

# National Cancer Advisory Board

## Genome-Wide Association Studies and the Road Ahead

**Stephen Chanock, M.D.**

Chief, Laboratory of Translational Genomics  
Director, Core Genotyping Facility  
Division of Cancer Epidemiology and Genetics

December 7, 2010

# Basic principle of genetic *association* studies in unrelated individuals

## Gene A

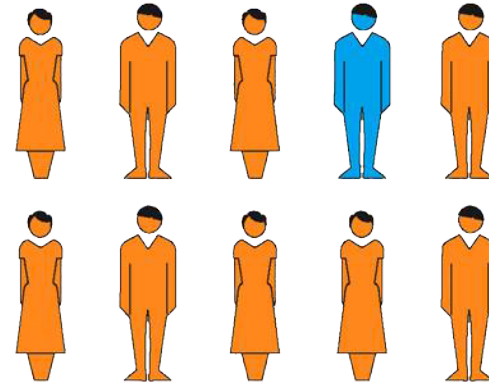


Affected

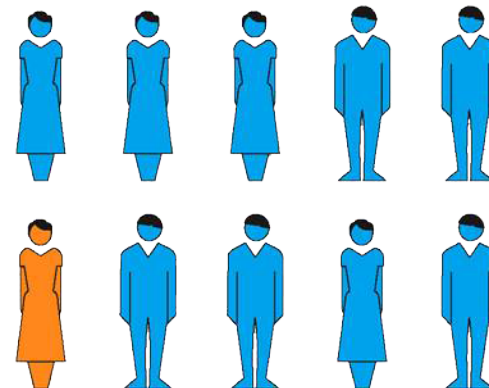


Unaffected

## Gene B

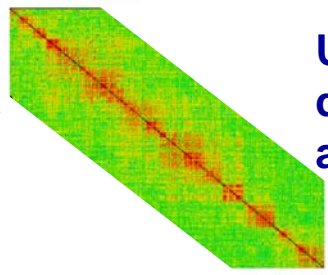
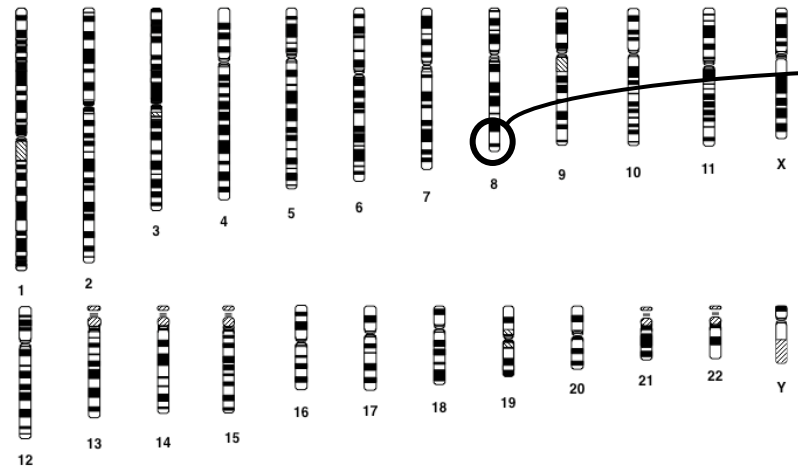


Affected



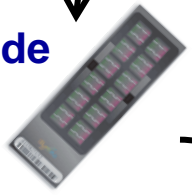
Unaffected

# Millions of common SNPs across the genome



Use linkage disequilibrium across the genome

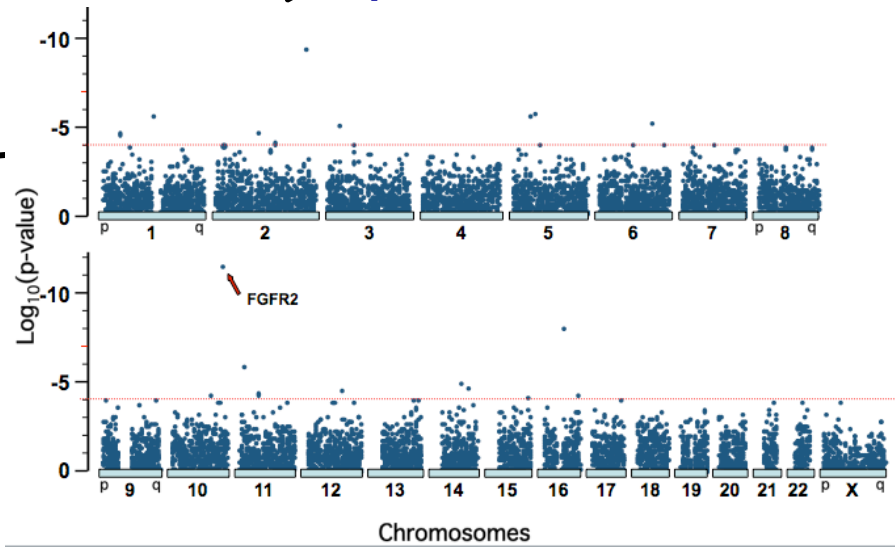
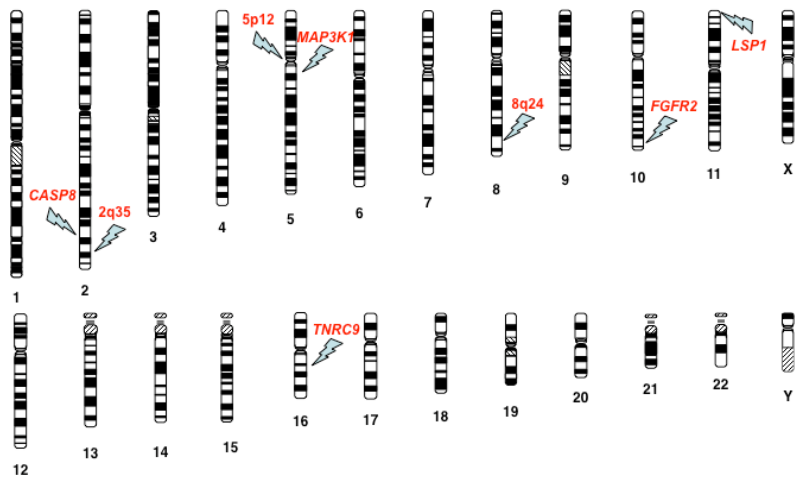
Selection of surrogate SNPs to capture common genetic variation



Genome wide SNP chips

Association with disease risk in discovery followed by replication studies

## Mapping of susceptibility loci



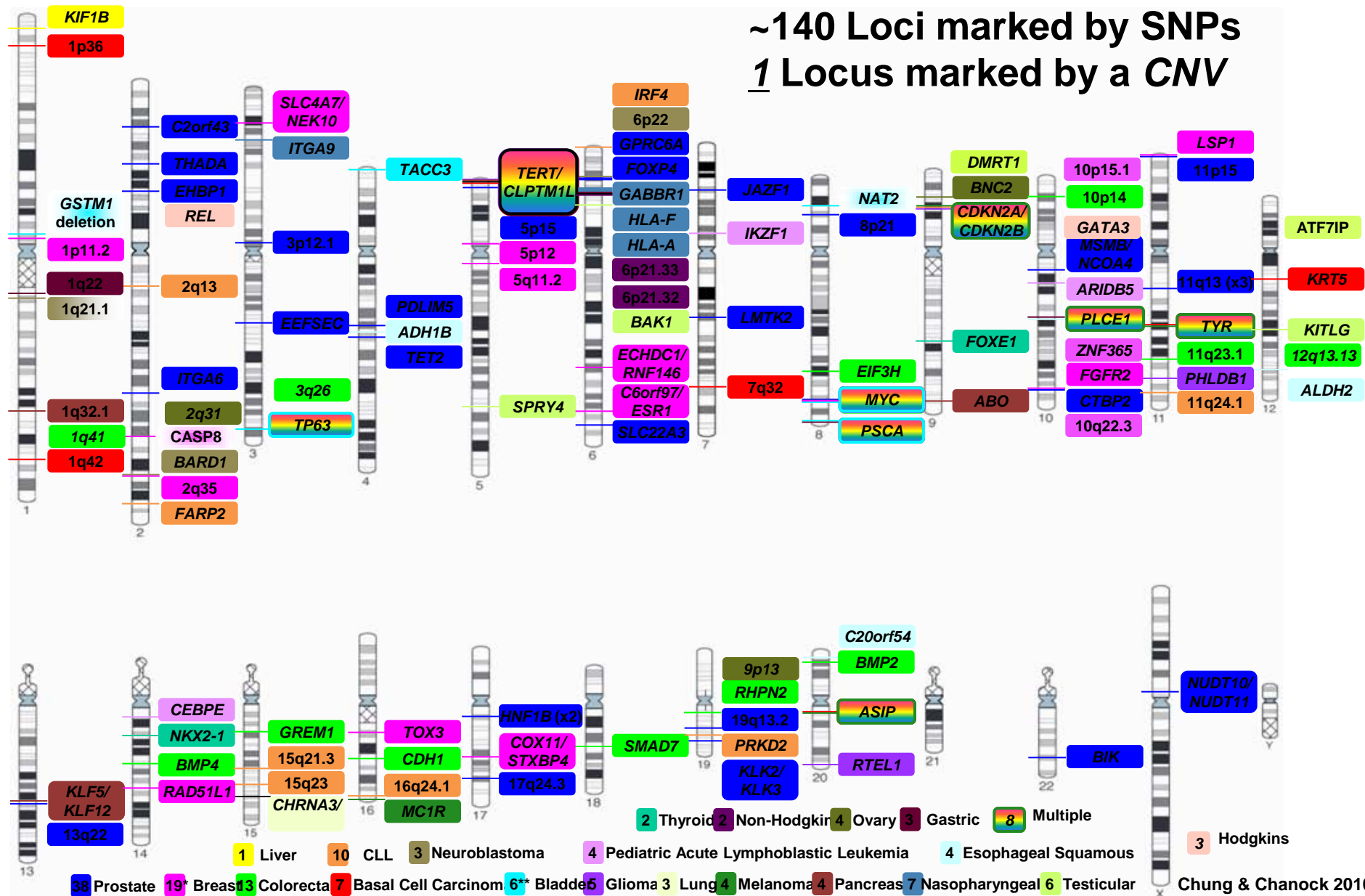
# “What’s past is prologue”

*William Shakespeare*

- In the mid 60’s, DCEG began collection of biospecimens from high risk families.
- This approach was extended in the 80’s to high-quality population studies.
- Scientific progress results from application of new technology.
- Remarkable opportunity for discovery of germ-line contributions to cancer.

