

UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de
Biosistemas



**Unveiling the Impact of Maternal
Intergenerational Effects on the
Methylome of Newborn Holstein Calves
Using Long-Read Sequencing**

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Adrián López Catalina

BSc in Biotechnology
MSc in Animal Genetics and Breeding, and Reproductive Biotechnology
MSc in Biostatistics

Madrid, 2025



UNIVERSIDAD POLITÉCNICA DE MADRID
Escuela Técnica Superior de Ingeniería
Agronómica, Alimentaria y de Biosistemas

**Doctoral Degree in Agro-environmental Technology for
Sustainable Agriculture**

**Unveiling the Impact of Maternal
Intergenerational Effects on the
Methylome of Newborn Holstein Calves
Using Long-Read Sequencing**

DOCTORAL THESIS

Submitted for the degree of Doctor by:

Adrián López Catalina

BSc in Biotechnology

MSc in Animal Genetics and Breeding, and Reproductive Biotechnology

MSc in Biostatistics

Under the supervision of:

Dr. Óscar González Recio

Madrid, 2025

Title: Unveiling the Impact of Maternal Intergenerational Effects on the Methyome of Newborn Holstein Calves Using Long-Read Sequencing.

Author: Adrián López Catalina

Doctoral Programme: Doctoral Degree in Agro-environmental Technology for Sustainable Agriculture

Thesis Supervision:

Dr. Óscar González Recio

External Reviewers:

Thesis Defense Committee:

Thesis Defense Date:

This thesis has been partially supported by **METALGEN**, funded by the Ministry of Economy and Competitiveness with reference number RTA2015-00022-C03-02. **RUMIGEN**: Towards improvement of ruminant breeding through genomic and epigenomic approaches. Funded by the Horizon 2020 programme of the European Union with reference 101000226 and Adrián López Catalina received a competitive **FPI-INIA** grant (FPI-SGIT2018-03).

*To all those whose actions allow my genes to be expressed in a favourable
environment.*

*They blended religion and art and science because, at base, science is no more
than an investigation of a miracle we can never explain, and art is an
interpretation of that miracle. – Ray Bradbury, The Martian Chronicles (1950).*

Acknowledgement

Dear reader, the language of science is English, but my feelings are best expressed in Spanish, so forgive me if this part is a rough translation. If you are here you are either looking for information because this thesis has randomly appeared on the internet or evaluating the work of many years. So please, allow me a few words free from judgement before we get to the serious part. You might think that the work you are about to read is mine, but that is not entirely true. I arrived at INIA without knowing how to open a file in R or how to navigate the CESGA terminal. Why don't they give these programs a nice interface...? My background was much more focused on wet lab work. So keep in mind that what you are about to read is the result of the support—both personal and professional—of many people.

I remember attending a talk for young researchers at the WCGALP in Rotterdam, where we were told that the most important thing to do in the early stages of a research career was to find a mentor. Having a PhD supervisor or someone overseeing your postdoc was not enough; you needed someone who would inspire you and push you to improve. I have no doubt that I found my mentor in Óscar, so my acknowledgements begin with him. Throughout my years here, I have felt constant support, I have been able to express my ideas and be heard, I have received guidance in my weaker areas, and I have been encouraged every step of the way. Thank you for the trust you have placed in me, for giving me the opportunity to share our results across the world, and above all, for pushing me to grow as a scientist. Now, if you mention the reference formatting one more time, just know that I will delete this paragraph.

I must also thank those who welcomed me on the other side of the world. The entire team at CSIRO and QAAFI went out of their way to ensure I kept learning and improving. Toni, Pamela, Juca, Yutao, Shuwen, Kate, Anna, Marina, and Jigme, thank you for making me feel so at home in a place so far from it.

It is important to surround yourself with people who remind you to pause and breathe from time to time. Over these nearly five years of my PhD, I have been lucky to cross paths with Mónica, María, Ramón, Alejandro, Anita, Adri, Paula, Eli, Silvia, Ester, Carlos, and Raúl. Thank you for the hours of laughter, the gossip, the advice, and for being part of this journey. I hope we continue to cross paths for many years to come—in many bars.

To Mum, Dad, Miguel, Andrea and Paola, for supporting me from the very beginning in venturing beyond the borders we knew, for sitting through talks and reading articles in English just to see what I am trying to start building. Thank you for allowing me to grow freely but always with your support.

Thank you once again, Eva. I've lost count of how many academic documents your name appears in. After the undergraduate dissertation, two master's theses, and this PhD, I don't know if I'll have another chance to dedicate a few lines to you—we may have run out of degrees to chase. How lucky I am to have a vet by my side who has taught me that cows exist in real life and are not just numbers and genes on a computer. Thank you for letting me share my worries and joys with you. Thank you for being my travel companion—both metaphorically and literally—and for packing your bags so we could live together for a few months 18,000 km from Spain. Thank you for helping me grow and for growing with me.

I'll stop now before this gets longer than the thesis itself. Thank you for reading my thanks, dear reader.

Agradecimientos

Querido lector, en las siguientes páginas estarás o bien buscando información sobre algo porque esta tesis te ha aparecido aleatoriamente en internet, o bien evaluando el trabajo de muchos años. Así que, por favor, deja que dedique unas palabras libres de juicio antes de que comience lo serio. Seguramente pienses que el trabajo que vas a leer es mío, pero no es así. Llegué al INIA sin saber abrir un archivo en R ni cómo enfrentarme a la terminal del CESGA. ¿Por qué no le ponen una interfaz bonita a estos programas...? Mi background era bastante más de wet lab. Así que ten en mente, que lo que vas a leer a continuación es fruto del apoyo (tanto personal como laboral) de muchas personas.

Recuerdo en una charla a jóvenes investigadores en el WCGALP de Róterdam donde nos decían que lo más importante que había que hacer en las primeras etapas de la carrera era encontrar un mentor. No era suficiente con tener un director de tesis o alguien que te supervise el posdoc, se necesitaba a alguien que te inspirara e impulsara a mejorar. Tengo claro que he encontrado a mi mentor en Óscar, así que mis agradecimientos empiezan por él. Durante mis años aquí he sentido un apoyo constante, he podido sentir respeto y escucha hacia mis ideas, refuerzo en mis puntos más débiles y un impulso contante. Gracias por la confianza depositada en mí, por ofrecerme la oportunidad de contar nuestros resultados en tantos puntos del mundo, y sobre todo por impulsarme a crecer como científico. Ahora que como me digas algo del formato de las referencias otra vez que sepas que borro este párrafo.

También tengo que agradecer a aquellas personas que me acogieron en la otra punta del mundo. A todo el equipo del CSIRO y QAAFI que estuvieron volcados en que siguiera aprendiendo y mejorando. Toni, Pamela, Juca, Yutao, Shuwen, Kate, Anna, Marina y Jigme, gracias por hacerme sentir tan cómodo en un lugar tan diferente y tan lejos de mi casa.

Es importante rodearse de personas que te recuerden parar y respirar de vez en cuando. Estos casi 5 años de tesis he tenido la suerte de cruzar caminos con Mónica, María, Ramón, Alejandro, Anita, Adri, Paula, Eli, Silvia, Ester, Carlos y Raúl. Gracias por las horas de risas, el marujeo, los consejos y por ser parte de este viaje. Espero que nos sigamos cruzando muchos años en muchos bares.

A mamá, papá, Miguel, Andrea y Paola por el apoyo desde el primer momento a salir a explorar qué hay fuera de las fronteras que conocíamos, por

tragarse charlas y artículos en inglés simplemente para ver lo que estoy intentando empezar a contruir. Gracias por permitirme y ayudarme a crecer con libertad pero siempre acompañado.

Gracias una vez más a Eva. Ya he perdido la cuenta de en cuantos documentos académicos aparece tu nombre. No sé si después del TFG, dos TFM y la Tesis tendré la oportunidad de volver a dedicarte unas líneas, creo que nos hemos quedado ya sin más títulos que perseguir. Qué suerte estar acompañado por una veterinaria que me haya enseñado que las vacas existen de verdad y no son sólo números y genes en un ordenador. Gracias por permitirme compartir mis agobios y alegrías contigo. Gracias por ser mi compañera de viaje, metafórica y literalmente, y hacer las maletas para irnos a vivir juntos unos meses a 18.000 km de España. Gracias por ayudarme a seguir evolucionando y a evolucionar conmigo.

Me callo ya, que va a ser más largo esto que la tesis. Gracias por leer mis gracias, lector.

Abstract

Maternal environmental conditions, such as metabolic stress induced by lactation, can influence offspring development through epigenetic mechanisms. This thesis investigates how maternal lactation status affects the DNA methylome of newborn Holstein calves, utilizing long-read sequencing technologies to enhance our understanding of intergenerational epigenetic inheritance in livestock. The study aims to (1) evaluate the reliability of Oxford Nanopore Technologies sequencing compared to bisulphite-based methods for livestock epigenomics, (2) identify differential methylation patterns between calves born to lactating and non-lactating cows, and (3) integrate epigenetic data into genomic selection models to improve breeding value and phenotypic predictions.

Blood and semen samples from Holstein calves and bulls were analyzed using Oxford Nanopore Technologies Sequencing and Reduced Representation Bisulphite Sequencing. Results revealed high concordance in methylation frequencies between methods. Oxford Nanopore Technologies sequencing provided broader genome coverage, while Reduced Representation Bisulphite Sequencing enriched promoter-associated guanine nucleotide sites islands. Calves from lactating cows exhibited 50,277 differentially methylated cytosines and 2,281 differentially methylated regions, many associated with immune response, metabolic pathways, and production traits such as milk yield. A co-methylation network analysis linked maternal lactation status to offspring methylation profiles and key phenotypic traits, including body capacity index and milk production efficiency. A recursive multi-omics model incorporating DNA methylation data improved the accuracy of genomic breeding value estimates by partitioning genetic and epigenetic variance.

This research contributes to livestock epigenetics by establishing robust methodologies for methylation profiling, identifying stress-induced epigenetic signatures, and demonstrating the potential of integrating epigenomics into genetic evaluations. Additionally, the development of a targeted methylation array comprising 48,351 guanine nucleotide sites (CpG) facilitates scalable screening for epigenetic biomarkers. These findings underscore the role of maternal environmental influences in shaping offspring phenotypes and offer novel strategies for enhancing dairy cattle resilience, productivity, and sustainability in response to climate and production challenges.

Key Findings:

- Oxford Nanopore Technologies sequencing complements Reduced Representation Bisulphite Sequencing for genome-wide methylation profiling.
- Maternal lactation induces epigenetic changes in offspring, affecting immune and metabolic pathways.
- Incorporating DNA methylation data into genomic selection models enhances breeding value predictions.

Implications:

This study pioneers the integration of epigenetic variation into livestock breeding, contributing to more sustainable and productive dairy cattle management.

Resumen

Las condiciones ambientales maternas, como el estrés metabólico asociado a la lactancia, pueden influir en el desarrollo de su descendencia a través de mecanismos epigenéticos. Esta tesis investiga cómo el estado de lactancia materna afecta el metiloma del ADN de terneras Holstein recién nacidas utilizando tecnologías de secuenciación de lecturas largas para mejorar la comprensión de la herencia epigenética intergeneracional en el ganado. Los objetivos del estudio fueron: (1) evaluar la fiabilidad de la secuenciación de Oxford Nanopore Technologies en comparación con métodos basados en bisulfito, (2) identificar patrones de metilación diferencial entre terneros nacidos de vacas lactantes y no lactantes, e (3) integrar datos epigenéticos en modelos de selección genómica para mejorar la predicción del valor genético.

Se analizaron muestras de sangre de terneras y semen de toros Holstein secuenciados mediante secuenciación por nanoporos y secuenciación por bisulfito. Se observó una alta correlación en las frecuencias de metilación entre ambos métodos. La secuenciación por nanoporos proporcionó una mayor cobertura del genoma, mientras que la secuenciación por bisulfito enriqueció islas CpG asociadas a regiones promotoras. Las terneras nacidas de vacas lactantes presentaron 50.277 citosinas diferencialmente metiladas y 2.281 regiones diferencialmente metiladas, muchas de ellas relacionadas con la respuesta inmune, rutas metabólicas y diferentes caracteres productivos, como la producción de leche. Mediante un análisis de redes de co-metilación se vinculó el estado de lactancia materna con los perfiles de metilación de su descendencia y con caracteres de importancia, como el índice de capacidad corporal y la eficiencia en la producción de leche. Finalmente, empleando un modelo multi-ómico recursivo se incorporaron datos de metilación del ADN, mejorando la precisión de la estimación del valor genético al separar la varianza aditiva de la varianza epigenética.

Esta investigación contribuye al campo de la epigenética animal al establecer metodologías robustas para la identificación y caractereización de patrones de metilación en el genoma, identificar patrones epigenéticos inducidos por el estrés y demostrar el potencial de integrar la epigenómica en las evaluaciones genéticas. Además, el desarrollo de un chip de metilación dirigido a 48.351 sitios CpG permite una rápida detección de biomarcadores epigenéticos. Estos resultados subrayan el papel de los factores ambientales maternos en la configuración de los fenotipos de la descendencia y ofrecen estrategias innovadoras

para mejorar la resiliencia, productividad y sostenibilidad del ganado lechero ante los desafíos climáticos y productivos.

Hallazgos clave:

- La secuenciación de Oxford Nanopore Technologies es complementaria a la secuenciación RRBS para describir la metilación del genoma completo.
- La lactancia materna induce cambios epigenéticos en la descendencia, afectando rutas inmunológicas y metabólicas.
- La incorporación de datos de metilación del ADN en modelos de selección genómica mejora la predicción del valor genético y el fenotipo de los animales.

Implicaciones:

Este estudio abre nuevas vías para la integración de la variación epigenética en la mejora genética del ganado, contribuyendo a una gestión más sostenible y productiva de la ganadería lechera.

Table of Contents

<i>Acknowledgement</i>	v
<i>Agradecimientos</i>	vii
<i>Abstract</i>	ix
<i>Resumen</i>	xi
Table of Contents	xiii
<i>List of Figures</i>	xvii
List of tables	xxi
<i>Abbreviations and Acronyms</i>	xxiii
1. General Introduction	27
1.1. Brief Introduction to Epigenetics.....	29
1.2. Epigenetic Inheritance and Reprogramming	31
1.3. Different Epigenetic Modifications	34
1.3.1. DNA Methylation.....	34
1.3.2. Noncoding RNA.....	35
1.3.3. Chromatin Structure and Histone Modification	36
1.3.4. Prions	37
1.4. Detection of Methylation Marks.....	38
1.4.1. Sodium Bisulphite-Based Methods	38
1.4.2. Enzymatic-Modification Based Methods	39
1.4.3. Biological Affinity-Based Methods.....	40
1.4.4. Nanopore Sequencing	40
1.4.5. Methylation Arrays.....	40
1.5. Epigenetics in Livestock	41
1.5.1. Epigenetic Marks as Biomarkers	42
1.5.2. Methylation Effects over Phenotypes of Interest	43
1.5.3. Impact of Nutrition on the Dairy Cattle Methylome	44
1.5.4. Addition of Epigenetic Information to Genomic Data	45
1.6. Genomic Selection and Integration of Epigenetic Information	48
1.7. The Current State of Spanish Dairy Cattle Breeding Programmes	52
1.8. Working Hypothesis and Objectives	55
1.8.1. Metabolic Stress During Dairy Cattle Lactation.....	55
1.8.2. The Particularity of the Holstein Cattle.....	56
1.8.3. Objectives	59
2. Materials and Methods	61
2.1. Materials.....	63
2.2. Methods.....	64

2.2.1.	DNA Extraction	64
2.2.2.	Sequencing	64
2.2.3.	Bioinformatics	65
2.2.4.	Statistical Methods	66
2.2.5.	Simulation of the Methylation Information	69
3.	Chapter 1 - Oxford Nanopore Sequencing as an Alternative to Reduced Representation Bisulphite Sequencing for the Identification of CpGs of Interest in Livestock Populations	73
3.1.	Abstract.....	75
3.2.	Introduction	75
3.3.	Material and methods.....	78
3.3.1.	Sample Collection, DNA Extraction and Sequencing	78
3.3.2.	Methylation Calling	79
3.3.3.	Differentially Methylated Sites	79
3.3.4.	Comparison of Called CpGs.....	80
3.4.	Results	80
3.4.1.	Correlation Between ONT and RRBS.....	80
3.4.2.	Genomic Position Distribution of the CpGs Covered by ONT and RRBS	81
3.4.3.	Differential Methylation Analyses	82
3.4.4.	QTL Associated with the DMCs	83
3.5.	Discussion	84
3.6.	Conclusion.....	87
3.7.	Funding.....	87
3.8.	Ethics Approval	87
3.9.	Acknowledgements.....	88
3.10.	Tables	89
3.11.	Figures	91
3.12.	References	100
4.	Chapter 2 - Stress-Induced Epigenetic Effects Driven by Maternal Lactation in Dairy Cattle: A Comethylation Network Approach	105
4.1.	Abstract.....	107
4.2.	Background.....	107
4.3.	Methods.....	110
4.3.1.	DNA Extraction	110
4.3.2.	DNA Sequencing	110
4.3.3.	Methylation Calling	111
4.3.4.	Differential Methylation Analyses	111
4.3.5.	Permutation of the Groups	112
4.3.6.	Gene Annotation and Description of Genomic Regions	112

4.3.7.	Enrichment Analyses.....	112
4.3.8.	Construction of the Comethylation Network.....	113
4.3.9.	Visualisation of the Network.....	114
4.4.	Results	114
4.4.1.	Sequencing Results.....	114
4.4.2.	Differential Methylation Analyses.....	115
4.4.3.	Comethylation Network	116
4.5.	Discussion	117
4.6.	Conclusions.....	124
4.7.	Acknowledgements.....	124
4.8.	Funding.....	124
4.9.	Availability of Data and Materials	125
4.10.	Ethics Approval	125
4.11.	Tables	126
4.12.	Supplementary Tables	126
4.13.	Figures	128
4.14.	Supplementary figures.....	133
4.15.	References	136
5.	Chapter 3 - A Recursive Model Approach to Include Epigenetic Effects in Genetic Evaluations Using Simulated DNA Methylation Effects	143
5.1.	Abstract.....	144
5.2.	Introduction	145
5.3.	Materials and Methods	147
5.3.1.	Data	147
5.3.2.	Methylation Pattern Analysis	148
5.3.3.	Simulation of Breeding Values and Traits	148
5.3.4.	Benchmark Model.....	150
5.3.5.	Multiomic Approach Using GOBLUP Model.....	151
5.3.6.	Model Assumptions and Remarks.....	152
5.4.	Results	153
5.4.1.	Overview of the Simulation.....	153
5.4.2.	Traditional Animal Model	154
5.4.3.	Epigenetic GOBLUP Model.....	154
5.4.4.	Comparison Between EBV From the Traditional and GOBLUP Models.....	155
5.4.5.	Implications and final remarks	157
5.5.	Conclusion.....	158
5.6.	Tables.....	160
5.7.	Figures	162

5.8. References.....	164
6. General Discussion	169
6.1. Introduction to the General Discussion	171
6.2. Identification of Epigenetic Marks in Livestock Populations	173
6.3. Maternal Lactation Influences on the Methylation Profile of Offspring	176
6.4. Integrating Epigenetic Data Into Breeding Programmes: Application of the GOBLUP Model	181
6.5. Limitations of This PhD Project.....	184
6.6. Final Thoughts and Future Research Lines.....	187
7. General Conclusions	191
8. References.....	195

List of Figures

Figure 1-1. Representation of epigenetic inheritance regarding the sex affected by the stressor. Source (Fitz-James and Cavalli, 2022)	32
Figure 1-2. Mechanisms of DNA methylation and demethylation. Open circles denote unmethylated CpGs and filled circles indicate methylated positions. Source (Ying and Chen, 2023).	35
Figure 1-3. Schematic models of mediation (A), moderation (B), and conditional process (C). Adapted from Igartua and Hayes (2021).	47
Figure 3-1. 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.	91
Figure 3-2. 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.	92
Figure 3-3. 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 ($=0$).	93
Figure 3-4. Barplots representing the distance to the TSS of the detected positions at 4x moving upstream to its left and downstream to its right.....	94
Figure 3-5. Barplots representing the features where the CpGs are detected at 4x, considering promoter regions those closer than 3kb to the TSS.....	94
Figure 3-6. 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).....	95
Figure 3-7. Density plots showing the position of the DMCs along the chromosomes detected by the different technologies and coverage filters.	95

Figure 3-8. 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.	96
Figure 3-9. Distribution of the genetic features associated with the DMCs detected by RRBS, and ONT at a minimum coverage of 4x and 7x.	97
Figure 3-10. Distance of the DMCs detected by each technique to the closest TSS.	97
Figure 3-11. 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.	98
Figure 3-12. 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.	99
Figure 3-13. 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.	99
Figure 4-1. Correlations between the methylation profiles of the six samples analysed. The upper triangle represents the Pearson correlation coefficient between the methylation frequencies of the samples by pair. The histogram on the diagonal illustrates the methylation frequency distribution of the marks throughout the genome. The lower triangle depicts a set of scatterplots showing the pairwise percentages of methylation in all samples, where the red line in the plot is from the linear regression fit and the green line is from the polynomial regression fit.	128
Figure 4-2. Genomic regions were split into 60 windows for each gene annotated in the cow genome to compare the methylation in different genic coordinates. The methylation profiles of the marks detected in each sample were -4,000 bp from the TSS and +4,000 bp from the TTS. The lines and dots represent the mean methylation, and the width of the bands represents the maximum and minimum values of the confidence intervals.	128
Figure 4-3. Distribution of the difference in methylation frequency for the DMCs and DMRs between the two study groups. The red vertical lines represent the	

delta value = 0.2 (minimum difference between the estimated means of each group).....	129
Figure 4-4. Venn diagram depicting the overlap of the genes associated with each DMC/DMR.	129
Figure 4-5. Distribution of the genetic features associated with each DMC/DMR depending on their status. Image a) depicts the distance of the mark to the transcription start site. b) represents the genetic feature with which the mark is associated.....	130
Figure 4-6. Enrichment analysis performed on the genes associated with the DMR and DMC regions using clusterProfiler and calves gestated by nonlactating cows as a reference group. The size of the dots represents the “GeneRatio”, defined as k/n , where k is the number of genes in our dataset that are annotated to the GO term and n is the total number of unique genes in the dataset.	131
Figure 4-7. Comethylation network between the regions and first neighbours of a) lactation status and b) body capacity index (ICAP) c) milk EBV of the calf. Yellow nodes represent genes associated with hypomethylated DMRs, and purple nodes represent those genes associated with hypermethylated DMRs. The white nodes are regions not detected as DMRs by DSS. The blue lines represent positive correlations, while the red lines indicate negative correlations. Larger nodes were regions that were annotated against promoter regions (-3,000 to 3,000 bp from the TSS).....	132
Supplementary Figure 4-8. Distribution of the difference in methylation frequency for the DMCs between the two groups to study by chromosome. Red vertical lines represent the delta value, 0.2.....	133
Supplementary Figure 4-9. Distribution of the difference in methylation frequency for the DMRs between the two groups to study by chromosome. Red vertical lines represent the delta value, 0.2.....	133
Supplementary Figure 4-10. Venn diagram depicting the overlap of the genes associated with each DMC/DMR for each iteration.....	134
Supplementary Figure 4-11. Comethylation network between the regions and the phenotypes of interest. Yellow nodes represent genes associated with hypomethylated DMRs, and purple nodes represent those genes associated with hypermethylated DMRs. The white nodes are regions not detected as DMRs by DSS. The blue lines represent positive correlations, while the red lines indicate	

negative correlations. Larger nodes were regions that were annotated against promoter regions (-3,000 to 3,000 bp from the TSS)..... 135

Figure 5-1. Average weighted methylation frequency of the samples sequenced by nanopore sequencing in genomic regions. Vertical lines limit the extension of the new windows where the sixty original windows were aggregated into new windows where the variability between an original window and its immediately following window was greater than 10%, resulting in thirteen windows whose mean and standard deviation were again estimated. 162

Figure 5-2. Scheme of how recursive effects of methylation over the phenotype are proposed in the generator (a) and statistical models (b). These models can be reparametrised to prove recursiveness (c), where b, is impacted by c' creating an effect of methylation marks on the phenotype combined with the animal's direct genetic effect. 162

Figure 5-3. Values of alpha across the genetic windows estimated by the epigenetic model. Each colour represents the alpha value in a different scenario of heritability of liability to methylation simulated. 163

Figure 5-4. Genetic correlation estimated for each heritability scenario. The y axis represents the pairwise correlation between two features and the x-axis the Pearson correlation coefficient. 163

Figure 6-1. Co-methylation network associating phenotypes of interest in animal production with different genomic regions. Source: López-Catalina et al. (2024). 172

Figure 6-2. Different combinations of coverage and their respective correlations (r) and sites detected (N). Correlation filtered for 1x, 5x, 10x, 15x, 20x and last image illustrates filtering ONT sites by 15x and Illumina for 1x..... 173

Figure 6-3. Epigenetic genome-wide association study performed using the RUMIGEN's epichip to associate maternal lactation status with the offspring's methylation status. The horizontal red line represents the FDR-adjusted p-value and the blue the Bonferroni-adjusted p-value. 180

List of tables

Table 3-1. 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.	89
Table 3-2. 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.....	90
Table 4-1. Summary statistics of sequencing performance. On the upper side of the table, calves born from nonlactating cows; on the bottom, calves born from lactating cows.	126
Supplementary Table 4-2. Summary statistics of sequencing performance. On the upper side of the table, calves born from nonlactating cows; on the bottom, calves born from lactating cows.....	126
Supplementary Table 4-3. Biological processes and their adjusted p-values associated with the genes used as input for the enrichment analysis.....	127
Table 5-1. Summary statistics of the methylation frequency for each of the thirteen genetic windows and the number of simulated SNPs with additive effect for the epigenetics-modulated trait and the liability to methylation.	160
Table 5-2. Estimated parameters by BLUP and GOBLUP for each heritability scenario. The variances and h^2 estimates and their HPD Intervals (95%) across scenarios for the epigenetics-modulated traits are presented in the third, fourth and fifth columns as well as the estimated average heritability and their	

standard deviations across the thirteen windows in the GOBLUP model in the second column.	161
---	-----

Abbreviations and Acronyms

5mC	5-methylcytosine
BLUP	Best Linear Unbiased Prediction
BS	Bisulphite Sequencing
CGI	CpG Island
ChIP	Chromatin Immunoprecipitation
CONAFE	Confederation of Spanish Holstein Associations
DMC	Differentially Methylated Cytosine
DMG	Differentially Methylated Gene
DMR	Differentially Methylated Region
DNMT	DNA methyltransferase
EBV	Estimated Breeding Value
EM-Seq	Enzymatic Methyl-Seq
EWAS	Epigenetic Genome-Wide Association Study
FDR	False Discovery Rate
GEBV	Genomic Estimated Breeding Value
GOBLUP	Genomic Omic Best Linear Unbiased Prediction
ICAP	Body Capacity Index
ICR	Imprinting Control Region
IGF-I	Insulin-like Growth Factor 1
IGF2	Insulin-like Growth Factor 2
IM€T	Economic Merit Index
miRNA	MicroRNA
NEB	Negative Energy Balance
ncRNA	Noncoding RNA
ONT	Oxford Nanopore Technologies
PCIT	Partial Correlation With Information Theory

PGC	Primordial Germ Cell
piRNA	Piwi-Interacting RNA
RRBS	Reduced Representation Bisulphite Sequencing
SCFA	Short-Chain Fatty Acids
siRNA	Small Interference RNA
SNP	Single Nucleotide Polymorphism
SSGBLUP	Single-Step Genomic BLUP
TEI	Transgenerational Epigenetic Inheritance
TF	Transcription Factor
TSS	Transcription Start Site
TTS	Transcription Termination Site
VWP	Voluntary Waiting Period
WGBS	Whole Genome Bisulphite Sequencing



The experimental work that led to results of this thesis was carried out in the “Departamento de Mejora Genética Animal” at the National Institute for Agricultural and Food Research and Technology (INIA), and was funded through the following research projects:

1. **METALGEN**. Funded by the Ministry of Economy and Competitiveness with reference number RTA2015-00022-C03-02.
2. **RUMIGEN**: Towards improvement of ruminant breeding through genomic and epigenomic approaches. Funded by the Horizon 2020 programme of the European Union with reference 101000226.

Predoctoral contract with a competitive **FPI-INIA** grant (FPI-SGIT2018-03).

1. General Introduction

1.1. Brief Introduction to Epigenetics

Over the past seven decades, genetic research has primarily focused on the heritability of DNA, the molecule responsible for conveying the information that shapes phenotypes, inherited from both paternal and maternal lineages (Daxinger and Whitelaw, 2012). DNA encodes the genetic instructions essential for the development, functioning, and reproduction of organisms (Bello and Gbolagade, 2017). These instructions are passed from parent to offspring, determining their phenotypic traits (Schuster, 2010). While heredity is mainly governed by DNA sequences, recent research has highlighted the role of transgenerational and intergenerational epigenetic inheritance, which allows the transmission of information without altering the underlying DNA sequence (Rothi and Greer, 2023). Thus, the perpetuation of life relies on two interdependent reservoirs of information: the linear DNA sequence, which is replicated during cell divisions, and the three-dimensional arrangement of molecules, recreated at the start of each generation through the mediation of epigenetic marks (Jose, 2018).

In 1942, Waddington introduced the term “epigenetics” to describe the interactions between an individual’s genome and their environment (Waddington, 1942). He emphasised the need for genomic flexibility to enable rapid adaptation to environmental changes and introduced the epigenetic landscape model, which illustrates how gene regulation influences development to ensure precise cell lineage determination. The prefix “epi-”, derived from Ancient Greek (ἐπι-), means “over”, “outside”, or “around” the root term. In this context, “epigenetics” refers to mechanisms that supplement traditional genetic inheritance without altering the DNA sequence itself.

While DNA sequences are widely recognised as the primary carriers of genetic information, recent studies have highlighted the significant role of inherited epigenetic information in shaping complex phenotypes across various species. Mechanisms of epigenetic inheritance regulate the transmission of environmental information across generations without inducing permanent changes in the genetic code (Norouzitallab et al., 2019). This emerging field not only provides insights into how organisms adapt to changing environments but also offers valuable perspectives on understanding pathologies influenced by parental environmental factors (Heijmans et al., 2009). The intricate interplay between DNA sequences and epigenetic marks presents a more holistic view of heritability, highlighting the

combined importance of genetic and epigenetic factors in shaping biological traits and responses to environmental stimuli (Burton and Greer, 2022).

Non-genetic inheritance mechanisms, including epigenetic alterations, significantly contribute to the phenotypic diversity observed in livestock populations alongside genetic inheritance (David et al., 2019). Exploring the interaction between genetics and epigenetics in livestock is vital for comprehending phenotypic diversity and adaptive mechanisms. Epigenetic modifications are pivotal in influencing key traits such as growth, development, and health outcomes in livestock (Ibeagha-Awemu and Khatib, 2023; Safdar and Ozaslan, 2023; Wang et al., 2023a). Research on ovine and caprine species across various climatic gradients has identified genetic distinctions related to environmental responsiveness, immune function, reproductive capabilities, and productivity, alongside genes with differential methylation patterns associated with lactation and muscle hypertrophy (Denoyelle et al., 2021). Developing a comprehensive pan-epigenome for bovine and ovine species could significantly enhance our understanding of the interactions between genomic and epigenomic factors in shaping biological phenotypes, thereby facilitating more precise trait prediction and optimising breeding methodologies (McKay et al., 2023). The complex interrelationship between genetic and epigenetic factors emphasizes the necessity of integrating both aspects in the management and breeding practices of livestock.

However, these findings should not be misconstrued as a neo-Lamarckian view of environmental epigenetics. While epigenetic alterations can be preserved from mother cells to daughter cells, they are not typically transmitted between individuals. Therefore, these modifications are better described as *transmittable* rather than *heritable* (Tian and Marjani, 2012). Understanding these foundational concepts of epigenetics is crucial for exploring its applications in livestock breeding. The subsequent sections delve into how these mechanisms function and their practical implications for improving dairy cattle production.

1.2. Epigenetic Inheritance and Reprogramming

Epigenetic inheritance bridges the gap between genetic coding and environmental influences. This section explores how these mechanisms transmit information across generations, setting the stage for targeted applications in livestock breeding.

Epigenetic inheritance must be considered separately from traditional notions of Mendelian inheritance. Within the conventional framework, inheritance refers to the systematic transmission of traits from progenitors to their descendants, preserved across successive generations. Concepts such as dominant and recessive alleles, codominance, and Mendel's laws of dominance, segregation, and independent assortment are constructs that do not apply to discussions of epigenetic inheritance (Martos et al., 2015).

Transgenerational epigenetic inheritance (TEI) is tightly linked to sex and generational lineage. By strict definition, TEI refers to the “germ line-mediated inheritance of epigenetic information between generations in the absence of direct environmental influences, leading to phenotypic variation” (Skinner, 2011). This definition excludes the transmission of epigenetic marks when direct environmental influences are present—a phenomenon known as intergenerational or parental epigenetic inheritance (Perez and Lehner, 2019).

In the context of intergenerational epigenetic inheritance, an organism passes an epigenetic modification, induced by internal or external stimuli, to its immediate offspring. This process differs by sex. In males, the stimulus affects both the individual and his germline, restricting the transmission of the epigenetic modification to the F1 generation. In contrast, a gestating female subjected to the same stimulus may influence up to three generations: herself (F0), the developing embryo (F1), and the germline of that embryo (F2). As a result, the stimulus can affect the mother, her child, and her grandchild. Under these conditions, TEI occurs only when the epigenetic modification persists into the F3 generation in females or the F2 generation in males.

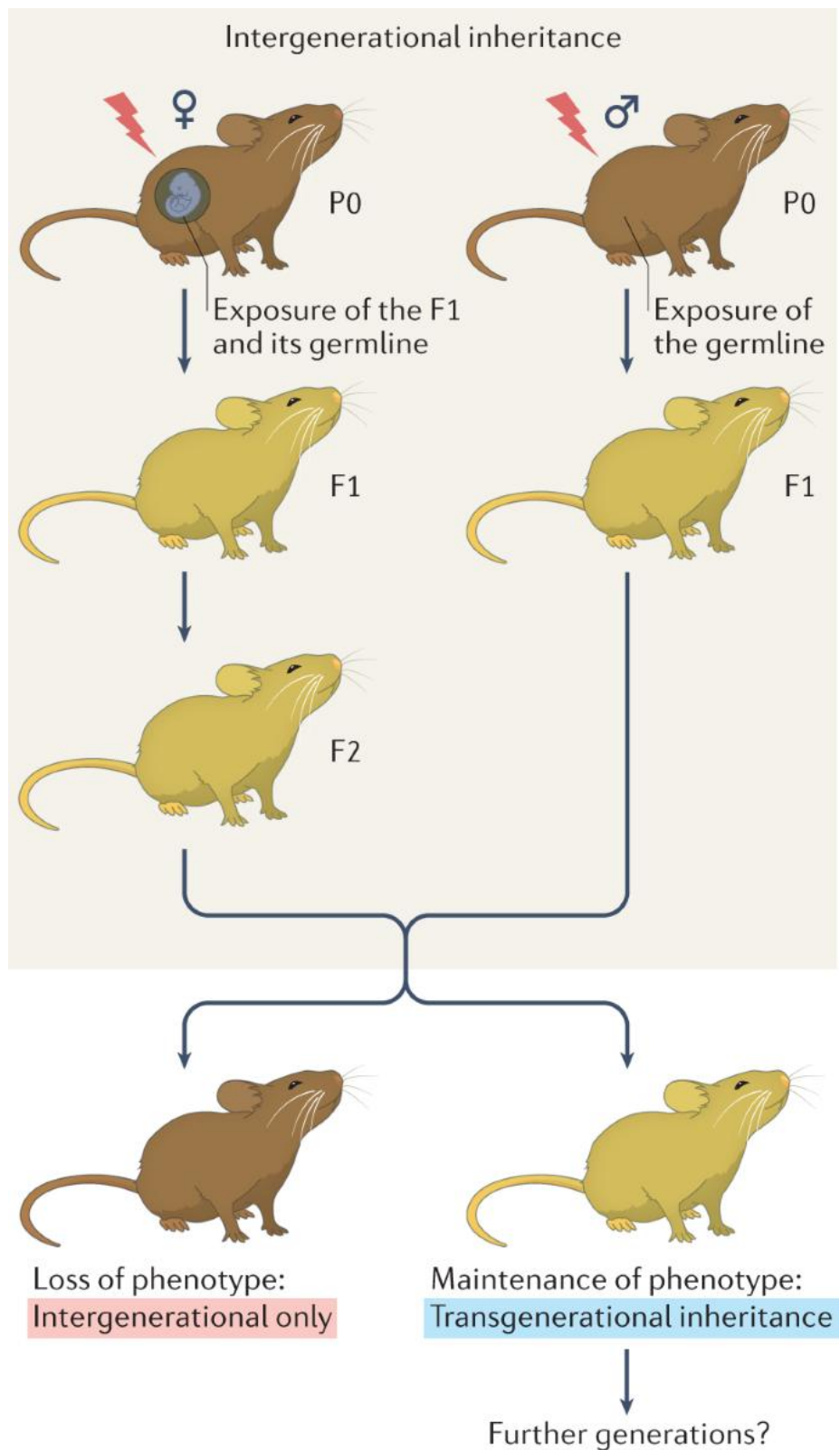


Figure 1-1. Representation of epigenetic inheritance regarding the sex affected by the stressor. Source (Fitz-James and Cavalli, 2022)

A key biological limitation to transgenerational epigenetic inheritance is epigenetic reprogramming. During development, animal genomes undergo two global waves of methylation reprogramming. The first wave occurs in primordial germ cells (PGCs), erasing epigenetic marks from imprinting control regions (ICRs), the X chromosome, and germline genes (Radford, 2018; Ying and Chen, 2023; Zeng and Chen, 2019). Whole-genome bisulphite sequencing has shown that global demethylation in PGCs happens during their expansion and migration, while specific sequences, such as imprints, are demethylated upon entry into the gonads (Seisenberger et al., 2012). The second wave of demethylation takes place during the pre-implantation stage of embryonic development, removing epigenetic marks inherited from maternal and paternal gametes to establish cellular totipotency. Certain regions, such as ICRs, are exceptions to this reprogramming. These events are crucial for establishing pluripotency and enabling proper germline development.

Despite these reprogramming processes, imprinted regions and transposable elements can evade demethylation, allowing some epigenetic modifications to persist across generations. Additionally, specific germ cell epigenetic modifications can survive embryonic development, facilitating the transmission of parental epigenetic marks to the next generation (Khatib et al., 2024; Slyvka et al., 2015). However, due to the extensive erasure of epigenetic marks, many studies have concluded that phenomena initially described as TEI are more accurately classified as intergenerational inheritance (Heard and Martienssen, 2014; Martos et al., 2015; Santilli and Boskovic, 2023). To definitively identify TEI effects, the same epimutation must be observed consistently across multiple generations, ruling out confounding factors that could influence offspring phenotypes.

Three major examples have described how maternal environmental factors can alter the offspring phenotypes: the endocrine disruptors in rats (Anway et al., 2005), the agouti locus in mice (Blewitt et al., 2006) and the Dutch famine in humans (Heijmans et al., 2008).

While epigenetic inheritance is crucial, its detection and characterisation rely heavily on sophisticated methodologies. The next section focuses on the various techniques used to uncover methylation marks and their implications for research.

1.3. Different Epigenetic Modifications

Epigenetic modifications can influence gene expression in different ways. While the primary focus of this thesis is on the impact of DNA methylation within the cattle genome, this section will also explore other epigenetic modifications to provide a broader context.

1.3.1. DNA Methylation

DNA methylation is a covalent modification wherein a methyl group is added to the fifth carbon of a cytosine, transforming it into a 5-methylcytosine (5mC). The regulation of methylation patterns is governed by the opposing actions of methylation and demethylation machineries (Ying and Chen, 2023). Methylation is catalysed by two family of DNA methyltransferases enzymes (DNMTs), DNMT1 which maintains methylation patterns during the DNA replication and DNMT3A and DNMT3B responsible for de novo methylation establishing methylation patterns (Chen and Zhang, 2020). DNMTs transfer a methyl group from the donor molecule S-adenosyl-L-methionine to the C-5 of the cytosine. This modification primarily occurs at CpG dinucleotides in CpG-rich regions known as CpG islands (CGI). CGI are interspersed DNA sequenced usually unmethylated that are associated with transcription initiation sites (Deaton and Bird, 2011). In mammals, methylation occurs within 60 to 80% of the CpG dinucleotides (Patil et al., 2014), characterised by cytosine bases adjacent to guanine bases and linked by a phosphodiester bond. In embryonic cells, brain tissue, and oocytes, non-CpG methylation is also present, manifesting in CpA, CpT, or CpC contexts (He and Ecker, 2015).

Methylation at CpG sites is associated with long-term gene silencing mechanisms, including the somatic silencing of germline genes, suppression of repetitive DNA elements, X-chromosome inactivation, and genomic imprinting in mammals (Angeloni et al., 2024; Zemach et al., 2010). In mammalian zygotes, demethylation occurs through both active and passive mechanisms. The paternal and maternal genomes undergo extensive demethylation prior to the first mitotic division (Guo et al., 2014). Active demethylation of the paternal genome occurs rapidly within 4 hours post-fertilisation, relying on Tet3-mediated oxidation, although the precise downstream mechanisms remain unclear (Santos et al., 2002). Passive demethylation occurs during the subsequent cleavage stages. This demethylation process is crucial for epigenetic reprogramming, differentiation,

and development. Additionally, de novo methylation happens specifically in the inner cell mass of blastocysts, potentially influencing early lineage specification (Abdalla et al., 2009).

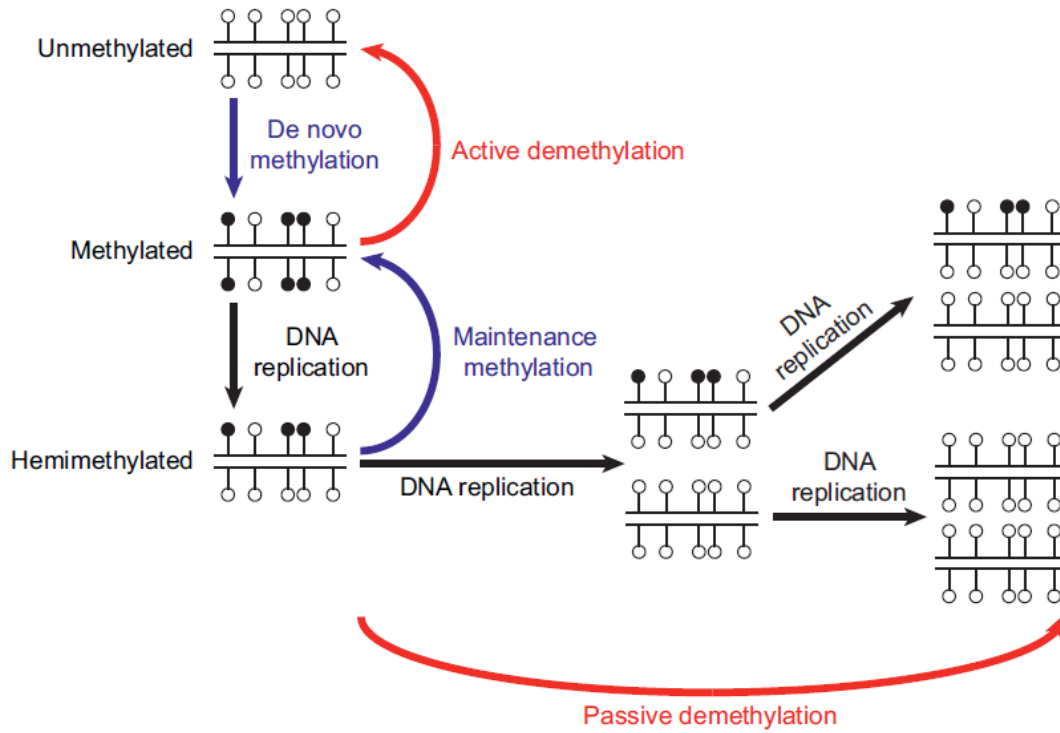


Figure 1-2. Mechanisms of DNA methylation and demethylation. Open circles denote unmethylated CpGs and filled circles indicate methylated positions. Source (Ying and Chen, 2023).

1.3.2. Noncoding RNA

The RNA molecules are fundamental to ensure the flow of genetic information, facilitating the processes of transcription and translation. These processes produce three primary classes of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). The mRNA molecules carry genetic instructions from DNA to synthesise proteins (protein-coding RNA), while rRNA and tRNA play supportive roles in this translation process.

Although approximately 90% of the eukaryotic genome is transcribed, only 1–2% consists of protein-coding RNA. The remaining portion comprises noncoding RNA (ncRNA), which is categorised into two groups based on size: small noncoding RNA (sncRNA), less than 200 nucleotides, and long noncoding RNA (lncRNA), greater than 200 nucleotides. sncRNA includes microRNA (miRNA), short-interfering RNA (siRNA), piwi-interacting RNA (piRNA), and small nucleolar RNA (snoRNA). These molecules play critical roles in mRNA degradation, transposon

silencing, germline development, and rRNA modification. In contrast, lncRNAs function as precursors to sncRNA and are involved in chromatin reprogramming and the regulation of mRNA levels.

Both small and long non-coding RNAs are pivotal for several biological processes, including genomic imprinting, dosage compensation, the cellular response to stress, and developmental patterning (Brosnan and Voinnet, 2009). These molecules frequently collaborate with regulatory proteins to establish intricate regulatory networks. The dysregulation of ncRNA biogenesis and expression has been associated with a variety of human diseases, notably cancer (Gomes et al., 2013). In livestock, ncRNA—specifically microRNAs and lncRNAs—has been recognised as a crucial factor influencing feed efficiency traits, such as residual feed intake and feed conversion ratios (Hu et al., 2022). In particular, lncRNAs, have been identified as playing significant roles in numerous biological processes, including developmental, metabolic, and immunological regulation (Weikard et al., 2017). In the context of animal products, lncRNAs are implicated in the deposition of adipose tissue, which subsequently influences meat quality and lean percentageessential economic traits within livestock production (Wang et al., 2021a).

1.3.3. Chromatin Structure and Histone Modification

Genomes consist of billions of nucleotides that must be compacted into a nucleus with a diameter of micrometres. The chromatin plays a crucial role in the condensation of genetic material, existing in two distinct forms. Euchromatin, characterised by an open configuration that permits access to transcriptional machinery, thereby facilitating gene expression. In contrast, heterochromatin represents a state in which the genetic sequences are extensively condensed, resulting in the suppression of transcriptional activity (Wallrath et al., 2014).

Condensation is achieved through the arrangement of histones, where an octamer of histones serves as the core of nucleosomes. Around this octamer, 147 base pairs of DNA are wrapped, forming two turns. The octamer is made of two subunits of core histones H2A, H2B, H3 and H4 and an additional H1 linker histone (Luger et al., 1997). Different histones residues can be modified affecting their condensation capability.

Histone modifications play a crucial role in defining chromatin structure and function, significantly impacting gene expression and nuclear organisation (Gelato

and Fischle, 2008). These post-translational changes, which include acetylation, methylation, phosphorylation, and ubiquitination, occur on histone tails and help establish different chromatin domains, specifically euchromatin and heterochromatin (Bártová et al., 2008). These modifications serve two main purposes: they influence either the interactions between DNA and nucleosomes or the interactions among nucleosomes themselves, by changing the charge of histones or adding physical groups at various residues. The secondary function, known as the docking mechanism, enhances the recruitment or repulsion of regulatory proteins through such modifications.

The pattern of histone modifications across the genome is non-random and closely associated with specific nuclear compartments. In RNA polymerase II-mediated transcription, certain histone modifications are linked to coding regions, and their misregulation may lead to abnormal intragenic transcription. Histone modifications are vital for maintaining the balance between cellular growth and differentiation, thus forming both active and repressive chromatin states, as well as poised chromatin regions that can rapidly respond to developmental and environmental signals (Albini et al., 2019).

1.3.4. Prions

Prions are proteins capable of adopting alternative, self-templating structures that can be transmitted across generations, constituting a distinctive form of epigenetic inheritance (Dennis and Garcia, 2022; Halfmann and Lindquist, 2010). Initially recognised for their association with rare neurodegenerative diseases in mammals, prions are now understood to be widespread in nature and are often linked to beneficial phenotypes, particularly in yeast. Unlike genetic mutations, prions arise at a higher rate, enabling cells to rapidly develop new phenotypes that may confer an adaptive advantage in response to swift environmental changes (Garcia and Jarosz, 2014).

These self-perpetuating protein states influence gene expression and cellular signalling, giving rise to diverse heritable traits without alterations in the underlying nucleic acids. In fungi, prions function as epigenetic determinants, modulating processes such as metabolism and gene regulation (Tuite and Serio, 2010). The parallels between mammalian and fungal prions suggest that these proteins are not merely biological anomalies but may represent a fundamental regulatory mechanism in cellular physiology.

In mammals, certain proteins exhibiting prion-like properties but lacking infectivity, referred to as prionoids, can acquire new functions upon conversion to a self-replicating conformation. This suggests that such a process may also regulate protein function in metazoans. Examples of functional prion-like proteins in the brain include the cytoplasmic polyadenylation element-binding protein (CPEB) and TIA-1 (Rayman and Kandel, 2017). Prions demonstrate the ability to link environmental stress to the emergence of novel traits, emphasising their potential role in evolution and adaptation. This dual capacity to provide stability while driving phenotypic innovation underscores their significance as a unique component of the epigenetic landscape.

1.4. Detection of Methylation Marks

DNA methylation marks can be explored by a wide range of techniques depending on the scope of the study. In this section, we focused on genome-wide DNA methylation profiling that can be subclassified in four mayor methods: sodium bisulphite-based methods, enzymatic modification-based methods, biological affinity-based methods and next generation methods.

1.4.1. Sodium Bisulphite-Based Methods

Sodium-bisulphite-based methods rely on the different sensitivities towards chemical modification of cytosine and 5-methylcytosine. The addition of sodium bisulphite leads to the deamination of the cytosine that becomes a uracil, 5mC are resistant to the deamination and the molecule remains the same after the treatment (Hayatsu, 2008; Hayatsu et al., 1970). After the bisulphite conversion, two PCR cycles are followed, primers are designed to be bisulphite-specific and flank methylation sites of interest. The product of the PCR would be an amplification of both methylated and unmethylated sequences. Unmethylated cytosines, transformed by the sodium bisulphite would appear as thymines in the sense strand and as adenines in the antisense strand. Methylated cytosines, protected by the 5-methyl group, would appear as cytosines after the PCR (Cokus et al., 2008; Frommer et al., 1992).

