Challenges and Opportunities in Translational Genomics

Leslie Biesecker, M.D.
National Human Genome Research Institute, NIH
Genetic architecture of disease

Rare variants

Common variants
Genetic architecture of disease

We hypothesize that large-scale sequencing will allow a comprehensive assessment of the architecture of the genetic component of disease.
Kinds of mutations

Penetrance vs. Frequency
Kinds of mutations

Frequency

Penetrance

[Graph showing the relationship between frequency and penetrance, with a downward curve.]
Kinds of mutations

Penetrance

Frequency

?
Translational genomics

- Genetic architecture of human disease
- Technologic advances
- New ways to ask and answer biomedical research questions
Translational genomics

- Genetic architecture of human disease
- Technologic advances
- New ways to ask and answer biomedical research questions
  - Pilot whole genome sequence acquisition of individual subjects as a clinical research tool from consent to the return of results
Sequencing technologies

NHGRI supported sequencing centers $\sim 2 \times 10^{11} \text{ bp/yr}$
Equivalent to 60 - 70 mammalian genomes/year
Solid phase sequencing

One million Megaclone™ beads arranged in a monolayer

Up to $10^9$ bp per run
$100,000$ genome
ClinSeq: A translational research project in clinical genomics

- Medical & Statistical Genetics
- NIH Intramural Sequencing Center
- NHLBI
- NIH Clinical Center
- ClinSeq

NIH Clinical Center

NIH Intramural Sequencing Center
Specific aims

1. Develop a robust infrastructure for the generation and use of LSMS in a clinical research setting

2. Use LSMS data to develop novel approaches to clinical biomedical research

3. To understand how to interact with subjects re LSMS

4. Genetically dissect a phenotype with complex genetic architecture
Initial approach

• Phenotype 1,000 subjects
  - Framingham risk score
  - 250 patients into each of four bins
    • Low, medium, high risk, known disease
• Initially sequence 400 candidate genes
  - 8,000,000 capillary reads
• Follow-up studies
• Interpret variants and validate some
• Return results
<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA1</td>
<td>CD40</td>
<td>HDAC2</td>
<td>MTP</td>
<td>ROS1</td>
</tr>
<tr>
<td>ABCG1</td>
<td>CD40LG</td>
<td>HMGR</td>
<td>MVK</td>
<td>RXRA</td>
</tr>
<tr>
<td>ABCG5</td>
<td>CDKN1A</td>
<td>HMOX1</td>
<td>MYLK</td>
<td>SAR1B</td>
</tr>
<tr>
<td>ABCG8</td>
<td>CDKN2A</td>
<td>HMOX2</td>
<td>NCCT</td>
<td>SCARB1</td>
</tr>
<tr>
<td>ACE</td>
<td>CDKN2B</td>
<td>ICAM1</td>
<td>NFKB1</td>
<td>SCD</td>
</tr>
<tr>
<td>ACTA2</td>
<td>CELSR2</td>
<td>IL18</td>
<td>NKCC2</td>
<td>SELL</td>
</tr>
<tr>
<td>ADH1C</td>
<td>CETP</td>
<td>IL8</td>
<td>NOS1</td>
<td>SELP</td>
</tr>
<tr>
<td>AGT</td>
<td>CFH</td>
<td>INSIG2</td>
<td>NOS3</td>
<td>SELS</td>
</tr>
<tr>
<td>AGTR1</td>
<td>CIITA</td>
<td>ITGA2</td>
<td>NOX1</td>
<td>SERPINE1</td>
</tr>
<tr>
<td>ADIPOR1</td>
<td>CRP</td>
<td>ITGB1</td>
<td>NPC1</td>
<td>SIRT1</td>
</tr>
<tr>
<td>ADIPOR2</td>
<td>CXCRC4</td>
<td>ITGB2</td>
<td>NPC1L1</td>
<td>SIRT3</td>
</tr>
<tr>
<td>ADIPOQ</td>
<td>CYBA</td>
<td>ITGB3</td>
<td>NPR1</td>
<td>SOAT1</td>
</tr>
<tr>
<td>ALOX</td>
<td>DGAT1</td>
<td>ITGB5</td>
<td>NRIH2</td>
<td>SOAT2</td>
</tr>
<tr>
<td>ALOX5AP</td>
<td>DGAT2</td>
<td>ITGB7</td>
<td>NRIH3</td>
<td>SOCS3</td>
</tr>
<tr>
<td>ANGPTL3</td>
<td>DUSP1</td>
<td>JAK3</td>
<td>NRF1</td>
<td>SOD2</td>
</tr>
<tr>
<td>ANGPTL4</td>
<td>ENPEP</td>
<td>KALRN</td>
<td>OLR1</td>
<td>SOD3</td>
</tr>
<tr>
<td>ANRIL</td>
<td>ESR1</td>
<td>KCNJ8</td>
<td>OR13G1</td>
<td>SORT1</td>
</tr>
<tr>
<td>APOA1</td>
<td>F13B</td>
<td>KCNMB1</td>
<td>P2RY12</td>
<td>SREBF1</td>
</tr>
<tr>
<td>APOA2</td>
<td>F2</td>
<td>KIF6</td>
<td>PAPPA</td>
<td>SREBF2</td>
</tr>
<tr>
<td>APOA5</td>
<td>F5</td>
<td>KL</td>
<td>PCSK9</td>
<td>STAT1</td>
</tr>
<tr>
<td>APOB</td>
<td>F7</td>
<td>LCAT</td>
<td>PDGFB</td>
<td>TAS2R50</td>
</tr>
<tr>
<td>APOBEC1</td>
<td>FABP2</td>
<td>LDLR</td>
<td>PDGFRB</td>
<td>TBX2A2R</td>
</tr>
<tr>
<td>APOC1</td>
<td>FABP4</td>
<td>LDLRAP1</td>
<td>PER1</td>
<td>TCF1</td>
</tr>
<tr>
<td>APOC2</td>
<td>FAM5C</td>
<td>LGALS2</td>
<td>PITX2</td>
<td>TCF7L2</td>
</tr>
<tr>
<td>APOC3</td>
<td>FAS</td>
<td>LPC</td>
<td>PLA2G4A</td>
<td>THBD</td>
</tr>
<tr>
<td>APOC4</td>
<td>FGB</td>
<td>LIPE</td>
<td>PLA2G7</td>
<td>THBS4</td>
</tr>
<tr>
<td>APOE</td>
<td>FGG</td>
<td>LIPG</td>
<td>PLAT</td>
<td>TIMP1</td>
</tr>
<tr>
<td>APM</td>
<td>FLI1</td>
<td>LPA</td>
<td>PLTP</td>
<td>TLR4</td>
</tr>
<tr>
<td>AR</td>
<td>FOS</td>
<td>LPL</td>
<td>PON1</td>
<td>TLR8</td>
</tr>
<tr>
<td>ARF</td>
<td>FTO</td>
<td>LRP6</td>
<td>PPARA</td>
<td>TNFRSF1A</td>
</tr>
<tr>
<td>ARG1</td>
<td>GAL</td>
<td>LRP8</td>
<td>PPARD</td>
<td>TNFSF4</td>
</tr>
<tr>
<td>ATF4</td>
<td>GALNT2</td>
<td>LTA</td>
<td>PPARG</td>
<td>TRIB1</td>
</tr>
<tr>
<td>BDKRB2</td>
<td>GAS6</td>
<td>LTA4H</td>
<td>PRDX2</td>
<td>TRIB3</td>
</tr>
<tr>
<td>BMI1</td>
<td>GATA2</td>
<td>LTC4S</td>
<td>PRDX3</td>
<td>UCP2</td>
</tr>
<tr>
<td>BSDL</td>
<td>GCKR</td>
<td>MB1</td>
<td>PRKG1</td>
<td>UCP3</td>
</tr>
<tr>
<td>C11orf2</td>
<td>GCLC</td>
<td>MEF2A</td>
<td>PSMA6</td>
<td>USF1</td>
</tr>
<tr>
<td>CALM1</td>
<td>GCLM</td>
<td>MLXIPL</td>
<td>PSRC1</td>
<td>VAMP8</td>
</tr>
<tr>
<td>CAPG</td>
<td>GDF5</td>
<td>MMAB</td>
<td>PTGIS</td>
<td>VEGF</td>
</tr>
<tr>
<td>CAV1</td>
<td>GJA4</td>
<td>MMP3</td>
<td>PTGS1</td>
<td>VLDLR</td>
</tr>
<tr>
<td>CCL2</td>
<td>GP1BA</td>
<td>MMP9</td>
<td>PTGS2</td>
<td>VNN1</td>
</tr>
<tr>
<td>CCR2</td>
<td>GP1BB</td>
<td>GP5</td>
<td>MOGAT1</td>
<td>RAP1B</td>
</tr>
<tr>
<td>CCR5</td>
<td>GP6</td>
<td>MOGAT2</td>
<td>RBKS</td>
<td></td>
</tr>
<tr>
<td>CD36</td>
<td>GPX1</td>
<td>MPO</td>
<td>ROMK</td>
<td></td>
</tr>
</tbody>
</table>
Progress

