



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	Identification and analysis of polymorphisms in miRNA genes and associated targets in <i>Bos taurus</i> and <i>Bos primigenius</i> for genomic selection in cattle
Author(s)	Braud, Martin
Publication Date	2017-02-21
Item record	http://hdl.handle.net/10379/6504

Downloaded 2024-03-13T09:46:35Z

Some rights reserved. For more information, please see the item record link above.



**Identification and analysis of polymorphisms in
miRNA genes and associated targets in *Bos taurus*
and *Bos primigenius* for genomic selection in cattle**

Martin Braud

Volume I of II

**A thesis submitted to National University of Ireland, Galway
for the degree of Doctor in Philosophy**

Submitted: February 2017

Primary Supervisor: Prof. Charles Spillane

Secondary Supervisor (Teagasc): Dr. Sinéad Waters

External Examiner: Dr. Emmeline Hill

Internal Examiner: Dr. Derek Morris

Funding: Walsh fellowship program, number: 2010062

Student ID: 10233139

Table of Content

1	General Introduction.....	1
1.1	THE ORIGIN OF CATTLE BREEDS	1
1.1.1	<i>The aurochs, the ancestor of modern cattle.....</i>	<i>1</i>
1.1.2	<i>The domestication of cattle</i>	<i>4</i>
1.1.3	<i>The origin of cattle breeds</i>	<i>6</i>
1.2	GENETIC AND GENOMIC METHODS DEVELOPMENT FOR CATTLE BREEDING AND SELECTION IMPROVEMENT	8
1.2.1	<i>Introduction to genetic marker in livestock breeding</i>	<i>8</i>
1.2.2	<i>Quantitative trait loci (QTL).....</i>	<i>9</i>
1.2.3	<i>Single Nucleotide Polymorphisms (SNPs) improvement of the genomic selection</i>	<i>10</i>
1.2.4	<i>SNP discovery in cattle</i>	<i>13</i>
1.2.5	<i>The International Dairy and Beef SNP Chip</i>	<i>15</i>
1.2.6	<i>Association studies.....</i>	<i>15</i>
1.2.7	<i>Genomic Selection.....</i>	<i>17</i>
1.3	CATTLE BREEDING IN IRELAND	18
1.3.1	<i>Dairy and beef production are an important part of Irish industry.....</i>	<i>18</i>
1.3.2	<i>Important cattle breeds</i>	<i>19</i>
1.3.2.1	Holstein-Friesian	19
1.3.2.2	Jersey	19
1.3.2.3	Aberdeen-Angus:.....	20
1.3.2.4	Limousin.....	20
1.3.2.5	Romagnola.....	20
1.3.2.6	Fleckvieh	20
1.3.3	<i>Irish programs for breeding improvement.....</i>	<i>21</i>
1.3.3.1	Dairy breeding programmes	21
1.3.3.2	Beef breeding programmes.....	24
1.4	THE ROLE OF MICRO-RNAs AND THEIR IMPORTANCE IN LIVESTOCK	25
1.4.1	<i>Discovery of microRNAs.....</i>	<i>25</i>
1.4.2	<i>MicroRNA structure.....</i>	<i>27</i>
1.4.3	<i>MicroRNA biogenesis</i>	<i>29</i>
1.4.4	<i>Role of miRNA in livestock economically important traits</i>	<i>30</i>
1.4.5	<i>Polymorphisms in miRNA</i>	<i>32</i>
1.5	BIOINFORMATICS TOOLS FOR IDENTIFICATION OF MIRNA TARGETS	34
1.5.1	<i>miRanda</i>	<i>34</i>
1.5.2	<i>Diana-microT.....</i>	<i>34</i>
1.5.3	<i>RNAhybrid.....</i>	<i>35</i>

1.5.4	<i>PicTar</i>	35
1.5.5	<i>TargetScan</i>	35
1.6	RESEARCH HYPOTHESIS AND OBJECTIVES.....	37
2	Genome-wide microRNA and microRNA binding site variation between extinct wild aurochs and modern cattle identifies candidate microRNA-regulated domestication genes.....	40
2.1	INTRODUCTION	40
2.2	MATERIALS AND METHODS	42
2.2.1	<i>B. primigenius</i> genome sequence data	42
2.2.2	Workflow for the identification and comparison of polymorphic binding sites.....	42
2.2.3	<i>B. primigenius</i> miRNA identification	44
2.2.4	Identification of <i>B. primigenius</i> 3' UTR sequences.....	44
2.2.5	Identification of miRNA complementary binding sites in 3' UTR using <i>TargetScan</i>	45
2.2.5.1	TargetScan results treatment pipeline.....	45
2.2.5.2	Score Site ratio (SSr): development of a metric for the comparison of polymorphic binding sites.....	47
2.2.6	Analyses of biological pathways involved in differential miRNA targeting of <i>Bos taurus</i> genes relative to <i>B. primigenius</i>	48
2.2.7	Analysis of transition/transversion ratios	48
2.3	RESULTS	49
2.3.1	Identification and conservation of microRNAs in the <i>B. primigenius</i> genome.....	49
2.3.2	Precursor microRNA genes have not been lost in the <i>B. taurus</i> genome relative to the wild <i>B. primigenius</i> genome.....	50
2.3.3	<i>Bta-mir-2893</i> contains a SNP between domesticated <i>B. taurus</i> and wild <i>B. primigenius</i> (<i>bpr-mir-2893</i>) that leads to differential targeting	51
2.3.4	The SNP present in <i>miR-2893</i> leads to loss and gain of gene targets between <i>B. taurus</i> and <i>B. primigenius</i> potentially dysregulating <i>PHYHIP</i> , <i>FADS2</i> and <i>FCRL1</i> expression	53
2.3.5	Neurodevelopment, fertility and pigmentation pathways are enriched for genes targeted by <i>miR-2893</i>	55
2.3.6	Comparison of <i>B. taurus</i> and <i>B. primigenius</i> 3' UTR sequences revealed a large number of DNA sequence polymorphisms.....	58
2.3.7	DNA sequence polymorphisms differentiating <i>B. taurus</i> and <i>B. primigenius</i> 3' UTRs generated altered miRNA binding sites.....	58
2.3.8	Using SSr scores to rank the 3' UTRs that had undergone the most extensive miRNA binding site modifications between <i>B. primigenius</i> and <i>B. taurus</i>	61

2.3.9	<i>Identification of networks, pathways, and functions enriched in genes having polymorphic miRNA targets sites between B. taurus and B. primigenius.....</i>	63
2.3.10	<i>The polymorphic miRNA binding sites in aurochs 3' UTRs were not enriched for transitions due to post-mortem deamination.....</i>	63
2.4	DISCUSSION	65
2.4.1	<i>Polymorphisms between B. primigenius and B. taurus affect miRNA binding to 3' UTRs</i>	65
2.4.2	<i>Neurological development is affected by the modification of miRNA binding between B. primigenius and B. taurus and could be linked to tameness.....</i>	66
2.4.3	<i>Change in metabolism during domestication could be linked to the SNP in mir-2893 and polymorphisms in miRNA-binding sites</i>	67
2.4.4	<i>Modification in miRNA gene and miRNA-binding sites play a role in potential immunologic changes during domestication.....</i>	68
2.4.5	<i>Reproduction, fertility and lactation are important traits in the extension of domestication of B. taurus and is linked to modification in miRNA and miRNA binding sites.....</i>	69
2.4.6	<i>Pigmentation pathway genes are differentially targeted between B. taurus and B. primigenius.....</i>	69
2.5	CONCLUSION	70
3	Identification of cattle breeds polymorphisms in miRNA sequences discover candidates for association studies.....	72
3.1	INTRODUCTION	72
3.2	MATERIALS AND METHODS	75
3.2.1	<i>Flowchart of identification of SNPs in miRNA for association study candidates</i>	75
3.2.2	<i>Identification of Single Nucleotide Polymorphism in miRNA genes.....</i>	75
3.2.3	<i>Prediction of the secondary structure of polymorphic miRNA</i>	77
3.2.4	<i>Construction of 3' UTRs alignment.....</i>	77
3.2.5	<i>Identification of the targets of the polymorphic miRNAs.....</i>	78
3.2.6	<i>Sorting and SSr-scoring of miRNAs variants.....</i>	78
3.2.7	<i>Pathway enrichments using IPA</i>	78
3.2.8	<i>Genotyping with the IDB SNP chip of 25000 cattle for frequency analysis.....</i>	81
3.2.9	<i>Genotyping for gene association.....</i>	81
3.2.9.1	<i>Data editing and phenotypes.....</i>	81
3.2.9.2	<i>Statistical analysis.....</i>	83
3.3	RESULTS AND DISCUSSION	84
3.3.1	<i>Identification of miRNA SNPs from dbSNP</i>	84
3.3.2	<i>Identification of miRNA SNPs for six cattle breeds</i>	84

3.3.3	<i>Pre-miRNA secondary structure modifications in variant miRNA</i>	88
3.3.4	<i>Identification of new target sites for polymorphic mature miRNA</i>	92
3.3.5	<i>Pathways enriched for miRNA targets are different between reference and variant miRNAs</i>	97
3.3.6	<i>QTL enrichment in polymorphic miRNA targets suggested influence of miRNA in cattle traits</i>	107
3.3.7	<i>Genotyping of seven miRNA polymorphisms showed difference in allele frequencies across 18 cattle breeds</i>	109
3.3.8	<i>Association analysis of seven miRNA SNPs in Holstein-Friesian breed</i>	110
3.3.9	<i>Analysis of QTLs and Pathway enrichment for mir-2419-3p and mir-1814b</i>	111
3.4	CONCLUSIONS.....	115
3.4.1	<i>mir-2419 presented two different SNPs which could have an influence on both dairy and beef cattle</i>	115
3.4.2	<i>Mir29e and mir-449b</i>	116
3.4.3	<i>Jersey specific SNPs in miRNA and population structure</i>	117
4	3' UTR variants in cattle breeds modify microRNA binding sites and display association with milk production traits	118
4.1	INTRODUCTION	118
4.2	MATERIALS AND METHODS	120
4.2.1	<i>Identification of Single Nucleotide Polymorphism in cattle genes 3' UTR</i>	120
4.2.2	<i>Construction of 3' UTRs alignment</i>	120
4.2.3	<i>Identification and analysis of polymorphic miRNA binding sites</i>	120
4.2.4	<i>Genotyping for gene association</i>	121
4.2.5	<i>Association study</i>	121
4.2.6	<i>Analysis of SNPs from 83 candidate genes for miRNA binding discovery</i>	121
4.3	RESULTS AND DISCUSSION	122
4.3.1	<i>Identification of 3' UTR SNPs in six cattle breeds</i>	122
4.3.2	<i>Polymorphism in 3' UTRs create or impair miRNA-binding sites</i>	124
4.3.3	<i>Pathways enriched for polymorphic 3' UTR are related to immune response</i>	131
4.3.4	<i>QTL enrichment in polymorphic 3' UTR</i>	134
4.3.5	<i>Genotyping of 25 3' UTR SNPs across 25 cattle breeds</i>	136
4.3.6	<i>Association analysis of genotyped SNPs in Holstein-Friesian breed</i>	138
4.3.7	<i>Identification of polymorphism in 3' UTR related to miRNA binding sites in candidate genes</i>	139
4.4	CONCLUSIONS.....	141

5	General discussion	144
6	Conclusion	150
7	List of Appendices (Volume II)	152
8	Publication.....	153
8.1	GENOME SEQUENCING OF THE EXTINCT EURASIAN WILD AUROCHS, BOS PRIMIGENIUS, ILLUMINATES THE PHYLOGEOGRAPHY AND EVOLUTION OF CATTLE	153
8.2	GENOME-WIDE MICRORNA BINDING SITE VARIATION BETWEEN EXTINCT WILD AUROCHS AND MODERN CATTLE IDENTIFIES CANDIDATE MICRORNA-REGULATED DOMESTICATION GENES 153	
	Bibliography	154

List of Figures

Figure 1 – Distribution of <i>Bos primigenius</i> subspecies.	2
Figure 2 – Skeleton of <i>Bos primigenius</i>	3
Figure 3 – Variation of size from aurochs to modern cattle through domestication.	4
Figure 4 – Phylogenetic representation of the <i>Bos</i> genus haplogroups.	7
Figure 5 – Example of SNP chip from Illumina.....	13
Figure 6 – Distribution of the main breeds in Ireland (Animal Identification and Movement (AIM) statistic report 2014).	18
Figure 7 – EBI averages by birth year for females and for bulls on active bull list in each year (ICBF annual report 2014).....	23
Figure 8 – Representation of the miRNA nomenclature with <i>bta-mir-2419</i> example.	28
Figure 9 – Biogenesis of miRNA from the pri-miRNA transcription to the binding of the mature RNA to the 3' UTR (Figure adapted from Kim et al. (2009))	29
Figure 10 – Workflow for the identification of polymorphic miRNA binding sites.	43
Figure 11 – Representation of the target sites locus restriction.	46
Figure 12 – Predicted pre-miRNA of mir-2893 in the <i>B. primigenius</i> (bpr) genome aligned to the <i>B. taurus</i> (bta) mature sequence of mir-2893.	50
Figure 13 – <i>B. taurus</i> miR-2893 targets enriched pathways.....	56
Figure 14 – <i>B. primigenius</i> miR-2893 targets enriched pathways.....	57
Figure 15 – Polymorphisms between <i>B. taurus</i> and <i>B. primigenius</i> in 3' UTR create or impair miRNA-binding sites.	60
Figure 16 – Pathways significantly enriched in genes having polymorphic miRNA-binding sites between <i>B. taurus</i> and <i>B. primigenius</i>	64
Figure 17 – Flowchart of identification of SNPs in miRNA for association study candidates.....	76

Figure 18 – Representation of the different groups of pathway enrichment.	80
Figure 19 – Distribution of polymorphic miRNA genes in genome.	85
Figure 20 – Heatmap of the impact of polymorphism in pre-miRNA on their thermodynamic stability.....	89
Figure 21 – Secondary structure of <i>mir-2456</i> and <i>mir-2895</i> is modified in their variants.	91
Figure 22 – Comparison of the number of targeted genes between the reference and variant miRNAs.	96
Figure 23 – Number of differentially enriched pathways in polymorphic miRNAs and number of their related targets showed a higher rate of lost pathways enrichment and lost targets.	98
Figure 24 – The miRNAs <i>mir-449b</i> and <i>mir-29e</i> variant targets differently IL-7 signalling pathway.....	101
Figure 25 – The miRNAs <i>mir-449b</i> and <i>mir-29e</i> variants targets differently Prolactin signalling pathway.....	103
Figure 26 – Allele frequencies of miRNA SNPs in cattle breeds.....	109
Figure 27 – Representation of carcass weight trait related QTL overlapping <i>mir-1814b</i> gene.	114
Figure 28 – Distribution of SNPs located in 3' UTR throughout cattle genome.	122
Figure 29 – Gene distribution across the different groups and breed.....	124
Figure 30 – Single nucleotide polymorphisms in <i>GIMAP1</i> 3' UTR through six cattle breeds creates lost and gain of miRNA binding sites.....	126
Figure 31 – Genotyping frequency of SNPs in miRNA binding sites across cattle breeds...	137

List of Tables

Table 1 – Number of predicted 3' UTR targets for mature miRNAs (bta-miR-769, bta-miR-940, bta-miR-2391, bta-miR-2469, bta-miR-2893) having polymorphic pre-miRNA between <i>B. taurus</i> and <i>B. primigenius</i>	52
Table 2 – Top ranked genes differentially targeted by bta-miR-2893 and bpr-miR-2893	54
Table 3 – Candidates domestication genes having polymorphic miRNA binding sites between <i>B. taurus</i> and <i>B. primigenius</i> presenting.....	62
Table 4 – Description of the six cattle breeds sequenced.	77
Table 5 – List of the SNPs identified with dbSNP 130 data.....	84
Table 6 – List of polymorphism in miRNA genes and their location in mature and seed region, in six cattle breeds	86
Table 7 – List of miRNAs having SNPs creating a seed region identical to another miRNA.....	87
Table 8 – List of polymorphic miRNA identified in cattle breeds and their predicted binding sites with no common target with the reference	93
Table 9 – Pathways enriched in unique miRNA variant targets.....	106
Table 10 – QTL enriched in multiple miRNA variant targets.....	108
Table 11 – Suggestive traits (p-value < 0.05) associated to the miRNA SNPs investigated	111
Table 12 – mir-1814b and mir-2419 are overlapping QTLs loci.....	112
Table 13– List of top genes having the most SNPs in their 3' UTR.....	123
Table 14 – Top ranked genes having the highest SSR score	127
Table 15 – Pathways significantly enriched with gene having polymorphic miRNA-binding sites	133

Table 16 – Traits associated to QTLs significantly enriched with genes having polymorphic miRNA binding sites	134
Table 17 – List of traits and SNPs tested for association with significant nominal p-value.....	139
Table 18 – List of top 10 polymorphic miRNA binding sites among a subset of 83 genes	140

Acknowledgment

Firstly, I would like to thanks Prof. Charles Spillane, Dr. Dermot Morris and Dr. Sinead Waters, for giving me the opportunity to do this PhD and for their help and support.

I would like to thanks Dr. Michael Mullen, Dr. Donagh Berry for integrating the SNPs identified into the IDB chips and providing the association results.

I would like to thanks Prof. David MacHugh, Dr. David Magee and Dr. Stephen Parks for giving me the opportunity to participate to the functional analysis of miRNAs during the *Bos primigenius* ancient DNA study.

I would like to thanks all the member of Charles Spillane lab, past and present with whom I spent a great time and have always been there to support me: Antoine, Peter R, Peter M, Sandesh, Mark, Duygu and Eva.

I also want to thanks Udo for his advices, cheering around coffee and keeping up my motivation. I also want to thanks all the great people I met in NUIG, Michel, Guillaume, Holly, Matthias Freddy and Lilit.

I want to thanks Imola for her love and help, going through the last year of my PhD, Nagy puszi koalackam!

Finally, I want to thanks my family who have always been there for me, my sisters Pauline and Armelle, my Dad and a special thought to my Mom.

Abstract

Cattle production is of high economic importance to agriculture in Ireland. For more than a decade, significant effort focused on the improvement of the selection of economically important traits in beef and dairy breeds. The application of genomics technologies has greatly improved our ability to detect genetic markers related to important traits.

This study identified novel Single nucleotide polymorphisms (SNP) between aurochs and cattle in micro-RNAs (miRNA) and miRNA-binding sites. This study also identified SNPs in miRNAs then SNPs in miRNA 3' untranslated region (3'UTR) targets between cattle breeds.

One polymorphic miRNA gene active region and 1606 polymorphic miRNA target genes have been identified between aurochs and modern cattle. These polymorphic miRNA and miRNA-binding sites, which arose since domestication, showed associations with important functions such as neurodevelopment and food metabolism that could have been selected by humans. These polymorphisms may represent potential DNA based biomarkers for improving these specific traits in cattle.

The screening of SNPs in miRNA between cattle breeds identified 89 miRNAs, which have different sequences compared to the reference miRNAs. Out of these 89, 25 had SNPs in their miRNA binding region. The analysis of change of targets and their related function highlighted three candidates of interest: *mir-2419*, *mir-449b* and *mir-29e*. These three candidate miRNAs have polymorphisms, which modify their binding properties to target genes, related to milk production. Some of these miRNA SNPs have been included into an association study on 914 Holsteins for which the traits of their 50,000 daughters had been quantified. However, only one SNP has been weakly associated with milk traits. This SNP is located in *mir-2419-3p* and is weakly associated with rear udder height.

The screening of polymorphisms in 3' UTR of cattle breeds allowed the identification of 726 genes containing 1097 different SNPs in their miRNA-binding sites. The analysis of these genes showed links with various functions of economic importance in cattle. The SNPs identified in the 3' UTRs of *TARBPI*, *GUCA2B* and *CYP3A5* were associated with milk production traits in a candidate genes association study on Holstein breed. Furthermore, a SNP in the 3' UTR of *PRKAG1* was identified, which created a change in a miRNA-binding site. This could have implications for milk

production, considering that *PRKAG1* was previously shown to be associated with this trait. Moreover, three other polymorphic miRNA-binding target genes, *BTS2B*, *GIMAP1* and *LAMTOR2* have been related to immune response in cattle. These SNPs identified in miRNA-binding sites warrant further investigation.

Analysis of miRNA-related polymorphisms in the aurochs ancestor and the commonly used beef and dairy breeds of cattle facilitated the discovery of candidate DNA markers useful for the better understanding and the improvement of traits in beef and dairy breeds. These novel SNPs could, (following validation on very large population of cattle) contribute to the national genome selection of beef and dairy breeding programme. These SNPs have been included on previous version of the International Dairy and Beef SNP-Chip (1 and 2) as well as the latest version (v3), used for the national genomic evaluations of beef and dairy herds in Ireland.

1 General Introduction

1.1 The origin of cattle breeds

1.1.1 The aurochs, the ancestor of modern cattle

Genomic analysis of the wild ancestor of domesticated animals is of major interest to decipher the domestication process. The improvement of genome sequencing methods allowed the access to genomic information of domesticated animals and their ancestors, facilitating a greater understanding of how natural and artificial selection leads to the variation in the modern domesticated bovine population (Groenen, 2012; Larkin, 2012; Schubert, 2014).

The Aurochs (*Bos primigenius*) is considered as the progenitor of modern taurine cattle (*B. taurus*) and zebu (*B. indicus*). The distribution of aurochs covered most part of Eurasia, from west Europe to India and Northern Africa as well as North and South Asia excluding Northern Russia (**Figure 1**). Three main subspecies have been identified based on archaeological artefacts. These include *B. primigenius primigenius* in Eurasia, which is suggested to be the progenitor of the taurine, *B. p. namadicus* in the Indian subcontinent thought to be the progenitor of the zebu and *B. p. opisthonomus* in Africa, which is considered to be the progenitor of some African breeds (Grigson, 1978, 1980).

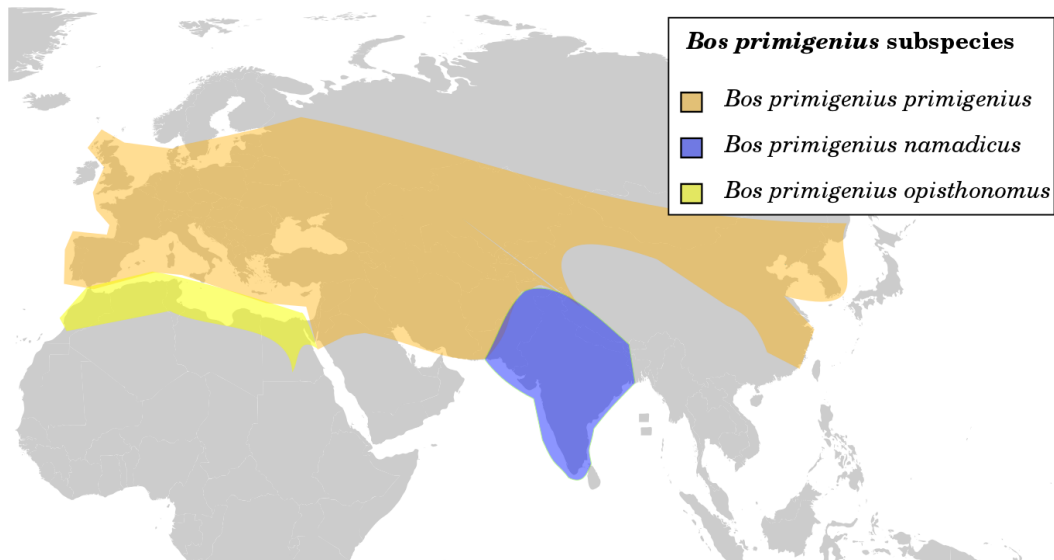


Figure 1 – Distribution of *Bos primigenius* subspecies. Based on archaeological discoveries, three subspecies of *B. primigenius* have been identified in Africa (*B. p. opisthonomus*, yellow), in Eurasia (*B. p. primigenius*, orange) and in the Indian subcontinent (*B. p. namadicus*, Blue). (Adapted from Van Vuure, 2005)

The morphology of aurochs is similar to modern cattle and was described by Gaius Julius Caesar in the *Commentarii de bello Gallico*: “[...] a little below the elephant in size, and of the appearance, the colour and shape of a bull. Their strength and speed are extraordinary, they spare neither man nor wild beast which they have espied”. The aurochs bull, has been estimated from bone remains to reach up to 2 m at the shoulder while the aurochs cow was around 1.50 m. The aurochs is recognisable from its horns as it displays a specific shape, curving forward from the skull and inward (**Figure 2**). The last recorded animal of the species died in 1627 in the area of the Jaktorów forest in Poland (Rokosz, 1995).



Figure 2 – Skeleton of *Bos primigenius*. This skeleton from Copenhagen museum (Marcus Sümnick, Creative Commons licence BY-SA 2.0) shows the specific horn shape that helped to distinguish wild aurochs from domestic cattle, along with size and bone structure.

The domestication of cattle from aurochs has been identified in two separate geographical areas. One event happened in the Near East, giving rise to the taurine lineage (*B. taurus*) including the European and African breeds when the second event took place in south Asia giving rise to zebuine lineage (*B. indicus*) including the Indian breeds (Loftus, 1994; Troy, 2001; Chen, 2010). Since these cattle domestication events, multiple cattle breeds have emerged through selection of traits of interest for different environment and human needs. For example, the remains of Neolithic domesticated bulls showed that these first animals were much smaller than the aurochs (Van Vuure, 2005), but later selection brought some modern breeds to a similar size to the aurochs (Figure 3).

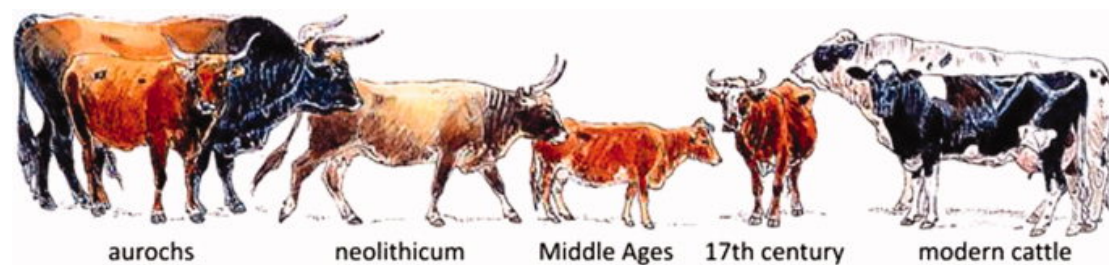


Figure 3 – Variation of size from aurochs to modern cattle through domestication. (Adapted from (Ajmone-Marsan, 2010)) Copyright © 2010 Wiley Periodicals, Inc.

1.1.2 The domestication of cattle

The earliest record of cattle domestication from archaeological material can be traced back to the 9th millennium BCE in the Near East and the 8th millennium BCE in the Indus valley. The first traces of domesticated cattle in Europe were found in southern Europe dating from the 7th millennium and from the 6th millennium in the rest of Europe (Ajmone-Marsan, 2010). The analysis of mitochondrial DNA (mtDNA) showed high diversity for *B. indicus* in the Indian sub-continent and for *B. taurus* in the so-called fertile crescent, supporting claims for early domestication occurring in these regions (Troy, 2001; Bradley, 2006; Edwards, 2007). A hyper-variable region in the mtDNA, called the D-loop, has been used to identify polymorphisms between different cattle breeds and confirmed the existence of two main haplogroups: haplogroups I for zebu and T for taurine (Achilli, 2009). The analysis and comparison of this mtDNA region between *B. indicus* and *B. taurus*, showed that differentiation occurred before domestication (Loftus, 1994). This study, identified and analysed two major hypervariable regions in the mtDNA D loop in European, African and Indian breeds.

They found a high level of divergence between Indian and European breeds while European and African breeds were closely related. They estimated that the divergence between Afro-European breeds and the Indian breeds occurred between 740,000 and 1.04 Myr. This divergence suggests that two independent domestication events occurred in different locations, and involving two different aurochs subspecies. Furthermore, the close relationship between European and African cattle breeds suggested a common ancestry, in opposition to what have been previously thought. The phenotypical resemblance of some African breeds to zebu is thought to be due to cross breeding from male zebu imported during the Arabic invasion (Ajmone-Marsan, 2010).

The origination of European cattle breeds from Near East aurochs, received further support from the sequencing of the mtDNA variable regions in 112 samples from *Bos* species (Edwards, 2007). These samples were composed of 83 aurochs-labelled bones and other *Bos* species, taurine and possibly bison or buffalos from archaeological sites across Europe and Near Eastern region. They identified a majority of P haplotype, the marker of European aurochs, but also a new haplotype designated as E from a German site. The phylogenetic tree, reconstructed from these sequences with taurine and zebu haplogroups, showed that the P haplogroup is closer to the T haplogroup (taurine) than the I haplogroup (zebu). The E haplogroup is clustering outside of these groups but is close to the T and P haplogroup clusters. Furthermore, they identified a T haplogroup in a sample from Syria labelled as *B. primigenius* from the archaeological record, dated between 10,700 and 10,300 years ago. This sample could correspond to an early domesticated taurine or an aurochs from the early Holocene. Even though the identity of the exact species could be debated, they presented it as a common European breed haplogroup, found in the suggested domestication origin region in early Neolithic, confirming the Near Eastern origin of modern European cattle.

Recent progress in ancient DNA handling and analysis techniques facilitated the first sequencing of the full nuclear genome (Park, 2015) and mitochondrial genome (Edwards, 2010) of *B. primigenius*. The contribution from this thesis to the Park *et al* analysis of the nuclear genome for the miRNA and miRNA binding sites is presented in the chapter 2 and will be discussed further in that chapter. “Nuclear sequencing revealed important new information regarding the origins and functional evolution of modern cattle, and discovered that the interface between early European domestic

populations and wild aurochs was significantly more complex than previously thought” (Park et al., 2015).

The genomes were extracted from a British aurochs bone pre-dating Neolithic period in Britain, and the sequence of the mtDNA showed variation corresponding to the P haplogroup. A phylogenetic comparison of the aurochs mtDNA coding sequence with the 148 cattle mtDNA sequences available in 2010 confirmed the clustering of the T and P haplotypes in different groups, showing that their divergence predated the divergence within the T haplogroup (Edwards, 2010). The analysis of the aurochs nuclear genome compared with different cattle breeds showed differences in genes related to the immune system, metabolism and the neurological system, which may be linked to domestication traits. Although this study suggested that the Northern European aurochs is an outgroup of domesticated cattle, it also showed level of admixture within domestic cattle herds from Northern Europe, mainly with British and Irish local breeds, when tested for post-domestication hybridisation on 79 breeds (Park, 2015). This showed that even if the major haplogroup of European cattle breeds originated from middle east, hybridisation between aurochs and ancient domesticated cattle happened and could have played a role in the selection of traits present in modern cattle.

