From Genome Scans To Quantitative Trait Nucleotides By The *A Posteriori* Granddaughter Design

J. I. Weller[*] and M. Ron[*]

Introduction

Until the year 2000, identification of the actual quantitative trait nucleotides (QTN) was possible only for microbes, plants and model organisms. At least four QTN have been identified and verified by multiple studies in farm animals. Of these, two are in dairy cattle, *DGAT1* and *ABCG2*. Ron and Weller (2007) presented a schematic strategy for QTN determination and validation in livestock.

SNP chips including tens of thousands of SNP covering the entire genome have been developed for all the major agricultural animal species. Tens of thousands of animals have already been genotyped for Infinium(R) BovineSNP50 BeadChip, which includes 54,001 SNPs approximately evenly spaced across the entire bovine genome. Genome-wide association studies (GWAS) have been completed for thousands of bulls in the largest dairy cattle populations. The results of the GWAS for the US Holstein population correspond to the results of daughter and granddaughter designs for the two QTN that have been determined (VanRaden et al. 2009). The markers with the lowest probability values for fat and protein concentration were closely linked to the QTN for *DGAT1* and *ABCG2*. Methods have been developed to compute genomic estimated breeding values based on GWAS without determination of the actual QTN responsible for the observed genetic variance (e. g., VanRaden 2008).

In view of the revolution of genomic selection in cattle a timely question would be: can GWAS data be exploited to determine QTN? We propose a new experimental design, denoted the "*a posteriori* granddaughter design" (APGD), that uses data obtained from GWAS for QTN determination.

Determination of concordance by the *a posteriori* granddaughter design

As noted by Ron and Weller (2007) the most convincing proof of QTN determination in farm animals is "concordance". That is to demonstrate for a group of animals that their genotypes for the putative QTN correspond to their inferred genotypes for the QTL. In dairy cattle QTL genotypes can be determined for sires with many progeny by application of either

[*] ARO, The Volcani Center, P. O. Box 6, Bet Dagan 50250 ISRAEL

a daughter or granddaughter design. The number of animals for which QTL genotype has been determined must be sufficiently large so that the hypothesis of obtaining concordance by chance within the QTL confidence interval can be rejected with reasonable statistical power. Approximately 15 animals with QTL genotype determined are sufficient to reject the hypothesis of concordance by chance (Ron and Weller 2007).

Once genomic regions containing segregating QTL have been determined by a GWAS, the QTL genotype of the sires with many sons included in the sample can be determined with high certainty as follows. For each of the fathers, haplotypes are determined based on the genotypes of their sons for a chromosomal segment including a SNP with a major effect on an economic trait. The sons can be assigned into two groups, based on which paternal haplotype was received. Since many tightly linked SNPs are genotyped, it should be possible to determine for nearly all sons which paternal haplotype was received (Baruch et al. 2006). Depending on the magnitude of the effect and the number of sons per family, it should be possible to determine with relative certainty if the sire was heterozygous or homozygous for the QTL, based on the difference between the mean daughter yield deviations of the two groups of sons (Israel and Weller 2004).

In the August, 2009, Interbull listing there were 8585 US Holstein bulls born between 1999 and 2004 with evaluations for milk production traits. Of these, 4145 were sons of 19 sires each with > 150 sons included in the sample. This sample is clearly sufficient to determine concordance for a putative QTN, and to determine haplotypes of the patriarchs, based on the genotypes of the sons (Baruch et al. 2006).

Application of the APGD can be illustrated with the following example. Of 912 bulls genotyped in Israel, 24 were sons of sire 829, who was heterozygous for the causative mutation in *ABCG2* on BTA6. Based on four heterozygous SNPs between positions 37.0 and 37.3 million bp, we were able to determine which paternal haplotype was passed to 22 of the sons. The difference between the means of the genetic evaluations of the two haplotype groups was 0.183$\pm$0.040% protein, which corresponds closely to the effect of 0.2% protein estimated by Cohen-Zinder *et al.* (2005). The largest effect associated with any of the individual markers in this chromosomal region was 0.147$\pm$0.024% protein.

Unlike the traditional granddaughter design, all families analyzed by the APGD will be informative. Concordance requires that a significant contrast should be obtained for all patriarchs heterozygous for the marker, and the same allele should have the positive effect in all families. Similarly, for those patriarchs homozygous for the putative QTN, it will still be possible to distinguish between their haplotypes based on flanking markers, and to determine which haplotype was passed to each son. If the contrast between the haplotypes is significant, the marker under analysis would be excluded as the QTN.

Useful byproducts obtained by application of the APGD

1. One of the weaknesses of single marker analysis of quantitative traits by linkage disequilibrium (LD) is that when a significant effect is detected, it will generally be biased due to multiple comparisons (Weller *et al.* 2005), and will also be confounded

with additive polygenic effects. That is, bulls with common SNP genotypes will also tend to be related. By application of the APGD, all effects not linked to the chromosomal segment analyzed should be randomly distributed among the two progeny groups.

