National Cancer Institute

Cancer Genetic Markers of Susceptibility: GWAS & Data Availability

Stephen Chanock, M.D.
Chief, Laboratory of Translational Genomics, DCEG, NCI
Director, Core Genotyping Facility, DCEG NCI

January 24, 2008

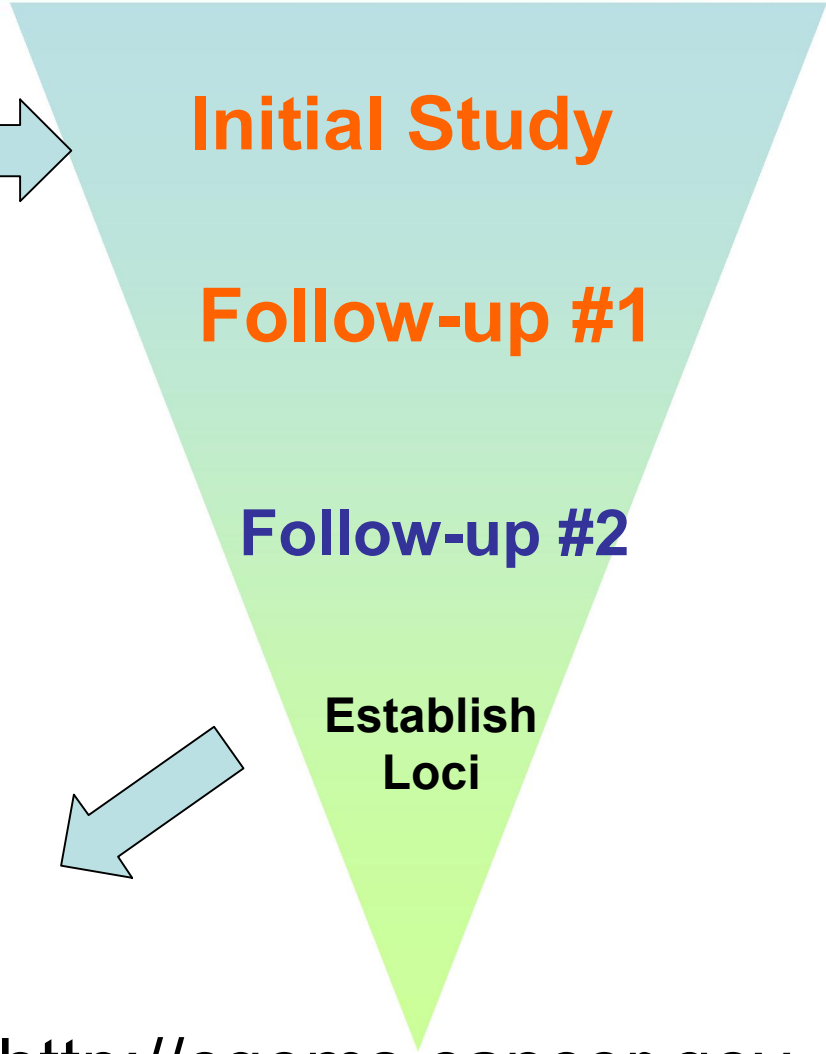
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
National Institutes of Health

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)

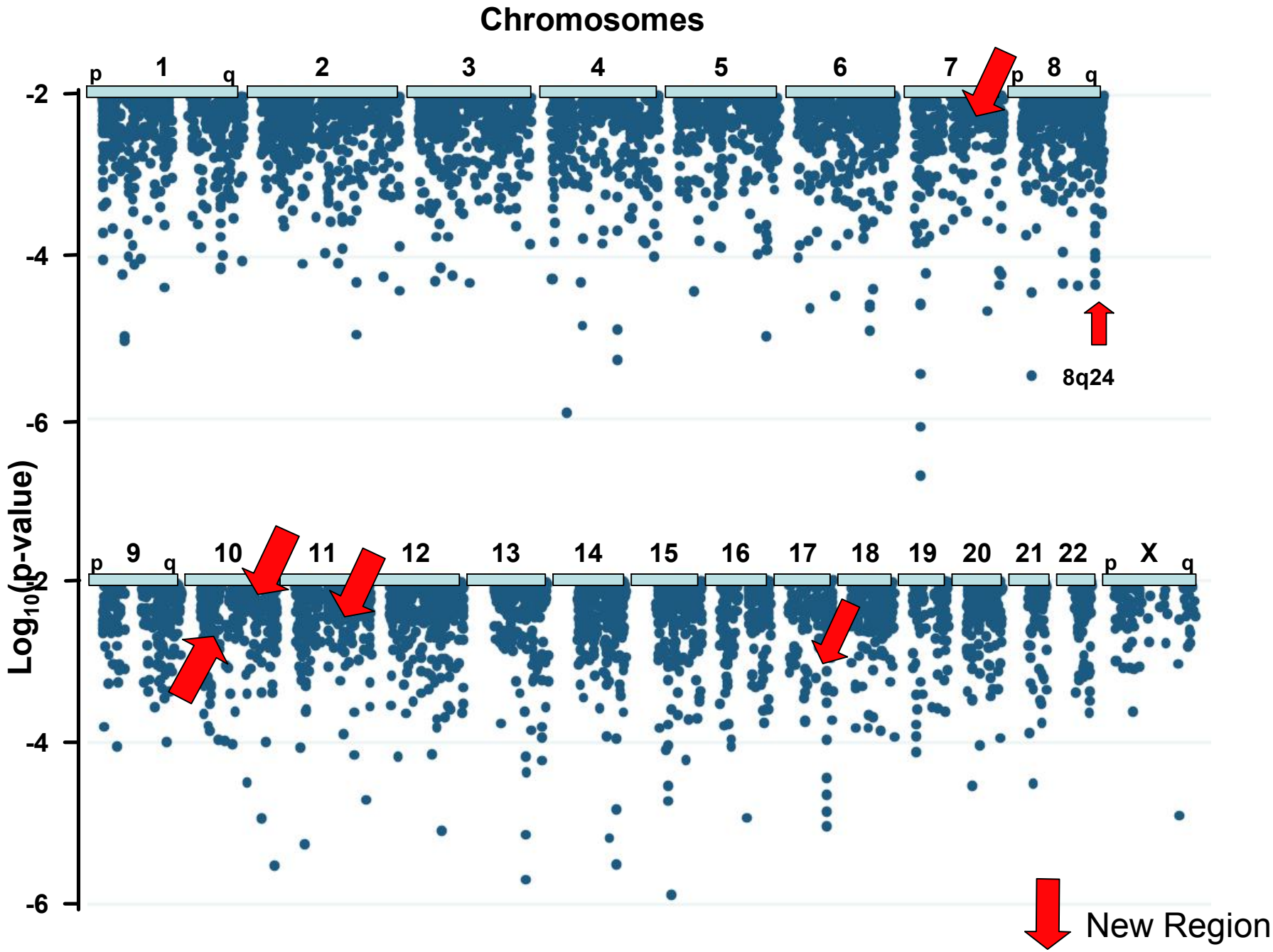
Analyze Long-Term Studies
- NCI PLCO Study
- Nurses' Health Study



