National Cancer Institute

How Do We Assess the Contribution of Complex Genotypes and Gene-Environment Interaction to the Population Burden of Common Diseases?

Cancer genetic epidemiology: what has worked and what has not worked?

Robert N. Hoover

Director
Epidemiology and Biostatistics Program
Division of Cancer Epidemiology and Genetics
National Cancer Institute

February 15, 2007

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Cancer Institute

Familial Cancer & Public Health

- Low-Prevalence, High-Penetrance Genes
  - High-Risk Groups – Primary and Secondary prevention
  - Understand Biology

National Cancer Institute

High-Penetrance Genes

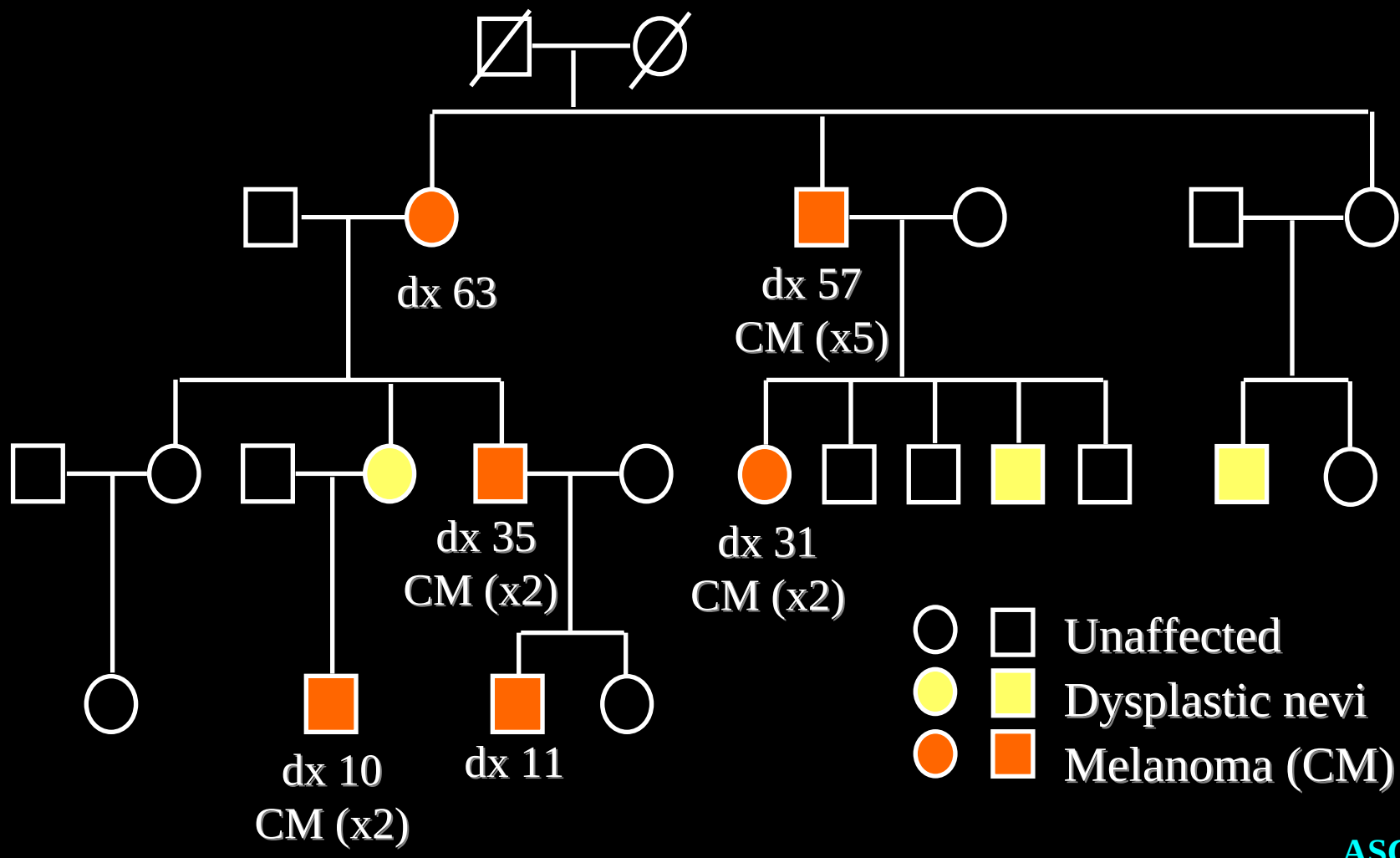
- Familial Syndromes
- “Loaded” Families

National Cancer Institute

Cloned Familial Tumor Suppressor Genes

Retinoblastoma	RB1	13q14	1986
Wilms' tumor	WT1	11p13	1990
Li-Fraumeni syndrome	p53	17p13	1990
Neurofibromatosis 1	NF1	17q11	1990
Neurofibromatosis 2	NF2	22q12	1993
von Hippel-Lindau syndrome	VHL	3p25	1993
Familial melanoma 1	p16	9p21	1994
Familial breast cancer 1	BRCA1	17q21	1994
Familial breast cancer 2	BRCA2	13q12	1995
Basal cell nevus syndrome	PTC	9q22	1996

Hereditary Cutaneous Melanoma



ASCO

National Cancer Institute

High-Penetrant Genes - Successes

- Breast Cancer: BRCA1+2
- Colon Cancer: FAP, HNPCC
- Melanoma: p16
- Prostate: 8 “putative” susceptibility genes

National Cancer Institute

Finding High-Penetrance Genes: Indices of Success

Characteristic	Contribution	
	Biological	Practical
Syndromes		
Other Conditions	√	
Unique Phenotypes	√	
"Loaded" Families	√	√
Common Tumors		√
Reasonable Survival		√
Consortia		√

National Cancer Institute

Susceptibility Genes: Low-penetrant, common variants

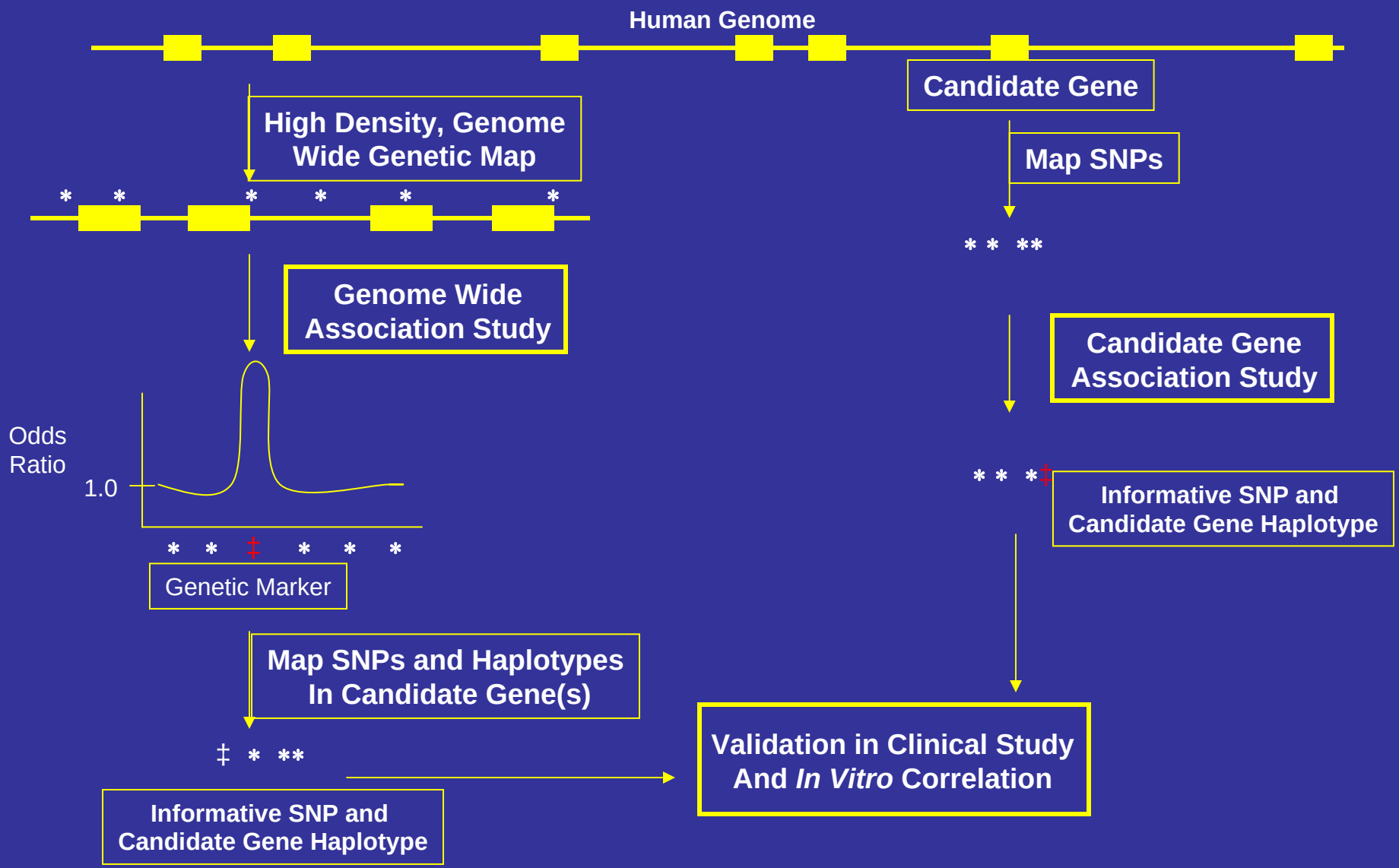
1. General: Carcinogenic mechanisms (e.g. - cell-cycle, DNA repair)
2. Specific: Modify effect of specific exposures
3. “Risk Factor” (e.g. - tobacco addiction, obesity

