Genome-wide association study to identify genomic regions and positional candidate genes associated with male fertility in beef cattle

H. Sweett1, P. A. S. Fonseca1, A. Suárez-Vega1, A. Livernois1,2, F. Miglior2 & A. Cánovas1*

Fertility plays a key role in the success of calf production, but there is evidence that reproductive efficiency in beef cattle has decreased during the past half-century worldwide. Therefore, identifying animals with superior fertility could significantly impact cow-calf production efficiency. The objective of this research was to identify candidate regions affecting bull fertility in beef cattle and positional candidate genes annotated within these regions. A GWAS using a weighted single-step genomic BLUP approach was performed on 265 crossbred beef bulls to identify markers associated with scrotal circumference (SC) and sperm motility (SM). Eight windows containing 32 positional candidate genes and five windows containing 28 positional candidate genes explained more than 1% of the genetic variance for SC and SM, respectively. These windows were selected to perform gene annotation, QTL enrichment, and functional analyses. Functional candidate gene prioritization analysis revealed 14 prioritized candidate genes for SC of which MAP3K1 and VIP were previously found to play roles in male fertility. A different set of 14 prioritized genes were identified for SM and five were previously identified as regulators of male fertility (SOD2, TCP1, PACRG, SPEF2, PRLR). Significant enrichment results were identified for fertility and body conformation QTLs within the candidate windows. Gene ontology enrichment analysis including biological processes, molecular functions, and cellular components revealed significant GO terms associated with male fertility. The identification of these regions contributes to a better understanding of fertility associated traits and facilitates the discovery of positional candidate genes for future investigation of causal mutations and their implications.

Abbreviations
- ssGBLUP: Single-step genomic BLUP
- WssGBLUP: Weighted single-step genomic BLUP
- SC: Scrotal circumference
- SM: Sperm motility
- FDR: False discovery rate
- AN: Angus
- SM: Simmental
- PI: Piedmontese
- GV: Gelbvieh
- CH: Charolais
- LM: Limousine
- HYS: Herd-year-season
- GALLO: Genomic Annotation in Livestock for positional candidate LOci
- MF: Molecular function
- BP: Biological process

1Department of Animal Biosciences, Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON N1G 2W1, Canada. 2Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, ON N1G 2W1, Canada. *email: acanovas@uoguelph.ca
As the human population continues to grow exponentially, so does the demand for animal proteins, such as beef and beef by-products. In 2014, the Intergovernmental Panel on Climate Change (IPCC) report estimated that by 2050 meat consumption will rise 21% per capita as a result of a larger human population, leading to business and economic benefits for beef producers. One way we can continue to match the demand of beef and beef by-products with the ever-growing population is through increasing beef cattle fertility. This can be attained through the use of genetics to select for more fertile bulls, thereby increasing breeding success rates and decreasing the number of replacement heifers. However, over the last few decades, fertility in beef cattle has decreased because of intense selection pressure on production traits. One of the reasons for this decrease in fertility results from the pleiotropic effect of genes and biological processes underlying fertility and production, such as cell proliferation and energy-conservation metabolism. Therefore, further research on the functional biology underlying fertility traits in cattle is required to prevent these undesirable side effects of selection.

Genetic studies have focused primarily on female fertility, which presents a significant challenge to improving beef cattle fertility because of the low heritability (0.01–0.10) of these reproductive traits. Fertility-related traits in bulls are moderately heritable (0.05–0.22) and significantly influenced by genetics. Therefore, improving our understanding of the genes and epigenetic modifications that contribute to bull fertility could improve reproduction success in beef cattle. Indeed, sires directly influence the fertilization process, the viability of the preimplantation embryo, as well as the conception rate. Sperm motility (SM) could be used as a proxy for identifying more fertile sires and improving reproduction success because it is a moderate to highly heritable trait (0.29–0.60) with significant components in the genome contributing to semen quality. With reduced motility in bull semen becoming a prominent concern in the breeding industry, it is essential to continue studying genomic regions, variants and functional genes that impact sperm motility.

Scrotal circumference (SC) is also a good indicator trait for sire fertility because it is highly correlated with testes weight, sperm output, and semen quality. Previous studies have suggested that a smaller SC is associated with lower fertility measures compared to bulls with a larger SC. Moreover, a positive association has been reported between SC and the percentage of live sperm, sperm number, and SM. Therefore, the identification of genetic markers linked to SC and SM could be used as a strategy to genetically improve reproductive performance. However, there is a negative genetic correlation between performance traits, such as feed efficiency, and SC, with more feed-efficient bulls having smaller SC. Thus, careful consideration must be taken into account when selecting for both production and fertility traits to avoid undesirable effects.

Relatively fewer GWAS for bull fertility traits have focused on crossbred beef cattle, which represent a significant proportion of the beef cattle population, compared to purebred cattle. This is in part due to multiple beef breeds with each breed association carrying out a separate genetic evaluation using different methods to calculate expected progeny deviation and the correspondence of SNP effects. In this study we used a sample size of 265 bulls with both SC and SM measurements. Numerous other male fertility GWAS in cattle have obtained interesting results with a very similar sample size. The objectives of the present study were (1) to identify positional candidate genes with additional functional evidence and their potential role in bull fertility, and (2) to find SNP windows significantly associated with SC and SM in crossbred beef bulls using a single-step approach, and to identify positional candidate genes with additional functional evidence and their potential role in bull fertility. To the best of our knowledge, this is the first study that uses a GWAS to analyze both SC and SM in crossbred beef bulls using a single-step approach.