Fine Mapping
Functional Studies
Validate Plausible Variants
Possible Clinical Testing

http://cgems.cancer.gov

CGEMS Prostate Cancer GWS



incidence density sampling

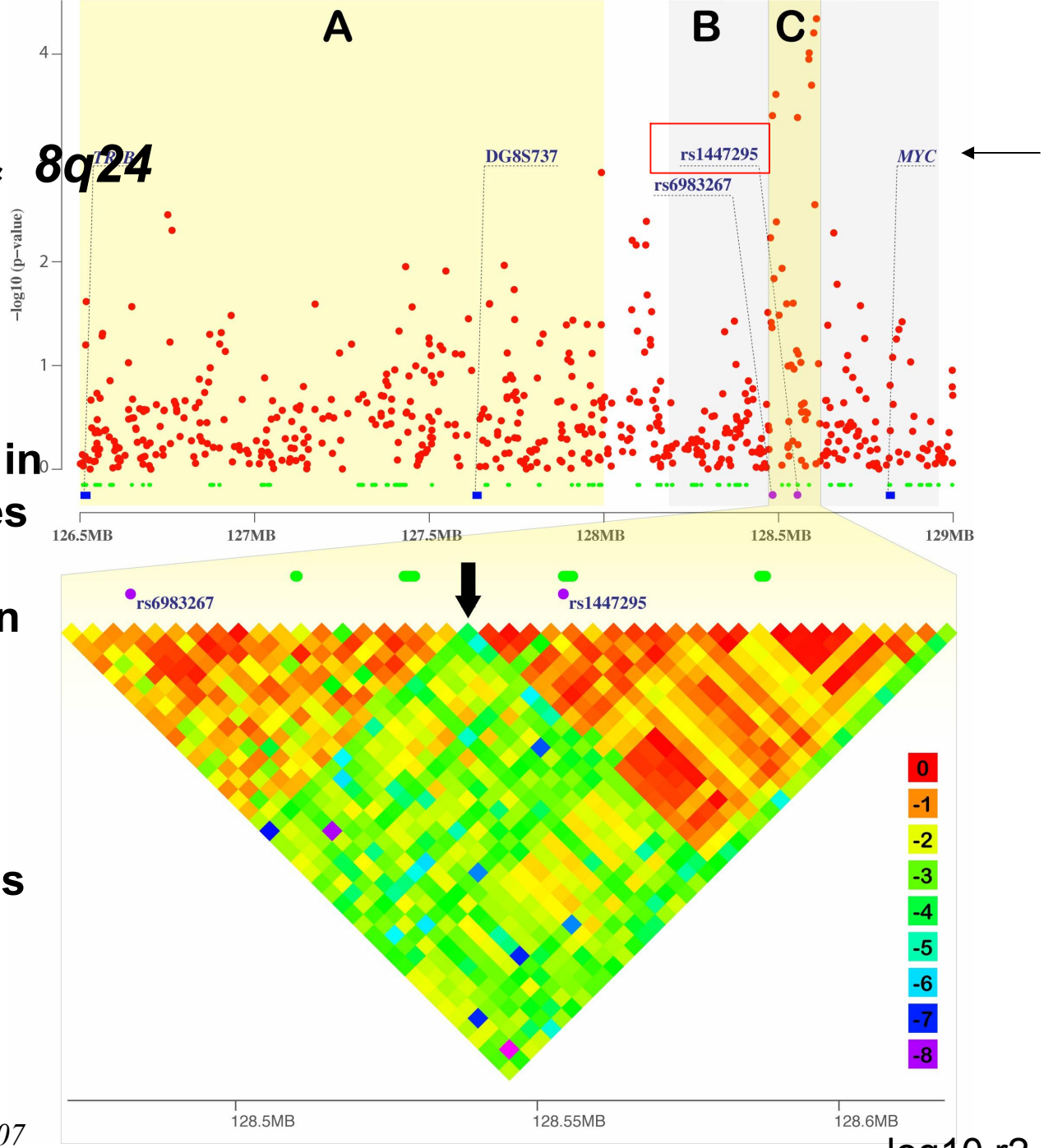
Prostate Cancer & 8q24

Rs1447295 replication in BPC3 & 3 Other Studies

Commonly Amplified in Prostate tumors

"Gene-poor region"

GWAS- multiple signals



Yeager et al Nat Genet 39:645-649, 2007

Replication Studies in CGEMS Prostate Cancer GWAS

	Subjects		rs6983267 Predisposing allele frequency		rs6983267 P-value	rs1447295 Predisposing allele frequency		rs1447295 P-value
			Cases	Cont.		Cases	Cont.	
PLCO	1157	1172	0.55	0.49	2.4×10^{-05}	0.14	0.10	9.8×10^{-05}
ACS	1151	1150	0.55	0.50	3.2×10^{-03}	0.12	0.08	2.7×10^{-05}
ATBC	896	894	0.57	0.51	1.9×10^{-03}	0.21	0.17	2.9×10^{-02}
FPCC	459	455	0.56	0.51	1.2×10^{-01}	0.12	0.07	4.4×10^{-03}
HPFS	636	625	0.57	0.51	1.0×10^{-02}	0.13	0.09	2.7×10^{-03}
ALL	4299	4296	0.56	0.50	9.4×10^{-13}	0.15	0.11	1.5×10^{-14}

