

Cancer Genetic Markers of Susceptibility

Stephen J Chanock, M.D.

November 29, 2006

<http://cgems.cancer.gov>



Mission of CGEMS



Conduct genome-wide SNP scans in

Prostate cancer (1 in 8 men)

Breast cancer (1 in 9 women)

Analyze and publish findings

Rapid sequential replication studies

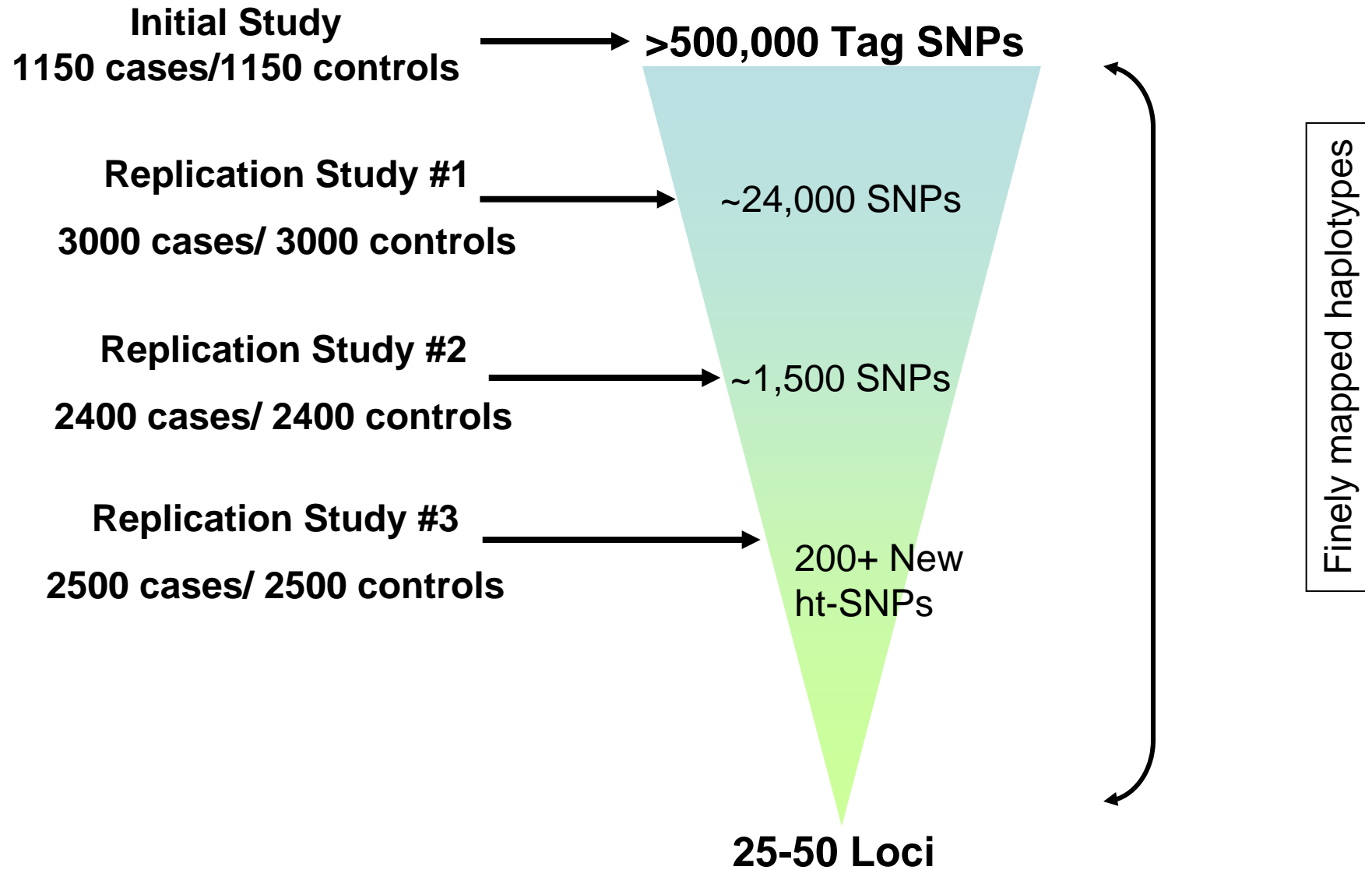
Aggressive timeline

Initial scan in nested case-control studies from

Prostate, Lung, Colon, Ovary (PLCO) Project

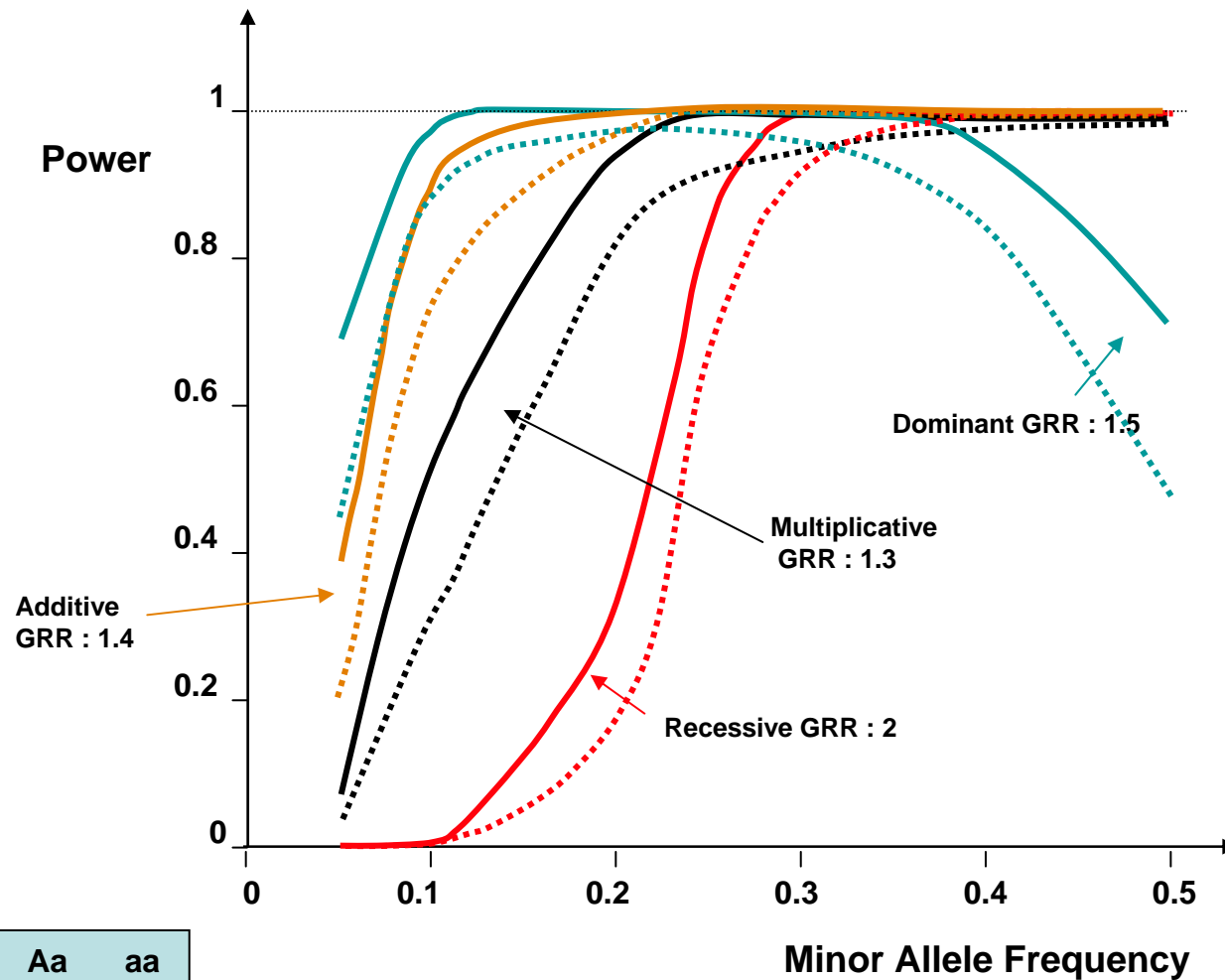
Nurses' Health Study

Replication Strategy for Prostate 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 $r^2 = 0.8$

CGEMS Scans



Prostate Cancer

Breast Cancer

Two Scans
Illumina

One Scan
Illumina

317k
(available)

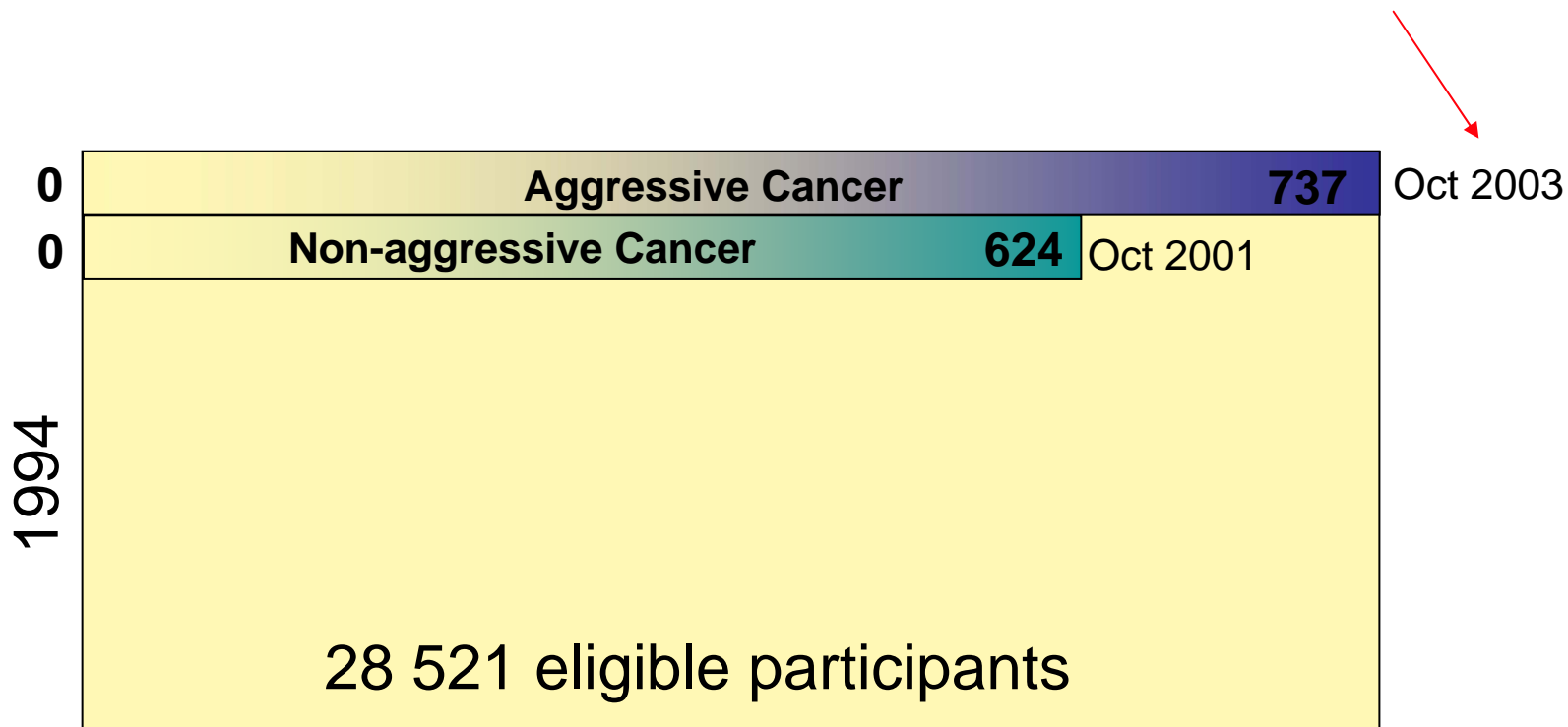
240k
(Feb 2007)

550k
(March 2007)

Aggressive Prostate Cancer

- High priority to examine early vs aggressive
- Cohort based studies (screening)
 - Bias towards early cases
- Enrich primary scan with >55% aggressive:45% early
 - Aggressive defined as:
 - Gleason \geq 7 +/-or Stage C/D
 - Follow-up studies in cohorts
 - Comparable distributions for early/advanced

Inclusion in CGEMS from PLCO of prostate cancer patients



Matching with controls was performed for 737 aggressive cases and 493 randomly selected non-aggressive cases.

