## Imputation of genotypes from low- to high-density genotyping platforms

D.P. Berry[1] J.F. Kearney[2]
[1]*Animal & Grassland Research and Innovation Center, Teagasc, Moorepark, Fermoy, Co. Cork.* [2]*Irish Cattle Breeding Federation. Bandon, Co. Cork.*

### Introduction

Genomic selection is increasing in popularity as a method of evaluating the genetic merit of animals. However, the acquisition of high density genotypes on individual females, necessary to obtain direct genomic values (DGVs), is currently prohibitively expensive for most individual farmers. The objective therefore, of this study was to quantify the accuracy of predicting or imputing genotypes from lower density, lower cost marker panels to higher density marker panels.

### Materials and Methods

Genotypes on 5,489 Holstein-Friesian AI bulls (n=4,318) and cows (n=1,171) using the Bovine50K marker panel (Illumina, San Diego, CA) were available. Following the removal of single nucleotide polymorphisms (SNPs) on the sex chromosome, of unknown position, and SNPs with >0.5% mendelian inconsistencies between parents-offspring, 51,602 SNPs remained. The commercially available lower density Bovine3K marker panel (Illumina, San Diego, CA) contains 2,900 SNPs. Only SNPs on the autosomes were retained and one SNP, of unknown position was also discarded leaving 2,730 SNPs.

Animals were separated into two groups: 1) reference group of animals born prior to 2006 (n=4,725), and 2) a test group of animals, >50% Holstein, born from 2006 onwards (n=764). Only the 2,730 SNPs from the Bovine3K (after editing) were retained in the test group of animals with the remaining SNPs on the Bovine50K in these animals to be imputed. Imputation was undertaken for each chromosome separately using the freely available software Beagle Version 3.1.0 (Browning and Browning, 2007). Genotype and allele concordance rate, defined as the average proportion of correctly imputed genotypes and alleles, respectively was used as a measure of accuracy of imputation.

To quantify the impact of imputation on the estimation of DGVs in the imputed animals, genomic prediction was undertaken for all traits in the Economic Breeding Index using the procedures outlined by Berry et al. (2009); the posterior probabilities of each imputed allele was used for the imputed animals. Animals included in the estimation of SNP effects were born prior to 2006 and had to have a reliability, less parental contribution, for the respective trait under investigation of ≥60%. A total of 41,609 SNPs with a minor allele frequency >0.02 and in Hardy-Weinberg equilibrium were included in the genomic prediction.

### Results and Discussion

Across all chromosomes and animals, the mean (standard deviation in parenthesis) genotype and allele concordance rate was 0.950 (0.044) and 0.974 (0.023), respectively. However, across chromosomes the mean genotype concordance rate varied from 0.930 to 0.959, while the mean allele concordance rate varied from 0.964 to 0.979.

Mean genotype concordance rate per animal varied from 0.843 to 0.994 but was negatively skewed; 98% of animals had a genotype concordance rate of ≥0.90, 57% had a genotype concordance rate of ≥0.95, and 3% had a genotype concordance rate of ≥0.99. Mean allele concordance rate per animal varied from 0.917 to 0.996 and was also negatively skewed; 97.6% had a genotype concordance rate of ≥0.95, and 4.8% had a genotype concordance rate of ≥0.99. The mean (standard deviation in parenthesis) genotype concordance rate per animal, for animals with no parent in the reference population (n=98), only one parent in the reference population (n=632), or both parents in the reference population (n=34) was 0.927 (0.022), 0.952 (0.015) and 0.991 (0.002), respectively; the respective statistics for the allele concordance rate was 0.962 (0.012), 0.975 (0.008) and 0.995 (0.001).

The differences between the DGVs predicted using the real or imputed genotypes were normally distributed for each trait. The standard deviation of the difference between the DGVs predicted using either the real or imputed genotypes varied from 0.13 (progeny carcass fat score) to 0.47 (locomotion) of the standard deviation of the DGVs of the 746 test group animals when estimated using their real genotypes. The regression of the DGVs of the test group of animals estimated using the real genotypes on the DGVs estimated using the imputed genotypes were close to unity and only differed ($P<0.05$) from unity for perinatal mortality, progeny carcass conformation and fat score, and locomotion. The correlation between the DGVs of the test group animals when their real or imputed genotypes were used varied from 0.92 (locomotion) to 0.99 (cow carcass weight and progeny carcass fat); the average was 0.96. The correlation between the EBI estimated from the weighted sum of the individual trait DGVs predicted using the real or imputed genotypes was 0.98.

### Conclusions

The cost of genomic selection can be considerably reduced by genotyping on the Bovine3K and, through the use of imputation algorithms, obtain 'in silico' genotypes on the Bovine50K.

### Acknowledgements

Genotypes made available through sharing with international partners as well as funding from national (RSF-06-0353; RSF-06-0428) and EU Framework 7 funding (http://www.robustmilk.eu) is acknowledged

### References

Berry, D.P. Kearney, J.F., & Harris B.L. (2009). *Interbull Bulletin 40. Uppsala, Sweden*

Browning, S.R. & Browning, B.L. (2007). *Amer. J. Human Gen. 81: 1084-1097*