Correlation scan: identifying genomic regions that affect genetic correlations applied to fertility traits

Babatunde S. Olasege[1,2], Laercio R. Porto-Neto[2], Muhammad S. Tahir[1,2], Gabriela C. Gouveia[3], Angela Cánovas[5], Ben J. Hayes[4] and Marina R. S. Fortes[1,4]

[1]The University of Queensland, School of Chemistry and Molecular Biosciences, Saint Lucia Campus, Brisbane, QLD, 4072, Australia

[2]CSIRO Agriculture and Food, Saint Lucia, QLD, 4067, Australia

[3]Animal Science Department, Veterinary School, Federal University of Minas Gerais, Belo Horizonte 31270-901, Brazil

[4]The University of Queensland, Queensland Alliance for Agriculture and Food Innovation (QAAFI), Saint Lucia Campus, Brisbane, QLD, 4072, Australia

[5]University of Guelph, Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, 50 Stone Rd E, Guelph N1G 2W1, ONT, Canada

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

Abstract

Reproductive traits are often genetically correlated. Yet, we don't fully understand the complexities, synergism, or trade-offs between male and female fertility. Here, we introduce correlation scan, a novel framework for identifying the drivers or antagonizers of the genetic correlation between male and female fertility traits across the bovine genome. The identification of these regions facilitates the understanding of the complexity of these traits. Although the methodology was applied to cattle phenotypes, using high-density SNP genotypes, the general framework developed can be applied to any species or traits, and it can easily accommodate genome sequence data.

Background

In animal genetics, insight into the nature of the genetic relationships between quantitative traits are important because they improve our understanding of complex traits and diseases [1, 2]. These relationships termed genetic correlations manifest when there is shared genetic influence between traits (i.e., pleiotropy) [3, 4]. or when there is non-random association between loci (i.e., linkage disequilibrium (LD)) [5, 6]. Estimated genetic correlations provide information on how genome-wide genetic effects align between two complex traits [7].

Understanding the interplay between the genomic variants and their effects on quantitative traits can yield insights to improve the prediction of genetic merit and the understanding of complex traits' biology [8-10]. Estimated genetic correlations have informed animal and crop breeding for many decades. For example, scrotal circumference is used as an indicator trait in beef cattle breeding because it is genetically correlated with female fertility traits [11]. Nevertheless, we still have a limited information of the regions across genome regulating the intersexual correlations between male and female fertility traits. Investigating these regions and leveraging on the resulting biological information could inspire new approaches in livestock breeding [12, 13].

Over the past 100 years, different methods have been employed to estimate the genetic correlation between traits [14-17]. Traditionally, these correlations are estimated from pedigree data. However, genome-wide single nucleotide polymorphisms (SNPs) are often used in recent times [18]. It is possible to estimate across-sex correlation between traits and this research niche continues to attract interest among quantitative geneticists [19-21]. The resulting estimates from both within and across-sex analyses range from -1 to +1, indicating the strength and magnitude of the correlation between traits [22]. Despite more than a century of research on estimating this parameter, it is only very recently that studies attempt to identify the region(s) in the genome that underpin genetic correlations between traits [23-25].

In theory, we propose that various genomic regions will contribute to the overall genetic correlation between complex traits. Further, some regions will be driving the genetic correlation while others might antagonize it. For instance, if the genetic correlation between two traits is 0.70, some regions will yield a significant and positive correlation, say 0.90, while other regions may antagonize the overall estimate, and in that region the correlation could be -0.50. Also, some genomic regions may be neutral, say 0.02 and not significant for the correlation between the studied traits. Identifying driver and antagonizing regions are of

particular interest if they are for two important traits which are unfavourably correlated, for example milk yield and fertility in dairy cattle. Identification of such regions could lead to more targeted genomic selection and rapid genetic gains for both traits. Current genomic tools have created a great opportunity to advance our knowledge of genetic correlations between complex traits, by investigating the regions in the genome that drive or antagonize these correlations.

Here, we introduce a framework termed "correlation scan", which uses a sliding window methodology to uncover the genomic regions driving and antagonizing genetic correlations in beef cattle. We applied the method to male and female fertility traits and showcase how the outcomes of this methodology can be interpreted in downstream analyses to gain further insight about the studied traits and their relationships. Reproductive traits are often genetically correlated, and yet we don't fully understand the complexities, synergism, or trade-offs between male and female fertility. To demonstrate the method, we used two pairs of reproductive traits with strong genetic correlations in two independent cattle populations from our previous study [26]. These traits are age at first *corpus luteum* (AGECL, i.e., female puberty) and serum levels of insulin growth hormone (IGF1 measured in bulls, IGF1b, or cows, IGF1c). These pairs of traits serve as example of a positive and a negative correlation between phenotypes measured in males and females, during pubertal development. The populations used in the study are formed by either Brahman (BB) cattle or Tropical Composite (TC) cattle, as described in our previous study [26].

Results

The total number of windows generated and analysed

Using the framework developed in our study (see Materials and Methods), genomic windows with their corresponding correlation estimates (**r**) for each pairwise trait in two beef cattle

populations were identified. The total number of windows generated for all pairwise traits in BB was 5,558 and the number in TC was 6,876. For all windows, the chromosome coordinates and the corresponding **r** estimates in each of the two populations are presented in Additional file 1 (Table S1-S2). The **r** estimates for all windows were plotted against their genomic position (i.e., midpoint between the start and end position of each window) (Figure 1). Results are presented separately per cattle population and for each pair of traits investigated.

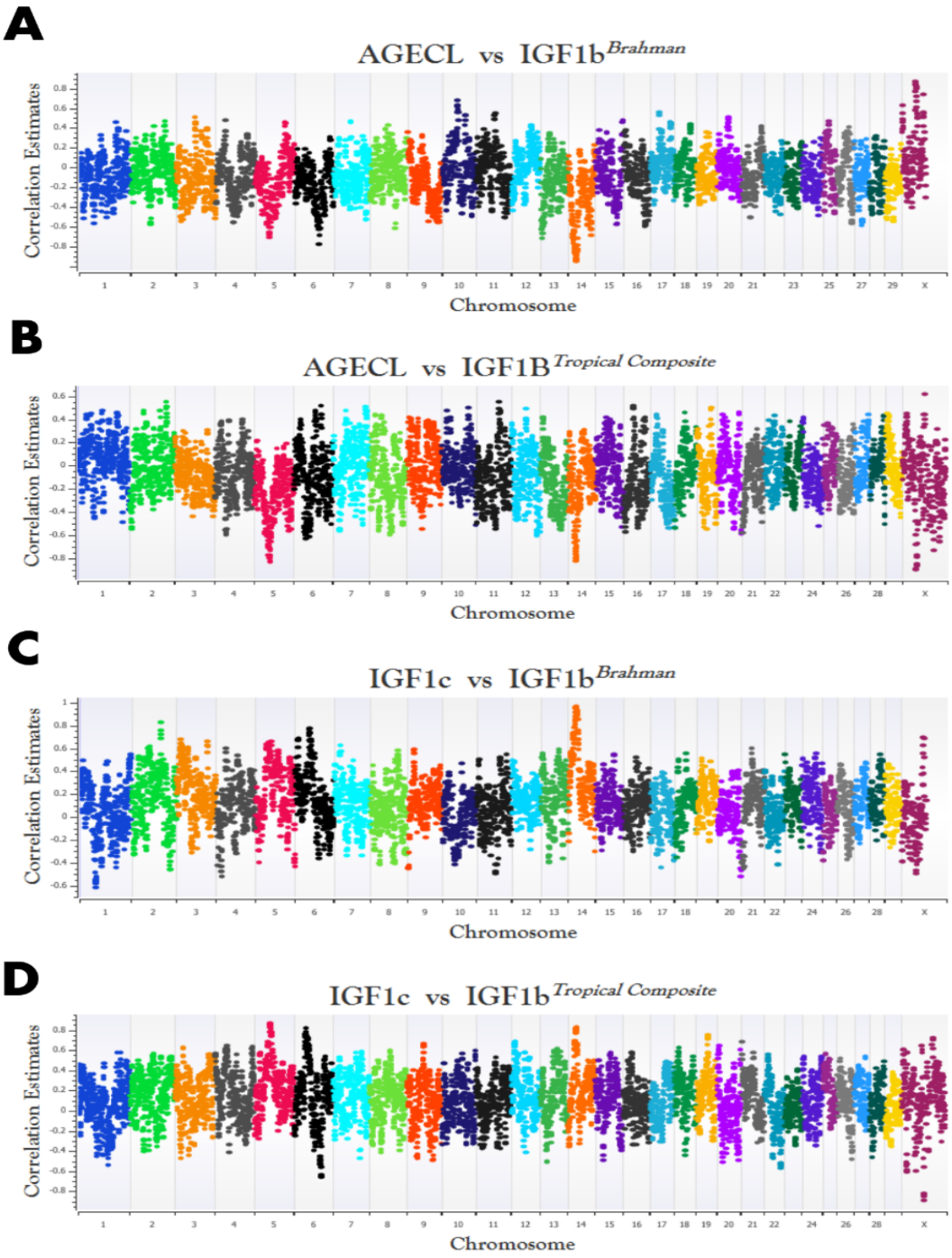

Figure 1. Genome plot of the regions driving and antagonizing trait correlations in Brahman (BB) and Tropical Composite (TC) for the pairwise traits (BB-AGECL vs IGF1b; **A**, TC-AGECL vs IGF1b; **B**, BB-IGF1c vs IGF1b; **C**, TC- IGF1c vs IGF1b; **D**). **AGECL**, age at first corpus; **IGF1**, serum levels of insulin growth hormone (measured in bulls, **IGF1b**, or cows, **IGF1c**). The correlation estimates were plotted on the y-axis and the genomic position (i.e., midpoint between the start and end position of each window) of each chromosome on the x-axis, according to the ARS_UCD1.2 bovine reference genome.

Driver, antagonizing and neutral genomic windows affecting genetic correlations between fertility traits: permutation test

In order to identify drivers, antagonizing and neutral regions across the bovine genome, we performed permutation test by randomly reshuffling the Single Nucleotide Polymorphisms (SNPs) effects in each window across all chromosomes in 1,000 iterations for each trait. Then, we applied our framework on the randomized SNP effects and observed the **r** estimate across each iteration for each window. In most cases, the maximum and the minimum **r** estimates for each window at each iteration (i.e., rand 1 to rand 1,000) range between ±0.20. Therefore, we considered neutral windows with no significant effect on the trait correlation as windows with -0.20≤ **r** ≤0.20 estimates. The genomic plots of the **r** estimates resulting from the permutation test (rand 500 only) in each population are presented in Additional file 2 (Fig. S1). Additional file 3 (Table S3-S6) shows the numbers of windows, their chromosome coordinates, and **r** estimates as well as the maximum and minimum r estimates across the 1,000 iterations for each pair of traits in each of the two populations.

As a result of the permutation test, we considered significant windows with **r** estimates >0.2 and **r** estimates <-0.2. These thresholds were used to define the significant windows or regions (i.e., driver and antagonizing) from non-significant (i.e., neutral) windows or regions. Depending on the overall genetic correlation between traits, driver and antagonizing windows can be deduced: in driver windows, the **r** estimate has the same direction, positive or negative, as the overall genetic correlation; in antagonizing windows it is the opposite.

The number of significant driver windows for the correlation between AGECL and IGF1b was 1,636 in BB and 1,914 in TC cattle. The number of significant windows for the antagonizing was 547 in BB and 898 in TC cattle, for AGECL vs IGF1b. For the correlation between IGF1c and IGF1b, the number of significant driver windows was 1,931 in BB and

2,549 in TC cattle. The antagonizing windows was 402 in BB and 587 in TC cattle (IGF1c vs IGF1b). The numbers of neutral windows were as follows: 3,375 in BB, and 4,064 in TC for AGECL vs IGF1b; and 3,225 in BB, and 3,740 in TC for IGF1c vs IGF1b. See Table 1 for details on numbers of windows in the two beef cattle populations. In addition, the lists of windows with their chromosomal coordinates for all driver, antagonizing, and neutral regions are presented in Additional file 4 (Table S7-S18).