Non aggressive : stage ≤ 2 (non invasive) and Gleason score ≤ 6
Aggressive : stage ≥ 3 (invasive) and Gleason score ≥ 7

Distribution of genotyped individuals used for the search of association

Prostate cancer status at start of CGEMS project		Number of times selected as controls				Total	
		0	1	2	3		
during follow-up	Always negative	0	1 082	22	1	1 105	"controls"
	Diagnosed with non-aggressive C.	461	26	1	0	488	} 1 177 cases
	Diagnosed with aggressive C.	673	16	0	0	689	
	Total	1 134	1 124	23	1	2 282	
		<div style="border: 1px solid black; padding: 5px; display: inline-block; background-color: yellow;">1 173 "controls"</div>					

dropped :
 1 XX DNA
 4 unexpected dup

dropped :
 1 XX DNA
 2 unexpected dup
 3 failed genotype

dropped :
 1 XX DNA
 4 failed genotype

Buccal Cell DNA and Infinium™ II: ACS:CGEMS Pilot

23 matched blood and buccal

Archived Buccal samples (2001/2002 in CPS-II)

Swish with Scope™ and store after centrifugation

Extracted simultaneously with Autopure (Gentra)

Target 50ng/uL by QDNA (picogreen)

4 outliers (0.5ng/uL- 35ng/uL)

HumanHap300 Infinium™ II protocol

Completion 99.02%

Concordance 99.96%

PLCO WGS QC

Removal of Inconsistent Genotypes

Low Completion Rate (<95%)

