National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Genome-Wide Association Studies in Cancer:

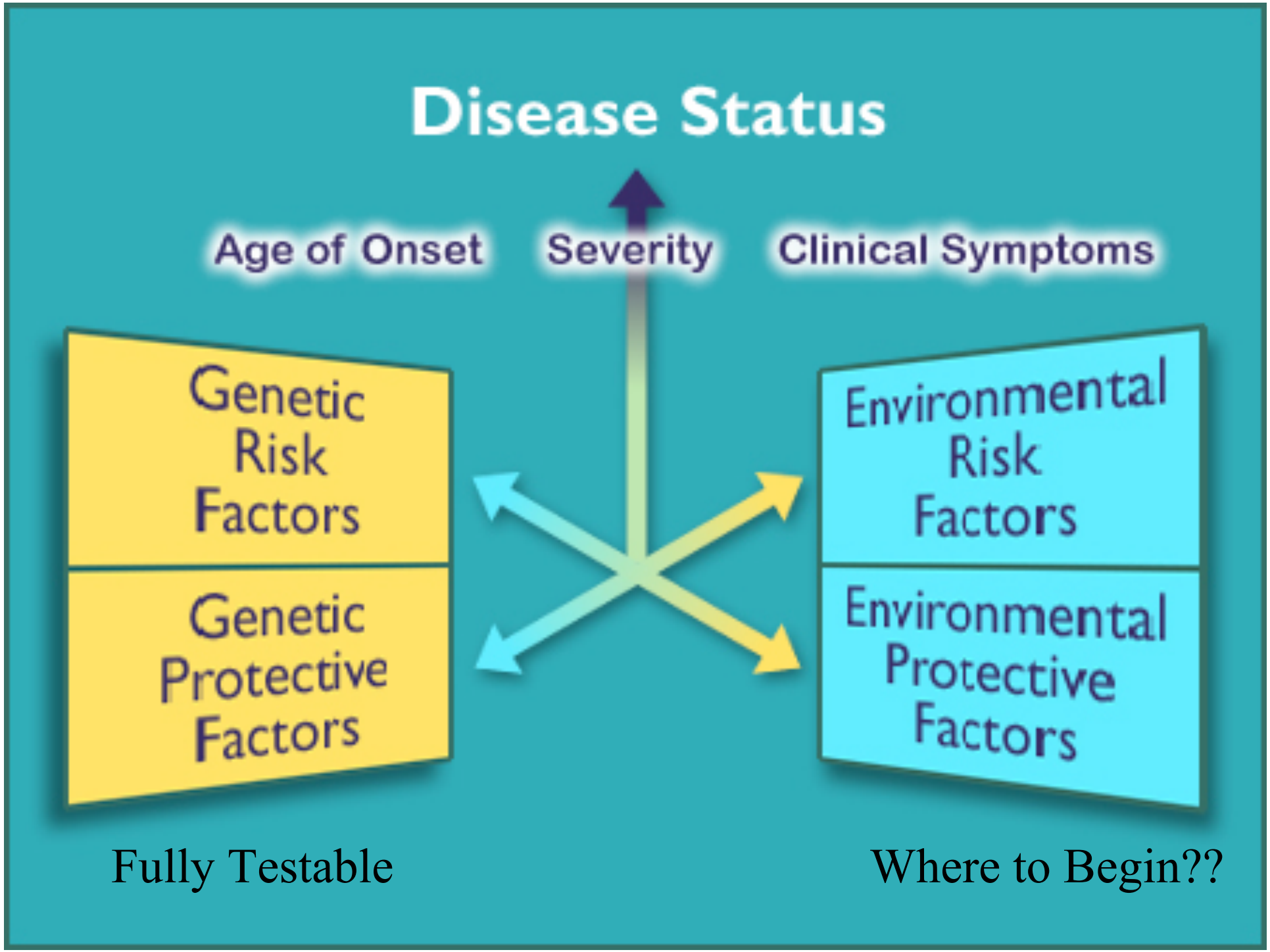
Stephen Chanock, M.D.

Chief of Laboratory of Translational Genomics, DCEG, NCI

Director, Core Genotyping Facility, DCEG, NCI

June 4, 2008

"What ever will we think about now that the genome project is almost complete?"


Disease Status
Age of Onset
Severity
Clinical Symptoms
Genetic Risk Factors
Genetic Protective Factors
Environmental Risk Factors
Environmental Protective Factors
Fully Testable
Where to Begin??

Promise of GWAS

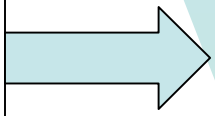
- Discovery of Common Markers in the Genome
  - 'Represents a portion of the genetic contribution'
- Opportunity to explore mechanism of biology
  - How and why cancer develops
- Outcomes
  - Etiology
  - Gene-Environment/Lifestyle Interactions
  - "Druggable" targets
- Establish genetic markers for:
  - Prevention
  - Intervention

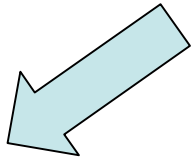
Identifying Genetic Markers for Prostate & Breast Cancer

CGEMS
Cancer Genetic Markers of Susceptibility

Genome-Wide Analysis
Public Health Problem
- **Prostate (1 in 8 Men)**
- **Breast (1 in 9 Women)**

Analyze Long-Term Studies
- **NCI PLCO Study**
- **Nurses' Health Study**





Initial Study

Follow-up #1

Follow-up #2

Establish Loci

Fine Mapping
Functional Studies
Validate Plausible Variants
Possible Clinical Testing