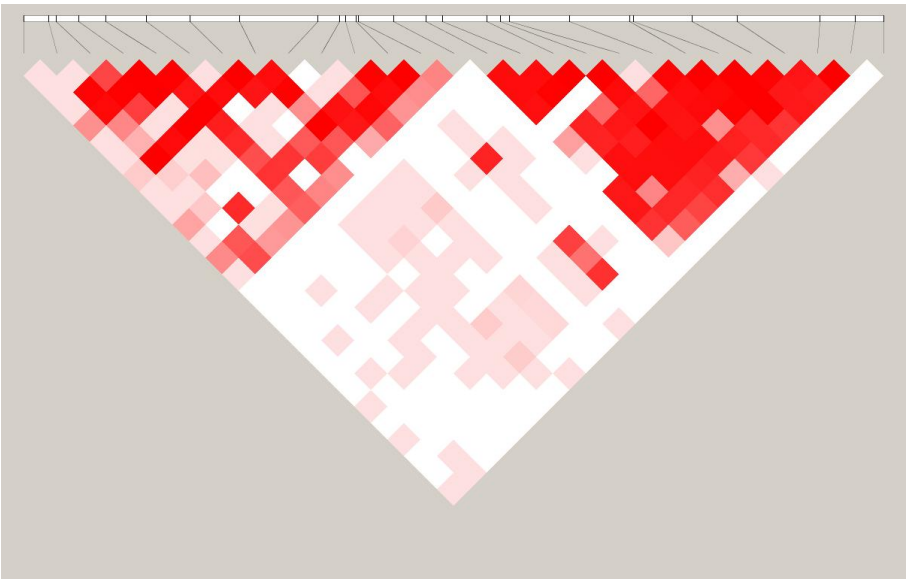
Estimated Odds Ratios Overall

Heterozygotes	1.26	1.43
Homozygotes	1.58	2.23

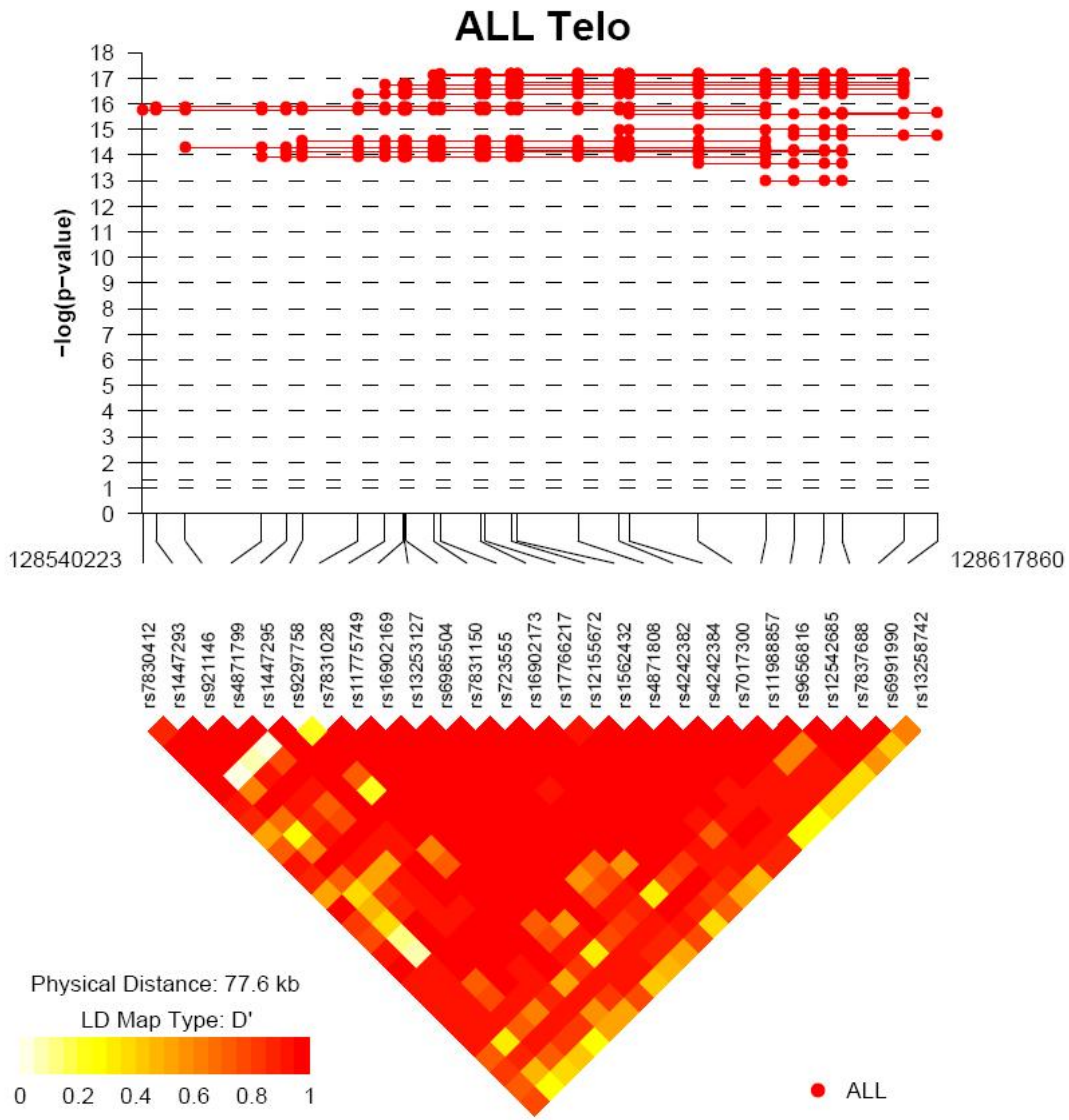
Population Attributable Risk of Prostate Cancer with 8q24 Loci in Caucasians

	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

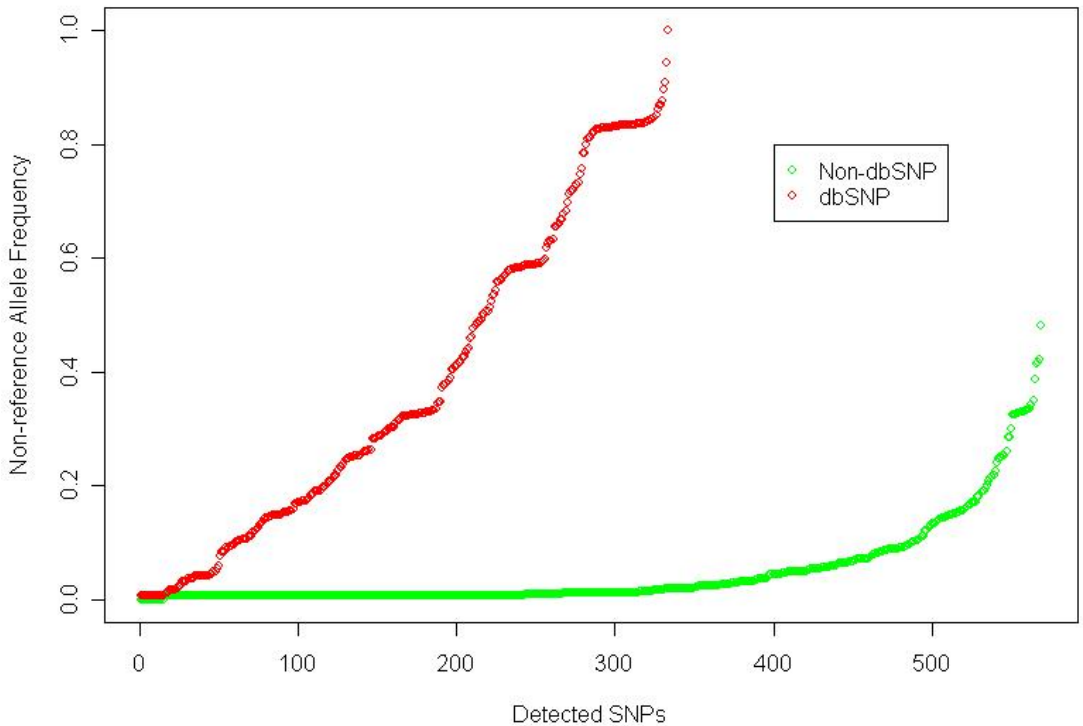
rs6983267 G: 21% **rs1447295 A: 7%**



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


ALL Telo
-log(p-value)
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
128540223
128617860
rs7830412
rs1447293
rs921146
rs4871799
rs1447295
rs9297758
rs7831028
rs11775749
rs16902169
rs13253127
rs6985504
rs7831150
rs723555
rs16902173
rs17766217
rs12155672
rs1562432
rs4871808
rs4242382
rs4242384
rs7017300
rs11988857
rs9656816
rs12542685
rs7837688
rs6991990
rs13258742
Physical Distance: 77.6 kb
LD Map Type: D'
0
0.2
0.4
0.6
0.8
1
ALL

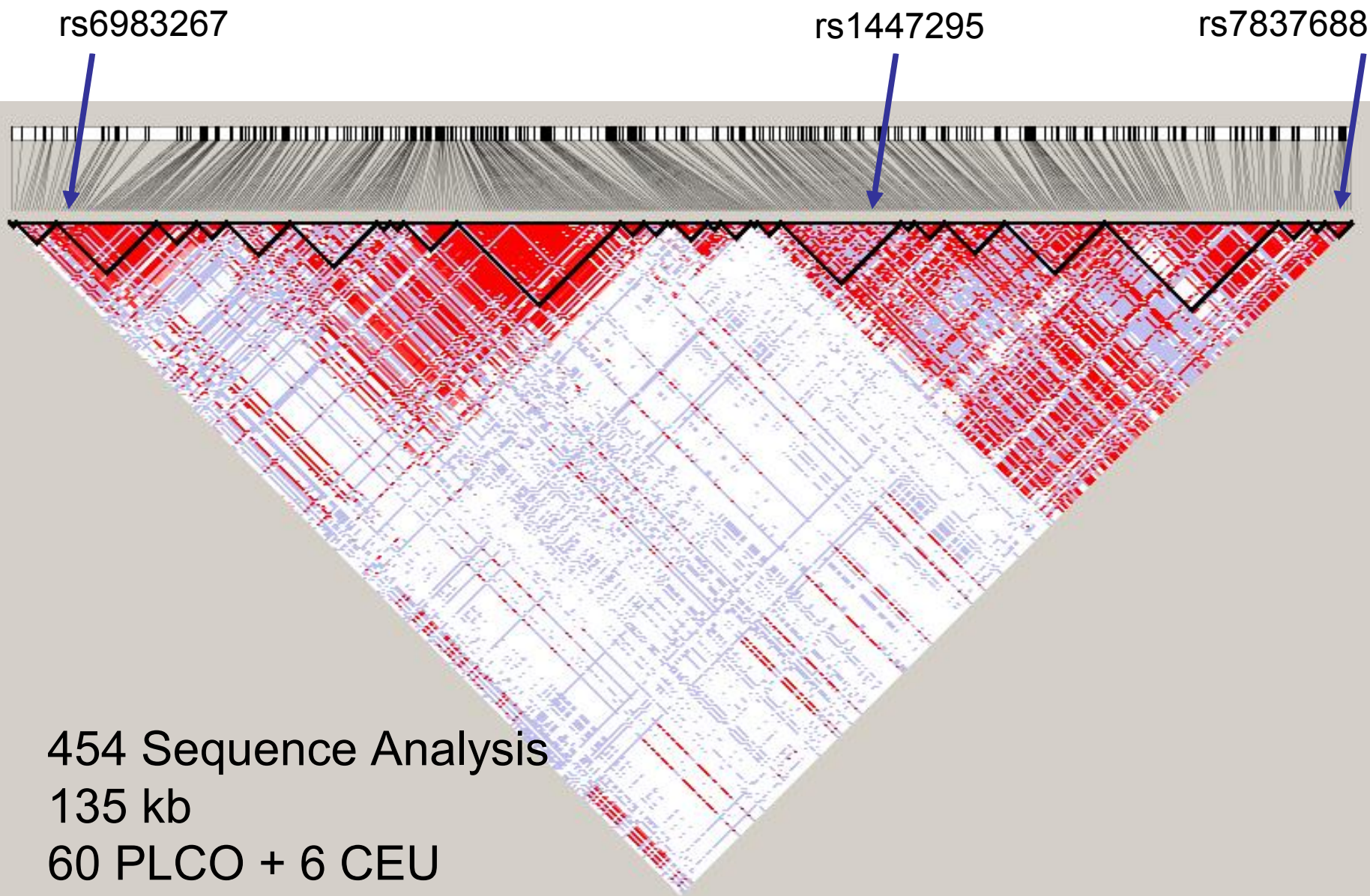
Discovery of Novel SNPs with Complete Sequence Analysis of 85 Individuals over 135 kb of 8q24