Table 1: The number of windows generated for the driver, antagonizing and neutral windows for each pairwise trait in Brahman and Tropical Composite population.

Pairwise Trait	**Number of windows (percentage to the total number of window)**			**Total number of windows**
	Driver (%)	Antagonizing (%)	Neutral (%)	
		Brahman		
AGECL vs IGF1b	1,636 (29.40%)	547 (9.84%)	3,375 (60.72%)	5,558
IGF1c vs IGF1b	1,931 (34.74%)	402 (7.23%)	3,225 (58.03%)	5,558
		Tropical Composite		
AGECL vs IGF1b	1914 (27.84%)	898 (13.06%)	4064 (59.10%)	6,876
IGF1c vs IGF1b	2549 (37.73%)	587 (8.54%)	3740 (54.39%)	6,876

AGECL, age at first *corpus;* **IGF1c**, serum levels of insulin growth hormone measured in cow; **IGF1b**, serum levels of insulin growth hormone measured in bulls

For the correlation between AGECL and IGF1b (overall genome-wide correlation of -0.65 (BB) and -0.55 (TC), see Table 2), the largest **r** estimate for the driver windows was -0.96 (bovine chromosome (BTA)14: 23.04 - 25.29Mb) in BB and -0.91 (BTAX: 39.76 - 42.86Mb) in TC. For the antagonizing windows, the largest **r** estimate was 0.87 (BTAX: 40.87 - 43.88Mb) in BB and 0.61 (BTAX: 66.62 - 69.622Mb) in TC.

For the correlation between IGF1c and IGF1b (overall genome-wide correlation of 0.86 (BB) and 0.93 (TC), see Table 2), the largest estimate for the driver windows was 0.97 (BTA14: 22.68 - 24.96Mb) in BB and 0.87 (BTA5: 46.13- 47.89Mb) in TC, while the estimate for the antagonizing was -0.62 (BTA1: 49.01 - 51.67Mb) in BB and -0.90 (BTAX: 65.64 - 68.39Mb) in TC. All **r** estimates are plotted in Figure 1.

Genes and quantitative trait loci (QTL) within driver and antagonizing regions across the two populations

Defining driver and antagonizing regions separately for each pair of traits, allowed us to identify the genes and QTLs within these regions for each of the two beef cattle populations. The percentage of the overlapping genes (Figure 2a) and QTLs (Figure 2b) across both populations was studied. The percentages of genes shared across the significant regions in BB and TC were calculated as a function of the total number of genes in BB or TC, respectively, and so they differ (Figure 2a and 2b).

The percentage of overlapping genes for each pair of traits in the two populations were as follows: for AGECL vs IGF1b driver regions, about 48% of the total number of genes annotated were shared between the two population, whereas, for the antagonizing regions, 22% of the gene annotated in BB were present in the TC population, and 26% of the genes annotated in TC were present in BB; for IGF1c vs IGF1b, the two populations shared about 58% of total number of genes annotated for the driver regions and about 17% were shared for the antagonizing regions.

The percentage of overlapping QTLs for each pair of traits in BB and TC population were as follows: for AGECL vs IGF1b driver regions, 52% of the QTLs annotated in BB were present in TC and 35% of the QTLs annotated in TC were present in BB, whereas, for the antagonizing regions, 20% of the QTLs annotated in BB were present in TC and 18% of the

QTLs annotated in TC were present in BB; for IGF1c vs IGF1b, 56% of the genes annotated in BB were present in TC, and 52% of the genes annotated in TC were present in BB, whereas, for the antagonizing regions, 29% of the genes annotated in BB were present in TC and 33% of the genes annotated in TC were present in BB population.

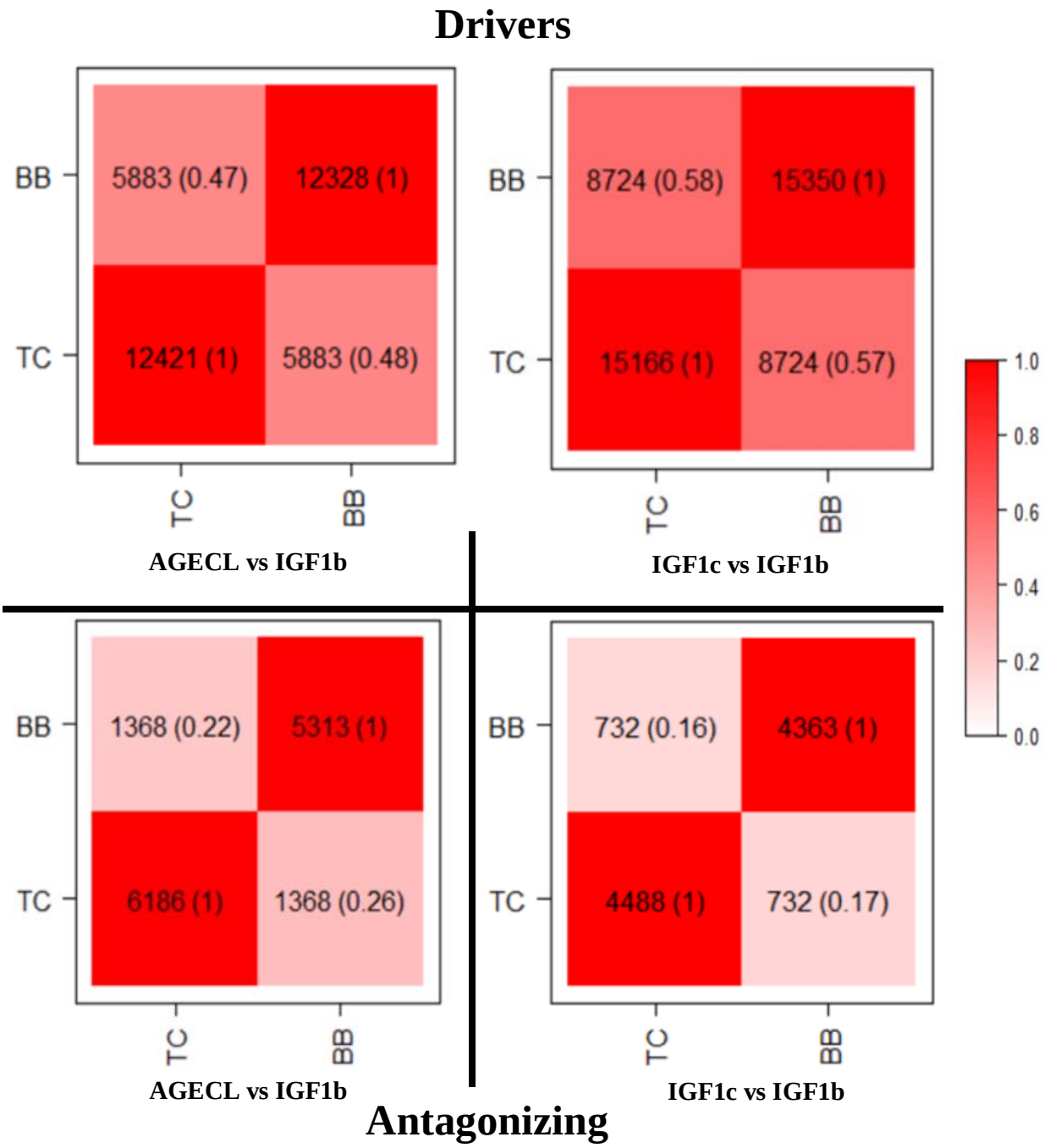

Figure 2a. Genes annotated in the significant (i.e., driver and antagonizing) genomic regions identified as explaining the genetic correlations between male and female fertility traits in Brahman (BB) and Tropical Composite (TC) population. The overlaps between the two studied populations are in the diagonal of each plot for each pair of traits within the driver (above) and antagonizing (below) regions. The darker the colour within the squares, the higher the percentage of shared genes or QTLs. **AGECL**, age at first corpus; **IGF1**, serum levels of insulin growth hormone (measured in bulls, **IGF1b**, or cows, **IGF1c**).

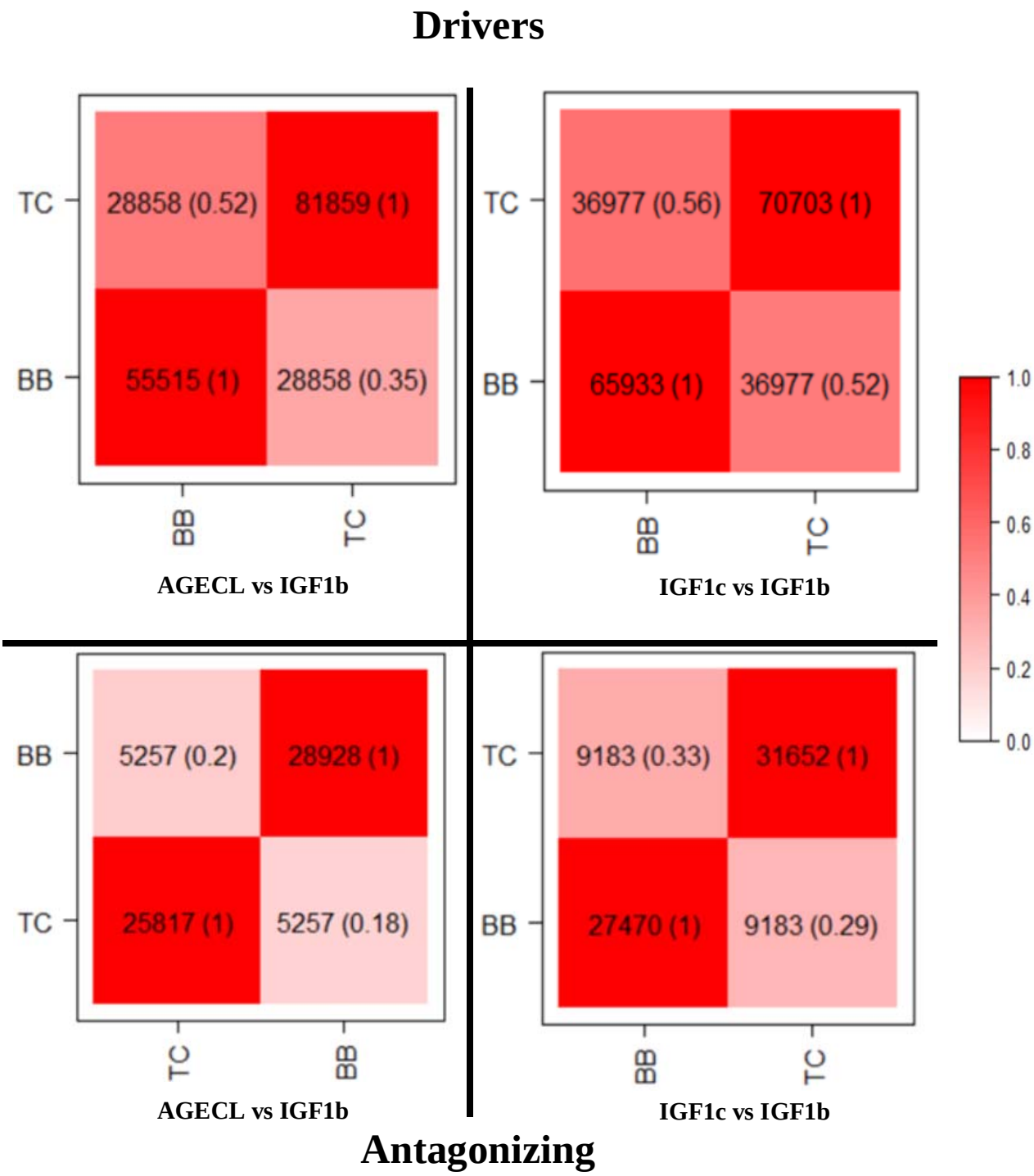

Figure 2a. QTLs annotated in the significant (i.e., driver and antagonizing) genomic regions identified as explaining the genetic correlations between male and female fertility traits in Brahman (BB) and Tropical Composite (TC) population. The overlaps between the two studied populations are in the diagonal of each plot for each pair of traits within the driver (above) and antagonizing (below) regions. The darker the colour within the squares, the higher the percentage of shared genes or QTLs. **AGECL**, age at first corpus luteum; **IGF1**, serum levels of insulin growth hormone (measured in bulls, **IGF1b**, or cows, **IGF1c**).

