Oxford nanopore sequencing as an alternative to reduced representation bisulphite sequencing for the identification of CpGs of interest in livestock populations

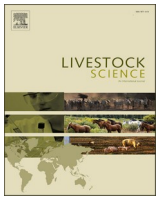
Adrián López-Catalina, Valentin Costes, Ramón Peiró-Pastor, Hélène Kiefer, Oscar González-Recio

Livestock Science 279 (2024) 105377

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



Oxford nanopore sequencing as an alternative to reduced representation bisulphite sequencing for the identification of CpGs of interest in livestock populations

Adrián López-Catalina [a,b], Valentin Costes [c], Ramón Peiró-Pastor [a], Hélène Kiefer [d,e], Oscar González-Recio [a,*]

[a] Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC), Crta. de la Coruña km 7.5, Madrid 28040 Spain
[b] Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, Madrid 28040, Spain
[c] R&D Department, ELIANCE, 149 rue de Bercy, Paris 75012, France
[d] Université Paris-Saclay, UVSQ, INRAE, BREED, Jouy-en-Josas 78350, France
[e] Ecole Nationale Vétérinaire d'Alfort, BREED, Maisons-Alfort 94700, France

HIGHLIGHTS

- Nanopore sequencing can simultaneously detect genomic variants and DNA methylation throughout the genome, allowing to investigate their joint role in traits of interest in livestock population.
- In comparison to RRBS, Nanopore sequencing detected a higher number of CpG sites located throughout the genome, not restricted to promoter sites, providing a broader and complementary view of methylation patterns.
- RRBS allowed for a larger coverage of promoter regions, facilitating a better definition of the methylation rate. Further research and investment endeavour could increase ONT sequencing depth.
- Both techniques identified differentially methylated genes linked to economically significant traits in animal breeding, suggesting potential biomarker applications.
- This study highlights the advantages of ONT sequencing for large-scale epigenotyping efforts in livestock, offering economic benefits and new avenues for research.

ARTICLE INFO

Keywords:
Cattle epigenetics
Methylation marks
Semen methylation
Nanopore sequencing

ABSTRACT

Epigenetic marks could potentially explain a portion of the phenotypic variance related to traits of interest in animal breeding models. DNA methylation is the most studied epigenetic mark, involving the addition of a methyl group to the fifth carbon of a cytosine, which transforms it into 5-methylcytosine. This mark is commonly associated with inhibiting gene expression without altering the nucleotide sequence, particularly when located within promoter regions. While bisulphite sequencing is the gold standard for detecting methylation marks, new techniques have emerged to address bisulphite's limitations. Some of these limitations include the confounding effects of bisulphite treatment on DNA methylation and C to T polymorphisms, as well as the inability to distinguish between DNA methylation and hydroxylation. In this context, we propose employing Nanopore sequencing to identify methylation marks. We conducted sequencing on six bull semen samples using both Nanopore and reduced representation bisulphite sequencing. CpGs were filtered based on two coverage thresholds (4x and 7x). Our results indicated that Nanopore methylation data exhibited a correlation exceeding 0.95 with bisulphite sequencing data. The correlation decreased to 0.67 when analysing only CpGs with methylation frequencies ranging from 0.1 to 0.9, due to the lack of enough coverage for ONT. We also examined differentially methylated cytosines identified by each sequencing method. The overlap between the two sets of DMCs and the associated genes was limited, as RRBS library preparation predominantly amplifies gene promoter regions, while ONT covers more intergenic regions. Interestingly, both methods highlighted differentially

* Corresponding author.
E-mail address: gonzalez.oscar@inia.csic.es (O. González-Recio).

https://doi.org/10.1016/j.livsci.2023.105377
Received 26 September 2023; Received in revised form 6 November 2023; Accepted 21 November 2023
Available online 22 November 2023

methylated genes and positions linked to quantitative trait loci of significant economic interest in animal breeding, suggesting promising applications for ONT in the field of agri-genomics.

1. Introduction

Genetic information has become an essential tool in the field of agri-genomics for diagnostic, research, and prediction of phenotypes of interest. However, state-of-the-art technologies enable us to go beyond genetics and delve into the fast-growing field of epigenetics. Discovering epigenetic marks can help detecting causal modifications associated not only with diseases or phenotypic traits of interest in personalized human medicine but also in agriculture and livestock.

The epigenome acts as a means of "communication" between the environment and the genome, modifying gene regulation and expression without altering the nucleotide sequence (Halušková et al., 2021). DNA methylation is an epigenetic modification involving the addition of a methyl group to the fifth carbon of a cytosine, creating a 5-methylcytosine (5mC). This process is mediated by DNA methyltransferases (DNMT1, DNMT3A, DNMT3B, and cofactor DNMT3L) (Miranda and Jones, 2007). Methylation can affect gene regulation by inhibiting gene transcription when located near regulatory regions or gene promoters, often associated with high CpG density regions known as CpG islands (Rakyan et al., 2011).

Genomic information alone explains only a portion of the phenotypic variance of traits of interest in animal production. Part of this variance could be embedded in the epigenome waiting to be harnessed (Mi et al., 2021; Ibeagha-Awemu and Zhao, 2015). Studying methylation marks across the genome of various livestock species could enhance the accuracy of models used in breeding programs. These marks hold special importance as some are modulated throughout the individual's life time (Cattaneoet al., 2022) and are proposed as biomarkers for phenotypes of interest, which may be valuable additions to breeding programs. Moreover, they can be influenced by both animal and environmental conditions even during stress periods, impacting on the phenotype of interest. Epigenetic marks may be altered by animal husbandry and nutrition practices. The combination of genetic and epigenetic information enhance the accuracy of predictive models.

Developing new tools for large-scale implementation is essential. One of these tools is the methylation chips that are specifically designed for livestock species. These methylation chips should encompass epigenetic marks linked to production and health traits of relevant importance in livestock populations. Traits like fertility, resistance to mastitis, heat stress, and metabolic stress play a pivotal role in ensuring dairy operation productivity and sustainability. An "epichip" would facilitate high-throughput DNA methylation analysis, providing a comprehensive grasp of epigenetic modifications associated with specific health conditions. The availability of methylation chips would also streamline large-scale genomic studies, enabling the integration of epigenetic data with other genomic information to advance our understanding of the intricate interplay between genetics, epigenetics, and the environment in animal production (Yousefi et al., 2013).

In dairy cattle, one such critical trait is male fertility. Epigenetic changes, such as methylation, are crucial to ensure proper spermatozoa differentiation. A primary global erasure of methylation marks from the sire genome is necessary in these cells to establish methylation marks in the foetus, continuing into adulthood (Carrell, 2012). Changes in the sperm methylome can correlate with atypical sperm epigenomes and conditions such as difficulties in fertilization, unfavorable pregnancy outcomes, reduced sperm count or abnormal sperm morphology, inadequate embryonic development, and metabolic disorders impacting the offspring (Perrier et al., 2018). Studying how methylation marks behave in fertile and sub-fertile bulls can shed light on mechanisms influencing cattle reproduction. Detecting these signals and unraveling their role in phenotype expression is relevant for mitigation and adaptation strategies in the agriculture and livestock sectors. Various strategies, including nutrition, environmental control, or gene editing, have been proposed to modulate or remove these marks.

Selecting and uncovering epigenetic marks to include in a methylation chip is challenging due to technical obstacles and cost implications. Whole genome bisulphite sequencing (WGBS) and reduced representation bisulphite sequencing (RRBS) are the most prevalent techniques for methylation studies (Zheng et al., 2021). These technologies use a sodium bisulphite treatment to convert the DNA methylation status into a C to T polymorphism (unmethylated C being changed to T while 5-mC is resistant to the transformation and remains a C). However, before sequencing, bisulphite-treated DNA undergoes a PCR amplification to determine cytosines that resisted the treatment and which ones were converted into uracils (Cokus et al., 2008; Meissner et al., 2005). This amplification, as well as the aggressive nature of bisulphite treatment, causing DNA fragment breaks, can lead to a conversion bias, over-representing some amplified sites and under-representing others, and thus introducing biases in the animal's epigenetic profile (Olova et al., 2018). Nanopore sequencing (ONT) represents an alternative approach that can address these limitations. With ONT, DNA strands can be sequenced without PCR amplification or chemical labeling of samples (Clarke et al., 2009). This technology serves as an alternative or complementary tool to bisulphite sequencing, enabling the study of epigenetic marks on an individual scale.

We previously used RRBS to identify fertility biomarkers in the bull sperm methylome (Costes et al., 2022). The objective of the present study was to investigate the potential of Nanopore sequencing as a complementary source of candidate CpGs related to bull fertility that could be used to develop the EpiChip and propose a pipeline that could be routinely implemented in the agri-genomics field.

