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

Although the genetic correlation between complex traits have been estimated for more than a century, only recently we have started to map and understand the precise localization of the genomic region(s) that underpin these correlations. Reproductive traits are often genetically correlated, and yet we don't fully understand the complexities, synergism, or trade-offs between male and female fertility. In this study, we used reproductive traits in two cattle populations to develop a novel framework termed correlation scan. This framework was used to identify regions associated with the genetic correlations between male and female fertility traits across the bovine genome. The traits used were age at first *corpus luteum* (AGECL) and serum levels of insulin growth hormone (IGF1 measured in bulls, IGF1b, or cows, IGF1c). The methodology developed herein used correlations of 500-SNP (single nucleotide polymorphism) effects in a 100-SNPs sliding window in each chromosome to identify regions in the genome that either drive (i.e., SNP effects on the same direction) or antagonize (i.e., SNP effects in the opposite direction) the genetic correlations between traits. We used a permutation test to confirm which regions of the genome harboured significant correlations. Hence, this framework can also identify neutral genomic regions with no effect on the pairwise trait studied. 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.

Keywords: Genomic correlation, drivers, antagonizing, RHOGDI, pathway analysis, QTLs

Author summary

In animal breeding, it is often common to estimate genetic correlations between economically important traits. These estimated correlations represent the average of the shared genetic similarities between traits across the genome. Despite this knowledge, we are yet to uncover the regions in the genome that explain the genetic correlations estimated. Targeting reproductive traits in cattle, we developed a new framework and used it to identify multiple regions across the genome that affect genetic correlations between male and female fertility traits. While some regions have no effect on these trait correlations, other loci drive or antagonize these relationships. We further subjected the identified regions to functional analysis and annotation for biological insights. Although the methodology was applied to cattle phenotypes, using high-density SNP genotypes, the general framework can be applied

to any species or traits. For example, the method could be used to identify genomic regions that explain the interplay between various mental illness phenotypes in humans.

Introduction

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) (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 S1 Table. 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.

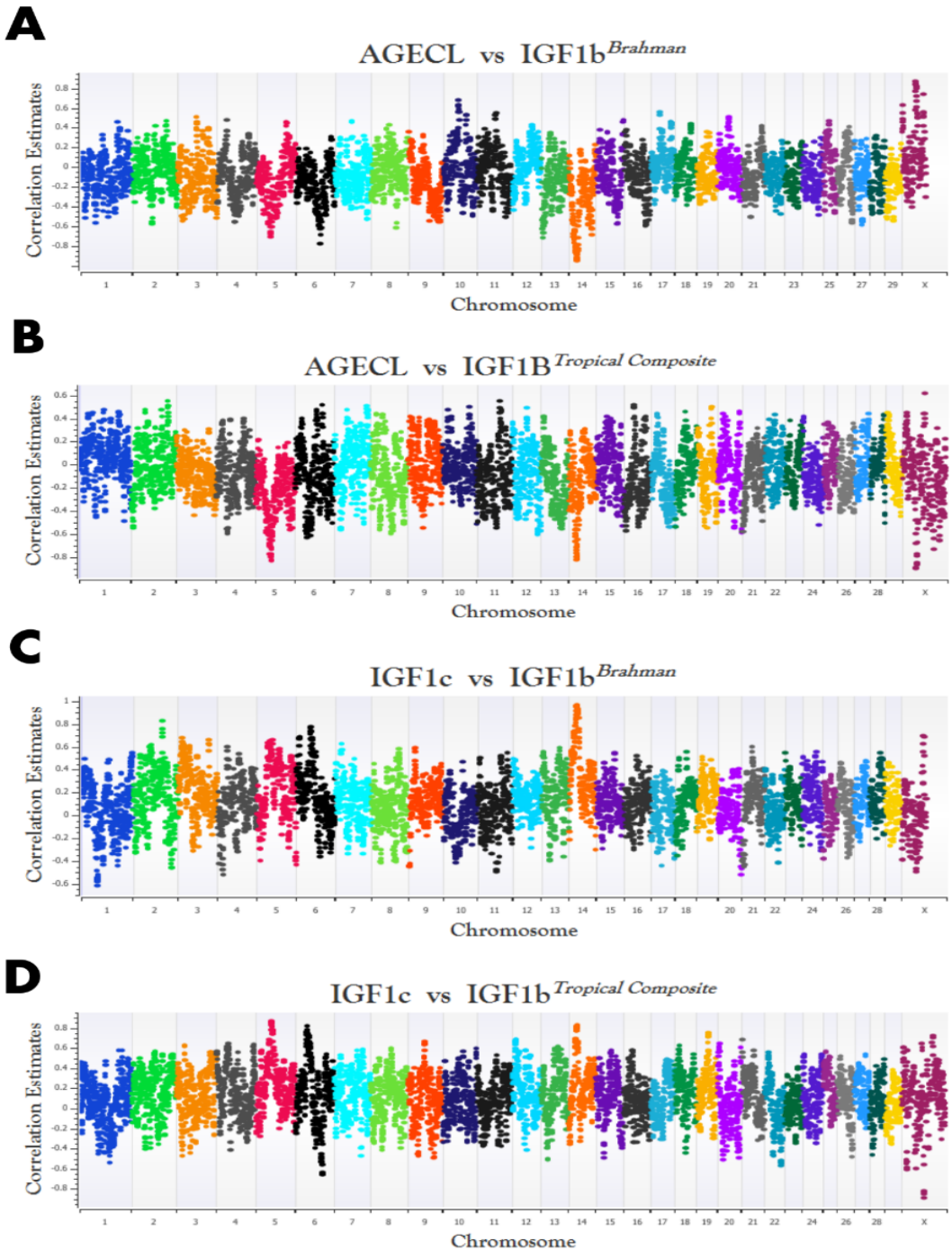


Figure 1. Genome plot of the regions driving and antagonizing trait correlations in Brahman (BB) and Tropical Composite (TC) for the pairwise traits (BB-AGECL vs IGF1b; **A**, TC-AGECL vs IGF1b; **B**, BB-IGF1c vs IGF1b; **C**, TC- IGF1c vs IGF1b; **D**). **AGECL**, age at first corpus; **IGF1**, serum 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.

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

134 In order to identify drivers, antagonizing and neutral regions across the bovine genome, we
135 performed permutation test by randomly reshuffling the Single Nucleotide Polymorphisms
136 (SNPs) effects in each chromosome in 1,000 iterations for each trait. Then, we applied our
137 framework on the randomized SNP effects and observed the **r** estimate across each iteration
138 for each window. In most cases, the maximum and the minimum **r** estimates for each window
139 at each iteration (i.e., rand 1 to rand 1,000) range between ±0.20. Therefore, we considered
140 neutral windows with no significant effect on the trait correlation as windows with -0.20≤ **r**
141 ≤0.20 estimates. The genomic plots of the **r** estimates resulting from the permutation test
142 (rand 500 only) in each population are presented in S1 Figure. S2 Table shows the numbers
143 of windows, their chromosome coordinates, and **r** estimates as well as the maximum and
144 minimum r estimates across the 1,000 iterations for each pair of traits in each of the two
145 populations.

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

152 The number of significant driver windows for the correlation between AGECL and IGF1b
153 was 1,636 in BB and 1,914 in TC cattle. The number of significant windows for the
154 antagonizing was 547 in BB and 898 in TC cattle, for AGECL vs IGF1b. For the correlation
155 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 S3 Table.

