



Candidate mutations used to aid the prediction of genetic merit for female reproductive traits in tropical beef cattle

Marina Rufino Salinas Fortes^{1,3*} , Charmaine Enculescu¹, Laercio R. Porto Neto², Sigrid A. Lehnert², Russell McCulloch², Ben Hayes³

¹ The University of Queensland, School of Chemistry and Molecular Biosciences, St Lucia, Brisbane, Australia.

² CSIRO Agriculture and Food, St Lucia, Brisbane, Australia.

³ The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, St Lucia, Brisbane, Australia.

ABSTRACT - In this study, we aimed to provide a wet laboratory validation for a set of single nucleotide polymorphisms (SNP), which had been identified as candidate functional variants *in silico*. Genotyping for candidate SNP was performed in Brahman and Tropical Composite cattle. After quality control, 29 SNP were first investigated individually for their association with female reproductive traits and then used as a panel for genomic predictions. The reproductive traits studied were age at first *corpus luteum* (AGECL; days), post-partum anoestrus interval (PPAI; days), and a binary trait that described if the cow had ovulated before weaning the first calf or not (PW, 0-1). Single nucleotide polymorphisms in six genes (*FOXA2*, *TRAF4*, *IRF2*, *IRF1*, *BPTF*, and *CPEB1*) were found to be significantly associated with reproduction traits. The genomic prediction method used was BayesR, to accommodate the 29 new SNP and compare their performance with predictions based on 50K genotypes (Illumina SNP chip). When new SNP and *PLAG1* mutation rs109231213 were included in the genomic predictions for female reproductive traits their accuracies improved. The best predictions were obtained by combining the new SNP and the 50K SNP using BayesR analysis, with a 4% improvement in accuracy. The proportion of the genetic variance explained by the new SNP together was 0.07 for AGECL, 0.03 for PPAI, and 0.02 for PW. It would be favourable to include these new SNP in future versions of bovine SNP chips to target selection for female reproductive traits. These new SNP are likely to improve genomic predictions for female reproductive traits in tropical beef cattle breeds, with varying degrees of *Bos indicus* content.

Key Words: anoestrus, BayesR, candidate genes, genomic prediction, heifer puberty

Introduction

Single nucleotide polymorphisms (SNP) are used to predict genomic estimated breeding values (GEBV). Genomic prediction holds great potential for beef cattle (Garrick, 2011; Meuwissen, et al., 2001) and has revolutionized dairy cattle breeding (Barkema, et al., 2015; Boichard, et al., 2016). The uptake of GEBV in beef cattle has been relatively slow. The multi-breed nature of the beef industry is a challenge for GEBV (Garrick, 2011; Rolf, et al., 2014).

Causative mutations could improve GEBV across breeds (Saatchi et al., 2014). However, causality is difficult to prove, and so, we turn our attention to “functional mutations”, such as non-synonymous SNP that alter the

peptide, or regulatory SNP that impact gene expression. Gene networks and pathways have been used to select functional mutations that yield more accurate GEBV than high-density SNP chip (Snelling et al., 2013). Selected SNP panels formed by functional mutations may improve the portability of GEBV across breeds and into crossbreds (Snelling et al., 2012; 2013). The use of causative mutations yielded an increase in accuracy of 2.5-3.7%, when using sequenced genomes (Meuwissen and Goddard, 2010). When using sequenced genomes, biological information regarding SNP can also improve accuracies (Perez-Enciso, et al., 2015). Single nucleotide polymorphisms panels composed of functional markers could aid across breed predictions and aid adoption in beef industries.

The concept of the present work was to select functional SNP mapped to transcription factors (TF) associated with female reproduction. These TF regulate the expression of a set of genes associated with post-partum anoestrus and pregnancy outcomes (Fortes et al., 2014a). Initially, 140 SNP mapped to TF and not represented on SNP chips were used to predict pregnancy outcomes. Selected TF originated from genome-wide association studies (GWAS; Hawken

Received: August 27, 2017

Accepted: May 8, 2018

*Corresponding author: m.fortes@uq.edu.au

Copyright © 2018 Sociedade Brasileira de Zootecnia. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

et al., 2012), microarray data (Fortes et al., 2014b), and RNA sequencing (Canovas et al., 2014). Genotypes for putative functional mutations ($n = 140$) were imputed in an independent crossbred population of 1,988 cows. These *in silico* genotypes were used in genomic predictions and accounted for 29% of additive variation in rebreeding outcome (Fortes et al., 2014a). The current study sought to provide a validation for the *in silico* trial by genotyping selected SNP.

Nonlinear genomic prediction methods that emphasize the discovery and use of functional mutations increased the accuracy of genomic prediction for dairy cattle (Kemper et al., 2015). We hypothesised that combining additional functional mutations and nonlinear methodologies could result in more accurate genomic predictions.

Material and Methods

Customized TaqMan[®] assays were designed to genotype 33 SNP in two populations of female cattle. TaqMan[®] genotyping was performed following manufacturer's instructions. The customized design of primers and probes for genotyping was carried out by uploading SNP flanking regions to the ThermoFisher appropriate online tool: Custom TaqMan[®] Assay Design Tool (<https://www.thermofisher.com/order/custom-genomic-products/tools/gene-expression/>). The selected SNP were mapped to the candidate genes *PBX1* (chromosome 3, at 4 Mb), *POU6F1* (chromosome 5, at 28 Mb), *SOX5* (chromosome 5, at 86 Mb), *PPARA* (chromosome 5, at 117 Mb), *IRF1* (chromosome 7, at 23 Mb), *FOXO3* (chromosome 9, at 42 Mb), *TBP* (chromosome 9, at 106 Mb), *LHX3* (chromosome 11, at 104 Mb), *FOXA2* (chromosome 13, at 41 Mb), *PLXNA2* (chromosome 16, at 77 Mb), *HNF1A* (chromosome 17, at 65 Mb), *TRAF4* (chromosome 19, at 20 Mb), *BPTF* (chromosome 19, at 49 Mb), *CPEB1* (chromosome 21, at 23 Mb), *CDC5L* (chromosome 23, at 17 Mb), *CUX1* (chromosome 25, at 35 Mb), and *IRF2* (chromosome 27, at 13 Mb). These SNP were selected from the 140 SNP previously described by Fortes et al. (2014a), because examination of sequenced genomes of representative sires showed that they were present in our validation populations. Sequenced genomes from Beef CRC sires were described before (Camargo et al., 2015) and were submitted to the 1000-bull genome project (Hayes et al., 2014).

