS announced a call for proposals to use these specimens in the Federal Register

Collaborative CDC-wide Proposal Objective

- Determine the prevalence of genotypes of public health importance

Criteria for Genetic Variants

Public Health Importance

- Known or hypothesized association with diseases of public health importance
- Role in pathways affecting multiple diseases
- Identified functional variants
- Relatively common (i.e., >2.0%)
- Previously described gene-environment or gene-gene interactions
- Relevant phenotypic data available in NHANES dataset
- No current use for clinical risk assessment or intervention

Challenges to Identifying Genes of Public Health Importance

- Gaps in information in the literature
- Methodological issues of many available studies
  - Selection bias, power, interaction
- Non-replication of gene-disease association

Public Health Significance of Proposal

- Prevalence of gene variants
  - Basis for estimating population attributable fraction in combination with measure of gene-disease association
  - Enable assessment of potential for screening population subgroups for susceptibility genes
  - Prevalence of combinations of variants in pathways and at different loci
- Examine gene-disease association, gene-environment and gene-gene interactions

Selected Pathways of Gene Variants
(87 variants of 57 genes)

- Nutrient Metabolism (e.g., folate and homocysteine, lipids, glucose, alcohol, vitamin D)
- Immune and Inflammatory responses (e.g., cytokines, receptors)
- Activation and detoxification pathways (e.g., drugs, carcinogens, environmental contaminants)
- DNA repair pathways (e.g. excision repair, environmental toxins)
- Hemostasis pathway and remodeling/extension (e.g. vasomotor) pathway
- Developmental (e.g., hearing loss)

Laboratory Methods

- Genotyping
  - Assessing Feasibility of External Laboratories to conduct high throughput, accurate, low cost, genotyping for >500,000 SNPs (~7500 specimens X 87 variants*)

*Total of gene variants available upon request

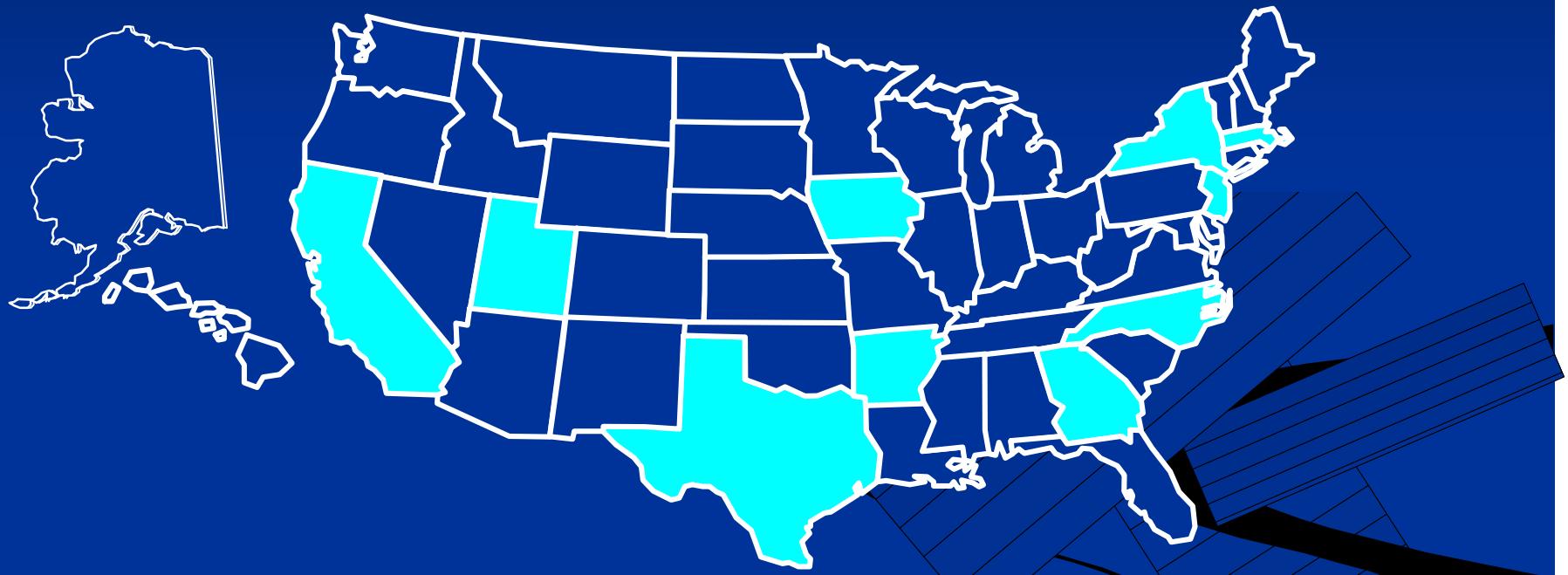
Next Steps

- Pending approval from NCHS
- Laboratory Selected
- Genotype-Phenotype analyses

CDC Working Group

ATSDR Olivia Harris **NCBDDD** Karen Abe, Cynthia Moore, Lorenzo Botto, Quanhe Yang **NCHSTP** Mary Reichler **NCID** Tom Hodge, Craig Hooper, Jai Lingappa, Janet McNicoll, Anne Dilley **NCEH** Amanda Brown, Peg Gallagher, Marta Gwinn, Omar Henderson, Bruce Lin, Mary Lou Lindegren, Julian Little, Karen Steinberg **NCCDPHP** Heidi Blanck, Wayne Giles, Ingrid Hall, Giuseppina Imperatore, Ann Malarcher **NIOSH** MaryAnn Butler, Ainsley Weston **PHPPO** Bin Chen **NIP** Scott Campbell **NCHS** Gerry McQuillan