2. Material and methods

2.1. Sample collection, DNA extraction and sequencing

A total of six semen samples were sequenced for this study using both RRBS and ONT techniques. The samples included semen from three fertile and three sub-fertile bulls. Approximately 20 million spermatozoa were lysed in the presence of 50 mM dithiothreitol and 0.2 mg/ml proteinase K, as described in (Costes et al., 2022). A DNA extraction was conducted using phenol:chloroform (1:1) and chloroform, then precipitated with ethanol and washed. The dried pellet was subsequently resuspended in TE buffer (10 mM Tris–HCl, pH 7.5, 2 mM EDTA), and the concentration was measured using a Qubit Fluorometer. The integrity of the genomic DNA was confirmed through agarose gel electrophoresis. All samples underwent sequencing using both RRBS and ONT methods to facilitate the comparative analyses.

Reduced Representation Bisulphite Sequencing. The DNA was digested using MspI, subjected to end-repair, and ligated with Illumina adapters. Subsequently, the DNA was size-selected using SPRIselect magnetic beads (Beckman-Coulter) and subjected to bisulphite conversion twice using the EpiTect kit (Qiagen). DNA was amplified using Pfu Turbo Cx hotstart DNA polymerase (Agilent). Libraries were quantified using a Qubit 2.0 Fluorometer and sequenced on an Illumina HiSeq4000 sequencer to generate 75 bp paired-end reads. Details on this protocol can be found in (Costes et al., 2022)

Nanopore. The samples were fragmented as recommended by Oxford Nanopore for MinION and GridION sequencing to an average length of 8kb using the Covaris g-TUBE. The sample was centrifuged at 8,600 rpm during 1 min to maximise library preparation yield according to the instructions provided by the manufacturer to fragment up to 15 μg of

mass in a total volume of 150 μl (https://www.covaris.com/g-tube-pr520079). Individual DNA libraries were prepared, starting with 3 μg of DNA and following the protocol from the SQK-LSK110 kit provided by the manufacturer. Subsequently, the library was sequenced using the GridION sequencer (Oxford Nanopore Technologies, ONT, Oxford, UK). Each sample was loaded onto an R9.4 flow cell. Basecalling was carried out using Guppy version 5.0.14 (nanoporetech.com/community), and methylation was called using the model dna_r9.4.1_450bps_modbases_5mc_cg_sup.cfg. A quality control threshold of Q-score > 10 was applied, and NanoPlot (v1.36.2) (De Coster et al., 2018) was used to perform quality control.

2.2. Methylation calling

RRBS. The methylation analyses were conducted using the Bismark software (v0.20.0) (Krueger and Andrews, 2011). Initially, the sample was aligned to the reference genome (ARS-UCD1.2) using Bowtie 1.2.1 (Langmead and Salzberg, 2012). CpGs in unplaced scaffolds were excluded. The remaining CpGs were assigned a methylation percentage using Bismark's methylation calling algorithm. Further details can be found in (Costes et al., 2022).

Nanopore. The analyses were executed using Nanopolish (v0.13.2) (Loman et al., 2015), following the recommended pipeline by its author. The fast5 and fastQ files obtained were indexed and then aligned against the Bos taurus reference genome (ARS-UCD1.2) using minimap2 v2.17-r941 (Li, 2018). Methylation sites were subsequently identified using the call_methylation function from Nanopolish, and the frequencies were extracted using the calculate_methylation_frequency.py script. This pipeline was established at the CESGA super-computing centre in Galicia, Spain.

2.3. Differentially methylated sites

The same differential methylation analysis was performed for all samples using the DSS package (Park and Wu, 2016) in R software. A filter was set to remove positions that were not present or enough covered in every analysed sample. These analyses were conducted at minimum coverage thresholds of 4x and 7x. Each cytosine was required to be covered above these thresholds in all the 6 samples individually, or filtered out otherwise. The list of regions present in every sample was then annotated to remove positions located more than 50 kb away from a transcription start site (Porto-Neto et al., 2014). To estimate the mean methylation frequencies for each group, the smoothing option was selected as 'True' as recommended by the authors for whole-genome analyses. The differentially methylated cytosines (DMC) were defined as significant when their p-value was lower than 0.05. Both the raw p-value and the false discovery rate (FDR) were considered at this step. Then, the resulting DMCs were filtered by delta (δ=0.2), which was the minimum value for the difference of the estimated methylation means for each group.

2.4. Comparison of called CPGS

ChipSeeker (Yu et al., 2015) was used to plot the CpGs with a coverage ≥ 4x surrounding the TSS to study the density of those around promoter sites. The distance of the CpGs to the closest TSS and the genomic features associated to them were also estimated. A kernel density estimation (KDE) was used to study the location of the CpGs in the chromosome.

The correlation between the methylation frequency of those CpGs sites detected by ONT and RRBS, was calculated using the aforementioned mentioned coverage thresholds of 4x and 7x. The agreement of genome-wide methylation detected by both techniques was analysed using the R package ChipSeeker without filtering by coverage. For this, methylation percentage in genomic regions, concordance between detected genes and detection of CpGs close to promoters were studied. Promoter regions were defined as −3000 and +3000 bp from ATG transcription-start codon.

ChipSeeker was used to find the proximity of the DMCs from each technique to TSS through the function *tagHeatmap,* the genomic annotation via *plotAnnoBar* and the density of DMCs in the proximal regions of the gene bodies using *plotPeakProf2.* Minimum coverages of 4x and 7x were tested.

The DMCs present in quantitative trait loci (QTL) were described using the files available in the Cattle QTL database (Hu et al., 2022). The QTLs were grouped by trait and plotted together to investigate the possible role of the epigenetic marks detected by each technique (Liao et al., 2019).

3. Results

3.1. Correlation between ONT and RRBS

The mean coverage was of 7x for Nanopore and 22x for RRBS. The average number of CpG sites called by each technique and their standard deviation were 2,467,573 ± 48,578.47 for bisulphite sequencing and 20,030,834 ± 1,150,110 for ONT. Nanopore sequencing detected 8.12x ± 0.50 more sites compared to RRBS when no filter was applied, 5.1 ± 1.64 for a filter of 4x and 1.87 ± 0.86 for a filter of 7x (Fig.1).

Table 1 shows the correlation between the methylation frequency detected by each method at different coverage thresholds. The more stringent the filter, the smaller number of sites detected, but the larger the correlation. A larger number of sites and a correlation greater than 0.95 was obtained at 4x in all the samples. Filtering by 4x provides with double the sites than by 7x and the correlation only decreased by 0.02. Removing sites with a methylation frequency lower than 0.1 or greater than 0.9 drops the correlations to 0.596 ± 0.05 and 0.651 ± 0.05 for 4x and 7x respectively. The 97 % of the sites detected were removed when filtering by coverage, from an average of 930,568 to 35,252 4x, and from 584,124 to 18,359 for 7x. These correlations are depicted in Fig. 2. The expected trend was a stronger correlation in the diagonal, proving an agreement between the frequencies called by both techniques. Larger agreement was observed at low or high methylation frequencies. The more stringent the filter used, the clearer the tendency, which suggests that a larger coverage threshold would help to improve the correlation at the expense of a lower number of CpG sites. A horizontal line marks the 0.5 frequency due to the relatively low coverage of the ONT data. The filter of a minimum coverage of the CpGs of 4x leads to the inclusion of many CpG sites with only 4 possible methylation frequencies, 0; 0.5; 0.75 and 1. This line denotes these combinations that did not happen in RRBS as the mean coverage was higher than ONT's.

3.2. Genomic position distribution of the CPGS covered by ONT and RRBS

The density of positions at transcription start sites (0 in X axis) was larger using ONT, also displaying a larger dispersion than RRBS (Fig. 3). CpG sites detected by RRBS appeared mainly in regions close to promoters, whereas ONT detected a larger number of CpGs intergenic regions (Fig.4).

Methylation marks act differently depending on their genomic position. Methylation in the promoter region is usually related to gene silencing whereas its role in other regions is more variable and could be related to the initiation of transcription from the alternative transcription start sites or the expression of ncRNAs and microRNAs, although this is still under study (Zhou et al., 2015). Nanopore sequencing detected a larger number of candidates in distal intergenic regions (>100 kb from the promoter) and intronic regions. Bisulphite sequencing revealed a larger proportion of promoter regions as expected from the restriction enzymes and size selection window used during the library preparation for RRBS (Figs. 4and 5).