The female cattle populations used in this study were from two breeds: Brahman and Tropical Composites. These populations and measured phenotypes have been previously described (Hawken et al., 2012; Johnston et al., 2009; 2010; 2013). Brahman heifers are considered pure

Bos indicus, while Tropical Composites are a *Bos indicus*-*Bos taurus* mixture breed.

We investigated three female reproductive traits measured early in life: age at first *corpus luteum* (AGECL; days), post-partum anoestrous interval (PPAI; days), and post-partum anoestrous interval with respect to weaning (PW; binary). The two traits relevant for post-partum anoestrous interval were measured after the first breeding season and reflect the outcomes of the first mating season. These heifers were joined in multiple sire mating systems, as two-year olds as per normal Australian beef industry practices. All phenotypes were measured using ultrasound imaging by experienced technicians and veterinarians. Ultrasound was used every other week to observe the first *corpus luteum* and annotate the age of the heifer on that occurrence (AGECL). After the first parturition, cows were scanned again to observe the first *corpus luteum* post-partum. The interval in days from parturition to first ovulation post-partum (calculated from the *corpus luteum* observation) was used to record PPAI and PW. The binary phenotype PW was a record of 1, if the cow ovulated prior to weaning, and 0, if the cow did not (it ovulated only after weaning) as previously described (Zhang et al., 2014). The ability to ovulate before weaning the calf is considered important because it shortens the interval between parturitions and is a signal that the cow has overcome lactational anoestrus. Lactational anoestrus is a major cause of prolonged post-partum anoestrus in *Bos indicus* herds.

A total of 1,221 Brahman cows had DNA available for genotyping, and all were used in this study. However, only 914 Brahman cows were measured for AGECL. Of these cows, 576 conceived in the first mating opportunity and, therefore, had the post-partum phenotypes PPAI and PW measured. A total of 895 Tropical Composite cows were genotyped for this project. Among the genotyped cows, 798 had AGECL measured, while 665 had PPAI and 668 had PW measurements. The calculation of the “zebu content”, or *Bos indicus* genetics, for both populations has been previously described (Porto-Neto et al., 2014). These are two independent populations that represent different breeds, with Brahman having much higher contents of *Bos indicus* genetics. The test of association was carried out one SNP at a time and separately per breed first, resulting in P-values and estimated SNP effects for Brahman and Tropical Composite cows. In a second analysis, data from both breeds were combined to estimate the SNP association across breeds. Association analyses were carried using the SNP & Variation Suite software from Golden Helix[™] (v7.6.8 Win64; Golden Helix, Bozeman, MT, USA <http://www.goldenhelix.com>). We used mixed models to fit

fixed effects of contemporary group and random effects of animal, SNP, and residual effects according to the following equation:

$$y = X\beta + Z_a u_a + s + e,$$

in which y is the vector of observations (AGECL, PPAI, or PW); β is the vector of fixed effects related to the observations by the incidence matrix X ; u_a is the vector of random additive genetic effects related to the observations by the incidence matrix Z_a ; s represents the SNP effect; and e is the vector of random residual effects. Contemporary groups (or cohorts) were formed by animals born in the same year and managed as a group in one of four research stations. Cohort had a significant effect for all studied traits as previously discussed (Barwick et al., 2009; Johnston et al., 2009).

In this mixed model, we used a genomic relationship matrix (GRM) built from high-density SNP chip genotype data. All animals had high-density SNP genotypes available, either directly genotyped or imputed from lower density genotypes. Cows were genotyped using the Illumina Bovine SNP50 BeadChip version 1. All SNP chips were assayed following the manufacturer's protocols, and repeated samples were included in the genotyping exercise for quality assurance. The BEAD STUDIO software (Illumina, Inc.) was used to determine genotype calls. Imputation was performed using a reference set of 917 animals genotyped with the high-density BovineHD chip. The imputation was performed using BEAGLE (Browning and Browning, 2009). Methods, number of animals used, and accuracy of imputation were described by Bolormaa et al. (2013).

The GRM was generated using the SNP & Variation Suite software from Golden Helix™, which implements the methods described by VanRaden (2008). Because we used a GRM for all our models and not pedigree matrices, we refer to the estimated heritabilities as “pseudo-heritability”. We used 3 GRM in this study, one built for Brahman only, another for Tropical Composites, and a third combining the Brahman and Tropical Composite genotypes in one GRM for across-breed analyses. When combining the two breeds, the method described by VanRaden (2008) estimated allele frequencies for one base population. Fixed effects used in the models considered effects of contemporary group within breed, and when breeds were combined in one analysis, the fixed effect of “breed” was also fitted in the equation. For all analyses, an interaction among cohort, farm of origin, and birth month was fitted as a fixed effect in the model. These fixed effects were deemed significant in previous analyses of this data (Hawken et al., 2012).

