Novel Findings in Genomics and Metabolomics in the ARIC study

Eric Boerwinkle
Boston, MA
May, 2017
Goals of Genetic Studies (of the Metabolome)

- Genes being novel predictors of disease
- Predictors vs Biomarkers and the principal of Mendelian Randomization
- Biology of the human metabolome
- Drug Target Discovery
- Gene x Environment Interaction
Maximizing Opportunity for Discovery

\[ \uparrow \text{Power, while controlling costs} \]

- Study Design (sample size, special populations, and families)
- Improved Analysis, e.g. SVs and Annotation
- Phenotype definition, especially endophenotypes
- Read-outs of G x E, e.g. microbiome, metabolome, methylation
Multi-Omics Integration

Achieving this vision, requires delivering large amounts of high quality data to the community in a timely manner.
The Atherosclerosis Risk in Communities (ARIC) Study

- 1,679 African-Americans (AAs) among the ARIC study
- 1,458 ARIC European Americans (EAs)

Having metabolomics data

Prediction of Incident Disease
Metabolomics in the ARIC study

<table>
<thead>
<tr>
<th>COMPOUND TYPE</th>
<th>COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Named Metabolites</td>
<td>361</td>
</tr>
<tr>
<td>Unnamed Metabolites</td>
<td>241</td>
</tr>
<tr>
<td><strong>Total Number of Measured Metabolites</strong></td>
<td><strong>602</strong></td>
</tr>
</tbody>
</table>

- **Detectable**
- **Repeatable**
- **ICC ≥0.6**
- **missingness<80%**

<table>
<thead>
<tr>
<th>COMPOUND TYPE</th>
<th>COUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid</td>
<td></td>
</tr>
<tr>
<td>Carbohydrate</td>
<td></td>
</tr>
<tr>
<td>Cofactors and vitamins</td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td></td>
</tr>
<tr>
<td>Lipid</td>
<td></td>
</tr>
<tr>
<td>No Super Pathway</td>
<td></td>
</tr>
<tr>
<td>Nucleotide</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td></td>
</tr>
<tr>
<td>Peptide</td>
<td></td>
</tr>
<tr>
<td>Xenobiotics</td>
<td></td>
</tr>
</tbody>
</table>

- **Total Number**
  - Named Metabolites: 118
  - Unnamed Metabolites: 86
  - **Total Number of Measured Metabolites**: 204
Why Multi-Ethnic Studies?

- Differences in Environment
- Differences in site frequency spectrum
- G x E
- Epidemiology of Disease
Advances in Genomics & Metabolome

**Genomics**
- Candidate Gene
- GWAS
- Exome Chip
- Whole Exome Seq
- Whole Genome Seq

**Metabolomics**
- Targeted Metabolomics
- Untargeted Metabolomics
- NMR
- GC-MS/LC-MS
- HILIC-MS
- CE-MS
Advances in Genomics & Metabolome

Genomics

Candidate Gene
**HAL, Histidine and Coronary Heart Disease**

**A**

- **Splice-5 (n = 1)**
- **R322X (n = 22)**
- **Splice-5 (n = 1)**

1 CDS End | 5 | 10 | 15 | 20 CDS Start

**B**

- **HAL**
- Histidine
- Histidine Ammonia-Lyase
- Trans-Urocanic Acid

**C**

- Mean of histidine levels

**African Americans in ARIC**

- LoF mutations in HAL
  - MAF = 0.01
  - \( P = 1.2 \times 10^{-13} \)

- High histidine levels
  - \( P = 1.9 \times 10^{-4} \)

- Low incident CHD risk
  - \( P = 0.05 \)

**European Americans in FHS**

- MAF = 0.0009
- \( P = 0.05 \)

Yu, *Circ Cardiovasc Genet*, 2015
Advances in Genomics & Metabolome
Common variants with p-value < $1.6 \times 10^{-10}$

Yu, PLoS Genet, 2014

**Genome-wide Significant Gene-Metabolite Pairs in 1,679 ARIC African Americans**
Genome-wide Significant Gene-Metabolite Pairs in 1,679 ARIC African Americans

Common variants with p-value < $1.6 \times 10^{-10}$

Yu, PLoS Genet, 2014
Hypothesis 1: NAT8 – N-acetylyornithine – chronic kidney disease?

Caucasians

- NAT8


risk allele: A
p = 4.0 × 10^{-66}

African Americans

- NAT8 (rs13538, missense)


N-acetylyornithine

Chronic kidney disease

N-acetylyornithine

Chronic kidney disease

cross-sectional? longitudinal?

Baseline eGFR (p = 2.7 × 10^{-14})
Incident CKD (p = 0.004)
Advances in Genomics & Metabolome

Genomics

- Candidate Gene
- GWAS
- Exome Chip
- Whole Exome Seq
Why Sequencing?

