National Cancer Institute

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Cancer Advisory Board

Genome-wide Association Studies

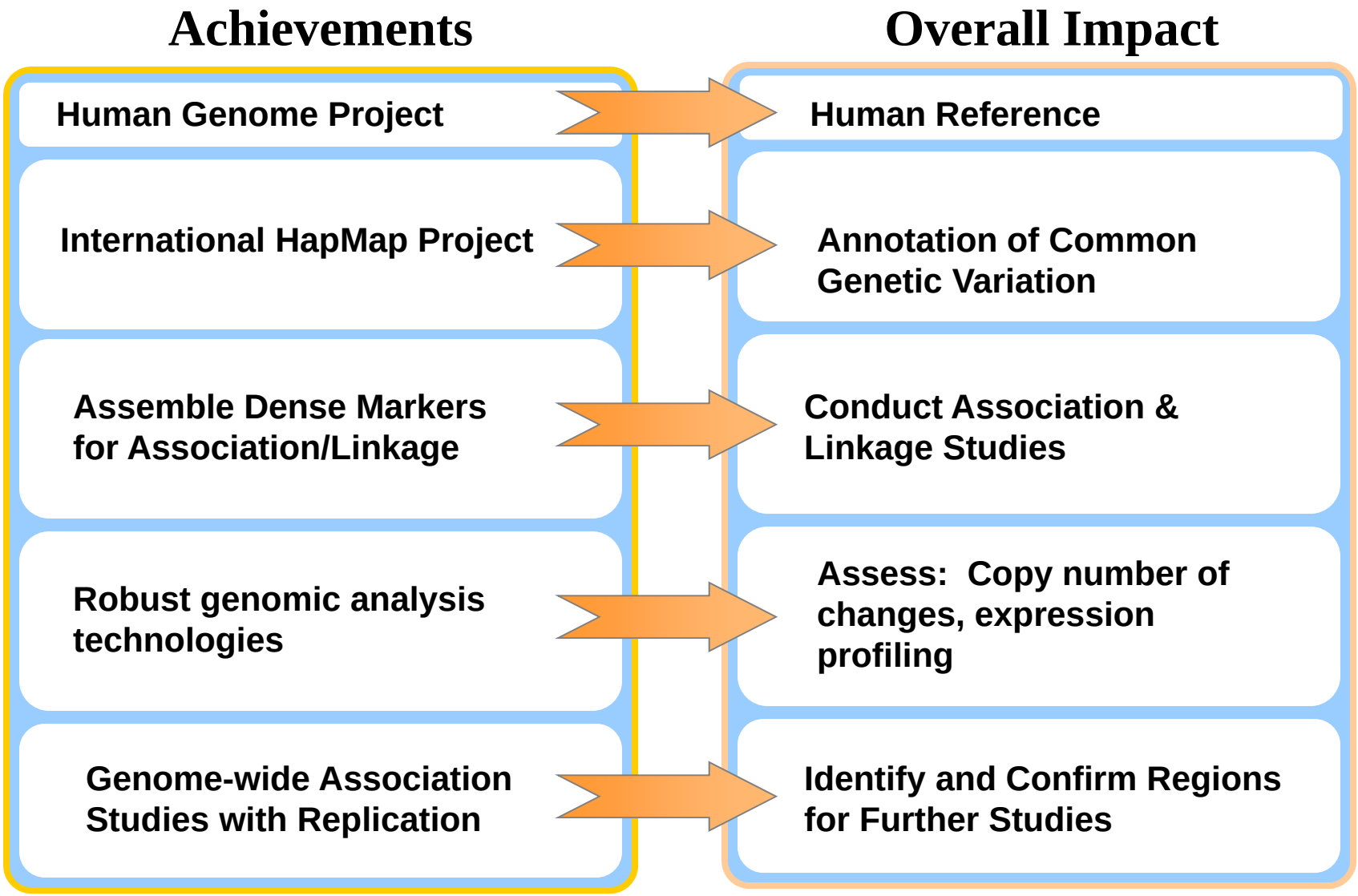
Stephen Chanock, M.D.

Chief, Laboratory of Translational Genomics, DCEG, NCI

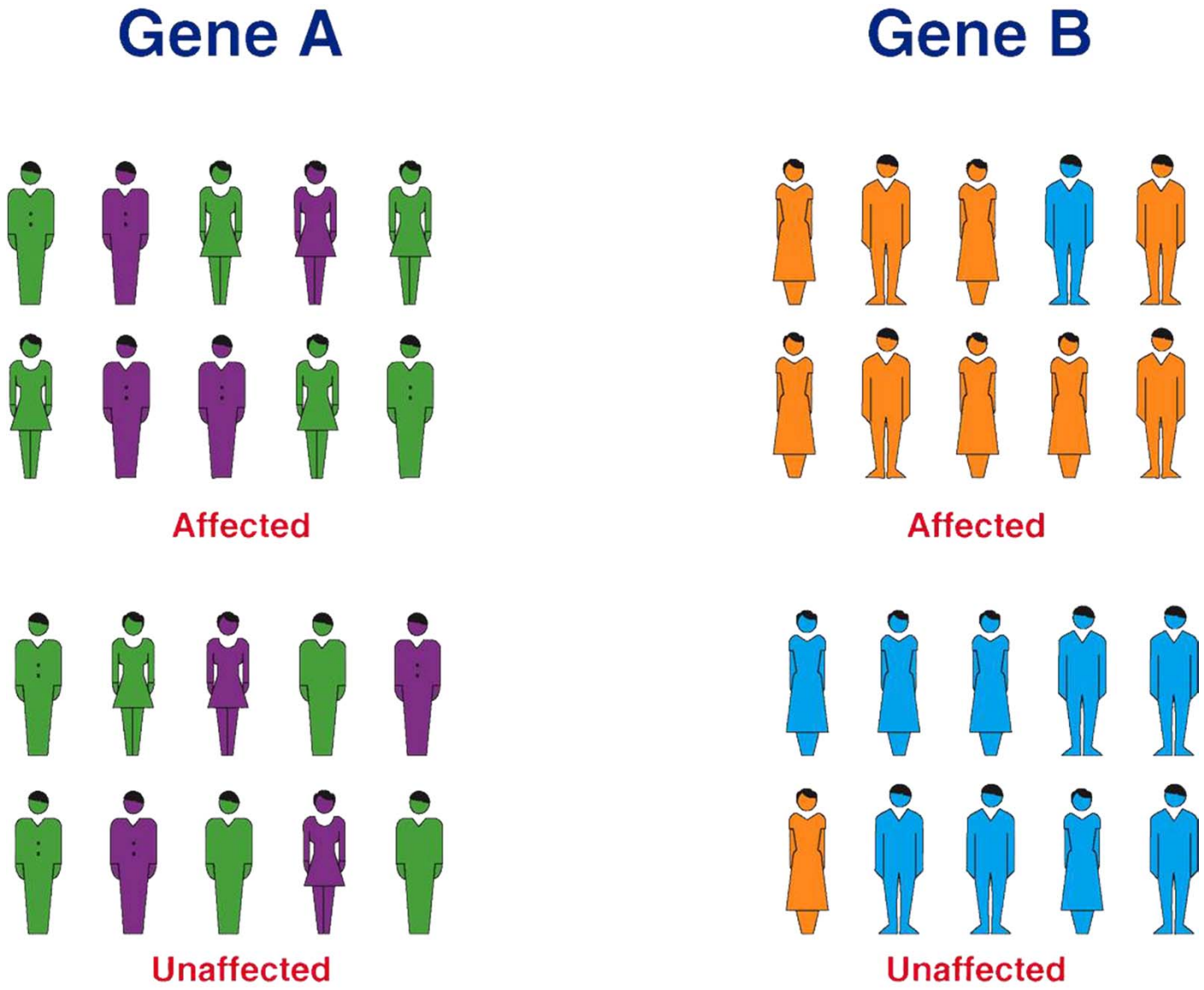
Director, Core Genotyping Facility, DCEG, NCI

December 9, 2008

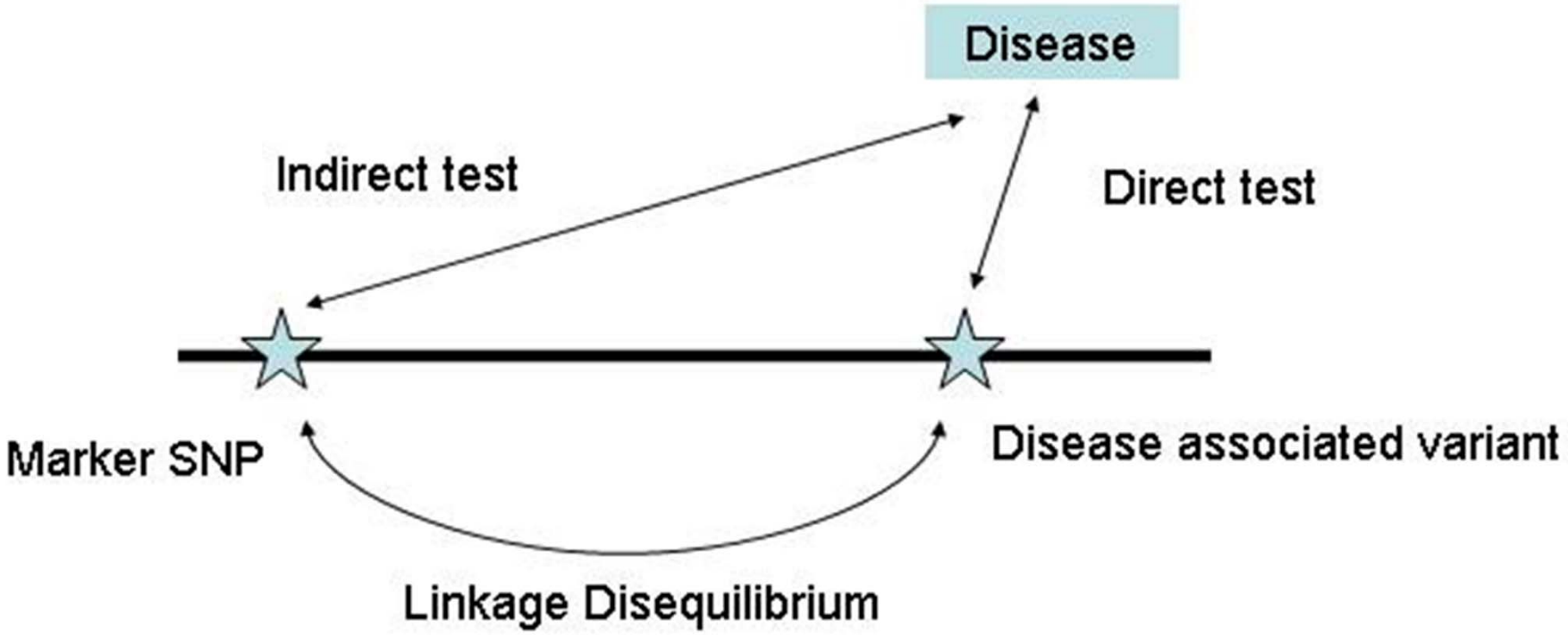
Milestones in Human Genomics & Disease Susceptibility



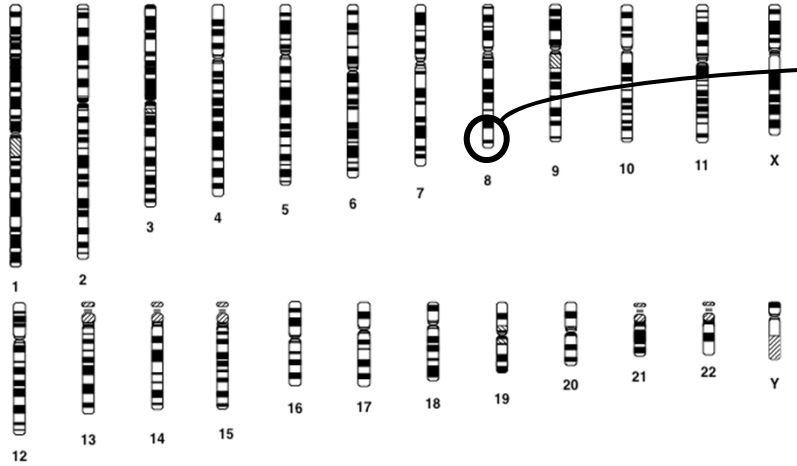
Basic Principle of Genetic Association Studies In Unrelated Individuals