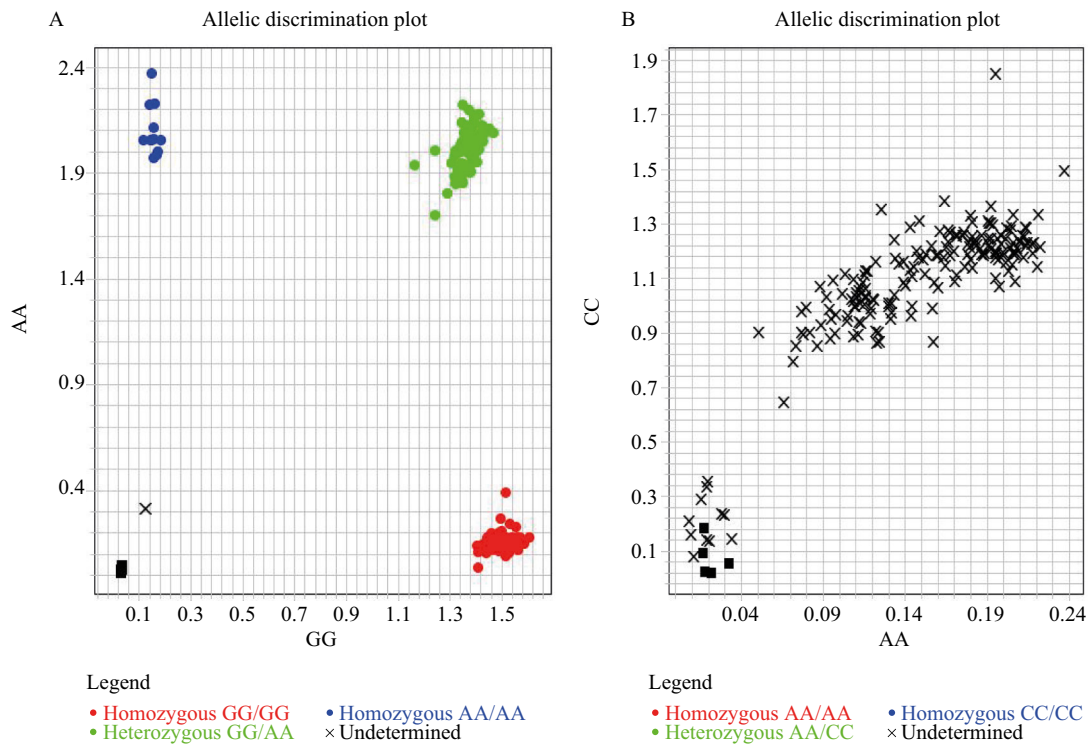
Two genomic prediction methods were evaluated: BayesR (a nonlinear method) and GBLUP (genomic best linear unbiased prediction). These methods have been compared in the analyses of dairy cattle data (Erbe et al., 2012; Kemper et al., 2015). In short, BayesR assumes that SNP effects are drawn from a mixture of normal distributions, with increasing variance, allowing moderate to large effects, while GBLUP assumes all SNP effects are derived from the same normal distribution, resulting in very small SNP effects for all SNP. The assumption of GBLUP is that all SNP have small effects, which are normally distributed in a classical infinitesimal model. The mixture of distributions allowed in BayesR accounts for higher variability of SNP effects, which corresponds to the hypothesis that some genes may have no association with the investigated phenotypes and other genes may have a moderate effect. The variance of SNP effects had four possible values: 0, $0.0001\sigma_g^2$, $0.001\sigma_g^2$, and $0.01\sigma_g^2$, in which σ_g^2 is the total additive genetic variance. A Gibbs sampling scheme was applied similar to that described by Erbe et al. (2012). The four possible values for SNP variance fit the rationale that the three reproductive traits under investigation are complex traits, with the majority of associated SNP having small effects. This genetic architecture was reported in an earlier GWAS (Hawken et al., 2012).

Accuracies of genome predictions were computed as a correlation between the estimated GEBV and the actual phenotypes divided by the square root of the heritability of the trait. Improvement of accuracies were considered significant if greater than one standard error in absolute terms. To calculate the accuracies of genome predictions, the dataset that included both breeds was randomly split into thirds to serve as “reference” and “validation” data. Then, three-fold cross validation was used to estimate the prediction accuracies.

Results

Four TaqMan® assays failed to produce useful genotypes. Example results for an efficient assay (Figure 1 A) and a failed assay (Figure 1 B) are provided. Results from the remaining 29 SNP assays were used in subsequent analyses.

In Brahman cows, the call rate for one SNP was very low (18%) and was, therefore, excluded from further analyses. Call rates for included SNP were 94.76% or higher. The lowest minor allele frequency (MAF) had a value of 0.43%, and only eight Brahman animals were heterozygous for this SNP. This near fixation of the major allele prevents the use



The fluorescence intensity for the VIC and FAM probes is plotted on the x and y axes. Note how the successful SNP plot has three clear clusters, one for each genotype group in A, while there are no clearly separated clusters, and alleles cannot be discriminated in B.

Figure 1 - Example of allelic discrimination plots for successful SNP (A) and failed SNP (B).

of this SNP in association and genomic prediction analyses. Both the SNP with the poor call rate and the SNP with the low MAF were excluded from further analyses in Brahman cows.

In Tropical Composite cows, allele frequencies were different from Brahman cows. For example, the MAF observed for a SNP mapped to *CUX1* in Brahman cows was 11.63%, while in Tropical Composites, the same minor allele was even less frequent (MAF = 1.54%). For ten SNP, the minor allele in one breed was actually the major allele in the other breed; these were rs464658757 mapped to *PPARA*, rs136542281 and rs136630122 mapped to *IRF1*, rs381517561 mapped to *LHX3*, rs208587911 mapped to *FOXA2*, rs132677230 in *PLXNA2*, rs717464147 and rs209560708 located in *TRAF4*, rs439699867 and rs377982205 in *BPTF*, and rs457608533 in *CDC5L*. In Tropical Composites, the lowest MAF was 1.54 %, and the lowest call rate was 83%; and so, we were able to use every SNP in the analyses for this breed.