Duplicates: HapMap & PLCO qc samples

Fitness for HW Proportion in controls

Exclusion Cut-off: <0.001

Re-Map SNP Positions

Examine adjacent bps of SNPs

Heterogeneity in Cases/Controls

Cryptic stratification

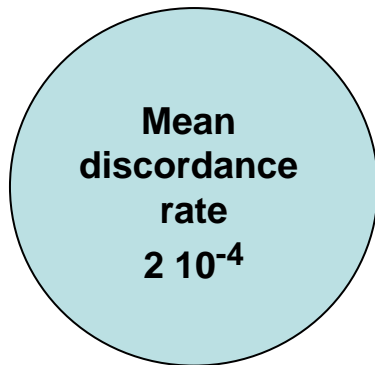
STRUCTURE (*Pritchard*)

Principal Component Analysis (*Price Nat Gen 2006*)

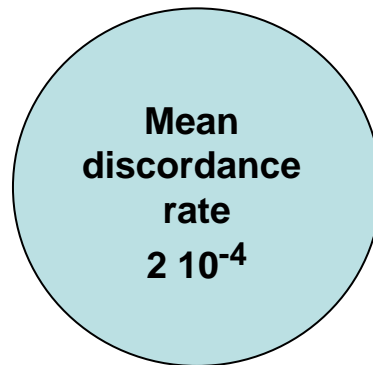
Study Center (9 for PLCO)

Discordance rate

PLCO
49 duplicate pairs



CEPH-CGEMS
74 duplicate pairs



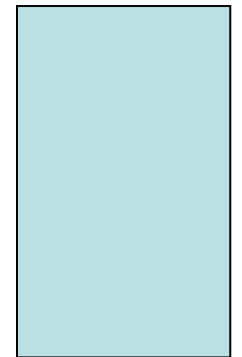
28 individuals
(with 24 duplicates)



