Is bigger better? Efficacy of HD (770K) panel

Written by Stephen D. Kachman
Department of Statistics
University of Nebraska–Lincoln

Introduction: The number of **single nucleotide polymorphisms** (SNP) used in the construction of **molecular breeding values** (MBV) has continued to grow as the cost of genotyping has dropped. The relative magnitude of the number of SNP available on individual chips is illustrated in Figure 1.

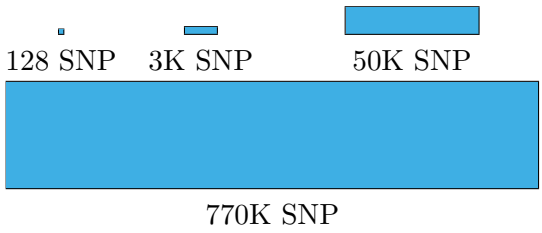

Figure 1: Relative sizes of 128, 3K, 50K, and 770K SNP panels.

Increasing the number SNP that are available has the potential of producing better tools for genomic prediction. However, the potential to do something doesn't necessarily mean that it will.

This paper will look to address the question "Does moving from a 50K panel of SNP to **high density** (HD) panel with 770K SNP make sense?". It will do this by, 1) examining the difference between information and data, 2) comparing the performance of MBV constructed using 50K and HD panels, and 3) considering where we may be headed with regard to genomic predictors.

Information versus data: A HD chip generates over 15 times the data on an animal compared to a 50K chip. The increase in data generated is coupled with a dramatically lower cost per genotyped SNP. What a HD chip doesn't do is generate 15 times the amount of information.

Information can be thought of as a measure of our ability to create a better predictor of future progeny performance. The fact that there is not a one-to-one correspondence between data and information is reflected in that we can fill in much of the data that would be generated by a HD chip using the genotypes from a 50K chip provided we have the correct supporting information.

Given a goal to produce better genomic predictors of future progeny performance, we can imagine what our ideal panel of SNP would be. Our ideal panel would include all the SNP that are economically relevant functional SNP and little else. That is all the SNP that have a functional impact of economic relevance without other SNP to complicate the picture. Unfortunately, we don't know which SNP those are. It seems likely that the total number of economically relevant functional SNP is relatively small.

We will now look to see if moving from a 50K panel to a HD panel has generated better genomic predictors.

Evaluation of the 50K and HD: One basic comparison of 50K and HD SNP panels is the relative performance of MBV constructed using the different sized panels. The relative performance of a MBV can be measured using the genetic correlation (r_g) between the MBV and the trait of interest. The MBV reliability, R^2, is the proportion of genetic variation accounted for by the MBV and is a function of the genetic correlation, $R^2 = r_g^2$.

The relative performance of MBV constructed using 50K and HD panels were examined in beef cattle by Warren Snelling and in dairy cattle by Dorian Garrick. The results, which they presented at 2012 BIF conference, are summarized in Tables 1 and 2. In both beef and dairy cat-

Table 1: Genetic correlations and standard errors for Cycle VII trained MBV.

	Trait	
Panel	Birth Wt.	REA
50K	0.52±0.05	0.30±0.12
HD	0.52±0.05	0.27±0.12

tle, the performance of MBV trained using HD panel was markedly similar to the performance of MBV trained using the 50K panel.

Table 2: Correlations for mixed breed trained MBV evaluated in the next generation.

	Trait		
Panel	Milk	Fat	Protein
	Holstein-Friesian (HF)		
50K	0.71	0.55	0.54
HD	0.71	0.58	0.57
	Jersey (J)		
50K	0.66	0.62	0.62
HD	0.66	0.59	0.58
	HFxJ cross		
50K	0.76	0.57	0.67
HD	0.75	0.59	0.69

Dorian Garrick also looked at the performance within breed and robustness across breeds of single breed trained MBV and didn't find evidence that the MBV trained using the HD panel outperformed the MBV trained using the 50K panel (Table 3). Overall, these results showed little if any benefit from using a HD panel over a 50K panel.

Table 3: Correlations for single breed trained milk volume MBV evaluated in the next generation.

	Evaluated		
Trained	HF	J	HFxJ
		50K	
HF	**0.69**	0.47	0.58
J	0.45	**0.56**	0.47
HFxJ	0.66	0.43	**0.60**
		HD	
HF	**0.70**	0.18	0.59
J	0.39	**0.59**	0.50
HFxJ	0.65	0.43	**0.62**

Improving predictors based on genomic information requires increasing the amount of information. The amount of information is due in large part to three components, 1) the number of phenotyped animals, 2) the number of genotyped animals, and 3) the number of genotyped SNP. Making good use of the resources available will require that the three types of information are kept in balance and as the previous results illustrate simply increasing one component doesn't guarantee that we will have better genomic predictors.

Where do we go from here?: To move closer to building a genomic predictor based on economically relevant functional SNP will require an integrated approach. DNA sequencing technologies will provide ever more detailed looks at an animal's genotype, technologies for measuring gene expression and regulation will provide information on which genes are active, gene networks will provide a means of looking at gene interactions, and system biology will be important as we move from the molecular to the phenotypic scale.

Ultimately, the value of any advance we make at the molecular level will be measured on the impact it has at the production level. Therefore, achieving our goal will depend on having a solid base of phenotyped and genotyped animals on which to build our genomic predictors.