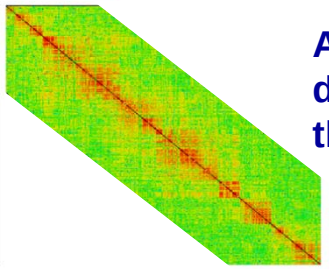


Genetic Association Testing: Finding ***Markers***



10 million SNPs across the genome

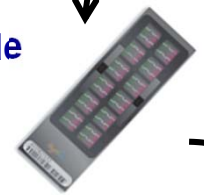




Areas of linkage disequilibrium across the genome

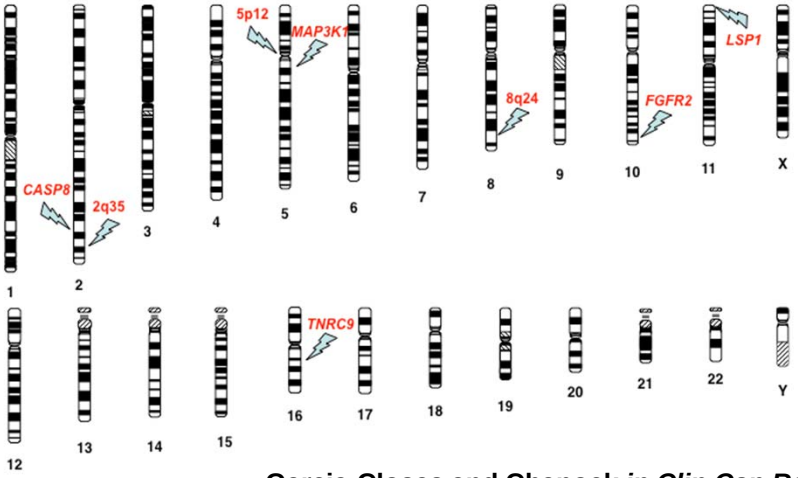
Selection of tagSNPs to capture common genetic variation in population under study

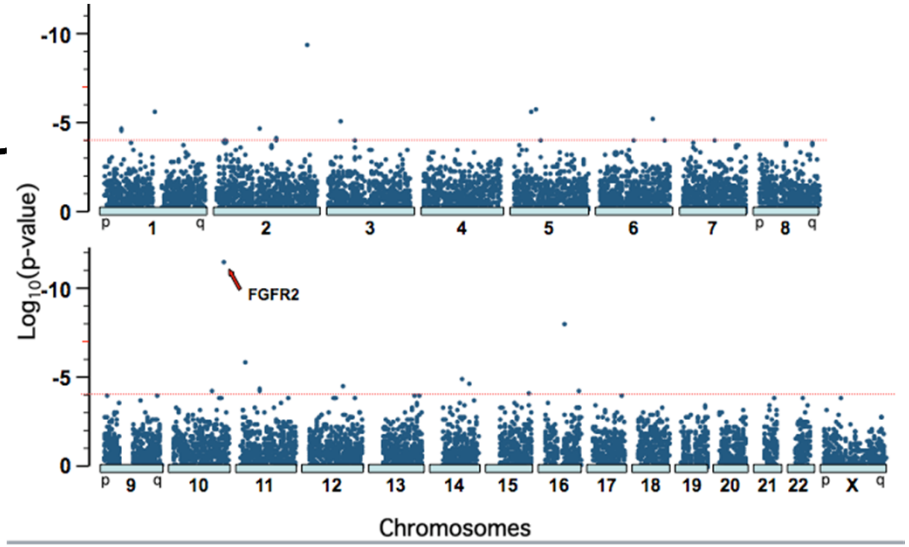
Genome wide SNP chips



Association with disease risk in discovery and replication studies

Mapping of susceptibility loci



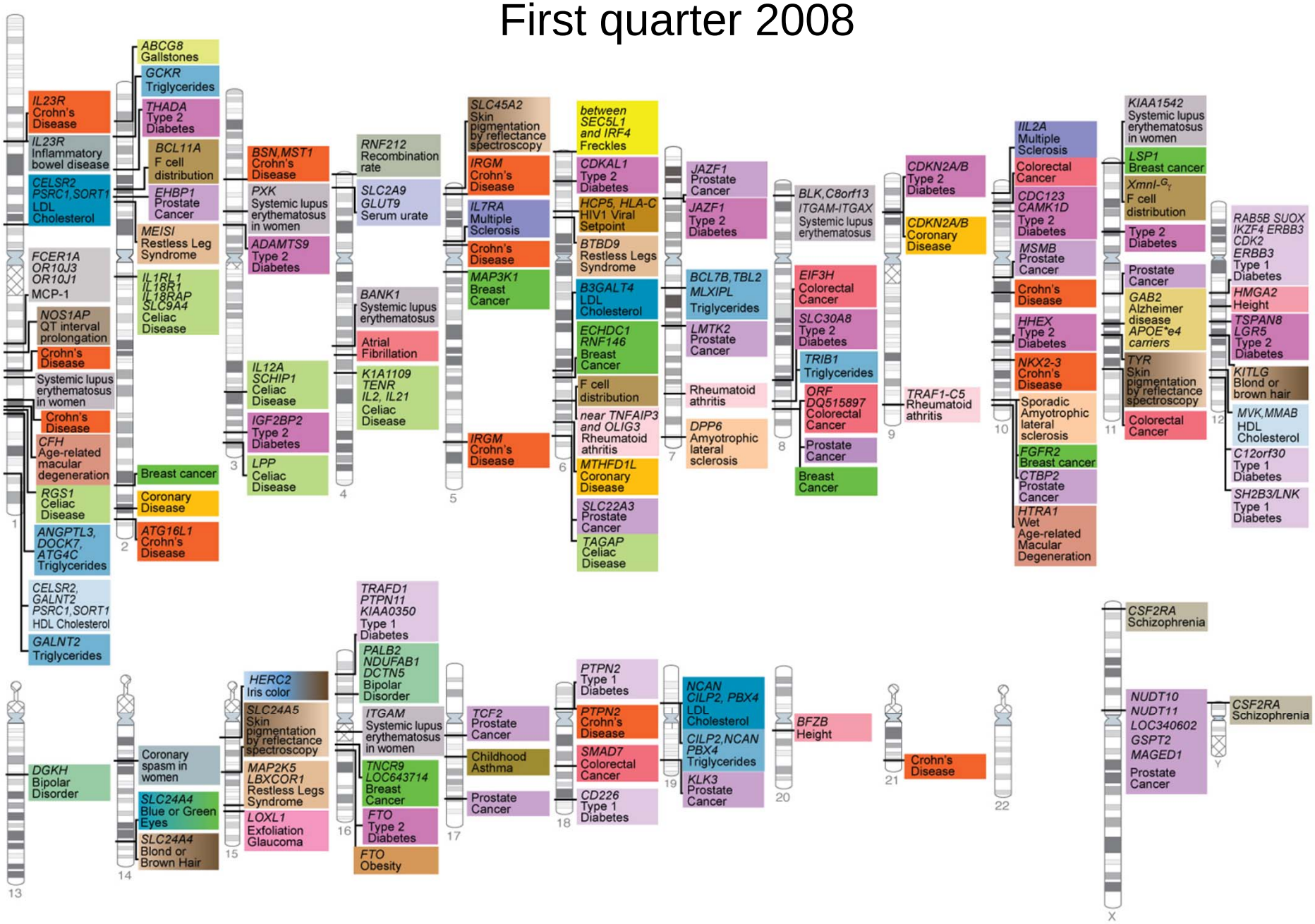


Garcia-Closas and Chanock *in Clin Can Res 2009*

Promise of Genome-wide Association Studies

1. **Discovery of New Regions in the Genome Associated with Diseases/Traits**
   - ***New "Candidate Genes"***
2. **Explore genes/pathways**
   - **Etiology**
   - **Gene-Environment/Lifestyle Interactions**
   - **"Druggable" targets**
3. **Establish utility of genetic markers for risk prediction**
   - **For individual or public health decisions**

First quarter 2008



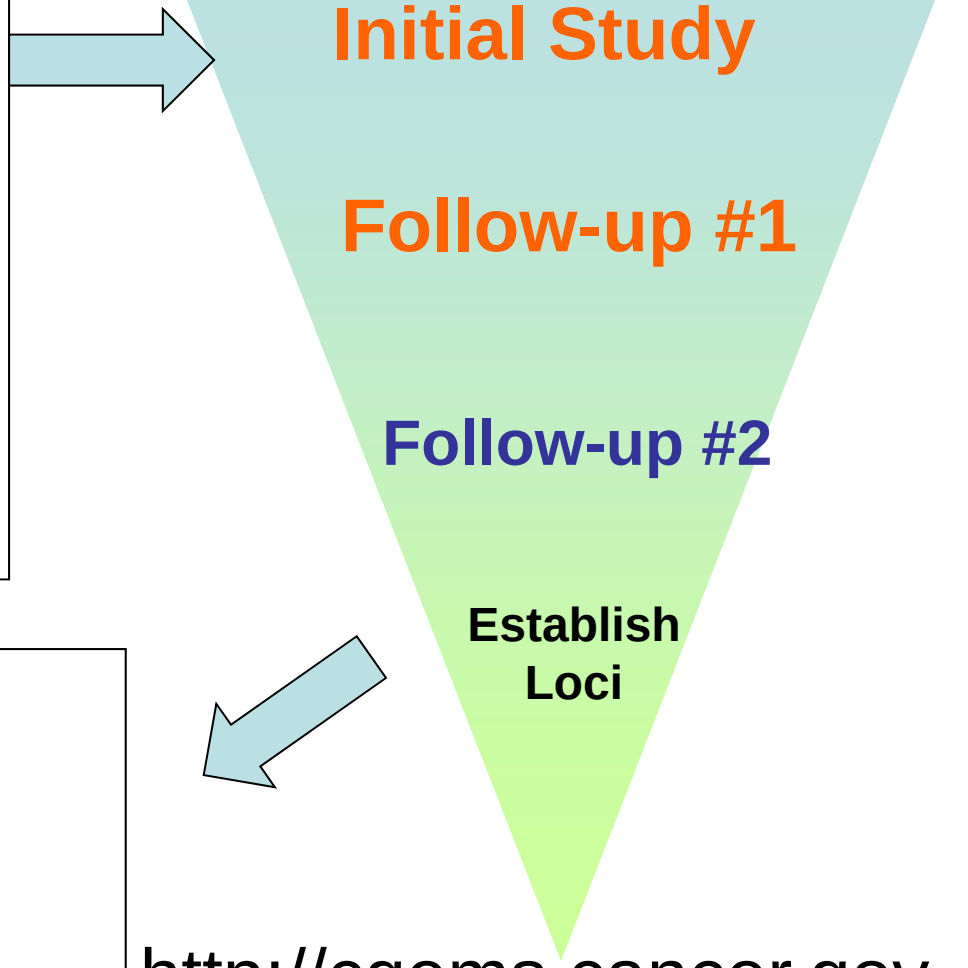
Manolio, Brooks, Collins, J. Clin. Invest., May 2008

