Genome-wide Scan For Positional And Functional Candidate Genes Affecting Milk Production Traits In Canadian Holstein Cattle

H. Li[*], Z. Wang[*], S. S. Moore[*], F. S. Schenkel[†], P. Stothard[*]

Introduction

The completion of the bovine genome sequence assembly has enabled the detection and mapping of genetic markers in linkage disequilibrium (*LD*) with genes that contribute to genetic variation in economically important traits. Information from these markers can be used to enhance genetic improvement of livestock populations. Genome-wide association studies (*GWAS*) are an efficient way to evaluate genetic variants throughout the entire genome with the goal of identifying positional candidate genes for target traits. These candidate genes may help to more quickly track down the causative mutations for targeted traits. The successful screening of *DGAT1*, for milk fat percentage in cattle, is a good example (Winter *et al.* 2004). The available BovineSNP50 BeadChip provides us with an opportunity to identify candidate genes that affect dairy milk production traits on the entire genome. With high density marker maps, candidate genes can be fine mapped within small confidence intervals, and the subsequent focus lies in searching for functional mutations within these intervals. The objectives of this study were 1) to conduct a *GWAS* to identify major positional candidate genes that affect dairy milk production traits; 2) to search potential functional candidate genes responsible for the targeted traits.

Material and methods

Animals and phenotypic data: *DNA* was extracted from semen samples of 647 proven bulls from the Semex Alliance, Canada. These bulls were born in North America between 1985 and 2002. The pedigree of these bulls consisted of 71 sires and 492 dams with an average paternal half-sib family size of 9.1. The estimated breeding values (*EBV*s) for milk yield (*MY*), fat yield (*FY*), protein yield (*PY*), fat percentage (*FP*) and protein percentage (*PP*) were obtained from the online April 2008 genetic evaluation files released by the Canadian Dairy Network (*CDN*). De-regressed *EBV*s were used as phenotypes in the analyses.

Genotyping and marker selection: high-throughput genotyping was carried out using the Illumina BovineSNP50 BeadChip. After removing *SNPs* with unknown chromosome assignments, *SNPs* located on chromosome X, *SNPs* with missing genotyping calls and *SNPs* with a minor allele frequency (*MAF*) <0.1, a total of 29,552 *SNPs* were used in this study.

Statistical analysis: *GWAS* was implemented in GenSel by using a combination of Bayesian models with different assumptions on the variance components (Fernando 2008; Kizilkaya *et al.* 2010). First, genetic variance and residual variance components were estimated using the BayesC method. These parameters were then fitted in BayesCPi in which posterior Pi was estimated. Finally, all these parameters were used in BayesB to estimate all marker effects simultaneously.

Positional and functional candidate gene study: Based on the marker allele substitution effect for each trait, the top 10 *SNP* markers were selected for positional candidate gene searches with a flanking window of 1 *Mb* on each side of the marker. Functional candidate

[*] Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, *AB*, Canada, T6G2P5
[†] Department of Animal and Poultry Science, University of Guelph, Guelph, *ON*, Canada, N1G2W1

genes were selected from the flanking regions using a semi-automated approach that involved gathering and examining gene function predictions and publications for each gene and their orthologues in other species.

Results and discussion

Marker effects were estimated simutanoeusly in BayesB analysis for five milk production traits (Figure 1). In total, 76.8% (*PP*) to almost 100% (*FY*) of the total phenotypic variance was explained by the 29,552 markers based on the de-regressed *EBV*s. For milk production traits with medium to high heritability, the top 250 *SNP* markers explained 48.4% (*PY*) to almost 99.7% (*FP*) of the total explained genetic variance (Table 1).