# Published GWAS Etiology Hits: 12.1.10

~140 Loci marked by SNPs  
1 Locus marked by a CNV



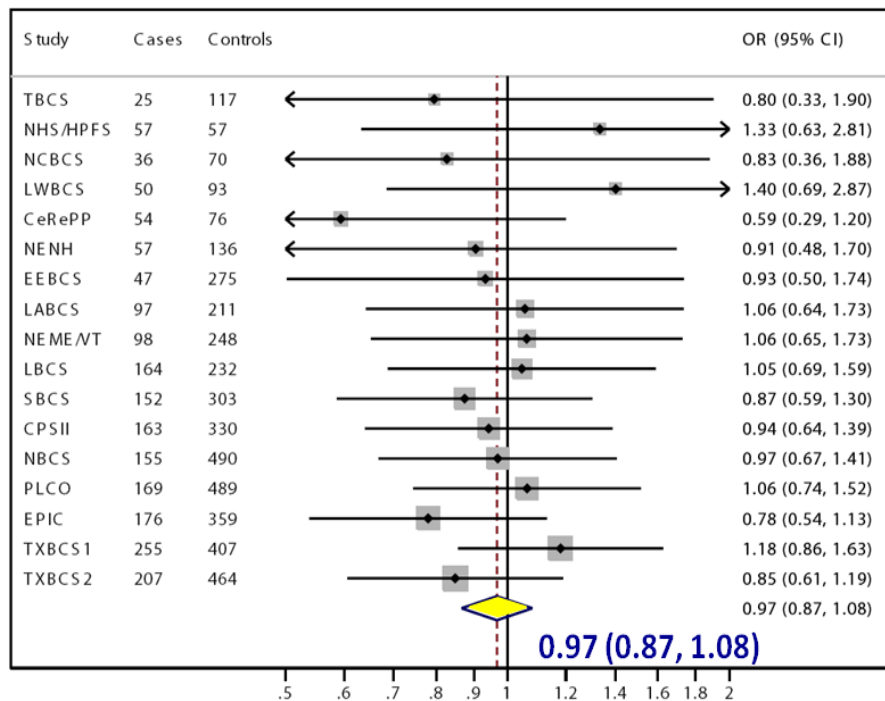
# Profile of Cancer GWAS Regions

---

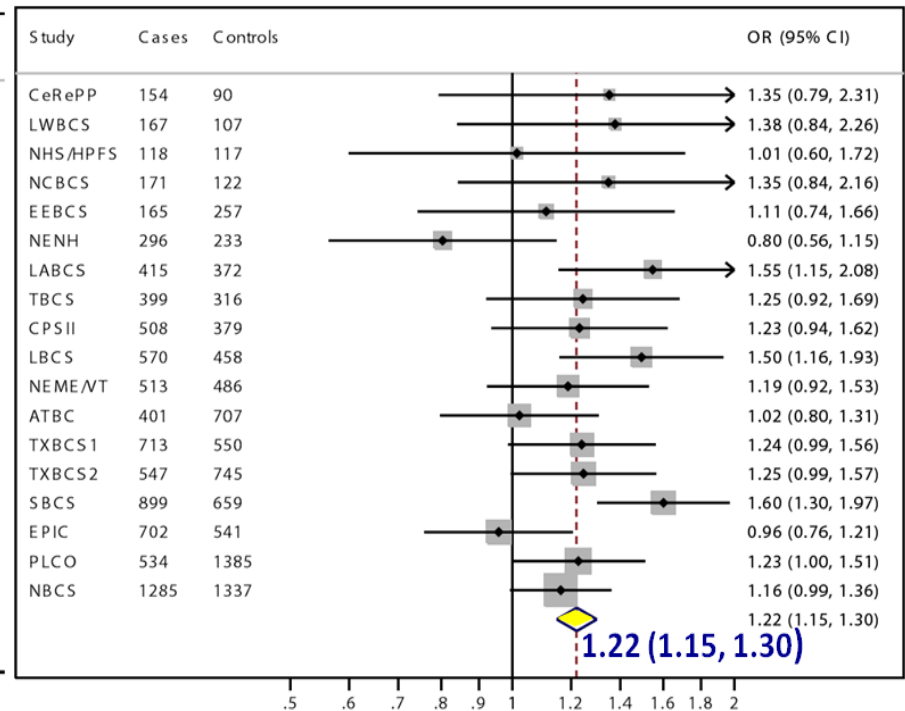
- Nearly all map to non-coding regions of the genome
- ~20% map to 'gene' deserts
- Estimated Effect Sizes (Odds Ratios) are small ( $< 1.5$ )
  - Exception is *KITLG* in Testicular Cancer
- Multi-Cancer Susceptibility Regions  
Examples: 8q24, 5p33.15 (*TERT*), 11q13

# Gene-Environment Interaction for Bladder Cancer Risk: Large International Study Shows *NAT2* Slow Acetylation Increases Risk Only for Smokers

## Never Smokers



## Ever Smokers



**P-interaction =  $2.8 \times 10^{-4}$**

**Rothman et al., *Nat Genet*, 2010**

# Prostate Cancer Risk Factors 2006

---

- Age
- Ethnic Background
- Family History





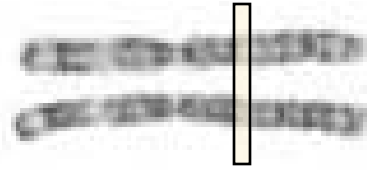
# Prostate Cancer Risk Factors 2010

---

- **Age**
- **Ethnic Background**
- **Family History**
- **Multiple common alleles**
  - *Each common variant provides a small contribution*
- **Inverse Relationship of Type 2 Diabetes & Prostate Cancer**

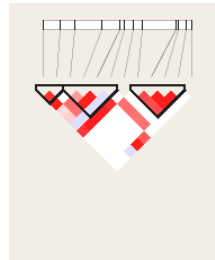
# Follow-up of GWAS Hits

1. Association Finding



GWAS or Linkage

2. Mapping Region



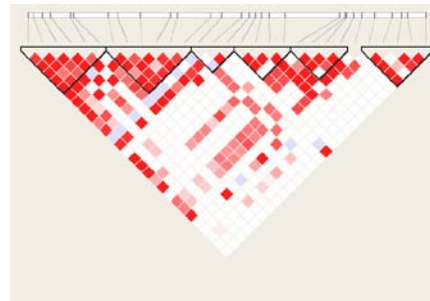
Regional Re-Sequence  
HapMap  
1000 Genomes Data

3. Assessment of Variants

```
ATGGAAAGGGAGCATATTACGCACTATGAAGCGGGGGAACGA
AGCATATTATGCACTATGAAGC
AAGGGAGCATATTACGCACTATGAAGCGGAGGAACGA
ATGGAAAGGGAGCATATTATGCAC GAGGAACGA
ATGGAAAGGGAGCATA ATGAAGCGGGGGAACGA
AAAGGGAG GCATATTACGCACTATGAAG AACGA
TATTATTGCACTATGAAGCGGAGGAACGA
ATGGAAAGGGAGCAT TATTATGCACTATGAAGCGGAGGAACGA
ATGGAAAGGGAGCATATTACGCACTAT AACGA
```

