Activation of cryptic splicing in bovine *WDR19* is associated with reduced semen quality and male fertility

Maya Hiltpold[1], Guanglin Niu[2], Naveen Kumar Kadri[1], Danang Crysnanto[1], Zih-Hua Fang[1], Mirjam Spengeler[3], Fritz Schmitz-Hsu[4], Christian Fuerst[5], Hermann Schwarzenbacher[5], Franz R. Seefried[3], Frauke Seehusen[6], Ulrich Witschi[4], Angelika Schnieke[2], Ruedi Fries[7], Heinrich Bollwein[8], Krzysztof Flisikowski[2], Hubert Pausch[1,$]

[1]Animal Genomics, ETH Zürich, 8315 Lindau, Switzerland

[2]Animal Biotechnology, TU München, 85354 Freising, Germany

[3]QualitasAG, 6300 Zug, Switzerland

[4]Swissgenetics, 3052 Zollikofen, Switzerland

[5]ZuchtData, 1200 Wien, Austria

[6]Institute of Veterinary Pathology, University of Zurich, 8057 Zurich, Switzerland

[7]Animal Breeding, TU München, 85354 Freising, Germany

[8]Clinic of Reproductive Medicine, University of Zurich, 8057 Zürich, Switzerland

[$]corresponding author: Hubert Pausch, hubert.pausch@usys.ethz.ch

Email-addresses: MH: maya.hiltpold@usys.ethz.ch

GN: guanglin.niu@wzw.tum.de

NKK: naveen.kadri@usys.ethz.ch

DC: danang.crysnanto@usys.ethz.ch

ZHF: zih-hua.fang@usys.ethz.ch

MS: mirjam.spengeler@qualitasag.ch

FSH: fsc@swissgenetics.ch

CF: fuerst@zuchtdata.at

HS: schwarzenbacher@zuchtdata.at

FRS: franz.seefried@qualitasag.ch

FSU: frauke.seehusen@uzh.ch

UW: uwi@swissgenetics.ch

AS: angelika.schnieke@wzw.tum.de

RF: ruedi.fries@tum.de

HB: hbollwein@vetclinics.uzh.ch

KF: flisikowski@wzw.tum.de

HP: hubert.pausch@usys.ethz.ch

Short title: WDR19 isoform reduces male fertility

Abstract

Cattle are ideally suited to investigate the genetics of male reproduction, because semen quality and fertility are recorded for all ejaculates of artificial insemination bulls. We analysed 26,090 ejaculates of 794 Brown Swiss bulls to assess ejaculate volume, sperm concentration, sperm motility, sperm head and tail anomalies and insemination success. The heritability of the six semen traits was between 0 and 0.26. Genome-wide association testing on 607,511 SNPs revealed a QTL on bovine chromosome 6 that was associated with sperm motility ($P = 2.5 \times 10^{-27}$), head ($P = 2.0 \times 10^{-44}$) and tail anomalies ($P = 7.2 \times 10^{-49}$) and insemination success ($P = 9.9 \times 10^{-13}$). The QTL harbors a recessive allele that compromises semen quality and male fertility. We replicated the effect of the QTL on fertility ($P = 7.1 \times 10^{-32}$) in an independent cohort of 2481 Brown Swiss bulls. The analysis of whole-genome sequencing data revealed that a synonymous variant (BTA6:58373887C>T, rs474302732) in *WDR19* encoding WD repeat-containing protein 19 was in linkage disequilibrium with the fertility-associated haplotype. WD repeat-containing protein 19 is a constituent of the intraflagellar transport complex that is essential for the physiological function of motile cilia and flagella. Bioinformatic and transcription analyses revealed that the BTA6:58373887 T-allele activates a cryptic exonic splice site that eliminates three evolutionarily conserved amino acids from WDR19. Western blot analysis demonstrated that the BTA6:58373887 T-allele decreases protein expression. We make the remarkable observation that, in spite of negative effects on semen quality and bull fertility, the BTA6:58373887 T-allele has a frequency of 24 % in the Brown Swiss population. Our findings are the first to uncover a variant that is associated with quantitative variation in semen quality and male fertility in cattle.

Author summary

In cattle farming, artificial insemination is the most commonly used method of breeding. To ensure high fertilization rates, ejaculate quality and insemination success are closely monitored in artificial insemination bulls. We analyse semen quality, insemination success and microarray-called genotypes at more than 600,000 genome-wide SNP markers in 794 bulls to identify a recessive allele that compromises semen quality. We take advantage of whole-genome sequencing to pinpoint a variant in the coding sequence of *WDR19* encoding WD repeat-containing protein 19 that activates a novel exonic splice site. Our results indicate that cryptic splicing in *WDR19* is associated with reduced male reproductive performance. This is the first report of a variant that contributes to quantitative variation in bovine semen quality.

Introduction

Reproduction plays a pivotal role in dairy and beef production. Delayed conception compromises profit and may lead to the unintended culling of cows and heifers [1–3]. Most cows and heifers are bred using cryoconserved semen from artificial insemination (AI) bulls. Because each AI bull is mated to many cows, factors contributing to conception can be partitioned accurately for males and females using multiple trait animal models [4,5].

The fertility of AI bulls can be quantified using insemination success adjusted for environmental and genetic effects [4,6]. However, the heritability of bull fertility is low [6–8]. Semen traits that are routinely recorded at AI centers such as have higher heritability than bull fertility [9-12]. Traits that are routinely assessed from fresh ejaculates include semen volume, sperm concentration, sperm motility, and the proportion of sperm with head and tail

anomalies. Computer-assisted and flow-cytometric sperm analyses sometimes complement the macroscopic and microscopic evaluations of semen samples [13].

Ejaculates that fulfill the quality requirements for AI [14] are diluted with cryoprotective semen extenders and filled in straws containing between 15 and 25 million spermatozoa per straw. In order to ensure uniform insemination success within and between bulls, the number of spermatozoa per straw is higher for ejaculates that fulfill the minimum requirements but contain some sperm with compensable defects [15].

Routinely recorded semen quality data facilitate investigation of the genetics of male fertility. SNP microarray-derived genotypes of AI bulls [16] can be imputed to the whole-genome sequence level using, e.g., the reference panel of the 1000 Bull Genomes Project consortium [17,18]. Large mapping cohorts with dense genotypes and semen traits provide high statistical power to detect male fertility-associated variants using genome-wide association testing.

Genome-wide association studies using microarray-derived genotypes provided evidence that inherited differences in semen quality are amenable to genome-wide association testing (e.g., [19–25]). However, low marker density resulted in large QTL confidence intervals that contained many genes and candidate causal variants. A missense variant (rs136195618) in the *PROP1* gene was postulated to be associated with bull fertility [26]. However, the effect of the missense variant on bull fertility was not validated in two independent populations [27]. Case-control association studies uncovered recessive variants in the *TMEM95*, *ARMC3* and *CCDC189* genes, that compromise semen quality and bull fertility in the homozygous state [28–30].

Here, we assess semen quality and fertility of 794 Brown Swiss (BSW) bulls using data from 26,090 routinely collected ejaculates. Genome-wide association testing revealed a QTL on bovine chromosome 6 that is associated with semen quality and fertility. Whole-genome sequencing, transcription and protein analyses enabled us to detect a cryptic splice site variant in *WDR19* that is associated with reduced semen quality and fertility.

Results

Semen quality and fertility of BSW bulls

After applying rigorous quality control on the phenotype data including the removal of ejaculates that did not fulfill the requirements for AI, we considered 26,090 ejaculates of 794 BSW bulls for genetic analyses (Table 1). Bull fertility was estimated from female non-return rates at 56 days after the insemination for 591 bulls. Average values for volume, sperm concentration, sperm motility, sperm head and tail anomaly scores, and number of sperm per insemination straw were calculated using between 8 and 162 ejaculates (median: 20) per bull. The average ejaculate volume was 3.93 ml (Table 1). On average, each ejaculate contained 1.32×10^9 sperm per ml and 86 % of the sperm were motile. Each ejaculate received scores between 0 and 3 indicating few and many sperm with head and tail anomalies, respectively. The vast majority of the ejaculates (93.4 and 87.6 %) had very few sperm with head and tail anomalies. On average, each ejaculate was diluted into 319 straws that contained 16.45 million spermatozoa per straw.

Table 1: Semen quality and fertility of 794 BSW bulls

	unit	min	mean (± sd)	max	repeatability	heritability
Ejaculates	n	8	32.86 ± 28.69	162		
Volume	ml	1.57	3.93 ± 0.94	7.31	0.41 ± 0.02	0.27 ± 0.05
Concentration	10^9/ml	0.47	1.32 ± 0.32	2.68	0.43 ± 0.02	0.25 ± 0.06
Motility	%	77.85	86.16 ± 2.33	92.27	0.44 ± 0.01	0.10 ± 0.04
Sperm number per straw	mio	15.00	16.45 ± 1.91	24.85	0.61 ± 0.01	0.12 ± 0.06
Tail anomalies	score	0.00	0.13 ± 0.26	1.53	0.43 ± 0.01	0.02 ± 0.03
Head anomalies	score	0.00	0.05 ± 0.14	0.90	0.38 ± 0.01	0[a]
Bull fertility	index	65	100.6 ± 9.37	127		
Inseminations	n	239	744 ± 1437	15690		

Heritability (h^2) and repeatability were estimated using pedigree-based relationship coefficients and taking into account genetic, non-genetic, environmental and permanent environmental effects (see Material and Methods). [a]The model for head anomalies converged at 10^{-9} as convergence criterion, for all other traits, the convergence criterion was 10^{-15}.

Marked phenotypic correlations were detected between the semen traits (Supporting File 1). Phenotypic correlations were particularly high between head and tail anomalies (r = 0.84), tail anomalies and sperm count per straw (r = 0.68), and motility and tail anomalies (r = -0.75). Motility was positively correlated with bull fertility (r = 0.22). Sperm count per straw, head, and tail anomalies were negatively correlated (r ≤ -0.2) with bull fertility. The repeatability of the semen traits was between 0.38 and 0.61, but the heritability was clearly lower indicating that permanent environmental effects markedly affect semen quality (Table 1). The heritability was close to zero for the proportion of sperm with either head or tail anomalies. The heritability was 0.27, 0.25, 0.12 and 0.10 for ejaculate volume, sperm concentration, sperm count per straw, and motility, respectively, suggesting that genome-wide association testing between semen quality and dense genotypes in 794 bulls should have sufficient statistical power to detect QTL with moderate to large effects [31].

A QTL for semen quality and fertility is located on bovine chromosome 6

To detect QTL for semen quality and fertility, we carried out genome-wide haplotype-based association tests that were based on additive and recessive modes of inheritance. We considered six semen traits and male fertility (Table 1), respectively, of 794 and 591 BSW bulls that had partially imputed genotypes at 607,511 SNPs. In order to take population stratification and the resulting inflation of false-positive association signals into account, we included the top ten principal components of the genomic relationship matrix in the statistical models. The average genomic inflation factor lambda was 1.29 ± 0.19 and 1.12 ± 0.07 for the additive and recessive model, respectively, indicating that this corrective measure was mostly successful.

Haplotype-based association testing (additive model) revealed a QTL on chromosome 6 (Supporting File 2) for sperm motility ($P = 4.1 \times 10^{-12}$), sperm head and tail anomalies ($P = 5.2 \times 10^{-16}$, $P = 1.3 \times 10^{-17}$), sperm count per straw ($P = 8.6 \times 10^{-12}$) and insemination success ($P = 1.5 \times 10^{-8}$). The strongest association signals resulted from three adjacent haplotypes with almost identical P values that were located between 57,335,668 and 57,993,128 bp. The QTL on chromosome 6 was not associated with ejaculate volume and sperm concentration. Only few haplotypes located on chromosomes other than BTA6 met the Bonferroni-corrected significance threshold for the seven traits analysed (Supporting File 2).

