National Cancer Institute

Genome Wide Association and Rare Cancers

Robert N. Hoover, M.D., Sc.D.
Director
Epidemiology and Biostatistics Program
Division of Cancer Epidemiology and Genetics
National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

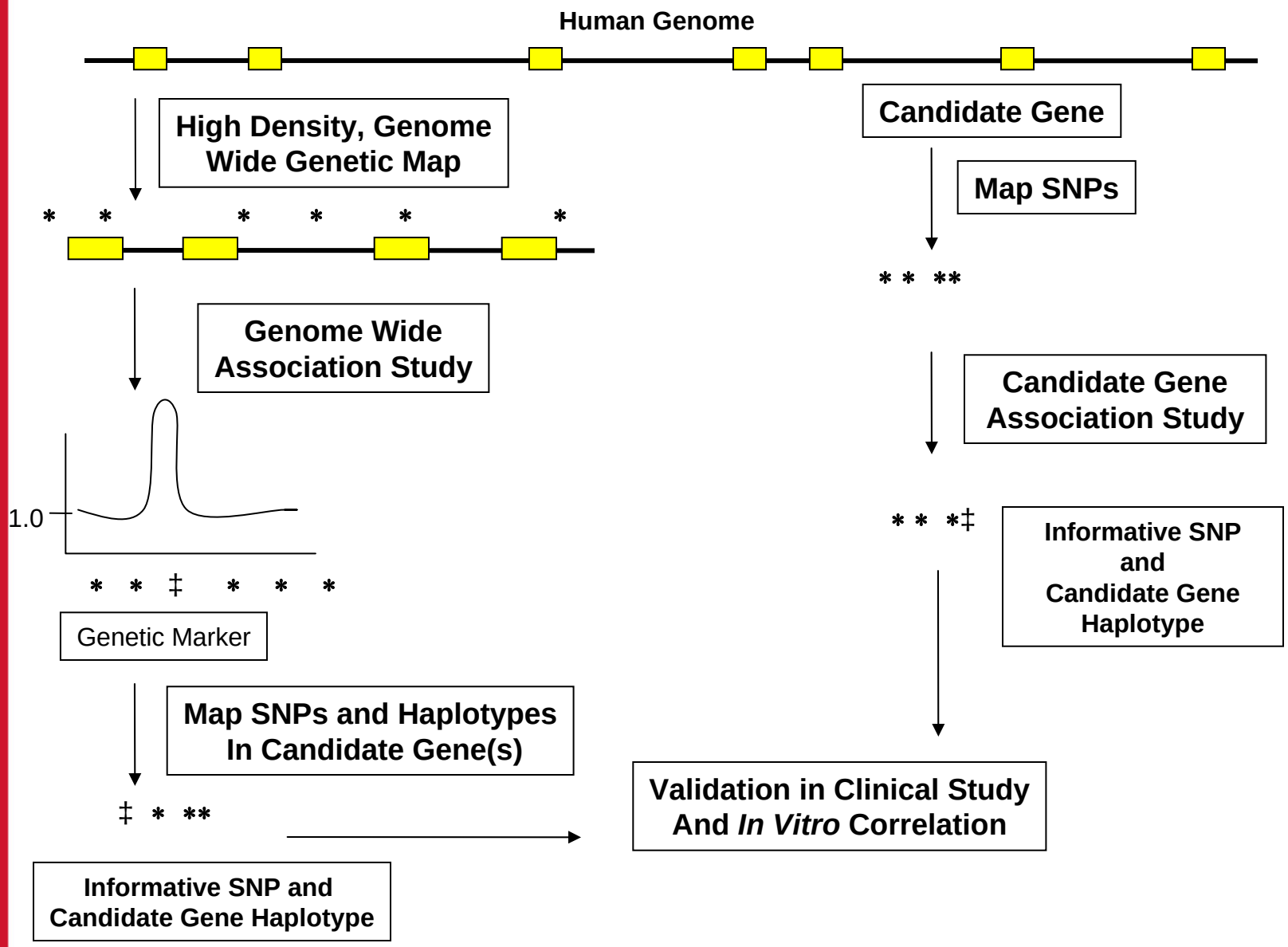
National Cancer Institute

Consortial Approaches to Rare Cancers

- Familial cancer registry consortia
- Case-control consortia
- Cohort Consortia
- Genome-wide association studies
- Biomarker studies (PCB, SC)
- Questionnaire studies (e.g. thyroid)