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

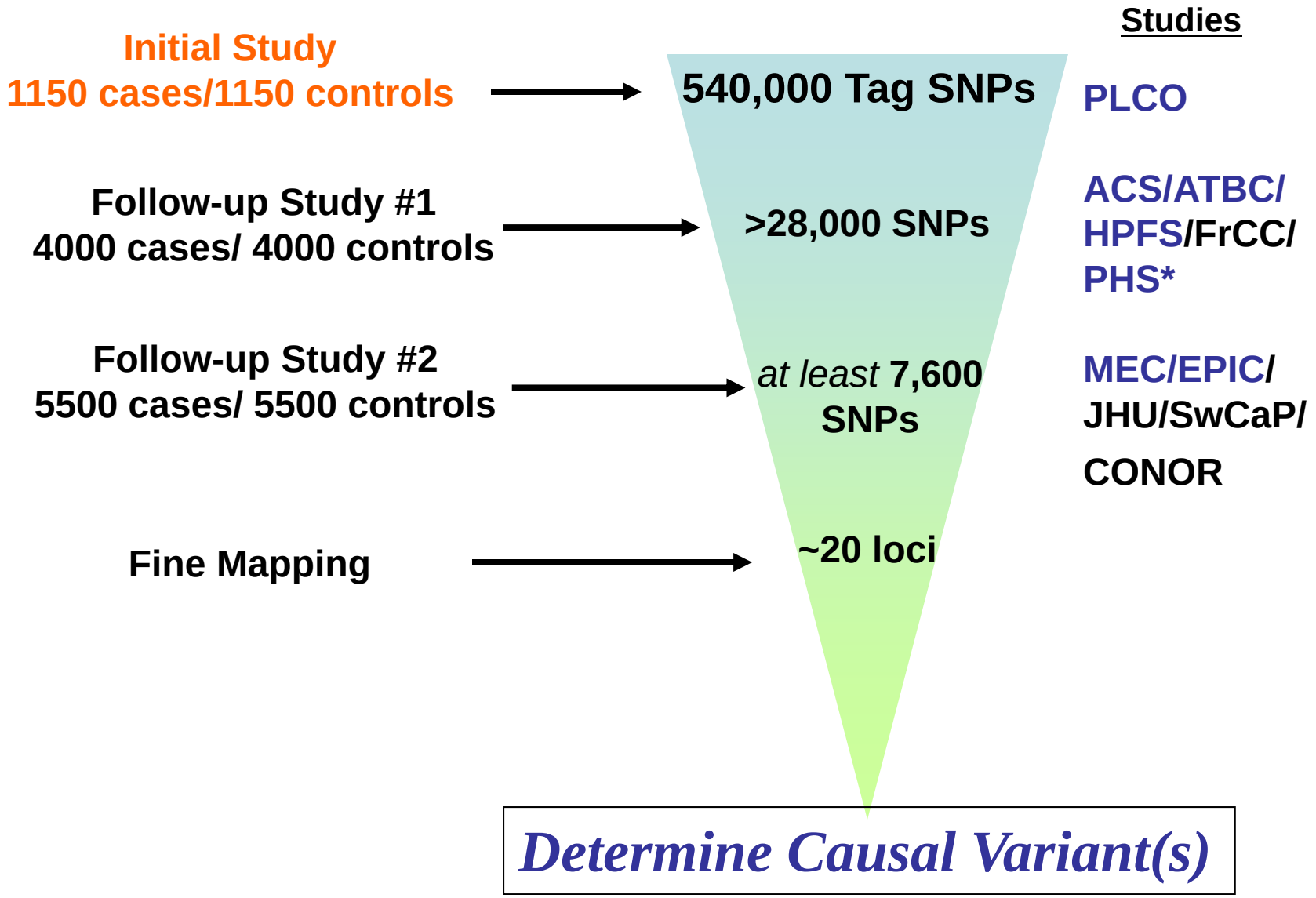
http://cgems.cancer.gov

Prostate Cancer Risk 2006

- Age
- Ethnic Background
- Family History

General Strategy for CGEMS Prostate GWAS

circa 2006

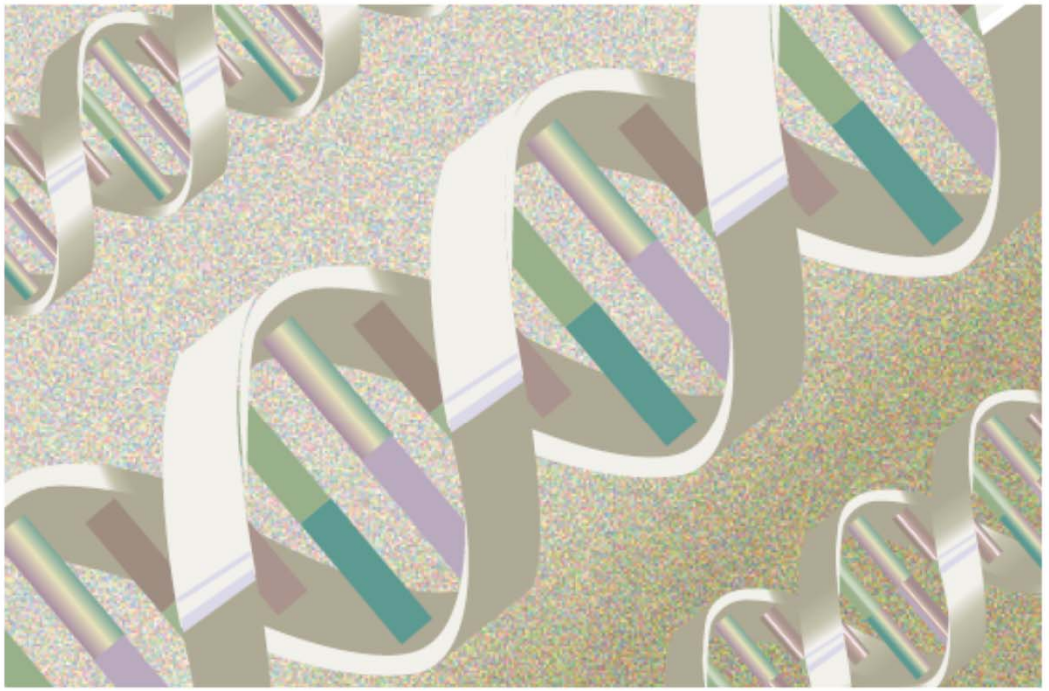


FEATURE

Replicating genotype–phenotype associations

What constitutes replication of a genotype–phenotype association, and how best can it be achieved?

NCI-NHGRI Working Group on Replication in Association Studies

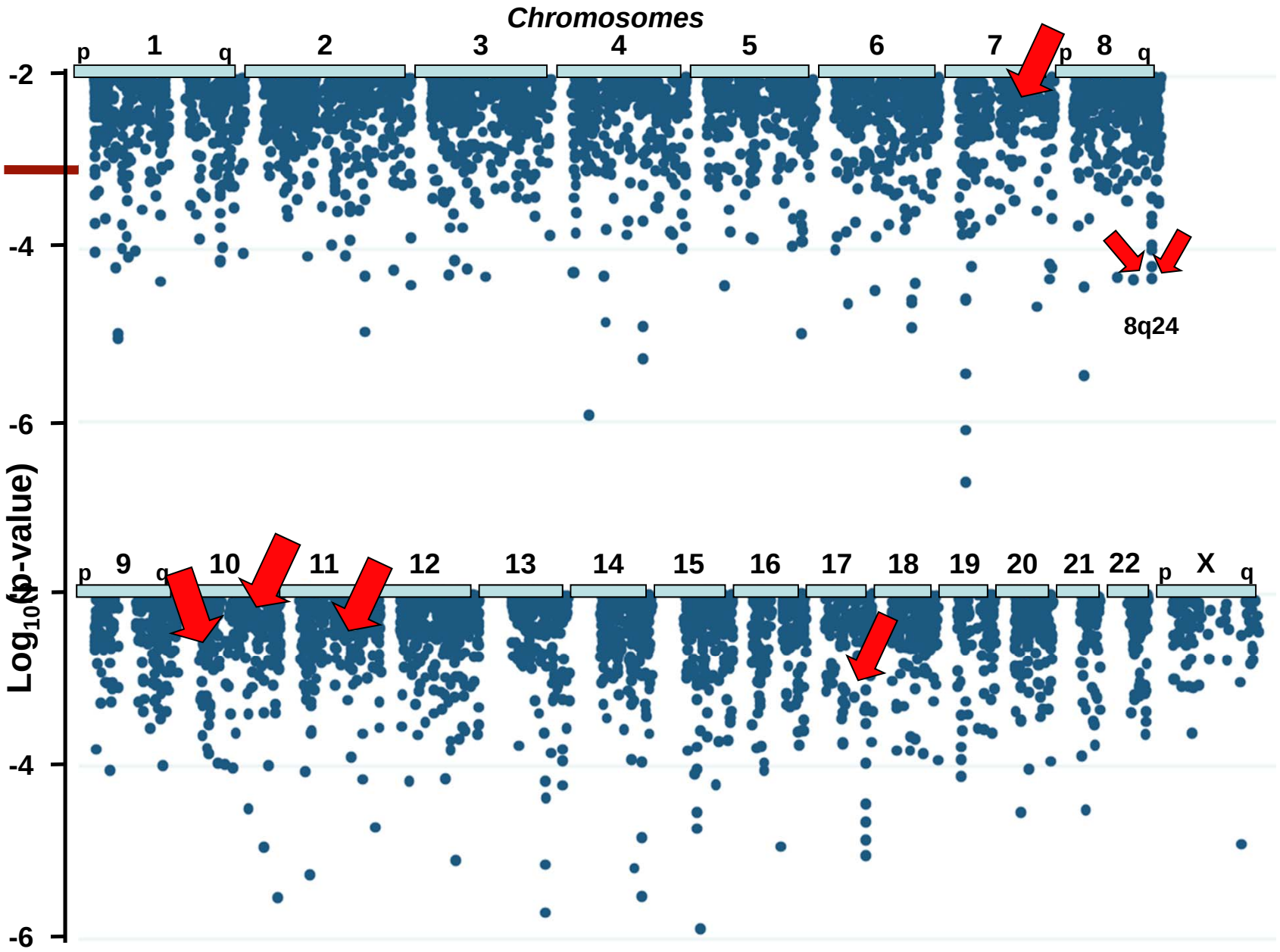


The study of human genetics has recently undergone a dramatic transition with the completion of both the sequencing of the human genome and the mapping of human haplotypes of the most common form of genetic variation, the single nucleotide polymorphism (SNP)[1–3]. In concert with this rapid expansion of detailed genomic information, cost-effective genotyping technologies have been developed that can assay hundreds of thousands of SNPs simultaneously. Together, these advances have allowed a systematic, even 'agnostic', approach to genome-wide interrogation, thereby relaxing the requirement for strong prior hypotheses.

So far, comprehensive reviews of the published literature, most of which reports work based on the candidate-gene approach, have demonstrated a plethora of questionable genotype–phenotype associations, replication of which has often failed in independent studies[4–7]. As the transition to genome-wide association studies occurs, the challenge will be to separate true associations from the blizzard of false positives attained through attempts to replicate studies because of issues in either the initial study or the attempted replication[4–6,32,33]. Small sample size is a frequent problem and can result in inappropriate conclusion from the literature because follow-up studies have not consistently analysed the same markers or those in perfect linkage dis-