The association of the QTL on chromosome 6 was more pronounced when the association tests were carried out assuming recessive inheritance (Figure 1). The strongest associations were detected for two overlapping haplotypes located between 57,538,068 and 57,993,128 bp. The P value of the top haplotype was 9.9×10^{-13}, 2.6×10^{-26}, 2.5×10^{-27}, 2.0×10^{-44}, and

7.2×10^{-49} for bull fertility, sperm count per straw, motility, head anomalies, and tail anomalies, respectively. The P value of the top haplotype was higher (i.e., less significant) for bull fertility than semen quality. This is partly due to the fact that the number of bulls with fertility records was lower than the number of bulls with semen quality records. After correcting for multiple testing, the top haplotype was not associated with ejaculate volume (P = 0.002) and sperm concentration (P = 0.65). The frequency of the top haplotype was 24 %. Of 794 AI bulls, 291 were heterozygous and 46 were homozygous for the top haplotype. The association signal on chromosome 6 was absent for all traits, when the association analysis was conditioned on the top haplotype indicating that the top haplotype fully accounts for the QTL.

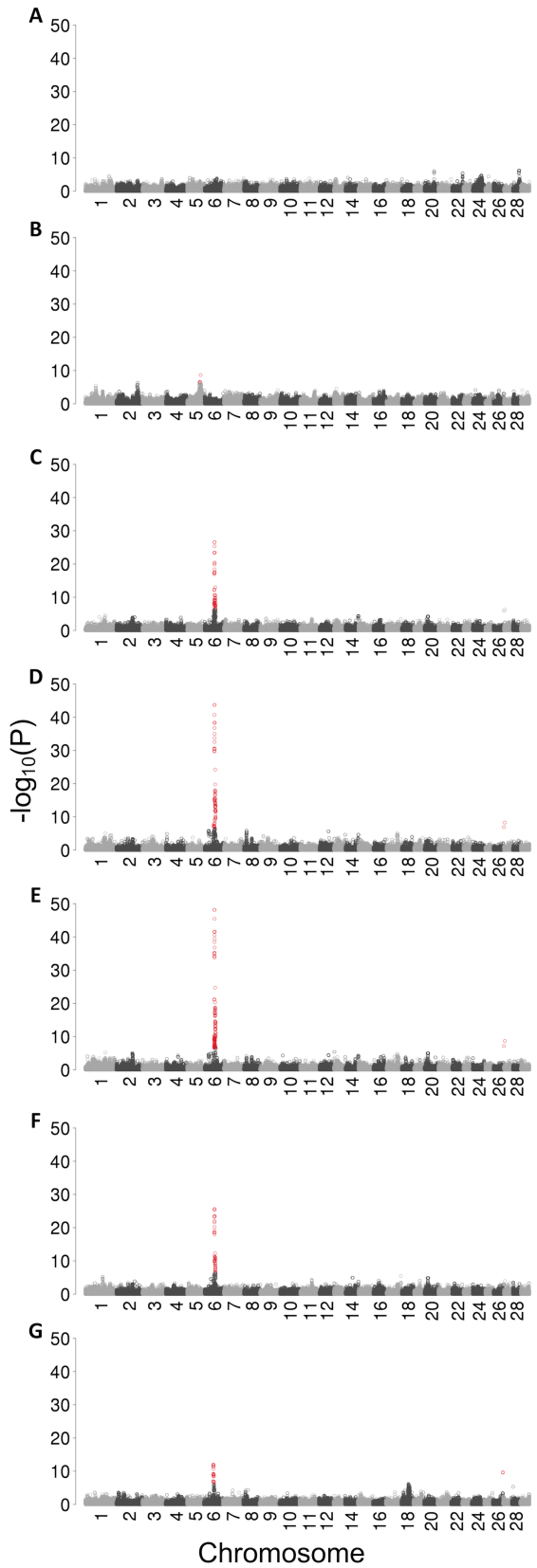

Figure 1: Detection of QTL for semen quality and fertility in BSW bulls. Manhattan plots representing the association ($-\log_{10}(P)$) of haplotypes with (A) ejaculate volume (genomic

inflation factor lambda = 1.25), (B) sperm concentration (lambda = 1.12), (C) sperm motility (lambda = 1.05), (D) sperm head anomalies (lambda = 1.10), (E) sperm tail anomalies (lambda = 1.12), (F) sperm per straw (lambda = 1.16), and (G) bull fertility (lambda = 1.06) assuming a recessive mode of inheritance. Red color indicates significantly associated haplotypes (P < Bonferroni-corrected significance threshold).

Homozygosity for the top haplotype on chromosome 6 is associated with reduced sperm motility and fertility, increased sperm head and tail anomalies, and more sperm per straw (Figure 2). The semen quality and fertility of heterozygous bulls were normal corroborating recessive inheritance. Semen quality of homozygous bulls was only slightly reduced compared to either heterozygous or non-carrier bulls. For instance, with an average proportion of 83 % motile sperm per ejaculate, homozygous bulls fulfill the minimum requirements for artificial insemination which is 70 % [32]. However, more ejaculates were rejected for AI due to less than 70 % motile sperm in homozygous than either heterozygous or non-carrier bulls (4.5 % vs. 1.7 %). Insemination straws of homozygous bulls contain 2.86 million additional sperm per straw. However, the fertility of homozygous bulls is reduced in spite of the compensation (Figure 2E).

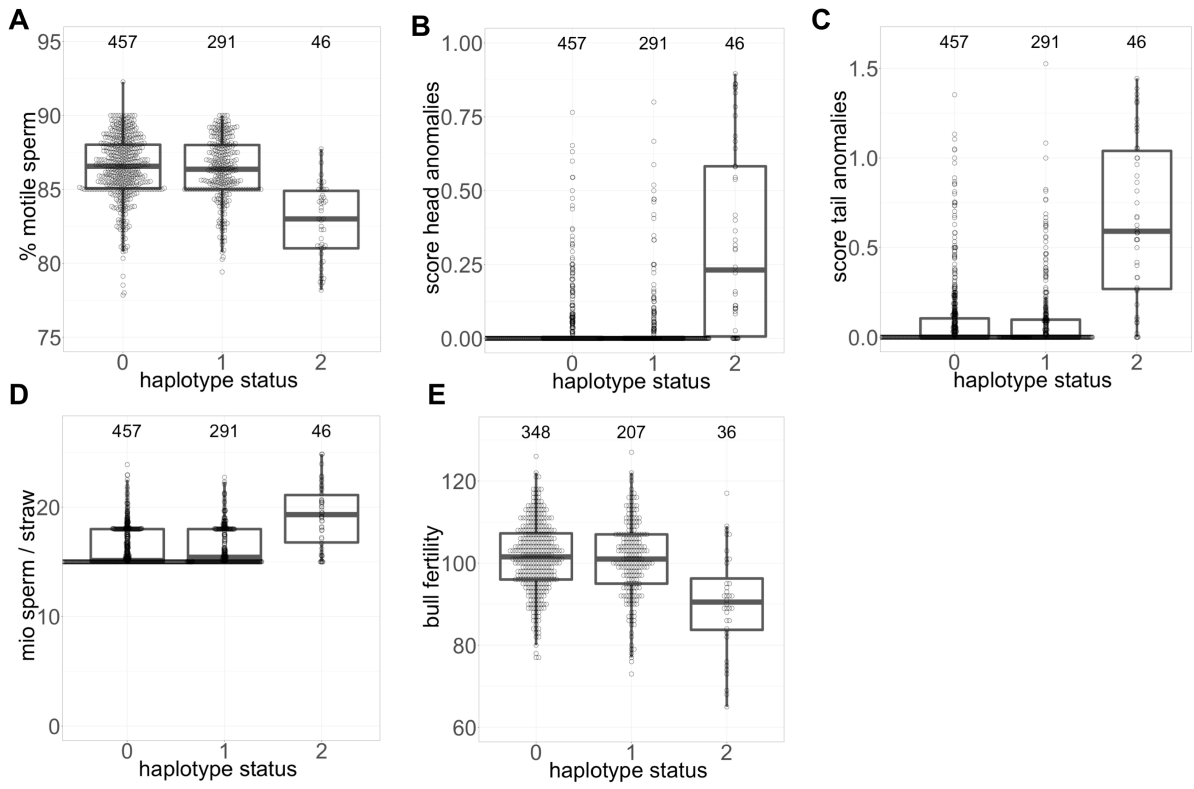

Figure 2: A recessive haplotype compromises semen quality and bull fertility. Boxplots representing the effect of the top haplotype on semen quality and fertility in non-carrier, heterozygous and homozygous bulls (haplotype status 0, 1, and 2, respectively). (A) Sperm motility of homozygous bulls is reduced by 1.6 phenotypic standard deviations (σ_p) (82.77 ± 2.59 % vs. 86.37 ± 2.14 %). Scores for sperm head (B) and tail (C) anomalies are increased by 2 and 2.1 σ_p in homozygous bulls. (D) The reduced semen quality of homozygous bulls is compensated with an increased number of sperm per dose (19.15 ± 2.83 vs. 16.29 ± 1.7 million). (E) Fertility of homozygous bulls is reduced by 1.2 σ_p (90.1 ± 12.3 vs. 101.27 ± 8.74).

Homozygous haplotype carriers share a 2.38 Mb segment on BTA6

Haplotypes that were significantly associated with either semen quality or bull fertility were detected in an 15 Mb interval (between 50 and 65 Mb) on chromosome 6. One hundred and five haplotypes that were significantly associated with at least two of the seven traits studied were located between 55,348,382 and 65,408,468 bp (Figure 3A). We detected haplotypes that had markedly lower P values (e.g., $P < 1 \times 10^{-30}$ for sperm head and tail anomalies, $P < 1 \times 10^{-15}$ for motility) than surrounding haplotypes between 56,922,962 and 58,293,842 bp. A deeper analysis of the genotypes of 46 bulls that were homozygous for the top haplotype showed that they share a 2.38 Mb segment of extended autozygosity (between 57,465,157 and 59,846,532 bp) (Figure 3B). This segment contains 23 genes (*KLF3*, *TLR10*, *TLR6*, *FAM114A1*, *ENSBTAG00000055220*, *TMEM156*, *KLHL5*, *WDR19*, *RFC1*, *KLB*, *RPL9*, *LIAS*, *UGDH*, *SMIM14*, *UBE2K*, *PDS5A*, *N4BP2*, *ENSBTAG00000049669*, *RHOH*, *CHRNA9*, *RBM47*, *NSUN7*, *APBB2*) (Figure 3C). Of these genes, 15 and 14 are expressed at medium to high levels (> 10 transcripts per million (TPM)) in the testes of newborn and mature males, respectively (Supporting File 3) including *NSUN7* encoding NOP2/Sun RNA Methyltransferase Family Member 7. Loss-of-function alleles in *NSUN7* have been associated with low sperm motility and impaired male fertility in mice and humans [33],[34].

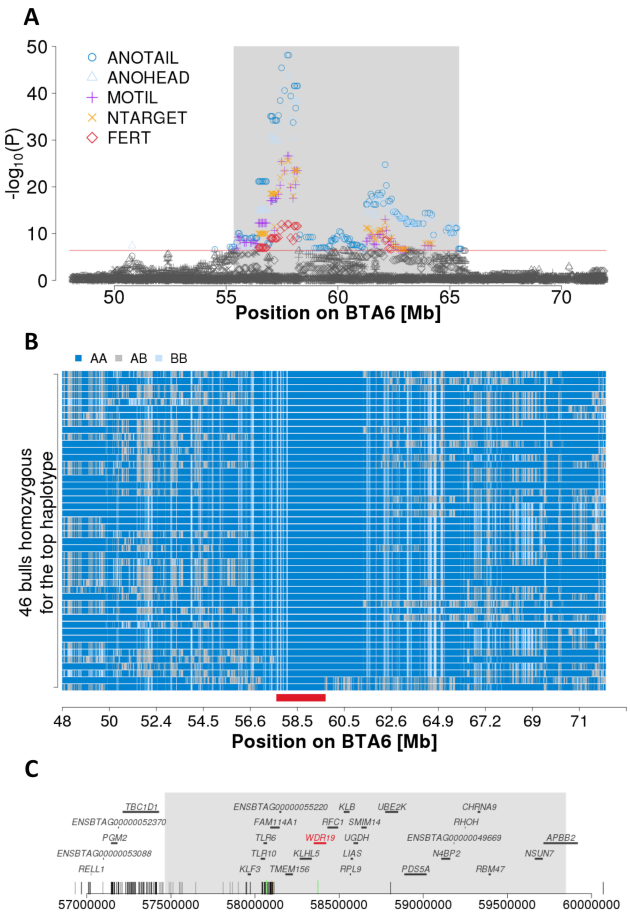

Figure 3: Detailed view of the associated region on chromosome 6. (A) Association (-$\log_{10}$(P)) of haplotypes located between 48 and 72 Mb on BTA6 with head and tail anomalies (ANOHEAD, ANOTAIL), motility (MOTIL), number of sperm per straw (NTARGET) and bull fertility (FERT) assuming recessive inheritance. The red line represents the Bonferroni-corrected significance threshold. The grey shaded area contains haplotypes that were associated with at least two traits. (B) A segment of extended autozygosity was detected in 46 bulls that were homozygous for the top-haplotype. Blue and pale blue represent homozygous genotypes (AA and BB), heterozygous genotypes are displayed in grey. The solid red bar indicates the common 2.38 Mb segment of extended autozygosity (from 57,465,157 to 59,846,532 bp). (C) The segment of extended autozygosity encompasses 23 protein-coding genes. Vertical bars at the bottom represent variants that are compatible with recessive inheritance, including 126 that are located within the segment of extended autozygosity. Dark

grey, light grey, green and orange bars represent intronic, intergenic, synonymous and missense variants, respectively.