Completion per locus > 50% (n = 903) is ~94%
Overall concordance with known data is 99.4%
HapMap = 99.3%, GWAS=99.43%

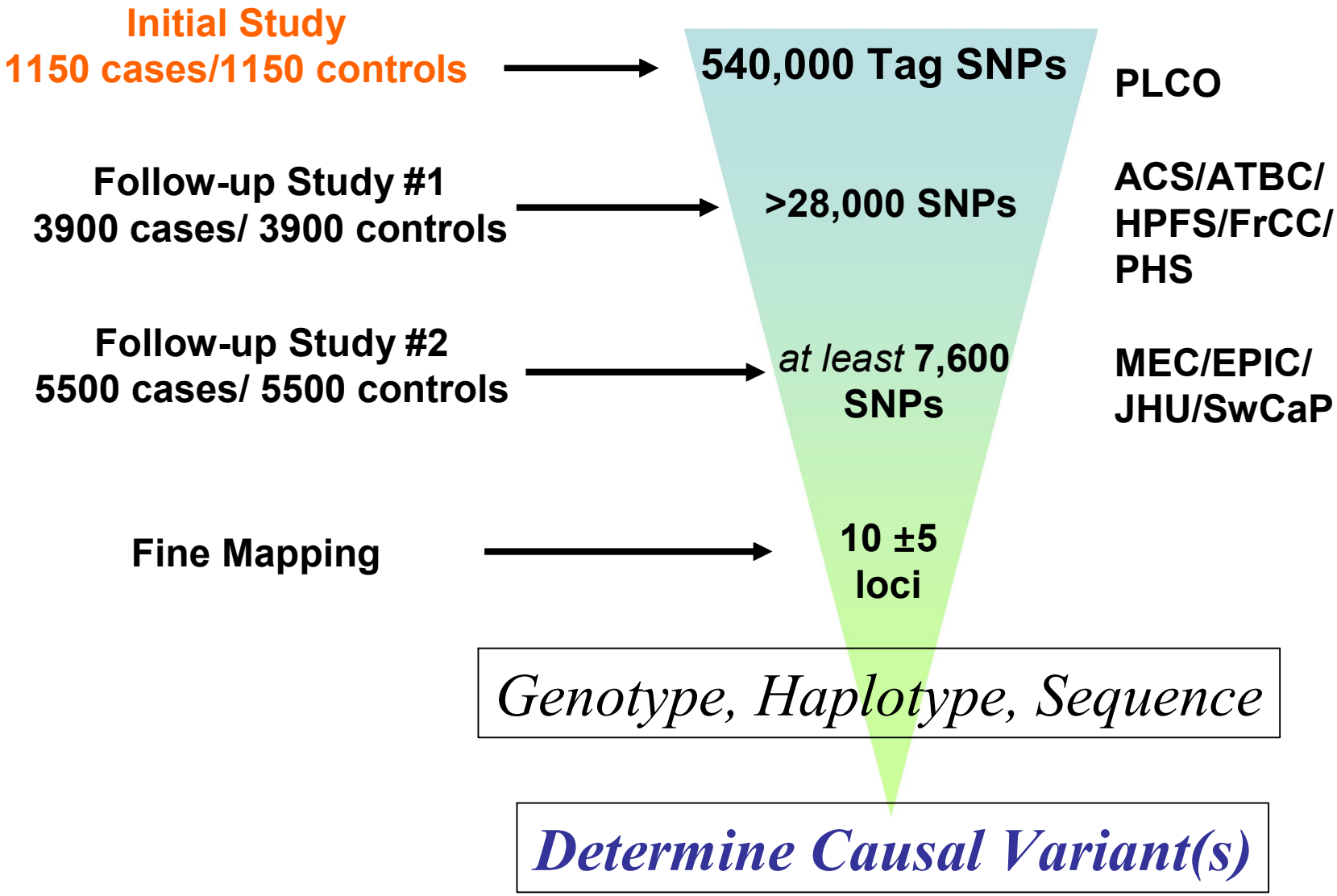
454 Long Range Targeted

40 PLCO Cases/40 cases + 5 HapMap CEU
chromosome 8 from 128473000 to 128609802



454 Sequence Analysis
135 kb
60 PLCO + 6 CEU

General Strategy for Prostate GWAS



Results of Replication #1

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

7 associated loci

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)
17q21	$1.5\ 10^{-10}$	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)

CGEMS Prostate Scan: Follow-up #1

Table 1. Results from the pooled trichotomous association analysis of 2,109 individuals with non-aggressive prostate cancer, 2,651 aggressive prostate cancer cases, and 5,133 controls.

					Non-aggressive[a] vs. control		Aggressive[b] vs. control		Initial GWAS[c]	
Region[f]	CHR	LOC[g]	X^{2h}	p	HET OR (95% CI)	HOM OR (95% CI)	HET OR (95% CI)	HOM OR (95% CI)	Rank	p
8q24	8	128586755	92.98	3.07E-19	1.41 (1.24-1.60)	1.86 (1.27-2.72)	1.66 (1.47-1.87)	2.22 (1.51-3.26)	116	1.12E-04
8q24	8	128482487	58.31	6.58E-12	0.79 (0.70-0.90)	0.64 (0.55-0.74)	0.78 (0.69-0.87)	0.66 (0.57-0.75)	300	3.92E-04
HNF1B	17	33172153	47.97	9.58E-10	0.72 (0.64-0.82)	0.66 (0.57-0.77)	0.85 (0.76-0.96)	0.72 (0.62-0.83)	384	5.21E-04
MSMB	10	51219502	62.85	7.31E-13	1.24 (1.10-1.39)	1.66 (1.42-1.95)	1.16 (1.04-1.29)	1.57 (1.36-1.81)	24,223	0.042
11q13	11	68751243	46.70	1.76E-09	0.78 (0.69-0.88)	0.65 (0.56-0.76)	0.91 (0.81-1.02)	0.71 (0.62-0.82)	2,439	0.004
CTBP2	10	126686862	37.12	1.70E-07	1.20 (1.07-1.34)	1.63 (1.33-1.99)	1.17 (1.05-1.30)	1.46 (1.22-1.76)	319	4.09E-04
JAZF1	7	27749803	31.76	2.14E-06	0.74 (0.66-0.83)	0.71 (0.55-0.90)	0.89 (0.80-0.98)	0.84 (0.67-1.05)	24,407	0.042

Thomas et al *Nature Genetics* in press

Looking deeper into the Replication: Promises of more loci?