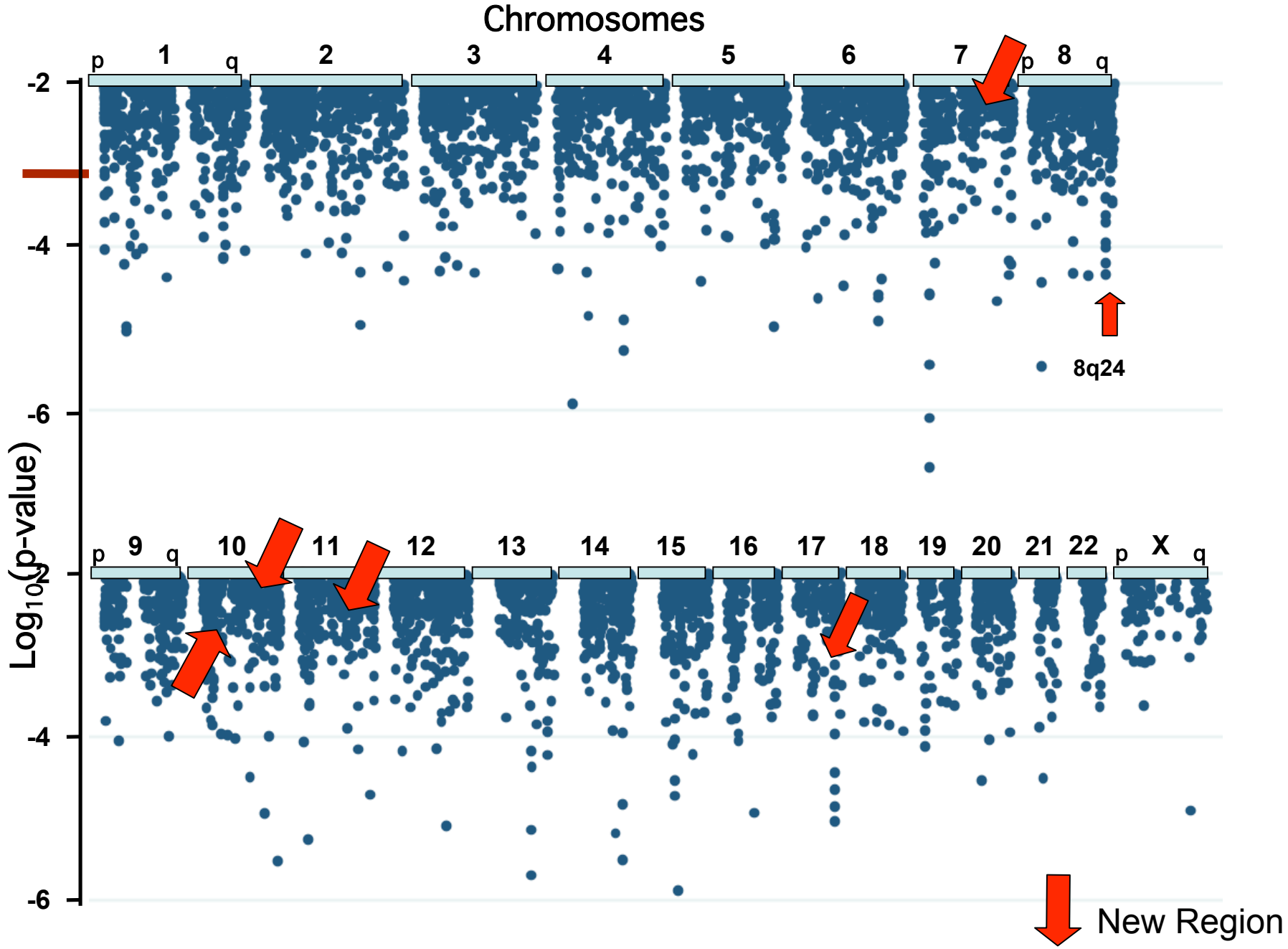
http://cgems.cancer.gov

Prostate Cancer Risk Circa..2006

The Enigma of a Common Disease

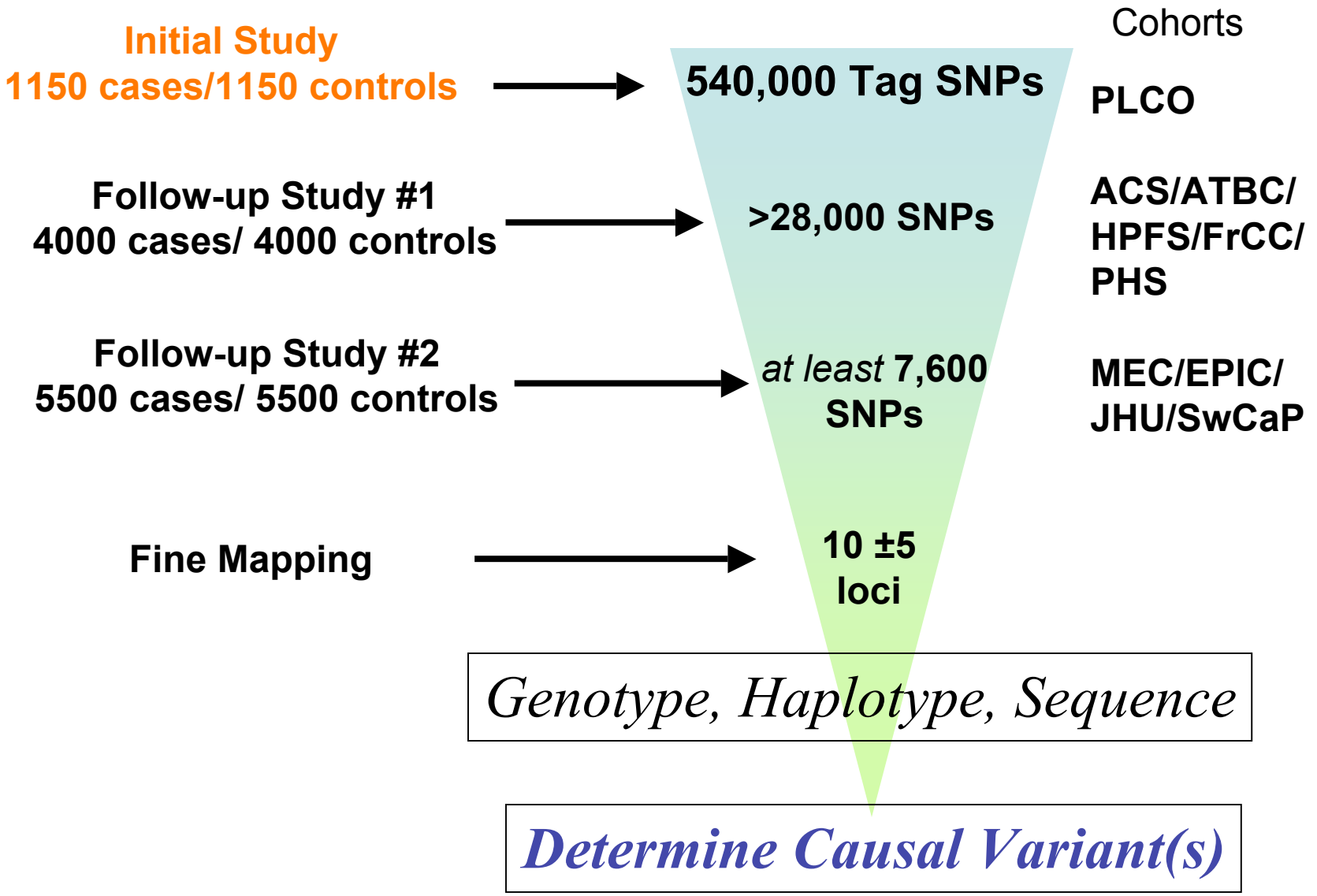
- Age
- Ethnic Background
- Family History
- One SNP- unknown function
  - Rs1447295 @ region 8q24 (no obvious gene)

CGEMS Prostate Cancer GWAS: Where are the True Signals Amidst the Blizzard of False Positives