Choose Variants for Testing

4. Secondary Association

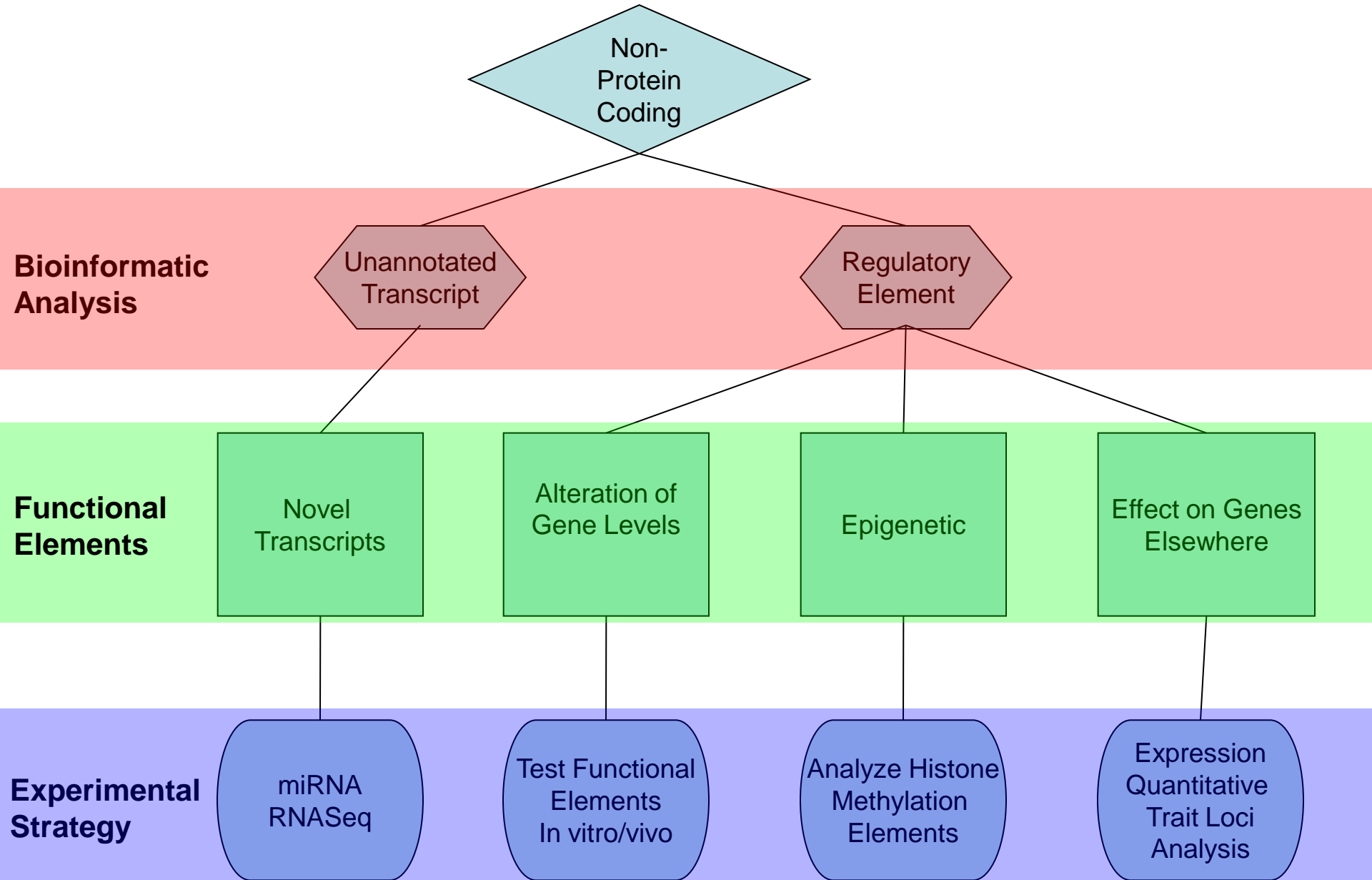


Follow-up Genotyping in Large Studies



**Nominating Optimal Variants for Functional Studies**

# The Majority of GWAS Map to Non-Coding Regions

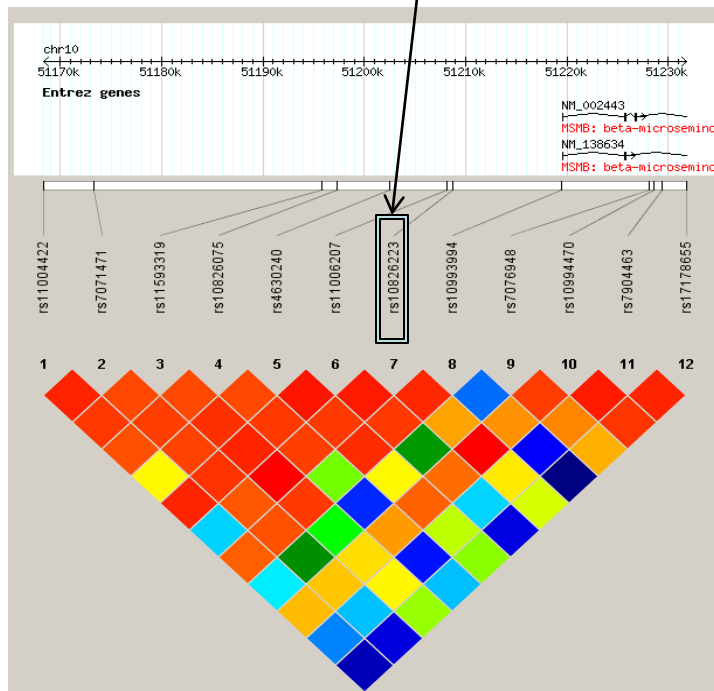


# 10q11.2 & Prostate Cancer

**Top Hit: rs10993994**  
**Promoter of *MSMB***

***MSMB* =  $\beta$ -microseminoprotein**  
**Prostate specific**  
**Serum marker in studies**

**20,000 subjects**



## Functional Analysis

**Risk Allele “T”**

**Lower expression levels**

**Reporter assays**

**Electromobility Shift Assays**

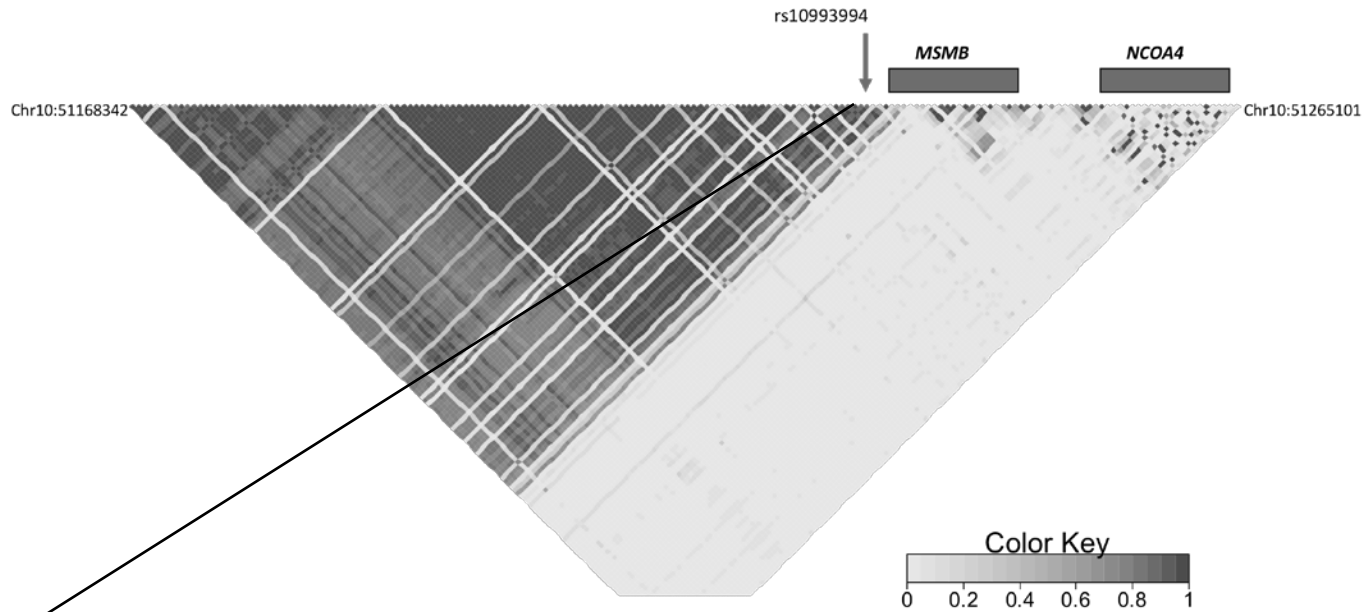
**Levels in Prostate Tissue**

**Tumor Tissue**

**TT Decrease Expression**  
**in NCI-60 Cell lines**

# 10q11.2 Could Be More Complex.....

## *MSMB and NCOA4*



RNA Expression *MSMB* and *NCOA4*