National Cancer Institute

Gene-Environment Studies: WHAT WORKS

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

Parallel Approaches To Identifying Genetic Determinants of Disease



National Cancer Institute

Review of Genetic Association Studies

- 603 associations of polymorphisms and disease
- 166 studied in at least three populations
- Only six seen reproducibly (>75% of studies)

Hirschhorn et al., Genetics in Medicine, 2002

National Cancer Institute

Interaction of NAT2 and Active Cigarette Smoking in Breast Cancer Risk

Study Year	# of Cases	Smoking Group	Subgroup	NAT2 ↑ risk group
1996	304	All	Postmenop	Slow
1997	466	Former	All	Rapid
1998	498	Recent	Postmenop	Rapid
2000	113	---	---	None
2000	177	All	Postmenop	Rapid
2001	149	All	All	Rapid
2002	442	Recent	≤50 yrs.	Slow

National Cancer Institute

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.

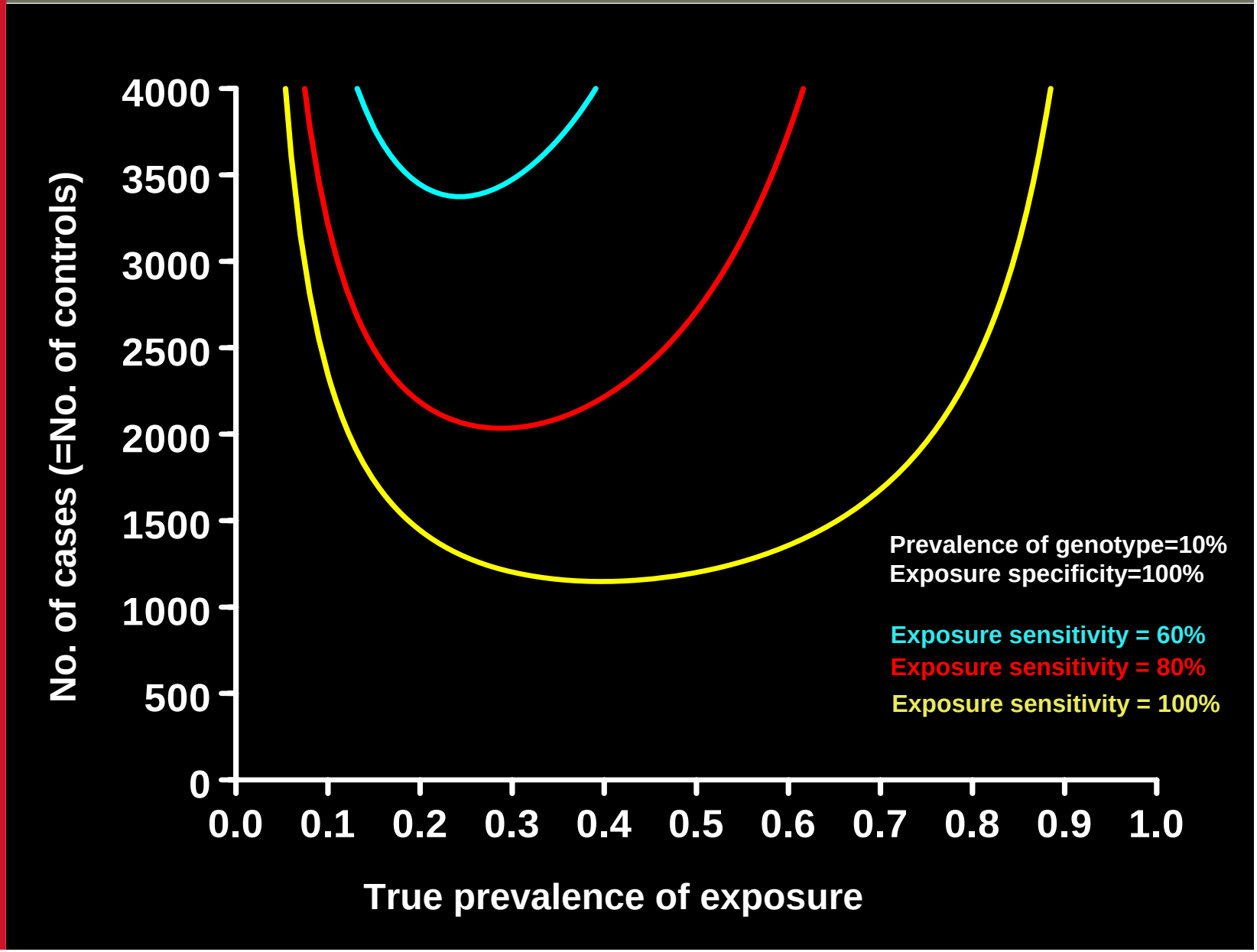
National Cancer Institute

Genes and Environment: The Dark Side

- Study Size
- Chance
- Bias

National Cancer Institute

Study Size to Detect a Two-Fold Interaction



National Cancer Institute

Genes and Environment: The Dark Side

- Study Size
- Chance
- Bias

National Cancer Institute

Simple Genetic Pathway: Environment Interaction

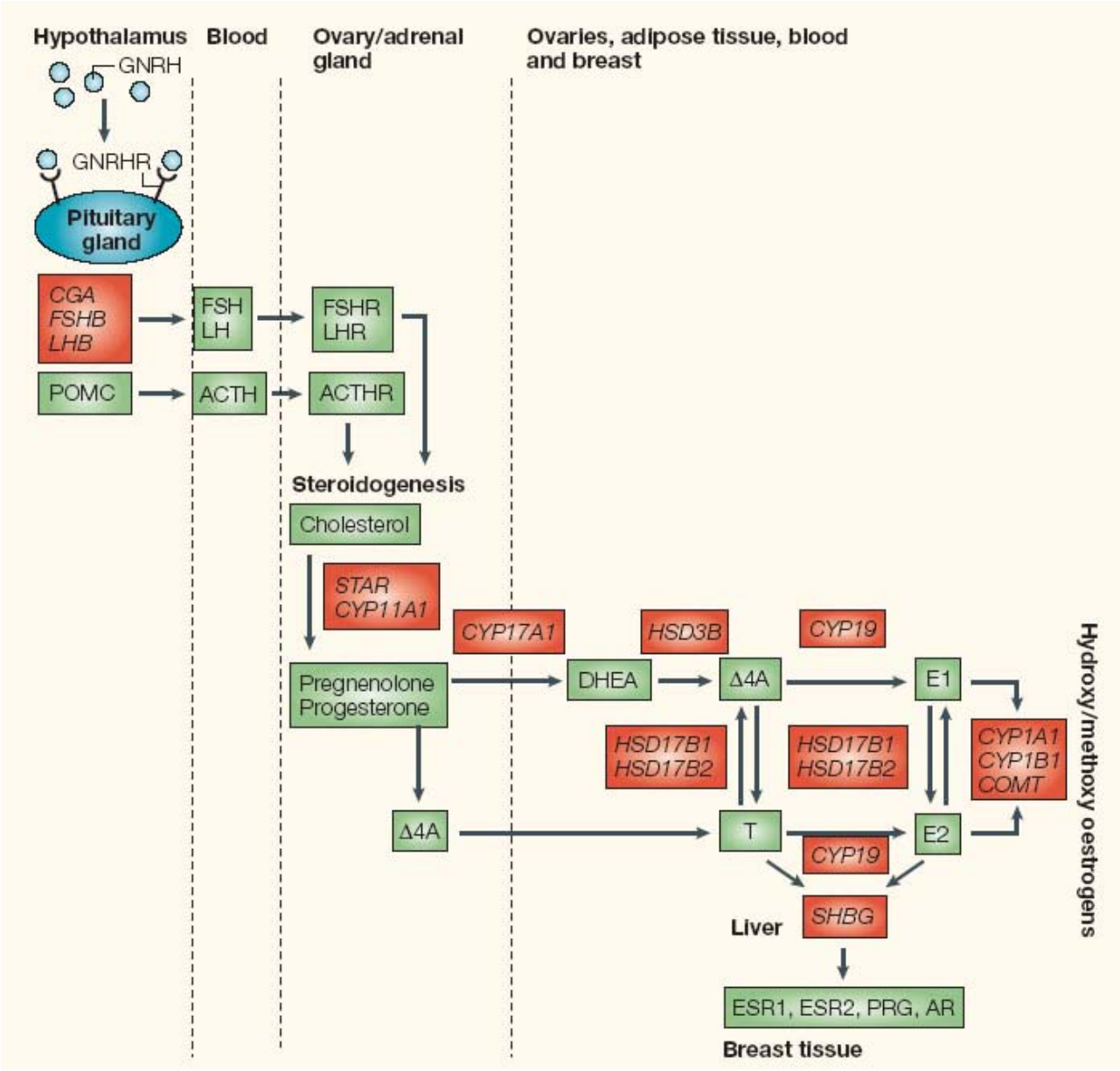
"Pathway" = 2 genes (10 SNPs each)
1 Exposure = 2 levels (low and high dose)