In Brahman cattle, AGECL had an estimated pseudo-heritability of 0.57 with a variance of 3.13×10^{-5} and a standard error of 5.59×10^{-3} . The proportion of the genetic variance explained by fixed covariates was 0.35. Three SNP were associated with AGECL, two mapped to the gene *IRF1* and one mapped to *CPEB1* ($P < 0.05$) (Table 1).

The analysis of 576 Brahman cows measured for PPAI resulted in an estimated pseudo-heritability of 0.43 with a variance of 2.40×10^{-5} and a standard error of 4.90×10^{-3} . The proportion of the genetic variance explained by fixed covariates was 0.20. For PPAI in Brahman, associated SNP ($P < 0.05$) were mapped to the genes *IRF1*, *BPTF*, and *IRF2* (Table 1).

Not surprisingly, the results for PW are similar to those obtained for PPAI, as these are correlated traits. The pseudo-heritability for PW was 0.56, and the variance was 1.63, while the standard error was 1.27. The proportion of variance explained by fixed covariates was 0.23. Two SNP were significant for PW, both mapped to *IRF1* ($P < 0.05$) and explaining between 0.82 and 1.20% of the genetic variance (Table 1). In Brahman, mutations in *IRF1* were associated with all studied traits.

In Tropical Composite cattle, AGECL had an estimated pseudo-heritability of 0.37, with a variance of 1.43×10^{-5} and a standard error of 3.78×10^{-3} . The proportion of the variance explained by fixed covariates was 0.09. The number of cows with phenotype records and genotypes was 798. In Tropical Composite cattle, 665 cows had genotypes and records for PPAI. The estimated pseudo-heritability for PPAI was 0.35 with a variance of 1.20×10^{-5} and a standard error of 3.48×10^{-3} . The proportion of variance explained by

fixed covariates was 0.09. For PW, the estimated pseudo-heritability was 0.30, with a variance of 0.75 and a standard error of 0.87. The proportion of the variance explained by fixed covariates was 0.08. In total, 668 Tropical Composite cows had genotypes and phenotypes for the PW analyses. Association results were significant for IRF1, with two SNP associated to PW (Table 2). Single nucleotide polymorphisms in *BPTF* and *IRF2* were associated with AGECL; *BPTF* was also significant for PPAI (Table 2).

Association results for each SNP in the analyses that considered the two breeds together are presented for AGECL, PPAI, and PW (Table 3). We present SNP associations with $P < 0.10$. When both breeds were analysed together, AGECL had an estimated pseudo-heritability of 0.59, with a variance of 3.58×10^{-5} and a standard error of 5.98×10^{-3} . The proportion of the variance explained by fixed covariates was 0.38.

In a previous study, these same cows were genotyped for a SNP mapped to the *PLAG1* gene: rs109231213 (Fortes et al., 2013). This *PLAG1* SNP was significantly associated with AGECL in the mentioned study. We included the *PLAG1* SNP genotype in the genomic prediction analyses. The genomic prediction method BayesR gives better results, especially when the new tested mutations that were mapped

to the 17 candidate genes targeted in this study and the *PLAG1* SNP are included in the model. The GBLUP models resulted in accuracies that were lower by one standard deviation (data not shown). The proportion of the variance explained by the 29 SNP together was 0.07 for AGECL, 0.03 for PPAI, and 0.02 for PW. The best predictions are from the combined new panel, with the BayesR analysis, with a 4% improvement in accuracy (Table 4).

Discussion

Four TaqMan® assays failed to produce meaningful results. TaqMan® is a robust methodology, but it may fail if there are other mutations nearby and/or if the population only has one of the tested alleles.

In a previous study, these same cows were genotyped for a SNP mapped to the *PLAG1* gene: rs109231213 (Fortes et al., 2013). The *PLAG1* SNP was mapped to a QTL on chromosome 14 that was associated with many traits of economic interest for beef and dairy cattle breeding. The SNP in this QTL was associated with height, weight, serum concentration of IGF1, and age at puberty (Fortes et al., 2013; Juma et al., 2016; Karim et al., 2011; Littlejohn et al., 2012; Nishimura et al., 2012; Pryce et al., 2012;

Table 1 - Brahman results: SNP association analyses

Trait	Gene (SNP)	P-value	Effect	Effect standard error	Proportion of variance explained (%)
AGECL	<i>PPARA</i> (rs464658757)	0.07	-11.43	6.39	0.37
AGECL	<i>IRF1</i> (rs134204419)	0.04	-11.95	5.90	0.47
AGECL	<i>IRF1</i> (rs135724080)	0.03	-12.86	5.93	0.55
AGECL	<i>CPEB1</i> (rs482358416)	0.01	-18.56	7.12	0.79
PPAI	<i>SOX5</i> (rs42615751)	0.05	-38.37	19.47	0.74
PPAI	<i>IRF1</i> (rs136542281)	0.04	14.22	6.99	0.79
PPAI	<i>IRF1</i> (rs136630122)	0.02	16.20	6.95	1.03
PPAI	<i>BPTF</i> (rs377982205)	0.03	13.42	6.20	0.89
PPAI	<i>IRF2</i> (rs109901025)	0.01	-23.06	8.67	1.34
PW	<i>IRF1</i> (rs136542281)	0.04	-0.07	0.03	0.82
PW	<i>IRF1</i> (rs136630122)	0.01	-0.08	0.03	1.20
PW	<i>IRF2</i> (rs109901025)	0.08	0.07	0.04	0.57

SNP - single nucleotide polymorphisms; AGECL - age at first *corpus luteum*; PPAI - post-partum anoestrous interval; PW - post-partum anoestrous interval with respect to weaning.