The density of CpGs increased around the TSS, and then a sudden

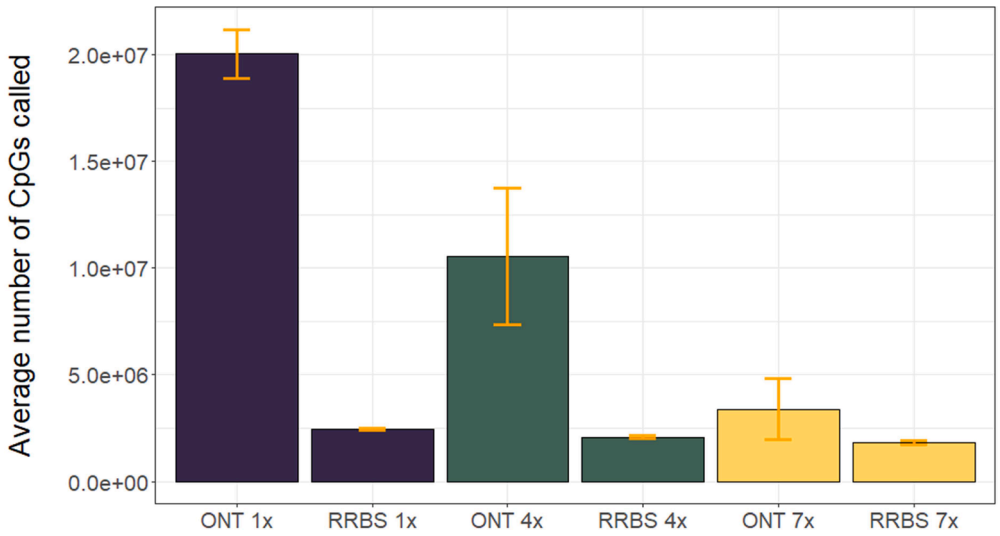


Fig. 1. Average CpG sites called (y-axis) by each technology (x-axis). Colours represent the coverage filter applied, dark blue for unfiltered files, green for a minimum coverage of 4x, and yellow for 7x. Orange vertical lines represent the standard deviation.

Table 1
Summary statistics for the number of common sites detected with both techniques at different coverages (4x or 7x), and the correlation between the methylation frequency from both methods. Rows gather the information by sample and coverage threshold. The first third column represents the correlation between the methylation frequency estimated by each techniques with the methylation frequency >0.1 and <0.9 filter applied. The fourth column represents number of sites detected by the coverage and thresholds filters shown in the second column. The fifth and sixth columns gather the same information without the methylation frequency filters.

Sample	Coverage	Correlation [0.1–0.9]	Sites filtered [0.1–0.9]	Correlation [0–1]	Sites [0–1]
1	4x	0.631	49,656	0.961	1,188,642
1	7x	0.683	26,620	0.962	759,621
2	4x	0.648	48,355	0.961	1,130,786
2	7x	0.7	26,675	0.956	720,805
3	4x	0.617	42,609	0.963	1,149,859
3	7x	0.666	22,427	0.961	749,179
4	4x	0.582	7,870	0.962	222,448
4	7x	0.642	3,851	0.961	133,258
5	4x	0.494	17,720	0.959	785,032
5	7x	0.552	7,408	0.953	445,967
6	4x	0.607	45,300	0.959	1,106,641
6	7x	0.66	23,175	0.958	695,911

drop followed by a plateau afterwards was observed (Fig. 6). The peaks in Fig. 6 represent the density of sites detected by each technique in genomic regions in proportion to the total amount of sites detected. This larger amount of CpGs around the TSS is likely related to the molecular mechanism of gene silencing, allowing the expression or repression of different genes in different tissues under different conditions. The total number of sites detected by ONT was larger than that of RRBS, with a higher proportion of positions detected in distal intergenic regions. ONT detected approximately 400,000 CpGs in the promoter region, whereas RRBS detected around 500,000 CpGs in the same region, corresponding to 5 % and 20 % of the sites, respectively (Fig. 5). The KDE shows a higher density of the marks at the end of the chromosomes, especially in the case of RRBS (Fig. 7). The higher the coverage filter, the more concentrated the sites appeared at the end of the chromosomes.

3.3. *Differential methylation analyses*

The analyses were performed using DSS to identify the DMCs related to bull fertility obtained using RRBS and ONT. Samples sequenced using RRBS 4x and RRBS 7x showed 4643 DMCs, while ONT 4x found 2001 DMCs and ONT 7 × 209 DMCs.

Coverages of 4x and 7x were compared to understand how increasing the sequencing coverage threshold could affect the detection of DMCs. Fig. 8 depicts the differences in the estimated means between fertile and subfertile bulls. A DMC having a negative difference is associated with a lower methylation frequency in the fertile group (hypomethylation of the DMC) meanwhile a positive difference is translated into a hypermethylation of that DMC with respect to the fertile group. The distribution of DMC hypomethylated in the fertile group called by RRBS gathered around −0.25, while the distribution of those detected by ONT was wider.

Fig. 9 shows that ONT detected more distal intergenic DMCs than RRBS, while the latter detected a larger proportion of marks associated with promoter regions. The results for the DMCs seem consistent with the results of the genomic location of the detected CpGs in the previous subsections. Fig. 10 depicts the distance of the DMCs to the closest TSS. RRBS tended to detect a larger number of marks closer to the TSS than ONT, which detected a larger proportion of differentially methylated cytosines located in distal regions. Differentially methylated cytosines were identified and assigned to their respective genes. RRBS4x called DMCs associated with 356 genes, accounting for approximately 39 % of the total differentially methylated genes. RRBS7x identified 312 genes, which represented around 35 % of the total differentially methylated genes. 853 genes were associated with the DMCs detected by ONT4x. Lastly, ONT7x called DMCs in 119 genes, making up approximately 13 % of the total genes detected in this work. Fig. 11 shows the overlap of those genes identified by both sequencing methods. Specifically, ONT4x identified 668 genes that were not detected by any other method,

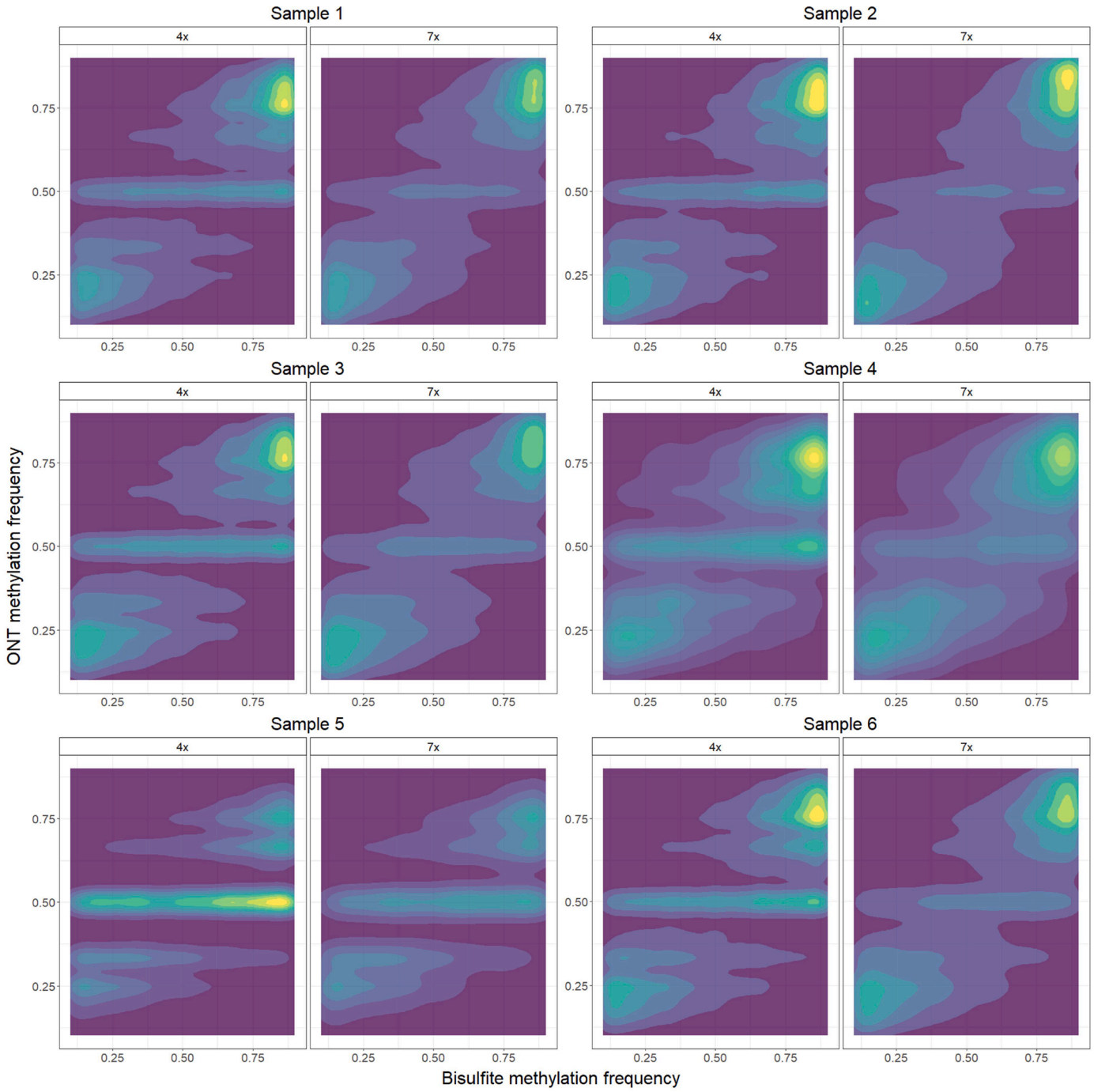


Fig. 2. Heatmaps representing the correlation between the methylation frequency called for the same position by both ONT and RRBS. Sites with a frequency >0.9 or <0.1 were filtered out. Each duo represents a sample with two different coverage filter applied: 4x and 7x. Y-axis represents the methylation frequency called by ONT and x-axis the one called by RRBS.

accounting for around 74 % of the genes identified by ONT4x. Nineteen genes were detected only by ONT7x. Fourty-seven genes were called only by RRBS4x. Similarly, RRBS7x called 9 genes that were not detected by other methods. In terms of agreement, a total of 75 genes were identified by all four methods, which represented approximately only 8 % of the total differentially methylated genes. ONT4x resulted on the lesser overlap with other DMC sets.