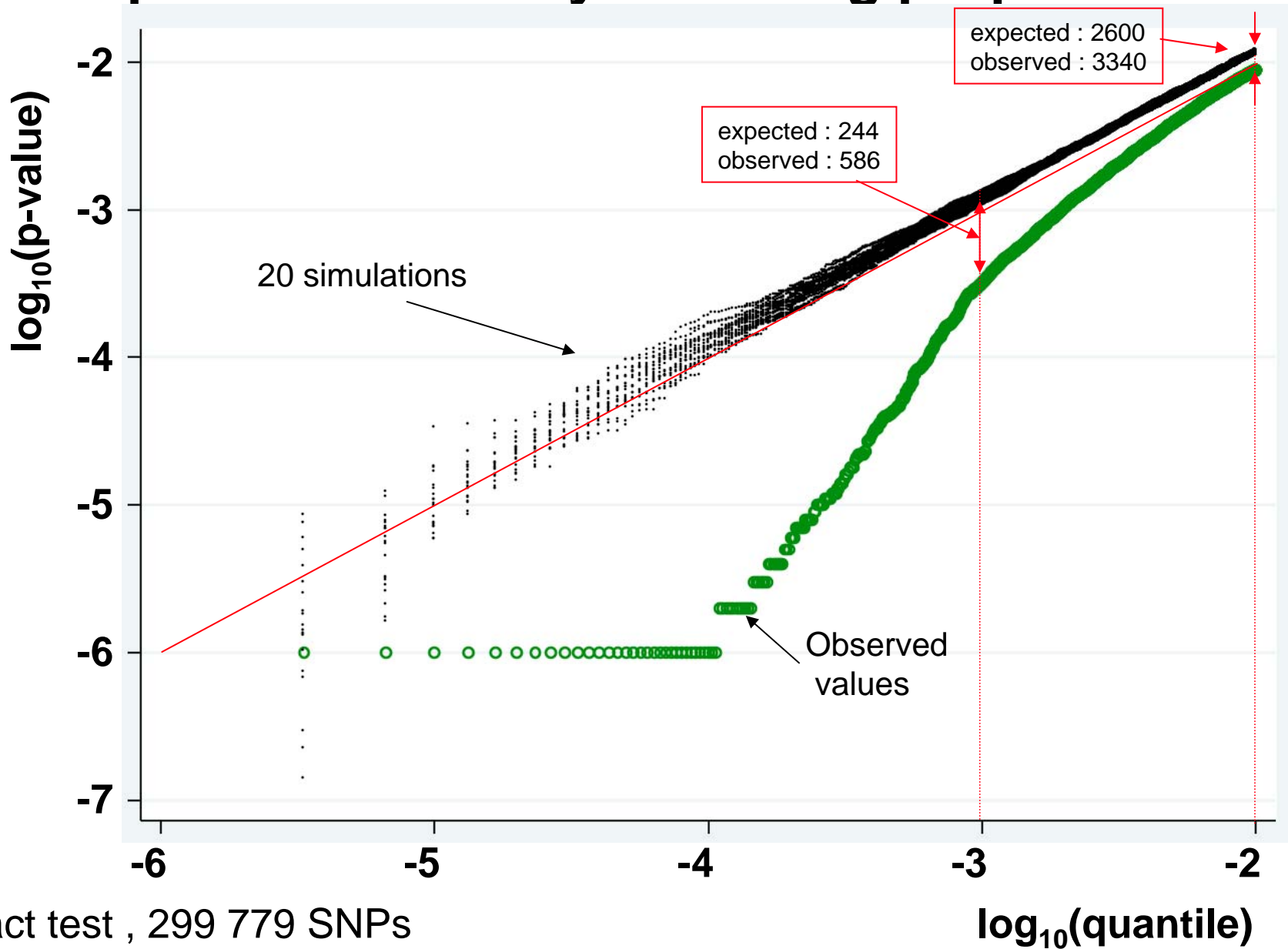
Mean
discordance
rate

$1.4 \cdot 10^{-3}$

**CEPH-
HapMap**

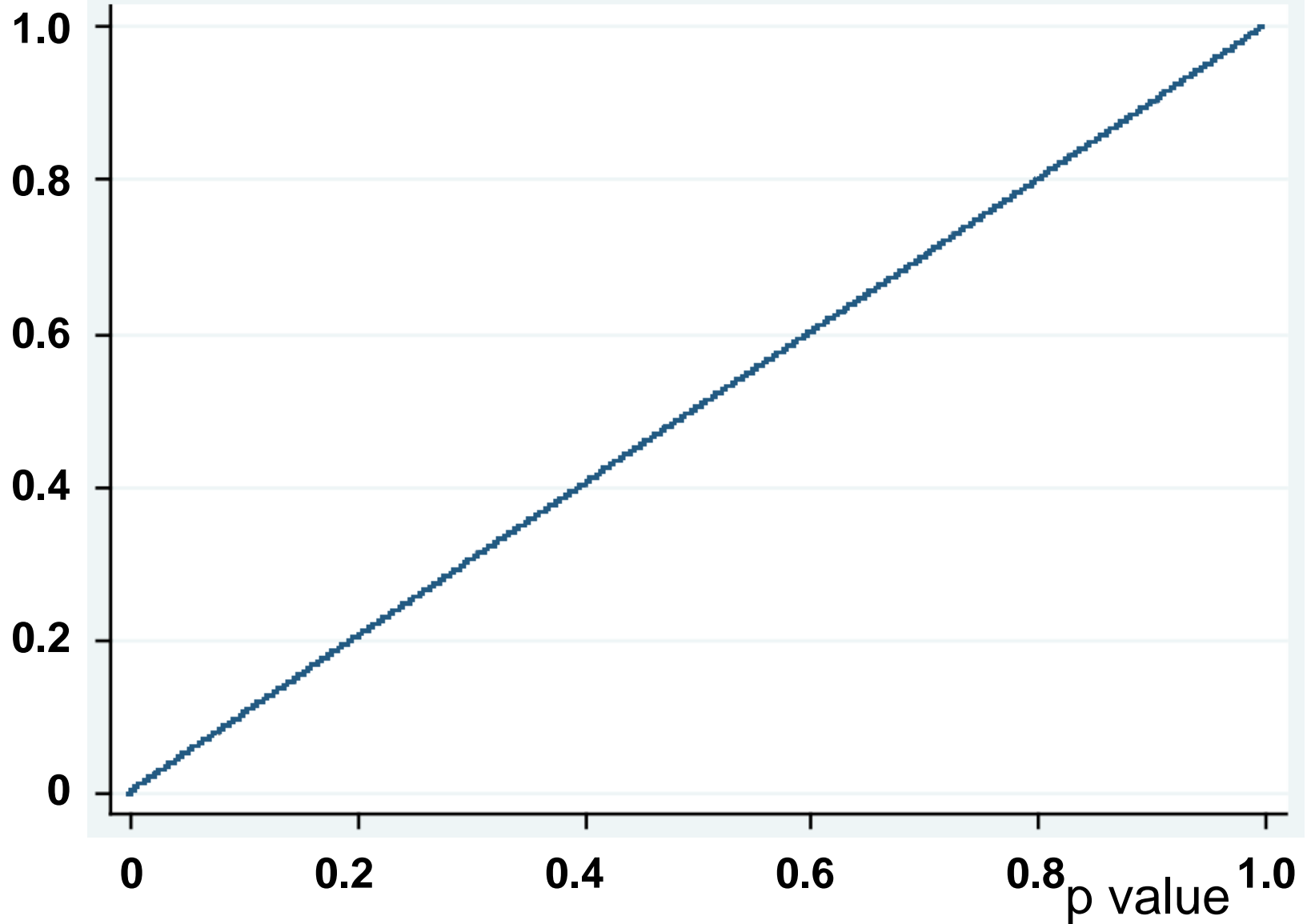


log-log quantile plot of p-value for Hardy-Weinberg proportion



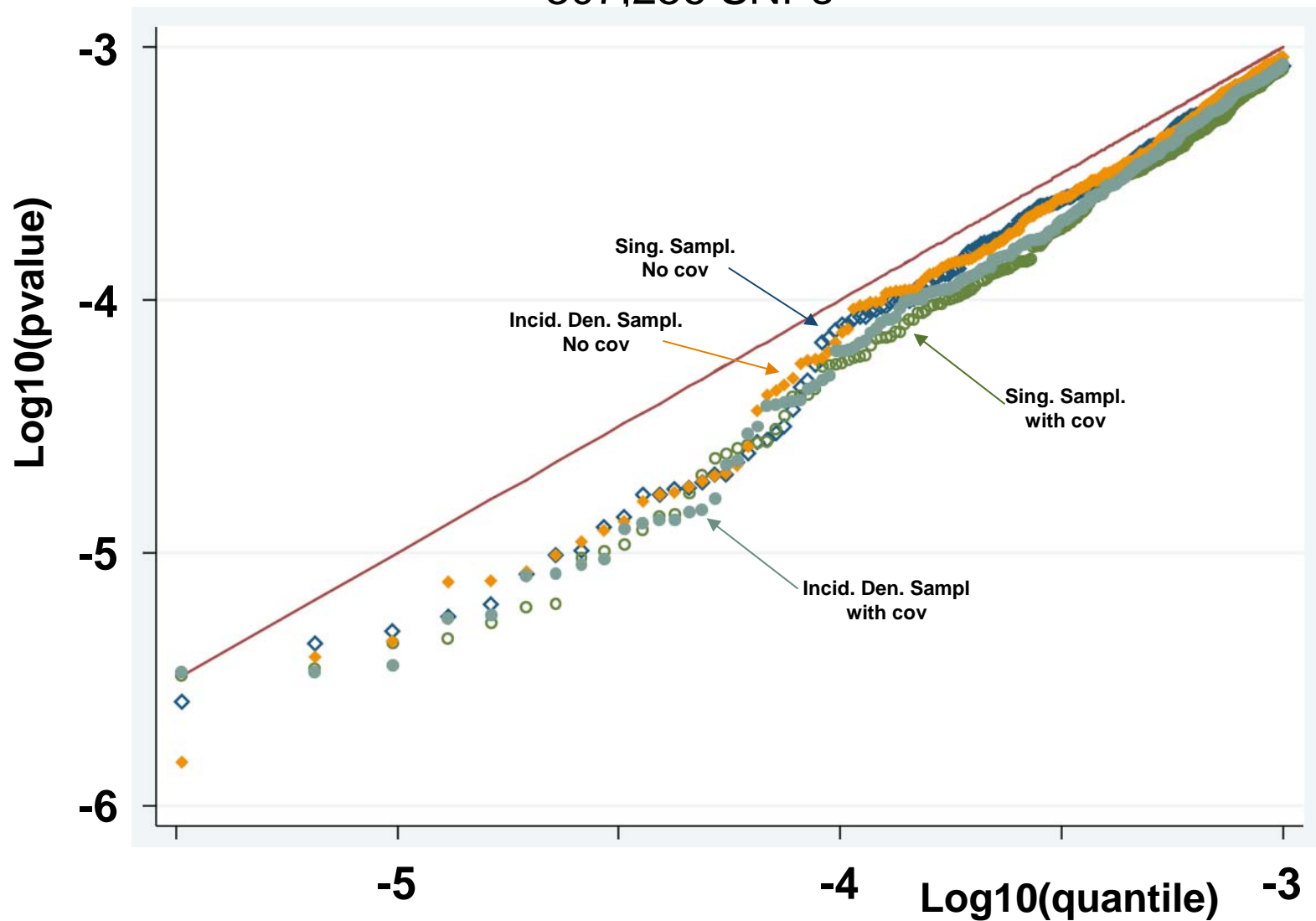
