Proc. Assoc. Advmt. Anim. Breed. Genet. 22:145-148

GENOME-WIDE UNCORRELATED TRAIT ANALYSIS IDENTIFIES PLEITROPIC MARKERS FOR DAIRY CATTLE IN AUSTRALIA

R. Xiang[1,2], I.M. MacLeod[2], S. Bolormaa[2,3] and M.E. Goddard[1,2]

[1] Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC3010, Australia

[2] AgriBio, Dept. Economic Development, Jobs, Transport & Resources, Bundoora, VIC3083, Australia

[3] Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW 2351, Australia

SUMMARY

Selection should favour alleles which increase profitability considering their effects across all important traits. Therefore, understanding pleiotropy is an important aim. Obviously if traits are genetically correlated they must share some causal variants but it is possible that even uncorrelated traits share some causal variants. Here we analyse 25 traits on Australian dairy cattle. The 25 raw traits (RTs), covering milk production, fertility, behaviour, somatic cell count and conformation, of 2841 bulls were used to calculate uncorrelated principal components (PCs) and Cholesky transformation traits (CT). Multi-trait meta-analyses of single-trait genome-wide association studies (GWAS) for RT, PC and CT in these bulls were validated in 6821 cows. We observed a positive relationship between heritability estimates and the number significant SNPs detected in RTs and CTs. However, there was no relationship between the phenotypic importance of PCs and the number of significant SNPs detected. The major dairy cattle locus DGAT1 not only affected dairy production traits, also had validated small effects on fertility, milk speed and temperament. Our results highlight the importance of using genetic information of all traits to maximise pleiotropy detection and prioritise multi-trait genetic markers for the dairy industry.

INTRODUCTION

The profitability of dairy farming depends on many traits including milk production, fertility, diseases, workability and conformation or type traits (Byrne et al., 2015). Therefore, genomic selection should target genetic variants that increase an economic combination of traits such as the balanced performance index (BPI). When identifying genetic markers, such as single nucleotide polymorphisms (SNPs), associated with economic traits, we need to know the effect of the marker on all economic traits not just those where the marker has the biggest effect. That is, we would like to understand the pleiotropic effects of genes across all important traits.

Widespread pleiotropic effects of SNPs have been observed in beef cattle (Bolormaa et al., 2014) and sheep (Bolormaa et al., 2016). If traits are genetically correlated there must be some genes that affect both traits. However, it is also possible that uncorrelated traits share some causal variants. Principal component (PC) analysis, producing a small number of uncorrelated traits, has been proposed for conducting multi-trait genetic analysis (Klei, Luca, Devlin, & Roeder, 2008). If genes act through a limited number of physiological pathways, principle component analysis might capture the most important pathways in the first few PCs leading to a simple picture of pleiotropy.

To further understand pleiotropy in the dairy cattle population, a dataset from the Australian Dairy Herd Improvement Scheme (ADHIS) with 25 traits recorded on 9662 animals was retrieved. These 25 raw traits (RTs), including milk production, survival, fertility, temperament and linear type traits, were used to construct uncorrelated PCs and Cholesky transformed traits (CTs) (Golub & Van Loan, 2012). RTs and generated PCs and CTs were analysed with multi-trait genome-wide association studies (GWAS).

MATERIALS AND METHODS

Analyses included genotype of 2841 bulls as the discovery population and 6821 cows as the validation population from the breeds Holstein, Jersey and Australian Red. The distribution of genomic relatedness of bulls and cows in three breeds were shown in Figure 1. SNPs were genotyped by Illumina BovineLD BeadChip (7K), Illumina Bovine SNP array (54K) and Illumina Bovine HD genotypes (777 K). All animals were imputed to HD genotypes using Fimpute (Sargolzaei, Chesnais, & Schenkel, 2014) and in total, 632,002 SNPs were used. SNPs with minor allele frequency <0.01 or significant departure from Hardy-Weinberg equilibrium (p<0.001) were filtered out. The 25 phenotypic traits of these animals (trait deviations for cows and daughter trait deviations for bulls) were from the April 2016 genetic evaluations from the DataGene. Daughter trait deviations were the average trait deviations of a bull's daughters and all phenotypes were pre-corrected for known fixed effects.