3.4. QTL associated with the DMCs

The possible role of the methylation of these DMCs was studied by positioning the detected DMC searched in the animal QTLdatabase. A total of 211 different QTLs out of 1579 were found associated to the DMCs detected by ONT4x (14 %), 117 out of 513 by ONT7x (23 %), 185 QTLs out of 1209 by RRBS4x (15 %) and 175 QTLs out of 1143 by RRBS7x (15 %). The majority of the QTLs annotated were related to traits of relevance in cattle production such as body and carcass weight, marbling score, milk yield, feed intake and calving easy. The QTLs detected are shown in Fig. 12. Out of the total 211 QTLs previously described, 30 were directly associated with bull fertility. Table 2 summarises the number of QTLs associated to different male fertility traits called by each technique. Most of them were related to scrotal circumference. Fig. 13 depicts the overlap between the calling made by each technique is depicted. ONT 4x was the technique that called a larger

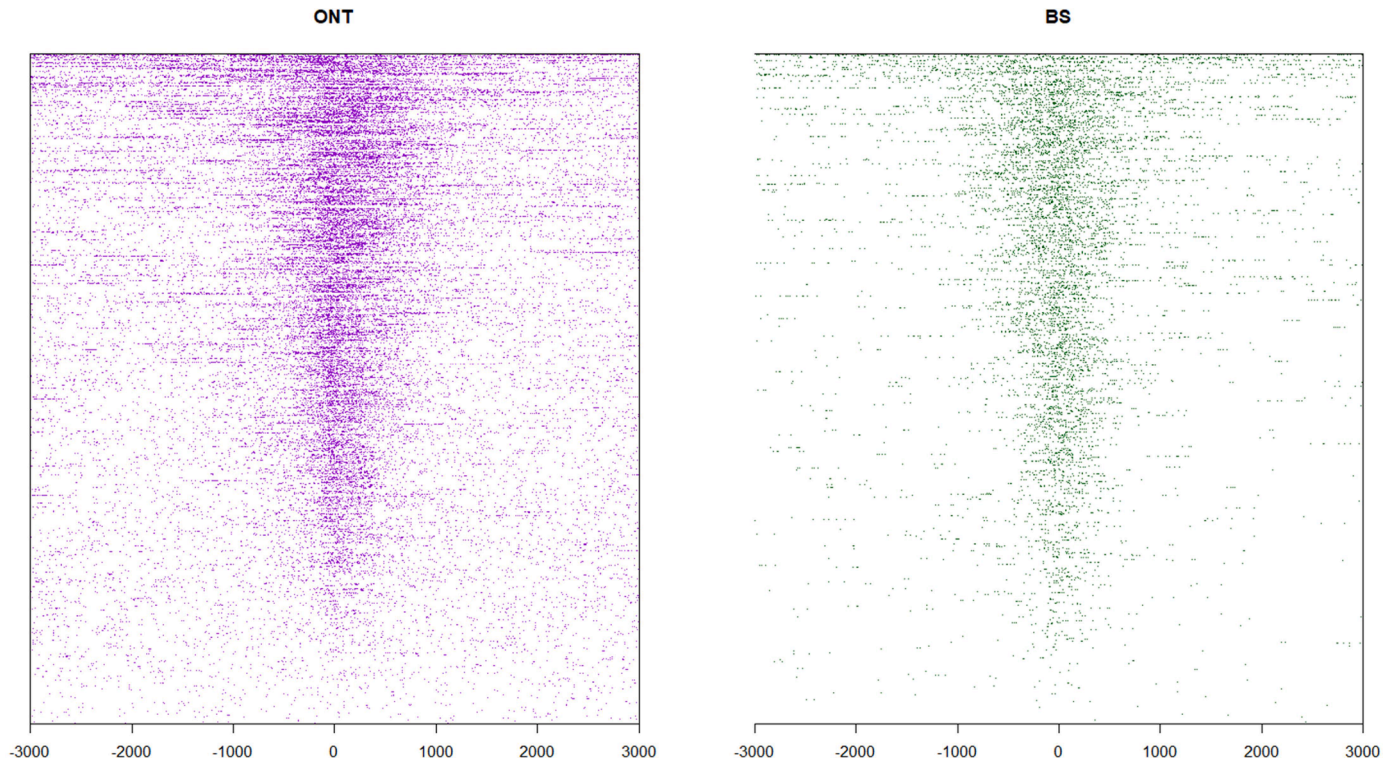


Fig. 3. Heatmaps for one representative sample (sample 1) representing the presence of a CpG candidate to be methylated using nanopore technology (purple) or RRBS (green), after filtering out sites with a sequencing depth lower than 4x. The x-axis denotes the position of the CpG relative to the transcription start site (=0).

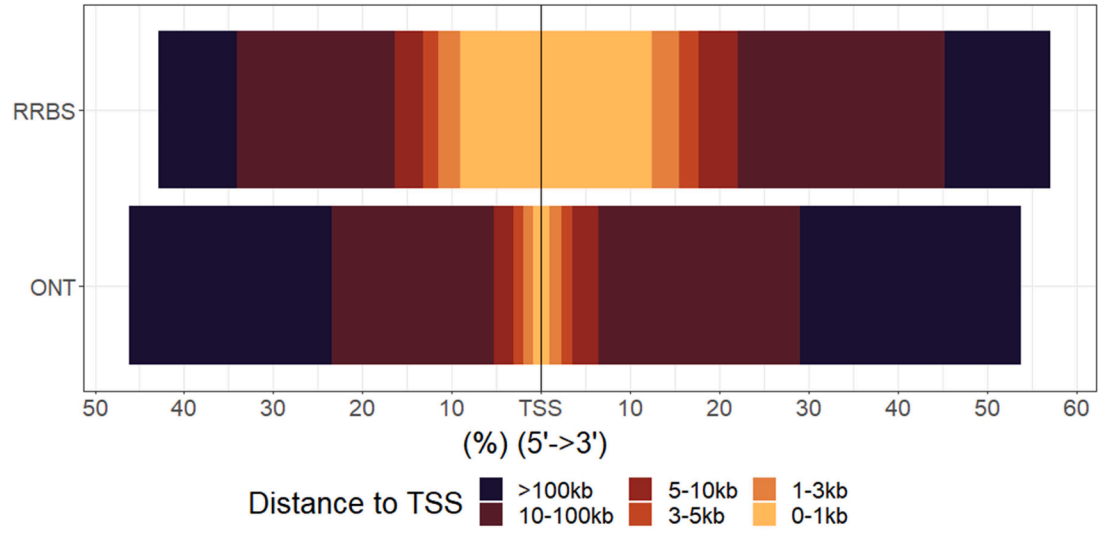


Fig. 4. Barplots representing the distance to the TSS of the detected positions at 4x moving upstream to its left and downstream to its right.

number of fertility related QTLs (48) while ONT 7x detected the fewer (12). RRBS 4x and RRBS 7x detected 38 and 33 respectively.

4. Discussion

This study compared the performance of ONT and RRBS for detecting methylation marks in the genome to be potential candidates to be included in a methylation chip. The results showed that ONT covered a much larger number of CpG sites than RRBS despite the mean coverage was three times larger for RRBS. Furthermore, ONT detected a mean of 8.12x more sites no coverage filter was applied, and 5.1x and 1.87x for coverage filters of 4x and 7x. The correlation between the methylation frequency measured by ONT and RRBS was also tested, resulting in a smaller number of sites detected at a higher coverage, but with a larger correlation. These correlations were >0.95 when comparing all the positions called by both techniques. However, intermediate methylation frequencies were not accurately estimated at a coverage <7x.

