Functional variation in the 1000 Genomes Project

Stephen Montgomery
smontgom@stanford.edu
montgomerylab.stanford.edu
Stanford University School of Medicine
Individual genomes contain 
~4-5 million variant sites
How do we interpret variation on a genome-scale?
Identifying that all individuals have dysfunctional genomes

All sequenced genomes contain rare protein-coding variants predicted to severely disrupt gene function.

LoF/PTV alleles in the 1000 Genomes

“We estimate that human genomes typically contain ~100 genuine LoF variants with ~20 genes completely inactivated. We identify rare and likely deleterious LoF alleles, including 26 known and 21 predicted severe disease–causing variants, as well as common LoF variants in nonessential genes.”

The distribution of variant sites per genome in 1000 Genomes Phase 3
How do we interpret variation on a genome-scale?
Integration of 1000 Genomes variants with ENCODE data to identify impactful non-coding variants
Differential selective constraints among noncoding categories

Khurana et al., Science, 2013
Interpreting the non-coding genome

Genetic studies of gene expression

SNP

Expression of nearby genes

Gene splicing/isoforms

Expression of distant genes

Cellular processes

Disease risk
eQTLs in the 1000 Genomes

Montgomery et al, PLoS Genetics, 2011
Montgomery et al, Genome Research, 2013
### eQTL genes discovered in six global populations in 1000 Genomes Phase 3

<table>
<thead>
<tr>
<th>Population</th>
<th>Sites Tested</th>
<th>FDR &lt; 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEU</td>
<td>7,651,459</td>
<td>744</td>
</tr>
<tr>
<td>CHB</td>
<td>7,094,961</td>
<td>1,182</td>
</tr>
<tr>
<td>GIH</td>
<td>7,913,992</td>
<td>1,186</td>
</tr>
<tr>
<td>JPT</td>
<td>7,073,650</td>
<td>1,223</td>
</tr>
<tr>
<td>LWK</td>
<td>11,413,786</td>
<td>1,021</td>
</tr>
<tr>
<td>YRI</td>
<td>11,236,023</td>
<td>1,064</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th>2,821 genes discovered in any population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>211 genes discovered in all populations</td>
</tr>
</tbody>
</table>

For each population, \( n = 69 \) and \( 16,122 \) genes tested.
Can we use whole genomes from global populations to identify causal variants within eQTLs?
Can we use whole genomes from global populations to identify causal variants within eQTLs?
Can we use whole genomes from global populations to identify causal variants within eQTLs?
Can we use whole genomes from global populations to identify causal variants within eQTLs?
Power of multi-ethnic panels to resolve causal variation!

African populations best at resolving functional variants
Future directions

INTERPRET

ALL THE VARIANTS
Acknowledgements

The 1000 Genomes Consortium

Adam Auton
Marianne DeGorter
Ekta Khurana
Tuuli Lappalainen
Daniel MacArthur

More info about my lab at montgomerylab.stanford.edu