The generation of PCs for the nth animal (u_n) was based on eigen-decomposition of $k=25$ RTs (g_n): $u_n = T'g_n$; Where u_n was a $k \times 1$ vector of PC scores for the animal n; T was an $k \times k$ matrix of eigenvectors such that the variance matrix of the PC Var(T'g) = D, a diagonal matrix of eigenvalues; g_n was an $k \times 1$ vector of RT for animal n. The CT scores for the nth animal (c_n) were calculated based on the Cholesky decomposition: $c_n = L^{-1}g_n$; where; L was the $k \times k$ matrix of the Cholesky factors which satisfied LL^t= V(g), the $k \times k$ covariance matrix (Golub & Van Loan, 2012); g_n was a k × 1 vector of RT for the animal n. Single-trait GWAS was performed in GEMMA (Zhou & Stephens, 2014) using data from the discovery population: $y = mean + fixed\ effects + SNP_i + GRM + e$; where y = vector of k RTs, PCs or CTs for bulls; fixed effects= breeds; SNP_i = that each SNP genotype was fitted as a covariate one at a time; a polygenic random effect described by the GRM= genomic relatedness matrix calculated from GEMMA based on all SNPs; e = error. A multi-trait meta-analysis based on either the 25 RTs, 25 PCs or 25 CTs followed previous procedures (Bolormaa et al., 2016; Bolormaa et al., 2014). SNPs that were significant in the discovery sample were tested in the validation sample using an index of traits that maximises the effect of the SNP (Bolormaa et al., 2016; Bolormaa et al., 2014). Single-trait GWAS in the validation population was also used to confirm SNP effects on individual RTs.

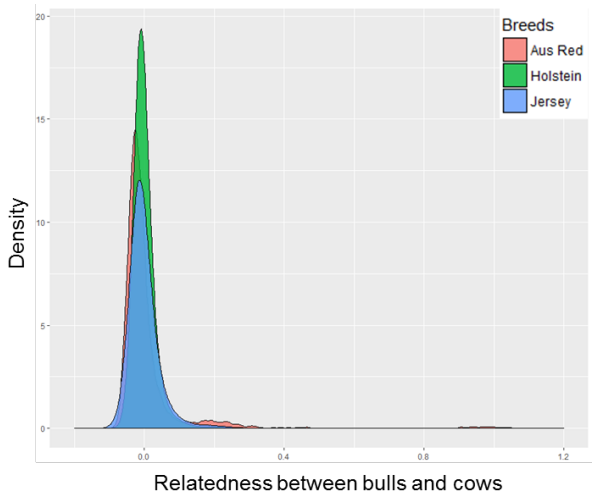


Figure 1. Density plot of the genomic relationship matrix between bulls and cows

RESULTS AND DISCUSSION

For both RTs and CTs, the number of significant ($P<1\times10^{-5}$) SNPs detected by single-trait GWAS generally increased with the estimated heritability of the phenotype because the power to detect effects increases with h^2 (Figure 2A,B). (The heritability of bull phenotypes is the proportion of variation in daughter trait deviation explained by all SNPs jointly). Consistent with previous reports (Kemper et al., 2015; MacLeod et al., 2016), the RT of milk, protein and fat yield had the highest heritability estimates (all h^2>0.8 and se=0.02) and the largest numbers of significant SNPs (more than 100) detected. Survival and fertility as reproductive complex traits had mid-range heritability estimates (both h^2>0.5 and se=0.03) with 27 and 31 significant SNPs detected, respectively. Mid-range heritability was also estimated for temperament and milk speed (both h^2>0.5 and se=0.03). However, single-trait GWAS only detected 6 and 13 significant SNPs for temperament and milk speed, respectively. The h^2 of likeability is 0.48 (se=0.03) with only four significant SNPs detected. The heritability estimates of dairy type traits ranged from 0.35 (rear legs set, se=0.04) to 0.69 (front teat placement, se=0.03). However, all type traits had a small number of significant SNP detected. Rear legs set had 15 significant SNPs and front teat placement had only 1 significant SNP. This is likely to be due to the complexity of the type traits, i.e. a large number of causal variants each with a very small effect. Our discovery sample size may not be large enough to capture highly significant SNPs.