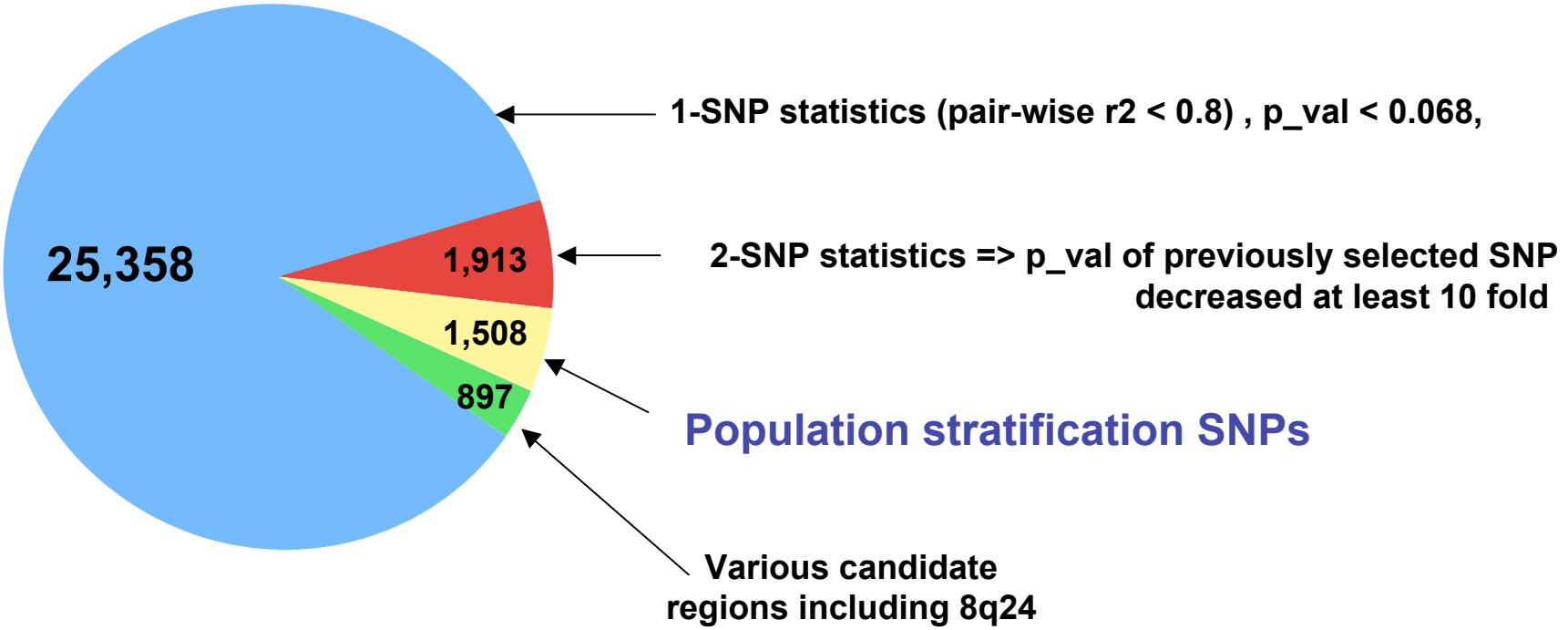
incidence density sampling

General Strategy for Prostate GWAS is Based on Replication, Replication, Replication



CGEMS prostate cancer stage 2

Selection of the SNPs to be taken to stage 2
Determining Real-estate to find the FEW true positives



SNPs distributed in 7608 distinct chromosomal regions

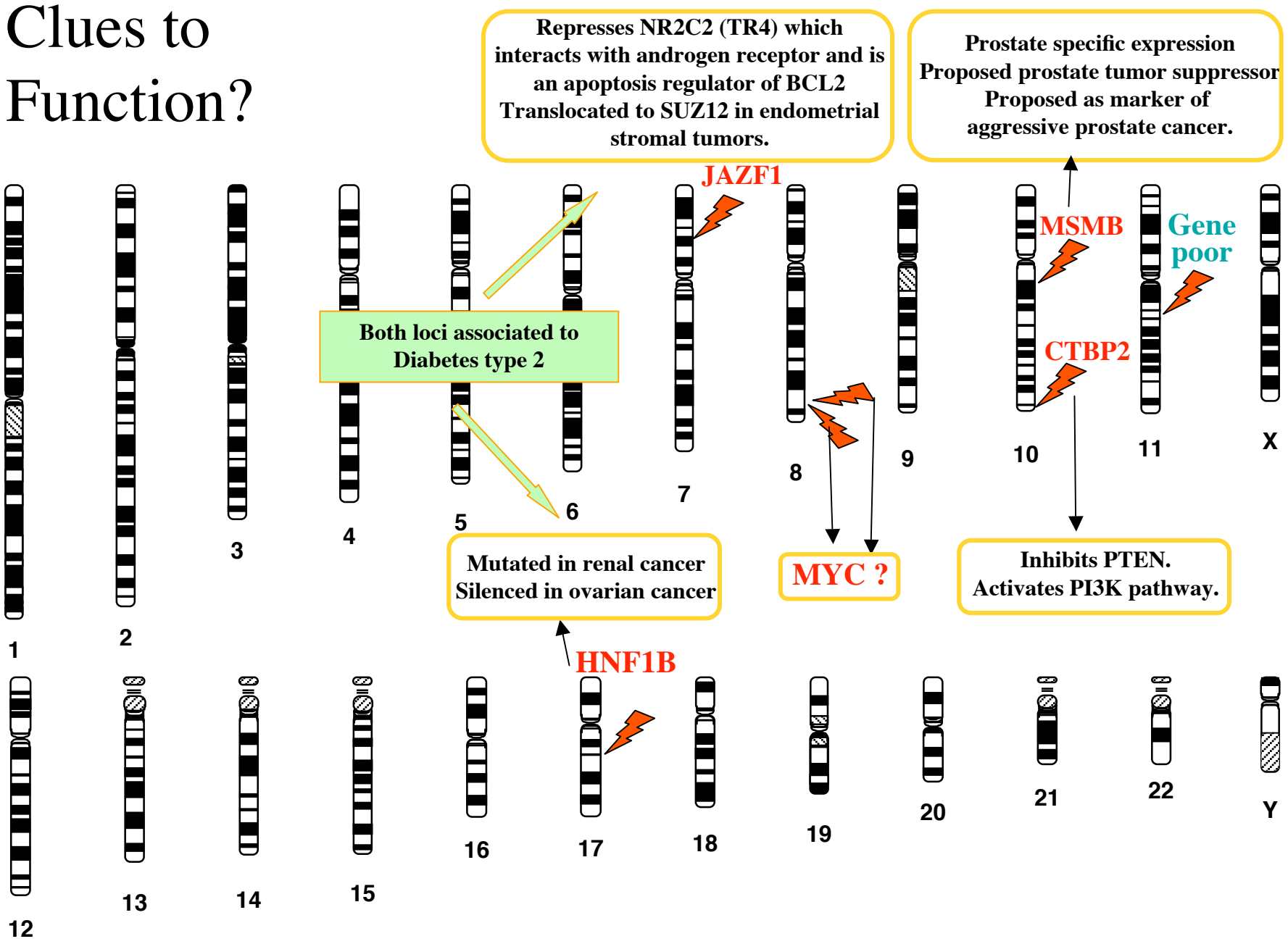
In a region the maximal distance between two adjacent SNPs is less than 100Kb