Two mayor techniques rely on this methodology, whole genome bisulphite sequencing (WGBS) and reduced representation bisulphite sequencing (RRBS). WGBS implements untargeted genome-wide sequencing of genomes, and has measured methylation of 95% of the CpGs in the human genome (Lister et al.,

2009). On the other hand, RRBS is a targeted methods that combines bisulphite sequencing with the use of restriction enzymes, commonly MspI is used to enrich areas with a large content of CpGs. This enzyme targets 5'CCGG3' sequences, mainly present in CGIs, and allows the digestion of both methylated and unmethylated CpGs. RRBS covers 80% of the CGIs and more than 50% of the promoters, reducing the sequencing requirement of gene expression analyses (Nakabayashi et al., 2023). While WGBS provides more information about the methylation status of a genome, its application is hindered by the high sequencing costs (Ziller et al., 2015), RRBS presents the advantage of yielding a larger coverage of the sequenced positions, reducing the experimental costs.

The main limitation of bisulphite-based methods is their amplification bias. PCR amplification and cloning steps during the preparation of libraries can affect data accuracy, skewing results and overestimating global methylation levels (Olova et al., 2018). This conversion only works under high temperature and low pH, leading to extreme conditions where up to 90% of the input DNA can end up degraded. When DNA resources are limited, this DNA loss can be problematic.

1.4.2. Enzymatic-Modification Based Methods

Enzymatic methods emerge to overcome the limitations of bisulphite-based methods. Enzymatic-methyl-seq (EM-Seq) also relies on the transformation of unmethylated cytosines into uracils, avoiding the use of bisulphite and its associated extreme conditions (Olova and Andrews, 2023; Vaisvila et al., 2021).

EM-seq can be conducted under mild conditions that do not degrade the DNA molecules. Three enzymes are employed to transform cytosines into uracils: ten-eleven translocation 2 (TET2), T4 phage β -glucosyltransferase (T4- β GT), and apolipoprotein mRNA editing enzyme, catalytic polypeptide-like 3 A (APOBEC3A) (Rauch and Pfeifer, 2023).

Methods based on enzymatic modification appear to outperform WGBS. EM-seq demonstrates higher mapping rates, lower duplication rates, increased coverage, and more consistent results across replicates compared to WGBS (Feng et al., 2020). EM-seq also demonstrates less bias in base composition and more accurate methylation detection, particularly in CHG and CHH contexts where WGBS tends to overestimate methylation levels.

1.4.3. Biological Affinity-Based Methods

Methylated DNA molecules can be affinity-purified using 5mC-specific antibodies or methylation-binding proteins that capture DNA-methylated sequences. On the other hand, unmethylated molecules can also be captured using specific protein domains that bind to unmethylated CpGs (Illingworth et al., 2008; Rauch et al., 2006; Rauch and Pfeifer, 2005; Weber et al., 2005).

Methylated DNA immunoprecipitation can be combined with various analytical approaches, including PCR for locus-specific analysis, microarrays, or high-throughput sequencing for genome-wide profiling (Mohn et al., 2009). This technique has been successfully applied to study DNA methylation patterns in diverse samples, including animal and plant tissues, as well as cancer cells (Weng et al., 2009).

1.4.4. Nanopore Sequencing

Oxford Nanopore Technologies (ONT) developed Nanopore sequencing, a third-generation sequencing technique that enables direct sequencing of DNA and RNA molecules by passing them through nanometer-sized protein channels known as nanopores (Branton et al., 2008; Stoddart et al., 2009). The process involves applying an electric potential to an electrolytic solution containing the nanopore, which drives the polynucleotides through the protein-based pore. As the molecule passes through, changes in the electrical current are measured, allowing for nucleotide identification without the need for PCR amplification or chemical labelling (Schmidt et al., 2020). Epigenetic modifications have an effect on current changes, which makes it possible to directly detect DNA methylation pattern. This technique offers advantages such as long read lengths, real-time sequencing, and the ability to detect epigenetic modifications. Nanopore sequencing has diverse applications, including genome assembly, pathogen identification, and antibiotic resistance monitoring (Wang et al., 2021b).

1.4.5. Methylation Arrays

The Illumina Infinium methylation arrays have become a popular and reliable method for detecting DNA methylation marks across the human genome (Wessely and Emes, 2012). The Infinium® HumanMethylation450 BeadChip allows for the assessment of 482,421 CpG sites selected primarily from promoter

regions, first exons, gene bodies and 3'UTRs and requires only 1 µg of genomic DNA as input (Weisenberger et al., 2008).

A bisulphite-based treatment is applied to the genomic DNA before its hybridisation. The primers are designed based on the correlation of the methylation status of neighbour CpGs. CpGs that are proximal in a range of 50 bp present the exact methylation status. A first step of single nucleotide primer extension separates unmethylated CpGs from methylated CpGs after bisulphite conversion. This extension incorporated ddNTP that are labelled and are later detected by sensors scanning the array. The methylation status of the analysed CpG is determined by the ratio of fluorescent signals from methylated and unmethylated CpGs (Rauch and Pfeifer, 2023).

The H2020 RUMIGEN project aimed to develop a comparable array for use in the bovine genome. INRAE, INIA, and QGG collaborated to identify candidate biomarkers linked to heat stress, mastitis, bull fertility, and metabolic stress. The methylation array, known as EpiChip, was developed between 2021 and 2023 and became available to project members in mid-2024.

By leveraging these techniques, researchers can uncover key epigenetic markers linked to livestock traits, a topic explored in the next section.

1.5. Epigenetics in Livestock

Epigenetics offers transformative insights into livestock management, linking environmental factors with phenotypic outcomes. This section illustrates how methylation patterns and other modifications impact livestock traits.

Epigenetic effects have been observed in different livestock species, including cattle, pigs, sheep, and poultry (Thompson et al., 2020). These epigenetic changes are influenced by environmental factors such as diet, toxicants, and infectious agents, potentially affecting traits like growth, disease resistance, and stress responses. The DNA molecule encompasses two interconnected layers of information: genetic and epigenetic. Recent studies have demonstrated that DNA variants influence epigenetic regulation through transcription factor activity, establishing a link between genetic and trait variation (Furey and Sethupathy, 2013; Stefansson et al., 2024). Understanding these genetic and epigenetic layers is crucial for deciphering normal development, disease mechanisms, and the impact of environmental influences on organisms.

1.5.1. Epigenetic Marks as Biomarkers

According to the World Health Organization (Organization and Safety, 2001), a biomarker is defined as a measurable substance in tissues and body fluids, such as chemicals, metabolites, enzymes, and other biochemical compounds, which provides information linking exposure to a chemical with internal dose and outcome, thereby contributing to risk assessment. Genetic biomarkers, based on DNA sequences, are stable indicators that aid in predicting disease susceptibility, diagnosing conditions, and guiding personalised treatments. By contrast, epigenetic biomarkers reflect changes induced by internal and external environmental stressors, making them valuable for monitoring the evolution of individual phenotypic variations (Wang and Ibeagha-Awemu, 2020). This dynamic responsiveness to both intracellular and extracellular conditions, as well as disease progression and environmental influences, offers epigenetic biomarkers a significant advantage in improving disease and production management (García-Giménez et al., 2016).

In human medicine, epigenetic biomarkers offer notable benefits over genetic biomarkers. In diseases such as cancer, cells acquire epigenetic plasticity, which alters chromatin architecture (Pattenden, 2022). Epigenetic biomarkers reflect both genetic background and environmental exposures, and they show a higher frequency of associations with pathology (>90–95%) compared to genetic mutations (<1%) (Skinner, 2024; Verma, 2012). Epigenetic alterations, particularly DNA methylation, can be detected through blood or faecal samples, enabling non-invasive diagnostic tools (Coppedè and Angela Lopomo, 2014).

Epigenetic markers are increasingly linked to livestock health, demonstrating potential as biomarkers for diagnostics, prognostics, predictions, and therapy monitoring. Environmental factors such as farm conditions, feed quality, pathogens, and stress can significantly impact livestock productivity, with these effects traceable through epigenetic markers. Changes in epigenetic mechanisms, particularly during embryonic development, are critical for ensuring healthier pregnancies and optimising farm management strategies. Early identification of these biomarkers can facilitate evaluations and interventions to improve offspring health and productivity. Epigenetic biomarkers are especially valuable for managing chronic, asymptomatic livestock diseases, including metabolic disorders, muscular degenerative diseases, chronic mastitis, subacute ruminal acidosis, and paratuberculosis (Dean et al., 2005; Luo et al., 2018).

1.5.2. Methylation Effects over Phenotypes of Interest

The study of methylation marks has shed light on the impact of epigenetics on livestock phenotypes of interest. Not only genetic alterations do explain much of the phenotypic variation observed in productive traits, but a portion of this variation remains elusive and may be attributed to the epigenome, which is yet to be fully characterised (Halušková et al., 2021). In the late 2010s, advancements in technology facilitated the characterisation of DNA methylation profiles across various livestock species and tissues, including fibroblasts in cattle (Korkmaz and Kerr, 2017), honeybee brains (Li et al., 2017), sheep ovaries (Zhang et al., 2017), blood samples from Creole cattle (Sevane et al., 2019), fibroblast from swine (Wang et al., 2020), and chicken (Han et al., 2020), among others.

The characterisation of the methylome enables the detection of specific CpG methylation marks associated with phenotypes of interest in farming practices. Multiple studies have identified epigenetic alterations in Holstein cows linked to milk production traits across different organs and tissues. For instance, the *EEF1D* gene was found to be hypomethylated during the dry period compared to early lactation in blood samples (Liu et al., 2017). In the liver and mammary gland, methylation of four CpG sites in the distal upstream regulatory region of the *CSN1S1* gene was associated with the frequency of daily milking, which influenced milk production (Nguyen et al., 2014). Additionally, following *E. coli* infection, methylation occurred in the upstream promoter of the *CSN1S1* gene within the liver and mammary gland, affecting the synthesis of $\alpha S1$ -casein (Vanselow et al., 2006). Blood samples revealed that lower-performing cows exhibited hypermethylation compared to high-yield cows, which correlated with reduced milk and protein yields (Wang et al., 2019). Moreover, a study identified 252 DMRs across various herd environments, with 72 DMRs specifically distinguishing between low- and high-yielding cows (Dechow and Liu, 2018).

These findings align with other research showing various epigenetic alterations in cattle in response to different stimuli and diseases. In Brahman cows and their offspring, prenatal stress induced by transportation resulted in 7,407 hypermethylated and 8,721 hypomethylated CpG sites in blood, with 1,205 differentially methylated cytosines located within promoter regions (Littlejohn et al., 2018). *Mycobacterium bovis* infection in Holstein–Friesian cattle led to a genome-wide DNA methylation profile of CD4+ T cells, revealing 760 differentially methylated regions (Doherty et al., 2016). Paratuberculosis in Holstein cows

caused significant epigenetic changes, with 2,000 DMCs and 205 DMRs in the ileum, and 6,394 DMCs and 3,946 DMRs in the ileum lymph node in response to *Mycobacterium avium* spp. paratuberculosis (Ibeagha-Awemu et al., 2020a, 2020b). Clinical mastitis in Holstein cattle resulted in altered methylation of CD4 in blood (Usman et al., 2016; Wang et al., 2013) and changed DNA methylation-regulated IL6R transcription in the mammary glands (Zhang et al., 2018). Further studies on *E. coli*- and *S. aureus*-induced mastitis in Holstein cows revealed extensive genome-wide. Additionally, *E. coli*-induced mastitis led to higher levels of H3K27me3 in silent genes within lymphocytes in subclinical *S. aureus* mastitis cattle compared to healthy cows (He et al., 2016).

Some livestock species are conceived during maternal lactation, a period marked by a significant negative energy balance (Esposito et al., 2014) and heightened nutritional demands. Calves gestated by lactating cows face competition for nutrients with their dam's mammary gland tissue, unlike calves gestated by non-lactating cows or heifers. This competition can have long-term negative effects on the offspring's milk production, lifespan, and metabolic efficiency compared to offspring gestated by non-lactating mothers (González-Recio et al., 2012). In the cited study, it was hypothesised that epigenetics may mediate these performance differences between groups. However, at the time, epigenotyping techniques were not widely available to confirm this hypothesis.

1.5.3. Impact of Nutrition on the Dairy Cattle Methylome

Nutrition plays a crucial role in the epigenetic regulation of gene expression in dairy cattle, particularly through DNA methylation (Lesta et al., 2023). Dietary methyl donor supplementation during late pregnancy and early lactation has the potential to enhance calf performance and immune response, although further research is required to fully understand the underlying mechanisms (Batistel, 2020). Optimising methyl donor nutrition not only enhances immediate productivity but also promotes beneficial epigenetic changes that can influence future generations (Pizano and Williamson, 2020). Thus, optimising dietary inputs in dairy cows has the potential to enhance health, productivity, and offspring development through epigenetic modifications. Key methyl donors, such as methionine, choline, betaine, and folate, are essential for one-carbon metabolism and the production of S-adenosylmethionine, which supports various cellular functions (McFadden et al., 2020). Inadequate nutrition can lead to metabolic disorders, reduced milk yield, and impaired fertility, highlighting the importance

of a balanced diet for optimal epigenetic regulation (Coleman et al., 2021; Murdoch et al., 2016). Supplementation with rumen-protected methionine has been shown to alter hepatic global DNA methylation and influence the methylation of specific gene promoters, such as PPARA, leading to improvements in lipid metabolism and immune function in periparturient dairy cows (Osorio et al., 2016).

Beyond methyl donors such as methionine, choline, and folate, short-chain fatty acids (SCFAs), particularly butyrate, play a critical role in modulating gene expression through epigenetic mechanisms. SCFAs, including acetate, propionate, and butyrate, are produced during microbial fermentation of dietary fibre in the rumen and provide up to 70% of the energy requirements for ruminants. While acetate and propionate serve as primary energy substrates, butyrate exhibits unique epigenetic properties as a histone deacetylase inhibitor (HDACi). By inhibiting HDACs, butyrate induces hyperacetylation of histones, leading to chromatin relaxation and altered gene expression (Wolffe and Guschin, 2000). This mechanism allows butyrate to regulate cellular processes such as differentiation, proliferation, and apoptosis, highlighting its functional role beyond energy metabolism (Myzak and Dashwood, 2006).

In bovine studies, butyrate treatment at physiological concentrations (2.5–10 mM) arrests cell cycles in the G1 phase, suppresses DNA synthesis, and induces apoptosis at higher concentrations (10 mM). These effects have been confirmed in bovine cell lines, where butyrate-induced histone hyperacetylation disrupts chromatin structure and DNA replication. Evidence also suggests potential interplay between histone acetylation and methylation, further reinforcing butyrate's role in regulating chromatin dynamics (Li, 2015).

The ability of butyrate to modulate gene expression and cell cycle progression underscores its importance as both a nutrient and an epigenetic regulator. Understanding these mechanisms provides opportunities to optimise rumen fermentation and enhance animal productivity, health, and resilience through dietary interventions. In addition, the therapeutic potential of butyrate as an HDACi, explored in human oncology, reflects its relevance as a modulator of cellular processes in cattle (Myzak and Dashwood, 2006).

1.5.4. Addition of Epigenetic Information to Genomic Data

Genomic information and genetic biomarkers such as between single nucleotide polymorphisms (SNPs) only explain part of the phenotypic variance and

the heritability of traits of interest in animal production. Integrating epigenetic information into prediction models has the potential to increase the accuracy of predictions for breeding values (Gonzalez-Recio et al., 2015; Ibeagha-Awemu and Khatib, 2023; Yakovlev, 2018). The potential of including epigenetic biomarkers in livestock breeding programmes is being explored with special emphasis in DNA methylation marks (Gonzalez-Recio et al., 2015; Ibeagha-Awemu and Zhao, 2015; Triantaphyllopoulos et al., 2016). These marks can provide insights into the heritability of traits and the environmental influences on gene expression, ultimately enhancing selection strategies (Magee et al., 2014). Phenotype is not only a reflection of the genetic sequence but also a dynamic interplay between genetic and epigenetic modifications, shaped by factors such as nutrition, stress, and management practices. These factors illustrate how the epigenome integrates historical and contemporary environmental influences, ultimately affecting livestock performance and adaptability. Understanding these interactions may lead to more resilient breeds, capable of thriving in diverse conditions while maintaining productivity (Gonzalez-Recio, 2012; Ibeagha-Awemu and Zhao, 2015).

The transmission of epigenetic biomarkers, either through intergenerational or transgenerational epigenetic inheritance, shows that individual phenotypic alterations might be affected by environmental effects on founder generations (Guerrero-Bosagna and Skinner, 2012). The reported associations SNPs and patterns of DNA methylation suggest a significant influence of methylation on the expression of genes that are related, ultimately leading to an impact on phenotypic outcomes (Banovich et al., 2014; Imgenberg-Kreuz et al., 2018).

Epigenetic information could be included in breeding programmes through the modification of the current mathematical models that estimate the GEBV. The contribution of epigenetic on phenotypes of interest in livestock is getting attention and could find an application in improving health related and productive traits in domestic animals (Bernstein et al., 2007; Monk et al., 2019). However, there is not enough data available to quantify the contribution of epigenetic marks to the heritability of productive and health traits in livestock populations. Statistical methods must be developed in order to ascertain the precise impact of epigenetic biomarkers on the variability of phenotypes (Wang and Ibeagha-Awemu, 2020).

Methylation marks influence the final outcome of phenotypes of interest (Hayes, 2009), but the way in which they act can differ. In addition to understanding how an effect operates, it is also important to consider when an effect operates. This is the focus of moderation analysis which explores the

conditions or circumstances under which the effect of X on Y changes. A moderator variable influences the strength or direction of the relationship between an independent variable and a dependent variable. Conditional process analysis integrates mediation and moderation by examining when a particular mechanism is at work. This involves testing whether an indirect effect is dependent on one or more moderator variables. The integration of mediation and moderation analyses allows researchers to examine more complex relationships between variables.

In a simple mediation model, the total effect of X on Y is partitioned into direct and indirect effects. The indirect effect represents the portion of the effect of X on Y that occurs through a mediator variable M , while the direct effect represents the effect of X on Y that is independent of M (Fig. 1-3. A). The indirect effect is quantified as the product of the path from X to M and the path from M to Y . The total effect is the sum of the direct and indirect effects. In more complex models with multiple mediators, there are specific indirect effects through each mediator and a total indirect effect equal to the sum of the specific indirect effects. Moderation analysis examines how the effect of one variable on another is changed by a third variable, the moderator (Fig. 1-3. B). Here, the moderator W modifies the relationship between X and Y . A third model, a conditional process model (Fig. 1-3. C), proposes a combination of the beforementioned two where the moderator W modifies the paths of a mediator model (Igartua and Hayes, 2021). These epigenetic influences provides a foundation for integrating such data into genomic selection models, which is discussed in the following section.

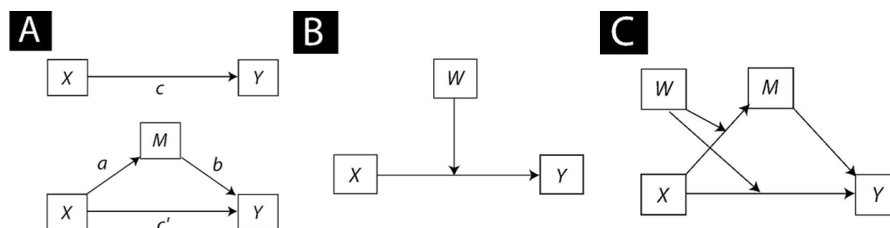


Figure 1-3. Schematic models of mediation (A), moderation (B), and conditional process (C). Adapted from Igartua and Hayes (2021).

1.6. Genomic Selection and Integration of Epigenetic Information

Building on the epigenetic insights discussed earlier, this section explores how genomic selection can benefit from integrating epigenetic data to enhance breeding outcomes.

The initial bovine genome reference assembly, which utilised the Hereford cow “Dominette” as its basis, was officially released in 2009 (Bovine Genome Sequencing and Analysis Consortium et al., 2009). Simultaneously, through resequencing efforts, sets of single-nucleotide polymorphisms numbering in the tens of thousands were uncovered, displaying polymorphisms across multiple cattle breeds (Van Tassell et al., 2008). These SNPs, complete with their mapped positions on the bovine genome assembly, facilitated the process of associating genomic positions with different traits and the reconstruction of historical population dynamics within diverse cattle breeds (The Bovine HapMap Consortium et al., 2009).

In the last two decades, efforts were made to improve livestock production and traits associated with both health and wellbeing of the animals. Genomic selection enables the selection of the parents of the next generation based on genomic estimated breeding values (GEBV) calculated as the sum of the effects of genetic markers across the whole genome over the expected phenotypes (Hayes et al., 2009). It needs a large reference population with a large density of markers to infer their effects and estimate GEBV. Once the effects are quantified, the information can then be implemented on a smaller population with less markers and predict their GEBV with larger accuracies even for young individuals who have not yet produced offspring (Scheifers and Weigel, 2012). The implementation of genomics in farming practises lead to changes in breeding programs and improved important traits such as meat quality, egg yield and milk yield and quality (Wang and Ibeagha-Awemu, 2020). Next-generation sequencing helped uncovering the molecular mechanism regulating gene expression related to productive and health traits (Georges et al., 2019; Gurgul et al., 2019). In the field of animal breeding, multivariate models play a crucial role in implementing a quantitative strategy based on an infinitesimal additive model. Progenitors are selected for the subsequent generation based on their genetic advantages, which are assessed through their genetic potential for various traits, including milk production,

somatic cell count, and reproductive success in dairy cattle (Gianola and Sorensen, 2004).

Genomic selection has brought a transformative impact on the dairy cattle industry by enhancing the accuracy and efficiency of breeding decisions (Hayes et al., 2009; Schefers and Weigel, 2012). In the last decade, the integration of genomic selection in dairy cattle has led to a remarkable enhancement in the pace of genetic improvement (Bouquet and Juga, 2013). Through the utilization of genomic information, breeders can pinpoint animals with the most favourable genetic characteristics at an early stage, facilitating quicker and more precise selection of top-tier individuals for breeding objectives. This technological advancement has not only streamlined the breeding process but has also contributed to the overall improvement of dairy cattle genetics and productivity.

This enhancement has resulted in consistent improvements in economically significant animal traits, as highlighted by Cole et al., (2021). In addition to genomic data, intermediary omics characteristics are now accessible at a reasonable cost. These characteristics include traits that have an impact on the expected genetic influence on the phenotype (Christensen et al., 2021). The integration of omics technologies presents an innovative approach to predicting an animal's phenotype by considering both genetic and environmental factors simultaneously.

Delineating the genetic influences from the non-genetic factors to estimate the heritability of a trait presents considerable challenges. Historically, it has been posited that the transmission of identical alleles results in the manifestation of an identical phenotype (Danchin et al., 2011). But epigenetic marks affect gene expression and thus the expected phenotypes of different individuals.

The non-genetic factors contributing to variations can be integrated into breeding programs by obtaining epigenetic information. To achieve this, epigenetic markers must be (1) temporary (intragenerational) or heritable (either intergenerational or transgenerational), (2) separated from the effects of DNA variants, and (3) their influence must be characterized as causal or consequential. These standards guarantee that epigenetic changes can be efficiently utilised to boost favourable characteristics in livestock, resulting in greater productivity and resilience. Intragenerational epigenetic diversity should be harnessed to enhance management and farming practices, and the marks that can be inherited either

inter or transgenerationally are those deemed suitable for inclusion in breeding programmes (Ibeagha-Awemu and Khatib, 2023).

DNA methylation profiles are subject to modulation by both hereditary and environmental factors, culminating in regions of the genome exhibiting variable methylation patterns. These variations in methylation induce phenotypic plasticity that may facilitate local adaptation to differing environmental contexts (Berbel-Filho et al., 2019). The identification of methylation quantitative trait loci has elucidated a portion of the interindividual variability inherent in DNA methylation (Teh et al., 2014). The interplay among epigenetics, genetics, and environmental factors has the potential to enhance the accuracy of breeding value estimations and mitigate biases encountered in animal breeding practices (Ghorbani and Shokri-Gharelo, 2019). Furthermore, the phenomenon of genotype-dependent methylation may elucidate the observed gene-by-environment ($G \times E$) interactions at specific loci, which could exert influence on various traits (Ma and Ordovas, 2017).

Understanding the role of methylation patterns poses a challenge due to the diverse impacts of methylation marks throughout the genome. CpG dinucleotides demonstrate a bimodal distribution in methylation characteristics, with the majority of cytosines exhibiting high methylation levels, with the exception of those located within CpG islands (Siegfried and Simon, 2010). The methylation status of CGIs has traditionally been associated with the regulation of gene expression, as methylation obstructs the binding of transcription factors to promoter regions (Weber et al., 2007). Conversely, hypermethylation in intragenic regions is thought to mitigate unintended gene body transcription rather than consistently correlate with suppression of gene expression (Illingworth et al., 2008; Illingworth and Bird, 2009). Future livestock improvement programs may benefit from considering epigenetic markers alongside genomic information.

Genomic data can be augmented with epigenetic information to enhance the comprehension of the mechanisms implicated in the regulation of phenotypic expressions (Ibeagha-Awemu and Zhao, 2015). The estimation of GEBVs necessitates an expansion to encompass not solely the cumulative effects of SNPs, but also the factors associated with epigenomics. It is required to delineate the genetic influences from the epigenetic effects to adequately incorporate their differential impacts on phenotypic expression and the estimation of trait heritability (Danchin et al., 2011). Omics data can be incorporated into genomic assessments via nested GBLUP models, in which the initial model measures the

impact of the omic feature on the trait and the additive residual breeding value, while the subsequent model evaluates the mediated breeding value that indicates the genetic impact on the omic feature (Legarra and Christensen, 2023).

The foundation of animal breeding rests upon the premise that the DNA sequence encodes the information necessary for the transmission of traits susceptible to selection for the enhancement of productive levels and health status of subsequent generations (Bonduriansky and Day, 2009). However, it is essential to acknowledge that non-genetic transmission also contributes significantly to the ultimate phenotypic expression of traits. This form of transmission is facilitated by any variable that is not inherent to the DNA sequence of the progenitors within the population (Danchin et al., 2011). The conditions experienced by the mother during the prenatal phase may significantly affect the phenotypes of the offspring, thereby altering the anticipated productive capabilities during the adult life stages of their progeny (Ibeagha-Awemu and Khatib, 2023).

Traditionally, the determination of phenotypic variation in the context of plant and animal breeding is accomplished through the estimation of variance components utilizing linear mixed models, which interrelate fixed and random effects alongside a residual to a specific phenotype (Henderson, 1984). Numerous models have been proposed to incorporate methylation data into linear mixed frameworks. Gianola and van Kaam, (2008) advocated for using an Reproducing Kernel Hilbert Spaces, wherein methylation is treated as an independent variable within linear mixed models, augmented by a vector of nonparametric epigenetic coefficients and a kernel matrix to encapsulate epigenetic affinities among individuals. Breiman, (2001) integrated methylation data within Random Forest models, devising classification trees predicated on a vector comprising both methylation and genotypic information. Varona et al., (2015) advanced the linear model to assess epigenetic variability by incorporating a vector of epigenetic effects, posited to be independent of genetic influences while sharing a common incidence matrix. Although not specifically for epigenetics, Christensen et al., (2021a) proposed a model, wherein the non-genetic effect exerts an influence on the phenotype, while not being reciprocally affected. They used metagenomic information as a case of study, although this can be extended to epigenetic factors.

Another study proposed various frameworks to quantify epigenetic variance, incorporating the utilization of epigenetic relationship matrices and recursive parameters to encapsulate the transmissibility of epigenetic markers (Graham et al., 2021). The integration of genomic and epigenetic information shows the ability

to improve the accuracy of breeding value evaluations and reduce bias estimation. Nevertheless, obstacles persist in accurately estimating genetic variance components in populations that are subjected to rigorous genomic selection (Jensen, 2016). Multiple-matrix animal models have been formulated to elucidate the genetic and non-genetic determinants of phenotypic variance, encompassing environmental, epigenetic, and cultural influences (Thomson et al., 2018). These sophisticated models are applicable to both natural and controlled populations, providing novel avenues to investigate the interaction between genetic and non-genetic elements in influencing evolutionary potential.

The incorporation of epigenetics into genomic selection models underscores the potential for improving livestock breeding programmes, the next section will explore the state of the Spanish breeding programme for Holstein-Friesian cattle and how epigenetic marks can be included into it.

1.7.The Current State of Spanish Dairy Cattle Breeding Programmes

The Spanish breeding programme for the Holstein breed is managed by CONAFE (Confederation of Spanish Holstein Associations), which operates under the support of the Ministry of Agriculture of Spain and the European Union. Recognised as a Qualified Centre for Animal Genetics in accordance with Regulation (EU) 2016/1012 and Royal Decree 45/2019, effective from March 3rd, 2020, CONAFE plays a crucial role in advancing the genetic improvement of Spanish Holstein cattle. CONAFE actively engages with farmers from various regions of Spain, ensuring that their needs are understood and that breeding models are adapted accordingly to address specific regional challenges. The confederation also plays a significant role in public science communication, disseminating advancements in the sector and helping farmers stay informed about the latest developments in livestock breeding. As part of its commitment to sustainability, CONAFE has begun to implement genetic evaluations for methane emission, aligning with the EU's goals for reducing greenhouse gas emissions. Moreover, CONAFE is a key partner in the H2020 RUMIGEN project, which aims to integrate methylation information into routine farming practices to enhance animal welfare and adaptability. The project is also working towards incorporating epigenetic data into breeding schemes, a move that could significantly improve

breeding programmes by considering not only genetic but also epigenetic factors, thereby increasing the resilience and sustainability of livestock populations.

To estimate genetic merit, CONAFE employs models such as BLUP, GBLUP, and SSGBLUP, which combine genetic information from both the animals and their relatives with phenotypic data. The Animal Model allows for the simultaneous evaluation of bulls and cows, using data from daughters to estimate the genetic merit of a bull while considering the contributions of the mothers. This family-based approach enhances the accuracy of genetic indices, particularly for bulls with limited daughter data, and ensures better comparability of indices across herds. The models also account for non-genetic effects, such as husbandry practices, number of lactations, age at parturition, and permanent environmental factors, to more accurately estimate the genetic value of the animals. For productive traits, such as milk, protein, and fat yield, additional non-genetic effects like month of birth and lactation status are also considered, while for conformation traits, the model includes factors such as herd-visit-grader, number of lactations, and the animal's age and lactation status.

CONAFE is responsible for calculating the ICO, a composite index that combines genetic values for both productive and morphological traits in Spanish Holstein cattle. The ICO aims to optimise animal health and farm profitability by integrating genetic merit across various traits, with a higher weight assigned to traits such as protein yield (21%) and a lower weight given to hoof health (4%). Traits considered in the ICO index include conformation, production characteristics, functional longevity, and somatic cell count, as well as reproductive efficiency, which is addressed through traits such as days open and calving ease. By considering these diverse factors, the ICO index strives to strike a balance between profitability, adaptability, and sustainability in breeding programmes. The index is calculated by standardising the genetic merit of different traits relative to their respective standard deviations, ensuring comparability across traits with varying units and variances. The components of the ICO index include milk yield (MY), fat yield (FY), protein yield (PY), the udder composite index (ICU), the feet and legs index (FLI), somatic cell count (SCC), and the hoof health index (HHI). The index is calculated as follows:

$$\begin{aligned}
ICO_{2023} = & 1000 + 1.5 * (11 * \frac{MY}{SD_{MY}} + 17 * \frac{FY}{SD_{FY}} + 21 * \frac{PY}{SD_{PY}} + 12 * \frac{ICU_{2019}}{SD_{ICU}} + 8 \\
& * \frac{FLI_{2019}}{SD_{FLI}} + 11 * \frac{Life\ Span - 100}{SD_{LS}} + 8 * \frac{(SCC - 100)}{SD_{SCC}} + 8 \\
& * \frac{Open\ Days - 100}{SD_{OD}} + 4 * \frac{HHI - 100}{SD_{HHI}})
\end{aligned}$$

Since the June 2019 evaluations, CONAFE has introduced four economic merit indices (IM€T), tailored to specific production systems: milk production, cheese production, farms practising grazing, and those adhering to organic production regulations. These indices incorporate the traits in the ICO index, as well as new traits such as milking speed (MS), calving ease (CE), and ICAP, which approximates the animal's live weight. Since 2023, methane (MET) emissions have also been included, reflecting the programme's ongoing commitment to environmental sustainability and animal welfare. For example, the economic merit index for milk production is calculated as follows:

$$\begin{aligned}
IM€T_{Milk} = & 100 + 0.16 * (28 * \frac{MY}{SD_{MY}} + 9 * \frac{FY}{SD_{FY}} + 16 * \frac{PY}{SD_{PY}} + 6 * \frac{ICU_{2019}}{SD_{ICU}} + 4 * \frac{FLI_{2019}}{SD_{FLI}} \\
& + 13 * \frac{Life\ Span - 100}{SD_{LS}} + 5 * \frac{(SCC - 100)}{SD_{SCC}} + 6 * \frac{Open\ Days - 100}{SD_{OD}} + 3 \\
& * \frac{MS - 100}{SD_{MS}} + 1 * \frac{HHI - 100}{SD_{HHI}} - 4 * \frac{ICAP}{SD_{ICAP}} + 3 * \frac{CE - 100}{SD_{CE}} + 2 \\
& * \frac{MET - 100}{SD_{MET}})
\end{aligned}$$

These indices are flexible and adaptable to the changing needs of the dairy sector. Methylation marks, if found relevant to improving milk yield, health traits or overall wellbeing could be included in them, either as a trait itself combining the methylation rate of positions associated with genes of interest or inside the animal models used to estimate the genetic merit of the animals for different traits. As the integration of methylation data becomes increasingly viable, Spanish breeding programmes are poised to lead innovations in sustainable dairy production.

1.8. Working Hypothesis and Objectives

1.8.1. Metabolic Stress During Dairy Cattle Lactation

The overlap of lactation peak with gestation in dairy cattle is a significant aspect of dairy management, influencing both milk production and reproductive strategies as well as impacting foetus development. As gestation progresses, particularly from 150 to 210 days post-conception, lactation persistency declines, especially in multiparous cows (Chen et al., 2024). This phenomenon creates a negative energy balance (NEB) on the adult cow, which is a prevalent issue during cow lactation, particularly in the early postpartum period (Knight, 2001). This condition arises when energy intake fails to meet the demands of milk production, leading to the mobilisation of fat reserves and associated health risks. Nutritional intake during late gestation is crucial; cows often exhibit reduced appetite, which can affect both lactation and gestation outcomes. This balance is vital to avoid metabolic disorders. Blood parameters indicate that lactating cows have different metabolic profiles compared to those in late gestation, highlighting the physiological changes that occur during this overlap (Pysera and Opalka, 2000).

High-producing dairy cows, especially those with high genetic merit, often experience an increased NEB because their feed intake is insufficient to meet the energy demands of pregnancy and lactation (Martens, 2023). NEB can lead to conditions such as ketosis and fatty liver syndrome, which are exacerbated by elevated non-esterified fatty acids (NEFA) levels (Štolcová et al., 2024). This negative balance can also impact the reproductive performance, resulting in delayed oestrus and lower conception rates (Mekuriaw, 2023). Calves born to cows with higher NEFA concentrations or oxidant status index show lower birth weight, increased reactive oxygen and nitrogen species, and altered inflammatory responses (Ling et al., 2018). Maternal metabolic stress can also affect oocyte and embryo quality, leading to changes in DNA methylation patterns and impaired embryo-endometrial signalling (Leroy et al., 2017).

During NEB, high-producing dairy cows mobilise fat stores through lipid mobilisation to meet their energy demands, which can lead to the production of ketone bodies via ketogenesis (Holtenius and Holtenius, 1996). These ketone bodies serve as an alternative energy source for both the cow and the developing calf, although excessive levels can negatively impact calf health. Moreover, NEB often results in insulin resistance, which affects nutrient availability for the calf,

potentially impairing its growth and development (Wathes et al., 2007). The levels of insulin-like growth factor 1 (IGF-I), a crucial hormone for growth, are also influenced by NEB, further impacting the calf's physiological development (Fenwick et al., 2008; Zulu et al., 2002). Consequently, the calf may adapt by optimising its metabolic pathways to utilize available nutrients more efficiently, ensuring its survival and growth despite the mother's compromised energy status.

Lactating cows demonstrate diminished foetal and placental weights in comparison to their non-lactating counterparts, which can be attributed to the hormonal and metabolic alterations that occur during lactation (Green et al., 2012). Elevated levels of plasma growth hormone and non-esterified fatty acids in lactating cows may impede foetal growth, whereas non-lactating cows exhibit increased concentrations of glucose and insulin, thereby facilitating enhanced foetal development. Nutritional deficiencies during the early stages of gestation in lactating cows result in suboptimal body condition scores, negatively influencing cow-calf performance and the immunological status of the progeny (Noya et al., 2019). The heightened metabolic requirements of lactation may put the cow's health at risk, potentially resulting in diminished milk quality and reduced growth rates in calves. Conversely, certain research findings imply that the lactation period does not markedly influence calf behaviour or growth performance, suggesting that additional factors may contribute to calf development. This underscores the intricate interplay between lactation and gestation in bovine species (Mohammed et al., 2018).

1.8.2. The Particularity of the Holstein Cattle

Over the past 50 years, the fertility of high-producing lactating dairy cows has declined, primarily due to intensive selection for increased milk production. The physiological and metabolic changes associated with high milk yields, such as decreased concentrations of glucose, insulin, and insulin-like growth factor I, and increased levels of circulating non-esterified fatty acids and ketone bodies, are linked to nutrient partitioning during negative energy balance (Lonergan et al., 2016). Pregnancy rates are around 55%-60% in heifers and 35% to 40% in lactating cows but can be increased with progesterone-based protocols (Guner et al., 2023; Lucy, 2001). These low rates might be linked to lack of follicle development, suboptimal reproductive tract environment or low oocyte quality (Walsh et al., 2011).

The low pregnancy rate in dairy cattle can be attributed to several interrelated factors, including genetics, management, and physiological issues. Genetic factors play a significant role, as low heritability of fertility traits and its negative correlations with milk production hinder genetic improvement in fertility (Lucy, 2007). Management practices, such as inadequate breeding programs and nutritional deficiencies, exacerbate reproductive challenges, leading to conditions like anovulatory cycles and irregular oestrous patterns. Additionally, the statistical genetic architecture of calving performance traits reveals heritable variability, indicating that specific genomic regions influence calving ease and perinatal mortality, which are critical for successful calving (Jakimowicz et al., 2022; Purfield et al., 2020).

As discussed above, conception occurs during the lactation peak in dairy cattle. Around 75% of the gestation of the calf overlaps with the lactation of the dam in a period of negative energy balance. Holstein cows have been intensively selected for milk yield, increasing from 5,000 to 12,000 kg per lactation (Charfeddine and Pérez-Cabal, 2017; Zehetmeier et al., 2014). Cows that are lactating and gestating at the same time need to partition the energy between the placenta and the mammary gland. Epigenetic effects on the gestated calves are expected to be different in lactating cows than in non-lactating pregnant cows. The implications for prenatal reprogramming of calves gestated by highly productive cows are likely to be more pronounced, given that they are gestated under more demanding metabolic conditions (Berry et al., 2008). Cows that are born to lactating dams exhibit inferior performance compared to those born to non-lactating dams. Specifically, calves born to dams in their first lactation produced 18 kg less than first calving cows, those born to dams in their second lactation produced 47 kg less, and those born to dams in their third lactation produced 91 kg less (González-Recio et al., 2012). The extent of the dam's lactation status leads to a higher disparity in the milk yield of her offspring when compared with calves born to primiparous cows. Individuals gestated during their mother's lactation, with milk yields below the 50th percentile, exhibited lower milk yield in their first lactation but similar functional lifespan compared to those gestated in absence of maternal lactation. Less productive cows tend to have fewer metabolic diseases and better reproductive performance, leading to longer retention in herds. The study showed that negative effects of the concurrence of lactation and gestation are more remarkable in high-producing cows.

To overcome this limitation, the voluntary waiting period (VWP) in dairy cattle can be adjusted so that the conception does not overlap with the lactation peak. This term refers to the interval between calving and the first eligible insemination, which can be adjusted based on individual cow characteristics. Research suggests that extending the VWP, particularly beyond the conventional range of 25–100 days, can improve reproductive performance, especially in primiparous cows. For example, cows with an extended VWP of 145–215 days exhibited higher first-service conception rates (67% vs. 51%) and required fewer inseminations per conception (1.6 vs. 2.0) compared to those with a conventional VWP (Rasmussen et al., 2023). Longer VWPs may also provide cows with sufficient time to recover their energy balance, potentially enhancing overall health and reducing the incidence of diseases (Rasmussen et al., 2024). However, while extended VWPs can yield improved fertility outcomes, studies indicate no significant effects on milk production or udder health, as evidenced by similar somatic cell counts and mastitis rates across varying VWP lengths (Ma et al., 2022). Therefore, the VWP serves as a valuable management tool that can be tailored to optimise reproductive efficiency and metabolic health in dairy herds (Burgers et al., 2023; Wang et al., 2024).

1.8.3. Objectives

This thesis addresses the challenge of identifying methylation signatures within the Holstein genome, with the objective of correlating the phenotypic traits of dams with the methylation profiles of their offspring, to incorporate epigenetic data into breeding programmes.

Principal Objectives of This Thesis

1. Compare Oxford Nanopore Sequencing results with the current gold standard, Bisulphite sequencing, and establish a pipeline to both sequence and extract methylation information from animal samples.
2. Associate maternal phenotypes with the methylation status of different genomic regions of their offspring.
3. Include methylation information in the linear mixed models used to estimate animal's breeding values in breeding programmes.

Specific Objectives

Chapter 1

- 1.1. Establish a protocol for sequencing and a comprehensive bioinformatics pipeline to facilitate analysis utilizing Oxford Nanopore Sequencing.
- 1.2. Conduct a comparative analysis of the methylation frequency of the CpG sites identified through ONT sequencing against the established gold standard by sequencing the same samples using both methodologies.
- 1.3. Investigate the distribution of genomic positions of the CpG sites encompassed by ONT and bisulphite sequencing.
- 1.4. Identify differentially methylated cytosines and provide annotations to ascertain the specific genomic features with which they are associated.

Chapter 2

- 2.1. Conduct comparative analyses of differential methylation, encompassing both CpGs and genomic regions, in order to determine the differences in the methylome of calves gestated by lactating dams versus those gestated by non-lactating dams.

- 2.2. Annotate the differentially methylated cytosines and regions to their corresponding genes and discuss their implication in embryo development, health and production of the progeny.
- 2.3. Develop a co-methylation network to correlate the lactation status of the dams with the whole-genome methylation profiles of their offspring, in addition to associating the phenotypes of the offspring with their respective methylomes.

Chapter 3

- 3.1. Simulate the methylation patterns of genotyped cows using the methylome of six calves sequenced via ONT as a reference.
- 3.2. Estimate an epigenetics-mediated breeding value for the aforementioned animals.
- 3.3. Develop a statistical framework that integrates genomic and methylation data to evaluate the genetic merit of individuals.
- 3.4. Compare the outcomes of this multi-omics model with those derived from traditional approaches used in animal breeding programmes.

2. Materials and Methods

The structure of this thesis comprises three primary sections. The first segment is dedicated to detailing the materials and methodologies employed in the comparison of RRBS, a widely accepted method for methylation mark detection, with Nanopore Sequencing, aimed at establishing an operational workflow and evaluating the sequencing output of ONT. In the subsequent section, an analysis is conducted to distinguish differentially methylated marks in the genomic composition of calves gestated by cows that are either lactating or non-lactating. Lastly, the focus of the final section is on the replication of the methylation pattern observed in the six samples from the preceding section on 13,000 genotyped animals.

2.1. Materials

Six semen samples provided by INRAE were used for the publication that composes chapter 1. Three came from fertile bull, while the other three came from sub-fertile bulls. For chapter 2, six blood samples were obtained from new-born calves from a commercial farm. Three of the calves were gestated by non-lactating heifers, the remaining three calves were gestated by cows on their third pregnancy and second lactation. Cows were unrelated. These six blood samples were also used for the analyses that compose the third publication of this thesis. Additionally, 13,183 cows were genotyped by CONAFE using the BovineSNP50 Illumina SNP array containing 54,609 SNPs.

The Ethics Committee of the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA-CSIC) 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 noninterventional study with reference CEEA-PH-2021-08. All methods were carried out in accordance with relevant guidelines and regulations from the European Union.

2.2. Methods

2.2.1. DNA Extraction

Semen Samples

Around 20 million spermatozoa were lysed from each semen sample at INRAE using 50mM dithiothreitol and 0.2 mg/ml proteinase K. The DNA was then extracted using a phenol-chloroform method to isolate the genomic DNA and the pellet was dried and resuspended in TE buffer (10 mM Tris-HCl, pH 7.5, 2 mM EDTA). The integrity of the six samples was confirmed through an agarose gel electrophoresis and the concentration was measured using a Qubit Fluorometer.

Blood Samples

The process of DNA extraction was carried out using the Monarch[®] HWM DNA Extraction Kit for Cells & Blood manufactured by New England Biolabs[®]. This kit was chosen for its capability to yield long DNA fragments, thus enhancing the overall sequencing output. To optimise the sequencing yield even further, the DNA underwent fragmentation to achieve a length of 8 kilobases utilizing a Covaris g-TUBE[®] by centrifuging the sample at 8,600 rpm, following the manufacturer's recommendations. Subsequently, the purity and concentration of the samples were determined utilizing a Qubit Fluorometer from Thermo Fisher Scientific.

2.2.2. Sequencing

Reduced-Representation Bisulphite Sequencing

The DNA samples were first digested with the restriction enzyme MspI, followed by end-repair and ligation with Illumina adapters. Size selection was performed using SPRIselect magnetic beads (Beckman-Coulter), and the DNA then underwent bisulphite conversion twice using the EpiTect kit from Qiagen. The amplified DNA was generated using Pfu Turbo Cx hotstart DNA polymerase (Agilent). The libraries were quantified using a Qubit 2.0 Fluorometer and sequenced on an Illumina HiSeq4000 sequencer, producing 75 bp paired-end reads.

Oxford Nanopore Sequencing

DNA samples were fragmented to an average length of 8 kb using a Covaris g-TUBE[®]. Library preparation was conducted using 3 µg of DNA and the SQK-

LSK110 kit. The sequencing was performed on a GridION sequencer, where each sample was loaded onto an R9.4 flow cell.

2.2.3. Bioinformatics

Methylation Calling from RRBS-Sequenced Samples

Methylation patterns from RRBS-sequenced samples were analysed using Bismark (v0.20.0) (Krueger and Andrews, 2011). Alignment of the samples to the reference genome (ARS-UCD1.2) was performed using Bowtie 1.2.1 (Langmead and Salzberg, 2012). During this process, CpGs located on unplaced scaffolds were excluded from further analysis. For the remaining CpGs, methylation frequencies were calculated using Bismark's methylation calling algorithm, following the methodology outlined by Costes et al. (2022).

Nanopore Base and Methylation Calling

The process used to perform the base and methylation calling of the semen samples from the first publication was as follows. Basecalling was carried out utilizing Guppy version 5.0.14, and methylation was detected utilising the model `dna_r9.4.1_450bps_modbases_5mc_cg_sup.cfg`. A quality control threshold of Q-score > 10 was enforced, with NanoPlot (v1.36.2) (De Coster et al., 2018) being employed for quality control assessment. The analyses were conducted using Nanopolish (v0.13.2) (Loman et al., 2015) in accordance with the recommended procedure by its developer. The FAST5 and FASTQ files acquired were organized and then aligned against the *Bos taurus* reference genome (ARS-UCD1.2) utilizing minimap2 v2.17-r941 (Li, 2018). Methylation sites were called using the `call_methylation` function from Nanopolish, and the frequencies were extracted utilizing the `calculate_methylation_frequency.py` script.

For the blood samples from chapters 2 and 3, each sample was sequenced individually within an R9.4 flow cell. Utilizing Guppy version 6.3.7, basecalling was performed with the `dna_r9.4.1_e8.1_modbases_5mc_cg_sup.cfg` model. Quality assessment of the samples was performed using NanoPlot (v1.36.2) (De Coster and Rademakers, 2023) to verify that both the Phred score and mean length of the samples met the required quality standard.

The analyses were conducted utilizing f5c version 1.1 (Gamaarachchi et al., 2020) on the Bracewell CSIRO HPC system located in Australia. Indexing of FAST5 and FASTQ files was executed through the `f5c index` function, followed by

alignment to the reference genome *Bos_taurus.ARS-UCD1.2.fa* using *minimap2*. Subsequently, the files were organized with *samtools* v1.16.1 (Li et al., 2009) to produce a bam file. Identification of candidate sites was carried out by extracting data from the FASTQ files employing the *call_methylation* function within the *f5c* software. The methylation frequency was then calculated using the *meth-freq* feature of *f5c*. The distribution of methylation frequencies obtained by *f5c* for each sample was assessed utilising *methykit* v3.16 (Akalın et al., 2012). To estimate the methylation frequency of CpGs identified in each sample near the transcription start site (TSS) and transcription termination site (TTS), the *ensembl genesBtau.gff3* annotation file (https://ftp.ensembl.org/pub/release-111/gff3/bos_taurus/) was used.

2.2.4. Statistical Methods

Differential Methylation Analyses

The R (version 4.2.1) (R Core Team, 2021) package *DSS* v2.46.0 (Park and Wu, 2016), specifically tailored for bisulphite-sequenced data, was utilised for estimating differentially methylated regions. Transformation of methylation files from *f5c* was done to adhere to *DSS* requirements, including retaining chromosome, CpG methylation site position, sequencing depth, and number of methylated reads. CpG sites were only considered when their coverage was $\geq 7X$, regardless of the methylation status.

The methylation object was then established using *makeBSseqData* function, and differential methylation at each CpG was assessed with *DMLtest*, with manual selection of sample groups. Smoothing was enabled for genome-wide analyses as advised, with DMCs identified using *callDML* function and a p-value threshold of 0.05. Delta was set at 0.2 for the minimum difference between group means, followed by filtering remaining sites with an FDR of 0.05. The same approach was applied to DMRs using *callDMR* function, which identifies statistically significant CpG regions and merges positions if at least 3 CpGs are within 50 base pairs.

Networks

Windows of 500 bp along the genome were created using *pycoMeth* (Snajder et al., 2023), estimating the log likelihood-ratio (llr) for methylation probability in

each window. Windows with llr values inside the range of -3.5 to 3.5 were discarded to eliminate ambiguities, and only windows present in all samples were retained.

The phenotypes of lactation and body capacity index (ICAP) and the estimated breeding value were provided by the Spanish Holstein Association, CONAFE. Both EBV and ICAP are relevant for assessing livestock genetics. The relationship between the regions was analysed using a PCIT analysis (Reverter and Chan, 2008). The PCIT algorithm comprises of two primary steps: (1) estimating partial correlations in each trio of nodes (genes, phenotypes, or DMRs) to assess the strength of the linear relationship when the first node pair is uncorrelated to the third node, and (2) utilizing the Data Processing Inequality theorem of Information theory to determine the tolerance level and eliminate the nonsignificant associations. The correlation matrix was visualized in Cytoscape (Shannon et al., 2003) using the *yFiles organic* layout, with nodes coloured based on methylation status and sized according to proximity to promoter regions.

