Opportunities and Challenges for Health Disparities Research in the Personal Genome Era

Carlos D. Bustamante
Department Biomedical Data Sciences and Genetics
Stanford University School of Medicine

cdbustam@stanford.edu
twitter: @cdbustamante
Key Driving Questions

* Why study diverse populations in clinical genomics research? What is the potential scientific gain?

* What have we learned from broadening representation thus far in GWAS/Mendelian?

* How do we properly design multi-ethnic studies so we maximize the power of discovery and interpretation?

* How do we modify protocols for recruitment, consent/enrollment process, and RoR in a multi- /trans- ethnic clinical genomic research setting?
COMPARING THE UNCOMPARABLE

The rarer a genetic variant is within a population, the less likely it is to be found in all ethnic groups. One hundred people were sampled from each population.

- Europeans*
- Europeans & Chinese†
- European & African‡

*Comparison of individuals of European descent in Utah and in Tuscany, Italy. † Han Chinese individuals from Beijing compared with Utah sample. ‡ Yoruba individuals from Ibadan, Nigeria, compared with Utah sample.
Broadening representation in genomics...
HLA B*5701 causes Abacavir hypersensitivity in HIV anti-retroviral therapy

**Incidence:** 3-5%

**Onset:** 4-6 wks after initiation of abacavir therapy

**Symptoms:**
Fever, skin rash, fatigue, GI symptoms (nausea, vomiting, diarrhea, abdominal pain), and respiratory tract symptoms (pharyngitis, dyspnea, or cough)

**Management:**
Discontinue abacavir
Do not re-start abacavir; severe symptoms will recur within hours, including life-threatening hypotension and death

<table>
<thead>
<tr>
<th>Population</th>
<th>Incidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati (India)</td>
<td>0.2</td>
</tr>
<tr>
<td>Massai (Kenya)</td>
<td>0.15</td>
</tr>
<tr>
<td>Utah (U.S. whites)</td>
<td>0.1</td>
</tr>
<tr>
<td>Tuscans (Italy)</td>
<td>0.05</td>
</tr>
<tr>
<td>Luhya (Kenya)</td>
<td>0.05</td>
</tr>
<tr>
<td>Mexican Americans</td>
<td>0.05</td>
</tr>
<tr>
<td>Chinese in Denver</td>
<td>0.05</td>
</tr>
<tr>
<td>Han Chinese</td>
<td>0.05</td>
</tr>
<tr>
<td>African Americans</td>
<td>0.05</td>
</tr>
<tr>
<td>Japanese from Tokyo</td>
<td>0.05</td>
</tr>
<tr>
<td>Yorubans</td>
<td>0.05</td>
</tr>
</tbody>
</table>
Early results suggest broadening ethnicity for GWAS works

Medical genetics continues to suffer from a European bias (Bustamante et al. Nature 2011), although that is slowly changing:

Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico

The genetics of Mexico recapitulates Native American substructure and affects biomedical traits

Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25

Melanesian Blond Hair Is Caused by an Amino Acid Change in TYRP1
**Multiethnic Genomewide Association (MEGA) Array**

- PAGE-II Custom Content, 42K
- Exonic Variants, 200K
- Functional Variants, 40K
- GWAS Scaffold, 360K
- African Power Diaspora Scaffold, 700K
- Human Core Scaffold, 300K
- Human Exome Array, 250K

(1.7M SNPs)
Lessons Learned thus far...

* 1000 Genomes and other population scale studies have demonstrated “Common variants are rare and shared; Rare variants are common and largely population private”

* Properly powered GWAS studies in understudied populations yield novel variants at previously associated genes (e.g., LDLR/PCSK9) and new genes underlying previously studied phenotypes (e.g., SLC16A11 in T2D for H/L).

* NHGRI efforts in diversifying medical and population genomics have led to important reagents for multi-ethnic GWAS (e.g., MEGA) and now sequencing (e.g., Centers for Common and Mendelian Diseases)

* Sequencing based population screening (e.g., CFTR sequencing in CA) yields a broad spectrum of alleles that are rare, many VUSs, and require additional clinical data for interpretation.

* Building diversity into CSER will likely yield new opportunities for biology and improve patient care/health outcomes for minority populations.
Key challenges ahead:

* Should representation in sequencing studies be proportional (65%, 15%, 10%, 5%, etc.) or stratified (25%, 25%, 25%, etc.)

* VUS rates are higher in non-white populations. Does this pose a challenge to genomic medicine and how do we address it?

* Inclusion doesn’t mean just ethnic diversity. How does SES, education, etc. impact enrollment, genome interpretation, and ROR?

* New technologies pose a risk to broadening health disparities. While overall improvement in health outcomes across populations is expected, rates of improvement could vary by race, ethnicity, SES, etc. Can we study this directly and develop countermeasures?