Genetics and Coronary Artery Disease

CAD As "A Genetic Disease"
Coronary Artery Disease (CAD)

- Affects 35.9% of people in Western Societies
- Has a heritibility of 65%
- Only a fraction of a percent is “Mendelian” (e.g., LDLR deletions)
- Does not have discrete phenotype categories (symptoms range from mild to dropping dead)
- No single factor, either genetic or environmental, is either necessary or sufficient for CAD
Coronary Artery Disease as A Complex Genetic Disease

No Single Locus Genotype (Or Any Other Non-Genetic Factor) Is Either Necessary Or Sufficient for Coronary Artery Disease
The Confoundment of Frequency and Apparent Causation in Systems of Interacting Factors

Factors That Are Rare Are More Strongly Associated With Phenotypic Variation Than Factors That Are Common

Let Frequency of A1 = 0.9, Frequency of A2 = 0.1, Frequency of B1 = 0.1, and Frequency of B2 = 0.9

Frequency in General Population = 0.09.
Frequency of the Disease Given A1 = Freq. (B1) = 0.1
Frequency of the Disease Given B1 = Freq. (A1) = 0.9
Focus on the ApoE Gene That Codes For A Critical Protein In Lipid Metabolism

ApoE Gene

- There are three common alleles coding for different amino acid sequences: ε2, ε3, and ε4
- These three alleles explain more variation in cholesterol levels than any other gene locus
ApoE Gene

Stengård et al. (1996) showed these same alleles at ApoE have a major impact on actual mortality due to CAD in a longitudinal study.

Epistasis Between ApoE and LDLR for LDL Cholesterol
Two Populations

- Frequency ApoE-4 Allele = 0.152
- Frequency ApoE-3 Allele = 0.77
- Frequency LDLR A2 Allele = 0.78
- Frequency ApoE-4 Allele = 0.95
- Frequency ApoE-3 Allele = 0.03
- Frequency LDLR A2 Allele = 0.50

Quantitative Genetic Components As a Function of Allele Frequencies: A. e4 allele at ApoE is Rare, A2 at LDLR Common; B. Reversed
A Retrospective Study of Older Adults from Rochester, MN on Factors Associated with Having vs. not Having CAD

Impact of Cholesterol

A Retrospective Study of Older Adults from Rochester, MN on Factors Associated with Having vs. not Having CAD

Impact of ApoE Genotype
The genotype with the highest incidence of CAD is the “good” genotype 2/3 with no “killer” alleles.

The “worst” genotype if you have high cholesterol is the “best” genotype if you have average cholesterol, and is the “average” genotype if you have low cholesterol.
Interactions Are Difficult to Study Because Large Samples Are Required and Because Even A Single Gene Displays So Much Variation That It Is Computationally Impossible To Examine All Possible Interactions Between The Gene With Other Genes or Environmental Factors.

Genetic Survey of Lipoprotein Lipase

LPL Has 10 Exons Over 30 kb of DNA on Chromosome 8p22

Sequenced 9,734 bp from the 3’ End of Intron 3 to the 5’ End of Intron 9

Sequenced:
24 Individuals from North Karelia, Finland (World’s Highest Frequency of CAD)
23 European-Americans from Rochester, Minnesota
24 African-Americans from Jackson, Mississippi

Found 88 Variable Sites, 87 of which were “SNPs” (Single Nucleotide Polymorphisms)

Ignored Singleton and Doubleton Sites and Variation Due to a Tetranucleotide Repeat, but Phased the Remaining 69 Polymorphic Sites by a Combination of Using Allele Specific Primer Pairs and “Haplotype Substraction”

The Phased Site Data Identified 88 Distinct Haplotypes, which in turn define 3,916 Distinct Genotypes.

NOTE: 3,916 IS MUCH GREATER THAN THE SAMPLE SIZE OF 71 INDIVIDUALS!
HOW DO YOU DO STATISTICS ON UNIQUE EVENTS?

One solution is to reduce the genetic state space in a manner that retains the most important genetic information.

There are many ways of doing this, but only one will be presented here: The HapMap (Haplotype Mapping) Project that is currently the primary focus of the human genome project.
What Is A Haplotype?