QQ plot for ~300k SNPs

Quantile

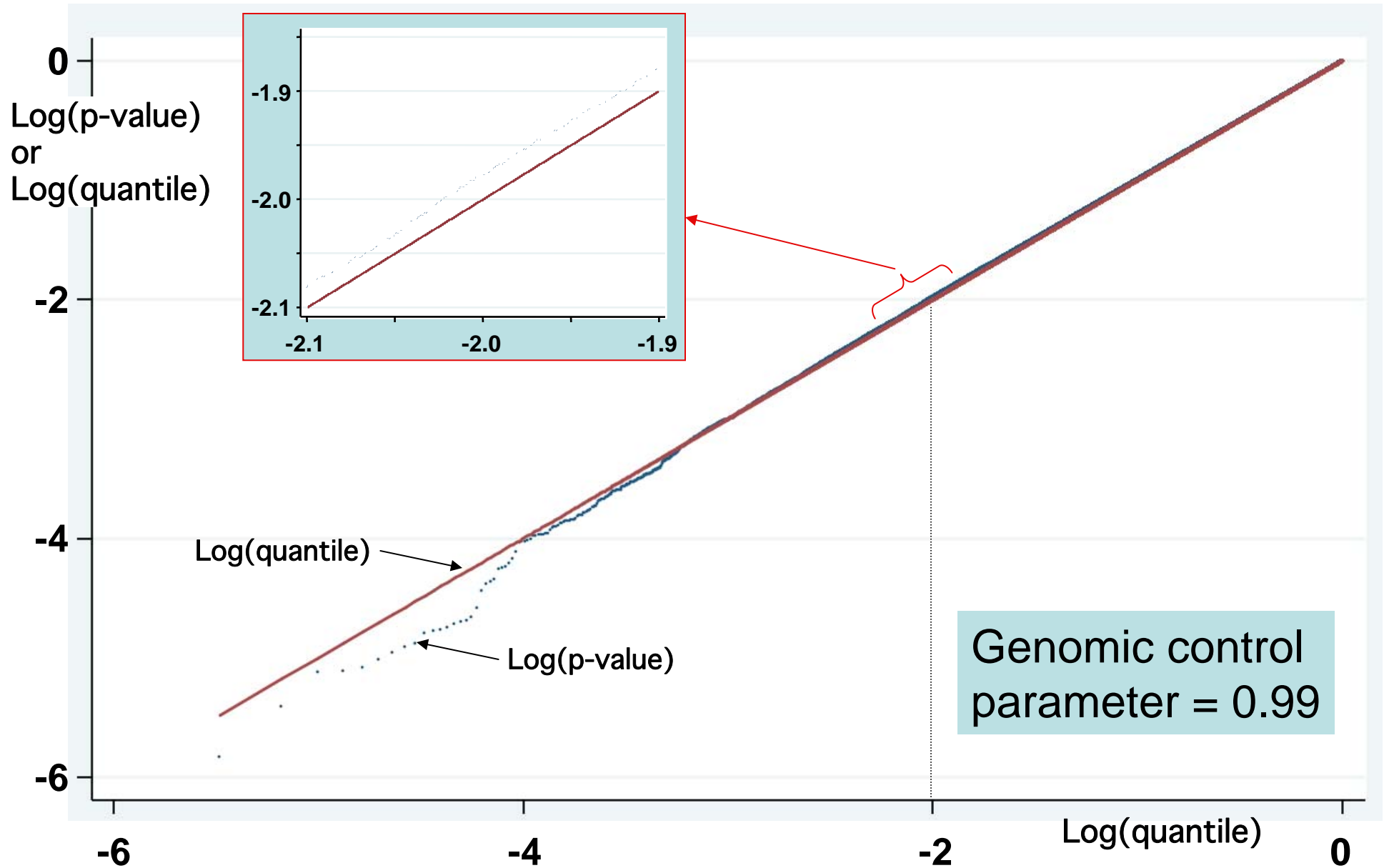


Log-Log quantile plot for p-value for the 4 statistical tests used

307,256 SNPs

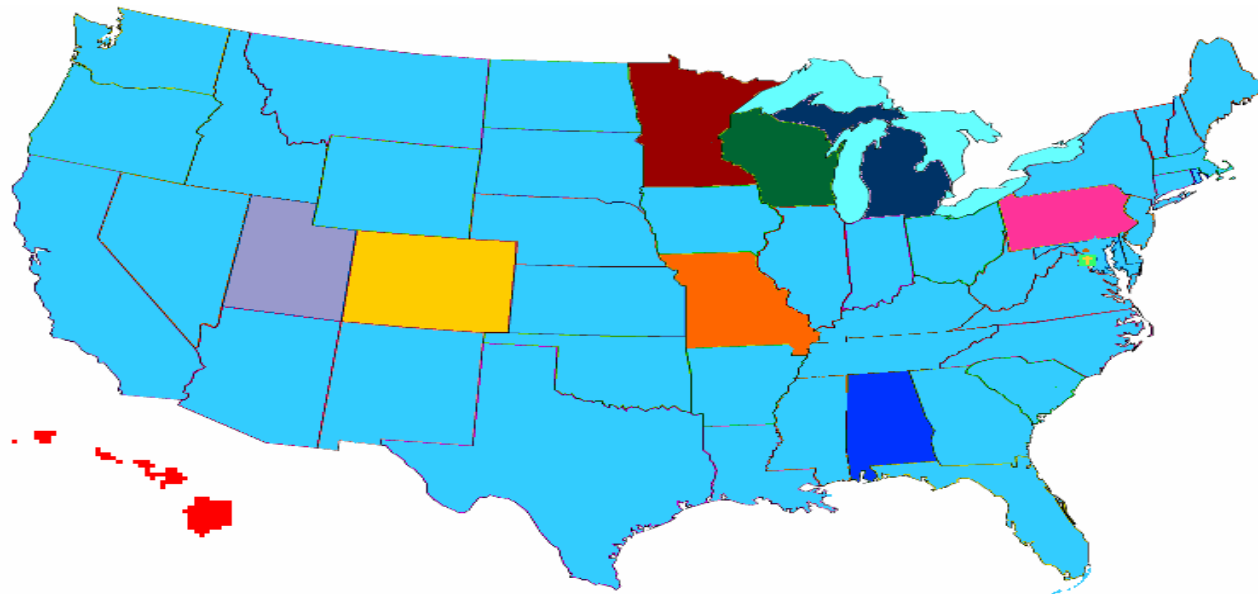


Log/log quantile plot of p value (observed)



