

CGEMS

Cancer Genetic Markers of Susceptibility

<http://cgems.cancer.gov>



The Promise of Genome-wide Association Studies (GWAS)

Stephen J. Chanock, M.D.

Division of Cancer Epidemiology and Genetics

Center for Cancer Research

December 1, 2006

Value of GWAS

- **Identification of promising low-penetrance, high-frequency susceptibility loci**
- **Evaluation of gene-gene interactions and genetic interactions with environmental exposures**
- **Tool for identifying novel mechanisms in cancer**
- **Foundation for strategies for prevention and intervention**

GWAS & NCI Priorities

- **Capitalize on revolution in genetics**
 - **Annotation of common genetic variation**
 - **Technology platforms**
- **Intramural capabilities at Core Genotyping Facility**
- **NCI investment in cohorts**
- **Informatics and access: NCICB (caBIG portal)**
- **Coordinate with NIH-wide activities**

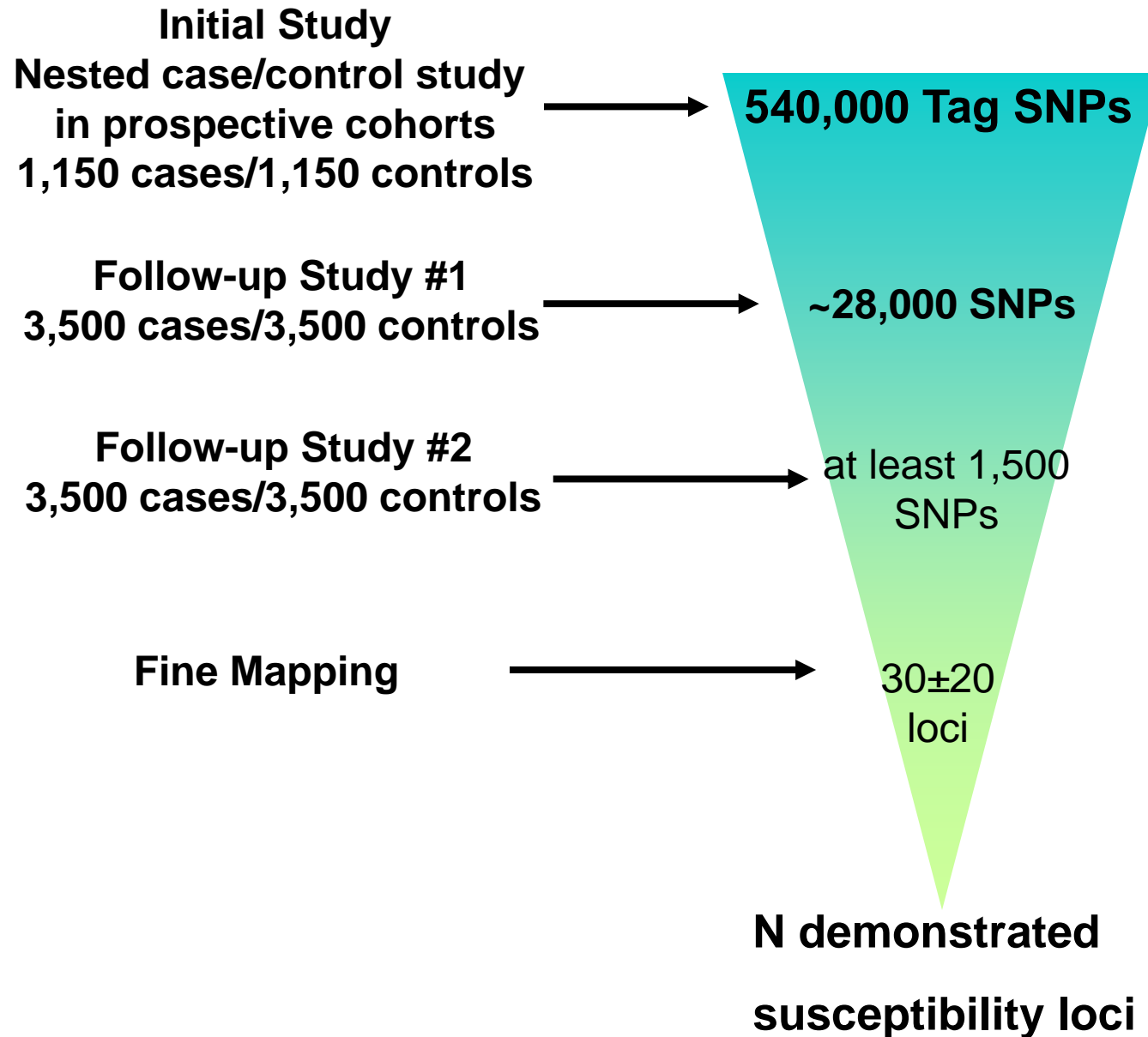
Mission of Cancer Genetic Markers of Susceptibility