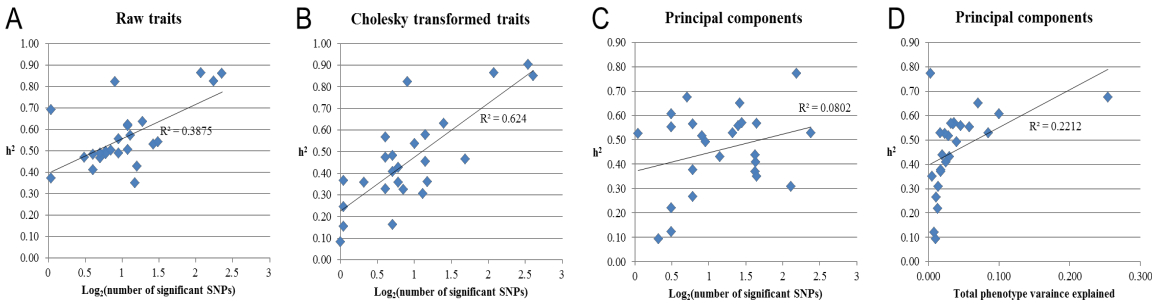


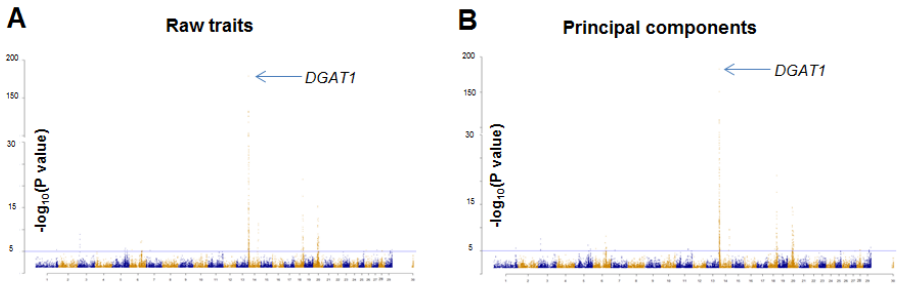
Figure 2. The relationship between heritability estimates and the number of significant SNPs detected by single-trait GWAS for RTs (A), CTs (B) and PC (C). D: The relationship between the heritability estimates of each PC and the total phenotypic variance explained by each PC

The 25 PCs, accumulatively explained 93% of the total phenotypic variances, showed a complicated pattern (Figure 2C,D). The first PC, which explained 25% of total phenotypic variances for all RTs, had a high estimation of heritability (0.67, se=0.03) but only 5 significant SNPs. This PC had loadings from many traits and perhaps this generates a very complex trait affected by many genes. On the other hand, PC18 with top factor loading related to milk fat yield, only contributed 1.7% of the variances to all traits, had a modest heritability (0.53, se=0.03), but had the largest number of significant SNPs (241) amongst the PCs. The last PC (PC25) with high positive factor loading for protein yield and high negative factor lading for milk yield, explained 0.03% of the variances in all traits, had a modest heritability (0.50±0.03) but l53 significant SNPs. Our results are consistent with a previous simulation study in humans where the genetic information of all PCs are important (Aschard et al., 2014). Thus, only considering a small number of PCs might cause loss of power for genetic analysis.

Thus the results do not support the hypothesis that genes act through a small number of common, physiological mechanisms. This is exemplified by SNPs within and near DGAT which

have significant effects on several PCs. This occurs because the effects of DGAT do not follow the pattern described by the overall genetic correlations. For instance, milk and fat yield are positively correlated but the allele of DGAT which increases milk decreases fat yield.