7 associated loci in CGEMS Prostate Cancer

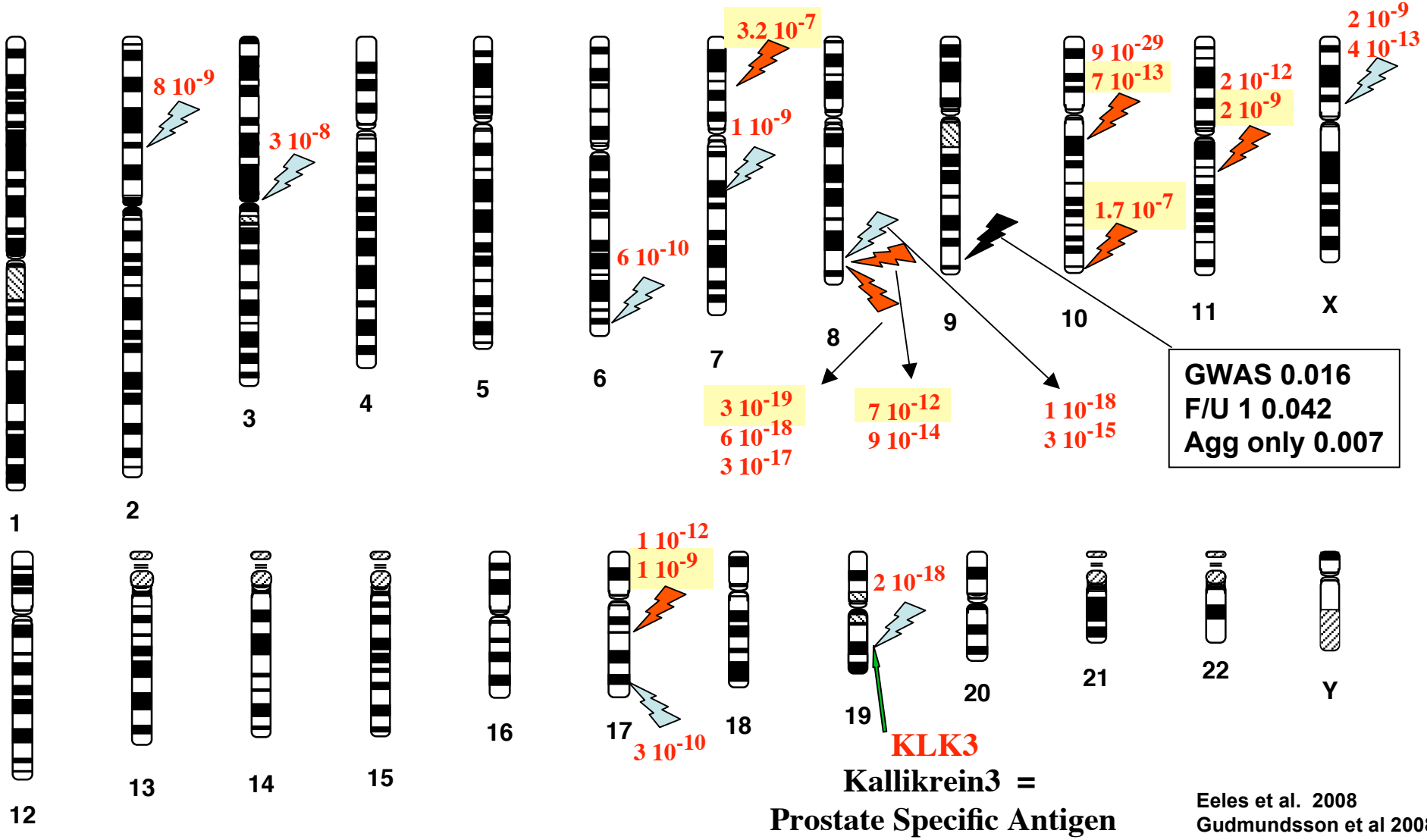
Region	p-value	Risk Allele Freq.	Odds ratios	
			Heterozygotes	Homozygotes
8q24 (loc1)	$6.7\ 10^{-16}$	0.1	1.49 (1.34-1.64)	1.83 (1.32-2.53)
10q11	$8.7\ 10^{-14}$	0.38	1.20 (1.10-1.31)	1.61 (1.42-1.81)
8q24 (loc2)	$4.7\ 10^{-13}$	0.50	1.13 (1.02-1.26)	1.46 (1.30-1.64)
17q12	$1.5\ 10^{-10}$	0.52	1.25 (1.13-1.34)	1.47 (1.31-1.65)
11q13	$4.1\ 10^{-10}$	0.50	1.18 (1.08-1.28)	1.48 (1.27-1.74)
10q26	$1.7\ 10^{-7}$	0.25	1.14 (0.94-1.38)	1.40 (1.16-1.69)
7p15	$3.2\ 10^{-7}$	0.76	1.18 (1.07-1.31)	1.54 (1.37-1.73)

Associated loci in CGEMS prostate stage 2

Clues to Function?



16^{+} published loci involved in prostate cancer susceptibility with significance $p < 5 \times 10^{-7}$