A synonymous variant in *WDR19* is in linkage disequilibrium with the top haplotype

We analysed whole-genome sequencing data of 42 BSW animals for which the status for the BTA6 top haplotype was known from SNP microarray genotypes. Of the 42 BSW animals, 3 were homozygous and 10 heterozygous carriers of the top haplotype. The remaining 29 bulls did not carry the top haplotype (Supporting File 4). The average sequencing depth of the 42 animals was 13.9-fold. Analysis of the sequencing read alignments and sequencing depth along chromosome 6 in three homozygous haplotype carriers using the *Integrative Genomics Viewer* [35] and *mosdepth* software [36] did not reveal large sequence variants that segregate with the fertility-associated haplotype.

As causal variants might reside outside of top haplotypes (e.g., [37]), we considered a 10 Mb interval (between 55 and 65 Mb, cf. Figure 3A) that contained haplotypes that were associated with at least two of the traits studied to identify variants that were compatible with recessive inheritance. Within this interval, we detected 56,081 variants (47,686 SNPs and 8,395 Indels) that were polymorphic in the 42 BSW animals. A variant filtration approach that takes into account potential sequencing errors, flawed genotypes in animals with low sequencing coverage, and misclassified haplotypes (see Material and Methods), yielded 824 variants that were compatible with the inheritance pattern of the top haplotype (Supporting File 5). Of the compatible variants, 126 were located within the 2.38 Mb segment of extended autozygosity.

The same filtration approach was applied to 57 larger sequence variants (insertions, deletions, duplications, inversions, translocations) that were detected between 55 and 65 Mb on BTA6 in the 42 BSW animals. However, none of these variants was compatible with the recessive inheritance of the top haplotype.

Because homozygosity for the top haplotype was associated with reduced semen quality and fertility, we hypothesized that the causal variant might reside in the coding sequence of a gene that is expressed in the testis tissue. Only four of 824 compatible variants were located in protein-coding regions: three synonymous variants in *TLR6* encoding Toll-like receptor 6 (rs68268274 at Chr6:58069459), in *WDR19* encoding WD repeat-containing protein 19 (rs474302732 at Chr6:58373887) and in *GABRA2* encoding Gamma-aminobutyric acid receptor subunit alpha-2 (rs42595907 at Chr6:64905616), and a missense variant in *FAM114A1* encoding family with sequence similarity 114 member A1 that had a SIFT score of 0.18 indicating that this amino acid change is tolerated (rs382246003 at Chr6:58099868, ENSBTAP00000044312.1:p.Asn22Lys) (Supporting File 5). The *TLR6*, *WDR19* and *FAM114A1* genes are within the 2.38 Mb segment of extended autozygosity. *TLR6*, *WDR19* and *FAM114A1* are expressed at 4, 16 and 27 TPM, respectively, in testicular tissue of mature bulls (Supporting File 3). The *GABRA2* gene is not expressed (TPM < 1) in testicular tissue of mature bulls. Moreover, the synonymous variant in *GABRA2* was more than 5 Mb from the segment of extended autozygosity, suggesting that it is less likely causal for the reduced semen quality and fertility of homozygous BSW bulls. We did not detect any sequence variants nearby the *NSUN7* gene neither in coding nor in non-coding sequences that were compatible with the inheritance pattern of the top haplotype (Figure 3C, Supporting File 5).

Using data from our in-house variant database and the latest variant discovery and genotyping run (run 7) of the 1000 Bull Genomes Project (http://www.1000bullgenomes.com/), we investigated the genotype distribution of the compatible variants in cattle from various breeds (Supporting File 5 & Supporting File 6). It turned out that the variants in *TLR6*, *GABRA2* and *FAM114A1* frequently occur in either heterozygous or homozygous state in cattle from various breeds including Fleckvieh. Such variants are not likely to be causal, because no QTL for semen quality and bull fertility had been detected in Fleckvieh nearby the top haplotype on bovine chromosome 6 [25,28]. In contrast, the Chr6:58373887 T-allele that is located in the coding sequence of *WDR19* was only detected in BSW cattle and in the heterozygous state in one bull (SAMEA5064547) of the Nordic Red Dairy cattle breed. The Nordic Red Dairy cattle breed has recently experienced considerable introgression of BSW haplotypes [38].

The BTA6:58373887 T-allele in *WDR19* activates cryptic splicing

A closer inspection of the BTA6:58373887C>T variant in the *WDR19* gene (ENSBTAG00000014512) revealed that it is located at the 3' end of exon 12, 8 bp from the splice donor site of intron 12. Although annotated as synonymous variant, we hypothesized that the T-allele might activate a novel exonic splice donor site (Figure 4A & B). *In silico* prediction corroborated that the BTA6:58373887 T-allele very likely activates cryptic splicing (prediction score: 0.88) that eliminates 9 bp from exon 12 and causes the in-frame deletion of 3 amino acids in the WD repeat-containing domain of WDR19. WDR19 is part of the intraflagellar transport complex A that is essential for the physiological function of motile cilia and flagella [39,40]. Thus, we considered the BTA6:58373887 T-allele in the *WDR19* gene as a plausible candidate mutation for the reduced semen quality and fertility of the BSW bulls.

To investigate functional consequences of the BTA6:58373887C>T variant, we sampled testicular tissue of one wild-type, one heterozygous and two homozygous mutant AI bulls from which between 24 and 93 ejaculates had been collected and analysed. The average proportion of motile sperm was 40 and 69 % in ejaculates of the two homozygous bulls. Sperm motility was higher in non-carrier (80 %) and heterozygous (84 %) bulls. The proportion of ejaculates that contained many sperm with either head or tail anomalies (scores 2 and 3) was clearly higher in the two homozygous bulls (40 and 88 %) than the heterozygous (0 %) or non-carrier bull (2 %).

Histological sections of testicular tissue of two homozygous bulls showed no apparent pathological structures that might cause reduced semen quality. We extracted RNA from testicular tissue in order to examine *WDR19* transcription by reverse transcription PCR (RT-PCR). Using primers located in exons 12 and 13, we obtained a 169 bp RT-PCR product in the wild-type bull (Figure 4A - C). In bulls homozygous for the mutant BTA6:58373887 T-allele, we primarily detected 160 bp RT-PCR products. Sequence analysis of the 160 bp RT-PCR products confirmed that the mutant T-allele activates a cryptic exonic splice donor site resulting in 9 bp shorter sequence of exon 12 (Figure 4C). We also detected a weak band at 169 bp in homozygous animals indicating the presence of the wild-type length RT-PCR product. Sequence analysis confirmed that bulls homozygous for the BTA6:58373887 T-allele express the wild-type length transcript at low levels. Both the wild-type and mutant RT-PCR products were detected at approximately similar amounts in the heterozygous bull.

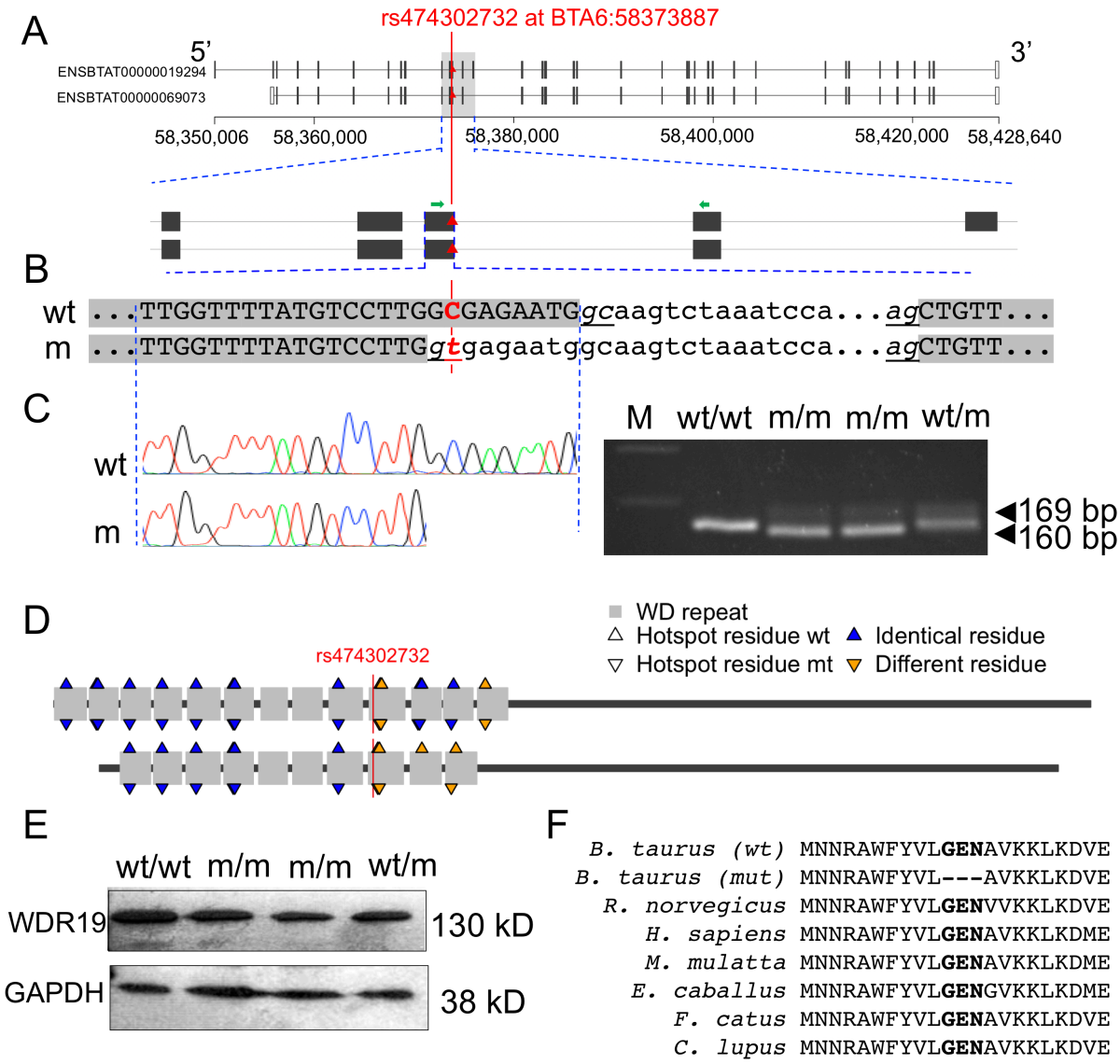

Figure 4: Effect of the BTA6:58373887 C/T variant on WDR19 mRNA and protein expression. (A) Two WDR19 isoforms with 1342 (ENSBTAT00000019294.6) and 1242 (ENSBTAT00000069073.1) amino acids are annotated in cattle. The BTA6:58373887C>T variant is located at the 3' end of exon 12 (ENSBTAT00000019294.6), 8 basepairs from the splice donor site of intron 12. The red triangle indicates the 58373887C>T variant, and green arrows indicate RT-PCR primers. (B) Genomic sequence surrounding the 58373887C>T-variant. Grey background indicates exons of the wild type (wt) and mutant (mt) transcripts, respectively. Underlined nucleotides indicate splice donor and splice acceptor sites. Upper and lower case letters indicate exonic and intronic nucleotides, respectively. (C) RT-PCR

analysis on testis tissue samples from wild type, mutant and heterozygous bulls. Resequencing of the transcripts confirmed that the BTA6:58373887 T-allele activates a novel exonic splice site that is predicted to eliminate 3 amino acids from the resulting protein. (D) The mutation is located in the WD repeat domain of WDR19. The long (ENSBTAT00000019294.6, upper figure) and short (ENSBTAT00000069073.1, lower figure) WDR19 isoform contains 13 and 10 WD repeats, respectively. Residues at predicted amino acid hotspot positions differ between the wild-type and mutated protein for both WDR19 isoforms. (E) Western Blot analysis in wild type, mutant and heterozygous bulls. GAPDH was used as the control. (F) Multi-species alignment of the WDR19 protein sequence. Bold type indicates residues that are missing in the mutated (mt) WDR19 protein. Protein sequences were obtained from Ensembl for *Bos taurus* (ENSBTAT00000019294.6), *Rattus norvegicus* (ENSRNOT00000003991.6), *Homo sapiens* (ENST00000399820.8), *Macaca mulatta* (ENSMMUT00000003922.4), *Equus caballus* (ENSECAT00000026479.2), *Felis catus* (ENSFCAT00000060823.2) and *Canis lupus familiaris* (ENSCAFT00000025574.4).

