eQTL mapping in Brown Swiss bulls to identify variants associated with male fertility

Author(s):
Mapel, Xena Marie; Hiltpold, Maya; Lloret i Villas, Audald; Pausch, Hubert

Publication date:
2022

Permanent link:
https://doi.org/10.3929/ethz-b-000633550

Originally published in:
(2273), https://doi.org/10.3920/978-90-8686-940-4_548

548. eQTL mapping in Brown Swiss bulls to identify variants associated with male fertility

X.M. Mapel*, M. Hiltpold, A. Lloret-Villas and H. Pausch

Animal Genomics, ETH Zürich, Universitätstrasse 2, 8092 Zürich, Switzerland; xenamarie.mapel@usys.ethz.ch

Abstract

Fertility is an essential component of the livestock industry. In cattle, numerous QTL for male reproductive success fall within regulatory regions. However, the effects of these loci have not been investigated in detail or on a large scale. Here, we assemble a sizeable cohort of mature bulls to detect expression quantitative trait loci (eQTL) and assess their effects on fertility-related genes. To do this, we sequenced genomes and total RNA from the testes of 72 bulls. We recovered 13,185,795 DNA sequence variants with minor allele frequency >5%, an average of 283,587,831 clean RNA reads per sample and 18,528 testis-expressed genes (TPM>0.2 in 75% of samples). In total, 2,178 genes had significant *cis*-eQTL at a false discovery rate of 5% (11.76% of expressed genes). Several genes associated with fertility, including *SPATA4* and *SPATA22* (which are responsible for reproductive processes), had significant *cis*-eQTL and variation in transcript abundance across genotypes.

Introduction

Efficient breeding operations require animals with consistent and reliable reproductive success. Artificial insemination centres often collect hundreds of ejaculates from a single bull which are then sold to a multitude of producers. Thus, bulls with impaired or complete loss of reproductive ability can be costly to both parties. The genetic basis of such subfertility is complex. Indeed, quantitative trait loci (QTL) associated with measures of male reproductive success have been identified in several breeds at varying frequencies (Hiltpold *et al.* 2020; Sweett *et al.* 2020). Further insight into these QTL and detection of additional expression QTL (eQTL) that impact the abundance of fertility-related genes can help producers select breeding bulls with high reproductive ability.

Complex traits – such as those associated with semen quality and male reproductive success – are influenced by a multitude of polymorphic sites, with many of those residing in regulatory regions (Xiang *et al.* 2019). For example, a QTL associated with sire non-return rate has been identified in a region that possibly regulates the expression of *SPATA16* (Hiltpold *et al.* 2021), a gene involved with spermatogenesis. Previously, testis eQTL were identified by the cGTEx consortium (https://cgtex.roslin.ed.ac.uk/). However, this dataset relied on samples that were collected at different points of time from various bovine sub-species and breeds and did not investigate the specific effects of testis eQTL. A uniform cohort with an abundance of sequence data would identify additional eQTL and enable the exploration of their effect on reproductive ability.

Here, we harness a cohort of Braunvieh (BV) bulls to identify *cis*-acting eQTL in male reproductive tissue and assess their effect on genes vital for reproduction. To do this, we relied on whole-genome sequence data and total RNA from the testis tissue of 72 bulls.

Materials & methods

Sampling, extraction, sequencing. We sampled testis tissue from 72 bulls (between 8 and 34 months of age) with various degrees of BV ancestry (i.e. Brown Swiss, Original BV, or a cross between one of these and a different breed). After regular slaughter, tissue was flash frozen with liquid nitrogen and kept at -80 °C until DNA and RNA extraction. Frozen testis tissue was homogenized with a Roche MagNA Lyser in 2M DTT

R.F. Veerkamp and Y. de Haas (eds)

DOI: 10.3920/978-90-8686-940-4_548, © X.M. Mapel *et al.* 2022

and RTL Buffer (QIAGEN). DNA and total RNA were extracted simultaneously with the Qiagen AllPrep Mini Kit (QIAGEN) following manufacture protocols. DNA concentration and quality was assessed with a Qubit 2.0 fluorometer (ThermoFisher) and RNA integrity was quantified with a RNA integrity number (RIN) inferred from a Bioanalyzer RNA 600 Nano assay (Agilent Tech.). Genomic libraries (paired-end, 150 bp) were sequenced on the Illumina NovaSeq6000. Total RNA libraries (paired-end, 150 bp) were prepared with the Illumina Stranded Total RNA kit and Illumina Ribo-Zero Plus, then sequenced with the Illumina NovaSeq6000. Library preparation and sequencing was conducted by the Function Genomic Centre Zurich.