Functional classification of QTLs within genomic regions that explain the genetic correlations between male and female fertility

To infer biological function and mine the existing literature, we examined the types of QTL (milk, reproduction, production, meat and carcass, health and exterior) present in the significant genomic regions identified above using GALLO [27]. The most frequent QTLs across all pairwise traits in the two populations for the driver and antagonizing regions were QTLs related to milk production, accounting for about 30-51% in most cases. This was followed by reproductive QTLs accounting for about 13-48% and production QTLs comprising 6-24%. Other QTL types (Exterior, health and meat and carcass) accounted for a relatively small proportion of QTLs in the significant regions (Figure 3a and 3b). In addition, we report the top 10 results for QTLs related to reproductive traits as these are relevant to our studied traits (Figure 3a and 3b). Among these reproductive QTLs, traits related to puberty (i.e., age at puberty, scrotal circumference) were prevalent in both populations.

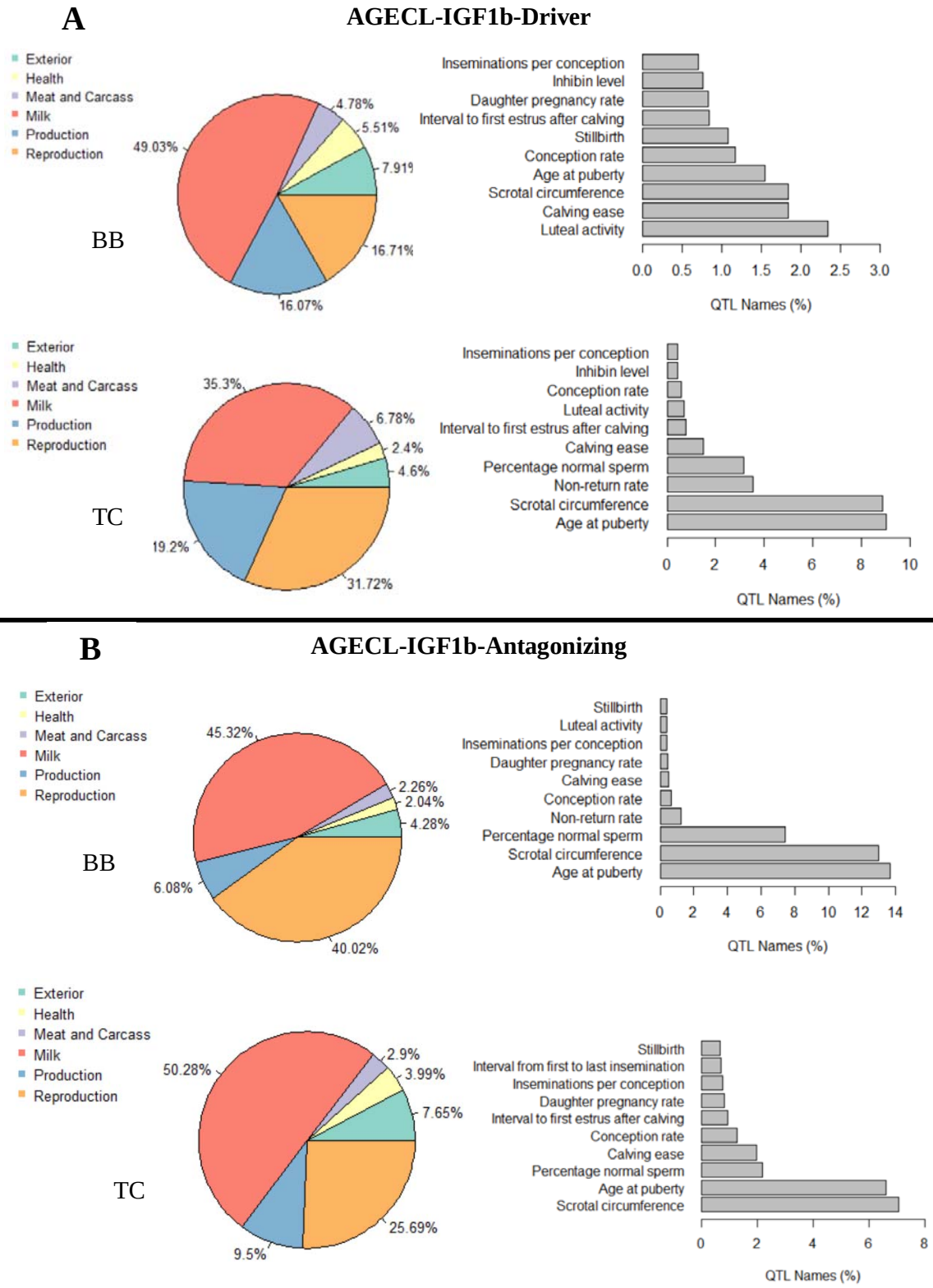

Figure 3a. Percentage of QTL type (pie chart) and trait related to reproduction QTLs (barplots) for the QTL annotation results obtained for (A) AGECL vs IGF1b - driver, (B) AGECL vs IGF1b- antagonizing in Brahman (BB) and Tropical Composite (TC) population. **AGECL**, age at first corpus luteum, **IGF1c**, **IGF1**, serum levels of insulin growth hormone (measured in bulls, **IGF1b**, or cows, **IGF1c**)

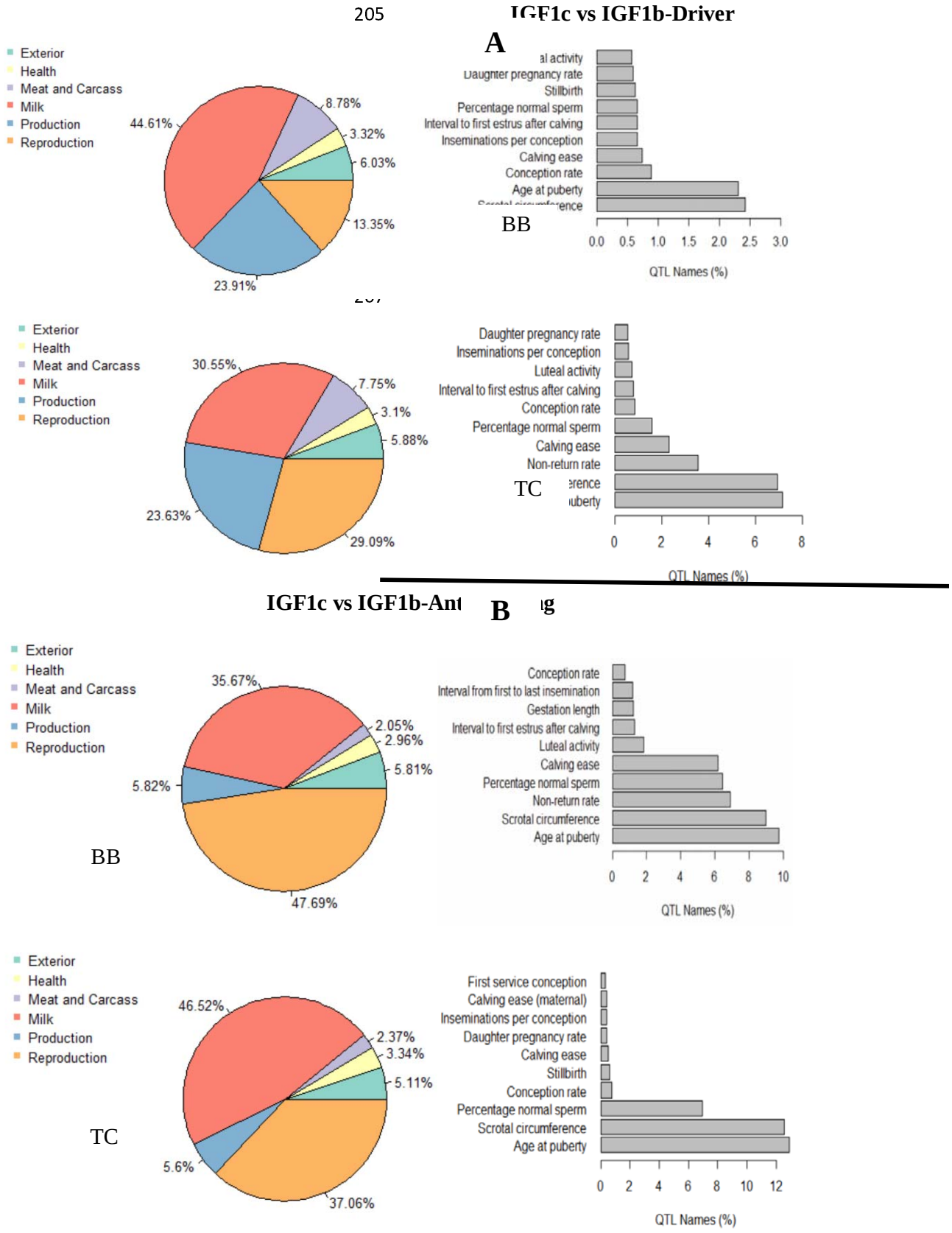

Figure 3b. Percentage of QTL type (pie chart) and trait related to reproduction QTLs (barplots) for the QTL annotation obtained for **(A)** IGF1c vs IGF1b- driver, (B) IGF1c vs IGF1b- antagonizing in Brahman (BB) and Tropical Composite (TC) population. **AGECL**, age at first corpus luteum, **IGF1**, serum levels of insulin growth hormone (measured in bulls, **IGF1b**, or cows, **IGF1c**)

QTL enrichment analysis

We performed a chromosome-wide QTL enrichment analysis to further test the significance of the QTLs identified for all the driver and antagonizing regions in each cattle population, for each trait pair using GALLO [27]. Enriched QTLs for the studied traits span across most QTL types, indicating the presence of complex genetic mechanisms. The results of the chromosome-wide QTLs enrichment (FDR-corrected p-value≤0.05) for the driver and antagonizing regions for all pairwise traits in each population are presented in Additional file 5 (Table S19-26).

For the driver regions, the number of QTLs enriched over a wide range of chromosomes for AGECL vs IGF1b were 233 and 144 in BB and TC beef cattle population, respectively. The number was 227 (BB) and 220 (TC) for IGF1c vs IGF1b. For AGECL vs IGF1b, the most enriched chromosome (no of enriched QTLs in parenthesis) was BTA5 (36) and BTA14 (18) in BB and TC, respectively. IGF1c vs IGF1b also followed similar pattern with the result above, with BTA5 (41) being the most enriched chromosome in BB and BTA14 (51) as the most in TC.

For the antagonizing regions, the number of QTLs enriched across the bovine chromosomes for AGECL vs IGF1b were 127 and 178 in BB and TC beef cattle population, respectively. The number was 179 (BB) and 195 (TC) for IGF1c vs IGF1b. For AGECL vs IGF1b, the most enriched chromosome was BTA17 (14) and BTA26 (21) in BB and TC, respectively. For IGF1c vs IGF1b, however, BTA14 (23) was the most enriched chromosome for these regions in BB, whereas, in TC, BTA14 (23) was the most enriched.

To identify the common results and shared biology between the driver and the antagonizing regions, we also investigated the overlaps of the QTL types associated with the studied trait (i.e., reproduction) in the two populations. The relationship between the top 10 enriched

reproductive QTLs in BB and TC are presented in Figure 4a and 4b. Irrespective of the trait pair, for the driver regions, the reproductive QTLs in BB in most cases overlap with those identified in TC. However, for the antagonizing regions, not all reproductive QTLs in BB were found in TC beef cattle population.

