Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Functional cluster analysis of genome wide associations for energy balance for cows in experimental herds in four European countries.

A.Tolkamp [*], M.P. Coffey[*], E. Wall[*], D.P. Berry‡, E. Strandberg§ and R.F. Veerkamp[†].
[*] Scotland's Rural College (SRUC), [†,] Wageningen UR Livestock Research, ‡
,Teagasc, §Swedish University of Agricultural Sciences

ABSTRACT: Genome-wide association studies (GWAS) for difficult-to-measure traits are generally limited by the sample size with accurate phenotypic data. The objective of this study was to utilise data on primiparous Holstein-Friesian dairy cows from experimental farms in Ireland, UK, the Netherlands and Sweden to identify genomic regions associated with energy balance which was modelled from dry matter intake (DMI), body condition score(BCS), live weight(LW) and fat and protein corrected milk yield (FPCM). Phenotypic data and 37,590 single nucleotide polymorphisms (SNPs) were available on up to 1,629 animals. Regions around significant SNPs identified by GWAS were analysed and functional clusters of nearby genes identified. Potential candidate genes are involved with functions linked to ion binding and proteolysis.
Keywords: energy balance; gwas; functional clustering

Introduction

In the past 2-3 decades, interest has shifted from primarily milk production towards functionality in dairy cattle. Energy balance has been shown to be genetically and phenotypically related to fertility (Wall et al, 2009) but it is unclear what is driving the relationship. Pryce et al (2002) showed there is a relationship between yield and fertility that is conditional upon body condition score (BCS). The role of feed intake and its contribution to energy balance (EB), health and fertility needs to be ascertained in a causal manner and identifying SNPs or QTLs with significant effects on these traits might help explain underlying pathways by their association with known genes.

The shift in breeding goals derives from the fact that genetic selection for yield increases feed intake but also results in a more negative energy balance (NEB) and more body tissue mobilisation during lactation (for reviews Veerkamp et al., 1993, Coffey et al., 2003, Dillon et al., 2006). Genetic selection for improvement in the feed utilisation complex has increased in interest because of its expected value in reducing methane emission of dairy cows (Wall et al., 2010) and also the need to reduce resource use per kg of food produced in an expanding human population.

Energy balance is difficult to accurately assess even in research herds where individual cow feed intake is measured. Individual cow feed intake is expensive to measure and so there has been a lot of interest in predictors of the feed utilisation complex, for example, body condition score (Koenen et al., 2001, Banos et al., 2004, Dechow et al., 2004) or other predictors of body energy state (Coffey et al., 2003). It is however, possible to model energy balance using predictor traits such as dry matter intake, live weight, body condition score and fat and milk yield (Banos et al, 2005).

Modern advances in genotyping technologies have led to the possibility of DNA based selection using SNP information to become economically viable even for expensive to measure traits such as energy balance or feed intake. The availability of high-throughput genotyping platforms with dense genome-wide markers make whole genome association studies a viable alternative to individual gene level analysis to explain the genetic variation in feed intake E.g. leptin, growth hormone or DGAT (Liefers et al., 2002, Banos et al., 2008, Oikonomou et al., 2009)

It is potentially possible to identify QTL and functional processes which underlie the genetic variation in the feed utilisation complex. However, the limiting factor in such analyses is the number of animals with accurate phenotypes in a single dataset. To overcome this limitation, in this study we collate cow genotypic and phenotypic information from research herds in Ireland, UK, the Netherlands and Sweden and related genetic markers across the genome to the phenotypes for the feed utilisation complex using a Bayesian stochastic search variable selection statistical approach.

The objectives, therefore, of this study were to discover functional groups of genes in close proximity to SNPs identified by genome wide association of energy balance.

Materials and Methods

Phenotypic data.

Phenotypic data were available on 2,031 Irish, 1,018 UK, 725 Dutch, and 225 Swedish Holstein-Friesian cows, but only first lactation records were selected from 1804 first lactation Holstein-Friesian cows with 66,116 test-day records collected up to 45 weeks (315 days) in lactation. Records were milk, fat and protein yield, live-weight (LW), body condition score (BCS), dry matter intake (DMI) and fat and protein corrected milk yield (FPCM). From these records average energy balance in weeks 1-44 was estimated as described in (Banos et al.,2012) for use in GWAS.