2. Unless the actual QTN was by chance included in the GWAS, the effects detected by LD analysis will underestimate the actual effect, because association between the SNP and the QTN will be incomplete. However, this will not be the case for the difference between two groups of sons that inherit the two paternal haplotypes.

3. The new analysis, based on contrasts of haplotypes within families, is sufficiently different from the preliminary LD analysis so that it can be considered a nearly independent confirmation of the effect detected.

4. Since a specific chromosomal region is targeted, the multiple comparisons problem is reduced, both with respect to the significance level required and bias.

The proposed scheme from GWAS to QTN determination in cattle

Even if 54,001 SNPs are included in the genome scan, very few will be QTN for detectable QTL. Thus, the criteria for selection of targets with the highest potential for QTN identification are:

1. Identification of a concise genomic region spanning < one million bp with several SNP markers with significant LD to the trait, and defined margins that show substantial decay of LD. Due to the huge number of SNP analyzed, the effect should be significant at the 5% level under the Bonferoni criterion of significance.
2. Priority should be given to regions that include segregating QTL based on meta-analyses (Khatkar *et al.* 2004; Salih and Adelson 2009).
3. For selection to be effective, the frequency of the economically favorable allele should be low.

As population-wide LD between a marker and a QTN can extend up to 100-500 Kbp (Sargolzaei *et al*. 2008), a region of this size may contain tens of genes. With the development of "next generation" sequencing technologies, tens of thousands of base pairs can be readily compared between individuals with opposing genotypes for the QTL (Mardis 2008). Thus the entire critical region of association delimited by margins that show substantial decay of LD can now be sequenced to determine putative QTN.

In Figure 1 the proposed scheme from GWAS to QTN identification in cattle is presented. The main features of this scheme are:

1. Application of a genome scan, using a dense matrix of SNP markers.
2. Estimation of the effects of each SNP on the economic traits.
3. Determination of target regions based on statistical analysis and previous data.

4. Determination of the physical map of genes in the target regions.
5. Determination of haplotypes of sires with many sons.
6. Deduction of QTL genotypes of these sires by the APGD.
7. Application of "next generation" sequencing technology to resequence sires with alternate QTL genotypes.
8. Determination of concordance for putative QTN.
9. Validation of the QTN effect by functional studies.

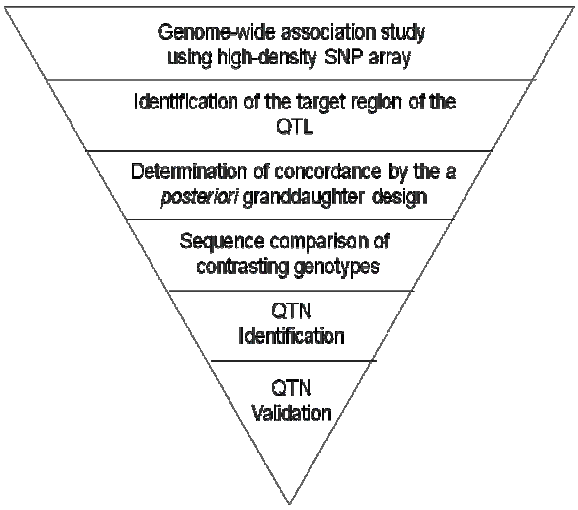

Figure 1: The proposed scheme from GWAS to QTN identification in cattle

Conclusions

Although genomic selection can increase rates of genetic gain without determination of the causative QTN, determination of these QTN should increase rates of genetic gain, and aid in the understanding of the mechanisms through which the trait is affected. Data obtained by GWAS can be used to deduce QTL genotypes of sires via application of the APGD for concordance testing of putative QTN. This together with next-generation sequencing technology will dramatically reduce costs for QTN determination. By complete genome sequencing of 20 sires with many AI sons it should be possible to determine concordance for all potential QTN detected by GWAS.

References

Baruch, E., Weller, J. I., Cohen-Zinder, M., et al. (2006). *Genetics* 172:1757-1765.

Cohen-Zinder, M., Seroussi, E., Larkin, D. M., et al. (2005). *Genome Res.*15:936-944.

Israel, C. and Weller, J. I. (2004). *Livest. Prod. Sci.* 85:189-199.

Mardis, E.R. (2008). *Trends Genet.* 24: 133–141.

Ron, M. and Weller, J. I. (2007). *Anim. Genet.* 38:429–439.

Salih, H. and Adelson, D. L. (2009). *BMC Genomics* 10:184.

Sargolzaei, M., Schenkel, F. S., Jansen, G. B., et al. (2008). *J. Dairy Sci.* 91:2106-2217.

VanRaden, P.M. (2008). *J. Dairy Sci.* 91: 4414–4423.

VanRaden, P. M, Van Tassell, C. P., Wiggans, G. R., *et al.* (2009). *J. Dairy Sci.* 92:16-24.

Weller, J. I., Shlezinger, M., and Ron M. (2005). *Genet. Sel. Evol.* 37:501–522.