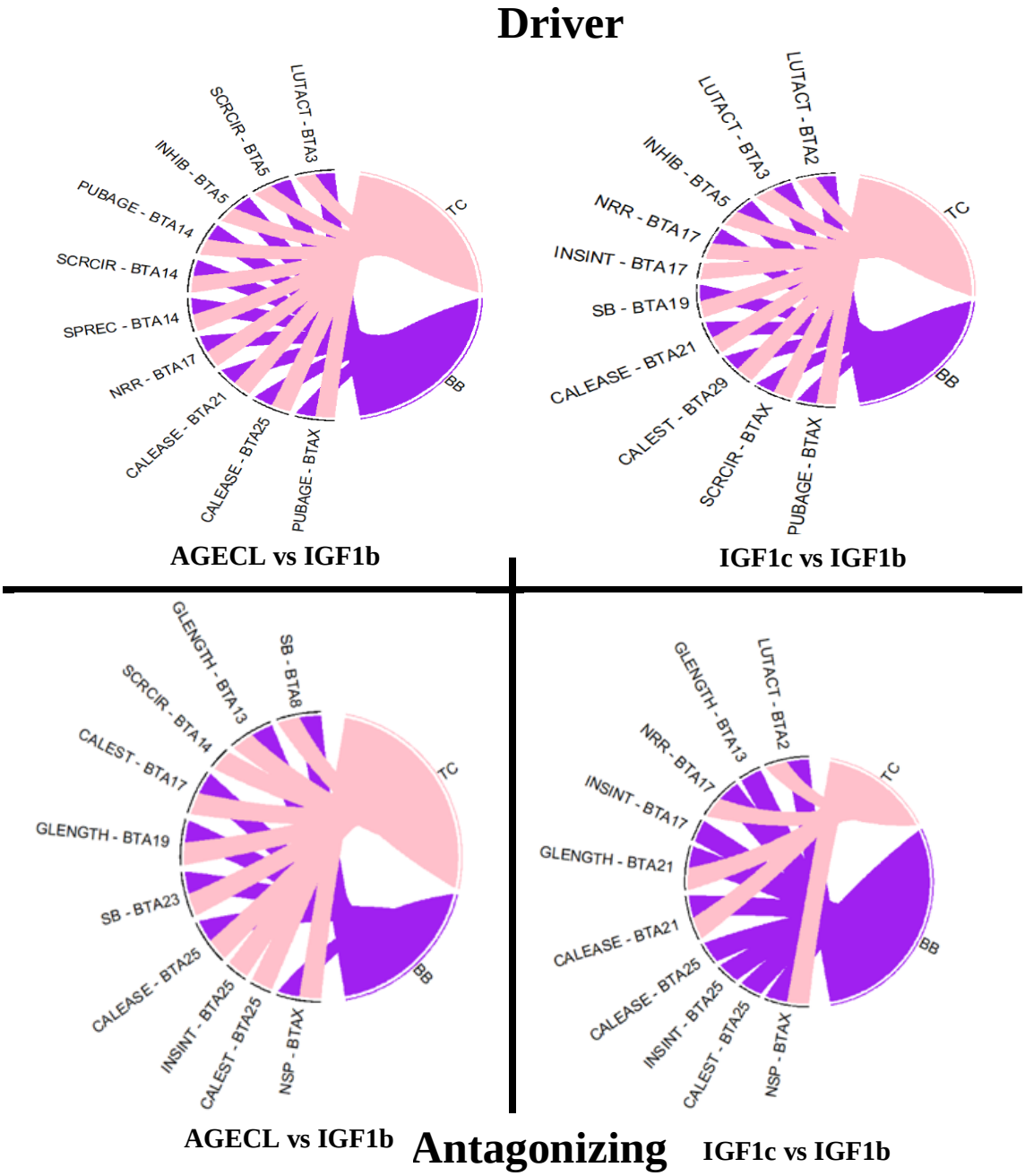

Figure 4. Chord plot showing the relationship between the top 10 enriched reproductive QTLs between Brahman (BB) and Tropical Composite (TC) for the driver (top) and the antagonizing (bottom) regions of the studied traits. **AGECL**, age at first corpus luteum, **IGF1c**, serum levels

of insulin growth hormone measured in cow; IGF1b. LUTACT, Luteal activity; SCRCIR, Scrotal circumference; INHIB, Inhibin level; PUBAGE, Age at puberty; SPREC, Sexual precocity; NRR, Non-return rate; CALEASE, Calving ease; INSINT, Interval from first to last insemination; SB, Still birth; CALEST, Interval to first estrus after calving; GLENGTH, Gestation length, NSP, Percentage normal sperm.

Functional enrichment analysis

Leveraging our methodology's directionality of gene effects with Ingenuity Pathway Analysis (IPA; http://www.ingenuity.com), we identified the enriched canonical metabolic pathways enriched at Benjamini–Hochberg corrected p-values (BH-*P*-value) of $p<0.01$. The graphical presentation of the canonical metabolic pathways predicted by IPA to be enriched and the proportion of driver and antagonizing genes in each pathway for all pairwise traits investigated in each population are illustrated in Additional file 6 (Fig. S2-S9). Although IPA provided information about whether the predicted pathways were being activated or inhibited based on our data, we remain cautious when interpreting our results since the **r** estimates are not the same as gene expression values, and IPA was originally designed to mine gene expression data.

The number of pathways enriched for AGECL vs IGF1b was 49 in BB and 68 in TC. For IGF1c vs IGF1b, the number of enriched pathways was 156 and 87 in BB and TC, respectively. For AGECL vs IGF1b in BB, the top 5 enriched canonical metabolic pathways were cardiac hypertrophy signaling (Enhanced), toll-like Receptor signaling, IL-6 Signaling, hepatic fibrosis signaling pathway and STAT3 Pathway. In TC population, the top 5 enriched canonical metabolic pathways were breast cancer regulation by stathmin1, signaling by Rho family GTPases, opioid signaling pathway, endocannabinoid developing neuron pathway, and CREB signaling in neurons.

For IGF1c vs IGF1b in BB, the top 5 enriched canonical metabolic pathways were cardiac hypertrophy signaling, CREB signaling in neurons, thrombin signaling, estrogen receptor

signaling, opioid signaling pathway and AMPK signaling. In TC, the top 5 enriched pathways were CREB signaling in neurons, cardiac hypertrophy signaling (enhanced), opioid signaling pathway, gαs signaling and breast cancer regulation by stathmin1. The enriched canonical metabolic pathways and all the genes involved in each pathway are available in Additional file 7 (Table S27-30).

Discussion

Complex phenotypes, including fertility, consist of multiple genetically correlated rather than independent traits [6]. The interplay between traits involve many genomic regions, usually in a large and polygenic regulatory network [28-31]. Genomic signals that regulate (i.e., drive or antagonize) complex traits are widely spread across the genome, including near many genes without significant effect on the phenotype or disease [29]. In the present post-genomic era, unravelling the genomic regions that regulate complex traits and the metabolic pathways associated with these phenotypes has become an important aspect of genetic studies in humans and animals [32]. In this study, we developed a novel framework termed correlation scan to reveal the significant regions that either drives or antagonize the genetic correlations between traits, across the genome. In addition, this method can also reveal genomic regions with no effect on the studied traits (neutral windows). The framework developed uses best linear unbiased prediction (BLUP) solution of SNP effects to estimate the local correlations between studied traits. Local correlations are based on sliding windows of 500-SNPs. We applied these sliding windows approach to reproductive traits measured in two populations and subject the outcomes, significant windows, to further analyses using GALLO [27]. and IPA (http://www.ingenuity.com) to gain further insight about the biology of studied traits and their relationships. Although the methodology was applied to beef cattle traits, using high-density SNP chip genotypes, the general framework can be applied to any species, any traits, and it can easily accommodate sequence level data.

Our results agreed with the established notion that multiple loci regulate reproductive traits [33-35]. Also, the mode of action of these loci and the magnitude of their effect varies across the genome. While some regions had no effect on the genetic correlations under investigation, other loci drive or antagonize the relationships between male and female fertility. The identification of driver and antagonizing loci creates opportunities to further understand the molecular mechanisms affecting quantitative traits. For example, correlations estimated from SNP effects have allowed researchers to construct gene networks [36]. Thereby, these types of approaches contribute to linking genotype with phenotype.

The two beef cattle populations investigated in this study are distinct in terms of their genetic composition. Brahman (BB) cattle are typically of *Bos indicus* origin whereas TC beef cattle emanated from the crossing between Bos *indicus* and Bos *taurus* breed [37]. Despite these differences, we found that a considerable number of annotated genes and QTLs driving trait correlation overlaps across breeds, although with variations in the size of SNP effects. This corroborates the findings of Bolormaa S, Pryce JE, Kemper KE, Hayes BJ, Zhang Y, Tier B, Barendse W, Reverter A and Goddard ME [38], where a substantial number of QTLs were found segregating in *Bos indicus* and composite cattle using the same dataset. In this present study, the top genomic signal driving trait correlation across all pairwise traits in BB were located on BTA14. The significant region contains a widely known and well-characterized QTL, including the *PLAG1* gene, reported to be associated with growth and reproductive traits in our populations and other studies [39-44]. In TC however, the top signal differs across traits and mostly spread across two or three chromosomes, although with considerable number of overlaps with BB. This could be partly due to the variations in the architecture of composite breed [45]. The genome of composite breeds usually contains new haplotypes emerging from generations of crossbreeding. Moreover, the contribution of the founder populations on chromosomes and specific genomic regions are usually unevenly distributed,

which most likely shapes the genome of composite breeds [45]. In short, differences between BB and TC are likely to impact the results of our analyses. Breed differences are expected, and so when two breeds share a similar result, it enhances our confidence in calling significant windows for the interplay between male and female fertility traits.

Most genomic regions antagonizing the genetic correlations between male and female fertility traits were located on chromosome X. Gene expression on chromosome X differs across-sex, resulting in genomic sexual conflict [46-48]. Genes in these antagonizing regions include *PO1FB, ZNF711, APOOL, HDX, DACH2, FAM133A,* among others. These genes are associated with different disorders including infertility, reproductive deficiencies, primary ovarian failure [49-51]. When some of these genes are over-expressed, it can dysregulate the cristae morphology of the mammalian mitochondria [52]. Understanding how these antagonizing genes interact to influence (in)fertility could help improve the reproductive potentials of beef cattle.

In animal production, more research is carried out on milk production-related traits, thereby creating large proportion of records for these traits in the cattle QTL database. These volumes of records can create a bias in the QTLs representativeness [27]. The QTL enrichment analysis allows testing the significance of the QTL representative using chromosome-wide approach to detect specific genomic region with many QTLs for a specific trait. For example, taking the driver regions for AGECL vs IGF1b in BB, the top enriched QTLs was found in BTA5, harbouring 36 QTLs. These QTLs comprised 8 different QTLs for reproduction (inhibin level, scrotal circumference, interval of first estrus after calving, gestation length, insemination per conception, conception rate, daughter pregnancy rate, and pregnancy rate). These 8-traits listed here have been found to be correlated with puberty (studied traits) in cattle. For instance, inhibin is regarded as a biomarker for sexual development because it regulates spermatogenesis in both beef and dairy bulls [53, 54]. Moderate genetic correlation

was found between inhibin and AGECL [55] and between inhibin and IGF1b [56] in BB. Scrotal circumference has also been found to be a moderate predictor of AGECL and IGF1b in BB [21, 26]. Thus, BTA5 may be a candidate region for fertility in BB beef cattle population. Other enriched QTLs out of the 36 mentioned above include 8 different production traits (average daily gain, metabolic body weight, length of productive life, body weight, rump width, body depth, residual feed intake, and net merit). These traits are related to feed efficiency in cattle. Improving feed efficiency of beef cattle is a major concern for beef producers. A recent study from Canal LB, Fontes PL, Sanford CD, Mercadante VR, DiLorenzo N, Lamb GC and Oosthuizen N [57] found that heifers that efficiently utilize feed attain puberty early than less feed efficient ones. Moreover, heifers that attain puberty at a relatively younger age have the potential to conceive early in life and be more productive throughout their lifetime [58]. In addition, IGF1 is an effective selection tools to improve feed efficiency and other production related traits, allowing breeders to preselect animals that can utilize feed efficiently [59, 60]. Other enriched QTLs for BTA5 in BB are related to exterior (7), milk (6), milk and carcass (5) and health (3) traits. Of note, the objective of most beef cattle breeding programs is to change the genetic merit of their cattle for many traits of interest [61]. The recurrent association of the BTA5 with multiple traits could suggest complex genetic mechanisms such as pleiotropy, epistasis, hitchhiking effects, linkage disequilibrium etc., regulating this chromosomal region [62, 63]. Therefore, breeders could target BTA5 to select multiple traits without any antagonistic effect on other traits listed herein.