Results from the pooled trichotomous association analysis

							Non-aggressive[a] vs. control		Aggressive[b] vs control		Initial GWAS[a]	
dbSNP ID[1]	Risk allele[1] (Freq)	LOCUS[5]	CHR	LOC[6]	X^2	p	HET OR (95% CI)	HOM OR (95% CI)	HET OR (95% CI)	HOM OR (95% CI)	Rank	p
rs4242382	A (0.12)	8q24	8	128867555	89.13	3.07E-19	1.38 (1.22-1.56)	1.58 (1.09-2.30)	1.65 (1.46-1.86)	2.39 (1.62-3.52)	116	1.12E-04
rs6983267	G (0.53)	8q24	8	128482487	57.98	6.58E-12	0.80 (0.71-0.90)	0.64 (0.55-0.74)	0.78 (0.69-0.88)	0.66 (0.57-0.76)	300	3.92E-04
rs4430796	A (0.54)	HNF1B	17	33172153	52.14	9.58E-10	0.72 (0.64-0.81)	0.66 (0.57-0.76)	0.85 (0.76-0.95)	0.71 (0.61-0.82)	384	5.21E-04
rs10993994	T (0.40)	MSMB	10	51295012	62.83	7.31E-13	1.21 (1.08-1.35)	1.64 (1.41-1.91)	1.16 (1.04-1.30)	1.58 (1.37-1.83)	24,223	0.042
rs10896449	G (0.52)	11q13	11	68751243	47.58	1.76E-09	0.79 (0.70-0.89)	0.65 (0.56-0.76)	0.90 (0.80-1.02)	0.71 (0.61-0.82)	2,439	0.004
rs4962416	C (0.27)	ZRANB1/TBP2	10	126868652	35.07	1.70E-07	1.16 (1.04-1.29)	1.60 (1.31-1.95)	1.17 (1.06-1.30)	1.46 (1.21-1.76)	319	4.09E-04
rs10486567	G (0.77)	JAZF1	7	27749803	33.39	2.14E-06	0.75 (0.67-0.83)	0.69 (0.54-0.88)	0.89 (0.80-0.99)	0.84 (0.67-1.06)	24,407	0.042
rs12771728	C (0.34)	10q23	10	89345293	30.84	1.01E-05	1.10 (0.99-1.23)	0.94 (0.78-1.12)	0.90 (0.81-1.00)	1.28 (1.09-1.50)	59,552	0.106
rs4961199	A (0.16)	CPNE3, CNR3	8	87650060	30.25	1.26E-05	1.18 (1.05-1.33)	1.88 (1.39-2.55)	0.98 (0.87-1.10)	1.74 (1.28-2.36)	7,464	0.012
rs4782726	A (0.18)	CDH13	16	81258834	27.95	2.50E-05	0.96 (0.85-1.07)	1.14 (0.85-1.54)	0.92 (0.82-1.03)	1.84 (1.43-2.37)	25,462	0.044
rs4072111	A (0.11)	IL16	15	79366194	27.75	1.19E-05	0.91 (0.80-1.03)	0.90 (0.52-1.54)	0.97 (0.86-1.10)	2.71 (1.75-4.20)	10,555	0.018
rs6982080	G (0.69)	8p21	8	29688861	27.55	2.28E-05	1.06 (0.95-1.18)	0.65 (0.53-0.78)	1.00 (0.90-1.11)	0.81 (0.68-0.97)	17,710	0.031

Heterogeneity in Signal: BCAC & CGEMS (NHS)

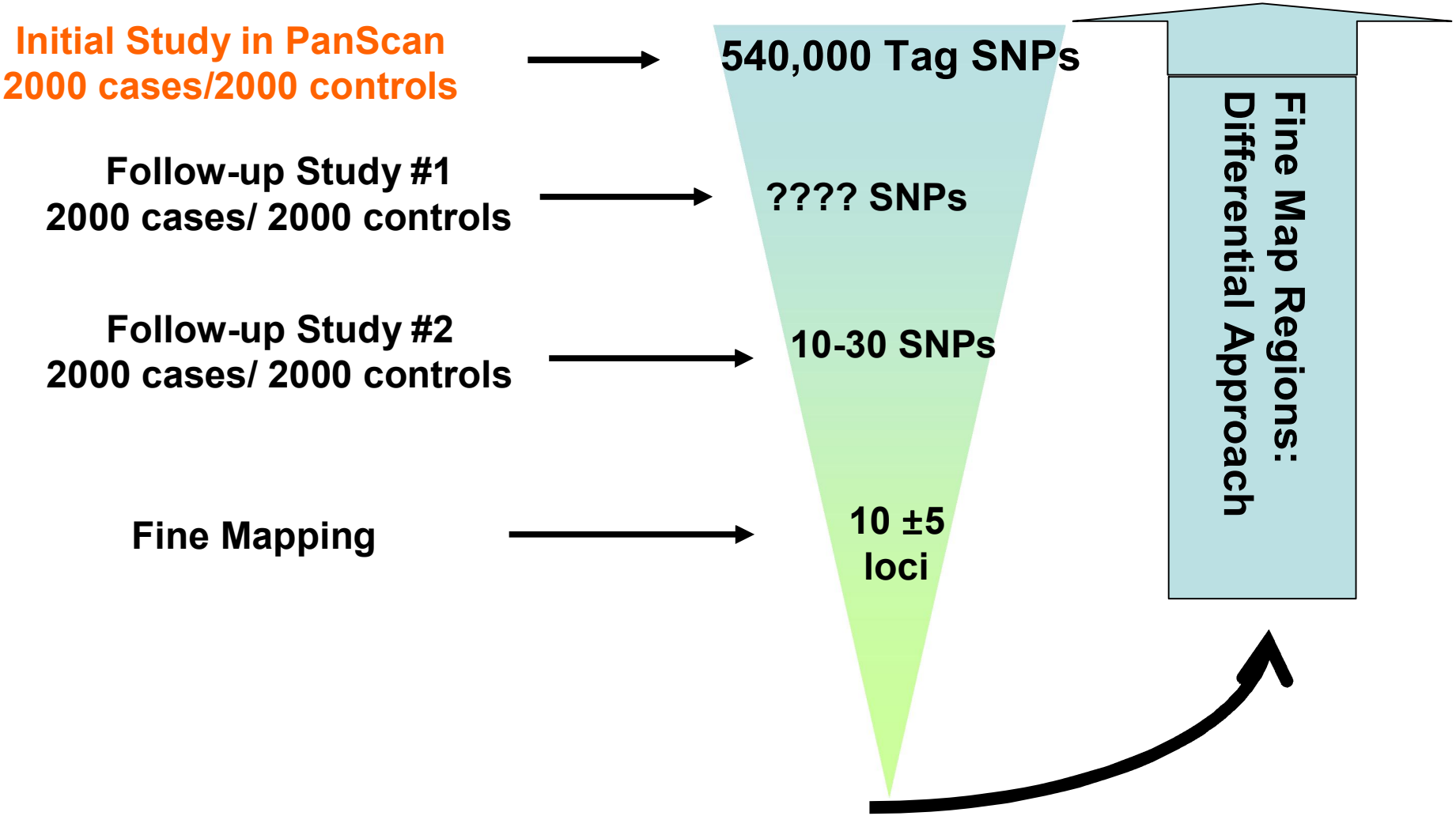
BCAC*- best hits CGEMS

SNP	GENE	CHR	TYPED IN CGEMS?	BEST IN 100kb REGION	p	RANK IN SCAN
rs1219648	**FGFR2**	**10**	**2.00E-06**			**1**
rs889312	MAP3K1	5	no	rs726501	0.012	6,340
rs3817198	LSP1	11	0.51	-	-	269,442
rs2107425	H19	11	no	rs217228	0.030	14,344
rs13281615		8	no	rs10098985	0.017	12,048
rs981782		5	no	rs4866929	7.30E-05	33
rs30099		5	0.88	-	-	462,781
rs4666451		2	no	rs12710697	0.046	24,741
rs3803662	TNRC9/LOC643714	16	0.05	-	-	27,025