Normal



Tumor Tissue

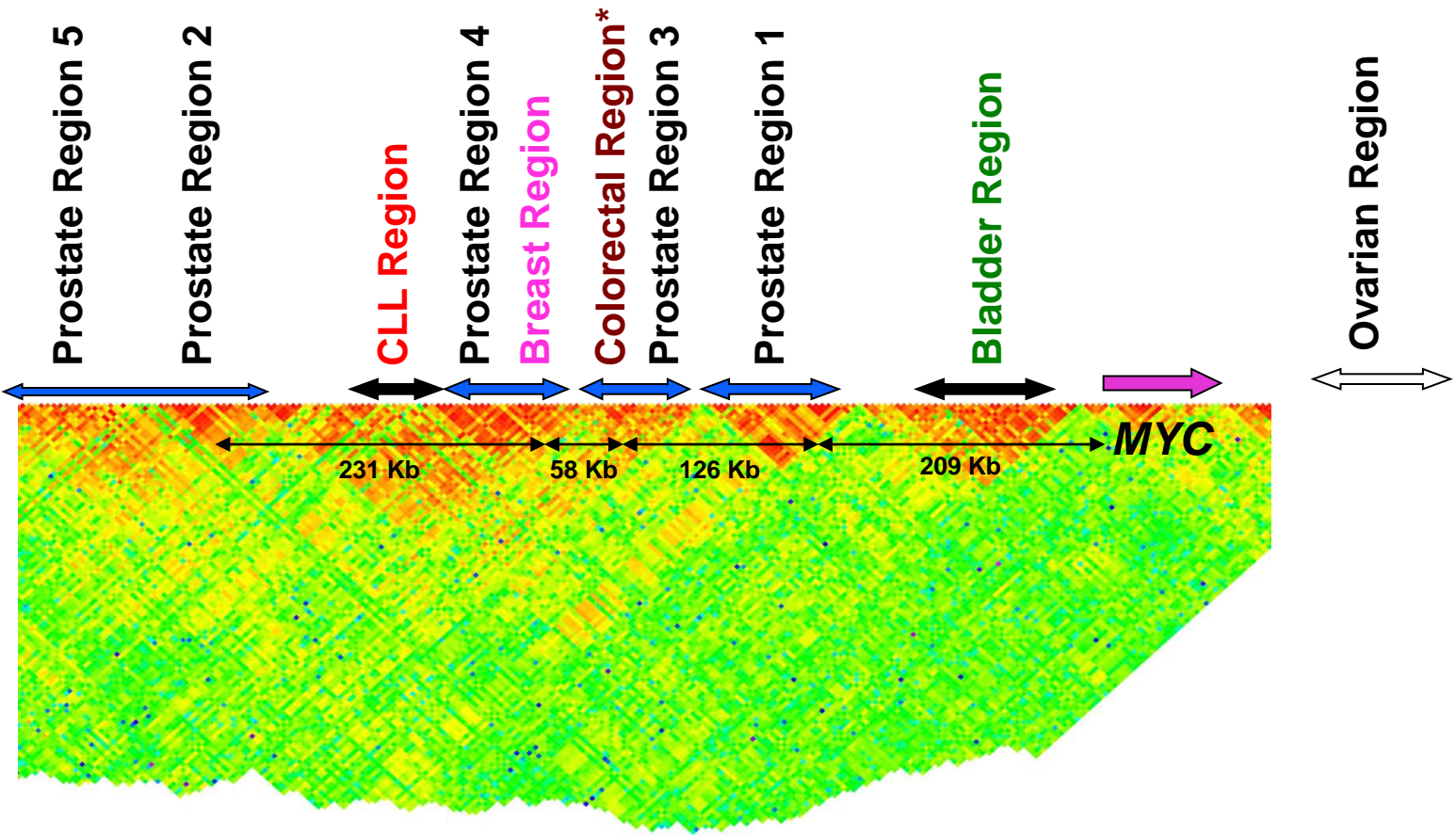


Anchorage Independent Growth is Specific to Prostate

*MSMB*- Suppression

*NCOA4*- Over-expression

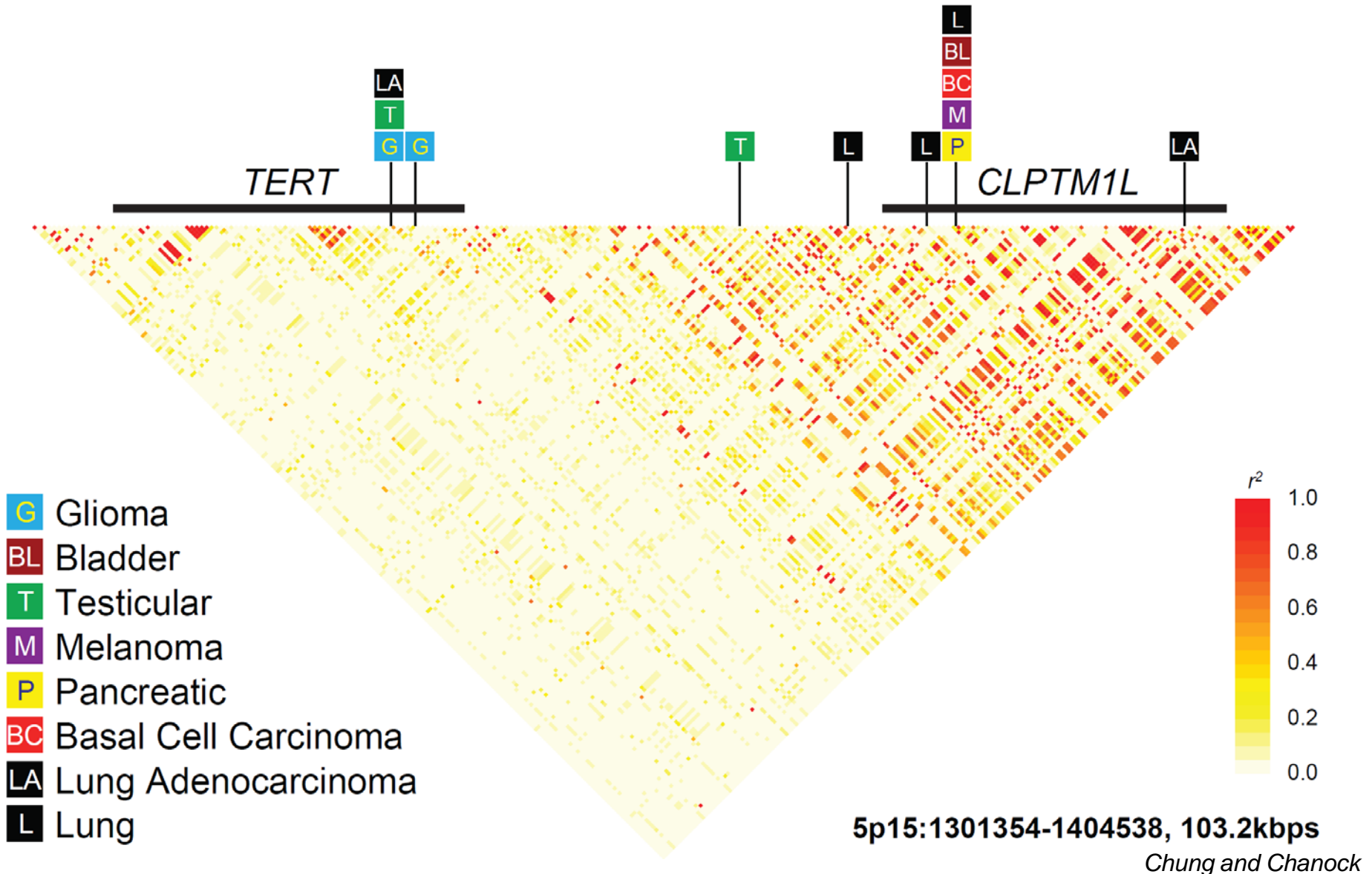
# Cancer Susceptibility Loci Across 8q24



\* Also Colorectal Adenoma

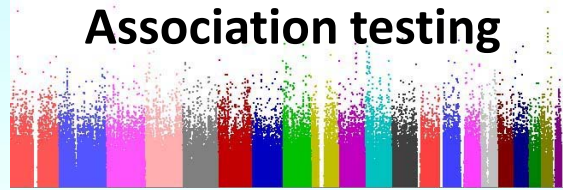


# GWAS Identifies 7 Cancers Associated with *TERT-CLPTM1L* Region on Chromosome 5p15.33





## Association testing



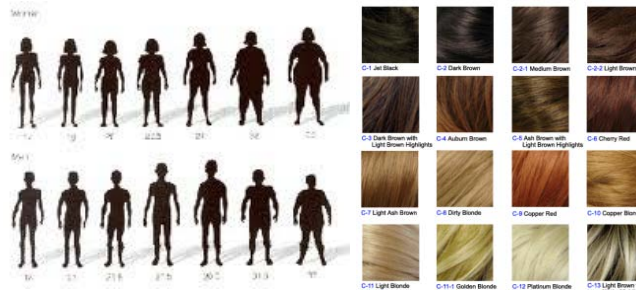
## Genome-wide association studies

### Behavioral traits



Tobacco  
Caffeine  
Alcohol

### Biometrics



Height, Weight, BMI,  
Menarche/Menopause

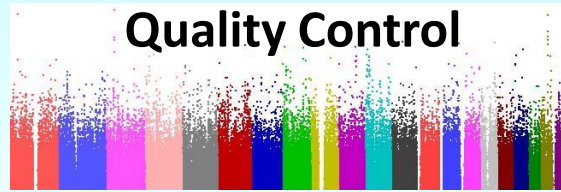
### Nutrient levels



Vitamins D, B<sub>12</sub>  
Carotene

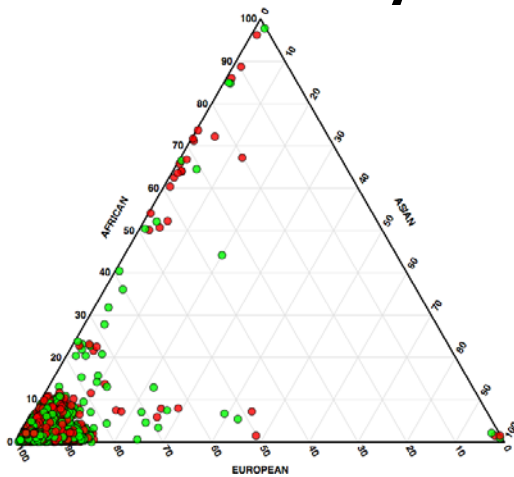
International Consortia (e.g., GIANT, SUNLIGHT)