- **Conduct GWAS in 2 cancers**
 - **Prostate (1 in 8 men)**
 - **Breast (1 in 9 women)**
- **Rapid sequential replication studies**
- **Aggressive timeline**
- **Initial scans in nested case-control studies**
 - **Prostate, Lung, Colon, Ovary (PLCO) Project**
 - **Nurses' Health Study**

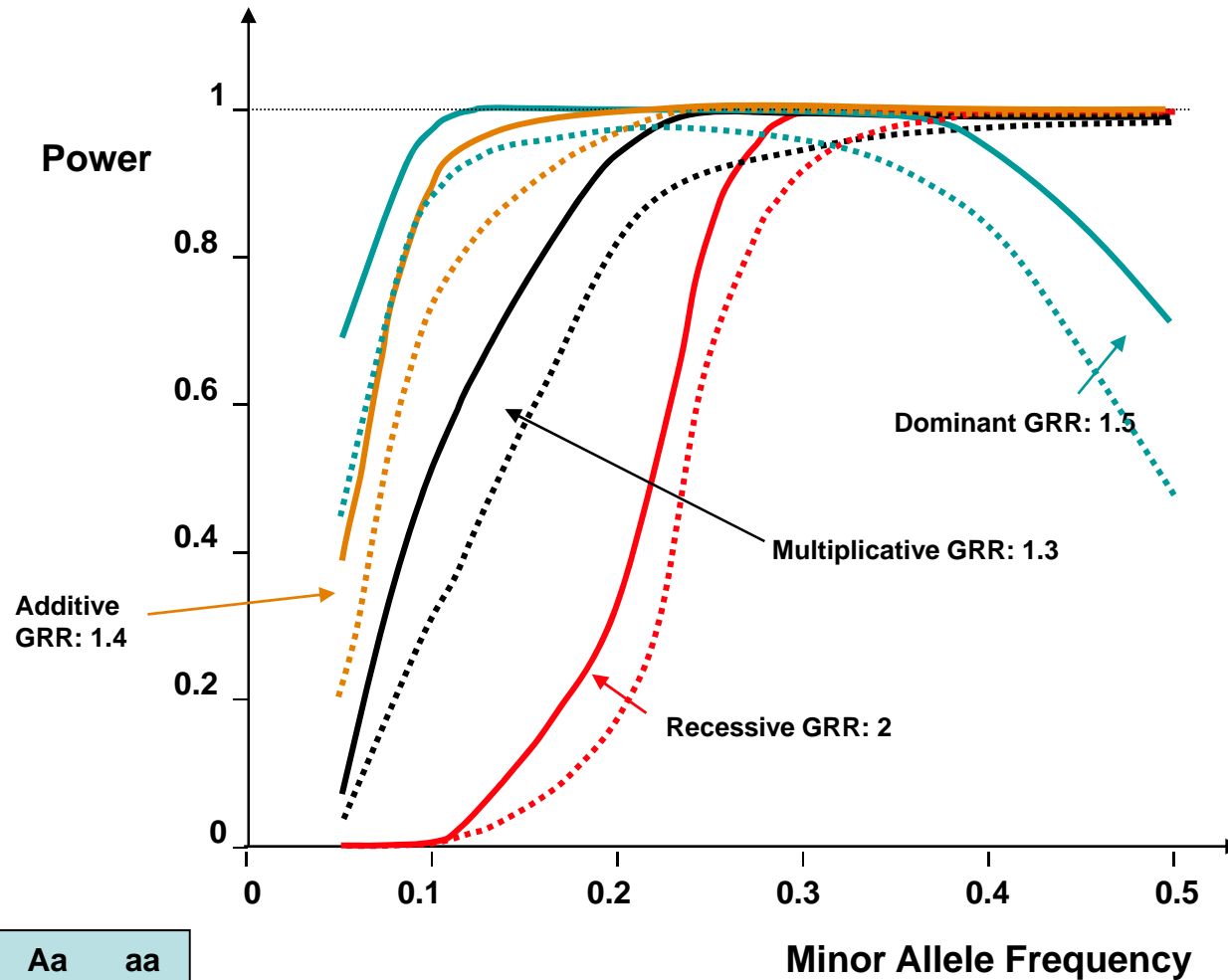
<http://cgems.cancer.gov>

Strategy for Prostate & Breast Cancer



Power of the First Two Phases of CGEMS

Point-wise significance 10^{-7} ; "genome wide" significance 0.05



	GRR	AA	Aa	aa
Recessive	2.0	1.0	1.0	2.0
Dominant	1.5	1.0	1.5	1.5
Additive	1.4	1.0	1.4	1.8
Multiplicative	1.3	1.0	1.3	1.69

Continuous line: power for direct detection ($r^2 = 1$)
 Dashed line: power for $r^2 = 0.8$

What is available for GWAS in 2006?