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

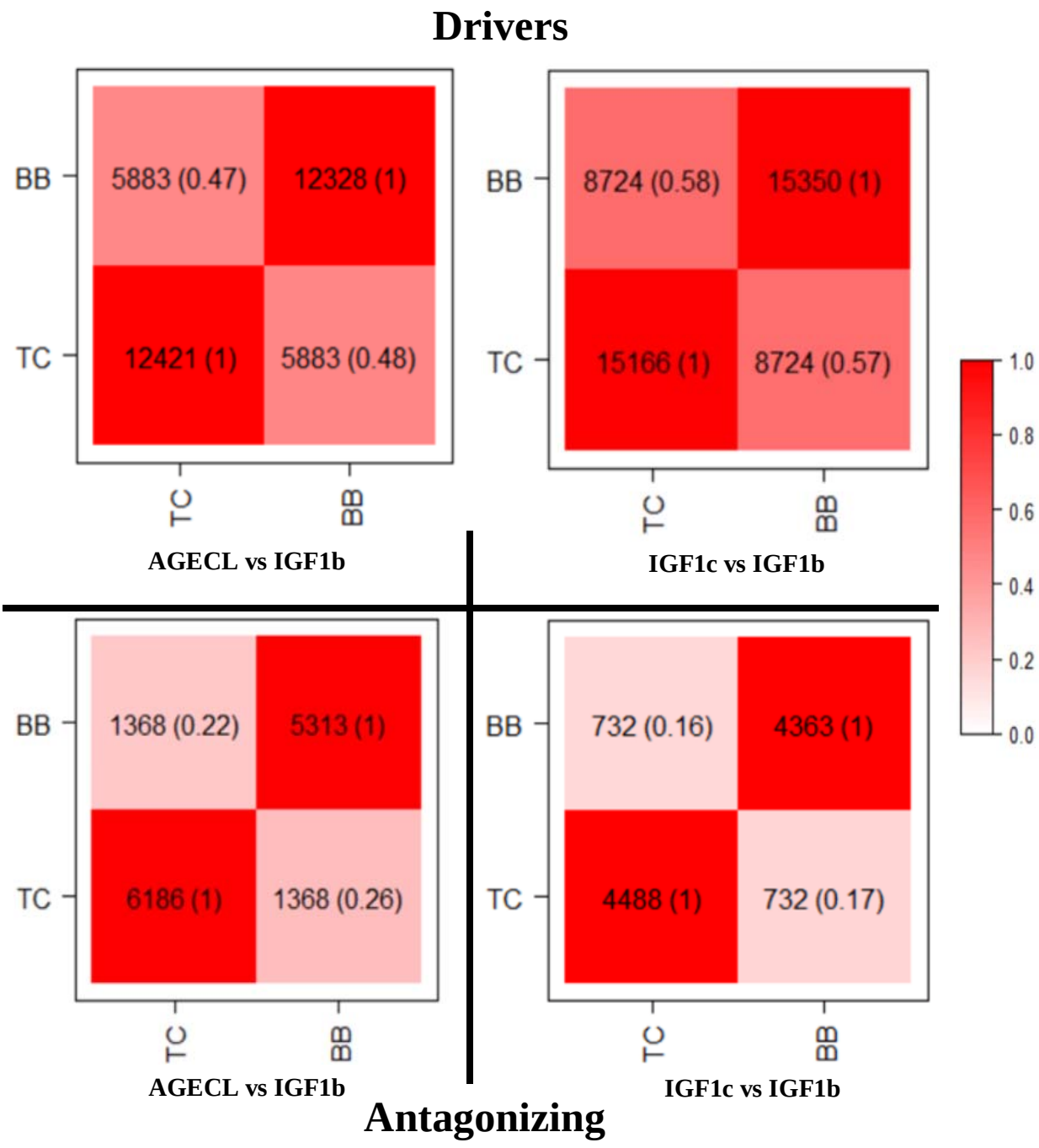
Defining driver and antagonizing regions separately for each pair of traits, allowed us to identify the genes and QTLs within these regions for each of the two beef cattle populations. The percentage of the overlapping genes (Figure 2a) and QTLs (Figure 2b) across both populations was studied. The percentages 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.

201



202

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

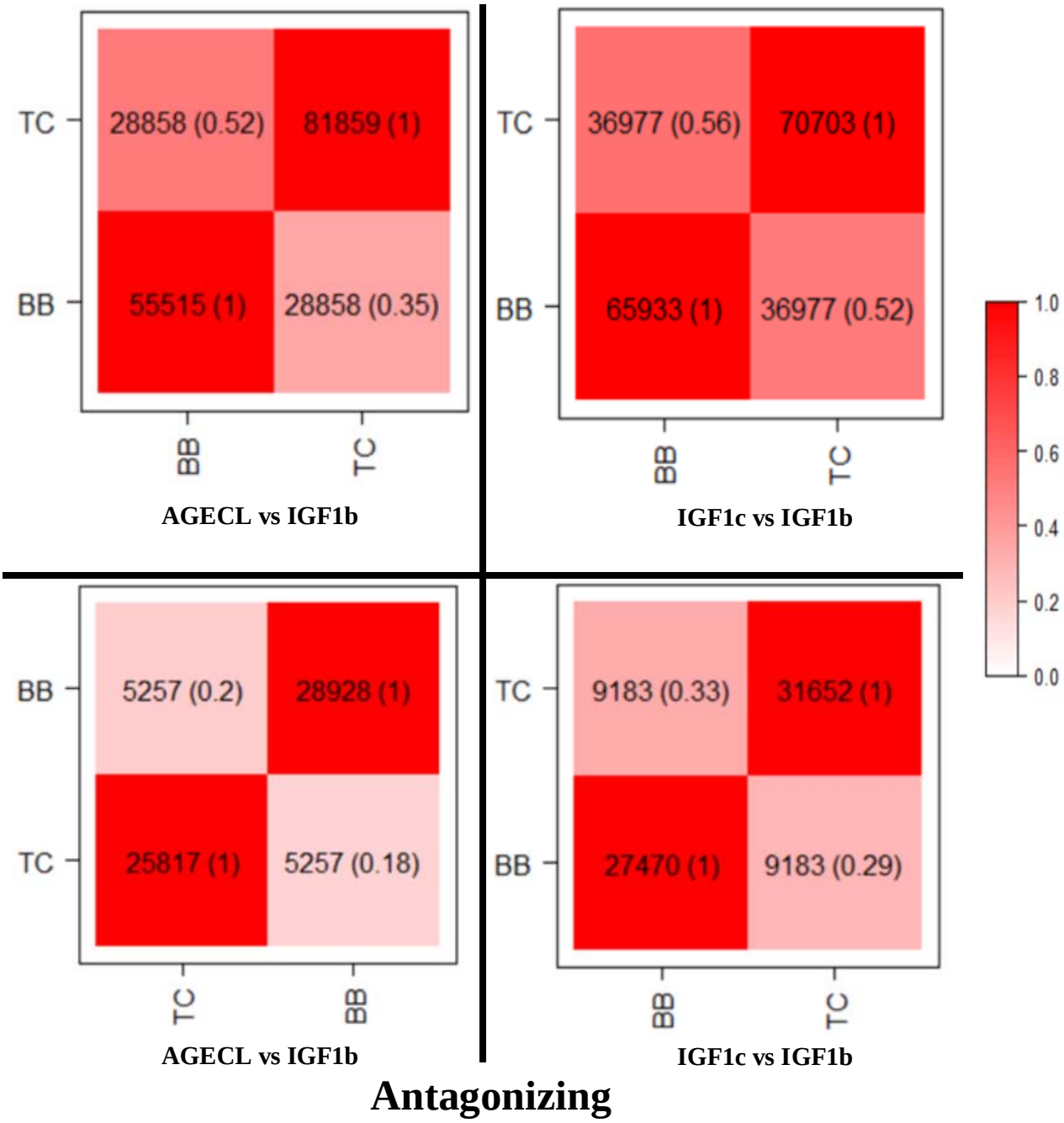
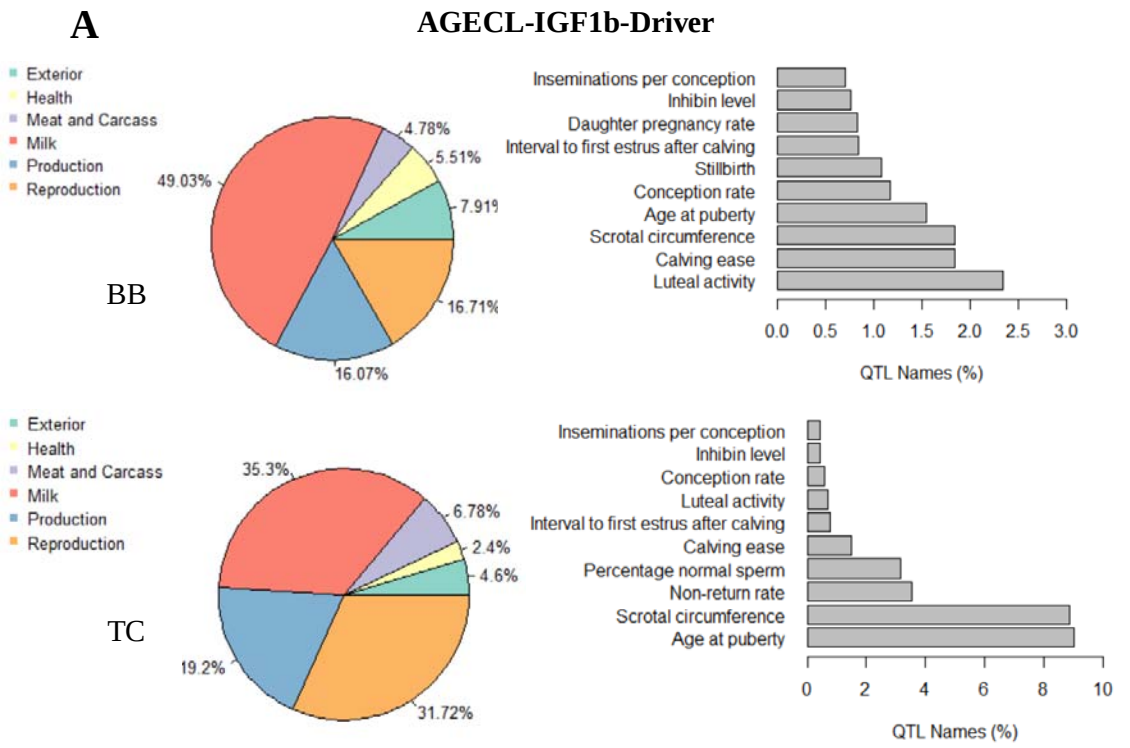