**Results**

**Scrotal circumference genome wide association study and QTL enrichment.** After quality control, 379,591 markers remained for analysis based on a call rate greater than 95% and a minor allele frequency greater than 5% with non-autosomal markers. The Manhattan plot in Fig. 1a shows the non-overlapping windows 1 Mb apart which explain the highest proportion of variance for SC. Eight windows explained more than 1% of the genetic variance for SC which were located on BTA9, BTA10, BTA20, BTA24 and BTA29, and explained 13.19% of the total genetic variance (Table 1). Thirty-two positional candidate genes were identified within these windows explaining more than 1% of the genetic variance for SC. ToppGene prioritization analysis revealed 14 of the 32 positional candidate genes were prioritized for spermatic-related processes and the windows which these genes are mapped within explained 9.76% of the total genetic variance for SC. The QTL annotation revealed several reproduction QTLs, annotated within the coordinates of the windows explaining > 1% of
Figure 1. Manhattan plot for the GWAS of (a) scrotal circumference and (b) sperm motility in crossbred beef cattle. The y-axis shows the proportion of variance explained by the non-overlapping windows and the x-axis indicates the chromosome number. The red line indicates the threshold for 1% of the variance explained by the windows.

Table 1. Significant associated windows explaining more than 1% of the variance for scrotal circumference and sperm motility in crossbreed beef cattle and the overlapping genes located within those windows.

*Prioritized candidate gene identified by GUILDify and ToppGene analysis.
the total genetic variance for SC, account for 7.85% of the QTLs annotated in those regions (Fig. 2a). The QTL enrichment analysis revealed 38 significant QTLs (FDR-corrected \(p\) value ≤ 0.05) on BTA9 and BTA10 annotated for traits related to exterior conformation (Table 2; Fig. 2b).

### Table 2. Significantly enriched QTLs identified annotated in the windows explaining more than 1% of the variance for scrotal circumference and sperm motility.

| Trait                  | QTL Type       | BTA | Number of QTLs | Expected number of QTLs | \(p\) value  | FDR-corrected \(p\) value |
|------------------------|----------------|-----|----------------|-------------------------|--------------|--------------------------|
| Scrotal circumference  | Udder depth    | 9   | 5              | 1.54                    | 3.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Body depth     | 9   | 5              | 1.36                    | 5.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Feet and leg conformation | 9   | 3              | 0.95                    | 1.60 \times 10^{-2} | 2.67 \times 10^{-2} |
|                        | PTA type       | 9   | 5              | 1.55                    | 2.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Udder attachment | 9  | 5              | 1.44                    | 4.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Strength       | 9   | 5              | 0.89                    | 1.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Teat placement—rear | 9  | 4              | 0.97                    | 3.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Udder cleft    | 9   | 5              | 1.51                    | 4.00 \times 10^{-3} | 9.38 \times 10^{-3} |
|                        | Stature        | 10  | 1              | 0.10                    | 5.00 \times 10^{-3} | 9.38 \times 10^{-3} |
| Sperm motility         | Body weight gain | 9  | 3              | 0.38                    | 1.00 \times 10^{-3} | 9.00 \times 10^{-3} |
|                        | Daughter pregnancy rate | 9  | 1              | 0.34                    | 4.50 \times 10^{-2} | 8.10E-02            |
|                        | Interval to first estrus after calving | 9   | 2              | 0.61                    | 2.50 \times 10^{-2} | 8.10 \times 10^{-2} |
|                        | Milk fat yield | 13  | 4              | 2.08                    | 4.30 \times 10^{-2} | 8.10 \times 10^{-2} |
|                        | Stature        | 13  | 3              | 1.13                    | 3.50 \times 10^{-2} | 8.10 \times 10^{-2} |

**Figure 2.** (a) Pie plot showing the percentage of each QTL class annotated in the windows explaining >1% of the total genetic variance for scrotal circumference. (b) Enriched traits identified in the QTL enrichment analysis for scrotal circumference. The area with the circles represents the number of observed QTLs for that class, while the color represents the \(p\) value scale (the darker the color, smaller the \(p\) value). The x-axis shows the richness factor for each QTL, representing the ratio of number of QTLs and the expected number of that QTL.

**Sperm motility genome wide association study and QTL enrichment.** The Manhattan plot in Fig. 1b shows the non-overlapping SNP windows 1 Mb apart that explain the highest proportion of variance for SM. Five windows explained more than 1% of the genetic variance for SM which were located on BTA9, BTA13, BTA20, and BTA24, and explained 7.17% of the total genetic variance (Table 1). Twenty-eight positional candidate genes were identified within these SNP windows explaining more than 1% of the genetic variance for SM. ToppGene prioritization revealed 14 of the 28 positional candidate genes were prioritized for spermatogenic-related processes and the windows which these genes are mapped within explained 7.16% of the total genetic variance for SM. The QTL annotation revealed several reproduction QTLs, annotated within the coordinates of the win-
dows explaining > 1% of the total genetic variance for SM, account for 8.11% of the QTLs annotated in these regions (Fig. 3a). The QTL enrichment analysis revealed 13 significant QTLs (FDR-corrected $p$ value ≤ 0.05) on BTA9 and BTA13 annotated for traits related to reproduction and exterior conformation (Table 2; Fig. 3b).