1.1.3 The origin of cattle breeds

As discussed previously, analysis of mitochondrial genome markers from different cattle breeds and the European wild aurochs showed different haplogroups, termed T, I, P, E, R and Q (**Figure 4**), corresponding to the different origins of the modern taurine (Achilli, 2009) and zebuine. Haplogroup T represents the modern taurine, including the sub-haplogroup T3, which includes most of the European breeds (Troy, 2001) while indicine cattle (zebu) belong to haplogroup I (Chen, 2010).

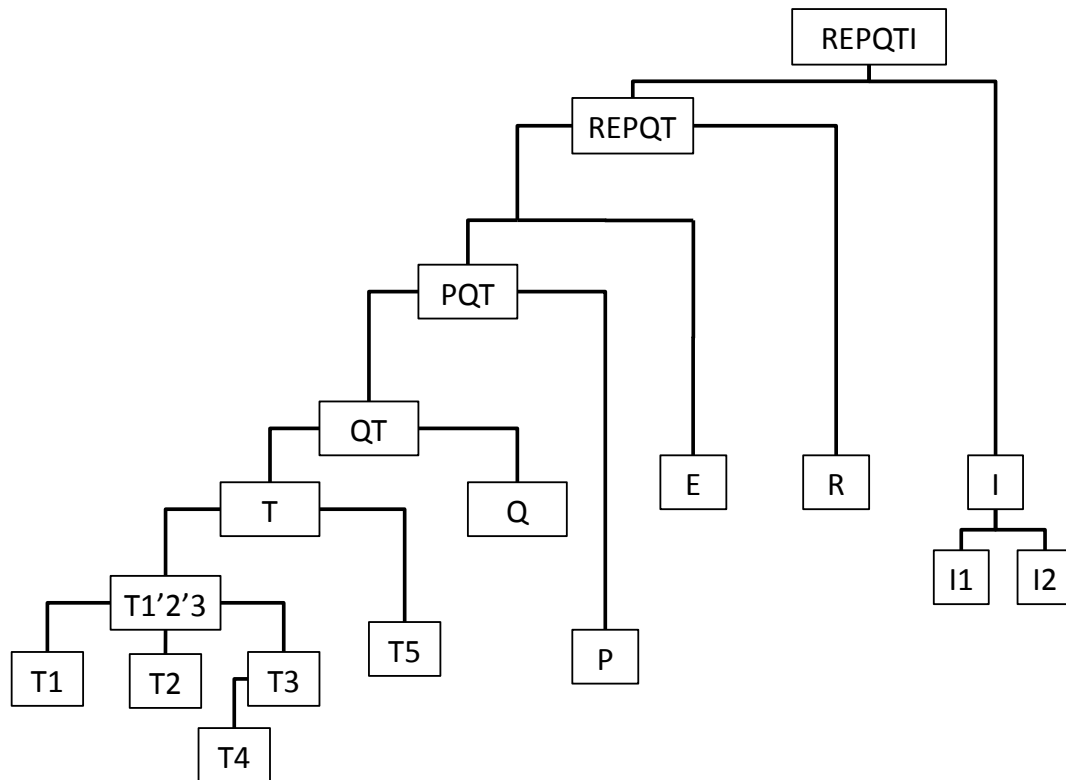


Figure 4 – Phylogenetic representation of the *Bos* genus haplogroups. The figure represents the phylogenetic tree of the different *Bos* species (extinct aurochs and modern cattle) based on haplogroups identified from mtDNA (adapted from Achilli 2009).

Among haplogroup T, haplotype T1, which contains six sub-clades designated T1a-T1f, is present to a small extent in European breeds but is highly represented in African and Creole breeds (Bonfiglio, 2012). However, not all European breeds are directly linked to haplotype T and its sub-clades. For example, haplogroup Q is a rare group, poorly differentiated from T3 sub-clade, which has been identified in Neolithic German, French, Iberian and Turkish cattle (Bonfiglio, 2010; Lopez-Oceja, 2015) as well as in modern local breeds from south-western China (Lai, 2006; Cortés, 2008). These data suggest that the presence of haplogroups Q and T in modern cattle, due to their evolution proximity, could have been part of the aurochs population from the Near East region, which was involved in the domestication event. Haplotype R is also a rare novel haplotype that has been identified in Italian breeds such as Agerolese or Romagnola. However this haplogroup diverges earlier in the phylogenetic tree than haplogroup P (specific to European aurochs), which has been suggested to result from

interbreeding between native Italian aurochs females and domesticated bulls, allowing the transmission of aurochs mitochondrial DNA to Italian breeds (Achilli, 2008).

1.2 Genetic and genomic methods development for cattle breeding and selection improvement

1.2.1 Introduction to genetic marker in livestock breeding

Early in the history of cattle husbandry, animal with traits recognisable visually such as sturdiness or behaviour for draught purpose were selected to improve cattle characteristics without knowledge of their genetics and genomics control (Ajmone-Marsan, 2010). To improve the traits and their transmission, breeders and evolutionary biologist developed methods of prediction. Quantitative genetics began by predicting how the mean value of a traits is transmitted, through the “breeder's equation”:

$$R = h^2 S$$

This equation represents the selection response (R) by including the relationship with heritability (h^2) expressed as the ratio of the phenotypic variability to the genetic variability and the selection differential (S) expressed as the average trait difference between the whole population and the selected parents generation (Slatkin, 1970; Roughgarden, 1972; Lande, 1976). Heritability is a useful statistic for quantifying the response to selection (Visscher, 2008). This equation is present in different forms representing more specific variables over the years and led to the introduction of Bayesian models (Sorensen, 2002).

The heritability of traits are variable, a general trend has emerged as morphological traits are more heritable than growth rate which itself is more heritable than reproductive traits (Hill, 2010). The identification of genetic and genomic variations is representing an important challenge to be able to understand, improve and maintain the different traits of economic interest in cattle and other livestock species.

To be able to measure this heritability and predict the transmission of traits, different techniques have been used. The first one was blood and protein test, giving some information on which were the parents of a given cattle, then linking recognised traits in the parent and the progeny to these first markers (Neimann-Sorensen, 1961; Stormont, 1967). Then the discovery of microsatellite, DNA motif of which repetitions are varying from one individual to another, became available as markers (Litt, 1989; Weber, 1989) and made the parentage identification and the trait recognition in cattle

more accurate (MacHugh, 1994). This new type of marker linked to heritable traits helped to identify specific locus related to a phenotype, improving the understanding of the underlying genomics.

1.2.2 Quantitative trait loci (QTL)

The quantitative trait locus (QTL) is the link between a DNA portion of the genome of a species and the variation in a quantitative trait. The need to have a better understanding of polygenic interaction on traits started to be answered with the methods to identify QTL. To map a QTL on a genome it is necessary to identify variations in the genome of inbred strains and detecting the phenotypic differences, using statistics methods like analysis of variance (ANOVA) or minimum likelihood (ML) methods (Mackay, 2001; Hill, 2010). Linkage disequilibrium (LD) is defined as the non-random association between two variable loci on a given chromosome. Two loci are in LD when the probability of the frequency of one allele at a locus depends on the probability of the presence of another allele at a different genetically linked locus.

(Lewontin, 1960). LD also allows the mapping of genes associated with complex traits. The LD also help to define haplotype block. A haplotype block is a section of the genome where all alleles are in strong LD such as an allele associated with a phenotype suggest that all the alleles of the haplotype block are associated (Slatkin, 2008).

The first markers identified for selection, which were based on blood typing, have been used in one of the first study on QTL identification in livestock (Neimann-Sorensen, 1961). Phenotypic differences among breeds and between domesticated species and their wild ancestor seemed to be caused essentially by small genetic changes. For example, the difference in size between white leghorn chickens and its wild relative, the red jungle fowl, could be explained by only four QTLs (Kerje, 2003). However, the effect of QTL on traits and epistatic gene interaction are biased due to the small number of individual analysed. Microsatellite markers, consisting of small sequence repeats of variable number along the genome, have also been used for association with traits in cattle such as the identification of QTL controlling milk production in dairy cattle (Georges, 1995). This method is now routinely used for the identification of QTLs in recorded traits through array chips, due to the increase of genetic markers identified (Berry, 2011). Another approach to identify QTL is by

investigating candidate genes, which have known functions and biological processes (Glazier, 2002).

The investigation of QTL in cattle, human and other animals produced a significant amount of data, which led to the creation of a database: Animal QTLdb (www.animalgenome.org/QTLdb/, 2016). The role of this database is to gather information from published work for livestock with the aim of facilitating comparative genome studies and the discovery of polygenic traits (Hu, 2007). Since the first release in 2006, the size of the database increased reaching 57,414 QTLs for six animal species: cattle, chicken, horse, pig, sheep and rainbow trout (Hu, 2016). The cattle database contains, at the last update, 42,019 QTLs from 646 publications, representing 482 different traits.

The analysis of QTL in livestock is an important topic to understand phenotypic variation in breeds and led to the discovery of causal genes affecting specific traits (Andersson, 2004). The double-muscle phenotype in cattle has been linked to the myostatin gene using genomic mapping in the muscular hypertrophy locus on chromosome 2 revealing a deletion of 11bp creating a premature stop codon (Grobet, 1997). The identification of a QTL in Holstein cattle, located on the chromosome six and linked to fat and protein concentration in milk, resulted in the localisation of a polymorphism in *ABCG2* gene (Cohen-Zinder, 2005). This candidate gene has been further studied and is of great importance in milk quality by influencing the transfer of xenobiotic such as enterolactone, riboflavin, flavonoids and antibiotics in milk (Wassermann, 2013; Otero, 2015).

Therefore, the identification of loci and their link to specific traits in cattle helped to improve selection by discovering causal polymorphisms. Furthermore the identification of specific polymorphisms related to a phenotype led to the notion of Quantitative Trait Nucleotide (QTN) (Ron, 2007).

1.2.3 Single Nucleotide Polymorphisms (SNPs) improvement of the genomic selection

The sequencing of cattle genome started as early as 2003 with the bovine genome sequencing project, following sequencing of human genome in 2001 (Venter, 2001). The bovine genome project was led by the Human Genome Sequencing Centre using BAC and whole genome sequencing of a Hereford cow and the first assemblies

(Btau3.1 and Btau4.0) have been published as part of the Bovine HapMap consortium (Consortium, 2009; Liu, 2009).

The Bovine HapMap consortium aimed to identify variation in cattle, using 497 animals from 19 breeds across the world. It helped to decipher the structure of cattle breeds from which they identified 50 necessary SNPs for parentage identification. At the same time the results from the consortium have been used to create a high density SNP assay for cattle (Matukumalli, 2009). This study led to the creation of the Illumina BovineSNP50 chip containing at the time 58,336 SNPs. The high density of SNPs present has resulted in a mean distance of 37kb between SNPs, which was below the estimated minimum (100kb) thus improving the amount of haplotype block under linkage disequilibrium identified, improving the accuracy of genome wide association studies.

Although the high density chip allowed the identification of many markers useful for cattle selection and association studies, it was hindered by the high cost of this platform. For that reason, algorithms have been developed to impute unknown animal genotypes from lower density genotypes based on high density genotype population and linkage disequilibrium (Druet, 2010; Daetwyler, 2011). This led to the creation of the lower density chip such as the Illumina BovineLD (v1.0: 6,909 SNPs), improving the previous GoldenGate Bovine3K (Boichard, 2012).

Following these first steps, other bovine genome assemblies such as UMD (Zimin, 2009) gave a new insight into the identification of genetic variation through SNPs. The last reference assemblies, UMD3.1 and Btau4.6, also allowed the assessment of the potential and accuracy of the different variation previously identified (Y. Hou, 2012). Moreover, the development of next generation sequencing (NGS) technology increased the potential number of animals sequenced, therefore the number of variants detectable. This increase in the identification of markers also increased QTLs identification, for example the identification of three major carcass weight QTLs in Japanese Black Cattle, which accounted for the third of the genetic variance, used an association study method based on the BovineSNP50 BeadChip (Nishimura, 2012). Another study, which was part of the 1000 bull genome project, sequenced 234 bulls with the aim to improve association studies and genomic prediction. They identified two QTLs associated with early-lactation fat content in Fleckvieh and Holstein-Friesian near two genes with strong links to this traits (Daetwyler, 2014).

These polymorphisms associated with linkage disequilibrium can predict variation in important traits and consequently the breeding value of candidates (Meuwissen, 2010). SNPs have become the predominant markers used for selection due to their relative low cost on a high number of individual and their higher accuracy for complex traits. By building a reference population from animals with measured traits and SNP based genotyping, an equation can be estimated to predict the breeding value of the animals. This estimation can be represented by model such as the Best Linear Unbiased Prediction (BLUP) for SNP effect normally distributed with small effect or BayesA assuming a small probability of moderate to large effect (Meuwissen, 2001). Then this equations can be used to predict breeding values of animals only based on their genotype (Hayes, 2013).

1.2.4 SNP discovery in cattle

The identification of polymorphisms in cattle represents an important challenge for understanding variations amongst breeds and improving their selection. This purpose has been facilitated with the use of SNP chips (**Figure 5**). The methodology itself is based on microarray technique. It uses DNA probe hybridisation by nucleotide complementarities. The probes are designed to identify specific SNPs. The first commercially available chip was produced by Affimetrix, which was developed for human genome applications and contained 1494 SNPs.



Figure 5 – Example of SNP chip from Illumina

The improvement of SNPs array techniques and developments in sequencing of whole genome in cattle allowed the use of these methods for bovine applications. Now many companies provide genotyping platforms for cattle:

- Illumina has three types of whole genotyping chips namely BovineHD (high definition), BovineSNP50 and BovineLD (Low definition) ('www.illumina.com', 2016). The BovineHD is a high definition chip containing 777,000 SNPs from beef and dairy breeds and spanning through the cattle genome in a uniform way. The BovineLD (v2.0) contains 7,931

SNPs, which have been selected for improved imputation to estimate breeding value, mainly focused on dairy breeds (Boichard, 2012). The BovineSNP50 (v3.0) contain 53,714 highly informative SNPs identified in the major dairy and beef breed (Matukumalli, 2009).

- Affymetrix provides two types of arrays for cattle, the Axiom® Genome-Wide BOS 1 Bovine Array and the Eureka™ Bovine Parentage Panel. The Axiom is a HD chip containing 640,000 markers from 20 different dairy and beef breeds (including zebu) aimed for performance trait identification and research (Rincon, 2011). Eureka is a much smaller chip of 122 SNPs designed for parentage identification ('www.affymetrix.com', 2016).
- Neogen is providing a range of SNPs related to performance traits through their GeneSeek Genomica™ (GGP) arrays commercialised by Illumina on their Infinium custom arrays. The GGP Chips are designed in both low definition, containing 42,000 markers and high definition, containing 150,000 markers. They offer chips designed specifically for dairy or beef and also on specific HD chip for the Angus breed (www.genomics.neogen.com, 2016).

1.2.5 The International Dairy and Beef SNP Chip

The advance in genomics, as mentioned previously and the requirement of deeper genetic information for parentage verification, the lack of a platform containing genetic tests for genetic malformations and disease and the high cost and lack of accuracy of the commercially available SNP chips, led to the development in 2012 by Teagasc and ICBF of a custom SNP chip for use by breed societies, DAFM, Irish farmers and researchers. The ultimate aim of developing this chip was to improve the genetic gain of the Irish beef and dairy cattle populations. The development of the custom SNP-chip, called International Dairy and Beef (IDB), aimed to reduce the cost of genotyping cattle for parentage verification by improving an Illumina low density chip (6,909 SNPs), designed to be able to impute the information from microsatellite (MS) genotyping, high density chip genotyping and facilitate the identification of congenital diseases (Mullen, 2013).

Until recently, MS were the only means of verifying the parentage of individuals, but the cheaper SNP genotyping, currently available on the IDB chip has tended to replace it. However, the familial link between animals can only be established if they are genotyped with the same process. To promote the transfer from MS to SNP genotyping, the IDB contributed 2,176 SNPs dedicated for the imputation of parentage from MS (McClure, 2012). Furthermore 5,500 SNPs were included to impute from high-density genotyping SNP-chip (HD) and 424 were included to identify diseases or major production genes variations such as myostatin or β casein. Finally, 1,873 SNPs were included for research purposes, containing some of the SNPs identified in chapter 3. A second version of the IDB was released in 2014 upon which extra novel SNPs have been included while also containing some of the SNPs identified in chapter 4. More recently, a third version of the chip, containing the previous SNPs, has become available and contains over 50,000 SNPs.

1.2.6 Association studies

With the multiple assemblies of cattle genome and the development of genotyping, the information to identify QTL has increased and has led to a refinement in the confidence interval towards Quantitative Traits Nucleotide (QTN) through Genome-wide association studies (GWAS). The mapping by linkage of causal genes has been widely used so far but shown to be weaker when multiple loci were involved. GWAS is a method which uses genome wide sequencing, to identify polymorphisms associated to

phenotypes in a population. This method allows the identification of markers which are more often used in marker-assisted selection (MAS). GWAS is based on population-wide LD, due to mutation, drift or selection, between genotyped variants in different loci as well as allele frequencies (Visscher, 2012). The GWAS methodology gained in popularity in the identification of QTL, due to a higher SNPs density available and cost reduction of the method (Berry, 2011).

Many studies have used GWAS to identify genes or region involved in the variation of traits and has a considerable impact on cattle breeding. A study on two dairy breeds, Holstein and Jersey, associated up to 461 SNPs in a GWAS with milk production traits, validated on 386 Holstein and 364 Jersey using the BovineSNP50 BeadChip (Pryce, 2010). This dairy breeds study allowed the identification of a haplotype, linked to a fertility QTL identified in both breeds in the chromosome 18. The results showed associations with milk protein and fat percentage for the candidate genes *DGAT1* and *ABCG2* variants, confirming causal mutations already identified in these genes. Another study used the same SNP chip on 313 Korean cattle breed and identified 5 SNPs associated with marbling score, back thickness and muscle eye area. The last trait is associated with a SNP located in *DVLI*, related to muscle development (Kim, 2011). An association analysis has been conducted for polymorphisms in imprinted genes in 848 Holstein-Friesian based on candidate genes for the evaluation of 17 SNPs. This study revealed that nine SNPs showed significant association with a wide range of traits from fertility to carcass weight. Most specifically two genes have been noted: *PEG3*, a paternally expressed gene, associated with calving and calf performance and *ZNF215*, a zinc finger protein associated with milk protein percentage, progeny carcass weight, culled cow carcass weight, angularity, body depth, rump width and animal stature (David A Magee, 2010). Two studies identified bovine tuberculosis-related QTLs. The first study identified three significant SNPs located in the chromosome 22 in the intron and upstream of the gene coding for the taurine transporter TauT and associated to susceptibility to tuberculosis (Finlay, 2012). The second more recent study identified a locus in Holstein-Friesian in chromosome 13, presenting a SNP block in strong LD and one SNP in chromosome 2, related to resistance to tuberculosis. SNPs in the locus of the chromosome 13 were linked to *PTPRT*, a gene coding for a receptor protein tyrosine phosphate. Furthermore, one SNP was present in

the intron of *MYO3B*, hence suggesting a polygenically regulated trait (Bermingham, 2014).

The use of GWAS in cattle has assisted in understanding the association between SNPs and economic traits of interest and in turn gene function in beef and dairy breeds. This new technique represents valuable data allowing improvement on parentage identification, selection and estimated breeding value in comparison with the previous micro-satellite method.

1.2.7 Genomic Selection

Genomic selection refers to selection decisions that are based on breeding values predicted using genome wide marker data such as SNPs (Meuwissen, 2001). The approach aims to increase the selection accuracy and accelerating genetic improvement by focusing on the SNPs the most strongly correlated to a given phenotype, although the genes and sequence variants directly affecting the phenotype remain largely unknown (Snelling, 2013).

Unlike phenotype based genetic evaluation, genomic prediction has the capacity to predict genetic merit of selection candidates at birth before phenotypes become available, which offers great promise in the prediction of genetic potential of selection candidates for traits that are difficult and expensive to measure such as feed efficiency and fertility. The success of genomic selection largely depends on the accuracy of the predicted genomic breeding values. However, genomic prediction accuracy in beef cattle is still not sufficiently high to allow selection of candidates without appropriate phenotypic measurement (Bolormaa, 2013).

Overall there are still challenge to overcome for genomic selection such as improving the accuracy of the genomic breeding value, integrating genetic information in genetic evaluation at national level and across countries as well as managing the genetic gain on a long term basis (Hayes, 2009). Inbreeding should also be considered in genomic selection as it can reduce the fitness of certain traits when improving a selected one. This fact is illustrated by the selection of milk traits in Holstein which led to a reduction of fertility associated with inbreeding (Mc Parland, 2007, 2009). However, the genome-wide selection method, assuming a high density of marker, has been shown to potentially reduce the inbreeding rate in comparison with traditional method like BLUP (Daetwyler, 2007).

1.3 Cattle breeding in Ireland

1.3.1 Dairy and beef production are an important part of Irish industry

The cattle industry is of high importance for the Irish economy. Beef exports represented €2.41 billion in 2015 consisting of 88% of the production (564,000 tonne produced) while dairy exports represented €3.24 billion, with 6,982 million litres produced (www.bordbia.ie, 2015). Cattle production has been steadily increasing since 1987 reaching a plateau around 1997 consisting of 110,000 cattle farms in the 2010 June census (Renwick, 2013) and the Animal Identification and Movement (AIM) counted at total of 6.13 million animals in 2014. This 2014 census reported seven main breeds in Ireland: Holstein-Friesian (FR), Limousin (LI), Charolais (CH), Aberdeen-Angus (AA), Hereford (HE), Simmental (SI) and Belgian-Blue (BB), with Holstein-Friesian accounting for almost 40% of the animals (**Figure 6**).

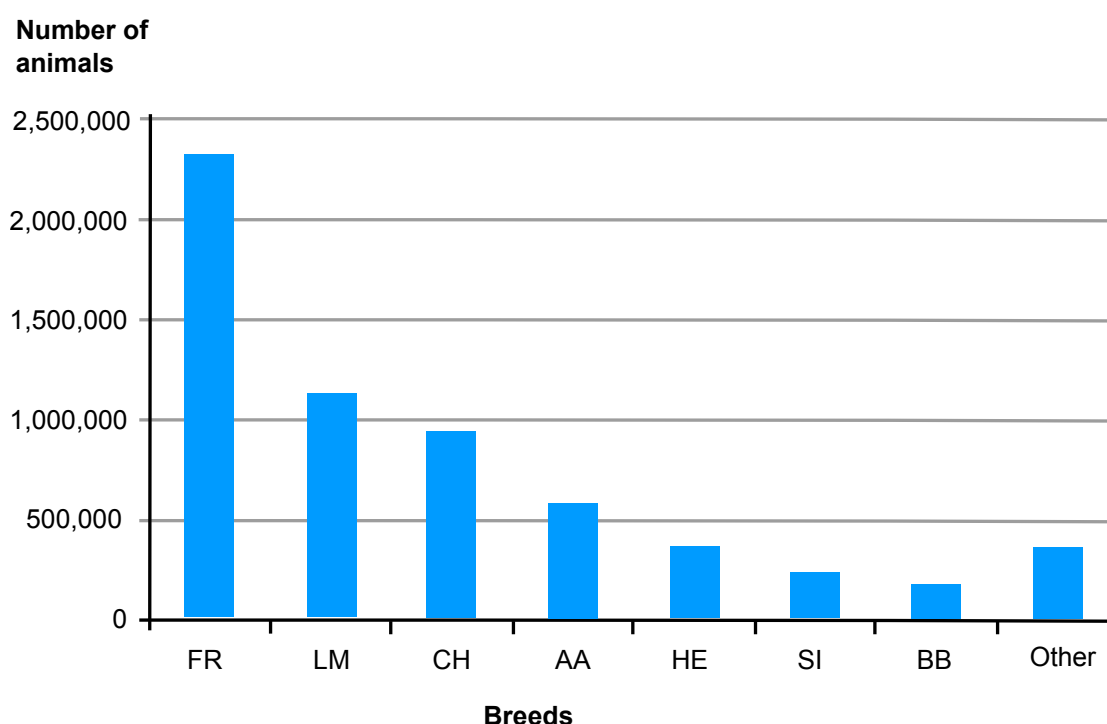


Figure 6 – Distribution of the main breeds in Ireland (Animal Identification and Movement (AIM) statistic report 2014). Total number of animals per breed. FR: Holstein-Friesian; LM: Limousin; CH: Charolais; AA: Aberdeen-Angus; HE: Hereford; SI: Simmental; BB: Belgian Blue

1.3.2 Important cattle breeds

In total, 990 cattle breeds are currently recognised worldwide, nevertheless, Food and Agriculture Organisation (FAO) reports on cattle diversity showed that many of these are endangered (Rischkowsky, 2007). However, high genomic diversity still exists and represents an important resource for preservation, and for potential future breeding programs.

The development of cattle breeds rose in the 19th century, giving birth to most part of the current modern breeds (Felius, 2011). The need to increase the production capacity in European cattle herds led to a diminution of the cattle breeds diversity used today, due to the selection of a small population of elite progenitors through artificial insemination (Georges, 1996). The main breeds used worldwide are of European origin with Holstein the breed with the widest distribution, currently present in 128 countries (Rischkowsky, 2007). In this thesis data from six main modern domesticated breeds have been used, namely:

1.3.2.1 Holstein-Friesian

Often abbreviated as Holstein, this dairy breed is the most widely distributed in the world. The statistics from the Irish Department of Agriculture, Food and the Marine for the year 2014 counted 2,320,533 heads representing 37.8% of the entire Irish cattle herds. The Holstein, breeds originated from the Netherland while the first Friesian in Ireland came from England and Scotland after the second world war. When the Holstein came from Canada in the mid 70's, this led to the actual Holstein-Friesian breed as present in current dairy herds (Dillon, 2008). This breed is considered to be the highest milk producer worldwide, having a black and white or red and white coat and is known for its easy handling due to its gentle behaviour. It has however been shown to be poorly resistant to high temperature and diseases (De Vaccaro, 1990; Kadzere, 2002). Furthermore, selection based on milk related traits in Holstein breeding programmes over the past few decades has led to an increase in inbreeding and a decrease in fertility (Royal, 2000; Mc Parland, 2007).

1.3.2.2 Jersey

The jersey breeds, originated from Jersey Island, UK, is the second widest distributed dairy breed in the world according to the FAO report. It is of smaller stature than the Holstein breed and is an ancestral breed, known for its high yield and richness of milk

as well as its grazing efficiency (White, 2001; Prendiville, 2010). Jersey is a robust breed easily adapted in different climates (Arrillaga, 1952) and can have a wide range of coat colour. The temperament of the cow is described as rather docile when the bull is considered as the least docile amongst the common cattle breeds (Oklahoma State University, 1997b). A study on the genetic structure of the Jersey breed in its original island population has found that even if no importation of other breeds has occurred, the estimated heterozygosity is high and comparable with other continental breeds but displaying a small amount of genetic diversity within the population (Chikhi, 2004). Similar results for heterozygosity in Jersey and the similarity with other common breeds has been previously shown in another study (Loftus, 1999).

1.3.2.3 Aberdeen-Angus:

Aberdeen-Angus, shortened as Angus, is a robust Scottish beef breed cattle formerly used as draught animals and characterised by being polled (hornless). This breed arises in Aberdeenshire, from local Scottish polled animals around the 19th century and is widely exported (Oklahoma State University, 1997a). The Angus is known for the high quality of its beef, with a distinctive fat marbling (Chambaz, 2003).

1.3.2.4 Limousin

The Limousin is a beef breed originated from the Limousin region in France. The breed has been widely exported since the 1960's and is found in 70 countries. The Limousin breed is known for its calving ease traits (Comerford, 1987) and its heavier carcass and meat quality (Chambaz, 2003).

1.3.2.5 Romagnola

The Romagnola is a grey breed, primarily used for draught but subsequently became a beef breed. The meat is considered to be of high quality. The inbreeding of the Romagnola due to its restricted population led to the apparition of genetic diseases such as cataract, lethal developmental dysplasia or pseudomyotonia which are actively investigated in gene tests (Murgiano, 2012, 2014; Testoni, 2012).

1.3.2.6 Fleckvieh

The Fleckvieh breed is a central European breed used for both milk and beef production (i.e. dual purpose) and originated from Austria. They are known for fast bull growth,

easy calving, udder health and resistance to different climates (*Fleckvieh Society of Australia*, 2011).

1.3.3 Irish programs for breeding improvement

To increase the production of dairy and beef cattle, different programs have been implemented in association with governmental bodies and farmers. The Irish Cattle Breeding Federation (ICBF), financed by farmers and the Department of Agriculture Food and Marine (DAFM), was established in 2000 with the objective to improve dairy and beef cattle production. Since then, ICBF has gathered valuable information about genetic markers, traits and bull performance from different breeds of cattle throughout Ireland. In addition, economic indexes have been developed in association with Teagasc such as the *Maternal Index*, *Terminal Index* and *Dairy Beef Index*, which assist the farmers in selecting better animals for breeding. This information has been collected and collated to create the Economic Breed Index (EBI) (Veerkamp, 2002). The EBI is based on the relative breeding index (RBI), which already consists of component data related to milk fat, protein yield and protein content but was lacking information about calving interval and further survival. These two last components, also related to milk production performance, have been incorporated to the EBI (Pryce, 2001). These parameters were linked to the economic value with the aim of tackling reduced fertility rates which was observed with the increase of milk yield in dairy cows (Royal, 2002). The GEN€ IRELAND program is focused on dairy cattle since 2005 and 2007 for beef with a rise in the latter since 2014, concentrating on testing young bulls for their progeny to increase their reproductive value. In 2010, the rise of genomic information and tools for cattle led to the creation of a new database in Ireland called IGenoP standing for International Genomic Partnership. The role of this database is to share Illumina SNP genotyping results obtained from different bi-lateral programs between ICBF and other countries. These programs, now interconnected and profiting from the findings of each other, have improved the genetic merit of dairy cattle in Ireland and are currently focusing now on beef breed improvement (Cromie, 2014).