Number of interactions = 1,370
Number of false positives (α=0.05) = 70

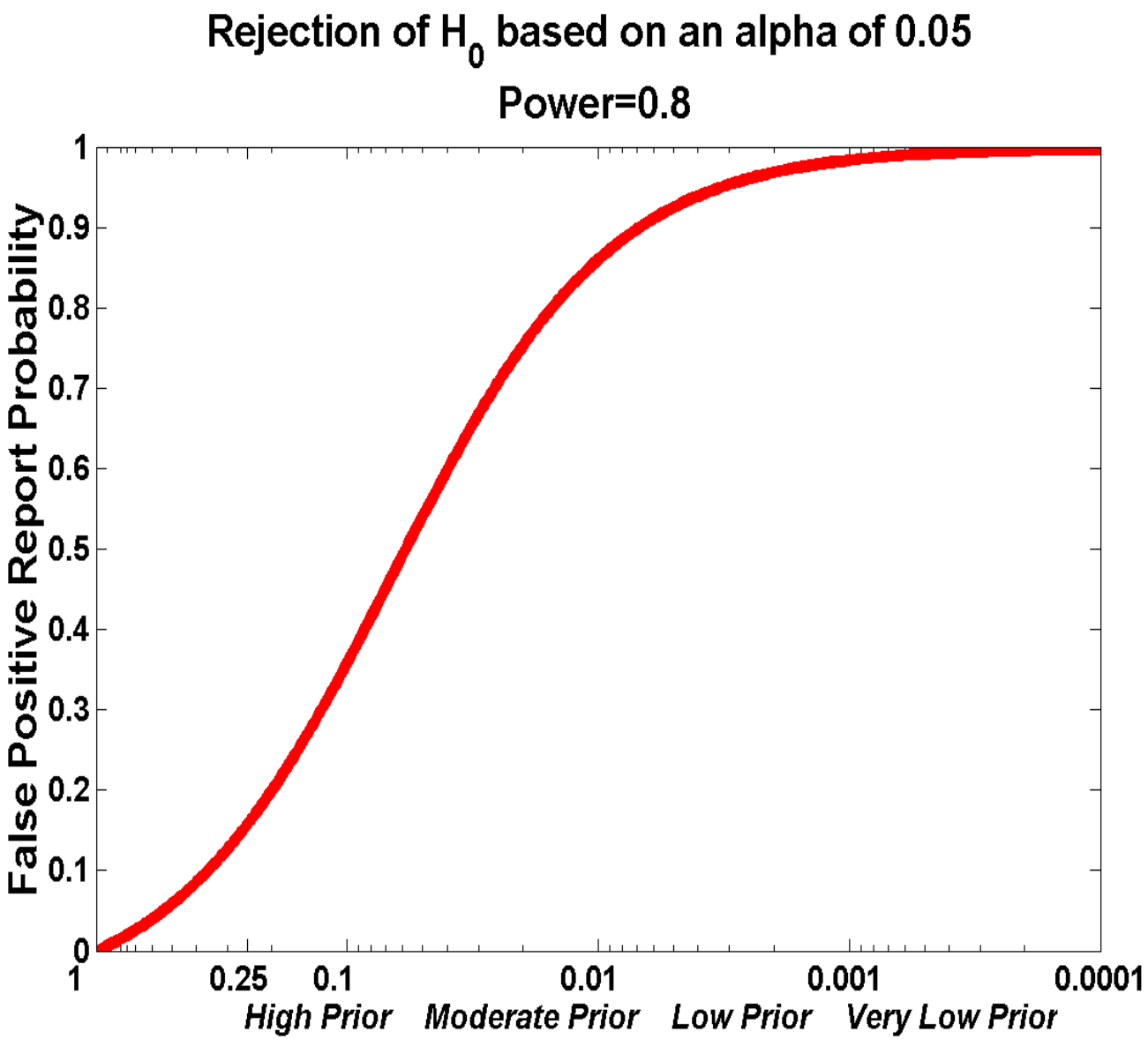
Remedies = p-value adjustment
REPLICATION