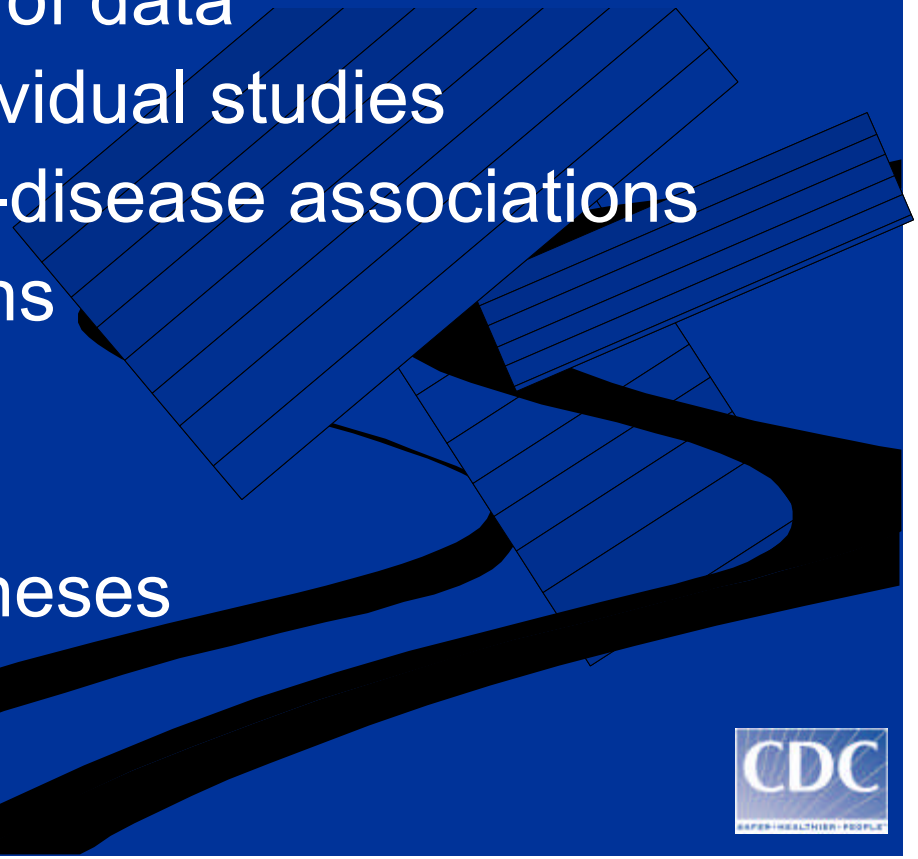
Centers for Birth Defects Research and Prevention