**Functional analysis of prioritized candidate genes for scrotal circumference and sperm motility.** The network constructed with the 14 prioritized candidate genes for SC and SM (total of 28 genes) using NetworkAnalyst had 1442 nodes (genes) related to 120 GO:BP (Table S2a), 78 GO:MF (Table S3a) and 69 GO:CC (Table S4a) terms. One hundred and twenty GO:BP were significant (FDR-corrected $p$ value ≤ 0.05) and 64 were identified as related to male fertility and reproduction (Table S2b). These 64 GO:BP terms were selected to construct the first module containing 732 genes (Table S2c). A second module was generated composed only by the prioritized candidate genes for SC and SM and its directly connected nodes leading to a network constructed of 20 genes and 26 GO:BP terms (Fig. 4a, Table S2d). None of the GO:BP in this module were significant (the smallest FDR-corrected $p$ value was 0.128), however nine were related to male fertility and reproduction of which regulation of MAPK cascade, spermatid differentiation, and regulation of hormone secretion were the most notable (Table 3). In this second module, five of the 20 genes were prioritized candidate genes for SM and one gene was prioritized for SC.

Of the 78 GO:MF, 75 were significant (FDR-corrected $p$ value ≤ 0.05) and 34 were identified as related to male fertility and reproduction functions (Table S3b). The first module for GO:MF terms was extracted using these 34 GO:MF terms containing 684 genes (Table S3c). A second module for GO:MF was generated composed only by the prioritized candidate genes for SC and SM and its directly connected nodes revealing a network constructed of 32 genes and 20 GO:MF terms (Fig. 4b, Table S3d). In this second module, 10 GO:MF were significant (FDR-corrected $p$ value ≤ 0.05) and six of these were related to male fertility and reproduction including, acetyltransferase activity, zinc ion binding, lipase activity, endonuclease activity, nuclease activity, and cation channel activity (Table 3). Five of the 32 genes in this second module for GO:MF were prioritized candidate genes for SM and four were prioritized candidate genes for SC.

Of the 69 GO:CC terms, 55 were significant (FDR-corrected $p$ value ≤ 0.05) and 28 were related to male fertility and reproduction functions and processes (Table S4b). These 28 GO:CC terms were selected to extract the first module for GO:CC which contained 992 genes (Table S4c). “Batch Selection” of the prioritized candidate genes for SC and SM was used to extract the second module for GO:CC which revealed a network containing 51 genes and 21 GO:CC terms (Fig. 4c, Table S4d). In this second module, five GO:CC were significant (FDR-corrected $p$ value ≤ 0.05) and three of these were related to male fertility and reproduction including, kinesin complex, spindle microtubule and cytosol (Table 3). Five of the 51 genes in the second GO:CC module were prioritized candidate genes for SM and seven genes were prioritized candidate genes for SC. A graphical representation of the functional analysis can be found in Fig. 5.

**Discussion**

The ability to identify cattle with greater reproductive performance would significantly improve the efficiency of the beef cattle industry. Thus, the discovery of genetic markers related to male fertility through GWAS can contribute to the evaluation of fertility traits, such as SC and SM, and their implications on male fertility. Spermatic
related traits, including SM are commonly measured in dairy cattle where semen is widely used for artificial insemination and measuring semen parameters is a standard practice. However, SM measurements are much less common in the beef industry where natural breeding is often used, which can be seen in our sample size. Testicular related traits, including SC have higher heritability values and correlations with fertility traits, although the majority of studies on SC have been on purebred bulls. Few SC GWAS have been conducted in crossbred beef cattle, which represents a significant proportion of the beef industry. To the best of our...
| GO category       | GO term                        | FDRa-corrected p value | Gene names |
|------------------|--------------------------------|------------------------|------------|
| Biological process | Regulation of MAPK cascade | 8.13 × 10^{-1}         | TCP1*, PIK3R1 |
|                  | Spermatid differentiation | 1                      | PRKR*       |
|                  | Regulation of hormone secretion | 1                  | ACACA      |
|                  | Acetyltransferase activity | 8.88 × 10^{-4}         | CDC27, JAK1, GRB2, PIK3R1 |
| Molecular function | Zinc ion binding | 2.12 × 10^{-3}         | JAK1, PIK3CA, PLK1, CDK1, PI4KA, SGK1, PRKACA |
|                  | Lipase activity             | 2.55 × 10^{-2}         | GRB2, MAP2K1, PLK1, MAP3K1*, PRKACA, FBX05* |
|                  | Endonuclease activity      | 2.55 × 10^{-2}         | PIK3CA, PLK1, CDK1, MAP3K1*, SGK1, PRKACA |
|                  | Nuclease activity           | 4.58 × 10^{-2}         | JAK1, MAP2K1, PLK1, CDK1, MAP3K1*, SGK1, PRKACA |
|                  | Cation channel activity    | 9.51 × 10^{-2}         | PLC2I, PLCB1, PLCB4* |
| Cellular component | Kinesin complex            | 1.53 × 10^{-2}         | CDC27, PLK1, CDK1 |
|                  | Spindle microtubule         | 1.53 × 10^{-2}         | RAC1, PIK3CA, GRB2, TRA8F, GNA14, MAP2K1, IRAK1, TLR4, TIBA*, PI4KA, HPS9A1A1, XRN1, SGK1, PRKACA, PRKR*, CD40, NPSR1, PPP2CA, RGS17* |
|                  | Cytosol                     | 4.42 × 10^{-2}         | JAK1, JAK3, TUBB4A, PLK1, CDK1, PPP2CA |