# Quality Control

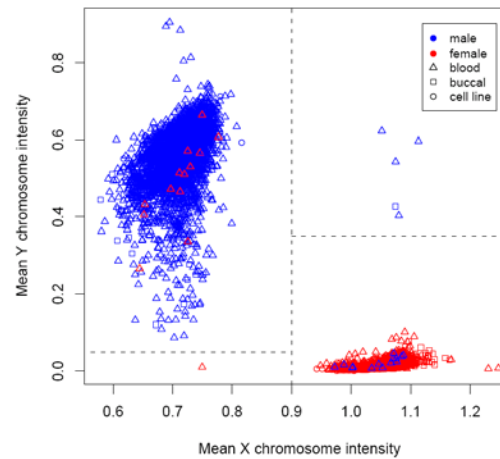


# Genome-wide association studies

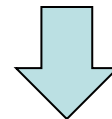
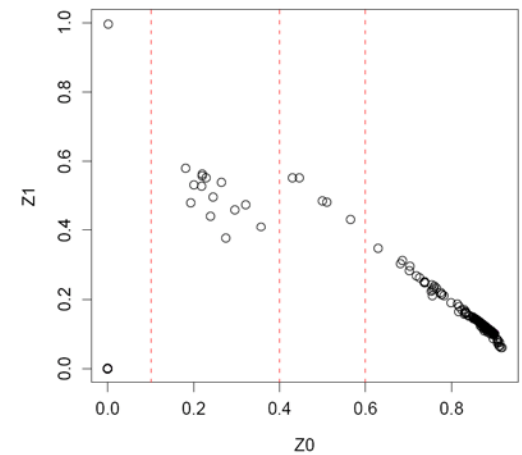
## Population genetics and ancestry



## Sex verification

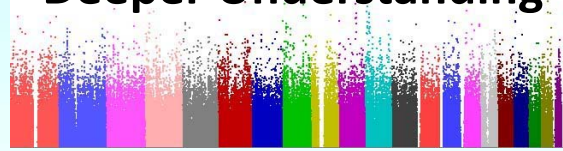


## Relationship testing



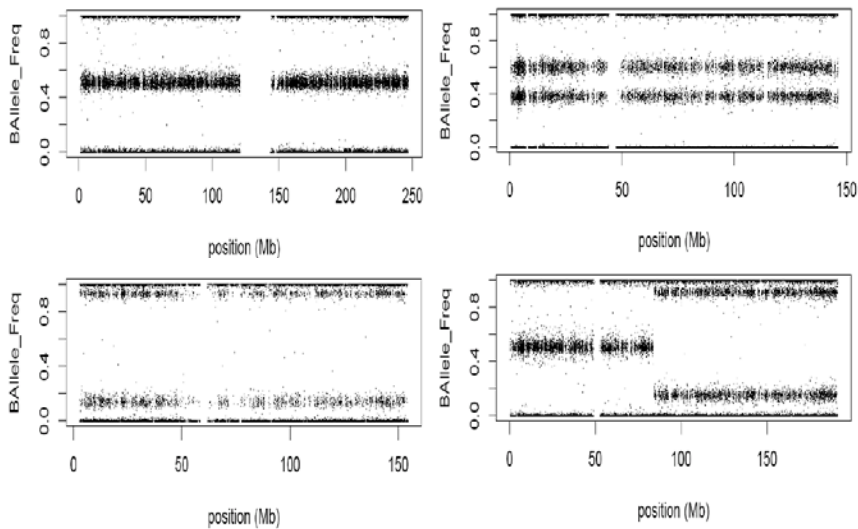
Pursuit of X-Chromosome Aneuploidy and Cancer Risk

# Deeper Understanding



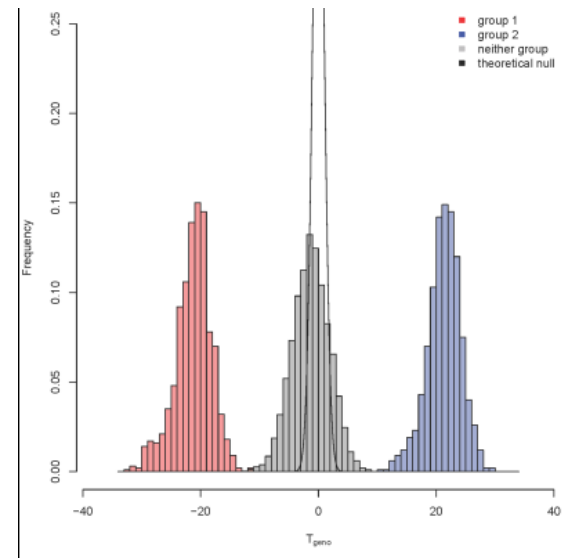
## Genome-wide association studies

Large chromosomal abnormalities, structural variation, aneuploidy in Germ-line DNA