Coverage analysis based on HapMap II Data

Build 20 MAF \geq 5%, $r^2 \geq$ 0.8 (pair-wise)

		CEU	YRI	JPT/CHB
Illumina	HumanHap300	80%	35%	40%
Illumina	HumanHap500	91%	58%	88%
Affymetrix*	500k Mapping	63*%	41%	63%

***77% (with 50k MegA)**

<http://tagzilla.nc.nih.gov>

CGEMS Scans



Prostate Cancer

Breast Cancer

Two Scans

One Scan

Illumina

Illumina

317k

(available)

240k

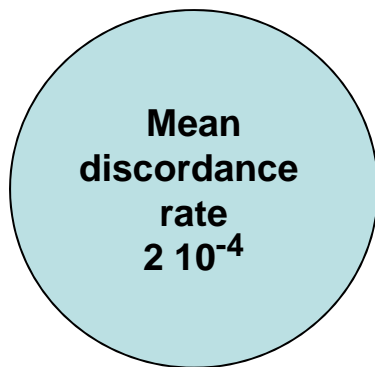
(February 2007)

550k

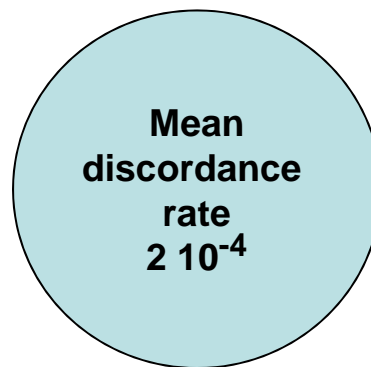
(March 2007)

Discordance Rates in Genotype Analysis

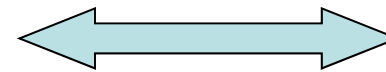
PLCO
49 duplicate pairs



CEPH-CGEMS
74 duplicate pairs

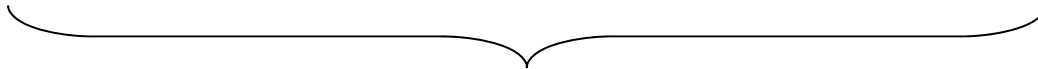
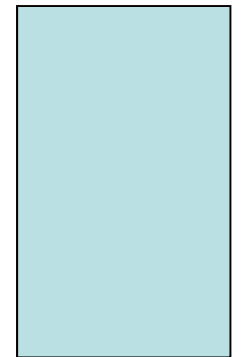


28 individuals
(with 24 duplicates)



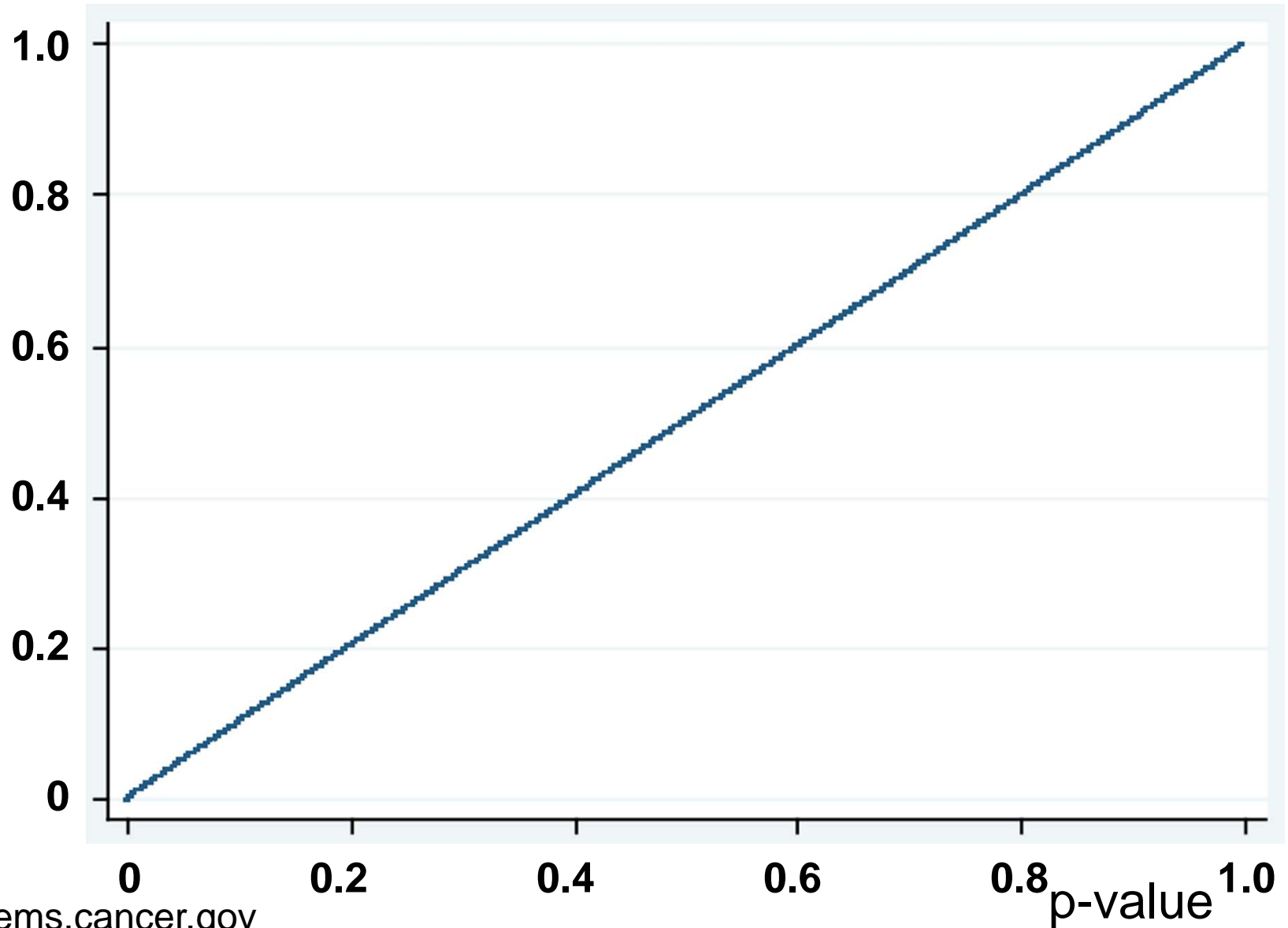
Mean
Discordance
rate
 $1.4 \cdot 10^{-3}$