Another interesting result from this study is the shared biology between the two breeds relative to the traits under study. Despite breed differences, the enriched reproductive QTLs driving the genetic correlations between male and female fertility are the same for the two cattle populations (Figure 4). Most of the enriched QTLs are related to reproductive traits

measured early life. A possible explanation could be that the reproductive phenotypes shared common fundamental biology in the two populations. For the antagonizing regions, however, most of the reproductive QTLs were breed specific depending on the trait pairwise. Perhaps, this could be partly explained by the diverse genetic composition of the two breeds. Understanding the genomic architectures driving these early-in-life male and female fertility traits and their known genomic antagonisms could foster effective selection for both traits in tropical breeds [64, 65].

The major challenge faced by researchers when analysing an overwhelmingly large amount of genomic data is how to extract meaningful mechanistic insights into the underlying biology characterizing the given trait under study. To increase the explanatory power of genomic studies, pathway analysis has become first choice, providing researcher with the ability to infer meanings to high-throughput genomic data [66]. Leveraging the directionality of gene effects from our method with IPA knowledge base, several biological pathways known to be involved in reproduction (i.e., studied trait) were significantly enriched for all pairwise traits investigated across the two breeds. These pathways include estrogen receptor signaling, p38 MAPK signaling, GnRH signaling, sperm motility, cAMP-mediated signaling, AMPK signaling, and androgen signaling. Although IPA provided information about the activation or inhibition state for the enriched canonical metabolic pathways with the use of the **r** estimates in place of the gene expression values, we are not sure if these pathways were being activated or inhibited since we don't have information about the expression values of the genes in these pathways. For example, Rho GDP Dissociation Inhibitor (RHOGDI) pathway was the only significant signaling pathway found to be inhibited across breeds in all pairwise traits investigated using IPA comparison analysis. The RHOGDIs (RHOGDIα, RHOGDIβ and RHOGDIγ) are well-characterized as a negative regulator of Rho GTPases [67]. These Rho GTPases play pivotal roles within the cell, including cell migration,

membrane trafficking, invasion, gene transcription, polarity, adhesion, cell survival and death; a process significantly involved in cancer initiation and metastasis [68, 69]. Once RHOGDI is inhibited, it induces constitutive activation of Rho GTPases, resulting in several malignant phenotype including tumour growth, angiogenesis, and invasive phenotypes [69, 70]. For instance, knocking out one of the three RHODGI genes resulted in a renal defect that progressively leads to death in adult mice, although embryonic development was not affected [71]. Togawa A, Miyoshi J, Ishizaki H, Tanaka M, Takakura A, Nishioka H, Yoshida H, Doi T, Mizoguchi A and Matsuura N [72] also found that male mice lacking RHOGDI1 were infertile with impaired spermatogenesis. The authors also reported problems of implantation in female mice due to this knockout. The knockout of two of the three RHOGDIs often results in a more severe phenotypes with additional immunological defects than when one of the RHOGDIs is disrupted [73]. Numerous studies have also reported that the RHOGDIs protein are involved in sperm movement, sperm capacitation and acrosome reaction, a process that is critical to occur for the sperm to interact and penetrate the egg for fertilization to take place [74-76]. Perhaps, this could be the major reason why signaling by Rho family GTPases were enriched in our metabolic pathway analysis. Notably, low reproduction performance is one of the major challenges facing beef producers in Northern Australia [77, 78]. Reproductive wastage is usually common, which is often a result of pregnancy failure and calf mortality [79, 80]. Given the role of the RHODGI pathway in reproduction, future studies could use gene expression data to investigate the genes involved in these pathways as a candidate region for infertility in cattle since we only use the **r** estimates in this study.

Conclusion

Overall, the framework developed in this study extends our knowledge about the regions driving and antagonizing correlations between quantitative traits. About 40% of the total genomic regions were identified as driving and antagonizing genetic correlations between

male and female fertility traits in the two population. These regions confirmed the polygenic nature of the traits being studied and pointed to genes of interest. Quantitative trait loci (QTL) and functional enrichment analysis revealed that many significant windows co-located with known QTLs related to milk production and fertility traits, especially puberty. In general, the enriched reproductive QTLs driving the genetic correlations between male and female fertility are the same for both cattle populations, while the antagonizing regions were population specific. Moreover, most of the antagonizing regions were mapped to the chromosome X. These results suggest regions of the chromosome X for further investigation into the trade-offs between male and female fertility. Although the methodology was applied to cattle phenotypes, using high-density SNP genotypes, the general framework developed can be applied to any species or traits, and it can easily accommodate genome sequence data.

Materials and Methods

Traits, genotypes and estimated genetic correlations

The traits used to demonstrate this methodology are a subset of traits from our previous study [26], where bivariate genetic correlations were estimated between 7 male and 6 female early-in-life reproductive phenotypes in two independent tropical beef cattle populations (BB and TC). The two female traits selected for this study are age at detection of the first *corpus luteum* (AGECL, days) and cows' blood concentration of insulin growth-factor 1, measured at 18 months of age (IGF1c). Only one male trait was selected: the blood concentration of insulin growth-factor 1, measured at 6 months of age (IGF1b). These traits are important in beef cattle fertility, especially during pubertal development. The estimated genomic correlations between the traits listed above in each population have been reported in our previous study [26]. These estimates and their corresponding standard error (S.E), number of SNPs and number of animals in each population are provided in Table 2. These traits were

selected because they had significant estimates of genomic correlation (i.e., traits with standard error (S.E) less than half of the size of the correlation) and different strength or direction of genetic relationships (i.e., negatively, and positively correlated traits). In brief, across-sex genetic correlations were estimated in a bivariate analysis using the linear mixed model approach. Firstly, the 770,000 genotypes were mapped to the new assembly of the bovine reference genome (ARS_UCD1.2, GenBank assembly accession GCA_002263795.2; [81]). After quality control filtering (i.e., excluding all SNPs with a minor allele frequency less than 5%), 554,712 and 686,626 SNPs remained for BB and TC datasets, respectively. Finally, bivariate genetic correlations were estimated using GIBBS2F90 [82], resulting to the estimates in Table 2.

Table 2: Genomic correlations estimates and their corresponding standard error (s.e), number of animals and number of SNPs estimates

Pairwise traits	No of animals	Number of SNP	Genetic correlation (s.e)
		Brahman	
AGECL vs IGF1b	AGECL- 980	554K	-0.65 (0.13)
	IGF1b- 964		
IGF1c vs IGF1b	IGF1c- 995	554K	0.86 (0.11)
	IGF1b- 964		
		Tropical Composite	
AGECL vs IGF1b	AGECL-996	686K	-0.55 (0.14)
	IGF1b- 998		
IGF1b vs IGF1b	IGF1c- 1015	686K	0.93 (0.11)
	IGF1b- 998		

AGECL, age at first *corpus;* IGF1c, serum levels of insulin growth hormone measured in cow; IGF1b, serum levels of insulin growth hormone measured in bulls; SNP, Single Nucleotide Polymorphisms; s.e, standard error

Overview of methods

For each trait considered in the two beef cattle populations, we estimated the genomic breeding values (gEBVs) of individuals using the GBLUP model:

$$\boldsymbol{y} = \boldsymbol{XB} + \boldsymbol{Za} + \boldsymbol{e}, \qquad \textbf{[1]}$$

where $\boldsymbol{y}$ is the vector of phenotypes, **X** is the incidence matrix of fixed effects, $\boldsymbol{B}$ is the vector of fixed effects, **Z** is an incidence matrix mapping phenotypes to animals, $\boldsymbol{a}$ is the vector of genomic breeding values (GEBVs) for each animal, and $\boldsymbol{e}$ is the vector of residuals. Vectors **a** and **e** are assumed to follow a normal distribution, thus $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G}\sigma^2\mathbf{g})$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma^2\mathbf{e})$. Matrix **G** is the genomic relationship matrix.

Following the estimation of GEBVs for each animal in each trait, we then back-solved these gEBVs to obtain SNP effects for all chromosomes following the method illustrated by Strandén I and Garrick D [83];

$$\hat{u} = \frac{1}{\sum_{j=1}^{m} 2p_j(1-p)} X'G^{-1}\hat{a} \qquad \textbf{[2]}$$

Where $\hat{\boldsymbol{u}}$ is the vector of estimated SNP effects; $\boldsymbol{m}$ is the number of SNPs; $\boldsymbol{p_j}$ is the allele frequency of the second allele of the jth marker; **X** is a matrix with gene contents for all markers; **G** is a genomic relationship matrix and $\hat{\boldsymbol{a}}$ is a vector of GEBVs.

Notably, the estimation of GEBVs using GBLUP model [84-86] and the back-solving for GEBVs to obtain SNP effects using equation **[2]** both fits all SNP simultaneously, thereby allowing the joint estimation of SNP effects. As a result, the SNP effects from this approach are not biased by LD and resulting effect sizes can be considered independent.

Using a chromosome-wide approach, we divided SNPs on the same chromosome into small sliding windows of 500 SNPs each and then estimated the correlation (**r**) between traits as being the correlation between the 500-SNP effects estimated for trait A and the 500-SNP effects estimated for trait B. We then moved 100 SNPs further from the start of the previous window to select the next 500-SNP window, which partially overlapped with previous

window, hence producing sliding windows that were 100 SNPs distant from the previous window. This was repeated for each trait pair, and for each chromosome, in a chromosome-by-chromosome approach. The resulting **r** estimates for all the chromosomes combined were denoted as $W_1 \ldots W_n$. The graphical illustration of this framework is presented in Figure 4. Moreover, the coordinates of the windows ($W_1 \ldots W_n$) were mapped to the ARS_UCD1.2 bovine reference genome. The signals across the genome were visualized with the **r** estimates of each window on the y-axis and genomic position (i.e the midpoint of the start and end position of each window) of each chromosome on the x-axis. The mapping to the bovine reference genome and plotting of the windows signals' graphs were done using SNP & Variation Suite v8.x Golden Helix [87].

Depending on the overall genetic correlation observed between the traits considered, the driver and antagonizing windows can be deduced. In this study, AGECL and IGF1b were negatively correlated. Hence, the driver windows were windows with significant and negative **r** estimates, while the antagonizings were windows with significant positive **r** estimates. For the positively correlated relationship between IGF1c and IGF1b, the driver windows were windows with significant and positive **r** estimates and the antagonizings were windows with significant and negative **r** estimates. The significance of each window was established with a permutation test, described in the next section.