National Cancer Institute

BPC3 Methods



National Cancer Institute



National Cancer Institute

Genes and Environment: The Dark Side

- Study Size
- Chance
- Bias

National Cancer Institute

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.**

National Cancer Institute

Emerging New Research Paradigm

<u>CONSORTIA</u>

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

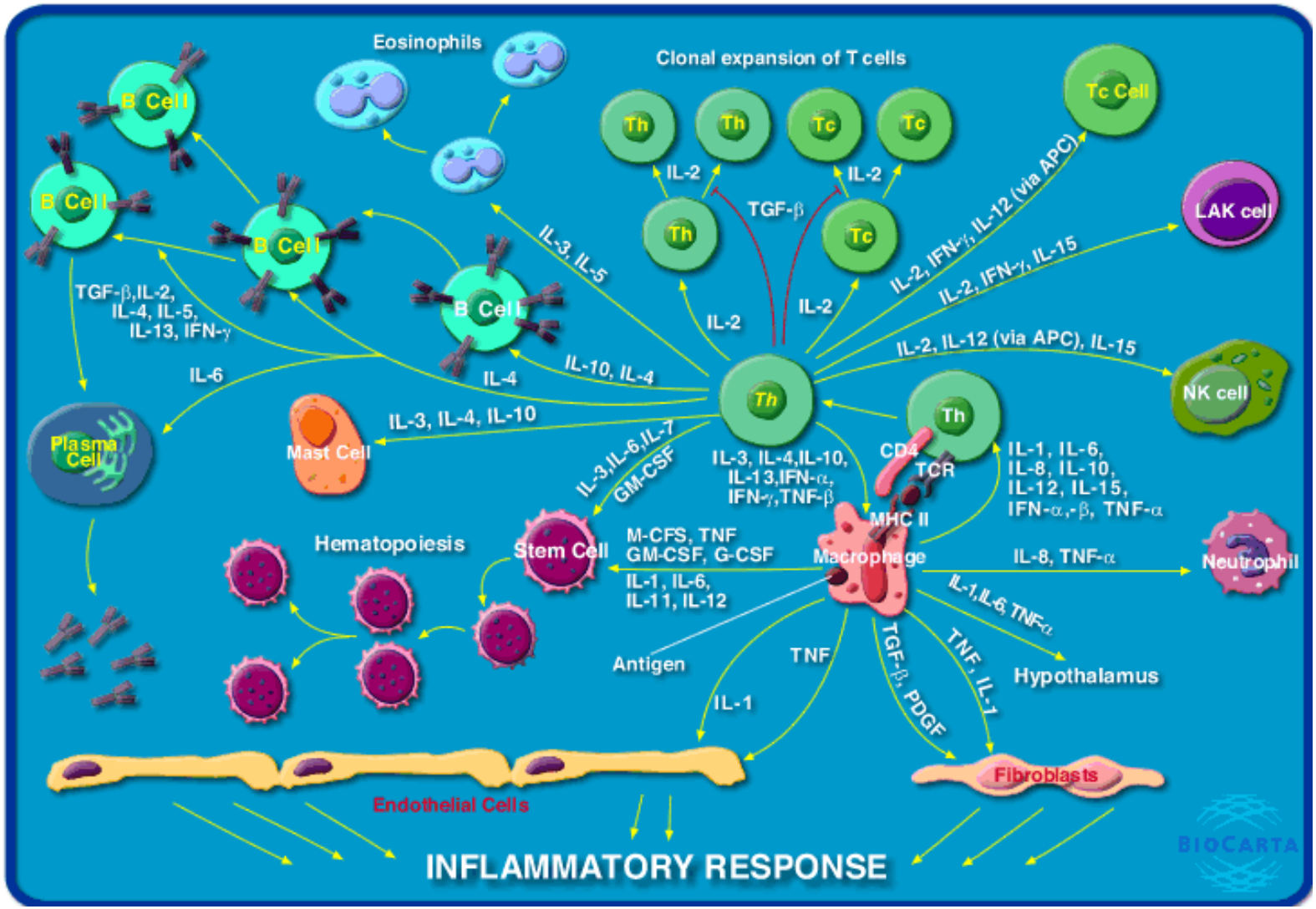
National Cancer Institute

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…

National Cancer Institute

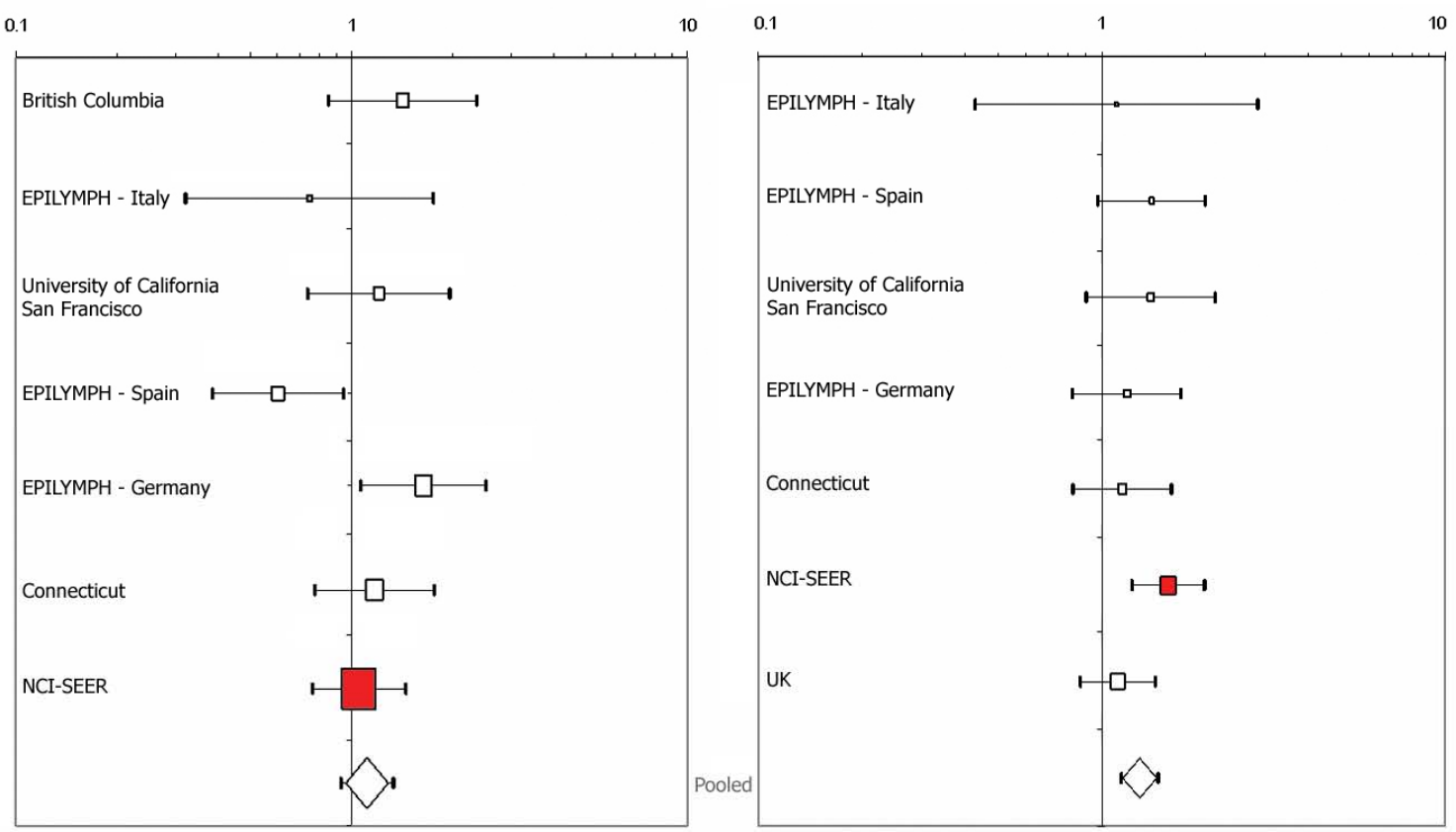
Genes in the Immune Pathway



InterLymph

International Lymphoma Epidemiology Consortium

21 member studies, over 18,000 cases of NHL



IL1B-511 variant

TNF G-308A variant

★ Participating center

Rothman N, Skibola C, Morgan G et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. Lancet Oncol. 2006 Jan;7(1):27-38.

National Cancer Institute

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

TNF-308 Genotype	Relative Risk	p-value
GG	1.00	baseline
GA	1.29	0.002
AA	1.65	0.005