Table 3. List of significant gene ontology terms associated with biological processes, molecular functions, and cellular components related to fertility identified by NetworkAnalyst using the protein–protein interaction network built by the prioritized candidate genes for scrotal circumference and sperm motility and the directly connected nodes. *False discovery rate. Prioritized candidate gene identified by GUILDify and ToppGene analysis.

knowledge, this is the first study that uses a GWAS to analyze both SC and SM measurements, simultaneously in Canadian crossbred beef cattle. The sample size in the present study (265) might result in a careful interpretation of the obtained results, however many other GWAS have obtained notable results with similar sample sizes, ranging from 41 to 692,15,18,23–27. For example, Hering et al.35 conducted a GWAS on 41 bulls with very poor SM and 279 bulls with excellent SM and identified nine candidate genes all with a strong relationship to sperm function. Buzanskas et al.36 used SC records on 392 bulls in a GWAS and identified the STATU2 gene which participates in multiple biological processes, including reproduction, developmental and immune systems. When taken together, the limited studies conducted in crossbred beef cattle, the high heritability of SC and SM, and the evaluation of these traits together results in a study that can bring new considerations for the current stage of the literature regarding male fertility traits in beef cattle.

The ssGBLUP method is based on an infinitesimal model and thus assumes equal variance for all SNP effects, posing an unrealistic situation for traits of economic interest, such as fertility35. Therefore, a weighted ssGBLUP (WssGBLUP) approached was used that combines pedigree, phenotype, and genotype data with the integration of different weights for markers in an iterative process to update SNP solutions. For SC, 8 windows, located on BTA9, BTA10, BTA24, and BTA29, explained more than 13.19% of the total genetic variance. For SM, 5 windows explained more than 7.17% of the genetic variance and were located on BTA9, BTA13, BTA20, and BTA24. Of these, BTA9 explained the highest proportion of the variance for SC (4.03%) and SM (2.76%), respectively. Chromosome-wise and genome-wise associations have previously been identified in BTA9 for SC and BTA24. Of these, BTA9 explained more than 13.19% of the total genetic variance. For SC, 8 windows, located on BTA9, BTA10, BTA24, and BTA29, explained more than 13.19% of the total genetic variance. For SM, 5 windows explained more than 7.17% of the genetic variance and were located on BTA9, BTA13, BTA20, and BTA24. Of these, BTA9 explained the highest proportion of the variance for SC (4.03%) and SM (2.76%), respectively. Chromosome-wise and genome-wise associations have previously been identified in BTA9 for SC at 420 days in Canchim beef cattle26. Moreover, a region on BTA9 in Nellore cattle was found to be associated with testicular hypoplasia, which is defined by a reduced size of both testicles, and consequently reduced scrotal circumference and sperm physiology35.

Of the 14 prioritized candidate genes identified for SC (SASH1, VIP, FBXO5, MTRF1L, RGS17, SLC24A5, SEMA6D, MAP3K1, SRPRA, TIRAR, DCPS, ST3GAL4, KIRREL3, CDON), two, MAP3K1 and VIP, were previously reported as related to male reproduction. Mitogen-activated protein kinase kinase kinase 1 (MAP3K1), was mapped within a window that explained 1% of the variance in SM (BTA20: 22.18–23.18 Mb) and was of interest for its anti- and pro-apoptotic functions in germ cells96. Guan et al.37 found the MAP3K1 gene was differentially expressed in testis from underfed and well-fed sexually mature sheep, indicating it could be a marker of germ cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. For SC, 8 windows, located on BTA9, BTA10, BTA24, and BTA29, explained more than 13.19% of the total genetic variance. For SM, 5 windows explained more than 7.17% of the genetic variance and were located on BTA9, BTA13, BTA20, and BTA24. Of these, BTA9 explained the highest proportion of the variance for SC (4.03%) and SM (2.76%), respectively. Chromosome-wise and genome-wise associations have previously been identified in BTA9 for SC at 420 days in Canchim beef cattle26. Moreover, a region on BTA9 in Nellore cattle was found to be associated with testicular hypoplasia, which is defined by a reduced size of both testicles, and consequently reduced scrotal circumference and sperm physiology35.

Of the 14 prioritized candidate genes identified for SC (SASH1, VIP, FBXO5, MTRF1L, RGS17, SLC24A5, SEMA6D, MAP3K1, SRPRA, TIRAR, DCPS, ST3GAL4, KIRREL3, CDON), two, MAP3K1 and VIP, were previously reported as related to male reproduction. Mitogen-activated protein kinase kinase 1 (MAP3K1), was mapped within a window that explained 1% of the variance in SM (BTA20: 22.18–23.18 Mb) and was of interest for its anti- and pro-apoptotic functions in germ cells96. Guan et al.37 found the MAP3K1 gene was differentially expressed in testis from underfed and well-fed sexually mature sheep, indicating it could be a marker of germ cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production. The vasoactive intestinal peptide (VIP, BTA9: 89.51–90.51 Mb window), another prioritized candidate gene mapped within a window for SC that explained 1.61% of the variance, is known to act directly on the testis, promoting the production of testosterone and cell apoptosis and therefore change SC and the efficiency of sperm production.
the development of normal sperm and male fertility. Previous studies involving PACRG (BTA9: 97.25–98.24 Mb window) knockdown in mice have discovered it plays a vital role in maintaining the functional stability of flagella indicating an important relation to sperm motility. Moreover, Guo et al. discovered SPEF2 (BTA20: 38.08–39.08 window) gene expression in the testes and sperm is regulated by alternative splicing and is thus, one of the determining factors of sperm motility. Lastly, PRLR (BTA20: 38.08–39.08 window) gene expression was identified in the bovine reproductive tract including, the testis, epididymis, spermatogonia, and differentiating germ cells, leading researchers to believe it may have an effect on spermatogenesis. This gene was also found to play a role in fertilization and survival rates through SNP-SNP interactions.

