National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Cancer Advisory Board

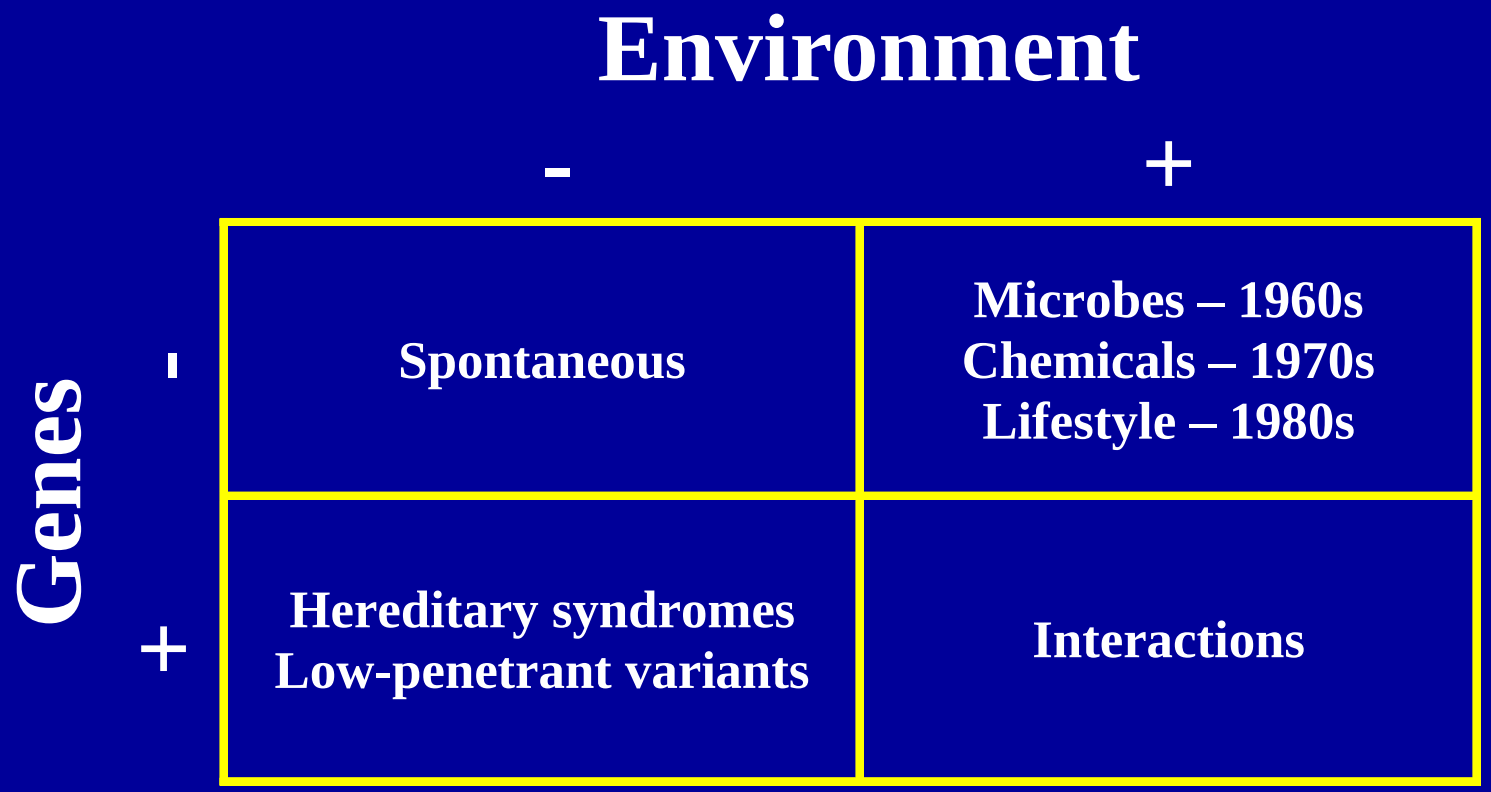
Genome-wide Association Studies: Introduction

Joseph F. Fraumeni, Jr., M.D.