National Cancer Institute

Parallel Approaches To Identifying Genetic Determinants of Disease



National Cancer Institute

State of Candidate Gene Association Studies in Cancer

Scope:

- PubMed Search ~300 genes out of 24,000 for cancer
- 2 or more references
- Less than 150 have ‘functional’ data
- Shift towards pathways & genes

National Cancer Institute

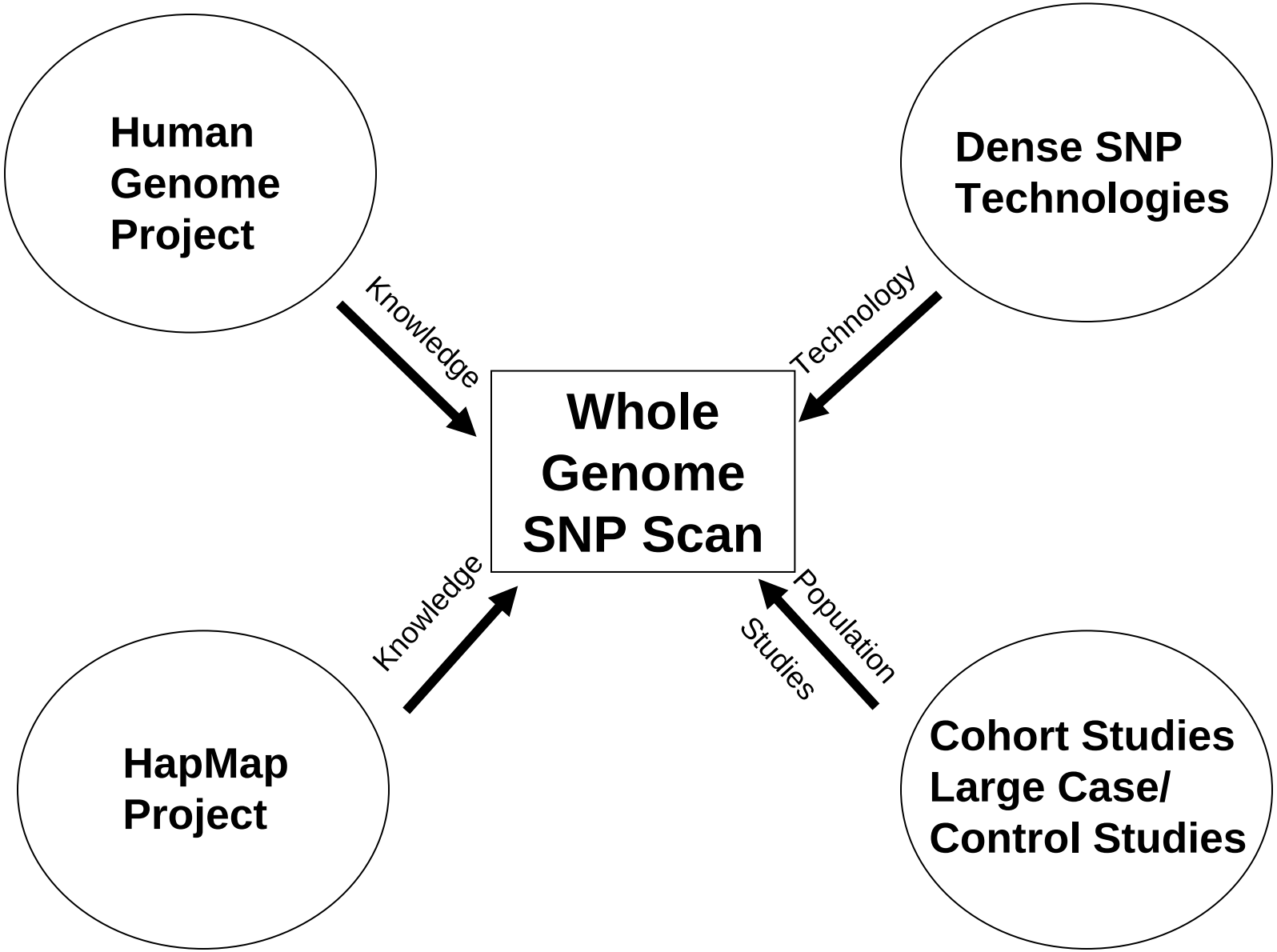