Five of the prioritized candidate genes for SC and SM have also been identified as playing important roles in the regulation of female fertility. Such genes include, semaphorin 6D, SRP receptor alpha subunit, Kirre like nephrin family adhesion molecule 3, acetyl-CoA Acetyltransferase 2, and proto-oncogene, G protein-coupled receptor (SEMA6D, SRPRA, KIRREL3, ACAT2, MAS1). Both SEMA6D (BTA10: 22.10–63.21 Mb) and SRPRA (BTA29: 29.35–30.35 Mb window) are expressed in the human female genital tract, with SEMA6D linked to known gonadal genes and SRPRA upregulated during pregnancy and involved in the transport of secretory and membrane proteins. On the other hand, the KIRREL3 (BTA29: 29.35–30.35 Mb window), ACAT2 (BTA9: 95.18–96.18 Mb window), and MAS1 (BTA9: 95.18–96.18 Mb window) genes have been previously studied in bovine. The KIRREL3 gene was found to be highly expressed in granulosa cells and may act as a metabolic messenger linking metabolism, body composition and fertility. The ACAT2 gene was associated with both daughter pregnancy rate and cow conception rate, and MAS1 gene has been suggested to play a role in the regulation of

Figure 5. Graphical representation of the functional analyses performed to identify the positional candidate genes with additional functional evidence for scrotal circumference and sperm motility.
Through functional analysis, three modules were created with the prioritized candidate genes for SC and SM, with GO:BP GO:MF, and GO:CC terms related to male fertility and reproduction. Even though nine GO:BP in the module (created via selection of the GO terms related to reproductive processes and the prioritized candidate genes for SC and SM) were related to male fertility and reproduction, due to the few genes contained in the network, only one or two genes were associated with these BPs, thus, none of the terms were significantly enriched. Despite this, it is worthy to highlight the most notable BPs in this network: regulation of MAPK cascade, spermatid differentiation, and regulation of hormone secretion. The MAPK cascade regulates spermatogenesis, sperm maturation, and the acrosome reaction, thereby playing an important role in male reproductive processes. One of the genes involved in the regulation of MAPK cascade is the T-complex protein 1 subunit alpha, TCP1, a prioritized candidate gene for SM and is involved in the assembly of actin and tubulin filaments. Actin filaments are present in mammalian germ cells and are involved in a number of changes that occur during spermatogenesis such as, the determination of cell shape, motility, maturation of spermatozoa, and capacitation. In particular, actin proteins change their distribution in the sperm head during maturation and control the balance between globular actin and fibrous actin. Another GO:BP, spermatid differentiation, is also an important part of spermatogenesis that involves the differentiation of spermatids into mature spermatozoon in the seminiferous tubules. The prolin receptor, a SM prioritized candidate gene, was found to be directly linked with spermatid differentiation. Moreover, spermatogenesis and reproductive success is completely dependent on the secretion of various hormones, including FSH, androgen, and testosterone. For example, testosterone levels and SC in bulls are known to be correlated, with peak testosterone levels being lower in bulls with smaller SC. Six of the significant GO:MF were related to male fertility and reproduction, including acetyltransferase activity, zinc ion binding, lipase activity, and cation channel activity. Acetyltransferases, specifically, histone acetyltransferases, act during spermatogenesis through the differentiation of spermatocytes. Zinc is incorporated into spermatozoa where it is bound by seminal fluid Zn-interacting proteins and plays a protective role for sperm chromatin decondensation and sperm motility. Moreover, a number of cation channels, including potassium and chloride channels, are involved in sperm motility, maturation, and the acrosome reaction. Three of the significant GO:CC terms were related to male fertility and reproduction, including, kinesin complex, spindle microtubule and cytoskeleton. A number of kinesins play key roles in mammalian spermatogenesis, including mitosis, meiosis, acrosome biogenesis, and tail formation. Moreover, kinesins, specifically kinesin-13 s, regulate the spindle microtubule dynamics and control spindle assembly and kinetochore-microtubule attachments. One example of the importance of the cytosol in male fertility is the cytosolic fraction of bovine spermatozoa, which exhibits tyrosine kinase activity associated with sperm capacitation and acrosome reaction. The GO:BP term spindle microtubule contained the TIRAP and RGS17 genes which are prioritized candidate genes for SC and PRLR, a prioritized candidate gene for SM, indicating that these genes could be important biomarkers of bovine fertility.

