Genome-Wide Association Studies in Cancer:

Stephen Chanock, M.D.  
Chief of Laboratory of Translational Genomics,  
DCEG, NCI  
Director, Core Genotyping Facility, DCEG, NCI  
June 4, 2008
"What ever will we think about now that the genome project is almost complete?"
Disease Status

Age of Onset  Severity  Clinical Symptoms

Genetic Risk Factors
Genetic Protective Factors

Environmental Risk Factors
Environmental Protective Factors

Fully Testable  Where to Begin??
Promise of GWAS

• Discovery of Common Markers in the Genome
  – ‘Represents a portion of the genetic contribution’
• Opportunity to explore mechanism of biology
  – How and why cancer develops
• Outcomes
  – Etiology
  – Gene-Environment/Lifestyle Interactions
  – “Druggable” targets
• Establish genetic markers for:
  – Prevention
  – Intervention
Identifying Genetic Markers for Prostate & Breast Cancer

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
Circa..2006

The Enigma of a Common Disease

• Age
• Ethnic Background
• Family History
• One SNP- unknown function
  – Rs1447295 @ region 8q24 (no obvious gene)
CGEMS Prostate Cancer GWAS: Where are the True Signals Amidst the Blizzard of False Positives

Chromosomes

Log_{10}(p-value)

New Region
General Strategy for Prostate GWAS is Based on Replication, Replication, Replication

Initial Study
1150 cases/1150 controls → 540,000 Tag SNPs

Follow-up Study #1
4000 cases/4000 controls → >28,000 SNPs

Follow-up Study #2
5500 cases/5500 controls → at least 7,600 SNPs

Fine Mapping → 10 ±5 loci

Genotype, Haplotype, Sequence

Cohorts
- PLCO
- ACS/ATBC/HPFS/FrCC/PHS
- MEC/EPIC/JHU/SwCaP

Determine Causal Variant(s)
Selection of the SNPs to be taken to stage 2
Determining Real-estate to find the FEW true positives

25,358

1-SNP statistics (pair-wise r^2 < 0.8), p_val < 0.068,

1,913

2-SNP statistics => p_val of previously selected SNP decreased at least 10 fold

1,508

Population stratification SNPs

897

Various candidate regions including 8q24

SNPs distributed in 7608 distinct chromosomal regions
In a region the maximal distance between two adjacent SNPs is less than 100Kb

CGEMS prostate cancer stage 2
### 7 associated loci in CGEMS Prostate Cancer

<table>
<thead>
<tr>
<th>Region</th>
<th>p-value</th>
<th>Risk Allele Freq.</th>
<th>Odds ratios</th>
</tr>
</thead>
<tbody>
<tr>
<td>8q24 (loc1)</td>
<td>6.7 $10^{-16}$</td>
<td>0.1</td>
<td>Heterozygotes: 1.49 (1.34-1.64)</td>
</tr>
<tr>
<td>10q11</td>
<td>8.7 $10^{-14}$</td>
<td>0.38</td>
<td>Heterozygotes: 1.20 (1.10-1.31)</td>
</tr>
<tr>
<td>8q24 (loc2)</td>
<td>4.7 $10^{-13}$</td>
<td>0.50</td>
<td>Heterozygotes: 1.13 (1.02-1.26)</td>
</tr>
<tr>
<td>17q12</td>
<td>1.5 $10^{-10}$</td>
<td>0.52</td>
<td>Heterozygotes: 1.25 (1.13-1.34)</td>
</tr>
<tr>
<td>11q13</td>
<td>4.1 $10^{-10}$</td>
<td>0.50</td>
<td>Heterozygotes: 1.18 (1.08-1.28)</td>
</tr>
<tr>
<td>10q26</td>
<td>1.7 $10^{-7}$</td>
<td>0.25</td>
<td>Heterozygotes: 1.14 (0.94-1.38)</td>
</tr>
<tr>
<td>7p15</td>
<td>3.2 $10^{-7}$</td>
<td>0.76</td>
<td>Heterozygotes: 1.18 (1.07-1.31)</td>
</tr>
</tbody>
</table>
Associated loci in CGEMS prostate stage 2

Clues to Function?