In order to quantify the abundance of the two WDR19 isoforms, we generated 63 million 150 bp sequencing reads from RNA prepared from testicular tissue sampled from one BSW bull (sample accession number SAMN14485268) homozygous for the BTA6:58373887 T-allele. Of 29 RNA-seq reads that overlap the junction of exons 12 and 13, 5 and 24 correspond to the wild-type and mutant transcript, respectively, indicating that the wild-type and mutant isoforms are expressed at a ratio of 1:5 (Supporting File 7).

The expression of *NSUN7* was intact in the BSW bull homozygous for the BTA6:58373887 T-allele (Supporting File 8).

Bioinformatic analysis of the sequences of the RT PCR products revealed that the altered sequence of exon 12 leads to the in-frame deletion of three evolutionarily conserved amino acids (positions 414 – 416) of the WDR19 protein (ENSBTAT00000019294.6) (Figure 4F). The affected amino acids are located in the tenth WD repeat unit. The elimination of three amino acids is predicted to alter hotspot residues of WDR19 that are supposed to be relevant for protein-protein interactions and the assembly of functional complexes [41] (Figure 4D).

Next, we analysed the effect of the alternative exon 12 sequence on the *WDR19* mRNA and protein expression in testicular tissue. Quantitative PCR analysis with primers located in exons 15 and 16 revealed no differences among the analysed samples indicating that the BTA6:58373887 T-allele does not affect the *WDR19* mRNA expression. However, western blot analysis revealed that the amount of WDR19 is reduced in testicular tissue of heterozygous and even stronger in homozygous mutant bulls, indicating that the BTA6:58373887 T-allele is associated with WDR19 protein production (Figure 4E).

The QTL on BTA6 can be validated in BSW bulls from Germany and Austria

In order to validate the effect of the fertility-associated haplotype in an independent population, we analysed genotype and fertility data of 2481 BSW bulls from Germany and Austria. The German, Austrian and Swiss BSW bulls are genetically connected. However, the bulls from Austria and Germany were housed at different AI centers than the Swiss bulls and their fertility was estimated using a separate genetic evaluation system, thus they qualify as validation cohort. Ejaculate data were not available for the German and Austrian bulls precluding the analysis of semen quality.

The haplotype-based association study (recessive model) yielded strong association of the QTL on BTA6 (Figure 5A). An inflation factor of 1.03 indicated that population stratification was properly taken into account. On BTA6, we detected 69 haplotypes between 55,423,341 and 64,927,963 bp that exceeded the Bonferroni-corrected significance threshold ($P < 4.4 \times 10^{-7}$). The most significantly ($P = 7.09 \times 10^{-32}$) associated haplotype was located between 57,335,668 and 57,624,763 bp. The frequency of this haplotype was 26% in the German and Austrian BSW population, i.e. slightly higher than the frequency of the top haplotype in the Swiss population. 127 and 1022 bulls from Germany and Austria were homozygous and heterozygous haplotype carriers, respectively.

The haplotype that showed the strongest association in the Swiss population (between 57,538,068 and 57,993,128 bp) was the second top haplotype in the German and Austrian BSW population and the P value was only slightly higher ($P = 7.8 \times 10^{-30}$) compared to the top haplotype. The fertility of homozygous bulls was lower than that of heterozygous and non-carrier bulls (Figure 5B). The association signal on chromosome 6 was absent when the association analysis was conditioned on the top haplotype from the Swiss population, indicating that the BTA6:58373887 T-allele also tags the QTL detected in the German and Austrian BSW population.

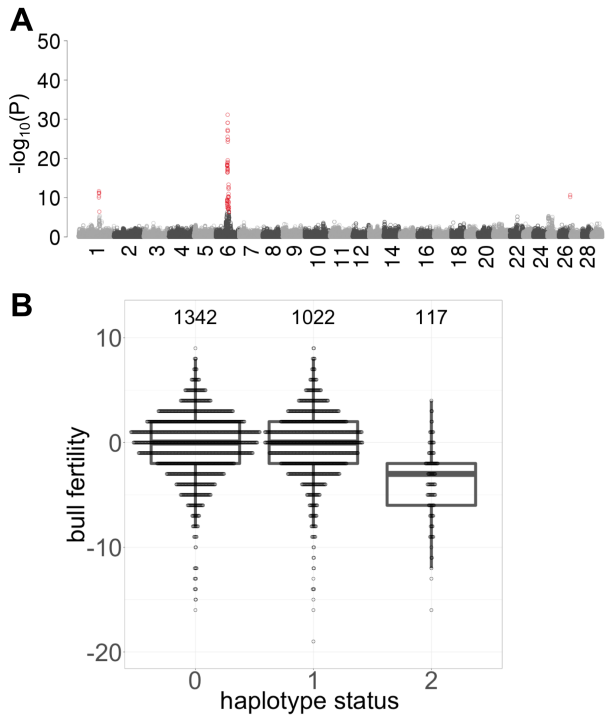

Figure 5: Detection of QTL for male fertility in 2481 Austrian and German BSW bulls. (A) Manhattan plot representing the association ($-\log_{10}(P)$) of 112,667 haplotypes with male fertility assuming a recessive mode of inheritance. Red color indicates significantly associated haplotypes ($P <$ Bonferroni-corrected significance threshold). (B) The fertility of homozygous haplotype carriers is reduced by 1.01 phenotypic standard deviations.

Discussion

Our haplotype-based association study uncovered a QTL on BTA6 for semen quality and fertility in BSW cattle. Response variables for genome-wide association testing were derived from 33 ejaculates per bull in average. Thus, our results confirm that variation in semen quality is partly due to genetic differences that can be identified in genome-wide association studies [42]. It is well known that bulls suffering from undetected diseases and stress may temporarily produce ejaculates that do not fulfill the requirements for artificial insemination [43,44]. Bulls that produce ejaculates of low quality despite being healthy might carry rare genetic variants that can be identified using case-control association testing [29,30]. We considered only ejaculates that fulfill minimum requirements for AI. The exclusion of ejaculates with low semen quality likely removes phenotypic outliers that could lead to false-positive association signals [45]. Low genomic inflation factors indicated that the results of our association studies were not enriched for spurious associations. However, we detected only very few haplotypes on chromosomes other than BTA6 that exceeded the Bonferroni-corrected significance threshold. Our rigorous approach to retain only ejaculates that comply with the requirements of AI may have reduced the amount of genetic variation that is amenable to association testing, thus possibly reducing the statistical power to detect QTL with small effects. Moreover, the heritability of semen quality was less than 0.26 for all traits considered, which agrees with previous results in cattle [9-12]. The repeatability was higher than heritability for all traits analysed indicating that the semen quality of AI bulls is markedly influenced by permanent environmental effects that may also capture non-additive genetic effects [46], which agrees with previous findings in boars and bulls [9],[47]. Detecting QTL with small to medium effects requires large mapping populations particularly for traits with low heritability [31]. Our mapping cohort consisted of 794 BSW bulls that were used for AI in the past 20 years. The size of this cohort is too small to detect QTL for

semen quality that are either rare or explain only a small fraction of the trait variation. Thus, our findings suggest that, apart from the QTL on BTA6, variants with small effects contribute to the genetic variance of semen quality from BSW bulls, otherwise they should have been detected in our study. The analysis of data from multiple countries might increase the statistical power to detect trait-associated variants and provide additional insight into the genetic determinants of quantitative variation in semen quality. However, this requires a uniform assessment of semen quality across AI centers.

The QTL detected on BTA6 acts in a recessive manner. Ejaculates of bulls homozygous for the top haplotype contain an increased proportion of sperm with head and tail anomalies and lower proportion of motile sperm. However, the semen quality of homozygous bulls is only slightly reduced in comparison to the average semen quality of BSW bulls. Thus, the BTA6 QTL contributes to the quantitative variation in semen quality and male fertility. Most ejaculates of bulls that are homozygous for the top haplotype fulfill the requirements for AI. Because the reduced semen quality of homozygous bulls is recognized during the routine quality assessment at the AI center, the insemination straws produced from their ejaculates contained more sperm per dose than standard doses. Increasing the number of sperm per dose typically increases the fertilization rates for ejaculates that contain sperm with compensable defects [15]. However, the insemination success of homozygous bulls is significantly reduced, indicating that an increased number of sperm per dose is not sufficient to enable normal fertility in homozygous haplotype carriers. Moreover, ejaculates from homozygous bulls are more often discarded due to insufficient semen quality than ejaculates produced by either heterozygous or non-carrier bulls. Because we did not consider discarded ejaculates to calculate average semen quality, the actual effect of the haplotype on semen quality might be greater than estimated in our study. Nevertheless, it is important to emphasize that

homozygous bulls are fertile. Therefore, the consequences of the BTA6 haplotype on male reproduction are less detrimental than, e.g., of recessive haplotypes that severely compromise semen quality and lead to male infertility [28–30].

We make the remarkable observation that the BTA6 haplotype compromising male fertility segregates in the BSW population at a frequency of 24 %. Balancing selection may maintain deleterious alleles at high frequency in natural populations [48]. However, if the haplotype has desirable effects on economically relevant traits, the effects must be very small; otherwise, they would have been detected in a recent study that carried out genome-wide association studies for more than 50 economically important traits in 4578 BSW bulls [49]. We cannot exclude the possibility that the haplotype affects traits that are not routinely recorded in BSW cattle. Hitchhiking with favourable alleles may also explain the high frequency of the deleterious haplotype on BTA6 [50]. Yet, the haplotype is more than 20 million basepairs from the *NCAPG* gene and the casein cluster, which are targets of selection for stature and milk production in many dairy cattle breeds [51,52]. Signatures of selection were not detected in BSW cattle nearby the male fertility-associated QTL on BTA6 [49,53]. Random drift and the frequent use of haplotype carriers in AI most likely propagated the haplotype in spite of its negative effect on male reproduction. Founder effects caused the frequent phenotypic manifestation of deleterious recessive alleles also in the Fleckvieh and Belgian Blue cattle breeds [28,54]. Another reason that likely contributed to the high haplotype frequency is the fact that reduced fertility becomes evident only in homozygous AI bulls. Nevertheless, close monitoring of the haplotype seems warranted in order to prevent a further increase in frequency and the emergence of many homozygous bulls with low semen quality and fertility.