1.3.3.1 Dairy breeding programmes

The program for cattle genetic improvement was first developed for dairy breeds, focusing mainly on Holstein-Friesian but also incorporating other dairy breed such as Jersey in an across-breed scheme. The first task was to collect information about cattle

traits in herds on a large scale, which led to the establishment of the ICBF database in 2002. Different traits related to dairy cattle have been introduced to the database since its foundation, to improve the Economic Breeding Index (EBI) for farmers. These traits firstly consist of production data such as milk yield were expanded over the years to include fertility and calving traits and also health traits such as somatic cell count, which represents a major indicator of infection in dairy breeds. It also included ancestry and genetic information. All this information facilitated improvement of the EBI index over the years increasing the value and genetic gain of the Irish herds (**Figure 7**).

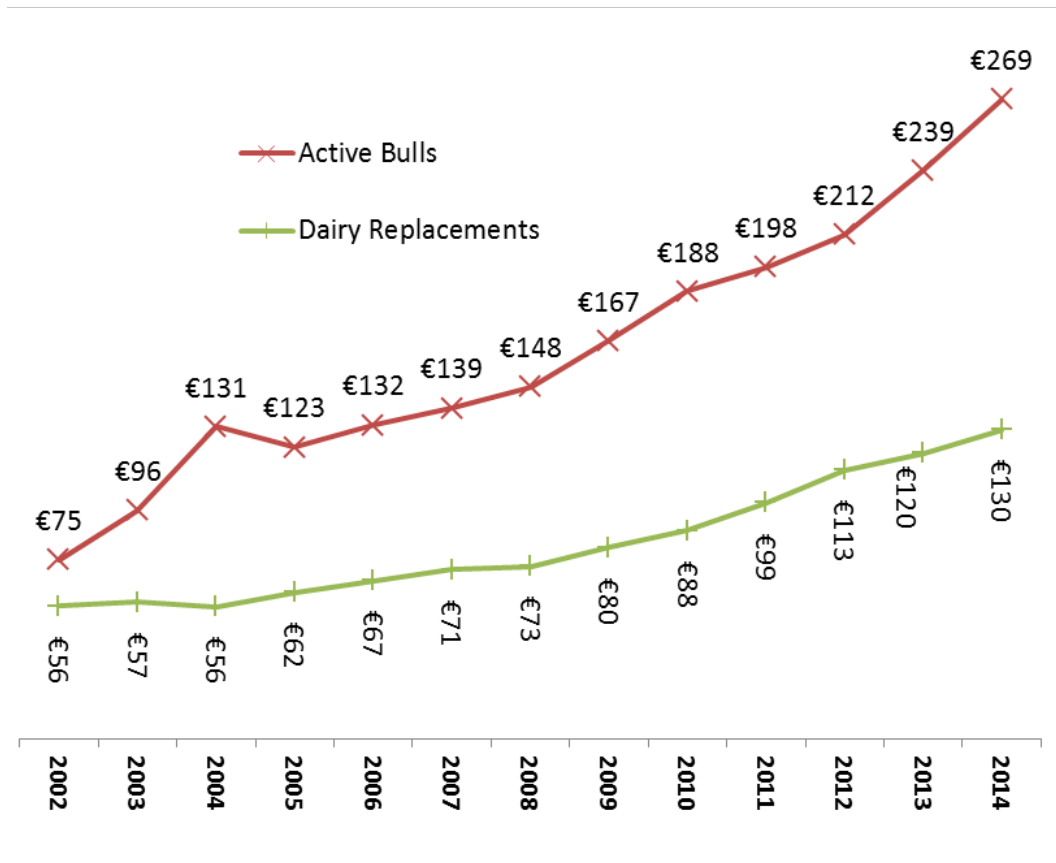


Figure 7 – EBI averages by birth year for females and for bulls on active bull list in each year (ICBF annual report 2014). The Economic Breeding Index (EBI) represents a metric primarily based on dairy cattle and regrouping data related to milk production. The graph shows the increase in value of the dairy cattle due to successful genomic selection programmes.

The rise of genomic analysis brought a new dimension to dairy cattle improvement. The information gathered allowed the improvement of bull selection for artificial insemination hence increasing the EBI, therefore increasing the productivity of the dairy offspring and the profitability for the farmers. In 2014, 59% of dairy inseminations were from genomically selected bulls, which directly correlated with an improvement in EBI. Furthermore, the genotyping of young heifers became a routine procedure helping the farmers to identify the best animal and at the same time, improving the genetic and genomic tools implemented.

1.3.3.2 Beef breeding programmes

The first step for genetic improvement for beef cattle was, as for the dairy cattle, to record the traits present in the herd and integrate them into a database. This was achieved in 2004 in the ICBF database and is still increasing at the present time. Traits specific to beef such as carcass weight but also fertility, calving, weaning and maternal weaning weight have been investigated following the same procedure as for dairy cattle. Specific services have been developed for beef breeders such as GROW[®] or Genetic Recording of Weanlings in 2002, to develop an across breeds score evaluating different weight traits for pedigree and non-pedigree animals, assisting the farmers to improve beef selection.

Interbeef is a program in development with the aim of gathering international information for evaluation of beef breeds and related traits. Furthermore, the decrease of the *Maternal Index* in beef is a major challenge for Irish farmers and led to a focus on maternal traits for genomic analysis leading to the launch of the Beef Data Genomic Program (BDGP) by DAFM in 2014. The BDGP is a six-year programme with €300M in funds available. The Maternal index, which represents the level of fertility, has been shown to decrease with the increase of beef trait performance. Such negative effects of selection have already been identified in dairy cows where the selection for high performance milk production traits has been linked to a decrease in fertility (Royal, 2002).

The aim of the BDGP is to improve beef herd management and genetic merit as well as a reduction of the amount of greenhouse gas produced. This will be achieved by collecting phenotypic and genetic data from suckler cow to select better animals and improve the breeding index.

Interbull is an organisation part of the International Committee for Animal Recording (ICAR) regrouping 17 countries around the world including Ireland. The Interbull centre is located at the Department of Animal Breeding and Genetics (HGEN) of the Swedish University of Agricultural Sciences. The role of Interbull is to collect data from different countries to improve the genetic merit of cattle through different tools. The Multiple Across Country Evaluation (MACE) is a tool gathering national evaluation data of different cattle breeds across countries and rank them according to their traits. Genomic Multiple Across Country Evaluation (GMACE) introduces the genomic information for across country evaluation of estimated breeding value (EBV)

increasing the efficiency of the evaluation. Intergenomics regroup the genomic evaluation from the different countries and is deregressing each country's GEBV to create multi-country results.

All these organisations aimed to increase the use of genomic evaluation of beef and dairy breeds to continue improving the economic value of cattle herds in Ireland.

1.4 The role of micro-RNAs and their importance in livestock

1.4.1 Discovery of microRNAs

The first mention of miRNA followed the observation of differences of expression of LIN-14 protein effect on cell lineage patterns during developmental stage of *C. elegans* (Ambros, 1984, 1987). In these studies, Ambros and Horvitz. observed that an allele of the gene *lin-14* caused a “gain of function” of *lin-14* leading to the developmental L1 larval stage phenotype and another allele caused the “loss of function” of *lin-14* leading to the developmental L2 larval stage phenotype.

These alleles leading to the gain and loss of function were linked respectively to the high presence of the LIN-14 protein in L1 stage cell and to an almost undetectable level in L2 stage (Ruvkun, 1989). Furthermore, a null mutation in *lin-4* gene led to the same “gain of function” and L1 stage phenotype (Chalfie, 1981) which suggested that *lin-4* gene is a regulator of *lin-14* (Ambros, 1989). The mutations in *lin-14*, leading to a gain and loss of function phenotypes, have been identified in the 3' UTR of the *lin-14* gene as insertions, deletions, and also point mutations (Wightman, 1991).

These findings motivated the investigation of the mechanism of *lin-14* regulation by *lin-4*. Two studies by Lee and Wightman, conducted the analysis of the molecular mechanism modifying the expression of *lin-14*. They found that the change of expression was due to post-transcriptional regulation. They showed that *lin-4* gene codes for two different untranslated RNAs of 22 and 61 nucleotides and presents complementary sequence with seven patterns in *lin-14* 3' UTR. They showed that the RNA pairing between *lin-14* and *lin-4* was sufficient for expressing the “loss” phenotype (Lee, 1993; Wightman, 1993).

Although the term miRNA was not mentioned in these studies, the characteristics and secondary structure of this small RNA were already described. In 2000 a second small RNA, *let-7*, showed an interaction with the 3' UTR of five genes including *lin-14*, modifying their expression and the regulation of developmental stage

of *C. elegans* (Reinhart, 2000). The first mention of the term micro-RNA was in a report on small interfering RNA in 2001 which regrouped miRNA identification from human and drosophila (Lagos-Quintana, 2001). The research following the first description of miRNAs in 1993, led to the identification of these small RNAs in many plants and animals and deciphered their molecular mechanisms (Hutvágner, 2001; Llave, 2002; Carrington, 2003).

1.4.2 MicroRNA structure

Further studies on miRNAs confirmed that they are small regulatory RNAs, ranging, in their active form (mature miRNA), between 16 to 26 nucleotides in length in cattle and are processed from a longer hairpin-loop structure called pre-miRNA. The pre-miRNA is processed from the primary transcript called pri-miRNA.

The nomenclature of miRNA, apart from the ones identified earlier such as *lin-4* or *let-7*, is composed of the prefix “mir” plus a number, for example *mir-206*, as described in the miRNA database miRBase (Griffiths-Jones, 2006). To distinguish between species, a three or four letter code can be added to specify to which species the miRNA is related, for example, cattle miRNA have the prefix ‘bta’ such as in *bta-mir-1*. The number of the miRNA between species is usually consistent with orthology between these miRNA, such as hsa-mir-101 (human) and mmu-mir-101 (mouse) are orthologues. When the miRNA is referenced as its shorter active form, it is written with a capital “R” and is not italicised such as miR-2419. The mature miRNAs, which are paralogous, with one or two nucleotides differences have an extra letter added such as miR-29a to miR-29e (five paralogs in cattle) and identical mature miRNAs which originated from different loci have a numeric suffix added such as miR-6526-1, miR-6526-2 and miR-6526-3.

Finally, the hairpin-loop structure of the pre-miRNA contains the mature miRNA in its stem where potentially two complementary active miRNAs are present. The mature miRNA from the 5’ end of the pre-miRNA is identified as 5p and the one from the 3’ end as 3p (**Figure 8**). The mature miRNA main active domain is the 6 nucleotides from position 2 to position 7 creating the binding region to the 3' UTRs and called seed region.

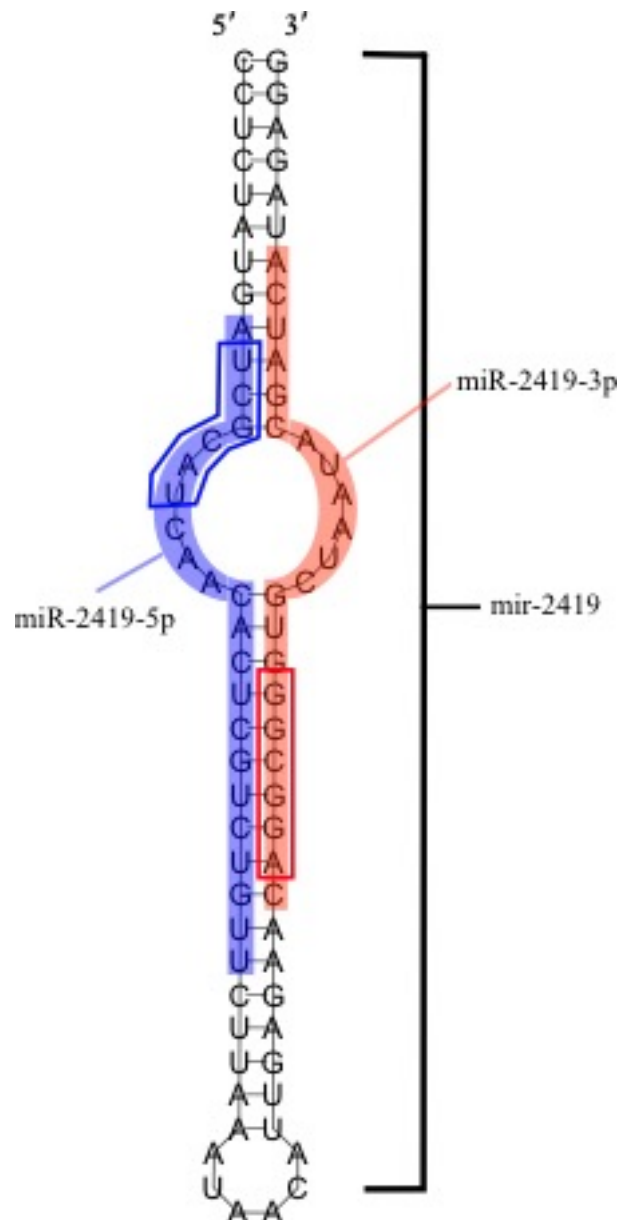


Figure 8 – Representation of the miRNA nomenclature with *bta-mir-2419* example. This representation of the product of *bta-mir-2419* gene shows the different nomenclature of the transcription products. The hairpin-loop is the pre-miRNA *mir-2419*, which will be cleaved to give the two mature miRNA *miR-2419-5p* from the 5' end (blue) and *miR-2419-3p* from the 3' end (red). The darker outline represents the main binding region (the seed region) in each of the mature miRNA at the 5' end of *miR-2419-5p* (dark blue line) and *miR-2419-3p* (dark red line). (Image: Martin Braud)

1.4.3 MicroRNA biogenesis

The miRNA genes are transcribed in animals by the polymerase Pol-II. The product obtained is the pri-miRNA, which can be composed of a cluster of miRNA hairpin loops of several kilobases long. This structure is then cleaved to release a single hairpin loop, the pre-miRNA. This function is processed by a nuclear RNase III, called Drosha, in association with the co-factor DiGeorge syndrome Critical Region gene 8 (DGCR8).

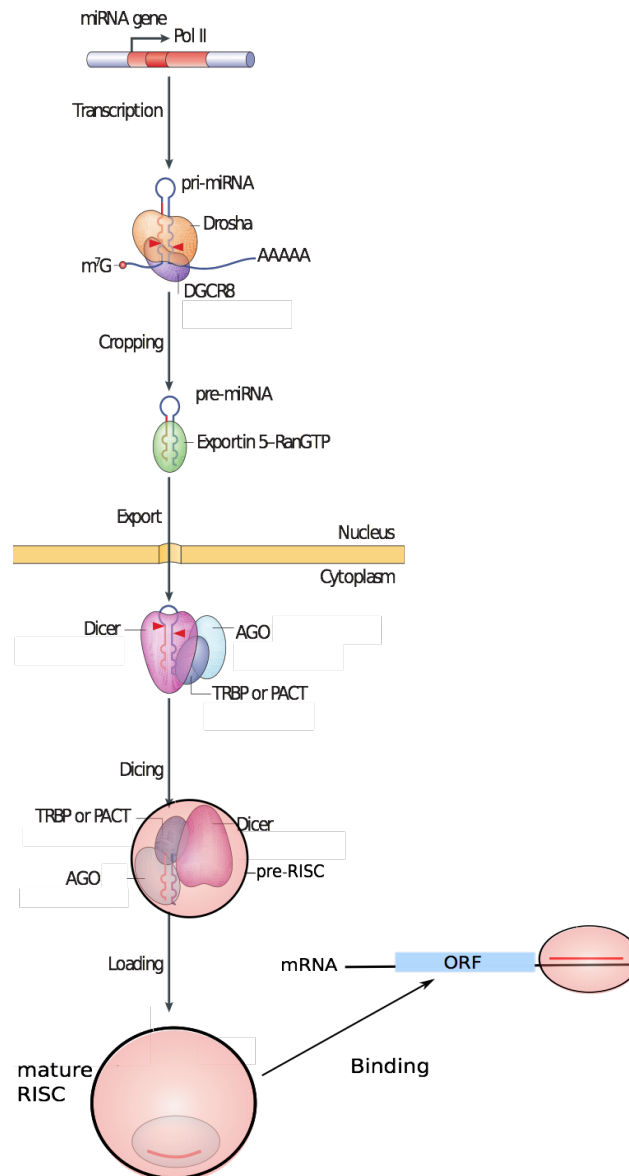


Figure 9 – Biogenesis of miRNA from the pri-miRNA transcription to the binding of the mature RNA to the 3' UTR (Figure adapted from Kim et al. (2009))

The pre-miRNA is then transported from the nucleus to the cytoplasm with the aid of exportin 5, which is mainly dedicated to this function. In the cytoplasm, the pre-miRNA is loaded in Dicer with two proteins TRBP and PACT which help to remove the loop structure, leaving the duplex of miRNA-3p and miRNA-5p and starting the formation of the RISC complex. The duplex is then loaded in an Argonaute protein (AGO) and one of the strands is eliminated forming the mature RISC complex. In most cases, only one specific strand of the miRNA duplex is selected but in some cases, the strand selected can be alternatively the 3p or the 5p strand (Yang, 2011). In human, only AGO2 seems to have the slicer activity, removing one of the strand but AGO1, 3 and 4 seems to be able to load miRNAs (**Figure 9**). The AGO protein helps the mature miRNA to bind to the 3' UTR of the target messenger RNA, by pairing to exact complementary sequence to the seed region. However, some part of the 3' end of the mature miRNA can also bind to the 3' UTR. The binding of the miRNA leads to the repression or the degradation of the messenger RNA (Kim, 2009).

1.4.4 Role of miRNA in livestock economically important traits

Many studies in livestock studied the role of different miRNAs and their expression for important economic traits. In Japanese black cattle breed, *miR-885* and *miR-196a* were found to be differentially expressed in slow-type and fast-type muscle (Muroya, 2013). It also has been shown in Piedmontese cattle breed that double-muscle phenotype is correlated with a higher expression of *mir-206* in comparison with Friesian breeds (Miretti, 2011). A study on chicken also showed that an overrepresentation of the mature miR-206 and miR-10b is associated with skeletal muscle development (T. Li, 2011). In Chinese cattle breed Quinchuan, highly expressed miRNAs (miR-1, -133a, -206, and -378), targeting genes involved in muscle development, have been identified in longissimus dorsi muscle. In the same study, some highly expressed miRNAs (miR-199a-3p, -154c, -320a and -432) related to back fat tissues have also been identified (Sun, 2014).

Moreover, other studies showed an association between mir-378 with back fat thickness (Jin, 2010; Morris, 2010). It also been shown that increase of miR-143 in *in-vitro* Holstein muscle culture enhanced the maturation of fibroblast-like adipocytes from intramuscular preadipocytes (H. Li, 2011). In Simmental beef, a study on subcutaneous and intra-muscular fat tissues showed an accumulation of miRNA expression and highlights four miRNAs (miR-143, miR-145, miR-2325c and miR-

2361) involved in different pathways and targeting genes which may regulate fat deposition (Wang, 2015).

Mammary gland and milk production traits are also of high impact in dairy production. The importance of 359 miRNAs as biomarkers has been cited for such traits in database gathering potential biomarkers for milk production and mastitis (Ogorevc, 2009). During different lactating periods, it has been shown that miRNA expression varies; 12 miRNAs are over expressed during the late lactating period and one during early lactation (Wang, 2012). MiRNAs have been shown to be present in micro-vesicle in milk, and suggested to have an influence on digestive metabolism and health in calves (Hata, 2010). Other studies shown that miRNAs are differentially expressed during different lactation stages and targets of these miRNAs suggested a role in immune system (Izumi, 2012; Y. Gu, 2012).

Mastitis is a major infection in dairy cattle leading to a reduction in milk production. Many studies have been performed on this infection in which miRNAs are suggested to be involved. Five miRNAs, targeting a leukocyte antigen gene (BOLA-DQA2) have been reported as differentially expressed between healthy and infected mammary glands (Hou *et al.*, 2012). Other studies reported differences in the expression of various miRNAs during udder infection confirming the role of miRNA in the response to mastitis (L. Li, 2012; Naeem, 2012). A more recent study on miRNA expression in peripheral blood vessels in healthy and mastitic Holstein cows showed a significant differences in the expression of 173 miRNAs and the most differentially expressed were related to immune pathways (Li, 2014). Apart from mastitis, miRNAs have also been identified to play a role in various infections. In the case of bovine tuberculosis, *mir-155* has been identified to be over-expressed at the early stage of infection in non-vaccinated cows when the vaccinated cows presented this over-expression only at a later stage after infection and could be used as an infection marker (Golby, 2014). It also has been shown that miRNAs play a role in heat-stressed Holsteins. The comparison of expression of these miRNAs showed a difference between heat-stressed and control animals and seems to be related to stress response and the immune system (Zheng, 2014).

MiRNAs have also been shown to be involved in fertility traits by regulating genes in oocytes, sperm and gonadal tissues (Hossain, 2012). Studies on expression level of miRNA in cattle ovaries showed that *mir-106* was more expressed in oocytes

than in cumulus-oocytes complex and granulosa cells (Miles, 2012) corresponding to the expression of the predicted target genes of miR-106 which was decreased in oocytes and not in cumulus-oocytes complex. The importance of miRNA in oocytes has been suggested as various miRNAs showed differential expression at different maturation stage of oocytes (Tesfaye, 2009; Tripurani, 2010; Abd El Naby, 2013). A study on Holstein bulls analysing low and high fertility animals, showed that seven miRNAs were differentially expressed in spermatozoa (Govindaraju, 2012). MiRNAs are also important in embryo development. Indeed, a miRNA profiling study on embryo early stage development showed that miR-21 and miR-130a expression increased from one-cell to eight-cell stage (Mondou, 2012). A study on placental development of cloned cattle showed a differential expression of numerous miRNAs, which have been linked to abnormal placental development (Liu, 2014). The miRNA family *mir-34* is composed of three members (a, b and c) and present in ovaries, testis, sperm oocyte and embryo of cattle, in different ratios. The study on *mir-34* members showed that *mir-34c* seemed to be the best candidate as a bull fertility biomarker (Tschermer, 2014).

1.4.5 Polymorphisms in miRNA

As mentioned previously, the discovery of miRNA has been directly linked to polymorphisms. Indeed, the first miRNA *lin-4* and its target *lin-14*, have been identified as such due to mutation in *lin-4* and in the 3' UTR of *lin-14*. Although the major characteristics of miRNAs for their annotation is their high level of conservation of the mature sequence across many species (Wang, 2005), polymorphisms in miRNA genes and 3' UTR target sites are not absent.

In Human, the screening of miRNA polymorphisms have been widely investigated even if it has been suggested to be rare (Saunders, 2007). The polymorphism related to miRNA genes and the targeted 3' UTRs have been shown to be involved in many diseases in Human (Sethupathy, 2008). A polymorphism in the 3' UTR of Slit and Trk-like 1 (*SLITRK1*) have been identified to be significantly associated with Tourette's syndrome and corresponded to a target site for miR-189 (Abelson, 2005). Another study on congenital cataracts identified a SNP in the seed region of miR-184, expressed in the central cornea. miR-184 is known to compete with miR-205, rescuing the knock down of *INPPL-1* by miR-205. The mutation in miR-184 impair its target binding leading to the repression of *INPPL-1* by miR-205 and the familial keratocornus syndrome provoking cataract (Hughes, 2011). A locus, which has

been linked to progressive hearing loss, led to the identification of SNPs in miR-96. These mutations, located in the seed region not only impaired the binding of the miRNA but also have been shown to destabilise the hairpin structure of the pre-miRNA and reduce the processing into mature miRNA. Five genes having a perfect match with wild type version of *mir-96* and expressed in the inner ear were selected and their *in-vitro* expression have been shown to increase with the mutant miR-96 in comparison to the wild type (Mencía, 2009). These three studies showed the role of polymorphisms in miRNA genes and 3' UTR targets in impairing binding sites and the subsequent dysregulation of the expression of genes leading to the modification of the phenotypes. It also showed that SNPs can also modify the secondary structure of the pre-miRNA impairing the processing into mature miRNA.

A wide range of studies in human showed the high impact of miRNAs and miRNA polymorphisms in different types of cancer such as breast, colon or prostate cancer, which made miRNA studies an important factor in oncogenetics (Xu, 2010; Zhang, 2011; Srivastava, 2012).

The SNPs in miRNA and miRNA targets have also been studied in livestock identifying their impact on productive traits. The most well-known study has been carried out in sheep. The Texel sheep has developed a trait in livestock called double-muscle phenotype, presenting higher muscle development than other breeds. In this study they identified a QTL associated with this trait in chromosome 2 using microsatellite sequencing of crossed Texel and Romanov sheep. The Romanov breed was used as a control as it was not presenting the double-muscle phenotype. The locus identified was refined and a specific marker was selected. This marker was known to be close to the *GDF8* gene, coding for the myostatin protein. Variants of *GDF8* were already known to create over-muscle phenotype in human, mouse and cattle. The analysis of the coding sequence did not show any SNPs, but the analysis of the flanking region revealed the presence of variants. They have been able to identify a SNP in the 3' UTR of *GDF8* and showed that this SNP created a binding-site for miR-1 and miR-106. This two miRNAs are highly expressed in muscle tissue, proving that these miRNAs were effectively inhibiting the expression of *GDF8*, dysregulating the control of muscle growth (Cloup, 2006).

The Texel sheep study, showed the potential impact of miRNA variations for livestock selection. Other studies also identified economically important traits such as

meat, milk and calving efficiency in pig and cattle, associated with miRNA and miRNA-binding site polymorphisms (J.-S. Lee, 2013; Sasaki, 2013; Mahmoudi, 2015).

1.5 Bioinformatics tools for identification of miRNA targets

The discovery of structure and function of miRNA and the increase in whole genome sequencing resulted in the development of prediction tool for the binding of miRNA to the 3' UTRs. There are many tools using different approaches for their identification and the main ones developed are miRanda, PicTar, Diana-microT, RNAhybrid and TargetScan, summarised below.

1.5.1 miRanda

The first published tool to identify miRNA targets is miRanda, published in 2004. It has been designed based on *Drosophila* data. The identification of the targets is made in three phases: sequence match, free energy calculation and evolutionary conservation. The sequence match is based on the Smith-Waterman algorithm but modified for the identification of complementary nucleotides and considering G:U wobble. It gives a score based on the complementarity of each nucleotide of the mature RNA to the 3' UTR, with few mismatches allowed and gap penalties. The free energy is calculated with the Vienna package to assure the thermodynamic stability of the RNA-RNA interaction (Wuchty, 1999). Then the conservation of the target is assessed by the comparison of the binding site between different species (different *Drosophilae* in this study). It requires that a miRNA match independently orthologous UTRs, the target sites need to have a higher identity score than a specified threshold and the position of the targets have to be similar when the UTRs between species are aligned (Enright, 2004).

1.5.2 Diana-microT

Shortly after miRanda, a second program, Diana-microT, was developed to predict miRNA-binding sites in human and mouse. Two criteria have been used for the identification of the target sites: the minimum energy of the complementary region and the loop and bulges present in the miRNA/miRNA-binding site duplex. They use experimentation on their first set of predicted miRNA-binding sites to refine their algorithm by adding a filter based on these experimental results (Kiriakidou, 2004).

1.5.3 RNAhybrid

RNAhybrid uses a former secondary structure prediction algorithm (Zuker, 1981) updated for two sequences. It is mainly focus on the minimum free energy (MFE) to determine stable hybridization. It filters out artefactual secondary structure and detects the seed region as a perfect match between nucleotides 2 to 7 of the miRNA and the target 3' UTR. For each site, the probability of the MFE and the number of sites identified for one miRNA is estimated to determine if this configurations can happen by chance (Rehmsmeier, 2004; Krüger, 2006).

1.5.4 PicTar

The prediction of targets by PicTar is firstly based on a region of seven nucleotides, called nucleus. They defined a “perfect nucleus” as a perfect pairing of the seven nucleotides from the first or the second nucleotide of the miRNA. They allow mismatches in the seed region as well as G:U wobbles if the free energy does not rise above a fixed value, defined then as an “imperfect nucleus”. Then PicTar calculates a maximum likelihood score for the binding of a miRNA to a 3' UTR target. The probabilities are calculated for each subsequence of a given 3' UTR to be a target by a miRNA. Then a score is also given based on other species orthologous 3' UTR by averaging the score of each species for a given nucleus, which give the final score for a 3' UTR to be targeted by a given miRNA (Krek, 2005).

1.5.5 TargetScan

TargetScan is a tool developed in its first version in 2003 (Lewis, 2003) and has been continuously updated, with the last version TargetScan7 released in August 2015 (Agarwal, 2015). The first version of TargetScan aimed to identify miRNA targets in vertebrates. The first algorithm includes RNA:RNA thermodynamic interaction with comparison across species for conservation and searches 3' UTR for a perfect 7 nucleotides match from 2nd nucleotide to 8th nucleotide of the miRNA, extending it as far as possible allowing G:U wobble but no mismatches. When the seed is identified, the 3' region of the miRNA is tested for matches in the 35 nucleotides following the seed region pairing. Finally it assigned a score based on the number of seed identified in the 3' UTR and the MFE (Lewis, 2003).

The following version (TargetScan2) added the recognition of adenosine flanking the seed region in the 3' UTR and reduced the primary seed region to 6

nucleotides. These adenosines were found to be present more often in conserved predicted miRNA-binding sites (Lewis, 2005). The next major change happened in version 4 in which many components for improving miRNA-binding sites detection have been added and led to the creation of the context score. This new version takes into account the increased repressive effect of more than one miRNA-binding site in a 3' UTR for a given miRNA, the presence of AU rich pattern flanking the seed region, the pairing of the 3' end of the miRNA from the 13th to 16th nucleotides, the distance of the sites to the stop codon and the preferential location of target sites in both ends of the 3' UTR. These features have been identified using microarray analysis of miRNA repressive activity on constructed 3' UTR containing these different features. The fraction of repression of each feature has been assessed and is constituting the context score, implemented from TargetScan4 reflecting the repressive potential of each miRNA target sites.

The version of TargetScan used in this thesis is TargetScan6.1. This version contains an updated calculation of the context score called context+. This new scoring, resulting from microarray analysis, assessed the role of Target Abundance (TA) and Sites-Pairing Stability (SPS). It was observed that a low SPS and high TA were correlated in mRNA repression by miRNA. These observations of expression led to the update of the context score to context+ score including both TA and SPS (Garcia, 2011). Recently a newer version, TargetScan7 has been released and assessed the site accessibility for miRNA-binding, leading to the new context++ score (Agarwal, 2015).

The long-term development of TargetScan facilitated a wide range of improvement compared with other methods. The many features of miRNA and miRNA targets identified for binding-sites prediction, the flexibility on conserved sites, allowing research of off-site targeting and the scoring system supported by experimental assays made TargetScan6.1 the most appropriate algorithm for the present study.