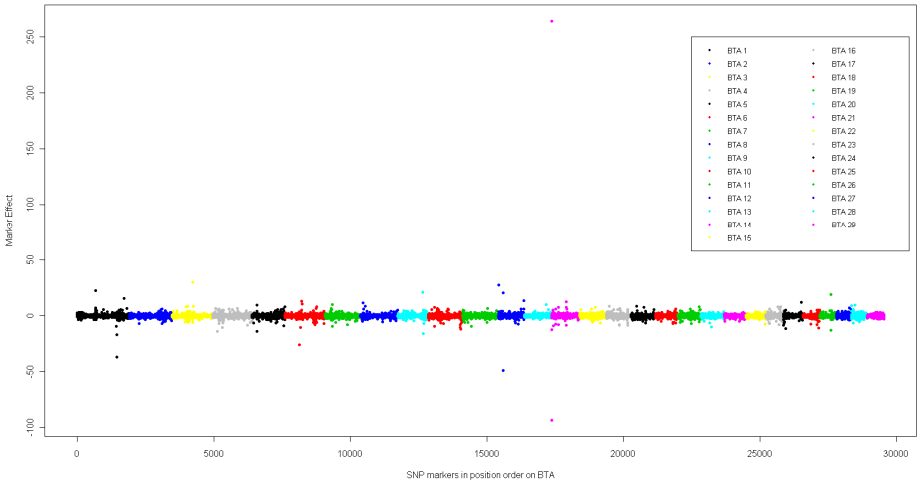

Figure 1: Allele substitution effects of *SNP* markers for *MY*

The results indicate that there are genes with large effects on milk production traits. The percentage traits (*FP* and *PP*) most likely have more genes with sizable effects than the yield traits (*FY* and *PY*) by comparing the genetic variance explained by the top 250 *SNP* markers. More markers with small to medium effects were identified for *FY* and *PY*.

Table 1: *GWAS* results for five milk production traits

Trait	Heritability[1]	Proportion of phenotypic variance a/c by all markers[2]	Proportion of explained genetic variance a/c by top 250 markers
MY	0.41	0.8546	0.8250
FY	0.34	0.9998	0.6602
PY	0.37	0.8508	0.4840
FP	0.50	0.7997	0.9967
PP	0.50	0.7682	0.8749

[1]Heritabilty values used by *CDN* in the genetic evaluations; For *FP* and *PP* the values are literature averages.
[2]Phenotypic variance is calculated based on de-regressed *EBV*s.

Results from positional candidate gene searches for the top 10 *SNPs* are presented here. In total, 108, 132, 107, 138 and 119 genes were examined for *MY*, *FY*, *PY*, *FP*, and *PP*,

respectively. Functional candidate genes were selected based on predicted and known functional information associated with the genes or their orthologues (Table 2).

Along with the confirmation of some candidate genes previously identified for milk production, *e.g. DGAT1*, *PPARGC1A* and *CYP11B1*, many new potential functional candidate genes were identified in this study. A few candidate genes were identified for more than one trait as the milk production traits are genetically correlated.

For *MY*, *MAPK15* on *BTA*14 may affect milk production through down-regulating transactivation of the glucocorticoid receptor, as glucocorticoid is an important hormone in maintaining milking (Saelzler *et al.* 2006). *CCDC80* on *BTA*1 has an effect on adipogenesis (Tremblay *et al.* 2009). *ACADM* on *BTA*3 influences lipid metabolism in rats (Saranteas *et al.* 2005).

Table 2: Identified positional and functional candidate genes of the top 10 *SNP* markers