PCIT

The PCIT algorithm is comprised of two principal steps: (1) the estimation of partial correlations within each triad of nodes (which may include genes, phenotypes, or DMRs) to ascertain the magnitude of the linear association when the initial node pair exhibits no correlation with the third node, and (2) the application of the Data Processing Inequality theorem from Information theory to compute the threshold of tolerance and eliminate associations that lack statistical significance. An association was deemed significant if the correlation between the initial two genes exceeded the established tolerance level. Regions exhibiting a correlation equal to or greater than 0.75 were preserved for further analytical procedures.

Gene Annotation and Enrichment Analyses

Gene annotation was conducted utilizing the *annotatePeak* function within the R package ChipSeeker (Yu et al., 2015). Annotation packages chosen were Org.Bt.eg.db (Carlson, 2019) and TxDb.Btaurus.UCSC.bosTau9.refGene, corresponding to the *Bos taurus* reference genome ARS-UCD1.2. A region spanning 3,000 base pairs in both the upstream and downstream directions of the promoters was defined as the transcription start site. Genomic positions were

annotated to the nearest gene/microRNA within 50 kb distance (Porto-Neto et al., 2014).

Enrichment analyses were carried out using the *compareCluster* function from the R package clusterProfiler (Wu et al., 2021), adjusted by the Benjamini–Hochberg procedure to control the false discovery rate with a p-value cut-off of 0.05. The *enrichGO* function was utilized for a gene ontology over-representation test. The networks' enrichment was executed utilizing g:Profiler (Raudvere et al., 2019), with the statistical domain scope set as 'only annotated genes' and the significance threshold as '*g:SCS threshold*' with an adjusted p-value threshold of 0.05.

GOBLUP

A multiomic approach was implemented adapting the procedure described in Christensen et al., (2021). This procedure considers that the productive trait is controlled by epigenetic and additive effects, and at the same time the epigenetic effect is controlled by additive effects and has a recursive mode of action on the productive trait. The model was named as Genomic Omic BLUP (GOBLUP). We adapted this model to include genomic-epigenomic effects as:

$$\mathbf{y} = \mathbf{X}'\boldsymbol{\beta} + \mathbf{Z}'_{\mathbf{m}}\mathbf{y}_{meth_w}\boldsymbol{\alpha} + \mathbf{Z}'\mathbf{a}_r + \mathbf{e} \quad (\text{Eq. 1})$$

$$\mathbf{y}_{meth_w} = \mathbf{X}'_{\mathbf{m}}\boldsymbol{\beta}_{\mathbf{m}} + \mathbf{M}'\mathbf{g}_{m_w} + \mathbf{e}_w \quad (\text{Eq. 2})$$

Terms in Eq [1] are described next: $\mathbf{y}$ is the phenotype of the epi-moderated trait (controlled by epigenetic effects), $\mathbf{X}$ is the incidence matrix that relates the systematic effects to the phenotypes (in this case only the population mean) with $\boldsymbol{\beta}$ being its effect, $\mathbf{Z}_{\mathbf{m}}$ is an incidence matrix that relates the methylation phenotype of each individual to their corresponding phenotype for the epi-moderated trait, with $\boldsymbol{\alpha}$ being the vector of the methylation effects on the epi-moderated trait, $\mathbf{Z}$ is the incidence matrix that relates each animal to its additive residual breeding value, $\mathbf{a}_r$ was the residual additive genetic effect assumed to be distributed as $\mathbf{a}_r \sim N(0, \mathbf{H}\sigma_{a_r}^2)$, $\mathbf{H}$ is the joint genomic and pedigree relationship matrix, and $\mathbf{e}$ is the vector of residual terms distributed as $N(0, \sigma_e^2)$.

In equation 2, $\mathbf{y}_{meth_w}$ was the vector containing the observed methylation frequency of individuals in window w , $\mathbf{X}_{\mathbf{m}}$ was the matrix that relates the systematic effects to the phenotypes (in this case only the population mean), $\boldsymbol{\beta}_{\mathbf{m}}$

are the methylation effects, $\mathbf{M}$ is the incidence matrix that relates the animals to the additive genetic effect for window w of the animal ($\mathbf{g}_{m_w}$) assumed to be distributed as $\mathbf{g}_{m_w} \sim N(0, \mathbf{H}\sigma_{m_w}^2)$ and $\mathbf{e}_w$ was the vector of residuals assumed to be distributed as $N(0, \sigma_{e_w}^2)$, with $\sigma_{m_w}^2$ and $\sigma_{e_w}^2$ being the additive genetic and residual variances. This model was applied to the phenotypic methylation observed in 13 genetic windows ($w=\{1...13\}$) (*i.e.* the methylation frequency of each window was analysed separately).

The final total estimated epigenetic value (EEV) was calculated as $EEV_i = \hat{a}_{r_i} + \hat{a}_{m_i}$, where the epi-moderated breeding value was $\hat{a}_{m_i} = \sum_{w=1}^{13} \hat{\alpha}_w \hat{g}_{i,m_w}$. The EEV includes the additive residual breeding value and the additive breeding value for the liability to methylation times the methylation effect of each window on the epi-moderated trait.

2.2.5. Simulation of the Methylation Information

The analysis was initiated by iteratively processing each gene entry within the *Bos taurus* gff3 dataset (https://ftp.ensembl.org/pub/release-111/gff3/bos_taurus/). Essential gene information, including gene names, chromosomes, start and end positions, and gene orientations, was extracted. The gene region was divided into thirteen consecutive, non-overlapping windows. These defined windows were then iterated through, and methylation data within each window was extracted. Windows 1 and 2 were associated with the promoter and TSS regions, whereas windows 10, 11, and 12 represented the gene body positions, which transcribes to protein and thus was referred to as the “core” region. All other regions were defined as the “coat” region. The mean methylation of each window (μ_w), and its standard deviation ($\sigma_{p,w}$), were estimated from the observed methylation in the six sequenced samples.

Two traits were simulated for the 13,183 animals with genotype data: (1) the liability to methylation at cytosine nucleotide followed by guanine nucleotide sites (CpG) as a purely additive trait, and (2) a trait (e.g. milk yield) with additive effect regulated by epigenetic effects through methylation. The SNP effects are moderated by the epigenetic effects, impacting the trait acting as the mediator.

For the first trait (liability to methylation), 10,000 SNPs were randomly selected and their effects (γ) were drawn from a gamma distribution with shape and scale hyperparameters of 0.42 and 5.4 (Hayes and Goddard, 2001). This

methylation effect refers to the genetic control of the individual on the susceptibility to present methylation on CpG sites.

The breeding value for susceptibility to methylation (mBV) for individual i was computed as

$$mBV_i = \sum_{j=1}^{j=10,000} x_{i,j} \cdot \gamma_j \quad (\text{Eq. 3})$$

where $x_{i,j}$, was the SNP_j genotype, γ_j was the SNP effect for liability to global DNA methylation.

The methylation frequency phenotype at each window (y_{meth_w}) was computed using the phenotypic variance observed in the sequenced individuals, as

$$y_{meth_{i_w}} = \mu_w + mBV_{i_w} + e(m)_{i_w},$$

where μ_w is the average methylation in window w , observed in the sequenced animals, mBV_{i_w} is the genetic value for liability to methylation distributed as $N(0, \sigma_{a_{meth_w}}^2)$, and $e(m)_{i_w}$ is the environmental residual effect on liability to methylation distributed as $N(0, \sigma_{e_{meth_w}}^2)$. The $\sigma_{a_{meth_w}}^2$ and $\sigma_{e_{meth_w}}^2$ are the corresponding variances for the genetic and environmental effects for the respective window w . These variances, and hence the corresponding effects, were adjusted to simulate three heritability scenarios $h^2 = 0.1$ (low), 0.3 (medium) and 0.8 (high).

The simulated trait, regulated through epigenetic effects acting as moderators, was generated as follows: First, the trait was assumed to be regulated by 10,000 SNP loci: 7,500 SNPs randomly selected from core regions to simulate the effect of the gene bodies on the phenotype; and 2,500 SNPs randomly selected outside the genic region. SNP effects (β) were also drawn from a gamma distribution with a shape and scale hyperparameters of 0.42 and 5.4 (Hayes and Goddard, 2001).

Rather than modelling a purely additive genetic effect, an epigenetics-moderated true breeding value incorporating both genetic and epigenetic effects, termed $epi_{mod}TBV_i$, was simulated. Here, the epigenetic effect moderates the influence of the SNP effects on the phenotype but does not independently

contribute to the production trait. We termed this phenotype as epi-moderated trait.

$$epi_{mod}TBV_i = epiTBV_{core_i} + epiTBV_{coat_i} \quad (\text{Eq. 4})$$

Where $epiTBV_{core_i}$ is the breeding value of individual i from core regions whose expression is modulated by methylation calculated as

$$epiTBV_{core_i} = \sum_{j=1} x_{i,j} * [\beta_j * (\frac{1}{2}(1 - y_{meth_{i,1}}) + \frac{1}{2}(1 - y_{meth_{i,2}}))] \quad (\text{Eq. 5})$$

And $epiTBV_{coat_i}$ is the breeding value from core regions not affected by methylation, calculated as:

$$epiTBV_{coat_i} = \sum_{j=1} x_{i,j} * \beta_j \quad (\text{Eq. 6})$$

where $x_{i,j}$ is the genotype for SNP j in animal i , β_j is the effect of the SNP j for the phenotypic trait, $y_{meth_{i,1}}$ is the methylation frequency for animal i in window 1 and $y_{meth_{i,2}}$ is the methylation frequency for animal i in window 2, which are the windows affecting the gene promoter. Note that $y_{meth_{i,1}}$ and $y_{meth_{i,2}}$ mimic the expression level of the gene regulated by methylation effects.

The phenotype for the epi-moderated trait (y) was simulated as:

$$y_i = \mu + epi_{mod}TBV_i + e_i \quad (\text{Eq. 7})$$

where μ is the population mean, $epi_{mod}TBV_i$ is the epigenetics-moderated true breeding value for the yield trait and e_i is the residual term following a distribution $N(0, \sigma_e^2)$, where σ_e^2 was set to obtain a $h_{prod}^2 = 0.3$ under a purely additive model (calculating the genetic variance from breeding values without considering methylation: $Var(\mathbf{X}\boldsymbol{\beta})$).

3. Chapter 1 - Oxford Nanopore Sequencing as an Alternative to Reduced Representation Bisulphite Sequencing for the Identification of CpGs of Interest in Livestock Populations

López-Catalina, A., Costes, V., Peiró-Pastor, R., Kiefer, H., and González-Recio, O. (2024). Oxford nanopore sequencing as an alternative to reduced representation bisulphite sequencing for the identification of CpGs of interest in livestock populations. *Livestock Science* 279, 105377.
doi: 10.1016/j.livsci.2023.105377

3.1. 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 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.

3.2. 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 (Halusková 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 (Cattaneo et 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 5mC 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.

3.3. Material and methods

3.3.1. 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.

3.3.2. 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.

3.3.3. 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 ($\delta=0.2$), which was the minimum value for the difference of the estimated methylation means for each group.

3.3.4. Comparison of Called CpGs

ChipSeeker (Yu et al., 2015) was used to plot the CpGs with a coverage $\geq 4x$ 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).

3.4. Results

3.4.1. 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 $\pm$ 48,578.47 for bisulphite sequencing and 20,030,834 $\pm$ 1,150,110 for ONT. Nanopore sequencing detected 8.12x $\pm$ 0.50 more sites compared to RRBS when no filter was applied, 5.1 $\pm$ 1.64 for a filter of 4x and 1.87 $\pm$ 0.86 for a filter of 7x (Fig. 3-1).

Table 3-1 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 $\pm$ 0.05 and 0.651 $\pm$ 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. 3-2. 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.

3.4.2. 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. 3-3). CpG sites detected by RRBS appeared mainly in regions close to promoters, whereas ONT detected a larger number of CpGs intergenic regions (Fig. 3-4).

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. 3-4 and 3-5).

The density of CpGs increased around the TSS, and then a sudden drop followed by a plateau afterwards was observed (Fig. 3-6). The peaks in Fig. 3-6 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. 3-5). The KDE shows a higher density of the marks at the end of the chromosomes, especially in the case of RRBS (Fig. 3-7). The higher the coverage filter, the more concentrated the sites appeared at the end of the chromosomes.

3.4.3. 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. 3-8 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. 3-9 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. 3-10 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. 3-11 shows the overlap of those genes identified by both sequencing methods. Specifically, ONT4x identified 668 genes that were not detected by any other method, accounting for around 74% of the genes identified by ONT4x. Nineteen genes were detected only by ONT7x. Forty-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.

3.4.4. 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. 3-12. Out of the total 211 QTLs previously described, 30 were directly associated with bull fertility. Table 3-2 summarises the number of QTLs associated to different male fertility traits called by each technique. Most of them were related to scrotal circumference. Fig. 3-13 depicts the overlap between the calling made by each technique is depicted. ONT 4x was the technique that called a larger number

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

3.5. 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 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 not restricted to promoter regions. This difference can be explained by the library preparation for each sequencing method. RRBS relies 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.

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., 5mC, 5hmc, 6mA and 4mC) 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 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.

3.6. 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 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.

3.7. 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.

3.8. Ethics Approval

The Ethics Committee of the Instituto Nacional de Investigación y Tecnología Agraria 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.

3.9. Acknowledgements

A. López-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).

3.10. Tables

Table 3-1. 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	49656	0.961	1188642
1	7x	0.683	26620	0.962	759621
2	4x	0.648	48355	0.961	1130786
2	7x	0.7	26675	0.956	720,805
3	4x	0.617	42609	0.963	1149859
3	7x	0.666	22427	0.961	749179
4	4x	0.582	7870	0.962	222448
4	7x	0.642	3851	0.961	133258
5	4x	0.494	17720	0.959	785032
5	7x	0.552	7408	0.953	445967
6	4x	0.607	45300	0.959	1106641
6	7x	0.66	23175	0.958	695911

Table 3-2. 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

3.11. Figures

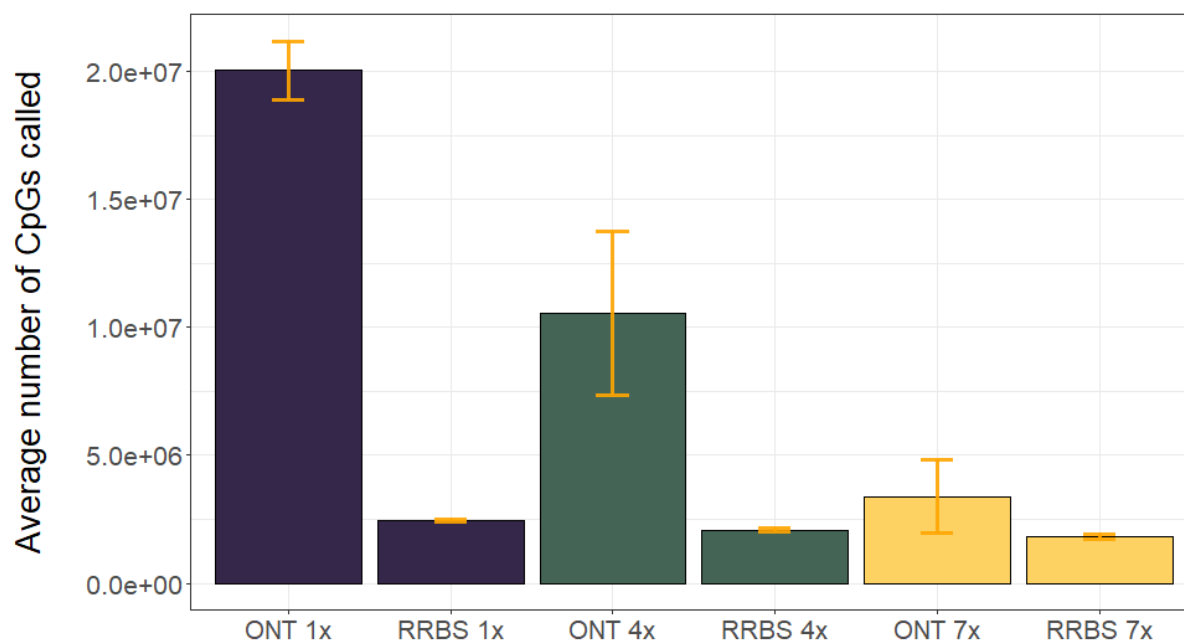


Figure 3-1. Average CpG sites called (y-axis) by each technology (x-axis). Colours represent the coverage filter applied, dark blue for unfiltered files, green for a minimum coverage of 4x, and yellow for 7x. Orange vertical lines represent the standard deviation.

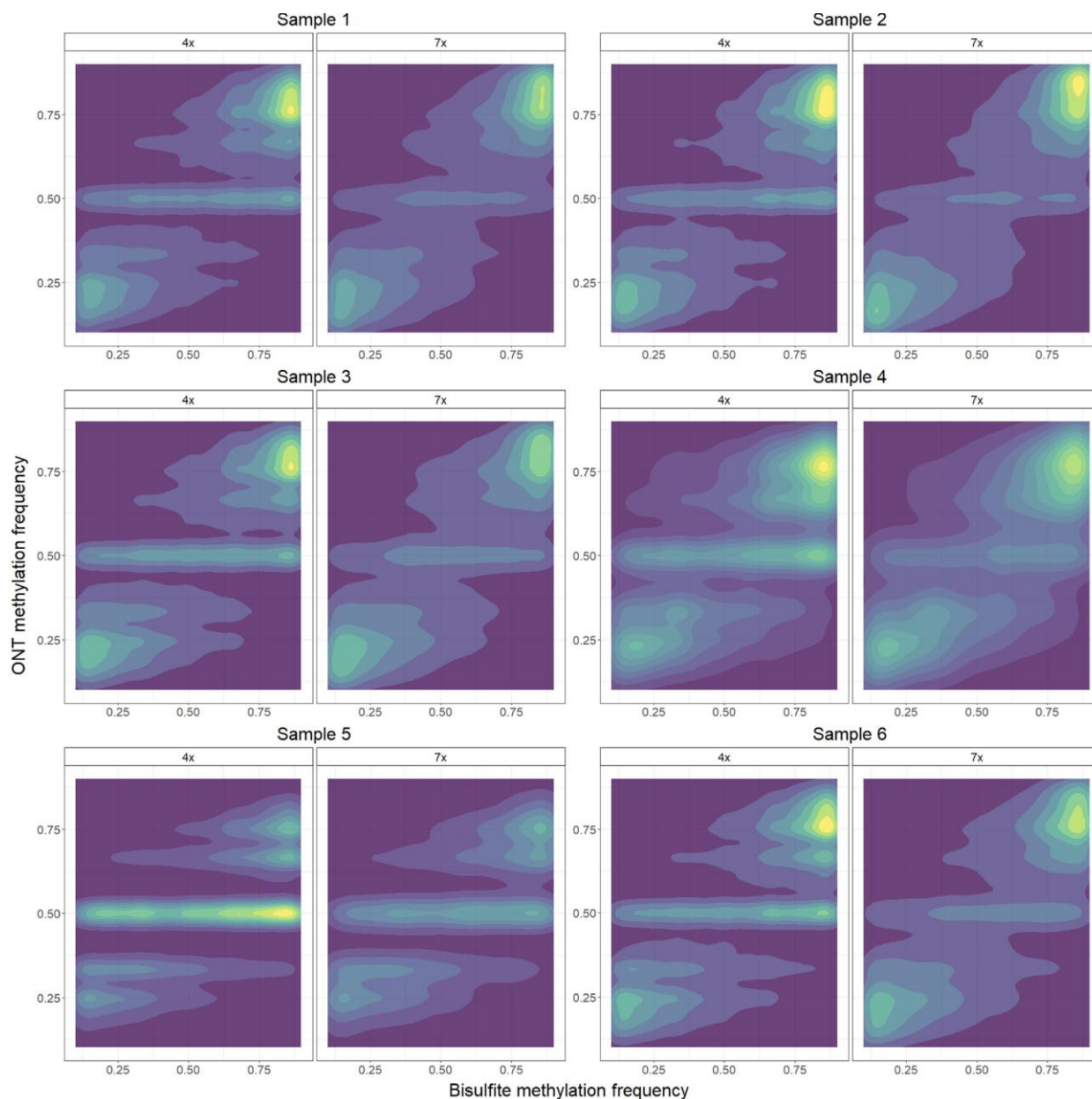


Figure 3-2. Heatmaps representing the correlation between the methylation frequency called for the same position by both ONT and RRBS. Sites with a frequency >0.9 or <0.1 were filtered out. Each duo represents a sample with two different coverage filter applied: 4x and 7x. Y-axis represents the methylation frequency called by ONT and x-axis the one called by RRBS.



Figure 3-3. 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).

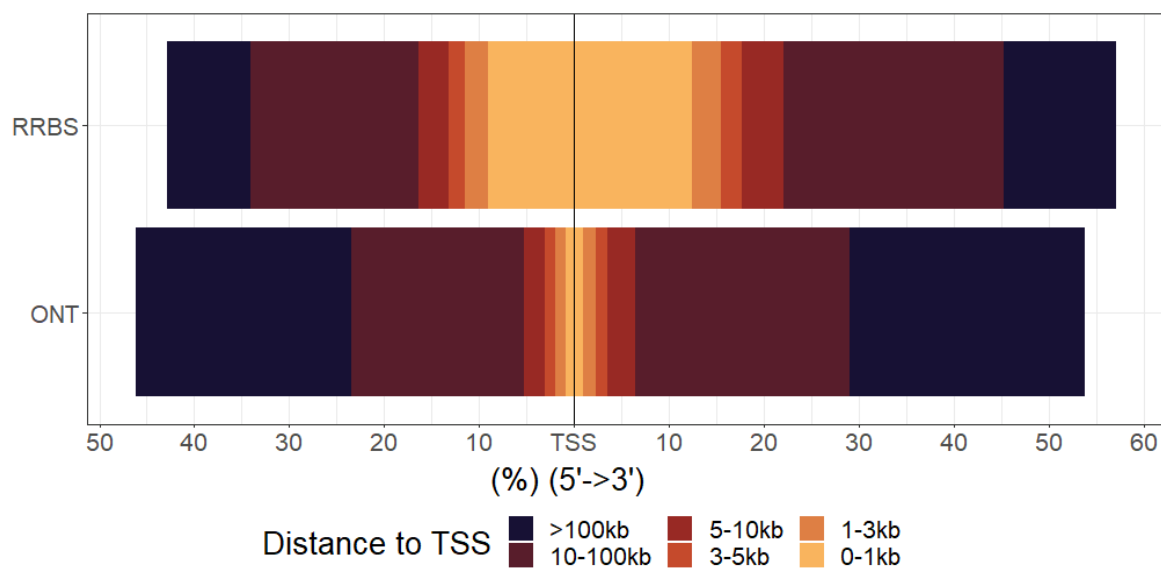


Figure 3-4. Barplots representing the distance to the TSS of the detected positions at 4x moving upstream to its left and downstream to its right.

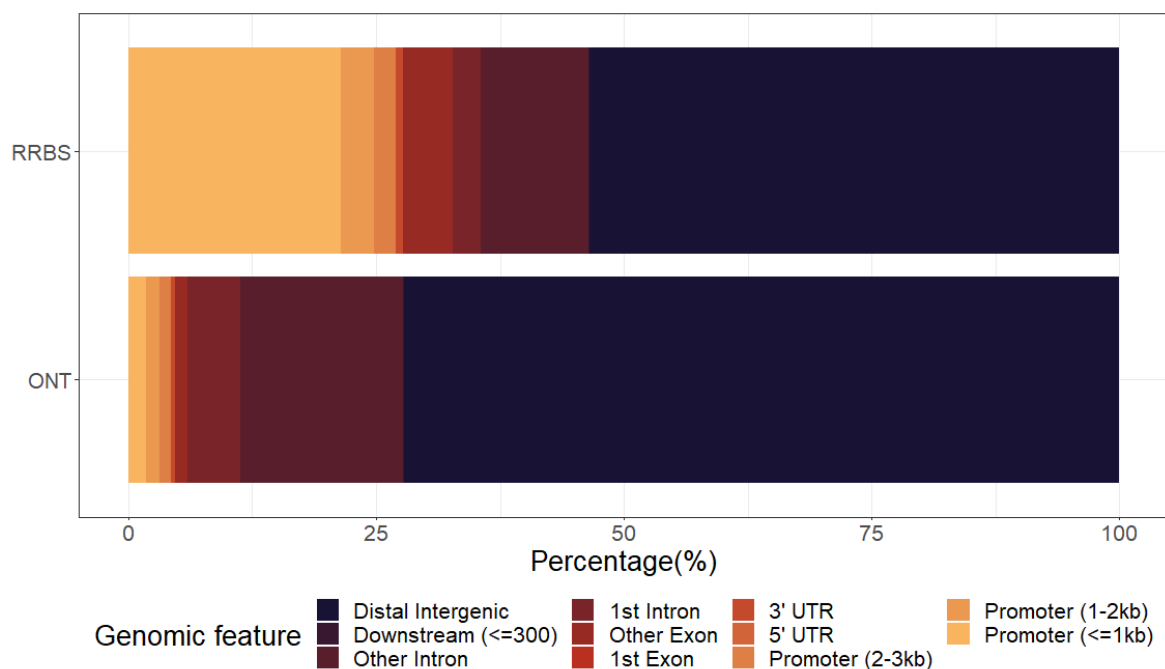


Figure 3-5. Barplots representing the features where the CpGs are detected at 4x, considering promoter regions those closer than 3kb to the TSS.

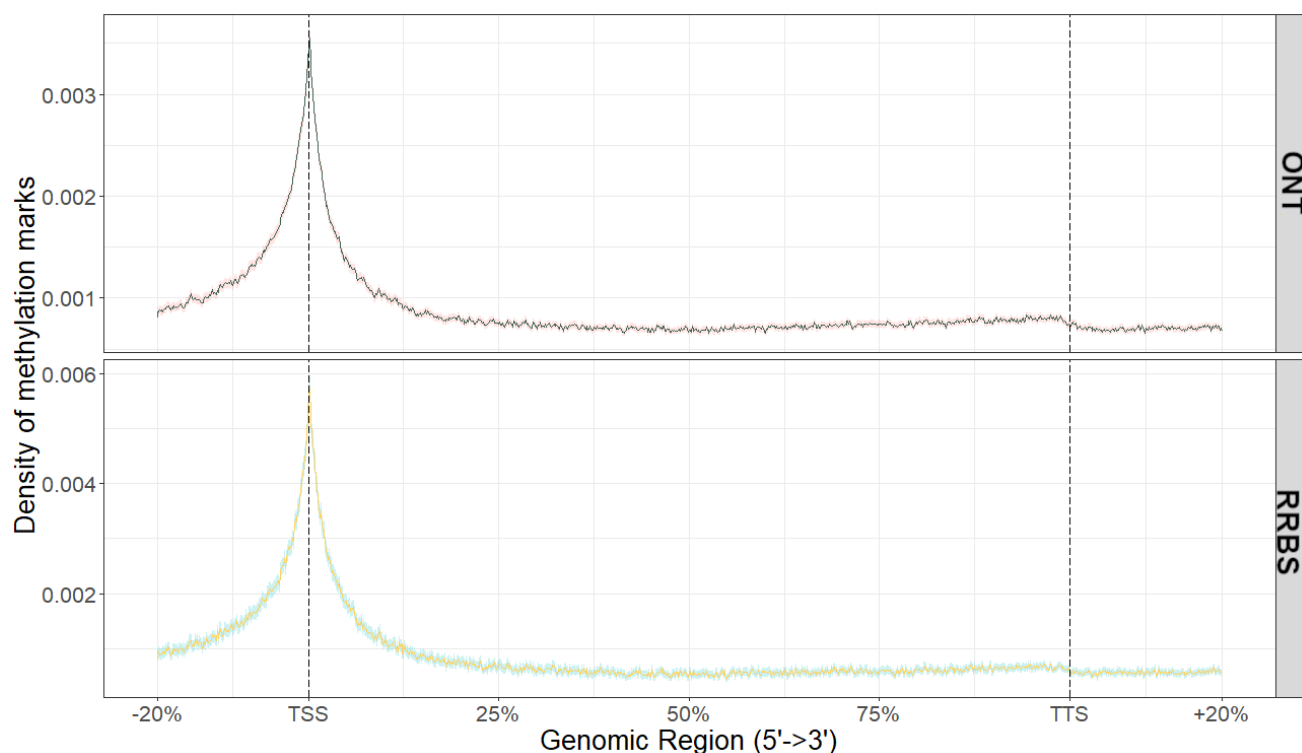


Figure 3-6. Density of the CpG positions detected by ONT in the upper side of the figure and RRBS in the bottom. The genomic region is windowed to depict the transcription start and termination sites (TSS and TTS) and the gene body (region in-between the TSS and the TTS).

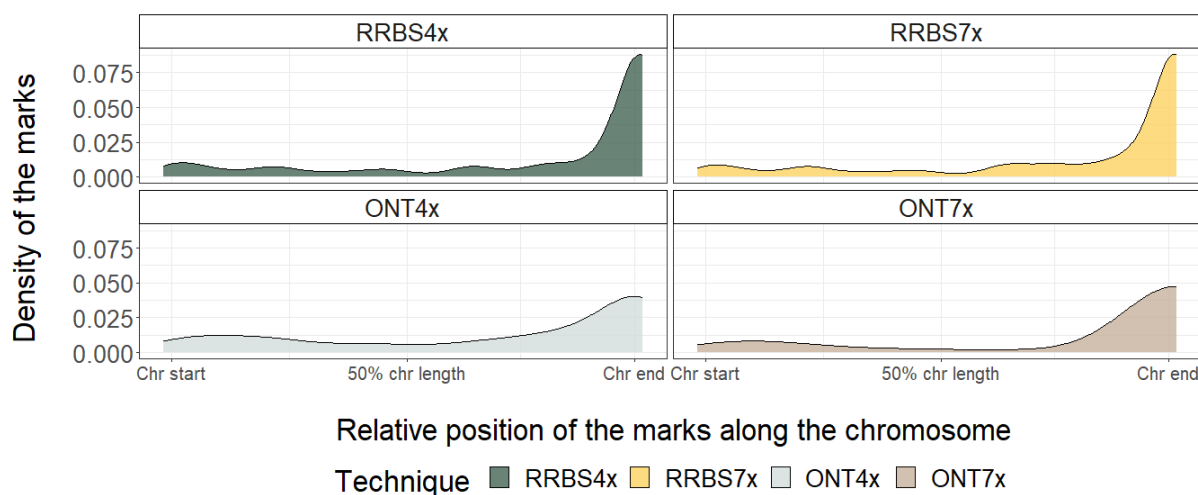


Figure 3-7. Density plots showing the position of the DMCs along the chromosomes detected by the different technologies and coverage filters.

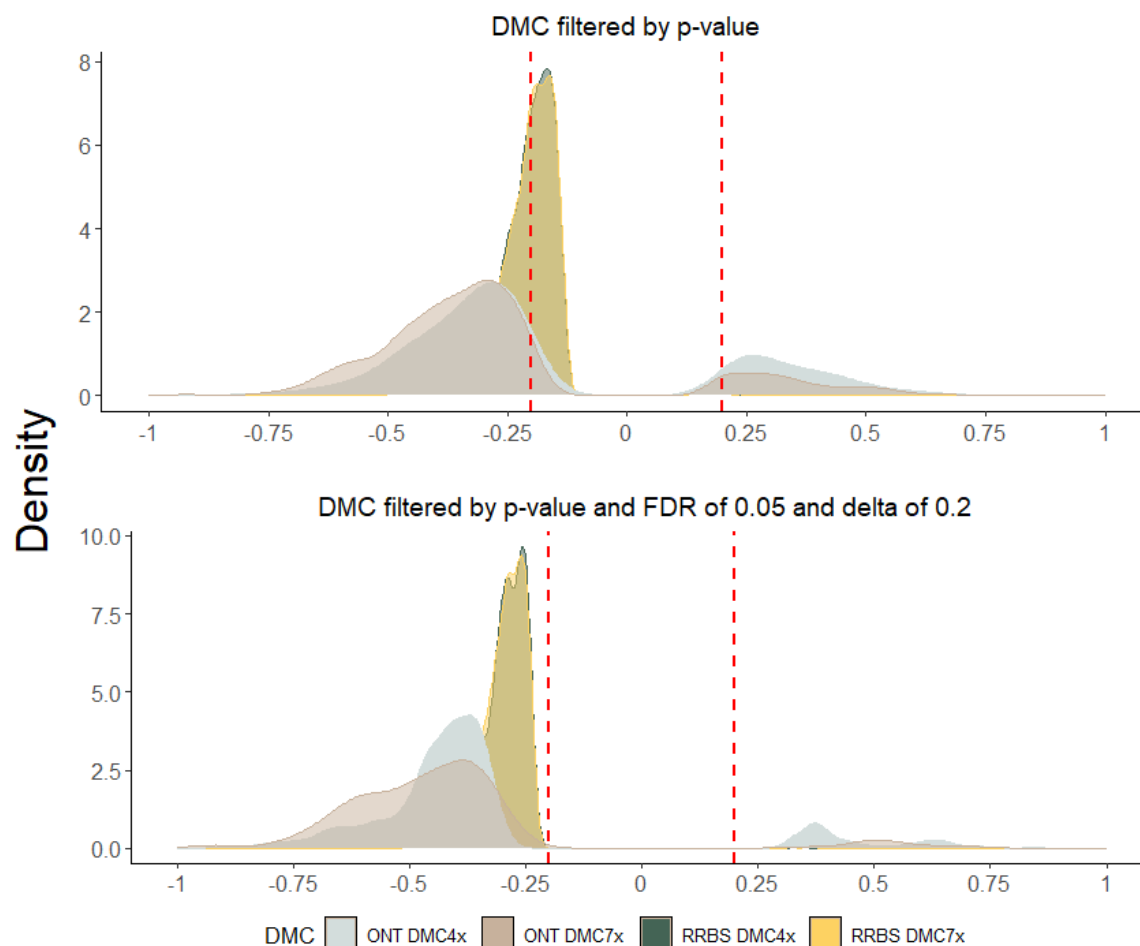


Figure 3-8. Density plot for the DMCs detected by ONT and RRBS filtering at 4x and 7x. Plot on the top illustrates the density of the DMCs filtered by a p-value of 0.05 while the one on the bottom has two additional filters. DMCs were only kept when having an FDR ≤ 0.05 and a delta > 0.2 . Each colour represents a different sequencing technique and a different coverage threshold.

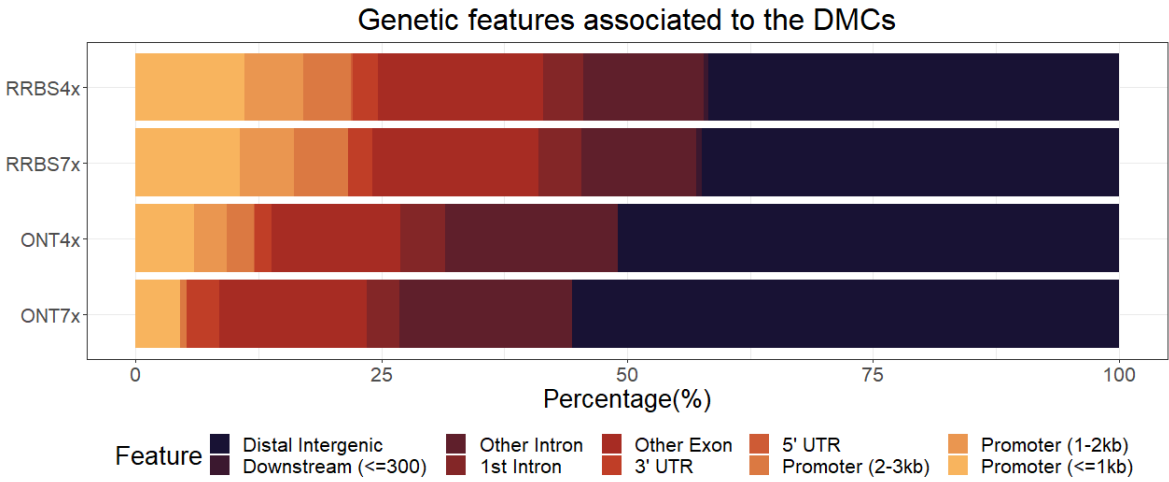


Figure 3-9. Distribution of the genetic features associated with the DMCs detected by RRBS, and ONT at a minimum coverage of 4x and 7x.

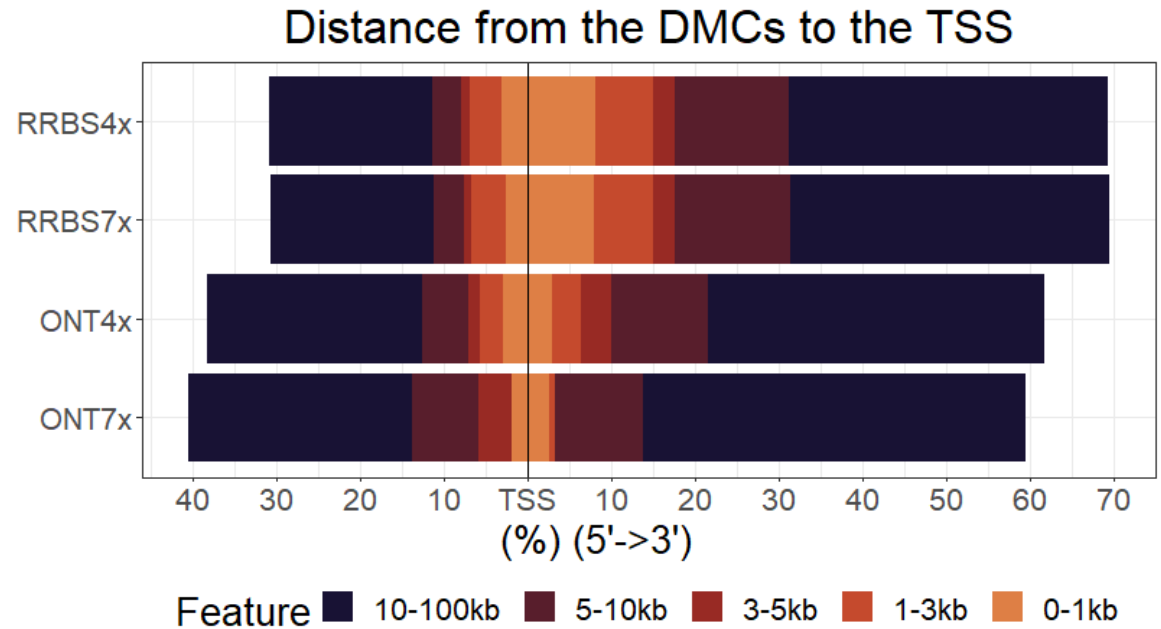


Figure 3-10. Distance of the DMCs detected by each technique to the closest TSS.

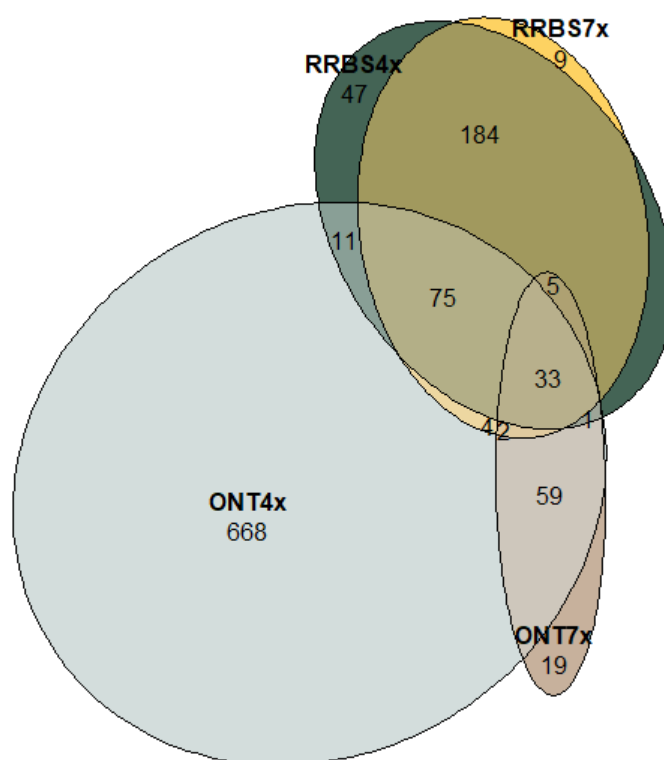


Figure 3-11. Euler diagram to represent the overlap of the genes detected by each technique. In grey, ONT by a minimum coverage depth of 4, in beige ONT by 7x. In green and yellow are the DMC-associated genes detected by RRBS4x and RRBS7x respectively.

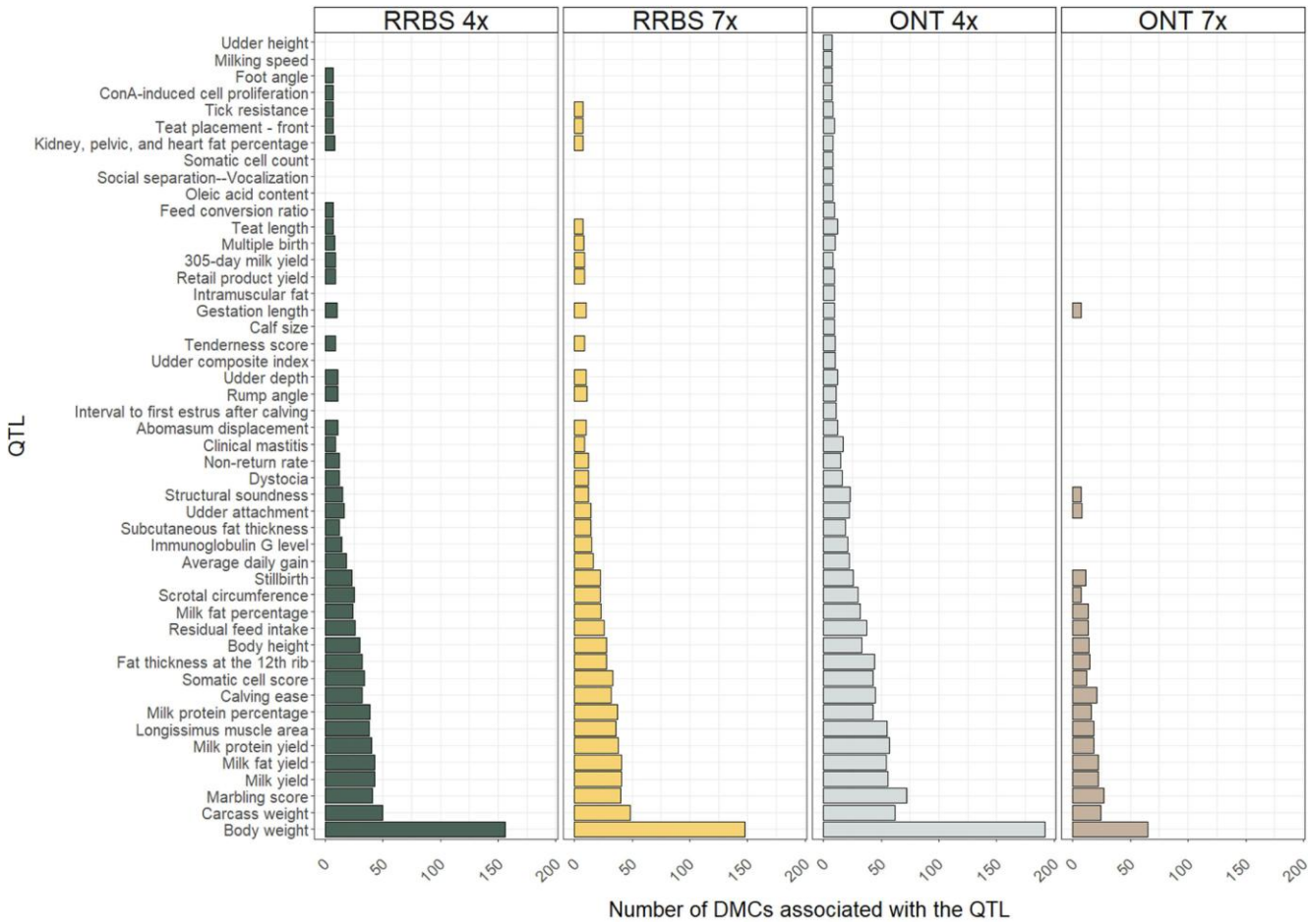


Figure 3-12. Barplot illustrating the number of DMC in QTLs associated to traits of interest. Each column represent a technique and coverage, and rows represent the QTL to which the methylation mark can be associated.

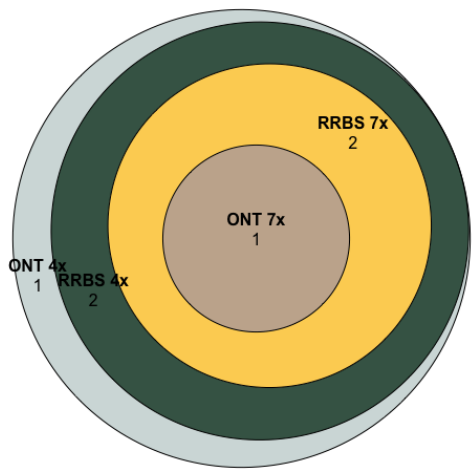


Figure 3-13. Euler diagram representing the overlap between the detected QTLs associated to bull fertility by each technique and coverage. The values represent the traits that have been detected by each technique/coverage.

3.12. 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.
- Halusková J., Holecková B., Stanicová 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., Prezelin, 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.

Schubeler, 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.

4. Chapter 2 - Stress-Induced Epigenetic Effects Driven by Maternal Lactation in Dairy Cattle: A Comethylation Network Approach

López-Catalina, A., Reverter, A., Alexandre, P. A., Nguyen, L. T., and González-Recio, O. (2024). Stress-induced epigenetic effects driven by maternal lactation in dairy cattle: a comethylation network approach. *Epigenetics* 19(1).
doi: 10.1080/15592294.2024.2381856

4.1. Abstract

Epigenetic marks do not follow the Mendelian laws of inheritance. The environment can alter the epigenotype of an individual when exposed to different external stressors. In lactating cows, the first stages of gestation overlap with the lactation peak, creating a negative energy balance that is difficult to overcome with diet. This negative energy balance could affect early embryo development that must compete with the mammary tissue for nutrients. We hypothesise that the methylation profiles of calves born to nonlactating heifers are different from those of calves born to lactating cows.

We found 50,277 differentially methylated cytosines and 2,281 differentially methylated regions between these two groups of animals. A comethylation network was constructed to study the correlation between the phenotypes of the mothers and the epigenome of the calves, revealing 265 regions associated with the phenotypes.

Our study revealed the presence of DMCs and DMRs in calves gestated by heifers and lactating cows, which were linked to the dam's lactation and the calves' ICAP and milk EBV. Gene-specific analysis highlighted associations with vasculature and organ morphogenesis and cell communication and signalling. These findings support the hypothesis that calves gestated by nonlactating mothers have a different methylation profile than those gestated by lactating cows.

4.2. Background

Breeding programs in animal production aim to increase both productivity and health-related traits. In dairy cattle, improvements in phenotypes of interest have been achieved due to the adoption of genetic selection and artificial selection (e.g. milk, protein and fat yields, fertility or welfare) (Bouquet and Juga, 2013). The genomic era provided new tools to further improve the statistical models used in animal breeding, increasing the accuracy of estimated breeding values (EBVs), even for young individuals who have not yet produced offspring (Scheifers and Weigel, 2012). The more information available, the fewer errors are made when selecting the parents of the next generation (Hayes et al., 2009). However, EBV accuracy is far from perfect yet, as statistical models cannot totally capture the complexity of biological systems, including the effect of environmental forces.

For instance, the environment surrounding an individual can modify their phenotype and that of their offspring via methylation marks (González-Recio, 2012). These external signals are usually stressors that put animal welfare at risk (Jensen, 2014). These acquired marks can persist throughout an individual's life (Hunter, 2012) and into adulthood and thus their study is relevant in the animal breeding field. In lactating dairy cows, the peak of lactation overlaps with the first stages of foetal development of subsequent gestation (González-Recio et al., 2012). During lactation, the cow mobilises large energy reserves, and the food intake does not provide the necessary energy to support both lactation and pregnancy. This leads to a negative energy balance that may challenge the metabolic demands for gestation, posing an external stressor for the embryo/foetus and potentially altering its epigenetic profile (Tremblay et al., 2018). This phenomenon does not occur in heifers (i.e., animals in their first pregnancy), as they have not undergone a previous lactation.

Some epigenetic modifications, such as methylation, involve a set of genetic features that do not follow the Mendelian laws of inheritance (Ospelt, 2022). In mammals, DNA methylation consists of the addition of a methyl group at the fifth carbon of a cytosine, creating 5-methylcytosine (5mC) (Miranda and Jones, 2007). This process is usually related to gene silencing, as it inhibits the binding of transcription factors (TFs) to promoter regions, preventing the start of DNA transcription into RNA (Jjingo et al., 2012). These marks are of crucial importance in cell differentiation and embryogenesis and are involved in the determination of cell types (Curradi et al., 2002) and imprinting (Kelsey and Feil, 2013). Nevertheless, the role of methylation marks in gene coding regions remains uncertain, as they are associated with both the protection of the genomic region from mutations and gene activation (Muyle et al., 2022; Zilberman, 2017). Methylation marks act on the DNA either silencing gene expression when appearing on promoter regions or enhancing it when overlapping with the gene body.

Methylation marks are just one mechanism that modulates phenotypes. Other regulators of gene expression include histone modifications, chromatin remodelling, and RNA-mediated silencing (Rodríguez-Orsorio et al., 2012). Histones, proteins in chromatin, undergo various posttranslational modifications that affect their interactions with nuclear proteins and DNA. During oocyte maturation, low acetylation levels maintain precise gene expression. In early embryo development, chromatin remodelling ensures proper nuclear architecture

for cellular differentiation. Before fertilization, male chromatin decondenses, while maternal chromatin remains stable but marked for replication and transcription. Gene silencing also occurs at the mRNA level via noncoding RNAs (ncRNAs), which include long ncRNAs (over 100 nt) and short ncRNAs (under 100 nt) (Brosnan and Voinnet, 2009). Long ncRNAs, like Xist RNA, ensure gene dosage balance (Ng et al., 2007), while short ncRNAs include siRNAs, miRNAs, and piRNAs, each playing roles in gene silencing and maintaining genome integrity (Brosnan and Voinnet, 2009; Iwasaki et al., 2015). In this work, we mainly focused on the silencing role of methylation marks, but these modifications also act as means of protecting the gene sequencing by hindering mutations or sequence alterations and upregulating gene expression (Ball et al., 2009; Rauch et al., 2009).

Recent advancements in technologies for detecting methylation marks have provided an opportunity for the animal breeding industry to complement genomic selection. Epigenetic information may increase the accuracy of genomic predictions of breeding values. If these marks persist across generations, some of them might be in linkage disequilibrium with SNPs and could be treated as (epi)genetic variants (Mccarthy and Hirschhorn, n.d.) that can be incorporated into genomic relationship matrices and statistical models. Those that are not associated with SNPs may provide new information and explain part of the phenotypic variability in the population.