From "Sequencing" Gel:
A C/T C A C/G C C T T A/T A T G

The “Genotype” of the Person is:
C/T C/G A/T

The Two Haplotypes (Haploid Types) Are:
1. C G A
2. T C T

The Genotype Data Is Compatible With
Any of the following haplotype pairs:
CCA/TGT or CGA/TCT or
CCT/TGA or CGT/TCA

Haplotype phasing

Experimental:
- Sperm amplification
- Cloning
- Allele-specific amplification
- DNA dilution
- Hybrid cell lines

Statistical inference:
Take advantage of the high structure (linkage disequilibrium) among SNPs
Haplotype inference: usual methodology

Region of interest

Sequencing/typing variant positions

aa gg tt cc

aa gt gc
gt cc
gt cc

Haplotype inference: usual methodology (diagram)

Population data

Haplotype structure

Reconstructed haplotypes

Statistical haplotype inference:

Methods:
- Start with 300 SNPs across 48 kb in KLK region
- Generate unphased data
- Phase it with different algorithms (default parameters):
  - Gerbil (EM on segments + ligation)
  - Arlequin (EM)
  - PHASE (Bayesian/coalescent)
- Determine performance
**Statistical haplotype inference:**

**Methods:** Determine performance

<table>
<thead>
<tr>
<th></th>
<th>errorR</th>
<th>SNPerrorR</th>
<th>switcherrorR</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLKB</td>
<td>0.023</td>
<td>0.049</td>
<td>0.033</td>
</tr>
<tr>
<td>KLKA</td>
<td>0.023</td>
<td>0.066</td>
<td>0.034</td>
</tr>
<tr>
<td>KLKB</td>
<td>0.095</td>
<td>0.094</td>
<td>0.103</td>
</tr>
<tr>
<td>KLKB</td>
<td>0.034</td>
<td>0.056</td>
<td>0.023</td>
</tr>
</tbody>
</table>

1. **TOTAL ERROR RATE**
   Percentage of individuals with incorrectly reconstructed haplotypes

2. **SNP ERROR RATE**
   Percentage of SNPs that are incorrectly assigned to haplotypes

3. **SWITCH ERROR RATE**
   Percentage of inter-SNP intervals that require a switch to maintain true phase

---

*Gerbil*  
*Arlequin*  
*Phase2*
One Solution To The Phase Problem Is To Only Look At Pairwise Phase (2 SNPs at a Time) That Can Be Phased Accurately. This Is The Approach of the HapMap Project.

Identifying Haplotype Blocks: The Four Gamete Test

Start With Genetic Variation At Only One Site: Two Gamete Types

A  G  A  G  C  G  C  G

Mutation At A Second Site Produces Three Gamete Types:

A  G  A  G  C  G  T

Recombination Produces Four Gamete Types

A  G  A  T  C  G  C  T

Haplotype Analysis Implies A Recombination Hotspot In LPL

4 Gamete Test Implies Uniform Recombination In LPL

Region of Overlap of the Inferred Intervals Of All 26 Recombination and Gene Conversion Events Not Likely to Be Artifacts.
Start With Genetic Variation At Only One Site: Two Gamete Types

```
A   G
A   G
C   G
C   G
```

Mutation At A Second Site Produces Three Gamete Types:

```
A   G
A   G
C   G
C   T
```

Second Mutation At Site Produces Four Gamete Types

```
A   T
A   G
C   G
C   T
```

<table>
<thead>
<tr>
<th>TYPE OF SITE</th>
<th>NUMBER OF NUCLEOTIDES</th>
<th>NUMBER POLYMORPHIC</th>
<th>% POLY. PER NUCLEOTIDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPG</td>
<td>198</td>
<td>19</td>
<td>9.6%</td>
</tr>
<tr>
<td>MONONUCLEOTIDE RUNS ≥ 5</td>
<td>456</td>
<td>15</td>
<td>3.3%</td>
</tr>
<tr>
<td>POLYMERASE α ARREST SITE ± 3 NUCLEOTIDES [TG(A/G)(A/G)GA]</td>
<td>264</td>
<td>8</td>
<td>3.0%</td>
</tr>
<tr>
<td>ALL OTHER NUCLEOTIDES</td>
<td>8,866</td>
<td>46</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

In-Likelihood Ratio Test of Homogeneity = 99.8, 3 df, p ≤ 1.75 x 10^-7

In-Likelihood Ratio Test of Homogeneity Within the Three Mutable Classes = 12.3, 2 df, p ≤ 0.002
Given Haplotype Blocks, Can Use Haplotype Trees For Genotype/Phenotype Associations

A Haplotype Tree is an Evolutionary Tree of The Genetic Variation Found In A Sample of Homologous DNA’s
  (Works best when there is little or no recombination within the DNA region)
Gene Tree  
*(all copies of homologous DNA coalesce to a common ancestral molecule)*

Haplotype Tree: Only The Branches Marked By A Mutational Change Are “Visible”