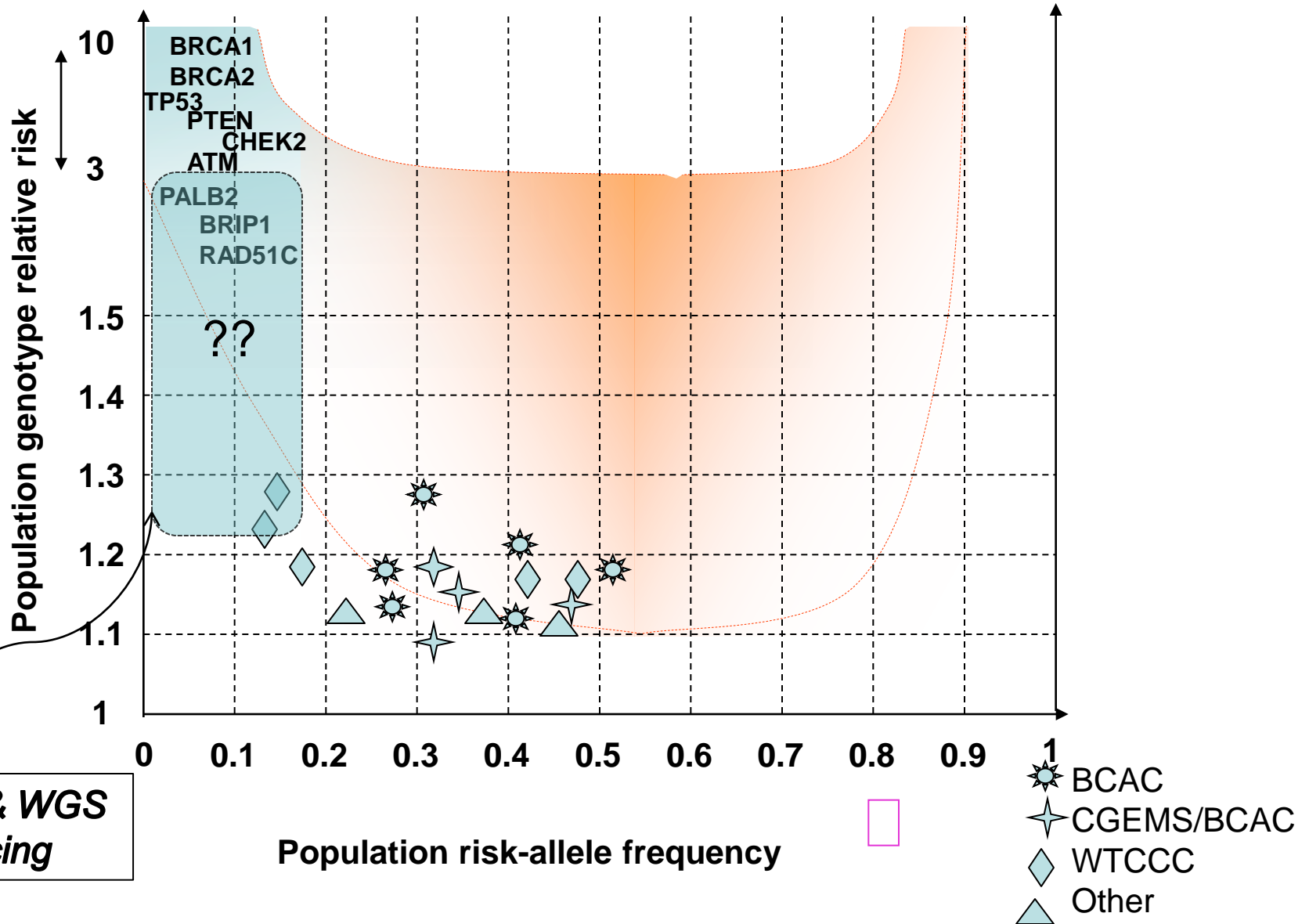
*Rodriguez-Santiago AJHG 2010*

Privacy & Confidentiality  
GWAS membership



*Jacobs Nature Genetics 2009*

# Genetic Predisposition to Breast Cancer European Population



# Risk prediction utility

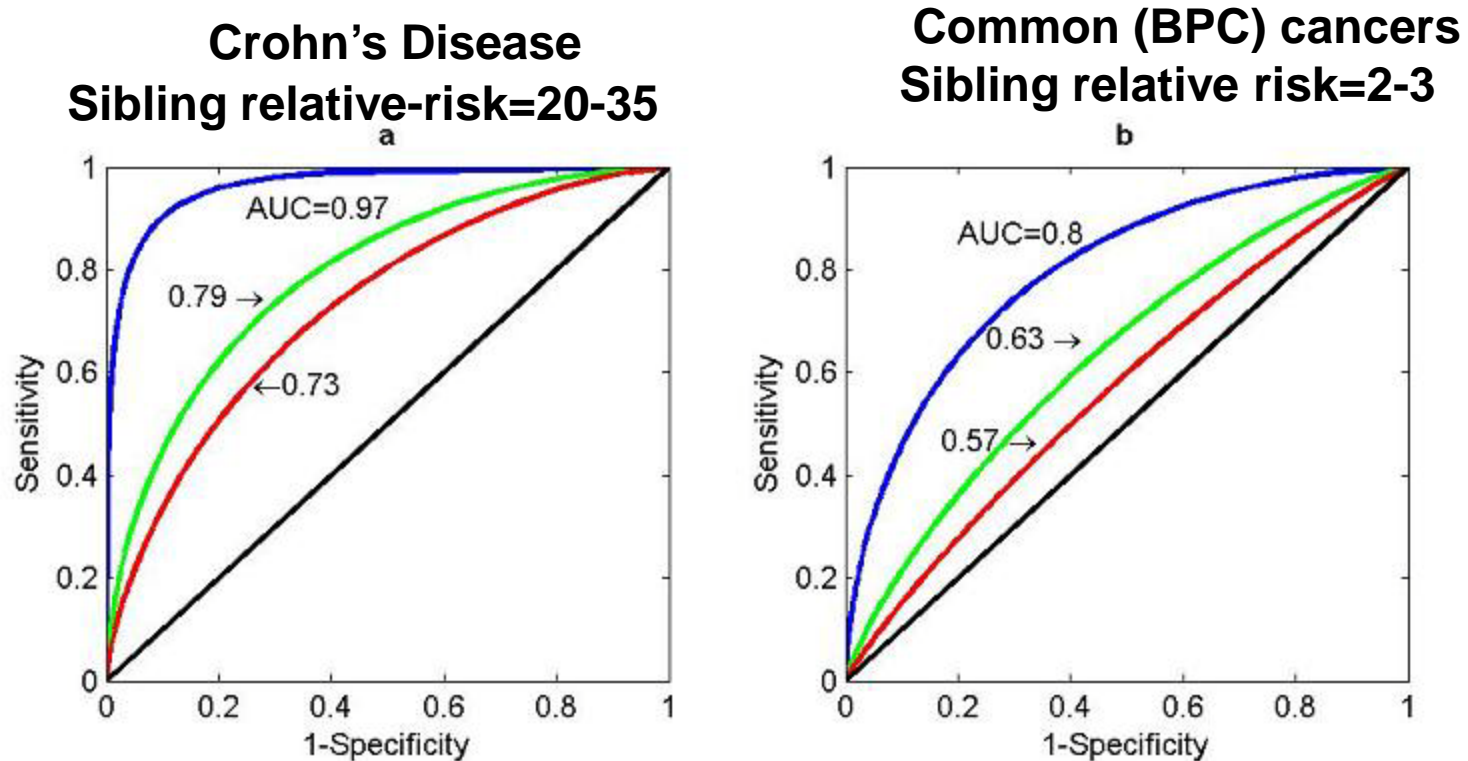
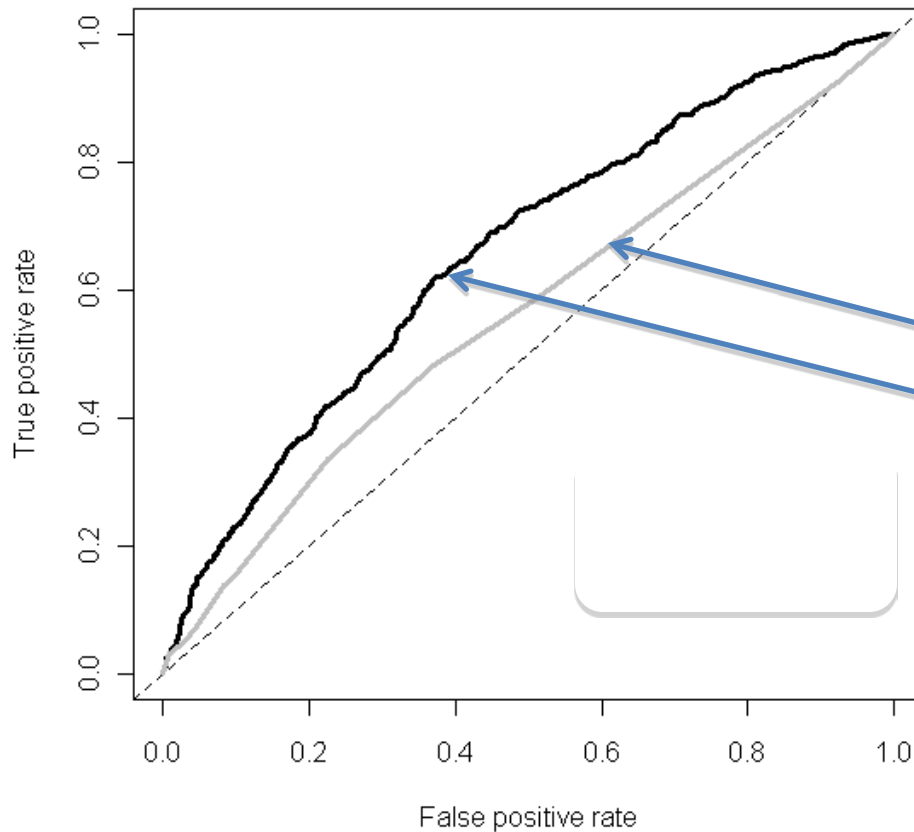


Figure 2 Receiver operator characteristics (ROC) curves for genetics risk models.

- Random
- Using known loci
- Using all estimated loci
- Ideal (if we could explain all heritability)

# Predicting Prostate Cancer Risk



**35 common SNPs from GWAS**  
**Men between 60 and 65 years**  
**NCI Breast and Prostate Cohort**  
**Consortium**

**Area Under the Curve**

**FH Alone= 0.52**

**FH + 35 SNPs= 0.66**

**ROC=Receiver Operator Curve**

# Status of Genome-Wide Association Studies

---

- ❖ **Discovery of New Regions in the Genome Associated with Diseases/Traits**
  - *New “Candidate Genes”*
- ❖ **Clues for Mechanistic Insights Using Common Variants**
  - *Etiology*
  - *Gene-Environment/Lifestyle Interactions*
  - *Outcomes & Pharmacogenomics*
- ❖ **Challenge of Genetic Markers for Risk Prediction for Individual or Public Health Decisions**
  - *Common Variants Represent a Fraction of the Genetic Contribution to Risk*
  - *Integration of Lifestyle/Environment*

# Acknowledgements



## NCI-DCEG

Joseph Fraumeni  
Gilles Thomas  
Robert Hoover  
Meredith Yeager  
Kevin Jacobs  
Nilanjan Chatterjee  
JuHyun Park  
Sholom Wacholder  
Mila Prokunina-Olsson  
Laufey Amundadottir  
Peggy Tucker  
Charles Chung

## NCI- OCG

Daniela Gerhard

## NCI-CCR

Mike Dean  
Hong Lou

## HSPH

David Hunter  
Pete Kraft

## CGEMS Studies

ACS (M Thun)  
ATBC (D Albanes)  
CAPS (H Gronberg/J Xu)  
CeRePP (O Cussenot)  
CONOR (L Vatten)  
EPIC (E Riboli)  
JHU (W Issacs/J Xu)  
MEC (B Henderson)  
PLCO (R Hayes)  
WHI (R Prentiss)

## DFCI

Matt Freedman  
Mark Pomerantz

## Bladder Cancer GWAS

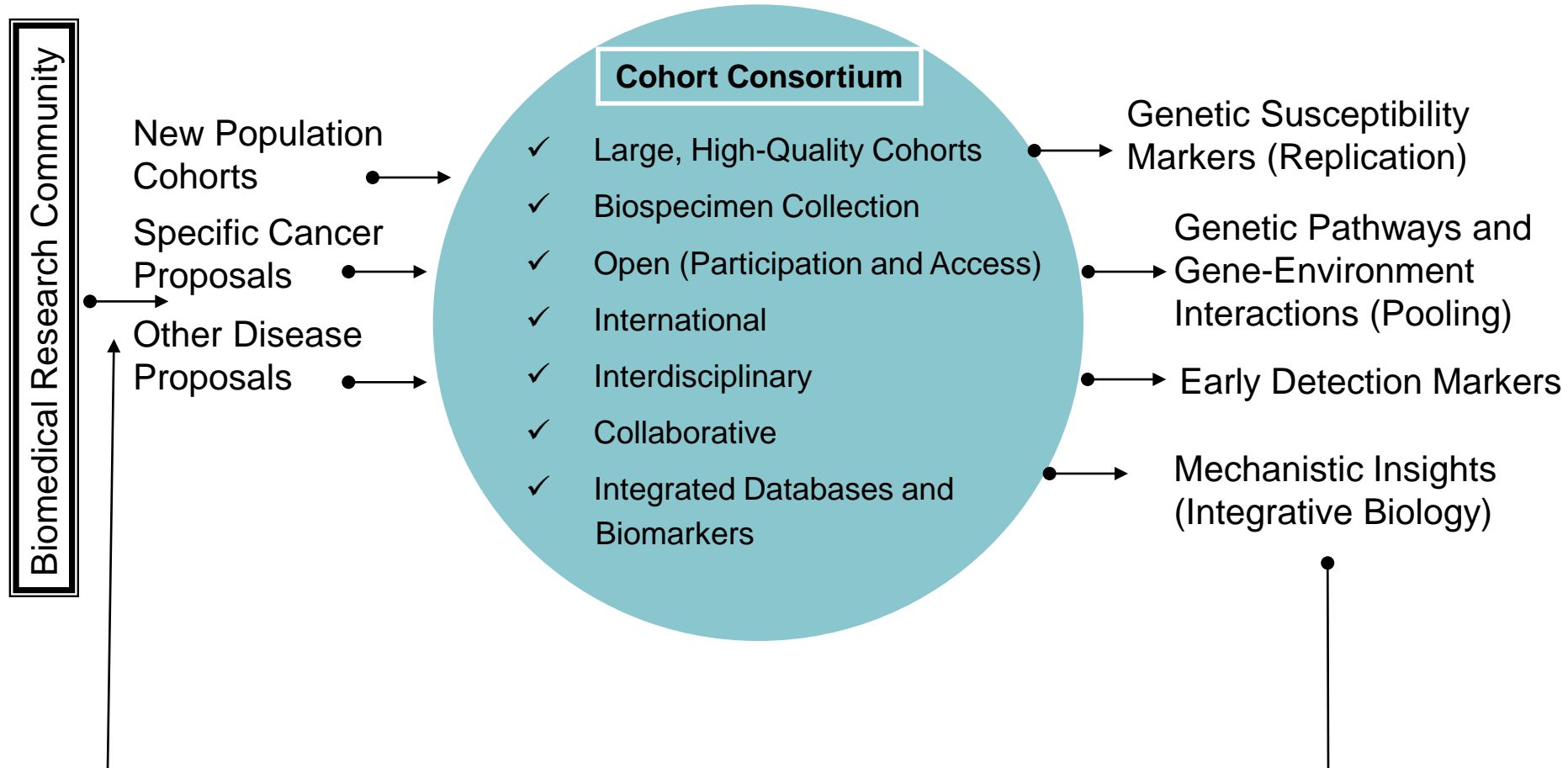
Nat Rothman  
Debra Silverman  
Montse Garcia-Closas