State of Candidate Gene Associations

- False Positives:
  - 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

Confluence of Opportunities to Identify Inherited Cancer Risk



National Cancer Institute

General Strategy for Prostate & Breast Cancer GWAS

Initial Study
1150 cases/1150 controls → **540,000 Tag SNPs**

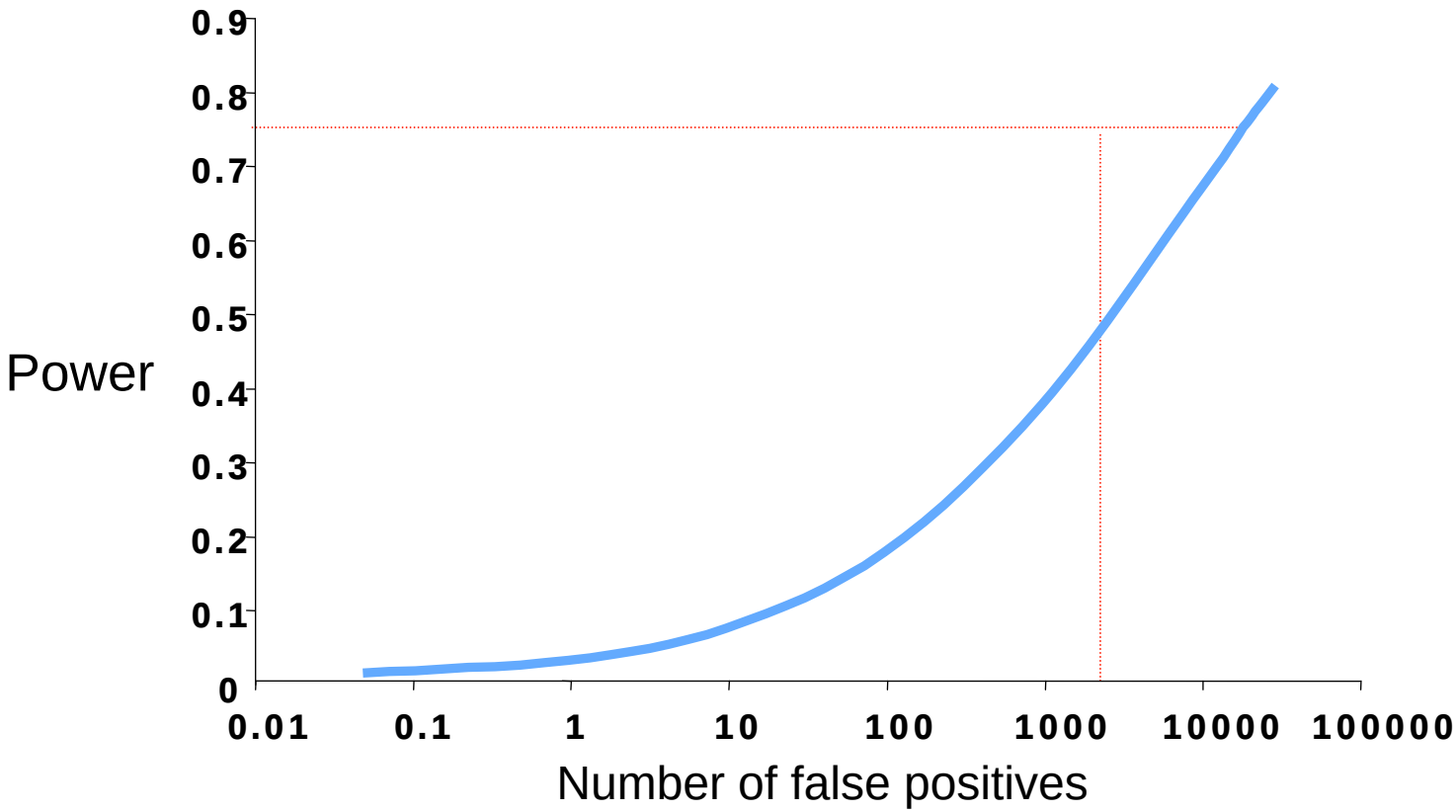
Follow-up Study #1
4500 cases/ 4500 controls → **~28,000 SNPs**

