

# Melding Epidemiology and Genomics

## Key Epidemiologic Challenges

Robert N. Hoover, M.D., Sc.D.

Director

Epidemiology and Biostatistics Program

December 1, 2006

“To realize fully the immense promise of the Human Genome Project, we must now extend its scientific and technological approaches to understand the complex gene-gene and gene-environment interactions that determine human health and disease.”

Dr. Francis Collins

# Genes and Environment

- Role of “susceptibility” genes in cancer
  - Detect susceptibility
  - Detect gene-environment interactions (and multiples)
  - Detect previously unrecognized carcinogens

# Cigarette Smoking, Genotype, and Breast Cancer

- >100 epidemiologic studies on smoking and breast cancer → inconsistent results
- Since 1995, 50 studies have examined this relationship in relation to a total of 11 susceptibility genes
- Some evidence in meta analyses, “however interpretation of the available literature is complicated by methodologic limitations, including small sample size, ..... which likely contributed to the inconsistent findings. These methodologic issues should be addressed in future studies .....

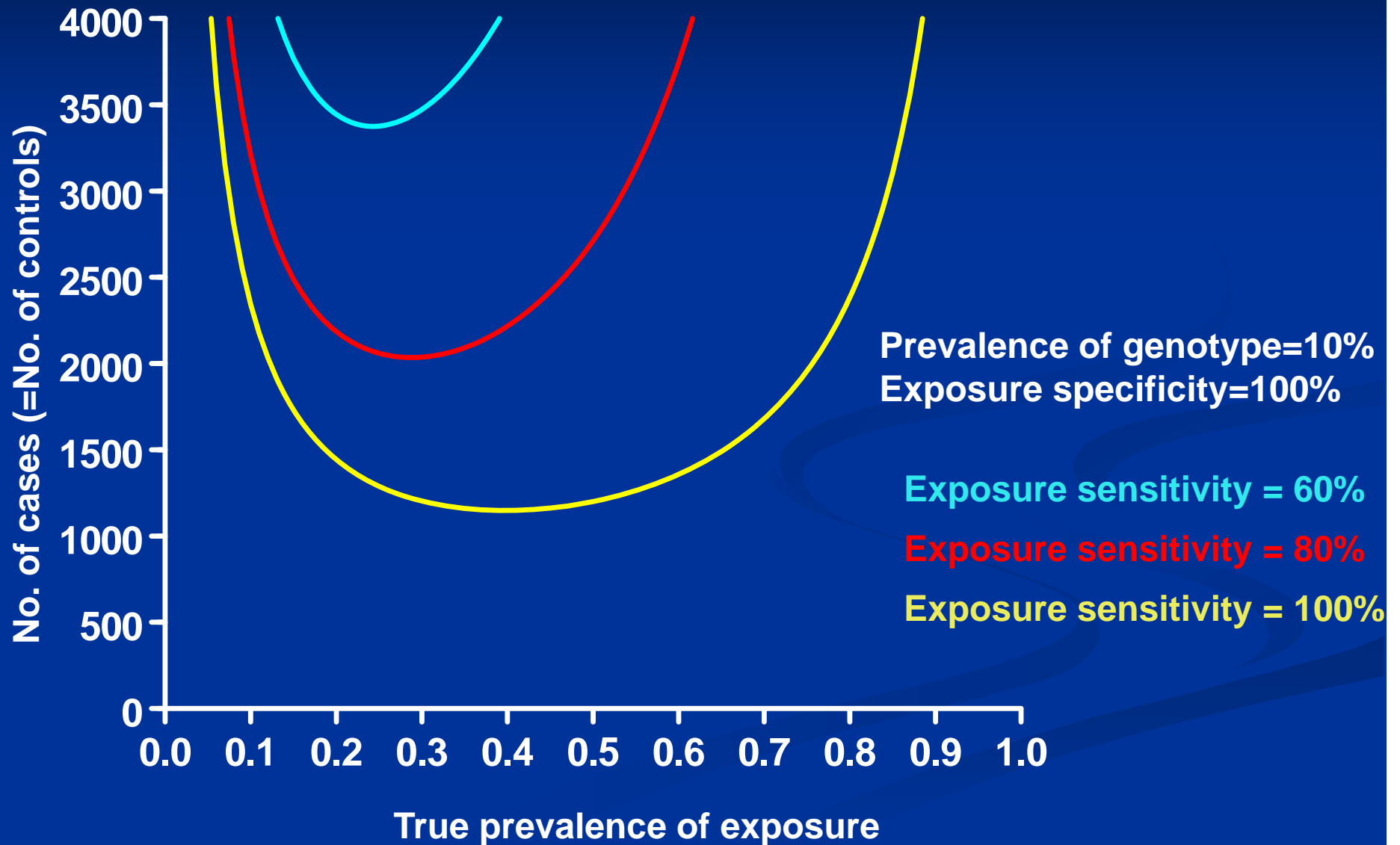
Cancer Epidemiol. Biomarkers Prev. 2006;15:602-11.