Division of Cancer Epidemiology and Genetics

December 9, 2008

Categories of Cancer Causation

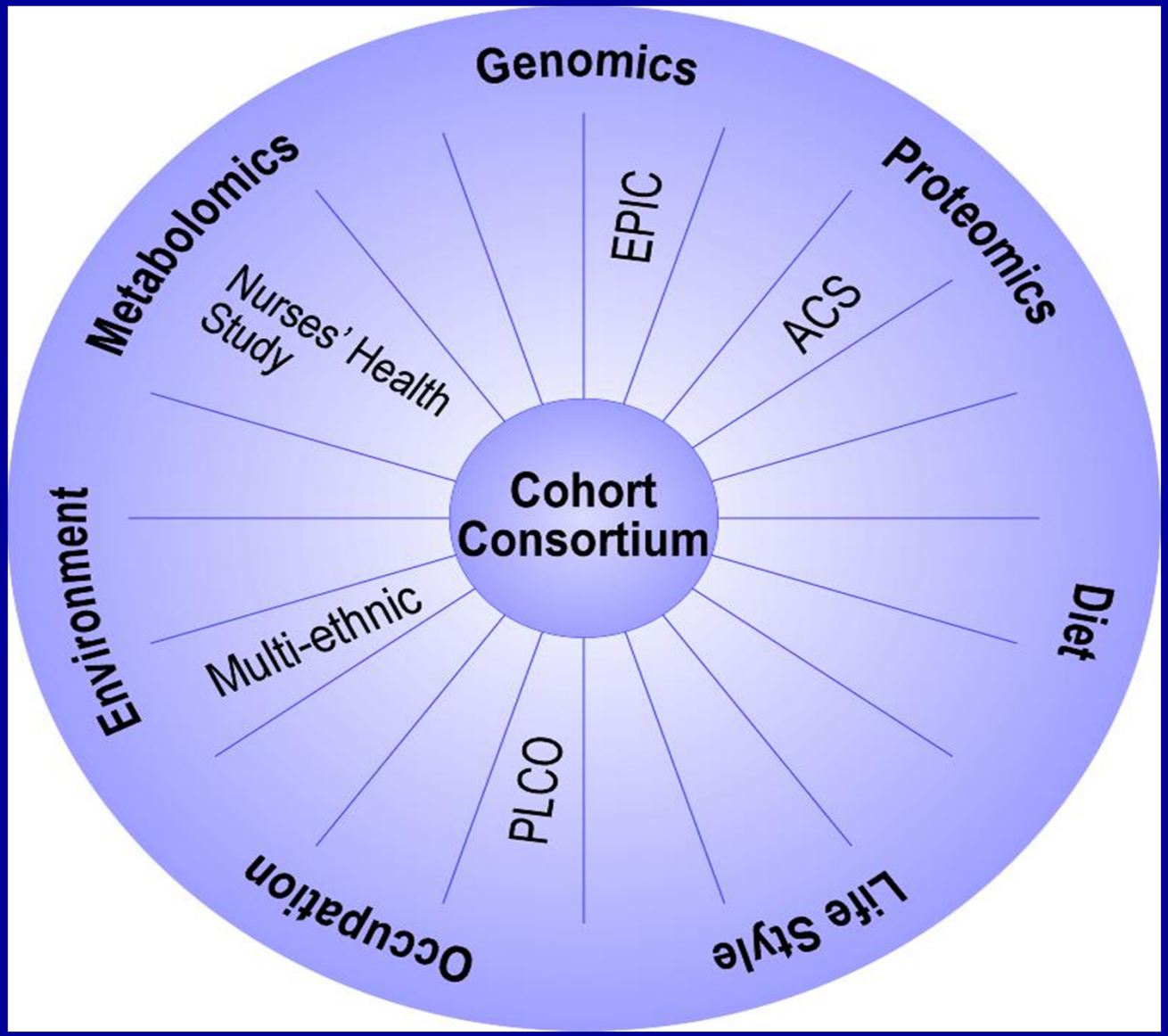


The Advent of GWAS at NCI

- A shift from candidate gene searches to systematic genome-wide scans (CGEMS)
- Large-scale association studies with staged replication strategies through intramural/extramural consortia with DCCPS
- Resequencing and fine-mapping of genomic loci to narrow the search for inherited genetic variants
- Downstream research with CCR: Functional studies, germline vs. somatic changes, pharmacogenomics, gene-environment interactions, risk prediction models, clinical applications
- Rapid dissemination of publicly available data (caBIG)

NCI Consortium of Cohorts

37 Population Cohorts with Nearly 4 Million Individuals (December 2008)



STATISTICS AND MEDICINE

Drinking from the Fire Hose — Statistical Issues in Genomewide Association Studies

David J. Hunter, M.B., B.S., and Peter Kraft, Ph.D.

Related article, page 443

"There have been few, if any, similar bursts of discovery in the history of medical research…"

The past 3 months have seen the publication of a series of stud… gene… mon… canc… and in this issue of the *Journal*, coronary artery disease (reported by Samani et al., pages 443–453). These genomewide association studies have been able to examine interpatient differences in inherited genetic variability at an unprecedented level of resolution, thanks to the development of microarrays, or chips, capable of assating the need for guessing which genes are likely to harbor variants lated to the disease. Some of these associations have been found in regions not even known to harbor genes, such as the 8q24 region, in which multiple variants have been found to be associated with prostate cancer.[2] Such findings promise to open up new avenues of research, through both the discovery of new genes rele-

The main problem with this strategy is that, because of the …ips, most studies are severely constrained in …er of samples …ited power to generate P values as small as 10^{-7}. In addition, most variants identified recently have been associated with modest relative risks (e.g., 1.3 for heterozygotes and 1.6 for homozygotes), and many true associations are not likely to exceed P values as extreme as 10^{-7} in an initial study. On the other hand, a "statistically significant" finding

Hunter DJ and Kraft P. N Engl J Med 2007; 357: 436-439