DNA sequence data. We used fastp (v0.19.4; Chen *et al.* 2018) with default parameters to remove adapter sequences and low-quality bases, and trim poly-G tails from raw DNA sequence data. After quality control, between 70,493,763 and 307,416,205 reads per sample were aligned to the ARS-UCD1.2 reference genome (Rosen *et al.* 2020) with the mem-algorithm from BWA (v0.717; Li 2013) and the -M flag. We used Sambamba (v0.2.2; Tarasov *et al.* 2015) to sort the aligned reads by coordinates, then combined the read-group specific BAM files to create sample specific sorted BAM files. The MarkDuplicates module from Picard tools (https://broadinstitute.github.io/picard/) was used to mark duplicated reads in the merged and coordinate sorted BAM files. We used mosdepth (v0.2.2; Pedersen and Quinlan 2018) to infer the sequencing coverage at a given genomic position, which was then used to estimate the average coverage per sample and chromosome. Only high-quality reads (mapping quality >10 or without SAM flag 1796) were considered when calculating the average coverage. We called sequence variants for the 72 samples with 24 additional samples from multiple breeds (Brown Swiss=6, Fleckvieh=1, Holstein=7, Original BV=5 Simmental=5), totalling to 96 samples. SNPs and indels were genotyped with the multi-sample variant discovery approach from the Haplotype Caller, Genomics DBImport and GenotypeGVCFs modules in GATK (McKenna *et al.*, 2010). We filtered variants according to GATK's best-practice guidelines. PLINK (v1.9; Purcell 2007) was used to produce a genomic relationship matrix (GRM) for 72 bulls and was constructed with variant genotypes that had minor allele frequency (MAF) >0.5%.

RNA sequence data. We used fastp to remove and trim low-quality bases, adapter sequences, poly-A tails and poly-G tails from the raw RNA sequence data. Filtered reads were aligned to the Ensembl gene annotation (release 104) for cattle with STAR (v2.7.9a; Dobin *et al.* 2013) using the recommended default parameters. This resulted in between 191,160,837 and 386,773,085 clean reads per sample (mean: 283,587,831±43,284,185). Gene abundance, quantified as transcripts per million (TPM), was calculated with the QTLtools 'quan' function with the --filter-mapping-quality and --filter-failed-qc flags to ignore low-quality reads (v1.3.1; Delaneau *et al.* 2017). Reads were mapped to 27,270 coding and non-coding genes from the most-recent Ensembl gene annotation for cattle.

eQTL analysis. We used QTLtools to identify *cis*-acting eQTL for testis expressed genes. Considering our small sample size, we filtered genomic variants to only include genotypes with MAF >5% in the 72 samples. We filtered the TPM matrix to only contain genes that had expression >0.2 TPM in 75% of samples. Age of the bull, RIN value, mass of the sampled testis, season of sampling and the first ten PCs of the GRM (to account for relatedness) were included as covariates for eQTL mapping. We used the *cis* function to identify eQTL within 1 Mb of the transcription start site (TSS) and implemented the --normal flag to rank-normal transform the gene counts, then performed 1000 permutations to infer beta-approximated *P*-values. To identify independent-acting loci and account for multiple testing, we calculated 5% false discover rate (FDR) and used the conditional mode in QTLtools. Allelic fold change (aFC) was calculated for variants of interest with methods described in Mohammadi *et al.* (2017).

Results & discussion

The 72 genomes had moderate sequence depth, which ranged from 6.66 to 27.94 (mean=12.84±4.23). Variant calling identified 20,401,236 variants with MAF>0.5% and 13,185,795 variants with MAF>5.0%. We detected 18,528 genes expressed with >0.2 TPM in 75% of samples.