# Genes and Environment: The Dark Side

- Study Size
- Chance
- Bias

# Study Size to Detect a Two-Fold Interaction



# Genes and Environment: The Dark Side

- Study Size
- **Chance**
- Bias



# Chance

- 24,000 Genes
- 3 Billion Base Pairs
- 8+ Million Common SNPs
- Gene<sup>(n)</sup>-Environment<sup>(n)</sup>

# Simple Genetic Pathway: Environment Interaction

“Pathway” = 2 genes (10 SNPs each)  
1 Exposure = 2 levels (low and high dose)

Numbers of Interactions = 1,370  
Number of false positives ( $\alpha=0.05$ ) = 70

Remedies = p-value adjustment  
**REPLICATION**

# Genes and Environment: The Dark Side

- Study Size
- Chance
- **Bias**

# BIAS

## Lung Cancer Risk and CYP2D6\*

|                       | Study 1           | Study 2          | Study 3         |
|-----------------------|-------------------|------------------|-----------------|
| Relative Risk         | 15.6 (4.8 – 55.9) | 6.1 (2.2 – 17.1) | 0.6 (0.3 – 1.2) |
| Epidemiologic Quality | Low               | Intermediate     | High            |
| (% participation)     | (?)               | (26%)            | (80%)           |

\* Risk of homozygous extensive metabolizers compared to homozygous poor metabolizers.

# Gene-Environment Studies: Needs

- Very large studies
- Replication, replication, replication (planned and coordinated)
- Rigorous, high-quality design, conduct, analysis
  - Genomics
  - Epidemiology
  - Statistics
  - Informatics
- Data sharing

# Emerging New Research Paradigm

## CONSORTIA

- Cohort
- Case-Control
- Epidemiologists, Clinical and Molecular Scientists
- Intensely Collaborative
  - Common Protocol and Methods
  - Coordinated Parallel and Pooled Analyses
- Data Sharing

# The Cohort Consortium



25+ cohorts, over 2.6 million individuals (1.2 million with DNA collected at baseline)

## BPC3

Cohorts: ATBC, CPS II, EPIC, HPFS, MEC, NHS, PHS, PLCO, WHS

Risk Factors: Hormone risk factors and hormones

Genes: 53 in steroid hormone and growth factor pathways

Cancer Sites: Breast & Prostate cancer

Cases with DNA: 6,000 breast and 7,000 prostate

Website:  
<http://epi.grants.cancer.gov/BPC3>

## CGEMS

Scan: PLCO, NHS

Replication: ATBC, CPS II, EPIC, HPFS, MEC, PHS, WHS, WHI

Risk factors: Same as BPC3 plus family history

Genes: Genome-wide Association Study (GWAS)

Cancer Sites: Breast & Prostate cancer

Cases with DNA: 8,850 breast and 8,000 prostate

Data Portal:  
<https://caintegrator.nci.nih.gov/cgems/>

## PanScan

Scan: ATBC, CLUE II, CPS II, EPIC, HPFS, NHS, NYUWHS, PHS, PLCO, SWHS, WHI, WHS

Replication: Pancreatic Cancer Case-control Consortium (PANC4)

Risk Factors: Tobacco, obesity, family history and diabetes

Genes: Genome-wide Association Study (GWAS)

Cancer Site: Pancreatic cancer

Cases with DNA: ~1,900 pancreatic

Website:  
Coming soon...