p trend = 0.000055

Rothman et al., Lancet Oncology, 2006

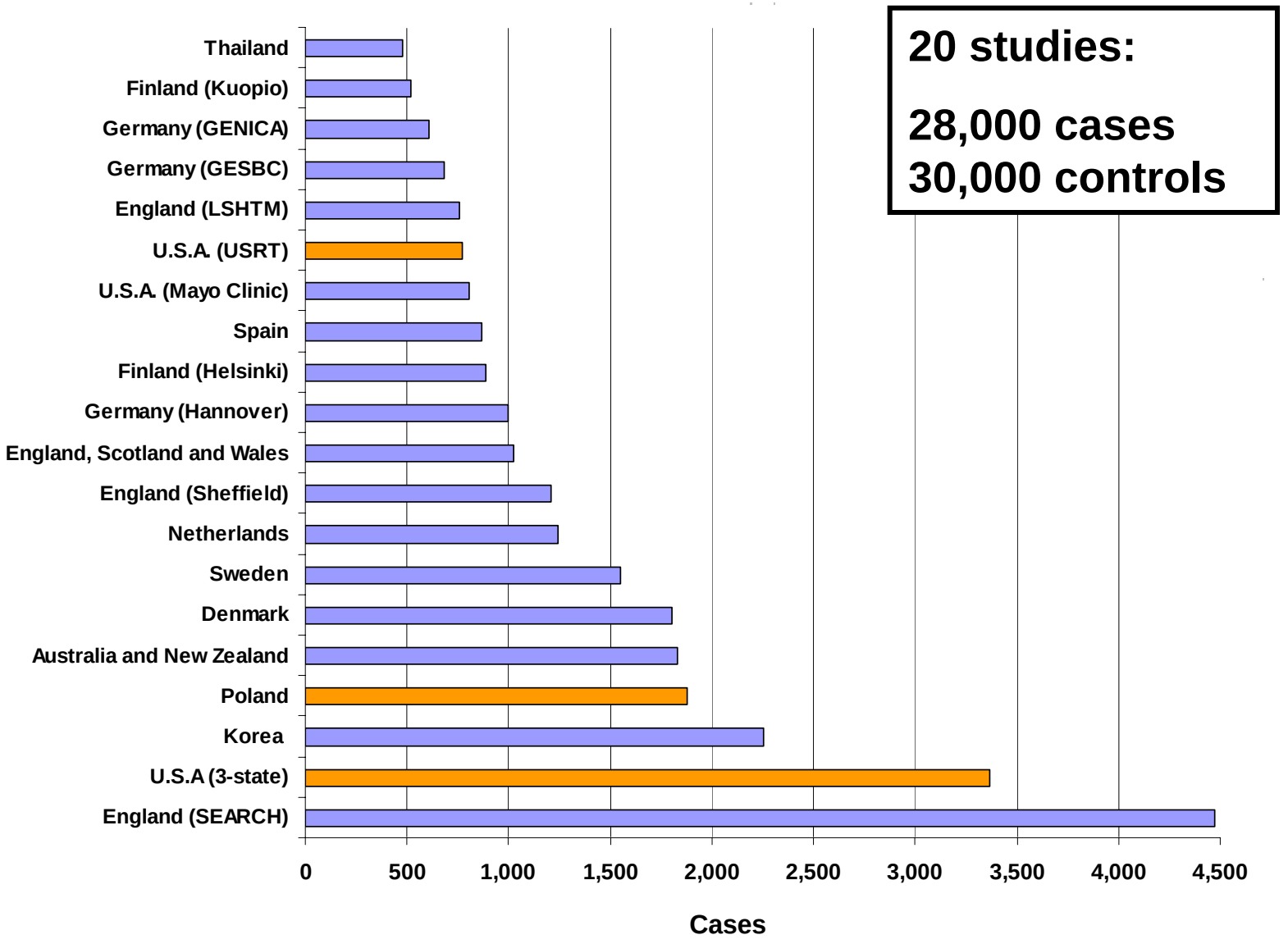
National Cancer Institute

Pathways of Interest in Breast Cancer

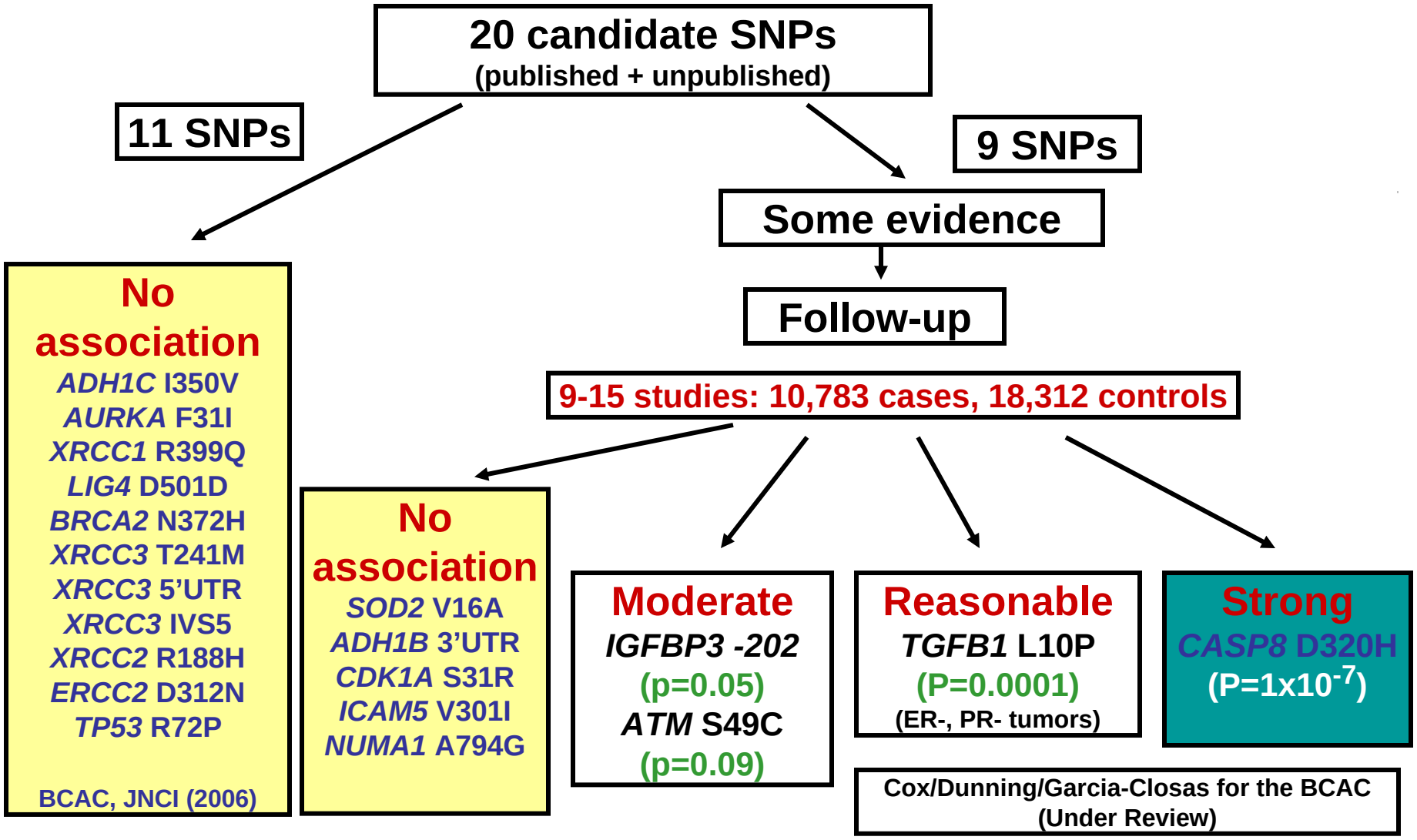
- **Established or possible risk factors:**
  - **Hormone biosynthesis, metabolism, and action**
  - **Obesity**
  - **Alcohol metabolism**
  - **Carcinogen metabolism**
  - **Inflammation**
- **Carcinogenic processes:**
  - **DNA repair, cell cycle control, and apoptosis**
  - **Cell signaling pathways**
  - **Telomere length**
- **Gene expression studies**
- **Somatic mutations**

National Cancer Institute

Breast Cancer Association Consortium

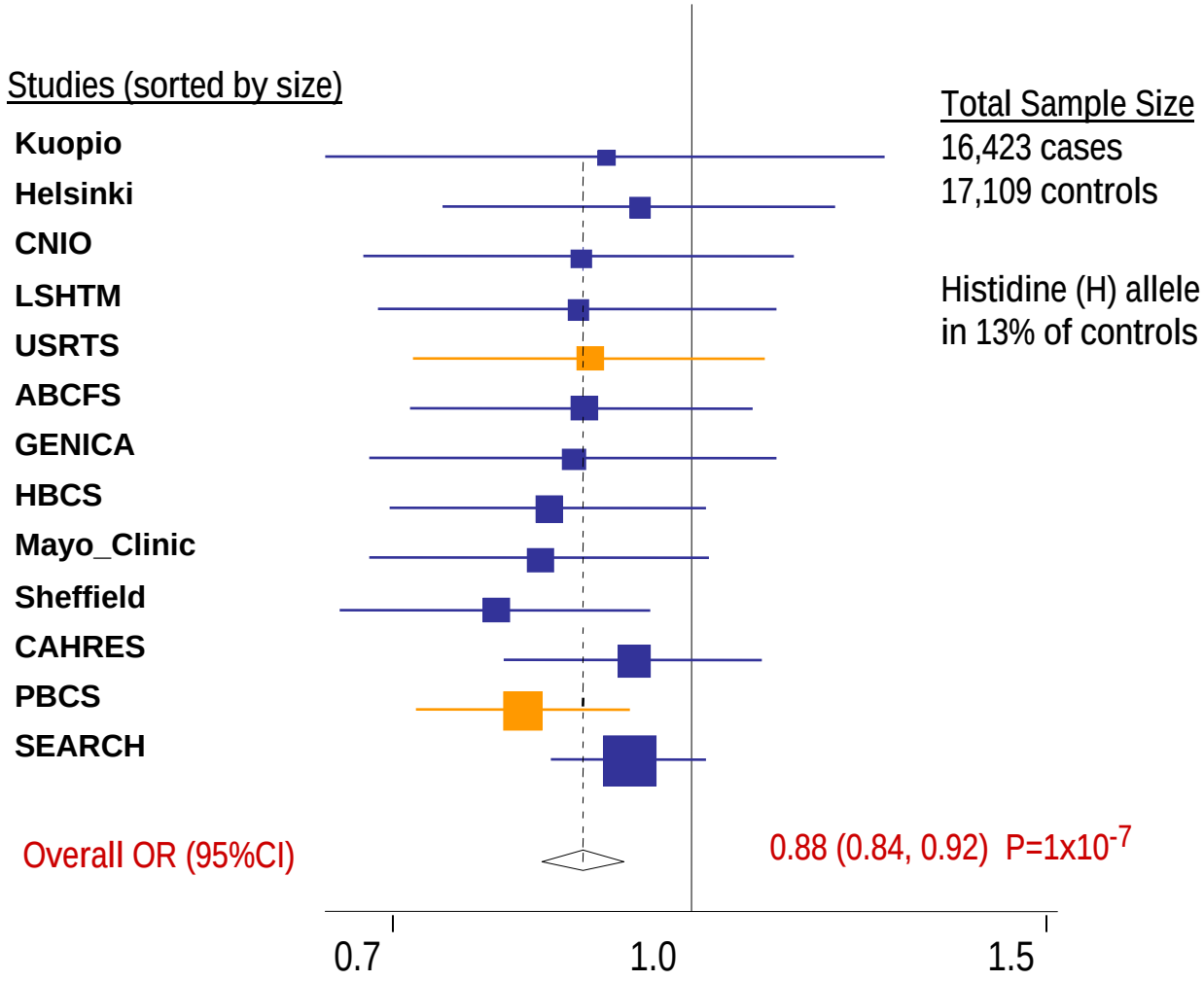


Breast Cancer Association Consortium: Findings to Date



National Cancer Institute

Caspase 8 *(CASP8)* D302H Variant Decreases Breast Cancer Risk



Cox A, Dunning A, García-Closas M for the BCAC (Under Review)*

* in alphabetical order

National Cancer Institute

Caspase 8 and Breast Cancer: Plausibility and Significance of Findings

- *CASP8* D302H is the first common variant with convincing evidence of an association with breast cancer.
- Caspase 8 is a critical initiator of death receptor mediated apoptosis.
- Follow-up:
  - Fine mapping to dissect genetic variants in *CASP8.*
  - Functional significance of variants.