Table 2 - Tropical Composites results: SNP association analyses

Trait	Gene (SNP)	P-value	Effect	Effect standard error	Proportion of variance explained (%)
AGECL	<i>BPTF</i> (rs377982205)	0.02	-16.85	7.49	0.63
AGECL	<i>IRF2</i> (rs109901025)	0.06	12.41	6.53	0.45
PPAI	<i>PLXNA2</i> (rs42599164)	0.07	-12.59	7.02	0.48
PPAI	<i>BPTF</i> (rs439699867)	0.06	14.87	7.88	0.54
PW	<i>IRF1</i> (rs134204419)	0.04	-0.06	0.03	0.65
PW	<i>IRF1</i> (rs135724080)	0.02	-0.07	0.03	0.78
PW	<i>BPTF</i> (rs439699867)	0.02	-0.06	0.03	0.76

SNP - single nucleotide polymorphisms; AGECL - age at first *corpus luteum*; PPAI - post-partum anoestrous interval; PW - post-partum anoestrous interval with respect to weaning.

Table 3 - Brahman and Tropical Composites results for SNP associations

Trait	Gene (SNP)	P-value	Effect	Effect standard error	Proportion of variance explained (%)
AGECL	<i>IRF1</i> (rs134204419)	0.02	-11.41	4.72	0.37
AGECL	<i>IRF1</i> (rs135724080)	0.01	-11.77	4.74	0.39
AGECL	<i>IRF2</i> (rs109901025)	0.09	8.37	4.98	0.18
PPAI	<i>FOXA2</i> (rs208587911)	0.04	10.61	5.24	0.37
PPAI	<i>TRAF4</i> (rs208903687)	0.03	12.06	5.45	0.44
PPAI	<i>IRF2</i> (rs109901025)	0.03	-11.61	5.18	0.46
PW	<i>SOX5</i> (rs42615751)	0.07	0.05	0.03	0.30
PW	<i>FOXA2</i> (rs208587911)	0.06	-0.04	0.02	0.32

SNP - single nucleotide polymorphisms; AGECL - age at first *corpus luteum*; PPAI - post-partum anoestrous interval; PW - post-partum anoestrous interval with respect to weaning.

Table 4 - Accuracy of genomic predictions

	r(GEBV,y*) ¹	Accuracy ²	Standard error
BAYESR 29 SNP + <i>PLAG1</i>			
AGECL	0.16	0.21**	0.02
PPAI	0.03	0.04	0.04
PW	0.05	0.07	0.03
BAYESR 50K SNP			
AGECL	0.28	0.37	0.03
PPAI	0.24	0.36	0.03
PW	0.2	0.27	0.03
BAYESR 50K + 29 SNP			
AGECL	0.31	0.41**	0.02
PPAI	0.24	0.36	0.02
PW	0.2	0.27	0.03

SNP - single nucleotide polymorphisms; AGECL - age at first *corpus luteum*; PPAI - post-partum anoestrous interval; PW - post-partum anoestrous interval with respect to weaning.

¹ Correlation of genomic prediction and phenotype corrected for fixed effects.

² Accuracy of genomic prediction, calculated as $r(\text{GEBV}, y^*) / \sqrt{h^2}$.

** Significant improvement in the accuracy of predictions.

Saatchi et al., 2014; Utsunomiya et al., 2013). The *PLAG1* mutation rs109231213 was included in genomic predictions described in this study.

Association analyses were carried in both breeds. In Brahman cattle, AGECL-estimated pseudo-heritability was in agreement with previous estimates of AGECL heritability in this Brahman population (Johnston et al., 2009). The effect predicted for associated SNP was between 11 and 19 days, which means that heifer puberty measured as AGECL could occur earlier in animals carrying the favourable variant.

The pseudo-heritability for PPAI was slightly lower than previous heritability estimates for the same population, which were between 0.51 and 0.52 (Johnston et al., 2010; Zhang et al., 2014). The difference in these heritability estimates could be due to the use of a pedigree relationship matrix in previous studies, while we used a GRM. More animals with phenotypes and pedigree were available, but not all those animals were genotyped. Because of genotype availability, about 100 cows from the previous studies were not included in our genomic analyses.

Single nucleotide polymorphisms with associations to PPAI were mapped to genes *SOX5*, *IRF1*, *BPTF*, and *IRF2*. The absolute effects of associated SNP were estimated between 13 and 39 days. This means that PPAI could be substantially reduced by selecting for the favourable alleles. The results for PW are similar to those obtained for PPAI, which is expected since these are correlated traits; one measured in days and another is binary simplification of the phenotype to indicate the ability of cows to ovulate prior to weaning. A higher heritability of PW (referred to as PPO), as compared to PPAI, was also observed in previous analyses of this population in which PW heritability was estimated from pedigree-based relationships as 0.62 (Zhang et al., 2014).

In Tropical Composite cattle, AGECL had an estimated pseudo-heritability of 0.37. This result is lower than that published before for the same population (Johnston et al., 2009). Again, herein we used a GRM and fewer animals than the study by Johnston and colleagues, which could explain the difference in results.

From single SNP analyses in Tropical Composites, genes *BPTF* and *IRF1* presented the most significant associations with these early-in-life indicators of female fertility. Gene *IRF1* emerges as a candidate for harbouring mutations associated to reproductive traits measured early in life, in both Brahman and Tropical Composite cows.

Gene *IRF1* (Interferon regulatory factor 1) is a transcription factor, which, in mammals, activates the expression of the cytokine interferon beta (Miyamoto et al., 1988). *IRF1* was subsequently shown to be a transcriptional activator and/or repressor of many target genes (a key regulator). *IRF1* regulates expression of target genes by binding to the interferon-stimulated response element. Interferon activity is known to regulate reproductive biology, more specifically early pregnancy stages (Mathew et al., 2016; Wiltbank et al., 2016). *IRF1* has also been shown to play a role in regulating post-translational modifications to tumour suppressor protein p53 (Dornan et al., 2004).

IRF1 is involved in the regulation of immune response, apoptosis, DNA damage, and tumour suppression; its link to reproductive ability has also been shown (Lim et al., 2016).