The distribution of the genomic positions for the detected CpGs was also analysed. As expected, ONT detected a larger proportion of methylated regions in intergenic regions, while RRBS mainly detected methylation sites in proximal regions to promoters. In absolute numbers, ONT detected an average of around 11 million methylation sites at $\geq$4x, 4 % of them in promoter regions, whereas RRBS detected an average of around 2 million sites at 4x (20 % in promoter regions). Bisulphite

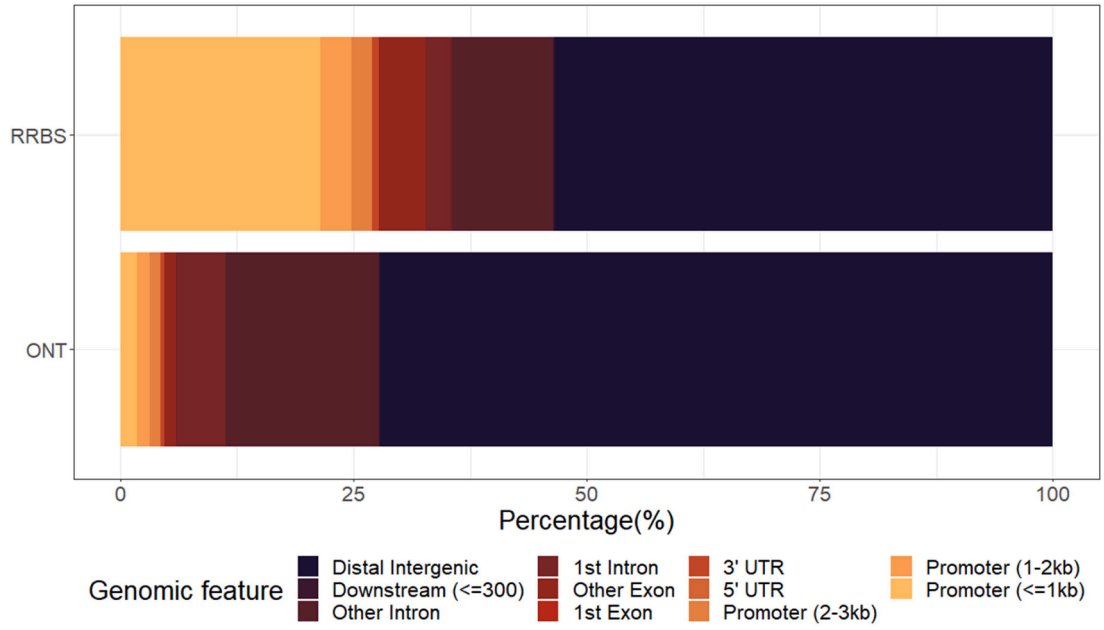


Fig. 5. Barplots representing the features where the CpGs are detected at 4x, considering promoter regions those closer than 3 kb to the TSS.

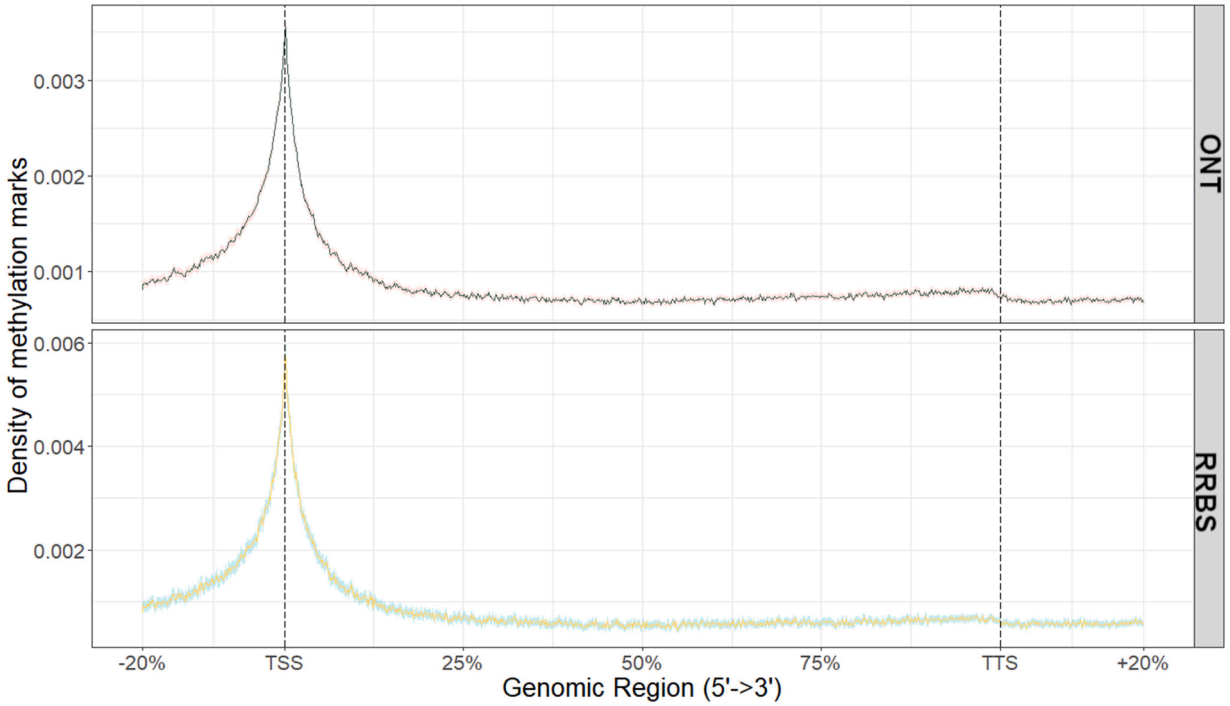


Fig. 6. Density of the CpG positions detected by ONT in the upper side of the figure and RRBS in the bottom. The genomic region is windowed to depict the transcription start and termination sites (TSS and TTS) and the gene body (region in-between the TSS and the TTS).

sequencing detected a larger proportion of sites in promoter regions but a similar number in absolute terms (ONT 421,881 vs RRBS 415,034). The role of DNA methylation in the promoter region is related to gene silencing, while the presence of these marks in intergenic regions is likely related to the initiation of transcription from the alternative transcription start sites or the expression of ncRNAs and microRNAs, so the quantification of the methylation status of those regions is of great interest to understand gene regulation and these marks can be used as potential biomarkers (Schübeler, 2015). Despite the larger sequencing depth for RRBS, the number of sites detected in promoter regions was similar to ONT.

An increased CpG density was found in the transcription start site, with a pronounced drop before and after it, where it showed a plateau. The CpG density pattern did not differ significantly between RRBS and ONT and behaved as expected in superior eukaryotes (Buitrago et al., 2021). The distribution for ONT differed slightly as this technique detected a larger proportion of CpGs in distal intergenic regions. The detection of marks in non-genic positions opens the opportunity to study DMCs that affect genes without pointing to the promoter regions. Nanopore sequencing offers some advantages over RRBS because it is

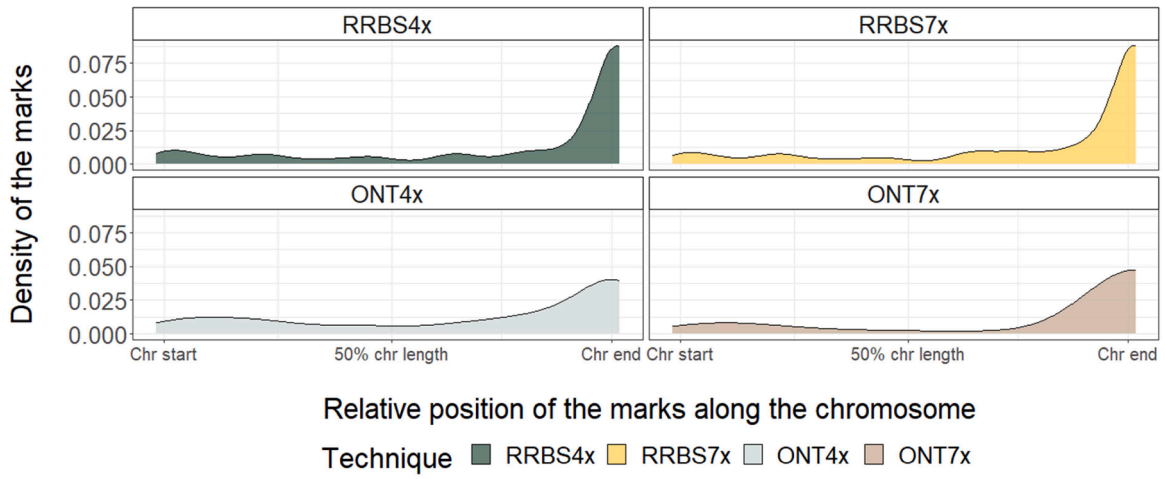


Fig. 7. Density plots showing the position of the DMCs along the chromosomes detected by the different technologies and coverage filters.