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:

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 cases, 6,160 prostate cases

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

PanScan

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

Replication: Pancreatic Cancer Case-Control Consortium (PANC4)

Risk Factors: Tobacco, obesity, family history and diabetes

Genes: Genome-wide Association Study (GWAS)

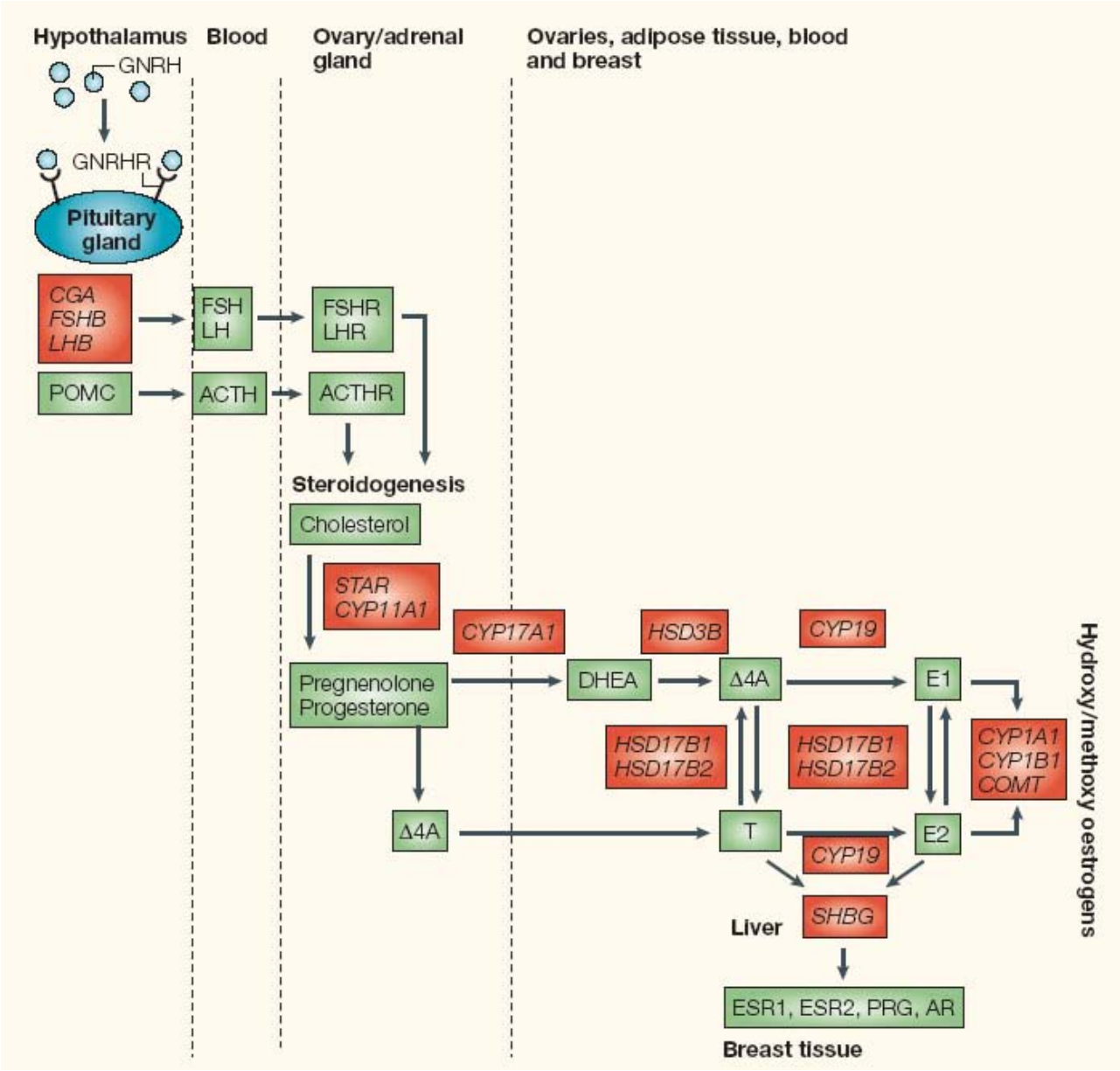
Cancer Sites: Pancreatic cancer

Cases with DNA: ~1,900 pancreatic cases

Website: http://epi.grants.cancer.gov/PanScan/

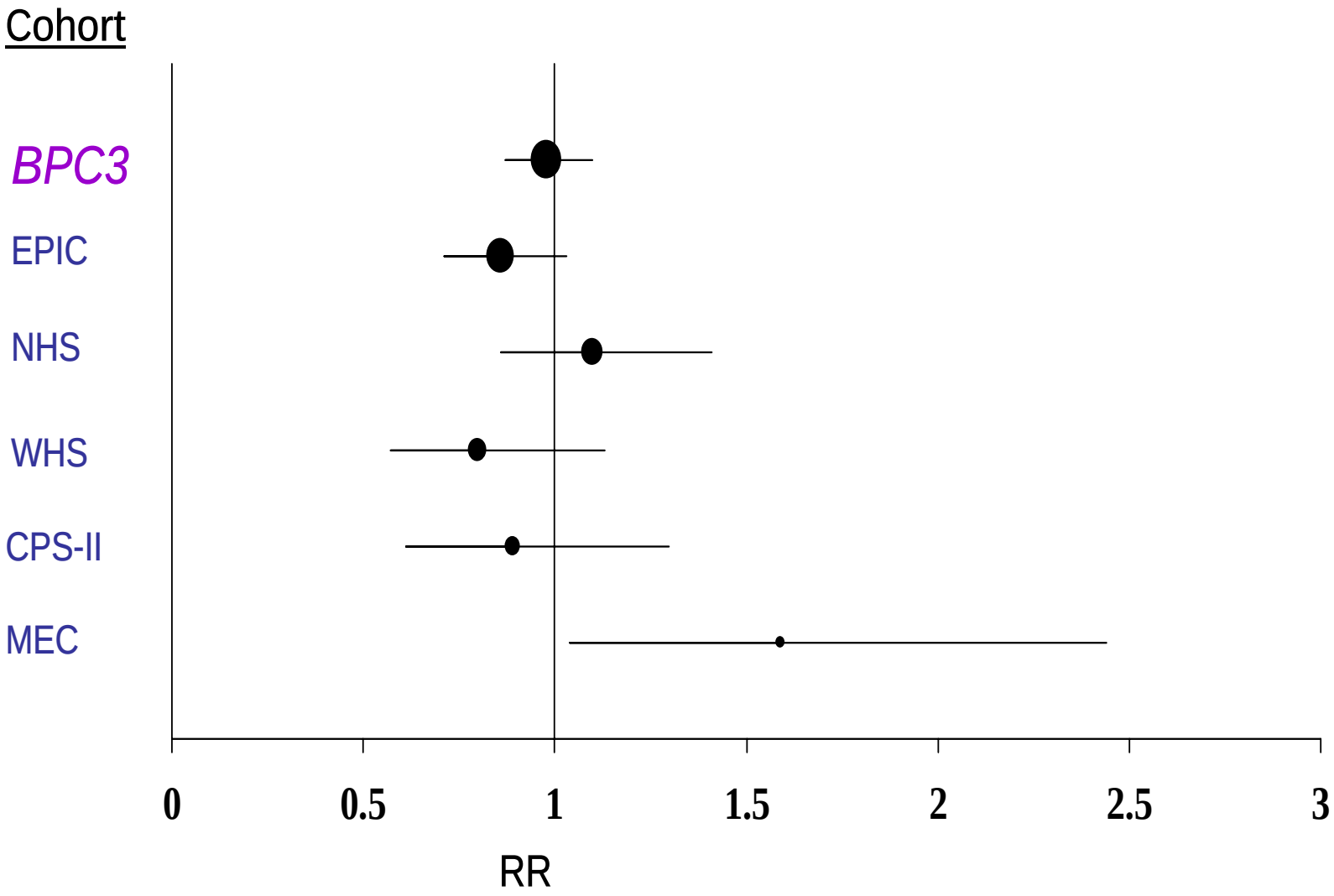
National Cancer Institute

BPC3 Methods



National Cancer Institute

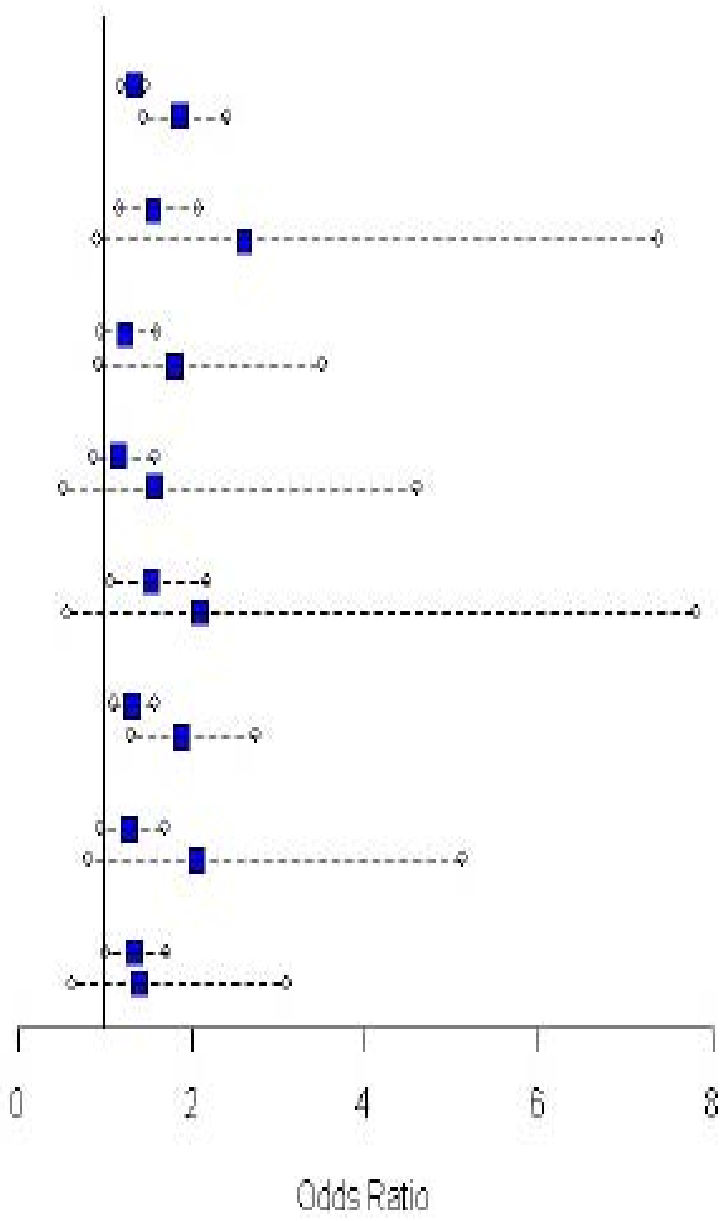
RR and 95% CI of Breast Cancer Associated With 312G Variant of HSD17B1