Figure 2a. QTLs annotated in the significant (i.e., driver and antagonizing) genomic regions identified as explaining the genetic correlations between male and female fertility traits in Brahman (BB) and Tropical Composite (TC) population. The overlaps between the two studied populations are in the diagonal 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.



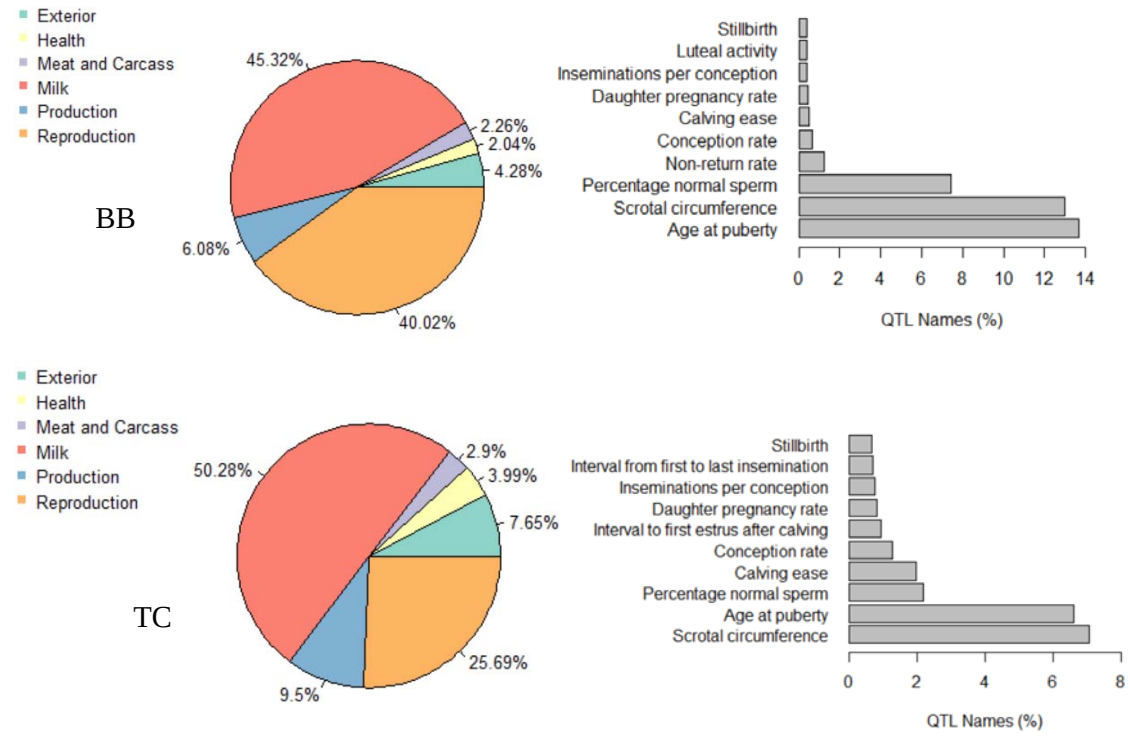
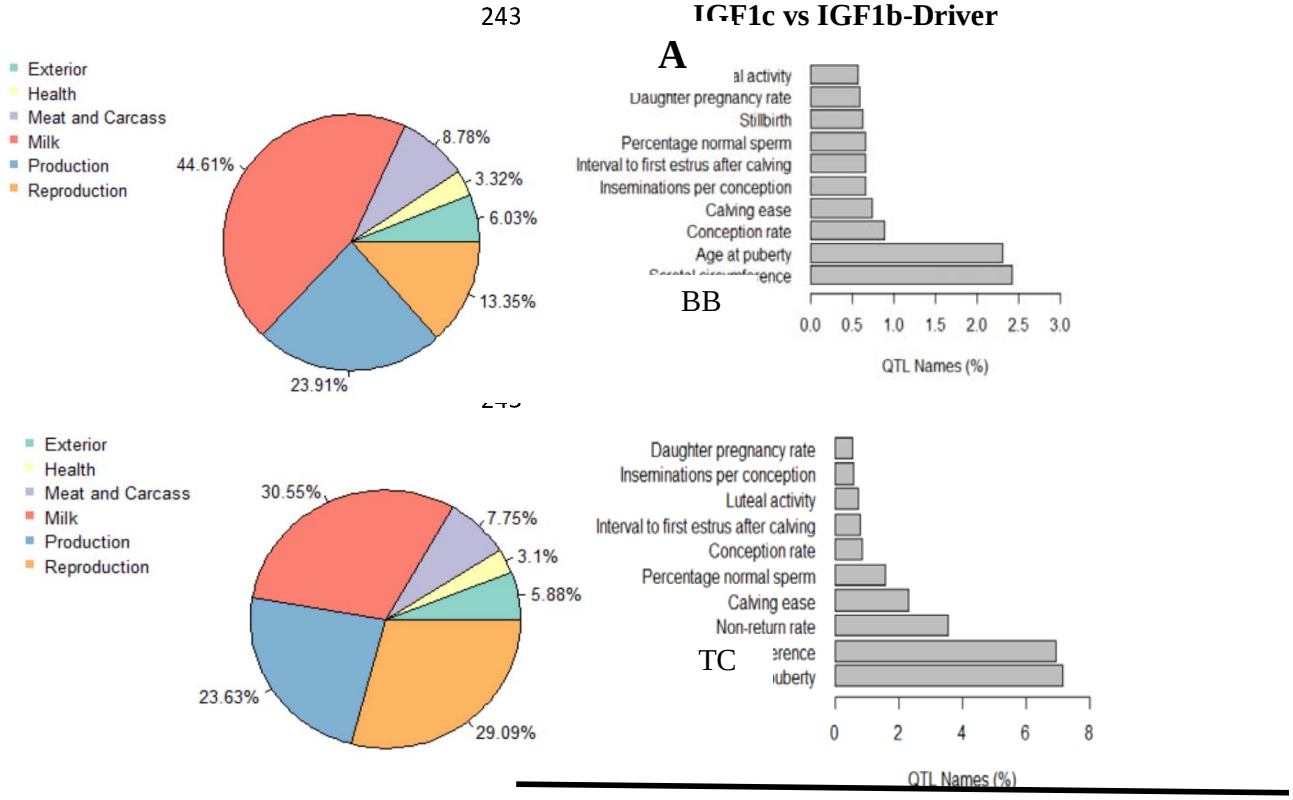