Chanock S, Manolio T, et al, Nature 2007; 447:655-660

CGEMS Prostate Cancer GWAS



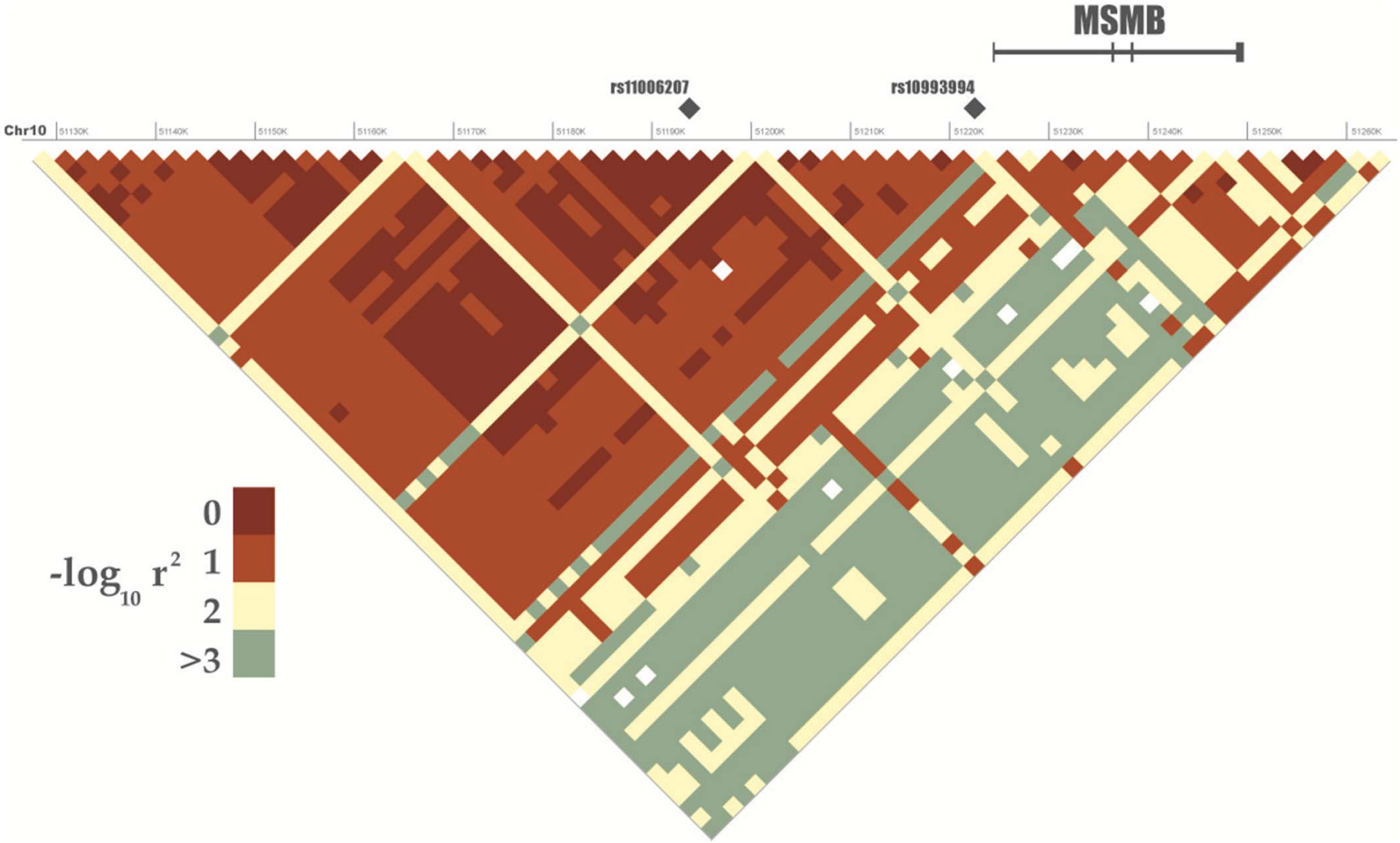
incidence density sampling

MSMB, b-microseminoprotein Chromosome 10

CGEMS
Cancer Genetic Markers of Susceptibility

- **Encodes MSP of the immunoglobulin binding factor family**
  - 10.7 kDa non-glycosylated cysteine rich protein
  - Synthesized by epidethilium of prostate
  - Secreted into seminal plasma
- **MSP and binding protein, PSPBP -- potential serum markers for early detection of high grade prostate cancer (*Bjartell et al Clin Can Res 2007; Reeves et al Clin Can Res 2006; Nam et al J Urol 2006*)**
- **Silenced by EZH2 in advanced, androgen-insensitive prostate cancer (*Beke et al Oncogene 2007*)**
- **"Best hit" rs10993994 -- promoter SNP alters gene expression *in vitro* (*Buckland et al Hum Mut 2005*)**

LD at *MSMB* – 15 SNPs tag this ~130kb region



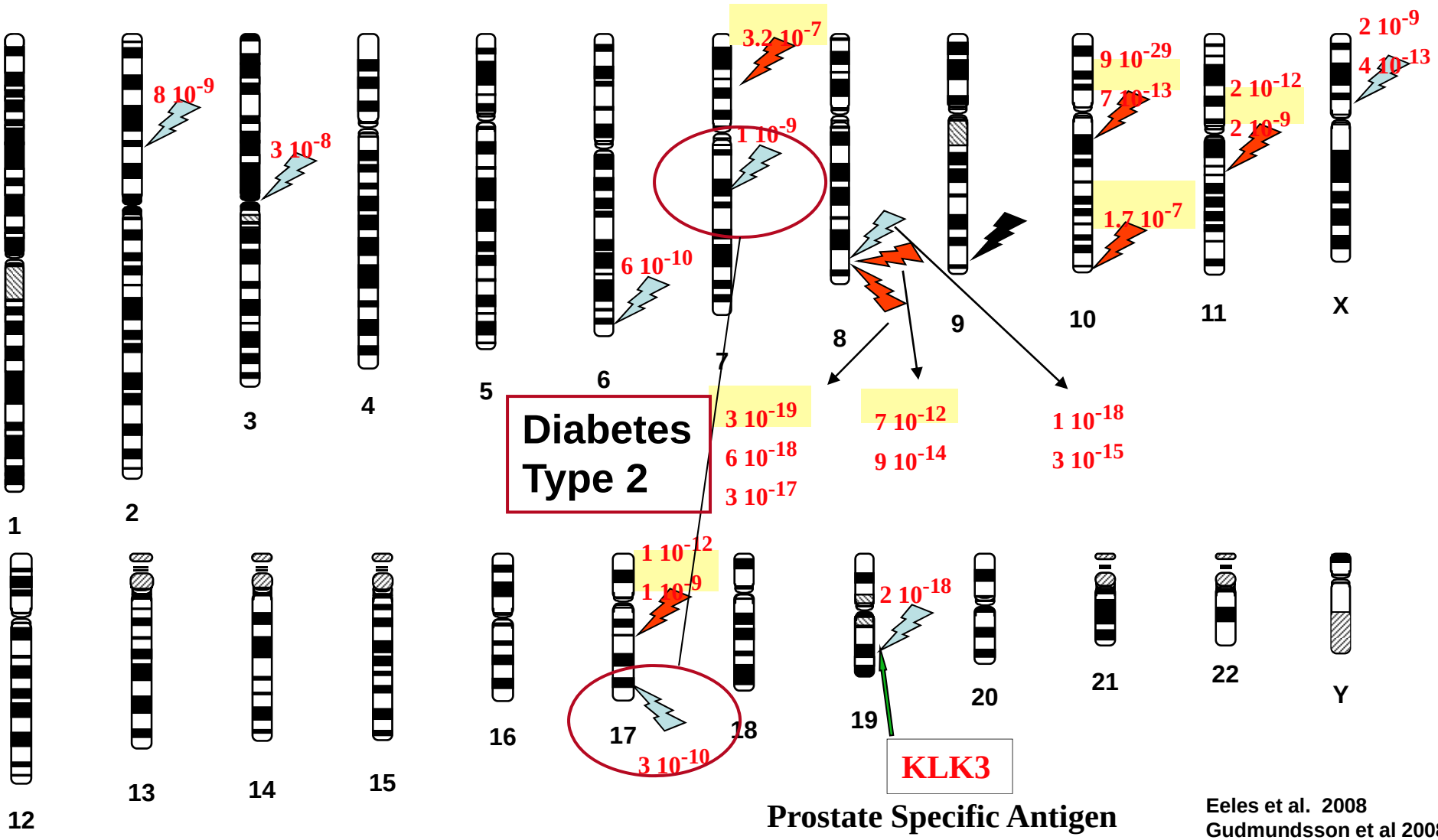
Association results, 6479 prostate cancer cases, ~6105 controls

Locus	Alleles	MAF	X^2, 2df	p	HetOR	95% CI	HomOR	95% CI
rs11004422	T,C	0.489	34.17	3.8E-08	1.08	0.97 – 1.20	1.42	1.25 – 1.60
rs7071471	C,T	0.457	40.31	1.8E-09	1.11	1.00 – 1.23	1.49	1.31 – 1.69
rs11593319	G,T	0.071	9.36	0.009	0.81	0.71 – 0.93	0.90	0.52 – 1.53
rs10826075	C,G	0.248	9.28	0.010	1.08	0.98 – 1.19	1.30	1.09 – 1.56
rs4630240	C,T	0.381	32.44	9.0E-08	0.79	0.72 – 0.87	0.71	0.62 – 0.82
rs11006207	C,T	0.464	58.39	2.1E-13	1.14	1.05 – 1.24	1.49	1.34 – 1.65
rs10826223	G,A	0.096	6.64	0.036	0.88	0.80 – 0.97	0.92	0.67 – 1.25
rs10993994	C,T	0.407	82.95	9.7E-19	1.20	1.11 – 1.30	1.64	1.47 – 1.82
rs7076948	T,C	0.376	8.09	0.012	1.03	0.94 – 1.13	1.21	1.06 – 1.39
rs10994470	G,A	0.036	1.04	0.596	0.92	0.78 – 1.10	0.74	0.23 – 2.33
rs7904463	C,T	0.327	5.77	0.056	1.06	0.98 – 1.14	1.15	1.02 – 1.30
rs17178655	G,A	0.211	1.42	0.491	0.99	0.90 – 1.09	0.88	0.72 – 1.08
rs10994675	G,A	0.416	9.12	0.011	1.07	0.99 – 1.16	1.17	1.06 – 1.30

Yeager et al. in submission 2008

16^{+} published loci involved in prostate cancer susceptibility

with significance $p < 5 \times 10^{-7}$



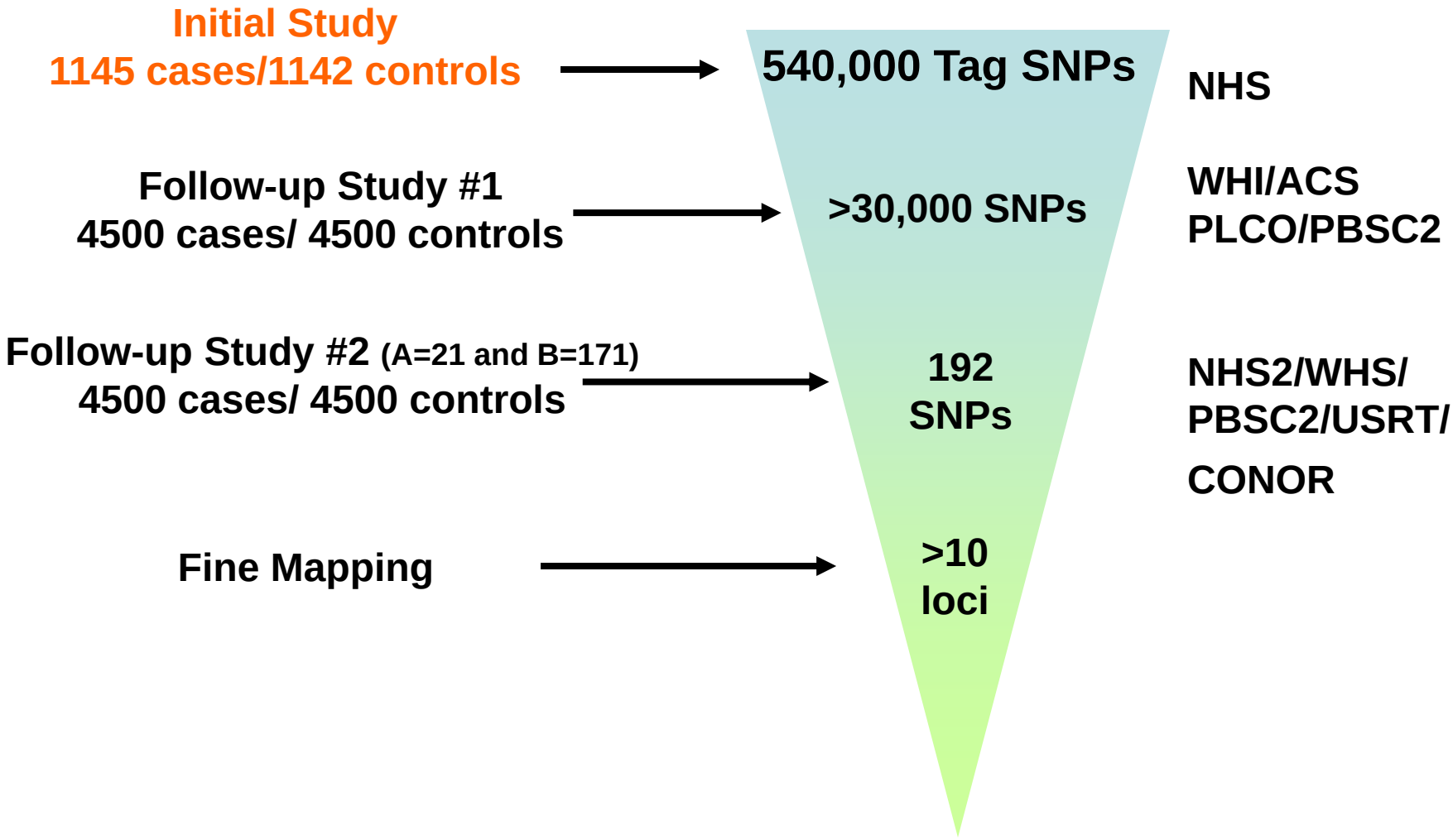
Eeles et al. 2008
Gudmundsson et al 2008
Haiman et al 2007