The genomic prediction method BayesR is a hierarchical method, which allows important DNA variants to be prioritised. The proportion of the variance explained by the 29 SNP together was 0.07 for AGECL, 0.03 for PPAI, and 0.02 for PW. These variances are higher than expected, given the low number of SNP in the tested panel. We tested the accuracy of genomic predictions using the selected SNP alone, the 50K chip SNP alone, and then all SNP together. The best predictions were from the combined new panel, with the BayesR analysis, with a 4% improvement in accuracy. This improvement in accuracy is comparable to the results obtained when BayesR was used in dairy breeds. In a combined dataset of Holstein and Jersey, BayesR also outperformed GBLUP models, and the increase in accuracies ranged from 5 to 15% for different milk traits (Erbe et al., 2012). BayesR improvement over GBLUP models is not a surprising result, as summarised and explained by Su et al. (2014). The significant improvement observed for AGECL might reflect the fact that a major QTL, namely the *PLAG1* region, is present for this trait. Bayesian models, such as BayesR, are reported to benefit more traits influenced by QTL of large effect (Legarra et al., 2011).

Simulation studies with full sequence data suggest that selecting variants that are close to (or are known) causative mutations can increase genomic prediction accuracies (Perez-Enciso et al., 2015; van den Berg et al., 2016). Our results agree with these simulation studies. In the foreseeable future, in which genomes from over a thousand bulls will be available (Hayes et al., 2014), understanding biology and prioritizing functional mutations will continue to be a relevant research effort.

In Australia, the current tool for female fertility selection is the use of GBLUP models to predict “days to calving”. The accuracy of predicting this female fertility trait might also be improved if the tested SNP were included in a genomic prediction framework, particularly if a BayesR analysis was used. Future work in this area could be promising, especially if additional putative functional SNP are discovered and incorporated in future versions of cattle SNP chips.

Conclusions

Mutations in five genes were found to be significantly associated with reproduction traits. The gene *IRF1* had significant SNP associations in both breeds and, therefore,

it emerges as a candidate for female reproductive traits measured early in life.

The genomic prediction results support the possibility that the described method to prioritise gene candidates to be incorporated in future SNP chips is worth exploring. The proposed SNP panel with 29 mutations results in reasonable accuracies of genomic prediction for *AGECL* on its own. However, it works even better when combined with the 50K chip data. The accuracy is improved for all traits when the 29 SNP are used together with 50K chip single nucleotide polymorphisms. The predictions were carried for Brahman and Tropical Composites. This approach is likely to serve all breeds with *Bos indicus* content.

Acknowledgments

The authors acknowledge Zoetis Inc. for the financial support regarding genotyping of SNP selected for this project.

References

- Barkema, H. W.; von Keyserlingk, M. A. G.; Kastelic, J. P.; Lam, T.; Luby, C.; Roy, J. P.; LeBlanc, S. J.; Keefe, G. P. and Kelton, D. F. 2015. Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. *Journal of Dairy Science* 98:7426-7445. <https://doi.org/10.3168/jds.2015-9377>
- Barwick, S. A.; Johnston, D. J.; Burrow, H. M.; Holroyd, R. G.; Fordyce, G.; Wolcott, M. L.; Sim, W. D. and Sullivan, M. T. 2009. Genetics of heifer performance in ‘wet’ and ‘dry’ seasons and their relationships with steer performance in two tropical beef genotypes. *Animal Production Science* 49:367-382. <https://doi.org/10.1071/EA08273>
- Boichard, D.; Ducrocq, V.; Croiseau, P. and Fritz, S. 2016. Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies* 339:274-277. <https://doi.org/10.1016/j.crv.2016.04.007>
- Bolormaa, S.; Pryce, J. E.; Kemper, K.; Savin, K.; Hayes, B. J.; Barendse, W.; Zhang, Y.; Reich, C. M.; Mason, B. A.; Bunch, R. J.; Harrison, B. E.; Reverter, A.; Herd, R. M.; Tier, B.; Graser, H. U. and Goddard, M. E. 2013. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. *Journal of Animal Science* 91:3088-104. <https://doi.org/10.2527/jas.2012-5827>
- Browning, B. L. and Browning, S. R. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics* 84:210-23. <https://doi.org/10.1016/j.ajhg.2009.01.005>
- Camargo, G. M. F.; Porto-Neto, L. R.; Kelly, M. J.; Bunch, R. J.; McWilliam, S. M.; Tonhati, H.; Lehnert, S. A.; Fortes, M. R. S. and Moore, S. S. 2015. Non-synonymous mutations mapped to chromosome x associated with andrological and growth traits in beef cattle. *BMC Genomics* 16:384. <https://doi.org/10.1186/s12864-015-1595-0>
- Canovas, A.; Reverter, A.; DeAtley, K. L.; Ashley, R. L.; Colgrave, M. L.; Fortes, M. R. S.; Islas-Trejo, A.; Lehnert, S.; Porto-Neto, L.; Rincon, G.; Silver, G. A.; Snelling, W. M.; Medrano, J. F. and Thomas, M. G. 2014. Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PLoS ONE* 9:e102551. <https://doi.org/10.1371/journal.pone.0102551>