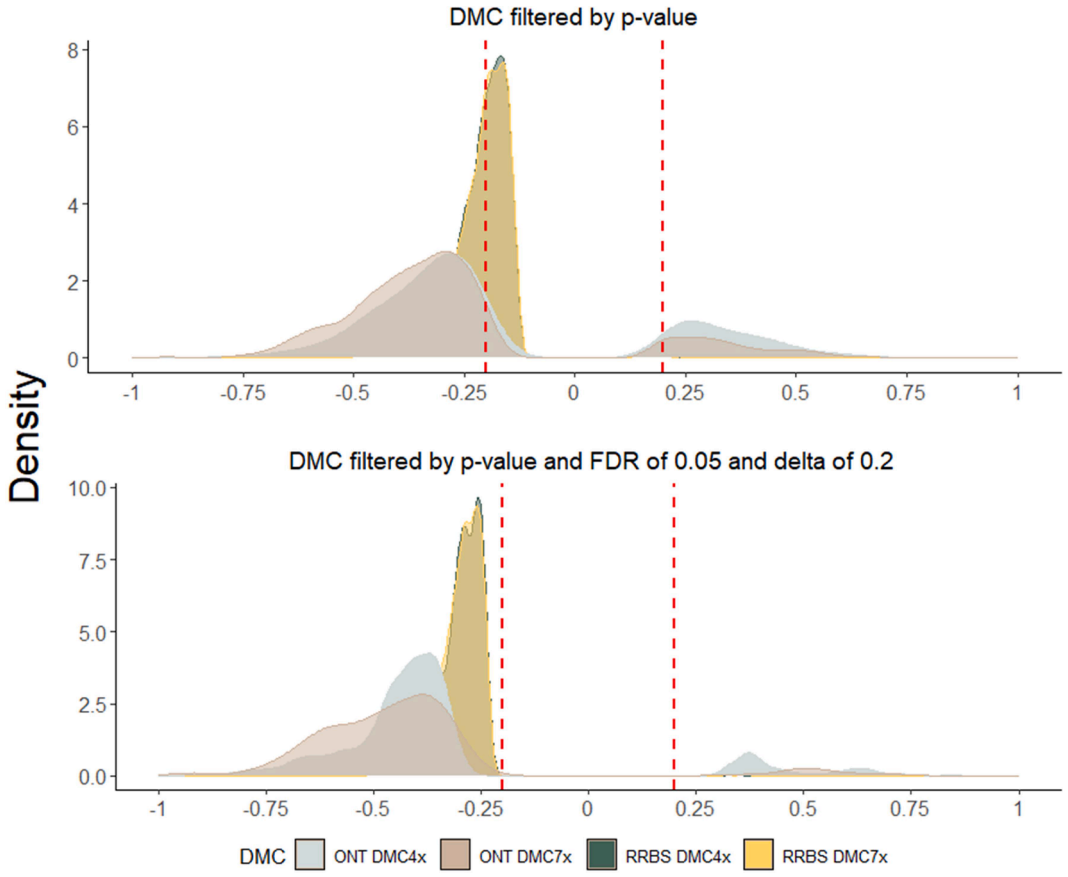


Fig. 8. Density plot for the DMCs detected by ONT and RRBS filtering at 4x and 7x. Plot on the top illustrates the density of the DMCs filtered by a p-value of 0.05 while the one on the bottom has two additional filters. DMCs were only kept when having an FDR $\leq$ 0.05 and a delta $>$ 0.2. Each colour represents a different sequencing technique and a different coverage threshold.

not restricted to promoter regions. This difference can be explained by the library preparation for each sequencing method. RRBS relays on the restriction enzyme MspI that targets 5′-CCGG-3, a sequence presence in CpG rich regions, generating a poor coverage of regions with a low density of CpGs (Lee et al., 2014). Other bisulphite sequencing alternatives, such as Whole Genome Bisulphite Sequencing (WGBS) do not target CpG-rich regions and could improve the RRBS bias (Beck et al., 2022). ONT can also differentiate between methylated cytosines and C/T SNPs, and it provides a more comprehensive picture of the epigenetic landscape of the genome (Viana et al., 2018). On the other hand, RRBS provides a larger mean sequencing depth within CpGs rich regions at a lower cost. A low sequencing depth of 4x from ONT still resulted in a DMC profile similar to that of RRBS without losing as many DMCs as when filtering at 7X. When annotating the DMCs to the closest gene, we observed that there was a low agreement between the genes detected by ONT4x and RRBS of around 37 % which opens a debate about using these techniques as complementary. This was probably due to the small sample size and the different mean depth in each method.

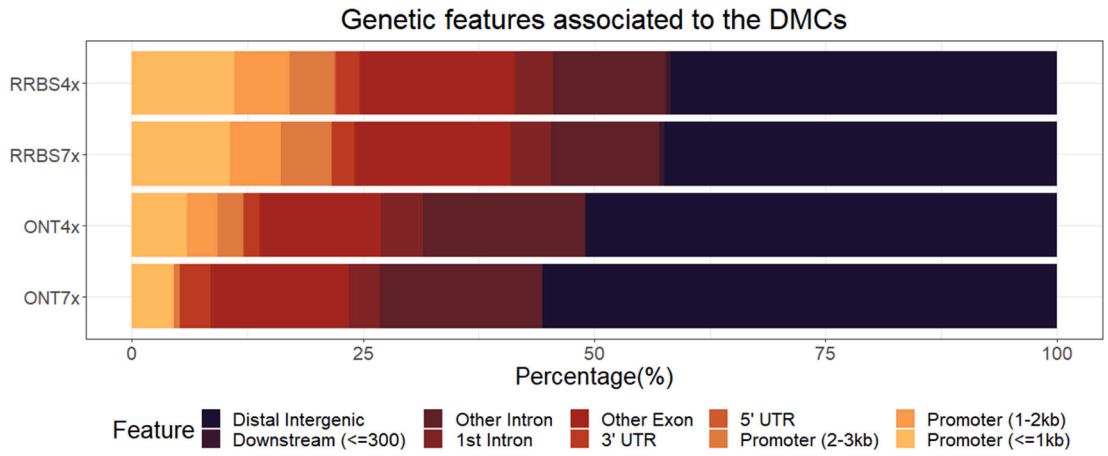


Fig. 9. Distribution of the genetic features associated with the DMCs detected by RRBS, and ONT at a minimum coverage of 4x and 7x.

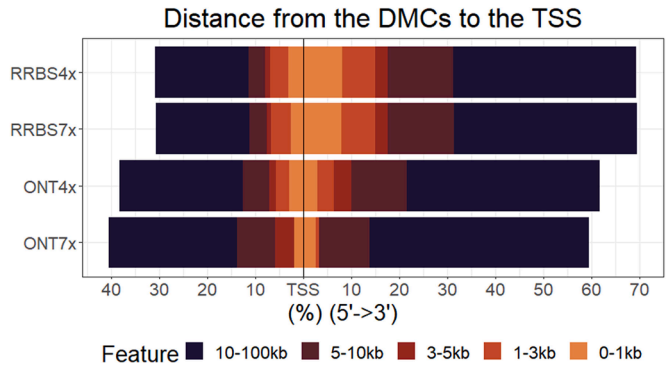


Fig. 10. Distance of the DMCs detected by each technique to the closest TSS.

Also, the different number of CpGs tested for each method, and the impact of multitest correction. Increasing the sequencing depth for ONT may lead to more consistent results between techniques in promoter regions.

The annotated genes by ONT4x and RRBS agreed on the traits associated to their respective QTLs. This is relevant for breeding programmes as new candidate positions of interest can be incorporated in the methylation chips with relative low coverage. Both techniques yielded similar QTLs associated to male fertility. ONT7x detected the fewest DMCs in QTLs related to bull fertility (12), constituting only 2.33 % of the total DMCs associated with QTLs identified by ONT7x. On the other hand, ONT4x identified 48 DMCs in QTLs associated with bull fertility traits (3.04 % of the total). Scrotal circumference was the QTL with the greatest representation in all four instances. This trait plays a crucial role in bull fertility and sperm quality. A larger scrotal circumference typically indicates a higher sperm-producing capacity, leading to increased sperm quantity. Moreover, it often correlates with improved sperm quality, as it allows a more efficient cooling of the testes, maintaining an optimal temperature for sperm production. A larger scrotal circumference is generally associated with larger semen quality and volume enhancing sperm performance and fertility (Menegassi et al., 2019; Almquist et al., 1976; Menon et al., 2011) .