Figure 3a. Percentage of QTL type (pie chart) and trait related to reproduction QTLs (barplots) for the QTL annotation results obtained for (A) AGECL vs IGF1b - driver, (B) AGECL vs IGF1b- antagonizing in Brahman (BB) and Tropical Composite (TC) population. **AGECL**, age at first corpus luteum, **IGF1c**,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



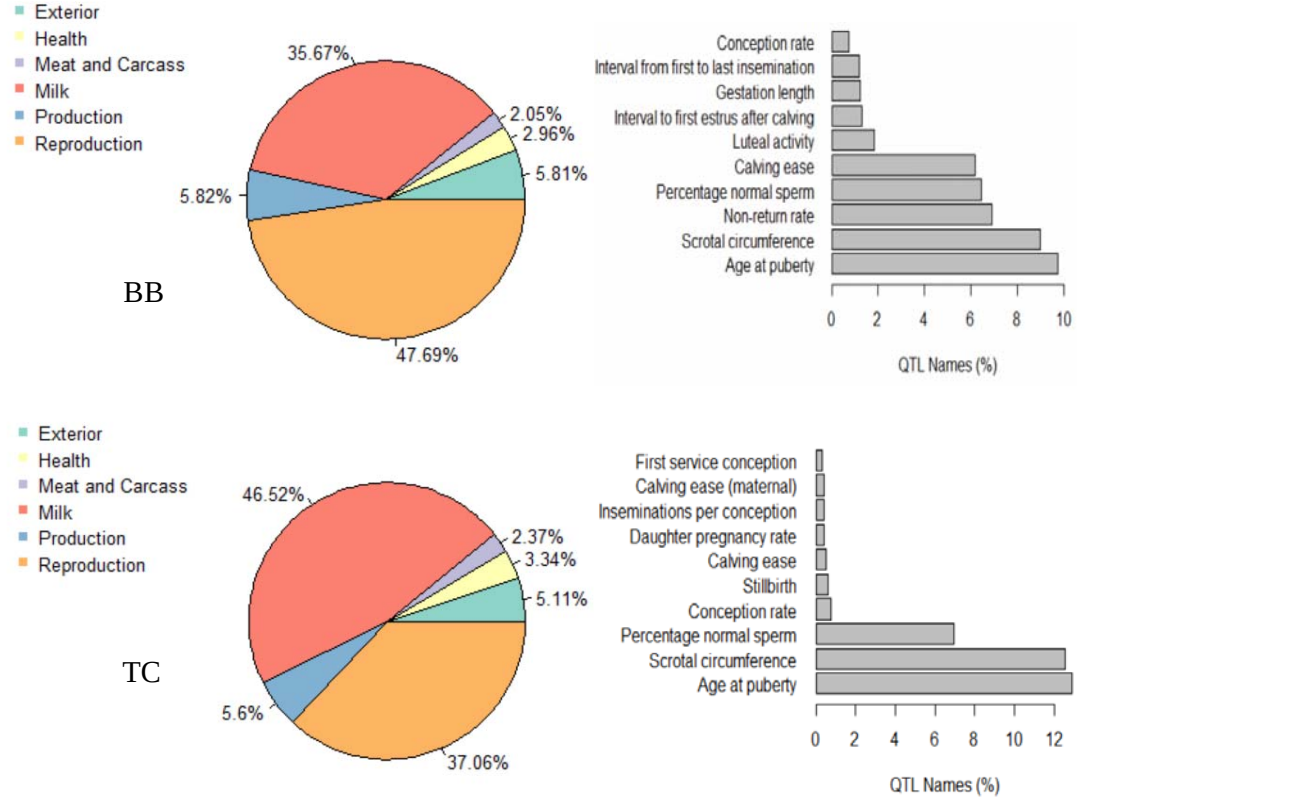


Figure 3b. Percentage of QTL type (pie chart) and trait related to reproduction QTLs (barplots) for the QTL annotation obtained for **(A)** IGF1c vs IGF1b- driver, (B) IGF1c vs IGF1b- antagonizing in Brahman (BB) and Tropical Composite (TC) population. **AGECL**, age at first corpus luteum, **IGF1**,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