Of the 18,528 expressed genes considered, 2,178 had at least one significant eVariant (11.76% of testis expressed genes; Figure 1). Each gene's top eVariant was often within 100 kb of the TSS (1,438 variants; 66.0%) and the average distance from the most-significant variant to the TSS was 149,656±229,236 bp (median: 40,824 bp). The number of genes with significant eQTL varied chromosome-by-chromosome (Figure 1B), with BTA18 having the most (165 genes with at least one eQTL). We detected 466,573 significant variants total (using 5% FDR), which is similar to the number of splice-site associated variants in the same cohort of bulls (over 495,000 sQTL; Kadri *et al.* 2021). The conditional analysis revealed that 86 genes had two independently acting eQTL (4.0% of genes with eQTL, 0.5% of all testis expressed genes). We identified more genes with significant *cis*-eQTL than the cGTEx consortium (cGTEx: 809 expressed genes with *cis*-eQTL in bovine testis), which is likely due to our larger sample size, deeper RNA sequencing coverage and more-uniform dataset.

Several genes that are associated with crucial male reproductive processes had *cis*-eQTL. For example, we detected *cis*-eQTL for two genes from the *SPATA* gene family – a gene-set responsible for spermatogenesis and fertilization (Figure 1a; Table 1). Expression of *SPATA4* was associated with a QTL on BTA27 (lead variant: G>A; position 7,704,939; adjusted $P=1.51\times10^{-9}$) while expression of *SPATA22* was associated with a QTL on BTA19 (lead variant: C>T position 24,111,667; adjusted $P=7.61\times10^{-7}$). Lead variants for both eQTL demonstrated additive effects: the abundance of each gene decreased in animals that carried the alternate allele, with an aFC of -0.24 and -0.36 for *SPATA22* and *SPATA4*, respectively (Figure 2). Abnormalities in *SPATA22* and *SPATA4* have been linked to infertility in multiple species (Ishishita 2013; Sujit 2019). Both genes were present in cGTEx, however they did not identify proximal eQTL. Future analyses will further inform if, and how, these and similar eQTL impact male reproductive ability in cattle.

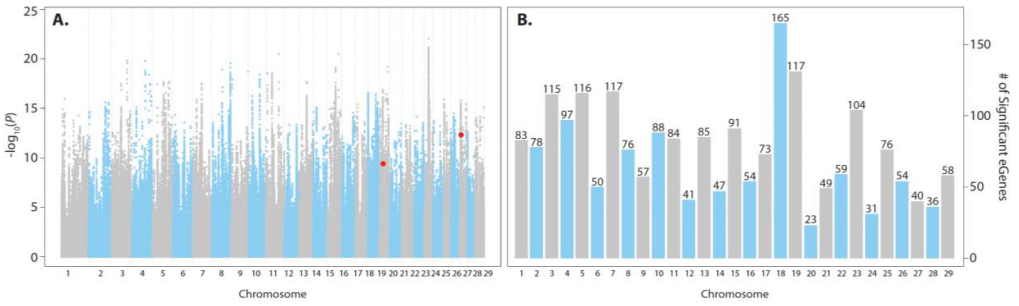

Figure 1. (A) Manhattan plot of genome-wide *cis*-eQTL with lead eVariants for *SPATA22* (chromosome 19) and *SPATA4* (chromosome 27) in red. (B) Number of testis-expressed genes with significant *cis*-eQTL per chromosome.

Table 1. Lead significant eVariants for *SPATA4* and *SPATA22*.