One of these advances is the detection of new epigenetic biomarkers (González-Recio, 2012). Oxford Nanopore Technologies (ONT) enables the simultaneous identification of a sample's genotype and epigenotype by generating FAST5 files (a native data format provided by nanopore sequencers) that capture the electric changes in the pore. These files are later transformed into nucleotide sequences (FASTQ files) through a process called basecalling. The sequences of nucleic acid patterns have a unique electrical profile that can be detected during the basecall process (Deamer et al., 2016). Methylation marks in CpGs change the original profile of cytosines, making them detectable by ONT (Simpson et al., 2017).

Based on these premises, we hypothesise that the methylation profile of calves born to nonlactating heifers may differ from that of calves born to lactating cows. Studying differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) could lead to the discovery of potential biomarkers of metabolic stress in calves and its implementation in breeding programs. Additionally, we inferred a novel comethylation network based on the partial

correlation and information theory (PCIT) algorithm (Reverter and Chan, 2008) to identify clusters of DMRs that share biological pathways, allowing us to reduce the dimensionality of the data by focusing on differentially methylated genes (DMGs).

4.3. Methods

4.3.1. DNA Extraction

Blood samples were drawn from six newborn Holstein-Friesian calves from a commercial farm by jugular venepuncture. Three of them were gestated by nonlactating heifers, and the remaining 3 were from multiparous cows during their third pregnancy and second lactation. All cows and calves were unrelated. Samples NL1, NL2 and NL3 were from calves born to nonlactating cows (NL), and L1, L2 and L3 were from lactating cows (L).

DNA extraction was conducted using the Monarch ® HWM DNA Extraction Kit for Cells & Blood (New England Biolabs®). This kit was selected to obtain long DNA fragments and improve the sequencing yield. To further increase the sequencing yield, the DNA was fragmented to a length of 8 kilobases using a Covaris g-TUBE® (Covaris 520,079) to centrifuge the sample at 8,600 rpm, as recommended by the manufacturer. The purity and concentration of the samples were measured using a Qubit Fluorometer (Thermo Fisher Scientific).

4.3.2. DNA Sequencing

Sequencing of the samples was performed using a GridION sequencer (Oxford Nanopore Technologies. ONT, Oxford, UK). Libraries were prepared starting from 3 µg of DNA following the protocol from the SQK-LSK110 kit provided by Oxford Nanopore Technologies. Each sample was sequenced individually in an R9.4 flow cell. Basecalling was performed using Guppy version 6.3.7 (https://community.nanoporetech.com/docs/prepare/library_prep_protocols/Guppy-protocol/v/gpb_2003_v1_revao_14dec2018/linux-guppy) with the model dna_r9.4.1_e8.1_modbases_5mc_cg_sup.cfg. Quality control of the samples was conducted using NanoPlot (v1.36.2) (De Coster et al., 2018) to ensure that the Phred score and mean length of the samples passed the quality threshold.

4.3.3. Methylation Calling

The analyses were performed using f5c version 1.1 in the Bracewell CSIRO HPC (Australia). FAST5 and FASTQ files were indexed using the f5c index function and then aligned to the reference genome Bos_taurus.ARS-UCD1.2.fa by minimap2 (Li, 2018). Files were sorted using samtools v1.16.1 (Li et al., 2009) to generate a bam file. The candidate sites were extracted from the FASTQ files using the function call_methylation from the f5c software. Then, the methylation frequency was determined using f5c's meth-freq. The distribution of the methylation frequencies obtained by f5c (Gamaarachchi et al., 2020) of each sample was estimated using methylkit v3.16 (Akalın et al., 2012). The ensembl genesBtau.gff3 annotation file (https://ftp.ensembl.org/pub/release-111/gff3/bos_taurus/) was used to estimate the methylation frequency of the CpGs detected in each sample in regions close to the transcription start site (TSS) and transcription termination site (TTS).

4.3.4. Differential Methylation Analyses

The R (version 4.2.1) (R Core Team, 2021) package DSS v2.46.0 (Park and Wu, 2016), which is specifically designed for bisulphite-sequenced data, was used to estimate differentially methylated regions. The methylation files generated from f5c were formatted to meet the requirements of DSS. This involved retaining only the chromosome, the position of the CpG methylation site, the sequencing depth of that CpG site, and the number of methylated reads. Only the sites that were present in all samples were retained, 12,160 CpGs were hence included in the following analyses. Files were filtered to only retain sites with CpGs with a sequencing depth greater than 7X, independent of their methylation status. The methylation object was then created using the function makeBSseqData, and the differential methylation at each CpG was calculated using DMLtest, manually selecting the groups to which each sample belongs. The smoothing option was set to TRUE, as recommended for genome-wide analyses. The DMCs were then called using the callDML function, setting a p-value threshold of 0.05. Delta, the minimum value for the difference between the estimated means of each group, was set at 0.2. The remaining sites were then filtered by a false discovery ratio (FDR) of 0.05. The same procedure was followed for the DMRs using the callDMR function. CallDMR detects regions with statistically significant CpGs and merges the position into regions if there are at least 3 CpGs within 50 base pairs.

Both DMCs and DMRs were separated into hypermethylated or hypomethylated features using calves born from NL heifers as a reference (DMR hypomethylated, DMR hypermethylated, DMC hypomethylated and DMC hypermethylated). Thus, we referred to a hypomethylated DMC/DMR when the methylation frequency was lower in the genome of calves from nonlactating heifers and hypermethylated when the position had a methylation frequency greater in this group than in the calves born from lactating cows. The genomic positions were annotated to the closest gene/microRNA with a maximum distance allowed of 50 kb (Porto-Neto et al., 2014). Regions that appeared associated with both hyper and hypomethylated genes were excluded to facilitate interpretation.

4.3.5. Permutation of the Groups

Permutation analyses were implemented to check for spurious associations. The differential methylation analyses were implemented on each of the 20 possible combinations of each sample belonging to the non-lactating or lactating groups given by the following formula $C_3(6) = \binom{6}{3} = \frac{6!}{3!(6!-3!)} = \frac{120}{6} = 20$ where the samples were interchanged between group of calves gestated by primiparous cows and calves gestated by multiparous cows. The minimum overlap of genes appearing in the four categories of DMC hypo- and hypermethylated and DMR hypo- and hypermethylated was sought, as well as a decreased number of genes overlapping between hypermethylated DMC and hypomethylated DMC.

4.3.6. Gene Annotation and Description of Genomic Regions

Gene annotation was performed using the function `annotatePeak` from the R package `ChipSeeker` (Yu et al., 2015). `Org.Bt.eg.db` (Carlson, 2019) and `TxDb.Btaurus.UCSC.bosTau9.refGene`, which correspond to the *Bos taurus* reference genome ARS-UCD1.2, were selected as annotation packages. A window of 3,000 base pairs both up- and downstream of the promoters was selected as the transcription start site. Each of the subset datasets was annotated individually, and the list of genes associated with those regions was extracted.

4.3.7. Enrichment Analyses

Enrichment analyses were performed on the previously mentioned datasets using the `compareCluster` function from the R package `clusterProfiler` (Wu et al., 2021) adjusted by the Benjamini–Hochberg procedure to control the false discovery

rate with a p-value cutoff of 0.05. The function “enrichGO” was applied to perform a gene ontology over-representation test.

To determine whether the significant regions are associated with genes involved in biological processes related to embryo development, well-being, or phenotypic traits of interest in the farming industry, an enrichment analysis was performed on the genes identified by the PCIT algorithm. Enrichment of the networks was performed using g:Profiler (Raudvere et al., 2019), the statistical domain scope was set as “only annotated genes” and the significance threshold as “g:SCS threshold” with an adjusted p-value threshold of 0.05.

4.3.8. Construction of the Comethylation Network

Windows of 500 bp along the genome were created using pycoMeth (Snajder et al., 2023). The log likelihood-ratio (llr) for methylation probability in each window was estimated. Those windows with a llr greater than -3.5 or lower than 3.5 were removed to avoid ambiguous sites. Missing data were not allowed (e.g., windows must appear in all 6 samples to be kept for the analyses).

The lactation status of the dam was included as a binomial explanatory variable to indicate whether the dam was lactating during gestation. The calves’ milk EBV and the genetic indices for body capacity index (ICAP) were also included to study their effect on the methylation status. The ICAP index incorporates traits like stature, chest width, body depth, and rump width into economic indexes by assigning them appropriate weights. Meanwhile, the EBV, or estimated breeding value, compares an animal's genetic merit for a specific trait to the population's average genetic value, enabling the ranking of animals according to their genetic potential. Both values were included in this work due to their significance in assessing and improving livestock genetics. All traits were provided by CONAFE, the Spanish Holstein Association, obtained during their genomic evaluations.

The relationship between the remaining regions was estimated using a PCIT analysis (Reverter and Chan, 2008). The PCIT algorithm consists of two main steps: (1) the partial correlations are estimated in every trio of nodes (genes, phenotypes or DMRs) to determine the strength of the linear relationship when the first node pair is uncorrelated to the third node, and (2) the Data Processing Inequality theorem of Information theory is used to calculate the tolerance level and discard the nonsignificant associations. An association was considered significant when the correlation of the first two genes of the gene was greater than

the tolerance level. Regions with a correlation greater than or equal to 0.75 were retained for subsequent analyses.

4.3.9. Visualisation of the Network

The correlation matrix was loaded into Cytoscape (Shannon et al., 2003), and the yFiles organic layout was used to visualise the correlations. This algorithm follows the force-directed layout approach, which handles the nodes as physical objects with repulsive forces and the edges (connections between the nodes) as springs, exerting forces in both directions. The aim is to minimise the forces exerted by both edges and nodes. Regions were annotated and named as the closest gene to them. Nodes are coloured according to the status of that region in the differential methylation analyses; hypermethylated regions were coloured in yellow, and hypomethylated regions were coloured in purple. The size of the gene-containing nodes is determined by the proximity to promoter regions and is greater if the region is annotated to a window of -3,000 to 3,000 bp to the promoter of the gene.

Phenotypic subnetworks of the first neighbours of lactation, EBV and ICAP, were extracted, and the nodes were separated according to their methylation status and their correlation with the phenotype.

4.4. Results

4.4.1. Sequencing Results

The sequencing yields are shown in Supplementary Table 4-2 and a summarization in Table 1. The mean length was ~8 kb, as expected and the Phred score was >11 for every sample.

A bimodal U-shaped distribution skewed towards 0% and 100% was observed for the methylation frequency. The diagonal of Figure 4-1 depicts these profiles, showing a tendency towards a 100% or 0% methylation frequency, with intermediate frequencies not being prevalent. The methylation profiles of the six samples were similar. The methylation pattern showed strong agreement between samples (upper triangle Figure 4-1). However, samples from NL calves showed slightly greater agreement between their methylation profiles ($0.94 < r < 0.95$) than did those from L calves ($r = 0.93$).

The methylation frequency in these regions was estimated for each of the 17,416 genes in the annotation file. It can be observed (Figure 4-2) that the methylation frequency decreased at positions closer to the TSS to allow gene expression. The values increased in the in-body gene region (from the TSS to the TTS) and decreased to approximately 60% from the TTS.

4.4.2. Differential Methylation Analyses

A total of 50,277 DMCs were detected showing a bimodal distribution in their delta values. This bimodality occurred as the delta values approached -1 or 1, reflecting whether the control group (nonlactating dams) exhibited a lower or higher estimated methylation mean compared to calves born from lactating cows. A delta value of -1 indicates complete unmethylation in the daughters of nonlactating heifers (NL samples) and complete methylation in calves gestated by lactating cows (L samples). Specifically, a hypomethylation pattern was observed in 54% of the DMCs in calves born from nonlactating heifers, while 46% of the DMCs displayed hypermethylation (Figure 4-3). Overall, the within chromosome distributions followed those of the genome-wide distribution, with the exception of chromosomes 24 and 28, which were hypermethylated in the nonlactating group, and chromosomes 8, 19, 25 and 29, which showed a greater proportion of hypomethylated sites (Supplementary Figure S4-8).

In terms of DMRs, 2,281 were detected with FDRs lower than 0.05 and a delta of 0.2 (Figure 3). On average, 59% of the regions were hypomethylated in the nonlactating group, and the remaining 41% were hypermethylated. At the chromosome level, chromosomes 20 and 24 seemed to be hypermethylated compared to those in the lactating group. Chromosomes 1, 8, 10 13, 14, 18, 19 23 and 29 appeared to be hypomethylated (Supplementary Figure S4-9).

A total of 4,704 genes were found to be associated with hypermethylated DMC; 4,658 with hypomethylated DMC; 859 with hypermethylated DMR and 1,119 with hypomethylated DMR. A Venn diagram (Figure 4-4) was generated to depict the overlap of the annotated genes. The four DMC and DMR groups shared 168 genes, but a total of 1,609 genes were associated with both hypermethylated and hypomethylated regions simultaneously. We found 1,784 genes associated with hypermethylated DMCs and 1,584 associated with hypomethylated DMCs.

Twenty permutations were drawn to cover the whole set of possible combinations for the six samples (Supplementary Figure S4-10). The scenario

where each animal was correctly assigned to its group, and its mirror image (with groups interchanged) had the lowest percentage (2.5%) of overlapping genes for DMR/DMC hypermethylation and hypomethylation. The total of overlapping genes was 168 (ranging from 175 to 283). Regarding hypoand hypermethylated DMCs, both combinations also had the lowest percentage of overlapping with 23.7% (1,609 genes), with a range of permutations from 1637 to 1920 in the permutation analyses. The Venn diagram depicts an overlap of the genes associated with different methylation statuses (DMC/DMR hypomethylated and DMC/DMR hypermethylated), which is due to the process of annotating the genomic positions to the closest gene. Two positions close to each other with different methylation statuses can be annotated to the closest gene by ChIPSeeker. A cytosine or region cannot be both hypoand hypermethylated at the same time, but a gene can be identified by two proximal regions, one hypoand the other hypermethylated. The permutation analysis supported the results from the association analyses, as permuted combinations increased the percentage of overlap between the four groups. As overlap can be interpreted as noise, the noise is increased for the permuted groups. For the same reason, the removal of overlapping genes is recommended in the quest for potential biomarkers.

After annotating DMCs and DMRs, we observed that a greater proportion of hypomethylated cytosines/regions in the nonlactating heifer group appeared in promoter regions (Figure 4-5). However, hypermethylated regions were more common in distal intergenic regions. Hypomethylated promoter regions in the nonlactating group may suggest an association with greater transcriptional activity of those genes compared to calves born to lactating cows.

The enrichment analysis of the genes associated with the set of DMCs showed that hypermethylated DMCs in NL calves were associated with biological processes involving anatomical structure, organ and cell morphogenesis and blood vessel and vasculature development. On the other hand, hypomethylated DMCs in NL calves were associated with genes involved in the positive regulation of stimuli, signals and communication as well as neural crest cell differentiation (Figure 4-6).

4.4.3. Comethylation Network

A total of 4,367,996 genetic windows, each spanning 500 base pairs, were identified by pycoMeth. Among those 19,722 regions were present in all the 6 samples, but this number decreased to 3,006 after filtering the methylation

frequency for $|\text{llr}| \geq 3.5$. This list served as the input for the PCIT algorithm, resulting in the estimation of 4,516,515 correlations, with 181,547 (4.01%) being statistically significant. To simplify the network, an additional file was created to retain the nodes with a correlation ≥ 0.7 (in absolute value). The enrichment analysis of the genes associated to these regions are shown in Supplementary Table 4-3. The enrichment analysis revealed a spectrum of enriched Gene Ontology (GO) terms. Noteworthy findings include the prevalence of terms associated with developmental processes, such as ‘multicellular organism development’, ‘anatomical structure development’, and ‘system development’. Additionally, terms related to cellular components and molecular functions, such as ‘cytoplasm’ and ‘protein binding’, were also found. The identification of terms related to both positive and negative regulation of biological processes highlights the regulatory networks governing developmental pathways. Also, the enrichment of terms such as ‘neurogenesis’, ‘nervous system development’, and ‘generation of neurons’ suggests a particular emphasis on neurodevelopmental processes. Furthermore, the presence of terms related to cell–cell communication and signalling underscores the importance of intercellular interactions in coordinating developmental events.

Supplementary figure S4-11 depicts the whole comethylation network. Figure 4-7 shows the first neighbours of lactation status of the dam (Figure 4-7a), ICAP (Figure 4-7b) and the calves’ milk EBV (Figure 4-7c). The calf EBVs for milk were associated with 124 regions, 114 regions were associated with ICAP, and 27 were associated with lactation status. The enriched genes found in the lactation status network were *ALCAM*, *ELF1* and *P2RY1*, in the ICAP network were *PAK1*, *SYT1*, *MYC*, *ATP6V0D2*, *COL1A2* and *HAS2* and in the milk EBV network were *KCNK1*, *MET*, *CIBAR1*, *PAK1*, *INHBA*, *SLF1*, *CHRM3* and *CIBAR1*.

4.5. Discussion

The study of the methylome in livestock populations could improve the accuracy of the prediction of phenotypes of interest by explaining alterations in gene expression. In this work, we proposed a comethylation network approach to determine how dam milk production impacts the methylation pattern of offspring. We also propose a relationship between the methylation status of the animal and its breeding values for body conformation and milk yield. This approach provides specific information at the molecular level, proposing new markers associated with traits of relevance in the livestock field. This information could be implemented in

future genetic evaluation following a quantitative approach to add valuable information to breeding programmes.

The proportion of methylation marks was similar among the six samples. A large proportion of the sites were close to complete methylation or unmethylation. This methylation pattern has already been reported in humans (Bell et al., 2010; Mao et al., 2021) and other animals (Liu et al., 2022) epigenetic studies where a U-shaped distribution was observed in the methylation frequency of the CpGs throughout the genome. The strong correlation between the methylation profiles of the six samples (Pearson correlation = 0.93-0.95) suggest that there is a population-specific methylation, which may reduce the inter-sample variability and hamper the detection of DMCs/DMRs.

The methylation frequency profile followed a similar distribution to that of other superior eukaryotes. A sudden drop closer to the TSS and promoter regions to avoid gene repression can be observed and an increase in methylation frequency up to 80% appears in the gene body regions (Zemach et al., 2010), as gene-body methylation is associated with increased gene transcription, and methylation in promoter regions is associated with gene silencing, which is necessary for the correct establishment of different cell lines (Bender et al., 1999; Greenberg and Bourc'his, 2019; Lewis et al., 1992).

We found differences in the distribution of the marks that were hypo- or hypermethylated in calves gestated by nonlactating heifers. Calves gestated by non-lactating cows present a larger proportion of hypomethylated DMCs and DMRs than L calves. Hypomethylated positions in the calves born from nonlactating cows seem to be closer to the TSS in a larger proportion than hypermethylated positions, meaning that these genes are most likely to be more highly expressed in the calves born from nonlactating heifers than in the calves born from lactating cows, although transcriptomic analyses would be necessary to validate this finding.

Previous publications have hypothesised that the productive level of mothers could impact the future performance of their daughters. Individuals whose initial stages of fetal development occurred without maternal lactation produced 52 kg more milk during their first lactation and lived 16 days longer than the average cow born to mothers who conceived while lactating (González-Recio et al., 2012). There were differences between the methylation pattern in blood samples of calves born to nonlactating Holstein heifers and lactating dams appeared when their feed

was supplemented with methyl donors (Bach et al., 2017). Bach et al. found no differences between the global methylation profile of calves born to non-lactating heifers and born to lactating cows. Regardless of this, they found 11 CpGs at promoter regions and 112 at gene region hypermethylated in calves gestated by heifers and 14 CpGs in promoter regions and 71 CpGs in gene regions hypomethylated in these animals. In our study, we identified 50,277 DMCs and 2,281 DMRs with a FDR of less than 0.05 and a delta greater than 0.2. Interestingly, 54% of the DMCs were hypomethylated in calves born from nonlactating cows. This proportion increased to 59% when considering the DMRs. These findings suggest a trend towards hypomethylation in the nonlactating group, which could have implications for gene expression and subsequent developmental processes. This hypomethylation pattern was observed across almost all chromosomes, indicating a widespread epigenetic alteration. However, exceptions were noted on chromosomes 20, 24, and 28. Specifically, chromosome 28 exhibited hypermethylation at DMCs for NL calves, chromosome 20 showed hypermethylation at DMRs, and chromosome 24 displayed hypermethylation at both DMCs and DMRs (Supplementary Figure 4-8). Our results aligned with the ones in the mentioned study, as we propose new DMC and DMRs associated to the effect of gestation and lactation and indicate that there are indeed differences in the methylation profile of calves gestated by non-lactating heifers and those gestated by lactating dams.

Our research revealed that genes linked to differentially methylated cytosines between NL and L animals showed significant enrichment for processes such as morphogenesis of anatomical structures, organ and cell morphogenesis, and vasculature development, particularly in cases of hypermethylated DMCs in NL calves. These processes are crucial in embryonic development, as they orchestrate the spatial and structural rearrangement of tissues necessary for precise organ formation (Scarpa and Mayor, 2016). Conversely, hypomethylated genes in NL were predominantly associated with the positive regulation of signaling pathways, response to stimuli, and cell communication. These findings highlight substantial differences in the methylation profiles between calves born to nonlactating cows and those born to lactating cows. The detected DMCs appear to be intricately linked to key developmental processes, which could have significant implications during early life stages and potentially influence adult physiology. Previous research demonstrated that metabolic deficits can adversely affect embryo development (Chaput and Sirard, 2020), and differences in the productivity of cows

born to nonlactating versus lactating cows have been observed (González-Recio et al., 2012) with NL producing more during longer lifetime. This suggests that variations in the methylation profiles of calves gestated by primiparous and multiparous cows may directly influence gene expression during embryogenesis. Identifying and modulating these stressors could be crucial for optimizing gestation conditions and ensuring proper embryonic development, but determination of the effects of this hypo/hypermethylation on gene expression needs to be determined first to act accordingly. By understanding and potentially correcting these epigenetic modifications, we can aim to secure the expected gene expression patterns, ultimately improving developmental outcomes and the productive life of the offspring.

One key novelty in our study was the use of a comethylation network that allowed us to identify key clusters reflecting the 3-way dynamics of methylation status, enriched biological process, and phenotype. Despite the limited sample size, which does not provide large statistical power and where large correlations are likely to appear by chance, the information theory approach embedded in the PCIT algorithm resulted in 4.01% of the 4,516,515 correlations being deemed significant. Importantly, the enriched network genes were associated with different developmental processes. This finding aligns with the previously discussed enrichment of the genes associated to the DMCs found between the two groups of animals.

The construction of a network opposed to the use of an Epigenome-wide association study (EWAS) or any other statistical method was based on the small sample size of this study. EWAS, as well as GWAS, needs a large consortium-based approach to reach a large number of individuals to obtain a robust statistical power (Flanagan, 2015). Most of available EWAS publications relied on data obtained by Illumina's arrays, biased for coding and promoter regions (Bultmann and Stricker, 2018). Using a network approach based on the PCIT algorithm poses an alternative to EWAS to characterize genetic networks associated with epigenetic information. Co-methylation networks generate a novel approach of deciphering epigenetic interaction of different genomic regions and productive traits (Marina et al., 2020).

Furthermore, our comethylation networks allowed us to visualise not only the correlation between methylated regions but also their interaction with traits of interest in animal husbandry, like EBVs or phenotypes or indexes. Networks with the direct neighbours of these traits are a visual guide for understanding whether a methylation marker affects the phenotype positively or negatively. In addition,

separating the regions into hypo- or hypermethylated regions provides more information on the mechanism of the interaction. In the networks, the way the methylation is affected by the phenotype differs for each trait. The lactation status of the dam can only unilaterally affect the methylation profile of the gestating calf because the embryo's methylation profile does not interact with the maternal phenotype. But the maternal uterine maternal phenotype determines the environment in which the embryo is developing. Conversely, for the calf's genetic values, there could be a two-way interaction since the calf's phenotype is determined by its genotype, which can be influenced by methylation marks. An increase in the phenotype of either the mother or the calf affects the methylation status of a given region if there is a positive correlation with that trait. For instance, *PRG3* appears to be hypermethylated in its promoter region in calves gestated by nonlactating heifers and has a negative correlation with lactation. This means that lactation is associated with a decrease in the methylation frequency of this gene in calves. *PRG3* is involved in the activation of granulocytes and the biosynthesis of histamines and has a strong relationship with the inflammatory response in pathological processes (Gray et al., 2022; Plager et al., 1999). If this methylation is not reversed during adulthood, calves gestated by nonlactating heifers might present a greater methylation frequency and thus lower gene expression of this gene than calves gestated by lactating cows. The gene *LOC100271685* belongs to the membrane-spanning 4-domain protein subfamily and was hypomethylated in its promoter region in calves gestated by nonlactating heifers. These proteins have been described in humans as being associated with the regulation of calcium influx in B-cell antigen receptors and involved in the signalling of different immune response processes (Zuccolo et al., 2010). *RHOF* was hypomethylated in calves gestated by nonlactating heifers and showed a positive correlation with calf milk EBV. This gene is involved in the amino acid transport and synthesis of proteins in milk and is strongly correlated with milk quality (Dai et al., 2018; Wang et al., 2018). The higher the milk EBV of the calf, the less methylated this gene appeared. A larger breeding value allows for a decreased methylation profile in regions associated with this gene, if happening in the promoter, this event would allow for a higher gene expression in NL calves. The ICAP of calves was positively correlated with the methylation status of the *SLC37A3* gene, which is associated with cow fertility and osteogenesis (Mohammadi et al., 2020). Calves with a larger ICAP showed more methylation in adjacent regions to this gene, translating into a decreased gene expression and worsened fertility and bone structure. The *KLHL1* gene was the only gene that

appeared in the three networks and was described as being involved in milk yield and milk quality in dairy cattle (Sun et al., 2019).

The enrichment analysis of the network revealed a variety of Gene Ontology (GO) terms linked to processes such as muscular, nervous, and system development. Terms like ‘multicellular organism development’, ‘system development’, and ‘anatomical structure development’ highlight the coordination of complex biological systems and anatomical structures in multicellular organisms. Additionally, ‘cytoplasm’, ‘protein binding’, and ‘cell junction’ terms suggest underlying cellular and molecular mechanisms. There is a noted emphasis on nervous system development with terms like ‘neurogenesis’ and ‘generation of neurons’. The importance of intercellular communication and signaling pathways is also highlighted. Lactation imposes metabolic stress on the gestated calves, impacting their methylation profiles during development. According to González et al. (González-Recio et al., 2012) these differences impacted the productive life of the calves in their adulthood. Further studies are necessary to determine if these differences persist throughout the animals' lifetimes or if they dissipate and if they could be linked to the phenotypic differences previously observed.

We found genes associated with each of the studied traits that were also enriched in the differential methylation analysis. *ALCAM* was hypomethylated in calves gestated by nonlactating heifers, with a negative correlation with the lactation status of the dam. This gene is involved in neural and branching organ development and haematopoiesis (Swart, 2002). Calves gestated by lactating cows will have an increased methylation status of regions associated with this gene, leading to decreased gene expression and a modification in the expected behaviour of these processes, which is partly explained by the mother's phenotype. Molecular analysis are necessary to understand the way methylation affects the expression of these genes. Nonetheless, if methylation is associated with gene silencing, NL calves would have a more proper embryo development, as L calves organ development is being impacted by these inhibitions. These comethylation networks also provide insight into how the phenotypes of calves are correlated with their own methylation profiles. The methylation of *INHBA* was negatively correlated with the EBV for milk so calves with larger milk EBV have lower methylation in CpGs close to this gene. This gene is associated with growth traits and feed efficiency in cattle (Pierce et al., 2020) and is highly expressed in animals with decreased residual feed intake (Chen et al., 2011), one of the selection objectives in dairy cattle (Stephansen et al., 2023). Calves with a higher EBV will have lower

methylation frequency in this genic region, allowing a increased expression of *INHBA*, and thus an increased growth and better residual feed intake than animals with a lower EBV. We found regions positively correlated with ICAP associated with the gene *COL1A2*, a gene involved in collagen synthesis that is underexpressed in cows with a high marbling score (Chen et al., 2019) and in double-muscled animals (Zhang et al., 2011).

Further studies with larger sample sizes are necessary to confirm that methylation marks in blood could be used as biomarkers to detect metabolic stress in calves using their dam phenotypes. The limited sample size of our study reduces its statistical power, indicating that our findings should be validated in a larger population for greater reliability. Consequently, the findings derived from the PCIT algorithm must be interpreted with caution. In order to deal with the limitation of a small sample size, we performed permutation analyses. This study introduces a novel methodology for investigating methylation marks and their associations with phenotypes, rather than aiming to identify new biomarkers. Future research should focus on the functional validation of these epigenetic changes at the phenotypic level. This involves conducting transcriptomic and gene expression studies to confirm the impact of these changes on various traits. Such validation is crucial for understanding the broader implications of our findings and their potential influence on breeding selection schemes. Verifying the effects of these epigenetic modifications would enhance the accuracy of candidate positions, which would allow new strategies to include epigenetic information in breeding programmes, ultimately improving desirable traits in the population. Interpreting these networks becomes challenging when extending beyond the scrutiny of individual genes at the molecular level. The adoption of a quantitative methodology would help to enhance the understanding and application of these findings. Including methylation information in breeding programmes requires new equations and models that need to be developed to account for the effect of potential methylation biomarkers on phenotypes of interest in animal production. Confounding effects also need to be considered. Nutritional factors could be considered in future studies, as the feed rations of heifers and cows are different, which could be a potential confounding factor. The number of lactations could also impact the methylome, even though in our work, all lactating cows were ~4yrs old and overcame their second lactation.

4.6. Conclusions

We demonstrated the presence of both DMCs and DMRs in calves gestated by heifers and lactating cows. The PCIT algorithm enabled the construction of a comethylation network that directly links the calves' ICAP and milk EBV with their methylome. This comethylation network approach linked the dam's lactation status with the methylation profile of the calves. Genes with differential methylation in L and NL calves were enriched in the following GO terms: vasculature and blood vessel development, cell and organ morphogenesis and different processes of signalling and cell communication. The methylation profile of regions close to these genes would alter the expected expression of genes associated with growth, embryo development and milk yield. Despite the limited sample size, our results suggest that the foetal and early development of calves is affected by the lactation status of the dam, and this may have important implications at rearing. Further research is necessary to integrate this information into traditional breeding models following a quantitative approach and to determine if the differentially methylated cytosines persist during the adult life of the animals.

4.7. Acknowledgements

The authors acknowledge the ideas and help of Juliana Afonso and Laercio R. Porto-Neto during the experimental design of the analyses. A. López-Catalina's research stay at CSIRO was funded by his predoctoral fellowship from the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (FPI-SGIT2018-03). The data were also analysed using the CESGA Supercomputing Infrastructure (Galicia, Spain).

4.8. Funding

A. López-Catalina predoctoral fellowship from Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, (FPI-SGIT2018-03). This project received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement no. 101000226. The authors thank the CRUE-CSIC agreement with Taylor & Francis Group for providing open access funding.

4.9. Availability of Data and Materials

The datasets generated and/or analysed during the current study are available in the National Center for Biotechnology Information (NCBI) repository under the accession number PRJNA981322.

4.10. Ethics Approval

The Ethics Committee of the Instituto Nacional de Investigación y Tecnología Agraria 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 noninterventional study with reference CEEA-PH-2021-08. All methods were carried out in accordance with relevant guidelines and regulations from the European Union and were reported in accordance with the ARRIVE guidelines.

4.11. Tables

Table 4-1. Summary statistics of sequencing performance. On the upper side of the table, calves born from nonlactating cows; on the bottom, calves born from lactating cows.

Sample	Mean length (kb)	Sequencing depth
NL1	10,205.90	7.27
NL2	7,253.20	10.67
NL3	9,013.90	11.06
L1	7,265.40	9.14
L2	8,078.10	6.79
L3	9,496.40	5.62

4.12. Supplementary Tables

Supplementary Table 4-2. Summary statistics of sequencing performance. On the upper side of the table, calves born from nonlactating cows; on the bottom, calves born from lactating cows.

Sample	Mean length (kb)	Mean quality	Number of reads	Total bases	Cov.	Methylation sites called at 7x
NL1	10,205.90	14	2,036,581	20,785,052,645	7.27	4,921,993
NL2	7,253.20	11.9	4,267,023	30,949,609,769	10.67	8,178,489
NL3	9,013.90	13.2	3,505,964	31,602,528,859	11.06	11,935,711
L1	7,265.40	11.1	3,594,990	26,118,876,849	9.14	4,252,453
L2	8,078.10	12.9	2,402,415	19,406,902,038	6.79	3,135,265
L3	9,496.40	12.4	1,692,000	16,067,916,820	5.62	1,220,971

Supplementary Table 4-3. Biological processes and their adjusted p-values associated with the genes used as input for the enrichment analysis.

Gene ontology term	Biological process	Adjusted p-value
GO:0007275	multicellular organism development	9.11×10^{-23}
GO:0048731	system development	6.92×10^{-22}
GO:0048856	anatomical structure development	2.81×10^{-21}
GO:0005737	cytoplasm	4.28×10^{-20}
GO:0048518	positive regulation of biological process	2.21×10^{-19}
GO:0032502	developmental process	4.10×10^{-19}
GO:0048522	positive regulation of cellular process	6.33×10^{-17}
GO:0009653	anatomical structure morphogenesis	1.39×10^{-16}
GO:0007399	nervous system development	3.68×10^{-16}
GO:0051239	regulation of multicellular organismal process	4.60×10^{-16}
GO:0005515	protein binding	7.93×10^{-15}
GO:0030054	cell junction	2.60×10^{-14}
GO:0048513	animal organ development	1.01×10^{-13}
GO:0048519	negative regulation of biological process	2.45×10^{-13}
GO:0048523	negative regulation of cellular process	3.17×10^{-13}
GO:0010646	regulation of cell communication	3.18×10^{-13}
GO:0023051	regulation of signalling	3.98×10^{-13}
GO:2000026	regulation of multicellular organismal development	4.77×10^{-13}
GO:0022008	neurogenesis	5.99×10^{-13}
GO:0009888	tissue development	8.68×10^{-13}
GO:0035295	tube development	2.13×10^{-12}
GO:0048699	generation of neurons	2.58×10^{-12}
GO:0048869	cellular developmental process	2.79×10^{-12}
GO:0030154	cell differentiation	2.79×10^{-12}

4.13. Figures

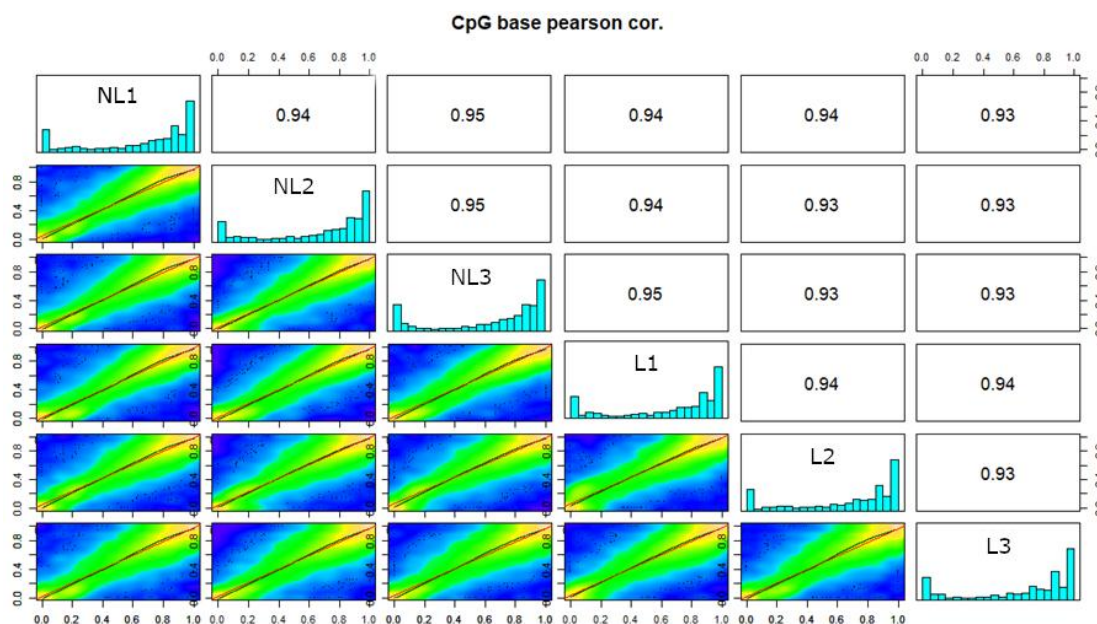


Figure 4-1. Correlations between the methylation profiles of the six samples analysed. The upper triangle represents the Pearson correlation coefficient between the methylation frequencies of the samples by pair. The histogram on the diagonal illustrates the methylation frequency distribution of the marks throughout the genome. The lower triangle depicts a set of scatterplots showing the pairwise percentages of methylation in all samples, where the red line in the plot is from the linear regression fit and the green line is from the polynomial regression fit.

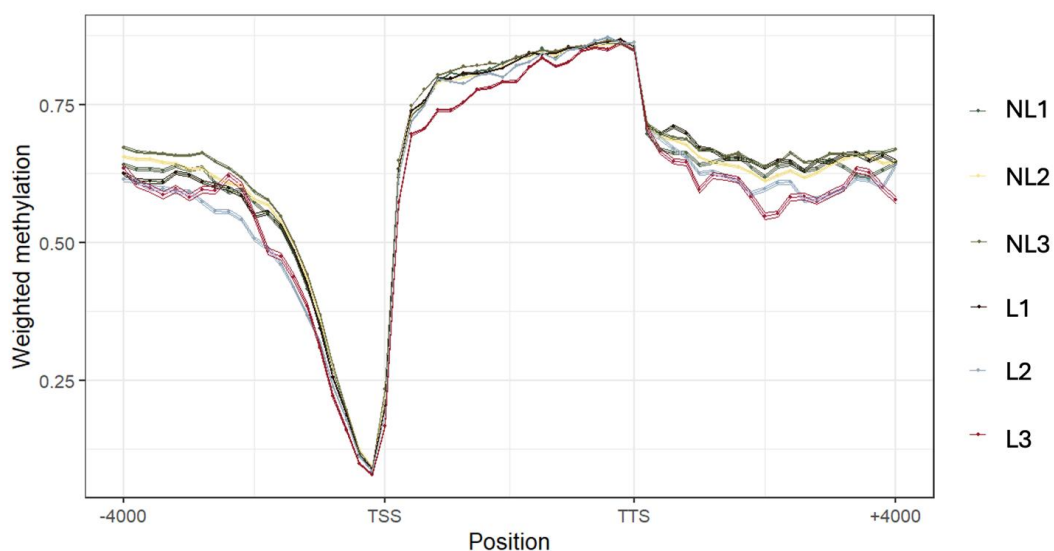


Figure 4-2. Genomic regions were split into 60 windows for each gene annotated in the cow genome to compare the methylation in different genic coordinates. The methylation profiles of the marks detected in each sample were -4,000 bp from the TSS and +4,000 bp from the TTS. The lines and dots represent the mean methylation, and the width of the bands represents the maximum and minimum values of the confidence intervals.

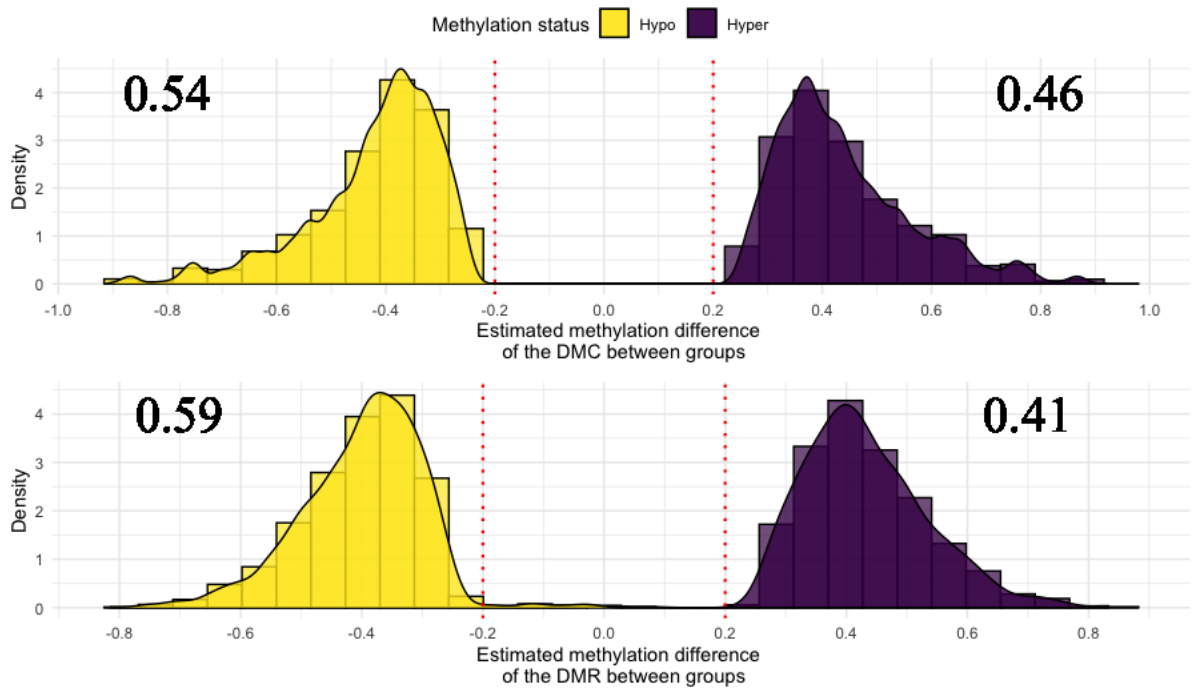


Figure 4-3. Distribution of the difference in methylation frequency for the DMCs and DMRs between the two study groups. The red vertical lines represent the delta value = 0.2 (minimum difference between the estimated means of each group).

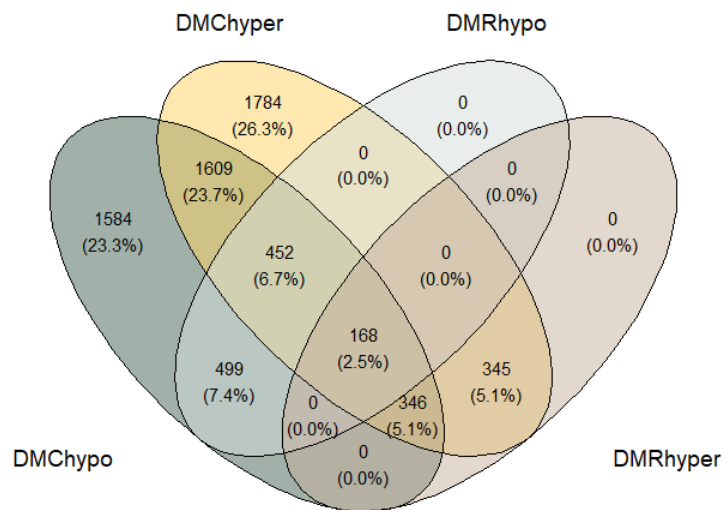


Figure 4-4. Venn diagram depicting the overlap of the genes associated with each DMC/DMR.

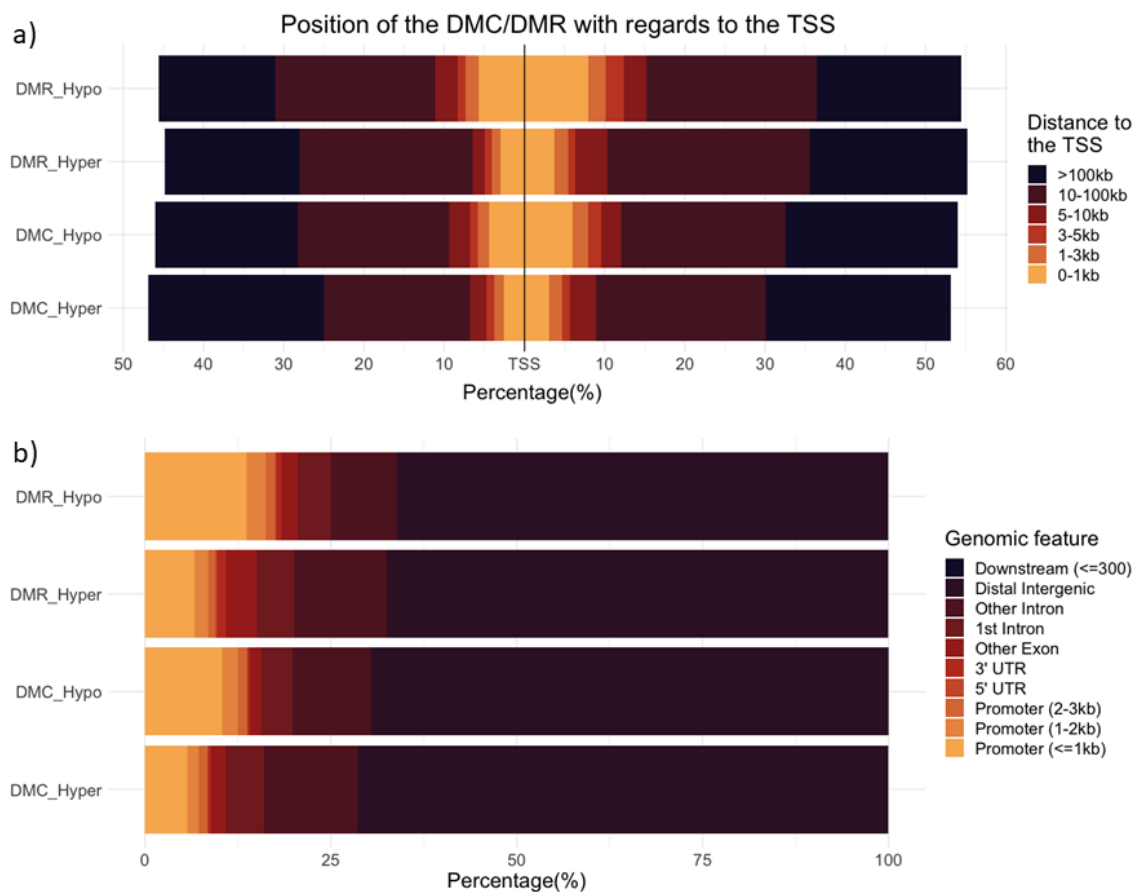


Figure 4-5. Distribution of the genetic features associated with each DMC/DMR depending on their status. Image a) depicts the distance of the mark to the transcription start site. b) represents the genetic feature with which the mark is associated.

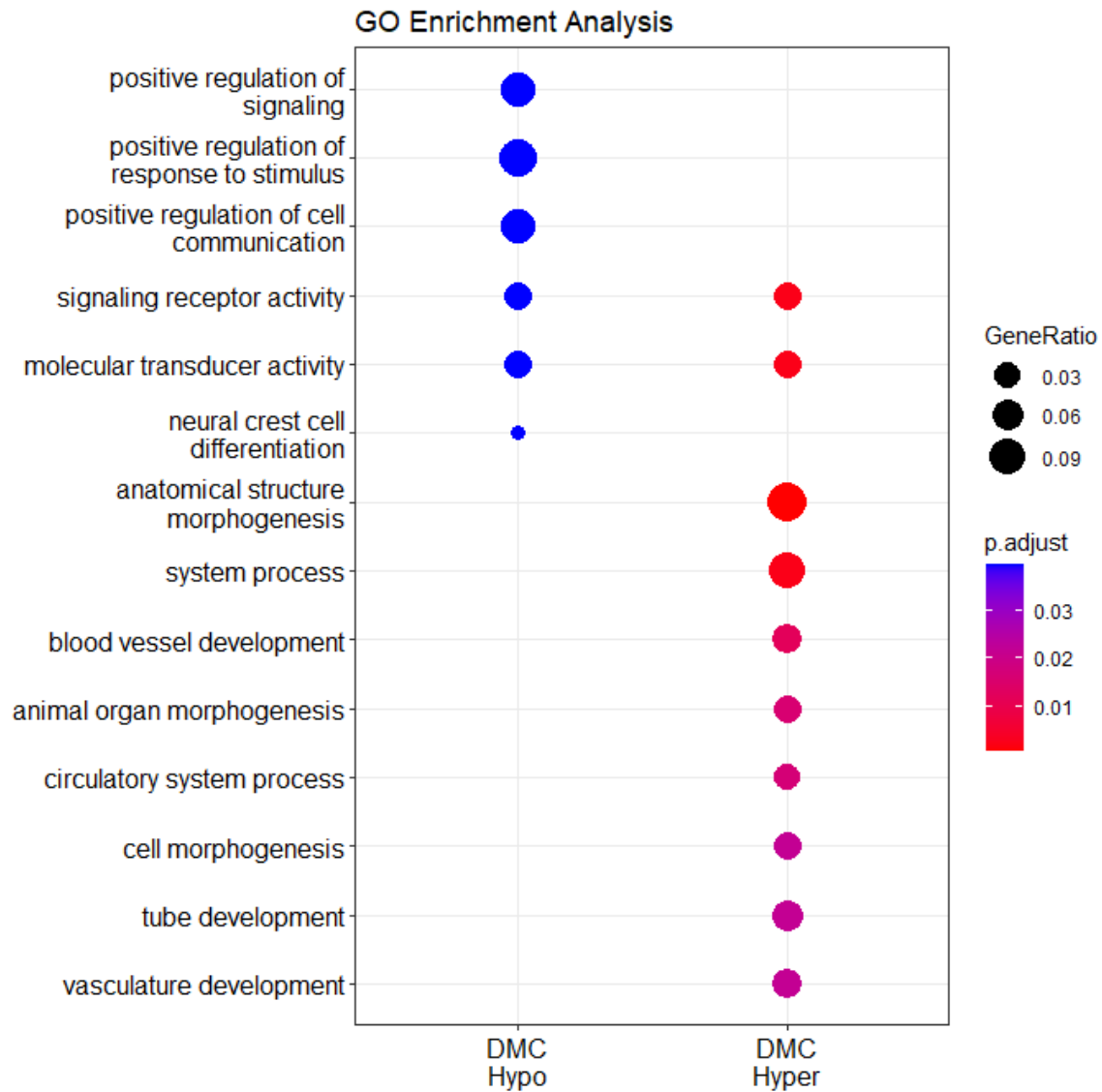


Figure 4-6. Enrichment analysis performed on the genes associated with the DMR and DMC regions using clusterProfiler and calves gestated by nonlactating cows as a reference group. The size of the dots represents the “GeneRatio”, defined as k/n , where k is the number of genes in our dataset that are annotated to the GO term and n is the total number of unique genes in the dataset.