Genotypic Data.

All animals with phenotypic information were genotyped with the two releases of the Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA) containing 54,001 single nucleotide polymorphisms (SNPs). SNPs that fulfilled the following criteria were included in the association study: 1) GCscore > 0.20 and GTscore > 0.55; 2) call rate > 0.95%; 3) minor allele frequency > 0.01 in each country; and 4) no extreme deviation from Hardy Weinberg Equilibrium (i.e., $\chi 2 < 600$). The GCscore and GTscore are quality measures on the genotype calls from the genotyping assay. After the quality control edits 37,590 SNPs remained. Checks for Mendelian inconsistencies between pedigree and SNP data were performed for all genotyped parent-offspring pairs and among sibs (Calus et al., 2011) and animals with suspected error pedigrees removed. After these edits 1629 animals with both phenotypes and genotypes remained in the data. Missing SNPs were imputed using Beagle (Browning and Browning, 2007). Chromosome number and locations of the SNPs on the BovineSNP50 were obtained from the UMD3.0 bovine genome assembly from the University of Maryland.

Genome-wide association analysis

The Bayesian Stochastic Search Variable Selection (**BSSVS**) model (described by Calus et al., 2008, Verbyla et al., 2010) was used to sample whether SNPs were linked to a QTL or not. The model used was:

$$y_i = \mu + animal_i + \sum_{j=1}^{nloc}\sum_{k=1}^{2} SNP_{ijk} + e_i$$

where y_i is the phenotypic record of animal i, μ is the overall mean, $animal_i$ is the random polygenic effect of animal i, nloc is number of SNP markers, SNP_{ijk} is a random effect for allele k at locus j of animal i, and e_i is a random residual for animal i. The vectors **animal** and **e** where assumed to be normally distributed, **animal** ~$\mathbf{N}(0, \mathbf{A}\sigma_u^2)$ and **e**~ $\mathbf{N}(0, \mathbf{I}\sigma_e^2)$, where **A** is the numerator relationship matrix and **I** is an identity matrix.

Functional clustering

A catalogue of genes within 500kbp of significant SNPs for energy balance were compiled using the UMD3.1 assembly from the ensemble database which was accessed via bioMART (Haider et al., 2009). A SNP was assumed to be significant if the Bayes factor (BF; Kass and Raftery, 1995) was >3.1, which is termed a 'substantial' effect.
The package DAVID (http://david.abcc.ncifcrf.gov/) was used to create functional annotation clusters: groupings of genes by similar functionality, and to test whether certain processes and functions were over represented around the significant genes using algorithms described in Huang et al. (2009)

Results and Discussion

Genome wide association of data collected across four experimental herds identified 73 SNPs from 18 chromosomes associated with energy balance. The GWAS gave 65 SNP with a BF>15 for any of the EB traits, but SNP were clearly different across lactation stages. For example, SNP on chromosome 5 showed a clear associated with EB in mid lactations, whereas the SNP on Chr8 showed a clearer association with EB in early lactation, whereas Chr10 and Chr24 had SNP that where mainly associated with EB at the end of lactation. The 5 SNP with the largest effect were two SNP on CHR5 both associated with EB3 (Hapmap38248-BTA-73843 (BF=75), BTB-00232920 (BF=212)), one SNP on CHR8 (ARS-BFGL-NGS-72418) affecting EB1 (BF=161), CHR26 ARS-BFGL-NGS-114308 affecting EB5 (BF=79) and ARS-BFGL-NGS-37484 on CHR21 associated with (BF=57)