**CEPH-
HapMap**



QQ Plot for p-values of ~300k SNPs in Prostate Scan 1A

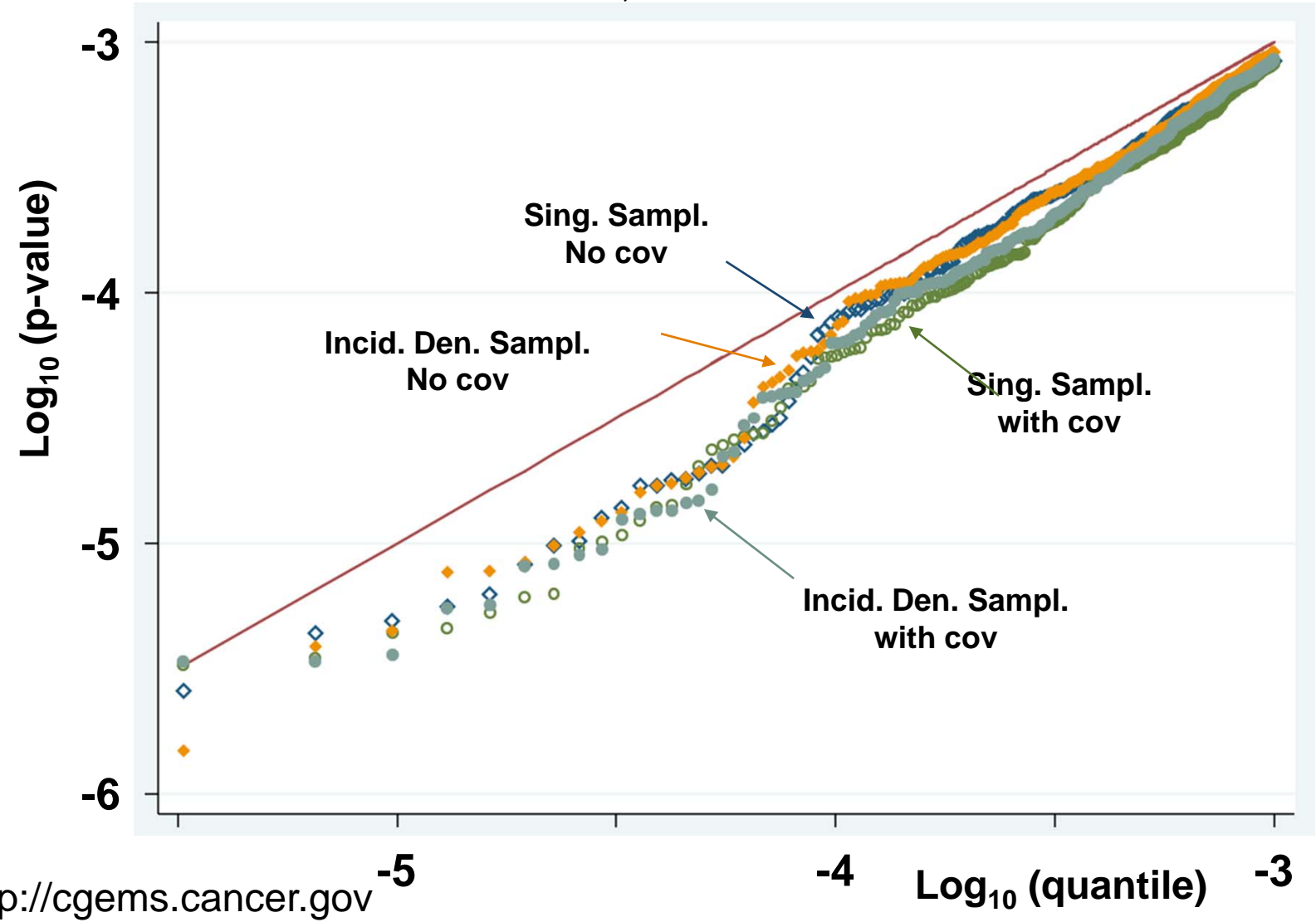
Quantile



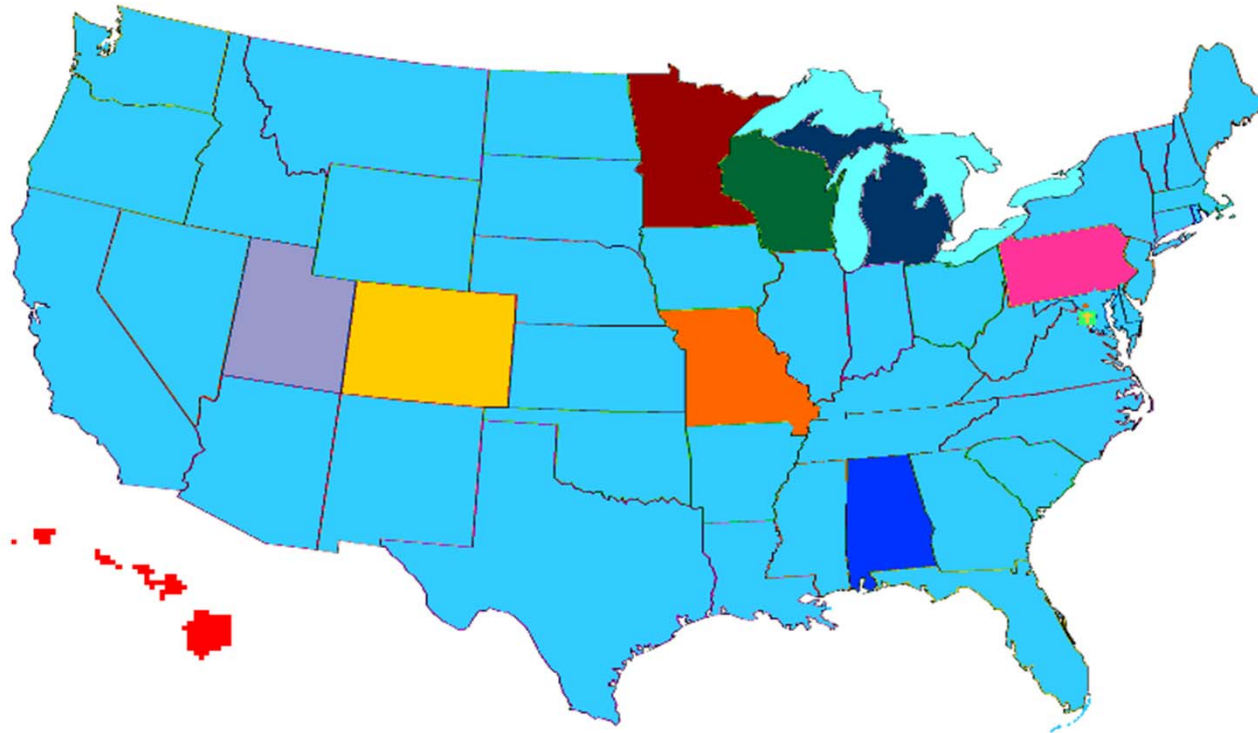
<http://cgems.cancer.gov>