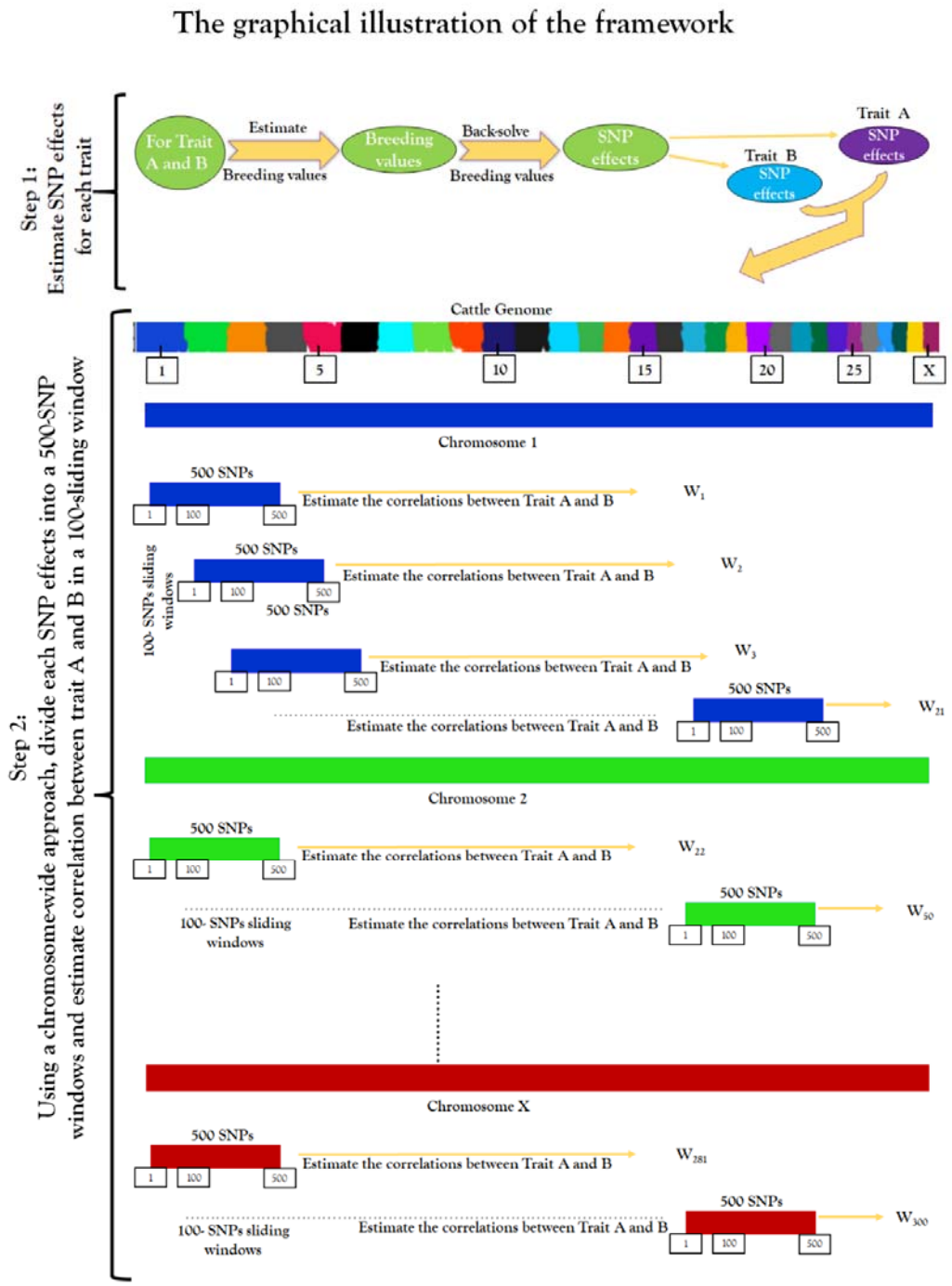

Figure 4. The graphical illustration of the sliding window framework. The framework involves 2 steps. Step 1 start from the estimation of genomic breeding values to the obtainment of SNP effects for each pairwise trait. Step 2 start from the estimation of 500-

SNP effects in a chromosome-wide approach to obtainment of the correlation estimate in a 100-sliding window.

Permutation test

To ensure the **r** estimates are not just noise but real signals, we performed permutation test by randomly reshuffling the SNP effects in each window across all the chromosomes in 1,000 iterations for each trait. Subsequently, we estimated correlations for 100-sliding windows of 500-SNP effects as described above. Finally, we observed the maximum and minimum **r** estimates for all the windows (W_1…….W_n) across the 1,000 iterations to reveal windows that were significant on the pairwise traits under investigation. Afterwards, we mapped the resulting **r** estimates for each window to the ARS_UCD1.2 bovine reference genome and plot the **r** estimates on the y-axis against the genomic position of each chromosome on the x-axis as described above. Consequently, significant windows were selected for the drivers and antagonizings genomic regions for each pairwise. Windows that were not significant were tagged "neutral windows" i.e., windows with no effect on the pairwise trait. Apart from using these windows to estimate genomic correlations and investigate the proportion of variance captured by these regions, they were excluded from other subsequent analyses. Finally, the **r** estimates of the significant windows for the driver and antagonizing regions were ranked from top to bottom in percentage (%) and the rank values were used solely for the purpose of subsequent downstream analyses. The ranking was done separately for the driver and antagonizing windows for each pairwise trait investigated in each population.

Gene and Quantitative Traits Loci (QTL) annotation

The significant windows along with their corresponding chromosome coordinates, **r** estimates and rank values for the driver and antagonizing regions that passed the specified threshold criteria following the permutation test in BB and TC were selected. The selected

windows were used for gene and QTL annotation using R package GALLO: Genomic Annotation in Livestock for positional candidate Loci (https://CRAN.R-project.org/package=GALLO) [27]. The .gtf annotation file corresponding to the bovine gene annotation from ARS-UCD1.2 assembly and the .gff file with the QTL information from cattle QTL Database (https://www.animalgenome.org/cgi-bin/QTLdb/index; [88, 89]), were used for gene and QTL annotation, respectively [27]. The two files use the same bovine reference genome (ARS-UCD1.2) to map the gene and QTLs. A remarkable advantage of GALLO is that the software retains all the information present in the input file when producing the output file. As a result, genes within each window can retain their **r** estimates and the rank values specific for their window.

The number and percentage of genes and QTLs annotated within a population (BB or TC) and the overlaps across populations (BB and TC) were investigated. Furthermore, we examined the QTLs representativeness and diversity to explain better the genomic content of the significant windows for the driver and antagonizing regions. Hence, the visualization of the percentage of cattle QTL types from cattle QTL database (i.e milk, reproduction, production, meat and carcass, health and exterior) were plotted using a pie chart by GALLO (27).

QTLs enrichment analysis

To further test the significance of the QTLs, a within population QTL enrichment analysis was conducted using a chromosome-based approach. The QTL enrichment analysis, using all the QTL information annotated within the significant windows for the driver and antagonizing regions, was performed using the qtl_enrich function from GALLO [90, 91]. Briefly, the observed number of QTLs for each trait in each annotated chromosome were compared with the expected number using a hypergeometric test approach in a 1,000 iteration

rounds of random sampling from the entire cattle QTL database. With this approach, a p-value for the QTL enrichment status of each annotated QTLs within the significant windows was estimated. These estimated p-values were corrected for multiple testing using a false discovery rate (FDR) of 5%. In addition, we used chord plots to reveal the relationships between the two breeds for the enriched reproductive QTLs based on the driver and antagonizing genomic regions.

Functional enrichment analysis

The annotated genes along with their corresponding **r** estimates and rank values for the significant driver and antagonizing windows for each pairwise trait in BB and TC populations were subjected to enrichment analysis using the commercial QIAGEN's Ingenuity Pathway Analysis (IPA; v.8.8, http://www.ingenuity.com). The IPA allows identifying overrepresented biological mechanism, metabolic pathways, and diseases and biological functions that are highly relevant to the traits of interest using the directionality of the submitted gene list [92, 93]. The outcome of our methodology indicates that genes within each window come with their directionalities, in this case, r estimates. Thus, we leveraged on the directionality of each gene by allowing the driver genes to be upregulated and antagonizing genes to be downregulated.

Summarily, a merged dataset containing gene identifiers that were significant for both the driver and antagonizing windows for each pairwise trait in each population and their corresponding **r** estimates and rank values were uploaded into IPA. The **r** estimates were used as the "Expr Log Ratio" and the rank values were used as p-values. The IPA software recognizes gene with positive signs (+) for "Expr Log Ratio" as upregulated genes and negative sign (-) as downregulated genes. We aim to allow the driver gene lists to have positive values for "Expr Log Ratio" and the antagonizing gene lists to be negative. Where

this is not achievable based on the original **r** estimates (i.e., AGECL vs IGF1b), we reversed the sign for the driver and antagonizing genes to meet this objective.

Of note, IPA can only analyse a maximum of 8,000 gene list. In most cases, the merged gene list for each pairwise trait in each population is often >8,000. Hence, we used the rank values as the cut-off to select the top ~80% genes from the driver and antagonizing gene list for the pathway analyses. Using a proportion of the gene list to infer biological pathways might result in the loss of some important biological information relevant to the trait of interest. We analysed the driver and antagonizing gene list separately for each pairwise trait in each population to ensure no important information was lost because of the cut-offs. Further, we compare the result of the separate analyses with the merged gene list from the ~80% cut-off.

The pathway analysis was conducted using the "Core Analysis" function implemented within IPA. In this analysis, associations were calculated using direct and indirect relationships among the gene lists. At first, the gene lists were mapped to human gene data. Genes without an associated gene symbol or gene annotation were subjected to an annotation by homology using BioMart application available in the Ensembl database (http://www.ensembl.org/biomart/martview/) [94, 95]. With this approach, we only considered non-annotated genes with percentage of identity ≥80% with human homolog. The final datasets used for the IPA analyses are presented in Additional file 8 (Table S31-34). Finally, the "Core Analysis" was used to identify canonical metabolic pathways enriched at Benjamini–Hochberg corrected p-values (B-H-P-value) of $p<0.01$).

Supporting information

Additional file 1: The number of windows, chromosome number, chromosome coordinates, and correlation estimates for each window for the two pairwise trait in Brahman and Tropical Composite population (Table S1-S2).

Additional file 2: Genome plots of the correlation estimates from the permutation test at 500 iterations in BB and TC for the two pairwise traits. The correlation estimates were plotted on the y-axis and the genomic position of each chromosome on the x-axis, according to the ARS_UCD1.2 bovine reference genome (Figure S1).

Additional file 3: The number of windows, chromosome number, chromosome coordinates, and correlation estimates as well as the maximum and minimum correlation estimate for each window for all trait pairwise in Brahman and Tropical Composite population following permutation test of 500-SNP effects in 100-SNP sliding windows at 1000 iterations (Table S3-S6).

Additional file 4: The number of windows, chromosome number, chromosome coordinates, correlation estimates and rank value for driver, antagonizing and the neutral regions that passed the threshold after permutation test in Brahman (BB) and Tropical Composite population for the studied trait (Table S7-18).

Additional file 5: The enriched QTLs of the driver and antagonizing regions for all trait pairwise in Brahman and Tropical Composite cattle. The enriched QTLs are rank based on the adj.pval (Table S19-26).

Additional file 6: Canonical pathways significantly enriched for all trait pairwise in Brahman and Tropical Composite population. Significantly enriched canonical pathways were identified using Benjamini-Hochberg p-values <0.01 (Figure S2-S9).

Additional file 7: The list of the significant enriched canonical metabolic pathways showing all the genes involved in each pathway for all trait pairwise in Brahman and Tropical Composite population. Significantly enriched canonical pathways were identified using

Benjamini-Hochberg p-values <0.01. Z-score >2 denote the activation of the pathway. Z-score <-2 indicate the inhibition of the pathway (Table S27-30).

Additional file 8: The final dataset used for Ingenuity Pathway Analysis (IPA) for all trait pairwise in Brahman and Tropical Composite cattle (Table S31-34).

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data used in this are available from the Cooperative Research Centre for Beef Genetic Technologies (Beef CRC). Data are available from https://www.beefcrc.com with the permission of Meat and Livestock Australia and the University of Queensland (if interested please contact the corresponding author).