Low effective population size and the resulting long-range linkage disequilibrium typically result in large QTL confidence intervals that contain many candidate causal variants. We detected significantly associated haplotypes that were in linkage disequilibrium with the top haplotype within a 15 Mb interval on chromosome 6 (Figure 3A). Multi-breed association analysis can be used to refine QTL because linkage disequilibrium is conserved over shorter distances across breeds than within breeds. However, association testing of semen quality and fertility across breeds was not possible because ejaculate data were only available for BSW bulls. Using whole-genome sequencing data of 42 BSW cattle with known haplotype status at the BTA6 QTL, we identified 824 variants that were compatible with the inheritance of the top haplotype. Variants segregating in breeds other than BSW can be discarded as candidate causal variants if the trait-associated mutation occurred after the formation of breeds. A filtration approach based on this assumption has been frequently applied to detect causal variants for recessive conditions in livestock and companion animals [37,55,56]. The removal of variants that segregate in breeds other than BSW would have excluded many of the 824 variants as potential candidate causal variants for the impaired fertility associated with homozygosity for the BTA6 QTL. Nevertheless, such a variant filtration approach does not take into account that deleterious alleles may segregate across populations [57–60]. It is possible that the QTL detected in our study also segregates in breeds other than BSW. However, QTL for semen quality and bull fertility were not detected in Fleckvieh cattle nearby the BTA6 QTL [25,28]. Thus, variants that occur at high frequency also in the Fleckvieh population are less likely to be causal for impaired reproductive performance. We applied an appropriate variant filtration approach to prioritize a variant (BTA6:58373887) in the coding sequence of the *WDR19* gene. Although annotated as synonymous variant, our analyses demonstrated that the BTA6:58373887 T-allele activates a novel exonic splice site that eliminates 3 amino acids from the resulting protein. Thus, our results indicate that

nondescript synonymous variants warrant close scrutiny because they might affect exonic splicing enhancers [61] or activate cryptic splice sites, which is not immediately apparent from standard variant annotation as it was the case for the BTA6:58373887 T-allele in *WDR19*.

The WD repeat-containing protein 19 is part of the intraflagellar transport complex A (IFT-A) that is essential for the assembly and physiological function of motile cilia and flagella [39,40,62,63]. Loss-of-function mutations in the WD repeat domain of the *WDR19* gene lead to severe ciliopathies in humans [64–67]. The consequences of these mutations on semen quality and male fertility are unknown, because affected individuals do not reach reproductive age [68]. The BTA6:58373887 T-allele is not a loss-of-function mutation because we detected the allele in the homozygous state in mature bulls, which were healthy apart from producing ejaculates with reduced semen quality and fertility. However, an excess of sperm with morphological anomalies and low motility indicates that spermatogenesis of bulls homozygous for the BTA6:58373887 T-allele is impaired. Our results show that *WDR19* is expressed in the testes of mature bulls. The BTA6:58373887 T-allele activates alternative splicing that eliminates 3 amino acids from the resulting protein. The loss of evolutionary conserved amino acids within the WD repeat domain alters hotspot residues at the surface of the WDR19 protein, which might compromise its secondary structure and interaction with other proteins including the constituents of the IFT-A complex [41,69]. We observed markedly reduced WDR19 protein expression in testicular tissue of heterozygous and homozygous bulls indicating that changes in the secondary structure may compromise protein stability and result in either protein degradation or post-translational modification of this WDR19 isoform. A lower amount of WDR19 may compromise spermiogenesis due to impaired intraflagellar transport thus causing abnormal sperm with low fertility in

homozygous bulls. We also detected the wild-type length transcript at low levels in bulls homozygous for the BTA6:58373887 T-allele indicating differential splicing. Factors controlling splice site selection may contribute to the marked phenotypic variability in semen quality in bulls homozygous for the BTA6:58373887 T-allele. The presence of a low amount of wild-type WDR19 protein seems to be sufficient to produce normal sperm. However, semen quality of individuals that exceed a certain threshold of mutant protein or undercut a certain threshold of wild-type protein might have a higher proportion of abnormal sperm, thus resulting in lower fertility.

This is the first report providing evidence that naturally occurring variation in *WDR19* is associated with reduced semen quality and fertility. However, variants affecting other constituents of the IFT-A complex have been shown to compromise male reproduction. Mice with a loss-of-function variant in the *IFT140* gene show multiple morphological aberrations of the sperm including short and/or bent sperm tail and abnormal heads [70]. Compound heterozygosity in human *IFT140* is associated with reduced sperm count and an excess of sperm with head and tail anomalies in an otherwise healthy individual [71]. This indicates that mutations of the IFT-A complex may reduce semen quality without compromising the general health, which agrees with our findings of a deleterious allele in bovine *WDR19* that affects semen quality and fertility in otherwise healthy bulls.

Methods

Ethics approval and consent to participate

Semen was collected at approved AI centers. Testis tissue was collected after regular slaughter at an approved slaughterhouse. The decision to slaughter the bulls was independent from our study. None of the authors of the present study was involved in the decision to slaughter the bulls. No ethics approval was required for the analyses.

Consent for publication

Breeding associations and AI centers provided written consent to the analyses performed and the publication of results and data.

Genotypes of Brown Swiss bulls

Genotypes of 4549 BSW bulls were provided by Swiss, German and Austrian breeding associations. A subset (N = 870) of the bulls was genotyped using the Illumina BovineHD (HD) bead chip that comprises 777,962 SNPs. All other bulls (N = 3679) were genotyped at approximately 50,000 SNPs using medium density (MD) chips (e.g., the Illumina BovineSNP50 bead chip that comprises 54,001 (version 1) or 54,609 (version 2) SNPs). The position of the SNPs corresponded to the ARS-UCD1.2-assembly of the bovine genome [72]. Quality control on the genotype data was carried out separately for the HD and MD datasets using the *plink* (version 1.9) software [73]. Animals and SNPs with more than 20 % missing genotypes were excluded from the data. We removed SNPs with minor allele frequency (MAF) less than 0.005 and SNPs for which the observed genotype distribution deviated significantly ($P < 0.00001$) from Hardy-Weinberg proportions. After quality control, 36,131 and 607,511 SNPs remained in the MD and HD dataset, respectively. Sporadically missing

genotypes were imputed in the MD and HD datasets separately using the *Beagle* (version 5.0) software [74]. Subsequently, we inferred haplotypes for both datasets using the *Eagle* (version 2.4) software [75]. We considered the HD genotypes of 870 bulls as a reference to impute MD genotypes to higher density using the haplotype-based imputation approach implemented in the *Minimac3* (version 2.0.1) software [76]. Following imputation, the genotype panels were merged to obtain the final dataset consisting of 4549 bulls genotyped at 607,511 autosomal SNPs. Principal components of the genomic relationship matrix were calculated using the *plink* (version 1.9) software [73].

Semen quality data of Swiss BSW bulls

The Swiss AI center Swissgenetics provided data on 70,990 ejaculates that were collected between January 2000 and March 2018 from 1,343 BSW bulls at the AI center in Mülligen, canton of Aargau, Switzerland. All ejaculates were collected as part of the breeding and reproduction service of the AI center to Swiss cattle farming. Semen quality was assessed by lab technicians immediately after ejaculate collection in order to identify and discard ejaculates of low quality. The parameters recorded for fresh ejaculates were semen volume (in ml), sperm concentration (million sperm cells per ml) quantified using photometric analysis, and sperm motility (percentage of sperm with forward motility) assessed visually using a heated-stage microscope at 200-fold magnification. Moreover, the presence of erythrocytes, leucocytes and other non-sperm cells was documented for each ejaculate. Each ejaculate received a score between 0 and 3 indicating the proportion of sperm with head and tail anomalies (0: no or very few anomalies, 1: less than 10 % sperm with anomalies, 2: between 10 and 30 % sperm with anomalies, 3: more than 30 % sperm with anomalies). Ejaculates that fulfilled minimum requirements for artificial insemination (semen volume above 1 ml, more than 300 million sperm per ml, at least 70 % motile sperm, no apparent

impurities and no excessive abnormalities of sperm heads and tails) were diluted using a Tris-egg yolk based extender, filled in straws containing between 15 and 25 million sperm cells and cryoconserved in liquid nitrogen.

We removed records of ejaculates that were pooled before semen analysis and considered only the first ejaculate per day that was collected from bulls between 400 and 1000 days of age (sometimes more than one ejaculate is collected per bull and day). Records for which either the interval between successive ejaculates or the semen collector were missing were excluded from our analysis. We retained only ejaculates that complied with minimum requirements for artificial insemination because insufficient semen quality might be attributable to rare genetic conditions [28,29]. Eventually, we retained only ejaculates of bulls for which at least 8 ejaculates were available. Our final dataset contained 26,090 ejaculates for 794 genotyped bulls (ø 33 ± 29 ejaculates per bull; median: 20) (Table 2). We considered six semen quality traits for our genetic investigations: ejaculate volume, sperm concentration, sperm motility, sperm head and tail anomaly score, and number of sperm filled per insemination straw.

Table 2: Quality control on the raw semen quality data

Filtering parameter	Number of ejaculates	Number of BSW bulls
Raw dataset	70,990	1343
Interval to preceding ejaculate known (excluding first ejaculate of bull in dataset)	69,647	1343
Age at semen collection between 400 and 1000 days	56,200	1302
Ejaculate volume recorded	47,956	1294
Fresh sperm motility recorded	47,941	1294
No (biological / technical) cause of rejection recorded	43,733	1250
Fresh sperm motility ≥ 70 %	43,571	1235
Ejaculate was not pooled before semen analysis	35,888	1224
Only first ejaculate per day	32,584	1221
Semen collector recorded	30,966	1201
Plausible sperm head and tail anomaly score	30,964	1201
Target number of sperm in insemination straw between 15 and 25 million	30,832	1201
At least 8 ejaculates collected per bull	29,751	946
Bull had (partially imputed) genotypes	26,090	794

Estimation of genetic parameters for semen quality

The heritability and repeatability of six semen quality traits was calculated using the average information restricted maximum likelihood (REML) estimation algorithm implemented in the *AIREMLF90* program [77]. We fitted the following univariate linear mixed model: $\mathbf{y} = \mathbf{1}\mu + \mathbf{h}a + \mathbf{d}f + \mathbf{Ct} + \mathbf{Sn} + \mathbf{Z}_u\mathbf{u} + \mathbf{Z}_p\mathbf{p} + \mathbf{e}$, where **y** is a vector of each phenotype tested, μ is the intercept (**1** is a vector of ones), **h** and **d** are vectors of age and interval between successive semen collections in days and a and f are their respective effects; **C**, **S**, $\mathbf{Z}_u$ and $\mathbf{Z}_p$ are incidence matrices relating collector (person collecting the ejaculate), season (4 seasons per year), random individual polygenic and random permanent environment effects to their respective records respectively, **t** is the vector of effects of different semen collectors, **n** is the vector of effects of different seasons, **u** is the vector of random individual polygenic effects assumed to be normally distributed ($N(0, \mathbf{A}\sigma_g^2)$), where **A** is the additive relationship matrix

estimated from the pedigree and σ_g^2 is the additive genetic variance), **p** is the vector of permanent environmental effects assumed to be normally distributed ($\mathrm{N}(0, \mathbf{I}\sigma_{\mathrm{pe}}^2)$, where σ_{pe}^2 is the permanent environment variance), and **e** is the vector of individual error terms assumed to be normally distributed ($\mathrm{N}(0, \mathbf{I}\sigma_{\mathrm{e}}^2)$, where σ_e^2 is the residual variance). Standard errors of heritability and repeatability were estimated using the Monte-Carlo method [78] implemented in *AIREMLF90*.

Semen quality phenotypes for association testing

Phenotypes for the association studies were the average values either from the filtered unadjusted data or from the residuals of the following linear model: $\mathbf{y} = \mathbf{1}\mu + \mathbf{h}\mathrm{a} + \mathbf{d}\mathrm{f} + \mathbf{Ct} + \mathbf{Sn} + \mathbf{e}$, where **y** is a semen quality parameter of each ejaculate, μ is the intercept (**1** is a vector of ones), **h** and **d** are the vectors of the bull's age (in days) when the ejaculate was collected and the interval in days to the preceding ejaculate respectively, a and f are their respective effects. **C** and **S** are incidence matrices relating collector (person collecting the ejaculate) and season (4 seasons per year), **t** is the vector of effects of different semen collectors, **n** is the vector of effects of different seasons, and **e** is a vector of random residuals. The correlation between the average values from the filtered data and the average values from the random residuals were high ($r > 0.90$) for all traits analysed. To facilitate a better interpretation of the effect estimates, we used the average values calculated from the filtered data as input variables throughout the manuscript.