Follow-up Study #2
3500 cases/ 3500 controls → ***at least* 1,500 SNPs**

Fine Mapping → **30 ±20 loci**

National Cancer Institute

Power of genome wide screen as a function of the number of retained false positive

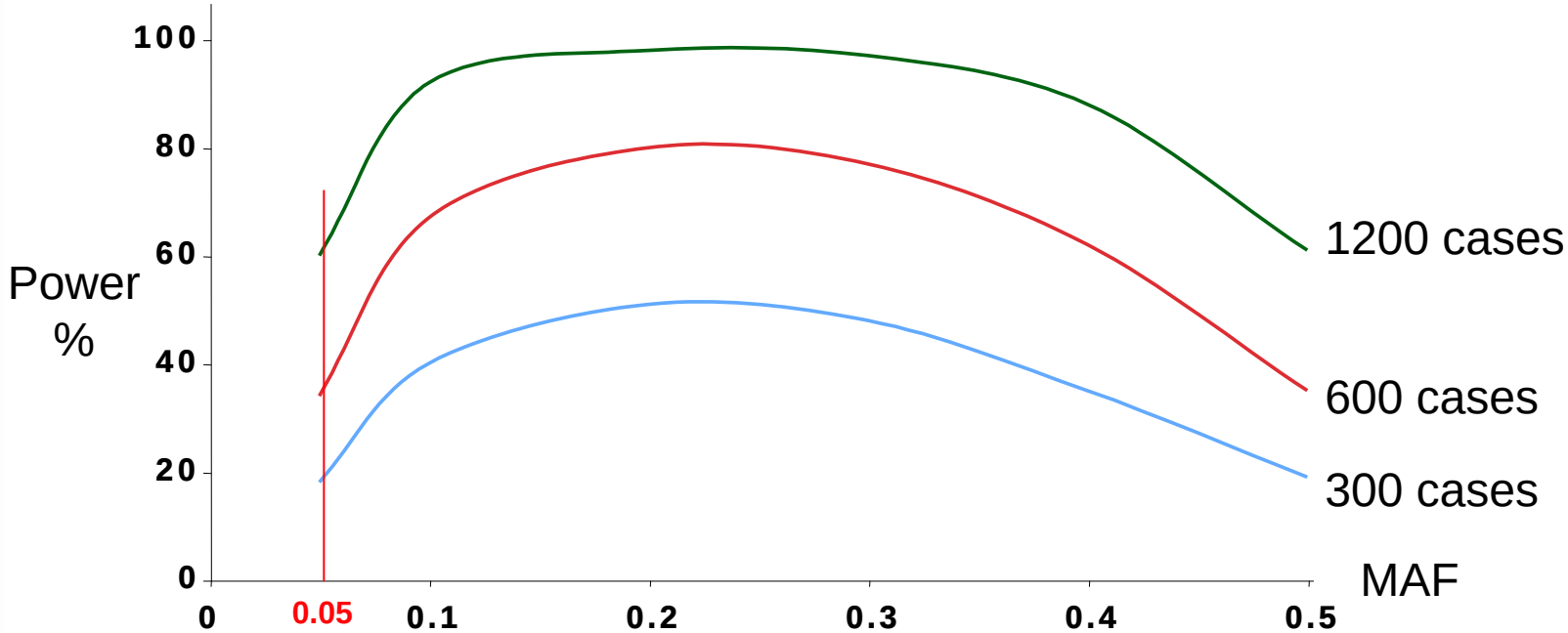


Model :
One susceptibility allele : MAF = 0.1 , Odds Ratio = 1.4
LD of typed marker with susceptibility marke : $r^2 = 0.8$
Number of cases/control pairs : 1,200
Number of markers types : 500,000

National Cancer Institute

2-Stage WGS Strategy

Power as a function of MAF and sample sizes typed in the first stage



Disease model
- Prevalence 1%