Easton Nature 2007
3 Stage Design
>28,000 cases/26,000 controls

http://cgems.cancer.gov
1145 cases/1142 controls

General Strategy for Pancreatic Cancer GWAS



GWAS at DCEG/NCI

Initial

Breast Cancer

Prostate Cancer

Pancreatic Cancer

Bladder Cancer

Lung Cancer

Follow-up

Breast Cancer
- CGEMS
- ER-negative BPC3

Prostate Cancer
- CGEMS
- Advanced Cancer BPC3

Bladder Cancer

Lung Cancer

Pancreatic Cancer

Colorectal Cancer

Kidney Cancer

Non-Hodgkin Lymphoma

Ovarian Cancer

Display of 'Pre-computed Analyses"
Request for Access to Raw Genotype Data
Results of Papers
Materials/Methods as per Nature 2007
New Tools

http://cgems.cancer.gov

NATIONAL CANCER INSTITUTE

CGEMS: caBIG Posting Pre-Computed Analysis

Pre-computed Analysis

No Restrictions

Raw Genotype

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

Registered Access

- **SF424**
- **Data Use Certificate**

http://cgems.cancer.gov/data

CGEMS
Cancer Genetic Markers of Susceptibility

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

Association Tests
8q24
Scan 1A
~300,000 SNPs

CGEMS SNP Association Finding Report

https://caintegrator.nci.nih.gov/cgems/searchAssociations.do

Viva La Voce Apple .Mac Amazon eBay Yahoo! News

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

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

Study: CGEMS Prostate Cancer WGAS Phase 1A

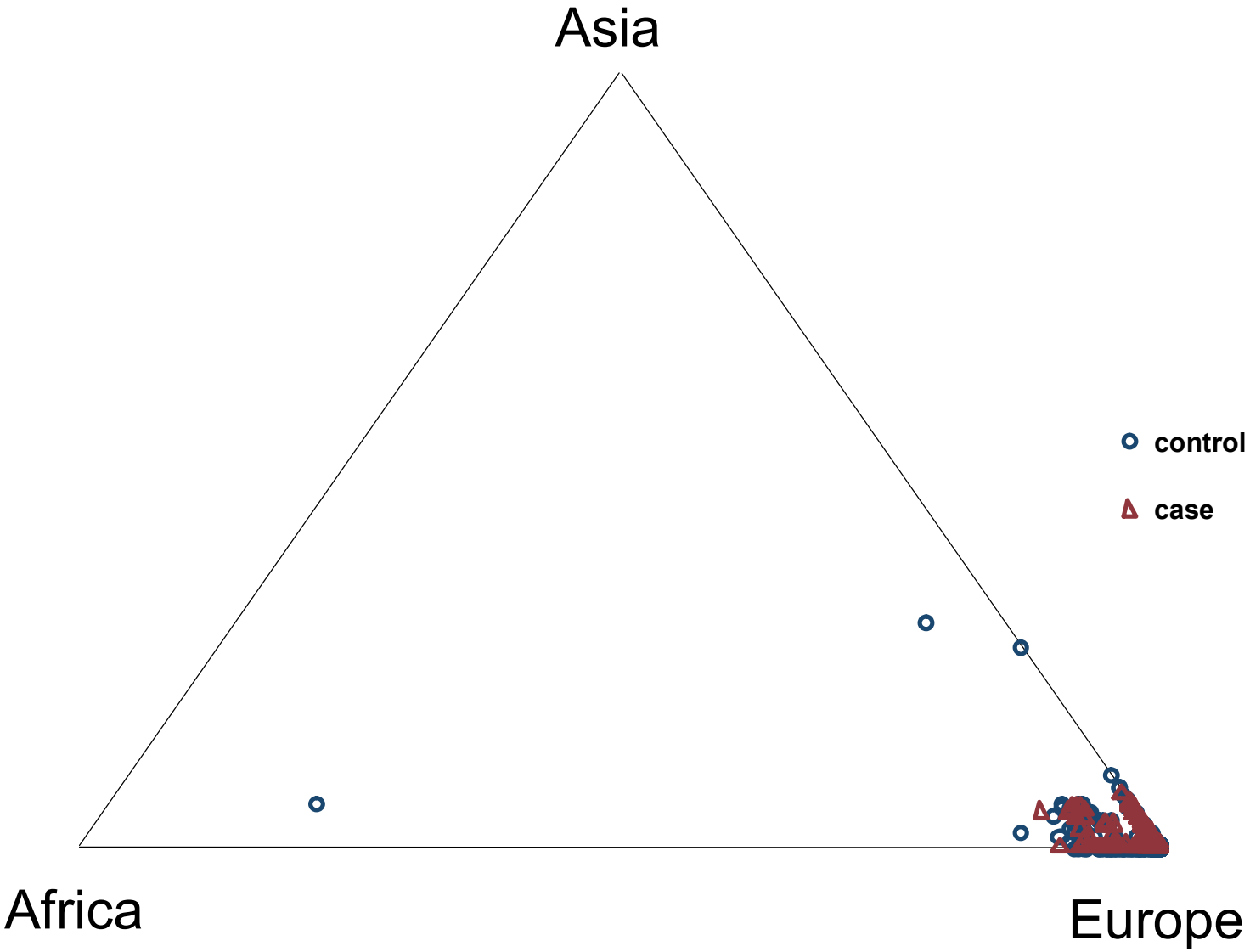
SNP Association Finding Report - (19 results)

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

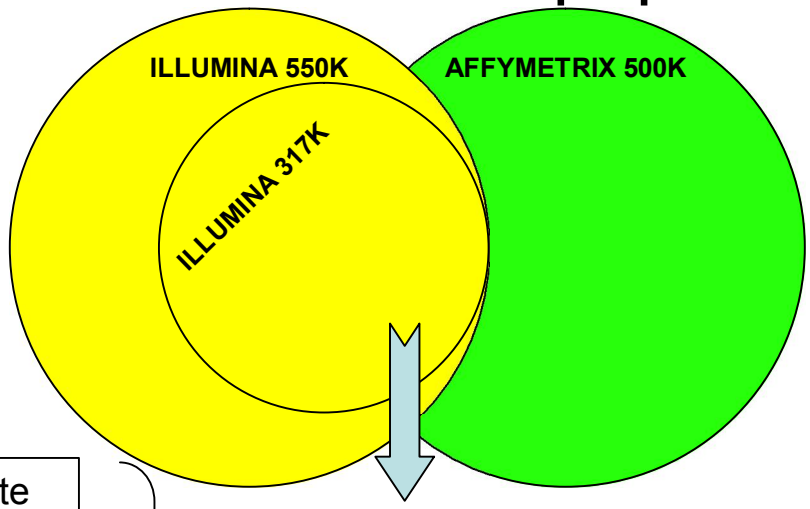
powered by caintegrator

http://cgems.cancer.gov
Available 10/06

Admixture coefficient in PLCO samples: Monitoring Population Structure



Selection of a set of SNPs for population stratification



Remove SNPs with call rate < 90% in either Illumina or Affymetrix

Remove untyped or monomorphic SNPs in YRI or (JPT+CHB)