Insemination success in the Swiss BSW population

Bull fertility (as at February 2019) was provided by Swissgenetics for 941 BSW bulls that had between 232 and 15,690 first inseminations with conventional (i.e., semen was not sorted for sex) frozen-thawed semen. Male fertility was estimated using a linear mixed model

similar to the one proposed by Schaeffer et al. [79]: $\mathbf{y}_{ijklmnopq} = \mu + \mathbf{MO}_i + \mathbf{PA}_j + \mathbf{PR}_k + (\mathbf{RS} \text{ x } \mathbf{RK})_{lm} + \mathbf{BE}_n + \mathbf{h}_o + \mathbf{s}_{lp} + \mathbf{e}_{ijklmnopq}$, where $\mathbf{y}_{ijklmnopq}$ is either 0 (subsequent insemination recorded within 56 days of the insemination) or 1 (no subsequent insemination recorded within 56 days after the insemination), μ is the intercept, $\mathbf{MO}_i$ is the insemination month, $\mathbf{PA}_j$ is the parity (heifer or cow), $\mathbf{PR}_k$ is the cost of the insemination straw, $(\mathbf{RS} \text{ x } \mathbf{RK})_{lm}$ is the combination bull's breed x cow's breed, $\mathbf{BE}_n$ is the insemination technician, h_o is the herd, $\mathbf{s}_{lp}$ is the fertility of the bull expressed in % deviation from the average non-return rate, and $\mathbf{e}_{ijklmnopq}$ is a random residual term. Bull fertility was subsequently standardized to a mean of 100 ± 12. Three bulls with very low fertility (i.e., more than 3 standard deviations below average) were not considered for our analyses because they might carry rare genetic conditions [28] and including such phenotypic outliers in genome-wide association testing might lead to spurious associations [45]. For the GWAS on bull fertility, we considered 591 bulls that had records for both male fertility and semen quality.

Male fertility in the German and Austrian BSW populations

Phenotypes for bull fertility (as at December 2017), were provided by ZuchtData EDV-Dienstleistungen GmbH, Austria, for 4617 BSW bulls from Germany and Austria that were used for 4,267,990 and 1,646,254 inseminations in cows and heifers, respectively. Bull fertility in the German and Austrian BSW populations is estimated using a multi-trait animal model that was proposed by Fuerst & Gredler [4]. The model includes a fixed effect for the service sire that represents bull fertility as deviation from the population mean. Four phenotypic outliers were not considered for subsequent analyses (see above). The final dataset included fertility records for 2481 bulls that also had (partially imputed) genotypes at 607,511 SNPs.

Haplotype-based association testing of phenotypes

Haplotype-based association testing was implemented in R using a sliding-window-based approach that we applied previously to investigate genetic conditions in cattle [58]. In brief, a sliding window of 50 contiguous SNPs corresponding to a haplotype length of ~200 kb was shifted along the autosomes in steps of 15 SNPs. Within each sliding window (N = 40,444), pre-phased haplotypes (see above) with frequency greater than 1 % were tested for association using the linear model $\mathbf{y} = \mu + \sum_{j=1}^{10} \mathbf{a}_j \mathbf{PC}_j + \mathbf{bHT} + \mathbf{e}$, where y is a vector of phenotypes (see above), μ is the intercept, $\mathbf{PC}_j$ are the top ten principal components of the genomic relationship matrix (see above), **a** and **b** are effects of the principal components and the haplotype (**HT**) tested, respectively, and **e** is a vector of residuals that are assumed to be normally distributed. Haplotypes were tested for association assuming either additive, dominant or recessive mode of inheritance. Recessive tests were carried out for haplotypes that were observed in the homozygous state in at least 1 % of the individuals. The genomic inflation factor lambda was calculated in R using the following formula: lambda = median(qchisq(1-p, 1)) / qchisq(0.5, 1) where p is a vector of P values.

Whole-genome sequencing and sequence variant genotyping

We used paired-end whole-genome sequencing reads (2x101 bp, 2x126 bp or 2x150 bp) of 42 BSW animals (3 homozygous, 10 heterozygous, 29 non-carrier) that were generated using Illumina HiSeq or NovaSeq instruments to identify candidate causal variants for the impaired reproductive performance of homozygous haplotype carriers. Some sequenced animals were key ancestors of the Swiss BSW population that had been sequenced previously [80].

Sequencing data of all animals are available from the European Nucleotide Archive (http://www.ebi.ac.uk/ena) under the accession numbers listed in Supporting File 4.

We removed from the raw data reads for which the phred-scaled quality was less than 15 for more than 15 % of bases, and trimmed adapter sequences using the *fastp* software [81]. Subsequently, the sequencing data were aligned to the bovine linear reference genome (ARS-UCD1.2) using the *BWA mem* algorithm [82]. Duplicates were marked and read alignments were coordinate sorted using the *Picard* tools software suite [83] and *Sambamba* [84], respectively. The average depth of the aligned sequencing reads was 13.9-fold and it ranged from 6.6 to 28.9-fold.

We discovered and genotyped SNPs and Indels from the linear alignments using the multi-sample variant calling approach implemented in the *Genome Analysis Toolkit* (*GATK*, version 4.1.0) [85]. Specifically, we followed *GATK*'s best practice recommendations for sequence variant discovery and filtration when variant quality score recalibration (VQSR) is not possible because the truth set of variants required for VQSR is not publicly available in cattle. Additional details on the applied variant calling and filtration approach can be found in Crysnanto et al. [86]. The *mosdepth* software (version 0.2.2) [36] was used to extract the number of reads that covered a genomic position.

Structural variants including large insertions, deletions, inversions, duplications and translocations were detected and genotyped in the sequenced BSW animals using the *delly* software with default parameter settings [87].

Identification of candidate causal variants

The status of the BTA6 top haplotype was determined for the 42 sequenced BSW animals using (partially) imputed SNP microarray-derived genotypes at 607,511 SNPs (see above). Sequence variant genotypes within a 10 Mb region on bovine chromosome 6 (from 55 to 65

Mb) encompassing the 2.38 Mb segment of extended homozygosity were filtered to identify variants compatible with recessive inheritance. In order to take into account possible haplotype phasing errors, inaccurately genotyped sequence variants and the undercalling of heterozygous genotypes due to low sequencing coverage, we applied a conservative filtering strategy. Specifically, we screened for variants that had following allele frequencies:

- ≥ 0.8 in three homozygous haplotype carriers (5 out of 6 alleles),
- between 0.4 and 0.6 in ten heterozygous haplotype carriers,
- ≤ 0.05 in 29 non-carriers of the haplotype.

This filtration identified 824 variants that were compatible with recessive inheritance. We annotated the candidate causal variants according to the Ensembl (release 95) annotation of the ARS-UCD1.2-assembly of the bovine genome using the *Variant Effect Predictor* (VEP) software tool [88]. The frequency and genotype distribution of candidate causal variants were also investigated in cattle from breeds other than BSW using polymorphism information from an in-house sequence variant database of 295 cattle (accession numbers and corresponding breeds are listed in Supporting File 6) and the most recent release (run 7) of the 1000 Bull Genomes Project (3078 cattle).

Quantification of transcript abundance in testis tissue

We downloaded between 47 and 58 million 2x150 bp paired-end sequencing reads from the ENA sequence read archive that were generated using RNA extracted from testicular tissue samples of three mature bulls and three newborn male calves of the Angus beef cattle breed (ENA accession numbers: SAMN09205186-SAMN09205191; [89]). The RNA sequencing data were pseudo-aligned to an index of the bovine transcriptome (ftp://ftp.ensembl.org/pub/release-98/fasta/bos_taurus/cdna/Bos_taurus.ARS-UCD1.2.cdna.all.fa.gz) and transcript abundance was quantified using the *kallisto* software

[90]. We used the R package tximport [91] to aggregate transcript abundances to the gene level.

Testicular tissue sampling

Testicular tissue was sampled at a commercial slaughterhouse from four AI bulls (two bulls homozygous for the mutant (mt) T-allele at Chr6:58373887, one heterozygous bull and one bull homozygous for the wild-type (wt) C-allele Chr6:58373887). At the time of slaughter, the bulls were between 595 and 778 days old and they had been kept under identical conditions at the AI center Swissgenetics. The haplotype status of the bulls was determined before slaughter using SNP microarray-derived genotypes. Tissue samples were frozen on dry ice immediately after collection and subsequently stored at –80 °C. Additionally, testis and epididymis tissues were formalin-fixed and paraffin-embedded. Sections of embedded tissue were stained with hematoxylin and eosin for microscopic examination.

Transcription analysis using RT-PCR

Total RNA from testicular tissue samples of the four bulls was extracted using Direct-zol RNA Mini Prep Kit (Zymo Research) according to the manufacturer's instructions. The integrity and concentration of RNA was analysed by agarose gel electrophoresis and using the Nanodrop ND-2000 spectrophotometer (Thermo Scientific), respectively. Total RNA was reverse transcribed using FastGene ScriptaseII (Fast Gene). RT-PCR was done with the GoTaq Polymerase (Promega) using forward primer (5'-ACGTGGAGCCCAACTTTGTA-3') and reverse primer (5'-AGTGCAGACGCATAGTCAGAA-3'). The RT-PCR products were separated on a 3 % agarose gel and the length of the products was analysed using Quantum (Vilber Lourmat). The sequence of the RT-PCR products was obtained using Sanger sequencing.

Quantitative RT-PCR

200 ng total RNA was used to synthesise complementary DNA (cDNA) using FastGene ScriptaseII (Fast Gene). Two-step qPCR experiments were performed using Fast SybrGreen MasterMix (Applied Biosystems) using forward primer (5'-TGATTATCGACATCCCGTCA-3') and reverse primer (5'-GTCTGGAATCTCATAGGTAG-3') and run on an ABI 7500 thermocycler (Applied Biosystems). Primer specificity and capture temperature were determined by melt curve analysis. The relative expression difference between the genotypes in all tissues was calculated for each sample (ΔΔCT). All cDNA samples were assayed in triplicate and relative expression levels normalised to the *GAPDH* reference gene.

Whole transcriptome sequencing and read alignment

A paired-end RNA library (2x150 bp) was prepared from total RNA from a testicular tissue sample of one BSW bull homozygous for the mutant T-allele at Chr6:58373887 using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA, USA). The library was sequenced using an Illumina NovaSeq6000 instrument. Quality control on 63,473,978 raw RNA sequencing reads was performed using the *fastp* software [81]. We removed adapter sequences and reads for which the phred-scaled quality was less than 15 for more than 15 % of bases. The filtered reads (N=63,138,436) were aligned to the ARS-UCD1.2 reference sequence and the Ensembl gene annotation (release 99) using the splice-aware read alignment tool STAR (version 2.7.3a) [92]. The *mosdepth* software (version 0.2.2) [36] was used to extract the number of RNA sequencing reads that covered a genomic position. The

RNA sequencing data have been deposited at the Sequence Read Archive of the NCBI under sample and study accession numbers SAMN14485268 and PRJNA616249, respectively.

Western blot

Total protein from testicular tissue was extracted using T-PER® Tissue Protein Extraction Reagent (Thermo Scientific) according to the manufacturer's instructions. Western blot was performed using iBind Western Blot System (Life Technologies). The bovine WDR19 protein was detected using rabbit abx316410 anti-human-WDR19 antibody (diluted 1:2000 in iBind solution) and horseradish peroxidase-labelled goat anti-rabbit sc-2005 (diluted 1:5000 in iBind solution). GADPH was detected using mouse monoclonal anti-GADPH #G8795, (diluted 1:3000) and rabbit anti-mouse IgG H&L (HRP) ab6728 (diluted 1:5000).

Bioinformatic analyses of the mutant allele and WDR19 topology prediction

The *NNSPLICE* software tool (https://www.fruitfly.org/seq_tools/splice.html, [93]) was used to predict putative splice sites within 1000 bp on either side of the BTA6:58373887C>T-variant. Multi-species alignment of the WDR19 protein sequence was performed using *Clustal Omega* [94,95]. The topology of bovine WDR19 isoforms ENSBTAT00000019294.6 (1342 amino acids) and ENSBTAT00000069073.1 (1242 amino acids) was predicted using the *WD40-repeat protein Structure Predictor* version 2.0 (http://www.wdspdb.com/wdsp/ [96]).

List of abbreviations

AI - artificial insemination; BSW – Brown Swiss; NRR - non-return rate; QTL – quantitative trait locus; SNP – single nucleotide polymorphism; IFT-A interflagellar complex A

Acknowledgements

We thank Braunvieh Schweiz for providing pedigree and genotype data of BSW bulls. We thank Swissgenetics for providing ejaculate data of BSW bulls. We thank Susanne Meese and Sarah Wyck for explaining semen collection and processing at Swissgenetics, and Sarah Wyck and Meenu Bhati for support in tissue sampling. We acknowledge the Arbeitsgemeinschaft Deutsches Braunvieh, Braunvieh Austria, Tierzuchtforschung Grub, Chair of Animal Breeding of TU München, the Institute of Animal Breeding from Bayerische Landesanstalt fuer Landwirtschaft and ZuchtData EDV Dienstleistungen GmbH for providing pedigree records, genotype and phenotype data for the Austrian and German BSW bulls. We thank the Functional Genomics Center Zurich for generating DNA and RNA sequencing data.