- Single susceptibility SNP with a linkage disequilibrium $r^2 = 0.8$ with 1 genotyped SNP
- Dominant transmission
- Genotype relative risk : 1.5

Study design
Cases = # Controls
Cases in stage 1 : as indicated
SNPs in stage 1 : 500,000
Cases in stage 2 : 2,000
SNPs in stage 2 : 25,000
Significance level 0.00002

Note: Significance level = 0.00002 => 10 false positives

National Cancer Institute

GWAS: What is Working

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

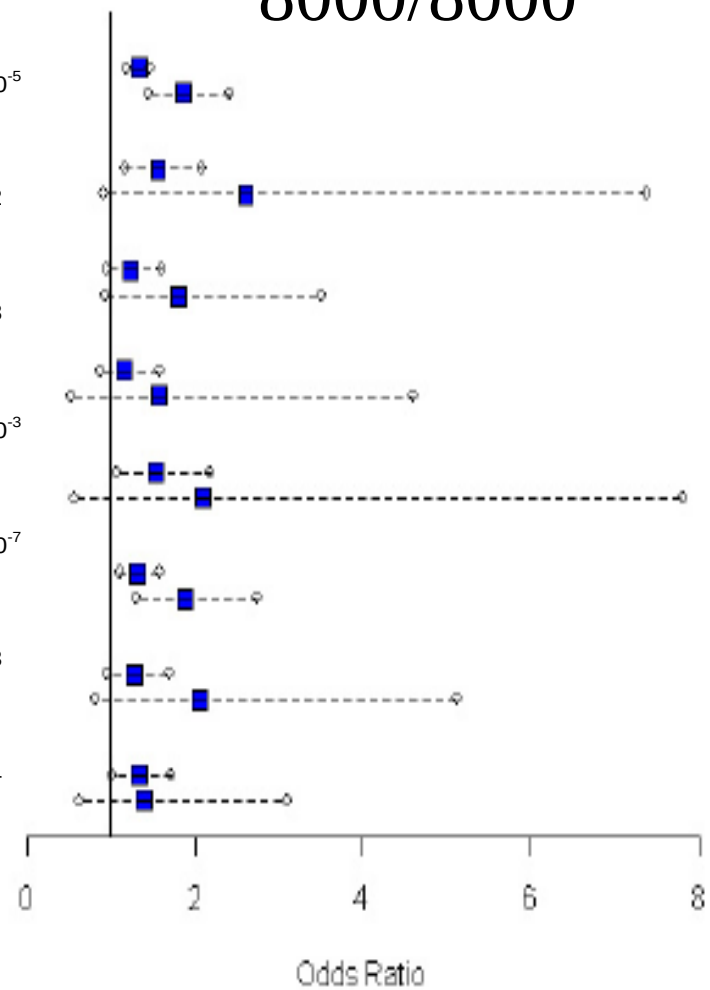
National Cancer Institute

Results: Overall

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)	