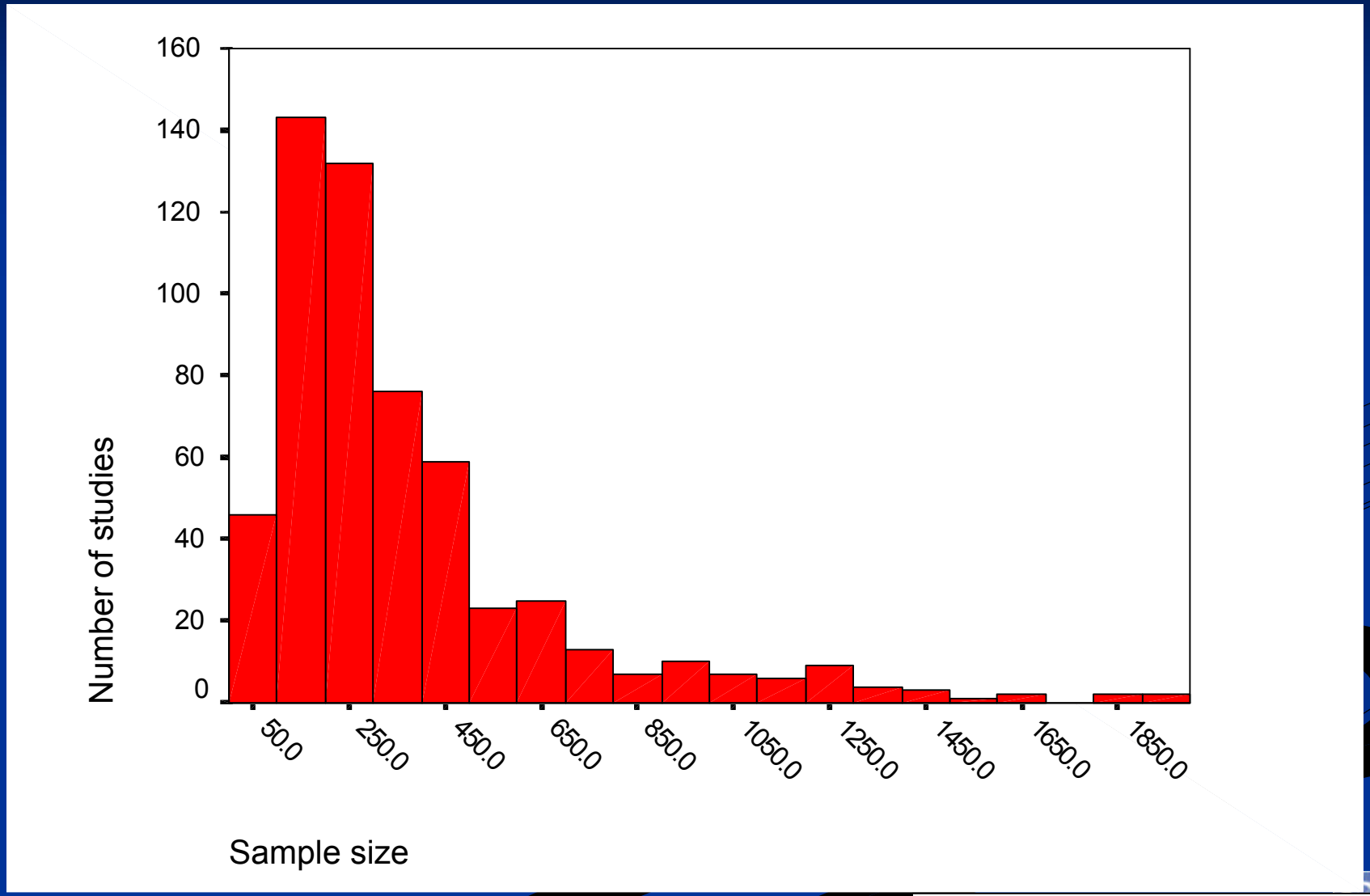
Multi-state Population-Based Case-Control Study Based on State-Based Birth Defects Registries

The Need for Integrating Epidemiologic Evidence: on Genes and Population Health

- Unmanageable amounts of data
- Small sample size of individual studies
- Small effect size of gene-disease associations
- Replication of associations
- Publication bias
- Heterogeneity
- Generate and test hypotheses