PLCO Recruitment Sites

Opportunity to look at geographic differences



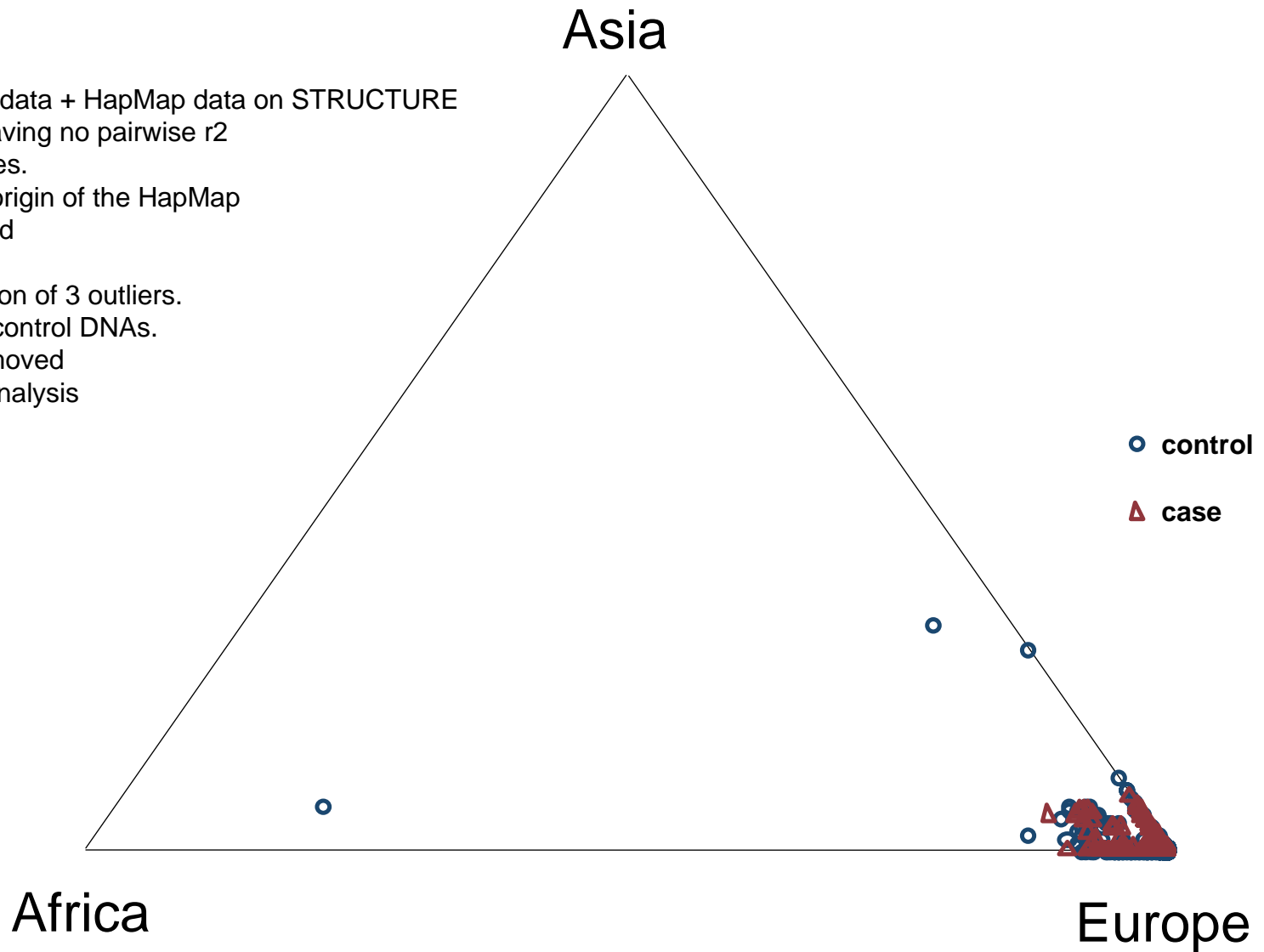
Admixture coefficient in PLCO samples

Method :

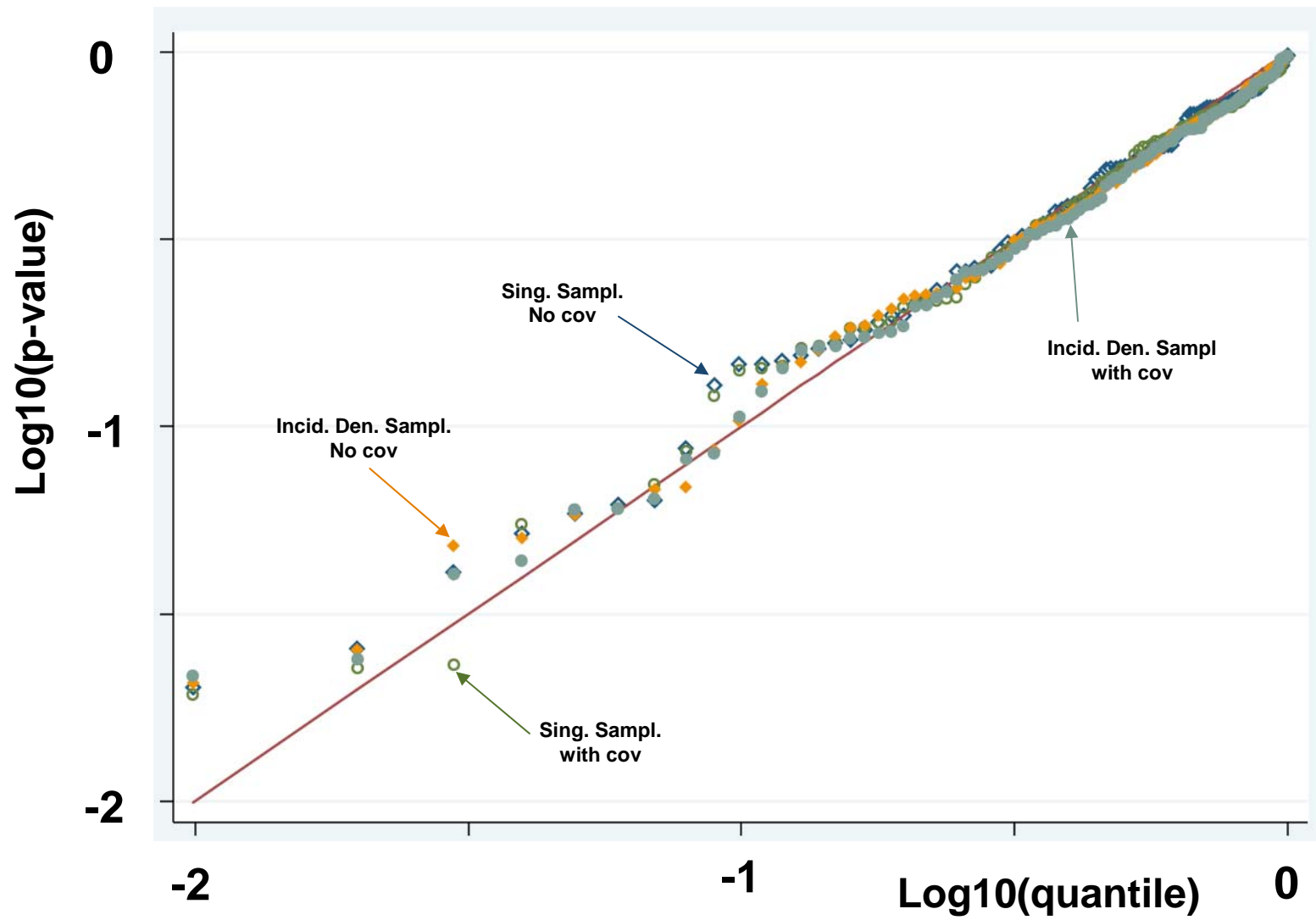
run merged PLCO data + HapMap data on STRUCTURE with 6000 SNPs having no pairwise r^2 and high F_{ST} values. The population of origin of the HapMap samples is specified

Result :