Log-Log Quantile Plot for p-values for the Four Statistical Tests Used

307,256 SNPs



PLCO Recruitment Sites Opportunity to Look at Geographic Differences



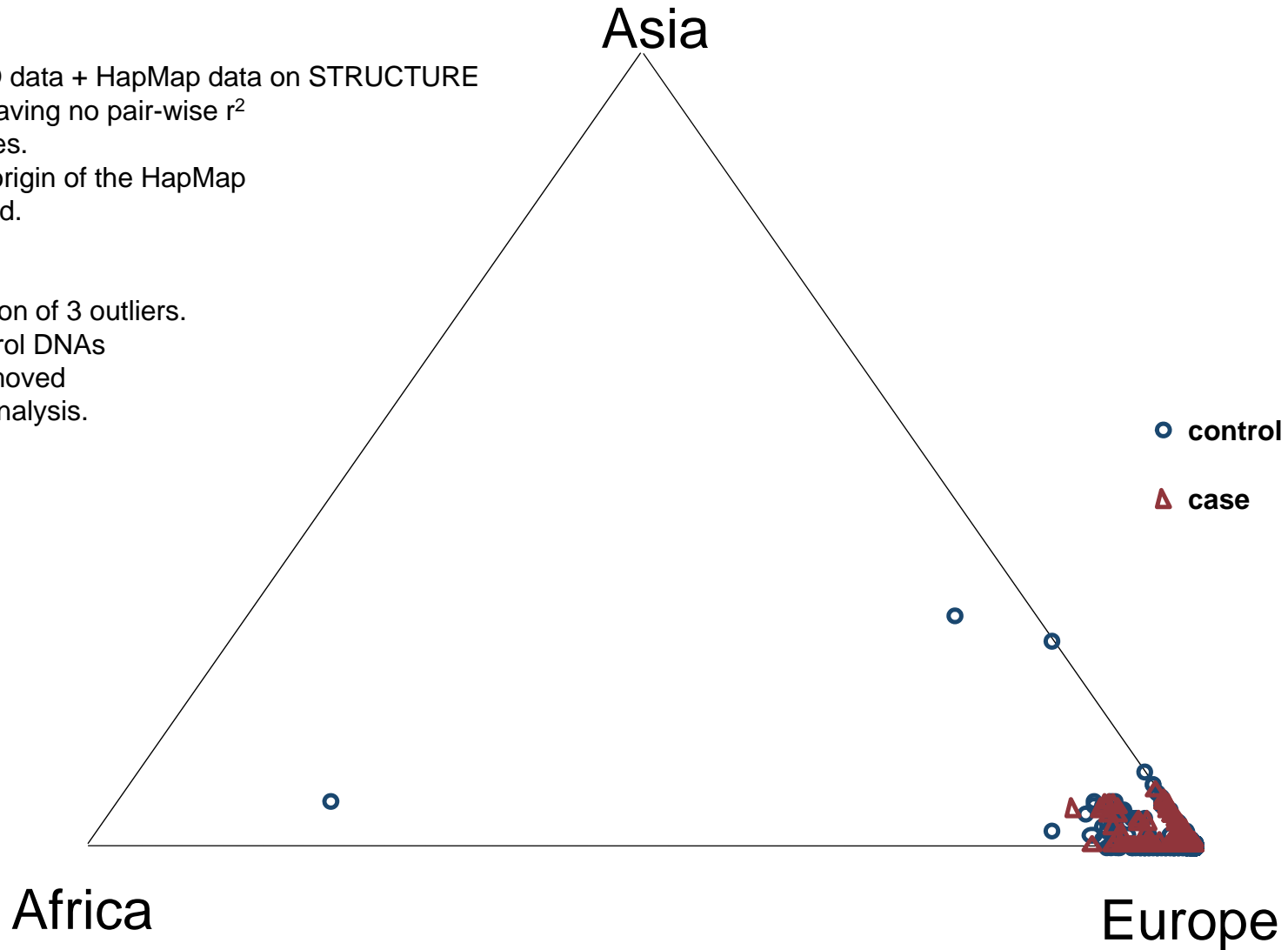
Admixture Coefficient in PLCO Prostate Study Samples