Of greater interest is whether those SNPs are associated with known genes or pathways that have plausible biological explanations in their role in energy balance. In total 26 genes were found within 500kbp of these SNPs. Two functional annotation clusters were found. (Figure 1). The first,(enrichment score 1.62) contained ion binding proteins and the second (enrichment score 1.22) contained peptidases and metallopeptidases. There was overlap in the genes found in each cluster. 11 genes were un clustered. These results are different from GWAS assessing DMI, LW and BCS individually which found genes linked to insulin, epidermal growth factor and tryptophan. (Veerkamp et al. 2012). Several genes were found to be of particular interest. Cluster one contained ENSBTAG00000018007 also known as NR2F2 which has been linked to glucose homeostasis in humans. Two genes (ENSBTAG00000016145, ENSBTAG00000023146) clustered in Annotation cluster two are metallopeptidases with Fertillin precursor motifs, which play a role in sperm-egg binding and fertility. The fact that the results were so different compared to traits commonly associated with energy balance; DMI, LW, and BCS, and in themselves are so diverse implies there are more, currently unknown factors affecting this energy balance.

Figure 1: Functional annotation clusters within 500 kbp significant SNPs for energy balance.

	Group Term	Count
Cluster 1	Metal ion binding	8
Enrichment Score: 1.624	Cation binding	8
	ion binding	8
	Zinc ion binding	5
	Transition metal ion binding	5
Cluster 2	Metallopeptidase activity	3
Enrichment Score: 1.224	Zing ion binding	5

Proteolysis	4
Peptidase activity, acting on L-animo acid peptides	3
Peptidase activity	3
Transition metal ion binding	3

Acknowledgements

This study is part of the RobustMilk project (http://www.robustmilk.eu), which is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services. Scottish government is recognised for continue funding of the Langhill lines of dairy cattle. Many thanks to Y. de Haas, M.P.L. Calus , H. Bovenhuis, and G. Banos for invaluable assistance in the data collection and analysis.

Literature Cited

Banos G, Brotherstone S and Coffey M.P., (2004). J Dairy Sci 87, 2669-2676.

Banos G, Coffey MP and Brotherstone S. (2005). J Dairy Sci. 88, 2226-2237

Banos G, Coffey M. P., Veerkamp R. F., Berry D.P, E.Wall. (2012) Anim. , 6:7, pp 1040–1048

Coffey M. P. , Simm G, Hill W. G. and Brotherstone S. (2003). J Dairy Sci 86, 2205-2212.

Dechow C. D., Rogers G. W., Klei L., Lawlor T. J. and VanRaden P. M. (2004). J Dairy Sci 87, 3534-3541

Dillon P, Berry D. P., Evans R. D., Buckley F and Horan B. (2006). Livest Sci. 99, 141-158.

Haider S, Ballester B, Smedley D, Zhang J. J., Rice P and Kasprzyk A. (2009). Nucleic Acids Res 37, W23-W27.

Huang DW, Sherman BT, Lempicki R. A., (2009) Nature Protoc.;4(1):44-57

Kass RE and Raftery AE (1995). J. Am. Statist. Assoc. 90, 773–795.

Huang D. W., Sherman B. T., Lempicki R. A.,(2009) Nucleic Acids Res. ;37(1):1-13

Koenen E. P. C., Veerkamp R. F., Dobbelaar P. and De Jong G. (2001). J Dairy Sci. 84, 1265-1270.

Liefers S. C., te Pas M. F. W., Veerkamp R. F. and van der Lende T. (2002). J Dairy Sci. 85, 1633-1638.

Oikonomou G, Angelopoulou K, Arsenos G, Zygoyiannis D and Banos G. (2009). Animal Genet 40, 10-17Pryce J. E., Coffey M. P., Brotherstone S. H., Wooliams J. A.(2002) J Dairy Sci. 85, 1590-1595

Veerkamp R. F., Coffey M. P., Berry D. P., de Haas Y., Strandberg E., Bovenhuis H., Calus M. P. L., Wall E. (2012). Anim. , 6:11, pp 1738–1749

Veerkamp R. F., Emmans G. C., Oldham J. D., and Simm G. (1993: Proceedings Fourth Zodiac Symposium, Wageningen, The Netherlands, April.

Wall E., Simm G. and Moran D. (2010). Anim. 4, 366-376.