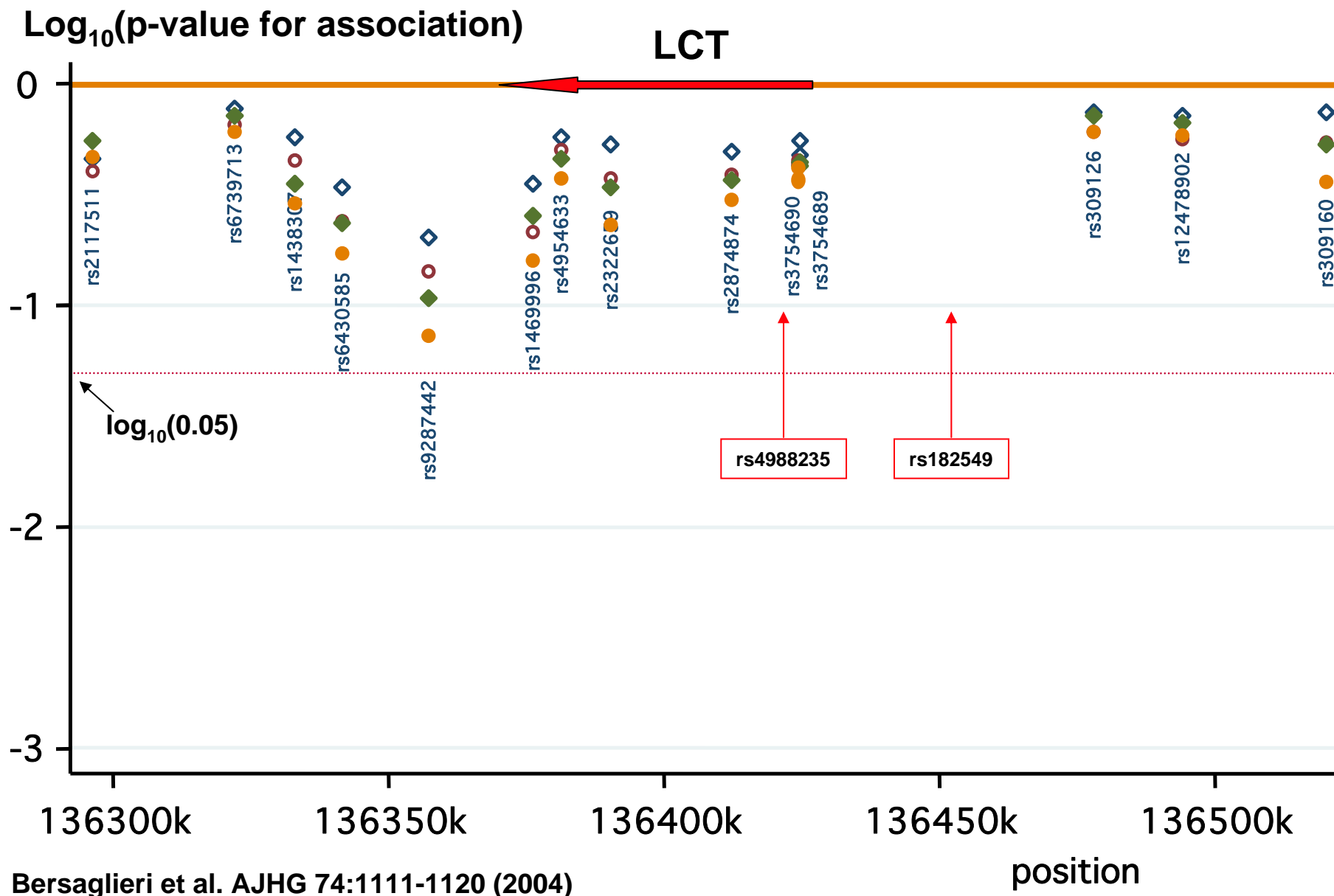
Reliable identification of 3 outliers. They are all three control DNAs. and have to be removed from subsequent analysis



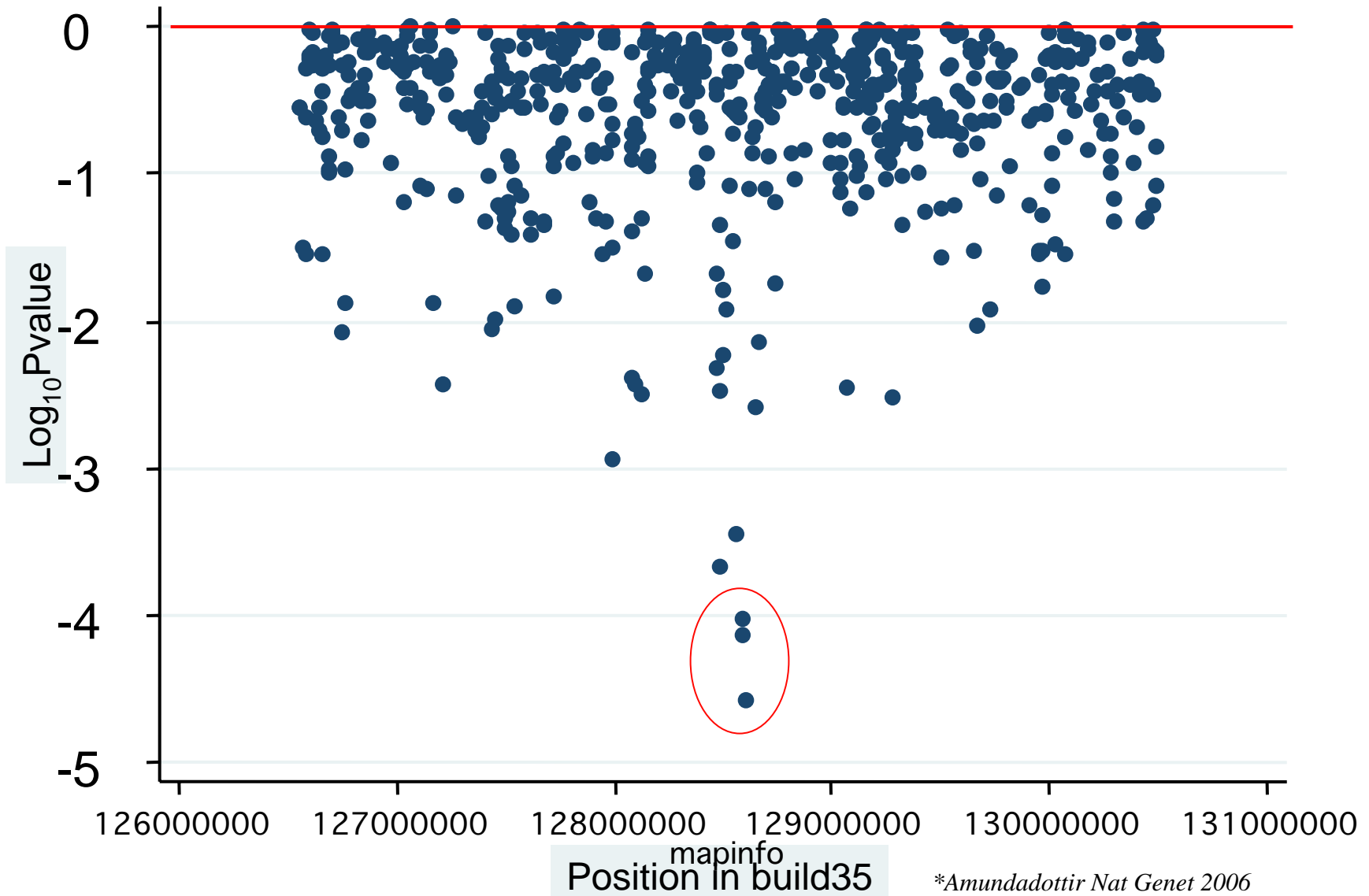
Log-Log quantile plot for p-values of 101 SNPs that differentiate the populations of South and North of Europe