- Dornan, D.; Eckert, M.; Wallace, M.; Shimizu, H.; Ramsay, E.; Hupp, T. R. and Ball, K. L. 2004. Interferon regulatory factor 1 binding to p300 stimulates DNA-dependent acetylation of p53. *Molecular and Cellular Biology* 24:10083-10098. <https://doi.org/10.1128/MCB.24.22.10083-10098.2004>
- Erbe, M.; Hayes, B. J.; Matukumalli, L. K.; Goswami, S.; Bowman, P. J.; Reich, C. M.; Mason, B. A. and Goddard, M. E. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science* 95:4114-4129. <https://doi.org/10.3168/jds.2011-5019>
- Fortes, M. R. S.; Kemper, K.; Sasazaki, S.; Reverter, A.; Pryce, J. E.; Barendse, W.; Bunch, R.; McCulloch, R.; Harrison, B.; Bolormaa, S.; Zhang, Y. D.; Hawken, R. J.; Goddard, M. E. and Lehnert, S. A. 2013. Evidence for pleiotropism and recent selection in the PLAG1 region in Australian beef cattle. *Animal Genetics* 44:636-647. <https://doi.org/10.1111/age.12075>
- Fortes, M. R. S.; Porto-Neto, L. R.; DeAtley, K. L.; Reverter, A.; Thomas, M. G.; Moore, S. S.; Lehnert, S. A. and Snelling, W. M. 2014a. Genetic markers in transcription factors of differentially expressed genes associated with post-partum anoestrus predict pregnancy outcome in an independent population of beef cattle. In: *World Congress of Genetics Applied to Livestock Production*, Vancouver, Canada.
- Fortes, M. R. S.; Suhaimi, A. H. M. S.; Porto-Neto, L. R.; McWilliam, S. M.; Flatscher-Bader, T.; Moore, S. S.; D'Occhio, M. J.; Meira, C. T.; Thomas, M. G.; Snelling, W. M.; Reverter, A. and Lehnert, S. A. 2014b. Post-partum anoestrus in tropical beef cattle: A systems approach combining gene expression and genome-wide association results. *Livestock Science* 166:158-166. <https://doi.org/10.1016/j.livsci.2014.06.017>
- Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. *Genetics Selection Evolution* 43:17. <https://doi.org/10.1186/1297-9686-43-17>
- Hawken, R. J.; Zhang, Y. D.; Fortes, M. R. S.; Collis, E.; Barris, W. C.; Corbet, N. J.; Williams, P. J.; Fordyce, G.; Holroyd, R. G.; Walkley, J. R. W.; Barendse, W.; Johnston, D. J.; Prayaga, K. C.; Tier, B.; Reverter, A. and Lehnert, S. A. 2012. Genome-wide association studies of female reproduction in tropically adapted beef cattle. *Journal of Animal Science* 90:1398-1410. <https://doi.org/10.2527/jas.2011-4410>
- Hayes, B. J.; MacLeod, I. M.; Daetwyler, H. D.; Bowman, P. J.; Chamberlain, A. J.; Vander Jagt, C. J.; Capitan, A.; Pausch, H.; Stothard, P.; Liao, X.; Schrooten, C.; Mullaart, E.; Fries, R.; Guldbandsen, B.; Lund, M. S.; Boichard, D. A.; Veerkamp, R. F.; VanTassell, C. P.; Gredler, B.; Druet, T.; Bagnato, A.; Vilkki, J.; Koning, D. J.; Santus, E. and Goddard, M. E. 2014. Genomic prediction from whole genome sequence in livestock: The 1000 bull genomes project. In: *10th World Congress of Genetics Applied to Livestock Production*, Vancouver, Canada.
- Johnston, D. J.; Barwick, S. A.; Corbet, N. J.; Fordyce, G.; Holroyd, R. G.; Williams, P. J. and Burrow, H. M. 2009. Genetics of heifer puberty in two tropical beef genotypes in northern Australia and associations with heifer- and steer-production traits. *Animal Production Science* 49:399-412. <https://doi.org/10.1071/EA08276>
- Johnston, D. J.; Barwick, S. A.; Fordyce, G. and Holroyd, R. G. 2010. Understanding the genetics of lactation anoestrus in Brahman beef cattle to enhance genetic evaluation of female reproductive traits. In: *9th WCGALP*, Leipzig, Germany.
- Johnston, D. J.; Barwick, S. A.; Fordyce, G.; Holroyd, R. G.; Williams, P. J.; Corbet, N. J. and Grant, T. 2013. Genetics of early and lifetime annual reproductive performance in cows of two tropical beef genotypes in northern Australia. *Animal Production Science* 54:1-15. <https://doi.org/10.1071/AN13043>
- Juma, A. R.; Damdimopoulou, P. E.; Grommen, S. V.; Van de Ven, W. J. and De Groef, B. 2016. Emerging role of PLAG1 as a regulator of growth and reproduction. *Journal of Endocrinology* 228:R45-56. <https://doi.org/10.1530/JOE-15-0449>
- Karim, L.; Takeda, H.; Lin, L.; Druet, T.; Arias, J. A. C.; Baurain, D.; Cambisano, N.; Davis, S. R.; Farnir, F.; Grisart, B.; Harris, B. L.; Keehan, M. D.; Littlejohn, M. D.; Spelman, R. J.; Georges, M. and Coppieters, W. 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics* 43:405-413. <https://doi.org/10.1038/ng.814>
- Kemper, K. E.; Reich, C. M.; Bowman, P. J.; Vander Jagt, C. J.; Chamberlain, A. J.; Mason, B. A.; Hayes, B. J. and Goddard, M. E. 2015. Improved precision of QTL mapping using a nonlinear bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution* 47:29. <https://doi.org/10.1186/s12711-014-0074-4>
- Legarra, A.; Robert-Granic, C.; Croiseau, P.; Guillaume, F. and Fritz, S. 2011. Improved lasso for genomic selection. *Genetics Research* 93:77-87. <https://doi.org/10.1017/S0016672310000534>
- Lim, R.; Tran, H. T.; Liang, S.; Barker, G. and Lappas, M. 2016. The transcription factor interferon regulatory factor-1 (IRF1) plays a key role in the terminal effector pathways of human preterm labor. *Biology of Reproduction* 94:32, 1-13. <https://doi.org/10.1095/biolreprod.115.134726>
- Littlejohn, M.; Grala, T.; Sanders, K.; Walker, C.; Waghorn, G.; Macdonald, K.; Coppieters, W.; Georges, M.; Spelman, R.; Hillerton, E.; Davis, S. and Snell, R. 2012. Genetic variation in PLAG1 associates with early life body weight and periparturient weight and growth in *Bos taurus*. *Animal Genetics* 43:591-594. <https://doi.org/10.1111/j.1365-2052.2011.02293.x>
- Mathew, D. J.; Lucy, M. C. and Geisert, R. D. 2016. Interleukins, interferons, and establishment of pregnancy in pigs. *Reproduction* 151:R111:R122. <https://doi.org/10.1530/REP-16-0047>
- Meuwissen, T. and Goddard, M. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185:623-U338. <https://doi.org/10.1534/genetics.110.116590>
- Meuwissen, T. H. E.; Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Miyamoto, M.; Fujita, T.; Kimura, Y.; Maruyama, M.; Harada, H.; Sudo, Y.; Miyata, T. and Taniguchi, T. 1988. Regulated expression of a gene encoding a nuclear factor, IRF-1, that specifically binds to IFN- β gene regulatory elements. *Cell* 54:903-913. [https://doi.org/10.1016/S0092-8674\(88\)91307-4](https://doi.org/10.1016/S0092-8674(88)91307-4)
- Nishimura, S.; Watanabe, T.; Mizoshita, K.; Tatsuda, K.; Fujita, T.; Watanabe, N.; Sugimoto, Y. and Takasuga, A. 2012. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese black cattle. *BMC Genetics* 13:40. <https://doi.org/10.1186/1471-2156-13-40>
- Perez-Enciso, M.; Rincon, J. C. and Legarra, A. 2015. Sequence- vs. Chip-assisted genomic selection: Accurate biological information is advised. *Genetics Selection Evolution* 47:43. <https://doi.org/10.1186/s12711-015-0117-5>
- Porto-Neto, L. R.; Reverter, A.; Prayaga, K. C.; Chan, E. K.; Johnston, D. J.; Hawken, R. J.; Fordyce, G.; Garcia, J. F.; Sonstegard, T. S.; Bolormaa, S.; Goddard, M. E.; Burrow, H. M.; Henshall, J. M.; Lehnert, S. A. and Barendse, W. 2014. The genetic architecture of climatic adaptation of tropical cattle. *PLoS ONE* 9:e113284. <https://doi.org/10.1371/journal.pone.0113284>
- Pryce, J. E.; Arias, J.; Bowman, P. J.; Davis, S. R.; Macdonald, K. A.; Waghorn, G. C.; Wales, W. J.; Williams, Y. J.; Spelman, R. J. and Hayes, B. J. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. *Journal of Dairy Science* 95:2108-2119. <https://doi.org/10.3168/jds.2011-4628>