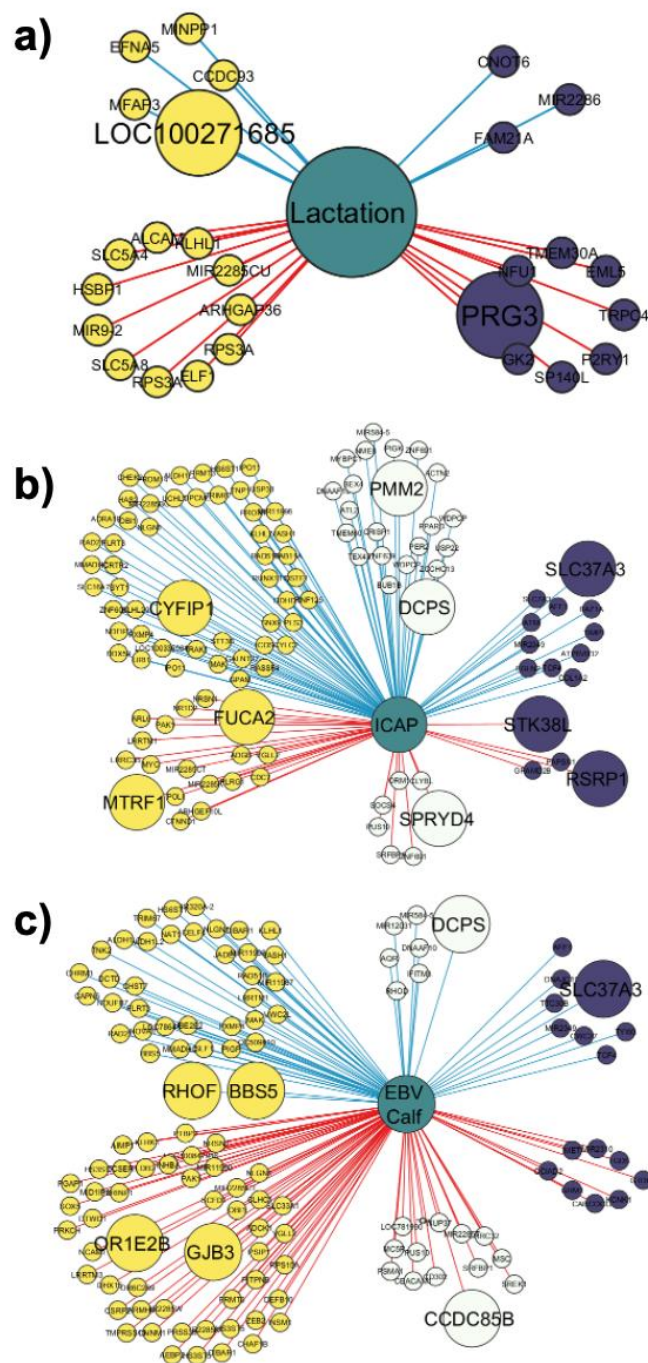
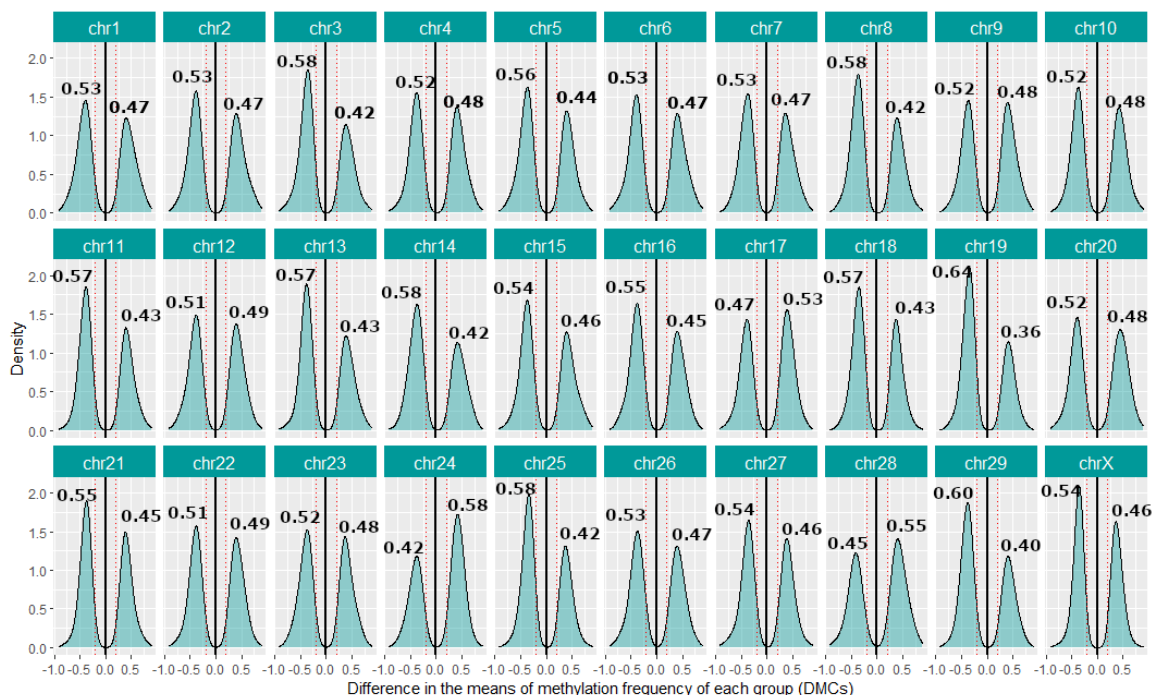
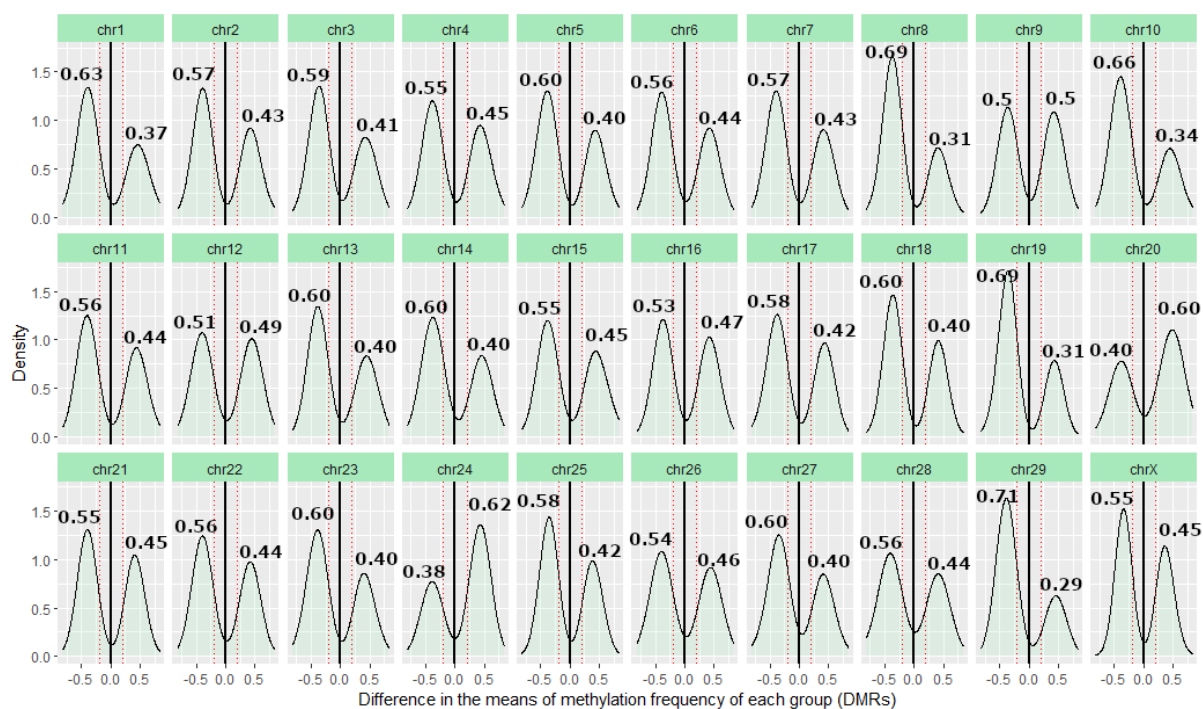


Figure 4-7. Comethylation network between the regions and first neighbours of a) lactation status and b) body capacity index (ICAP) c) milk EBV of the calf. Yellow nodes represent genes associated with hypomethylated DMRs, and purple nodes represent those genes associated with hypermethylated DMRs. The white nodes are regions not detected as DMRs by DSS. The blue lines represent positive correlations, while the red lines indicate negative correlations. Larger nodes were regions that were annotated against promoter regions (-3,000 to 3,000 bp from the TSS).

4.14. Supplementary figures

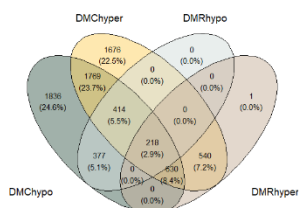


Supplementary Figure 4-8. Distribution of the difference in methylation frequency for the DMCs between the two groups to study by chromosome. Red vertical lines represent the delta value, 0.2.

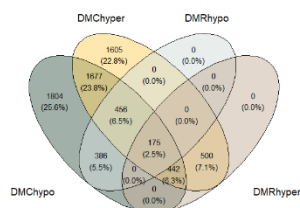


Supplementary Figure 4-9. Distribution of the difference in methylation frequency for the DMRs between the two groups to study by chromosome. Red vertical lines represent the delta value, 0.2.

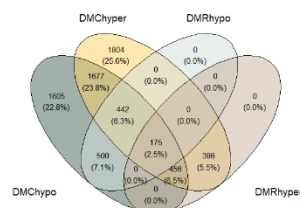
Venn diagram for iteration number: 1



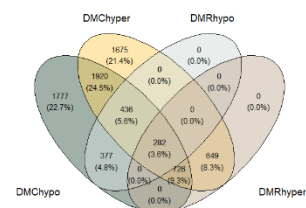
Venn diagram for iteration number: 2



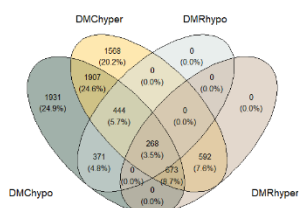
Venn diagram for iteration number: 3



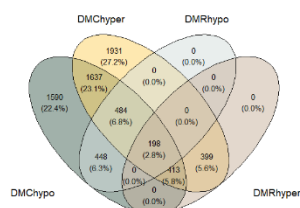
Venn diagram for iteration number: 4



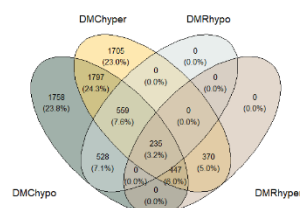
Venn diagram for iteration number: 5



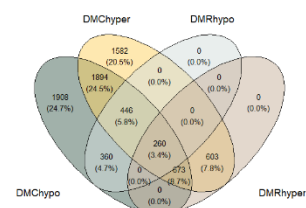
Venn diagram for iteration number: 6



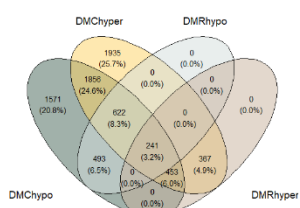
Venn diagram for iteration number: 7



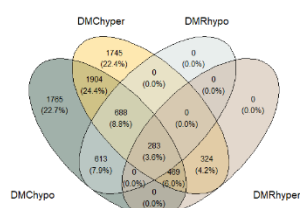
Venn diagram for iteration number: 8



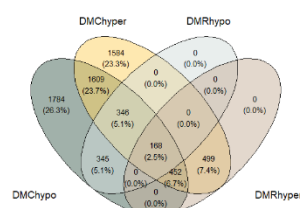
Venn diagram for iteration number: 9



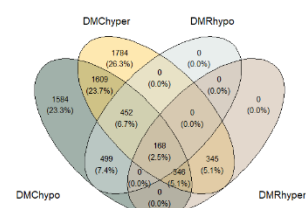
Venn diagram for iteration number: 10



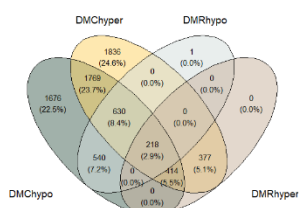
Venn diagram for iteration number: 11



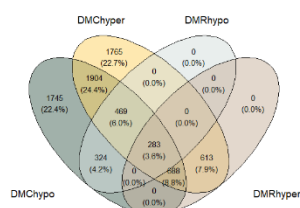
Venn diagram for iteration number: 12



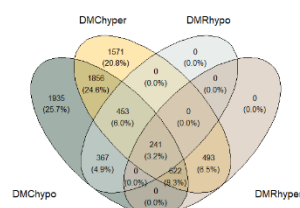
Venn diagram for iteration number: 13



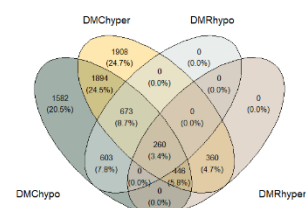
Venn diagram for iteration number: 14



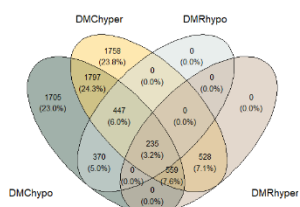
Venn diagram for iteration number: 15



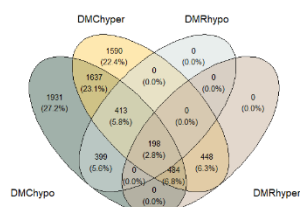
Venn diagram for iteration number: 16



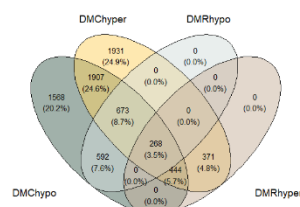
Venn diagram for iteration number: 17



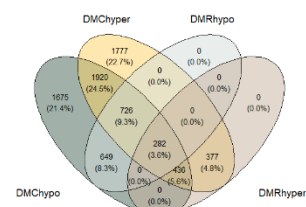
Venn diagram for iteration number: 18



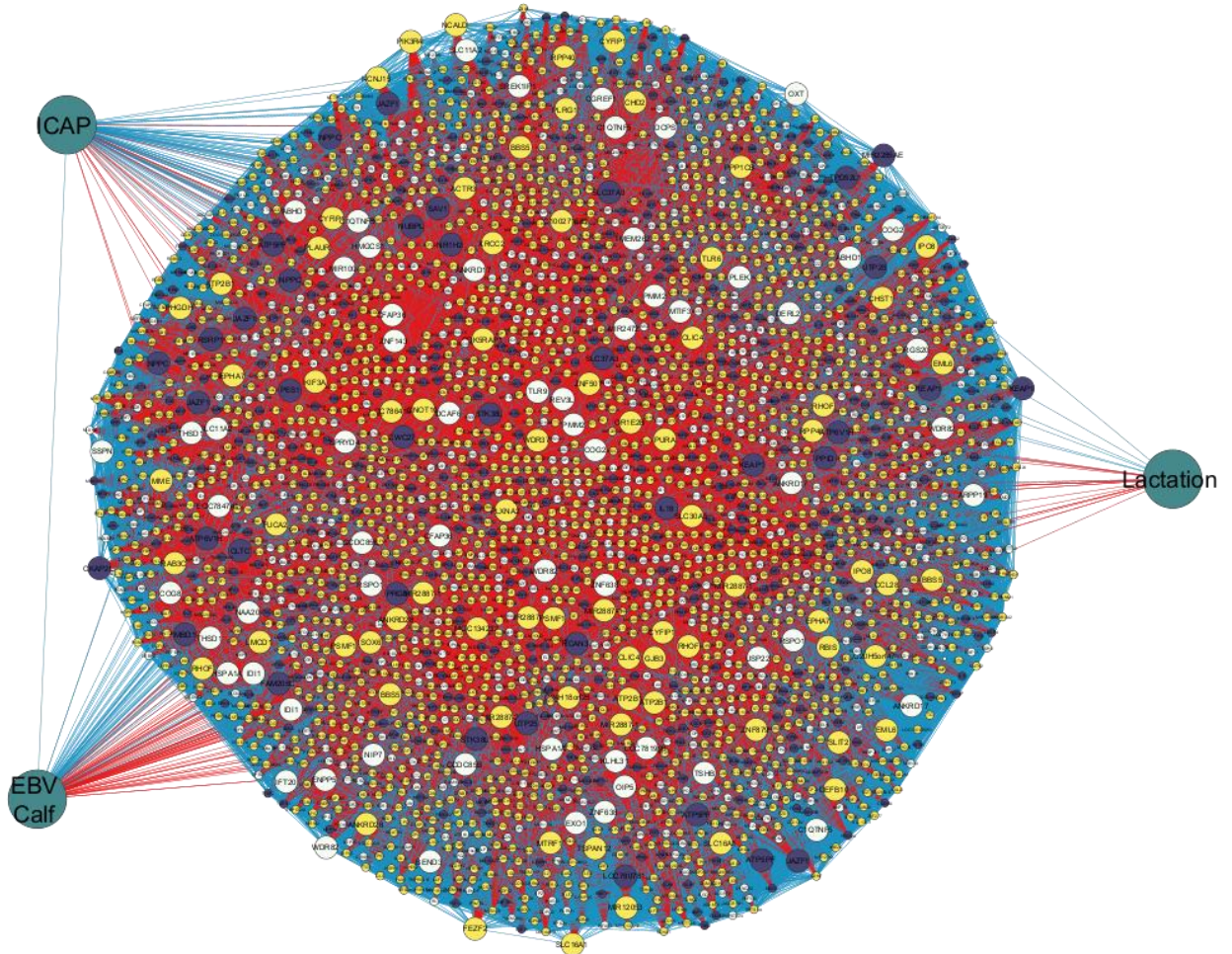
Venn diagram for iteration number: 19



Venn diagram for iteration number: 20



Supplementary Figure 4-10. Venn diagram depicting the overlap of the genes associated with each DMC/DMR for each iteration.



Supplementary Figure 4-11. Comethylation network between the regions and the phenotypes of interest. Yellow nodes represent genes associated with hypomethylated DMRs, and purple nodes represent those genes associated with hypermethylated DMRs. The white nodes are regions not detected as DMRs by DSS. The blue lines represent positive correlations, while the red lines indicate negative correlations. Larger nodes were regions that were annotated against promoter regions (-3,000 to 3,000 bp from the TSS).

4.15. References

- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology* 13, 1–9. <https://doi.org/10.1186/GB-2012-13-10-R87/FIGURES/6>
- Bach, A., Aris, A., Guasch, I., 2017. Consequences of supplying methyl donors during pregnancy on the methylome of the offspring from lactating and non-lactating dairy cattle. <https://doi.org/10.1371/journal.pone.0189581>
- Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E.M., Park, I.-H., Xie, B., Daley, G.Q., Church, G.M., 2009. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 27, 361–368. <https://doi.org/10.1038/nbt.1533>
- Bell, C.G., Teschendorff, A.E., Rakyan, V.K., Maxwell, A.P., Beck, S., Savage, D.A., 2010. Genome-wide DNA methylation analysis for diabetic nephropathy in type 1 diabetes mellitus. *BMC Medical Genomics* 3, 1–11. <https://doi.org/10.1186/1755-8794-3-33/TABLES/2>
- Bender, C.M., Gonzalgo, M.L., Gonzales, F.A., Nguyen, C.T., Robertson, K.D., Jones, P.A., 1999. Roles of Cell Division and Gene Transcription in the Methylation of CpG Islands. *Molecular and Cellular Biology* 19, 6690. <https://doi.org/10.1128/MCB.19.10.6690>
- Bouquet, A., Juga, J., 2013. Integrating genomic selection into dairy cattle breeding programmes: a review. *animal* 7, 705–713. <https://doi.org/10.1017/S1751731112002248>
- Brosnan, C.A., Voinnet, O., 2009. The long and the short of noncoding RNAs. *Current Opinion in Cell Biology, Nucleus and gene expression* 21, 416–425. <https://doi.org/10.1016/j.ceb.2009.04.001>
- Bultmann, S., Stricker, S.H., 2018. Entering the post-epigenomic age: back to epigenetics. *Open Biology* 8, 180013. <https://doi.org/10.1098/rsob.180013>
- Carlson, M., 2019. org.Bt.eg.db: Genome wide annotation for Bovine. R package version 3.8.2 [WWW Document]. URL <https://bioconductor.org/packages/release/data/annotation/html/org.Bt.eg.db.html> (accessed 3.7.23).

Chaput, C., Sirard, M.A., 2020. Embryonic response to high beta-hydroxybutyrate (BHB) levels in postpartum dairy cows. *Domestic Animal Endocrinology* 72, 106431. <https://doi.org/10.1016/J.DOMANIEND.2019.106431>

Chen, D., Li, W., Du, M., Cao, B., 2019. Adipogenesis, fibrogenesis and myogenesis related gene expression in longissimus muscle of high and low marbling beef cattle. *Livestock Science* 229, 188–193. <https://doi.org/10.1016/j.livsci.2019.09.032>

Chen, Y., Gondro, C., Quinn, K., Herd, R.M., Parnell, P.F., Vanselow, B., 2011. Global gene expression profiling reveals genes expressed differentially in cattle with high and low residual feed intake. *Animal Genetics* 42, 475–490. <https://doi.org/10.1111/j.1365-2052.2011.02182.x>

Curradi, M., Izzo, A., Badaracco, G., Landsberger, N., 2002. Molecular Mechanisms of Gene Silencing Mediated by DNA Methylation. *Molecular and Cellular Biology* 22, 3157–3173. <https://doi.org/10.1128/MCB.22.9.3157-3173.2002/ASSET/D6CDD365-FCDD-4870-A89F-35192507AACE/ASSETS/GRAPHIC/MB0921995008.JPEG>

Dai, W., Zou, Y., White, R.R., Liu, J., Liu, H., 2018. Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Funct Integr Genomics* 18, 125–140. <https://doi.org/10.1007/s10142-017-0580-x>

De Coster, W., D'Hert, S., Schultz, D.T., Cruets, M., Van Broeckhoven, C., 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. <https://doi.org/10.1093/BIOINFORMATICS/BTY149>

Deamer, D., Akeson, M., Branton, D., 2016. Three decades of nanopore sequencing. *Nature Biotechnology* 2016 34:5 34, 518–524. <https://doi.org/10.1038/nbt.3423>

Flanagan, J.M., 2015. Epigenome-Wide Association Studies (EWAS): Past, Present, and Future, in: Verma, M. (Ed.), *Cancer Epigenetics: Risk Assessment, Diagnosis, Treatment, and Prognosis*. Springer, New York, NY, pp. 51–63. https://doi.org/10.1007/978-1-4939-1804-1_3

Gamaarachchi, H., Lam, C.W., Jayatilaka, G., Samarakoon, H., Simpson, J.T., Smith, M.A., Parameswaran, S., 2020. GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis. *BMC Bioinformatics* 21, 1–13. <https://doi.org/10.1186/S12859-020-03697-X/FIGURES/6>

González-Recio, O., 2012. Epigenetics: A new challenge in the post-genomic era of livestock, *Frontiers in Genetics*. Frontiers.

González-Recio, O., Ugarte, E., Bach, A., 2012. Trans-Generational Effect of Maternal Lactation during Pregnancy: A Holstein Cow Model. *PLOS ONE* 7, e51816. <https://doi.org/10.1371/JOURNAL.PONE.0051816>

Gray, A.L., Pun, N., Ridley, A.J.L., Dyer, D.P., 2022. Role of extracellular matrix proteoglycans in immune cell recruitment. *International Journal of Experimental Pathology* 103, 34–43. <https://doi.org/10.1111/iep.12428>

Greenberg, M.V.C., Bourc'his, D., 2019. The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology* 20:10 20, 590–607. <https://doi.org/10.1038/s41580-019-0159-6>

Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92, 433–443. <https://doi.org/10.3168/JDS.2008-1646>

Hunter, R.G., 2012. Epigenetic effects of stress and corticosteroids in the brain. *Frontiers in Cellular Neuroscience* 6, 1–20. <https://doi.org/10.3389/FNCEL.2012.00018/BIBTEX>

Iwasaki, Y.W., Siomi, M.C., Siomi, H., 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu Rev Biochem* 84, 405–433. <https://doi.org/10.1146/annurev-biochem-060614-034258>

Jensen, P., 2014. Behaviour epigenetics – The connection between environment, stress and welfare. *Applied Animal Behaviour Science* 157, 1–7. <https://doi.org/10.1016/J.APPLANIM.2014.02.009>

Jjingo, D., Conley, A.B., Yi, S.V., Lunyak, V.V., King Jordan, I., 2012. On the presence and role of human gene-body DNA methylation. *Oncotarget* 3, 462–474. <https://doi.org/10.18632/ONCOTARGET.497>

Kelsey, G., Feil, R., 2013. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368. <https://doi.org/10.1098/RSTB.2011.0336>

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., Bird, A., 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell* 69, 905–914. [https://doi.org/10.1016/0092-8674\(92\)90610-O](https://doi.org/10.1016/0092-8674(92)90610-O)

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/BIOINFORMATICS/BTY191>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078. <https://doi.org/10.1093/BIOINFORMATICS/BTP352>

Liu, S., Tengstedt, A.N.B., Jacobsen, M.W., Pujolar, J.M., Jónsson, B., Lobón-Cervià, J., Bernatchez, L., Hansen, M.M., 2022. Genome-wide methylation in the panmictic European eel (*Anguilla anguilla*). *Molecular Ecology* 31, 4286–4306. <https://doi.org/10.1111/MEC.16586>

Mao, Y., Huang, P., Wang, Y., Wang, M., Li, M.D., Yang, Z., 2021. Genome-wide methylation and expression analyses reveal the epigenetic landscape of immune-related diseases for tobacco smoking. *Clinical Epigenetics* 13, 1–14. <https://doi.org/10.1186/S13148-021-01208-0/FIGURES/6>

Marina, H., Reverter, A., Gutiérrez-Gil, B., Alexandre, P.A., Porto-Neto, L.R., Suárez-Vega, A., Li, Y., Esteban-Blanco, C., Arranz, J.-J., 2020. Gene Networks Driving Genetic Variation in Milk and Cheese-Making Traits of Spanish Assaf Sheep. *Genes* 11, 715. <https://doi.org/10.3390/genes11070715>

Mccarthy, M.I., Hirschhorn, J.N., n.d. Genome-wide association studies: potential next steps on a genetic journey. <https://doi.org/10.1093/hmg/ddn289>

Miranda, T.B., Jones, P.A., 2007. DNA methylation: The nuts and bolts of repression. *Journal of Cellular Physiology* 213, 384–390. <https://doi.org/10.1002/JCP.21224>

Mohammadi, A., Alijani, S., Rafat, S.A., Abdollahi-Arpanahi, R., 2020. Genome-Wide Association Study and Pathway Analysis for Female Fertility Traits in Iranian Holstein Cattle. *Annals of Animal Science* 20, 825–851.

Muyle, A.M., Seymour, D.K., Lv, Y., Huettel, B., Gaut, B.S., 2022. Gene Body Methylation in Plants: Mechanisms, Functions, and Important Implications for Understanding Evolutionary Processes. *Genome Biology and Evolution* 14. <https://doi.org/10.1093/GBE/EVAC038>

Ng, K., Pullirsch, D., Leeb, M., Wutz, A., 2007. Xist and the order of silencing. *EMBO reports* 8, 34–39. <https://doi.org/10.1038/sj.embor.7400871>

Ospelt, C., 2022. A brief history of epigenetics. *Immunology Letters* 249, 1–4. <https://doi.org/10.1016/J.IMLET.2022.08.001>

Park, Y., Wu, H., 2016. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 32, 1446–1453. <https://doi.org/10.1093/BIOINFORMATICS/BTW026>

Pierce, C.F., Speidel, S.E., Coleman, S.J., Enns, R.M., Bailey, D.W., Medrano, J.F., Cánovas, A., Meiman, P.J., Howery, L.D., Mandeville, W.F., Thomas, M.G., 2020. Genome-wide association studies of beef cow terrain-use traits using Bayesian multiple-SNP regression. *Livestock Science* 232, 103900. <https://doi.org/10.1016/j.livsci.2019.103900>

Plager, D.A., Loegering, D.A., Weiler, D.A., Checkel, J.L., Wagner, J.M., Clarke, N.J., Naylor, S., Page, S.M., Thomas, L.L., Akerblom, I., Cocks, B., Stuart, S., Gleich, G.J., 1999. A Novel and Highly Divergent Homolog of Human Eosinophil Granule Major Basic Protein*. *Journal of Biological Chemistry* 274, 14464–14473. <https://doi.org/10.1074/jbc.274.20.14464>

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 Selection Evolution* 46, 1–5. <https://doi.org/10.1186/1297-9686-46-22/FIGURES/2>

R Core Team, 2021. R: A language and environment for statistical computing. [WWW Document]. URL <https://www.R-project.org/>. (accessed 11.15.21).

Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., Pfeifer, G.P., 2009. A human B cell methylome at 100–base pair resolution. *Proceedings of the National Academy of Sciences* 106, 671–678. <https://doi.org/10.1073/pnas.0812399106>

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* 47, W191–W198. <https://doi.org/10.1093/NAR/GKZ369>

Reverter, A., Chan, E.K.F., 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. <https://doi.org/10.1093/BIOINFORMATICS/BTN482>

Rodriguez-Osorio, N., Dogan, S., Memili, E., 2012. Epigenetics of Mammalian Gamete and Embryo Development, in: *Livestock Epigenetics*. John Wiley & Sons, Ltd, pp. 3–25. <https://doi.org/10.1002/9781119949930.ch1>

Scarpa, E., Mayor, R., 2016. Collective cell migration in development. *Journal of Cell Biology* 212, 143–155. <https://doi.org/10.1083/JCB.201508047>

Schefers, J.M., Weigel, K.A., 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers* 2, 4–9. <https://doi.org/10.2527/AF.2011-0032>

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., Timp, W., 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods* 14:4 14, 407–410. <https://doi.org/10.1038/nmeth.4184>

Snajder, R., Leger, A., Stegle, O., Bonder, M.J., 2023. pycoMeth: a toolbox for differential methylation testing from Nanopore methylation calls. *Genome Biology* 24, 83. <https://doi.org/10.1186/s13059-023-02917-w>

Stephansen, R.B., Martin, P., Manzanilla-Pech, C.I.V., Gredler-Grandl, B., Sahana, G., Madsen, P., Weigel, K., Tempelman, R.J., Peñagaricano, F., Parker Gaddis, K.L., White, H.M., Santos, J.E.P., Koltjes, J.E., Schenkel, F., Hailemariam, D., Plastow, G., Abdalla, E., VandeHaar, M., Veerkamp, R.F., Baes, C., Lassen, J., 2023. Novel genetic parameters for genetic residual feed intake in dairy cattle using time series data from multiple parities and countries in North America and Europe. *Journal of Dairy Science* 106, 9078–9094. <https://doi.org/10.3168/jds.2023-23330>

Sun, H.Z., Plastow, G., Guan, L.L., 2019. Invited review: Advances and challenges in application of feedomics to improve dairy cow production and health. *Journal of Dairy Science* 102, 5853–5870. <https://doi.org/10.3168/jds.2018-16126>

Swart, G.W.M., 2002. Activated leukocyte cell adhesion molecule (CD166/ALCAM): Developmental and mechanistic aspects of cell clustering and cell migration. *European Journal of Cell Biology* 81, 313–321. <https://doi.org/10.1078/0171-9335-00256>

Tremblay, R., Dufort, I., Sirard, M.A., 2018. Metabolic stress induces modifications in the epigenetic program of preimplantation bovine embryos. *Molecular reproduction and development* 85, 117–127. <https://doi.org/10.1002/MRD.22941>

Wang, X., Zhang, L., Jin, J., Xia, A., Wang, C., Cui, Y., Qu, B., Li, Q., Sheng, C., 2018. Comparative transcriptome analysis to investigate the potential role of miRNAs in milk protein/fat quality. *Sci Rep* 8, 6250. <https://doi.org/10.1038/s41598-018-24727-y>

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2, 100141. <https://doi.org/10.1016/J.XINN.2021.100141>

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. <https://doi.org/10.1093/BIOINFORMATICS/BTV145>

Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. https://doi.org/10.1126/SCIENCE.1186366/SUPPL_FILE/ZEMACH.SOM.PDF

Zhang, Y.-Y., Zan, L.-S., Wang, H.-B., Qing, L., Wu, K.-X., Quan, S.-A., Li, C.-Q., Zhong, X., Wang, C.-J., 2011. Differentially expressed genes in skeletal muscle tissues from castrated Qinchuan cattle males compared with those from intact males. *Livestock Science* 135, 76–83. <https://doi.org/10.1016/j.livsci.2010.06.070>

Zilberman, D., 2017. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biology* 18, 1–3. <https://doi.org/10.1186/S13059-017-1230-2/METRICS>

Zuccolo, J., Bau, J., Childs, S.J., Goss, G.G., Sensen, C.W., Deans, J.P., 2010. Phylogenetic Analysis of the MS4A and TMEM176 Gene Families. *PLOS ONE* 5, e9369. <https://doi.org/10.1371/journal.pone.0009369>

5. Chapter 3 - A Recursive Model Approach to Include Epigenetic Effects in Genetic Evaluations Using Simulated DNA Methylation Effects

López-Catalina, A., Ragab, M., Reverter, A., and González-Recio, O. (2025). A recursive model approach to include epigenetic effects in genetic evaluations using simulated DNA methylation effects. *Journal of Animal Breeding and Genetics*. doi:10.1111/jbg.12925

5.1. Abstract

The advancement of epigenetics has highlighted DNA methylation as an intermediate-omic influencing gene regulation and phenotypic expression. With emerging technologies enabling the large-scale and affordable capture of methylation data, there is growing interest in integrating this information into genetic evaluation models for animal breeding. This study used methylome information from 6 dairy cows to simulate the methylation profile of 13,183 genotyped animals. The liability to methylation was treated as an additive trait, while a trait mediated by methylation effects was also simulated. A multiomic model (GOBLUP) was adapted to incorporate methylation data in genomic and genetic evaluations, using the traditional BLUP method as a benchmark. The GOBLUP accurately recovered heritability estimates for the liability to methylation in all low, medium and high heritability scenarios and was consistent at estimating the heritability for the epigenetics-modulated trait of interest at a low-medium heritability of 0.14. The genetic variance recovered by the BLUP model was influenced by the h^2 of the liability to methylation, and part of the methylation variance for the phenotypic trait was captured as additive. The h^2 of the phenotypic trait partially relies on the h^2 value for the methylation windows in the traditional model. A newly proposed estimated epigenetic value (EEV) combines the traditional additive genetic information from genotyping arrays with epigenetic information. The correlation between the traditional EBV and EEV was high (0.92 to 0.99 depending on the scenario), but the correlation of the EEV with the true breeding value was higher than the correlation between the traditional EBV and the TBV (0.85 vs 0.75, 0.71 vs 0.66 and 0.61 vs 0.62 depending on the scenario). This study demonstrates that the GOBLUP multiomic recursive model can effectively separates additive and epigenetic variances, enabling improved breeding decisions by accounting for genetic liability to DNA methylation. This enables more informed breeding decisions, optimizing selection for desired traits. Emerging sequencing techniques offer new opportunities for cost-effective simultaneous acquisition of genetic and epigenetic data, further enhancing breeding accuracy.

5.2. Introduction

The widespread adoption of genomic selection over the past decade has significantly increased the rate of genetic gain in livestock populations, resulting in consistent improvements in phenotypes that are economically important (Cole et al., 2021). In addition to genomic information, intermediary omics features can now be obtained at an affordable cost. These features include traits that influence the expected genetic effect on the phenotype (Christensen et al., 2021). The incorporation of omics technologies provides novel opportunities to predict phenotypes and breeding values considering both genetic and epigenetic factors. These opportunities allow for more comprehensive evaluations that address complex traits influenced by multiple biological pathways.

Examples of these omics applications in livestock systems include the study of milk metabolites (metabolomics) to diagnose heat-stressed animals (Tian et al., 2016); transcriptomics to identify different immune responses to subclinical mastitis pathogens (Bisutti et al., 2023); metagenomics to explore the relationship between rumen microbial composition and methane emissions (Saborío-Montero et al., 2021) or the association of different phenotypes of interest in livestock practices with regions of the methylome in dairy cattle (López-Catalina et al., 2024). Another important intermediate omics feature is epigenetics. Epigenetics, the study of changes in gene expression not caused by DNA sequence alterations, is gaining importance in livestock breeding (Gonzalez-Recio et al., 2015; Kim and Kim, 2021). It encompasses mechanisms like DNA methylation, histone modifications, and non-coding RNAs. These epigenetic marks can respond to environmental factors and influence phenotypes, and have been associated with the problem of the missing heritability in livestock traits (Ibeagha-Awemu and Zhao, 2015).

The emergence of next-generation sequencing techniques, such as Nanopore sequencing, has enabled the simultaneous retrieval of both DNA sequences and modified bases in the same sequencing process (Wang et al., 2021) even at a low sequencing depth (González-Recio et al., 2023), simplifying the acquisition of genomic information. Methylation arrays are also being developed to profile CpG methylation in livestock. Epigenetic information, particularly DNA methylation, can also be considered as an intermediate feature with a direct effect on the phenotype (Ospelt, 2022). DNA methylation, first discovered in the late 1940s (Hotchkiss, 1948), involves the addition of a methyl group to the fifth carbon of cytosine, resulting in 5-methylcytosine (5mC) (Miranda and Jones, 2007).

Nowadays, it is also known that non-cytosine modifications such as N6-methyladenine (6mA) and 5-hydroxymethyluracil (5hmU) are also implicated in epigenetic regulation, suggesting a broader spectrum of epigenetic marks beyond traditional cytosine modifications (Lowe et al., 2021). Methylated regions are usually associated with gene silencing, as they inhibit the binding of transcription factors to gene promoter regions, preventing the start of DNA transcription (Jjingo et al., 2012). Epigenetic processes also play a critical role in tissue development, suppressing the expression of unnecessary genes and ensuring accurate cell line determination (Curradi et al., 2002). DNA methylation patterns can be influenced by both genetic and environmental factors, leading to variably methylated regions in the genome. These methylation differences can result in plastic phenotypes that may become locally adapted across environments (Berbel-Filho et al., 2019), highlighting the potential for integrating epigenetic data to enhance selective breeding efforts. Methylation quantitative trait loci have been identified, explaining some of the inter-individual variation in DNA methylation (Teh et al., 2014). Considering the interaction between epigenetics, genetics, and environment can improve the estimation of breeding values and reduce bias in genetic evaluations (Ghorbani and Shokri-Gharelo, 2019). Moreover, genotype-dependent methylation might elucidate observed gene-by-environment ($G \times E$) interactions that influence various traits (Ma and Ordovas, 2017). While transgenerational epigenetic inheritance in mammals is still debated, epigenetic research in livestock species is growing, albeit limited compared to human studies (Gonzalez-Recio et al., 2015; Ibeagha-Awemu and Zhao, 2015).

Future livestock improvement programs may benefit from considering epigenetic markers alongside genomic information. Recent research highlights the importance of incorporating epigenetic effects in animal breeding models. Epigenetic factors can contribute to non-Mendelian variability in traits and might have inter or trans-generational effects. Studies have proposed models to estimate epigenetic variance, including the use of epigenetic relationship matrices and recursive parameters to capture transmissibility of epigenetic markers (Varona et al., 2015). Omics information can be integrated into genomic evaluations through nested GBLUP models, where the subrogated model estimates the effect of the omics feature on the trait and the additive residual breeding value, and the main model estimates the mediated breeding value reflecting the genetic influence on the omics feature (Legarra and Christensen, 2023).

The integration of genomic and epigenetic information can potentially improve breeding value estimations and reduce biases. However, challenges remain in estimating genetic variance components in populations undergoing intense genomic selection (Jensen, 2016). Multiple-matrix animal models have been developed to disentangle genetic and non-genetic causes of phenotypic variance, including environmental, epigenetic, and cultural effects (Thomson et al., 2018). These advanced models can be applied to both wild and experimental populations, offering new opportunities to understand the interplay between genetic and non-genetic factors in shaping evolutionary potential.

Here, we propose an epigenetic model based on the equations proposed by Christensen et al. (2021), where the methylation frequency affects the phenotype but not the opposite. The objective of this study was to propose new models to separate the epigenetic effects from the additive genetic effects and discuss the implication that adding methylation information could have on breeding programmes. This approach aims to better address the complexity of trait heritability and improve the predictive ability of genetic evaluations in livestock.

5.3. Materials and Methods

5.3.1. Data

Blood samples from six unrelated Holstein-Friesian calves were sequenced using an Oxford Nanopore GridION device equipped with R9.4 flow cells. Three of the calves were born to lactating cows, while the remaining three were born to non-lactating cows. All samples were collected from a single commercial farm and were part of a larger set of samples from the EU H2020 RUMIGEN project. The animals were handled following ethical standards and were not involved in any experimental procedure. Samples were chosen at random. DNA libraries were prepared using the LSK-110 kit. Basecalling was conducted using Guppy v6.3.7 (only available to Oxford Nanopore Technologies customers via their community site <https://community.nanoporetech.com>) with the model dna_r9.4.1_e8.1_modbases_5mc_cg_sup.cfg. Methylation sites and methylation frequency were identified using F5C (Gamaarachchi et al., 2020). Only positions with a minimum coverage of 7x were retained for downstream analyses. Additionally, genotype data from 13,183 animals from various commercial farms

genotyped with the BovineSNP50 Illumina SNP array (54,609 SNPs) were included.

5.3.2. Methylation Pattern Analysis

The analysis was initiated by iteratively processing each gene entry within the *Bos taurus* gff3 dataset (https://ftp.ensembl.org/pub/release-111/gff3/bos_taurus/). Essential gene information, including gene names, chromosomes, start and end positions, and gene orientations, was extracted. The gene region was divided into thirteen consecutive, non-overlapping windows. These defined windows were then iterated through, and methylation data within each window was extracted. Windows 1 and 2 were associated with the promoter and TSS regions, whereas windows 10, 11, and 12 represented the gene body positions (Fig. 5-1), which transcribes to protein and thus was referred to as the “core” region. All other regions were defined as the “coat” region. The mean methylation of each window (μ_w), and its standard deviation ($\sigma_{p,w}$), were estimated from the observed methylation in the six sequenced samples.

5.3.3. Simulation of Breeding Values and Traits

Two traits were simulated for the 13,183 animals with genotype data: (1) the liability to methylation at cytosine nucleotide followed by guanine nucleotide sites (CpG) as a purely additive trait, and (2) a trait (e.g. milk yield) with additive effect regulated by epigenetic effects through methylation. The SNP effects are mediated by the epigenetic effects, impacting the trait acting as the mediator.

For the first trait (liability to methylation), 10,000 SNPs were randomly selected and their effects (γ) were drawn from a gamma distribution with shape and scale hyperparameters of 0.42 and 5.4 (Hayes and Goddard, 2001). This methylation effect refers to the genetic control of the individual on the susceptibility to present methylation on CpG sites.

The breeding value for susceptibility to methylation (mBV) for individual i was computed as

$$mBV_i = \sum_{j=1}^{j=10,000} x_{i,j} \cdot \gamma_j \quad (\text{Eq. 1})$$

where $x_{i,j}$, is the SNP_j genotype and γ_j is the SNP effect for liability to global DNA methylation.

The methylation frequency phenotype at each window (y_{meth_w}) was computed using the phenotypic variance observed in the sequenced individuals, as $y_{meth_{i_w}} = \mu_w + mBV_{i_w} + e(m)_{i_w}$, where μ_w is the average methylation in windows w , observed in the sequenced animals, mBV_{i_w} is the genetic value for liability to methylation distributed as $N(0, \sigma_{a_{meth_w}}^2)$, and $e(m)_{i_w}$ is the environmental residual effect on liability to methylation distributed as $N(0, \sigma_{e_{meth_w}}^2)$. The $\sigma_{a_{meth_w}}^2$ and $\sigma_{e_{meth_w}}^2$ are the corresponding variances for the genetic and environmental effects for the respective window w . These variances, and hence the corresponding effects, were adjusted to simulate three heritability scenarios $h^2 = 0.1$ (low), 0.3 (medium) and 0.8 (high).

The simulated trait, regulated through epigenetic effects acting as moderators, was generated as follows: First, the trait was assumed to be regulated by 10,000 SNP loci: 7,500 SNPs randomly selected from core regions to simulate the effect of the gene bodies on the phenotype; and 2,500 SNPs randomly selected outside the genic region. SNP effects (β) were also drawn from a gamma distribution with a shape and scale hyperparameters of 0.42 and 5.4 (Hayes and Goddard, 2001).

Rather than modelling a purely additive genetic effect, an epigenetics-modulated true breeding value incorporating both genetic and epigenetic effects, termed $epiTBV_i$, was simulated. Here, the epigenetic effect moderates the influence of the SNP effects on the phenotype but does not independently contribute to the production trait.

$$epiTBV_i = epiTBV_{core_i} + epiTBV_{coat_i} \quad (\text{Eq. 2})$$

Where $epiTBV_{core_i}$ is the breeding value of individual i from core regions whose expression is mediated by methylation calculated as

$$epiTBV_{core_i} = \sum_{j=1} x_{i,j} * [\beta_j * (\frac{1}{2}(1 - y_{meth_{i,1}}) + \frac{1}{2}(1 - y_{meth_{i,2}}))] \quad (\text{Eq. 3})$$

And $epiTBV_{coat_i}$ is the breeding value from core regions not affected by methylation, calculated as:

$$epiTBV_{coat_i} = \sum_{j=1} x_{i,j} * \beta_j \quad (\text{Eq. 4})$$

where $x_{i,j}$ is the genotype for SNP j in animal i , β_j is the effect of the SNP j for the phenotypic trait, $y_{meth_{i,1}}$ is the methylation frequency for animal i in window 1 and $y_{meth_{i,2}}$ is the methylation frequency for animal i in window 2, as described above. Note that $y_{meth_{i,1}}$ and $y_{meth_{i,2}}$ mimic the expression level of the gene regulated by methylation effects.

The phenotype for the epigenetics-modulated trait (y_{prod}) was simulated as:

$$y_{prod_i} = \mu_{prod} + epiTBV_i + e_i \quad (\text{Eq. 5})$$

where μ_{prod} is the population mean, $epiTBV_i$ is the epigenetics-modulated true breeding value for the yield trait as described above and e_i is the residual term following a distribution $N(0, \sigma_e^2)$ where σ_e^2 was set to obtain a $h_{prod}^2 = 0.3$ under a purely additive model (calculating the genetic variance from breeding values without considering methylation: $Var(\mathbf{X}\boldsymbol{\beta})$).

5.3.4. Benchmark Model

A traditional animal model was used to estimate variance components and animal breeding values for the simulated epigenetics-modulated trait using the population mean as the only systematic effect and the animal and residual effect as random:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (\text{Eq. 6})$$

where $\mathbf{y}$ is the phenotype of interest, $\mathbf{X}$ is an incidence matrix that relates the systematic effects to the phenotypes, $\mathbf{Z}$ is an incidence matrix that relates the animal random effect to the phenotype, $\boldsymbol{\beta}$ are the fixed effects (in this case only for the population mean), $\mathbf{a}$ is the additive estimated breeding values assumed to be distributed as $\mathbf{a} \sim N(0, \mathbf{H}\sigma_a^2)$, and $\mathbf{e} \sim N(0, \sigma_e^2)$ is the vector of residuals, with σ_a^2 and σ_e^2 being the additive genetic and residual variances, and $\mathbf{H}$ being the genomic and pedigree relationship matrix from a ssGBLUP model as described in Legarra et al. (2014). The Blupf90+ suite of programs from Misztal et al. (2014) were used to solve the model.

5.3.5. Multiomic Approach Using GOBLUP Model

Alternatively, a multiomic approach was implemented adapting the procedure described in Christensen et al. (2021). This procedure considers that the epigenetics-modulated trait is controlled by epigenetic and additive effects, and at the same time the epigenetic effect is controlled by additive effects and has a recursive mode of action on the epigenetics-modulated trait. The model was named as Genomic Omic BLUP (GOBLUP) in Christensen et al. (2021). We adapted this model to include genomic-epigenomic effects as:

$$\mathbf{y} = \mathbf{X}'\boldsymbol{\beta} + \mathbf{Z}'_{\mathbf{m}}\mathbf{y}_{meth_w}\boldsymbol{\alpha} + \mathbf{Z}'\mathbf{a}_r + \mathbf{e} \quad (\text{Eq. 7})$$

$$\mathbf{y}_{meth_w} = \mathbf{X}'_{\mathbf{m}}\boldsymbol{\beta}_{\mathbf{m}} + \mathbf{M}'\mathbf{g}_{m_w} + \mathbf{e}_w \quad (\text{Eq. 8})$$

Terms in Eq [7] are described next: $\mathbf{y}$ is the phenotype of the epigenetics-modulated trait (controlled by epigenetic effects), $\mathbf{X}$ is the incidence matrix that relates the systematic effects to the phenotypes (in this case only the population mean) with $\boldsymbol{\beta}$ being its effect, $\mathbf{Z}_{\mathbf{m}}$ is an incidence matrix that relates the methylation phenotype of each individual to their corresponding phenotype for the epigenetics-modulated trait random effects to the phenotype, with $\boldsymbol{\alpha}$ being the vector of the methylation effects on the epigenetics-modulated trait, $\mathbf{Z}$ and $\mathbf{M}$ are incidence matrices that related each animal to its phenotype, $\mathbf{a}_r$ is the additive residual breeding value (non-affected by epigenetic effects) and assumed to be distributed as $\mathbf{a}_r \sim N(0, \mathbf{H}\sigma_{a_r}^2)$, and $\mathbf{e}$ is the vector of residual terms distributed as $N(0, \sigma_e^2)$.

In Eq [8], $\mathbf{y}_{meth_w}$ is the vector containing the observed methylation frequency of individuals in window $w=\{1...13\}$, $\mathbf{X}_{\mathbf{m}}$ is the matrix that relates the systematic effects to the phenotypes (in this case only the population mean), $\mathbf{Z}'$ is an incidence matrix that relates the animal effect to the phenotype $\mathbf{y}_{meth_w}$; $\mathbf{g}_{m_w}$ is the additive genetic effect for window w assumed to be distributed as $\mathbf{g}_{m_w} \sim N(0, \mathbf{H}\sigma_{m_w}^2)$ and $\mathbf{e}_w$ is the vector of residuals assumed to be distributed as $N(0, \sigma_{e_w}^2)$, with $\sigma_{m_w}^2$ and $\sigma_{e_w}^2$ being the additive genetic and residual variances. This model was applied to the phenotypic methylation observed in the 13 windows described above (*i.e.* the methylation frequency of each window was analysed separately).

The final total estimated epigenetic value (EEV) was calculated as $EEV_i = \hat{a}_{r_i} + \hat{a}_{m_i}$, where the omics-mediated breeding value was $\hat{a}_{m_i} = \sum_{w=1}^{13} \hat{a}_w \hat{g}_{i,m_w}$. The

EEV includes the additive residual breeding value and the additive breeding value for the liability to methylation times the methylation effect of each window on the epigenetics-modulated trait.