The QTL analysis revealed reproduction QTLs account for 7.85% and 8.11% of the QTLs annotated in the windows explaining >1% of the total genetic variance for SC and SM, respectively. These reproduction QTLs included, reproductive efficiency, age at first calving, calving ease, daughter pregnancy rate, interval from first to last insemination, fertility index, conception rate, and interval to first estrus after calving. The largest proportion of QTLs in this study were related to milk production for both SC and SM (31.94% and 54.05%, respectively). A QTL enrichment analysis was conducted as the simple bias of investigation in the QTL database for cattle can result in a larger proportion of records in the database. The enrichment analysis for the windows explaining >1% of the total genetic variance for SC and SM revealed QTLs annotated for exterior conformation traits, including five different body depth, three different feet and leg conformation, five different strength, and five different stature QTLs annotated within the candidate windows for SC, and three different stature QTLs and three different body weight gain QTLs annotated within the candidate windows for SM (Table 2; Table S5). In a study conducted by Schenkel et al., SC was found to be genetically correlated (P < 0.05) to average daily gain, ultrasound backfat thickness, mid-test metabolic weight and hip height in 13,151 bulls. A QTL enrichment analysis was conducted as the simple bias of investigation in the QTL database for cattle, therefore udder traits may be used in conjunction with body conformation traits for the enrichment analysis. Therefore, bulls with larger SC have a larger body size and faster growth. It is suggested that these regions may be regulating both fertility and conformation traits, however the biological mechanisms associated with this correlation is not well understood. Another enriched QTLs for SC included udder depth, udder attachment, teat placement-rear, and udder cleft (Table 2; Table S5). Cows with tightly attached udders and proper teats tend to remain in a herd longer, are easier to nurse, and less susceptible to mastitis, therefore udder traits may be used in conjunction with body conformation traits for the indirect selection of longevity in beef cattle. The QTL enrichment analysis for SM also revealed QTLs related to fertility on BTA9, one for daughter pregnancy rate and two different QTLs for interval to first estrus after calving (Table 2; Table S5). That is consistent with other studies that reveal a genetic correlation between male and female fertility traits. For example, Johnston et al. found SM at 18 months of age to be highly correlated with female reproduction traits in Brahman and Tropical Composite cows, such as conception (0.53 and 0.72, respectively) and pregnancy rate (0.58 and 0.95, respectively). Gargantini et al. identified a genetic correlation between yearling SC and age at puberty and pregnancy rate in heifers was −0.57 and 0.35, respectively. Thus, BTA9 may be a candidate region for bovine fertility. A study conducted by McClure et al. identified 41 SC QTLs in Angus cattle, of which three QTLs were identified on BTA9. The overlap between the candidate regions identified in the present study and previous studies reinforces the association of these genomic regions with the regulation of genes and biological processes responsible for male fertility traits, including VIP and SOD2 genes, mapped on BTA9, and the regulation of MAPK cascade and spermatid differentiation.
To conclude, functional analysis for the prioritized candidate genes identified in this study revealed significant GO terms associated with biological processes and molecular functions related to male fertility and reproduction. For scrotal circumference, both MAP3K1 and VIP identified genes control testis function and could be used as potential biomarkers of spermato genesis and apoptosis. For sperm motility, the SOD2, TCP1, PACRG, SPEG, and PRLR genes were related to sperm concentration, development, and motility. These results help to better understand the genetic bases of scrotal circumference and sperm motility specifically in crossbred beef bulls and revealed positional candidate genes with additional functional evidence that might ultimately improve bull genomic prediction for these traits. Moreover, these candidate regions, specifically those mapped on BTA9, have known genetic correlations to other economically important traits including, conformation, female fertility, and udder structure. However, future research on these candidate genes and their impact on bull fertility is warranted.

Materials and methods

Population structure and phenotypic data. Data used in this study are from animals cared for under protocols approved by the University of Guelph Animal Care Committee, which follows guidelines of the Canadian Council on Animal Care (1993).

The population for this study came from the Ontario Beef Research Centre; the University of Guelph research farm located in Elora, ON, Canada and consisted of 265 crossbred bulls. At the time of collection, bulls had an average age of 384 days and average weight of 555 kg. The predominant breeds (and corresponding average composition of that breed in the test group, see Table S1) of these crossbred bulls were Angus (AN: 52%), Simmental (SM: 24%), Piedmontese (PI: 6.6%), Gelbvieh (GV: 6.3%), Charolais (CH: 3.8%), and Limousine (LM: 1.1%).

The fertility traits used in this study were SC and SM. Scrotal circumference was measured by palpating the testes into the lower half of the scrotum and measuring the greatest circumference using a looped tape as described by Awada et al. Semen ejaculates were collected between 12 and 15 months of age immediately after the SC measurement using electroejaculation (Pulsator IV- Auto Adjust Electro-Ejaculator: Lane Manufacturing, Inc, Denver, CO). Sperm motility was visually estimated immediately and ejaculates were then extended, cooled, placed in liquid nitrogen and thawed as previously described by Pursel & Johnson. Sperm motility was assessed as described by Awada et al. Briefly, sperm were evaluated on the CASA (Integrated Visual Optical System (IVOS) CASA System, Hamilton Thorne, Inc., Beverly, MA) sampled from 1 × 10⁶ spermatozoa/ml in BTS and a four-chamber standard count analysis slide (Leja products B.V. Luzenestraat 10, 2153 GN Nieuw-Vennep, The Netherlands), at 37 °C.

Genotyping and quality control. Genotyping was completed from 265 animals using the Affymetrix Genechip Bovine Genome High Density Array, which included 648,874 SNPs. Marker coordinates were converted to the new bovine genome assembly ARS-UCD 1.2. Quality control was performed using Plink and the following criteria were used for the exclusion of SNPs: non-autosomal regions; minor allele frequency < 0.05; and a call rate < 95%.

Genome-wide association study. The programs of BLUPF90 family were used for the weighted single-trait genomic BLUP (WssGBLUP) analysis. The GWAS results were reported as the proportion of the variance explained by non-overlapping genomic windows of 1 Mb.

