**Genotype Concordance Between Low-Coverage Whole Genome and High-Coverage Exome Sequencing: Results from the UK10K Study**

Hou-Feng Zheng, on behalf of the UK10K Consortium Cohorts Group  
Contacts: hou.zheng@mail.mcgill.ca  
Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, Jewish General Hospital, McGill University, Montreal, Quebec, Canada.  
The data is publicly available ([www.uk10k.org](http://www.uk10k.org))

**Background:** Given the high cost of large, next-generation sequencing studies, there is considerable debate as to the optimal study design to accurately capture rare variants. While low-coverage whole genome sequence provides an accessible survey of genotypes across the entire genome, including regulatory regions, it is unclear if low-coverage studies yield sufficient genotypic accuracy to justify their utility.

**Aim:** To compare the genotype concordance rates of low coverage sequencing and high depth sequencing in the same individuals.

**Methods:**
- **UK10K Project** ([http://www.uk10k.org/](http://www.uk10k.org/))  
  2432 whole-genome sequencing samples  
  Median depth per variant = 6x  
- **TwinsUK Project** ([http://www.twinsuk.ac.uk/](http://www.twinsuk.ac.uk/))  
  494 whole-exome sequencing samples  
  Median Depth per variant = 65.6x  

Variant calls from 22 autosomes were compared using VCFtools

157 samples overlapped  
178,112 sites in common  
177,922 sites reported  
190 sites had missing genotypes (0.1%)

**Results:**
The overall genotype concordance rate was 99.41%. Table 1 showed the concordance rate for reference and non-reference allele; Table 2 showed the concordance rate for homozygotes and heterozygotes. Concordance rates increased with the allele depth and allele frequency, yet remained high even for alleles of low frequency (99.1% for variants with a non-reference allele frequency between singletons and 1%) (Figure 1). Discordance rates were also observed to be higher at specific genomic locations.

**Figure 1**

Matches VS Allele Frequency  
RR: homozygosity of reference allele  
AA: homozygosity of non-reference (alternative) allele

**Table 1**

<table>
<thead>
<tr>
<th>variant type</th>
<th>number of mismatches</th>
<th>number of matches</th>
<th>concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>287</td>
<td>177635</td>
<td>175718</td>
</tr>
<tr>
<td>RA</td>
<td>1469</td>
<td>126768</td>
<td>1963379</td>
</tr>
<tr>
<td>AA</td>
<td>58706</td>
<td>14595478</td>
<td>99.60%</td>
</tr>
</tbody>
</table>

Concordance for reference and non-reference allele

**CONCLUSION**
The concordance rate between low coverage sequencing and high depth sequencing is high. These findings have important implications for sequencing study design and suggest that low coverage sequencing is able to accurately call genotypes.

**Acknowledgements and Financial support**