1.6 Research hypothesis and objectives

Research hypothesis

The cattle industry is of high economic importance in Ireland through beef and dairy production. Milk exports is the second highest agricultural product exported, representing 28% of these exportations. The foundation of ICBF was dedicated to improve cattle breeding and the value of cattle products by identifying valuable allele for the selection of superior animals and the improvement of genetic gain in the population. SNPs in micro-RNA binding sites, responsible for positive agronomic traits such as the double-muscling phenotype in sheep, highlighted the potential importance of miRNA as a tool for MAS. Furthermore, the advance in genetics and genomics allowed the full cattle genome sequencing leading to improvement of association studies. These factors facilitated the identification and characterisation of genes and polymorphisms explaining important traits variations and helped to improve cattle selection and breeding. The analysis of non-coding RNAs and particularly miRNAs is a rapidly developing field which already showed their importance in animal models like fly or mouse, is extensively studied in human and presented potential application for livestock.

Research hypothesis

In the thesis three chapters are presenting the research done on miRNA and miRNA binding sites variants between *Bos taurus* and *Bos primigenius* and also between cattle beef and dairy breeds, with the objective to investigate to the following hypothesis:

Genetic variants in miRNA and miRNA binding sites between *Bos taurus* and *Bos primigenius* are responsible for modification of the regulation of genes involved in the domestication process.

Genetic variants in miRNA sequences between beef and dairy cattle breeds modify the regulation of genes involved in important economic traits and thus could be used as candidate for genomic selection.

Genetic variants in 3' UTR between cattle breeds modify miRNA binding sites in these genes and are involved in the differential expression of important economic

traits, thus they could serve as useful candidates for genomic selection.

Research objectives:

- Identify the variants between aurochs and modern cattle in miRNA and miRNA targets

In this project, the access to the genome of aurochs (*Bos primigenius*), the ancestor of modern cattle, is of importance to discover variants which arise in cattle that could be specific to domestication and breeding process. For that purpose, the comparison of consensus sequences of *Bos taurus* and *Bos primigenius* has been done to identify the variants located in miRNA sequences and miRNA target sequences. To identify the miRNA targets a review of existing tools has been done and TargetScan has been selected. An algorithm has been developed to identify the targets having polymorphisms between the two species and to summarise and rank the genes based on the multiple results given by TargetScan. The ranking system helped to prioritise the genes having variants which could modify the most their function. Furthermore, the functions of these genes has been investigated using enrichment analysis on pathways and QTLs databases, publically available.

- Identify variants between different cattle breeds in miRNA genes and predict their function

The access to breed specific SNP is important for the deconvolution of the variants which are specific to beef breeds or dairy breeds. To investigate the variant present in miRNA sequences in different breeds, the algorithm developed for the analysis of *Bos taurus* and *Bos primigenius* has been adapted. The adaptation of the algorithm was orientated to compare more than two species. In this case the different breeds were treated in TargetScan as different species to be able to distinguished the different sequences. The same ranking system has been used and similar enrichments on pathways and QTLs have been done to decipher the functions of the polymorphic miRNAs. The ranking system has also been used to select the best candidates to be included into genotyping and association study of national breeding programmes such as the International Dairy and Beef SNP chip.

- Identify variants between different cattle breeds in miRNA targets and predict their functions

Finally, the tools developed in the two previous research objectives have been adapted

to the variation in binding sequences of miRNA targets and predict their potential function and importance in breeding. The same cattle breeds have been used and the algorithm for variant targets comparison has been modified in the similar way than in second objectives to consider multiple groups. The list of genes having variants in their binding sites has been ranked in the same way and used for enrichment in pathways and QTLs. The best candidates based on the ranking system has been used in the breeding programmes genotyping and association study.

2 Genome-wide microRNA and microRNA binding site variation between extinct wild aurochs and modern cattle identifies candidate microRNA-regulated domestication genes

2.1 Introduction

The wild aurochs is considered as the ancestor of modern cattle. Many studies on microsatellite and mitochondrial genome retrace the evolution of modern cattle, deciphering the origin of European cattle as progeny of middle-eastern aurochs (Bradley, 2006; Edwards, 2007, 2010). Since their domestication from the wild aurochs (*Bos primigenius*) some 10,000 years ago, cattle (*Bos taurus*) have been continuously exposed to both natural and artificial selection (Bradley, 2006; Magee, 2014). These processes, coupled with mutation and genetic drift, have produced a multitude of breeds of extant humpless taurine and humped zebu cattle breeds (Food and Agriculture Organisation, 2015). However, the underlying genomic factors (including genes, regulatory elements and DNA sequence differences (polymorphisms)) contributing to the traits selected during the domestication process remain largely unknown (Elsik, 2009; Canavez, 2012; Kemper, 2012; Womack, 2012; Utsunomiya, 2015; Gutiérrez-Gil, 2016).

Recent studies have highlighted the importance of microRNAs (miRNAs)—short non-coding RNAs that post-transcriptionally regulate gene expression—in regulating a wide range of biological processes in mammalian species including domestic livestock (McDanel, 2009; Liu, 2010; Fatima, 2013; Wang, 2013). Non-coding miRNA genes are initially transcribed to produce primary and then precursor miRNA molecules. The precursor miRNA is then processed into the short mature miRNA which is the molecule that can target and down regulate expression of other transcripts (Kim, 2009). In mammals, miRNA targeting has predominantly been associated with the 3' UTR region of transcripts derived from open reading frames (ORFs), typically leading to down-regulation through triggering RNA degradation, RNA instability and/or reduced translation (Winter, 2009; Krol, 2010; Huntzinger, 2011).

DNA sequence polymorphisms occurring within the targeted miRNA-binding site of genes can have influence phenotypic traits of economic importance in mammalian livestock (Clop, 2006; Sasaki, 2013; Hou, 2015; Rong, 2015; An, 2016). It also has been shown in a study on rabbit domestication that regulatory region polymorphisms are likely to play an important role in the domestication process in the modification of traits in a polygenic fashion (Carneiro, 2014). Consequently, it has been hypothesized that polymorphisms affecting miRNA gene and miRNA-mediated gene regulation are associated with pre- and post-domestic differences in *B. taurus*, particularly for traits under natural and artificial selection within the ~10,000 years timeframe of *B. taurus* domestication (Larson, 2014, 2014; MacHugh, 2016).

Using sequence data from the recently published *B. primigenius* genome (Park, 2015) and sequenced genomes of modern-day *B. taurus* (Zimin, 2009), a ranking metric [Score Site ratio (SSr)] has been developed to identify the miRNA-targeted bovine genes that display the greatest sequence difference in their miRNA binding sites between domesticated taurine cattle breeds and wild aurochs. Using this approach, genes displaying high levels of miRNA binding site DNA sequence differences (polymorphisms) between orthologous gene pairs have been identified, where one gene is from the wild aurochs and the other is from domesticated taurine cattle. Genes displaying post-domestication variation in miRNA binding sites are shown to be involved in mammalian growth and development (e.g., ephrin signalling and androgen signalling) and cellular function and metabolism (e.g., cholesterol biosynthesis). It has been considered that some of the miRNA binding site polymorphisms identified here, which are either absent or specific to the aurochs genome, modulate expression of candidate domestication genes associated with traits that have been selected during or subsequent to domestication.

The aim of this study is to test the hypothesis that genetic variants in miRNA and miRNA binding sites between *Bos taurus* and *Bos primigenius* are responsible for dysregulation of genes involved in the domestication process. The first objective is to identify the SNPs between *B. primigenius* and *B. taurus* located in miRNA and miRNA binding sites. The second objective is to identify the targeted genes which are different between the two species and evaluate their potential as marker of domestication.

The results of this chapter have been published in two articles. The results of the analysis of the variation in miRNA sequence between *B. primigenius* and *B. taurus* have been included in the functional analysis section of the aurochs genome study done

by Park et al. and published in *Genome Biology* (Park, 2015). The results of the analysis of miRNA binding sites in polymorphic 3'UTR between *B. primigenius* and *B. taurus* have been published in *Frontiers in Genetics* (Braud, 2017).

2.2 Materials and methods

2.2.1 *B. primigenius* genome sequence data

All *B. primigenius* genome sequence data was retrieved from the Sequence Read Archive database (accession number SRX1266623). The laboratory and bioinformatics methods used for DNA extraction, purification and sequence analysis of the CPC98 aurochs specimen have been described previously (Park, 2015). The genome sequence data is from an archaeologically verified *B. primigenius* specimen retrieved in 1998 from Carsington Pasture Cave in Derbyshire, England, and radiocarbon dated to $6,738 \pm 68$ calibrated (cal.) years before present (yBP). The authenticity of the CPC98 genome sequence was verified using a number of different approaches that are detailed by Park et al. (2015). The high-quality single nucleotide polymorphisms (SNPs) that were used for the analyses described here were identified using a stringent computational workflow as fully described in Park et al. (2015).

2.2.2 Workflow for the identification and comparison of polymorphic binding sites

To identify miRNA binding sites which are polymorphic between *B. primigenius* and *B. taurus*, a workflow has been developed. This workflow can consider SNPs in miRNA mature sequences and also SNPs in 3' UTR. If miRNAs are, considered the row headers in the final file will contain the miRNA names and if the 3' UTR are considered the row headers will contain the gene name.

The pipeline is divided into three main components (**Figure 10**). The first component (A) consists of the identification of the sequences needed for the analysis namely: the 3' UTR reference sequence and variant sequence, here respectively *B. primigenius* and *B. taurus* and similarly the miRNA mature sequences and the seed region sequences. The second component (B) is the identification of the binding sites and the scoring using TargetScan on the entire set of sequences identified. Finally, the third component (C) is the sorting of the TargetScan results and the final scoring with the SSr method.

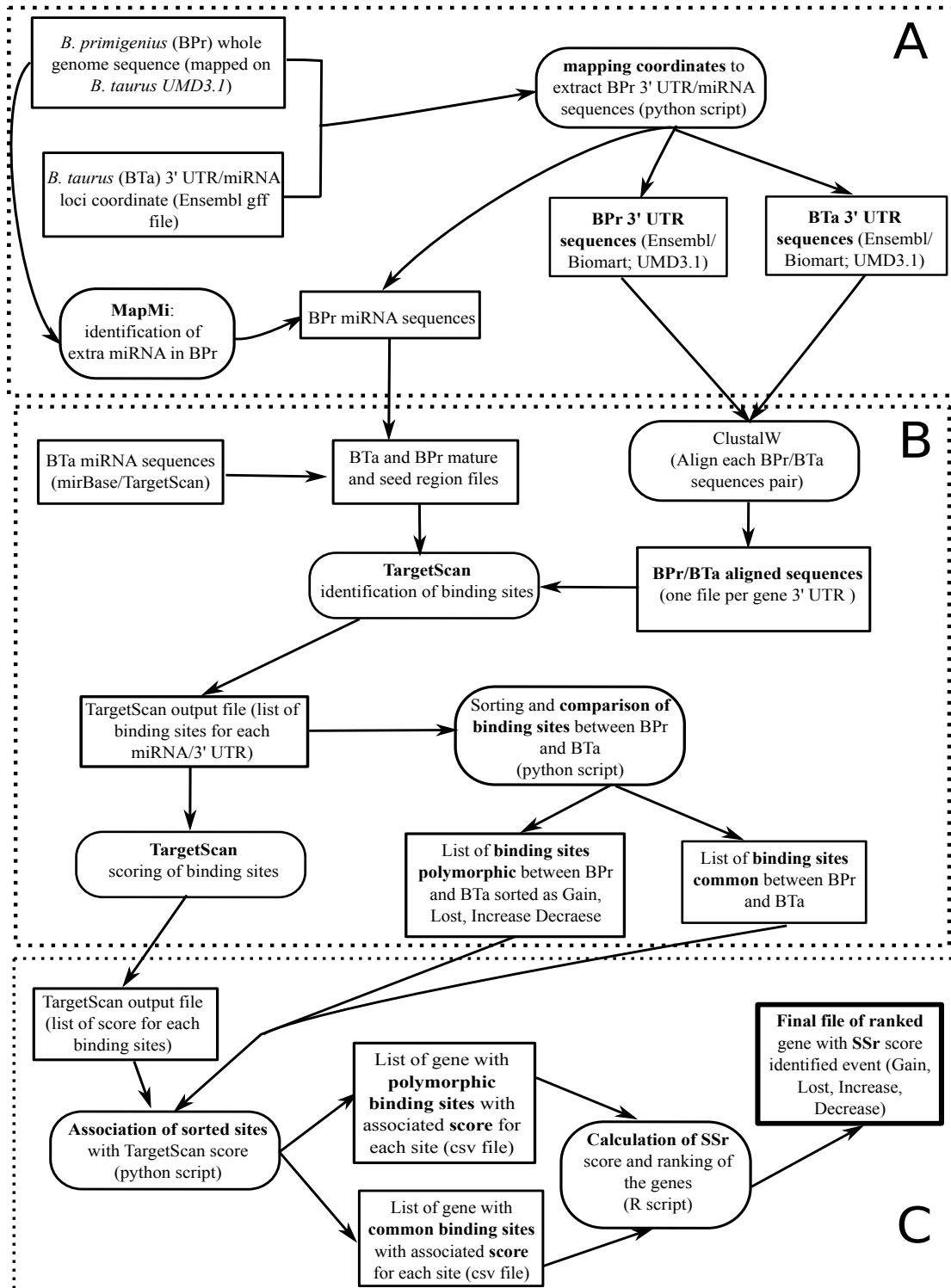


Figure 10 – Workflow for the identification of polymorphic miRNA binding sites.

The workflow is taking both reference (here *B. primigenius*) and variant (here *B. taurus*) miRNA and/or 3' UTR sequences to identify the miRNA binding sites. Section A identifies and reconstructs the variant sequences, section B identifies the binding sites base on these sequences and section C sorts the bindings sites for each gene/miRNA and calculate for each of them the corresponding SSr score.

2.2.3 *B. primigenius* miRNA identification

B. primigenius miRNAs were identified by mapping the *B. taurus* miRNA coordinates onto whole genome sequence alignments between the *B. taurus* genome (UMD3.1) and the *B. primigenius* consensus sequence. The pre-miRNA sequence of *B. primigenius* and *B. taurus* were compared using a python script and the polymorphisms were identified. To verify if the polymorphic sequence formed a stable hairpin loop each sequence was analysed with RNAfold (Hofacker, 2003) to predict its secondary structure. The minimum free energy, measuring the stability of the predicted RNA structure, given by RNAfold were also used to compare polymorphic hairpins between *B. taurus* and *B. primigenius*.

To identify mammalian miRNAs in the genome of *B. primigenius* that are unannotated in the genome of *B. taurus*, the mature sequences of the mammalian miRNAs (across all available species) were used in conjunction with the *B. primigenius* consensus sequence and putative miRNA genes were identified using MapMi (Guerra-Assuncao, 2010). The putative *B. primigenius* miRNA genes were compared to the closest mammalian homologues as defined by MapMi scores. These putative genes were then searched for in the *B. taurus* genome aligned sequence and the miRNA having polymorphisms in their mature sequences were filtered in (Figure 10-A).

2.2.4 Identification of *B. primigenius* 3' UTR sequences

To identify miRNA binding sites in target genes, 3' UTR sequences and associated bovine genome coordinates were obtained from Ensembl using the Biomart retrieval tool (Hubbard, 2002; Kinsella, 2011) and the *B. taurus* UMD3.1 reference genome assembly (Zimin, 2009). The information for the 3' UTR sequence from *B. taurus* UMD3.1 annotation has been used to predict and retrieve the 3' UTR sequences from the *B. primigenius* genome consensus sequence, using the exact coordinates associated with each *B. taurus* 3' UTR (Figure 10-A). Then each *B. taurus* and *B. primigenius* 3' UTR pairs has been aligned with clustalW and formatted to be treated with TargetScan. All *B. taurus* and *B. primigenius* 3' UTR pairs were then used to identify sequence polymorphisms within putative miRNA binding site for *B. taurus* and *B. primigenius*, using the differential results from TargetScan through a custom script screening each 3' UTRs pair for nucleotide differences in the sequence. In addition, the SNP data from the sequencing of 49 *B. taurus* cattle samples from the aurochs genome study have been used to confirm the fixation of the SNPs identified between *B. taurus* and *B.*

primigenius across different breeds (Angus, Holstein, Jersey, Limousin, Romagnola, Fleckvieh, N'Dama).

2.2.5 Identification of miRNA complementary binding sites in 3' UTR using TargetScan

To identify miRNA target binding sites in *B. taurus* and *B. primigenius* 3' UTR sequences and compare them, the TargetScan 6.1 software package was used (Lewis, 2005). Targetscan has been previously shown to be an accurate predictor of miRNA binding sites (Baek, 2008; Ulitsky, 2010). The prediction score for each miRNA binding site incorporates the evolutionary conservation of a predicted miRNA binding site among different species (Friedman, 2009). It also takes into account the presence of many structural motifs in mRNA sequences related to miRNA binding to infer a particular score for each site (**Figure 10-B**).

2.2.5.1 TargetScan results treatment pipeline

The TargetScan 6.1 package consists of two main Perl scripts: `targetscan_60.pl` and `targetscan_60_context_scores.pl`. The first script, `targetscan_60.pl`, identifies only the primary seed region recognition motif composed of six nucleotides (positions 2 to 7 in the mature miRNA) and one additional flanking nucleotide (position 1 or 8 of the mature miRNA) to identify four types of seed region (six nucleotides, seven nucleotides (6 + 1 adenosine), seven nucleotides, and eight nucleotides (7 + 1 adenosine)). The script `targetscan_60.pl` takes for inputs the sequence alignments of the 3' UTRs and the list of miRNA seed regions for *B. taurus* and *B. primigenius* miRNAs. For each alignment the script returns an output containing the miRNA seed regions that have been identified in the 3' UTR and their corresponding positions (**Figure 10-B**).

The output of this script was parsed using a custom Python script that reads through the output files and extracts the information contained in each line of the file. This includes the binding site nucleotide location in the 3' UTR and the group of species that have a site at the same position in the 3' UTR multiple alignments. The *B. taurus* and *B. primigenius* seed regions that were identified, yet containing ambiguous nucleotides in the *B. primigenius* 3' UTR were removed from the analysis. Any seed sites overlapping with a previously identified seed site for the same miRNA were also removed (**Figure 11**).

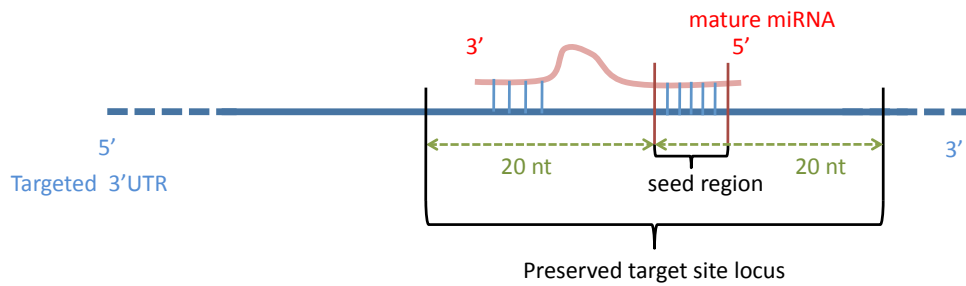


Figure 11 – Representation of the target sites locus restriction. When a given miRNA is identified to have overlapping target in the same locus, only one site is retained. This figure shows the distance restriction of 20 nt, flanking the last seed region nucleotide, set to avoid overlapping sites for an identical miRNA.

The second Targetscan Perl script, `targetscan_60_context_scores.pl`, generates a score for each site taking into account the position of the seed region in the 3' UTR, its nucleotide composition, the pairing of the miRNA in its 13-16 position with the UTR and the composition of the flanking region of the binding site. The Targetscan context score corresponds to a prediction, based on experimental miRNA repression, expressed as log-ratios (Grimson, 2007; Garcia, 2011). For this script, the results files from the first Targetscan script (catalogued as either a gain, a loss, an increase or a decrease of miRNA target sites), the alignments and the list of mature miRNA sequences specific to either *B. taurus* or *B. primigenius* were used as inputs.

To parse the results, a custom Python script was used to generate two results files: a file containing genes with miRNA binding sites polymorphism between aurochs and domestic taurine cattle and a second file with shared miRNA binding sites across the two taxa (**Figure 10-B**). All results in which ambiguous nucleotides were present in the UTR target region (complementary to the 3' end region of the mature miRNA) were removed. For miRNAs with multiple binding sites, the score for each gene/miRNA pair was generated by summing the individual context scores (calculated by TargetScan for each site).

2.2.5.2 Score Site ratio (SSr): development of a metric for the comparison of polymorphic binding sites

The analysis of miRNA binding sites produces a large amount of results, with each miRNA potentially targeting several genes (Enright, 2004; Chen, 2006). Similarly, a gene can be targeted by more than one miRNA and one miRNA can have multiple binding sites in one given gene. To compare the effect of a SNP in a binding site or in miRNA seed region, it is necessary to give an overall weight for each polymorphic gene or miRNA. This weight will allow a ranking of the results highlighting the genes most affected by the polymorphisms including an overall value of a given target compared to the number of targets for other miRNAs in a given gene. To rank the targeted genes, an additional score, termed the Score-Site ratio (*SSr*) has been developed. The *SSr* facilitated the analysis to which extent miRNA binding sites polymorphisms between *B. taurus* and *B. primigenius*, relative to those that were common to both *B. taurus* and *B. primigenius*, affected the potential expression of targeted genes. The *SSr* was calculated by summing all context scores for each predicted miRNA binding site across each 3' UTR and multiplying this sum by the ratio of the number of polymorphic sites over the number of common sites between *B. taurus* and *B. primigenius* (**Figure 10-C**). The *SSr* statistic is calculated using the following equation:

$$SSr = \sum_{i=1}^n score_i \times \frac{psn}{csn}$$

where n is the number of miRNA binding sites identified for a particular gene 3'UTR; $score_i$ is the score for each predicted site; psn (polymorphic site number) represents the number of miRNA binding sites polymorphic between *B. taurus* and *B. primigenius* for a particular gene; and csn (common site number) represents the number of miRNA binding sites common to *B. taurus* and *B. primigenius* for a particular gene.

The *SSr* score was used to rank genes according to the extent of their miRNA site polymorphisms differentiating *B. primigenius* and *B. taurus*. This provided a basis to compare the genes according to the extent of dysregulation of their miRNA binding sites between aurochs and modern cattle. In addition, the *SSr* statistic provides a weighting for each gene, which combines both the number of different miRNA binding sites (ratio of number of sites) and their possible impact (sum of TargetScan scores) between *B. taurus* and *B. primigenius* and facilitates normalization of the number of discovered sites, which can depend on the size of the 3' UTR. The calculation of the ratio of number of sites in the *SSr* score helped to emphasise genes having

polymorphisms creating a new binding sites in comparison with the reference gene. If the considered gene presented many other non-polymorphic binding sites, the score will be then reduced and the gene will be ranked down. Alternatively, when it contains few to no non-polymorphic binding sites, the related gene with a new polymorphic binding site will be ranked up.

2.2.6 Analyses of biological pathways involved in differential miRNA targeting of *Bos taurus* genes relative to *B. primigenius*

To identify the biological function that are the most affected by miRNA binding site variation between the 3' UTRs of *B. taurus* and *B. primigenius* genes, Ingenuity Pathway Analysis (IPA – www.ingenuity.com) was used. *B. taurus* genes with variant miRNA binding sites (when compared with *B. primigenius*) were used as the input to IPA. The IPA core analysis evaluates enrichment of the selected genes in different pathways, returning a *P*-value for each pathway identified, which was then corrected for multiple used to rank enriched pathways and identify the ones that are statistically significant (*P*-value < 0.05). A pathway score was then assigned to each significant pathway corresponding to the sum of the *SSr* score for each gene associated with a particular pathway.

2.2.7 Analysis of transition/transversion ratios

To determine if SNPs identified between *B. taurus* and *B. primigenius* could be due to post-mortem deamination the transition/transversion ratio (ti/tv) were calculated. To calculate the ti/tv in the 3' UTR miRNA binding site SNPs, the fixed SNPs identified in the miRNA binding sites were sorted in two mutational classes based on the variants observed: a transition class (A↔G, C↔T) and a transversion class (A↔C, A↔T, C↔G, G↔T). Using a custom python script, the number of variants were summed for each group and the ti/tv ratio was calculated.

2.3 Results

2.3.1 Identification and conservation of microRNAs in the *B. primigenius* genome

To identify and determine the extent of conservation of precursor microRNA genes between domesticated cattle (*Bos taurus*) and wild aurochs (*Bos primigenius*), all 672 annotated *B. taurus* pre-microRNAs (<http://miRBase.org/>, release 18) were mapped using whole genome alignments of *B. primigenius* consensus sequence and *B. taurus* genome (UMD3.1) and miRNA annotation file from miRBase.

The majority (554, 82.44%) of annotated domestic *B. taurus* pre-miRNAs were identified to have 100% coverage of the *B. primigenius* draft genome. In contrast, 32 (4.76%) *B. taurus* pre-miRNAs were not detected in the *B. primigenius* consensus sequence. Of those pre-miRNAs that mapped to the consensus sequence of *B. primigenius*, 508 (91.7%) were identical to their corresponding *B. taurus* pre-miRNA sequence. Of the remaining 46, 23 contained ambiguous nucleotides in the *B. primigenius* sequence. The other remaining 23 contained a single SNP difference between the pre-miRNA in *B. taurus* and the predicted pre-miRNA in *B. primigenius*, with the exception of three pre-miRNAs that had more than one SNP difference; bta-mir-769 (two SNPs), bta-mir-2417 (three SNPs) and bta-mir-2284d (three SNPs).

All of the 23 *B. primigenius* pre-miRNA sequence that have been predicted and containing a SNP, formed a stable hairpin loop structure comparable to their corresponding *B. taurus* pre-miRNA sequences. Out of the 23 polymorphic pre-miRNAs, five predicted *B. primigenius* pre-miRNAs have been identified to contain SNPs in their mature miRNA sequences (bta-mir-769, bta-mir-940, bta-mir-2391, bta-mir-2469 and bta-mir-2893) relative to their *B. taurus* genome homologues. Of these five, one miRNA, bta-mir-2893, found on the plus strand of chromosome 17 (17:73346080-73346140) contains a SNP at position 7 of the 7mer seed region (position 8 of the mature miRNA) (**Figure 11**).

bta-mir-2893	1	-GUGGAGGAGAAUGCCCGGGG-----	20
		.	
bpr-mir-2893	1	CGUGGAGGCGAAUGCCCGGGCGUGCGGGCGUGCGGGCUGGGCUGGCA	50
bta-mir-2893	20	-----	20
bpr-mir-2893	51	GCUCCUCCUC	61

Figure 12 – Predicted pre-miRNA of mir-2893 in the *B. primigenius* (bpr) genome aligned to the *B. taurus* (bta) mature sequence of mir-2893. The presence of a SNP, in position 8 (A/C) of the mature miRNA (position 7 of the seed region), leads to a modification of the miRNA targeting between *B. primigenius* and *B. taurus*.

2.3.2 Precursor microRNA genes have not been lost in the *B. taurus* genome relative to the wild *B. primigenius* genome

To determine whether any conserved microRNAs had been lost in the *B. taurus* genome during the domestication process from *B. primigenius*, a pipeline has been developed to test for loss of conserved miRNAs in domestic *B. taurus* relative to wild *B. primigenius*. All mammalian microRNAs that did not have an annotated gene family member in the *B. taurus* genome were retrieved, and putative pre-miRNA sequences were searched for in the genomic consensus sequence of the *B. primigenius* using MapMi (Guerra-Assuncao, 2010) and the mature miRNA sequences of each mammalian member of the gene family annotated in miRBase (<http://miRBase.org/>, release 18). The settings used allowed three mismatches in the mature miRNA sequence, a maximum of 0.05% low complexity sequence, and a minimum MapMi score of 40.

344 mammalian miRNA-families were identified as having no annotated miRNA-family member in the *B. taurus* genome. Of these 344 microRNA families, 59 miRNA-families (matching to 74 mature sequences) were identified as having 153 putative miRNA genes in the *B. primigenius* genome. Furthermore, all of the 153 putative miRNA genes in the *B. primigenius* genome had a minimum free energy of less than -22 kcal/mol, which is in the range of the including cutoff value employed in miRNA identification studies (Lai, 2003; Bonnet, 2004; Wang, 2005). 111 of these putative *B. primigenius* miRNAs had at least one SNP in the mature sequence and nine contained a single SNP in the seed region (each SNP found in position 7 of the seed)

when compared to their closest mammalian homologue. One *B. primigenius* miRNA had an indel in its mature sequence compared to its closest mammalian homologue.

To determine whether these 153 putative *B. primigenius* miRNA genes were genuinely absent from the domesticated *B. taurus* genome, or are just un-annotated in the *B. taurus* genome, the consensus sequence from the genome of *B. taurus* (UMD3.1) was retrieved. Each *B. taurus* sequence from the appropriate region for each miRNA was then tested to identify whether it had a putative pre-miRNA structure. Each of the 153 *B. taurus* sequences was found to form a stable hairpin structure (RNAfold, MFE < -22 kcal/mol). The predicted pre-miRNAs from both *B. taurus* and *B. primigenius* were aligned. Fifteen of the 153 (un-annotated) putative pre-miRNAs identified contained SNPs between *B. taurus* and *B. primigenius* (a further 21 contain ambiguous nucleotides in the *B. primigenius*). In all cases, none of the SNPs in the pre-miRNAs were located in the mature sequence region of the miRNA gene. This result indicates that these 153 predicted pre-miRNAs in the *B. taurus* genome represented un-annotated miRNA genes rather than miRNA genes that were once in the genome of *B. primigenius* genome but lost in the *B. taurus* genome during domestication. Hence, there was no evidence of any loss of precursor miRNA genes from the *B. primigenius* genome relative to the *B. taurus* genome.

2.3.3 *Bta-mir-2893* contains a SNP between domesticated *B. taurus* and wild *B. primigenius* (bpr-mir-2893) that leads to differential targeting

As indicated, only five predicted *B. primigenius* pre-miRNAs contain SNPs in the mature miRNA sequences (bta-mir-769, bta-mir-940, bta-mir-2391, bta-mir-2469 and bta-mir-2893) relative to domesticated *Bos taurus*. To determine the extent of any differential miRNA targeting due to polymorphisms in mature miRNAs between *B. primigenius* and *B. taurus*, the predicted 3' UTRs were extracted from the *B. primigenius* consensus sequence using the *B. taurus* 3' UTRs annotation as a guide. Using TargetScan target detection software with a strict 5' seed pairing requirement, all potential 3' UTR targets were identified for bta-mir-769, bta-mir-940, bta-mir-2391, bta-mir-2469 and bta-mir-2893, using the corresponding *B. primigenius* and *B. taurus* mature sequences (Table 1).