A WssGBLUP method was used to estimate SNP effects. The observed phenotypes of SC and SM were used as dependent variables in a single-trait animal model:

\[ y = Xb + Z_\alpha a + e \]

where \( y \) represented a vector of observed phenotypes for animals (SC or SM); \( X \) is the incidence matrix of fixed effects; \( b \) is the vector of fixed effects, \( Z_\alpha \) is the incidence matrix of additive genetic effects; \( a \) is the random vector of additive genetic effects; and \( e \) is the vector of residual effects. Fixed effects for SC included 25 levels of herd-year-season (HYS), body weight, age as a third-degree polynomial, and breed composition for the most prevalent breeds (AN, SM, PI, GV, CH, and LM). Fixed effects for SM included 25 levels of HYS, age, and breed composition for the most prevalent breeds.

The solutions of the SNP effects (\( \hat{\alpha} \)) were obtained using the AIREMLF90 algorithm with two iterations, as proposed by Wang et al. For each iteration of the algorithm \( D_n = I \) and \( G_n = ZD_nZ^\top \lambda \), where \( D \) is a diagonal matrix of weights for SNP variances, \( G \) is the weighted genomic relationship matrix, \( Z \) is a matrix relating genotypes of each locus, \( \lambda \) is a variance ratio, and \( n \) is the iteration number; the breeding values (\( \hat{\alpha}_i \)) were calculated using single-trait genomic best linear unbiased predictor (sGBLUP); SNP effects were calculated via \( \hat{\alpha}_i = D_n^{(i)}D_n^{(i)\top}Z_n^{(i)}\hat{\lambda}_n^{(i)} \); SNP weights were calculated via \( d_n^{(i)} = \hat{a}_i^2/Z_n^{(i)}\hat{\lambda}_n^{(i)} \); SNP effects were obtained using \( D_n^{(i)} = \frac{tr(D_n^{(i)})}{tr(D_n^{(i)\top}D_n^{(i)})} \) so the additive genetic variance remains constant; and the Genomic matrix was recalculated via \( G_n^{(i)} = ZD_n^{(i)}Z^\top \) to obtain the SNP effects. These iterations were used to calculate the proportion of variance explained by non-overlapping windows.

The proportion of variance explained by non-overlapping windows was estimated using the PostGSS90 algorithm by summing the variance of SNPs within 1 Mb. Windows that explained greater than 1% of the genetic variance for SC and SM were selected for QTL and gene annotation, conducted using R (Version 4.0.0; R Core Team, 2020) and the R package: Genomic Annotation in Livestock for positional candidate LOci (GALLO—https://github.com/pablobio/GALLO). The .gff file with the QTL information from Animal QTL Database, using the same reference genome (ARS-UCD1.2) to map the QTLs, were used for gene and QTL annotation, respectively. A QTL
enrichment analysis was also conducted using the GALLO R package for all the QTL information annotated within the candidate windows using a chromosome-based enrichment analysis. Briefly, a bootstrap approach was used to compare the observed number of QTL for each trait in each chromosome annotated using GALLO with the expected number for each trait estimated using 1000 iteration rounds of random sampling from the whole Cattle QTLdb. Using this approach, a \( p \) value for the QTL enrichment status of each annotated trait within the candidate windows was calculated. Additionally, the \( p \) values were corrected for multiple testing using FDR (5%).

**Gene prioritization analysis.** Functional candidate gene prioritization was conducted using the ToppGene Suite. A trained list of genes associated with keywords outlined by Fonseca et al. using GUILDify software and a species-specific (Homo sapiens) interaction network, including, “scrotal circumference,” “scrotal,” “testicular,” “testis,” “testes,” “sperm,” “semen,” “spermatozoa,” “spermatogenesis,” and “fertility,” made up the “trained list” of genes. From this analysis, the top 100 genes ranked using an algorithm based on network topology on GUILDify were used to create the “trained” gene list. This “trained” gene list was used in conjunction with the “test” gene list containing the genes within windows explaining greater than 1% of the genetic variance for SC and SM. An annotation-based prioritization analysis through a multivariate approach was conducted using ToppGene. Gene Ontology terms for molecular function (MF), biological process (BP), and cellular component (CC); human and mouse phenotypes; metabolic pathways; Pubmed publications; and diseases were utilized to retrieve functional information for the trained list and for the list of positional candidate genes (those genes annotated within windows explaining more than 1% of the genetic variance for the trait). Overall \( p \) values were obtained using a combination of the \( p \) values obtained from the intermediate values from the above functional information using a random sampling of 5000 genes from the whole genome for each annotation information. Significant prioritized genes were selected based on a FDR 5% multiple correction, meaning these genes have a functional profile that is significantly similar to the functional profile of the “trained” gene list. As the “trained” gene list is known to be associated with fertility, by the guilty by association principle, it is probable that the prioritized genes will also be associated with fertility.

**Functional analysis.** A protein–protein interaction network analysis was performed in order to identify interactions between the positional candidate genes and other genes in the genome that are relevant to fertility. Therefore, the prioritized candidate genes for both SC and SM were inputted together into NetworkAnalyst 3.0 and the STRING interactome protein–protein interaction database was used with a confidence score cutoff of 900. A second-order interaction network was generated containing 1442 nodes (genes) with 4888 edges. Then, gene ontology (GO) enrichment analyses, including the three main categories biological process (BP), molecular functions (MF), and cellular components (CC), were performed using the genes comprising the network. Functional evidence of the relationship between the significant GO terms (FDR-corrected \( p \) value ≤ 0.05) and the target phenotypes (SC and SM) was identified. The GO terms related to reproductive processes were selected to extract the first module from this network composed only by the nodes (genes) associated with the selected GO:BP, GO:MF and GO:CC. Next, a second module was extracted, from the previous sub-network. This second module was composed only by the positional prioritized candidate genes for SC and SM and its direct connected nodes. The GO:BP, GO:MF and GO:CC enriched for this second module were analyzed for its association with reproductive processes.