Eeles et al. 2008
Gudmundsson et al 2008
Haiman et al 2007

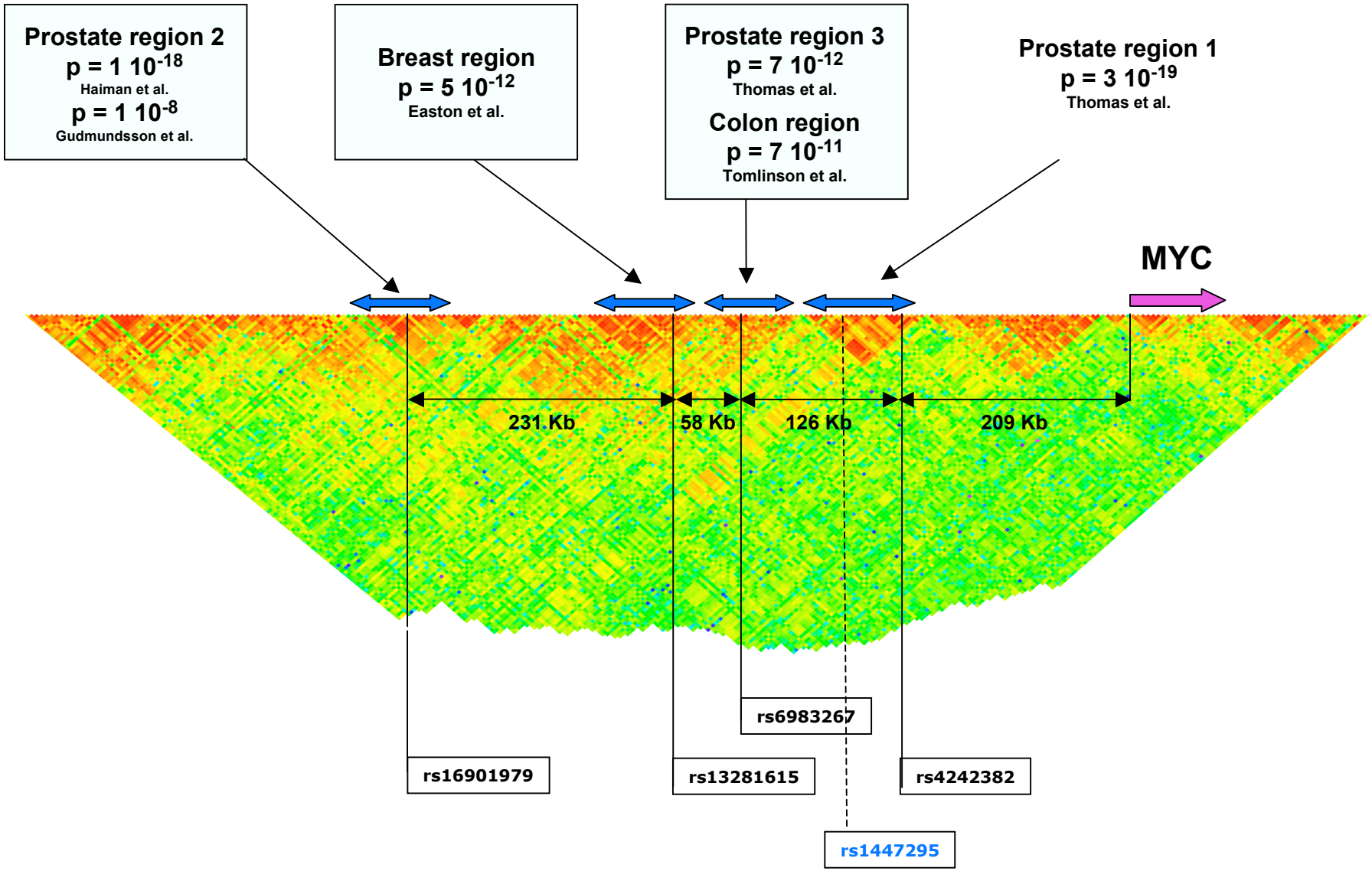
Additional variants – March 2008

	CGEMS	CRUK	deCODE
8q24*	X	X	X
HNF1B (17q12)	X	X	X
MSMB (10q11)	**X**	**X**	
17q24		X	X
NUDT10/11 (Xp11)		**X**	**X**
JAZF1 (7p15)	**X**		
CTBP2 (10q26)	**X**		
11q13	**X**		
CPNE3 (8q21)	**X**		
IL16 (15q25)	**X**		
CDH13 (16q23)	**X**		
SLC22A3 (6q25)		**X**	
3p12		**X**	
LMTK2 (7q21)		**X**	
KLK2,3 (19q13)		**X**	
2p15			**X**

Prostate Cancer Risk 2008

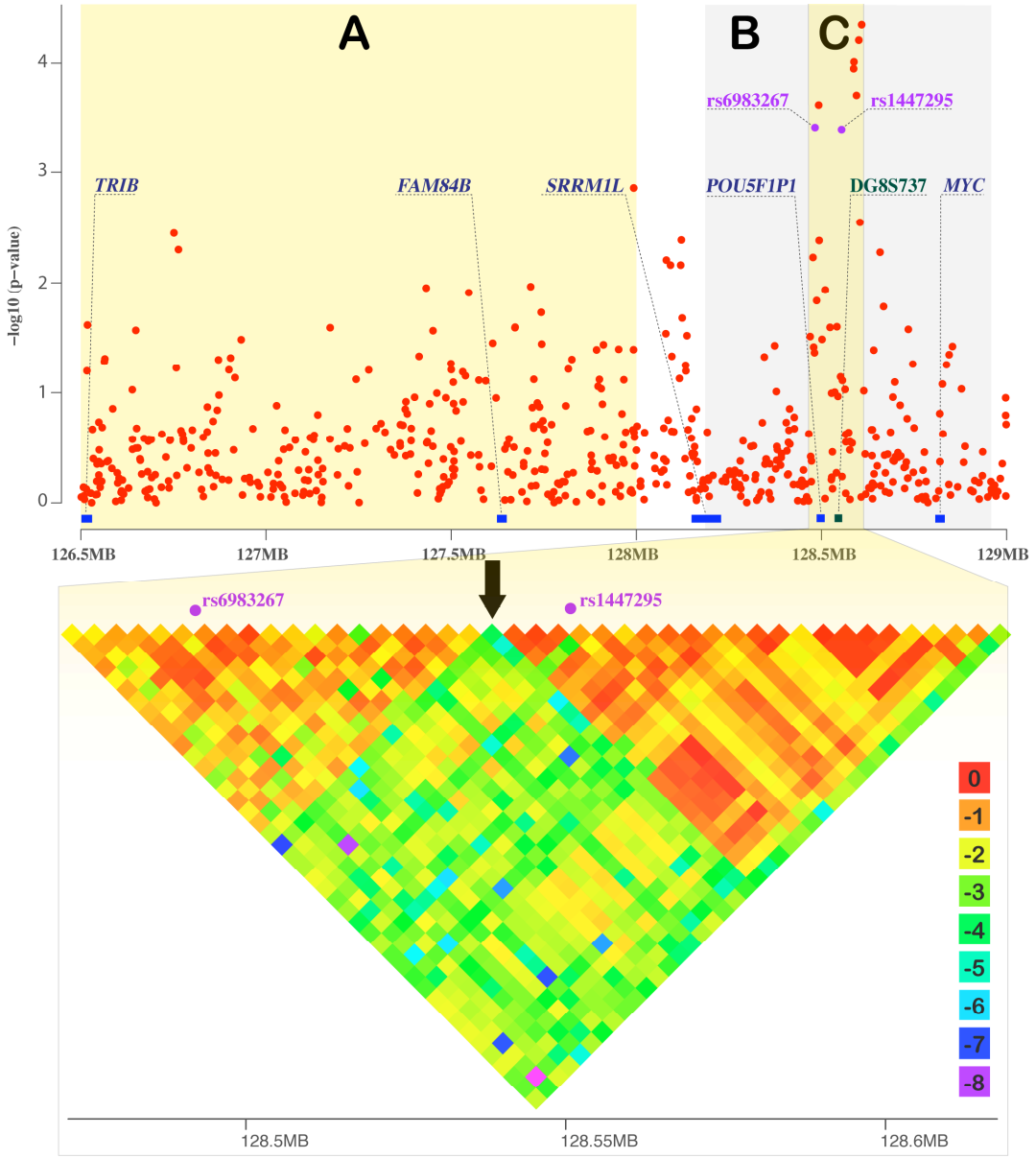
- Age
- Ethnic Background
- Family History
- Genetic markers
  - 16 Regions of the Genome!!!