• Enrollment began January, 2007
• Nearly 350 patients enrolled mid May
• >900,000 sequence reads to date
• Results already returned
• N=75 ClinSeq
  - A50S
  - A391T
  - T726I
  - A391T
  - A391T
  - A391T
  - P685L
  - Y188X
  - A391T
  - V827I
  - A50S
  - T726I
  - R744Q:A391T
  - A391T

• N=5 HapMap
  - A391T

Initial ClinSeq
LDLR variants
Medical history

- 65 yo female
- High cholesterol diagnosed at 25 years
- RX: atorvastatin, ezetimibe, hctz, lisinopril, niacin
- Coro Ca\(^{++}\) 1,726
- Chol 172, Trig 50, HDL 75, LDL 88
- Family members diagnosed & treatment started
What we will accomplish

• Develop molecular, bioinformatic & medical approaches for clinical LSMS
• Comprehensively dissect the genetic architecture of a phenotype
• Pilot approaches to personalized health care
• Ascertain the subject’s views of LSMS
What we will accomplish

• Develop molecular, bioinformatic & medical approaches for clinical LSMS
• Comprehensively dissect the genetic architecture of a phenotype
• Pilot approaches to personalized health care
• Ascertain the subject’s views of LSMS
• Proceed to exome or genome sequencing - fully consented!
Public resource

• Clinical data deposited in dbGAP
• Sequences in trace archive
• Lymphoblast cell lines available
• Plasma/serum available (limited)
Hypothesis generating clinical research

- Collect patients with broadly defined clinical phenotype
- Sort on genomic attributes
- Refine phenotype of selected patients
Collaborators

• **NISC**
  - Jim Mullikin, Bob Blakesly, Gerry Bouffard, Pedro Cruz, Nancy Hanson, Morgan Park, Alice Young

• **NHGRI**
  - Eric Green, Flavia Facio, Paul Goboume, Jennifer Johnston, Teri Manolio, Jamie Teer, Clesson Turner, Alec Wilson

• **NHLBI**
  - Richard Cannon, Andrew Arai, Paul Hwang, Toren Finkel, Vandana Sachdev, Bob Shamburek

• **NIHCC**
  - Alan Remaley
Results flow to subjects: Variants of clear clinical significance
Results flow to subjects: Variants of uncertain clinical significance