A New Method for Exploring Genome-wide Associations Applied to Cattle Puberty

M. R. S. Fortes[*], A. Reverter[‡], Y. Zhang[†], E. Collis[‡], S. H. Nagaraj[‡], N. N. Jonsson[*], W. Barris[‡], R. J. Hawken[‡]

Introduction

Genome wide approaches are now commonly used to examine the genetics underlying complex phenotypes. These approaches yield a varying number of loci associated with the phenotype, depending on significance thresholds and interpretation of results is less than straightforward. There is a consequential need to develop new analytical approaches to maximise the value of genome wide association studies (GWAS). A systemic approach is fitted for complex phenotypes (Mackay, Stone et al. 2009). It allows a large number of associations by using relaxed significance thresholds, while simultaneously reducing the number of false positives.

Here, we describe a new systemic approach applied to a GWAS conducted in Brahman cattle. We call the approach an association weight matrix (AWM). The use of AWM is two-fold: it estimates correlations between phenotypes and it explores gene to gene interactions across the genome. We demonstrate the application of AWM for complex phenotypes to examine the genetic drivers of puberty in Brahman cattle.

Material and methods

We used data from 843 Brahman cows from a herd bred by the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC). This *Bos indicus* herd was described in detail previously (Barwick, Johnston et al. 2009; Johnston, Barwick et al. 2009). The GWAS considers 22 phenotypes measured on three occasions: when the mean age of animals was 18 months (T1); at the time of observation of the first corpus luteum (CL) and when the mean age was 24 months (T2). At T1, 8 phenotypes were measured including: live weight (WT, kg), hip height (HH, cm), serum concentration of insulin-like growth factor I (IGF-I, ng/mL), average daily weight gain (ADG, kg/day), body condition score (CS, score 1-10), scanned *Longissimus dorsi* area (SEMA, cm2), scanned fat depth at the P8 site (SP8, mm) and scanned fat depth measured between the last 2 ribs (SRIB, mm). At first CL, which was detected through regular ovarian scans, we recorded the cow's age (AGECL), live weight (WTCL, kg) and subcutaneous fat depth at the P8 site (FATCL, mm). We consider AGECL as a phenotype for time of puberty. Also, the presence or absence of a CL close to the first day of joining was recorded (CLJOIN). At T2, the measurements taken at T1 were repeated.

Cooperative Research Centre for Beef Genetic Technologies, Australia.
[*] School of Veterinary Science, The University of Queensland, Australia
[‡] CSIRO Livestock Industries, Queensland Bioscience Precinct, Brisbane, Australia
[†] Animal Genetics and Breeding Unit, University of New England, Armidale, Australia

A full description of these phenotypic measurements is published elsewhere (Barwick, Johnston et al. 2009; Johnston, Barwick et al. 2009). In addition, heifers that reached puberty prior to the first mating season, conceived and calved had another phenotype measured: post partum anoestrus interval (PPAI), defined as the interval, in days, between calving and first CL after calving. Also, post partum anoestrus interval with respect to weaning time (PW), a related binary phenotype was recorded. The BovineSNP50 Bead Chip (Illumina 2008) was used to genotype 843 cows. The additive effect of a SNP on each phenotype was calculated by regression analysis after fitting an animal full mixed-model with components estimated using the ASREML software (Gilmour AR 2006). Fixed effects included were contemporary groups, herd of origin, sex of calf, month of calving and sire of calf. SNP were fitted as random effects.

The association results obtained with ASREML were used to construct an association weight matrix (AWM), commencing with the selection of associated SNP. The criteria to select SNP that would represent genes on the AWM include the significance ($p<0.05$) of the allele substitution effect and the SNP genomic position. Our selection criteria were developed to favour genes harbouring SNP with significant association across related phenotypes. The AWM was constructed with as many rows as selected SNP and as many columns as phenotypes. Each AWM cell {i,j} contains a value corresponding to the z-score normalized additive effect of the i-th SNP on the j-th phenotype. The AWM approach explores phenotypic correlations column-wise and gene correlations row-wise. Column-wise, Pearson correlations between AGECL and the other 21 phenotypes were calculated using the SNP effect values. The results of these SNP-based correlations were compared with the genetic correlations, estimated via pedigree-based restricted maximum likelihood (REML), established for the same populations previously (Johnston, Barwick et al. 2009). Row-wise, the AWM explores the correlations between SNP effects to predict gene interactions forming a network. In the network, genes are nodes connected by edges representing a significant correlation that was established by the PCIT algorithm (Reverter and Chan 2008). We studied the predicted gene interactions to identify genetic drivers of puberty. To provide an *in-silico* validation for gene-gene interactions predicted via AWM, we performed regulatory sequence analysis of genes predicted to interact with key transcription factors (TFs), an estrogen receptor (ESRRG) and a TF of fat metabolism (PPARG) using Genomatix (http://www.genomatix.de/).

Results and discussion

On average, our Brahman cows were pubertal at 25 month of age (AGECL), which is within the expected range for *Bos indicus* breeds (Abeygunawardena and Dematawewa 2004). Averages, standard deviations and definitions for each phenotype are summarized in Table 1. AGECL heritability is estimated at 0.57 (Johnston, Barwick et al. 2009), indicating its high genetic component that we aimed to dissect with the new AWM approach.