Received: 10 August 2020; Accepted: 16 October 2020
Published online: 18 November 2020

**References**

1. Cerri, C. C. et al. Assessing the carbon footprint of beef cattle in Brazil: a case study with 22 farms in the State of Mato Grosso. *J. Clean Prod.* **112**, 2593–2600. https://doi.org/10.1016/j.jclepro.2015.10.072 (2016).

2. Revell, B.J. One man’s meat...2050? Ruminations on future meat demand in the context of global warming. *J. Agric. Econ.* **66**, 573–614. https://doi.org/10.1111/1477-9552.12121 (2015).

3. Thundathil, J. C., Dance, A. L. & Kastelic, J. P. Fertility management of bulls to improve beef cattle productivity. *Theriogenology* **86**, 397–405. https://doi.org/10.1016/j.theriogenology.2016.04.054 (2016).

4. Nguyen, L. T. et al. STAT6, PBX2, and PBRM1 emerge as predicted regulators of 452 differentially expressed genes associated with puberty in Brahman heifers. *PloS One* https://doi.org/10.1371/journal.pone.0120251 (2014).

5. Fonseca, P. A. S. et al. Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. *PloS ONE* https://doi.org/10.1371/journal.pone.0205295 (2018).

6. Diao, M. M. et al. SNP detection using RNA-sequences of candidate genes associated with puberty in cattle. *Anim. Genet.* https://doi.org/10.1111/age.12420 (2016).

7. Berry, D. P., Wall, E. & Pryce, J. E. Genetics and genomics of reproductive performance in dairy and beef cattle. *Animal* **8**, 105–121. https://doi.org/10.1017/S1751731114000743 (2014).
50. Lehti, M. S. & Sironen, A. Formation and function of sperm tail structures in association with sperm motility defects. *Biol. Reprod.* 97, 522–536. https://doi.org/10.1093/biolre/iox096 (2017).

51. Guo, F. et al. Alternative splicing, promotor, methylation, and functional SNPs of sperm flagella 2 gene in testis and mature spermatozoa of Holstein bulls. *Reprod. 147*, 241–252. https://doi.org/10.1530/REP-13-0343 (2014).

52. Pratt, S. L. et al. Identification of bovine prolactin in seminal fluid, and expression and localization of the prolactin receptor and prolactin-inducible protein in the testis and epididymis of bulls exposed to ergot alkaloids. *Theriogenology* https://doi.org/10.1016/j.theriogenology.2014.10.031 (2014).

53. Huang, W., Mikhail, D., Bindrim, A. C. & Khatib, H. Interactions of the bovine placental lactogen and prolactin receptor genes are associated with fertility traits in cattle. *Anim.* 3, 1743–1745. https://doi.org/10.1515/17517311109998026 (2009).

54. Iliaard, S. et al. Array-CGH diagnosis in ovarian failure: identification of new molecular actors for ovarian physiology. *J. Ovarian Res.* https://doi.org/10.1186/s13048-016-0272-5 (2016).

55. Whitington, C. M. et al. Transcriptomic changes in the pre-implantation uterus highlight nutritive biology of the developing marsupial embryo. *Sci. Rep.* https://doi.org/10.1038/s41598-018-20744-z (2018).

56. Coyral-Castel, S. et al. KIRREL is differentially expressed in adipose tissue from ‘fertile+’ and ‘fertile−’ cows: in vitro role in ovary?. *Reprod.* 155, 181–196. https://doi.org/10.1530/REP-17-0649 (2018).

57. Cochran, S. D., Cole, J. B., Null, D. J. & Hansen, P. J. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genet.* https://doi.org/10.1186/s1471-2156-14-49 (2013).

58. Schenkel, F. S., Miller, S. P. & Wilton, J. W. Genetic parameters and breed differences for feed efficiency, growth, and body composi-...
Acknowledgements
The authors acknowledge financial support from the (FDE.13.17) Sustainable Beef and Forage Science Cluster funded by the Canadian Beef Cattle Check-Off, Beef Cattle Research Council (BCRC), Alberta Beef Producers, Alberta Cattle Feeder’s Association, Beef Farmers of Ontario, La Fédération des Producteurs de bovins du Québec, and Agriculture and Agri-Food Canada’s Canadian Agricultural Partnership. This study was also supported by the Ontario Ministry of Agriculture, Food, and Rural Affairs (OMAFRA), Ontario Ministry of Research and Innovation, Agriculture and Agri-Food Canada (AAFC), and Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. Hannah Sweett was supported by the OMAFRA Highly Qualified Personnel Scholarship Program.

Author contributions
A.C. was responsible for the conceptualization, experimental design, and theoretical discussions. P.A.S.F., A.S.-V., and A.L. were responsible for data curation. A.C., H.S. and P.A.S.F. conducted the GWAS, gene annotation, QTL annotation and functional analysis. A.C. was responsible for funding acquisition. All authors have read and agreed to the published version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-75758-3.

Correspondence and requests for materials should be addressed to A.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020