255 **QTL enrichment analysis**

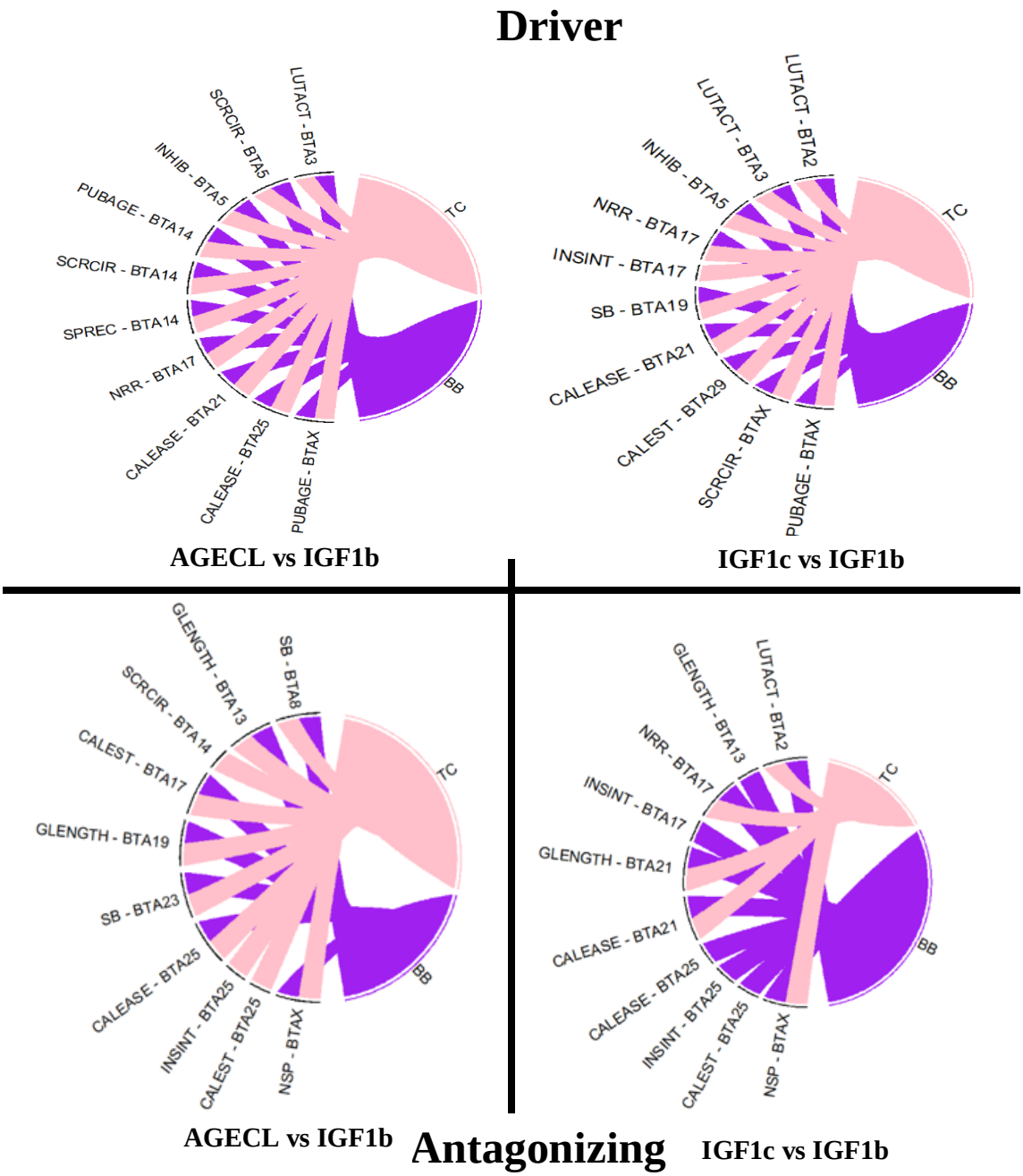
256 We performed a chromosome-wide QTL enrichment analysis to further test the significance
257 of the QTLs identified for all the driver and antagonizing regions in each cattle population,
258 for each trait pair using GALLO (27). Enriched QTLs for the studied traits span across most
259 QTL types, indicating the presence of complex genetic mechanisms. The results of the
260 chromosome-wide QTLs enrichment (FDR-corrected p-value≤0.05) for the driver and
261 antagonizing regions for all pairwise traits in each population are presented in S4 Table.

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

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

275 To identify the common results and shared biology between the driver and the antagonizing
276 regions, we also investigated the overlaps of the QTL types associated with the studied trait
277 (i.e., reproduction) in the two populations. The relationship between the top 10 enriched
278 reproductive QTLs in BB and TC are presented in Figure 4a and 4b. Irrespective of the trait

279 pair, for the driver regions, the reproductive QTLs in BB in most cases overlap with those
280 identified in TC. However, for the antagonizing regions, not all reproductive QTLs in BB
281 were found in TC beef cattle population.



282
283
284 **Figure 4.** Chord plot showing the relationship between the top 10 enriched reproductive QTLs
285 between Brahman (BB) and Tropical Composite (TC) for the driver (top) and the antagonizing
286 (bottom) regions of the studied traits. **AGECL**, age at first corpus luteum, **IGF1c**, serum levels
287 of insulin growth hormone measured in cow; IGF1b. LUTACT, Luteal activity; SCRCIR,
288 Scrotal circumference; INHIB, Inhibin level; PUBAGE, Age at puberty; SPREC, Sexual

289 precocity; NRR, Non-return rate; CALEASE, Calving ease; INSINT, Interval from first to
290 last insemination; SB, Still birth; CALEST, Interval to first estrus after calving; GLENGTH,
291 Gestation length, NSP, Percentage normal sperm.
292

293 **Functional enrichment analysis**

294 Leveraging our methodology's directionality of gene effects with Ingenuity Pathway
295 Analysis (IPA; http://www.ingenuity.com), we identified the enriched canonical metabolic
296 pathways enriched at Benjamini–Hochberg corrected p-values (BH-*P*-value) of $p<0.01$. The
297 graphical presentation of the canonical metabolic pathways predicted by IPA to be enriched
298 and the proportion of driver and antagonizing genes in each pathway for all pairwise traits
299 investigated in each population are illustrated in S2A-H Figure**.** Although IPA provided
300 information about whether the predicted pathways were being activated or inhibited based on
301 our data, we remain cautious when interpreting our results since the **r** estimates are not the
302 same as gene expression values, and IPA was originally designed to mine gene expression
303 data.

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

312 For IGF1c vs IGF1b in BB, the top 5 enriched canonical metabolic pathways were cardiac
313 hypertrophy signaling, CREB signaling in neurons, thrombin signaling, estrogen receptor
314 signaling, opioid signaling pathway and AMPK signaling. In TC, the top 5 enriched pathways

315 were CREB signaling in neurons, cardiac hypertrophy signaling (enhanced), opioid signaling
316 pathway, gαs signaling and breast cancer regulation by stathmin1. The enriched canonical
317 metabolic pathways and all the genes involved in each pathway are available in S5 Table.

318

319 **Discussion**

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

364 BB and TC are likely to impact the results of our analyses. Breed differences are expected,
365 and so when two breeds share a similar result, it enhances our confidence in calling
366 significant windows for the interplay between male and female fertility traits.

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

376 In animal production, more research is carried out on milk production-related traits, thereby
377 creating large proportion of records for these traits in the cattle QTL database. These volumes
378 of records can create a bias in the QTLs representativeness (27). The QTL enrichment
379 analysis allows testing the significance of the QTL representative using chromosome-wide
380 approach to detect specific genomic region with many QTLs for a specific trait. For example,
381 taking the driver regions for AGECL vs IGF1b in BB, the top enriched QTLs was found in
382 BTA5, harbouring 36 QTLs. These QTLs comprised 8 different QTLs for reproduction
383 (inhibin level, scrotal circumference, interval of first estrus after calving, gestation length,
384 insemination per conception, conception rate, daughter pregnancy rate, and pregnancy rate).
385 These 8-traits listed here have been found to be correlated with puberty (studied traits) in
386 cattle. For instance, inhibin is regarded as a biomarker for sexual development because it
387 regulates spermatogenesis in both beef and dairy bulls (53, 54). Moderate genetic correlation
388 was found between inhibin and AGECL (55) and between inhibin and IGF1b (56) in BB.

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

408 Another interesting result from this study is the shared biology between the two breeds
409 relative to the traits under study. Despite breed differences, the enriched reproductive QTLs
410 driving the genetic correlations between male and female fertility are the same for the two
411 cattle populations (Figure 4). Most of the enriched QTLs are related to reproductive traits
412 measured early life. A possible explanation could be that the reproductive phenotypes shared
413 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 *et al.* (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.