Prostate Cancer Risk 2008

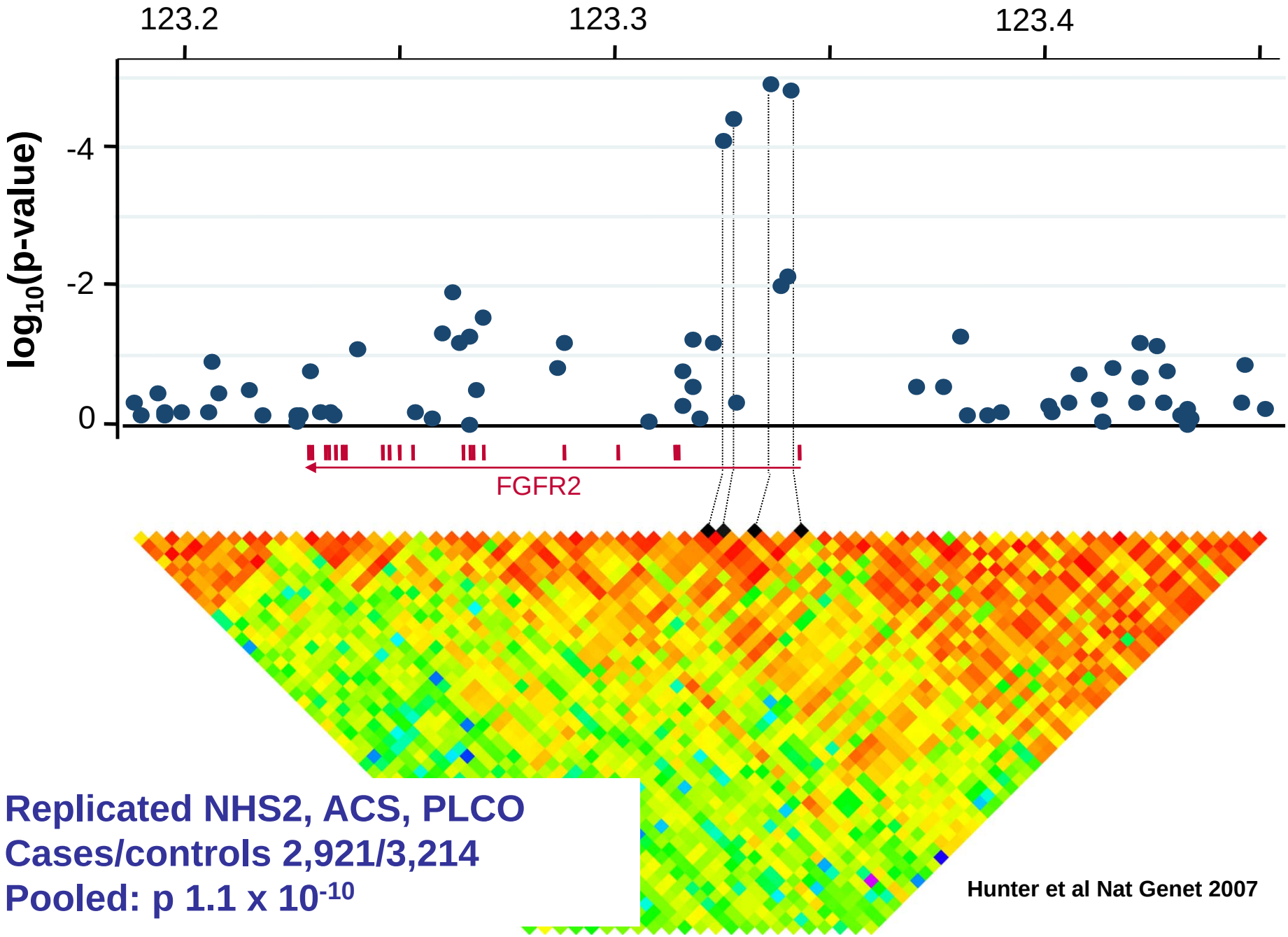
- **Age**
- **Ethnic Background**
- **Family History**
- **Genetic markers**
  - **16 Regions of the Genome!!!**

CGEMS – Cancer Genetic Markers of Susceptibility

General Strategy for Breast Cancer GWAS



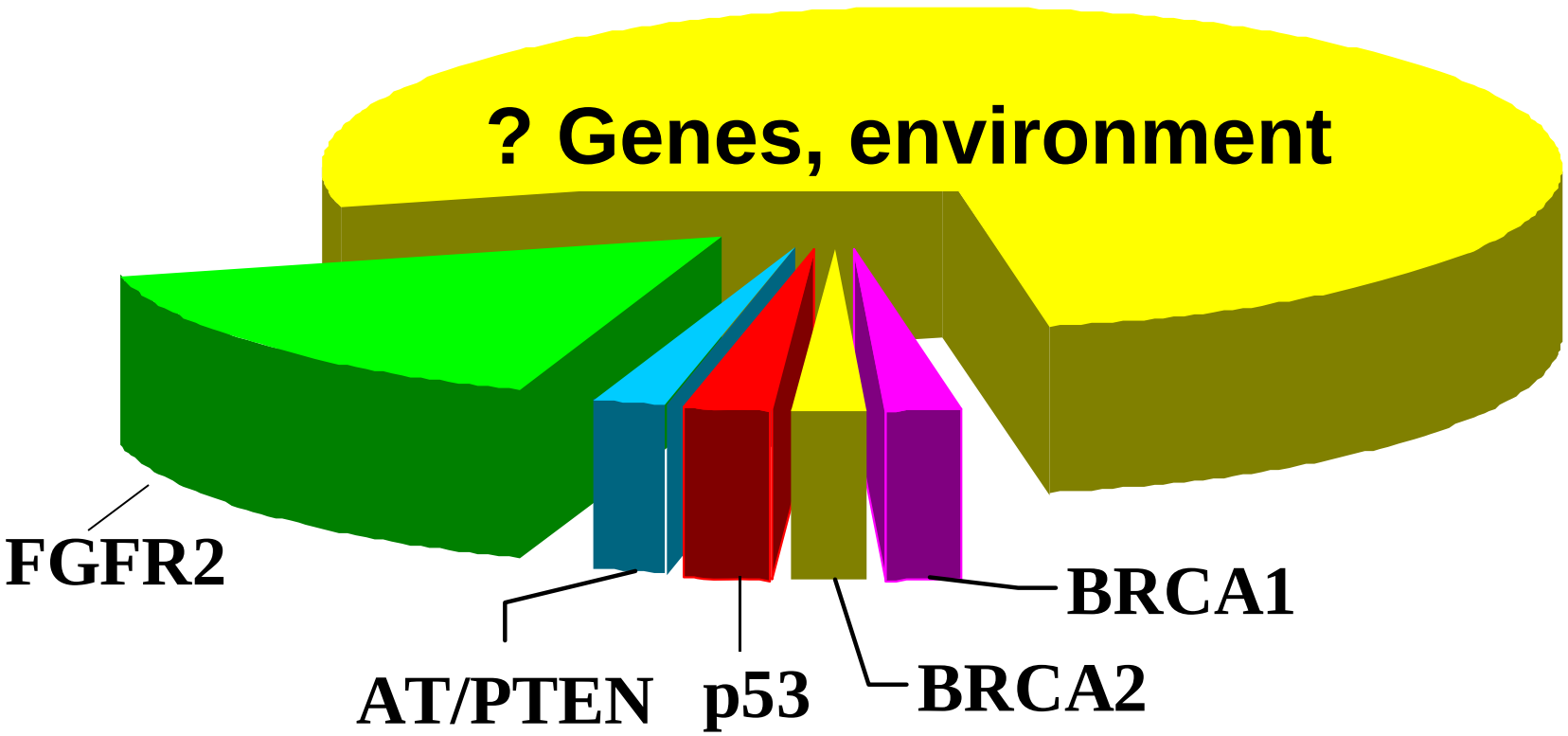
FGFR2 Signals in GWAS: Intron 2



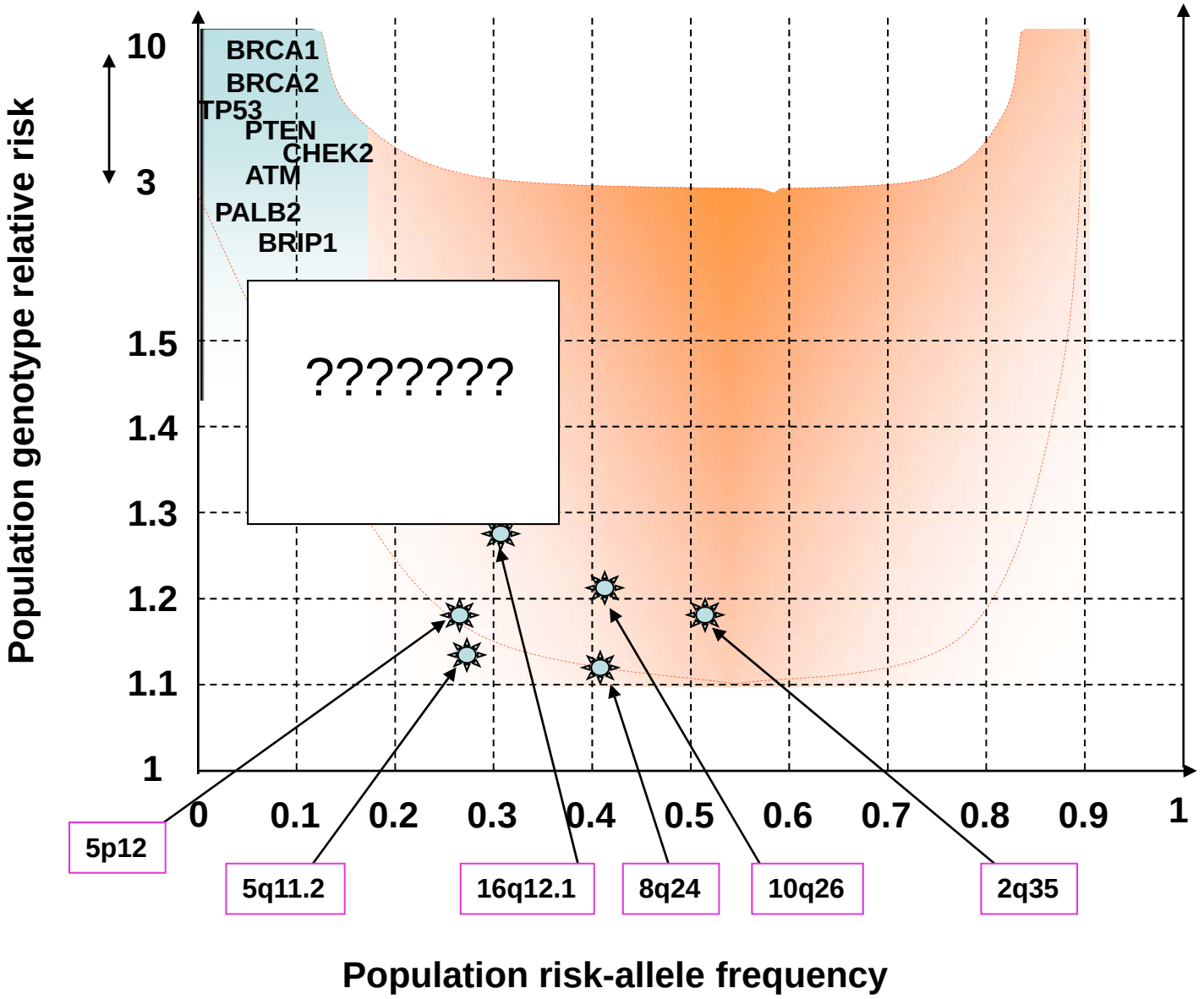
Replicated NHS2, ACS, PLCO
Cases/controls 2,921/3,214
Pooled: p 1.1 x 10^{-10}