Cancer susceptibility loci in the 8q24 region



Yeager et al Nature Genetics 2007

Discovery of ALL Variants



Roche/454 next-gen sequencing analysis

50X coverage, ~140kb

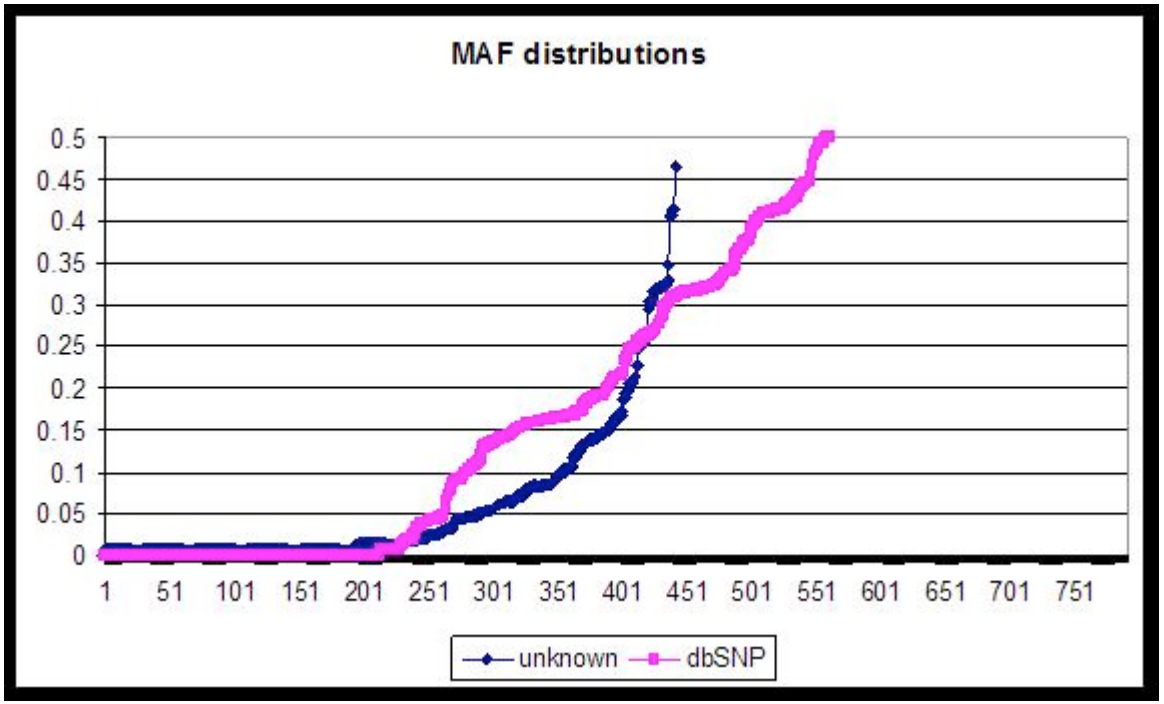
40 prostate cancer cases

40 controls

7 individuals from a CEPH family in which the at-risk haplotype is segregating (ARG)

Polymorphism identification in 87 Caucasians (40 cases, 39 controls & 8 CEU)

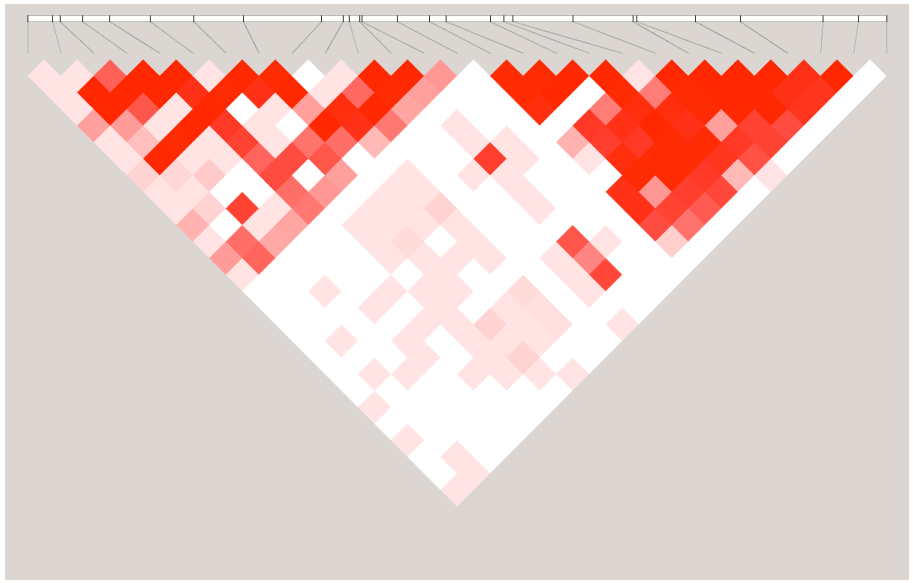
	Non-dbSNP	dbSNP
# monomorphic	n/a	213
# polymorphic	442	349
Minimum MAF	0.006	0.000
Maximum MAF	0.464	0.500
Mean MAF	0.060	0.142
Median MAF	0.013	0.101



Population Attributable Risk of Prostate Cancer with 8q24 Loci in Caucasians

	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.

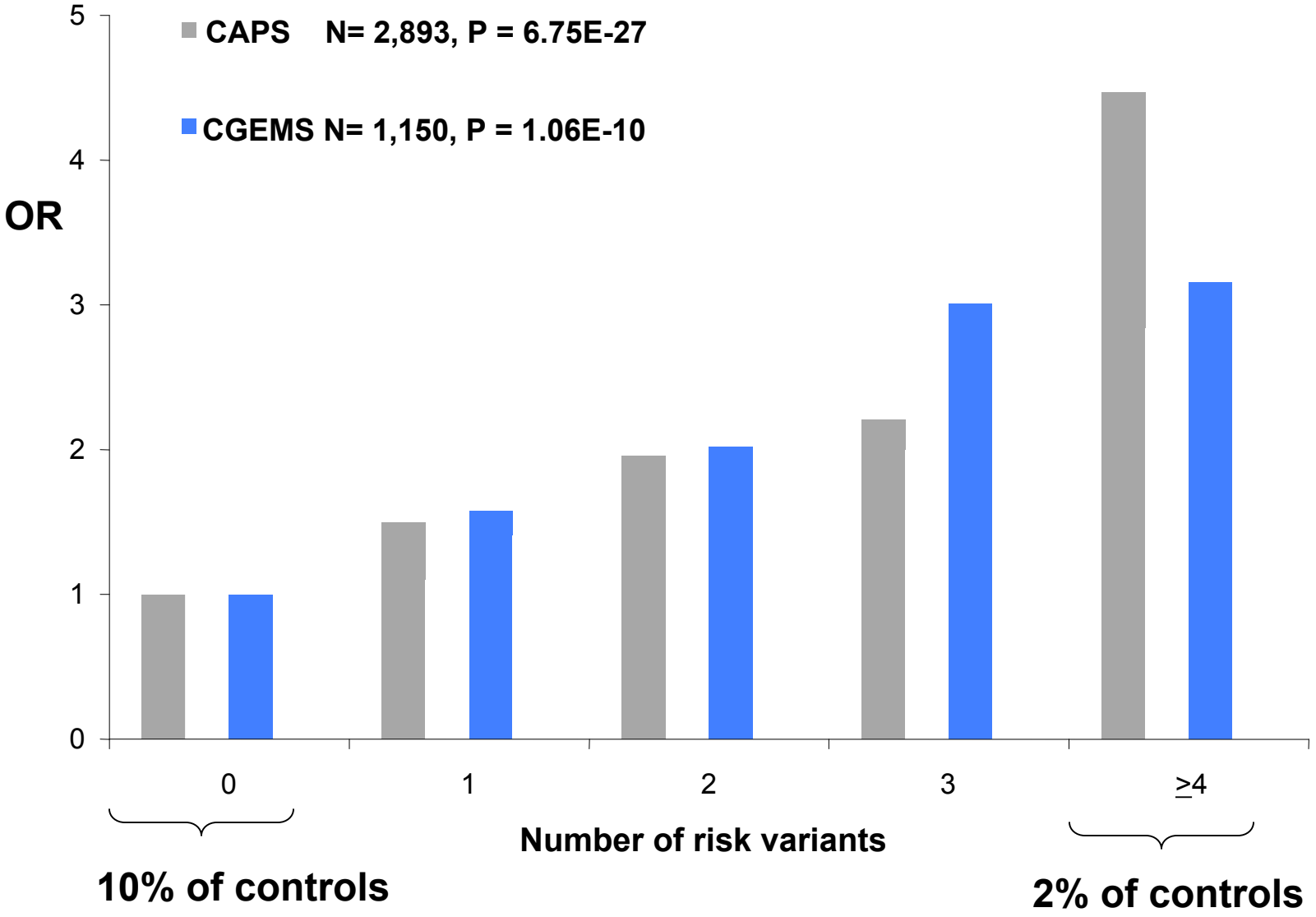
What variants to include in risk scores?