The study of epigenetic marks associated with welfare and production traits is expected to grow in livestock science. Currently, bisulphite sequencing has been the preferred technique to study whole genome methylation status. Nanopore sequencing offers new possibilities to study epigenetics marks in animals, because this technology allows genotyping and epigenotyping simultaneously, and the long-reads are easier to map and phase. Besides, ONT sequencing allows to differentiate between different types of DNA methylation marks (e.g., 5-mC, 5-hmc, 6-mA and 4-mC) which is not possible with bisulphite sequencing (Liu et al., 2021). Nanopore sequencing allows a PCR-free sequencing, avoiding the potential bias associated with this amplification needed for bisulphite sequencing. It must be pointed out that RRBS targets CpG rich regions as it targets CCGG sites, which reduces the possibility of detecting other candidate methylated positions (Sun et al., 2015), although with the advantage of achieving a high sequencing depth at a lower cost. DNA shearing for ONT sequencing is optional and this fragmentation does not generate a bias, as it only generates fragments of a desired size, without targeting any specific region. However, the targeted fragmentation of RRBS poses a limitation to this study, as regions with a large proportion of CpGs (e.g. CpG islands) will be over-represented in contrast to regions with fewer CpGs (e.g. distal intergenic regions). This difficults the comparison of CpGs located away from

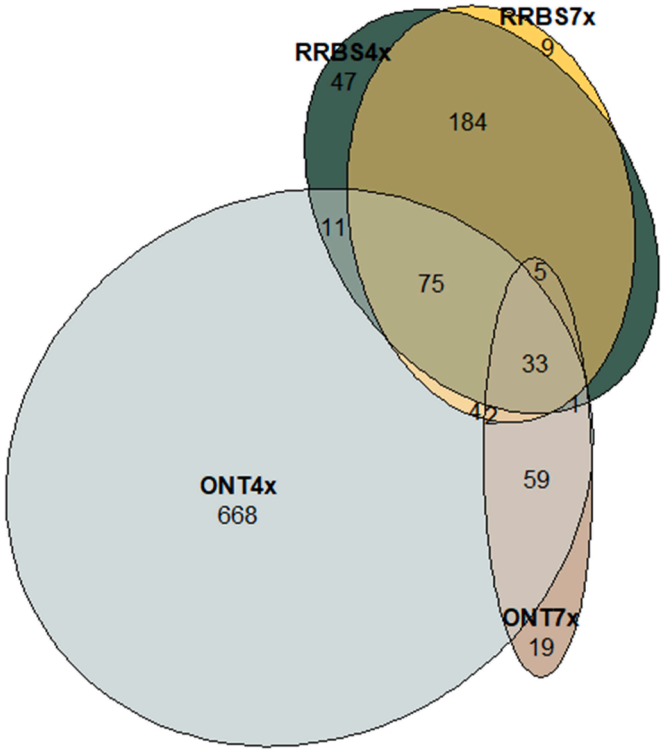


Fig. 11. Euler diagram to represent the overlap of the genes detected by each technique. In grey, ONT by a minimum coverage depth of 4, in beige ONT by 7x. In green and yellow are the DMC-associated genes detected by RRBS4x and RRBS7x respectively.

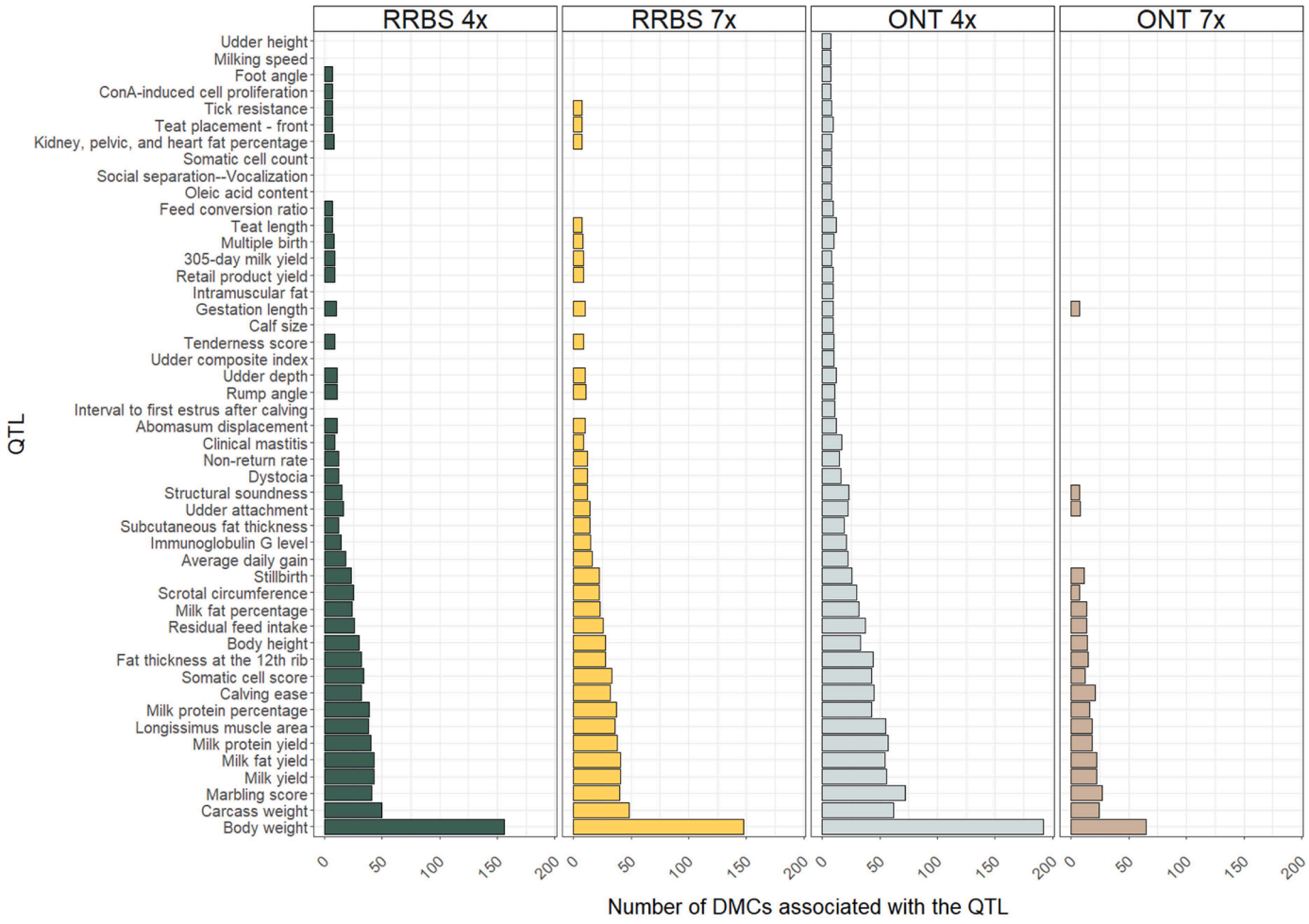


Fig. 12. Barplot illustrating the number of DMC in QTLs associated to traits of interest. Each column represent a technique and coverage, and rows represent the QTL to which the methylation mark can be associated.

Table 2
Number of CpGs associated to bull fertility QTLs that have been called by Oxford Nanopore Sequencing and Reduced representation bisulphite sequencing at coverage thresholds of 4x and 7x. Each column represents the values called by each technique and each row illustrates a QTL In terms of relative figures, RRBS4x proved to be the method that identified a greater percentage of DMCs linked to bull fertility QTLs (3.14 % of the total QTLs), whereas ONT7x was the one that identified a smaller percentage of these QTLs (2.34 %). ONT4x and RRBS7x, on the other hand, identified 3.04 % and 2.89 %, respectively.

QTL	ONT 4x	ONT 7x	RRBS 4x	RRBS 7x
Inseminations per conception	3	0	1	1
Interval from first to last insemination	2	0	1	1
Paired testes volume	1	0	1	0
Paired testes weight	1	0	1	0
Percentage abnormal sperm	2	0	1	1
Percentage live sperm after osmotic stress	1	1	1	1
Percentage live sperm after thawing	2	1	2	2
Pregnancy rate	1	0	0	0
Scrotal circumference	30	7	25	22
Semen volume	1	0	1	1
Sperm average path velocity	1	0	1	1
Sperm motility	3	3	3	3

promoter or genic regions between the methylation frequency estimated by RRBS and ONT. A future ideal approach would add sequencing depth to the ONT data and would include WGBS data to correctly compare intergenic regions. Here, we proposed the complementarity of RRBS and ONT, where RRBS increases the coverage in promoter regions and ONT explores other regions of the genome unreachable to RRBS due to its chemistry.

Our study showed encouraging results to use ONT sequencing in livestock species to add new candidate positions to those proposed by RRBS. By using ONT positions, we can describe new DMCs that RRBS cannot reach due to the use of restriction enzymes and size selection. This information can be integrated to propose biomarkers related to different health and productive traits of relevance in animal production. These positions can be included in a methylation array whose use could be of use in farming practices and breeding programmes.