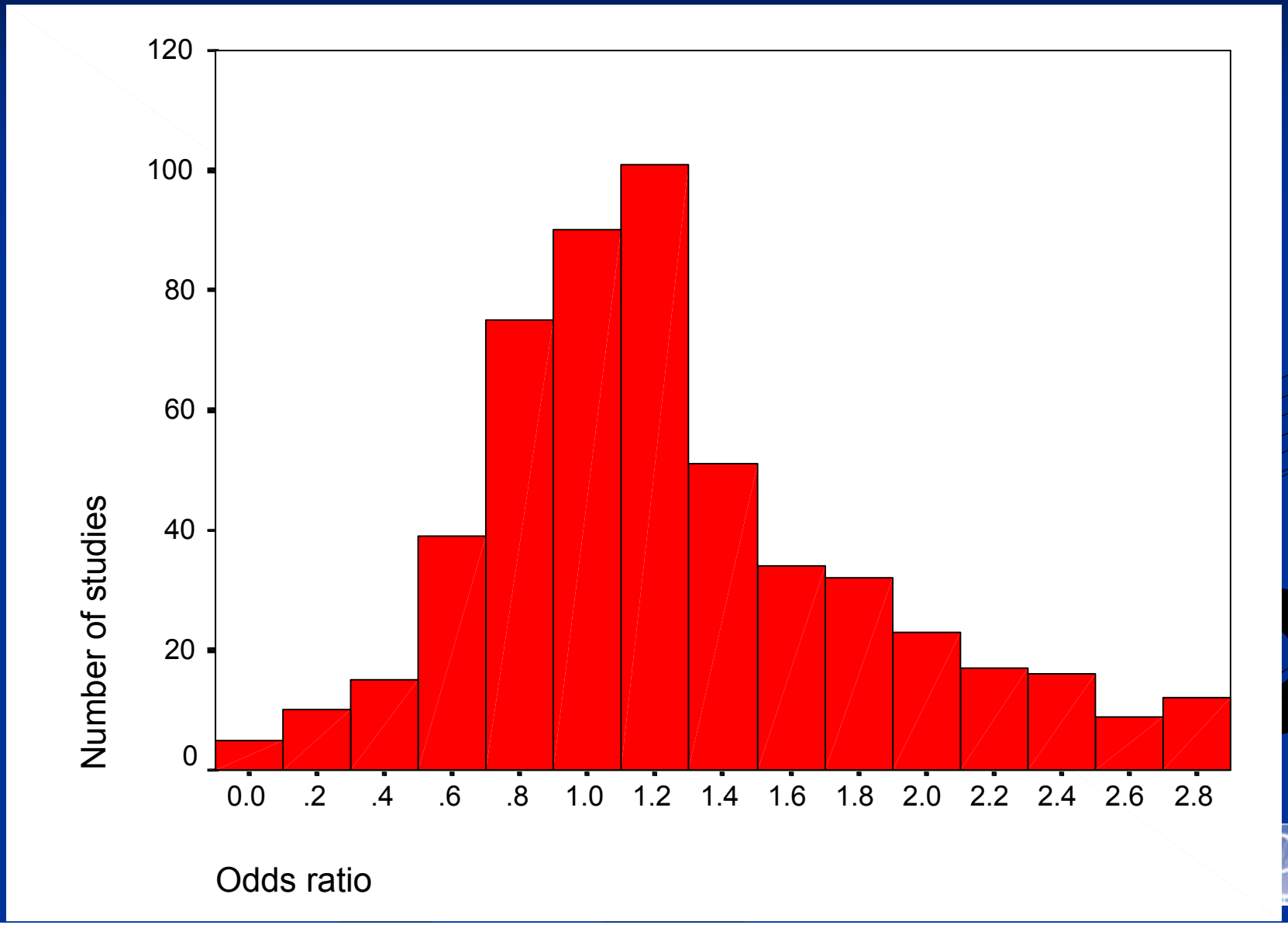


Small sample size of individual studies



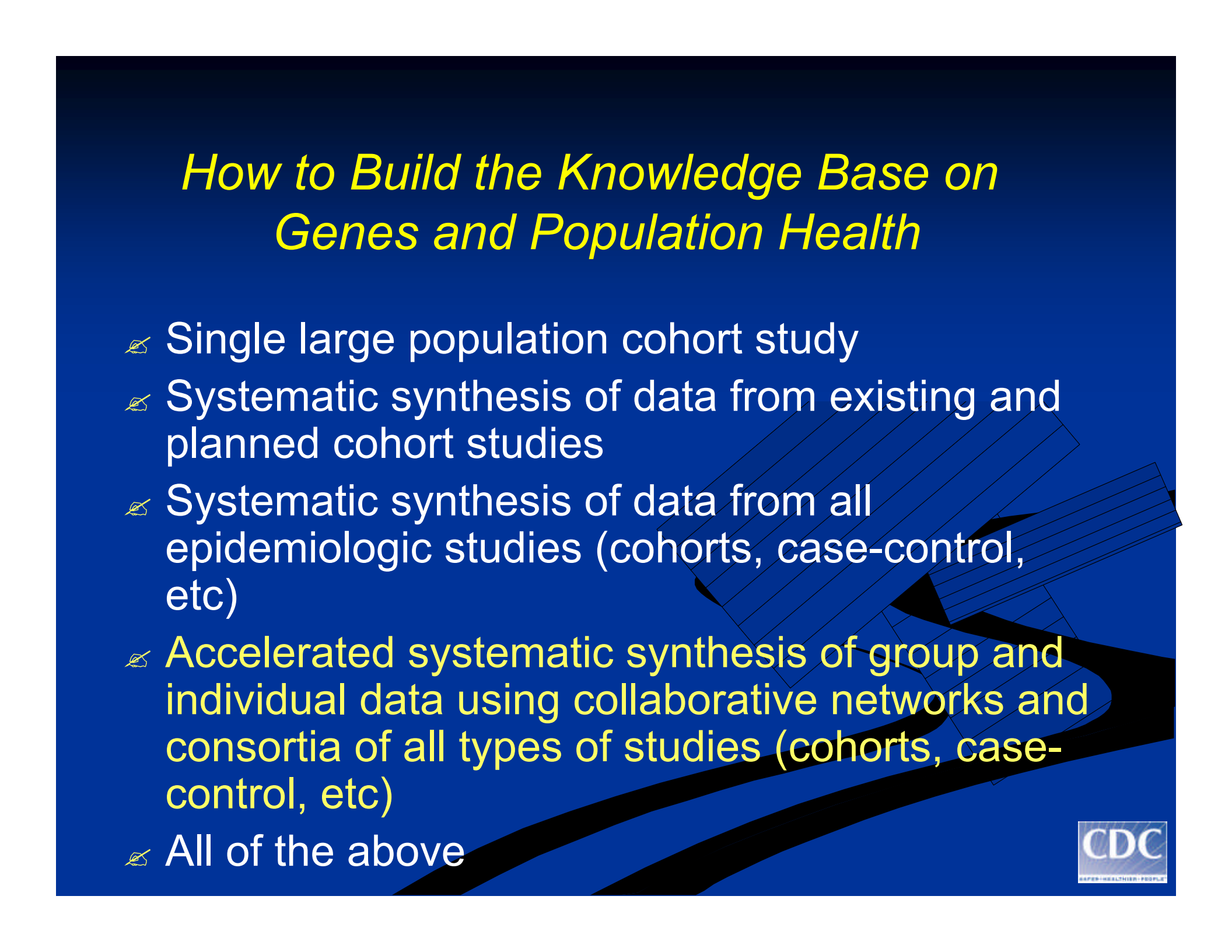
Ioannidis, Trends Mol Med 2003

Small effect sizes in individual studies



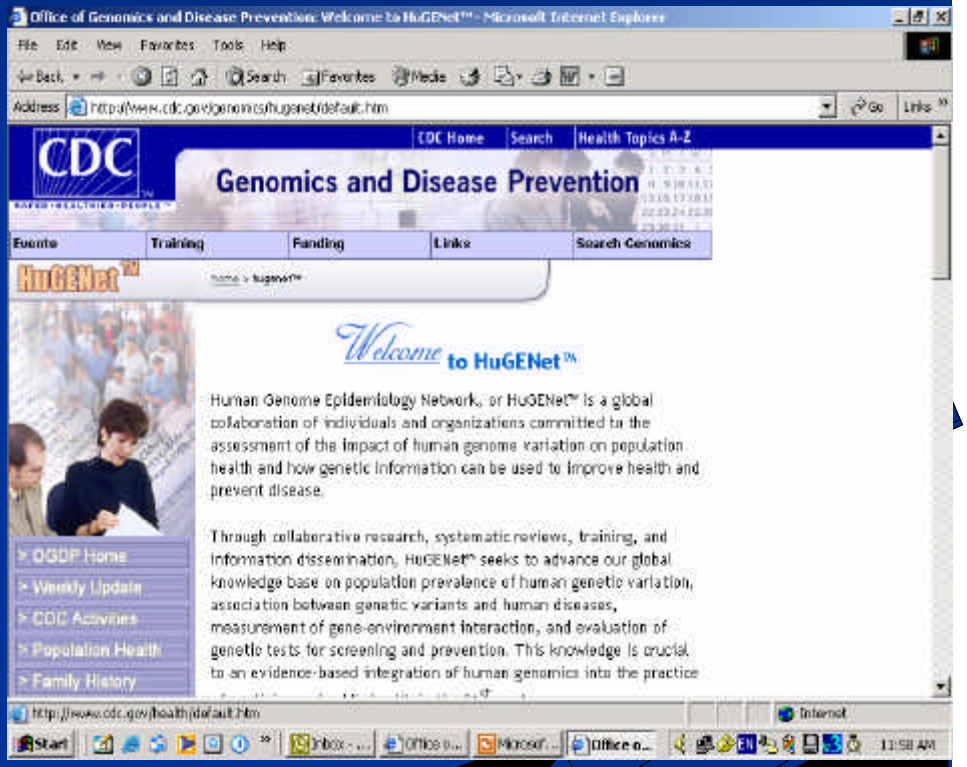
How to Build the Knowledge Base on Genes and Population Health

- Single large population cohort study
- Systematic synthesis of data from existing and planned cohort studies
- Systematic synthesis of data from all epidemiologic studies (cohorts, case-control, etc)
- Accelerated systematic synthesis of group and individual data using collaborative networks and consortia of all types of studies (cohorts, case-control, etc)
- All of the above



Human Genome Epidemiology Network (HuGE Net)