Hunter et al Nat Genet 2007

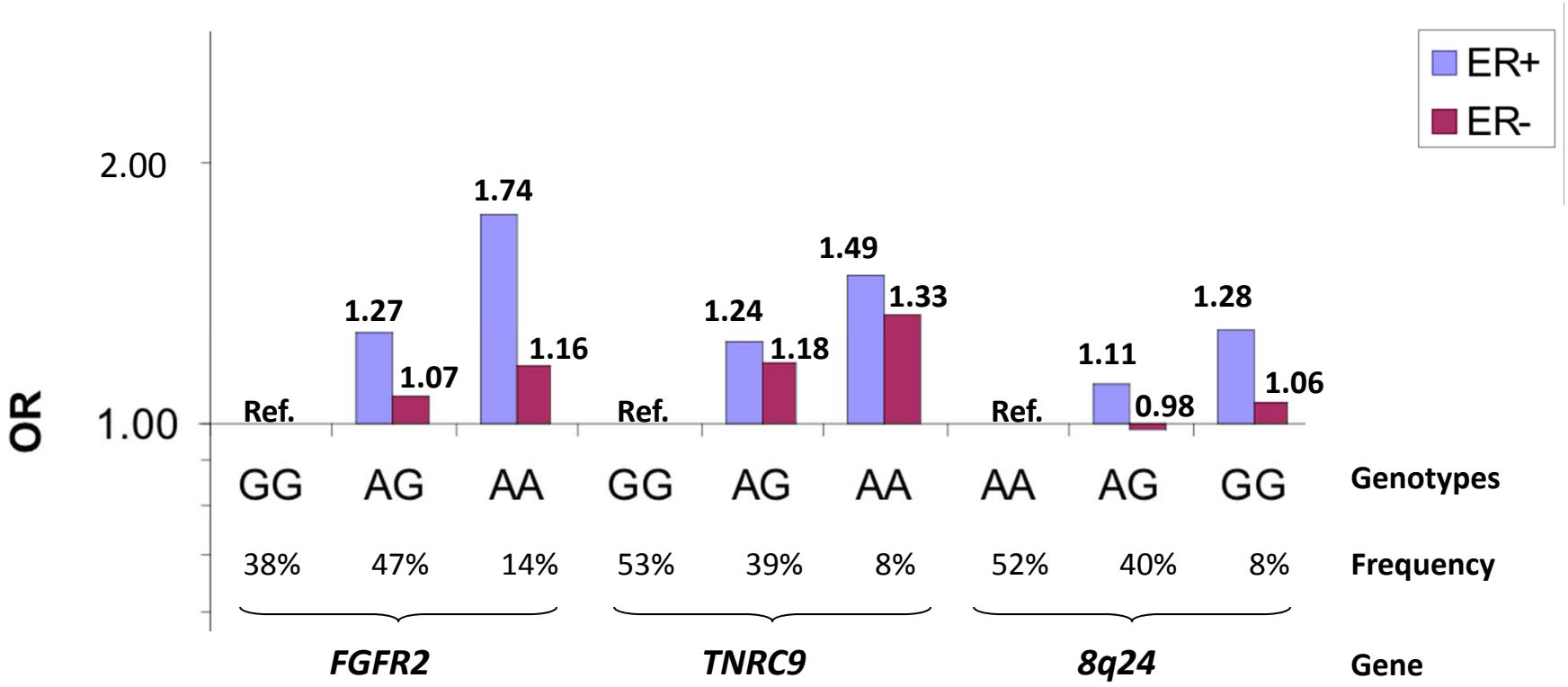
Inherited Susceptibility to Breast Cancer



