Immunogenomic Data Management Methods

Jill A. Hollenbach & Steven J. Mack
Immunogenomics Data-Analysis Working Group

16th International HLA and Immunogenetics Workshop
Immunogenomic Data Management Methods

Goal: overview of community data-management practices

First Phase of Project:
- Surveys of current HLA & KIR data-management practices
  - Data-capture and Transmission Practices
  - Ambiguity Resolution Practices
  - Data Analyses

Second Phase of Project:
- Determine the effects (if any) of current practices on common data applications
  - Registry Searches
  - Disease Association
  - Population Studies
Immunogenomic Data Management Methods

1° Survey: Data-capture and Transmission Practices

For each participating group:

Which data generation methods are employed?
How are genotype data stored and viewed?
How are stored genotype & allele data formatted?
How are ambiguities managed?
What non-genetic data are stored?
How are data transmitted?
How does your group manage public data?
Do you carry out analyses of this data?
What tools would simplify your data-management?
Immunogenomic Data Management Methods

2° survey: HLA Ambiguity Resolution Practices

Distribute synthetic ambiguous datasets for:
- Sub-saharan African, Amerindian, & “admixed” populations
- Include “easy” and “difficult” to resolve data

Participants resolve the ambiguity and document the process(es) used

This survey will also collect data on the influence of:
- Geographic origin of the population
- Typing resolution of the distributed data
- Format used to record ambiguity
- Organization of alleles and genotypes

Survey attitudes about local methods and potential automated and/or standard methods
## Most Ambiguity Resolution Practices are Undocumented

### Sample X

<table>
<thead>
<tr>
<th>Genotype</th>
<th>HLA-B allele 1</th>
<th>HLA-B allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotype 3</td>
<td>27:09</td>
<td>44:02:01:01, 44:02:01:02S, 44:02:03, 44:19N, 44:23N, 44:24, 44:27, 44:33</td>
</tr>
<tr>
<td>Genotype 4</td>
<td>27:09</td>
<td>44:02:02, 44:11</td>
</tr>
</tbody>
</table>

**Ambiguity Resolution**

- 27:03 44:02
- 27:05 44:02
Example Documented Ambiguity Resolution

HLA-B allele 1                      HLA-B allele 2

Genotype 1


Genotype 2


Genotype 3

27:09

Genotype 4

27:09

4-digit filtering, Elimination of rare alleles, Binning alleles identical over exons 2&3  (B*4402 & 4427)

B*27:05

Regional population-level frequency data

Unambiguous data
3º survey: Primary Analytical Practices

What basic analyses are routinely performed?
   Haplotype Estimation/Linkage Disequilibrium
   Hardy-Weinberg

What programs/tools are used?

How are these data stored?

How are these data used?
Participation

Project will start in January 2011

Surveys and datasets will be distributed online

Details and downloads will be available from:
www.igdawg.org/workshop.html

For more information or to participate, contact:
sjmack@chori.org, jhollenbach@chori.org