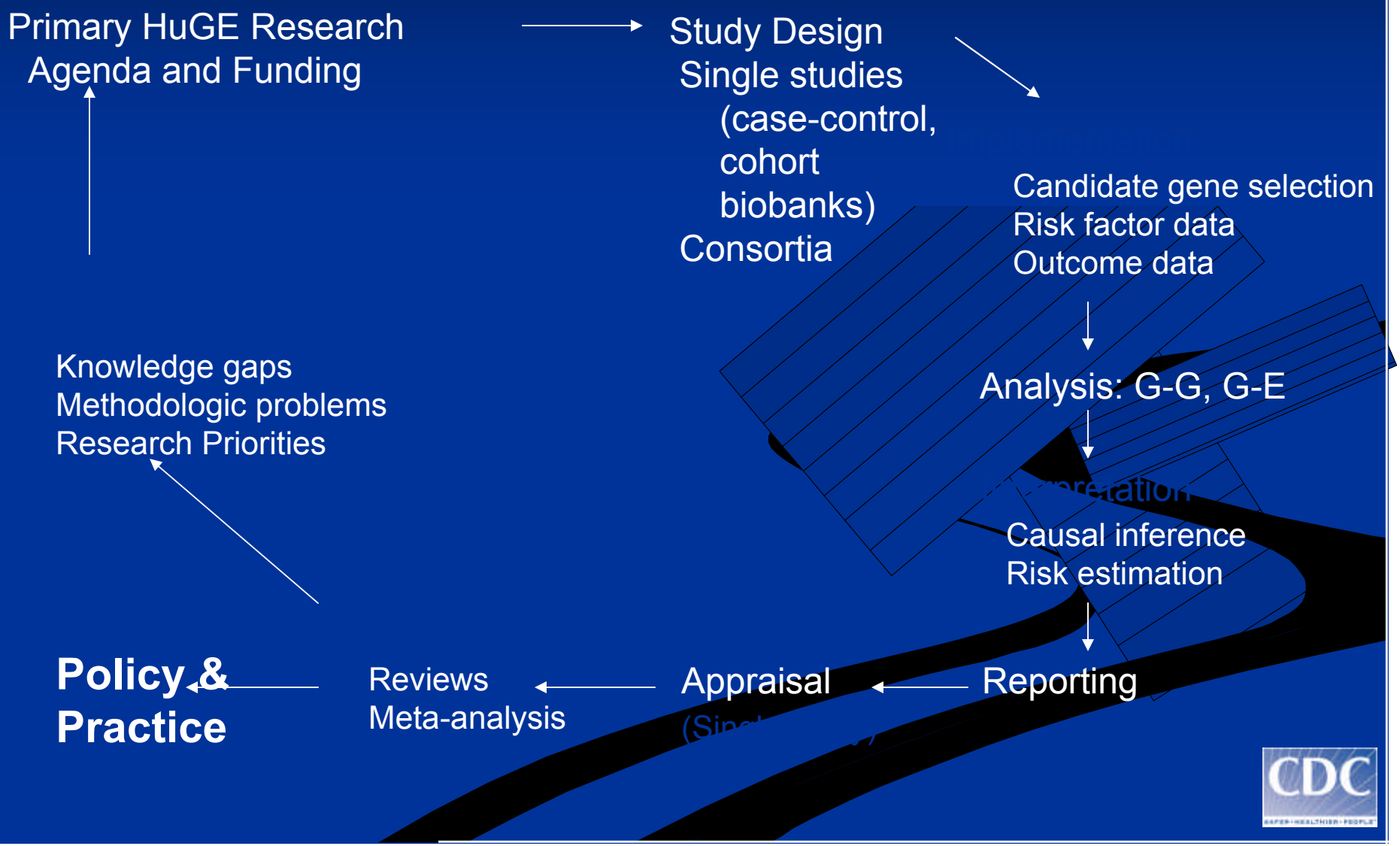
- **Global collaboration of individuals and organizations to assess population impact of genomics and how it can be used to improve health and prevent disease**
  - Information Exchange
  - Training and Technical Assistance
  - Knowledge Base Development
  - Information Dissemination