Table 1 – Number of predicted 3' UTR targets for mature miRNAs (bta-miR-769, bta-miR-940, bta-miR-2391, bta-miR-2469, bta-miR-2893) having polymorphic pre-miRNA between *B. taurus* and *B. primigenius*.

miRNA	<i>B. taurus</i>	<i>B. primigenius</i>	Overlap (number of target)	Overlap (% of <i>B. taurus</i> target)	Overlap (% of <i>B.</i> <i>primigenius</i> target)
mir-769	573	572	570	99.5	99.7
mir-940	1246	1241	1233	99.0	99.0
mir-2391	349	344	343	98.3	98.3
mir-2469	712	705	701	98.5	98.5
mir-2893	1148	1077	610	53.1	53.1

The overlap of the predicted targets between *B. taurus* and *B. primigenius*, for bta-miR-769, bta-miR-940, bta-miR-2391 and bta-miR-2469, corresponded to 98.3% to 99.7% of all targets. In contrast, the same overlap, in the case of bta-miR-2893/bpr-miR-2893 (where position 7 of the seed region differs) corresponded to 53.1% of all *B. taurus* targets and 56.6% of all *B. primigenius* targets, indicating that polymorphisms in the bta-miR-2893 leads to a divergence of targeting of genes and pathways between *B. primigenius* and *B. taurus*.

2.3.4 The SNP present in miR-2893 leads to loss and gain of gene targets between *B. taurus* and *B. primigenius* potentially dysregulating *PHYHIP*, *FADS2* and *FCRL1* expression

Using the workflow incorporating Targetscan and the SSr ranking system which has been developed, the genes having the greatest difference in their miR-2893 targets between *B. taurus* and *B. primigenius* were identified. 1030 new miR-2893 targets have arisen in 931 *B. taurus* genes. Among these 931, 558 have orthologous genes in *B. primigenius*, which are not targeted at all by the *B. primigenius* mir-2893 variant. Furthermore, 940 miR-2893 binding sites in 849 genes were lost in *B. taurus* in comparison with *B. primigenius* orthologous genes, among which 474 genes did not have any binding sites at all for miR-2893 in *B. taurus* in comparison with the *B. primigenius* orthologues. These genes have been ranked using the SSr described previously (**Supplementary table 1**). Considering the highest SSr ranked genes, three genes stand out as important candidates having functions, which could have been selected during domestication, namely *PHYHIP*, *FADS2* and *FCRL1*. *PHYHIP* and *FADS2* genes were identified as not targeted in *B. taurus* by miR-2893. In contrast, *FCRL1* is newly targeted in *B. taurus* by miR-2893. The miR-2893 regulatory polymorphisms for these three genes could lead to an upregulation of *PHYHIP* and *FADS2* and a downregulation of *FCRL1* (**Table 2**). The gene *FADSI*, a paralog of *FADS2*, had been shown to be dysregulated between the domesticated chicken and wild fowl (Li *et al.*, 2012). The dysregulation via mutations in miR-2893 could have played a role in the artificial selection transition from *B. primigenius* to *B. taurus*.

Table 2 – Top ranked genes differentially targeted by bta-miR-2893 and bpr-miR-2893

PS: Polymorphic Sites; CS: Common Sites; TgS sum: TargetScan score sum of all sites identified; Gain: unique sites in *B. taurus*; Lost: unique site in *B. primigenius*; Increase: extra sites in *B. taurus*; Decrease: extra sites in *B. primigenius*

Ensembl ID	gene name	PS	CS	TgS sum	SSR	event
ENSBTAG00000023551	<i>MOB3C</i>	3	1	-0.591	-1.773	Decrease
ENSBTAG00000015572	<i>YTHDC1</i>	2	1	-0.823	-1.646	Increase
ENSBTAG00000011539	<i>PCGF2</i>	3	1	-0.541	-1.623	Increase
ENSBTAG00000002985	<i>KCNMB1</i>	3	1	-0.519	-1.557	Lost
ENSBTAG00000003595	<i>FAM216B</i>	2	1	-0.732	-1.464	Increase
ENSBTAG00000015757	<i>PHYHIP</i>	3	0	-0.484	-1.452	Lost
ENSBTAG00000003595	<i>FAM216B</i>	2	1	-0.719	-1.438	Decrease
ENSBTAG00000000080	<i>GRWD1</i>	3	1	-0.467	-1.401	Increase
ENSBTAG00000007415	<i>SLC7A8</i>	2	1	-0.697	-1.394	Increase
ENSBTAG00000015200	<i>L2HGDH</i>	2	1	-0.672	-1.344	Decrease
ENSBTAG00000018999	<i>MGRN1</i>	4	1	-0.326	-1.304	Decrease
ENSBTAG00000026242	<i>BAALC</i>	2	1	-0.631	-1.262	Increase
ENSBTAG00000039335	<i>ARRDC2</i>	2	1	-0.61	-1.22	Decrease
ENSBTAG00000007415	<i>SLC7A8</i>	3	1	-0.404	-1.212	Decrease
ENSBTAG00000016805	<i>SGMS2</i>	2	1	-0.603	-1.206	Increase
ENSBTAG00000000097	<i>KIAA0494</i>	2	1	-0.593	-1.186	Increase
ENSBTAG00000021799	<i>RCN3</i>	2	1	-0.584	-1.168	Decrease
ENSBTAG00000014944	<i>TBXA2R</i>	2	1	-0.566	-1.132	Increase
ENSBTAG00000015200	<i>L2HGDH</i>	2	1	-0.548	-1.096	Increase
ENSBTAG00000004556	<i>SLC2A3</i>	2	1	-0.544	-1.088	Increase
ENSBTAG00000047314	<i>PPIE</i>	2	1	-0.542	-1.084	Increase
ENSBTAG00000013607	<i>WDR1</i>	2	1	-0.534	-1.068	Decrease
ENSBTAG00000021969	<i>KLHL29</i>	2	1	-0.524	-1.048	Decrease
ENSBTAG00000005891	<i>FCRL1</i>	2	0	-0.521	-1.042	Gain
ENSBTAG00000033679	<i>HLCS</i>	2	1	-0.51	-1.02	Increase
ENSBTAG00000027654	<i>EIF4EBP1</i>	2	1	-0.507	-1.014	Decrease
ENSBTAG00000017251	<i>SLC26A8</i>	2	1	-0.504	-1.008	Increase
ENSBTAG00000003209	<i>BLCAP</i>	2	1	-0.503	-1.006	Increase
ENSBTAG00000010775	<i>POGK</i>	2	1	-0.503	-1.006	Increase
ENSBTAG00000004603	<i>PHC2</i>	2	1	-0.499	-0.998	Increase
ENSBTAG00000014916	<i>GJC2</i>	3	1	-0.328	-0.984	Increase
ENSBTAG00000021879	<i>VCL</i>	2	1	-0.479	-0.958	Decrease
ENSBTAG00000018362	<i>TMEM109</i>	2	1	-0.465	-0.93	Increase
ENSBTAG00000015505	<i>FADS2</i>	3	0	-0.296	-0.888	Lost

2.3.5 Neurodevelopment, fertility and pigmentation pathways are enriched for genes targeted by miR-2893

The pathways which were enriched in genes, targeted by bta-miR-2893 and its *B. primigenius* variant, were identified using IPA software (**Figure 13, Figure 14**). Genes having polymorphic targets for miR-2893 were significantly enriched in 48 pathways in *B. primigenius* and 89 in *B. taurus*, with 32 pathways in common. More specifically, amongst the 57 pathways specific to *B. taurus* some pathways clearly related to the domestication process. For instance, the gain of *B. taurus*-specific targets by miR-2893 were implicated in axonal guidance signalling and androgen signalling. Moreover, the *B. primigenius*-specific pathways, which were potentially less targeted by miR-2893 in *B. taurus*, are sucrose degradation and melanocyte development and pigmentation signalling. Reduced targeting by miR-2893 of these pathways could therefore have been selected through the process of domestication and subsequent artificial selection.



The green bars represent the pathways significantly enriched uniquely in *B. taurus*. The values on the top of the bar represent the sum of the SSr score of each targets involved in the pathways

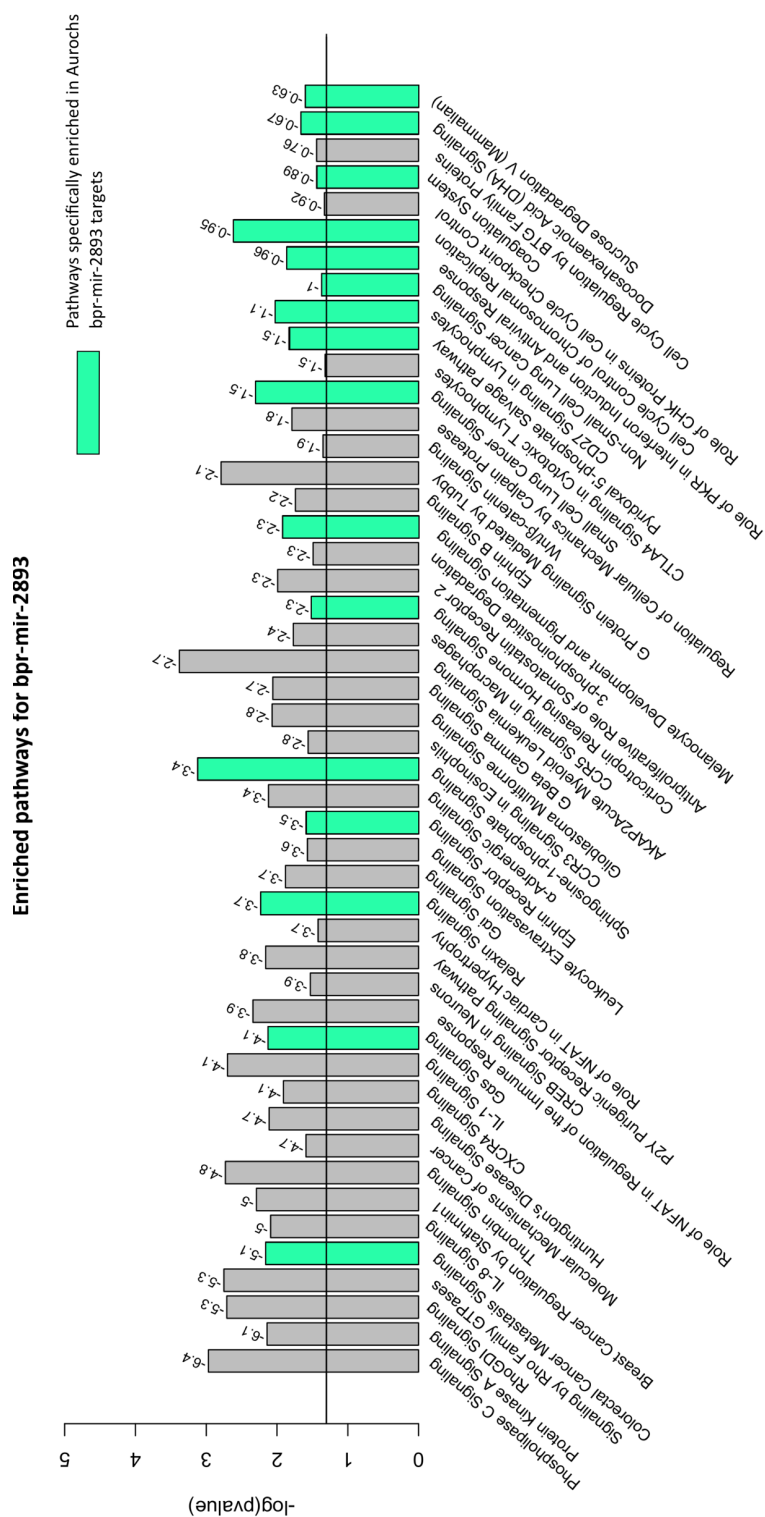


Figure 14 – *B. primigenius* miR-2893 targets enriched pathways

The green bars represent the pathways significantly enriched uniquely in *B. primigenius*. The values on the top of the bar represent the sum of the SSr score of each targets involved in the pathways

2.3.6 Comparison of *B. taurus* and *B. primigenius* 3' UTR sequences revealed a large number of DNA sequence polymorphisms

Of the 19,994 annotated protein-coding genes in the taurine cattle genome (assembly UMD3.1), 11,761 had an annotated 3' UTR (59%). To identify all SNPs differentiating the 3' UTRs of *B. taurus* and *B. primigenius* genes, we compared these 11,761 3' UTRs sequences from *B. taurus* with the corresponding mapped regions in *B. primigenius*. It has been determined that 3,066 (26%) of these genes contained one or more SNPs, and identified a total of 6,355 SNPs that were different between *B. taurus* and *B. primigenius* in the 3,066 3' UTRs. Analysis of the numbers of SNPs in the 3' UTR of each gene yielded a mean of 1.92 SNPs per 3' UTR, with a median value of 1.00 SNPs per 3' UTR and a maximum of 27 *B. taurus*/*B. primigenius* SNPs for the methionine adenosyltransferase I, alpha gene (*MAT1A*).

2.3.7 DNA sequence polymorphisms differentiating *B. taurus* and *B. primigenius* 3' UTRs generated altered miRNA binding sites

Of the 3,066 3' UTRs exhibiting SNPs differentiating the two taxa, a total of 2,634 SNPs located in miRNA binding sites within 1,620 genes were identified. Notably, DNA sequence polymorphisms in the 3' UTR of the protein kinase, AMP-activated, gamma 3 non-catalytic subunit gene (*PRKAG3*) produced the highest number of miRNA binding site changes across all of these 1,620 genes. For the *PRKAG3* 3' UTR, the analysis demonstrated that 23 different miRNAs had a modification of their binding potential (gain or loss). These 23 miRNAs were associated with 15 distinct miRNA-binding sites that differ between the *B. taurus* and *B. primigenius* *PRKAG3* genes.

In the analysis of polymorphisms in 3' UTRs, individual miRNA binding sites were classified as a “gain” when SNPs between *B. taurus* and *B. primigenius* generated new miRNA binding sites in *B. taurus* 3' UTRs, and as a “lost” when they lead to loss of an existing miRNA binding site in the *B. taurus* genome. In some cases, a polymorphic miRNA binding site can be considered simultaneously as a “gain” and a “loss” when one or more SNPs generate new binding sites for one particular miRNA and eliminate a binding site for another miRNA. An additional scenario were considered when a miRNA gains a new binding site in *B. taurus* but also had another common binding site between *B. taurus* and *B. primigenius*. Such miRNA binding site copy number “amplification” were categorised as an “increase,” while a reduction of binding site copy number was categorised as a “decrease.”

Among the 1,620 genes with SNPs in the 3' UTR, 892 genes had lost their miRNA binding sites (including up to 13 losses for the membrane-spanning 4-domains, subfamily A, member 7 gene (*MS4A7*) in the *B. taurus* lineage. Moreover, 885 genes that had gained a miRNA binding site in *B. taurus*, which is not present in *B. primigenius* have been identified (including a gain of up to 11 new miRNA binding sites for the 3-hydroxy-3-methylglutaryl-CoA synthase 2 gene (*HMGCS2*). Furthermore, an overlap of 322 genes that have both a gain and loss of miRNA binding sites in *B. taurus* has been identified. Finally, 306 genes in *B. taurus* had a number of miRNA-binding sites that have increased in number for at least one miRNA (up to seven miRNA binding sites for the ubiquitin specific peptidase 33 gene, *USP33*), while 250 genes had a decrease in miRNA binding site copy number (up to 11 miRNA binding sites for the *MAT1A* gene in comparison to *B. primigenius* (**Figure 15**).

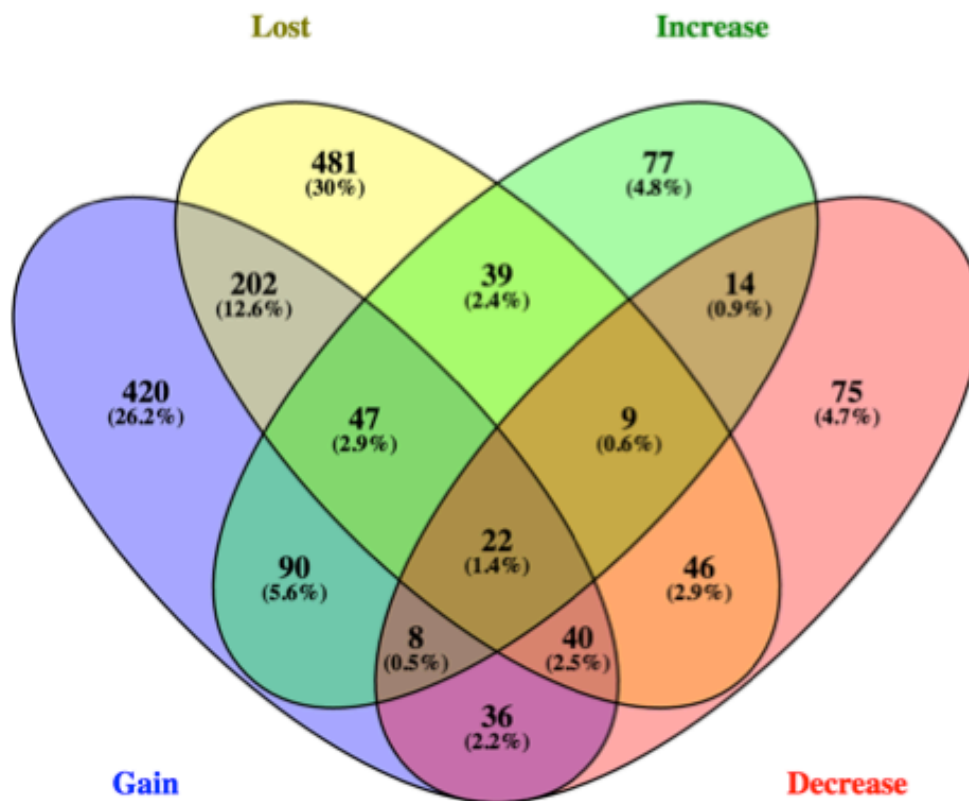


Figure 15 – Polymorphisms between *B. taurus* and *B. primigenius* in 3' UTR create or impair miRNA-binding sites. The Venn diagram represents the number of genes having polymorphisms in their 3' UTR between *B. taurus* and *B. primigenius*, modifying miRNA binding sites. Gain (blue) or lost (yellow) corresponded respectively to the gain or loss of all miRNA-binding sites identified for a given miRNA. Increase (green) and decrease (red) corresponded respectively to an augmentation or diminution of the number of target sites for a given miRNA. The intersection in loss and gain represented these SNPs in 3' UTR which created a binding site for one miRNA and impaired it for another one in the same gene sometimes for the same binding site.

2.3.8 Using SSr scores to rank the 3' UTRs that had undergone the most extensive miRNA binding site modifications between *B. primigenius* and *B. taurus*

The hypothesis in this study is that selection processes (both natural and artificial) during and after domestication may generate major changes in the miRNA binding sites of genes whose expression contributes to variation in complex traits in *B. taurus*. To identify which miRNA targeted genes and cellular pathways that displayed the most miRNA targeting divergence between *B. taurus* and *B. primigenius*, the SSr score have been used to rank the different genes (**Supplementary table 2**). Notably, several of the top 10 SSr-ranked genes were shown to play key roles in important metabolic, physiological and reproductive processes. These included: (1) the guanylate cyclase activator 2B (*GUCA2B*) gene, in which the binding sites for five miRNAs (bta-miR-541, bta-miR-2433, bta-miR-1777a, bta-miR-2454, bta-miR-2327) were absent in *B. taurus* but present in *B. primigenius* with only one common binding site between both; (2) the mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase (*HMGCS2*) gene, in which the binding sites for 11 miRNAs were present in *B. taurus* but absent in *B. primigenius* with also the number of binding sites increased for two miRNAs in *B. taurus*; and (3) the membrane-spanning 4-domains subfamily A (*MS4A7*) gene, in which the binding sites for 13 miRNAs were absent in *B. taurus* and present in *B. primigenius* representing the most important loss of potential binding of miRNAs of the entire dataset. Three metabolically relevant genes: glutathione peroxodase type 5 (*GPX5*), kappa casein-3 (*CSN3*) and S-adenosylmethionine synthetase type-1 (*MAT1A*), that displayed DNA sequence polymorphism differences between the *B. taurus* and *B. primigenius* lineages in their associated miRNA binding sites have also been identified (**Table 3**).

Table 3 – Candidates domestication genes having polymorphic miRNA binding sites between *B. taurus* and *B. primigenius*

Gene name	Rank	SSr	MiRNA targeting polymorphic site	Event
<i>GUCA2B</i>	1	-5.200	bta-miR-541:bta-miR-2433:bta-miR-1777a:bta-miR-2454:bta-miR-2327	Lost:5
<i>MS4A7</i>	5	-1.156	bta-miR-17-5p:bta-miR-2284p:bta-miR-221:bta-miR-222:bta-miR-93:bta-miR-20b:bta-miR-433:bta-miR-20a:bta-miR-302d:bta-miR-106:bta-miR-302b:bta-miR-302c:bta-miR-106b	Lost:13
<i>HMGCS2</i>	10	-0.905	bta-miR-301b:bta-miR-301a:bta-miR-2375:bta-miR-454:bta-miR-152:bta-miR-148b:bta-miR-148a:bta-miR-130a:bta-miR-130b:bta-miR-2317:bta-miR-2329-3p:bta-miR-199a-5p:bta-miR-199b	Increase:2 Gain:11
<i>PRKAG3</i>	16	-0.659	bta-miR-2302:bta-miR-345-5p:bta-miR-2311:bta-miR-2309:bta-miR-2329-3p:bta-miR-542-5p:bta-miR-2917:bta-miR-141:bta-miR-2330:bta-miR-200a:bta-miR-1343:bta-miR-2397:bta-miR-429:bta-miR-2285b:bta-miR-200c:bta-miR-200b:bta-miR-491:bta-miR-2888:bta-miR-873:bta-miR-2383:bta-miR-764:bta-miR-369-3p:bta-miR-2442	Increase:6 Decrease:1 Gain:6 Lost:10
<i>GPX5</i>	28	-0.298	bta-miR-677:bta-miR-449d:bta-miR-26c:bta-miR-33a:bta-miR-2294:bta-miR-33b:bta-miR-18b:bta-miR-18a:bta-miR-217:bta-miR-2461-5p:bta-miR-2319b	Increase:1 Gain:8 Lost:2
<i>MAT1A</i>	30	-0.293	bta-miR-2459:bta-miR-495:bta-miR-423-3p:bta-miR-2388:bta-miR-2406:bta-miR-2427:bta-miR-2450b:bta-miR-194:bta-miR-2888:bta-miR-2327:bta-miR-491:bta-miR-152:bta-miR-148b:bta-miR-1777a:bta-miR-148a:bta-miR-2454:bta-miR-2433:bta-miR-2416:bta-miR-1434	Increase:1 Decrease:1 1 Gain:2 Lost:5
<i>CSN3</i>	53	-0.193	bta-miR-2312:bta-miR-2301:bta-miR-1603:bta-miR-2337	Gain:1 Lost:3

2.3.9 Identification of networks, pathways, and functions enriched in genes having polymorphic miRNA targets sites between *B. taurus* and *B. primigenius*

Ingenuity Pathway Analysis (IPA) knowledgebase was used to gain further insight into the biological functions, canonical pathways and biological networks enriched by the genes with polymorphic miRNA binding sites between *B. taurus* and *B. primigenius*. For this, the 1,606 polymorphic 3' UTRs identified have been uploaded to the IPA server to identify functions, canonical pathways and networks based on their associated *p*-value (as generated by IPA). Using this approach, 25 networks, 73 pathways and 500 functions, significantly enriched in *B. taurus* relative to *B. primigenius* have been identified.

The two main networks identified are metabolism and neurological diseases. Among the significantly enriched canonical pathways differentially regulated by the miRNAs were pathways involved in immunobiology (for example, the *IL1*, *CXCL4* and NFAT pathways), physiology and development (for example, the ephrin signalling pathway and the *PTEN* and *PI3K/AKT* signalling pathways), and reproductive physiology (for example, the androgen signalling pathway) (**Figure 16**).

2.3.10 The polymorphic miRNA binding sites in aurochs 3' UTRs were not enriched for transitions due to post-mortem deamination

The analysis of ancient DNA sequence data can be difficult due to post-mortem deamination leading to C→T or G→A transition substitutions (Hofreiter, 2001; Briggs, 2007). To investigate whether the 2,634 SNPs (that were polymorphic in the miRNA binding sites in aurochs 3' UTRs) could be enriched for transitions due to post-mortem deamination, the ti/tv ratio was calculated for the set of 2,634 SNPs located in miRNA binding sites and was determined to be 1.93:1.00. Previously, for the entire CPC98 aurochs genome, 2,009,261 biallelic SNPs were examined (73.3 % of which were homozygous) and it has been determined a transition to transversion (ti/tv) ratio of 2.19:1.00 (Park, 2015), which is very similar to ti/tv ratios obtained for female and male Holstein genome sequences: 2.18:1.00 and 2.16:1.00, respectively (Köks, 2013, 2014). The lower ti/tv ratio detected for the 3' UTR miRNA SNPs is consistent with published ti/tv ratios calculated for mammalian 3' UTRs (Chen, 2015), and suggests that post-mortem deamination is not a confounding factor for SNPs within miRNA binding sites located in aurochs 3' UTRs.

binding sites between *B. taurus* and *B. primigenius*. The values on the top of the bar represent the mean of the SSr score of each targets involved in the pathways

2.4 Discussion

The analysis of whole genome sequence data from mammalian livestock and their wild ancestors have revealed genetic loci and genetic changes associated with the domestication process (Groenen, 2012; Carneiro, 2014; Schubert, 2014). Indeed, these studies have highlighted how natural and artificial selection, together with genetic drift, have shaped post-domestication patterns of genetic variation, that contribute, in part, to the wide range of phenotypic variations observed in modern breeds such as tameness, increased fertility, decreased body size and coat colour modification (Vigne, 2011). Notably, recent work has shown that miRNAs play a key role in regulating a wide range of biological processes at the post-transcriptional level in mammals while DNA sequence variation in the miRNA-binding sites of mammalian genes can also influence phenotypic variation. Therefore, for the present study, the comparison of miRNA gene and miRNA-binding sites of domestic *B. taurus* and their wild ancestor, *B. primigenius*, could reveal novel DNA sequence polymorphisms differentiating the two lineages and which may underlie key traits that have been subject to selection during and after the domestication process.

2.4.1 Polymorphisms between *B. primigenius* and *B. taurus* affect miRNA binding to 3' UTRs

By comparing genes annotated in *B. primigenius*, mapped on to the *B. taurus* genome, many SNPs which were correlated with miRNA binding sites and fewer SNPs associated with changes in miRNA genes have been identified. A major polymorphism between *B. primigenius* and *B. taurus* in the seed region of *bta-mir-2893* leading to major difference between *B. primigenius* and *B. taurus* regarding the genes targeted by this miRNA have been found. There is almost no study mentioning *mir-2893* in any species. However, the study mentioned by miRBase, which identified *bta-mir-2893* showed that this miRNA is expressed in ovaries of domestic *B. taurus* (Hossain, 2009). The SNP identified in *mir-2893* in comparison with *B. primigenius* was specific to *Bos taurus* species as was fixed across 7 different *B. taurus* breeds. Although the seed region of mature *bta-miR-2893* was polymorphic between *B. primigenius* and *B. taurus*, it showed some conservation between *B. primigenius* and humans orthologous *hsa-miR-3197* (miRBase BLAST online tool). This conservation supported the genuine activity of this miRNA in *B. primigenius* and consequently the modification of activity in the *B. taurus* orthologue. Furthermore, the same miRBase online BLAST searches showed

that there is no identified orthologue for *bta-mir-2893* across all referenced species. Thus, it led one to consider that this miRNA was spurious and that its characteristic of miRNA is to be widely conserved across species (Altuvia, 2005). However, the fact that this miRNA has been identified in an expression study on a specific organ in cattle could give more confidence that this miRNA was providing some functional role, especially as it was variable with the cattle ancestor. It was reasonable to suggest that the SNP in the seed region of miR-2893 that appeared in *B. taurus* could be link with the domestication process of *B. primigenius*.

More extensively, miRNA binding site polymorphisms in the 3' UTRs of 1606 genes have also been identified. The novel SSr ranking system based on TargetScan have been developed to identify the 3' UTRs which had undergone the most miRNA binding site divergence between *B. taurus* and *B. primigenius*. This facilitated identification of genes which had undergone the greatest amount of miRNA regulatory modifications between *B. primigenius* and *B. taurus*. These genes were then analysed for any enrichment in relation to pathways, networks and functions.

By analysing the ranked genes and pathways, five main traits, have been identified which could have been altered during the domestication of ancestral *B. primigenius* via miRNA regulation, leading to current *B. taurus* breeds. Polymorphisms that differ between *B. primigenius* and domestic *B. taurus*, related to miRNA regulation were preferentially related to immunology, metabolism, physiology, development and fertility, suggesting that these processes played a major role in the domestication of ancestral *B. primigenius* via miRNA regulation, leading to development of modern *B. taurus*.

2.4.2 Neurological development is affected by the modification of miRNA binding between *B. primigenius* and *B. taurus* and could be linked to tameness

Modifications to neurological development likely led to tameness and improved the handling of *B. taurus* by increased docility. One gene and two pathways which were related to neurological regulation and signalling and showed high miRNA regulation polymorphism between *B. primigenius* and *B. taurus* have been identified. The SNP in *bta-mir-2893* leads to differential targets for *PHYHIP*, which is no longer a *bta-mir-2893* target in *B. taurus* while it had three putative sites in *B. primigenius*. The protein produced by *PHYHIP* is a substrate of DYRK1A, involved in neurodevelopment (Murakami, 2009). The axonal guidance pathways, enriched in genes targeted by *bta-*

mir-2893, and the ephrin pathways, which were enriched in genes having miRNA target sites polymorphisms, have also been implicated in brain development (Soskis, 2012). These modifications between *B. primigenius* and *B. taurus* tended to show brain development modifications which could have resulted from selection for docile animals.