BPC3

8000/8000



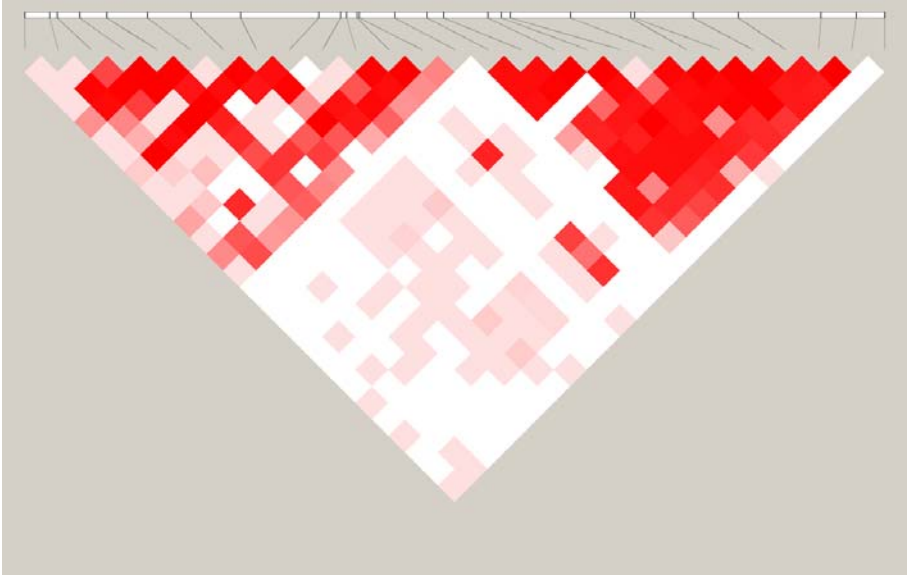
Schumacher FR et al., Cancer Res. 2007 Apr 1;67(7):2951-6.