- Rapid pace of identification of new variants
- 2-3 years more to “complete” discovery for common alleles in common diseases
- Until then we are operating with a subset of common risk-associated variants

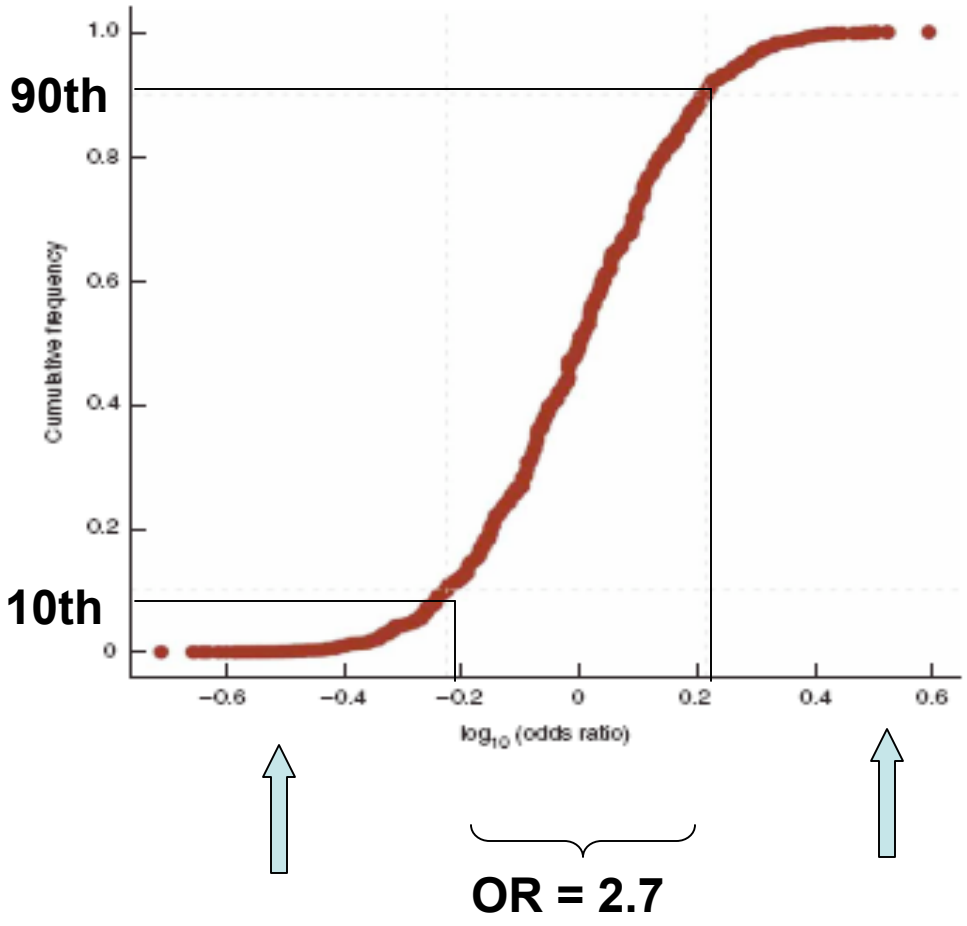
- Under the radar…..copy number variants, “rare” variants i.e <5% allele frequency

Genetic Gold Rush???

Cumulative effect of 5 risk variants (8q, 17q) on prostate cancer risk
Zheng et al, NEJM January, 2008



7-SNP CGEMS risk score:
"Cumulative" Population Attributable Risk (PAR) = 107% !!!



Thomas et al, 2008

How do we know there are many more variants to find?

- ***Current variants only account for a small fraction of the effect of family history***
  - **BCAC Breast Cancer SNPs account for less than 5%**
- **Current GWAS underpowered for low risk alleles**
- **Some known alleles have not shown up in GWAS**
- **Growing experience with pooling across GWAS datasets**
  - **e.g. Diabetes type II, Crohn's disease**

GWAS Studies: Just the Start……

This is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning.

Sir Winston Churchill @ Lord Mayor's Luncheon,
Mansion House following the victory at El Alameinin North Africa
London, 10 November 1942.

Follow-up to GWAS Studies

- Fine Mapping of Notable Regions
  - **Genotyping & Sequencing**
  - **Bio-informatics (exclude common CNV)**
- Analysis of Population Genetics
- Functional Determination of Causal Variant(s)
- Exploration of Pathways
  - **Etiology**
  - **Drug Targets**
- Design Issue for Analysis in Clinical Evaluation
  - **Population-based studies**
  - **Careful Clinical Studies**

Functional Analyses: Laboratory of Translational Genomics, DCEG, NCI

- Determine Plausibility of Finding
  - *Can we explain the effect?*
  - Molecular Phenotype
    - Correlation of *in vitro* changes with germ-line variant(s)
    - Cell line or tissue work with germ-line analyses
- Correlation with **Somatic Alterations**
  - Association of germ-line with somatic observations
  - Driver mutation

What Next?