2.4.3 Change in metabolism during domestication could be linked to the SNP in *mir-2893* and polymorphisms in miRNA-binding sites

Metabolism may have also been an important process to be modified during and after the domestication due, primarily, to changes in the diet of animals within increasingly organised agricultural systems over the millennia; a process that continues today (Noe-Nygaard, 2005). Four genes and two pathways have been found related to metabolism and have undergone extensive modification between *B. primigenius* and *B. taurus*. *FASD2*, an enzyme involved in fatty acid metabolism, had lost its *mir-2893* targets in *B. taurus* which could led to its upregulation and increased metabolic flux of this enzyme (Schaeffer, 2006). *HMGCS2*, which encodes mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase, is targeted by miRNAs in *B. taurus* but not in *B. primigenius* and ranked in our top ten genes. *HMGCS2* is involved in ketone metabolism, regulation of which varies depending on the food intake (Kostiuk, 2010). The overexpression of *HMGCS* in transgenic mice led to an increase of ketone bodies and a decrease of lipolysis in adipose tissue (Valera, 1994).

In addition, miRNA-binding site differences between *B. taurus* and *B. primigenius* in the *MAT1A* gene (methionine adenosyltransferase) have been identified. *MAT1A* is exclusively expressed in mature liver cells in mammals where it is involved in the regulation of hepatic function (Mato, 2002). *MAT1A* also played a role in S-adenosylmethionine formation, which has been linked, in knock-down mice, to hepatic regeneration (Chen, 2004). Polymorphisms in *MAT1A* were also related to the level of homocysteine in dietary fatty acids (Huang, 2012).

Finally, the SSr-ranking method identified *PRKAG3* as showing miRNA-binding polymorphisms between *B. taurus* and *B. primigenius*; previous work has shown that polymorphisms in this genes have been associated with differences in glycogen levels (Milan, 2000) and meat quality in domestic livestock (Lindhahl, 2004; Roux, 2006). Moreover, the insulin receptor signalling pathway and valine degradation pathways were also highly enriched in genes having polymorphisms in miRNA binding

sites, both of which were involved in metabolism. The different metabolic genes and pathways identified as under differential miRNA regulatory selection in *B. taurus* versus *B. primigenius* could have had a role in the evolution of *B. taurus* feeding strategies, growth rate and meat quality improvement.

2.4.4 Modification in miRNA gene and miRNA-binding sites play a role in potential immunologic changes during domestication

Under a model of domestication, increased confinement and interaction with humans within settlements would be expected to result in concomitant exposure of early domesticates to novel pathogens, leading to increased selection pressure on genes regulating immune functions (Cleaveland, 2001). In this context one highly ranked gene, and one entire pathway, involved in immunity and related to *mir-2893* variant between *B. primigenius* and *B. taurus* have been identified. *FCRL1* is newly targeted in *B. taurus* by miR-2893. This gene is involved in immune response and has been shown to result to autoimmune disease if overexpressed (Baranov, 2012). The NFAT pathway enriched for targets of the polymorphic miR-2893 is also involved in innate immunity, which can allow fast responses to infection (Fric, 2012). A modification in this pathway could have resulted of selection due to the increase in infection in herds of greater size and intensification in domesticated cattle. Furthermore, analysis of variation in miRNA-binding sites also identified the immunology-related gene, *MS4A7*, which lost the most miRNA binding site in *B. taurus* in comparison to *B. primigenius*. *MS4A7* is part of the *MS4A* cell surface protein family (Liang, 2001), the members of which are expressed in B lymphocytes (Pant, 2006; Zuccolo, 2013). Moreover, *MS4A7* polymorphism have been linked to differential expression of this gene in human B cells in response to glucocorticoid treatment (Maranville, 2011), while a recent study has shown that a CpG epigenetic polymorphism in 3' UTR of *MS4A7* have been associated with the metabolism of coprostanol in humans, which is linked to cardiovascular disease (Irvin, 2014). Polymorphism in 3' UTR between *B. primigenius* and *B. taurus*, differentially targeted by miRNAs, have also been found enriched in the NFAT pathway.

2.4.5 Reproduction, fertility and lactation are important traits in the extension of domestication of *B. taurus* and is linked to modification in miRNA and miRNA binding sites

Fertility is of key importance in *B. taurus* improvement to allow a higher birth rate and more effective reproduction while maintaining milk (lactation) yields. Two genes and five pathways have been identified of relevance to fertility, which are highly ranked in relation to SNPs in gene binding sites. Glutathione peroxidase-5 gene (*GPX5*) is targeted by eight additional miRNAs in *B. taurus*. The level of the GPX5 protein expressed played an important role in spermatozoa DNA integrity maintenance (Chabory, 2010). Moreover, *GUCA2B*, which is the highest ranked gene, encodes a fertility-associated protein binding to GUCY2C, a membrane protein, which had a role in the regulation of the concentration of cGMP. *GUCA2B* is also involved in the fluid balance in intestine and kidney (Sindić, 2006) and the variation in expression of *GUCA2B* in mice has been shown to have an effect on pregnancy and fertility (McConaha, 2011). The androgen signalling pathways, enriched in genes containing miRNA-binding sites polymorphisms, had impacts on both bull and cow fertility. Conjugated linoleic acid have been shown to increase the expression of *PTEN* expression and *PI3K/AKT* regulating roles in granulosa cell proliferation and steroidogenesis in buffalo which could affect the ovulation and therefore the fertility (Sharma, 2012). The IL-8 pathway is activated during theca cell differentiation and in pre-ovulatory follicle differentiation (Walsh, 2012). The Platelet-derived growth factor signalling pathway has also been shown to be of high importance in development and especially in embryogenesis (Hoch, 2003). In addition, one of the highest ranked genes showing loss of miRNA targets (rank 52/1606), *CSN3*, encodes kappa casein-3 which is an important protein for milk nutrition quality and milk processing (Comin, 2008).

2.4.6 Pigmentation pathway genes are differentially targeted between *B. taurus* and *B. primigenius*

A significant gene enrichment for melanocyte development and pigmentation signalling has been identified to be associated with miRNA regulatory polymorphism between *B. primigenius* and *B. taurus*. Such miRNA-regulatory effects on this pathway can be linked to the coat colour changes that appear during domestication. The genes involved in these pathways have also been linked to tameness, consistent with the link between domestication and coat colour changes (Linderholm, 2013). The coat colour

change was not an unknown phenomenon in the domestication process. Various studies on the process of domestication of mammals observed this phenomenon. A study focus on coat colour related genes in horse showed that variation in colour appeared during domestication, the analysis of these genes suggested Eurasian steppe origin of horse domestication (Ludwig, 2009). Similarly a study comparing the genome of the wild boar and the domesticated pig found region under positive selection overlapping gene related to coat colour as well as muscle development with a variant of the gene having low frequency in a white pig breed (Amaral, 2011). It was also interesting to mention a study on silver fox aiming to reproduce domestication event solely on tameness selection. The process led to morphological modification as curled tail, floppy ear and also coat colour changes (Trut, 2009). Similarly, a study on mink showed that animals with aberrant coat colour were tamer which has been linked to a change of serotonin level in the hypothalamus (Trapezov, 2008).

2.5 Conclusion

The sequencing of the *B. primigenius* genome allows the identification of genetic differences arising from domestication and artificial selection of taurine *B. taurus* from *B. primigenius*. Genetic polymorphisms affecting miRNA regulation can have dramatic effects on livestock traits. MiRNA regulatory polymorphisms affecting key domestication and economic traits would have been selected for during domestication and subsequent artificial selection. In this study, miRNA regulatory polymorphisms that were different between *B. primigenius* and *B. taurus*, which represented candidate domestication mutations that helped define our current domestic *B. taurus* breeds, have been identified.

This study demonstrated that, small changes in miRNA seed regions or miRNA binding sites have led to major changes to the regulation of miRNA-targeted genes and/or biological functions within the short time-scale of *B. taurus* domestication. This observation was in line with a study in rabbit domestication, which suggested that small modifications in regulatory domain were more likely to be marker of domestication rather than drastic changes in coding region (Carneiro, 2014).

It has been identified that 1606 protein coding genes and one miRNA gene seed region had SNPs between *B. primigenius* and *B. taurus* that led to dramatic changes in miRNA gene targeting, affecting pathways and traits related to these genes. Important traits subject to altered miRNA regulation in *B. taurus* relative to *B. primigenius* have

been identified, including neurological development (axonal guidance), fertility (androgen signalling, ephrin pathways), immune characteristics (*FCRL1*, and NFAT pathways) and metabolism (*HMGCS2*), insulin receptor and sucrose degradation pathways. All of these miRNA-regulated pathways were highly associated with traits selected during the 10,000 years of domestication and can be considered as potential candidates for breeding improvement in *B. taurus*. A similar study on horse domestication, sequenced two ancient horse DNA pre-dating the horse domestication (Schubert, 2014). By comparing these sequences with modern horse and wild horse they identified candidate genes for domestication related to various function which suggested artificial selection by man for its own need but also behavioural traits linked with tameness. This is could be linked with candidate variant binding sites, discovered in this chapter, mentioning neurological development regulation. Furthermore they emphasised deleterious variant arising with inbreeding linked with selection which was also a concern in cattle selection where inbreeding led to decrease in production, fertility (Smith, 1998) or survivability (Thompson, 2000).

The results from the analysis of variant in miRNA between *B. taurus* and *B. primigenius* were published in *Genome Biology* (Park et al. 2015) as part of the functional analysis (**Section 8.1**). The results from the analysis of the variant in miRNA binding sites in 3' UTR were published in *Frontiers in Genetic* (Braud et al. 2017) (**Section 8.2**).

3 Identification of cattle breeds polymorphisms in miRNA sequences discover candidates for association studies

3.1 Introduction

MicroRNAs (miRNAs) are non-coding RNAs, the active form of which is a mature miRNA that ranges in length from 16 to 26 nucleotides. According to miRBase, the average length for miRNAs in cattle is 21 nucleotides. The first microRNAs to be discovered was described in *C. elegans* as a small RNA named *lin-4*, which regulated the expression of the LIN-14 protein and had antisense complementarities to the 3' UTR region of the *lin-14* gene (Lee, 1993; Wightman, 1993). Since their initial discovery in nematodes, miRNAs have been identified in many animals and plants (Lagos-Quintana, 2001, 2003; Lee, 2001; Llave, 2002; Reinhart, 2002; Kozomara, 2013).

There are currently 797 miRNA genes identified in cattle (miRBase 21), coding for 784 unique mature miRNAs (miRNA paralogues had the same mature sequence). MiRNAs have been shown to regulate gene expression in animals by complementary binding to the 3' UTR region of gene transcripts, leading to the degradation of the targeted transcript and/or inhibition of the mRNA translation (Carthew, 2009; Chekulaeva, 2009). The RISC complex is a ribonucleoprotein, composed of the mature miRNA sequence and an ARGONAUTE protein, which function was to facilitate the binding of the miRNA to its target sequence and the degradation of the messenger RNA (Kim, 2009). The miRNA binds its target to a six nucleotide region in positions 2 to 7 of the mature miRNA, called the seed region, which has perfect complementarity with its target. It has also been observed that complementarity of nucleotide 8 and the presence of an adenosine at position 1 also enhances the target binding, triggering a higher repressive effect (Lewis, 2005). It also has been shown than the region between nucleotides 13 to 16 of the mature miRNA can also be complementary to the 3' UTR binding sites, again enhancing the strength of the down-regulation (Grimson, 2007).

Studies on microRNAs in humans, livestock and other animal systems have increased drastically in the last decade in livestock and in other organisms (Fatima, 2013). Understanding of the role of miRNAs in the regulation of gene expression and their impacts on various traits such as development, cell death, inflammatory response

or cancer, have become a key part in mammals genetics (Ambros, 2004; Iorio, 2012; O'Connell, 2012).

MiRNAs played significant roles in the regulation of important genes and pathways expression in domesticated livestock. MiRNAs have been shown to have roles in many different economically important traits in cattle such as muscle, fat content, milk production, immunity and reproduction. The majority of studies conducted in livestock were focused on differential expression associated with a specific trait. For example, overexpression of *mir-206* was involved in the double-muscle phenotype in cattle (Miretti, 2011) and *mir-885* and *mir-196a* were differentially expressed in specific muscle type in Japanese black cattle breed (Muroya, 2013). It has also been reported that an increase of *mir-143* in Holstein muscular tissue enhanced the maturation of adipocytes (McConaha, 2011) and four other miRNAs were linked to development of subcutaneous and intra-muscular fat in Simmental beef (Wang, 2015).

The expression of miRNA also varies in relation to mammary gland maintenance and milk production traits. During different lactating period in cattle, it has been shown that 12 miRNAs were up-regulated during the fresh lactating period and one during early lactation (Wang, 2012). Many other studies have reported differences of expression in various miRNAs during udder infection showing the role of miRNA in the response to mastitis and immunity (L. Li, 2012; Naeem, 2012; Li, 2014). MiRNAs have also been shown to be involved in fertility traits by regulating genes in oocytes, sperm and gonadal tissues (Hossain, 2012).

The identification of polymorphisms and their link to different phenotype in livestock is an important subject for selection. A lot of effort has been put into sequencing and genotyping multiple breeds of livestock together with gathering phenotypical information associated with these genetic variations (Charlier, 2008; Consortium, 2009; D A Magee, 2010; Waters, 2011; Berry, 2012; Mullen, 2012). The identification of variations in miRNAs is still in an early stage and of their role in cattle is poorly understood. However, multiple studies have shown that polymorphisms in both miRNA genes and miRNA-binding sites can impair miRNA-binding properties and could therefore affect gene regulation. For instance, some SNPs in miRNAs have been associated with production traits in pig breeds where SNPs in *mir-206* and *mir-133b* were associated with different muscle and fat traits (J.-S. Lee, 2013). Moreover, recent studies identified polymorphisms in miRNA genes for vertebrates, including

livestock, using publicly available data. These studies showed that cattle have the greatest number of polymorphisms in miRNA genes when compared with 14 other species of interest (Jevsinek Skok, 2013; Zorc, 2015). Therefore, genetic differences in miRNA sequences between different cattle breeds, which could modify the regulation of genes involved in important functions and livestock traits, could represent novel candidates for selection during breeding programs.

To test this hypothesis, the present study aimed to identify polymorphisms in miRNAs of cattle and investigate their potential impact on traits of economic importance, based on their predicted targets. SNPs between six different breeds of cattle and the reference cattle genome were obtained from a collaboration with USDA as part of the aurochs sequencing project (Park, 2015). From this list of SNPs, the SNPs in all cattle miRNA binding sites were identified. The SSR scoring system was used to prioritise the miRNA target polymorphism and identify potential candidates, some of which have been included in the International Dairy Breed SNP chip (IDB) (Mullen, 2013) in collaboration with Teagasc and the Irish Cattle Breeding Federation. The genotyping with IDB gave the allelic ratio of the selected SNPs for a wider range of breeds. The SNPs were included in a second Teagasc study to be genotyped as part of an association study on Holstein-Friesian bulls for which daughters' phenotypes had been determined by ICBF.

3.2 Materials and Methods

3.2.1 Flowchart of identification of SNPs in miRNA for association study candidates

The SNPs data were retrieved from dbSNP and USDA cattle sequencing. The pipeline of identification and categorisation of the miRNA variant and their target is similar to the one developed in Chapter 2. The genotyping on 25 breeds was part of the IDB project led by Teagasc and ICBF. The association study with the identified miRNA variant from dbSNP is part of a study on Holstein bull led by Teagasc (**Figure 17**).

3.2.2 Identification of Single Nucleotide Polymorphism in miRNA genes

To identify polymorphisms in miRNA genes, data from dbSNP, available in 2010 (dbSNP build 130, BTau.4.1) and containing 9,493,419 SNPs, were used. MiRNA general feature format file (gff) were then retrieved from miRBase and Ensembl (BTau.4.1). The SNP loci were compared to the miRNA gene loci and those which overlapped with pre-miRNA transcripts were extracted. Each polymorphic pre-miRNA was reconstructed using a custom python script. Finally, the miRNA in which the polymorphisms were located in the mature sequence themselves were filtered in and the mature sequence and seed region were extracted for inclusion in the corresponding input TargetScan file (**Supplementary table 3, Supplementary table 4**). For the genotyping, which was performed subsequently the BTau4.1 coordinate were converted to the more recent UMD3.1.

In addition, a second set of SNPs identified in the resequencing of 45 animals from six European cattle breeds from low coverage sequencing (Park, 2015) were used (**Table 4**). These new set provided a total of 24,032,577 SNPs, a much higher number compared with those identified in dbSNP in 2010, as well as breed information. Comparison with the sequences of pre-miRNA, mature miRNA and seed regions were conducted as described previously.

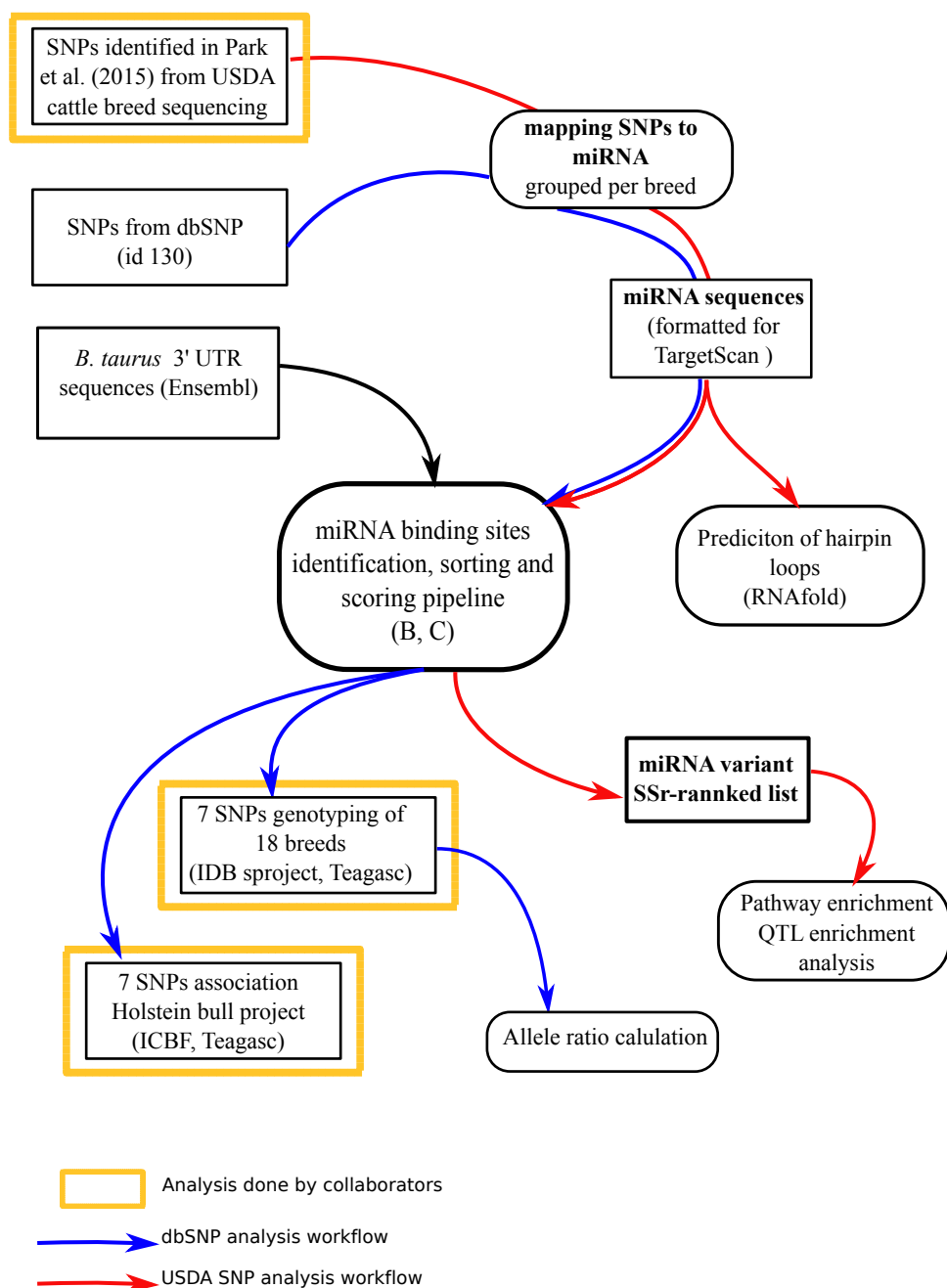


Figure 17 – Flowchart of identification of SNPs in miRNA for association study candidates

This represents the workflow leading to the identification and characterisation of SNPs in miRNA sequences. The pipeline consists of miRNA binding site identification (Chapter 2, Figure10-B) and scoring (Chapter 2, Figure10-C).

Table 4 – Description of the six cattle breeds sequenced. The number of animals is the total of individuals, which were sequenced and from which the SNPs were extracted by Park *et al* (Park, 2015).

breed	Number of animals	type
Angus (AN)	10	beef
Limousin (LM)	10	beef
Romagnola (RO)	3	beef
Fleckvieh (FL)	1	dual purpose
Holstein (HO)	10	dairy
Jersey (JE)	10	dairy

3.2.3 Prediction of the secondary structure of polymorphic miRNA

The hairpin loop secondary structure of pre-miRNA is important for mature miRNA production. It has been shown that polymorphisms in the hairpin loop can modify their stability and impair the maturation of pre-miRNA (Khvorova, 2003; Schwarz, 2003; Zeng, 2003; Iwai, 2005). Indeed differences in thermodynamic stability have been shown to modify the loading efficiency of pre-miRNA by the ARGONAUTE protein (Gu, 2011). To determine if polymorphisms in cattle miRNA genes can modify the hairpin loop structure of pre-miRNA, the secondary structures of each breed variants and reference pre-miRNAs were predicted using RNAfold (Hofacker, 2003) by determining the lowest minimum free energy (MFE) structure for each miRNA. The free energy was quantifying the stability of a DNA or RNA secondary structure hence the lowest one is the most stable. The MFE have been used to compare the differences produced by the SNPs in the hairpin structures.

3.2.4 Construction of 3' UTRs alignment

To identify miRNA target sites, TargetScan software was used. Alignments of 3' UTR were recovered and formatted as in Chapter 2 and used as the basis for TargetScan analysis.

3.2.5 Identification of the targets of the polymorphic miRNAs

As noted previously, the variant and the reference mature miRNA and seed region files were created and formatted similarly to those provided by TargetScan. Sequences with identical seed regions were grouped into families. For each polymorphism of each breed identified in seed region the variant sequence for each breed was added to the reference seed region file. For TargetScan to be able to group the similar binding sites, a unique identifier number for each breed was introduced instead of the species number usually used by TargetScan (*B. taurus* was originally tagged 9913). This unique identifier is designated as the breed tag namely: Holstein: 9913.1; Jersey: 9913.2; Angus: 9913.3; Limousin: 9913.4; Romagnola: 9913.5; Fleckvieh: 9913.6 while the reference sequenced was kept as 9913. Furthermore, the miRNA names in the file were given prefixes corresponding to the breed abbreviation (and “bta” for the reference) to allow each sequences to be distinguished. The identification of the binding sites was done using the same pipeline described in Chapter 2 (**Figure 10-B**).

3.2.6 Sorting and SSr-scoring of miRNAs variants

The results obtained from TargetScan were treated with the same pipeline described in Chapter 2 (**Figure 10-C**) with minor modification of the python and R scripts to handle more than two different taxa (here the six breed variants). The SSr scoring was similar, considering each variant miRNA in comparison with the reference one and summing the score of all targeted genes for each miRNA.

3.2.7 Pathway enrichments using IPA

To identify the likely functions of each miRNA, enrichment of the targeted genes in different biological pathways was conducted using the IPA software. The software was provided with a list of targets, common and polymorphic, for each miRNA variant and each corresponding reference sequence. The *p-value* obtained for each analysis of miRNA target enrichment was adjusted using the Benjamini-Hochberg method. The scoring and ranking of the pathways was performed as in Chapter 2. As more data could be derived from the use of multiple miRNA variants, the pathways were grouped into four categories (**Figure 18**):

Pathways specific to a single miRNA: these are enriched in novel targets (Lost or Gain) of one unique variant miRNA. These pathways were not identified as significant level in other specific variant miRNA targets (Lost or Gain) nor in the reference miRNA.

Pathway specific to multiple miRNAs: are similar to the single variant miRNA described previously except that more than one variant miRNAs of the same target type (Gain or Lost) is involved.

Pathway specific and common: are similar to the specific pathways for single or multiple miRNA except that they are also targeted by another variant miRNA (Gain) and the corresponding reference one (Lost).

Gain and Lost common pathways: these are enriched in novel targets for different miRNAs variants for Lost and Gain targets.

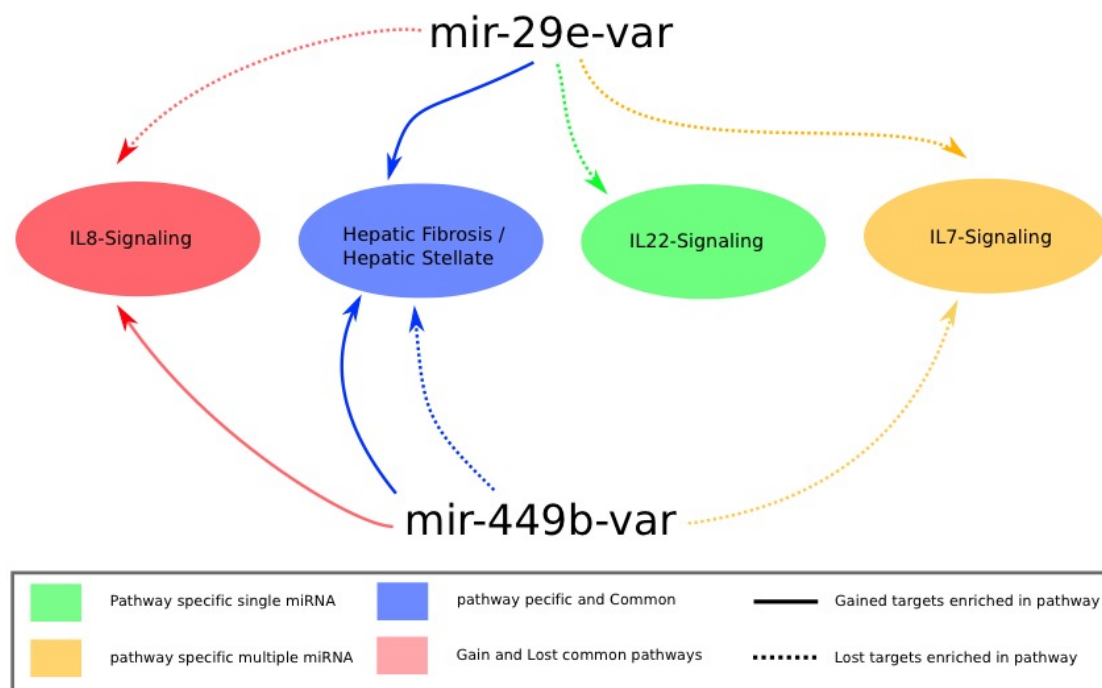


Figure 18 – Representation of the different groups of pathway enrichment. This is an example of how pathways enriched in genes differentially targeted by variant form of *mir-449b* and *mir-29e* were grouped. The pathways were grouped, depending of the type of targets they are enriched with. The type of targeted genes and enriched pathways were defined as “Lost” if the binding sites were all impaired in a given miRNA variant identified in comparison with the reference one. Similarly, a targeted gene and enriched pathway were defined as “Gained” if a new binding site arise from the variation of a given miRNA and no binding sites were identified in the reference miRNA. Pathways enriched in genes targeted both by variant and reference miRNA are defined as “Common”.

3.2.8 Genotyping with the IDB SNP chip of 25000 cattle for frequency analysis

Genotyping analysis was conducted in collaboration with Teagasc with the first version of the IDB project, integrating the seven SNPs identified in this chapter from the dbSNP data. These SNPs have been genotyped across 18 breeds (**Supplementary table 5**) using a custom made SNPchip. This chip was used to analyse 25000 animals, in association with the farmers.

3.2.9 Genotyping for gene association

A gene association analysis was also conducted in collaboration with Teagasc as part of a dairy cattle association study. A total of seven SNPs identified during the *in silico* analysis were genotyped across 914 Holstein-Friesian sires. The SNP genotyping was carried out commercially using the Sequenom MassArray® iPLEX Gold assay (Sequenom, San Diego, CA, USA). As a quality control measure, 25 animals, originating from 25 different samples of extracted genomic DNA were genotyped twice for all SNPs. Mean concordance across all SNPs and all duplicates was 99.8%. Where discordance existed, the genotype for the sample in question was discarded. The following method was developed and conducted by the Teagasc collaborators.

3.2.9.1 Data editing and phenotypes

An iterative algorithm was used to simultaneously discard both SNPs and individuals with poor genotype call rates (Waters et al 2011). This resulted in genotypes of 66 individuals and two SNPs being discarded.

Daughter yield deviations (expressed on the scale of predicted transmitting ability; PTA) and PTAs, as well as associated data reliabilities, for a range of performance traits evaluated by the Irish Cattle Breeding Federation in the January 2009 domestic genetic evaluations, were available for inclusion in the analysis. Models used in genetic evaluations in Ireland, as well as variance components, are summarized in detail by Berry et al. (Berry, 2007). Daughter yield deviations (DYD) for 305-day milk, fat and protein yield as well as geometric mean SCS (\log_e somatic cell count) are estimated in Ireland using a repeatability animal model across the first five lactations. PTA for calving interval and survival are estimated using a multi-trait animal model, including data from the first three lactations. PTA for milk yield is used to adjust survival for differences in genetic merit of milk yield; hence, this survival trait is functional survival. PTA for cow carcass weight, progeny carcass weight, progeny carcass fat score and progeny carcass conformation score, measured at slaughter, are

estimated in a multi-trait animal model that includes weaning weight, live-weight of the animal between 300 and 600 days of age, feed intake, and skeletal and muscular linear traits. Cows slaughtered between 875 and 4,000 days of age were included in the evaluation of cow carcass weight while male progeny slaughtered between 300 and 1,200 days of age and female progeny slaughtered between 300 and 875 days of age were included in the evaluation of the remaining three carcass traits. Genetic evaluations for linear type traits are undertaken as part of a joint evaluation in the UK and Ireland. The estimated breeding values (EBVs) were standardized to the mean and standard deviation of the base population. PTAs were de-regressed using the procedure outlined by Berry et al (Berry, 2009):

$$\tilde{\mathbf{y}} = \mathbf{R}(\mathbf{R}^{-1} + \mathbf{A}^{-1})\hat{\mathbf{a}}$$

Where $\tilde{\mathbf{y}}$ is the deregressed PTA, $\hat{\mathbf{a}}$ is a vector of PTAs from the genetic evaluations, \mathbf{R} is a diagonal matrix where each element is $\frac{1}{(\text{Reliability of the respective animal} - 1)}$, and \mathbf{A} is the numerator relationship matrix.