# Established Case-Control Consortia

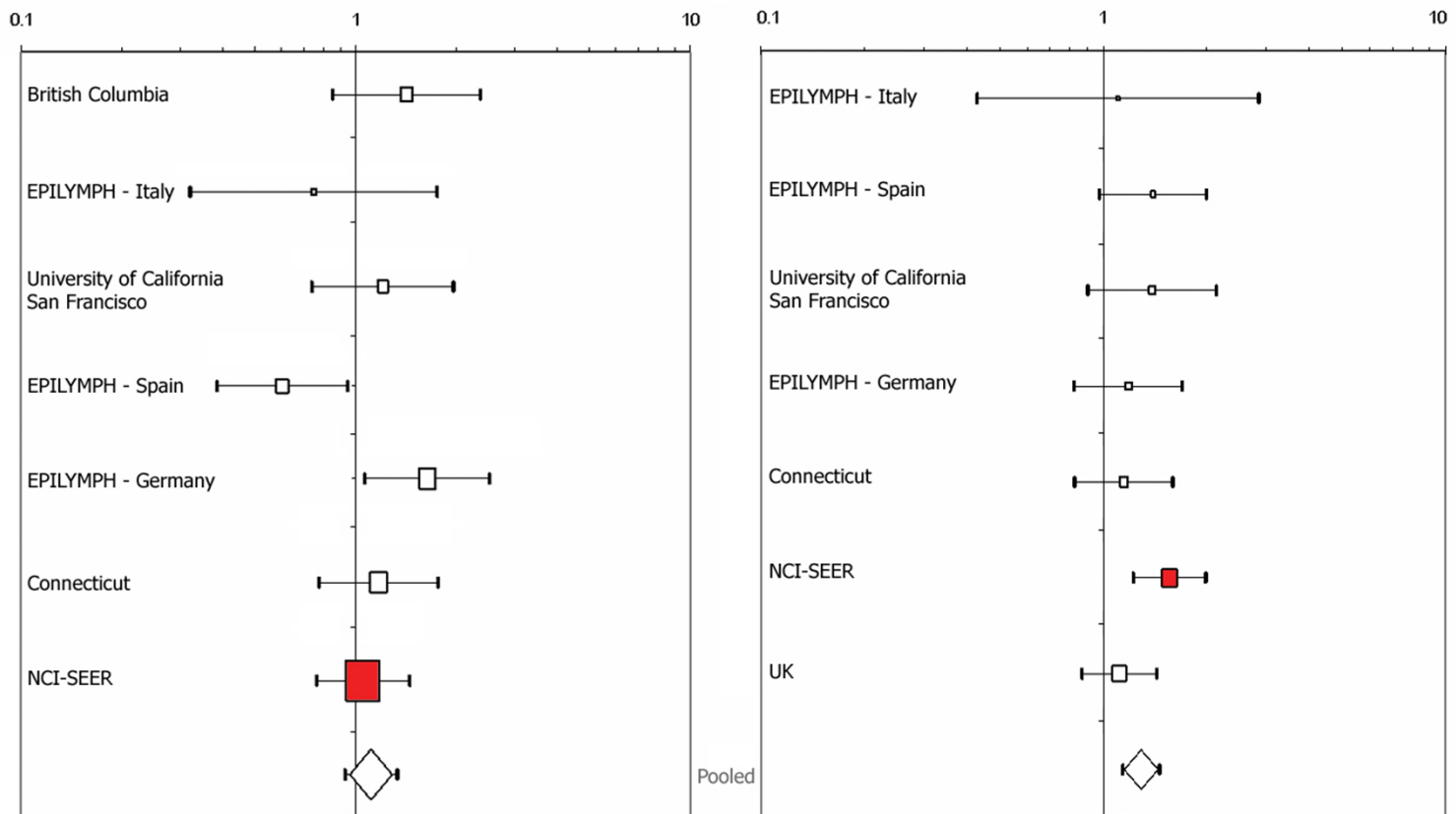
- Bladder Cancer Consortium
- Brain Tumor Epidemiology Consortium (BTEC)
- Epidemiology of Endometrial Cancer Consortium (E2C2)
- Esophageal Adenocarcinoma and Barrett's Esophagus Consortium (BEACON)
- International Consortium on Lymphoma Epidemiologic Studies (InterLymph)
- International Consortium on Prostate Cancer Genetics (ICPCG)
- International Head and Neck Cancer Consortium (INHANCE)
- International Lung Cancer Consortium (ILCCO)
- International Genetic Melanoma Consortium (GenoMel)
- Molecular Epidemiology of Colorectal Cancer (MECC)
- Pacific Ovarian Cancer Research Consortium
- Pancreatic Cancer Case-Control Consortium (PANC4)
- Prostate Cancer Genetics Study (CaP Genes)
- Western Pancreatic Cancer Consortium
- and others...



# InterLymph

International Lymphoma Epidemiology Consortium

21 member studies, over 18,000 cases of NHL



IL1B-511 variant

TNF G-308A variant

Analysis of *TNF-308* and Risk of DLBC  
Lymphoma:  
*Pooled Analysis from 7 Studies*

| <u><i>TNF-308 Genotype</i></u> | <u>Relative Risk</u> | <u>P-value</u> |
|--------------------------------|----------------------|----------------|
| GG                             | 1.00                 | baseline       |
| GA                             | 1.29                 | 0.002          |
| AA                             | 1.65                 | 0.005          |

p trend = 0.000055

# Summary

- Exciting, unprecedented opportunities for insights into genetic pathways and environmental interactions that determine human health and disease
- Daunting, unprecedented challenges to exploiting these opportunities
- Emerging science and research paradigms allowing us to overcome these challenges