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 IGF1b- 964	554K	-0.65 (0.13)
IGF1c vs IGF1b	IGF1c- 995 IGF1b- 964	554K	0.86 (0.11)
		Tropical Composite	

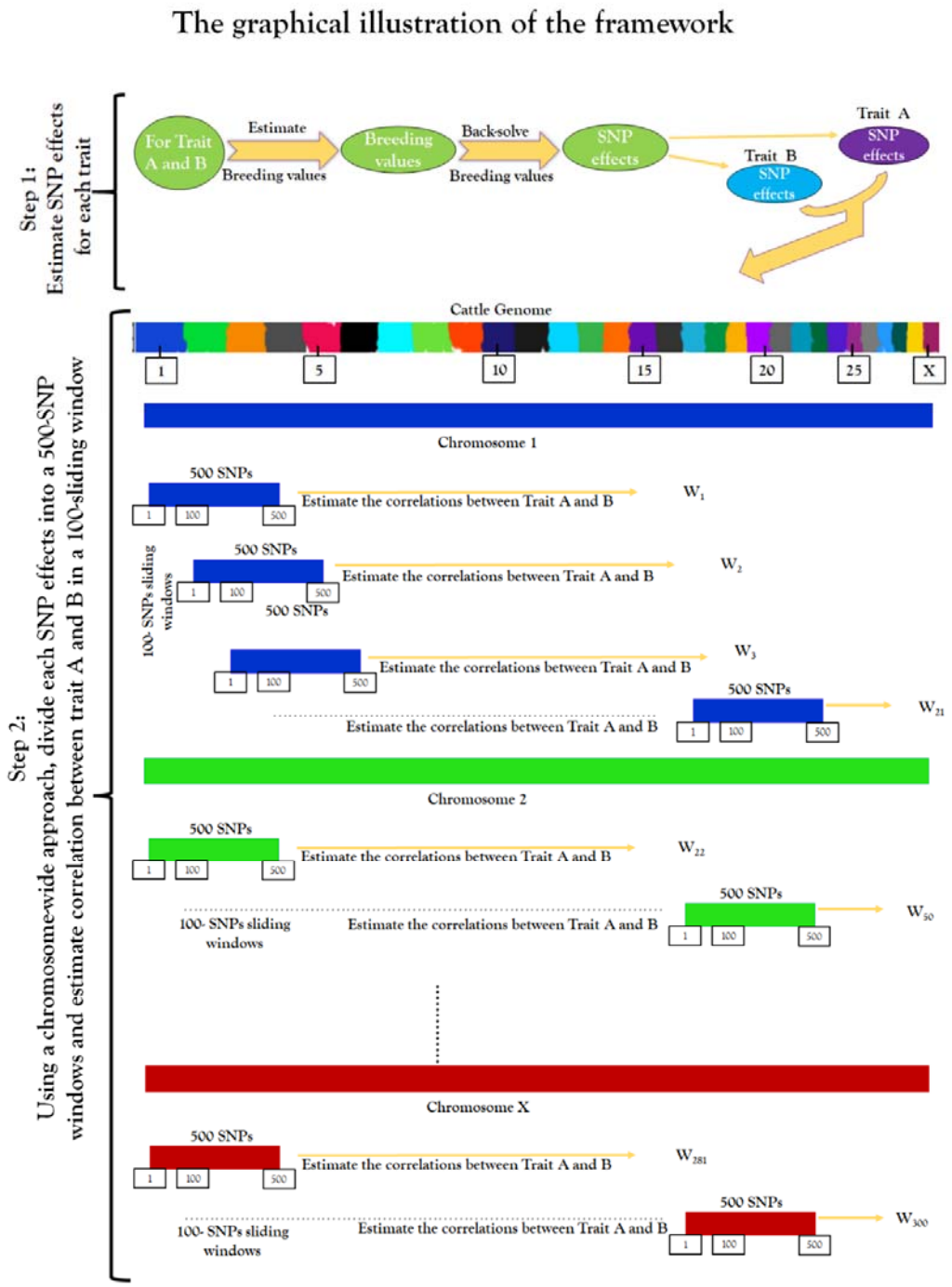
AGECL vs IGF1b	AGECL-996 IGF1b- 998	686K	-0.55 (0.14)
IGF1b vs IGF1b	IGF1c- 1015 IGF1b- 998	686K	0.93 (0.11)

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 and then back-solved these gEBVs to obtain SNP effects for all chromosomes using GCTA (83). 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 (84).

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.



512

513 **Figure 4.** The graphical illustration of the sliding window framework. The framework
514 involves 2 steps. Step 1 start from the estimation of genomic breeding values to the
515 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 chromosome 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; (85, 86)), 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 (87, 88). 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

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

570 ***Functional enrichment analysis***

571 The annotated genes along with their corresponding **r** estimates and rank values for the
572 significant driver and antagonizing windows for each pairwise trait in BB and TC populations
573 were subjected to enrichment analysis using the commercial QIAGEN's Ingenuity Pathway
574 Analysis (IPA; v.8.8, http://www.ingenuity.com). The IPA allows identifying
575 overrepresented biological mechanism, metabolic pathways, and diseases and biological
576 functions that are highly relevant to the traits of interest using the directionality of the
577 submitted gene list (89, 90). The outcome of our methodology indicates that genes within
578 each window come with their directionalities, in this case, r estimates. Thus, we leveraged on
579 the directionality of each gene by allowing the driver genes to be upregulated and
580 antagonizing genes to be downregulated.

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

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

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

598 The pathway analysis was conducted using the "Core Analysis" function implemented within
599 IPA. In this analysis, associations were calculated using direct and indirect relationships
600 among the gene lists. At first, the gene lists were mapped to human gene data. Genes without
601 an associated gene symbol or gene annotation were subjected to an annotation by homology
602 using BioMart application available in the Ensembl database
603 (http://www.ensembl.org/biomart/martview/) (91, 92). With this approach, we only
604 considered non-annotated genes with percentage of identity ≥80% with human homolog. The
605 final datasets used for the IPA analyses are presented in S6 Table. Finally, the "Core
606 Analysis" was used to identify canonical metabolic pathways enriched at Benjamini–
607 Hochberg corrected p-values (B-H-P-value) of $p<0.01$).

608 **Supporting information**

609 **S1 Table. The number of windows, chromosome number, chromosome coordinates, and**
610 **correlation estimates for each window for the two pairwise trait in Brahman and**
611 **Tropical Composite population (XLSX).**

S2 Table. 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 (XLSX).

S3 Table. 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) population for AGECL vs IGF1b (XLSX).

S4 Table. 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 (XLSX).

S5 Table. 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 (XLSX).

S6 Table. The final dataset used for Ingenuity Pathway Analysis (IPA) for all trait pairwise in Brahman and Tropical Composite cattle (XLSX).

S1 Figure. 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 (DOC).

S2A-H Figure. 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 (PDF).

Data availability

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).

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, et al. Epigenomics and genotype-phenotype association analyses reveal conserved genetic architecture of complex traits in cattle and human. BMC biology. 2020;18(1):1-16.
2. Zhang Y, Lu Q, Ye Y, Huang K, Liu W, Wu Y, et al. 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, et al. Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. PloS one. 2014;9(7):e102551.
4. Fonseca PAdS, Id-Lahoucine S, Reverter A, Medrano JF, Fortes MS, Casellas J, et al. Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. PLoS One. 2018;13(10):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(10):567-81.

7. Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. Nature genetics. 2020;52(8):859-64.
8. Mackay TF, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics. 2009;10(8):565-77.
9. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nature genetics. 2016;48(7):709-17.
10. Price AL, Spencer CC, Donnelly P. Progress and promise in understanding the genetic basis of common diseases. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1821):20151684.
11. Lunstra D, Cundiff L. Growth and pubertal development in brahman-, boran-, tuli-, belgian blue-, hereford-and angus-sired f1 bulls. Journal of Animal Science. 2003;81(6):1414-26.
12. Fang L, Sahana G, Ma P, Su G, Yu Y, Zhang S, et al. Use of biological priors enhances understanding of genetic architecture and genomic prediction of complex traits within and between dairy cattle breeds. BMC genomics. 2017;18(1):1-12.
13. Neyhart JL, Lorenz AJ, Smith KP. Multi-trait improvement by predicting genetic correlations in breeding crosses. G3: Genes, Genomes, Genetics. 2019;9(10):3153-65.
14. Grotzinger AD, Rhemtulla M, de Vlaming R, Ritchie SJ, Mallard TT, Hill WD, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nature human behaviour. 2019;3(5):513-25.
15. Henderson C. Recent developments in variance and covariance estimations. Journal of Animal Science. 1986;63(1):208-16.
16. Lee SH, Van der Werf JH. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics. 2016;32(9):1420-2.
17. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nature genetics. 2018;50(2):229-37.
18. Sodini SM, Kemper KE, Wray NR, Trzaskowski M. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. Genetics. 2018;209(3):941-8.
19. Collet JM, Fuentes S, Hesketh J, Hill MS, Innocenti P, Morrow EH, et al. Rapid evolution of the intersexual genetic correlation for fitness in Drosophila melanogaster. Evolution. 2016;70(4):781-95.
20. Connallon T, Matthews G. Cross-sex genetic correlations for fitness and fitness components: connecting theoretical predictions to empirical patterns. Evolution letters. 2019;3(3):254-62.
21. Raidan FS, Porto-Neto LR, Reverter A. Across-sex genomic-assisted genetic correlations for sex-influenced traits in Brahman cattle. Genetics Selection Evolution. 2019;51(1):41.
22. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics. 2015;47(3):291-5.
23. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local genetic correlation gives insights into the shared genetic architecture of complex traits. The American Journal of Human Genetics. 2017;101(5):737-51.
24. Werme J, van der Sluis S, Posthuma D, de Leeuw C. LAVA: An integrated framework for local genetic correlation analysis. bioRxiv. 2021:2020.12. 31.424652.
25. Zhang Y, Lu Q, Ye Y, Huang K, Liu W, Wu Y, et al. SUPERGNOVA: local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits. Genome biology. 2021;22(1):1-30.
26. Olasege B, Tahir M, Gouveia G, Kour J, Porto-Neto L, Hayes B, et al. Genetic parameter estimates for male and female fertility traits using genomic data to improve fertility in Australian beef cattle.
27. Fonseca PA, Suárez-Vega A, Marras G, Cánovas Á. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience. 2020;9(12):giaa149.

715 28. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering
716 of regulatory networks in human B cells. Nature genetics. 2005;37(4):382-90.
717 29. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to
718 omnigenic. Cell. 2017;169(7):1177-86.
719 30. Chen B-S, Yang S-K, Lan C-Y, Chuang Y-J. A systems biology approach to construct the gene
720 regulatory network of systemic inflammation via microarray and databases mining. BMC Medical
721 Genomics. 2008;1(1):1-22.
722 31. Satokangas I, Martin S, Helanterä H, Saramäki J, Kulmuni J. Multi-locus interactions and the
723 build-up of reproductive isolation. Philosophical Transactions of the Royal Society B.
724 2020;375(1806):20190543.
725 32. Skelly DA, Raghupathy N, Robledo RF, Graber JH, Chesler EJ. Reference trait analysis reveals
726 correlations between gene expression and quantitative traits in disjoint samples. Genetics.
727 2019;212(3):919-29.
728 33. Antonarakis SE, Chakravarti A, Cohen JC, Hardy J. Mendelian disorders and multifactorial
729 traits: the big divide or one for all? Nature Reviews Genetics. 2010;11(5):380-4.
730 34. Goddard M, Kemper K, MacLeod I, Chamberlain A, Hayes B. Genetics of complex traits:
731 prediction of phenotype, identification of causal polymorphisms and genetic architecture.
732 Proceedings of the Royal Society B: Biological Sciences. 2016;283(1835):20160569.
733 35. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery,
734 estimation and prediction analysis of complex traits using a Bayesian mixture model. PLoS genetics.
735 2015;11(4):e1004969.
736 36. Fortes MR, Reverter A, Zhang Y, Collis E, Nagaraj SH, Jonsson NN, et al. Association weight
737 matrix for the genetic dissection of puberty in beef cattle. Proceedings of the National Academy of
738 Sciences. 2010;107(31):13642-7.
739 37. Bolormaa S, Hayes B, Hawken R, Zhang Y, Reverter A, Goddard M. Detection of chromosome
740 segments of zebu and taurine origin and their effect on beef production and growth. Journal of
741 animal science. 2011;89(7):2050-60.
742 38. Bolormaa S, Pryce JE, Kemper KE, Hayes BJ, Zhang Y, Tier B, et al. Detection of quantitative
743 trait loci in Bos indicus and Bos taurus cattle using genome-wide association studies. Genetics
744 Selection Evolution. 2013;45(1):1-12.
745 39. Fortes M, Kemper K, Sasazaki S, Reverter A, Pryce J, Barendse W, et al. Evidence for
746 pleiotropism and recent selection in the PLAG 1 region in A ustralian B eef cattle. Animal genetics.
747 2013;44(6):636-47.
748 40. Juma AR, Damdimopoulou PE, Grommen SV, Van de Ven WJ, De Groef B. Emerging role of
749 PLAG1 as a regulator of growth and reproduction. Journal of Endocrinology. 2016;228(2):R45-R56.
750 41. Karim L, Takeda H, Lin L, Druet T, Arias JA, Baurain D, et al. Variants modulating the
751 expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nature
752 genetics. 2011;43(5):405-13.
753 42. Koufariotis L, Hayes B, Kelly M, Burns B, Lyons R, Stothard P, et al. Sequencing the mosaic
754 genome of Brahman cattle identifies historic and recent introgression including polled. Scientific
755 reports. 2018;8(1):1-12.
756 43. Littlejohn M, Grala T, Sanders K, Walker C, Waghorn G, Macdonald K, et al. Genetic variation
757 in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus.
758 Animal Genetics. 2012;43(5):591-4.
759 44. Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, et al. Genome-wide
760 association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN
761 for stature in Japanese Black cattle. BMC genetics. 2012;13(1):1-11.
762 45. Paim TdP, Hay E, Wilson C, Thomas M, Kuehn L, Paiva S, et al. Dynamics of genomic
763 architecture during composite breed development in cattle. Animal genetics. 2020;51(2):224-34.
764 46. Mank JE. Population genetics of sexual conflict in the genomic era. Nature Reviews Genetics.
765 2017;18(12):721-30.