Remove SNPs with P-values for HW proportion < 0.01

50 374 SNP

40 829 SNP

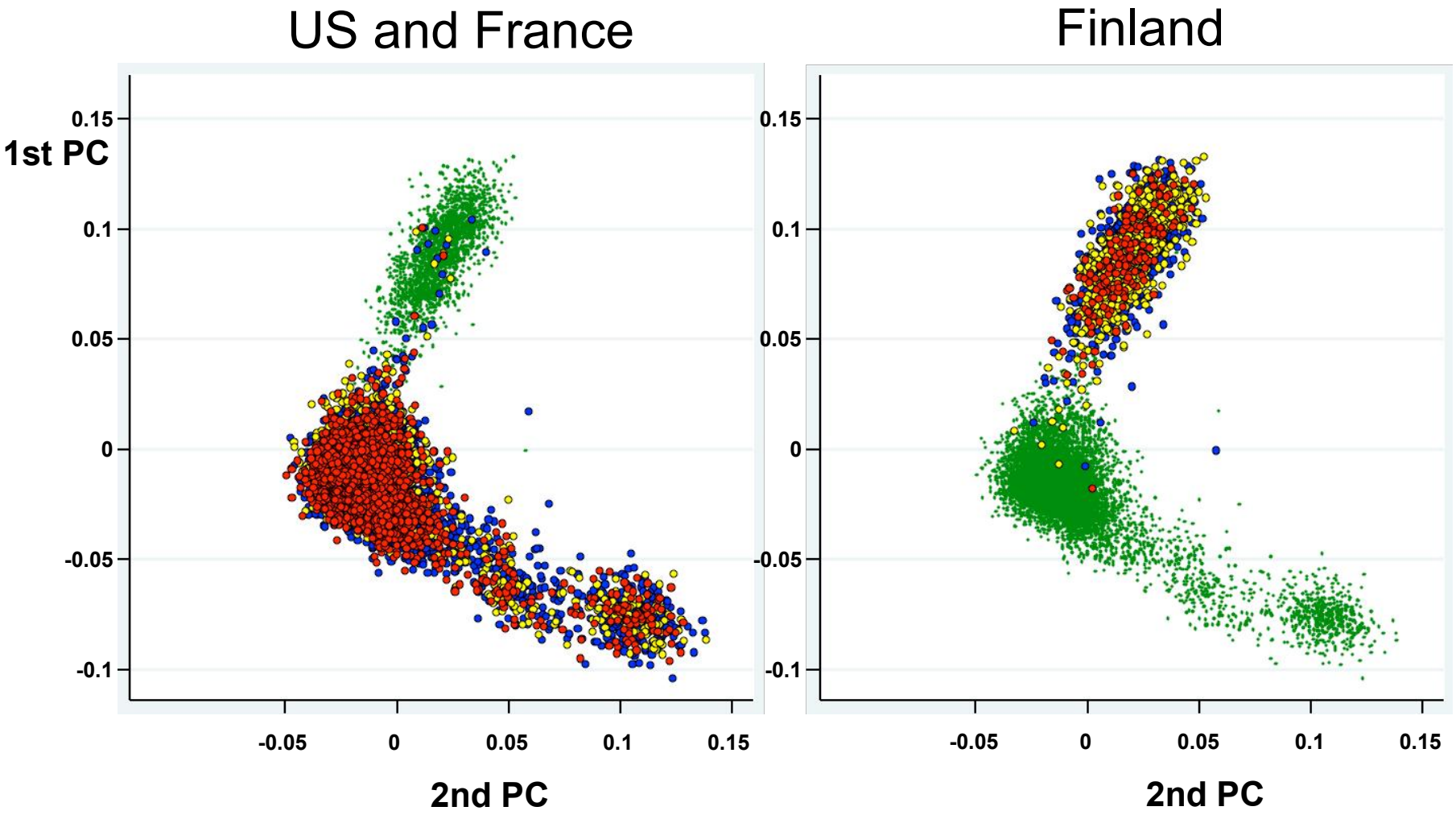
Select a set of SNPs with pairwise $r^2 < 10^{-3}$

10 095 SNPs

These SNPs are expected to provide reliable genotypes and will be included in the SNP set typed in most GWAS.

When ancestral populations known optimization possible :
Pfaff et al. Genetic Epi 26:305-315(2004)

PCA on all Studies.
1st and 2nd PC



Blue = control
Yellow = non-aggressive
Red = aggressive
Green = other studies

CGEMS Data Portal

CGEMS
Cancer Genetic Markers of Susceptibility

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

Home | Login | Contact Us | Visit the CGEMS Home Site

CGEMS/data portal
Cancer Genetic Markers of Susceptibility

Core Genotyping Facility | Office of Cancer Genomics | Division of Cancer Epidemiology and Genetics | caBIG

About CGEMS Data | Browse Data | Bulk Data Downloads | Cite Data | Data Access

Browse Data:

Select a Study:

- CGEMS Breast Cancer WGAS (Illumina 550K)
- CGEMS Prostate Cancer Follow-up scan 1
- CGEMS Prostate Cancer WGAS (Illumina 317K and 240K)

Select a Version:

Version 1.0

Study Description:

The first prostate follow-up scan of The Cancer Genetic Markers of Susceptibility (CGEMS) study,which is being conducted to test markers identified as promising in the genome wide scan stage

Select a Dataset:

- (●) Association Finding
- ( ) Population Frequency
- ( ) Subjects Data (login required)

Reset | Submit

powered by cointegrator

Home | Text-Only Version | Contact Us | Policies | Accessibility | Site Map

FIRSTGOV

CGEMS Prostate Follow-up 1 Association Analysis Result Example

CGEMS
Cancer Genetic Markers of Susceptibility

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

Home | Login | Contact Us | Visit the CGEMS Home Site

CGEMS/data portal
Cancer Genetic Markers of Susceptibility

Core Genotyping Facility | Office of Cancer Genomics | Division of Cancer Epidemiology and Genetics | caBIG

About CGEMS Data | Browse Data | Bulk Data Downloads | Cite Data | Data Access

Study: CGEMS Prostate Cancer Follow-up scan 1 Version: Version 1.0

SNP Association Finding Report - (10 results)

Save Results

dbSNP ID	Chromosome	Physical Position (bp)	Associated Genes	Analysis Name	p-value	Whole Genome Rank	Estimated Odds Ratios OR1		Estimated Odds Ratios OR2	
							Heterozygote Risk	Homozygote Risk	Heterozygote Risk	Homozygote Risk
rs4242382	8	128586755		Aggressive and non-aggressive versus control (trichotomous), genotype-specific effect model, adjusted, ALL (ACS, HPFS, FPCC, ATBC, PLCO)	3.065E-19	1	1.406 (1.2367-1.5985)	1.8572 (1.2688-2.7183)	1.6588 (1.4721-1.8691)	2.2226 (1.5135-3.2639)
rs4242384	8	128587736		Aggressive and non-aggressive versus control (trichotomous), genotype-specific effect model, adjusted, ALL (ACS, HPFS, FPCC, ATBC, PLCO)	1.127E-18	2	1.3896 (1.2216-1.5805)	1.8003 (1.2314-2.6319)	1.6561 (1.4693-1.8665)	2.1949 (1.4963-3.2198)
rs1447295	8	128554220		Aggressive and non-aggressive versus control (trichotomous), genotype-specific effect model, adjusted, ALL (ACS, HPFS, FPCC, ATBC, PLCO)	1.362E-17	3	1.3446 (1.1826-1.5287)	1.911 (1.2913-2.8283)	1.6073 (1.4271-1.8101)	2.4448 (1.6553-3.6107)
rs7837688	8	128608542		Aggressive and non-aggressive versus control (trichotomous), genotype-specific effect model, adjusted, ALL (ACS, HPFS, FPCC, ATBC, PLCO)	4.632E-17	4	1.3096 (1.1509-1.4902)	1.8902 (1.2818-2.7875)	1.6136 (1.4318-1.8186)	2.3649 (1.6043-3.4861)
rs11988857	8	128601055		Aggressive and non-aggressive versus control (trichotomous), genotype-specific effect model, adjusted, ALL (ACS, HPFS, FPCC, ATBC, PLCO)	8.69E-16	5	1.3032 (1.1501-1.4768)	1.8985 (1.3481-2.6736)	1.5321 (1.3652-1.7193)	2.1921 (1.5495-3.1013)