## Renal Cancer GWAS

M Purdue  
P Brennan  
M Lathrop  
M Johansson



# Cohort Consortium Strategic Plan



# Mapping Complex Diseases: Identifying Genetic Markers

**Comprehensive Catalog of Genetic  
Variation Enables 'Agnostic' Search**

## **Frequency of SNP Markers**

Common	>10%
Uncommon	1-10%
Rare	< 1%

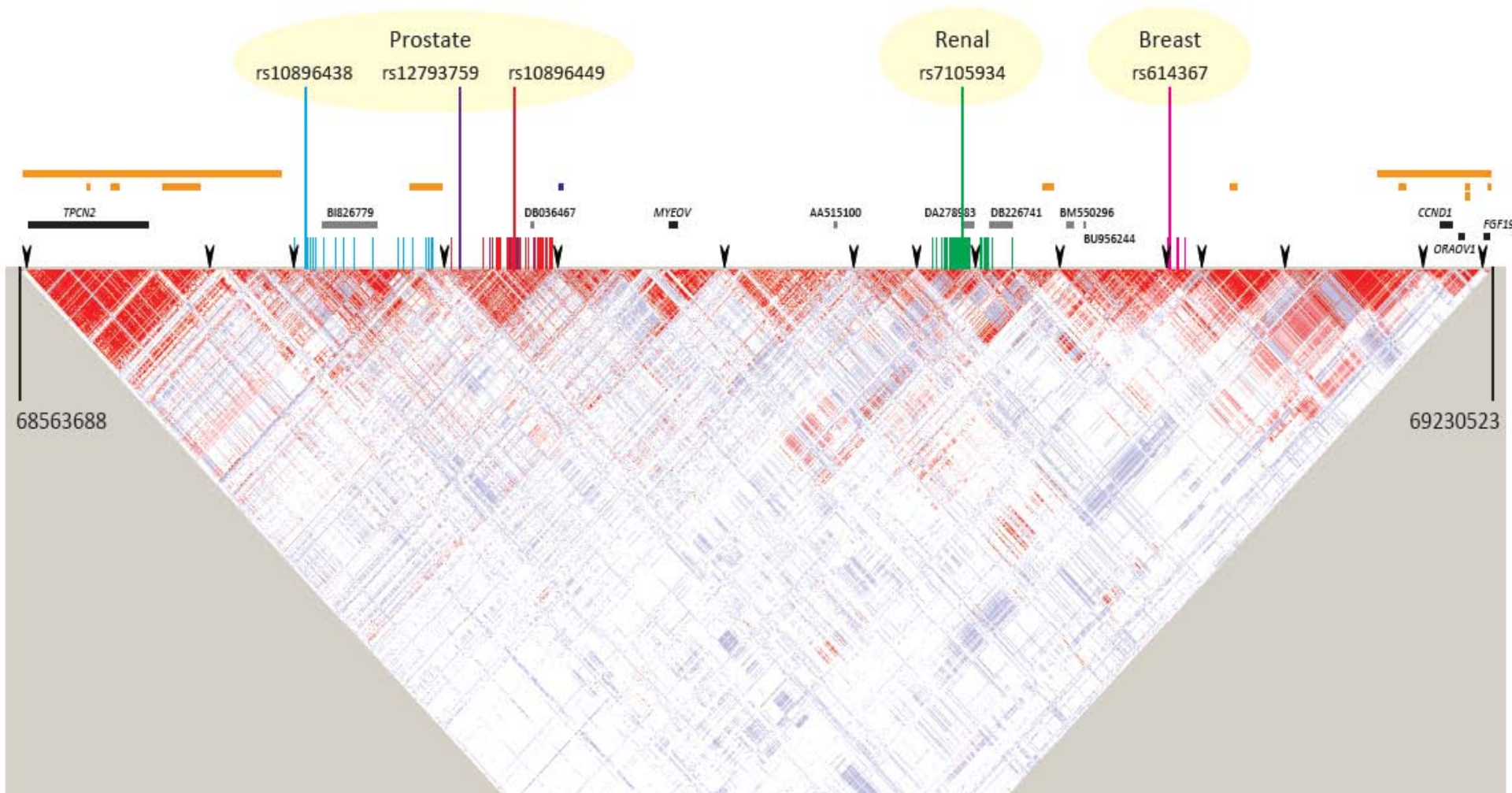
## **Study Design for Discovery**

GWAS
GWAS or Sequencing
Linkage or Sequencing

**SNP= Single Nucleotide Polymorphism**

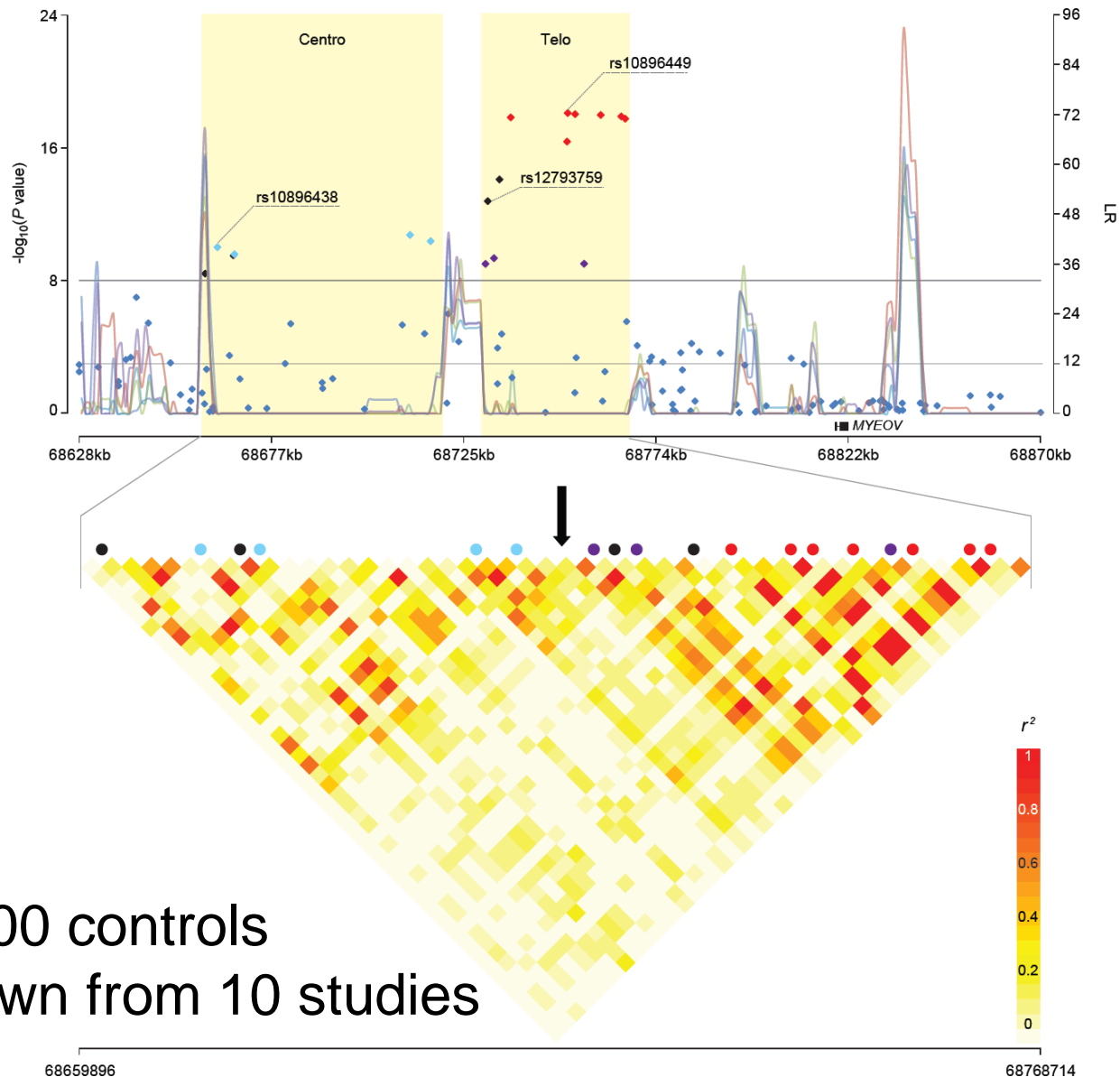


# 11q13: Multi-Cancer Susceptibility Region



660 kb across the region

# Recombination hotspots and LD structure in 11q13: GWAS Region with no candidate genes



Sampled 900 controls  
5 times drawn from 10 studies

68659896

68768714

# 2.5 M SNP Chip

- Beta tested at CGF
- High performance at CGF (~750 samples)
  - Completion rates of 99% with concordance > 99%
- 2.5 M Array
  - Average MAF=10.4% (range 3-15%)
  - Based on 3 algorithms (UofM, Broad and WTSI)
  - New common regions not covered on prior chips
- 2.5 M S (2<sup>nd</sup> half of 5 M)
  - Design begins in November 2010
  - Projected beta test in Spring 2011