Competing interests

None

Funding

BSO was supported by funding from Meat and Livestock Australia (L.GEN.1710; Female Fertility Phenobank and L.GEN.1818; Bull fertility update: historical data, new cohort and advanced genomics) and top-up scholarship from CSIRO.

Authors' contributions

OBS and LRC conceived the idea. OBS wrote the code, conducted the analyses, and wrote the manuscript. MST and GCC contributed to the analyses, manuscript writing and editing. AC conceive the downstream biological analyses. LR, BJH, and MRS supervised the study. All authors read, edited, and approved the final manuscript.

Acknowledgements

This research was conducted using the legacy database of the Cooperative Centre for Beef Genetics Technologies and their core partners including Meat and Livestock Australia. The authors acknowledge all the participants of the Cooperative Centre for Beef Genetics Technologies for their efforts in conducting the field trials and recording the phenotypes. We also acknowledge the mentorship of Professor Steve Moore to Babatunde Olasege in his PhD.

References

1. Liu S, Yu Y, Zhang S, Cole JB, Tenesa A, Wang T, McDaneld TG, Ma L, Liu GE, Fang L: **Epigenomics and genotype-phenotype association analyses reveal conserved genetic architecture of complex traits in cattle and human.** *BMC biology* 2020, **18:**1-16.
2. Zhang Y, Lu Q, Ye Y, Huang K, Liu W, Wu Y, Zhong X, Li B, Yu Z, Travers BG: **Local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits.** *bioRxiv* 2020.
3. Cánovas A, Reverter A, DeAtley KL, Ashley RL, Colgrave ML, Fortes MR, Islas-Trejo A, Lehnert S, Porto-Neto L, Rincón G: **Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle.** *PloS one* 2014, **9:**e102551.
4. Fonseca PAdS, Id-Lahoucine S, Reverter A, Medrano JF, Fortes MS, Casellas J, Miglior F, Brito L, Carvalho MRS, Schenkel FS: **Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle.** *PLoS One* 2018, **13:**e0205295.
5. Lynch M, Walsh B: **Genetics and analysis of quantitative traits.** 1998.
6. Van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR: **Genetic correlations of polygenic disease traits: from theory to practice.** *Nature Reviews Genetics* 2019, **20:**567-581.
7. Ning Z, Pawitan Y, Shen X: **High-definition likelihood inference of genetic correlations across human complex traits.** *Nature genetics* 2020, **52:**859-864.
8. Mackay TF, Stone EA, Ayroles JF: **The genetics of quantitative traits: challenges and prospects.** *Nature Reviews Genetics* 2009, **10:**565-577.

687 9. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA: **Detection and interpretation of**
688 **shared genetic influences on 42 human traits.** *Nature genetics* 2016, **48:**709-717.
689 10. Price AL, Spencer CC, Donnelly P: **Progress and promise in understanding the genetic basis**
690 **of common diseases.** *Proceedings of the Royal Society B: Biological Sciences* 2015,
691 **282:**20151684.
692 11. Lunstra D, Cundiff L: **Growth and pubertal development in brahman-, boran-, tuli-, belgian**
693 **blue-, hereford-and angus-sired f1 bulls.** *Journal of Animal Science* 2003, **81:**1414-1426.
694 12. Fang L, Sahana G, Ma P, Su G, Yu Y, Zhang S, Lund MS, Sørensen P: **Use of biological priors**
695 **enhances understanding of genetic architecture and genomic prediction of complex traits**
696 **within and between dairy cattle breeds.** *BMC genomics* 2017, **18:**1-12.
697 13. Neyhart JL, Lorenz AJ, Smith KP: **Multi-trait improvement by predicting genetic correlations**
698 **in breeding crosses.** *G3: Genes, Genomes, Genetics* 2019, **9:**3153-3165.
699 14. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, Ip HF, Marioni RE,
700 McIntosh AM, Deary IJ: **Genomic structural equation modelling provides insights into the**
701 **multivariate genetic architecture of complex traits.** *Nature human behaviour* 2019, **3:**513-
702 525.
703 15. Henderson C: **Recent developments in variance and covariance estimations.** *Journal of*
704 *Animal Science* 1986, **63:**208-216.
705 16. Lee SH, Van der Werf JH: **MTG2: an efficient algorithm for multivariate linear mixed model**
706 **analysis based on genomic information.** *Bioinformatics* 2016, **32:**1420-1422.
707 17. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R,
708 Zacher M, Furlotte NA: **Multi-trait analysis of genome-wide association summary statistics**
709 **using MTAG.** *Nature genetics* 2018, **50:**229-237.
710 18. Sodini SM, Kemper KE, Wray NR, Trzaskowski M: **Comparison of genotypic and phenotypic**
711 **correlations: Cheverud's conjecture in humans.** *Genetics* 2018, **209:**941-948.
712 19. Collet JM, Fuentes S, Hesketh J, Hill MS, Innocenti P, Morrow EH, Fowler K, Reuter M: **Rapid**
713 **evolution of the intersexual genetic correlation for fitness in Drosophila melanogaster.**
714 *Evolution* 2016, **70:**781-795.
715 20. Connallon T, Matthews G: **Cross-sex genetic correlations for fitness and fitness**
716 **components: connecting theoretical predictions to empirical patterns.** *Evolution letters*
717 2019, **3:**254-262.
718 21. Raidan FS, Porto-Neto LR, Reverter A: **Across-sex genomic-assisted genetic correlations for**
719 **sex-influenced traits in Brahman cattle.** *Genetics Selection Evolution* 2019, **51:**41.
720 22. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale
721 BM: **LD Score regression distinguishes confounding from polygenicity in genome-wide**
722 **association studies.** *Nature genetics* 2015, **47:**291-295.
723 23. Shi H, Mancuso N, Spendlove S, Pasaniuc B: **Local genetic correlation gives insights into the**
724 **shared genetic architecture of complex traits.** *The American Journal of Human Genetics*
725 2017, **101:**737-751.
726 24. Werme J, van der Sluis S, Posthuma D, de Leeuw C: **LAVA: An integrated framework for**
727 **local genetic correlation analysis.** *bioRxiv* 2021:2020.2012. 2031.424652.
728 25. Zhang Y, Lu Q, Ye Y, Huang K, Liu W, Wu Y, Zhong X, Li B, Yu Z, Travers BG: **SUPERGNOVA:**
729 **local genetic correlation analysis reveals heterogeneous etiologic sharing of complex**
730 **traits.** *Genome biology* 2021, **22:**1-30.
731 26. Olasege B, Tahir M, Gouveia G, Kour J, Porto-Neto L, Hayes B, Fortes M: **Genetic parameter**
732 **estimates for male and female fertility traits using genomic data to improve fertility in**
733 **Australian beef cattle.**
734 27. Fonseca PA, Suárez-Vega A, Marras G, Cánovas Á: **GALLO: An R package for genomic**
735 **annotation and integration of multiple data sources in livestock for positional candidate**
736 **loci.** *GigaScience* 2020, **9:**giaa149.

737 28. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A: **Reverse engineering**
738 **of regulatory networks in human B cells.** *Nature genetics* 2005, **37:**382-390.
739 29. Boyle EA, Li YI, Pritchard JK: **An expanded view of complex traits: from polygenic to**
740 **omnigenic.** *Cell* 2017, **169:**1177-1186.
741 30. Chen B-S, Yang S-K, Lan C-Y, Chuang Y-J: **A systems biology approach to construct the gene**
742 **regulatory network of systemic inflammation via microarray and databases mining.** *BMC*
743 *Medical Genomics* 2008, **1:**1-22.
744 31. Satokangas I, Martin S, Helanterä H, Saramäki J, Kulmuni J: **Multi-locus interactions and the**
745 **build-up of reproductive isolation.** *Philosophical Transactions of the Royal Society B* 2020,
746 **375:**20190543.
747 32. Skelly DA, Raghupathy N, Robledo RF, Graber JH, Chesler EJ: **Reference trait analysis reveals**
748 **correlations between gene expression and quantitative traits in disjoint samples.** *Genetics*
749 2019, **212:**919-929.
750 33. Antonarakis SE, Chakravarti A, Cohen JC, Hardy J: **Mendelian disorders and multifactorial**
751 **traits: the big divide or one for all?** *Nature Reviews Genetics* 2010, **11:**380-384.
752 34. Goddard M, Kemper K, MacLeod I, Chamberlain A, Hayes B: **Genetics of complex traits:**
753 **prediction of phenotype, identification of causal polymorphisms and genetic architecture.**
754 *Proceedings of the Royal Society B: Biological Sciences* 2016, **283:**20160569.
755 35. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM: **Simultaneous discovery,**
756 **estimation and prediction analysis of complex traits using a Bayesian mixture model.** *PLoS*
757 *genetics* 2015, **11:**e1004969.
758 36. Fortes MR, Reverter A, Zhang Y, Collis E, Nagaraj SH, Jonsson NN, Prayaga KC, Barris W,
759 Hawken RJ: **Association weight matrix for the genetic dissection of puberty in beef cattle.**
760 *Proceedings of the National Academy of Sciences* 2010, **107:**13642-13647.
761 37. Bolormaa S, Hayes B, Hawken R, Zhang Y, Reverter A, Goddard M: **Detection of chromosome**
762 **segments of zebu and taurine origin and their effect on beef production and growth.**
763 *Journal of animal science* 2011, **89:**2050-2060.
764 38. Bolormaa S, Pryce JE, Kemper KE, Hayes BJ, Zhang Y, Tier B, Barendse W, Reverter A,
765 Goddard ME: **Detection of quantitative trait loci in Bos indicus and Bos taurus cattle using**
766 **genome-wide association studies.** *Genetics Selection Evolution* 2013, **45:**1-12.
767 39. Fortes M, Kemper K, Sasazaki S, Reverter A, Pryce J, Barendse W, Bunch R, McCulloch R,
768 Harrison B, Bolormaa S: **Evidence for pleiotropism and recent selection in the PLAG 1**
769 **region in A ustralian B eef cattle.** *Animal genetics* 2013, **44:**636-647.
770 40. Juma AR, Damdimopoulou PE, Grommen SV, Van de Ven WJ, De Groef B: **Emerging role of**
771 **PLAG1 as a regulator of growth and reproduction.** *Journal of Endocrinology* 2016, **228:**R45-
772 R56.
773 41. Karim L, Takeda H, Lin L, Druet T, Arias JA, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart
774 B: **Variants modulating the expression of a chromosome domain encompassing PLAG1**
775 **influence bovine stature.** *Nature genetics* 2011, **43:**405-413.
776 42. Koufariotis L, Hayes B, Kelly M, Burns B, Lyons R, Stothard P, Chamberlain A, Moore S:
777 **Sequencing the mosaic genome of Brahman cattle identifies historic and recent**
778 **introgression including polled.** *Scientific reports* 2018, **8:**1-12.
779 43. Littlejohn M, Grala T, Sanders K, Walker C, Waghorn G, Macdonald K, Coppieters W, Georges
780 M, Spelman R, Hillerton E: **Genetic variation in PLAG1 associates with early life body weight**
781 **and peripubertal weight and growth in Bos taurus.** *Animal Genetics* 2012, **43:**591-594.
782 44. Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, Sugimoto Y,
783 Takasuga A: **Genome-wide association study identified three major QTL for carcass weight**
784 **including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle.** *BMC genetics* 2012,
785 **13:**1-11.