CGEMS Data Access Request:

CGEMS
Cancer Genetic Markers of Susceptibility

Data Access Committee
- 9 Voting Federal Employees
- 2 External Study PIs
- 3 Support

Biweekly Meeting

Requires Unanimous Vote

As of 1/24/08:

34 Requests Received

29 Requests Approved with User Accounts Issued

Data Access Request/SF424

CGEMS
Cancer Genetic Markers of Susceptibility

Requirements:

One PI for One Institution
One Signing Official per Institution

Data can not be shared between Institutions

Strict Security Requirements

Compliant with the Trans-NIH GWAS Policy

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

2. DATE SUBMITTED
Applicant Identifier
3. DATE RECEIVED BY STATE
State Application Identifier
4. Federal

1. * TYPE OF SUBMISSION
☐ Pre-application ☐ Application
☐ Changed/Corrected Application

5. APPLICANT INFORMATION
* Organizational DUNS: 04-670-5849
* Legal Name: Looney Tunes University
Department: Genetics
Division: Genetic Quackery
* Street1: 1234 ATCG Way
Street2: Rm 101
* City: Lostin
County:
* State: MD
* ZIP Code: 20877
* Country: USA

Person to be contacted on matters involving this application
Prefix: Mr. * First Name: Daffy Middle Name: D. * Last Name: Duck Suffix: Ph.D
* Phone Number: 555-555-5555 Fax Number: 222-222-2222 Email: laughs@silly.edu

6. * EMPLOYER IDENTIFICATION (EIN) or (TIN):
7. * TYPE OF APPLICANT:

8. * TYPE OF APPLICATION: ☒ New
☐ Resubmission ☐ Renewal ☐ Continuation ☐ Revision
Other (Specify):
Small Business Organization Type
☐ Women Owned ☐ Socially and Economically Disadvantaged

If Revision, mark appropriate box(es).
☐ A. Increase Award ☐ B. Decrease Award ☐ C. Increase Duration
☐ D. Decrease Duration ☐ E. Other (specify)
* Is this application being submitted to other agencies? Yes ☐ No ☐
What other Agencies?

9. * NAME OF FEDERAL AGENCY:
10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER:
TITLE:

11. * DESCRIPTIVE TITLE OF APPLICANT'S PROJECT:
Role of genetic variation in Cartoon Structure

12. * AREAS AFFECTED BY PROJECT (cities, counties, states, etc.)

13. PROPOSED PROJECT:
* Start Date * Ending Date
14. CONGRESSIONAL DISTRICTS OF:
a. * Applicant b. * Project

15. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION
Prefix: Mr. * First Name: Daffy Middle Name: D. * Last Name: Duck Suffix: Ph.D
Position/Title: Principal Investigator * Organization Name: Looney Tunes University
Department: Genetics
Division: Genetic Quackery
* Street1: 1234 ATCG Way
Street2: Rm 101
* City: Lostin
County:
* State: MD
* ZIP Code: 20877
* Country: USA
* Phone Number: 555-555-5555 Fax Number: 222-222-2222 * Email: laughs@silly.edu

OMB Number: 4040-0001
Expiration Date: 04/30/2008

SF424 FOR CGEMS

SUBMIT BY

- Mail
- Fax
- PDF by email

Single request per disease site

National Cancer Institute
Cancer Genetic Markers of Susceptibility
Data Access Request Form

Appendix 1: Project Summary and Statement of Intent

Project Summary

Project Title:

Name of CGEMS data set being requested (Separate requests should be made for each study):

☐ **Breast Cancer (NHS)** ☐ **Prostate Cancer (PLCO)**

Please enter the required Statement of Intent in the area below:
It should be a brief description of the proposed research suitable for dissemination to the public and may include a statement of objectives and methods to be employed. This summary must not include any proprietary/ confidential information. Please limit the summary to <= 200 type-written words.

CGEMS DUC- Required Signature

CGEMS
Cancer Genetic Markers of Susceptibility

The Principal Investigator and this Institution make the following assurances:
The CGEMS dataset will be used solely in connection with the research intent.
The Institution has considered any participant protection issues and agrees that the research can go forward.
No efforts will be made to attempt to identify the participants and their relatives in this Research.
There will be no transfer of restricted CGEMS data to any other investigators/institutions (except research assistants, students, post-doctoral fellows, etc., who answer directly to the PI and are bound by this DUC).
To prevent accidental transfer of restricted CGEMS data, adequate security controls (e.g. no network accessibility, secure lap tops and mass storage devices) are in place.
To immediately any reportable events to the CGEMS DAC in writing: unintentional identification of research participant; unauthorized release of data; accidental compromise of data security.
To respect the pre-competitive nature of CGEMS datasets in considering claims of intellectual property derived from use of these data and to adhere to the tenets of NIH's Best Practices for the Licensing of Genomic Inventions.
To report publications on an annual basis.
To acknowledge the CGEMS dataset in any publication or presentation resulting from the use of the dataset.
By entering name and date, and submitting this form, I certify agreement to the terms and conditions specified in the Data Use Certification.

Requesting Investigator
Name: _____ Title: _____
Signature: _____ Date: _____
Requesting Institution
Institutional Signing Official:
Name: _____ Title: _____
Signature: _____ Date: _____

GLU: Genetics Library Utilities

CGEMS
Cancer Genetic Markers of Susceptibility

- GLU provides command-line tools for
  - the management of large amounts of SNP genotype data
  - check genotype quality
  - test for association between SNP markers and phenotypes
- Designed for large scale GWAS
- Ample for smaller data sets
- TaqMan to mid-size genotyping (Illumina OPA)

http://cgf.nci.nih.gov

GLU: Library

CGEMS
Cancer Genetic Markers of Susceptibility

- Scalable genotype data storage and management
- Manage a trillion genotypes (10^{12})
  - One hundred thousand samples (10^5)
  - Ten million of genotypes (10^7)
- Queries on data and metadata
- Transformations
- Import
- Export

http://cgf.nci.nih.gov

GLU:Utilities

CGEMS
Cancer Genetic Markers of Susceptibility

- Data conversion
- Data management
- Genotype quality control
- Association testing
- Tag SNP selection
- Genomic metadata

Acknowledgements

CGEMS & DCEG

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

HSPH

David Hunter
Peter Kraft
David Cox
Sue Hankinson

ACS

Michael Thun
Heather Feigelson
Eugenia Calle

NCICB

Ken Buetow
Carl Schaefer
Subhah Madhavan
Liming Yang

CeRePP, France

Olivier Cussenot
Geraldine Cancel-Tassin
Antoine Valeri

Wellcome Trust, UK

Mark Minichiello

NPHI, Finland

Jarmo Virtamo

Wash. U., St Louis

Gerald Andriole

Follow-up to GWAS Discoveries: Just the beginning…..

CGEMS
Cancer Genetic Markers of Susceptibility

Fine Mapping of Notable Regions

- Genotyping & Sequencing
- Bio-informatics

Analysis of Population Genetics

Functional Determination of Causal Variant(s)

Exploration of Pathways

- Etiology
- Drug Targets

Design Issues for Clinical Evaluation

- Population-based studies
- Sequence of Clinical Studies

Validation of Clinical Use

CGEMS Data Use Certification (DUC)

CGEMS
Cancer Genetic Markers of Susceptibility

- Articulates the investigator's agreement to the conditions for responsible use of the CGEMS datasets

- Terms of the Data Access Agreement:
  - -Requesting Investigator
  - -Research Use
  - -Non-Identification
  - -Non-Transferability
  - -Intellectual Property
  - -Publication
  - -Research Use Reporting
  - -Dissemination of Research Results and Acknowledgements
  - -Non-Endorsement Indemnification
  - -Termination and Violations