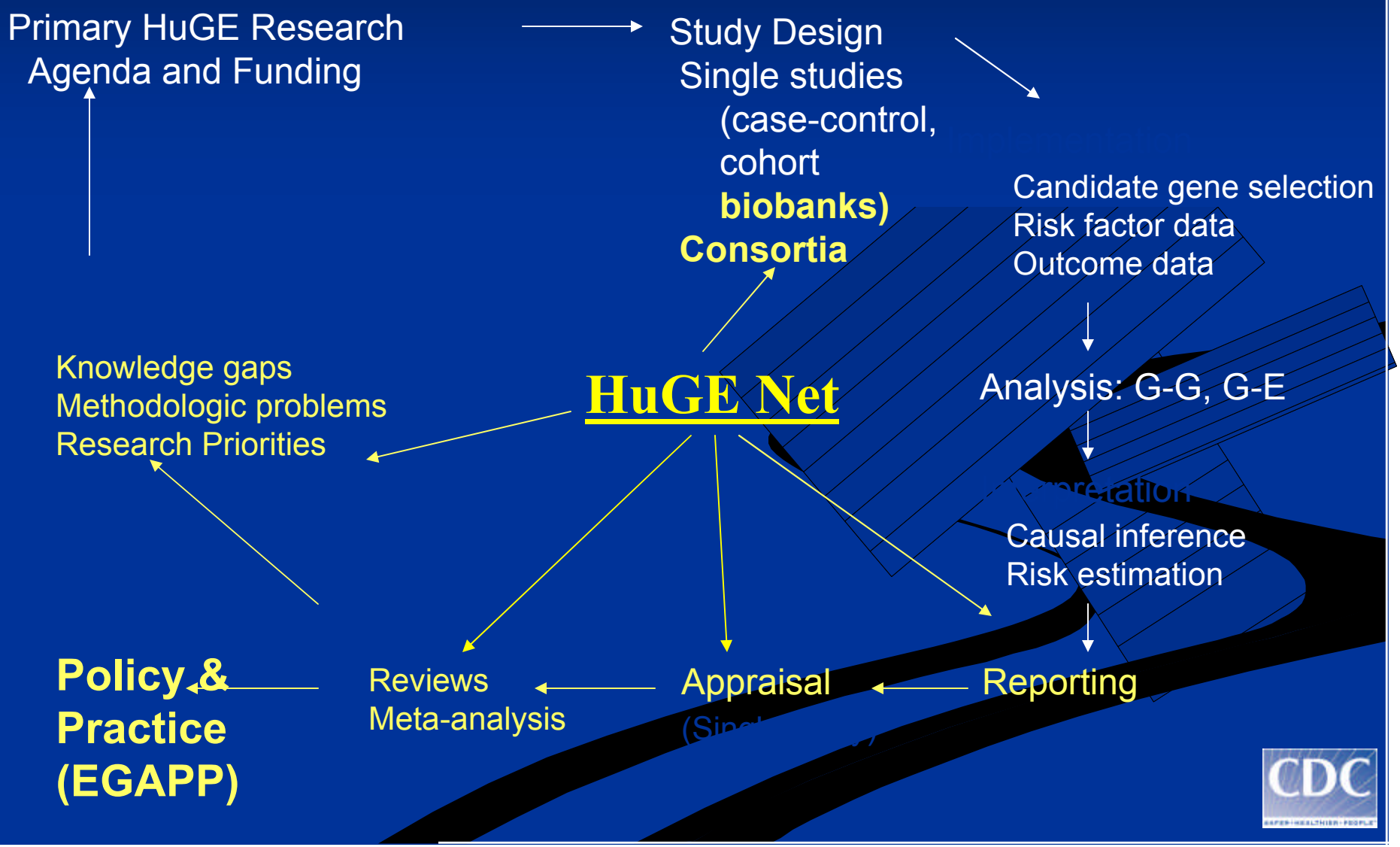
Selected HuGE Net Activities

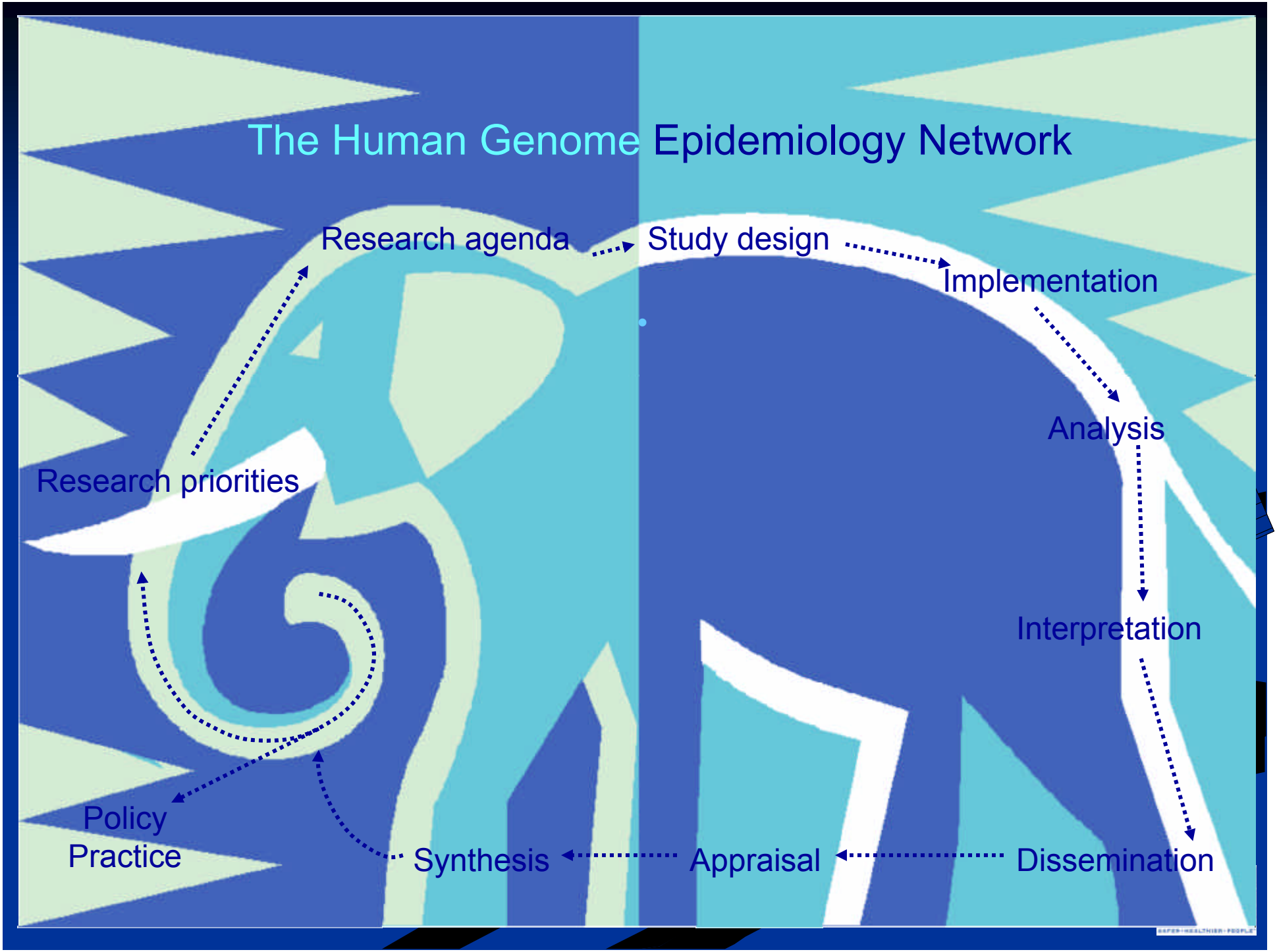
- HuGE Studies Database
- HuGE Reviews
- Methodology/Training Workshops
- International Biobank/Cohort Study meeting
- "Network of Networks"
- Connecting epidemiologic information to evidence-based evaluation of genomic applications in practice and prevention (EGAPP)