Method:

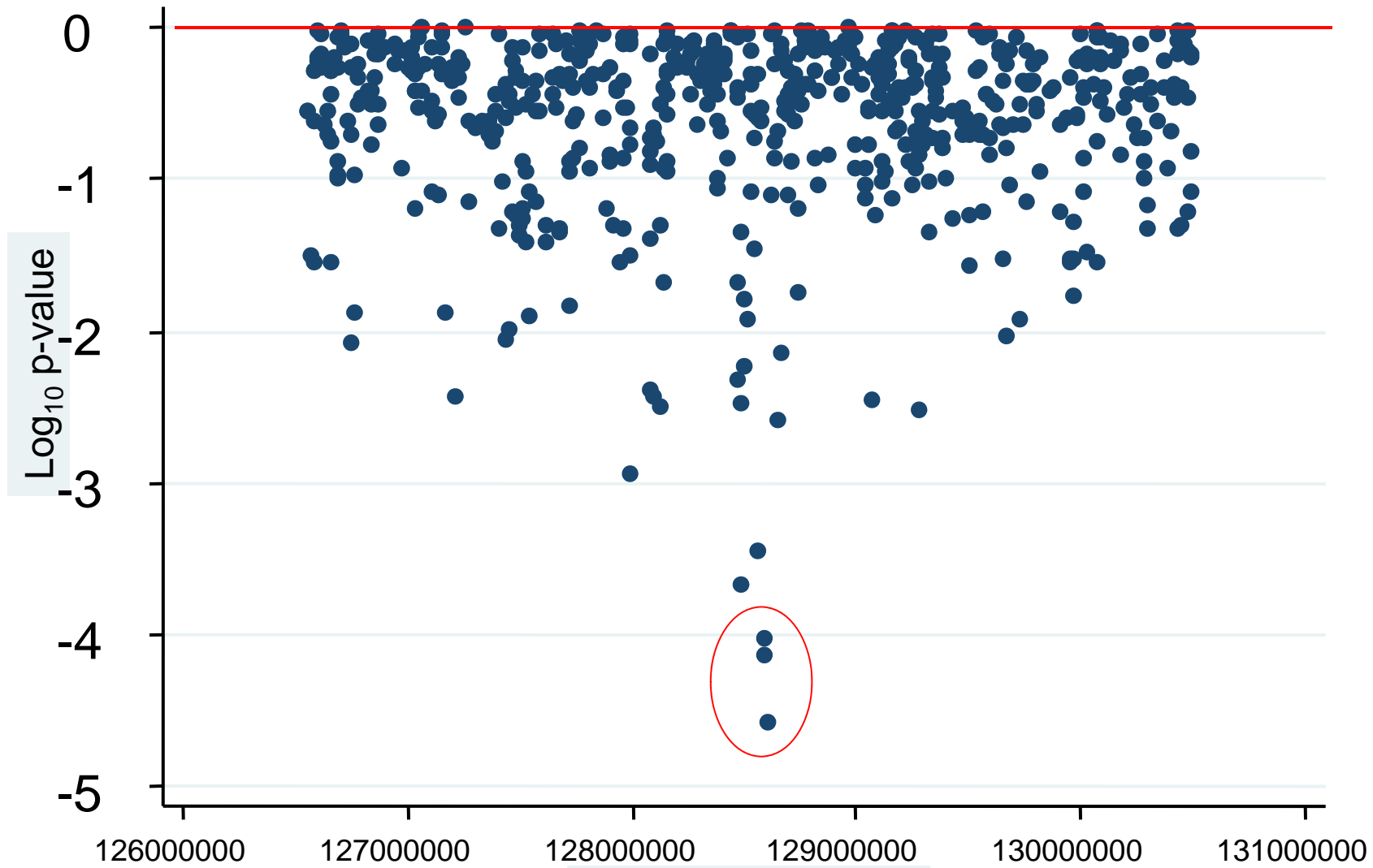
Run merged PLCO data + HapMap data on STRUCTURE with 6,000 SNPs having no pair-wise r^2 and high F_{ST} values. The population of origin of the HapMap samples is specified.