National Cancer Institute

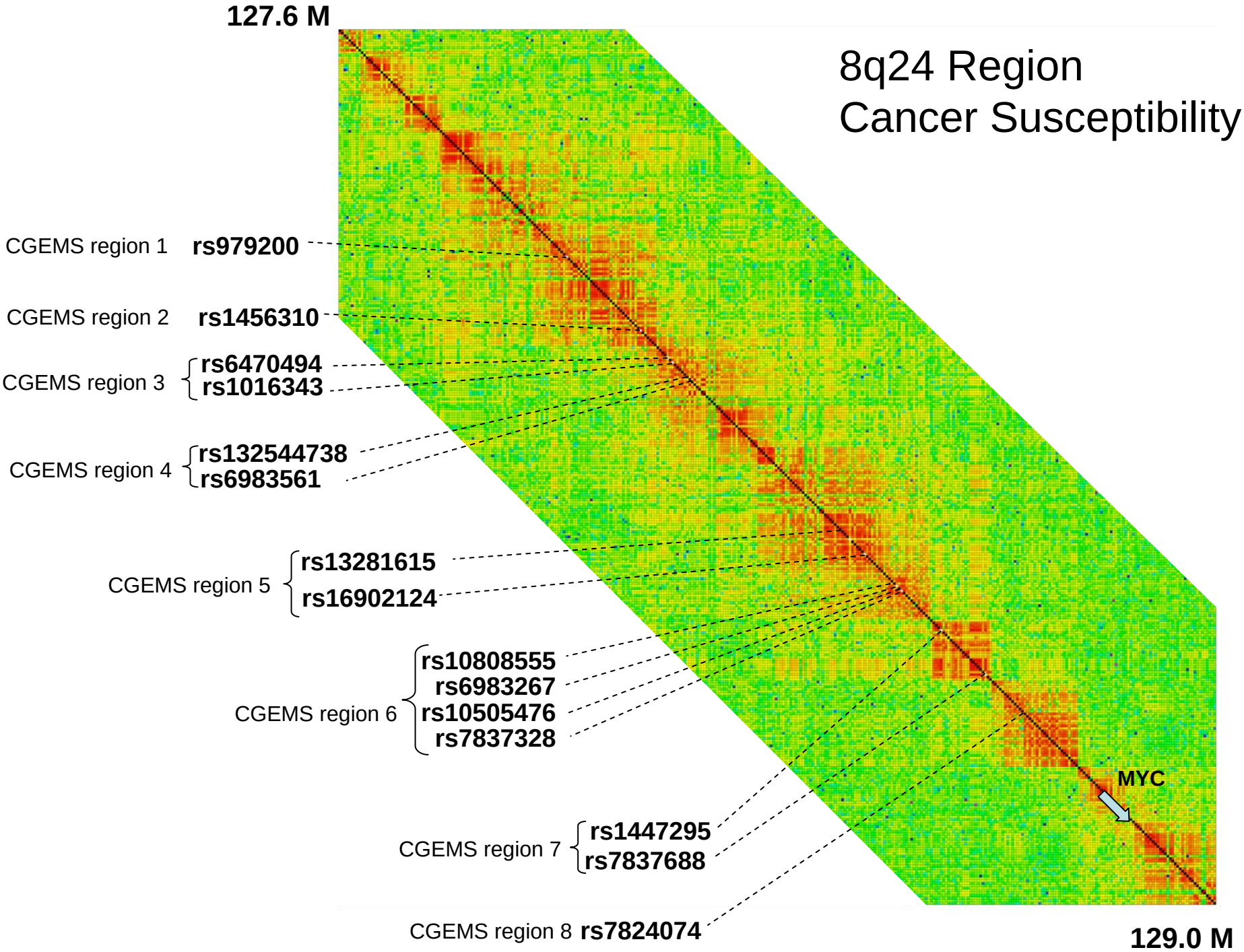
Population Attributable Risk of Prostate Cancer and 8q24 SNPs

	Joint PAR	PAR rs1447295	PAR rs6983267
ALL	0.284	0.085	0.209
ACS	0.255	0.094	0.192
ATBC	0.251	0.052	0.157
FPCC	0.306	0.096	0.091
HPFS	0.249	0.085	0.180
PLCO	0.347	0.086	0.276

rs6983267 G: 21% **rs1447295 A: 7%**



- Suggests that both SNPs contribute substantially to the population burden of prostate cancer.


127.6 M
8q24 Region
Cancer Susceptibility
CGEMS region 1
rs979200
CGEMS region 2
rs1456310
CGEMS region 3
rs6470494
rs1016343
CGEMS region 4
rs132544738
rs6983561
CGEMS region 5
rs13281615
rs16902124
CGEMS region 6
rs10808555
rs6983267
rs10505476
rs7837328
MYC
CGEMS region 7
rs1447295
rs7837688
CGEMS region 8
rs7824074
129.0 M