766 47. Parsch J, Ellegren H. The evolutionary causes and consequences of sex-biased gene
767 expression. Nature Reviews Genetics. 2013;14(2):83-7.
768 48. Sayadi A, Barrio AM, Immonen E, Dainat J, Berger D, Tellgren-Roth C, et al. The genomic
769 footprint of sexual conflict. Nature ecology & evolution. 2019;3(12):1725-30.
770 49. Bione S, Rizzolio F, Sala C, Ricotti R, Goegan M, Manzini M, et al. Mutation analysis of two
771 candidate genes for premature ovarian failure, DACH2 and POF1B. Human reproduction.
772 2004;19(12):2759-66.
773 50. Jedidi I, Ouchari M, Yin Q. Sex chromosomes-linked single-gene disorders involved in human
774 infertility. European journal of medical genetics. 2019;62(9):103560.
775 51. Okten G, Gunes S, Onat OE, Tukun A, Ozcelik T, Kocak I. Disruption of HDX gene in premature
776 ovarian failure. Systems biology in reproductive medicine. 2013;59(4):218-22.
777 52. Weber TA, Koob S, Heide H, Wittig I, Head B, van der Bliek A, et al. APOOL is a cardiolipin-
778 binding constituent of the Mitofilin/MINOS protein complex determining cristae morphology in
779 mammalian mitochondria. PloS one. 2013;8(5):e63683.
780 53. Kaneko H, Noguchi J, Kikuchi K, Akagi S, Shimada A, Taya K, et al. Production and endocrine
781 role of inhibin during the early development of bull calves. Biology of reproduction. 2001;65(1):209-
782 15.
783 54. Phillips DJ. Activins, inhibins and follistatins in the large domestic species. Domestic animal
784 endocrinology. 2005;28(1):1-16.
785 55. Johnston D, Corbet N, Barwick S, Wolcott ML, Holroyd R. Genetic correlations of young bull
786 reproductive traits and heifer puberty traits with female reproductive performance in two tropical
787 beef genotypes in northern Australia. Animal Production Science. 2014;54(1):74-84.
788 56. Corbet N, Burns B, Johnston D, Wolcott ML, Corbet D, Venus B, et al. Male traits and herd
789 reproductive capability in tropical beef cattle. 2. Genetic parameters of bull traits. Animal Production
790 Science. 2012;53(2):101-13.
791 57. Canal LB, Fontes PL, Sanford CD, Mercadante VR, DiLorenzo N, Lamb GC, et al. Relationships
792 between feed efficiency and puberty in Bos taurus and Bos indicus-influenced replacement beef
793 heifers. Journal of Animal Science. 2020;98(10):skaa319.
794 58. Perry GA, Cushman R. Effect of age at puberty/conception date on cow longevity. Veterinary
795 Clinics: Food Animal Practice. 2013;29(3):579-90.
796 59. Bunter KL, Cai W, Johnston DJ, Dekkers JC, Bunter K. Selection to lower residual feed intake
797 in pigs produces a correlated response in juvenile insulin-like growth factor-I (IGF-1) concentration.
798 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
799 growth and development traits of Nanyang cattle in different growth period. Yi Chuan= Hereditas.
800 2006;28(8):927-32.
801 61. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A multi-trait, meta-
802 analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle.
803 PLoS genetics. 2014;10(3):e1004198.
804 62. Hackinger S, Zeggini E. Statistical methods to detect pleiotropy in human complex traits.
805 Open biology. 2017;7(11):170125.
806 63. Id-Lahoucine S, Molina A, Cánovas A, Casellas J. Screening for epistatic selection signatures:
807 A simulation study. Scientific reports. 2019;9(1):1-5.
808 64. Crowley J, Evans R, Mc Hugh N, Kenny D, McGee M, Crews Jr D, et al. Genetic relationships
809 between feed efficiency in growing males and beef cow performance. Journal of animal science.
810 2011;89(11):3372-81.
811 65. Purfield DC, Evans RD, Berry DP. Breed-and trait-specific associations define the genetic
812 architecture of calving performance traits in cattle. Journal of animal science. 2020;98(5):skaa151.
813 66. Nguyen T-M, Shafi A, Nguyen T, Draghici S. Identifying significantly impacted pathways: a
814 comprehensive review and assessment. Genome biology. 2019;20(1):1-15.

67. Bozza WP, Zhang Y, Hallett K, Rosado LAR, Zhang B. RhoGDI deficiency induces constitutive activation of Rho GTPases and COX-2 pathways in association with breast cancer progression. Oncotarget. 2015;6(32):32723.
68. Hall A. Rho family gtpases. Biochemical Society Transactions. 2012;40(6):1378-82.
69. Humphries B, Wang Z, Yang C. Rho GTPases: Big Players in Breast Cancer Initiation, Metastasis and Therapeutic Responses. Cells. 2020;9(10):2167.
70. Vega FM, Ridley AJ. Rho GTPases in cancer cell biology. FEBS letters. 2008;582(14):2093-101.
71. Shibata S, Nagase M, Yoshida S, Kawarazaki W, Kurihara H, Tanaka H, et al. Modification of mineralocorticoid receptor function by Rac1 GTPase: implication in proteinuric kidney disease. Nature medicine. 2008;14(12):1370-6.
72. Togawa A, Miyoshi J, Ishizaki H, Tanaka M, Takakura A, Nishioka H, et al. Progressive impairment of kidneys and reproductive organs in mice lacking Rho GDIα. Oncogene. 1999;18(39):5373-80.
73. Ishizaki H, Togawa A, Tanaka-Okamoto M, Hori K, Nishimura M, Hamaguchi A, et al. Defective chemokine-directed lymphocyte migration and development in the absence of Rho guanosine diphosphate-dissociation inhibitors α and β. The Journal of Immunology. 2006;177(12):8512-21.
74. Huta Y, Nitzan Y, Breitbart H. Ezrin protects bovine spermatozoa from spontaneous acrosome reaction. Theriogenology. 2020;151:119-27.
75. Shi Z-H, Zhao C, Wu H, Liu X-M. Expression of RhoGDI alpha in human testes and sperm and its correlation with the success rate of IVF. Zhonghua nan ke xue= National Journal of Andrology. 2011;17(4):325-9.
76. Wang L, Chen W, Zhao C, Huo R, Guo X-J, Lin M, et al. The role of ezrin-associated protein network in human sperm capacitation. Asian journal of andrology. 2010;12(5):667.
77. Burns B, Fordyce G, Holroyd R. A review of factors that impact on the capacity of beef cattle females to conceive, maintain a pregnancy and wean a calf—Implications for reproductive efficiency in northern Australia. Animal Reproduction Science. 2010;122(1-2):1-22.
78. McLean I, Holmes P, Counsell D. Final Report: The Northern beef report, 2013 Northern beef siutation analysis (B. COM. 0348). Meat and Livestock Australia, Sydney. 2013.
79. Chang AZ, Swain DL, Trotter MG. Calf loss in northern Australia: a systematic review. The Rangeland Journal. 2020;42(1):9-26.
80. Fordyce G. Pregnancy rates achieved by mating bulls with different percentages of morphologically normal sperm. Bullpower Delivery of adequate normal sperm to site of fertilisation Project NAP3. 2005;117:142-51.
81. Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik CG, Tseng E, et al. De novo assembly of the cattle reference genome with single-molecule sequencing. GigaScience. 2020;9(3):giaa021.
82. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee D, editors. BLUPF90 and related programs (BGF90). Proceedings of the 7th world congress on genetics applied to livestock production; 2002.
83. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics. 2011;88(1):76-82.
84. Bozeman M. Golden Helix, Inc. SNP & Variation Suite™[Software]. [(Version 8.x)].

Available from http://www.goldenhelix.com.
85. Hu Z-L, Park CA, Reecy JM. Developmental progress and current status of the Animal QTLdb. Nucleic acids research. 2016;44(D1):D827-D33.
86. Hu Z-L, Park CA, Wu X-L, Reecy JM. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic acids research. 2013;41(D1):D871-D9.
87. Lam S, Miglior F, Fonseca P, Gómez-Redondo I, Zeidan J, Suárez-Vega A, et al. Identification of functional candidate variants and genes for feed efficiency in Holstein and Jersey cattle breeds using RNA-sequencing. Journal of dairy science. 2021;104(2):1928-50.

866 88. Sweett H, Fonseca P, Suarez-Vega A, Livernois A, Miglior F, Cánovas A. Genome-wide
867 association study to identify genomic regions and positional candidate genes associated with male
868 fertility in beef cattle. Scientific reports. 2020;10(1):1-14.
869 89. Krämer A, Green J, Pollard Jr J, Tugendreich S. Causal analysis approaches in ingenuity
870 pathway analysis. Bioinformatics. 2014;30(4):523-30.
871 90. Medici V, Kieffer DA, Shibata NM, Chima H, Kim K, Canovas A, et al. Wilson Disease:
872 Epigenetic effects of choline supplementation on phenotype and clinical course in a mouse model.
873 Epigenetics. 2016;11(11):804-18.
874 91. Cardoso TF, Quintanilla R, Castelló A, González-Prendes R, Amills M, Cánovas Á. Differential
875 expression of mRNA isoforms in the skeletal muscle of pigs with distinct growth and fatness profiles.
876 BMC genomics. 2018;19(1):1-12.
877 92. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and
878 Bioconductor: a powerful link between biological databases and microarray data analysis.
879 Bioinformatics. 2005;21(16):3439-40.