Result:

Reliable identification of 3 outliers. They are all 3 control DNAs and have to be removed from subsequent analysis.



Log₁₀ p-value of the 4 d.f. χ^2 test Plotted Against the Position of the 8q24 SNP (rs#1447295)* in Build 35



*Amundadottir Nat Genet 2006
*Freedman PNAS 2006

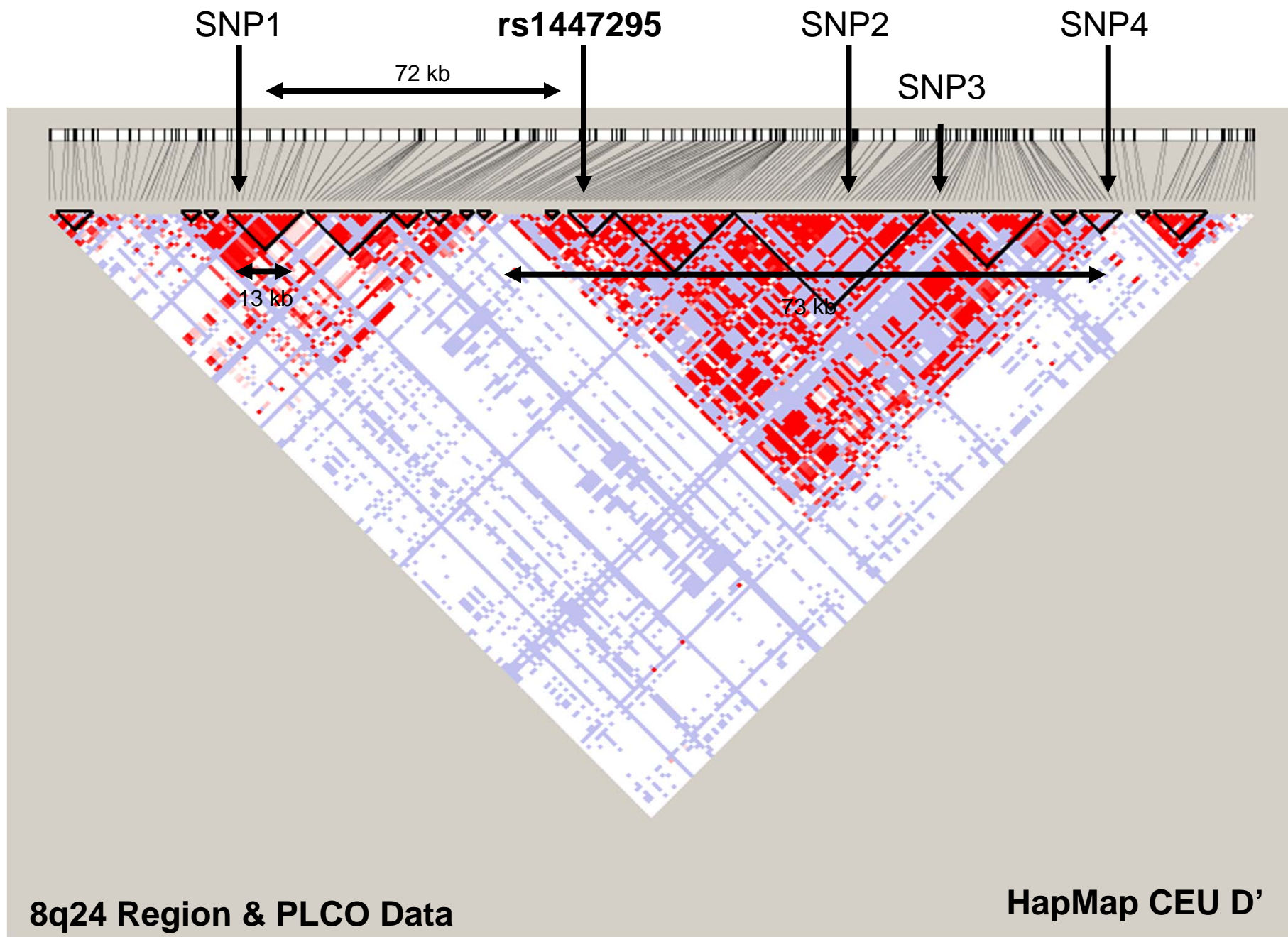
Prostate Scan 8q24 Region



rs number	susceptibility allele	allele frequency	Genotype RR for Indolent		Genotype RR for aggressive	
			Heterozyg.	Homozig.	Heterozyg.	Homozig.
rs1447295	A	0.1	1.08	1.45	1.24	1.46
rs4242382	A	0.1	1.13	1.39	1.27	1.39
rs7017300	C	0.13	1.14	1.63	1.17	1.37
rs7837688	T	0.1	1.14	1.36	1.26	1.54

Key Findings:

1. Comparable risk as original reports in Nat Genet and PNAS
2. Comparable risk for BPC3 (~6,500 cases/controls)
3. Discovery of 1 and perhaps 2 additional loci



Value-added Analysis in CGEMS



Opportunity to Investigate

- **Determinants of risk factors**
 - **BMI, smoking, hormone levels**
- **Multi-SNP analysis**
- **Gene: Gene interactions**
 - **Explore pathways**

CGEMS: caBIG Posting Pre-computed Analysis



National Cancer Institute

U.S. National Institutes of Health | www.cancer.gov



CGEMS

Cancer Genetic Markers of Susceptibility



Division of Cancer
Epidemiology
and Genetics

[Home](#) | [Browse Data](#) | [Bulk Data Download](#) | [Feedback](#)



This is the home page of the Cancer Genetic Markers of Susceptibility (CGEMS) data access. The following links provide information on the [project](#) and [background](#). The CGEMS study design uses cases and controls drawn from well designed epidemiological studies of prostate and breast cancer. DNA from these subjects is being used to generate genotypes to perform a Genome-Wide Association Study (GWAS) on over 500,000 genetic variants to determine their role in cancer susceptibility.

CGEMS Prostate Scan Phase 1

A GWAS has been conducted in a large, national study in the U.S.A., the Prostate, Lung, Colorectal, and Ovary study ([PLCO](#)). The analysis includes 1,177 subjects who developed prostate cancer during the observational period and 1,105 individuals who did not develop prostate cancer during the same time period. The prostate scan is being conducted in two parts, Phase 1A and Phase 1B