- Rolf, M. M.; Decker, J. E.; McKay, S. D.; Tizioto, P. C.; Branham, K. A.; Whitacre, L. K.; Hoff, J. L.; Regitano, L. C. A. and Taylor, J. F. 2014. Genomics in the United States beef industry. *Livestock Science* 166:84-93. <https://doi.org/10.1016/j.livsci.2014.06.005>
- Saatchi, M.; Schnabel, R. D.; Taylor, J. F. and Garrick, D. J. 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* 15:442. <https://doi.org/10.1186/1471-2164-15-442>
- Snelling, W. M.; Cushman, R. A.; Fortes, M. R. S.; Reverter, A.; Bennett, G. L.; Keele, J. W.; Kuehn, L. A.; McDanel, T. G.; Thallman, R. M. and Thomas, M. G. 2012. Physiology and endocrinology symposium: How single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. *Journal of Animal Science* 90:1152-65. <https://doi.org/10.2527/jas.2011-4581>
- Snelling, W. M.; Cushman, R. A.; Keele, J. W.; Maltecca, C.; Thomas, M. G.; Fortes, M. R. S. and Reverter, A. 2013. Breeding and genetics symposium: Networks and pathways to guide genomic selection. *Journal of Animal Science* 91:537-552. <https://doi.org/10.2527/jas.2012-5784>
- Su, G.; Christensen, O. F.; Janss, L. and Lund, M. S. 2014. Comparison of genomic predictions using genomic relationship matrices built with different weighting factors to account for locus-specific variances. *Journal of Dairy Science* 97:6547-6559. <https://doi.org/10.3168/jds.2014-8210>
- Utsunomiya, Y. T.; Carmo, A. S.; Carvalheiro, R.; Neves, H. H. R.; Matos, M. C.; Zavarez, L. B.; O'Brien, A. M. P.; Solkner, J.; McEwan, J. C.; Cole, J. B.; Van Tassell, C. P.; Schenkel, F. S.; Silva, M. V. D.; Porto Neto, L. R.; Sonstegard, T. S. and Garcia, J. F. 2013. Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC Genetics* 14:52. <https://doi.org/10.1186/1471-2156-14-52>
- van den Berg, I.; Boichard, D.; Guldbandsen, B. and Lund, M. S. 2016. Using sequence variants in linkage disequilibrium with causative mutations to improve across-breed prediction in dairy cattle: A simulation study. *G3-Genes Genomes Genetics* 6:2553-2561. <https://doi.org/10.1534/g3.116.027730>
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91:4414-4423. <https://doi.org/10.3168/jds.2007-0980>
- Wiltbank, M. C.; Baez, G. M.; Garcia-Guerra, A.; Toledo, M. Z.; Monteiro, P. L.; Melo, L. F.; Ochoa, J. C.; Santos, J. E. and Sartori, R. 2016. Pivotal periods for pregnancy loss during the first trimester of gestation in lactating dairy cows. *Theriogenology* 86:239-53. <https://doi.org/10.1016/j.theriogenology.2016.04.037>
- Zhang, Y. D.; Johnston, D. J.; Bolormaa, S.; Hawken, R. J. and Tier, B. 2014. Genomic selection for female reproduction in Australian tropically adapted beef cattle. *Animal Production Science* 54:16-24. <https://doi.org/10.1071/AN13016>