Copies of Homologous DNA
ApoE Haplotypes

Making a simple cladogram or haplotype tree from a few of haplotypes

Sites: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

1 C-A--TG-GCGA----TCCA-GC
2 C-A--TG-GCAA----TCCA-GC
3 C-A--TG-GCAA----TCCA-GT
4 C-A--CGGGCAA----TCCA-GT

1 11 23 6 4
The Apo-protein E Haplotype Tree Estimated With The Program TCS Using Statistical Parsimony

Tree Scanning:
Cut Each Branch In The Tree To Create A Series of One Locus/Two Allele Genotypes That Are Examined For Phenotypic Associations. Additional Rounds of Cuts Can Be Made If Needed.
Tree Scanning:
Cut Each Branch In The Tree To Create A Series of One Locus/Two Allele Genotypes That Are Examined For Phenotypic Associations. Additional Rounds of Cuts Can Be Made If Needed.

Results of Tree Scanning The Apoprotein E Haplotype Tree

Peak Significant TC at the 5% Level With Correction for Multiple Comparisons
Peak Significant HDL at the 5% Level With Correction for Multiple Comparisons
Peak Significant lnTG at the 5% Level With Correction for Multiple Comparisons
Peak Significant lnApoE at the 5% Level With Correction for Multiple Comparisons

* Inference Not Robust To Ambiguity in Tree
Haplotype Trees Solve The Problem of “Too Much Variation”, But A Simpler Solution is to Analyze Each Polymorphic Nucleotide Separately.

So, Is It Worth The Effort Of Doing A Haplotype Analysis?

### SNP Analysis of The Apoprotein E Data

<table>
<thead>
<tr>
<th>Sample</th>
<th>Ln TG</th>
<th>TC</th>
<th>HDL-C</th>
<th>Ln ApoE</th>
</tr>
</thead>
<tbody>
<tr>
<td>J Females</td>
<td>4075</td>
<td></td>
<td>3937, 4036, 4075</td>
<td></td>
</tr>
<tr>
<td>J Males</td>
<td></td>
<td>3937, 4036, 4075</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N Females</td>
<td>832, 3937, 4075</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N Males</td>
<td>624</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Females</td>
<td>4075</td>
<td></td>
<td>624, 1998, 3937, 4075, 4951</td>
<td></td>
</tr>
<tr>
<td>R Males</td>
<td>832, 1998, 3937, 4075, 4951</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fewer Effects Found With Less Statistical Significance, Compared To Tree Scan.
Homoplasy Was Often The Cause of The Low Power Of SNP Analyses

E.g., Site 560 Shows Much Homoplasy:

Branches Defined By SNP 560 Have Significant Phenotypic Associations

SNP 560 Has No Significant Effects

Even When Both Analyses Seem The Same, They Are Not.

E.g., SNP’s 3937, 4036, and 4075 All Are Significantly Associated With lnApoE Levels In Jackson Males and Females.

the tree scan finds the phenotype of lnApoE is significantly associated with branches 7-8 (defined by SNP 3937), 6-15 (defined by SNP 4036), and 7-13 (defined by SNP 4075)
Even When Both Analyses Seem The Same, They Are Not.

Tree Scanning Makes It Clear That These Three SNP’s or Branches Define Two Different Associations!

Tree Scanning Has Already Been Extended To Multivariate Traits and To Simultaneous Scans At Two Loci To Search For Non-Additive Interactions.

It Can Therefore Be A Powerful Tool In Studying Complex Genetic Architectures Involving Multiple Loci With Epistasis
Complexity is both a challenge and an opportunity. About a 5-fold range of incidence variation and about a 100-fold range of incidence variation.
Taking Interactions Into Account Taps Into More Biological Information and Allows More Individualized Risk Assessments for CAD and Other Complex Diseases.
Taking Interactions Into Account Allows More Individualized Treatment and Environmental Intervention for CAD and Other Complex Diseases.
Genomics and CAD


Genomics and CAD

More Individualized Risk Assessment and Environmental Interventions Will Increasingly Shift The Medical Attention on CAD From Therapeutic Medicine to Preventative Medicine — A Shift From Focusing On Disease to Focusing on Health.
Genomic Approaches to Common Chronic Disease

U. of Texas
Houston, TX
James E. Hixson (Component 1)

Eric Boerwinkle (Component 2)
Myriam Fornage
Craig Hanis
Andrei Rodin

U. of Michigan
Ann Arbor, MI
Charles F. Sing (PI, Component 4)
Sharon L. Kardia
Kathy L. Klos
Sara Hamon
Alan R. Templeton

Northwestern U.
Kiang Liu

U. of Alabama
Heather McCreath
O. Dale Williams