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)

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/) (2432 whole-genome sequencing samples) Median depth per variant = 6x

TwinsUK Project (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%)

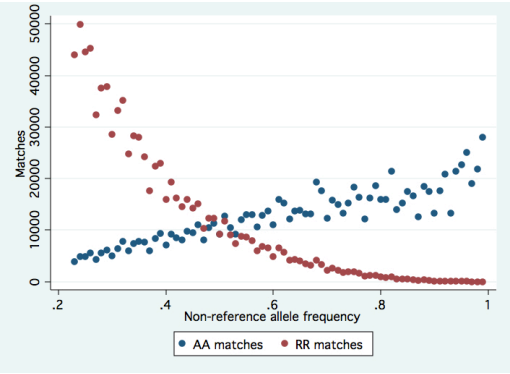
Table 1

	reference allele	non-reference allele
matches	177635	175718
mismatches	287	1917
total	177922	177635
concordance	99.84%	98.92%

Concordance for reference and non-reference allele

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 Alelle Frequency
RR: homozygosity of reference allele
AA: homozygosity of non-reference (alternative) allele

Table 2

variant type	number of mismatches	number of matches	concordance
RR	58706	14595478	99.60%
RA	33303	1963379	98.33%
AA	14659	1267678	98.86%

Concordance for homozygote and heterozygote
RR: homozygosity of reference allele
RA: heterozygosity
AA: homozygosity of non-reference (alternative) 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