Table 1: Definitions and descriptive statistics for the 22 phenotypes measured

Phenotype		Description	Average	St. dev.
CL	CLJOIN (0 or 1)	presence of CL on joining	0.43	0.50
	AGECL (days)	age at first CL	750.60	142.10
	FATCL (mm)	P8 fat at first CL	4.47	2.19
	WTCL (kg)	live weight at first CL	334.40	44.80
T1	ADG (kg/day)	average daily gain	0.61	0.15
	CS (score 1-10)	condition score	8.30	1.40
	SEMA(cm2)	area of eye muscle	44.10	6.60
	HH (cm)	hip height	127.40	4.90
	IGF (ng/ml)	serum IGF level	182.60	84.30
	SP8 (mm)	subcutaneous P8 fat	3.70	1.90
	SRIB (mm)	rib fat	2.00	1.00
	WT (kg)	liveweight	287.60	43.80
T2	ADG (kg/day)	average daily gain	0.14	0.23
	CS (score 1-10)	condition score	7.40	1.40
	SEMA(cm2)	area of eye muscle	44.10	8.80
	HH (cm)	hip height	132.40	4.90
	IGF (ng/ml)	serum IGF level	215.40	92.30
	SP8 (mm)	subcutaneous P8 fat	3.10	1.80
	SRIB (mm)	rib fat	1.90	1.00
	WT (kg)	liveweight	320.00	58.70
PPAI	PPAI (days)	post partum anestrus interval	179.98	108.57
	PW (0 or 1)	presence of CL before or after waning	0.90	0.84

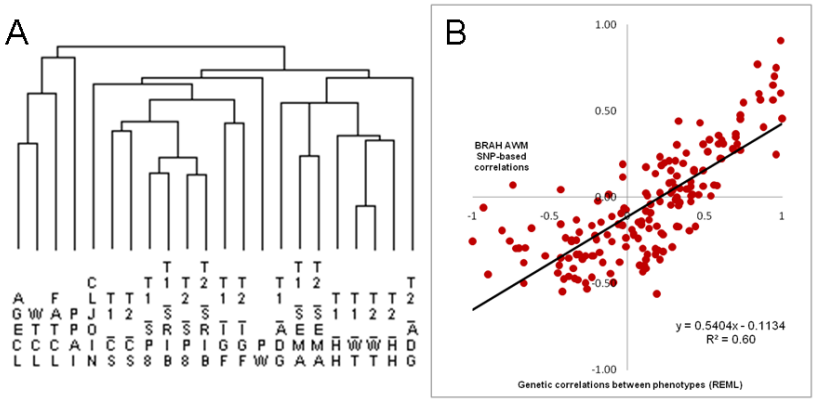

Figure 1: A) Tree of the hierarchical clustering of phenotypes based on the AWM matrix; B) Comparison between the genetic correlations estimated by REML (x-axis) and those estimated based on SNP effects (y-axis)

Column-wise we used AWM to estimate correlations between phenotypes, shown as a hierarchical tree cluster (Figure 1, A). AWM-driven correlations were similar to genetic correlations established previously (Figure 1, B). Row-wise, we used AWM to predict gene-gene interactions and build a network for cattle puberty. Figure 2 illustrates a subset of the AWM network, showing the first neighbors of 3 important genes: ESRRG, PPARG and a zinc finger (ZNF462) previously associated with puberty in humans (Perry, Stolk et al. 2009). Amongst ESRRG and PPARG first neighbors, 27% have a corresponding binding site, providing evidence for our AWM predictions.

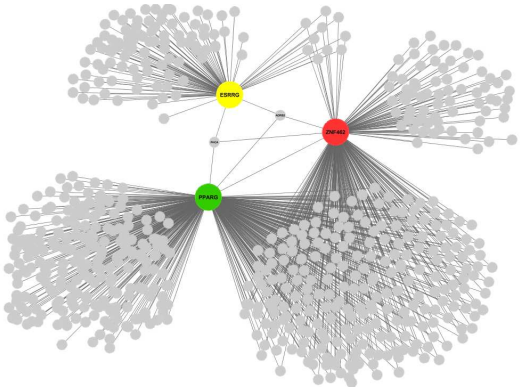

Figure 2: Subset of the AWM-predicted gene network for cattle puberty.

Conclusion

AWM is a hypothesis generating process, appropriate for application to complex phenotypes such as puberty. Our results show that AWM recapitulates the known biology behind puberty; captures experimentally validated binding sites, and identifies novel gene-gene interactions for further investigation.

References

Abeygunawardena, H. and Dematawewa, C. M. (2004). Anim Reprod Sci 82-83: 373-87.

Barwick, S. A., Johnston, D. J. Burrow, H. M. et al. (2009). Animal Production Science 49(5-6): 367-382.

Gilmour A. R., Gogel, B., Welham S., et al. (2006). ASReml, user guide. Release 2.0 Hemel Hempstead. VSN International Ltd.

Johnston, D. J., Barwick, S. A., Corbet, N. J. et al. (2009). Animal Production Science 49(5-6): 399-412.

Mackay, T. F., Stone, E. A. and Ayroles, J. F. (2009). Nat Rev Genet 10(8): 565-77.

Perry, J. R., Stolk, L., Franceschini, N. et al. (2009). Nat Genet.

Reverter, A. and Chan, E. K. (2008). Bioinformatics 24(21): 2491-7.