Gene	Position	# Variants in *Cis* window	aFC	Nominal *P*-value	Adjusted β *P*-value
SPATA4	27_7704939	16,450	-0.36	4.74×10^{-13}	1.51×10^{-9}
SPATA22	19_24111667	15,736	-0.24	3.99×10^{-10}	7.61×10^{-7}

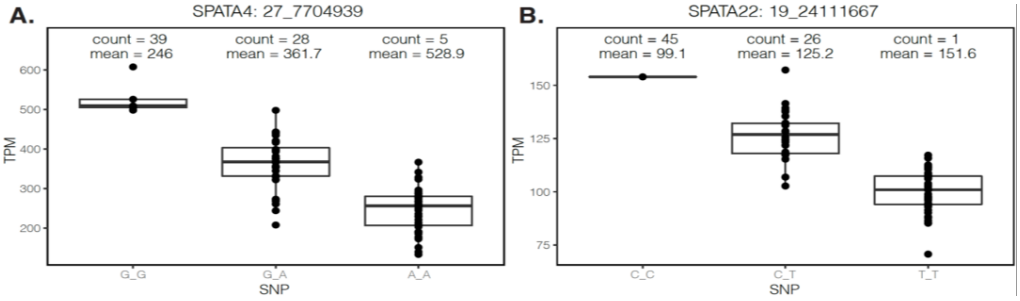

Figure 2. Abundance of *SPATA4* (A) and *SPATA22* (B) for different eVariant genotypes.

Conclusions

Here, we report the first analysis of a testis eQTL study in bulls and identify a large number of eQTL that warrant further investigation. Testing for *trans*-eQTL can uncover additional variants of interest. Furthermore, transcriptome-wide association analyses may help prioritize variants that are associated with variation in male reproduction.

Acknowledgements

Funding for this research was provided by an ETH Research Grant and Swissgenetics.

References

Chen S., Zhou Y., Chen Y., and Gu J. (2018) Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Delaneau O, Ongen H, Brown A, Fort A, Panousis NI, *et al.* (2017) Nat. Commun. 8:15452.10.1038/ncomms15452

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, *et al.* (2013) Bioinformatics. 29:15–21. https://doi.org/10.1093/bioinformatics/bts635

Hiltpold M, Niu G, Kadri NK, Crysnanto D, Fang ZH, *et al.* (2020) PLoS Genet. 16(5):e1008804. https://doi.org/10.1371/journal.pgen.1008804

Hiltpold, M, Kadri NK, Janett F, Witschi U, Schmitz-Hsu F, *et al.* (2021) BMC Genom. 22(1):1–19. https://doi.org/10.1186/s12864-021-07523-3

Ishishita S, Inui T, Matsuda Y, Serikawa T, and Kitada K (2013) Exp. Anim. 62(3):219-227. https://doi.org/10.1538/expanim.62.219

Kadri NK, Mapel XM, and Pausch H (2021) Commun Biol 4, 1206. https://doi.org/10.1038/s42003-021-02725-7

Li H (2013) http://arxiv.org/abs/1303.3997.

McKenna A, Hanna M. Banks E, Sivachenko A, Cibulskis K, *et al.* (2010) Genome Res. 20(9):1297-1303. https://doi.org/10.1101/gr.107524.110

Mohammadi P, Castel SE, Brown AA, and Lappalainen T (2017) Genome res., 27(11):1872. 1884. https://doi.org/10.1101/gr.216747.116

Pedersen BS and Quinlan AR (2018) Bioinformatics 34:867–868. https://doi.org/10.1093/bioinformatics/btx699

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, *et al.* (2007) Am. J. of Hum. Genet.,81. https://doi.org/10.1086/519795

Rosen BD, Bickhart DM, Schnabel RD, Koren S, Elsik, CG *et al.* (2020) Gigascience 9: giaa021. https://doi.org/10.1093/gigascience/giaa021

Sujit KM, Singh V, Trivedi S, Singh K, and Gupta G (2020) Andrology, 8(3):602-609. https://doi.org/10.1111/andr.12742

Sweett H, Fonseca PAS, Suárez-Vega A, Livernois A, Miglior F, *et al.* (2020) Sci Rep 10:20102. https://doi.org/10.1111/andr.12742

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, and Prins P (2015) Bioinformatics 31:20322034. https://doi.org/10.1093/bioinformatics/btv098

Xiang R., Van Den Berg I, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, *et al.* (2019) PNAS. 116(39):19398–19408. https://doi.org/10.1093/bioinformatics/btv098.