- More Scans in Each Disease
  - Subtypes
  - Specific Populations: Breast cancer in AA
- In progress GWAS
  - Aggressive adult cancers
    - Pancreas, brain, ovary, esophagus, renal, bladder, melanoma
  - Rare/Pediatric
    - Neuroblastoma, childhood leukemia, osetogenic sarcoma
  - Ample follow-up for mapping/function
- Risk Assessment- Suitable Reporting
  - Public Health and Personal Decisions
- Next-Generation Sequencing

CGEMS: caBIG Posting
Pre-Computed Analysis

Pre-computed Analysis Post 4 Months Before Publication

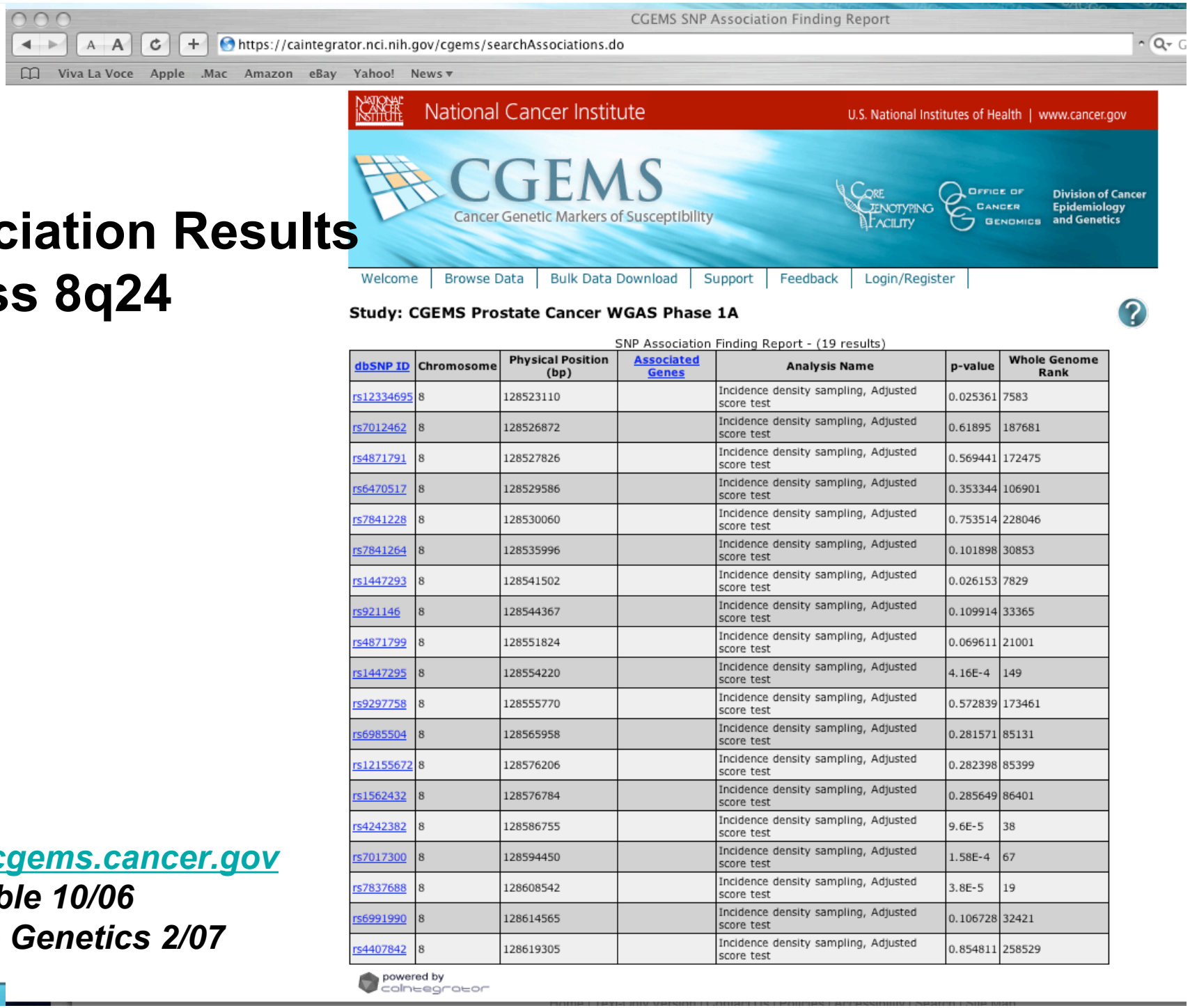
No Restrictions

Raw Genotype
- **Case/control**
- **Age (in 5 yrs)**
- **Family Hx (+/-)**

Registered Access
- **SF424**
- **Data Use Certificate**

http://cgems.cancer.gov/data

Association Results Across 8q24



dbSNP ID	Chromosome	Physical Position (bp)	Associated Genes	Analysis Name	p-value	Whole Genome Rank
rs12334695	8	128523110		Incidence density sampling, Adjusted score test	0.025361	7583
rs7012462	8	128526872		Incidence density sampling, Adjusted score test	0.61895	187681
rs4871791	8	128527826		Incidence density sampling, Adjusted score test	0.569441	172475
rs6470517	8	128529586		Incidence density sampling, Adjusted score test	0.353344	106901
rs7841228	8	128530060		Incidence density sampling, Adjusted score test	0.753514	228046
rs7841264	8	128535996		Incidence density sampling, Adjusted score test	0.101898	30853
rs1447293	8	128541502		Incidence density sampling, Adjusted score test	0.026153	7829
rs921146	8	128544367		Incidence density sampling, Adjusted score test	0.109914	33365
rs4871799	8	128551824		Incidence density sampling, Adjusted score test	0.069611	21001
rs1447295	8	128554220		Incidence density sampling, Adjusted score test	4.16E-4	149
rs9297758	8	128555770		Incidence density sampling, Adjusted score test	0.572839	173461
rs6985504	8	128565958		Incidence density sampling, Adjusted score test	0.281571	85131
rs12155672	8	128576206		Incidence density sampling, Adjusted score test	0.282398	85399
rs1562432	8	128576784		Incidence density sampling, Adjusted score test	0.285649	86401
rs4242382	8	128586755		Incidence density sampling, Adjusted score test	9.6E-5	38
rs7017300	8	128594450		Incidence density sampling, Adjusted score test	1.58E-4	67
rs7837688	8	128608542		Incidence density sampling, Adjusted score test	3.8E-5	19
rs6991990	8	128614565		Incidence density sampling, Adjusted score test	0.106728	32421
rs4407842	8	128619305		Incidence density sampling, Adjusted score test	0.854811	258529

http://cgems.cancer.gov
Available 10/06
Nature Genetics 2/07

Acknowledgements

CGEMS & DCEG
Gilles Thomas
Robert Hoover
Joseph Fraumeni
Daniela Gerhard
Richard Hayes
Kevin Jacobs
Meredith Yeager
Sholom Wacholder
Nilanjan Chatterjee
Zhaoming Wang
Xiang Deng
Nick Orr
Robert Welch
Sonja Berndt
Kai Yu
Margaret Tucker
Jesus Gonzalez Bosquet
Jiyoung Ahn

HSPH
David Hunter
Peter Kraft
Fred Schumacher
W Willett
J Ma

PLCO
Chris Berg
E David Crawford

ACS
Michael Thun
Heather Feigelson
Eugenia Calle
Carmen Rodriguez

CeRePP, France
Olivier Cussenot
Geraldine Cancel-Tassin
Antoine Valeri

NPHI, Finland
Jarmo Virtamo

Wash. U., St Louis
Gerald Andriole
Adam Kybell