The GOBLUP model was solved using the Gibbsf90 program from the Blupf90+ family (Miszta et al., 2014) running 200,000 iterations with a burn-in period of 20,000. Both the traditional and GOBLUP models were evaluated under the three heritability scenarios simulated for the liability to methylation (low, medium and high). Pearson correlations between traditional estimated breeding value (EBV), additive residual EBV, additive genetic effect, EEV and the epigenetics-modulated breeding value using the cor function from R (R Core Team, 2021). Plots were generated using the ggplot2 package in R (Wickham, 2009).

5.3.6. Model Assumptions and Remarks

The generator model used for the simulation and the statistical model (GOBLUP) were intentionally different. The generator model tries to simplify the biological reality behind methylation effects in the most plausible mode of action. It was assumed that the methylation frequency of the promoter region (windows 1 and 2) hinders gene expression, decreasing the SNP effect in the gene body region. We represent this in Eq. 3 where β_j is evenly multiplied by the phenotypic methylation frequency of windows 1 and 2. For instance, given an original SNP effect of $\beta_j = 100$, if the methylation frequency of said promoter regions are $y_{meth_1} = 0.6$ and $y_{meth_2} = 0.8$, and following $\beta_j * \left(\frac{1}{2}(1 - y_{meth_{i,1}}) + \frac{1}{2}(1 - y_{meth_{i,2}}) \right)$, the new SNP mediated-effect was $100 * \left(\frac{1}{2}(1 - 0.6) + \frac{1}{2}(1 - 0.8) \right) = 30$. On the other hand, the statistical model captures and separates the effects of a large data set where the biological reality is not described exhaustively.

Figure 5-2 illustrates the difference between our mentioned models following the nomenclature proposed by Hayes (2009). Figure 5-2.a represents the generator model. In this framework, the genotype (X) influences the methylation effects (M) through a , as these effects are partly modulated by the animal's genetics. Simultaneously, X impacts the phenotype (Y) via c' through b , effectively reducing the SNP effect. Figure 5-2.b depicts the statistical model underpinning GOBLUP. Here, the key distinction lies in the methylation effects (M) not modifying the original SNP effect. Instead, M independently contributes to the final phenotype (Y) through b , while c' represents the unchanged direct influence

of X on Y. Figure 5-2.c shows the reparametrised form of the generator model. In this case, the parameters are redefined as $b' = -c \cdot b$ and $c'' = c'$. This formulation highlights the direct interaction between methylation effects and the genetic contribution to the phenotype, providing an equivalent yet simplified interpretation of the relationships in Figure 5-2.a, and supporting the use of a recursive model.

5.4. Results

5.4.1. Overview of the Simulation

The whole-genome methylation profile was assessed using data from six ONT-sequenced samples. Methylation frequencies showed an average ranging from 0.50 to 0.70 at approximately 4 kb upstream from the TSS, followed by a sharp decline to approximately 0.10 in the promoter region, which facilitates gene expression regulation. Methylation levels increased sharply to over 0.80 within the gene body, specifically defined as the "core" region between the TSS and TTS, contributing to mutation protection and enhanced gene expression (Jones, 2012; Sriraman et al., 2020). Beyond the TTS, a slight decrease in methylation frequencies returned values to around 0.70, consistent with our prior observations in López-Catalina et al. (2024), and similar to findings in mammalian species (Su et al., 2014; Gao et al., 2023). This data was used to create the thirteen genomic windows (Figure 5-1). The average methylation frequencies in each window and their standard deviations are shown in Table 5-1. Regions associated with repression of gene expression (windows 1 and 2) exhibited frequencies of 0.62 ± 0.01 and 0.54 ± 0.02 , respectively. Frequencies decreased progressively across windows 3 to 8 (0.47 ± 0.03 , 0.41 ± 0.03 , 0.34 ± 0.02 , 0.26 ± 0.02 , 0.18 ± 0.01 , and 0.11 ± 0.01 , respectively), with window 9 showing a frequency of 0.09 ± 0.01 . The core regions (windows 10, 11, and 12) had frequencies of 0.20 ± 0.02 , 0.62 ± 0.03 , and 0.82 ± 0.04 , respectively, while window 13 had a frequency of 0.64 ± 0.03 , as illustrated in Figure 5-1 and Table 5-1.

A summary of the SNPs located in each window and region is shown in Table 5-1. The differences in window lengths reflect the varying quantities of SNPs found within each individual window. For example, a higher number of SNPs with an effect on the epigenetics-modulated trait were located in window 12, compared to SNPs associated with methylation liability.

5.4.2. Traditional Animal Model

The data were analysed with the traditional BLUP model, ignoring methylation. This benchmark model estimated the heritability of the epigenetics-modulated trait across the different methylation scenarios obtaining different heritability estimates depending on the simulated heritability for liability to methylation: 0.15 when the heritability of liability to methylation was low (0.10), 0.18 when this heritability was medium (0.3), and 0.25 for the scenario for high (0.8) heritability to methylation (Table 5-2). Increased heritability of liability to methylation resulted in larger estimated genetic variance for the epigenetics-modulated trait. Note that the simulated additive variance for the epigenetics-modulated trait was kept constant across scenarios. Thus, this implies that part of the epigenetic variance of the epigenetics-modulated trait was captured as additive genetic variance. However, the traditional model was not able to detect interactions between SNPs of the two simulated traits. This feature is equivalent to that described by Hill (2017) who demonstrated that epistatic effects generate additive genetic variance. A formal analysis would be necessary to prove that epigenetic effects also generate additive variance, but our results suggest that variance from methylation effects are also captured by traditional models as additive genetic variance.

5.4.3. Epigenetic GOBLUP Model

The proposed epigenetic GOBLUP model assumes methylation effects to be independent from additive genetic effects influencing the epigenetics-modulated trait. The heritability estimate obtained for liability to methylation and the epigenetics-modulated trait were in agreement with the simulated parameters (mean heritability of liability to methylation across windows were 0.10, 0.32 and 0.85 respectively, and 0.14 for the epigenetics-modulated trait) (Table 5-2). Consistent variance components of the epigenetics-modulated trait were observed across scenarios, with additive genetic variance at 1.01 and residual variance at 6.09. The estimated additive residual breeding values ($\hat{\mathbf{a}}_r$) were also consistent across the scenarios with different heritabilities for the liability to methylation. We expected $\hat{\mathbf{a}}_r$ to be similar across scenarios, as the GOBLUP model captured the direct genetic value of the animal which was the same in all simulated scenarios. On the other hand, the omics-mediated breeding value ($\hat{\mathbf{a}}_m$) showed differences in each scenario as the heritability of methylation was different in each one of them.

The findings indicate that GOBLUP is capable of distinguishing between the epigenetic influences and the direct genetic effects present in individuals.

The average values of the epigenetic (methylation) effects ($\hat{\alpha}$), were negative for windows 1 and 2: -29.59 and -28.25 at $h^2 = 0.1$, -29.71 and -28.23 at $h^2 = 0.3$, and -31.14 and -28.52 at $h^2 = 0.8$ (Fig. 5-3). These negative estimates support our hypothesis and the generator model, highlighting the repressive impact of methylation on gene expression in promoter regions. Our results demonstrate that the larger and more negative effect estimates of methylation on the epigenetics-modulated trait were in windows 1 and 2, as expected. A higher methylation frequency on these genomic positions would negatively affect the epigenetic estimated breeding value of our population. In the remaining windows, the standard deviation of $\hat{\alpha}$ included 0 or were negligible. These results support our simulation, where the methylation effects on the phenotype were only considered for windows 1 and 2.

5.4.4. Comparison Between EBV From the Traditional and GOBLUP Models

The correlation between the traditional EBV and our proposed EEV was consistently greater than 0.90 (Fig. 5-4) across all heritability scenarios, with stronger correlations observed as the heritability of liability to methylation decreased. In scenarios of low to moderate methylation heritability, the correlation between our proposed EEV and the traditional EBV exceeded 0.98. However, with high heritability of liability to methylation, this correlation slightly decreased to 0.92.

The correlation of the additive residual breeding value ($\hat{\mathbf{a}}_r$) with the traditional EBV was lower than that obtained with the proposed EEV under moderate to high heritability for methylation liability, but the opposite was true for the low heritability scenario. Nonetheless, EEVs showed a higher correlation with the epigenetics-modulated TBV than EBV, except on the scenario low heritability scenario ($h_{meth}^2 = 0.1$), where the correlations were similar. An analogous trend was observed for the correlation of the epigenetic-mediated breeding value ($\hat{\mathbf{a}}_m$) with EEV and EBV. The epigenetics-modulated TBV had moderate correlations (~ 0.5 in the three scenarios) with $\hat{\mathbf{a}}_r$ and more variable correlations with $\hat{\mathbf{a}}_m$. This was affected by the heritability of liability to methylation: the higher this heritability, the stronger the correlation. An

explanation can be provided: $\hat{\mathbf{a}}_{\mathbf{m}}$ is calculated as $\hat{a}_m = \sum_{i=1} \hat{a}_w \hat{g}_{i,m_w}$ where the distribution of $\mathbf{g}_{m_w}$ depends on $\sigma_{m_w}^2$. Thus, higher heritability of methylation liability resulted in greater variance in the methylation windows. In scenarios with higher heritability for the liability to methylation, $\hat{\mathbf{a}}_{\mathbf{m}}$ significantly influences EEV rather than to $\hat{\mathbf{a}}_{\mathbf{r}}$. The variance of $\hat{\mathbf{a}}_{\mathbf{m}}$ increases with the heritability of liability to methylation, having a larger impact in the estimation of the EEV than $\hat{\mathbf{a}}_{\mathbf{r}}$. Consequently, liability to methylation impacts the epigenetic value of the individual in a more pronounced manner, especially with higher heritability. The correlation between $\hat{\mathbf{a}}_{\mathbf{r}}$ and $\hat{\mathbf{a}}_{\mathbf{m}}$ was ~ 0 in the three scenarios. This was also expected as these traits were assumed independent in the simulation.

These results indicated that EEVs obtained via GOBLUP showed higher correlations with the epigenetics-modulated TBV than the traditional EBVs obtained using BLUP, according to Christensen et al. (2021). Moreover, our findings confirmed that $\text{cor}(\hat{\mathbf{a}}_{\mathbf{r}}, \hat{\mathbf{a}}_{\mathbf{m}})$ approximates to 0 across the three heritability scenarios. This supports the conclusion that the multiomic strategy effectively separates additive genetic components from epigenetically mediated effects, although this may also be due to the assumptions in the generator model.

Epigenetics has been proposed as an explanation to the missing heritability conundrum due to its role in the inheritance of chromatin conformation and gene expression, as well as its involvement in the transgenerational inheritance of complex traits (Petronis, 2010). The issue of missing heritability points to the limited portion of actual heritability explained by genetic variants used in traditional models (Manolio et al., 2009; Zuk et al., 2012). However, the concept of epigenetic transgenerational inheritance in mammals has been misrepresented over the past decade, with few studies providing concrete evidence of its existence (Khatib et al., 2024). Epigenetic inheritance is known to be an intergenerational process, affecting up to three generations in females and two in males, due to the two waves of epigenetic reprogramming during embryonic development (Reik et al., 2001; Radford, 2018; Rotondo et al., 2021). Consequently, epigenetic inheritance does not follow conventional Mendelian laws. Given these considerations and our results, it appears that epigenetics may not hold the key to resolving the missing heritability problem, as traditional models might already capture a portion of epigenetic variance.

5.4.5. Implications and final remarks

Traditionally, phenotypic variation in plant and animal breeding is explained by estimating variance components through linear mixed models, which relate fixed and random effects and a residual to a phenotype Henderson (1984). Several models have been proposed to incorporate methylation data into these models. For instance, Gianola and van Kaam (2008) suggested using Reproducing Kernel Hilbert Spaces (RKHS), treating methylation as an independent variable in linear mixed models, adding a vector of nonparametric epigenetic coefficients and a kernel matrix to represent epigenetic similarities between individuals. (Breiman, 2001) included methylation information in Random Forest models, constructing classification trees based on a vector of methylation and genotypic information. Varona et al. (2015) expanded the linear model to estimate epigenetic variability by including a vector of epigenetic effects, assumed independent from genetic effects but sharing the same incidence matrix.

In this study, we focused on epigenetic marks within specific genomic regions, investigating their potential impact on phenotypes through gene expression inhibition or reduction. The epigenetic GOBLUP model separated epigenetic from additive genetic effects, allowing for independent estimation of additive residual breeding values and omics-mediated breeding values. These contribute to the EEV, enabling tailored selection intensities for specific breeding goals.

Distinguishing between these two types of genetic merit allows specific strategies in breeding programs. Greater emphasis on a_r in breeding objectives supports gradual and sustained trait improvement over time, akin to traditional breeding programmes. Alternatively, a stronger selection intensity can be applied to a_m for a quicker response in the mediated epigenetics-modulated trait of interest, provided that methylation effects on the phenotype enhance fitness under the experienced environmental conditions. This strategy is particularly effective when the environmental conditions inducing methylation (e.g., stressor A) are similar to the conditions faced by the population (e.g., stressor B). Under such circumstances, selecting for a_m allows populations to react more rapidly to specific external stressors while maintaining adaptive relevance.

Selection indices may adequately balance these different genetic characteristics in the breeding programs. The use of new sequencing techniques such as EpiGLOWS (González-Recio et al., 2023) is able to capture both genetic and

epigenetic information in the same sequencing process at low coverage, with a strong correlation with the actual genotyping arrays, easing the inclusion on methylation information in breeding programmes.

This study used simulated data, which simplifies the complex biological reality of DNA methylation. Admittedly, our simulation did not encompass all biological factors influencing methylation and its phenotypic effects, nor did it cover methylation across all genomic regions. The simulation assumed independence between genetic susceptibility to methylation and environmental factors, which may oversimplify their intricate relationship. Similarly, our focus on methylation as a repressor of gene expression potentially overlooked methylation events in other genomic areas or any positive effects within gene bodies.

Despite these limitations, the main goal of our study was to test the integration of methylation data into breeding programmes using the GOBLUP model, rather than exploring methylation's impact on different genomic regions. Future work will utilise larger datasets, such as those from the H2020 Rumigen Project's Epigenetic arrays, providing real methylation data and new biomarkers to better address these complexities.

5.5. Conclusion

The multiomic approach offered the possibility to distinguish the phenotypic variance explained by methylation marks from that explained by additive genetic effects. The proposed GOBLUP model for epigenetics allowed a comprehensive decomposition of variance component by accurately differentiating genetic and methylation variance. Additionally, it also generated two different genetic merits for the animals: an epigenetic merit and a direct additive genetic merit.

This distinction enables breeding programmes to apply tailored selection strategies, focusing on specific components of genetic merit based on the environmental influence on genetic regulation of traits. This approach is particularly valuable for addressing genotype-by-environment ($G \times E$) interactions, where environmental conditions impact epigenetic regulation. By leveraging these insights, breeding decisions can be optimised, enhancing selection outcomes under diverse management scenarios.

The emerging availability of technologies capable of capturing methylation data will drive efforts to estimate the epigenetic merit of breeding animals. Such

advances have the potential to improve the reliability of estimated breeding values and enable data-informed management practices. Moving forward, breeding programmes will need to develop targeted strategies to integrate methylation information effectively, leveraging its potential to enhance selection accuracy and adaptability in response to environmental challenges.

5.6. Tables

Table 5-1. Summary statistics of the methylation frequency for each of the thirteen genetic windows and the number of simulated SNPs with additive effect for the epigenetics-modulated trait and the liability to methylation.

Window	Mean methylation frequency	Standard deviation	Number of SNP for the epigenetics-modulated trait	Number of SNPs for liability to methylation
1	0.6179	0.0147	0	105
2	0.5357	0.0227	0	28
3	0.4702	0.0341	0	7
4	0.4112	0.0288	0	22
5	0.3423	0.0242	0	19
6	0.2559	0.0222	0	9
7	0.1843	0.0139	0	11
8	0.1145	0.0091	0	13
9	0.0873	0.0055	0	22
10	0.2029	0.0231	398	144
11	0.6205	0.0257	364	128
12	0.8191	0.0389	6738	2483
13	0.6414	0.0252	0	175
Non-genic region	--	--	2500	6834

Table 5-2. Estimated parameters by BLUP and GOBLUP for each heritability scenario. The variances and h2 estimates and their HPD Intervals (95%) across scenarios for the epigenetics-modulated traits are presented in the third, fourth and fifth columns as well as the estimated average heritability and their standard deviations across the thirteen windows in the GOBLUP model in the second column.

Trait	Scenario (h ²)	$\widehat{h^2}_{methylation}$ GOBLUP	$\hat{\sigma}^2_{gen\ mediated\ trait}$	$\hat{\sigma}^2_{res\ mediated\ trait}$	$\hat{h}^2_{retrieved\ mediated\ trait}$
Traditional BLUP	0.10	-	1.17 (0.99-1.34)	6.63 (6.44-6.81)	0.15 (0.13-0.17)
	0.30	-	1.41 (1.22-1.60)	6.49 (6.31-6.69)	0.18 (0.16-0.20)
	0.80	-	2.00 (1.78-2.22)	6.16 (5.98-6.35)	0.25 (0.22-0.27)
Epigenetic model (GOBLUP)	0.10	0.102 ± 0.01	1.01 (0.86-1.17)	6.09 (5.92-6.26)	0.14 (0.12-0.16)
	0.30	0.315 ± 0.01	1.01 (0.86-1.17)	6.09 (5.92-6.26)	0.14 (0.12-0.16)
	0.80	0.846 ± 0.01	1.01 (0.86-1.17)	6.09 (5.91-6.25)	0.14 (0.12-0.16)

5.7. Figures

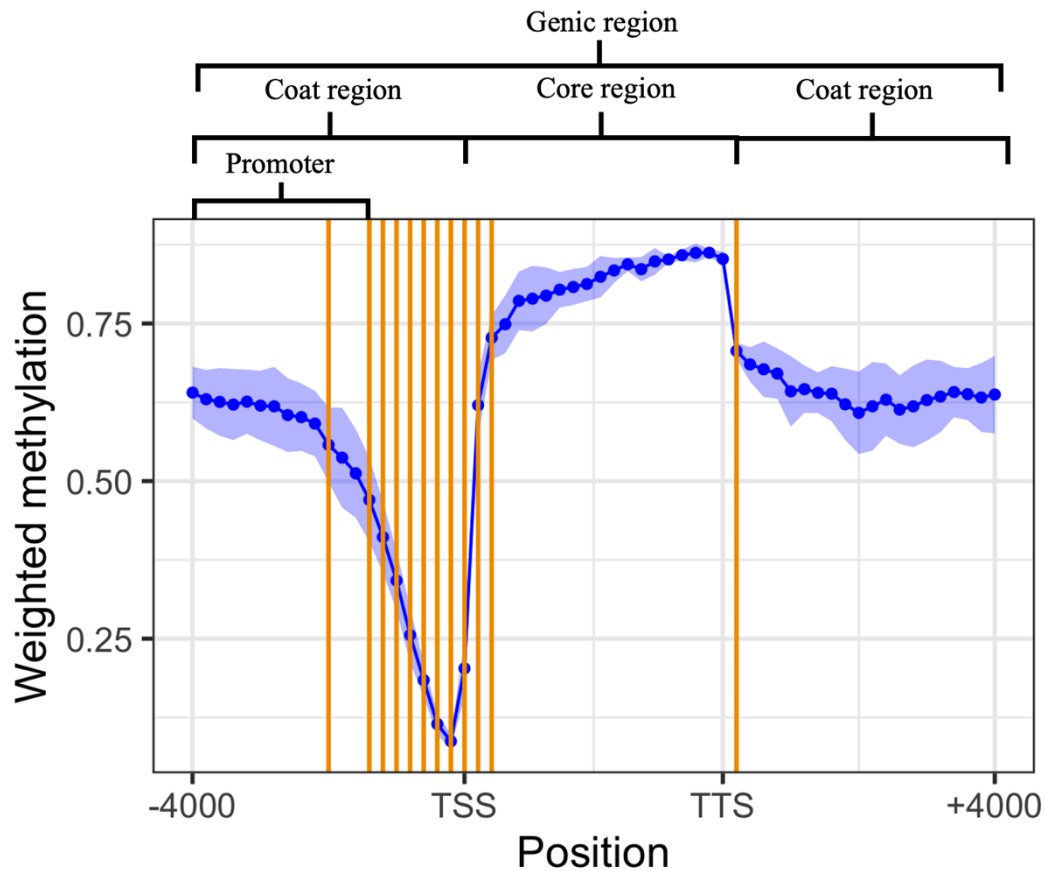


Figure 5-1. Average weighted methylation frequency of the samples sequenced by nanopore sequencing in genomic regions. Vertical lines limit the extension of the new windows where the sixty original windows were aggregated into new windows where the variability between an original window and its immediately following window was greater than 10%, resulting in thirteen windows whose mean and standard deviation were again estimated.

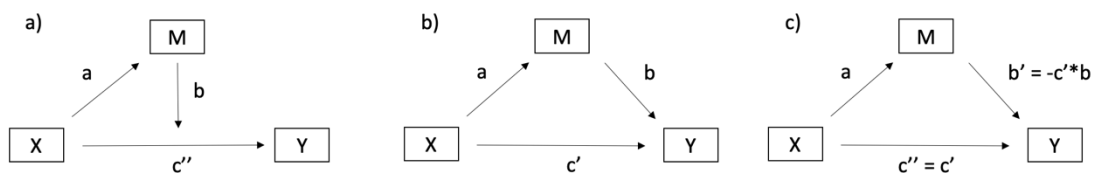


Figure 5-2. Scheme of how recursive effects of methylation over the phenotype are proposed in the generator (a) and statistical models (b). These models can be reparametrised to prove recursiveness (c), where b, is impacted by c' creating an effect of methylation marks on the phenotype combined with the animal's direct genetic effect.

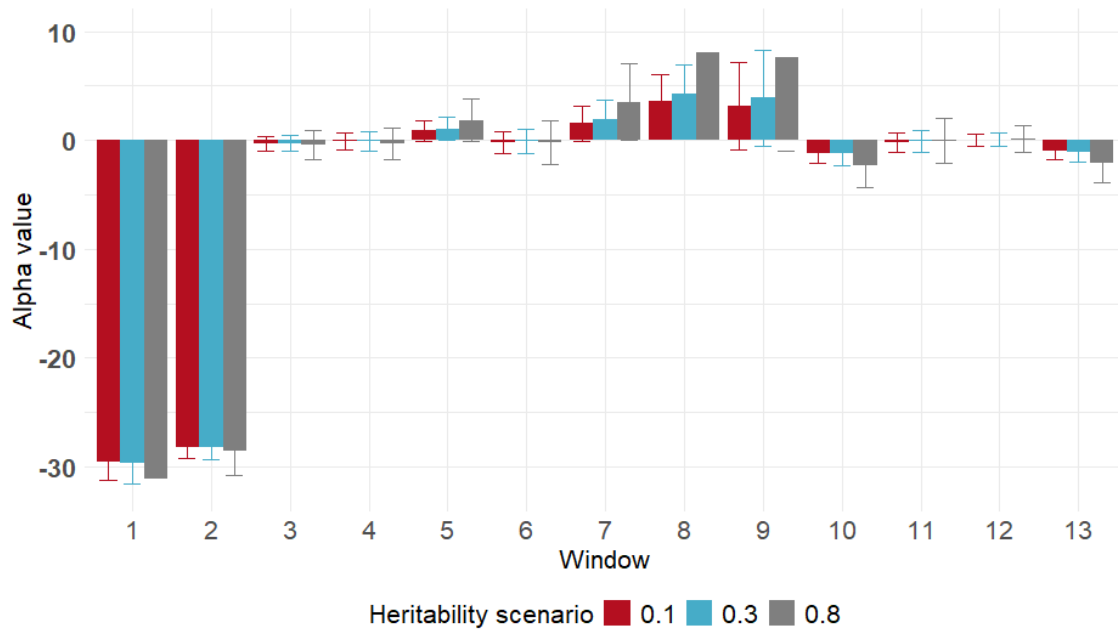


Figure 5-3. Values of alpha across the genetic windows estimated by the epigenetic model. Each colour represents the alpha value in a different scenario of heritability of liability to methylation simulated.

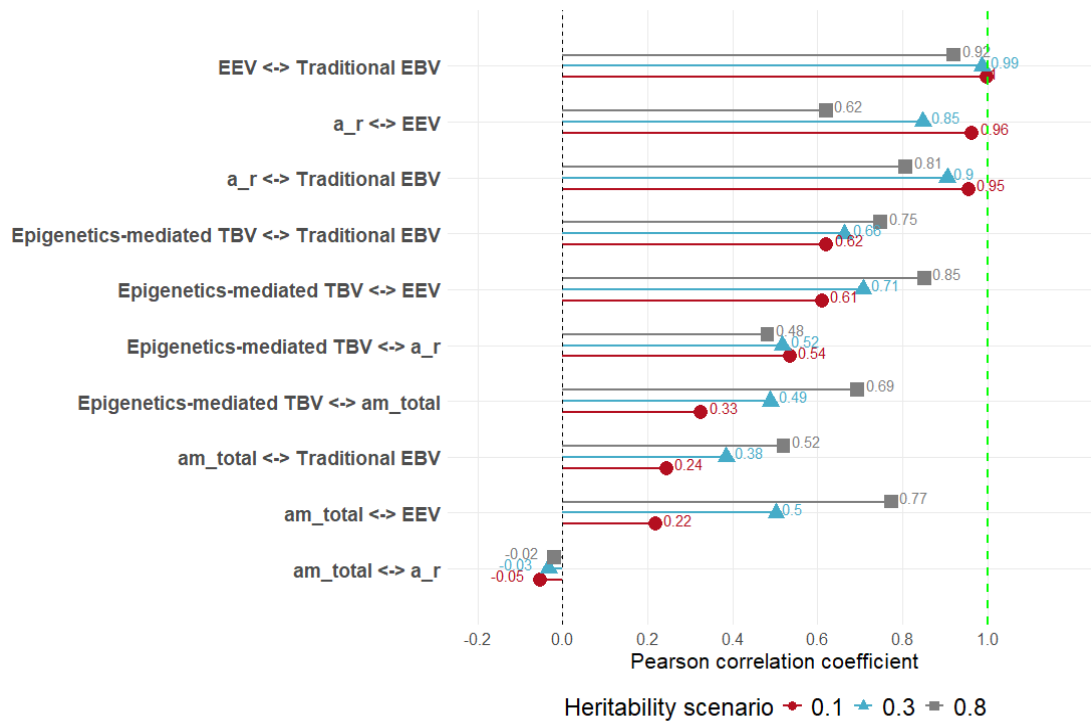


Figure 5-4. Genetic correlation estimated for each heritability scenario. The y axis represents the pairwise correlation between two features and the x-axis the Pearson correlation coefficient.

5.8. References

- Berbel-Filho, W. M., D. Rodríguez-Barreto, N. Berry, C. Garcia De Leaniz, and S. Consuegra. 2019. “Contrasting DNA Methylation Responses of Inbred Fish Lines to Different Rearing Environments.” *Epigenetics* 14: 939–948. <https://doi.org/10.1080/15592294.2019.1625674>.
- Bisutti, V., N. Mach, D. Giannuzzi, et al. 2023. “Transcriptome-Wide Mapping of Milk Somatic Cells Upon Subclinical Mastitis Infection
- Gamaarachchi, H., C. W. Lam, G. Jayatilaka, et al. 2020. “GPU Accelerated Adaptive Banded Event Alignment for Rapid Comparative Nanopore Signal Analysis.” *BMC Bioinformatics* 21: 1–13. <https://doi.org/10.1186/S12859-020-03697-X/FIGURES/6>.
- Gao, Y., Y. Chen, J. Qiao, J. Huang, and L. Wen. 2023. “DNA Methylation Protocol for Analyzing Cell-Free DNA in the Spent Culture Medium of Human Preimplantation Embryos.” *STAR Protocols* 4: 102247. <https://doi.org/10.1016/j.xpro.2023.102247>.
- Ghorbani, R., and R. Shokri-Gharelo. 2019. “Computational-Based Approaches in Epigenetic Research: Insights From Computational Tools, Mathematical Models, and Machine Learning Methods.” *Journal of Applied Biology & Biotechnology* 7: 79–83. <https://doi.org/10.7324/JABB.2019.70114>.
- Gianola, D., and J. B. C. H. M. van Kaam. 2008. “Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits.” *Genetics* 178: 2289–2303. <https://doi.org/10.1534/genetics.107.084285>.
- González-Recio, O., A. López-Catalina, R. Peiró-Pastor, A. Nieto-Valle, M. Castro, and A. Fernández. 2023. “Evaluating the Potential of (Epi) genotype-By-Low Pass Nanopore Sequencing in Dairy Cattle: A Study on Direct Genomic Value and Methylation Analysis.” *Journal of Animal Science and Biotechnology* 14: 98. <https://doi.org/10.1186/s40104-023-00896-3>.
- Gonzalez-Recio, O., M. A. Toro, and A. Bach. 2015. “Past, Present and Future of Epigenetics Applied to Livestock Breeding.” *Frontiers in Genetics* 6: 305. <https://doi.org/10.3389/fgene.2015.00305>.

Hayes, A. F. 2009. "Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium." *Communication Monographs* 76: 408–420. <https://doi.org/10.1080/03637750903310360>.

Hayes, B., and M. E. Goddard. 2001. "The Distribution of the Effects of Genes Affecting Quantitative Traits in Livestock." *Genetics, Selection, Evolution* 33: 1–21. <https://doi.org/10.1186/1297-9686-33-3-209>.

Henderson, C. R. 1984. *Applications of Linear Models in Animal Breeding*. Guelph, ON, USA: University of Guelph.

Hill, W. G. 2017. "Conversion' of Epistatic Into Additive Genetic Variance in Finite Populations and Possible Impact on Long-Term Selection Response." *Journal of Animal Breeding and Genetics* 134: 196–201. <https://doi.org/10.1111/jbg.12270>.

Hotchkiss, R. D. 1948. "The Quantitative Separation of Purines, Pyrimidines, and Nucleosides by Paper Chromatography." *Journal of Biological Chemistry* 175: 315–332. [https://doi.org/10.1016/S0021-9258\(18\)57261-6](https://doi.org/10.1016/S0021-9258(18)57261-6).

Ibeagha-Awemu, E. M., and X. Zhao. 2015. "Epigenetic Marks: Regulators of Livestock Phenotypes and Conceivable Sources of Missing Variation in Livestock Improvement Programs." *Frontiers in Genetics* 20: 302. <https://doi.org/10.3389/FGENE.2015.00302>.

Jensen, J. 2016. "Estimation of Genetic Variance in the Age of Genomics." *Journal of Animal Breeding and Genetics* 133: 333. <https://doi.org/10.1111/jbg.12235>.

Jjingo, D., A. B. Conley, S. V. Yi, V. V. Lunyak, and I. King Jordan. 2012. "On the Presence and Role of Human Gene-Body DNA Methylation." *Oncotarget* 3: 462–474. <https://doi.org/10.18632/ONCOTARGET.497>.

Jones, P. A. 2012. "Functions of DNA Methylation: Islands, Start Sites, Gene Bodies and Beyond." *Nature Reviews. Genetics* 13: 484–492. <https://doi.org/10.1038/nrg3230>.

Khatib, H., J. Townsend, M. A. Konkel, G. Conidi, and J. A. Hasselkus. 2024. "Calling the Question: What Is Mammalian Transgenerational Epigenetic Inheritance?" *Epigenetics* 19: 2333586. <https://doi.org/10.1080/15592294.2024.2333586>.

Kim, D.-Y., and J.-M. Kim. 2021. “Multi-Omics Integration Strategies for Animal Epigenetic Studies — A Review.” *Animal Bioscience* 34: 1271–1282. <https://doi.org/10.5713/ab.21.0042>.

Legarra, A., and O. F. Christensen. 2023. “Genomic Evaluation Methods to Include Intermediate Correlated Features Such as High-Throughput or Omics Phenotypes* *Presented as Part of the Breeding and Genetics Symposium: Beyond Genetic Markers—Additional Data to Improve Long-Term Selection Held at the ADSA Annual Meeting, June 2022.” *JDS Communication* 4: 55–60. <https://doi.org/10.3168/jdsc.2022-0276>.

Legarra, A., O. F. Christensen, I. Aguilar, and I. Misztal. 2014. “Single Step, a General Approach for Genomic Selection.” *Livestock Science* 166: 54–65. <https://doi.org/10.1016/j.livsci.2014.04.029>.

López-Catalina, A., A. Reverter, P. A. Alexandre, L. T. Nguyen, and O. González-Recio. 2024. “Stress-Induced Epigenetic Effects Driven by Maternal Lactation in Dairy Cattle: A Comethylation Network Approach.” *Epigenetics* 19: 2381856. <https://doi.org/10.1080/15592294.2024.2381856>.

Lowe, P., R. Olinski, and A. Ruzov. 2021. “Evidence for Noncytosine Epigenetic DNA Modifications in Multicellular Eukaryotes: An Overview.” In *DNA Modifications: Methods and Protocols*, edited by A. Ruzov and M. Gering, 15–25. New York, NY: Springer. https://doi.org/10.1007/978-1-0716-0876-0_2.

Ma, Y., and J. M. Ordovas. 2017. “The Integration of Epigenetics and Genetics in Nutrition Research for CVD Risk Factors.” *Proceedings of the Nutrition Society* 76: 333–346. <https://doi.org/10.1017/S0029665116000823>.

Manolio, T. A., F. S. Collins, N. J. Cox, et al. 2009. “Finding the Missing Heritability of Complex Diseases.” *Nature* 461: 747–753. <https://doi.org/10.1038/nature08494>.

Miranda, T. B., and P. A. Jones. 2007. “DNA Methylation: The Nuts and Bolts of Repression.” *Journal of Cellular Physiology* 213: 384–390. <https://doi.org/10.1002/JCP.21224>.

Misztal, I., S. Tsuruta, D. Lourenco, I. Aguilar, A. Legarra, and Z. Vitezica. 2014. *Manual for BLUPF90 Family of Programs*, 199. Athens, Greece: Athens University.

Ospelt, C. 2022. “A Brief History of Epigenetics.” *Immunology Letters* 249: 1–4. <https://doi.org/10.1016/J.IMLET.2022.08.001>.

Petronis, A. 2010. “Epigenetics as a Unifying Principle in the Aetiology of Complex Traits and Diseases.” *Nature* 465: 721–727. <https://doi.org/10.1038/nature09230>.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radford, E. J. 2018. “Exploring the Extent and Scope of Epigenetic Inheritance.” *Nature Reviews. Endocrinology* 14: 345–355. <https://doi.org/10.1038/s41574-018-0005-5>.

Reik, W., W. Dean, and J. Walter. 2001. “Epigenetic Reprogramming in Mammalian Development.” *Science* 293: 1089–1093. <https://doi.org/10.1126/science.1063443>.

Rotondo, J. C., C. Lanzillotti, C. Mazziotta, M. Tognon, and F. Martini. 2021. “Epigenetics of Male Infertility: The Role of DNA Methylation.” *Frontiers in Cell and Development Biology* 9: 9624. <https://doi.org/10.3389/fcell.2021.689624>.

Saborío-Montero, A., M. Gutiérrez-Rivas, A. López-García, et al. 2021. “Holobiont Effect Accounts for More Methane Emission Variance Than the Additive and Microbiome Effects on Dairy Cattle.” *Livestock Science* 250: 104538. <https://doi.org/10.1016/j.livsci.2021.104538>.

Sriraman, A., T. K. Debnath, B. Xhemalce, and K. M. Miller. 2020. “Making It or Breaking It: DNA Methylation and Genome Integrity.” *Essays in Biochemistry* 64: 687–703. <https://doi.org/10.1042/EBC20200009>.

Su, J., Y. Wang, X. Xing, J. Liu, and Y. Zhang. 2014. “Genome-Wide Analysis of DNA Methylation in Bovine Placentas.” *BMC Genomics* 15:12. <https://doi.org/10.1186/1471-2164-15-12>.

Teh, A. L., H. Pan, L. Chen, et al. 2014. “The Effect of Genotype and in Utero Environment on Interindividual Variation in Neonate DNA Methylomes.” *Genome Research* 24: 1064–1074. <https://doi.org/10.1101/gr.171439.113>.

Thomson, C. E., I. S. Winney, O. C. Salles, and B. Pujol. 2018. “A Guide to Using a Multiple-Matrix Animal Model to Disentangle Genetic and Nongenetic Causes of Phenotypic Variance.” *PLoS One* 13: e0197720. <https://doi.org/10.1371/journal.pone.0197720>.

Tian, H., N. Zheng, W. Wang, et al. 2016. “Integrated Metabolomics Study of the Milk of Heat-Stressed Lactating Dairy Cows.” *Scientific Reports* 6: 24208. <https://doi.org/10.1038/srep24208>.

Varona, L., S. Munilla, E. F. Mouresan, A. González-Rodríguez, C. Moreno, and J. Altarriba. 2015. “A Bayesian Model for the Analysis of Transgenerational Epigenetic Variation.” *G3: Genes, Genomes, Genetics* 5: 477–485. <https://doi.org/10.1534/g3.115.016725>.

Wang, Y., Y. Zhao, A. Bollas, Y. Wang, and K. F. Au. 2021. “Nanopore Sequencing Technology, Bioinformatics and Applications.” *Nature Biotechnology* 39: 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x>.

Wickham, H. 2009. “ggplot2: Elegant Graphics for Data Analysis.” *ggplot2*. <https://doi.org/10.1007/978-0-387-98141-3>.

Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander. 2012. “The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability.” *Proceedings of the National Academy of Sciences of the United States of America* 109: 1193–1198. <https://doi.org/10.1073/pnas.1119675109>.

6. General Discussion

This discussion synthesises the findings of the study, contextualising them within the broader field of epigenetics and livestock breeding. It highlights the key implications of the research, explores limitations, and proposes future research directions to bridge existing gaps.

6.1. Introduction to the General Discussion

The conventional perspective of biology is changing. Genes are no longer the blueprint for life; instead, the intricate interaction between genetic factors and environmental influences indicates that our biological predispositions are not strictly dictated by genetics but are influenced by a multitude of factors throughout an individual's lifespan. Our epigenome, the sum of the epigenetic marks along the genome, is significantly shaped by environmental factors, impacted by elements such as nutrition, stress, or exposure to toxicants, which may have ramifications for offspring and subsequent generations (Pogribny and Rusyn, 2013; Tiffon, 2018).

Owing to significant advancements in technology and the decrease in costs of different sequencing methods, genomic and –omics data has become increasingly accessible. Not only is genomic information readily available, but epigenetic modifications can also be easily detected and analysed, elucidating how environmental determinants impact gene expression and contribute to the complexity of biological systems.

This section provides a summary and critical analysis of the research conducted during this PhD project. It focuses on three main topics: (1) establishing a comprehensive workflow to detect genome-wide methylation marks, including both a wet lab benchmark and a bioinformatics pipeline; (2) examining the association between methylation rates across different genomic regions and phenotypic traits; and (3) exploring how methylation marks can be integrated into breeding programmes.

In **Chapter 1**, we performed a comparative analysis of CpG sites identified using Oxford Nanopore Sequencing and Reduced Representation Bisulphite Sequencing. The correlation between the methylation frequencies estimated by these two methodologies exceeded 0.95 across overlapping CpG sites, demonstrating a high degree of concordance in the methylation patterns detected. This study established the methodology underpinning subsequent analyses in this thesis and has since been adopted as the standard approach for detecting methylation marks in our laboratory.

Chapter 2 explored the influence of dams' lactation status on the genome-wide methylation profile of their offspring. As illustrated in Fig. 6-1, this relationship is complex, with maternal phenotypes contributing to the methylation status of various genomic regions. Additionally, the individual's phenotype influences its own methylation profile, and methylation at one genomic position can modulate the methylation rate at other positions. To incorporate these findings into breeding programmes, we propose integrating methylation data into linear mixed models for estimating breeding values, employing a multi-omics approach.

Chapter 3 addresses the incorporation of methylation information into traditional statistical models for genetic evaluation and genetic parameters estimations, highlighting the advantages of this integration. Conventional genetic methodologies predominantly estimate breeding values using genomic information (Dekkers, 2012). However, the emerging field of epigenetics introduces an additional layer of complexity that significantly influences phenotypic expression yet remains underrepresented in traditional genetic models (Wang et al., 2012). The integration of epigenetic data through the GOBLUP framework enables a clearer distinction between genetic effects and epigenetic modifications, thereby improving the precision and accuracy of estimated breeding values.

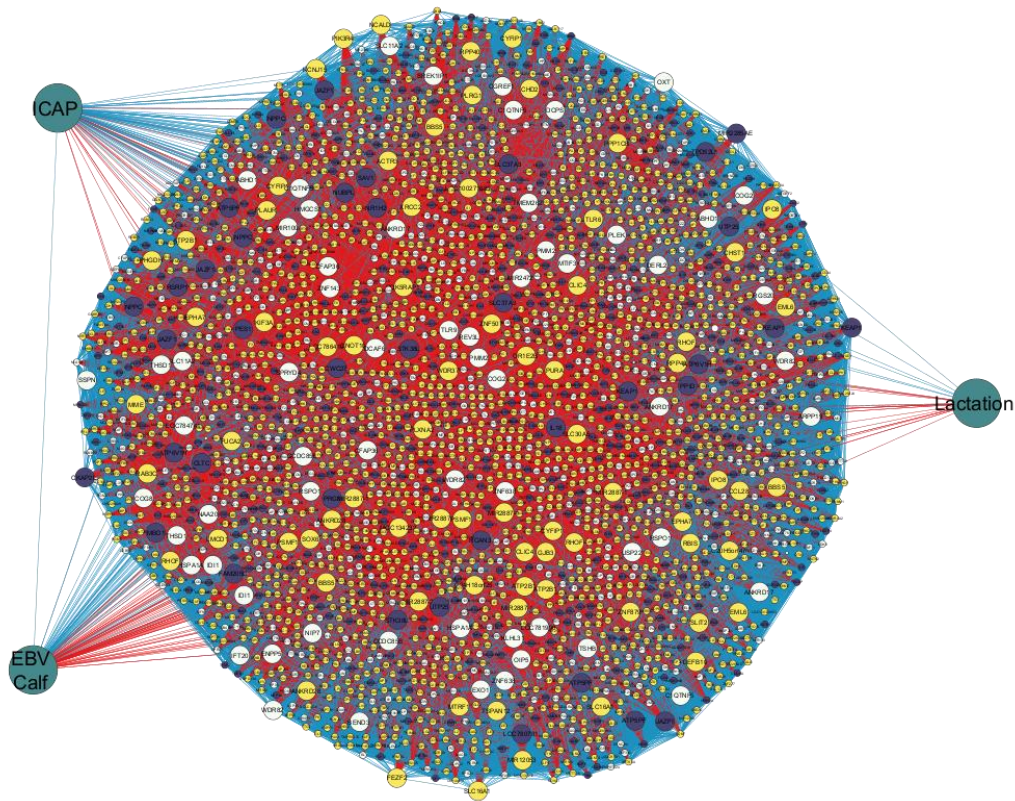


Figure 6-1. Co-methylation network associating phenotypes of interest in animal production with different genomic regions. Source: López-Catalina et al. (2024).

6.2. Identification of Epigenetic Marks in Livestock Populations

The first significant challenge encountered during this PhD project was the implementation of a novel sequencing method to detect genome-wide methylation information. When the project began in 2020, scientific literature on epigenetic studies utilising Nanopore Sequencing was scarce, particularly concerning livestock species. This limited availability of resources required a comprehensive and systematic evaluation of every step of the process, from identifying the most effective DNA extraction kit to optimising the bioinformatics pipelines for data analysis. Each component had to be rigorously tested and refined to maximise yield and ensure that the sequencing data provided the most accurate and valuable insights possible.

To evaluate the accuracy and compatibility of different sequencing methods for methylation analysis, we sequenced a single sample using whole-genome bisulphite sequencing and Oxford Nanopore Technology. We then compared the correlation of methylation frequencies detected by each method across five coverage thresholds: 1x, 5x, 10x, 15x, and 20x (Fig. 6-2). Using WGBS at an approximate coverage of 30x and a Q-score of 38, we identified 52,409,895 CpG sites. For ONT sequencing, employing the now-discontinued SQK-LSK109 kit and an R9.4 flow cell, we were able to identify 21,694,649 CpG sites with a mean coverage of 7x and a Q-score of 9.9 (López-Catalina et al., 2022).

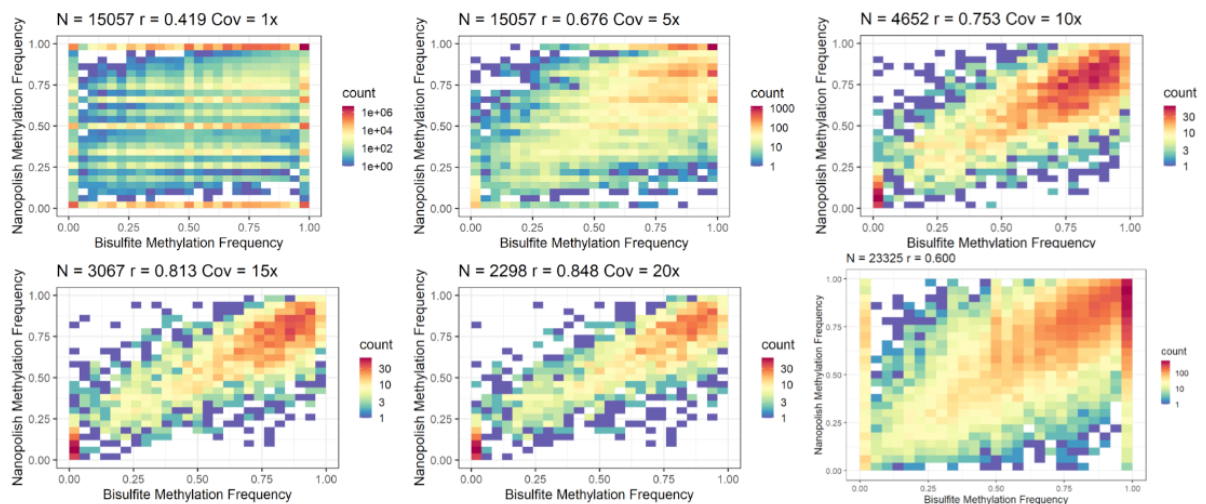


Figure 6-2. Different combinations of coverage and their respective correlations (r) and sites detected (N). Correlation filtered for 1x, 5x, 10x, 15x, 20x and last image illustrates filtering ONT sites by 15x and Illumina for 1x.

This initial comparison, conducted in 2021, positioned ONT as a promising yet still developing technique, competing with the established WGBS. At the time, ONT required improvements in Q-score and coverage. By 2024, advancements in ONT technology significantly improved its performance, as evidenced by the analysis presented in **Chapter 1**. Upgrading to the SQK-LSK110 library preparation kit and the Guppy base-calling software (version 5.0.14) markedly enhanced ONT's capabilities. These upgrades enabled the detection of more CpGs with ONT than with Reduced Representation Bisulphite Sequencing, with 8.12x more sites identified by ONT. However, despite its broader coverage, ONT's average coverage per site (7x) remained lower than the 22x achieved by RRBS. Notably, ONT demonstrated its strength in identifying CpGs in distal intergenic regions, whereas RRBS was limited to promoter regions. Nonetheless, both methods identified differentially methylated cytosines in promoter regions associated with similar traits, highlighting their complementary potential in epigenetic analyses.

The synergy of both methodologies discussed in **Chapter 1** provides a constructive outlook on the animal breeding field. In this research, CpGs within diverse genomic regions were identified. This data aligns with one of the primary objectives of the H2020 RUMIGEN project, the creation of an epi-genotyping array tailored for *Bos taurus*. Methylation arrays represent a powerful tool in livestock breeding during the post-genomic era, enabling the analysis of epigenetic modifications that influence traits without altering the underlying DNA sequence. These arrays can be particularly useful when integrated with genomic selection strategies, as they provide insights into how environmental factors may affect gene expression and, consequently, phenotypic traits.

The integration of methylation and SNP arrays is expected to enhance the accuracy of genomic selection and promote precisely tuned molecular genetic breeding in livestock, potentially leading to improved management interventions and performance outcomes. For instance, the mammalian methylation array has enabled epigenome-wide association studies across multiple ruminant species, identifying age-related methylation sites (Arneson et al., 2022). These arrays complement existing SNP genotyping arrays, which have been widely used in animal genetics and breeding (Wang et al., 2023b). Species-specific arrays are expected to have a greater impact, as they allow for the selection of critical CpG positions associated with phenotypes that enhance the health, production, or adaptation of a given species.

Additionally, we contributed to RUMIGEN's Work Packages 6, "Epigenetic biomarkers to improve selection and precision farming"; WP7, "Impact of pre- and post-natal farming practices on foetal programming and epigenetic regulation"; and WP8, "Integration/modelling: towards sustainable and socially acceptable breeding programmes and management practices". The publication forming the first chapter of this doctoral thesis facilitated the identification of a list of candidate CpG sites associated with metabolic stress, heat stress, mastitis and semen fertility in bulls. These CpG positions were integrated with those identified by INRAE, which sequenced the same samples using RRBS. The complementarity of the two techniques, as highlighted in the publication, enabled the development of a more comprehensive list of candidate CpGs associated with various traits of interest in animal breeding.

This list was subsequently expanded with CpGs of interest reported in the literature to delineate the genomic coordinates affected by differential methylation in cattle. The resulting list was instrumental in creating the first epi-genotyping array for dairy cattle, developed in collaboration with Illumina. This array includes 48,351 CpGs, approximately 22.5% of which were identified by our team using Oxford Nanopore Sequencing. The first arrays arrived in mid-to-late 2024 and are currently being used.

The introduction of these methylation arrays represents a significant advancement in reducing the costs associated with epi-genotyping. As of December 2024, the cost per sample using the array is approximately €65 for ~50,000 positions. In contrast, genome-wide methylation site detection using Nanopore Sequencing at a 10x coverage requires three samples per PromethION flow cell, amounting to approximately €250 per sample, including DNA extraction and library preparation. This cost-effectiveness underscores the array's potential to accelerate epigenetic research and its application in livestock breeding programmes.

While the cost of genome-wide methylation detection using Nanopore Sequencing is substantially higher than that of methylation arrays, the advantages it offers can justify the expense in certain contexts. Nanopore Sequencing provides unparalleled genome-wide coverage, enabling the identification of CpG sites across the entire genome, including intergenic and regulatory regions often overlooked by pre-designed arrays. This is particularly valuable in exploratory studies where the genomic regions of interest are not yet

well characterized, as it facilitates the discovery of novel epigenetic markers associated with traits of interest.

For breeding programmes, where cost-effectiveness is paramount, methylation arrays offer a highly practical solution for routine large-scale epi-genotyping. The high costs associated with genome-wide sequencing currently render it less applicable in this context. However, alternative approaches such as low-pass sequencing (González-Recio et al., 2023) hold promise for significantly reducing costs, potentially enabling genome-wide applications to become a routine component of breeding programmes. Conversely, during research and development phases, where the primary objectives include uncovering the underlying mechanisms of epigenetic regulation and identifying novel biomarkers, the comprehensive insights provided by genome-wide detection methods such as Nanopore Sequencing are invaluable. These methods allow for the exploration of epigenetic patterns across the entire genome, facilitating the discovery of new regulatory regions and traits. Ultimately, the decision to employ either methylation arrays or genome-wide sequencing is contingent on the specific goals of the study, the available resources, and the stage of implementation within the breeding pipeline. Combining these technologies strategically may offer the best balance between cost-effectiveness and discovery potential.