Three multi-trait meta-analyses were performed based on either all RTs or PCs or CTs. The 3 meta-analyses largely detected the same significant SNPs as they are all approximations to a full multi-trait analysis. They also detected many more significant SNPs than single-trait GWAS using the same threshold ($P<1\times10^{-5}$ and FDR<0.01) (Figure 3A,B).



C

DGAT1.Chr14:1.8M

RT.Prot	RT.Fat	RT.Milk	RT.SCC	RT.SurvD	RT.Fert	RT.Temp	RT.MSpeed	RT.Stat	RT.Like	RT.Angul	RT.Bone	RT.CentL	RT.ChestW	RT.ForeA	RT.MuzW	RT.PinSet	RT.PinW	RT.RSet	RT.RearAH	RT.RearAW	RT.TeatPF	RT.UdTex	RT.OType	RT.Mamm
-12.8	16.3	-17.6			2.9	1.8	-1.5							3.0			-1.1		-2.1	-2.2				
PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25
	2.0			1.1	1.5	2.2	-2.0	-3.1	2.6	4.9	5.0	7.3	-2.6	-7.1	5.1	-7.1	-16.3	11.8		-4.1		2.5	7.7	3.3
CT.Prot	CT.Fat	CT.Milk	CT.SCC	CT.SurvD	CT.Fert	CT.Temp	CT.MSpeed	CT.Stat	CT.Like	CT.Angul	CT.Bone	CT.CentL	CT.ChestW	CT.ForeA	CT.MuzW	CT.PinSet	CT.PinW	CT.RSet	CT.RearAH	CT.RearAW	CT.TeatPF	CT.UdTex	CT.OType	CT.Mamm
-12.8	23.6	-1.9	-1.3		4.5	1.6	-1.2	3.7			1.4	1.4	-2.2	1.7	-1.1				-1.5		2.0		-4.1	2.0

Figure 3. A-B: Manhattan plot of multi-trait meta-analysis. C: t values with absolute values >1 of DGAT1 across traits

DGAT1 was the most significant locus in the multi-trait analysis (Figure 3A,B) with effects on many RT, PCs and CTs. Along with the strong effects on production traits, DGAT1 also had small but validated effects on fertility, milk speed, temperament and type RTs, which are important information for the breeders (Figure 3C). This highlights the advantage of conducting multi-trait analysis in extending knowledge for unknown effects of known loci. Most SNPs did not have a significant effect on many traits as DGAT did but this may indicate a lack of power rather than a lack of pleiotropic effects. If the example of DGAT is repeated for other loci it is important because it indicates that SNPs with a small effect on one trait may be detected by their large effect on another trait.

REFERENCES

Aschard, H., Vilhjálmsson, B. J., et al. (2014). *Am. J. Hum. Genet.,* ***94***(5), 662-676.
Bolormaa, S., Hayes, B. J., et al. (2016). *BMC Genomics,* ***17***(1), 1.
Bolormaa, S., Pryce, J. E., et al. (2014). *PLoS Genet,* ***10***(3), e1004198.
Byrne, T., Martin-Collado, D., et al. (2015). *Proc. Assoc. Advmt. Breed. Genet,* ***21***, 21-24.
Golub, G. H., & Van Loan, C. F. (2012). *Matrix computations* (Vol. **3**): JHU Press.
Kemper, K. E., Reich, C. M., et al. (2015). *Genet. Sel. Evol.,* ***47***(1), 1.
Klei, L., Luca, D., et al. (2008). *Genet. Epidemiol.,* ***32***(1), 9-19.
MacLeod, I., Bowman, P., et al. (2016). *BMC Genomics,* ***17***(1), 1.
Sargolzaei, M., Chesnais, J. P., et al. (2014). *BMC Genomics,* ***15***(1), 1.
Yang, J., Lee, S. H., et al. (2011). *Am. J. Hum. Genet.,* ***88***(1), 76-82.
Zhou, X., & Stephens, M. (2014). *Nat. Methods,* ***11***(4), 407.