786 45. Paim TdP, Hay E, Wilson C, Thomas M, Kuehn L, Paiva S, McManus C, Blackburn H: **Dynamics**
787 **of genomic architecture during composite breed development in cattle.** *Animal genetics*
788 2020, **51:**224-234.
789 46. Mank JE: **Population genetics of sexual conflict in the genomic era.** *Nature Reviews Genetics*
790 2017, **18:**721-730.
791 47. Parsch J, Ellegren H: **The evolutionary causes and consequences of sex-biased gene**
792 **expression.** *Nature Reviews Genetics* 2013, **14:**83-87.
793 48. Sayadi A, Barrio AM, Immonen E, Dainat J, Berger D, Tellgren-Roth C, Nystedt B, Arnqvist G:
794 **The genomic footprint of sexual conflict.** *Nature ecology & evolution* 2019, **3:**1725-1730.
795 49. Bione S, Rizzolio F, Sala C, Ricotti R, Goegan M, Manzini M, Battaglia R, Marozzi A, Vegetti W,
796 Dalpra L: **Mutation analysis of two candidate genes for premature ovarian failure, DACH2**
797 **and POF1B.** *Human reproduction* 2004, **19:**2759-2766.
798 50. Jedidi I, Ouchari M, Yin Q: **Sex chromosomes-linked single-gene disorders involved in**
799 **human infertility.** *European journal of medical genetics* 2019, **62:**103560.
800 51. Okten G, Gunes S, Onat OE, Tukun A, Ozcelik T, Kocak I: **Disruption of HDX gene in**
801 **premature ovarian failure.** *Systems biology in reproductive medicine* 2013, **59:**218-222.
802 52. Weber TA, Koob S, Heide H, Wittig I, Head B, van der Bliek A, Brandt U, Mittelbronn M,
803 Reichert AS: **APOOL is a cardiolipin-binding constituent of the Mitofilin/MINOS protein**
804 **complex determining cristae morphology in mammalian mitochondria.** *PloS one* 2013,
805 **8:**e63683.
806 53. Kaneko H, Noguchi J, Kikuchi K, Akagi S, Shimada A, Taya K, Watanabe G, Hasegawa Y:
807 **Production and endocrine role of inhibin during the early development of bull calves.**
808 *Biology of reproduction* 2001, **65:**209-215.
809 54. Phillips DJ: **Activins, inhibins and follistatins in the large domestic species.** *Domestic animal*
810 *endocrinology* 2005, **28:**1-16.
811 55. Johnston D, Corbet N, Barwick S, Wolcott ML, Holroyd R: **Genetic correlations of young bull**
812 **reproductive traits and heifer puberty traits with female reproductive performance in two**
813 **tropical beef genotypes in northern Australia.** *Animal Production Science* 2014, **54:**74-84.
814 56. Corbet N, Burns B, Johnston D, Wolcott ML, Corbet D, Venus B, Li Y, McGowan M, Holroyd R:
815 **Male traits and herd reproductive capability in tropical beef cattle. 2. Genetic parameters**
816 **of bull traits.** *Animal Production Science* 2012, **53:**101-113.
817 57. Canal LB, Fontes PL, Sanford CD, Mercadante VR, DiLorenzo N, Lamb GC, Oosthuizen N:
818 **Relationships between feed efficiency and puberty in Bos taurus and Bos indicus-**
819 **influenced replacement beef heifers.** *Journal of Animal Science* 2020, **98:**skaa319.
820 58. Perry GA, Cushman R: **Effect of age at puberty/conception date on cow longevity.**
821 *Veterinary Clinics: Food Animal Practice* 2013, **29:**579-590.
822 59. Bunter KL, Cai W, Johnston DJ, Dekkers JC, Bunter K: **Selection to lower residual feed intake**
823 **in pigs produces a correlated response in juvenile insulin-like growth factor-I (IGF-1)**
824 **concentration.**
825 60. Gao X, Xu X-R, Ren H-Y, Zhang Y-H, Xu S-Z: **The effects of the GH, IGF-I and IGF-IBP3 gene on**
826 **growth and development traits of Nanyang cattle in different growth period.** *Yi Chuan=*
827 *Hereditas* 2006, **28:**927-932.
828 61. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, Tier B, Savin K, Hayes BJ,
829 Goddard ME: **A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for**
830 **stature, fatness and reproduction in beef cattle.** *PLoS genetics* 2014, **10:**e1004198.
831 62. Hackinger S, Zeggini E: **Statistical methods to detect pleiotropy in human complex traits.**
832 *Open biology* 2017, **7:**170125.
833 63. Id-Lahoucine S, Molina A, Cánovas A, Casellas J: **Screening for epistatic selection signatures:**
834 **A simulation study.** *Scientific reports* 2019, **9:**1-5.

835 64. Crowley J, Evans R, Mc Hugh N, Kenny D, McGee M, Crews Jr D, Berry D: **Genetic**
836 **relationships between feed efficiency in growing males and beef cow performance.** *Journal*
837 *of animal science* 2011, **89:**3372-3381.
838 65. Purfield DC, Evans RD, Berry DP: **Breed-and trait-specific associations define the genetic**
839 **architecture of calving performance traits in cattle.** *Journal of animal science* 2020,
840 **98:**skaa151.
841 66. Nguyen T-M, Shafi A, Nguyen T, Draghici S: **Identifying significantly impacted pathways: a**
842 **comprehensive review and assessment.** *Genome biology* 2019, **20:**1-15.
843 67. Bozza WP, Zhang Y, Hallett K, Rosado LAR, Zhang B: **RhoGDI deficiency induces constitutive**
844 **activation of Rho GTPases and COX-2 pathways in association with breast cancer**
845 **progression.** *Oncotarget* 2015, **6:**32723.
846 68. Hall A: **Rho family gtpases.** *Biochemical Society Transactions* 2012, **40:**1378-1382.
847 69. Humphries B, Wang Z, Yang C: **Rho GTPases: Big Players in Breast Cancer Initiation,**
848 **Metastasis and Therapeutic Responses.** *Cells* 2020, **9:**2167.
849 70. Vega FM, Ridley AJ: **Rho GTPases in cancer cell biology.** *FEBS letters* 2008, **582:**2093-2101.
850 71. Shibata S, Nagase M, Yoshida S, Kawarazaki W, Kurihara H, Tanaka H, Miyoshi J, Takai Y,
851 Fujita T: **Modification of mineralocorticoid receptor function by Rac1 GTPase: implication**
852 **in proteinuric kidney disease.** *Nature medicine* 2008, **14:**1370-1376.
853 72. Togawa A, Miyoshi J, Ishizaki H, Tanaka M, Takakura A, Nishioka H, Yoshida H, Doi T,
854 Mizoguchi A, Matsuura N: **Progressive impairment of kidneys and reproductive organs in**
855 **mice lacking Rho GDIα.** *Oncogene* 1999, **18:**5373-5380.
856 73. Ishizaki H, Togawa A, Tanaka-Okamoto M, Hori K, Nishimura M, Hamaguchi A, Imai T, Takai
857 Y, Miyoshi J: **Defective chemokine-directed lymphocyte migration and development in the**
858 **absence of Rho guanosine diphosphate-dissociation inhibitors α and β.** *The Journal of*
859 *Immunology* 2006, **177:**8512-8521.
860 74. Huta Y, Nitzan Y, Breitbart H: **Ezrin protects bovine spermatozoa from spontaneous**
861 **acrosome reaction.** *Theriogenology* 2020, **151:**119-127.
862 75. Shi Z-H, Zhao C, Wu H, Liu X-M: **Expression of RhoGDI alpha in human testes and sperm and**
863 **its correlation with the success rate of IVF.** *Zhonghua nan ke xue= National Journal of*
864 *Andrology* 2011, **17:**325-329.
865 76. Wang L, Chen W, Zhao C, Huo R, Guo X-J, Lin M, Huang X-Y, Mao Y-D, Zhou Z-M, Sha J-H: **The**
866 **role of ezrin-associated protein network in human sperm capacitation.** *Asian journal of*
867 *andrology* 2010, **12:**667.
868 77. Burns B, Fordyce G, Holroyd R: **A review of factors that impact on the capacity of beef**
869 **cattle females to conceive, maintain a pregnancy and wean a calf—Implications for**
870 **reproductive efficiency in northern Australia.** *Animal Reproduction Science* 2010, **122:**1-22.
871 78. McLean I, Holmes P, Counsell D: **Final Report: The Northern beef report, 2013 Northern**
872 **beef siutation analysis (B. COM. 0348).** *Meat and Livestock Australia, Sydney* 2013.
873 79. Chang AZ, Swain DL, Trotter MG: **Calf loss in northern Australia: a systematic review.** *The*
874 *Rangeland Journal* 2020, **42:**9-26.
875 80. Fordyce G: **Pregnancy rates achieved by mating bulls with different percentages of**
876 **morphologically normal sperm.** *Bullpower Delivery of adequate normal sperm to site of*
877 *fertilisation Project NAP3* 2005, **117:**142-151.
878 81. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, Rowan TN, Low WY, Zimin
879 A, Couldrey C: **De novo assembly of the cattle reference genome with single-molecule**
880 **sequencing.** *GigaScience* 2020, **9:**giaa021.
881 82. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee D: **BLUPF90 and related programs**
882 **(BGF90).** In *Proceedings of the 7th world congress on genetics applied to livestock*
883 *production*. 2002: 743-744.
884 83. Strandén I, Garrick D: **Derivation of equivalent computing algorithms for genomic**
885 **predictions and reliabilities of animal merit.** *Journal of dairy science* 2009, **92:**2971-2975.

886 84. Aguilar I, Legarra A, Cardoso F, Masuda Y, Lourenco D, Misztal I: **Frequentist p-values for**
887 **large-scale-single step genome-wide association, with an application to birth weight in**
888 **American Angus cattle.** *Genetics Selection Evolution* 2019, **51**:1-8.
889 85. VanRaden PM: **Efficient methods to compute genomic predictions.** *Journal of dairy science*
890 2008, **91**:4414-4423.
891 86. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC,
892 Martin NG, Montgomery GW: **Common SNPs explain a large proportion of the heritability**
893 **for human height.** *Nature genetics* 2010, **42**:565-569.
894 87. Bozeman M: **Golden Helix, Inc. SNP & Variation Suite™[Software]. [(Version 8.x)].**

895 **Available from http://www.goldenhelix.com.**
896 88. Hu Z-L, Park CA, Reecy JM: **Developmental progress and current status of the Animal**
897 **QTLdb.** *Nucleic acids research* 2016, **44**:D827-D833.
898 89. Hu Z-L, Park CA, Wu X-L, Reecy JM: **Animal QTLdb: an improved database tool for livestock**
899 **animal QTL/association data dissemination in the post-genome era.** *Nucleic acids research*
900 2013, **41**:D871-D879.
901 90. Lam S, Miglior F, Fonseca P, Gómez-Redondo I, Zeidan J, Suárez-Vega A, Schenkel F, Guan L,
902 Waters S, Stothard P: **Identification of functional candidate variants and genes for feed**
903 **efficiency in Holstein and Jersey cattle breeds using RNA-sequencing.** *Journal of dairy*
904 *science* 2021, **104**:1928-1950.
905 91. Sweett H, Fonseca P, Suarez-Vega A, Livernois A, Miglior F, Cánovas A: **Genome-wide**
906 **association study to identify genomic regions and positional candidate genes associated**
907 **with male fertility in beef cattle.** *Scientific reports* 2020, **10**:1-14.
908 92. Krämer A, Green J, Pollard Jr J, Tugendreich S: **Causal analysis approaches in ingenuity**
909 **pathway analysis.** *Bioinformatics* 2014, **30**:523-530.
910 93. Medici V, Kieffer DA, Shibata NM, Chima H, Kim K, Canovas A, Medrano JF, Islas-Trejo AD,
911 Kharbanda KK, Olson K: **Wilson Disease: Epigenetic effects of choline supplementation on**
912 **phenotype and clinical course in a mouse model.** *Epigenetics* 2016, **11**:804-818.
913 94. Cardoso TF, Quintanilla R, Castelló A, González-Prendes R, Amills M, Cánovas Á: **Differential**
914 **expression of mRNA isoforms in the skeletal muscle of pigs with distinct growth and**
915 **fatness profiles.** *BMC genomics* 2018, **19**:1-12.
916 95. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, Huber W: **BioMart and**
917 **Bioconductor: a powerful link between biological databases and microarray data analysis.**
918 *Bioinformatics* 2005, **21**:3439-3440.