- **JAZF1**
  - Represses NR2C2 (TR4) which interacts with androgen receptor and is an apoptosis regulator of BCL2.
  - Translocated to SUZ12 in endometrial stromal tumors.

- **CTBP2**
  - Gene poor

- **MSMB**
  - Inhibits PTEN.
  - Activates PI3K pathway.

- **HNF1B**
  - Both loci associated to Diabetes type 2
  - Mutated in renal cancer
  - Silenced in ovarian cancer

- **Myc** ?

- **Diabetes type 2**
  - Both loci associated to
16+ published loci involved in prostate cancer susceptibility

with significance p < 5 x 10^{-7}

KLK3
Kallikrein3 = Prostate Specific Antigen

Eeles et al. 2008
Gudmundsson et al 2008
Haiman et al 2007
### Additional variants – March 2008

<table>
<thead>
<tr>
<th>Location</th>
<th>CGEMS</th>
<th>CRUK</th>
<th>deCODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8q24*</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>HNF1B (17q12)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MSMB (10q11)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>17q24</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>NUDT10/11 (Xp11)</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>JAZF1 (7p15)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTBP2 (10q26)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11q13</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPNE3 (8q21)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL16 (15q25)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDH13 (16q23)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC22A3 (6q25)</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3p12</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LMTK2 (7q21)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLK2,3 (19q13)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2p15</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Prostate Cancer Risk 2008

- Age
- Ethnic Background
- Family History
- Genetic markers
  - 16 Regions of the Genome!!!
Cancer susceptibility loci in the 8q24 region

**Prostate region 2**
- **p = 1 \times 10^{-18}**
  - Haiman et al.
- **p = 1 \times 10^{-8}**
  - Gudmundsson et al.

**Breast region**
- **p = 5 \times 10^{-12}**
  - Easton et al.

**Prostate region 3**
- **p = 7 \times 10^{-11}**
  - Tomlinson et al.

**Colon region**
- **p = 7 \times 10^{-12}**
  - Easton et al.

**Prostate region 1**
- **p = 3 \times 10^{-19}**
  - Thomas et al.

Markers:
- rs13281615
- rs16901979
- rs6983267
- rs1447295
- rs4242382
- rs1264816
- rs2298320
- rs6047775
- rs1051730
- rs1447295
- rs1447295
- rs4242382
- rs1447295

Genetic distances (in Kb):
- 231 Kb
- 58 Kb
- 126 Kb
- 209 Kb
Roche/454 next-gen sequencing analysis

50X coverage, ~140kb

40 prostate cancer cases

40 controls

7 individuals from a CEPH family in which the at-risk haplotype is segregating (ARG)

Yeager et al Nature Genetics 2007

Discovery of ALL Variants
Polymorphism identification in 87 Caucasians (40 cases, 39 controls & 8 CEU)

<table>
<thead>
<tr>
<th></th>
<th>Non-dbSNP</th>
<th>dbSNP</th>
</tr>
</thead>
<tbody>
<tr>
<td># monomorphic</td>
<td>n/a</td>
<td>213</td>
</tr>
<tr>
<td># polymorphic</td>
<td>442</td>
<td>349</td>
</tr>
<tr>
<td>Minimum MAF</td>
<td>0.006</td>
<td>0.000</td>
</tr>
<tr>
<td>Maximum MAF</td>
<td>0.464</td>
<td>0.500</td>
</tr>
<tr>
<td>Mean MAF</td>
<td>0.060</td>
<td>0.142</td>
</tr>
<tr>
<td>Median MAF</td>
<td>0.013</td>
<td>0.101</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Joint PAR</th>
<th>PAR rs1447295</th>
<th>PAR rs6983267</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>0.284</td>
<td>0.085</td>
<td>0.209</td>
</tr>
<tr>
<td>ACS</td>
<td>0.255</td>
<td>0.094</td>
<td>0.192</td>
</tr>
<tr>
<td>ATBC</td>
<td>0.251</td>
<td>0.052</td>
<td>0.157</td>
</tr>
<tr>
<td>FPCC</td>
<td>0.306</td>
<td>0.096</td>
<td>0.091</td>
</tr>
<tr>
<td>HPFS</td>
<td>0.249</td>
<td>0.085</td>
<td>0.180</td>
</tr>
<tr>
<td>PLCO</td>
<td>0.347</td>
<td>0.086</td>
<td>0.276</td>
</tr>
</tbody>
</table>

- Suggests that both SNPs contribute substantially to the population burden of prostate cancer.
What variants to include in risk scores?