The data generated from these scans can be accessed through this portal. The first posting includes data from Phase 1A of the prostate cancer scan and includes:

- Association test results for over 300,000 SNPs
- Frequency and descriptive statistics on these SNPs
- Individual phenotypic and genotypic data for the study participants and control samples. Note that these data can only be made available to eligible investigators after a registration process ([link](#)).

The results of Phase 1B will be available in February 2007.

[Browse Data](#)

[Bulk Data Download](#)

For more information on:

- [About CGEMS Study](#)
- [How to use the CGEMS data portal](#)
- [Register to access raw data](#)



Click the question mark icon for context sensitive help throughout the application.

CGEMS updates:

- This release, Version 1.0, was deployed on Oct 10, 2006.
- The current dataset in use was deployed on Oct 10, 2006

**Pre-computed Analysis
No Restrictions**

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

<http://cgems.cancer.gov>

Association Tests

8q24

Scan 1A

~300,000 SNPs



Welcome | Browse Data | Bulk Data Download | Support | Feedback | Login/Register

Study: CGEMS Prostate Cancer WGAS Phase 1A

SNP Association Finding Report - (19 results)

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>

GWAS in Pancreatic Cancer: PanScan Objectives

- **Identify loci associated with pancreatic cancer**
 - **1,200 cases and 1,200 controls drawn from 12 cohort studies**
- **Define susceptibility loci for common genetic variants (MAF > 5%)**
- **Follow-up studies in cohort and case control studies**
- **Public access for data**
 - **Pre-computed association testing – Open**
 - **Raw genotype data with limited phenotype data-registered access**

What is down the road?

2-4 Year Forecast

- **Cheaper and denser SNP technologies**
 - Better coverage of genome

4-8 Year Forecast

- **Whole Genome Sequencing**
 - Replace SNPs
 - Magnification of Challenge of Confidentiality
 - Challenge to Epidemiologic Rigor

Follow-up of GWAS: Steps to Clinical Implementation

- **Fine mapping of notable regions**
- **Functional determination of causal variants**
- **Design issue for analysis in clinical studies**
 - **Population-based studies**
 - **Sequence of clinical studies**
- **Validation criteria**