From HuGE Research to Synthesis & Dissemination for Policy and Practice

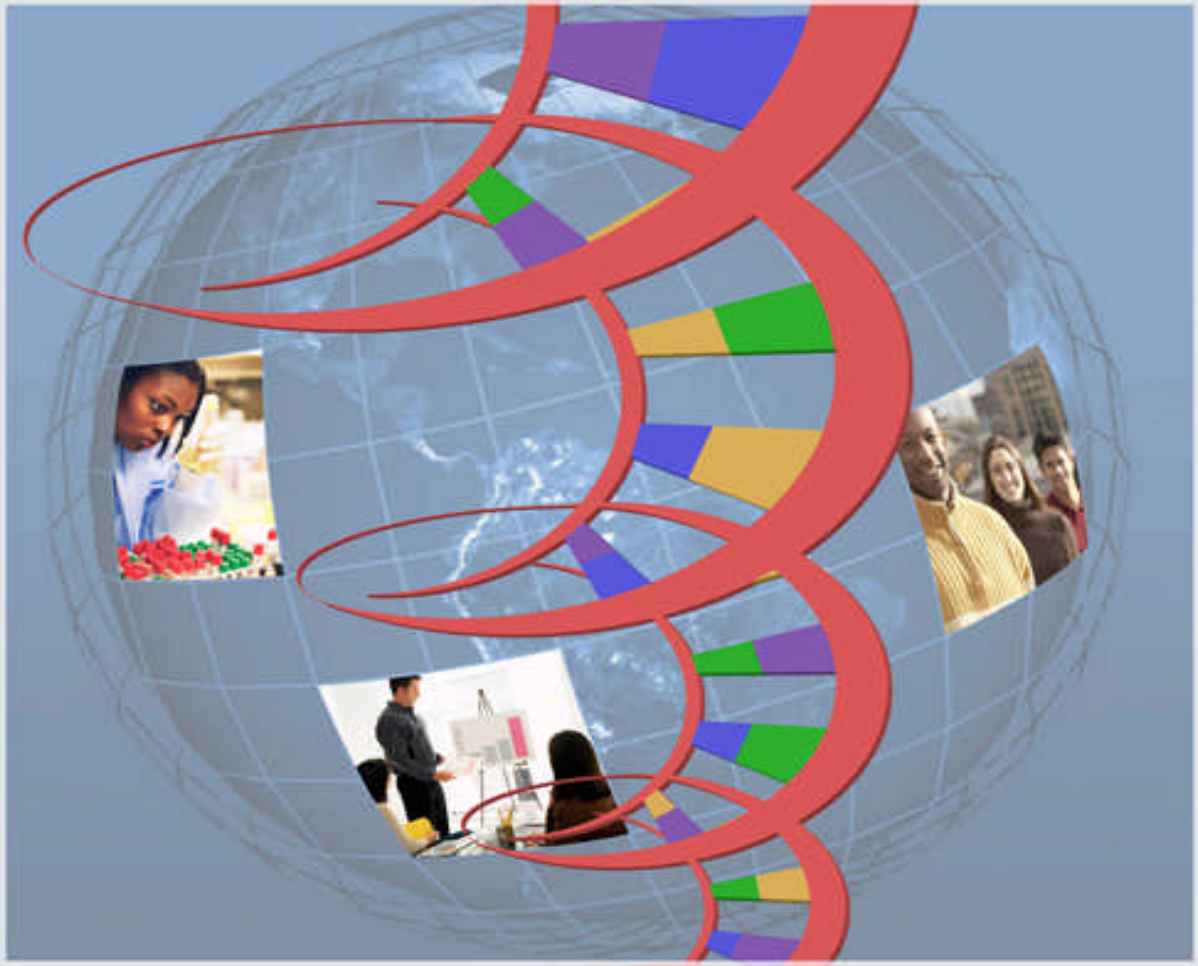


From HuGE Research to Synthesis & Dissemination for Policy and Practice




The Human Genome Epidemiology Network
Research agenda
Study design
Implementation
Analysis
Interpretation
Dissemination
Appraisal
Synthesis
Policy
Practice
Research priorities

International Biobank and Cohort Studies: Developing a Harmonious Approach



Sheridan Buckhead Hotel
Atlanta, GA
February 7-8, 2005

CDC

CDC

Biobank Meeting Discussions and Outcomes

- STROBE like statement for publishing studies derived from biobanks
- General report on the design and conduct of biobanks
- Online knowledge base with register of studies and tools (e.g. protocols, informed consent)
- Future periodic meetings with information sharing

Outline: Three Messages

- Global Collaboration in Biobanks/Population-based Cohort Studies
  - One cohort study in one country is not enough!
- Integration of all Human Genome Epidemiologic Data (cohort, case-control, other)
  - There is more than one way to get there!
- Development of Evidence-based Processes that use Human Genome Epidemiologic Data
  - Direct link from epidemiology to practice!