Genetic Predisposition to Breast Cancer
European population

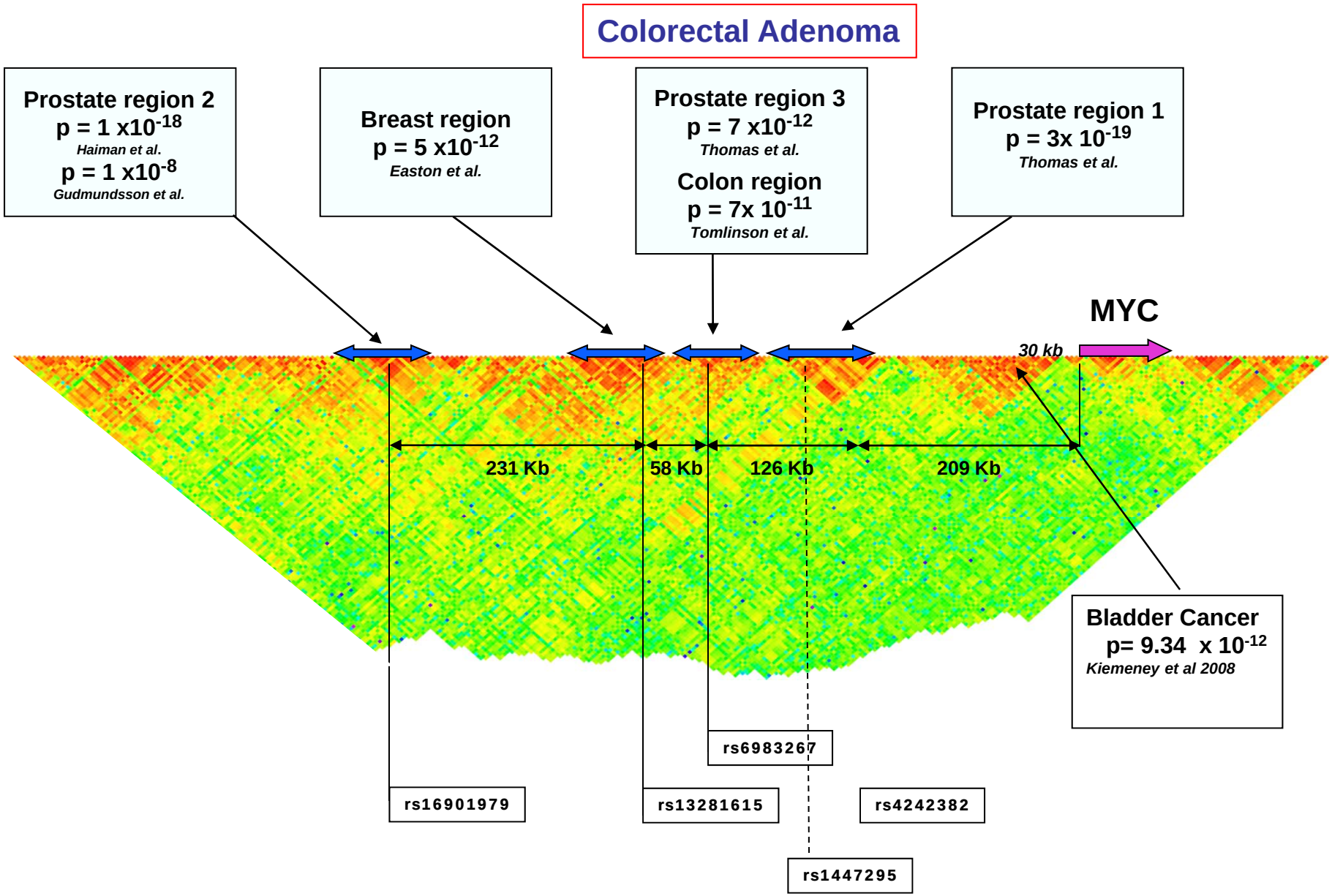


Susceptibility Loci Modified by Estrogen Receptor Status

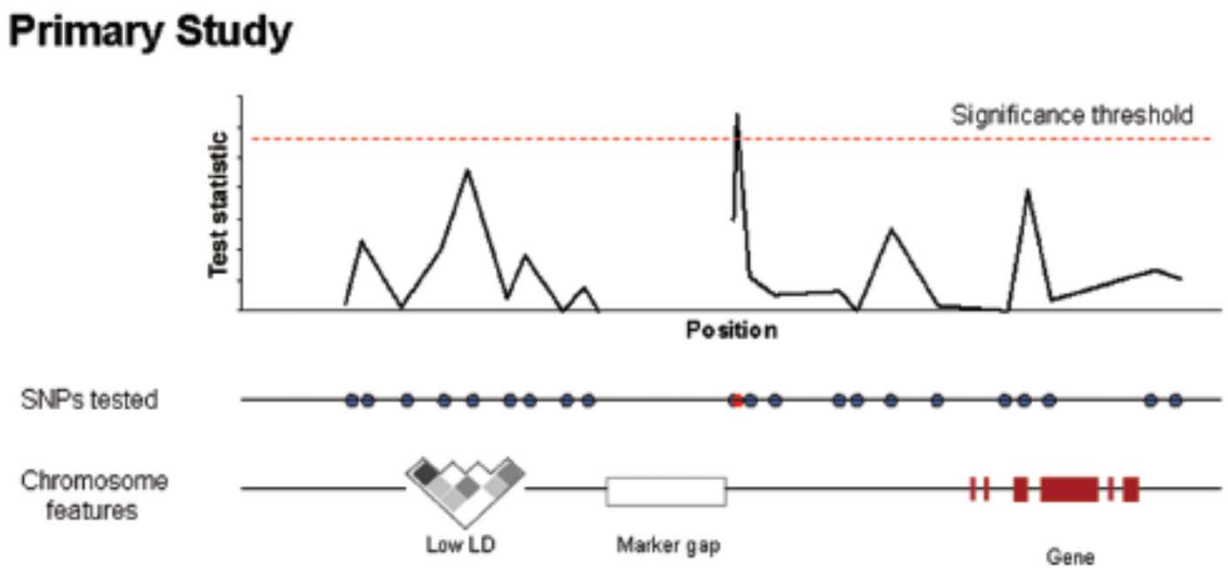


Garcia-Closas et al. PLoS Genetics 2008

Cancer susceptibility loci in the 8q24 region



Replication and Mapping



Begin Functional Analyses

Clarke et al AJHG 2007

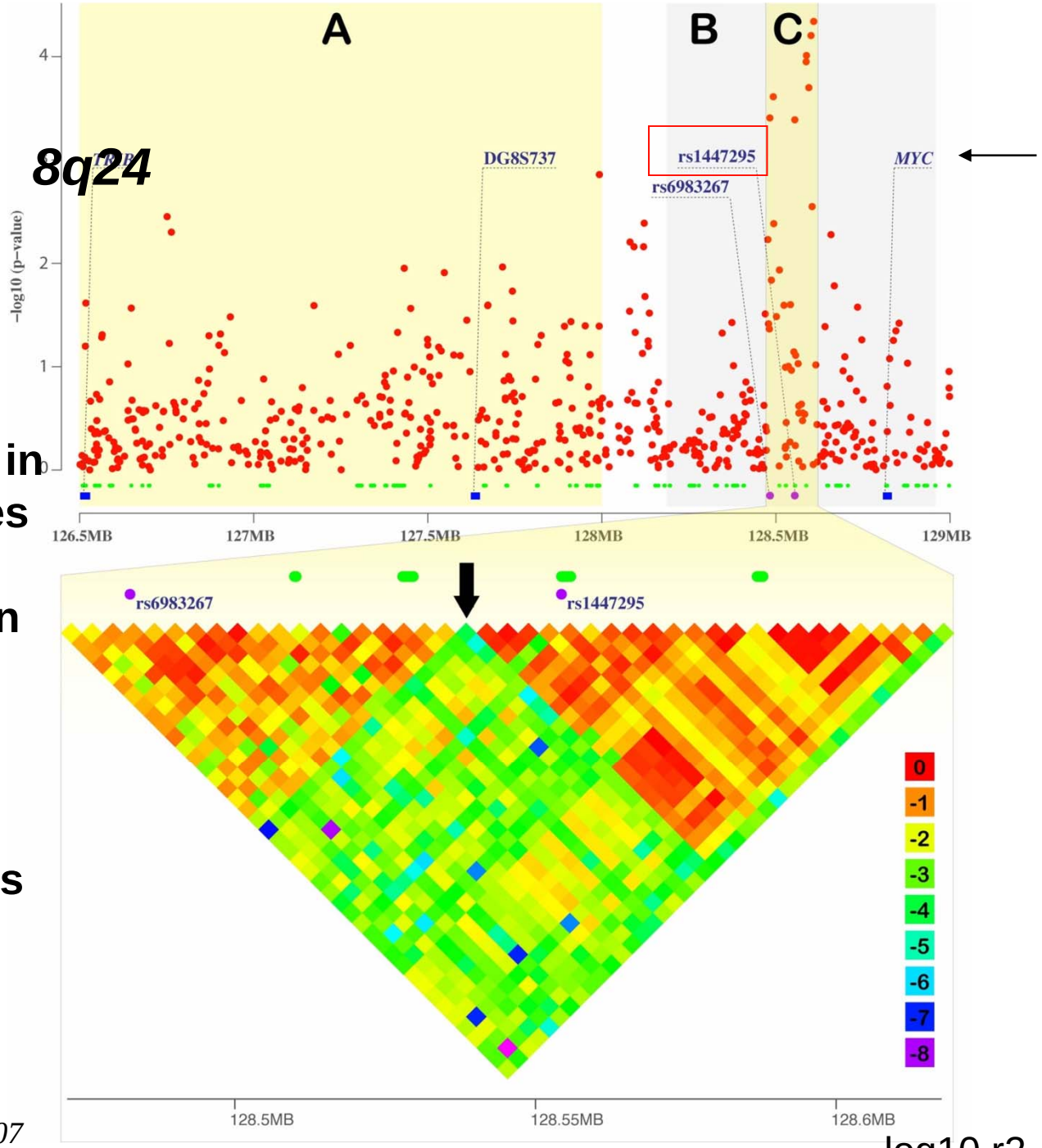
Prostate Cancer & 8q24

Rs1447295 replication in BPC3 & 3 Other Studies

Commonly Amplified in Prostate tumors

“Gene-poor region”

GWAS- multiple signals

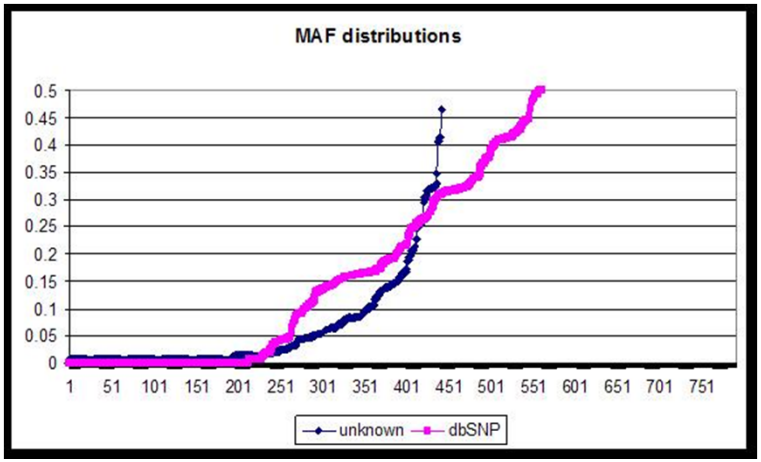


Yeager et al Nat Genet 39:645-649, 2007

454 Quality Control and genetic analyses

- Concordance with GWAS data (for 79 PLCO samples), 39 SNPs, >99% per locus and per sample
- Genotype calls for 791 SNPs

	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



Resequence analysis: Discovery of ALL Variants

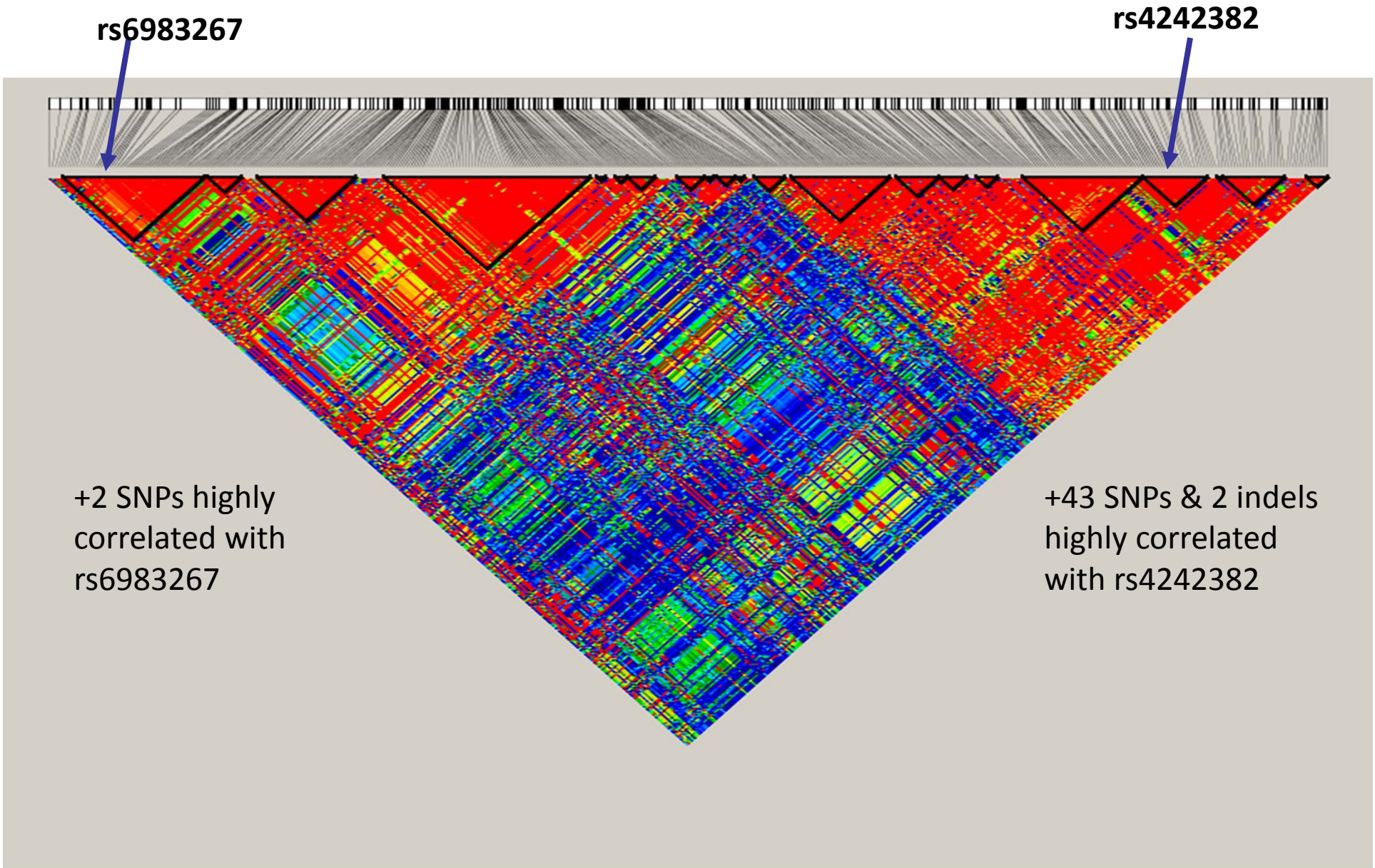
Table 2. Comparative Coverage of Variants in 8q24 Regions 1 and 3

	# bins monitored	# SNPs monitored	# bins not monitored	# SNPs not monitored
All SNPs (n=454)	114	454	0	0
dbSNP (n=299)	80	410	34	44
HapMap (n=174)	53	353	62	101
Novel (n=155)	34	44	80	410

Further Genotyping to Nominate Best SNPs for Functional Analysis

Yeager et al 2008

Refining Linkage Disequilibrium across 2 Regions of 8q24



Initial observations on 8q24 Prostate Cancer Susceptibility

- All common (> 1% MAF) variants identified within 136 kb region
- Two SNPs highly correlated with rs6983267
  - Associated with prostate, colon, ovarian cancers
- 43 SNPs and 2 indels highly correlated with rs4242382
  - Associated with prostate cancer in Caucasian men
- Each of these variants is equally as likely to be responsible for disease as the initially-reported SNPs

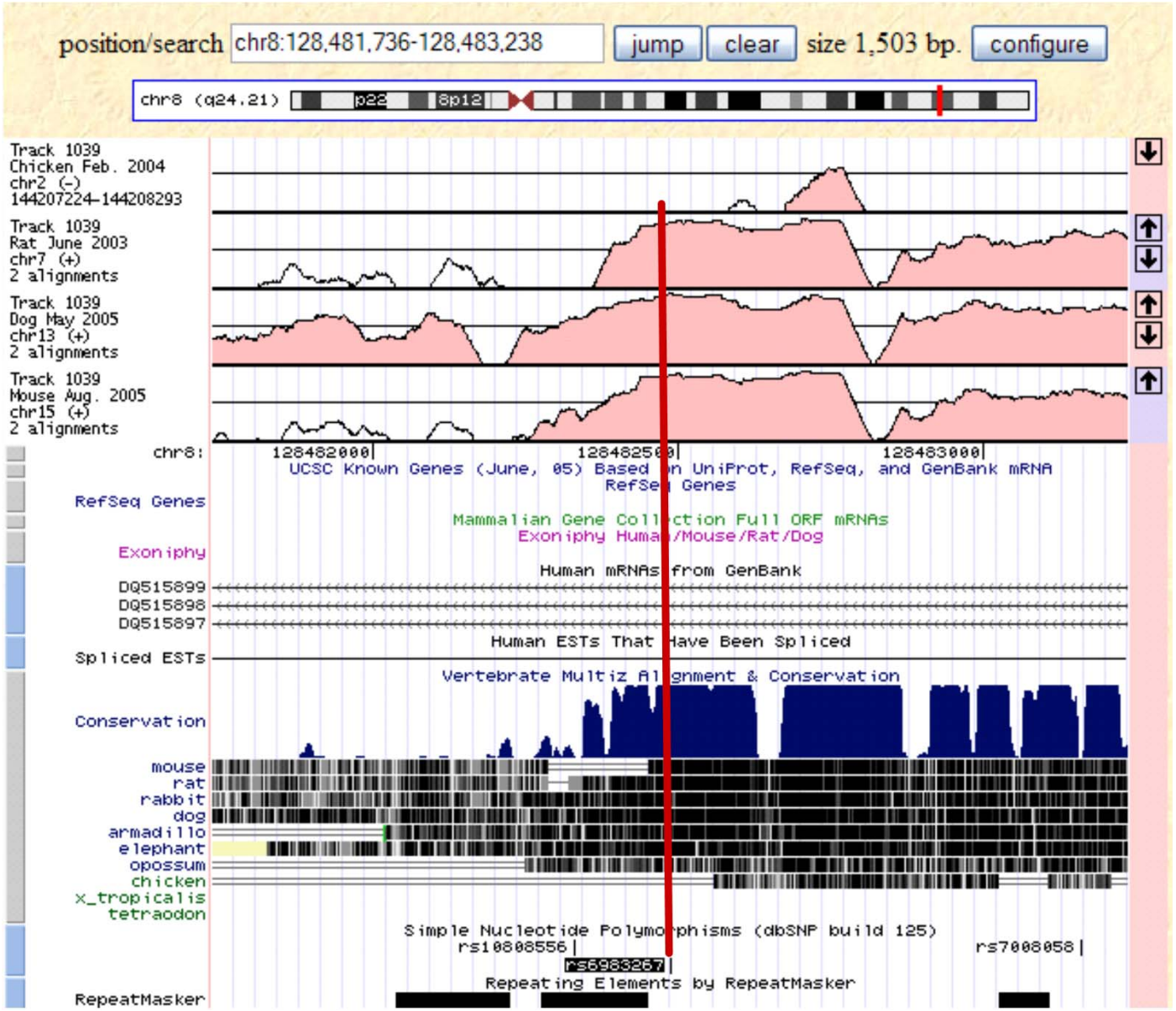
rs6983267 – possible candidate for causal/contributory variant