National Cancer Institute

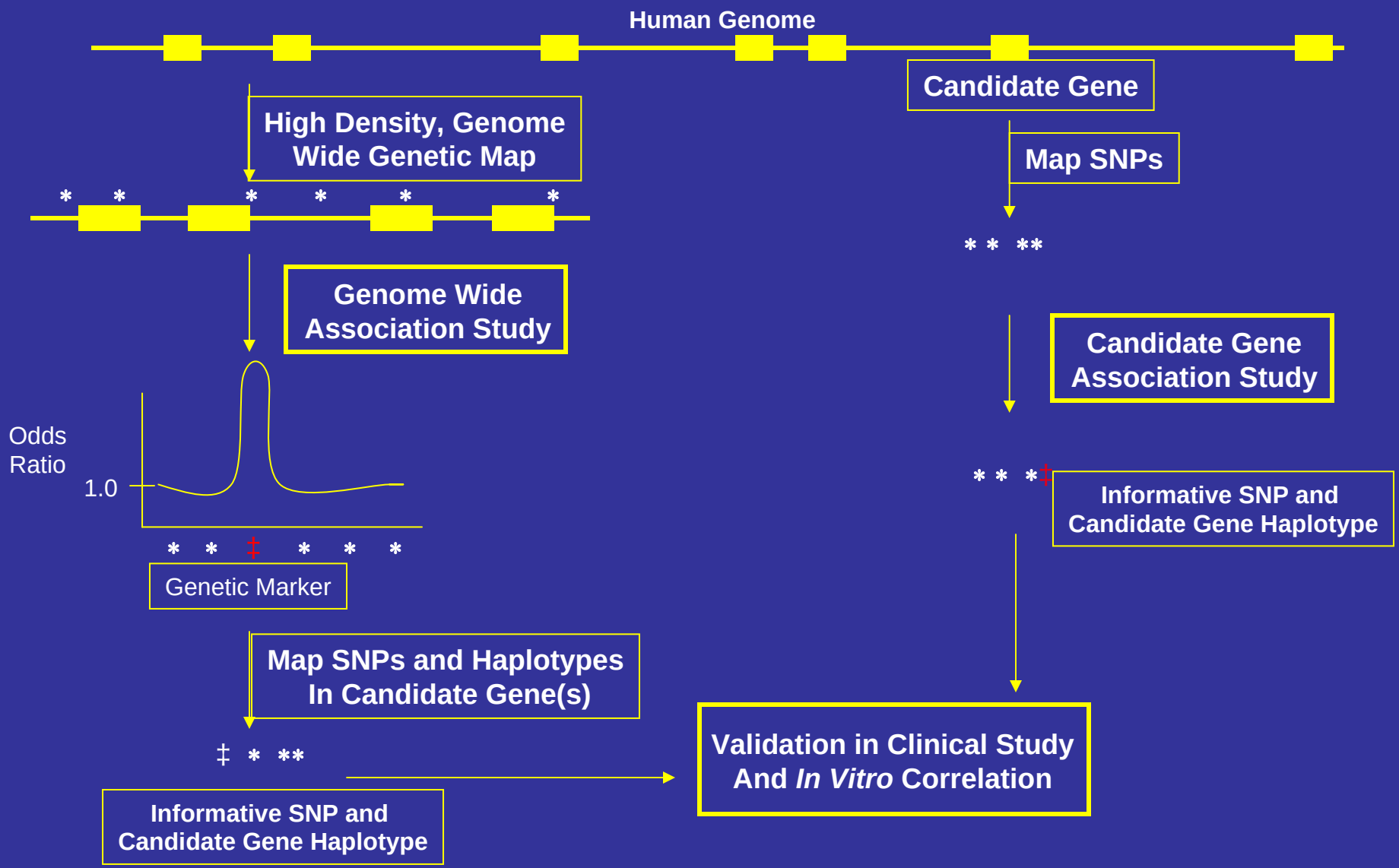
Results: Overall in BPC3

Cohort	Genotype	Cases / Controls	OR (99%CI)	P-value
All	CC	5,566 / 6,666	Ref.	4.00×10^{-19}
(p_{het}=0.483)	AC	2,064 / 1,842	1.33 (1.20-1.46)	
	AA	279 / 175	1.87 (1.44-2.42)	
ACS	CC	871 / 955	Ref.	2.63×10^{-5}
	AC	238 / 166	1.56 (1.17-2.08)	
	AA	21 / 9	2.61 (0.92-7.37)	
ATBC	CC	606 / 623	Ref.	0.012
	AC	312 / 260	1.23 (0.95-1.60)	
	AA	45 / 25	1.81 (0.94-3.51)	
EPIC	CC	551 / 869	Ref.	0.258
	AC	169 / 233	1.17 (0.87-1.58)	
	AA	12 / 12	1.57 (0.53-4.59)	
HPFS	CC	495 / 545	Ref.	3.63×10^{-3}
	AC	157 / 114	1.53 (1.07-2.19)	
	AA	11 / 6	2.09 (0.56-7.80)	
MEC	CC	1,426 / 1,565	Ref.	2.58×10^{-7}
	AC	728 / 614	1.32 (1.11-1.58)	
	AA	146 / 88	1.89 (1.30-2.75)	
PHS	CC	801 / 1,123	Ref.	0.013
	AC	200 / 220	1.27 (0.96-1.69)	
	AA	21 / 15	2.06 (0.83-5.12)	
PLCO	CC	816 / 986	Ref.	0.014
	AC	260 / 235	1.33 (1.02-1.72)	
	AA	23 / 20	1.39 (0.63-3.10)	