Having established the presence and relevance of epigenetic marks in livestock, the discussion now turns to the specific maternal factors influencing these marks, particularly the impact of lactation.

6.3. Maternal Lactation Influences on the Methylation Profile of Offspring

The second chapter examined the peculiarities of gestation in dairy cattle. The metabolic stress experienced by the foetus arises from the concurrent occurrence of lactation and gestation in multiparous cows. This temporal overlap not only presents significant challenges to the metabolic well-being of the cows but also impacts the long-term viability and productivity of their progeny (Ceciliani et al., 2018; Sundrum, 2015). Lactating cows must sustain milk production, and insufficient nutritional intake can exacerbate these stresses, perpetuating a cycle of adverse health outcomes for both the dam and her offspring (Knight, 2001).

Stress levels have been linked to changes in DNA methylation and gene expression profiles in cows, influencing pathways related to nervous system

activity and immune function (Del Corvo et al., 2020). Factors such as parental age, maternal metabolism, and in vitro culture conditions are known to affect epigenetic programming in gametes and embryos, resulting in similar transcriptomic and methylation patterns (Wu and Sirard, 2020). DNA methylation has also been implicated in the acute regulation of milk protein gene expression during mastitis and mammary involution, as well as in potential transgenerational effects on milk production (Singh et al., 2012).

One of the most significant challenges lies in understanding the pleiotropic effects of methylation changes. Epigenetic modifications often influence multiple pathways, and altering methylation at a single locus could have cascading effects on unrelated traits. Advances in multi-omic integration and systems biology approaches could help disentangle these complex interactions, providing a clearer roadmap for targeted epigenetic interventions.

In this chapter, we contributed to the existing literature by quantifying the impact of maternal lactation on the methylome of gestated calves. We assessed the influence of maternal milk lactation, along with calves' estimated breeding values and individual body condition scores, on the methylome through the construction of a co-methylation network (Fig. 6-1). In **Chapter 2**, we identified a total of 50,277 differentially methylated cytosines and 2,281 differentially methylated regions between calves born to lactating dams and those born to non-lactating cows, substantiating our preliminary hypothesis regarding the influence of metabolic stress on embryonic development proposed by González-Recio et al. (2012). Annotation of these differentially methylated sites revealed that approximately 20% of the hypomethylated DMCs and DMRs in the non-lactating group were located in promoter regions, which may be associated with increased gene expression compared to calves born to lactating cows.

In immune-related genes such as *PRG3*, hypermethylation in calves from non-lactating dams suggests suppressed granulocyte activation and inflammatory responses (Gray et al., 2022). This could compromise immune resilience, making animals more susceptible to infections. Reducing methylation (demethylation) in these regions may activate immune-related pathways, potentially improving the animal's capacity to respond to environmental challenges such as pathogens or stress. Similarly, for genes like *LOC100271685*, where hypomethylation correlates with altered immune response signalling, maintaining balanced methylation levels may prevent hyperactivation of immune pathways (Zuccolo et al., 2010), which could lead to autoimmune issues or chronic inflammation.

In the context of *ALCAM*, which plays roles in neural development and haematopoiesis (Swart, 2002), hypomethylation observed in calves from non-lactating dams suggests enhanced plasticity in these pathways. While this might benefit adaptability in early development, prolonged hypomethylation could dysregulate neural or organ development. Controlled methylation adjustments could fine-tune these processes, promoting robust developmental outcomes without overexpression of potentially deleterious pathways.

Production traits, particularly those linked to milk yield and quality, also exhibit strong associations with methylation. For instance, the positive correlation between hypomethylation of *RHOF* and milk EBV, suggests that reduced methylation is associated with the genetic potential of milk protein synthesis and amino acid transport. Further demethylation of this region could enhance milk yield and quality, offering a practical target for improving dairy productivity. However, excessive demethylation could lead to metabolic imbalances, potentially increasing susceptibility to conditions like ketosis or milk fever. Therefore, fine-tuning methylation rather than complete demethylation may yield optimal results.

Similarly, *INHBA* and *KLHL1* represent promising targets for balancing growth, feed efficiency, and milk production (Pierce et al., 2020; Sun et al., 2019). The negative correlation between *INHBA* methylation and milk EBV suggests that demethylating this region could enhance production efficiency, while the regulatory role of *KLHL1* in multiple networks highlights its potential as a key target for interventions. Future research should explore how modulating methylation in these genes affects not only production traits but also energy balance, reproductive success, and longevity.

Methylation patterns associated with fertility and structural traits also hold significant potential. For instance, *SLC37A3* and *COL1A2* demonstrated positive correlations between methylation and ICAP. Increased methylation in *SLC37A3* could enhance fertility and skeletal development, addressing two critical traits in dairy cattle. Similarly, methylation of *COL1A2*, associated with collagen synthesis and muscle development, could improve bone structure and durability, indirectly supporting longevity and resilience in high-yielding cows. However, hypermethylation in these genes should be approached with caution to avoid unintended suppression of beneficial pathways, suggesting that a moderate increase in methylation might be the most advantageous.

The decision to methylate or demethylate specific regions must consider the interplay between methylation levels, gene function, and trait expression. For health-related genes like *PRG3*, demethylation may activate suppressed immune pathways, enhancing disease resistance. Conversely, over-demethylation of production-related genes such as *RHOF* could lead to overexpression, potentially resulting in metabolic overload. A balanced approach is therefore critical, aiming to achieve optimal methylation levels that promote health and productivity without compromising long-term viability.

The differential methylation patterns observed in key genes underscore the importance of understanding the functional consequences of methylation and demethylation on molecular and phenotypic traits. Epigenetic modifications like DNA methylation do not alter the genetic code but influence gene expression, offering a reversible mechanism to optimise traits without permanent genetic changes. However, the decision to methylate or demethylate specific regions requires careful evaluation of the biological context and downstream effects.

The development of high-throughput epi-genotyping arrays, such as the epichip described in this thesis, represents a transformative opportunity to explore these questions at scale. These tools will enable the identification of critical methylation sites associated with multiple traits and provide a foundation for precision epigenetic editing. Techniques such as CRISPR-dCas9-based epigenome editing (Cai et al., 2023; Policarpi et al., 2024) could allow researchers to experimentally manipulate methylation levels at specific loci, directly assessing their functional impacts on traits of interest.

These epigenotyping arrays facilitated the expansion of our hypothesis by increasing the sample size to 482 animals with recorded maternal lactation status. These animals were also genotyped using traditional arrays, and an epigenetic genome-wide association study (EWAS) was performed. While the model requires further refinement, these preliminary results identified 144 significant CpGs associated with maternal lactation status (Fig. 6-3). Functional annotation of these CpGs using the PANTHER Classification System (Mi et al., 2013; Thomas et al., 2022) revealed their involvement in critical biological processes, including regulation, response to stimuli, developmental processes, cellular processes, and biological regulation. These findings offer novel insights into the potential molecular mechanisms by which maternal lactation status influences the offspring's epigenetic landscape and highlight the relevance of integrating epigenetic data into studies of complex traits in livestock.

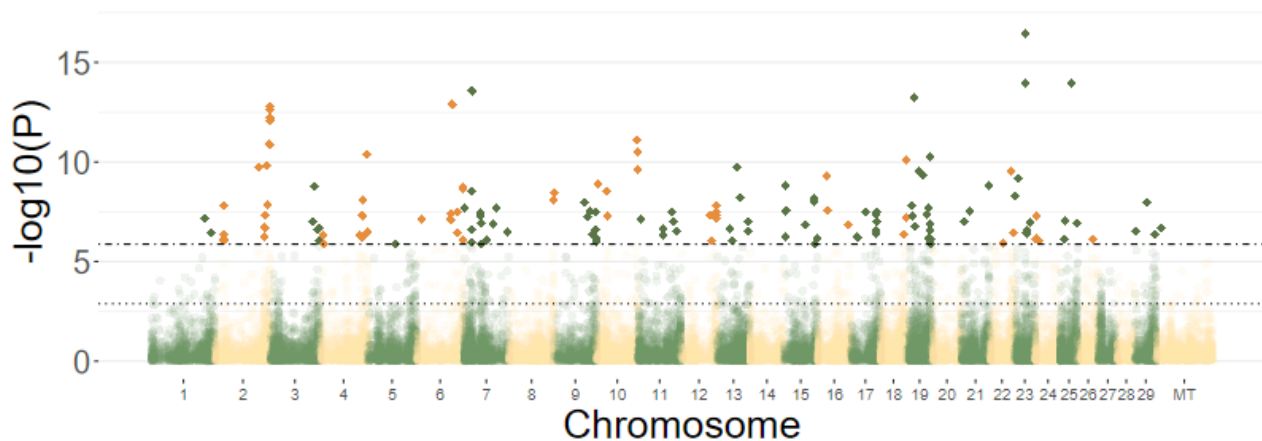


Figure 6-3. Epigenetic genome-wide association study performed using the RUMIGEN's epichip to associate maternal lactation status with the offspring's methylation status. The horizontal red line represents the FDR-adjusted p-value and the blue the Bonferroni-adjusted p-value.

The findings in this study do not fully satisfy the requirements to categorise the observed epigenetic modifications as TEI if we follow the criteria outlined by Khatib et al. (2024) for defining transgenerational epigenetic inheritance that require: (1) the inheritance of phenotypes by the first unexposed generation following environmental factors applied to the F0 generation, (2) the inheritance of epigenetic marks in the first unexposed generation as a result of these environmental factors, (3) the persistence of the same epimutations across multiple generations, (4) observable gene expression changes in subsequent generations, and (5) confirmation of germ cell involvement in each generation. Instead, the study investigates intergenerational epigenetic inheritance, as both the F0 dams and their F1 offspring were exposed to the same external stressors during gestation. Therefore, we cannot yet confirm the inheritance of these marks across generations. Rather, our findings reflect their transmissibility within a single generation, highlighting the potential for environmental factors to influence the epigenome of the directly exposed offspring.

We demonstrated molecular-level differences in the methylation profiles of calves born to lactating versus non-lactating cows. Nonetheless, interpreting these networks within the context of breeding programmes remains challenging. Epigenetic alterations are influenced by multiple factors beyond those analysed, including environmental conditions and direct genetic effects on an animal's own methylome. Furthermore, methylation alone does not fully account for phenotypic

variation; other epigenetic mechanisms, environmental, nutritional, and health-related factors likely contribute to these traits. The integration of this information should adopt a quantitative perspective, recognising that a large number of methylation effects exert only small impacts on traits of interest. Additionally, non-genetic components must also be considered, akin to the infinitesimal model proposed by Fisher (Barton et al., 2017; Fisher, 1918). This approach facilitates the incorporation of epigenetic data into breeding strategies while accounting for the complexity and multifactorial nature of phenotype determination. Incorporating epigenetic information into breeding programmes represents a paradigm shift in livestock management. By integrating methylation data with traditional SNP-based genomic evaluations, breeding schemes can account for both genetic predispositions and environmentally influenced modifications. This approach could lead to more sustainable and adaptable dairy systems, balancing productivity with health and welfare. Future research should also explore the transgenerational effects of methylation changes, assessing their stability and heritability across generations to ensure long-term benefits.

In conclusion, the ability to modulate methylation levels offers exciting possibilities for improving dairy cattle health and productivity. Whether to methylate or demethylate specific regions depends on the biological context and the desired outcomes for health, production, or adaptability. Advances in epigenotyping technologies and molecular editing tools will pave the way for precision interventions, fostering a new era of molecular genetic breeding in livestock.

The observed effects of maternal lactation on offspring methylation profiles underscore the need to integrate such data into breeding programmes. The next subsection explores how this integration could transform genetic evaluations.

6.4. Integrating Epigenetic Data Into Breeding Programmes: Application of the GOBLUP Model

To incorporate the findings from **Chapter 2** into breeding programmes, we propose integrating methylation data into linear mixed models for estimating breeding values through a multi-omic approach. **Chapter 3** details the methods and benefits of incorporating epigenetic information into traditional breeding programmes.

The inclusion of DNA methylation data offers a transformative potential for enhancing the precision of genetic evaluations. While traditional models, such as BLUP, primarily rely on genomic data, they often overlook the additional complexity introduced by epigenetic modifications. This gap can be addressed through models like GOBLUP, as proposed by Christensen et al. (2021), which incorporates methylation data into multi-omic evaluations. By accounting for both genetic and epigenetic contributions to phenotypic variance, this approach can significantly improve the accuracy of selection decisions, particularly for traits influenced by environmental and epigenetic factors.

Despite the potential benefits, traditional breeding programmes have yet to fully utilise epigenetic information, particularly regarding the influence of environmental and epigenetic factors on phenotypic expression. While most methylation marks are reset during gametogenesis and early embryonic development to acquire totipotency, some marks are retained and can influence the offspring's phenotype even without direct exposure to the original environmental stressor (Morgan et al., 1999). In this context, methylation marks could be treated similarly to SNPs in genomic evaluations, as they are *inherited* by the offspring rather than directly *transmitted* from one generation to the next. A classic example of this transgenerational epigenetic inheritance is the agouti viable yellow A^{vy} allele in mice, which causes variations in coat colour due to the insertion of a retrotransposon (Wolff et al., 1998). Some mice inherit this allele from their mothers but show wild type colour as the promoter is silenced by methylation marks.

In human genetics, a well-documented instance of external stressors influencing health and epigenetics is the Dutch Famine of 1944-1945, during which the German occupation of the Netherlands led to widespread food and fuel shortages (Roseboom et al., 2011). Prenatal exposure to this famine has been linked to numerous adverse health outcomes, including lower birth weight, increased risks of cardiovascular and metabolic disorders later in life, higher susceptibility to schizophrenia, and transgenerational effects such as elevated neonatal adiposity (Roseboom et al., 2011). Furthermore, Heijmans et al. (2008) demonstrated that famine exposure during early gestation resulted in persistent hypomethylation of the insulin-like growth factor 2 (*IGF2*) gene in offspring, with these epigenetic changes observed even six decades later. This was among the first findings to highlight how transient environmental exposures during critical developmental windows can leave long-lasting epigenetic imprints.

Although the concept of TEI remains a subject of debate within the scientific community, with calls for more rigorous studies to confirm its mechanisms and scope, examples like the methylation changes in *IGF2* during the Dutch Famine and the epigenetic regulation of the *A^{vy}* allele in mice underscore the importance of this research. It is still necessary to found and explore inheritable epigenetic marks in livestock populations, although these findings encourage continued exploration of epigenetic inheritance and its implications for health, resilience, and phenotypic plasticity (Khatib et al., 2024), and enable the opportunity as adding epigenetic information into breeding decisions whether the marks are maintained during the next generations, adding a cumulative epigenetic gain to the current genetic objectives.

Traditional genetic models estimate breeding values primarily using genomic markers (Dekkers, 2012). However, they fail to address the role of epigenetics in phenotypic expression. This omission limits the precision of breeding value estimates for traits with multifactorial inheritance. By incorporating epigenetic data, GOBLUP provides a robust framework for distinguishing between purely genetic effects and those mediated by epigenetic modifications (Wang et al., 2012). The integration of epigenetic data through the GOBLUP framework facilitates a more nuanced differentiation between the effects attributable purely to genetic factors and those resulting from epigenetic modifications, thereby significantly enhancing the precision and accuracy of EBVs.

The concept of epigenetic influences is crucial in addressing the “missing heritability” problem in relation to various livestock characteristics, where genetic markers alone are insufficient to fully explain the heritability of complex traits with multifactorial inheritance patterns (Ibeagha-Awemu and Zhao, 2015; Trerotola et al., 2015). This problem highlights the discrepancy between the heritability estimates from traditional quantitative methods, GWAS studies, and the real heritability of complex traits (Sandoval-Motta et al., 2017). The difference between the gaps of variance assumed by different methods is referred to as “missing” (Matthews and Turkheimer, 2022). The contribution of rare genetic variants has been largely underestimated due to conventional models that do not account for allele frequency in affected populations, leading to recalibrated estimates that suggest a much higher heritability contribution from these variants (Nagao, 2024). Some researchers argue that the problem can be largely dissolved by recognizing that traditional methods include non-additive genetic variance and transgenerational epigenetic factors, which GWAS do not account for (Bourrat and

Lu, 2017). It has been proposed in human genetics that cultural factors may also interact with epigenetic mechanisms, suggesting that both biological and cultural heritages shape developmental outcomes and contribute to the missing heritability of complex diseases (Wallace, 2017). Thus, epigenetics serves as a crucial link between environmental influences and genetic expression, potentially elucidating the complexities of heritability in various traits and diseases. However, our results in **Chapter 3** highlighted that traditional models for estimating heritability might already account for some of the variance attributed to epigenetics, suggesting that epigenetics alone may not resolve the missing heritability problem. Even though, a larger sample size and real data would be necessary to deepen in this hypothesis. The incorporation of epigenetic components has significant potential to account for some of this unexplained variability, particularly for traits strongly influenced by environmental factors, thereby fostering a more comprehensive understanding of genetic inheritance.

By distinguishing the additive genetic value from the omics-mediated breeding value, breeding programmes are enabled to implement targeted selection pressures that are specifically tailored to particular environmental or management conditions (Gonzalez-Recio et al., 2015). For example, emphasising the epigenetic component may yield beneficial outcomes in environments marked by rapidly changing conditions, which foster a strong interaction between epigenetic factors and environmental conditions. This approach introduces a novel framework that could potentially facilitate more rapid and adaptive genetic responses to challenges such as climate change, thereby significantly enhancing the resilience of livestock populations.

6.5. Limitations of This PhD Project

Integrating methylation epigenetic marks into statistical models offers intriguing possibilities for enhancing breeding programmes, though it also presents significant challenges. One approach is to treat methylation status at specific genomic sites as a phenotype acting with recursion over the trait, as this thesis shows in **Chapter 3**. This involves measuring methylation levels and using them as selection criteria alongside other traits. However, the impact on selection accuracy may be limited, and cost-effective methylation measurement methods are essential for practical implementation.

Models including the reset coefficient, representing the gain and loss rates of methylation frequency (Gonzalez-Recio, 2012), need to be explored to include the epigenetic reprogramming of the offspring cells. Some alternatives such as those from Slatkin (2009) to include epigenetic variation and from Tal et al. (2010) to estimate the epigenetic covariance between relatives have already been proposed.

Another promising strategy is incorporating methylation data into an epigenetic relationship matrix (T). A Bayesian mixed model using a T matrix could quantify the covariance between individuals based on epigenetic similarities. Methylation profiles could inform the construction of this matrix, with similarity measures reflecting methylation patterns (Varona et al., 2015). To refine this approach, specific methylation marks could be weighted based on their stability and heritability, ensuring that more stable marks exert greater influence on the matrix. Additionally, the model could account for methylation-specific reset coefficients, reflecting the proportion of marks erased during meiosis, thereby improving the representation of transgenerational inheritance.

Still, efforts should focus on characterising the environmental factors that influence the methylation rates of genomic regions relevant to animal breeding. By understanding how external forces modulate gene expression, it may be possible to develop new management and husbandry practices that promote a favourable methylome without necessarily incorporating this information into breeding schemes. Recent advancements in the field of agri-epigenomics have demonstrated that the epigenome is highly responsive not only to genetic factors but also to nutrition. Such findings highlight the potential to enhance the expression of beneficial genes or silence detrimental ones through targeted nutritional and environmental interventions.

Epigenetics is a rapidly advancing field with significant potential in agricultural sciences, offering solutions to future challenges such as rapid adaptation to abrupt climate changes, enhanced disease resistance, and improved resilience in livestock populations. While this study provides valuable insights into the potential of integrating epigenetic information into livestock breeding, several limitations must be acknowledged to contextualise the findings and identify areas for future improvement. 1) DNA methylation is a highly tissue-specific process, meaning that findings from one tissue type cannot always be generalised to others. Methylation profiles can vary considerably across different tissues due to their distinct biological functions, and this specificity may limit the interpretation of epigenetic modifications in relation to the phenotypes of interest. While there is

some methylation specificity within tissues, this is mainly related to cellular differentiation. We hypothesize that there still could be some global methylation shared between different tissues. In this study, the choice of tissue types was strongly influenced by practical and ethical constraints, as the research was conducted in collaboration with commercial farms and without access to an experimental facility. Chapter 1 employed semen samples from six bulls (three sub-fertile and three fertile), while Chapter 2 focused on blood samples from six calves (three born to lactating cows and three to non-lactating cows). Tissues such as the brain may be more biologically relevant for some of the processes discussed, however it was not feasible to collect such samples because it requires the sacrifice of animals—an approach incompatible with commercial farm operations and animal welfare standards. Blood was therefore selected as a non-invasive, ethically appropriate, and logistically feasible alternative for field-based studies. Although some tissue-specificity is expected, we proposed to use blood as a proxy for global methylation. Some of the epi-markers influenced by environmental stressors can be shared across tissues. Certainly, the effect of this methylation can differently affect the cell phenotype, specific to each tissue. In this case, two tissues can methylate similarly, but only one will express the phenotype. Therefore, there would still be some association between the methylation pattern in blood, and the animal phenotype. More importantly, the main objective of Chapter 2 was not to investigate the underlying biological mechanisms in depth, but rather to explore the potential of using blood-derived DNA methylation as a source of epigenetic biomarkers that could be integrated into animal breeding programmes. In this context, blood provides a highly practical matrix, suitable for large-scale applications and longitudinal sampling within livestock production systems. Notably, some of the biomarkers identified in this study have already contributed to the design of a custom methylation array (EpiChip), which is currently being used within the RUMIGEN project to support population-wide epigenetic studies.

2) the small sample size in both chapters introduces challenges in the statistical power and extrapolation of the results. While these numbers are typical in exploratory livestock studies due to resource constraints, they limit the ability to detect subtle epigenetic changes and increase the risk of type I (false positive) and type II errors (false negative). Additionally, the “large p , small n ” problem, where the number of CpG sites (p) vastly exceeds the number of samples (n), complicates the identification of true epigenetic signals. This imbalance can lead to overfitting and highlights the need for robust validation in independent datasets. The limited

sample size could be addressed in future studies through collaborative data-sharing initiatives, thereby enhancing statistical power

These limitations underscore the need for continued methodological advancements, larger and more diverse sample sizes, and innovative approaches to studying epigenetic inheritance in livestock species. Despite these challenges, this work provides a solid foundation for further exploration of epigenetic mechanisms in animal breeding. Addressing these limitations opens new avenues for future research, which are discussed in the concluding section.

6.6. Final Thoughts and Future Research Lines

Epigenetic information has now emerged as a valuable resource that can be utilised within the domain of animal sciences, providing researchers with a deeper understanding of the complex biological mechanisms at play. In this contemporary post-genomic era, characterised by a significant shift in research focus that extends beyond genomics to encompass a wide array of disciplines including epigenetics, structural variant analysis, proteomics, metabolomics, metagenomics, and a multitude of other -omics fields (Debnath et al., 2010); the acquisition of all this diverse and intricate information can, however, impose considerable financial burdens on research institutions and laboratories. Nanopore Sequencing introduces a new methodology for the sequencing of biological samples, as it facilitates the simultaneous obtaining of whole-genome information, methylome, and even the variome (Romagnoli et al., 2023; Watson et al., 2022) of an individual organism during a single sequencing run, while maintaining a remarkably low sequencing depth that ensures the resultant data exhibits a high degree of reliability and accuracy (González-Recio et al., 2023).

Epigenetic studies present a transformative opportunity for the development of innovative methodologies concerning selection and adaptation, while simultaneously acknowledging the enduring benefits of cumulative and sustained genetic advancements that traditional breeding strategies can provide; however, these studies also propose new and potentially more effective approaches that enable animals to swiftly adjust to and thrive under abrupt and challenging environmental conditions (Devaskar and Raychaudhuri, 2007; Gotoh, 2014)

In this thesis, our primary objective was to facilitate the introduction of novel sources of information that could significantly enhance the animal breeding sector, all while considering and addressing the diverse and evolving needs of this

particular sector. Due to the anticipated increase in global temperatures, livestock populations will face major challenges in terms of their strength and ability to recover, highlighting the need for immediate focus on their adaptability. Animals must have the ability to quickly adapt to the escalating effects of climate change. Furthermore, the emergence of new diseases, which are anticipated to correlate with these shifting climatic conditions, places the health and well-being of animals at considerable risk, thereby underscoring the urgency of the situation (Hoffmann, 2010; Naskar et al., 2015; Prasad and Sejian, 2015).

We propose both wet-lab and dry-lab pipelines for sequencing and analysing animal samples using nanopore technologies equivalent to the current gold standard bisulphite sequencing. To perform differential methylation and enrichment analyses between groups, we used the R packages DSS and ChIPseeker as well as g:profiler. Using the PCIT algorithm and Cytoscape, the thesis determined the correlation between the methylation profiles of different genomic regions of calves and the lactation status of their dams. We concluded that the lactation status of the cow had an impact on the methylome of her offspring in a molecular approach. The application of these results in the breeding sector is challenging because a variety of factors influence the methylation frequency of a region or cytosine. We have proposed a quantitative approach to incorporate these findings into breeding programmes, where epigenetic information is incorporated into the estimation of breeding values. We found that our multi-omics approach allows us to split breeding value into direct breeding value and omics-mediated breeding value, allowing the application of two different selection intensities.

The findings detailed in this thesis are aligned with the objectives of Work Packages 6, 7, and 8 within the H2020 RUMIGEN project. These aims include identifying epigenetic markers to enhance selective breeding and precision agriculture, examining the impact of pre- and post-natal farming practices on foetal programming and epigenetic regulation, and integrating/modelling sustainable and socially responsible breeding programs and management practices. Collaborations with external research teams and published studies have enabled the expansion of our epigenetics research across various animal species, including water buffalo (Cascone et al., 2024a; Cascone et al., 2024b). The development of the EpiChip has provided a larger dataset for analysing methylation patterns from birth to adulthood, facilitated by a three-year monitoring period with regular blood sample collection from calves. Furthermore,

the use of Nanopore Sequencing has enabled international collaborations to determine and characterise the structural variome of different cattle species.

In conclusion, this study advances our understanding of epigenetic influences in livestock, particularly the intergenerational effects of maternal lactation on offspring. While challenges remain, the integration of epigenetic data into breeding programmes represents a promising frontier in animal genetics, offering the potential to enhance productivity, sustainability, and adaptability in dairy cattle.

7. General Conclusions

1. The study confirms the feasibility of long-read sequencing for epigenetics. Oxford Nanopore sequencing allows direct detection of methylation without bisulphite conversion, simplifying protocols and preserving DNA integrity, especially useful in livestock research settings.
2. Oxford Nanopore Technologies is a valid alternative to RRBS. ONT provided broader genome coverage and showed high concordance with Reduced Representation Bisulphite Sequencing.
3. Maternal lactation status significantly alters offspring DNA methylation. Calves born to lactating cows exhibited over 50,000 differentially methylated cytosines and 2,281 differentially methylated regions, indicating a strong influence of maternal metabolic stress on the epigenome of the newborn calf.
4. These methylation changes are associated with relevant traits in dairy production. The DMRs were enriched in genes linked to immune function, metabolism, and production traits such as milk yield and body capacity index, suggesting potential long-term phenotypic changes if not reverted during adulthood.
5. Maternal metabolic stress induces transmissible-like epigenetic marks, acting as an intergenerational effect with a detectable epigenetic footprint in the F1 generation.
6. Actual trans-generational inheritance requires further confirmation with specific experimental design.
7. A recursive model integrating methylation data allows accounting for variance due to epigenetic effects.
8. The GOBLUP model allowed for proper partitioning of additive and epigenetic variance due to methylation effects.
9. Incorporating DNA methylation into genomic selection models via a GOBLUP approach faintly improved the accuracy of estimated breeding values in comparison to the traditional BLUP model.

8. References

- Abdalla, H., Yoshizawa, Y., Hochi, S., 2009. Active demethylation of paternal genome in mammalian zygotes. *J Reprod Dev* 55, 356–360. <https://doi.org/10.1262/jrd.20234>
- Akalın, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology* 13, 1–9. <https://doi.org/10.1186/GB-2012-13-10-R87/FIGURES/6>
- Albini, S., Zakharova, V., Ait-Si-Ali, S., 2019. Chapter 3 - Histone Modifications, in: Palacios, D. (Ed.), *Epigenetics and Regeneration, Translational Epigenetics*. Academic Press, pp. 47–72. <https://doi.org/10.1016/B978-0-12-814879-2.00003-0>
- Angeloni, A., Fissette, S., Kaya, D., Hammond, J.M., Gamaarachchi, H., Deveson, I.W., Klose, R.J., Li, W., Zhang, X., Bogdanovic, O., 2024. Extensive DNA methylome rearrangement during early lamprey embryogenesis. *Nat Commun* 15, 1977. <https://doi.org/10.1038/s41467-024-46085-2>
- Anway, M.D., Cupp, A.S., Uzumcu, M., Skinner, M.K., 2005. Epigenetic Transgenerational Actions of Endocrine Disruptors and Male Fertility. *Science* 308, 1466–1469. <https://doi.org/10.1126/science.1108190>
- Arneson, A., Haghani, A., Thompson, M.J., Pellegrini, M., Kwon, S.B., Vu, H., Maciejewski, E., Yao, M., Li, C.Z., Lu, A.T., Morselli, M., Rubbi, L., Barnes, B., Hansen, K.D., Zhou, W., Breeze, C.E., Ernst, J., Horvath, S., 2022. A mammalian methylation array for profiling methylation levels at conserved sequences. *Nat Commun* 13, 783. <https://doi.org/10.1038/s41467-022-28355-z>
- Banovich, N.E., Lan, X., McVicker, G., Geijn, B. van de, Degner, J.F., Blischak, J.D., Roux, J., Pritchard, J.K., Gilad, Y., 2014. Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. *PLOS Genetics* 10, e1004663. <https://doi.org/10.1371/journal.pgen.1004663>
- Barton, N.H., Etheridge, A.M., Véber, A., 2017. The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* 118, 50–73. <https://doi.org/10.1016/j.tpb.2017.06.001>
- Bártová, E., Krejčí, J., Harničarová, A., Galiová, G., Kozubek, S., 2008. Histone Modifications and Nuclear Architecture: A Review. *J Histochem Cytochem* 56, 711–721. <https://doi.org/10.1369/jhc.2008.951251>
- Batistel, F., 2020. 145 Nutritional Advances in Fetal and Neonatal Development: Methyl Donors. *Journal of Animal Science* 98, 122–123. <https://doi.org/10.1093/jas/skaa054.209>
- Bello, H.K., Gbolagade, K.A., 2017. A Survey of Human Deoxyribonucleic Acid. *Current Journal of Applied Science and Technology* 1–10. <https://doi.org/10.9734/BJAST/2017/32463>
- Berbel-Filho, W.M., Rodríguez-Barreto, D., Berry, N., Garcia De Leaniz, C., Consuegra, S., 2019. Contrasting DNA methylation responses of inbred fish lines to different rearing environments. *Epigenetics* 14, 939–948. <https://doi.org/10.1080/15592294.2019.1625674>
- Bernstein, B.E., Meissner, A., Lander, E.S., 2007. The Mammalian Epigenome. *Cell* 128, 669–681. <https://doi.org/10.1016/j.cell.2007.01.033>

- Berry, D.P., Lonergan, P., Butler, S.T., Cromie, A.R., Fair, T., Mossa, F., Evans, A.C.O., 2008. Negative Influence of High Maternal Milk Production Before and After Conception on Offspring Survival and Milk Production in Dairy Cattle. *Journal of Dairy Science* 91, 329–337. <https://doi.org/10.3168/jds.2007-0438>
- Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., Whitelaw, E., 2006. Dynamic Reprogramming of DNA Methylation at an Epigenetically Sensitive Allele in Mice. *PLOS Genetics* 2, e49. <https://doi.org/10.1371/journal.pgen.0020049>
- Bonduriansky, R., Day, T., 2009. Nongenetic Inheritance and Its Evolutionary Implications. *Annual Review of Ecology, Evolution, and Systematics* 40, 103–125. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173441>
- Bouquet, A., Juga, J., 2013. Integrating genomic selection into dairy cattle breeding programmes: a review. *animal* 7, 705–713. <https://doi.org/10.1017/S1751731112002248>
- Bourrat, P., Lu, Q., 2017. Dissolving the Missing Heritability Problem. *Philosophy of Science* 84, 1055–1067. <https://doi.org/10.1086/694007>
- Bovine Genome Sequencing and Analysis Consortium, Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Elnitski, L., Guigó, R., Hamernik, D.L., Kappes, S.M., Lewin, H.A., Lynn, D.J., Nicholas, F.W., Reymond, A., Rijnkels, M., Skow, L.C., Zdobnov, E.M., Schook, L., Womack, J., Alioto, T., Antonarakis, S.E., Astashyn, A., Chapple, C.E., Chen, H.-C., Chrast, J., Câmara, F., Ermolaeva, O., Henrichsen, C.N., Hlavina, W., Kapustin, Y., Kiryutin, B., Kitts, P., Kokocinski, F., Landrum, M., Maglott, D., Pruitt, K., Sapojnikov, V., Searle, S.M., Solovyev, V., Souvorov, A., Ucla, C., Wyss, C., Anzola, J.M., Gerlach, D., Elhaik, E., Graur, D., Reese, J.T., Edgar, R.C., McEwan, J.C., Payne, G.M., Raison, J.M., Junier, T., Kriventseva, E.V., Eyraes, E., Plass, M., Donthu, R., Larkin, D.M., Reecy, J., Yang, M.Q., Chen, L., Cheng, Z., Chitko-McKown, C.G., Liu, G.E., Matukumalli, L.K., Song, J., Zhu, B., Bradley, D.G., Brinkman, F.S.L., Lau, L.P.L., Whiteside, M.D., Walker, A., Wheeler, T.T., Casey, T., German, J.B., Lemay, D.G., Maqbool, N.J., Molenaar, A.J., Seo, S., Stothard, P., Baldwin, C.L., Baxter, R., Brinkmeyer-Langford, C.L., Brown, W.C., Childers, C.P., Connelley, T., Ellis, S.A., Fritz, K., Glass, E.J., Herzig, C.T.A., Iivanainen, A., Lahmers, K.K., Bennett, A.K., Dickens, C.M., Gilbert, J.G.R., Hagen, D.E., Salih, H., Aerts, J., Caetano, A.R., Dalrymple, B., Garcia, J.F., Gill, C.A., Hiendleder, S.G., Memili, E., Spurlock, D., Williams, J.L., Alexander, L., Brownstein, M.J., Guan, L., Holt, R.A., Jones, S.J.M., Marra, M.A., Moore, R., Moore, S.S., Roberts, A., Taniguchi, M., Waterman, R.C., Chacko, J., Chandrabose, M.M., Cree, A., Dao, M.D., Dinh, H.H., Gabisi, R.A., Hines, S., Hume, J., Jhangiani, S.N., Joshi, V., Kovar, C.L., Lewis, L.R., Liu, Y.-S., Lopez, J., Morgan, M.B., Nguyen, N.B., Okwuonu, G.O., Ruiz, S.J., Santibanez, J., Wright, R.A., Buhay, C., Ding, Y., Dugan-Rocha, S., Herdandez, J., Holder, M., Sabo, A., Egan, A., Goodell, J., Wilczek-Boney, K., Fowler, G.R., Hitchens, M.E., Lozado, R.J., Moen, C., Steffen, D., Warren, J.T., Zhang, J., Chiu, R., Schein, J.E., Durbin, K.J., Havlak, P., Jiang, H., Liu, Y., Qin, X., Ren, Y., Shen, Y.,

- Song, H., Bell, S.N., Davis, C., Johnson, A.J., Lee, S., Nazareth, L.V., Patel, B.M., Pu, L.-L., Vattathil, S., Williams, R.L., Curry, S., Hamilton, C., Sodergren, E., Wheeler, D.A., Barris, W., Bennett, G.L., Eggen, A., Green, R.D., Harhay, G.P., Hobbs, M., Jann, O., Keele, J.W., Kent, M.P., Lien, S., McKay, S.D., McWilliam, S., Ratnakumar, A., Schnabel, R.D., Smith, T., Snelling, W.M., Sonstegard, T.S., Stone, R.T., Sugimoto, Y., Takasuga, A., Taylor, J.F., Van Tassell, C.P., Macneil, M.D., Abatepaulo, A.R.R., Abbey, C.A., Ahola, V., Almeida, I.G., Amadio, A.F., Anatriello, E., Bahadue, S.M., Biase, F.H., Boldt, C.R., Carroll, J.A., Carvalho, W.A., Cervelatti, E.P., Chacko, E., Chapin, J.E., Cheng, Y., Choi, J., Colley, A.J., de Campos, T.A., De Donato, M., Santos, I.K.F. de M., de Oliveira, C.J.F., Deobald, H., Devinoy, E., Donohue, K.E., Dovic, P., Eberlein, A., Fitzsimmons, C.J., Franzin, A.M., Garcia, G.R., Genini, S., Gladney, C.J., Grant, J.R., Greaser, M.L., Green, J.A., Hadsell, D.L., Hakimov, H.A., Halgren, R., Harrow, J.L., Hart, E.A., Hastings, N., Hernandez, M., Hu, Z.-L., Ingham, A., Iso-Touru, T., Jamis, C., Jensen, K., Kapetis, D., Kerr, T., Khalil, S.S., Khatib, H., Kolbehdari, D., Kumar, C.G., Kumar, D., Leach, R., Lee, J.C.-M., Li, C., Logan, K.M., Malinverni, R., Marques, E., Martin, W.F., Martins, N.F., Maruyama, S.R., Mazza, R., McLean, K.L., Medrano, J.F., Moreno, B.T., Moré, D.D., Muntean, C.T., Nandakumar, H.P., Nogueira, M.F.G., Olsaker, I., Pant, S.D., Panzitta, F., Pastor, R.C.P., Poli, M.A., Poslusny, N., Rachagani, S., Ranganathan, S., Razpet, A., Riggs, P.K., Rincon, G., Rodriguez-Osorio, N., Rodriguez-Zas, S.L., Romero, N.E., Rosenwald, A., Sando, L., Schmutz, S.M., Shen, L., Sherman, L., Southey, B.R., Lutzow, Y.S., Sweedler, J.V., Tammen, I., Telugu, B.P.V.L., Urbanski, J.M., Utsunomiya, Y.T., Verschoor, C.P., Waardenberg, A.J., Wang, Z., Ward, R., Weikard, R., Welsh, T.H., White, S.N., Wilming, L.G., Wunderlich, K.R., Yang, J., Zhao, F.-Q., 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324, 522–528. <https://doi.org/10.1126/science.1169588>
- Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S.B., Krstic, P.S., Lindsay, S., Ling, X.S., Mastrangelo, C.H., Meller, A., Oliver, J.S., Pershin, Y.V., Ramsey, J.M., Riehn, R., Soni, G.V., Tabard-Cossa, V., Wanunu, M., Wiggin, M., Schloss, J.A., 2008. The potential and challenges of nanopore sequencing. *Nat Biotechnol* 26, 1146–1153. <https://doi.org/10.1038/nbt.1495>
- Breiman, L., 2001. Random Forests. *Machine Learning* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brosnan, C.A., Voinnet, O., 2009. The long and the short of noncoding RNAs. *Current Opinion in Cell Biology, Nucleus and gene expression* 21, 416–425. <https://doi.org/10.1016/j.ceb.2009.04.001>
- Burgers, E.E.A., Goselink, R.M.A., Bruckmaier, R.M., Gross, J.J., Jorritsma, R., Kemp, B., Kok, A., van Knegsel, A.T., 2023. Effect of voluntary waiting period on metabolism of dairy cows during different phases of the lactation. *Journal of Animal Science* 101, skad194. <https://doi.org/10.1093/jas/skad194>
- Burton, N.O., Greer, E.L., 2022. Multigenerational epigenetic inheritance: Transmitting information across generations. *Seminars in Cell &*

- Developmental Biology, Special Issue: Mesoderm differentiation in vertebrate development and regenerative medicine by Fiona Wardle / Special Issue: Genomic and epigenetic stability and inheritance by Karen Yuen 127, 121–132. <https://doi.org/10.1016/j.semcd.2021.08.006>
- Cai, R., Lv, R., Shi, X., Yang, G., Jin, J., 2023. CRISPR/dCas9 Tools: Epigenetic Mechanism and Application in Gene Transcriptional Regulation. *Int J Mol Sci* 24, 14865. <https://doi.org/10.3390/ijms241914865>
- Carlson, M., 2019. org.Bt.eg.db: Genome wide annotation for Bovine. R package version 3.8.2 [WWW Document]. URL <https://bioconductor.org/packages/release/data/annotation/html/org.Bt.eg.db.html> (accessed 3.7.23).
- Cascone, I., López-Catalina, A., Albarella, S., Ciotola, F., Cosenza, G., Salzano, A., Peretti, V., González-Recio, Ó., 2024a. Methylation Study in the Mediterranean Italian River Buffalo (*Bubalus bubalis*): Preliminary Results, in: 25th International Colloquium on Animal Cytogenetics and Genomics. Naples, Italy.
- Cascone, I., López-Catalina, A., González-Recio, Ó., Pugliano, M., D'Anza, E., Albarella, S., Ciotola, F., Peretti, V., 2024b. An Overview of DNA Methylation in *Mytilus galloprovincialis* through Nanopore Technology, in: 25th International Colloquium on Animal Cytogenetics and Genomics. Naples, Italy.
- Charfeddine, N., Pérez-Cabal, M.A., 2017. Effect of claw disorders on milk production, fertility, and longevity, and their economic impact in Spanish Holstein cows. *Journal of Dairy Science* 100, 653–665. <https://doi.org/10.3168/jds.2016-11434>
- Chen, Y., Steeneveld, W., Frankena, K., Leemans, I., Aardema, H., Vos, P.L.A.M., Nielen, M., Hostens, M., 2024. Association between days post-conception and lactation persistency in dairy cattle. *Journal of Dairy Science* 107, 5794–5804. <https://doi.org/10.3168/jds.2023-24282>
- Chen, Z., Zhang, Y., 2020. Role of Mammalian DNA Methyltransferases in Development. *Annu Rev Biochem* 89, 135–158. <https://doi.org/10.1146/annurev-biochem-103019-102815>
- Christensen, O.F., Börner, V., Varona, L., Legarra, A., 2021. Genetic evaluation including intermediate omics features. *Genetics* 219, iyab130. <https://doi.org/10.1093/genetics/iyab130>
- Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., Jacobsen, S.E., 2008. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219. <https://doi.org/10.1038/nature06745>
- Cole, J.B., Dürr, J.W., Nicolazzi, E.L., 2021. *Invited review*: The future of selection decisions and breeding programs: What are we breeding for, and who decides? *Journal of Dairy Science* 104, 5111–5124. <https://doi.org/10.3168/jds.2020-19777>
- Coleman, D.N., Alharthi, A.S., Liang, Y., Lopes, M.G., Lopreiato, V., Vailati-Riboni, M., Lóor, J.J., 2021. Multifaceted role of one-carbon metabolism on immunometabolic control and growth during pregnancy, lactation and

- the neonatal period in dairy cattle. *Journal of Animal Science and Biotechnology* 12, 27. <https://doi.org/10.1186/s40104-021-00547-5>
- Coppedè, F., Angela Lopomo, R.S., 2014. Genetic and epigenetic biomarkers for diagnosis, prognosis and treatment of colorectal cancer. *World Journal of Gastroenterology* 20, 943–956. <https://doi.org/10.3748/wjg.v20.i4.943>
- Costes, V., Chaulot-Talmon, A., Sellem, E., Perrier, J.-P., Aubert-Frambourg, A., Jouneau, L., Pontlevoy, C., Hozé, C., Fritz, S., Boussaha, M., Le Danvic, C., Sanchez, M.-P., Boichard, D., Schibler, L., Jammes, H., Jaffrézic, F., Kiefer, H., 2022. Predicting male fertility from the sperm methylome: application to 120 bulls with hundreds of artificial insemination records. *Clinical Epigenetics* 14, 54. <https://doi.org/10.1186/s13148-022-01275-x>
- Danchin, É., Charmantier, A., Champagne, F.A., Mesoudi, A., Pujol, B., Blanchet, S., 2011. Beyond DNA: integrating inclusive inheritance into an extended theory of evolution. *Nat Rev Genet* 12, 475–486. <https://doi.org/10.1038/nrg3028>
- David, I., Canario, L., Combes, S., Demars, J., 2019. Intergenerational Transmission of Characters Through Genetics, Epigenetics, Microbiota, and Learning in Livestock. *Front. Genet.* 10. <https://doi.org/10.3389/fgene.2019.01058>
- Daxinger, L., Whitelaw, E., 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* 13, 153–162. <https://doi.org/10.1038/nrg3188>
- 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. <https://doi.org/10.1093/BIOINFORMATICS/BTY149>
- De Coster, W., Rademakers, R., 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39, btad311. <https://doi.org/10.1093/bioinformatics/btad311>
- Dean, W., Lucifero, D., Santos, F., 2005. DNA methylation in mammalian development and disease. *Birth Defects Res C Embryo Today* 75, 98–111. <https://doi.org/10.1002/bdrc.20037>
- Deaton, A.M., Bird, A., 2011. CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022. <https://doi.org/10.1101/gad.2037511>
- Debnath, M., Prasad, G.B.K.S., Bisen, P.S., 2010. Molecular Diagnosis in the Post Genomic and Proteomic Era, in: Debnath, M., Prasad, G.B.K.S., Bisen, P.S. (Eds.), *Molecular Diagnostics: Promises and Possibilities*. Springer Netherlands, Dordrecht, pp. 97–109. https://doi.org/10.1007/978-90-481-3261-4_7
- Dechow, C.D., Liu, W.-S., 2018. DNA methylation patterns in peripheral blood mononuclear cells from Holstein cattle with variable milk yield. *BMC Genomics* 19, 744. <https://doi.org/10.1186/s12864-018-5124-9>
- Dekkers, J.C.M., 2012. Application of Genomics Tools to Animal Breeding. *Curr Genomics* 13, 207–212. <https://doi.org/10.2174/138920212800543057>
- Del Corvo, M., Bongiorno, S., Stefanon, B., Sgorlon, S., Valentini, A., Ajmone Marsan, P., Chillemi, G., 2020. Genome-Wide DNA Methylation and Gene

- Expression Profiles in Cows Subjected to Different Stress Level as Assessed by Cortisol in Milk. *Genes* 11, 850. <https://doi.org/10.3390/genes11080850>
- Dennis, E.M., Garcia, D.M., 2022. Biochemical Principles in Prion-Based Inheritance. *Epigenomes* 6, 4. <https://doi.org/10.3390/epigenomes6010004>
- Denoyelle, L., de Villemereuil, P., Boyer, F., Khelifi, M., Gaffet, C., Alberto, F., Benjelloun, B., Pompanon, F., 2021. Genetic Variations and Differential DNA Methylation to Face Contrasted Climates in Small Ruminants: An Analysis on Traditionally-Managed Sheep and Goats. *Front. Genet.* 12. <https://doi.org/10.3389/fgene.2021.745284>
- Devaskar, S.U., Raychaudhuri, S., 2007. Epigenetics – A Science of Heritable Biological Adaptation. *Pediatr Res* 61, 1–4. <https://doi.org/10.1203/pdr.0b013e31805cddb8>
- Doherty, R., Whiston, R., Cormican, P., Finlay, E.K., Couldrey, C., Brady, C., O’Farrelly, C., Meade, K.G., 2016. The CD4+ T cell methylome contributes to a distinct CD4+ T cell transcriptional signature in *Mycobacterium bovis*-infected cattle. *Sci Rep* 6, 31014. <https://doi.org/10.1038/srep31014>
- Edvardsson Rasmussen, A., Båge, R., Holtenius, K., Strandberg, E., von Brömssen, C., Åkerlind, M., Kronqvist, C., 2023. A randomized study on the effect of an extended voluntary waiting period in primiparous dairy cows on fertility, health, and culling during first and second lactation. *J Dairy Sci* 106, 8897–8909. <https://doi.org/10.3168/jds.2023-23470>
- Esposito, G., Irons, P.C., Webb, E.C., Chapwanya, A., 2014. Interactions between negative energy balance, metabolic diseases, uterine health and immune response in transition dairy cows. *Animal Reproduction Science* 144, 60–71. <https://doi.org/10.1016/j.anireprosci.2013.11.007>
- Feng, S., Zhong, Z., Wang, M., Jacobsen, S.E., 2020. Efficient and accurate determination of genome-wide DNA methylation patterns in *Arabidopsis thaliana* with enzymatic methyl sequencing. *Epigenetics & Chromatin* 13, 42. <https://doi.org/10.1186/s13072-020-00361-9>
- Fenwick, M.A., Fitzpatrick, R., Kenny, D.A., Diskin, M.G., Patton, J., Murphy, J.J., Wathes, D.C., 2008. Interrelationships between negative energy balance (NEB) and IGF regulation in liver of lactating dairy cows. *Domestic Animal Endocrinology* 34, 31–44. <https://doi.org/10.1016/j.domaniend.2006.10.002>
- Fisher, R.A., 1918. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh* 52, 399–433. <https://doi.org/10.1017/S0080456800012163>
- Fitz-James, M.H., Cavalli, G., 2022. Molecular mechanisms of transgenerational epigenetic inheritance. *Nature Reviews Genetics* 2022 23:6 23, 325–341. <https://doi.org/10.1038/s41576-021-00438-5>
- Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* 89, 1827. <https://doi.org/10.1073/PNAS.89.5.1827>
- Furey, T.S., Sethupathy, P., 2013. Genetics Driving Epigenetics. *Science* 342, 705–706. <https://doi.org/10.1126/science.1246755>