VISTA

Enhancer prediction

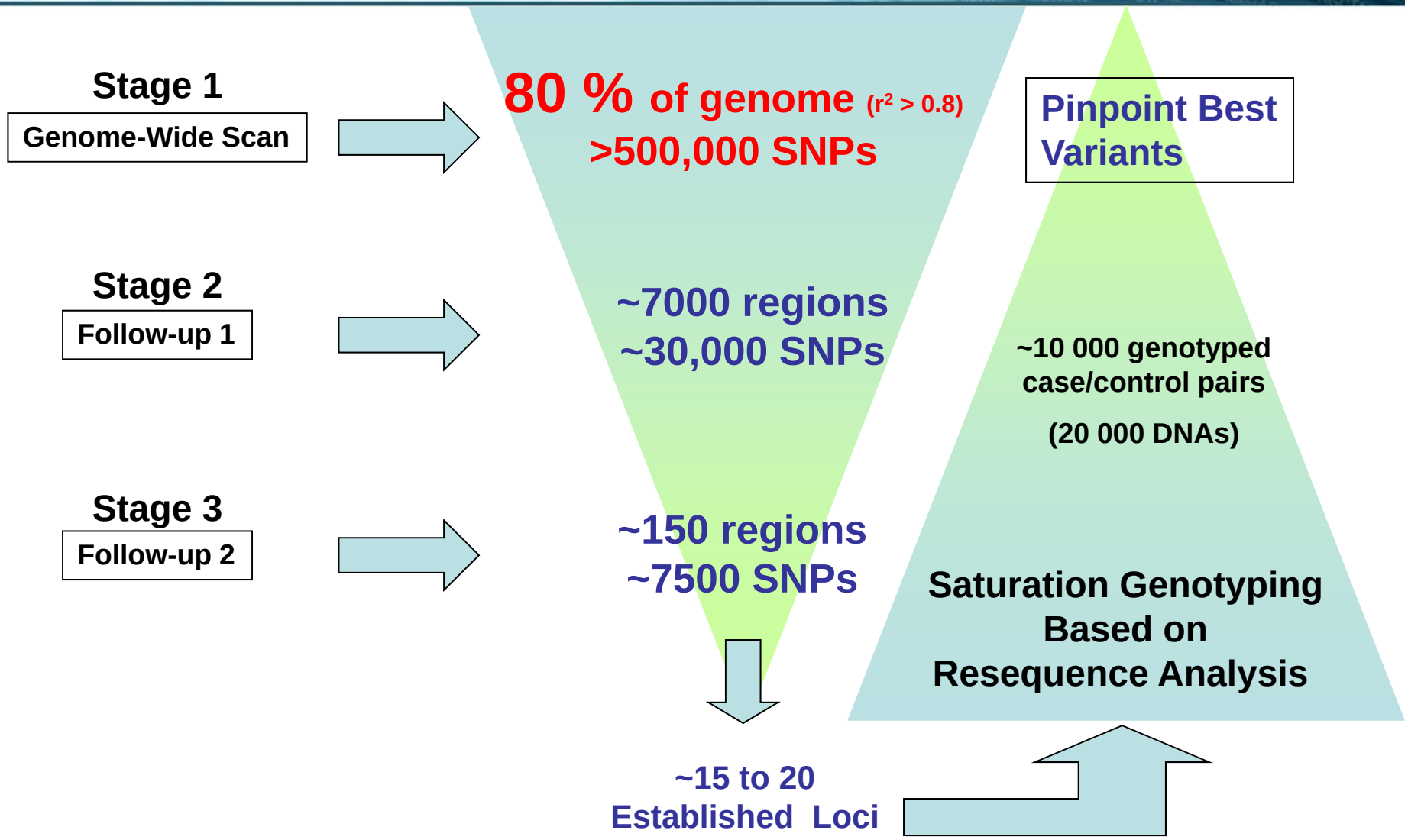
Regulatory potential

Mammalian conservation



Nominating Variants for Functional Analysis: Refining the GWAS and Replication

CGEMS
Cancer Genetic Markers of Susceptibility



GWAS at NCI

DCEG

CGEMS

- **Breast cancer**
- **Prostate cancer**

PanScan I & II

Lung cancer

Bladder cancer

Kidney cancer

NHL

Upper GI

- ***Gastric cancer***
- ***Esophageal cancer***

Collaborative

BPC3

- **Aggressive prostate cancer**
- **ER (-) breast cancer**

Brain tumor

Testicular

African American

- **Prostate cancer**
- **Breast cancer**

nature genetics

Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer

COGENT Study[1]

Genome-wide association (GWA) studies have identified multiple loci at which common variants modestly influence the risk of developing colorectal cancer (CRC). To enhance power to identify additional loci with similar effect sizes, we conducted a meta-analysis of two GWA studies, comprising 13,315 individuals genotyped for 38,710 common tagging SNPs. We undertook replication testing in up to eight independent case-control series comprising 27,418 subjects. We identified four previously unreported CRC risk loci at 14q22.2 (rs4444235, ***BMP4***; $P = 8.1 \times 10^{-10}$), 16q22.1 (rs9929218, ***CDH1***; $P = 1.2 \times 10^{-8}$), 19q13.1 (rs10411210, ***RHPN2***; $P = 4.6 \times 10^{-9}$) and 20p12.3 (rs961253; $P = 2.0 \times 10^{-10}$). These findings underscore the value of large sample series for discovery and follow-up of genetic variants contributing to the etiology of CRC.

Whereas inherited susceptibility is responsible for ~35% of all CRC[1], high-risk germline mutations in *APC*, the mismatch repair (MMR) genes, *MUTYH* (*MYH*), *SMAD4*, *BMPR1A* and *STK11/LKB1* account for <6% of all cases[2]. Recent GWA studies have validated the hypothesis that part of the heritable risk is caused by common, low-risk variants, identifying CRC susceptibility loci mapping to 8q24 (rs6983267)[3,4], 8q23.3 (rs16892766, *EIF3H*)[5], 10p14 (rs10795668)[5], 11q23 (rs3802842)[6], 15q13 (rs4779584)[7] and 18q21 (rs4939827, *SMAD7*)[6,8].

GWA studies are not contingent on prior information concerning candidate genes or pathways, and thereby have the ability to identify important variants in hitherto unstudied genes. However, the effect sizes of individual variants, the need for stringent thresholds for establishing statistical significance, and financial constraints on numbers of variants that can be followed up inevitably constrain study power. We recently published two separate GWA studies for CRC. To augment the power to detect additional CRC risk loci, we have conducted a meta-analysis of data from these studies and followed up the best supported associations in large sample sets. This analysis, in conjunction with a replication study using eight independent case-control series, has enabled us to identify four new loci predisposing to CRC. This brings to ten the number of independent loci conclusively associated with CRC risk, and provides additional insight into the genetic architecture of inherited susceptibility to CRC.

RESULTS

Meta-analysis of genome-wide association scans

The GWA studies were both conducted by centers in London and Edinburgh, and were both based on designs involving two-phase strategies and using samples from UK populations (Table 1 and Supplementary Table 1 online). The London phase 1 was based on genotyping 940 cases with familial colorectal neoplasia and 965 controls ascertained through the Colorectal Tumour Gene Identification (CoRGI) consortium for 555,352 SNPs using the Illumina HumanHap550 BeadChip Array. Phase 1 in the Edinburgh study consisted of genotyping 1,012 early-onset (aged ≤55 years) Scottish CRC cases and 1,012 controls for 555,510 SNPs using the Illumina HumanHap300 and HumanHap240S arrays. After applying quality control filters, the following data were available: London phase 1, 547,487 SNP genotypes from 922 familial neoplasia cases (614 with CRC and 308 with high-risk colorectal adenomas) and 927 controls; Edinburgh phase 1, 548,586 SNP genotypes from 980 CRC cases and 1,002 controls.

London phase 2 was based on genotyping 2,873 CRC cases and 2,871 controls ascertained through the National Study of Colorectal Cancer Genetics (NSCCG), whereas Edinburgh phase 2 was based on genotyping 2,057 cases and 2,111 controls. For phase 2, the London and Edinburgh samples were genotyped for a common set of SNPs: the 14,982 SNPs most strongly associated with colorectal neoplasia from London phase 1; the 14,972 most strongly associated SNPs from Edinburgh phase 1 (432 of these SNPs were common to both the London and Edinburgh lists of most strongly associated SNPs); and 13,186 SNPs showing the strongest association with CRC risk from a joint analysis of all CRC cases and controls from both phase 1 data sets (that were not already included in any of the preceding categories). Therefore, phase 2 was based on genotyping 42,708 SNPs in total. After applying quality control filters, the following data were available: London phase 2, 38,715 polymorphic SNPs in 2,854 cases and 2,822 controls; Edinburgh phase 2, 38,710 polymorphic SNPs in 2,024 cases and 2,092 controls. Overall, there were 38,710 polymorphic SNPs common to all four data sets (phases 1 and 2 in London and Edinburgh).

Prior to undertaking the meta-analysis of phases 1 and 2, we searched for potential errors and biases in the four case-control series.

[1]A full list of authors and affiliations is provided at the end of this paper.

Received 6 August; accepted 17 September; published online 16 November 2008; doi:10.1038/ng.262

Established prior to Meta-Analysis in Colorectal Cancer

- 8q24
  - Also Prostate Region 3
  - Colorectal Adenoma
- 8q23.3 (*EIF3H*)
- 10p14
- 11q23
- 15q13
- 18q21 (*SMAD7*)

New Loci

- 14q22 (*BMP4*)
- 16q22.1 (*CDH1*)
- 19q13.1 (*RHPN2*)
- 20p12.3

Novel GWAS Findings in Cancer

- Lung
  - 15q24/25.1 (smoking behavior)
  - 5p15.33 (*TERT or CLPTM1L*)
  - 6p21.33
- Melanoma
  - 20q11.22 (*ASIP*, also basal cell carcinoma)
  - 11q14-q21 (*TYR*, also basal cell carcinoma)
- Neuroblastoma
  - 6q22

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.

Next Steps?

- **Mapping Loci (>45 in cancer alone)**
  - **Identify "best variants" for biological investigation**
- **Functional Analysis of Variants**
  - **Provide plausibility**
  - **Explore pathways**
- **New GWAS Needed for Outcomes**
  - **So far, etiology and outcome have very little overlap**
  - **Different regions drive cancer progression**
- **GWAS for Etiology in Comparable Environmental Exposures**
  - **Role of gene-environment interaction(s)**

Assessing GWAS Findings

- **Risk Assessment-**
  - **What Constitutes Suitable Reporting of Data**
    - **Absolute Risk vs. Relative Risk**
- **Pharmacogenomics**
  - **New GWAS Studies**
- **Value for Public Health and Personal Decisions**
- **Consider New and Old Paradigms**
    - **Capitalize on Available Studies**
    - **Biospecimens on ALL current and new studies**

Acknowledgements

CGEMS & DCEG

Gilles Thomas
Robert Hoover
David Hunter
Richard Hayes
Joseph Fraumeni
Daniela Gerhard
Demetrios Albanes
Kevin Jacobs
Meredith Yeager
Sholom Wacholder
Nilanjan Chatterjee
Zhaoming Wang
Xiang Deng
Nick Orr
Regina Ziegler
Kai Yu
Margaret Tucker
Jesus Gonzalez Bosquet
Montse Garcia-Closas

HSPH
Peter Kraft
David Cox
Sue Hankinson
Fred Schumacher

PLCO
Chris Berg
E David Crawford
Catherine McCarty

WHI
Ross Prentice
Rebecca Jackson
Charles Kooperberg
Rowan Chlebowski

PBCS
Jolanta Lissowska
Beata Peplonska

ACS
Michael Thun
Heather Feigelson
Eugenia Calle
Ryan Diver

CeRePP, France
Olivier Cussenot
Geraldine Cancel-Tassin
Antoine Valeri

NPHI, Finland
Jarmo Virtamo

Wash. U., St Louis
Gerald Andriole
Adam Kybell