rs#1447295

Parallel Approaches To Identifying Genetic Determinants of Disease

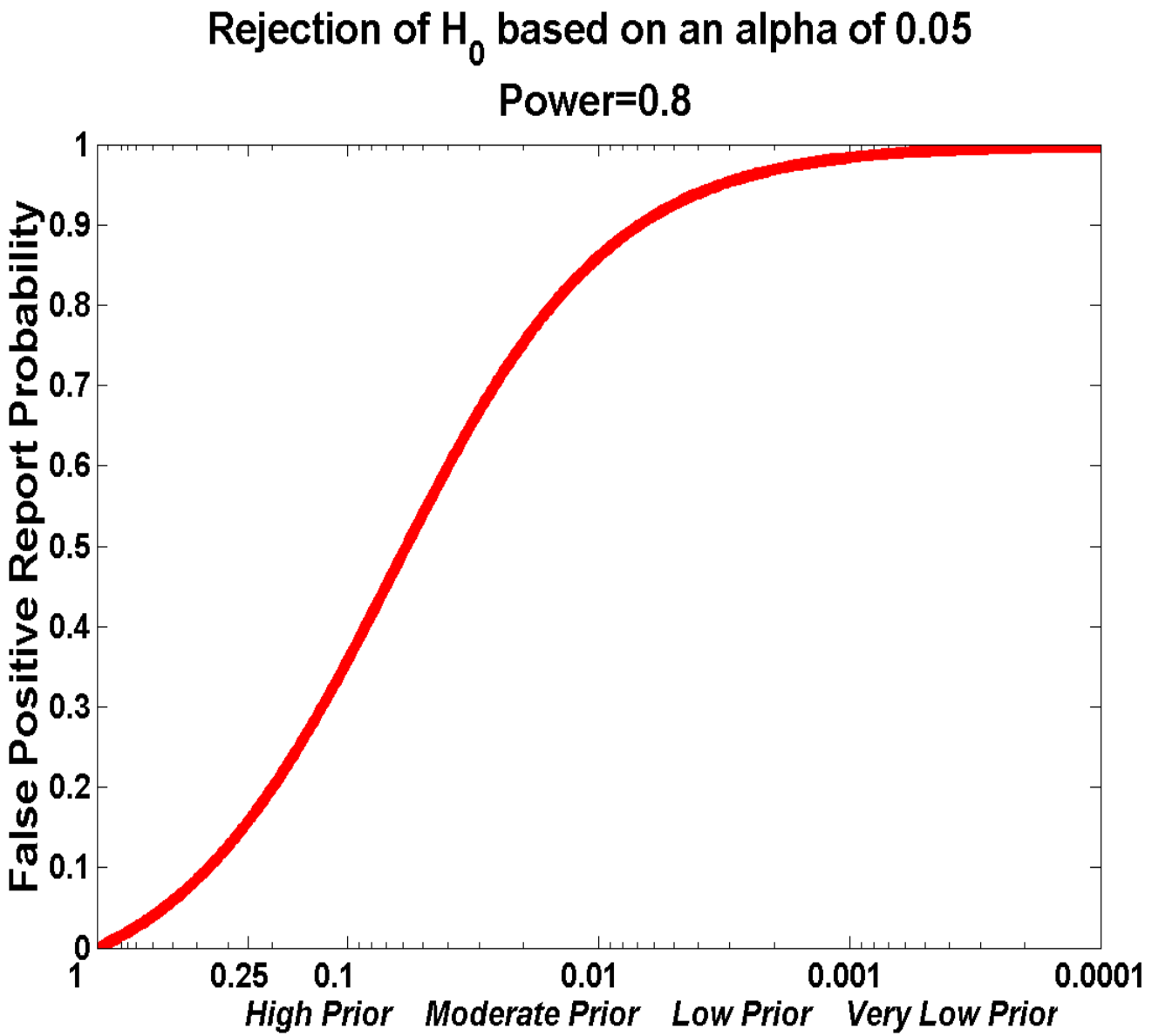


National Cancer Institute

Chance

- 24,000 Genes
- 3 Billion Base Pairs
- 8+ Million Common SNPs
- $Gene^{(n)}$-$Environment^{(n)}$

National Cancer Institute



National Cancer Institute

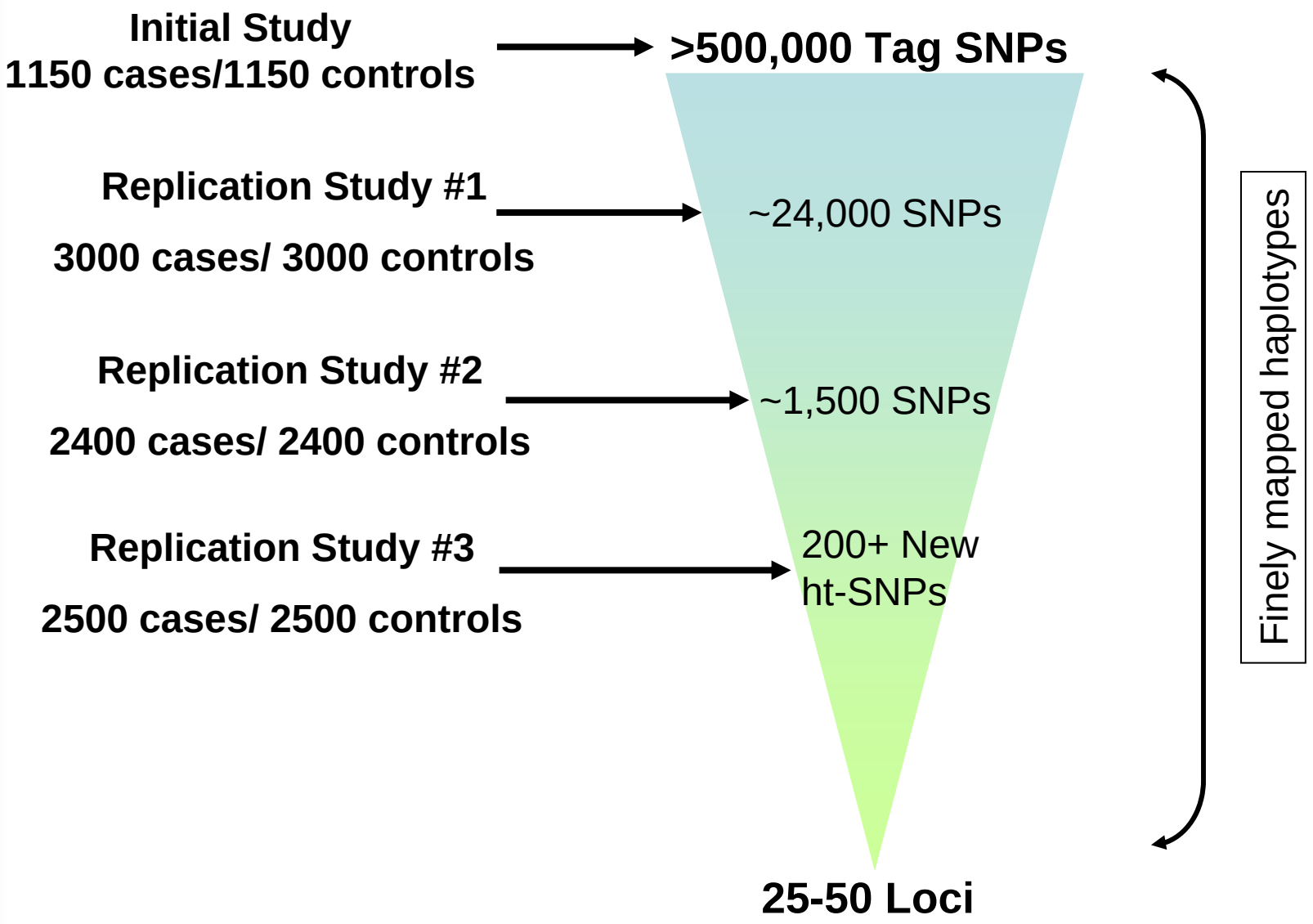
	# of cases	# of SNPs
Tier 1	443	198,000
Tier 2	332	1800

- We identified 11 SNPs that were associated with PD ($P < 10^{-1}$) in both tier 1 and tier 2 samples and had the same direction of effect." (Maraganore et al)

"In this issue, four investigative teams …have sought to replicate the findings from a GWA study of PD by Maraganore et al. Taken together these four studies appear to provide substantial evidence that none of the SNPs originally featured as potential PD loci are convincingly replicated and that all may be false positives."

National Cancer Institute

Replication Strategy for Prostate Cancer



http://cgems.cancer.gov

National Cancer Institute

Gene-Environment Studies: WHAT WORKS

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

National Cancer Institute

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

National Cancer Institute

National Cancer Institute

Finding High-Penetrance Genes: Indices of Success (continued)

Palladin Gene and Pancreatic Cancer (example)

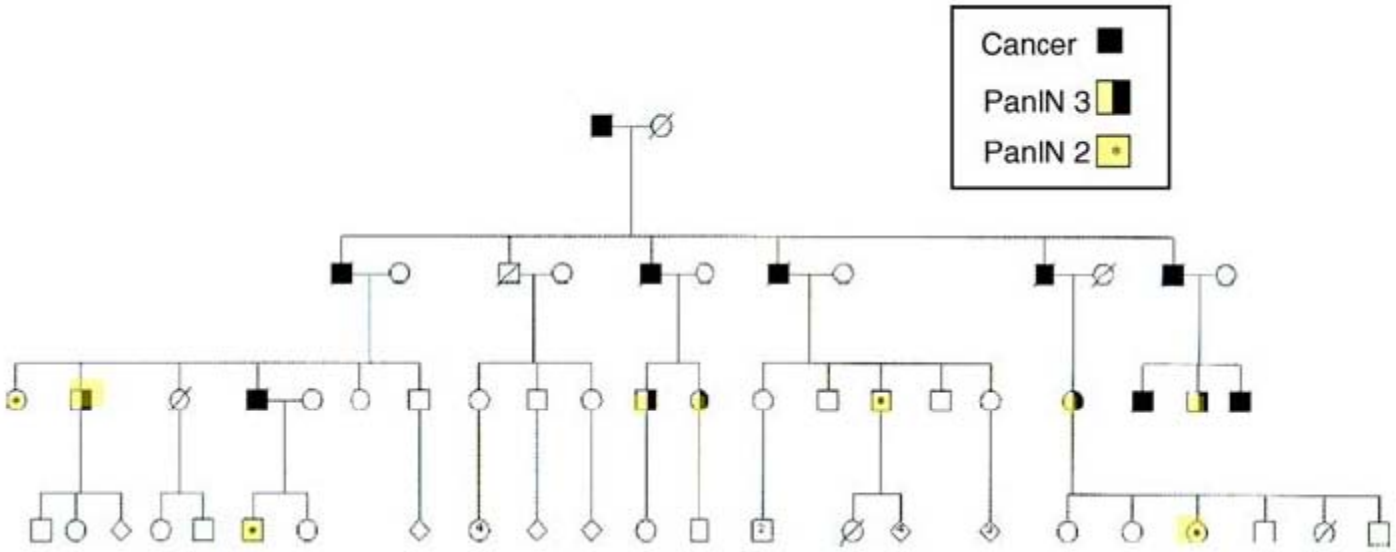


Figure: Nine members of this family were diagnosed with pancreatic cancer and nine with pancreatic precancer (five with carcinoma in situ [PanIN 3] and four with low-grade dysplasia [PanIN 2]

National Cancer Institute

Genetic Susceptibility

- Heritability e.g., twin studies
- Family History Risk Factor
- Striking Familial Aggregations
- Biologic Pathway Speculation
- Agnostic Gene-Hunting (GWAS)