• Rapid pace of identification of new variants
• 2-3 years more to “complete” discovery for common alleles in common diseases
• Until then we are operating with a subset of common risk-associated variants

• Under the radar.....copy number variants, “rare” variants i.e <5% allele frequency
Genetic Gold Rush???

Cumulative effect of 5 risk variants (8q, 17q) on prostate cancer risk

- **CAPS**  N= 2,893, P = 6.75E-27
- **CGEMS** N= 1,150, P = 1.06E-10

**OR**

- 0% of controls
- 10% of controls
- 2% of controls

**Number of risk variants**

- 0
- 1
- 2
- 3
- ≥4
7-SNP CGEMS risk score:
“Cumulative” Population Attributable Risk (PAR) = 107% !!!

Thomas et al, 2008
How do we know there are many more variants to find?

• Current variants only account for a small fraction of the effect of family history
  – BCAC Breast Cancer SNPs account for less than 5%
• Current GWAS underpowered for low risk alleles
• Some known alleles have not shown up in GWAS
• Growing experience with pooling across GWAS datasets
  – e.g. Diabetes type II, Crohn’s disease
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 Alamein in North Africa London, 10 November 1942.
Follow-up to GWAS Studies

- Fine Mapping of Notable Regions
  - Genotyping & Sequencing
  - Bio-informatics (exclude common CNV)
- Analysis of Population Genetics
- Functional Determination of Causal Variant(s)
- Exploration of Pathways
  - Etiology
  - Drug Targets
- Design Issue for Analysis in Clinical Evaluation
  - Population-based studies
  - Careful Clinical Studies
Determine Plausibility of Finding
  – *Can we explain the effect?*
  – Molecular Phenotype
    • Correlation of *in vitro* changes with germ-line variant(s)
    • Cell line or tissue work with germ-line analyses

Correlation with *Somatic Alterations*
  – Association of germ-line with somatic observations
  – Driver mutation
What Next?

• More Scans in Each Disease
  – Subtypes
  – Specific Populations: Breast cancer in AA

• In progress GWAS
  – Aggressive adult cancers
    • Pancreas, brain, ovary, esophagus, renal, bladder, melanoma
  – Rare/Pediatric
    • Neuroblastoma, childhood leukemia, osetogenic sarcoma
  – Ample follow-up for mapping/function

• Risk Assessment- Suitable Reporting
  – Public Health and Personal Decisions

• Next-Generation Sequencing
CGEMS: caBIG Posting
Pre-Computed Analysis

Pre-computed Analysis
Post 4 Months Before Publication
No Restrictions

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

Registered Access
SF424
Data Use Certificate

http://cgems.cancer.gov/data

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.

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

http://cgems.cancer.gov/data
Association Results Across 8q24

http://cgems.cancer.gov
Available 10/06
Nature Genetics 2/07

CCEMS SNP Association Finding Report - (19 results)
Acknowledgements

CGEMS & DCEG
Gilles Thomas
Robert Hoover
Joseph Fraumeni
Daniela Gerhard
Richard Hayes
Kevin Jacobs
Meredith Yeager
Sholom Wacholder
Nilanjan Chatterjee
Zhaoming Wang
Xiang Deng
Nick Orr
Robert Welch
Sonja Berndt
Kai Yu
Margaret Tucker
Jesus Gonzalez Bosquet
Jiyoung Ahn

HSPH
David Hunter
Peter Kraft
Fred Schumacher
W Willett
J Ma

ACS
Michael Thun
Heather Feigelson
Eugenia Calle
Carmen Rodriguez

PLCO
Chris Berg
E David Crawford

CeRePP, France
Olivier Cussenot
Geraldine Cancel-Tassin
Antoine Valeri

NPHI, Finland
Jarmo Virtamo

Wash. U., St Louis
Gerald Andriole
Adam Kybell