5. Conclusion

This study shows that RRBS and ONT are complementary techniques to obtain candidate biomarkers for the design of methylation arrays. Nanopore sequencing provided a larger amount of methylation marks far from the promoters and in distal intergenic regions, in addition to the

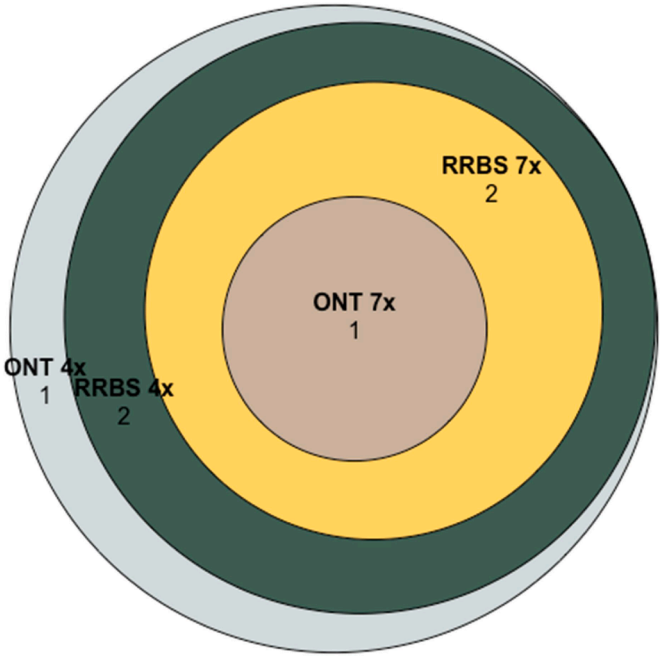


Fig. 13. Euler diagram representing the overlap between the detected QTLs associated to bull fertility by each technique and coverage. The values represent the traits that have been detected by each technique/coverage.

CpG islands also detected by RRBS. Nanopore also detected methylated candidates in distal intergenic regions, although at a lower sequencing depth than RRBS. Nonetheless, both techniques detected DMC in promoter regions associated to similar traits. A set of DMCs associated with male fertility were detected. These potential candidates exhibit distinct epigenetic signatures that may contribute to bull fertility traits. These identified candidates are being included in the development of an epigenotyping chip to facilitate large-scale genotyping efforts. This tool could be routinely used in breeding programmes and research. The role of the epigenetic marks proposed needs to be carefully studied to understand how silencing genes affect the phenotypes and to which degree do so. This chip will enable large-scale screening to disentangle the epigenetic landscape associated with bull fertility and other traits, opening new pathways for reproductive research and improved breeding strategies.

Data availability

The data that support the findings of this study are available on request from the corresponding author.

Authors' contributions

ALC and OGR sequenced the samples using Nanopore Technologies. ALC, OGR and RPP analysed the ONT data. VC and HK sequenced the samples using RRBS and analysed the bisulphite data. ALC wrote the first version of the manuscript. All authors helped to write the final version of the manuscript. The authors read and approved the final manuscript.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101000226.

Open Access funding provided thanks to the CRUE-CSIC agreement with Elsevier.

Ethics approval

The Ethics Committee of the Instituto Nacional de Investigación y Tecnología Agraría y Alimentaria (INIA) waived the need for ethics approval and the need to obtain consent for the collection, analysis and publication of the retrospectively obtained and anonymized data for this non-interventional study with reference CEEA-pH-2021–08. All methods were carried out in accordance with relevant guidelines and regulations from the European Union.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgements

A. Lopez-Catalina predoctoral fellowship from Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, (FPI-SGIT2018-03).

Data was analysed using the CESGA Supercomputing Infrastructure (Galicia, Spain).

References

Almquist, J.O., Branas, R.J., Barber, K.A., 1976. Postpuberal changes in semen production of Charolais bulls ejaculated at high frequency and the relation between testicular measurements and sperm output. J. Anim. Sci. 42, 670–676.

Beck, D., Ben Maamar, M., Skinner, M.K, 2022. Genome-wide CpG density and DNA methylation analysis method (MeDIP, RRBS, and WGBS) comparisons. Epigenetics 17, 518–530.

Buitrago, D., Labrador, M., Arcon, J.P., Lema, R., Flores, O., Esteve-Codina, A., et al., 2021. Impact of DNA methylation on 3D genome structure. Nat. Commun. 12 (1), 1–17, 2021 12.

Carrell, D.T., 2012. Epigenetics of the male gamete. Fertil. Steril. 97, 267–274.

Cattaneo, L., Laporta, J., Dahl, G.E., 2022. Programming effects of late gestation heat stress in dairy cattle. Reprod. Fertil. Develop. 35, 106–117.

Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., Bayley, H., 2009. Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. 4, 265–270, 2009 44.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., et al., 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature 452, 215–219.

Costes, V., Chaulot-Talmon, A., Sellem, E., Perrier, J.P., Aubert-Frambourg, A., Jouneau, L., et al., 2022. Predicting male fertility from the sperm methylome: application to 120 bulls with hundreds of artificial insemination records. Clinic. Epigenetics 14, 1–27.

De Coster, W., D\'Hert, S., Schultz, D.T., Cruts, M., Van Broeckhoven, C, 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34, 2666–2669.

Halušková J., Holečková B., Staničová J. DNA methylation studies in cattle. Available from: https://doi.org/10.1007/s13353-020-00604-1.

Hu, Z.L., Park, C.A., Reecy, J.M., 2022. Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 50, D956–D961.

Ibeagha-Awemu, E.M., Zhao, X., 2015. Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. Front. Genet. 0, 302.

Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9 (4), 357–359, 2012 9.

Lee, Y.K., Jin, S., Duan, S., Lim, Y.C., Ng, D.P., Lin, X.M., et al., 2014. Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. Biol. Proced. Online 16, 1.

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100.

Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z., WebGestalt, Zhang B., 2019. 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 47, W199–W205.

Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., et al., 2021. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol. 22, 1–33.

Loman, N.J., Quick, J., Simpson, J.T., 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. Methods 12 (8), 733–735, 2015 12:

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 33, 5868–5877.

Menegassi, S.R.O., Pereira, G.R., McManus, C., Roso, V.M., Bremm, C., Koetz, C., et al., 2019. Evaluation and prediction of scrotal circumference in beef bulls. Theriogenology 140, 25–32.

Menon, A.G., Barkema, H.W., Wilde, R., Kastelic, J.P., Thundathil, J.C., 2011. Associations between sperm abnormalities, breed, age, and scrotal circumference in beef bulls. Can. J. Vet. Res. 75, 241–247.
Mi, S., Chen, S., Li, W., Fang, L., Yu, Y, 2021. Effects of sperm DNA methylation on domesticated animal performance and perspectives on cross-species epigenetics in animal breeding. Animal Front. 11, 39–47.
Miranda, T.B., Jones, P.A., 2007. DNA methylation: the nuts and bolts of repression. J. Cell. Physiol. 213, 384–390.
Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R.V., Branco, M.R., et al., 2018. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. Genome Biol. 19, 1–19.
Park, Y., Wu, H., 2016. Differential methylation analysis for BS-seq data under general experimental design. Bioinformatics 32, 1446–1453.
Perrier, J.P., Sellem, E., Prézelin, A., Gasselin, M., Jouneau, L., Piumi, F., et al., 2018. A multi-scale analysis of bull sperm methylome revealed both species peculiarities and conserved tissue-specific features. BMC Genomics 19.
Porto-Neto, L.R., Kijas, J.W., Reverter, A, 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genetics Select. Evol. 46, 1–5.
Rakyan, V.K., Down, T.A., Balding, D.J., Beck, S., 2011. Epigenome-Wide Association Studies for common human diseases. Nat. Rev. Genet. 12, 529.
Schübeler, D., 2015. Function and information content of DNA methylation. Nature 517, 321–326, 2015 517:7534.
Sun, Z., Cunningham, J., Slager, S., Kocher, J.P., 2015. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. Epigenomics 7, 813.
Viana, A.G.A., Martins, A.M.A., Pontes, A.H., Fontes, W., Castro, M.S., Ricart, C.A.O., et al., 2018. Proteomic landscape of seminal plasma associated with dairy bull fertility. Sci. Rep. 8.
Yousefi P., Huen K., Schall R.A., Decker A., Elboudwarej E., Quach H., et al. Considerations for normalization of DNA methylation data by Illumina 450K BeadChip assay in population studies. https://doi.org/104161/epi26037. 2013;8: 1141–52.
Yu, G., Wang, L.G., He, Q.Y., 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383.
Zheng, X., Wu, Q., Wu, H., Leung, K.S., Wong, M.H., Liu, X., et al., 2021. Evaluating the consistency of gene methylation in liver cancer using bisulfite sequencing data. Front. Cell Dev. Biol. 9, 1022.
Zhou, Z.Y., Li, A., Wang, L-G, Irwin, D.M., Liu, Y.H., Xu, D., et al., 2015. DNA Methylation Signatures of Long Intergenic Noncoding RNAs in Porcine Adipose and Muscle Tissues OPEN. Nature Publishing Group [Internet]. Available from: www.nature.com/scientificreports.