References

1. Esslemont RJ, Kossaibati MA, Allcock J. Economics of fertility in dairy cows. BSAP Occasional Publication. 2001;26: 19–29. doi:10.1017/S0263967X00033565

2. Inchaisri C, Jorritsma R, Vos PLAM, van der Weijden GC, Hogeveen H. Economic consequences of reproductive performance in dairy cattle. Theriogenology. 2010;74: 835–846. doi:10.1016/j.theriogenology.2010.04.008

3. De Vries A. Economic Value of Pregnancy in Dairy Cattle. Journal of Dairy Science. 2006;89: 3876–3885. doi:10.3168/jds.S0022-0302(06)72430-4

4. Fuerst C, Gredler B. Genetic evaluation for fertility traits in Austria and Germany. Interbull Bulletin. 2009; 3.

5. Norman HD, Hutchison JL, VanRaden PM. Evaluations for service-sire conception rate for heifer and cow inseminations with conventional and sexed semen. Journal of Dairy Science. 2011;94: 6135–6142. doi:10.3168/jds.2010-3875

6. Nadarajah K, Burnside EB, Schaeffer LR. Genetic Parameters for Fertility of Dairy Bulls. Journal of Dairy Science. 1988;71: 2730–2734. doi:10.3168/jds.S0022-0302(88)79866-5

7. Tiezzi F, Maltecca C, Penasa M, Cecchinato A, Bittante G. Short communication: Genetic analysis of dairy bull fertility from field data of Brown Swiss cattle. Journal of Dairy Science. 2013;96: 7325–7328. doi:10.3168/jds.2013-6885

8. Rezende FM, Nani JP, Peñagaricano F. Genomic prediction of bull fertility in US Jersey dairy cattle. Journal of Dairy Science. 2019;102: 3230–3240. doi:10.3168/jds.2018-15810

9. Mathevon M, Buhr MM, Dekkers JCM. Environmental, Management, and Genetic Factors Affecting Semen Production in Holstein Bulls. Journal of Dairy Science. 1998;81: 3321–3330. doi:10.3168/jds.S0022-0302(98)75898-9

10. Druet T, Fritz S, Sellem E, Basso B, Gérard O, Salas-Cortes L, et al. Estimation of genetic parameters and genome scan for 15 semen characteristics traits of Holstein bulls. Journal of Animal Breeding and Genetics. 2009;126: 269–277. doi:10.1111/j.1439-0388.2008.00788.x

11. Berry DP, Eivers B, Dunne G, McParland S. Genetics of bull semen characteristics in a multi-breed cattle population. Theriogenology. 2019;123: 202–208. doi:10.1016/j.theriogenology.2018.10.006

12. Burren A, Joerg H, Erbe M, Gilmour AR, Witschi U, Schmitz-Hsu F. Genetic parameters for semen production traits in Swiss dairy bulls. Reproduction in Domestic Animals. 2019;54: 1177–1181. doi:10.1111/rda.13492

13. Rodríguez-Martínez H. State of the art in farm animal sperm evaluation. Reprod Fertil Dev. 2006;19: 91–101. doi:10.1071/RD06104

14. Vincent P, Underwood SL, Dolbec C, Bouchard N, Kroetsch T, Blondin P. Bovine Semen Quality Control in Artificial Insemination Centers. In: Hopper RM, editor. Bovine Reproduction. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2014. pp. 685–695. doi:10.1002/9781118833971.ch74

15. Thundathil JC, Dance AL, Kastelic JP. Fertility management of bulls to improve beef cattle productivity. Theriogenology. 2016;86: 397–405. doi:10.1016/j.theriogenology.2016.04.054

16. García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proc Natl Acad Sci USA. 2016;113: E3995–E4004. doi:10.1073/pnas.1519061113

17. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nature Genetics. 2014;46: 858–865. doi:10.1038/ng.3034

18. Pausch H, MacLeod IM, Fries R, Emmerling R, Bowman PJ, Daetwyler HD, et al. Evaluation of the accuracy of imputed sequence variant genotypes and their utility for causal variant detection in cattle. Genetics Selection Evolution. 2017;49: 24. doi:10.1186/s12711-017-0301-x

19. Fortes MRS, DeAtley KL, Lehnert SA, Burns BM, Reverter A, Hawken RJ, et al. Genomic regions associated with fertility traits in male and female cattle: Advances from microsatellites to high-density chips and beyond. Animal Reproduction Science. 2013;141: 1–19. doi:10.1016/j.anireprosci.2013.07.002

20. Puglisi R, Gaspa G, Balduzzi D, Severgnini A, Vanni R, Macciotta NPP, et al. Genomewide analysis of bull sperm quality and fertility traits. Reproduction in Domestic Animals. 2016;51: 840–843. doi:10.1111/rda.12747

21. Qin C, Yin H, Zhang X, Sun D, Zhang Q, Liu J, et al. Genome-wide association study for semen traits of the bulls in Chinese Holstein. Animal Genetics. 2017;48: 80–84. doi:10.1111/age.12433

22. Fortes MRS, Reverter A, Kelly M, McCulloch R, Lehnert SA. Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. Andrology. 2013;1: 644–650. doi:10.1111/j.2047-2927.2013.00101.x

23. Hering DM, Olenski K, Kaminski S. Genome-wide association study for poor sperm motility in Holstein-Friesian bulls. Animal Reproduction Science. 2014;146: 89–97. doi:10.1016/j.anireprosci.2014.01.012

24. Han Y, Peñagaricano F. Unravelling the genomic architecture of bull fertility in Holstein cattle. BMC Genetics. 2016;17: 143. doi:10.1186/s12863-016-0454-6

25. Ferenčaković M, Sölkner J, Kapš M, Curik I. Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. Journal of Dairy Science. 2017;100: 4721–4730. doi:10.3168/jds.2016-12164

26. Lan XY, Peñagaricano F, DeJung L, Weigel KA, Khatib H. Short communication: A missense mutation in the PROP1 (prophet of Pit 1) gene affects male fertility and milk production traits in the US Holstein population. Journal of Dairy Science. 2013;96: 1255–1257. doi:10.3168/jds.2012-6019

27. Pausch H, Wurmser C, Reinhardt F, Emmerling R, Fries R. Short communication: Validation of 4 candidate causative trait variants in 2 cattle breeds using targeted sequence imputation. Journal of Dairy Science. 2015;98: 4162–4167. doi:10.3168/jds.2015-9402

28. Pausch H, Kölle S, Wurmser C, Schwarzenbacher H, Emmerling R, Jansen S, et al. A Nonsense Mutation in TMEM95 Encoding a Nondescript Transmembrane Protein Causes Idiopathic Male Subfertility in Cattle. PLoS Genet. 2014;10. doi:10.1371/journal.pgen.1004044

29. Pausch H, Venhoranta H, Wurmser C, Hakala K, Iso-Touru T, Sironen A, et al. A frameshift mutation in ARMC3 is associated with a tail stump sperm defect in Swedish Red (Bos taurus) cattle. BMC Genetics. 2016;17: 49. doi:10.1186/s12863-016-0356-7

30. Iso-Touru T, Wurmser C, Venhoranta H, Hiltpold M, Savolainen T, Sironen A, et al. A splice donor variant in CCDC189 is associated with asthenospermia in Nordic Red dairy cattle. BMC Genomics. 2019;20: 286. doi:10.1186/s12864-019-5628-y

31. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 2009;10: 381–391. doi:10.1038/nrg2575

32. Weitze KF. 7 Andrologie beim Bullen. In: Walter B, Holzmann A, editors. Veterinärmedizinische Andrologie: Physiologie und Pathologie der Fortpflanzung bei männlichen Tieren. Stuttgart, Deutschland: Schattauer Verlag; 2001. pp. 119–214.

33. Harris T, Marquez B, Suarez S, Schimenti J. Sperm motility defects and infertility in male mice with a mutation in Nsun7, a member of the Sun domain-containing family of putative RNA methyltransferases. Biol Reprod. 2007;77: 376–382. doi:10.1095/biolreprod.106.058669

34. Khosronezhad N, Hosseinzadeh Colagar A, Mortazavi SM. The Nsun7 (A11337)-deletion mutation, causes reduction of its protein rate and associated with sperm motility defect in infertile men. J Assist Reprod Genet. 2015;32: 807–815. doi:10.1007/s10815-015-0443-0

35. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. Nat Biotechnol. 2011;29: 24–26. doi:10.1038/nbt.1754

36. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018;34: 867–868. doi:10.1093/bioinformatics/btx699

37. Pausch H, Schwarzenbacher H, Burgstaller J, Flisikowski K, Wurmser C, Jansen S, et al. Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. BMC Genomics. 2015;16: 312. doi:10.1186/s12864-015-1483-7

38. Zhang Q, Calus MPL, Bosse M, Sahana G, Lund MS, Guldbrandtsen B. Human-Mediated Introgression of Haplotypes in a Modern Dairy Cattle Breed. Genetics. 2018;209: 1305–1317. doi:10.1534/genetics.118.301143

39. Dam TJP van, Townsend MJ, Turk M, Schlessinger A, Sali A, Field MC, et al. Evolution of modular intraflagellar transport from a coatomer-like progenitor. PNAS. 2013;110: 6943–6948. doi:10.1073/pnas.1221011110

40. Efimenko E, Blacque OE, Ou G, Haycraft CJ, Yoder BK, Scholey JM, et al. Caenorhabditis elegans DYF-2, an Orthologue of Human WDR19, Is a Component of the Intraflagellar Transport Machinery in Sensory Cilia. MBoC. 2006;17: 4801–4811. doi:10.1091/mbc.e06-04-0260

41. Wang Y, Hu X-J, Zou X-D, Wu X-H, Ye Z-Q, Wu Y-D. WDSPdb: a database for WD40-repeat proteins. Nucleic Acids Res. 2015;43: D339–D344. doi:10.1093/nar/gku1023

42. Taylor JF, Schnabel RD, Sutovsky P. Review: Genomics of bull fertility. animal. 2018;12: s172–s183. doi:10.1017/S1751731118000599

43. Barth AD, Waldner CL. Factors affecting breeding soundness classification of beef bulls examined at the Western College of Veterinary Medicine. Can Vet J. 2002;43: 274–284.