Lactase region



Log₁₀Pvalue 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

Characteristics of the SNPs demonstrating the strongest signal of association in 8q24

	position rsnumber (b.35)	MAF	Pval	HW controls	completion rate	
298.	rs4242382 128586755	.14	.7604	1	1	
299.	rs7017300 128594450	.18	.1629	1	1	
300.	rs7837688 128608542	.14	.8663	.999		
301.	rs1447295 128554220	.14	.6012			

1

Linkage disequilibrium (r^2) with **rs1447295** of the SNPs demonstrating the strongest signal of association

rs#	position (b. 35)	rs1447295	r2 with passoc
rs4242382	128586755	.94	.00007
rs7017300	128594450	.71	.00009
rs7837688	128608542	.84	.00003
rs1447295	128554220	-	.0003

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 (~6500 cases/controls)
3. Discovery of 1 and perhaps 2 additional loci

Value-Added Analysis CGEMS



Opportunity to investigate

- Gene:environment
 - Covariates: BMI, smoking, serum levels
- Multi-SNP Analysis
- Gene:gene interactions
 - Explore pathways
- Follow-up in cohort studies in CGEMS

<http://cgems.cancer.gov>



National Cancer Institute

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



CGEMS

Cancer Genetic Markers of Susceptibility

[Contact Us](#)

[Site Map](#)

[Search](#)



Division of Cancer
Epidemiology
and Genetics

[About CGEMS](#)

[News & Announcements](#)

[Resource Room](#)

[CGEMS Intranet](#)



Cancer Genetic Markers of Susceptibility Project

The Cancer Genetic Markers of Susceptibility (CGEMS) is a three-year, \$14 million initiative that will identify genetic alterations that make people susceptible to prostate and breast cancer. Scientists involved will use DNA available from five large studies of prostate cancer and five large studies of breast cancer to “scan” the genome for common genetic variations between patients who have these cancers and controls who do not have cancer.

[Learn more >>](#)

[Background](#)

Spotlight

[Cancer Genetic Markers of Susceptibility \(CGEMS\)](#)

February 13, 2006

NCI begins studies to identify genetic risk factors for prostate and breast cancer. [more](#)

[DCEG and CGF Collaborate on CGEMS Initiative](#)

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**

Association Finding

CGEMS Search Associations

[Login](#)

- [About](#)
- [Browse Data](#)
- [Bulk Data Download](#)
- [Feedback](#)

Study: CGEMS Prostate Scan 1

Search Association Findings by:

Analysis Name:

Analysis Abstract:

[?](#)

AND

Genomic Location: Chromosome

(Based on Genome build 35) From bp

To bp

HUGO Gene Symbol List
(Fill in new line separated list or upload a list file with max. 100 gene symbols via Browse... button)

Browse...

dbSNP Identifier List
(Fill in new line separated list or upload a list file with max. 1000 dbSNP ids via Browse... button)

Browse...

AND

p-value <=

Whole Genome Rank <=

Association Finding Report

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

 **CGEMS**
Cancer Genetic Markers of Susceptibility

 CORE GENOTYPING FACILITY  OFFICE OF CANCER GENOMICS Division of Cancer Epidemiology and Genetics

CGEMS SNP Association Finding Report

[Login](#)

- [About](#)
- [Browse Data](#)
- [Bulk Data Download](#)
- [Feedback](#)

Study: CGEMS Prostate Scan 1

SNP Association Finding Report - (38 findings)

dbSNP ID	Chromosome	Physical Position (bp)	Associated Genes	Analysis Name	p-value	Whole Genome Rank
rs2803291	1	1924487	KIAA1751	score test	0.90626	273931
rs4648592	1	1833196	GNB1	score test	0.042421	12776
rs6603793	1	1590522	HSPC182	score test	0.475174	143777
rs6675798	1	1216520	B3GALT6 Cab45 LOC388581	score test	0.98977	298748
rs6681938	1	1813382	GNB1	score test	0.200469	60571
rs11721	1	1192554	B3GALT6 TNFRSF18 TNFRSF4 Cab45	score test	0.951451	287492

Population Frequency Report

EMS SNP Population Frequency Report

[Login](#)

[About](#)
[Browse](#)
[Data](#)
[Bulk Data](#)
[Download](#)
[Feedback](#)

Study: CGEMS Prostate Scan 1

SNP Frequency Report - (39 findings)

dbSNP ID	Chromosome	Physical Position (bp)	Associated Genes	Population	Completion Rate (N/M)	Hardy Weinberg pValue	Allele	Allele Count (Frequency)	Genotype	Genotype Count (Frequency)
rs12726255	1	1089873		CONTROL	100 (1104/1104)	0.14233	A G	1890(0.856) 318(0.144)	AA AG GG	815(0.738) 260(0.236) 29(0.026)
rs6687776	1	1070488	FLJ20584	CONTROL	100 (1104/1104)	0.595514	C T	1830(0.829) 378(0.171)	CC CT TT	761(0.689) 308(0.279) 35(0.032)
rs4970357	1	1116987	LOC254099	CONTROL	100 (1104/1104)	0.066365	A C	1985(0.899) 223(0.101)	AA AC CC	898(0.813) 189(0.171) 17(0.015)
rs2474460	1	1876208	CAGLP LOC339456 LOC339457	CONTROL	99.9094 (1103/1104)	0.809619	G A	1125(0.51) 1081(0.49)	GG GA AA	289(0.262) 547(0.496) 267(0.242)
rs4970420	1	1146396	FLJ36119	CONTROL	99.3659 (1097/1104)	0.771619	G A	1770(0.807) 424(0.193)	GG GA AA	712(0.649) 346(0.315) 39(0.036)
rs7531583	1	1738322	FLJ13052	CONTROL	99.0036 (1093/1104)	0.618021	G A	1663(0.761) 523(0.239)	GG GA AA	629(0.575) 405(0.371) 59(0.054)
rs3820011	1	1920355	KIAA1751	CONTROL	99.8188 (1102/1104)	0.153441	G T	1592(0.722) 612(0.278)	GG GT TT	565(0.513) 462(0.419) 75(0.068)

Acknowledgements



NCI

Gilles Thomas
Robert Hoover
Joseph Fraumeni
Daniela Gerhard
Kevin Jacobs
Zhaoming Wang
Meredith Yeager
Robert Welch
Richard Hayes
Sholom Wacholder
Nilanjan Chatterjee
Kai Yu
Margaret Tucker
Marianne Rivera-Silva

HSPH

David Hunter
Peter Kraft

ACS

Heather Feigelson
Carmen Rodriguez
Eugene Calle
Michael Thun