- Gamaarachchi, H., Lam, C.W., Jayatilaka, G., Samarakoon, H., Simpson, J.T., Smith, M.A., Parameswaran, S., 2020. GPU accelerated adaptive banded event alignment for rapid comparative nanopore signal analysis. *BMC Bioinformatics* 21, 1–13. <https://doi.org/10.1186/S12859-020-03697-X/FIGURES/6>
- Garcia, D.M., Jarosz, D.F., 2014. Rebels with a cause: molecular features and physiological consequences of yeast prions. *FEMS Yeast Research* 14, 136–147. <https://doi.org/10.1111/1567-1364.12116>
- García-Giménez, J.L., Ushijima, T., Tollefsbol, T.O., 2016. Chapter 1 - Epigenetic Biomarkers: New Findings, Perspectives, and Future Directions in Diagnostics, in: García-Giménez, J.L. (Ed.), *Epigenetic Biomarkers and Diagnostics*. Academic Press, Boston, pp. 1–18. <https://doi.org/10.1016/B978-0-12-801899-6.00001-2>
- Gelato, K.A., Fischle, W., 2008. Role of histone modifications in defining chromatin structure and function. *Biol Chem* 389, 353–363. <https://doi.org/10.1515/BC.2008.048>
- Georges, M., Charlier, C., Hayes, B., 2019. Harnessing genomic information for livestock improvement. *Nat Rev Genet* 20, 135–156. <https://doi.org/10.1038/s41576-018-0082-2>
- Ghorbani, R., Shokri-Gharelo, R., 2019. Computational-based approaches in epigenetic research: Insights from computational tools, mathematical models, and machine learning methods. *J App Biol Biotech* 7, 79–83. <https://doi.org/10.7324/JABB.2019.70114>
- Gianola, D., Sorensen, D., 2004. Quantitative Genetic Models for Describing Simultaneous and Recursive Relationships Between Phenotypes This article is dedicated to Arthur B. Chapman, teacher and mentor of numerous animal breeding students and disciple and friend of Sewall Wright. *Genetics* 167, 1407–1424. <https://doi.org/10.1534/genetics.103.025734>
- Gianola, D., van Kaam, J.B.C.H.M., 2008. Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits. *Genetics* 178, 2289–2303. <https://doi.org/10.1534/genetics.107.084285>
- Gomes, A.Q., Nolasco, S., Soares, H., 2013. Non-coding RNAs: multi-tasking molecules in the cell. *Int J Mol Sci* 14, 16010–16039. <https://doi.org/10.3390/ijms140816010>
- Gonzalez-Recio, O., 2012. Epigenetics: A New Challenge in the Post-Genomic Era of Livestock. *Front. Genet.* 2. <https://doi.org/10.3389/fgene.2011.00106>
- González-Recio, O., López-Catalina, A., Peiró-Pastor, R., Nieto-Valle, A., Castro, M., Fernández, A., 2023. Evaluating the potential of (epi)genotype-by-low pass nanopore sequencing in dairy cattle: a study on direct genomic value and methylation analysis. *J Animal Sci Biotechnol* 14, 98. <https://doi.org/10.1186/s40104-023-00896-3>
- Gonzalez-Recio, O., Toro, M.A., Bach, A., 2015. Past, present and future of epigenetics applied to livestock breeding. *Front. Genet.* 6. <https://doi.org/10.3389/fgene.2015.00305>
- González-Recio, O., Ugarte, E., Bach, A., 2012. Trans-Generational Effect of Maternal Lactation during Pregnancy: A Holstein Cow Model. *PLOS ONE* 7, e51816. <https://doi.org/10.1371/JOURNAL.PONE.0051816>

- Gotoh, T., 2014. Potential of the application of epigenetics in animal production. *Anim. Prod. Sci.* 55, 145–158. <https://doi.org/10.1071/AN14467>
- Graham, J.R., Johnson, J.S., Araujo, A.C., Howard, J.T., Brito, L.F., 2021. PSXV-9 Transgenerational epigenetic variance for production and reproduction traits in maternal-line pigs. *Journal of Animal Science* 99, 258–259. <https://doi.org/10.1093/jas/skab235.472>
- Gray, A.L., Pun, N., Ridley, A.J.L., Dyer, D.P., 2022. Role of extracellular matrix proteoglycans in immune cell recruitment. *International Journal of Experimental Pathology* 103, 34–43. <https://doi.org/10.1111/iep.12428>
- Green, J.C., Meyer, J.P., Williams, A.M., Newsom, E.M., Keisler, D.H., Lucy, M.C., 2012. Pregnancy development from day 28 to 42 of gestation in postpartum Holstein cows that were either milked (lactating) or not milked (not lactating) after calving. <https://doi.org/10.1530/REP-11-0461>
- Guerrero-Bosagna, C., Skinner, M.K., 2012. Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. *Molecular and Cellular Endocrinology* 354, 3–8. <https://doi.org/10.1016/J.MCE.2011.10.004>
- Guner, B., Erturk, M., Dursun, M., Ozturk, B., Yilmazbas-Mecitoglu, G., Keskin, A., Dikmen, S., Gumen, A., 2023. Effect of oestrous expression prior to timed artificial insemination with sexed semen on pregnancy rate in dairy cows. *Reproduction in Domestic Animals* 58, 342–348. <https://doi.org/10.1111/rda.14293>
- Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., Walsh, C.P., Li, J., Tang, F., Xu, G.-L., 2014. Active and passive demethylation of male and female pronuclear DNA in the mammalian zygote. *Cell Stem Cell* 15, 447–459. <https://doi.org/10.1016/j.stem.2014.08.003>
- Gurgul, A., Miksza-Cybulska, A., Szmatoła, T., Jasielczuk, I., Piestrzyńska-Kajtoch, A., Fornal, A., Semik-Gurgul, E., Bugno-Poniewierska, M., 2019. Genotyping-by-sequencing performance in selected livestock species. *Genomics* 111, 186–195. <https://doi.org/10.1016/j.ygeno.2018.02.002>
- Halfmann, R., Lindquist, S., 2010. Epigenetics in the Extreme: Prions and the Inheritance of Environmentally Acquired Traits. *Science* 330, 629–632. <https://doi.org/10.1126/science.1191081>
- Halušková, J., Holečková, B., Staničová, J., 2021. DNA methylation studies in cattle. *J Appl Genetics* 62, 121–136. <https://doi.org/10.1007/s13353-020-00604-1>
- Han, W., Xue, Q., Li, G., Yin, J., Zhang, H., Zhu, Y., Xing, W., Cao, Y., Su, Y., Wang, K., Zou, J., 2020. Genome-wide analysis of the role of DNA methylation in inbreeding depression of reproduction in Langshan chicken. *Genomics* 112, 2677–2687. <https://doi.org/10.1016/j.ygeno.2020.02.007>
- Hayatsu, H., 2008. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis — A personal account. *Proceedings of the Japan Academy, Series B* 84, 321–330. <https://doi.org/10.2183/pjab.84.321>
- Hayatsu, H., Wataya, Y., Kai, K., 1970. Addition of sodium bisulfite to uracil and to cytosine. *J. Am. Chem. Soc.* 92, 724–726. <https://doi.org/10.1021/ja00706a062>

- Hayes, A.F., 2009. Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs* 76, 408–420. <https://doi.org/10.1080/03637750903310360>
- Hayes, B., Goddard, M.E., 2001. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution* 2001 33:3 33, 1–21. <https://doi.org/10.1186/1297-9686-33-3-209>
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J., Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92, 433–443. <https://doi.org/10.3168/JDS.2008-1646>
- He, Y., Ecker, J.R., 2015. Non-CG Methylation in the Human Genome. *Annual Review of Genomics and Human Genetics* 16, 55–77. <https://doi.org/10.1146/annurev-genom-090413-025437>
- He, Y., Song, M., Zhang, Yi, Li, X., Song, J., Zhang, Yuan, Yu, Y., 2016. Whole-genome regulation analysis of histone H3 lysin 27 trimethylation in subclinical mastitis cows infected by *Staphylococcus aureus*. *BMC Genomics* 17, 565. <https://doi.org/10.1186/s12864-016-2947-0>
- Heard, E., Martienssen, R.A., 2014. Transgenerational Epigenetic Inheritance: Myths and Mechanisms. *Cell* 157, 95–109. <https://doi.org/10.1016/j.cell.2014.02.045>
- Heijmans, B.T., Tobi, E.W., Lumey, L.H., Slagboom, P.E., 2009. The epigenome: Archive of the prenatal environment. *Epigenetics* 4, 526–531. <https://doi.org/10.4161/epi.4.8.10265>
- Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E., Lumey, L.H., 2008. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences* 105, 17046–17049. <https://doi.org/10.1073/pnas.0806560105>
- Henderson, C.R., 1984. *Applications of Linear Models in Animal Breeding*. University of Guelph.
- Hoffmann, I., 2010. Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics* 41, 32–46. <https://doi.org/10.1111/j.1365-2052.2010.02043.x>
- Holtenius, P., Holtenius, K., 1996. New Aspects of Ketone Bodies in Energy Metabolism of Dairy Cows: A Review. *Journal of Veterinary Medicine Series A* 43, 579–587. <https://doi.org/10.1111/j.1439-0442.1996.tb00491.x>
- Hu, G., Do, D.N., Davoudi, P., Miar, Y., 2022. Emerging Roles of Non-Coding RNAs in the Feed Efficiency of Livestock Species. *Genes (Basel)* 13, 297. <https://doi.org/10.3390/genes13020297>
- Ibeagha-Awemu, E.M., Bhattarai, S., Dedemaine, P.-L., Wang, M., McKay, S.D., Zhao, X., Bissonnette, N., 2020a. PSVIII-15 Genome wide DNA methylation analysis reveals role of DNA methylation in cow's ileal response to *Mycobacterium avium* subsp. *paratuberculosis*. *Journal of Animal Science* 98, 260–261. <https://doi.org/10.1093/jas/skaa278.471>
- Ibeagha-Awemu, E.M., Bhattarai, S., Dedemaine, P.-L., Wang, M., McKay, S.D., Zhao, X., Bissonnette, N., 2020b. 63 DNA methylome wide profile associates differentially methylated loci and regions with cow's ileal lymph node

- response to *Mycobacterium avium* subsp. *paratuberculosis*. *Journal of Animal Science* 98, 39–40. <https://doi.org/10.1093/jas/skaa278.071>
- Ibeagha-Awemu, E.M., Khatib, H., 2023. Chapter 29 - Epigenetics of Livestock Health, Production, and Breeding, in: Tollefsbol, T.O. (Ed.), *Handbook of Epigenetics* (Third Edition). Academic Press, pp. 569–610. <https://doi.org/10.1016/B978-0-323-91909-8.00041-4>
- Ibeagha-Awemu, E.M., Zhao, X., 2015. Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Frontiers in Genetics* 0, 302. <https://doi.org/10.3389/FGENE.2015.00302>
- Igartua, J.-J., Hayes, A.F., 2021. Mediation, Moderation, and Conditional Process Analysis: Concepts, Computations, and Some Common Confusions. *The Spanish Journal of Psychology* 24, e49. <https://doi.org/10.1017/SJP.2021.46>
- Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C., Bird, A., 2008. A Novel CpG Island Set Identifies Tissue-Specific Methylation at Developmental Gene Loci. *PLOS Biology* 6, e22. <https://doi.org/10.1371/journal.pbio.0060022>
- Illingworth, R.S., Bird, A.P., 2009. CpG islands – ‘A rough guide.’ *FEBS Letters, Prague Special Issue: Functional Genomics and Proteomics* 583, 1713–1720. <https://doi.org/10.1016/j.febslet.2009.04.012>
- Imgenberg-Kreuz, J., Almlöf, J.C., Leonard, D., Alexsson, A., Nordmark, G., Eloranta, M.-L., Rantapää-Dahlqvist, S., Bengtsson, A.A., Jönsen, A., Padyukov, L., Gunnarsson, I., Svenungsson, E., Sjöwall, C., Rönnblom, L., Syvänen, A.-C., Sandling, J.K., 2018. DNA methylation mapping identifies gene regulatory effects in patients with systemic lupus erythematosus. *Annals of the Rheumatic Diseases* 77, 736–743. <https://doi.org/10.1136/annrheumdis-2017-212379>
- Jakimowicz, M., Szyda, J., Zarnecki, A., Jagusiak, W., Morek-Kopeć, M., Kosińska-Selbi, B., Suchocki, T., 2022. Genome-Wide Genomic and Functional Association Study for Workability and Calving Traits in Holstein Cattle. *Animals* 12, 1127. <https://doi.org/10.3390/ani12091127>
- Jensen, J., 2016. Estimation of genetic variance in the age of genomics. *J Anim Breed Genet* 133, 333. <https://doi.org/10.1111/jbg.12235>
- Jose, A.M., 2018. Replicating and Cycling Stores of Information Perpetuate Life. *BioEssays* 40, 1700161. <https://doi.org/10.1002/bies.201700161>
- Khatib, H., Townsend, J., Konkeli, M.A., Conidi, G., Hasselkus, J.A., 2024. Calling the question: what is mammalian transgenerational epigenetic inheritance? *Epigenetics* 19, 2333586. <https://doi.org/10.1080/15592294.2024.2333586>
- Knight, C.H., 2001. Lactation and gestation in dairy cows: flexibility avoids nutritional extremes. *Proceedings of the Nutrition Society* 60, 527–537. <https://doi.org/10.1079/PNS2001115>
- Korkmaz, F.T., Kerr, D.E., 2017. Genome-wide methylation analysis reveals differentially methylated loci that are associated with an age-dependent increase in bovine fibroblast response to LPS. *BMC Genomics* 18, 405. <https://doi.org/10.1186/s12864-017-3796-1>

- Krueger, F., Andrews, S.R., 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571. <https://doi.org/10.1093/BIOINFORMATICS/BTR167>
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012 9:4 9, 357–359. <https://doi.org/10.1038/nmeth.1923>
- Legarra, A., Christensen, O.F., 2023. Genomic evaluation methods to include intermediate correlated features such as high-throughput or omics phenotypes* *Presented as part of the Breeding and Genetics Symposium: Beyond Genetic Markers—Additional Data to Improve Long-Term Selection held at the ADSA Annual Meeting, June 2022. *JDS Communications* 4, 55–60. <https://doi.org/10.3168/jdsc.2022-0276>
- Leroy, J.L.M.R., Bie, J., Jordaens, L., Desmet, K., Smits, A., Marei, W.F.A., Bols, P.E.J., Hoeck, V.V., 2017. Negative energy balance and metabolic stress in relation to oocyte and embryo quality: an update on possible pathways reducing fertility in dairy cows. *Anim Reprod* 14, 497–506. <https://doi.org/10.21451/1984-3143-AR992>
- Lesta, A., Marín-García, P.J., Llobat, L., 2023. How Does Nutrition Affect the Epigenetic Changes in Dairy Cows? *Animals* 13, 1883. <https://doi.org/10.3390/ani13111883>
- Li, C., 2015. Nutritional Epigenomics, in: Khatib, H. (Ed.), *Molecular and Quantitative Animal Genetics*, New York Academy of Sciences Series. John Wiley & Sons, Incorporated, Newark, pp. 215–225.
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/BIOINFORMATICS/BTY191>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078. <https://doi.org/10.1093/BIOINFORMATICS/BTP352>
- Li, Y., Zhang, L.-Z., Yi, Y., Hu, W.-W., Guo, Y.-H., Zeng, Z.-J., Huang, Z.-Y., Wang, Z.-L., 2017. Genome-wide DNA methylation changes associated with olfactory learning and memory in *Apis mellifera*. *Sci Rep* 7, 17017. <https://doi.org/10.1038/s41598-017-17046-1>
- Ling, T., Hernandez-Jover, M., Sordillo, L.M., Abuelo, A., 2018. Maternal late-gestation metabolic stress is associated with changes in immune and metabolic responses of dairy calves. *Journal of Dairy Science* 101, 6568–6580. <https://doi.org/10.3168/jds.2017-14038>
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322. <https://doi.org/10.1038/nature08514>
- Littlejohn, B.P., Price, D.M., Neuendorff, D.A., Carroll, J.A., Vann, R.C., Riggs, P.K., Riley, D.G., Long, C.R., Welsh, T.H., Jr., Randel, R.D., 2018. Prenatal transportation stress alters genome-wide DNA methylation in suckling

- Brahman bull calves^{1,2}. *Journal of Animal Science* 96, 5075–5099. <https://doi.org/10.1093/jas/sky350>
- Liu, X., Yang, J., Zhang, Q., Jiang, L., 2017. Regulation of DNA methylation on EEF1D and RPL8 expression in cattle. *Genetica* 145, 387–395. <https://doi.org/10.1007/s10709-017-9974-x>
- Loman, N.J., Quick, J., Simpson, J.T., 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods* 2015 12:8 12, 733–735. <https://doi.org/10.1038/nmeth.3444>
- Lonergan, P., Fair, T., Forde, N., Rizos, D., 2016. Embryo development in dairy cattle. *Theriogenology* 86, 270–277. <https://doi.org/10.1016/j.theriogenology.2016.04.040>
- López-Catalina, A., Bach, A., Gutiérrez-Rivas, M., Peiró-Pastor, R., González-Recio, Ó., 2022. 544. An approach to study the association between the blood cell methylome with feed efficiency traits, in: *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. Wageningen Academic Publishers, pp. 2257–2260. https://doi.org/10.3920/978-90-8686-940-4_544
- Lucy, M.C., 2007. Fertility in high-producing dairy cows: reasons for decline and corrective strategies for sustainable improvement. *Soc Reprod Fertil Suppl* 64, 237–254. <https://doi.org/10.5661/rdr-vi-237>
- Lucy, M.C., 2001. Reproductive Loss in High-Producing Dairy Cattle: Where Will It End? *Journal of Dairy Science* 84, 1277–1293. [https://doi.org/10.3168/jds.S0022-0302\(01\)70158-0](https://doi.org/10.3168/jds.S0022-0302(01)70158-0)
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260. <https://doi.org/10.1038/38444>
- Luo, C., Hajkova, P., Ecker, J.R., 2018. Dynamic DNA methylation: In the right place at the right time. *Science* 361, 1336–1340. <https://doi.org/10.1126/science.aat6806>
- Ma, J., Kok, A., Goselink, R.M.A., Lam, T.J.G.M., Kemp, B., Knegsel, A.T.M. van, 2022. Udder health of dairy cows with an extended voluntary waiting period from calving until the first insemination. *Journal of Dairy Research* 89, 271–278. <https://doi.org/10.1017/S0022029922000516>
- Ma, Y., Ordovas, J.M., 2017. The integration of epigenetics and genetics in nutrition research for CVD risk factors. *Proceedings of the Nutrition Society* 76, 333–346. <https://doi.org/10.1017/S0029665116000823>
- Magee, D.A., Spillane, C., Berkowicz, E.W., Sikora, K.M., MacHugh, D.E., 2014. Imprinted loci in domestic livestock species as epigenomic targets for artificial selection of complex traits. *Animal Genetics* 45, 25–39. <https://doi.org/10.1111/age.12168>
- Martens, H., 2023. Invited Review: Increasing Milk Yield and Negative Energy Balance: A Gordian Knot for Dairy Cows? *Animals* 13, 3097. <https://doi.org/10.3390/ani13193097>
- Martos, S.N., Tang, W., Wang, Z., 2015. Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease. *Progress in Biophysics and Molecular Biology* 118, 44–54. <https://doi.org/10.1016/j.pbiomolbio.2015.02.011>

- Matthews, L.J., Turkheimer, E., 2022. Three legs of the missing heritability problem. *Studies in History and Philosophy of Science* 93, 183–191. <https://doi.org/10.1016/j.shpsa.2022.04.004>
- McFadden, J.W., Girard, C.L., Tao, S., Zhou, Z., Bernard, J.K., Duplessis, M., White, H.M., 2020. Symposium review: One-carbon metabolism and methyl donor nutrition in the dairy cow. *J Dairy Sci* 103, 5668–5683. <https://doi.org/10.3168/jds.2019-17319>
- McKay, S., Murdoch, B.M., Hagen, D.E., 2023. 53 Establishing a Pan-Epigenome for Cattle and Sheep. *Journal of Animal Science* 101, 48–49. <https://doi.org/10.1093/jas/skad281.059>
- Mekuriaw, Y., 2023. Negative energy balance and its implication on productive and reproductive performance of early lactating dairy cows: review paper. *Journal of Applied Animal Research* 51, 220–228. <https://doi.org/10.1080/09712119.2023.2176859>
- Mi, H., Muruganujan, A., Thomas, P.D., 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 41, D377–386. <https://doi.org/10.1093/nar/gks1118>
- Mohammed, H.H., Ibrahim, M.Y., Saleem, A.-S.K., Fakhry, A., Youniss, I., 2018. A study on Calf's Behavior and Growth Performance in Relation to Gestation Period and Lactation Season. *SVU-International Journal of Veterinary Sciences* 1, 60–68. <https://doi.org/10.21608/svu.2018.19857>
- Mohn, F., Weber, M., Schübeler, D., Roloff, T.-C., 2009. Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol* 507, 55–64. https://doi.org/10.1007/978-1-59745-522-0_5
- Monk, D., Mackay, D.J.G., Eggermann, T., Maher, E.R., Riccio, A., 2019. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat Rev Genet* 20, 235–248. <https://doi.org/10.1038/s41576-018-0092-0>
- Morgan, H.D., Sutherland, H.G.E., Martin, D.I.K., Whitelaw, E., 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23, 314–318. <https://doi.org/10.1038/15490>
- Murdoch, B.M., Murdoch, G.K., Greenwood, S., McKay, S., 2016. Nutritional Influence on Epigenetic Marks and Effect on Livestock Production. *Frontiers in Genetics* 7.
- Myzak, M.C., Dashwood, R.H., 2006. Histone deacetylases as targets for dietary cancer preventive agents: lessons learned with butyrate, diallyl disulfide, and sulforaphane. *Curr Drug Targets* 7, 443–452. <https://doi.org/10.2174/138945006776359467>
- Nagao, Y., 2024. Contribution of rare variants to heritability of a disease is much greater than conventionally estimated: modification of allele distribution model. *J Hum Genet* 69, 663–668. <https://doi.org/10.1038/s10038-024-01281-2>
- Nakabayashi, K., Yamamura, M., Hasegawa, K., Hata, K., 2023. Reduced Representation Bisulfite Sequencing (RRBS), in: Hatada, I., Horii, T. (Eds.), *Epigenomics: Methods and Protocols*. Springer US, New York, NY, pp. 39–51. https://doi.org/10.1007/978-1-0716-2724-2_3

- Naskar, S., Gowane, G.R., Chopra, A., 2015. Strategies to Improve Livestock Genetic Resources to Counter Climate Change Impact, in: Sejian, V., Gaughan, J., Baumgard, L., Prasad, C. (Eds.), *Climate Change Impact on Livestock: Adaptation and Mitigation*. Springer India, New Delhi, pp. 441–475. https://doi.org/10.1007/978-81-322-2265-1_25
- Nguyen, M., Boutinaud, M., Pétridou, B., Gabory, A., Pannetier, M., Chat, S., Bouet, S., Jouneau, L., Jaffrezic, F., Laloë, D., Klopp, C., Brun, N., Kress, C., Jammes, H., Charlier, M., Devinoy, E., 2014. DNA Methylation and Transcription in a Distal Region Upstream from the Bovine AlphaS1 Casein Gene after Once or Twice Daily Milking. *PLOS ONE* 9, e111556. <https://doi.org/10.1371/journal.pone.0111556>
- Norouzitallab, P., Baruah, K., Vanrompay, D., Bossier, P., 2019. Can epigenetics translate environmental cues into phenotypes? *Science of The Total Environment* 647, 1281–1293. <https://doi.org/10.1016/j.scitotenv.2018.08.063>
- Noya, A., Casasús, I., Ferrer, J., Sanz, A., 2019. Long-Term Effects of Maternal Subnutrition in Early Pregnancy on Cow-Calf Performance, Immunological and Physiological Profiles during the Next Lactation. *Animals* 9, 936. <https://doi.org/10.3390/ani9110936>
- Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R.V., Branco, M.R., Reik, W., 2018. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology* 19, 1–19. <https://doi.org/10.1186/S13059-018-1408-2/TABLES/3>
- Olova, N.N., Andrews, S., 2023. Whole Genome Methylation Sequencing via Enzymatic Conversion (EM-seq): Protocol, Data Processing and Analysis. <https://doi.org/10.1101/2023.10.07.561361>
- Organization, W.H., Safety, I.P. on C., 2001. Biomarkers in risk assessment : validity and validation. World Health Organization.
- Osorio, J.S., Jacometo, C.B., Zhou, Z., Luchini, D., Cardoso, F.C., Loor, J.J., 2016. Hepatic global DNA and peroxisome proliferator-activated receptor alpha promoter methylation are altered in peripartal dairy cows fed rumen-protected methionine. *Journal of Dairy Science* 99, 234–244. <https://doi.org/10.3168/JDS.2015-10157>
- Park, Y., Wu, H., 2016. Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics* 32, 1446–1453. <https://doi.org/10.1093/BIOINFORMATICS/BTW026>
- Patil, V., Ward, R.L., Hesson, L.B., 2014. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* 9, 823–828. <https://doi.org/10.4161/epi.28741>
- Pattenden, S.G., 2022. Epigenetic plasticity in tumor biology. *BioEssays* 44, 2200092. <https://doi.org/10.1002/bies.202200092>
- Perez, M.F., Lehner, B., 2019. Intergenerational and transgenerational epigenetic inheritance in animals. *Nat Cell Biol* 21, 143–151. <https://doi.org/10.1038/s41556-018-0242-9>
- Pierce, C.F., Speidel, S.E., Coleman, S.J., Enns, R.M., Bailey, D.W., Medrano, J.F., Cánovas, A., Meiman, P.J., Howery, L.D., Mandeville, W.F., Thomas, M.G.,

2020. Genome-wide association studies of beef cow terrain-use traits using Bayesian multiple-SNP regression. *Livestock Science* 232, 103900. <https://doi.org/10.1016/j.livsci.2019.103900>
- Pizano, J.M., Williamson, C.B., 2020. Nutritional Influences on Methylation, in: Noland, D., Drisko, J.A., Wagner, L. (Eds.), *Integrative and Functional Medical Nutrition Therapy: Principles and Practices*. Springer International Publishing, Cham, pp. 269–284. https://doi.org/10.1007/978-3-030-30730-1_18
- Pogribny, I.P., Rusyn, I., 2013. Environmental Toxicants, Epigenetics, and Cancer, in: Karpf, A.R. (Ed.), *Epigenetic Alterations in Oncogenesis*. Springer, New York, NY, pp. 215–232. https://doi.org/10.1007/978-1-4419-9967-2_11
- Policarpi, C., Munafò, M., Tsagkris, S., Carlini, V., Hackett, J.A., 2024. Systematic epigenome editing captures the context-dependent instructive function of chromatin modifications. *Nat Genet* 56, 1168–1180. <https://doi.org/10.1038/s41588-024-01706-w>
- 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 Selection Evolution* 46, 1–5. <https://doi.org/10.1186/1297-9686-46-22/FIGURES/2>
- Prasad, C.S., Sejian, V., 2015. Climate Change Impact on Livestock Sector: Visioning 2025, in: Sejian, V., Gaughan, J., Baumgard, L., Prasad, C. (Eds.), *Climate Change Impact on Livestock: Adaptation and Mitigation*. Springer India, New Delhi, pp. 479–489. https://doi.org/10.1007/978-81-322-2265-1_26
- Purfield, D.C., Evans, R.D., Berry, D.P., 2020. Breed- and trait-specific associations define the genetic architecture of calving performance traits in cattle. *Journal of Animal Science* 98, skaa151. <https://doi.org/10.1093/jas/skaa151>
- Pysera, B., Opalka, A., 2000. The effect of gestation and lactation of dairy cows on lipid and lipoprotein patterns and composition in serum during winter and summer feeding. *J. Anim. Feed Sci.* 9, 411–424. <https://doi.org/10.22358/jafs/68061/2000>
- R Core Team, 2021. R: A language and environment for statistical computing. [WWW Document]. URL <https://www.R-project.org/>. (accessed 11.15.21).
- Radford, E.J., 2018. Exploring the extent and scope of epigenetic inheritance. *Nat Rev Endocrinol* 14, 345–355. <https://doi.org/10.1038/s41574-018-0005-5>
- Rasmussen, A.E., Holtenius, K., Båge, R., Strandberg, E., Åkerlind, M., Kronqvist, C., 2024. Customized voluntary waiting period before first insemination in primiparous dairy cows: Effect on milk production, fertility, and health. *Journal of Dairy Science* 107, 9558–9571. <https://doi.org/10.3168/jds.2023-24593>
- Rauch, T., Li, H., Wu, X., Pfeifer, G.P., 2006. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Research* 66, 7939–7947. <https://doi.org/10.1158/0008-5472.CAN-06-1888>

- Rauch, T., Pfeifer, G.P., 2005. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Laboratory Investigation* 85, 1172–1180. <https://doi.org/10.1038/labinvest.3700311>
- Rauch, T.A., Pfeifer, G.P., 2023. Chapter 8 - Methods for Analyzing DNA Cytosine Modifications Genome-wide, in: Tollefsbol, T.O. (Ed.), *Handbook of Epigenetics* (Third Edition). Academic Press, pp. 123–135. <https://doi.org/10.1016/B978-0-323-91909-8.00030-X>
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* 47, W191–W198. <https://doi.org/10.1093/NAR/GKZ369>
- Rayman, J.B., Kandel, E.R., 2017. Functional Prions in the Brain. *Cold Spring Harb Perspect Biol* 9, a023671. <https://doi.org/10.1101/cshperspect.a023671>
- Reverter, A., Chan, E.K.F., 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497. <https://doi.org/10.1093/BIOINFORMATICS/BTN482>
- Romagnoli, S., Bartalucci, N., Vannucchi, A.M., 2023. Resolving complex structural variants via nanopore sequencing. *Front. Genet.* 14. <https://doi.org/10.3389/fgene.2023.1213917>
- Roseboom, T.J., Painter, R.C., Abeelen, A.F.M. van, Veenendaal, M.V.E., Rooij, S.R. de, 2011. Hungry in the womb: What are the consequences? Lessons from the Dutch famine. *Maturitas* 70, 141–145. <https://doi.org/10.1016/j.maturitas.2011.06.017>
- Rothi, M.H., Greer, E.L., 2023. From correlation to causation: The new frontier of transgenerational epigenetic inheritance. *BioEssays* 45, 2200118. <https://doi.org/10.1002/bies.202200118>
- Safdar, M., Ozaslan, M., 2023. Epigenetic Markers for Animal Health and Productivity in Livestock. *The Eurasia Proceedings of Health, Environment and Life Sciences* 1–6. <https://doi.org/10.55549/ephels.63>
- Sandoval-Motta, S., Aldana, M., Martínez-Romero, E., Frank, A., 2017. The Human Microbiome and the Missing Heritability Problem. *Front Genet* 8, 80. <https://doi.org/10.3389/fgene.2017.00080>
- Santilli, F., Boskovic, A., 2023. Mechanisms of transgenerational epigenetic inheritance: lessons from animal model organisms. *Current Opinion in Genetics & Development* 79, 102024. <https://doi.org/10.1016/j.gde.2023.102024>
- Santos, F., Hendrich, B., Reik, W., Dean, W., 2002. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev Biol* 241, 172–182. <https://doi.org/10.1006/dbio.2001.0501>
- Schefers, J.M., Weigel, K.A., 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. *Animal Frontiers* 2, 4–9. <https://doi.org/10.2527/AF.2011-0032>
- Schmidt, J., Blessing, F., Fimpler, L., Wenzel, F., 2020. Nanopore Sequencing in a Clinical Routine Laboratory: Challenges and Opportunities. *Clin Lab* 66. <https://doi.org/10.7754/Clin.Lab.2019.191114>

- Schuster, P., 2010. Genotypes and Phenotypes in the Evolution of Molecules, in: *Evolutionary Genomics and Systems Biology*. John Wiley & Sons, Ltd, pp. 121–152. <https://doi.org/10.1002/9780470570418.ch7>
- Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., Reik, W., 2012. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol Cell* 48, 849–862. <https://doi.org/10.1016/j.molcel.2012.11.001>
- Sevane, N., Martínez, R., Bruford, M.W., 2019. Genome-wide differential DNA methylation in tropically adapted Creole cattle and their Iberian ancestors. *Animal Genetics* 50, 15–26. <https://doi.org/10.1111/age.12731>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research* 13, 2498–2504. <https://doi.org/10.1101/gr.1239303>
- Siegfried, Z., Simon, I., 2010. DNA methylation and gene expression. *WIREs Systems Biology and Medicine* 2, 362–371. <https://doi.org/10.1002/wsbm.64>
- Singh, K., Molenaar, A.J., Swanson, K.M., Gudex, B., Arias, J.A., Erdman, R.A., Stelwagen, K., 2012. Epigenetics: a possible role in acute and transgenerational regulation of dairy cow milk production. *Animal* 6, 375–381. <https://doi.org/10.1017/S1751731111002564>
- Skinner, M.K., 2024. Epigenetic biomarkers for disease susceptibility and preventative medicine. *Cell Metabolism* 36, 263–277. <https://doi.org/10.1016/j.cmet.2023.11.015>
- Skinner, M.K., 2011. Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. *Epigenetics* 6, 838–842. <https://doi.org/10.4161/epi.6.7.16537>
- Slatkin, M., 2009. Epigenetic Inheritance and the Missing Heritability Problem. *Genetics* 182, 845–850. <https://doi.org/10.1534/GENETICS.109.102798>
- Slyvka, Y., Zhang, Y., Nowak, F.V., 2015. Epigenetic effects of paternal diet on offspring: emphasis on obesity. *Endocrine* 48, 36–46. <https://doi.org/10.1007/s12020-014-0328-5>
- Snajder, R., Leger, A., Stegle, O., Bonder, M.J., 2023. pycoMeth: a toolbox for differential methylation testing from Nanopore methylation calls. *Genome Biology* 24, 83. <https://doi.org/10.1186/s13059-023-02917-w>
- Stefansson, O.A., Sigurpalsdottir, B.D., Rognvaldsson, S., Halldorsson, G.H., Juliusson, K., Sveinbjornsson, G., Gunnarsson, B., Beyter, D., Jonsson, H., Gudjonsson, S.A., Olafsdottir, T.A., Saevarsdottir, S., Magnusson, M.K., Lund, S.H., Tragante, V., Oddsson, A., Hardarson, M.T., Eggertsson, H.P., Gudmundsson, R.L., Sverrisson, S., Frigge, M.L., Zink, F., Holm, H., Stefansson, H., Rafnar, T., Jonsdottir, I., Sulem, P., Helgason, A., Gudbjartsson, D.F., Halldorsson, B.V., Thorsteinsdottir, U., Stefansson, K., 2024. The correlation between CpG methylation and gene expression is driven by sequence variants. *Nat Genet* 1–8. <https://doi.org/10.1038/s41588-024-01851-2>
- Stoddart, D., Heron, A.J., Mikhailova, E., Maglia, G., Bayley, H., 2009. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a

- biological nanopore. *Proceedings of the National Academy of Sciences* 106, 7702–7707. <https://doi.org/10.1073/pnas.0901054106>
- Štolcová, M., Bartoň, L., Řehák, D., 2024. Milk components as potential indicators of energy status in early lactation Holstein dairy cows from two farms. *animal* 18, 101235. <https://doi.org/10.1016/j.animal.2024.101235>
- Sun, H.Z., Plastow, G., Guan, L.L., 2019. Invited review: Advances and challenges in application of feedomics to improve dairy cow production and health. *Journal of Dairy Science* 102, 5853–5870. <https://doi.org/10.3168/jds.2018-16126>
- Swart, G.W.M., 2002. Activated leukocyte cell adhesion molecule (CD166/ALCAM): Developmental and mechanistic aspects of cell clustering and cell migration. *European Journal of Cell Biology* 81, 313–321. <https://doi.org/10.1078/0171-9335-00256>
- Tal, O., Kisdi, E., Jablonka, E., 2010. Epigenetic Contribution to Covariance Between Relatives. *Genetics* 184, 1037–1050. <https://doi.org/10.1534/genetics.109.112466>
- Teh, A.L., Pan, H., Chen, L., Ong, M.-L., Dogra, S., Wong, J., MacIsaac, J.L., Mah, S.M., McEwen, L.M., Saw, S.-M., Godfrey, K.M., Chong, Y.-S., Kwek, K., Kwoh, C.-K., Soh, S.-E., Chong, M.F.F., Barton, S., Karnani, N., Cheong, C.Y., Buschdorf, J.P., Stünkel, W., Kobor, M.S., Meaney, M.J., Gluckman, P.D., Holbrook, J.D., 2014. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res* 24, 1064–1074. <https://doi.org/10.1101/gr.171439.113>
- The Bovine HapMap Consortium, Gibbs, R.A., Taylor, J.F., Van Tassell, C.P., Barendse, W., Eversole, K.A., Gill, C.A., Green, R.D., Hamernik, D.L., Kappes, S.M., Lien, S., Matukumalli, L.K., McEwan, J.C., Nazareth, L.V., Schnabel, R.D., Weinstock, G.M., Wheeler, D.A., Ajmone-Marsan, P., Boettcher, P.J., Caetano, A.R., Garcia, J.F., Hanotte, O., Mariani, P., Skow, L.C., Sonstegard, T.S., Williams, J.L., Diallo, B., Hailemariam, L., Martinez, M.L., Morris, C.A., Silva, L.O.C., Spelman, R.J., Mulatu, W., Zhao, K., Abbey, C.A., Agaba, M., Araujo, F.R., Bunch, R.J., Burton, J., Gorni, C., Olivier, H., Harrison, B.E., Luff, B., Machado, M.A., Mwakaya, J., Plastow, G., Sim, W., Smith, T., Thomas, M.B., Valentini, A., Williams, P., Womack, J., Woolliams, J.A., Liu, Y., Qin, X., Worley, K.C., Gao, C., Jiang, H., Moore, S.S., Ren, Y., Song, X.-Z., Bustamante, C.D., Hernandez, R.D., Muzny, D.M., Patil, S., San Lucas, A., Fu, Q., Kent, M.P., Vega, R., Matukumalli, A., McWilliam, S., Sclep, G., Bryc, K., Choi, J., Gao, H., Grefenstette, J.J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Goddard, M.E., Hayes, B.J., Bradley, D.G., Barbosa da Silva, M., Lau, L.P.L., Liu, G.E., Lynn, D.J., Panzitta, F., Dodds, K.G., 2009. Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324, 528–532. <https://doi.org/10.1126/science.1167936>
- Thomas, P.D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.-P., Mi, H., 2022. PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* 31, 8–22. <https://doi.org/10.1002/pro.4218>

- Thompson, R.P., Nilsson, E., Skinner, M.K., 2020. Environmental epigenetics and epigenetic inheritance in domestic farm animals. *Animal Reproduction Science* 220, 106316. <https://doi.org/10.1016/j.anireprosci.2020.106316>
- Thomson, C.E., Winney, I.S., Salles, O.C., Pujol, B., 2018. A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. *PLOS ONE* 13, e0197720. <https://doi.org/10.1371/journal.pone.0197720>
- Tian, X.C., Marjani, S.L., 2012. Epigenetics of Cloned Preimplantation Embryos of Domestic Animals, in: *Livestock Epigenetics*. John Wiley & Sons, Ltd, pp. 27–42. <https://doi.org/10.1002/9781119949930.ch2>
- Tiffon, C., 2018. The Impact of Nutrition and Environmental Epigenetics on Human Health and Disease. *International Journal of Molecular Sciences* 19, 3425. <https://doi.org/10.3390/ijms19113425>
- Trerotola, M., Relli, V., Simeone, P., Alberti, S., 2015. Epigenetic inheritance and the missing heritability. *Human Genomics* 9. <https://doi.org/10.1186/S40246-015-0041-3>
- Triantaphyllopoulos, K.A., Ikononopoulos, I., Bannister, A.J., 2016. Epigenetics and inheritance of phenotype variation in livestock. *Epigenetics & Chromatin* 2016 9:1 9, 1–18. <https://doi.org/10.1186/S13072-016-0081-5>
- Tuite, M.F., Serio, T.R., 2010. The prion hypothesis: from biological anomaly to basic regulatory mechanism. *Nat Rev Mol Cell Biol* 11, 823–833. <https://doi.org/10.1038/nrm3007>
- Usman, T., Yu, Y., Wang, Y., 2016. P2001 CD4 promoter hyper methylation is associated with lower gene expression in clinical mastitis cows and vice versa in the healthy controls. *Journal of Animal Science* 94, 38. <https://doi.org/10.2527/jas2016.94supplement438x>
- Vaisvila, R., Ponnaluri, V.K.C., Sun, Z., Langhorst, B.W., Saleh, L., Guan, S., Dai, N., Campbell, M.A., Sexton, B.S., Marks, K., Samaranayake, M., Samuelson, J.C., Church, H.E., Tamanaha, E., Corrêa, I.R., Pradhan, S., Dimalanta, E.T., Evans, T.C., Williams, L., Davis, T.B., 2021. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* 31, 1280–1289. <https://doi.org/10.1101/gr.266551.120>
- Van Tassell, C.P., Smith, T.P.L., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C., Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* 5, 247–252. <https://doi.org/10.1038/nmeth.1185>
- Vanselow, J., Yang, W., Herrmann, J., Zerbe, H., Schuberth, H.-J., Petzl, W., Tomek, W., Seyfert, H.-M., 2006. DNA-remethylation around a STAT5-binding enhancer in the α S1-casein promoter is associated with abrupt shutdown of α S1-casein synthesis during acute mastitis. <https://doi.org/10.1677/jme.1.02131>
- Varona, L., Munilla, S., Mouresan, E.F., González-Rodríguez, A., Moreno, C., Altarriba, J., 2015. A Bayesian Model for the Analysis of Transgenerational Epigenetic Variation. *G3 Genes|Genomes|Genetics* 5, 477–485. <https://doi.org/10.1534/g3.115.016725>

- Verma, M., 2012. Epigenetic Biomarkers in Cancer Epidemiology, in: Dumitrescu, R.G., Verma, M. (Eds.), *Cancer Epigenetics: Methods and Protocols*. Humana Press, Totowa, NJ, pp. 467–480. https://doi.org/10.1007/978-1-61779-612-8_28
- Waddington, C.H., 1942. The epigenotype. *Endeavour* 1, 18–20.
- Wallace, R., 2017. Cultural Epigenetics: On the Missing Heritability of Complex Diseases, in: Wallace, R. (Ed.), *Computational Psychiatry: A Systems Biology Approach to the Epigenetics of Mental Disorders*. Springer International Publishing, Cham, pp. 37–70. https://doi.org/10.1007/978-3-319-53910-2_2
- Wallrath, L.L., Vitalini, M.W., Elgin, S.C.R., 2014. Heterochromatin: A Critical Part of the Genome, in: Workman, J.L., Abmayr, S.M. (Eds.), *Fundamentals of Chromatin*. Springer, New York, NY, pp. 529–552. https://doi.org/10.1007/978-1-4614-8624-4_13
- Walsh, S.W., Williams, E.J., Evans, A.C.O., 2011. A review of the causes of poor fertility in high milk producing dairy cows. *Animal Reproduction Science* 123, 127–138. <https://doi.org/10.1016/j.anireprosci.2010.12.001>
- Wang, J., Hua, G., Chen, J., Cui, K., Yang, Z., Han, D., Yang, X., Dong, X., Ma, Y., Cai, G., Zhang, Y., Li, J., Tai, Y., Da, L., Li, X., Ma, L., Ma, Q., Li, R., Liu, J., Darwish, H.Y.A., Wu, K., Rong, W., Liu, W., Zhao, Y., Deng, X., 2023a. Epigenetic mechanism of Gtl2-miRNAs causes the primitive sheep characteristics found in purebred Merino sheep. *Cell & Bioscience* 13, 190. <https://doi.org/10.1186/s13578-023-01142-z>
- Wang, J.-H., Zhao, Q.-Y., Zhou, Y.-L., Shi, L.-Y., Wang, C.-D., Yu, Y., 2023b. Application and prospect of gene chip in genetic breeding of livestock and poultry. *Yi Chuan* 45, 1114–1127. <https://doi.org/10.16288/j.ycz.23-233>
- Wang, L., Sun, H.Z., Guan, L.L., Liu, J.X., 2019. *Short communication*: Relationship of blood DNA methylation rate and milk performance in dairy cows. *Journal of Dairy Science* 102, 5208–5211. <https://doi.org/10.3168/jds.2018-15869>
- Wang, L., Xie, Y., Chen, W., Zhang, Y., Zeng, Y., 2021a. The role of long noncoding RNAs in livestock adipose tissue deposition — A review. *Anim Biosci* 34, 1089–1099. <https://doi.org/10.5713/ab.21.0006>
- Wang, M., Feng, S., Ma, G., Miao, Y., Zuo, B., Ruan, J., Zhao, S., Wang, H., Du, X., Liu, X., 2020. Whole-Genome Methylation Analysis Reveals Epigenetic Variation in Cloned and Donor Pigs. *Front. Genet.* 11. <https://doi.org/10.3389/fgene.2020.00023>
- Wang, M., Ibeagha-Awemu, E.M., 2020. Impacts of Epigenetic Processes on the Health and Productivity of Livestock. *Front Genet* 11, 613636. <https://doi.org/10.3389/fgene.2020.613636>
- Wang, X.S., Zhang, Y., He, Y.H., Ma, P.P., Fan, L.J., Wang, Y.C., Zhang, Y.I., Sun, D.X., Zhang, S.L., Wang, C.D., Song, J.Z., Yu, Y., 2013. Aberrant promoter methylation of the CD4 gene in peripheral blood cells of mastitic dairy cows. *Genet. Mol. Res.* 12, 6228–6239. <https://doi.org/10.4238/2013.December.4.10>
- Wang, Y., Ipema, A., Goselink, R., Burgers, E., Gross, J., Bruckmaier, R., Kemp, B., Knegsel, A. van, 2024. Effects of calving interval of dairy cows on

- development, metabolism, and milk performance of their offspring. *Journal of Dairy Science* 107, 9934–9947. <https://doi.org/10.3168/jds.2024-24885>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Yuru, Au, K.F., 2021b. Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology* 2021 39:11 39, 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Wang, Zhong, Wang, Zuoheng, Wang, J., Sui, Y., Zhang, J., Liao, D., Wu, R., 2012. A quantitative genetic and epigenetic model of complex traits. *BMC Bioinformatics* 13, 274. <https://doi.org/10.1186/1471-2105-13-274>
- Wathes, D.C., Fenwick, M., Cheng, Z., Bourne, N., Llewellyn, S., Morris, D.G., Kenny, D., Murphy, J., Fitzpatrick, R., 2007. Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. *Theriogenology, Proceedings of the International Conference on Farm Animal Reproduction* 68, S232–S241. <https://doi.org/10.1016/j.theriogenology.2007.04.006>
- Watson, C.M., Holliday, D.L., Crinnion, L.A., Bonthron, D.T., 2022. Long-read nanopore DNA sequencing can resolve complex intragenic duplication/deletion variants, providing information to enable preimplantation genetic diagnosis. *Prenatal Diagnosis* 42, 226–232. <https://doi.org/10.1002/pd.6089>
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schübeler, D., 2005. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37, 853–862. <https://doi.org/10.1038/ng1598>
- Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., Schübeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39, 457–466. <https://doi.org/10.1038/ng1990>
- Weikard, R., Demasius, W., Kuehn, C., 2017. Mining long noncoding RNA in livestock. *Anim Genet* 48, 3–18. <https://doi.org/10.1111/age.12493>
- Weisenberger, D.J., Berg, D.J.V.D., Pan, F., Berman, B.P., Laird, P.W., 2008. Comprehensive DNA Methylation Analysis on the Illumina ® Infinium ® Assay Platform.
- Weng, Y.-I., Huang, T.H.-M., Yan, P.S., 2009. Methylated DNA immunoprecipitation and microarray-based analysis: detection of DNA methylation in breast cancer cell lines. *Methods Mol Biol* 590, 165–176. https://doi.org/10.1007/978-1-60327-378-7_10
- Wessely, F., Emes, R.D., 2012. Identification of DNA methylation biomarkers from Infinium arrays. *Front Genet* 3, 161. <https://doi.org/10.3389/fgene.2012.00161>
- Wolff, G.L., Kodell, R.L., Moore, S.R., Cooney, C.A., 1998. Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J* 12, 949–957.
- Wolffe, A.P., Guschin, D., 2000. Review: chromatin structural features and targets that regulate transcription. *J Struct Biol* 129, 102–122. <https://doi.org/10.1006/jsbi.2000.4217>

- Wu, C., Sirard, M.-A., 2020. Parental Effects on Epigenetic Programming in Gametes and Embryos of Dairy Cows. *Front. Genet.* 11. <https://doi.org/10.3389/fgene.2020.557846>
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2, 100141. <https://doi.org/10.1016/J.XINN.2021.100141>
- Yakovlev, A.F., 2018. Epigenetic Effects in Livestock Breeding. *Russ J Genet* 54, 897–909. <https://doi.org/10.1134/S1022795418080148>
- Ying, Z., Chen, T., 2023. Chapter 2 - Mechanisms of DNA Methylation and Demethylation During Mammalian Development, in: Tollefsbol, T.O. (Ed.), *Handbook of Epigenetics (Third Edition)*. Academic Press, pp. 11–26. <https://doi.org/10.1016/B978-0-323-91909-8.00009-8>
- 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. <https://doi.org/10.1093/BIOINFORMATICS/BTV145>
- Zehetmeier, M., Gandorfer, M., Hoffmann, H., Müller, U.K., de Boer, I.J.M., Heißenhuber, A., 2014. The impact of uncertainties on predicted greenhouse gas emissions of dairy cow production systems. *Journal of Cleaner Production, Towards eco-efficient agriculture and food systems: Selected papers from the Life Cycle Assessment (LCA) Food Conference, 2012*, in Saint Malo, France 73, 116–124. <https://doi.org/10.1016/j.jclepro.2013.09.054>
- Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919. https://doi.org/10.1126/SCIENCE.1186366/SUPPL_FILE/ZEMACH.SOM.PDF
- Zeng, Y., Chen, T., 2019. DNA Methylation Reprogramming during Mammalian Development. *Genes* 10, 257. <https://doi.org/10.3390/genes10040257>
- Zhang, Y., Li, F., Feng, X., Yang, H., Zhu, A., Pang, J., Han, L., Zhang, T., Yao, X., Wang, F., 2017. Genome-wide analysis of DNA Methylation profiles on sheep ovaries associated with prolificacy using whole-genome Bisulfite sequencing. *BMC Genomics* 18, 759. <https://doi.org/10.1186/s12864-017-4068-9>
- Zhang, Yunfang, Zhang, X., Shi, J., Tuorto, F., Li, X., Liu, Yusheng, Liebers, R., Zhang, L., Qu, Y., Qian, J., Pahima, M., Liu, Ying, Yan, M., Cao, Z., Lei, X., Cao, Y., Peng, H., Liu, S., Wang, Y., Zheng, H., Woolsey, R., Quilici, D., Zhai, Q., Li, L., Zhou, T., Yan, W., Lyko, F., Zhang, Ying, Zhou, Q., Duan, E., Chen, Q., 2018. Dnmt2 mediates intergenerational transmission of paternally acquired metabolic disorders through sperm small non-coding RNAs. *Nat Cell Biol* 20, 535–540. <https://doi.org/10.1038/s41556-018-0087-2>
- Ziller, M.J., Hansen, K.D., Meissner, A., Aryee, M.J., 2015. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat Methods* 12, 230–232. <https://doi.org/10.1038/nmeth.3152>
- Zuccolo, J., Bau, J., Childs, S.J., Goss, G.G., Sensen, C.W., Deans, J.P., 2010. Phylogenetic Analysis of the MS4A and TMEM176 Gene Families. *PLOS ONE* 5, e9369. <https://doi.org/10.1371/journal.pone.0009369>

- Zulu, V.C., Nakao, T., Sawamukai, Y., 2002. Insulin-Like Growth Factor-I as a Possible Hormonal Mediator of Nutritional Regulation of Reproduction in Cattle. *Journal of Veterinary Medical Science* 64, 657–665. <https://doi.org/10.1292/jvms.64.657>