# ***IMPUTATION STUDY @ CGF***

750 Caucasians (250 of ACS, ATBC & PLCO)

2.5 M SNP array

1.0M Omni array

1.0M Duo or 660w array

Affymetrix 6.0

100 AA & 75 Chinese with 2.5 M and 1 M/660w

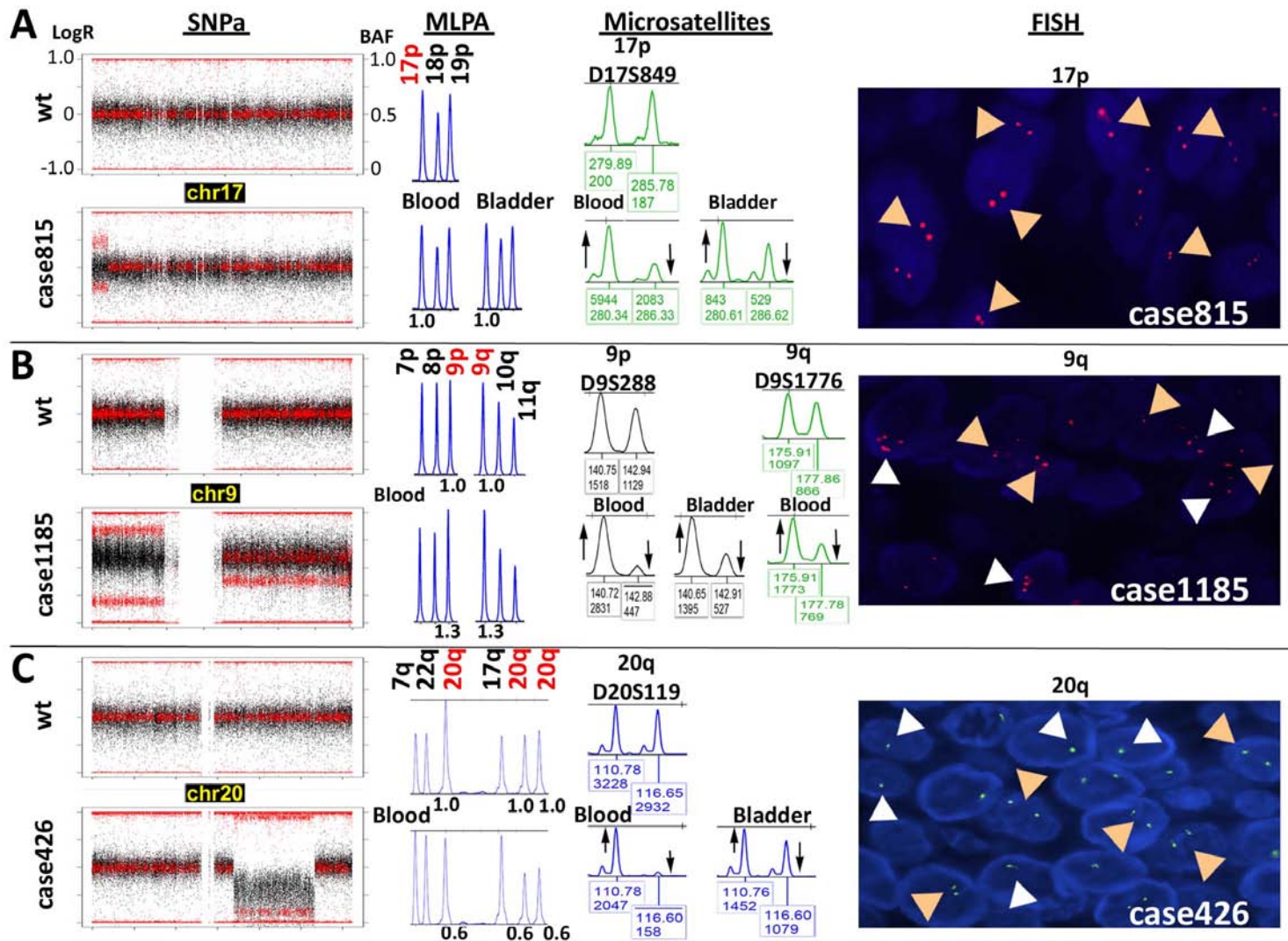
Deposit in dbGaP in Winter of 2010

Approximately 3.4 M genotyped SNPs in CEU for  
analysis with 1000 CEU data (new download)



# Validation for 42 events: 100% Validation

Allelic imbalance &  
 - normal dosage UPD  
 - abnormal dosage CNV



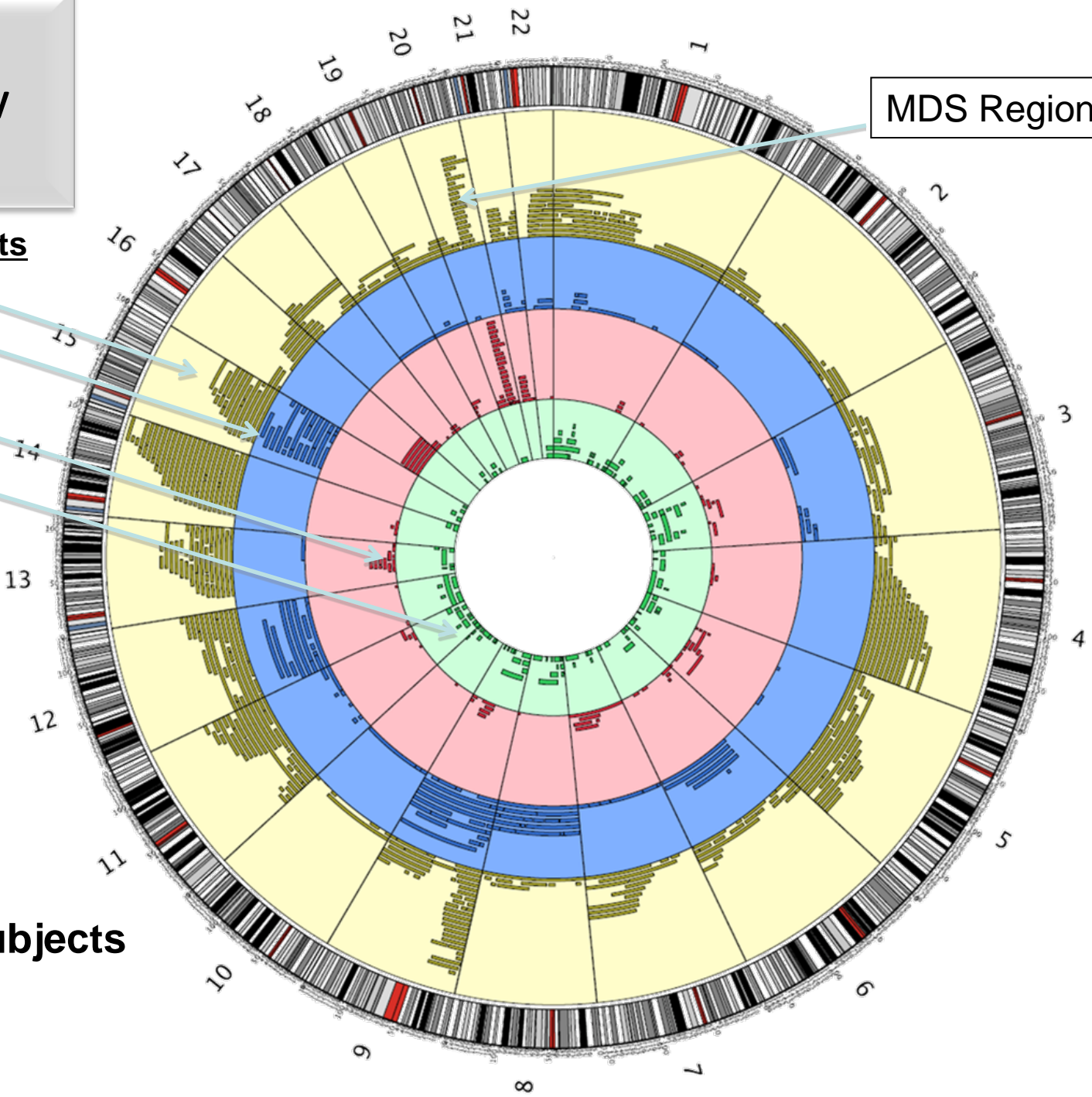
5 / 6 events: also found in bladder tissue; early developmental events



**Large Structural  
Events (>5 Mb)  
by  
Chromosome**

MDS Region

- Germline Mosaic Events**  
Copy Neutral LOH  
Duplication/trisomy  
Deletion  
UPD



**50,520 Elderly Subjects**  
**≥550,000 SNPs**  
**Illumina Infinium**