Parental contribution to the reliability of each DYD or PTA was removed using the approach of (Harris, 1998):

$$\tilde{R} = \frac{R_{\text{TRAD}} - R_{\text{PA}}}{R_{\text{TRAD}} \cdot R_{\text{PA}} + 1 - 2R_{\text{PA}}}$$

Where \tilde{R} is the reliability less the parental contribution, R_{TRAD} is the reliability from the traditional genetic evaluation (i.e., includes information from all relatives), R_{PA} is parental average reliability.

Only sires with a reliability (less parental contribution) of >60% for the trait under investigation were retained for inclusion in the association analysis. A total of 742 sires fulfilled these criteria for inclusion in the analysis of milk, fat and protein yield as well as milk fat and protein concentration; the number of sires included in the association analysis with SCS, calving interval and survival was 703, 501, and 477, respectively. The number of sires with a reliability of >60% for the carcass traits was 446 and the number of sires with a reliability of >60% for the linear type traits varied from 484 to 551. Out of the seven SNPs, only five passed the genotyping platform criteria while the two others were considered to present discordance.

3.2.9.2 Statistical analysis

The association between each SNP and performance was quantified using weighted mixed models in ASREML (Gilmour, 2009) with genotyped individual included as a random effect and average expected relationships amongst individuals accounted for through the numerator relationship matrix. Year of birth (divided into 5 yearly intervals) and percent Holstein of the individual sire were included as fixed effects in the model. In all instances the dependent variable was DYD for milk, yield, fat yield, protein yield and SCS and de-regressed PTA for the remaining traits, weighted by their respective reliability less the parental contribution.

Genotype was included in the analysis as a continuous variable coded as the number of copies of a given allele.

In total, 90 statistical tests (i.e., 5 SNPs by 18 traits) were undertaken but not all were independent because of covariance among traits. Banos et al. (2008) grouped phenotypes based on similarities. Using this approach, the traits were categories into 5 groups: milk production, udder health, fertility, carcass and linear type traits implying 25 different statistical tests ignoring linkage disequilibrium among SNPs.

3.3 Results and Discussion

3.3.1 Identification of miRNA SNPs from dbSNP

To identify polymorphisms in miRNA, all published cattle polymorphisms were collected using the publically available SNP data from the NCBI dbSNP website (id:130, 2010) and screened with a custom script using the annotation of mature miRNA from miRBase. Seven SNPs located in seven pre-miRNAs were identified (**Table 5**). Among these seven SNPs, two were located in the seed regions, one in miR-29e and one in miR-2313-3p seed region. In addition, one SNP was located in the mature sequence of miR-2419-3p but not in the seed region. The four other SNPs were located in the hairpin loop of the *mir-2329*, *mir-2284s*, *mir-1179* and *mir-1814b*. These seven SNPs have been used for the genotyping analysis (Mullen, 2013) and association studies, prior the access to the following USDA polymorphisms in cattle breeds.

Table 5 – List of the SNPs identified with dbSNP 130 data

miRNA name	chromosome	locus	ref	alt	dbSNP id	miRNA seq
bta-mir-29e	ch16	73969907	T	C	rs41825418	mature(5p) seed
bta-mir1179	ch21	19206160	A	G	rs41974457	hairpin
bta-mir-1814b	ch22	4893712	G	A	rs41990658	hairpin
bta-mir-2284s	ch12	53400159	T	G	rs43430325	hairpin
bta-mir-2313	ch15	32789501	C	T	rs41761413	mature(3p) seed
bta-mir-2329	ch18	47837451	T	G	rs41885631	hairpin
bta-mir-2419	ch4	68649844	G	A	rs43400521	mature(3p)

3.3.2 Identification of miRNA SNPs for six cattle breeds

As mentioned previously, 24,032,577 SNPs data from USDA cattle breeds genome sequencing was retrieved. A total of 125 SNPs located in 89 miRNAs genes were identified. In addition to the four SNPs in hairpin loop identified previously, 96 SNPs out of 125 have also been identified exclusively in hairpin loops in 71 miRNAs (i.e. not in the mature sequence).

Among the remaining 29 SNPs located in 25 mature miRNAs, 14 are in the seed region of 12 miRNAs, in addition of the two already identified (**Figure 19**). The SNP data presented each variation related to each animal for which the breed is known. Hence

each SNP identified in relation to the corresponding breed was selected, in order to reconstruct breed specific miRNAs (**Table 6, Supplementary Table 6**). Note that all 7 SNPs identified in dbSNP build130, were present in the USDA data.

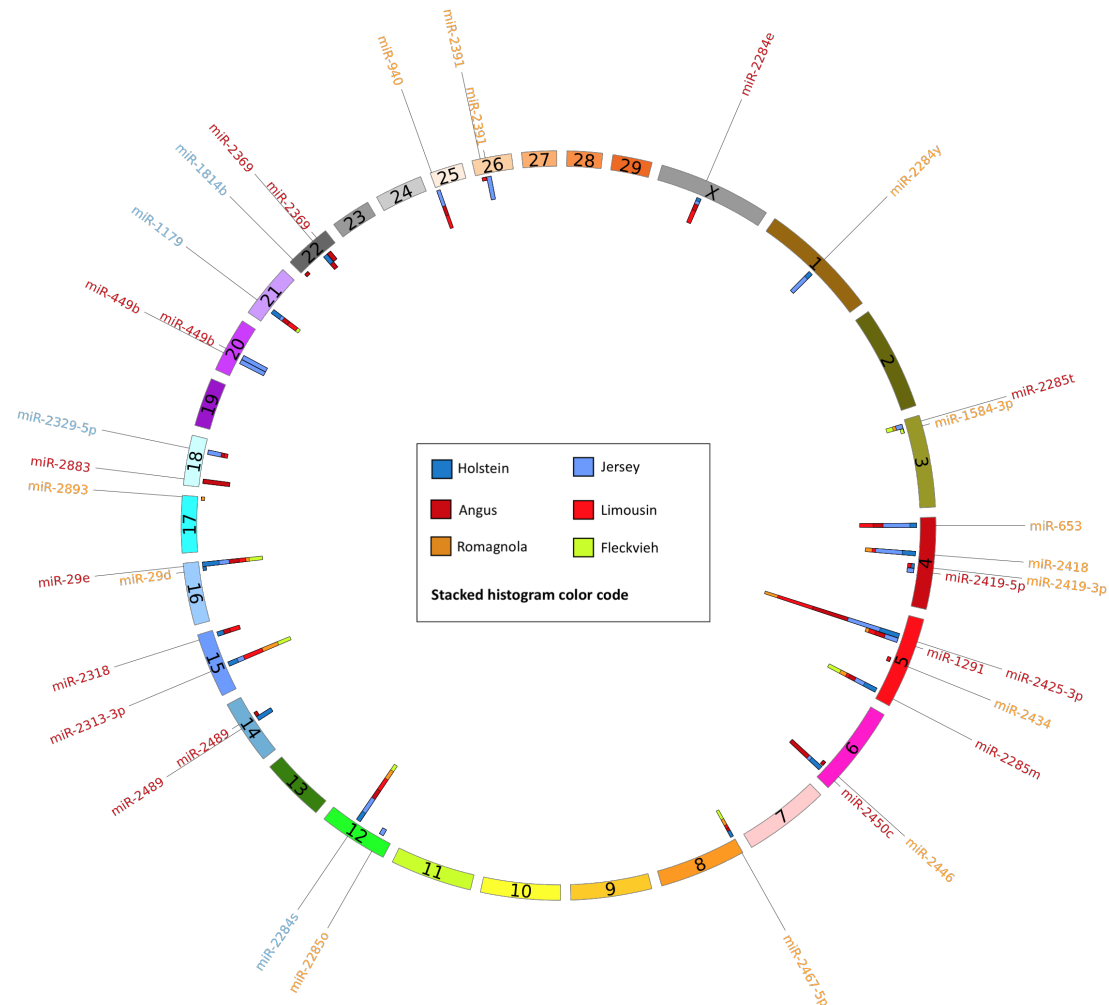


Figure 19 – Distribution of polymorphic miRNA genes in genome. Representation of the polymorphic miRNAs through cattle genome (UMD3.1) and their SNP coverage based on USDA cattle whole genome sequencing. MiRNAs with SNPs in their seed regions are labelled in red, those with SNPs in the mature sequence but not in seed region are labelled in orange and those having SNP (from dbSNP) in the pre-miRNA but neither in the seed region nor in the mature sequence are labelled in blue.

Table 6 – List of polymorphism in miRNA genes and their location in mature and seed region, in six cattle breeds

miRNA name	Breeds	number of SNPs	location in miRNA	SNPs chr:loci	Reference allele	Variant allele
mir-1291	LM, AN, FL, JE	1	seed	5:31283924	C	T
mir-1584	FL, JE, RO	1	mature	3:9881823	G	A
mir-2284e	LM, AN, HO, JE	1	seed	X:61016918	C	T
mir-2284y	HO, JE	1	mature	1:83105454	G	A
mir-2285m	FL, AN, HO,RO,	1	seed	5:111012132	C	T
mir-2285o	JERSEY	1	mature	12:16533036	G	T
mir-2285t	FL, JE, RO	1	seed	3:7929201	C	T
mir-2313	LM, HO, FL,JE,RO	1	seed	15:34628957	C	T
mir-2318	LM, AN, HO	1	seed	15:78356674	A	C
mir-2369	HO, AN	3	mature	22:40172551	C	T
			seed	22:40172555	A	G
				22:40172560	A	C
mir-2391	LM	1	mature	26:10964348	A	G
	JE	1	mature	26:10964356	C	G
mir-2419	LM, HO, RO	1	mature	4:66346486	T	C
	JE	1	seed	4:66346535	C	T
mir-2425	LM, AN, HO, JE,	1	seed	5:30251690	G	C
mir-2434	LM	1	mature	5:64962222	C	A
mir-2446	AN	1	mature	6:99976591	C	T
mir-2450c	HO, AN, JE	1	seed	6:107741122	G	A
mir-2467	FL, AN, HO,RO,	1	mature	8:1172076	A	G
mir-2489	HO	1	seed	14:48094694	C	T
	AN, HO	1	seed	14:48094695	A	G
mir-2883	AN,HO,RO,LM	1	seed	18:8166103	C	T
	LM	1	seed	18:8166104	G	A
mir-2893	RO	1	mature	17:73346081	G	A
mir-29d	HO	1	mature	16:77562609	A	G
mir-29e	FL, AN, HO, RO,	1	seed	16:77562055	T	C
mir-449b	JE	2	mature	20:23967292	G	T
mir-653	LM, AN, HO, JE	1	mature	4:10653589	A	G
mir-940	LM, AN, JE	1	mature	25:1792942	C	T

List of miRNAs containing SNPs located in mature miRNA or seed region and the related breeds. HO: Holstein, JE: Jersey, AN: Angus, LM: Limousin, FL: Fleckvieh, RO: Romagnola.

Interestingly, the SNPs present in some of the seed regions of cattle miRNA corresponded to other reference miRNA seed region in cattle and other species (**Table 7**). Of the 16 miRNAs with SNPs in their seed regions, eight of them have identical seed regions to other miRNA genes, suggesting a potential conservation of function. For example, the SNP in *mir-29e* modified the reference seed region into the seed region of the reference *mir-29a*, *mir-29b*, *mir-29c* and *mir-29d* across 12 species including cattle. The polymorphisms in miRNA genes are present in cattle and more precisely are different among breeds, which could suggest a link between selection of animals and variation in miRNA. Moreover, correlation of variation in the seed regions with other miRNA suggested that these polymorphic miRNAs could be still functional and therefore modify the related function.

Table 7 – List of miRNAs having SNPs creating a seed region identical to another miRNA

Polymorphic miRNA ID	breeds	miRNA with identical seed	species
miR-2489	HO	miR-2372	bta
miR-2284y	JE HO	miR-2284g miR-	bta
miR-2318	HO AN LM	miR-675-3p	mmu
miR-29e	AN HO LM FL	miR-29c miR-29b	eca hsa xtr mml bta cfa
miR-2313-3p	RO JE FL LM HO	miR-4277 miR-3600	hsa bta
miR-2434	LM	miR-3602	bta
miR-940	JE AN LM	miR-664 miR-456	bta gga

List of polymorphic *B. taurus* miRNA, which SNPs in the seed region creating an identical seed region to another miRNA in cattle and other species referenced in TargetScan. Each miRNA mentioned are found in all species listed except for miR-3600 and miR-644 corresponding to bta, miR-456 corresponding to gga and miR-4277 corresponding to hsa; eca: *Equus caballus*; has: *Homo sapiens*; xtr: *Xenopus tropicalis*; mml: *Macaca mulatta*; bta: *Bos taurus*; cfa: *Canis familiaris*; mmu: *Mus musculus*; mdo: *Monodelphis domestica*; ptr: *Pan troglodytes*; gga: *Gallus gallus*; rno: *Rattus norvegicus*; oan: *Ornithorhynchus anatinus*.

3.3.3 Pre-miRNA secondary structure modifications in variant miRNA

Pre-miRNAs with polymorphisms in their sequences had their secondary structure predicted with RNAfold and the corresponding MFEs were compared. The average MFE for the variant pre-miRNAs ranged from $-9.4 \text{ kcal.mol}^{-1}$ to $-51 \text{ kcal.mol}^{-1}$ while the average MFE for the reference pre-miRNAs ranged from $-11.3 \text{ kcal.mol}^{-1}$ to $-52.1 \text{ kcal.mol}^{-1}$ (**Supplementary table 6** – List of polymorphism in miRNA genes and their location for hairpin, mature and seed region, in six cattle breeds

Supplementary table 7). The comparison of MFE between the reference miRNA and the reconstructed variant for the different breeds were carried out by calculating the absolute ΔMFE for each pair. 21 unique miRNA gene variants out of 89 screened, showed an increase in absolute MFE value of more than 1 kcal.mol^{-1} . In contrast, 42 miRNA gene variants showed a reduction in MFE by more than 1 kcal.mol^{-1} (**Figure 20**). Among these, three miRNAs (mir-2489, mir-2467, mir-2419) had ΔMFE values which could be either positive or negative depending on the SNP found in each breed. In the case of mir-2467 and mir-2419, both showed an opposite value of ΔMFE , respectively positive (+1,8) and negative (-2.6) in the variant form found in the genome of Jersey cattle compared to those of the other breeds. These two miRNAs had SNP specific to Jersey breeds, mir-2467 had one SNP in its hairpin region while all other breeds tested had two, and mir-2419 had a SNP specific to Jersey. Finally, 14% of the miRNAs hairpin variant (13/92) had a $\Delta\text{MFE} > 5$ (6/13) or $\Delta\text{MFE} < -5$ (7/13). A study on SNP in miRNA suggested that a difference of 2 kcal.mol^{-1} could affect the processing of the pre-miRNA (Gong, 2012). This show that the polymorphisms modified the stability of the hairpin structure of the pre-miRNA and could impair its maturation process even with slight changes in stability.

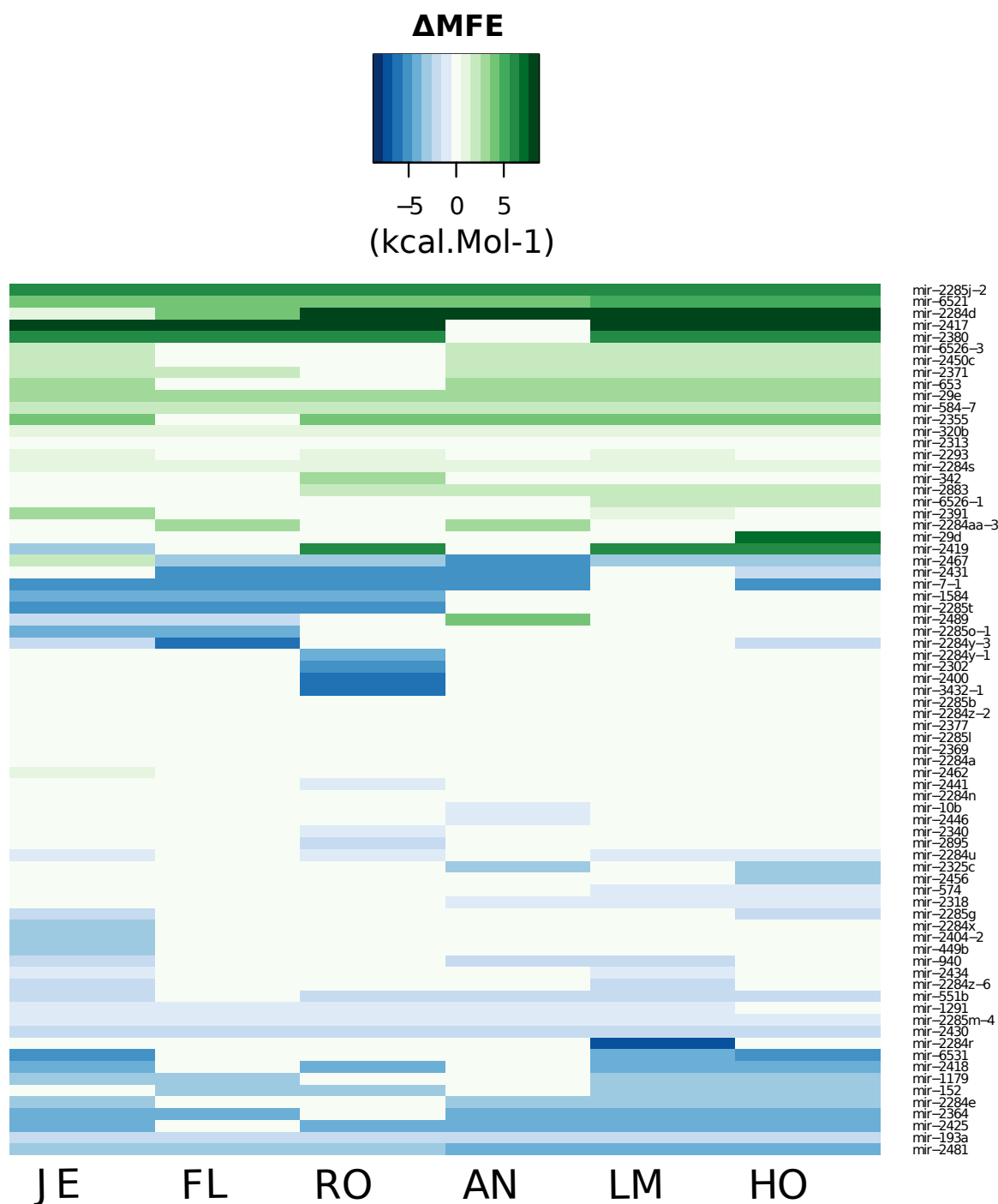


Figure 20 – Heatmap of the impact of polymorphism in pre-miRNA on their thermodynamic stability

The heatmap represents all polymorphic pre-miRNA in each breed with a difference of minimum free energy (ΔMFE). This value represents the difference of absolute MFE between the reference and the variant pre-miRNA. Breed abbreviation: JE: Jersey, FL: Fleckvieh, RO: Romagnola, AN: Angus, LM: Limousin, HO: Holstein.

The predicted hairpin structures of 50 variant miRNAs differed from the reference one. It was noticed loss or gains of internal loops and bulges (with modification of the size of these elements) which could alter the maturation of the miRNAs. For example, *mir-2456* and *mir-2895* are the two miRNAs which underwent the most modification in their hairpin loop structure due to SNPs (**Figure 21**). These two variant structures prediction present extra hairpin loops included in the stem, which could interfere with the processing of these two miRNAs. Indeed a study showed that secondary structure and thermodynamic stability changes in small hairpin RNA processed by Dicer have an impact for their biogenesis and thus their repressive potential (Krol, 2004). Another study on pre-miRNA structures showed that mismatches affected the pre-processing by Dicer. The presence of bulges or internal loops in pre-miRNA appeared to increase the cutting precision of Dicer reducing the amount of non-canonical miRNA and off-set targeting, following the “loop counting rule” (S. Gu, 2012). Finally, two studies showed the impact of SNPs in miRNA hairpin loops in the processing of miRNAs creating off-targets and modifying the regulated functions (Sun, 2009; Xiong, 2013). Our results showed that the polymorphisms in cattle pre-miRNA can modify the structure and the thermodynamic stability, potentially impairing the targeting function of these miRNAs.

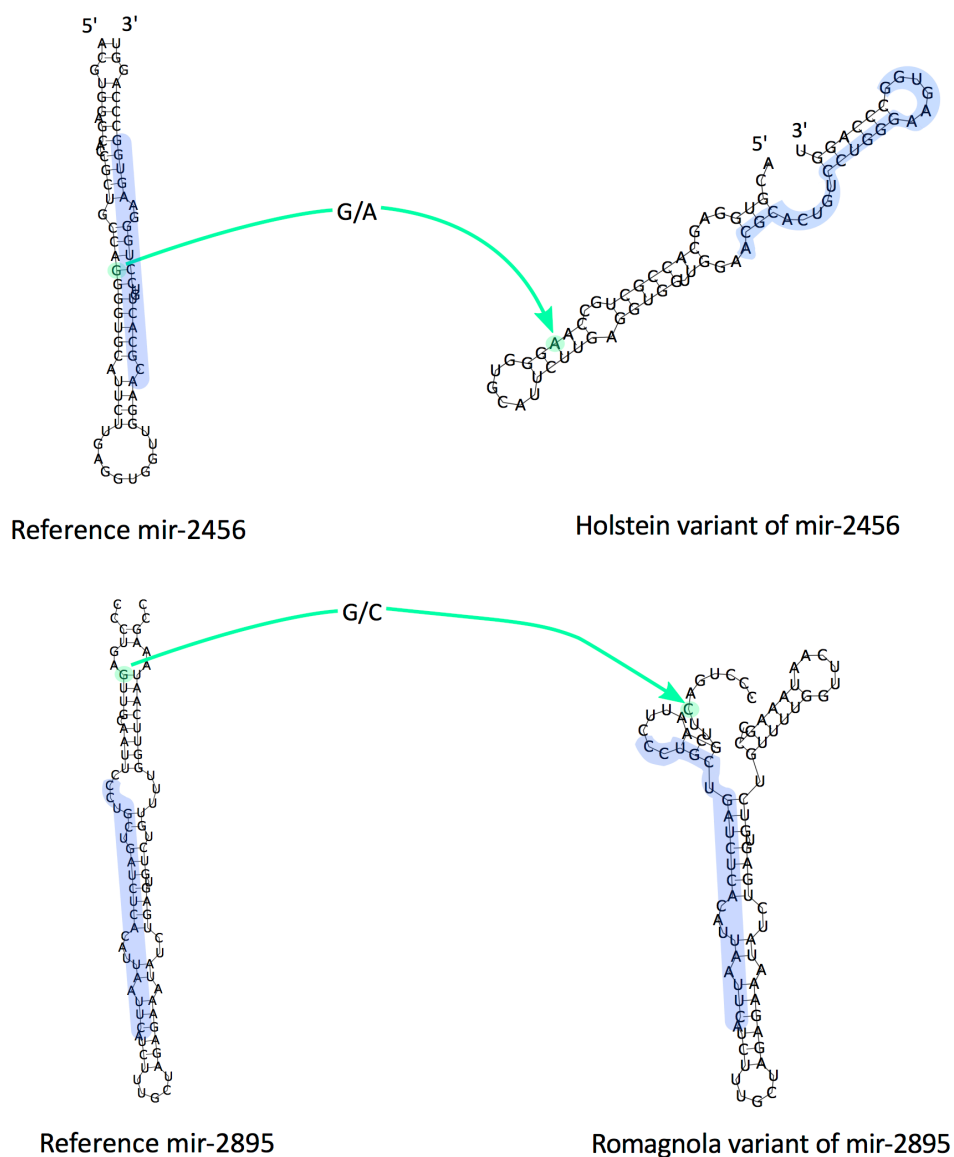


Figure 21 – Secondary structure of *mir-2456* and *mir-2895* is modified in their variants. Based on the RNAfold analysis, it has found that both *mir-2456* and *mir-2895* had a hairpin loop structure which is highly disrupted, leading to the possibility that their functions are similarly perturbed. The polymorphic bases are highlighted in pale green, and the mature miRNA in pale blue (as defined in miRBase).

3.3.4 Identification of new target sites for polymorphic mature miRNA

To further investigate the role of polymorphisms in miRNA, the SNPs located in mature miRNA sequences were examined, focusing on those which had the potential to modify the targeted genes and thereby disrupt their biological function. To identify these changes of targets, the TargetScan algorithm was used to predict the targets of the reference and of the corresponding polymorphic miRNAs and to calculate the scores associated with. After filtering based on the score, a total of 12,025 unique sites, which were newly targeted by 23 polymorphic miRNAs and 7117 unique sites from which targeting is lost due to polymorphisms in 21 miRNAs have been identified (**Supplementary table 8**). Among the polymorphic miRNAs, 10 mature miRNAs (miR-1291, miR-2489, miR-29e, miR-2369, miR-2883, miR-2450c, miR-2285m, miR-2419-5p, miR-2284e, miR-2425-3p) had no or few common target sites with their reference one (**Table 8**). This meant that the SNPs in these miRNAs totally switched their binding sites in comparison to those targeted by the corresponding reference miRNAs.

Table 8 – List of polymorphic miRNA identified in cattle breeds and their predicted binding sites with no common target with the reference

1 - Gain				
(breed)miRNA ID	PS	CS	score sum	SSr
HO-miR-2883	1400	0	-164.25	-229943
AN-miR-2883	1398	0	-164.01	-229281.79
RO-miR-2883	1398	0	-164	-229267.81
LM-miR-2285m	1346	0	-141.07	-189876.18
HO-miR-2285m	1344	0	-140.9	-189364.22
FL-miR-2285m	1343	0	-140.87	-189183.04
RO-miR-2285m	1343	0	-140.87	-189183.04
AN-miR-2285m	1343	0	-140.86	-189177.67
JE-miR-2285m	1342	0	-140.63	-188728.14
JE-miR-2419-5p	1240	0	-137.99	-171103.88
RO-miR-1291	960	0	-170.57	-163751.04
JE-miR-1291	959	0	-170.57	-163574.71
AN-miR-1291	960	0	-170.28	-163469.76
LM-miR-1291	958	0	-170	-162860
HO-miR-29e	909	1	-170.07	-155256.26
RO-miR-29e	910	0	-170.32	-154992.11
FL-miR-29e	910	0	-170.32	-154991.2
JE-miR-29e	909	1	-170.27	-154772.7
AN-miR-29e	909	0	-170.22	-154733.62
LM-miR-29e	909	1	-170.13	-154644.53
AN-miR-2450c	781	1	-100.52	-78509.24
HO-miR-2450c	780	0	-100.24	-78186.42
JE-miR-2450c	779	0	-100.15	-78019.97
LM-miR-2883	883	0	-88.1	-77792.3
LM-miR-2284e	764	0	-95.64	-73072.78
AN-miR-2284e	762	0	-95.44	-72722.99
JE-miR-2284e	761	0	-95.29	-72516.45
HO-miR-2284e	761	0	-95.29	-72513.41
AN-miR-2425-3p	638	0	-98.32	-62725.61
RO-miR-2425-3p	638	0	-98.31	-62719.87
JE-miR-2425-3p	638	0	-98.3	-62717.31
LM-miR-2425-3p	637	0	-98.27	-62596.72
HO-miR-2425-3p	637	0	-98.26	-62593.53
HO-miR-2489	790	0	-76.58	-60501.36
HO-miR-2369	578	0	-82.16	-47486.17
AN-miR-2369	578	0	-81.92	-47346.87

2 - Lost

(breed)miRNA ID	PS	CS	score sum	SSr
AN-miR-1291	1243	0	-262.73	-326570.9
JE-miR-1291	1243	0	-262.73	-326570.9
LM-miR-1291	1243	0	-262.73	-326570.9
RO-miR-1291	1243	0	-262.73	-326570.9
HO-miR-2489	870	0	-109.55	-95307.63
AN-miR-29e	755	0	-122.28	-92322.16
FL-miR-29e	755	0	-122.28	-92322.16
RO-miR-29e	755	0	-122.28	-92322.16
JE-miR-29e	754	1	-122.21	-92145.59
LM-miR-29e	754	1	-122.21	-92145.59
HO-miR-29e	754	1	-122.21	-92145.59
AN-miR-2369	445	0	-64.33	-28625.52
HO-miR-2369	445	0	-64.33	-28625.52
AN-miR-2883	249	0	-56.65	-14106.6
HO-miR-2883	249	0	-56.65	-14106.6
RO-miR-2883	249	0	-56.65	-14106.6
LM-miR-2883	248	0	-56.48	-14007.78
HO-miR-2450c	214	0	-46.5	-9949.93
JE-miR-2450c	214	0	-46.5	-9949.93
AN-miR-2285m	225	0	-43.9	-9876.6
FL-miR-2285m	225	0	-43.9	-9876.6
HO-miR-2285m	225	0	-43.9	-9876.6
JE-miR-2285m	225	0	-43.9	-9876.6
LM-miR-2285m	225	0	-43.9	-9876.6
RO-miR-2285m	225	0	-43.9	-9876.6
AN-miR-2450c	213	1	-46.33	-9867.23
JE-miR-2419-5p	103	0	-25.99	-2677.07
AN-miR-2284e	103	0	-23.01	-2370.03
HO-miR-2284e	103	0	-23.01	-2370.03
JE-miR-2284e	103	0	-23.01	-2370.03
LM-miR-2284e	103	0	-23.01	-2370.03
AN-miR-2425-3p	73	0	-18.76	-1369.63
HO-miR-2425-3p	73	0	-18.76	-1369.63
JE-miR-2425-3p	73	0	-18.76	-1369.63
LM-miR-2425-3p	73	0	-18.76	-1369.63
RO-miR-2425-3p	73	0	-18.76	-1369.63

1- List of ranked miRNA variant which gained sites. **2-** List of the ranked miRNA variant which lost sites. PS: number of polymorphic sites; CS: number of common sites; score sum: sum of the TargetScan score of the PS; SSr: Site Score ratio used for ranking the miRNAs.