Trait	Gene Name
MY	*BTA*1: *CCDC80*, *PLSCR4*; *BTA*3: *RABGGTB*, *ACADM*; *BTA*9: *KATNA1*; *BTA*12: *TPT1*; *BTA*14: *DGAT1*, *MAPK15*, *CYP11B1*
FY	*BTA*1: *ACPP*; *BTA*2: *KYNU*; *BTA*3: *PTGFR*, *ELTD1*; *BTA*4: *ZPBP*; *BTA*5: *LOC786490*; *BTA*7: *HAND1*; *BTA*14: *DGAT1*, *MAPK15*; *BTA*27: *RAB11FIP1*, *GOT1L1*, *ADRB3*, *STAR*, *PPAPDC1B*
PY	*BTA*6: *DHX15*, *PPARGC1A*, *SOD3*; *BTA*12: *GTF2F2*, *NUFIP1*, *TPT1*; *BTA*14: *DGAT1*, *MAPK15*, *CYP11B1*; *BTA*23: *SIRT5*, *TBC1D7*; *BTA*26: *DNAJC10*, *PNLIP*, *PNLIPRP2*
FP	*BTA*5: *SLC2A3*, *LOC786521*, *LRP6*, *LOC786490*, *DUSP16*; *BTA*7: *EFNA5*; *BTA*6: *DKK2*, *SGMS2*, *CYP2U1*, *HADH*, *NPFFR2*; *BTA*8*: SECISBP2*; *BTA*12: *UCHL3*, *TBC1D4*, *KLF12*; *BTA*20: *PRKAA1*, *PTGER4*, *OXCT1*; *BTA*14: *DGAT1*, *MAPK15*, *CYP11B1*, *KCNK9*
PP	*BTA*6: *GRSF1*, *IGJ*, *GC*, *SLC4A4*; *BTA*10: *APH1B*, *RAB8B*, *PPIB*, *LACTB*; *BTA*14: *DGAT1*, *MAPK15*, *CYP11B1*; *BTA*17: *TAOK3*; *BTA*20: *RPL37*

For *FY*, *HAND1* on *BTA*7 is an important candidate gene, as a regulator of lipid secretion (Elferink *et al.* 1995). *STAR* on *BTA*27 encodes a mitochondrial cholesterol delivery protein which can decrease intracellular lipids (Ning *et al.* 2009). Other candidate genes are involved in the plasma membrane recycling system or encode receptors or membrane glycoproteins involved in fat secretion.

For *PY*, *DHX15* on *BTA*6 functions in ribosome biogenesis and is required for release of lariat-intron from the spliceosome (Martin *et al.* 2002; Bohnsack *et al.* 2009). *GTF2F2* on *BTA*12 is a subunit of transcription factor *IIF* and is important in the suppression of activated transcription (Sopta *et al.* 1989). *NUFIP1* on *BTA*12 also works as a transcription factor to activate transcription by *RNA* polymerase II (Cabart *et al.* 2004).

For *FP*, *KCNK9* on *BTA*14 encodes a protein located in the membrance that affects aldosterone secretion (Brenner *et al.* 2008). In addition, 5 candidate genes located on *BTA*6 are associated in some way with fat metobolism. They are *DKK2*, *SGMS2*, *CYP2U1*, *HADH* and *NPFFR2*. *UCHL3* and *TBC1D4* on *BTA*12 are involved in adipogenesis in adipocytes which may have a role in fat secration in milk (Gridley *et al.* 2005; Suzuki *et al.* 2009). *PRKAA1*, *PTGER4* and *OXCT1* located on *BTA*20 may have effects on fat metabolism through integrating transcription networks in energy signalling, adipocyte differentiation and

expression of metabolic enzymes (Tsuboi *et al.* 2004; Yang *et al.* 2008; MacDonald *et al.* 2009). *SLC2A3* and *LOC786521* on *BTA*5 also have effects on adipose metabolism and energy balance (Ganguly *et al.* 2008; Shen *et al.* 2009). *LRP6* on *BTA*5 encodes the low-density lipoprotein receptor-related protein 6 and is required for normal mouse mammary gland development and may influence fat secretion in milk (Tomaszewski *et al.* 2009).