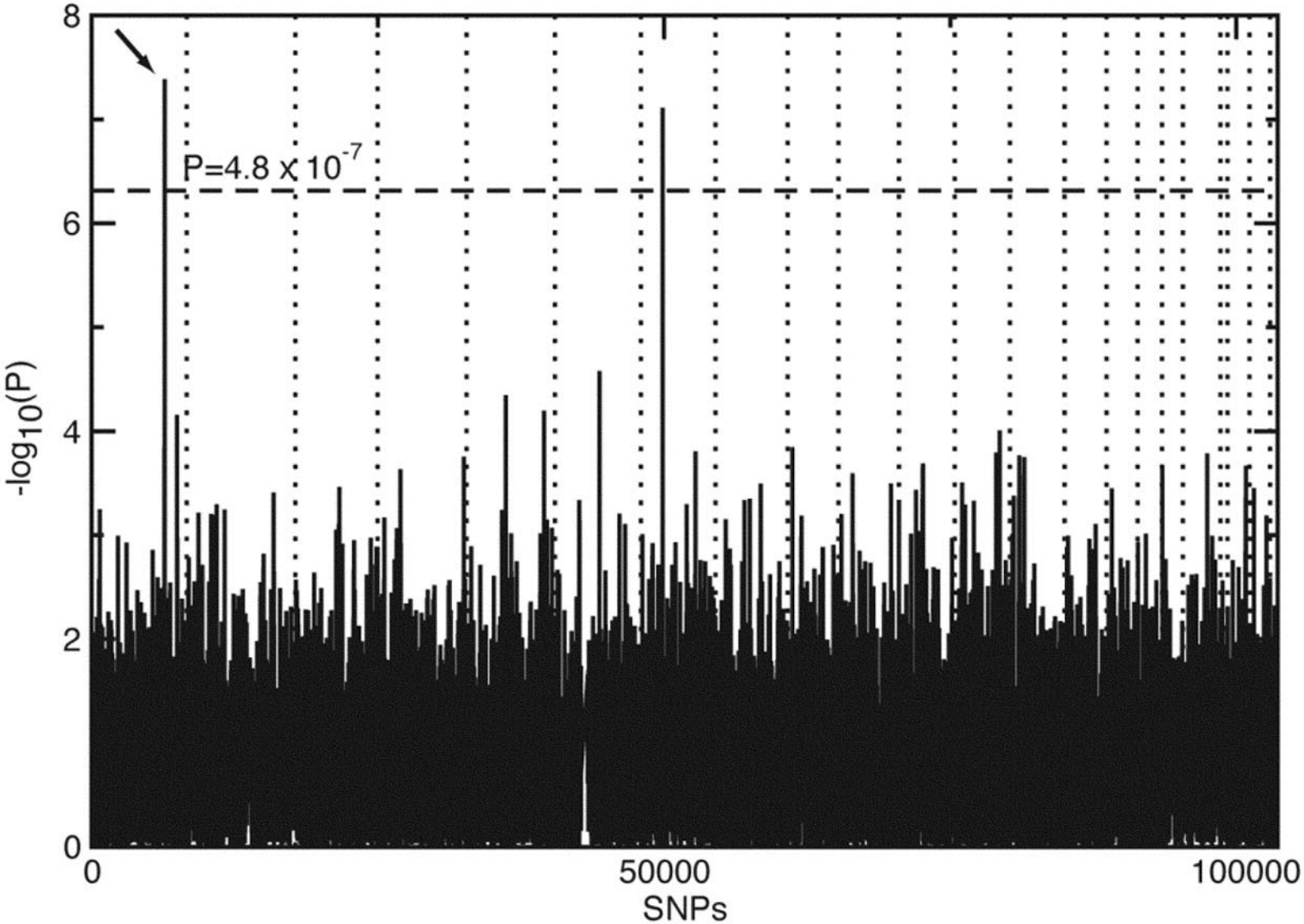
National Cancer Institute

Rare Tumors Present Challenges

- Numbers
  - Initial vs. subsequent stages of scan
  - Replication studies
- Quality (Epi)
- Examples:
  - ARMD
  - Cambridge breast cancer
  - PanScan
- Strategies for what to relax and in what order is complicated

National Cancer Institute

Complement Factor H Gene and Macular Degeneration



Science. 2005 April 15; 308:385

National Cancer Institute

Challenges and Lessons from Consortial Studies

- Data Sharing
  - Internal
  - External
- Authorship
- Junior Investigators
- Informatics
- Decision-making
- (Big Science vs. Not)