44. Ax RL, Gilbert GR, Shook GE. Sperm in Poor Quality Semen from Bulls During Heat Stress Have a Lower Affinity for Binding Hydrogen-3 Heparin1. Journal of Dairy Science. 1987;70: 195–200. doi:10.3168/jds.S0022-0302(87)79994-9

45. Auer PL, Reiner AP, Leal SM. The effect of phenotypic outliers and non-normality on rare-variant association testing. Eur J Hum Genet. 2016;24: 1188–1194. doi:10.1038/ejhg.2015.270

46. Vitezica ZG, Reverter A, Herring W, Legarra A. Dominance and epistatic genetic variances for litter size in pigs using genomic models. Genetics Selection Evolution. 2018;50: 71. doi:10.1186/s12711-018-0437-3

47. Gruhot T, Gray K, Brown V, Huang Y, Kachman SD, Spangler ML, et al. Genetic relationships among sperm quality traits of Duroc boars collected during the summer season. Animal Reproduction Science. 2019;206: 85–92. doi:10.1016/j.anireprosci.2019.05.012

48. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, et al. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. PLoS Genet. 2014;10. doi:10.1371/journal.pgen.1004049

49. Fang Z-H, Pausch H. Multi-trait meta-analyses reveal 25 quantitative trait loci for economically important traits in Brown Swiss cattle. BMC Genomics. 2019;20: 695. doi:10.1186/s12864-019-6066-6

50. Chun S, Fay JC. Evidence for Hitchhiking of Deleterious Mutations within the Human Genome. PLoS Genet. 2011;7. doi:10.1371/journal.pgen.1002240

51. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50: 362–367. doi:10.1038/s41588-018-0056-5

52. Pausch H, Emmerling R, Gredler-Grandl B, Fries R, Daetwyler HD, Goddard ME. Meta-analysis of sequence-based association studies across three cattle breeds reveals 25 QTL for fat and protein percentages in milk at nucleotide resolution. BMC Genomics. 2017;18. doi:10.1186/s12864-017-4263-8

53. Signer-Hasler H, Burren A, Neuditschko M, Frischknecht M, Garrick D, Stricker C, et al. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. Genet Sel Evol. 2017;49. doi:10.1186/s12711-017-0358-6

54. Sartelet A, Druet T, Michaux C, Fasquelle C, Géron S, Tamma N, et al. A Splice Site Variant in the Bovine RNF11 Gene Compromises Growth and Regulation of the Inflammatory Response. PLoS Genet. 2012;8. doi:10.1371/journal.pgen.1002581

55. Drögemüller M, Jagannathan V, Becker D, Drögemüller C, Schelling C, Plassais J, et al. A Mutation in the FAM83G Gene in Dogs with Hereditary Footpad Hyperkeratosis (HFH). PLoS Genet. 2014;10. doi:10.1371/journal.pgen.1004370

56. McClure MC, Bickhart D, Null D, VanRaden P, Xu L, Wiggans G, et al. Bovine Exome Sequence Analysis and Targeted SNP Genotyping of Recessive Fertility Defects BH1, HH2, and HH3 Reveal a Putative Causative Mutation in SMC2 for HH3. PLOS ONE. 2014;9: e92769. doi:10.1371/journal.pone.0092769

57. Schwarzenbacher H, Burgstaller J, Seefried FR, Wurmser C, Hilbe M, Jung S, et al. A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. BMC Genomics. 2016;17: 400. doi:10.1186/s12864-016-2742-y

58. Pausch H, Ammermüller S, Wurmser C, Hamann H, Tetens J, Drögemüller C, et al. A nonsense mutation in the COL7A1 gene causes epidermolysis bullosa in Vorderwald cattle. BMC Genetics. 2016;17: 149. doi:10.1186/s12863-016-0458-2

59. Bhati M, Kadri NK, Crysnanto D, Pausch H. Assessing genomic diversity and signatures of selection in Original Braunvieh cattle using whole-genome sequencing data. BMC Genomics. 2020;21: 27. doi:10.1186/s12864-020-6446-y

60. Reinartz S, Distl O. Short communication: Lethal mutations in Vorderwald cattle through Montbéliarde incrossings. Journal of Dairy Science. 2020;103: 613–618. doi:10.3168/jds.2019-17213

61. Blencowe BJ. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends in Biochemical Sciences. 2000;25: 106–110. doi:10.1016/S0968-0004(00)01549-8

62. Wei Q, Zhang Y, Li Y, Zhang Q, Ling K, Hu J. The BBSome controls IFT assembly and turnaround in cilia. Nat Cell Biol. 2012;14: 950–957. doi:10.1038/ncb2560

63. Iomini C, Li L, Esparza JM, Dutcher SK. Retrograde Intraflagellar Transport Mutants Identify Complex A Proteins With Multiple Genetic Interactions in Chlamydomonas reinhardtii. Genetics. 2009;183: 885–896. doi:10.1534/genetics.109.101915

64. Coussa RG, Otto EA, Gee H-Y, Arthurs P, Ren H, Lopez I, et al. WDR19: An ancient, retrograde, intraflagellar ciliary protein is mutated in autosomal recessive retinitis pigmentosa and in Senior-Loken syndrome. Clinical Genetics. 2013;84: 150–159. doi:10.1111/cge.12196

65. Lee JM, Ahn YH, Kang HG, Ha IS, Lee K, Moon KC, et al. Nephronophthisis 13: implications of its association with Caroli disease and altered intracellular localization of WDR19 in the kidney. Pediatr Nephrol. 2015;30: 1451–1458. doi:10.1007/s00467-015-3068-8

66. Bredrup C, Saunier S, Oud MM, Fiskerstrand T, Hoischen A, Brackman D, et al. Ciliopathies with Skeletal Anomalies and Renal Insufficiency due to Mutations in the IFT-A Gene WDR19. The American Journal of Human Genetics. 2011;89: 634–643. doi:10.1016/j.ajhg.2011.10.001

67. Fehrenbach H, Decker C, Eisenberger T, Frank V, Hampel T, Walden U, et al. Mutations in WDR19 encoding the intraflagellar transport component IFT144 cause a broad spectrum of ciliopathies. Pediatr Nephrol. 2014;29: 1451–1456. doi:10.1007/s00467-014-2762-2

68. Arts H, Knoers N. Cranioectodermal Dysplasia. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993. Available: http://www.ncbi.nlm.nih.gov/books/NBK154653/

69. Xu C, Min J. Structure and function of WD40 domain proteins. Protein Cell. 2011;2: 202–214. doi:10.1007/s13238-011-1018-1

70. Zhang Y, Liu H, Li W, Zhang Z, Zhang S, Teves ME, et al. Intraflagellar transporter protein 140 (IFT140), a component of IFT-A complex, is essential for male fertility and spermiogenesis in mice. Cytoskeleton. 2018;75: 70–84. doi:10.1002/cm.21427

71. Wang X, Sha Y, Wang W, Cui Y, Chen J, Yan W, et al. Novel IFT140 variants cause spermatogenic dysfunction in humans. Molecular Genetics & Genomic Medicine. 2019;7: e920. doi:10.1002/mgg3.920

72. 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. doi:10.1093/gigascience/giaa021.

73. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4: 7. doi:10.1186/s13742-015-0047-8

74. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. The American Journal of Human Genetics. 2018;103: 338–348. doi:10.1016/j.ajhg.2018.07.015

75. Loh P-R, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48: 1443–1448. doi:10.1038/ng.3679

76. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012;44: 955–959. doi:10.1038/ng.2354

77. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). Proceedings of the 7th world congress on genetics applied to livestock production. 2002. pp. 743–744.

78. Houle D, Meyer K. Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood. J Evol Biol. 2015;28: 1542–1549. doi:10.1111/jeb.12674

79. Schaeffer LR. Evaluation of Bulls for Nonreturn Rates Within Artificial Insemination Organizations. Journal of Dairy Science. 1993;76: 837–842. doi:10.3168/jds.S0022-0302(93)77409-3

80. Baes CF, Dolezal MA, Koltes JE, Bapst B, Fritz-Waters E, Jansen S, et al. Evaluation of variant identification methods for whole genome sequencing data in dairy cattle. BMC Genomics. 2014;15: 948. doi:10.1186/1471-2164-15-948

81. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34: i884–i890. doi:10.1093/bioinformatics/bty560

82. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:13033997 [q-bio]. 2013 [cited 4 Nov 2019]. Available: http://arxiv.org/abs/1303.3997

83. Picard Tools - By Broad Institute. [cited 4 Nov 2019]. Available: https://broadinstitute.github.io/picard/

84. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015;31: 2032–2034. doi:10.1093/bioinformatics/btv098

85. DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–498. doi:10.1038/ng.806

86. Crysnanto D, Wurmser C, Pausch H. Accurate sequence variant genotyping in cattle using variation-aware genome graphs. Genetics Selection Evolution. 2019;51: 21. doi:10.1186/s12711-019-0462-x

87. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28: i333–i339. doi:10.1093/bioinformatics/bts378

88. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biology. 2016;17: 122. doi:10.1186/s13059-016-0974-4

89. Gao Y, Li S, Lai Z, Zhou Z, Wu F, Huang Y, et al. Analysis of Long Non-Coding RNA and mRNA Expression Profiling in Immature and Mature Bovine (Bos taurus) Testes. Front Genet. 2019;10. doi:10.3389/fgene.2019.00646

90. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34: 525–527. doi:10.1038/nbt.3519

91. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2016;4. doi:10.12688/f1000research.7563.2

92. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21. doi:10.1093/bioinformatics/bts635

93. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved Splice Site Detection in Genie. Journal of Computational Biology. 1997;4: 311–323. doi:10.1089/cmb.1997.4.311

94. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7: 539. doi:10.1038/msb.2011.75

95. Madeira F, Park Y mi, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47: W636–W641. doi:10.1093/nar/gkz268

96. Ma J, An K, Zhou J-B, Wu N-S, Wang Y, Ye Z-Q, et al. WDSPdb: an updated resource for WD40 proteins. Bioinformatics. 2019;35: 4824–4826. doi:10.1093/bioinformatics/btz460

Supporting Files

Supporting File 1: Phenotypic correlations (off-diagonal) and heritability (diagonal) of the traits studied. Correlation and heritability of seven traits relevant for semen quality and fertility.

Supporting File 2: Detection of QTL for semen quality and fertility in BSW bulls. Manhattan plots representing the association ($-\log_{10}(P)$) of haplotypes with (A) ejaculate volume (genomic inflation factor lambda = 1.67), (B) sperm concentration (lambda = 1.29), (C) sperm motility (lambda = 1.35), (D) proportion of sperm with head anomalies (lambda = 1.10), (E) proportion of sperm with tail anomalies (lambda = 1.14), (F) sperm per straw (lambda = 1.22), and (G) bull fertility (lambda = 1.22) assuming an additive mode of inheritance. Red color indicates significantly associated haplotypes (P < Bonferroni corrected significance threshold).

Supporting File 3: Expression of genes located within the segment of extended homozygosity in testis tissue. Transcripts per million (TPM) in testis tissue of three mature bulls (grey) and three newborn male calves (black). The horizontal line represents the median expression (5.9 TPM) of 22,372 genes. To improve readability, the expression for *RPL* is only shown in the inset.

Supporting File 4: Accession numbers of 42 BSW animals. The numbers listed indicate accession numbers from the sequence read archive of the European Nucleotide Archive (http://www.ebi.ac.uk/ena). The second and third column indicates the status of the sequenced animals with regard to the top haplotype and the sequencing coverage, respectively.

Supporting File 5: Variants compatible with recessive inheritance. Functional consequences of 824 variants were predicted using the VEP software tool. The frequency of the alternate allele is presented for homozygous, heterozygous and non-carrier animals of the BSW cattle breed (n = 42), as well as for animals from breeds other than BSW (n = 253) that are part of our in-house variant database.

Supporting File 6: Accession numbers of 295 animals from different cattle breeds. The numbers listed indicate accession numbers from the sequence read archive of the European Nucleotide Archive (http://www.ebi.ac.uk/ena). The second column indicates the breed of the sequenced animals (BSW – Brown Swiss; FV – Fleckvieh; HOL – Holstein; NRC – Nordic Red Dairy Cattle; OBV – Original Braunvieh; TGV – Tyrolean Grey).

Supporting File 7: Activation of cryptic splicing through the BTA6:58373887 T-allele. Screen captures of IGV outputs from testis RNAseq alignments of a BSW bull (SAMN14485268) homozygous for the mutant (mt) T-allele (A) and two control bulls from the Angus breed (SAMN09205187, SAMN09205188) that are homozygous for the wild-type (wt) C-allele at Chr6:58373887 (B, C). The red bar indicates nine nucleotides that are truncated from exon 12 of *WDR19* in the BSW bull (A) due to cryptic splicing activated by the T-allele. A low number of sequence reads corresponding to the wild-type transcript were also detected in the bull homozygous for the BTA6:58373887 T-allele.

Supporting File 8: Expression of *NSUN7*. Expression of *NSUN7* quantified using testis RNAseq alignments of a BSW bull (SAMN14485268) homozygous for the mutant (mt) T-allele (A) and two control bulls from the Angus breed (SAMN09205187, SAMN09205188) that are homozygous for the wild-type (wt) C-allele at Chr6:58373887 (B, C). The number of

reads covering a genomic position were extracted from coordinate sorted BAM files using the *mosdepth* software and subsequently divided by the total number of reads (in million) mapped to transcripts.

Supporting File 9: Data of 794 bulls used to detect the QTL on BTA6. The archive contains phased genotypes (haplotypes_6) at 28,872 SNPs located on chromosome 6 (markers_6), the top 20 principal components (evecs) and the average sperm motility of 794 BSW bulls.

Supporting File 10: R script used to perform the haplotype-based association testing. This archive contains the R script that was used to carry out the haplotype-based association testing, a README file that provides information how the script can be applied to analyse the raw data provided in Supporting File 7 and a Jupyter notebook file (GWAS.html) that explains how to process the output file.