For *PP*, genes that have functions in the secretory pathway were identified on *BTA*10. *APH1B* affects gamma secretase activity (Serneels *et al.* 2005). *RAB8B* is involved in regulated secretion and polarized membrance transport (Hattula *et al.* 2002). The product of *PPIB* is most likely located at the endoplamic reticulum (*ER*) which is associated with the secretaory pathway (Caroni *et al.* 1991). *RPL37* on *BTA*20 encodes a ribosomal protein which could affect translation (Wool *et al.* 1995). The *RNA*-binding protein *GRSF1* located on *BTA*6 regulates eukaryotic protein synthesis (Park *et al.* 1999).

Conclusion

This study employed the BovineSNP50 BeadChip genotyping information for *GWAS* of milk production traits using Bayesian methods. Many positional and functional candidate genes were identified for the milk production traits across the whole genome. The results confirmed most previously identified genes with large effects for milk production traits and identified several novel candidate genes which may have functional association with the analyzed traits. These positional functional candidate genes should be further studied to assess their impact on the traits in question.

References

Bohnsack, M. T., R. Martin, *et al.* (2009). *Mol Cell* 36(4): 583-592.
Brenner, T. and K. M. O'Shaughnessy (2008). *Am J Physiol-Endoc M* 295(6): E1480-E1486.
Cabart, P., H. K. Chew, *et al.* (2004). *Oncogene* 23(31): 5316-5329.
Caroni, P., A. Rothenfluh, *et al.* (1991). *J Biol Chem* 266(17): 10739-10742.
Elferink, R. P. J. O., R. Ottenhoff, *et al.* (1995). *J Clin Invest* 95(1): 31-38.
Fernando, R. L., and D. J. Garrick (2008). *http://taurus.ansci.iastate.edu/gensel*.
Ganguly, A. and S. U. Devaskar (2008). *Am J Physiol-Endoc M* 294(6): E1144-E1151.
Gridley, S., W. S. Lane, *et al.* (2005). *Cell Signal* 17(1): 59-66.
Hattula, K., J. Furuhjelm, *et al.* (2002). *Mol Biol Cell* 13(9): 3268-3280.
Kizilkaya, K., R. L. Fernando, *et al.* (2010). *J Anim Sci* 88(2): 544-551.
MacDonald, M. J., M. J. Longacre, *et al.* (2009). *Diabetologia* 52(6): 1087-1091.
Martin, A., S. Schneider, *et al.* (2002). *J Biol Chem* 277(20): 17743-17750.
Ning, Y. X., Q. M. Bai, *et al.* (2009). *Atherosclerosis* 204(1): 114-120.
Park, Y. W., J. Wilusz, *et al.* (1999). *P Natl Acad Sci USA* 96(12): 6694-6699.
Saelzler, M. P., C. C. Spackman, *et al.* (2006). *J Biol Chem* 281(24): 16821-16832.
Saranteas, T., N. Zotos, *et al.* (2005). *Eur J Anaesth* 22(3): 222-226.
Serneels, L., T. Dejaegere, *et al.* (2005). *P Natl Acad Sci USA* 102(5): 1719-1724.
Shen, J. J., L. H. Huang, *et al.* (2009). *Mol Endocrinol* 23(1): 113-123.
Sopta, M., Z. F. Burton, *et al.* (1989). *Nature* 341(6241): 410-414.
Suzuki, M., R. Setsuie, *et al.* (2009). *Endocrinology* 150(12): 5230-5239.
Tomaszewski, M., F. J. Charchar, *et al.* (2009). *Arterioscl Throm Vas* 29(9): 1316-1321.
Tremblay, F., T. Revett, *et al.* (2009). *J Biol Chem* 284(12): 8136-8147.
Tsuboi, H., Y. Sugimoto, *et al.* (2004). *Biochem Bioph Res Co* 322(3): 1066-1072.
Winter, A., A. Alzinger, *et al.* (2004). *Genomics* 83(1): 172-180.
Wool, I. G., Y. L. Chan, *et al.* (1995). *Biochem Cell Biol* 73(11-12): 933-947.
Yang, J., S. Maika, *et al.* (2008). *Biochem Bioph Res Co* 370(2): 248-253.