McCarthy, Nat Reviews Genet, 2009
Defining LOF

• Variants predicted to trigger nonsense-mediated decay (NMD)

• Categories:
  1) Premature stop codon-introducing
  2) Disrupt essential splice site
  3) Insertion/deletion frameshifts (indel)

• Additional Subdivision:
  • Full: all known protein coding transcripts
  • Partial: affecting only a fraction of known coding transcripts

Image via: http://combio.berkeley.edu/people/ed/rust/
Annotating LOF

- 8,554 ARIC Study participants
  - 5,718 EA and 2,836 AA
  - (4,277 disc and 4,277 repl)
- Variant filtering:
  - Single-exon genes
  - Non protein-coding genes
  - Affect all gene isoforms
  - Terminal gene exon

- 36,787 LOF sites in 11,922 genes

- Average per individual:
  - Heterozygous (homozygous)

<table>
<thead>
<tr>
<th>LOF type</th>
<th>Initial</th>
<th>After filtering</th>
<th>% Filtered out</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop gain</td>
<td>19,759</td>
<td>14,076</td>
<td>28.7%</td>
</tr>
<tr>
<td>Splice</td>
<td>10,634</td>
<td>8,843</td>
<td>16.8%</td>
</tr>
<tr>
<td>Frame Shift</td>
<td>33,703</td>
<td>13,868</td>
<td>58.8%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>64,096</td>
<td>36,787</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOF type</th>
<th>AA</th>
<th>EA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stop gain</td>
<td>27.3 (2.1)</td>
<td>21.1 (2.2)</td>
</tr>
<tr>
<td>Splice</td>
<td>16.7 (1.9)</td>
<td>9.6 (1.8)</td>
</tr>
<tr>
<td>Frame Shift</td>
<td>36.1 (4.4)</td>
<td>22.6 (3.1)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>80.1 (8.4)</td>
<td>53.3 (7.1)</td>
</tr>
</tbody>
</table>

FHS: phs000651.v4.p9; CHS: phs000667.v1.p1; ARIC phs000668.v1.p1
324 single LoF variants (MAF ≥ 5%), 1285 genes with cMAC ≥ 7 included, p-value < 1.3 × 10^{-7}

SLCO1B1, Hexadecanedioate & Heart Failure

Discovery stage (ARIC AAs)

LoF mutations in SLCO1B1

MAF = 0.025
P = 2.2 × 10^{-9}

MAF = 0.023
P = 1.0 × 10^{-4}

High hexadecanedioate levels

High risk of incident HF

HR = 1.22
P = 3.0 × 10^{-7}

HR = 1.09
P = 0.005

Replication stage (ARIC AAs and EAs)

HR = 1.29, P = 0.05
(ARIC AAs and EAs)

**Possible Mechanism of the Association**

250 mg/kg/day hexadecanedioate feeding

<table>
<thead>
<tr>
<th></th>
<th>European-Americans (n = 1,551)</th>
<th>African-Americans (n = 2,448)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Systolic BP</strong></td>
<td>1.30 (0.43) Beta (SE) 0.002</td>
<td>3.38 (0.61) Beta (SE) 4E-8</td>
</tr>
<tr>
<td><strong>Diastolic BP</strong></td>
<td>0.74 (0.26) Beta (SE) 0.004</td>
<td>1.33 (0.360 Beta (SE) 2E-4</td>
</tr>
</tbody>
</table>

Advances in Genomics & Metabolome

Genomics

Candidate Gene

GWAS

Exome Chip

Whole Exome Seq

Whole Genome Seq
The genomes were annotated by ANNOVAR based on the RefSeq database.

- **Intronic**: 58%
- **Intergenic**: 35%
- **Other**: 7%

- **ncRNA**: 45%
- **5'UTR or 3'UTR**: 21%
- **Splicing**: 17%
- **Upstream or downstream**: 16%
- **Exonic**: 7%
$2.0 \times 10^{15}$ sequenced bases

US corn production in 2014: $1.3 \times 10^{15}$ kernels

From G. Abecasis
Aggregate by annotated functional motif for low frequency and rare SNVs (MAF ≤ 5%):
- Sliding window
- Regulatory domain
- First intron

Asparagine:
- A non-essential amino acid;
- Biosynthesis/diet intake;
- Required for development and function of the brain.

2 significant sliding windows (p < 4e-8), 6kb downstream of AGA

rs11131799, the most significant common variant

Functional coding variants (MAF < 5%) in AGA aggregately affect asparagine levels
**AGA gene:**
- Aspartylglucosaminidase
- Cleaves *asparagine* from N-acetylglucosamine.

**rs11131799 (1st intron of AGA):**
- Predicted promoter;
- Influences the expression levels of AGA (p = 0.01).
Multi-Omics Integration

Achieving this vision, requires delivering large amounts of high quality data to the community in a timely manner.
Identification of “Causal” Pathways among the Serum Metabolome

• As shown above, the principal of Mendelian randomization can lend credence to claims of causal inference.
• This principal of Mendelian randomization can extend to information across the genome.
• We (Yazdani, 2016) have combined the principal of genome-wide Mendelian randomization with Directed Acyclic Graph algorithms (GDAG).
In total, 9 metabolites have direct effects on triglyceride levels.

Yazdani et. al, (2016). Metabolomics
Acknowledgments

University of Texas Health Science Center at Houston
Bing Yu
Mandana Yazdani
Elena Feofanova
Paul de Vries
Xiaoming Liu
Alanna Morrison

Baylor College of Medicine
Ginger Metcalf
Fuli Yu
Donna Muzny
Richard Gibbs

Framingham Heart Study
Robert Gerszten
Eugene Rhee
Dan Levy
Vasan Ramachandran

Atherosclerosis Risk in Communities Study
Christie Ballantyne
Josef Coresh