Moreover, these ten miRNAs showed few common gene targets between the variant and the reference alleles (**Figure 22**). For example, *mir-2489* had two different mature miRNA variants for Holstein-Friesian and Angus while the Holstein variant had a similar number of targets compare with the reference and the Angus variant showed a reduction in the number of targeted genes. Furthermore, the miRNAs were ranked following the SSr score described in the material and method. According to the SSr ranking, the top polymorphic miRNA which gained targets was mir-2883 and the top one for loss of targets was mir-1291. This analysis confirmed that polymorphisms in miRNA had a strong impact on the number of potential targeted genes (Gong, 2012; Jevsinek Skok, 2013). Moreover, SNPs for a given miRNA showed differences between the investigated breeds, potentially leading to specific gene expression and modification of biological pathways regulation linked with selection and breeding.

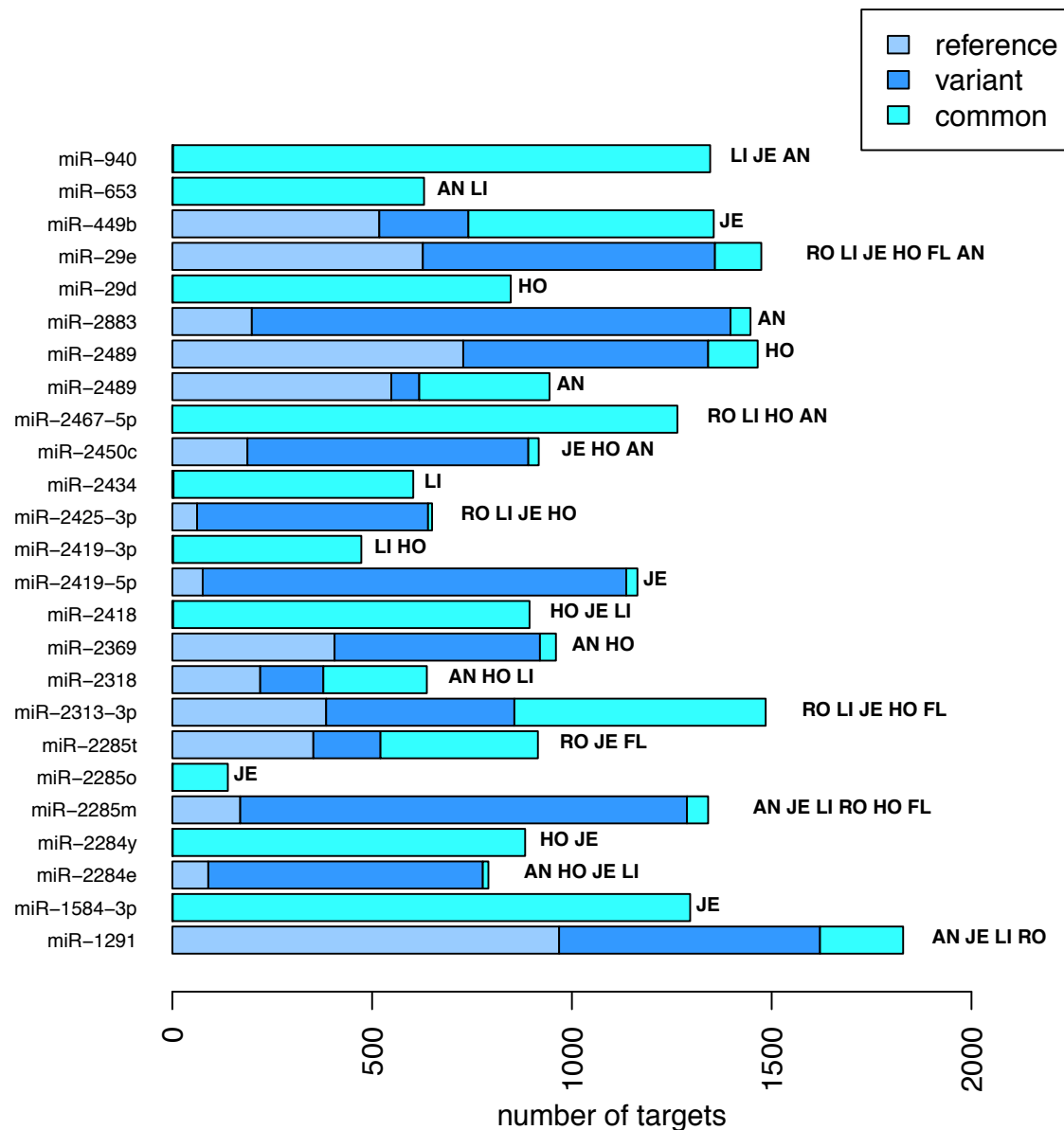


Figure 22 – Comparison of the number of targeted genes between the reference and variant miRNAs. The stacked histogram represents the number of genes targeted for each miRNA. The name of the breeds to which the variant is related is mention at the end of each bar. The light blue square represents the genes targeted in both variant and reference miRNAs when the medium blue represents the number of targets for the reference allele and the dark blue represents the targets for the variant allele. In the case of miR-2489, the two variant Angus (AN) and Holstein (HO) are different, hence represented separately. The miRNAs showing mainly common targets (light blue bar) are miRNAs having variation outside of their seed region therefore do not present any difference of targets based on TargetScan algorithm.

3.3.5 Pathways enriched for miRNA targets are different between reference and variant miRNAs.

To identify which function could be affected by the change of targets due to SNPs in mature region of miRNAs, the potential enrichment of these targets in biological pathways was investigated. To identify the pathways targeted by different miRNAs, Ingenuity Pathway Analysis (IPA) software was used and the significantly enriched pathways, corrected with Benjamini-Hochberg method (adjusted p-value < 0.01), were compared between the reference and the variant. Of the 25 miRNAs, which displayed SNPs in their mature sequences, nine polymorphic miRNAs showed differences in the number of pathways enriched in targets in comparison with their corresponding reference.

Only miRNA having SNPs in their seed region showed differential enrichment pathways for their targets. To investigate the extent of modification due to miRNA polymorphisms, the pathways targeted by a given reference miRNA were compared to the pathways targeted by its variant. In total 39 pathways were specifically enriched in targets of miRNA variant and 132 pathways were specifically enriched in targets of reference miRNAs (**Supplementary table 9**). Thirteen pathways were significantly enriched only for targets of the variant miRNAs and 73 pathways were significantly enriched for targets of the reference miRNAs. Moreover, among the 73 pathways enriched only in reference miRNA targets, 21 were targeted by more than one miRNA. By applying the score system used to rank the polymorphic miRNAs, a score has been associated to each pathway enriched corresponding of the average score of all variant binding sites of genes involved in each pathway.

Among the miRNA differentially targeted pathways, the variants for miR-449b and miR-29e were the most represented. They respectively lost target enrichment for 30 and 40 unique pathways out of the 73 pathways enriched only in reference miRNA targets. These pathways corresponded to 135 targets for miR-449b and 82 targets for miR-29e out of 373 unique genes differentially targeted by a unique variant miRNA and involved in significantly enriched pathways (**Figure 23**).

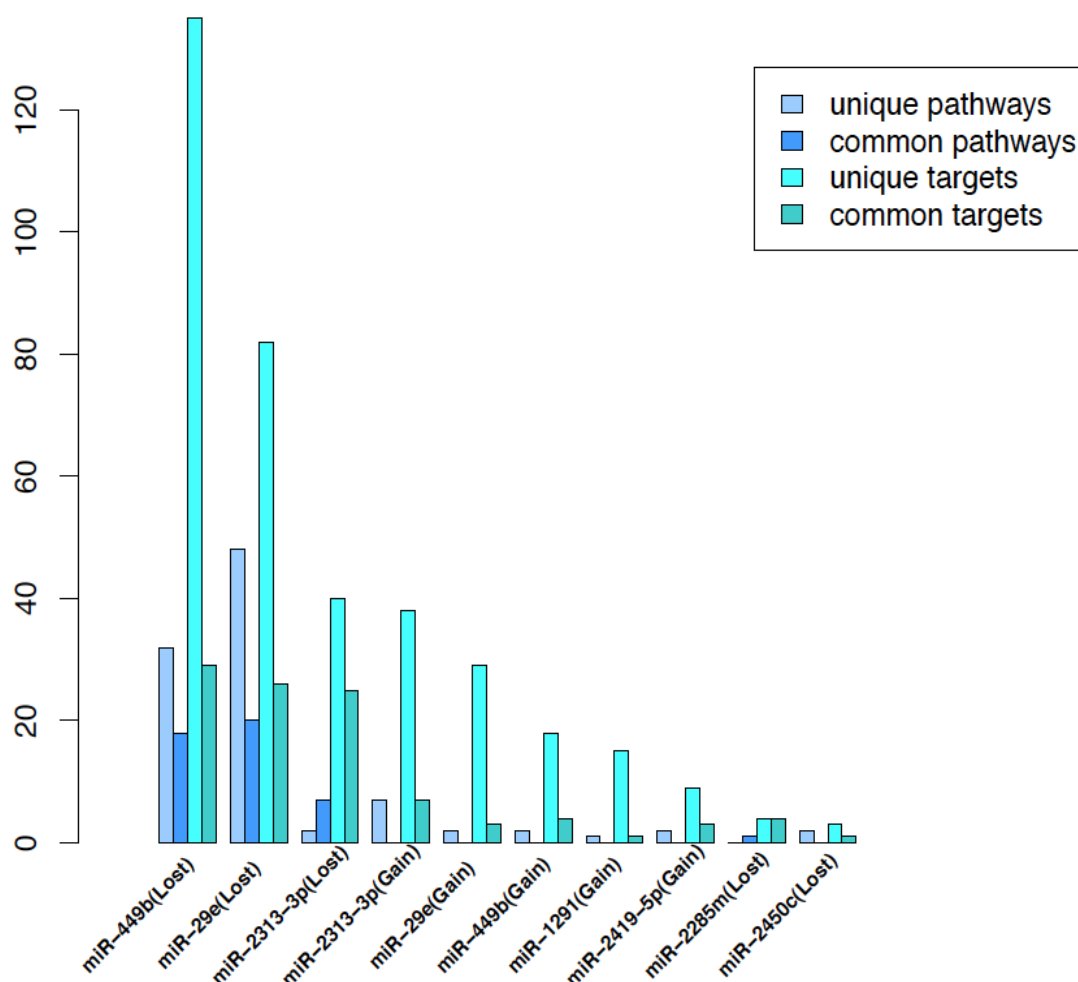


Figure 23 – Number of differentially enriched pathways in polymorphic miRNAs and number of their related targets showed a higher rate of lost pathways enrichment and lost targets. Representation of the variant miRNAs with significantly enriched pathways differing from those of their reference miRNAs. “Gain” represented the pathways enriched for gained targets and “Lost” the pathways enriched for lost targets. Unique pathways represented pathways significantly enriched only by one specific variant miRNA targets and common represent pathway enriched by more than one variant miRNAs targets.

Moreover, the variants of the miR-449b/miR-29e pair lost target enrichment for 15 pathways out of the 21 differentially targeted by more than one variant miRNA. Similarly, the miR-2313-3p variant, pairing with either miR-29e or miR-449b, lost targets for 7 pathways. Among the 15 pathways enriched in targets for miR-29e and miR-449b pair, four pathways may be of interest to the dairy breed, which were present in both mir-449b (Jersey specific) and mir-29e variants. These four pathways were, in score ranking order: Interleukin-17 (IL-17) Signalling, ErbB Signalling, p70S6K Signalling and Prolactin Signalling pathways.

Out of all pathways significantly enriched, IL-17 had the highest score for miR-29e targets and these targets represented more than 10% of the total molecules involved in this pathway (14.8% for miR-449b targets). Based on the representation of the pathway of IL-17 Signalling, it was observed that lost targets of miR-29e and miR-449b variant did not totally overlap, four lost targets were specific to miR-29e and five were specific to miR-449b variant allele, which could suggest a complementarity of these two miRNAs (**Figure 24**). The IL-17 is a family of homodimeric glycoprotein, conserved through species and produced by T lymphocyte involved in the inflammatory response (Moseley, 2003). IL-17 has a role in immune response and has been shown (as well as Interleukin 6, also present in the ranking) to be highly expressed in milk somatic cell during mastitis, induced by *S. aureus* in cattle and correlated to milk yield (Tao, 2007). Another study in dry-off dairy cow, a sensitive period for infection, investigated the gene expression level during *Streptococcus* infection leading to mastitis. They observed an increase of IL-17 expression supporting the inflammatory role of this protein during mastitis (Bruno, 2010). The role of IL-17 in defence against *E. coli* intrammary infection has been investigated. It was observed an increase of expression of IL-17 inducing a pro-inflammatory and anti-microbial response during mastitis (Roussel, 2015). Finally, another study on mastitis in dairy cow integrated GWAS with transcriptomic study aiming to identify pathways directly linked to the mastitis immune response. They found IL-17 signalling pathway as one of the highest ranked, linked with an eQTL related to cell movement and cell-to-cell signalling (Lewandowska-Sabat, 2012). This showed that the control of expression of the IL-17 pathway is important during mastitis in dairy cow, therefore highlighting the role of *mir-29e* and *mir-449b* as potential regulator of this pathway.

ErbBs were part of a receptor tyrosine kinase, and their signalling pathway is involved in organogenesis of heart and neural cell, muscular neurosynapses and

epithelial morphogenesis (Burden, 1997; Yarden, 2001). The different ErbBs form heterodimers acting as receptors for neuregulin, a member of the epithelial growth factor molecules (EGF). The ErbBs have been shown to be involved in mammary gland development each one at different stage of development (Schroeder, 1998). The EGF receptors have also been identified in cattle mammary tissue and lactating glands (Spitzer, 1987; Sheffield, 1998).

The p70 ribosomal protein kinase was a major substrate of mammalian target of rapamycin (mTOR) involved in cell proliferation and growth (Dann, 2007; Shin, 2011). Myostatin has been shown to inhibit the Akt/mTOR/p70S6K protein synthesis pathway, which mediates both differentiation in myoblasts and hypertrophy in myotubes (Trendelenburg, 2009). Moreover mTOR/p70S6K have been shown to be down-modulated by IL-6, modifying the myogenic differentiation (Pelosi, 2014). Moreover ErbB2, involved in breast cancer through angiogenesis, has been shown to increase the vascular endothelial growth factor production by activating p70S6K (Klos, 2006).

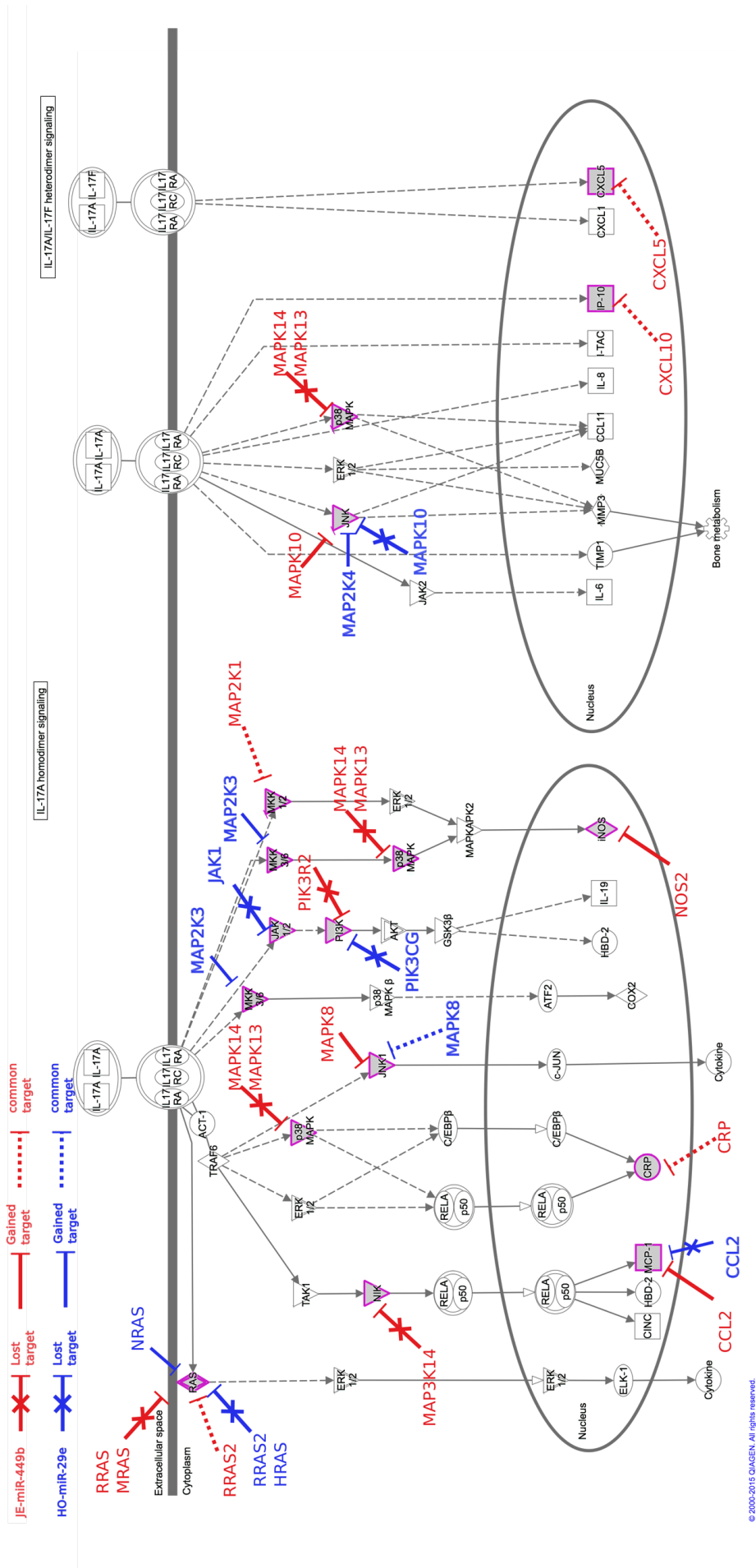


Figure 24 – The miRNAs mir-449b and mir-29e variant targets differently IL-7 signalling pathway. The pathway for Interleukin 17 signalling was significantly enriched in lost targets for the variant haplotype of *mir-449b* and *mir-29e*. The crossed line corresponds to the gene targeted by the reference allele, the dashed line corresponds to genes targeted by both alleles and the straight line corresponds to the genes targeted by the variant allele (Adapted from IPA graph).

Finally, the Prolactin pathway is also enriched for lost targets for miRNA-449b and miR-29e (**Figure 25**), many of which are shared with the IL-17, ErBb and P70S6k pathways. Prolactin had an important role in mammary gland development and differentiation and had been reported to play a role in the maintenance of mammary gland function (Forsyth, 1986; Accorsi, 2002). It should be noted that the role of Prolactin in lactogenesis, is not as important as that of growth hormone in ruminants compared to other mammals like rodents (Knight, 1986; Madon, 1986), but it is still likely to play a role in this process. Indeed, a study on photoperiod length related to prolactin and milk yield in dairy cows, suggested a link between the sensitivity and responsiveness to prolactin during transition to lactation and increase of the subsequent milk yield (Auchtung, 2005). It has also been shown that prolactin presence had an effect on fatty acid synthesis, mediated by leptin in mammary glands of dairy cattle and also increased the level of expression of α -casein and β -lactoglobulin (Feuermann, 2004).

The different genes targeted by the *mir-29e* and *mir-449b* variants when compared to the reference form showed that important pathways for dairy cattle are enriched for these targets. Moreover, the fact that the *mir-449b* variant was specific to the Jersey breed suggests that these SNPs could be good candidates for selection during breeding for related traits. Furthermore, the fact that both variants were targeting the same pathways, strengthened the potential relaxation of the repression effect of these pathways *in vivo*.

As shown in the representation of the IL-17 and Prolactin pathways, the enrichment is significant only for one allele (reference) of miRNAs. However, some targets are unique or common to the other allele (variant) but do not present a significant enrichment. For example, in the prolactin signalling pathway, the reference allele of miR-449b targets 11 genes (significantly enriched; $p\text{-value} < 0.01$) when the variant one of miR-449b targets only six genes from which five targets were common to both alleles. In the latter case of the variant, the $p\text{-value}$ for the enrichment showed that it is not significant.

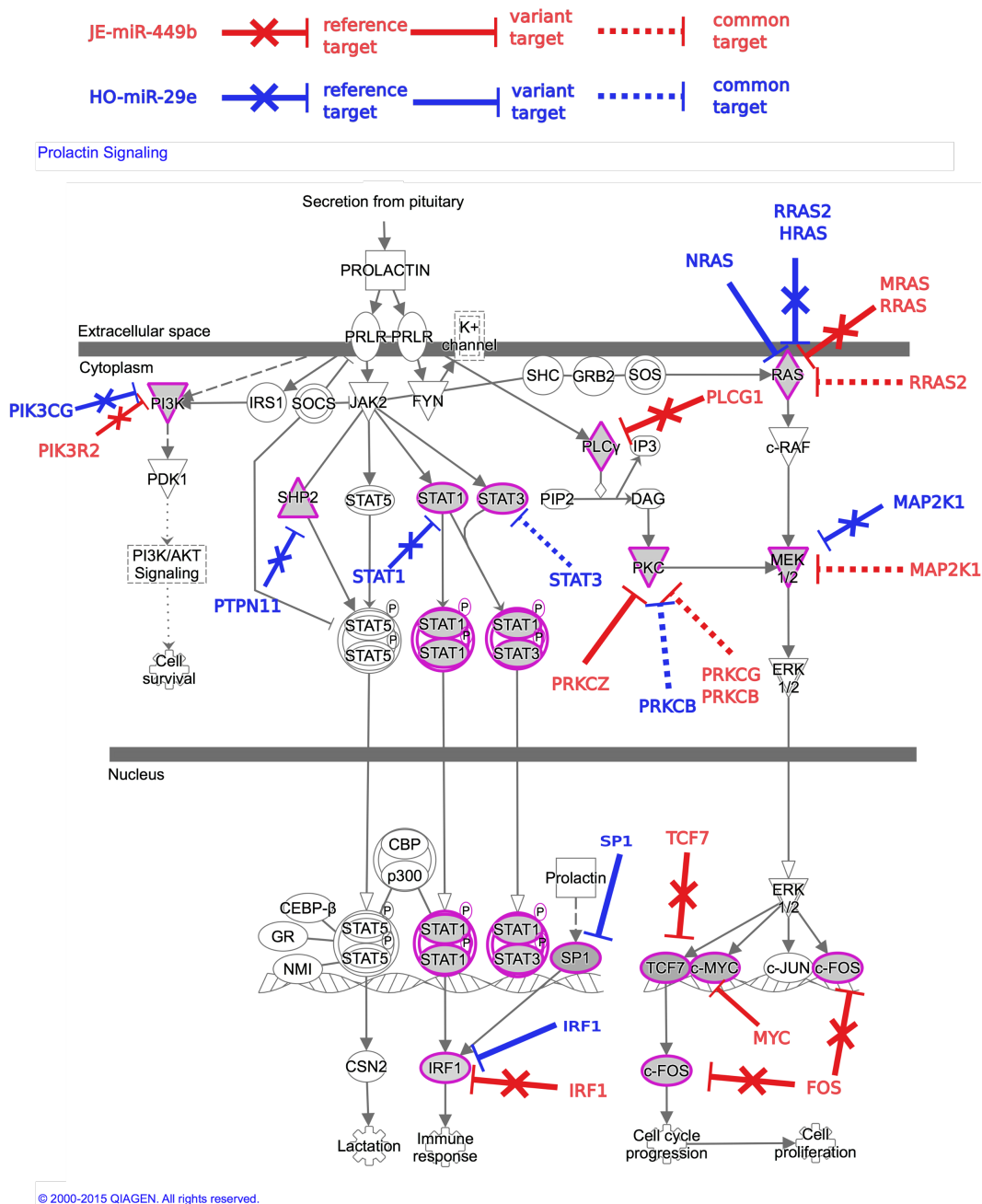


Figure 25 – The miRNAs mir-449b and mir-29e variants targets differently Prolactin signalling pathway. The pathway for Prolactin signalling is significantly enriched in lost targets for the variant haplotype of *mir-449b* and *mir-29e*. The crossed line corresponded to the gene targeted by the reference haplotype, the dashed line corresponded to genes targeted by both haplotype and the straight line corresponded to the genes targeted by the variant haplotype.

As described previously the pathways differently enriched for multiple variant miRNAs varied, and in most cases only few targets were shared between the reference

and the variant miRNA. However, 18 pathways were identified as enriched in targets of one specific miRNA allele (either the reference or variant form) but did not contained any targets at all for the other allele form. Out of these 18 pathways, five pathways were enriched in variant targets and 13 were enriched for the reference targets (**Table 9**). It was found that 12 pathways were targeted by miR-29e and one by miR-449b, out of the 13 pathways only targeted by the reference alleles. Among these pathways, the presence of Dopamine Receptor Signalling pathway was particularly interesting. This pathway was targeted by the two miRNAs (miR-29e and miR-449b) and hence was included into the “Pathways specific for multiple miRNAs” group. However, only the miR-29e reference allele uniquely targeted this pathway while both alleles of miR-449b had targets in the pathway with only the reference allele showing enrichment.

It was also interesting to note for these “unique-allele type” targeted pathways, the presence of significant miRNA variant allele targeted enrichments, which were absent from the pathways specific for multiple miRNAs. The two miRNAs variant alleles corresponded to miR-2419-5p, which targeted four pathways (Role of JAK family kinase in IL-6-type Cytokine Signalling, Inhibition of Angiogenesis by TSP1, Production of Nitric Oxide and Reactive Oxygen Species in Macrophages, S-adenosyl-L-methionine Biosynthesis) and miR-1291, which targeted one pathway (Actin Nucleation by ARP-WASP Complex). It was also found that the miR-2419-5p variant was also specific to the Jersey breed.

The pathway enrichment assisted in understanding the function potentially dysregulated by polymorphisms in miRNA genes. It also helped to refine the selected targets to important ones for breed selection. The main findings from the pathway enrichment analysis, is the miRNA pair miR-449b and miR-29e variants. Together they differentially target genes involved in four pathways related to mammary gland maintenance (IL-17 pathway) and development (ErBb, Prolactin and P70S6K pathways). Moreover, considering that the allele of miR-449b was identified in a dairy breed, Jersey, this is supporting a functional role of *mir-29e* and *mir-449b* toward milk production traits and their variant as important markers for dairy genomic selection.

Table 9 – Pathways enriched in unique miRNA variant targets List of pathways targeted uniquely by one miRNA allele, the group are as described in the method: PSS: Pathway specific for single miRNA; PSM: Pathways specific for multiple miRNAs; GLCP: Gain and Lost common pathways

Pathways	type	miRNA	breeds	Pathway group
Role of JAK family kinases in IL-6-type Cytokine Signalling	Gain	miR-2419-5p	JE	GLCP
Ovarian Cancer Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Role of JAK1 and JAK3 in yc Cytokine Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
G-Protein Coupled Receptor Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Inhibition of Angiogenesis by TSP1	Gain	miR-2419-5p	JE	PSS
Telomerase Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Chronic Myeloid Leukaemia Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Gai Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Angiopoietin Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Role of JAK2 in Hormone-like Cytokine Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
PEDF Signalling	Lost	miR-449b	JE	PSS
Histidine Degradation VI	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
IL-15 Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
cAMP-mediated signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSS
Dopamine Receptor Signalling	Lost	miR-29e	JE, RO, FL, HO, AN, LM	PSM
Actin Nucleation by ARP-WASP Complex	Gain	miR-1291	RO, JE, LI, AN	PSS
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	Gain	miR-2419-5p	JE	GLCP
S-adenosyl-L-methionine Biosynthesis	Gain	miR-2419-5p	JE	PSS

3.3.6 QTL enrichment in polymorphic miRNA targets suggested influence of miRNA in cattle traits

In addition to pathway enrichment analysis, aimed to identify potential functions, which could be deregulated, Quantitative Trait Locus (QTL) enrichments were also investigated to identify the potential effects of miRNA polymorphism in cattle traits. For that purpose, the data for QTL and the associated traits provided by CattleQTLdb (www.animalgenome.org) were retrieved, counting a total of 9,180 QTLs (April, 2014). The enrichment was conducted using the hypergeometric test, corrected by Benjamini-Hochberg methods ($p\text{-value} < 0.01$). Among the 32 miRNAs which had SNPs in their mature sequence, 12 had differentially target gene enrichments for QTLs between the reference and the variants; these 12 miRNAs were miRNAs having SNPs in their seed regions. In total 56 QTLs were enriched by miRNA variants and reference haplotype target genes. Among these QTLs, 21 are enriched only in targets gained by variant miRNA and 26 were enriched in targets lost by variant miRNAs (only targeted by the reference haplotype). 14 QTLs, which were uniquely enriched for variants miRNA targets and 15 QTLs which were uniquely enriched with genes not targeted anymore by variant miRNAs, were also identified (**Supplementary table 10**). Moreover, three QTLs (Spleen percentage, Blood creatinine level and Meat colour L*) were enriched only in gained targets for more than one variant miRNA (**Table 10**).

In the latter set of QTLs, two traits, spleen percentage and meat colour L* (lightness) were significantly enriched for the same group of miRNA variants (miR-2425-3p, miR-2284e). Although these two QTLs showed significant enrichment, only one gene was located in this QTL: *PITPNB* (phosphoinositol transfer protein- β : chr17 69488675-69640595), explaining the corrected $p\text{-value}$ of 0.

The spleen percentage QTL was identified in the Jersey and Limousin back-cross study (McCabe, 2012) and reported in QTLdb, however it was mentioned in the QTLdb file that this QTL was associated with a weak $p\text{-value}$ ($p\text{-value} < 0.1$).

The locus associated with meat colour L* however, mentioned as LDL (L reflectance parameter on *M.longissimus thoracis et lamborum*) in the study of Esmailizadeh et al. (Esmailizadeh, 2011) on the same animals, was significant ($p\text{-value} < 0.05$) and corresponded to meat brightness. The gene *PITPNB* codes for a protein binding and transferring phosphatidylinositol and phosphatidylcholine between membranes. It has been identified as a potential candidate in a chicken QTL (chromosome 15) associated with fatness (Jennen, 2004) and two splice variants of this

gene have been shown to be highly expressed in heart and liver of chicken in a study on meat quality (Jorge, 2010). It should be noted that the function of the PITPNB protein function is not well characterized and seemed to have different roles and expression pattern in different mammalian species (Venuti, 1988; Alb, 2002; Patel, 2013). The presence of polymorphic miR-2425-3p and miR-2284e binding sites suggested a role in the control of the expression of *PITPNB* on meat quality related to transfer of certain lipids. The blood creatinine level QTL was also enriched for targets of multiple miRNAs (miR-29e and miR-2419-5p). Interestingly, the variant of miR-2419-5p, which was unique to the Jersey breed, is predicted to target *HIBCH*, which was found to be located in a QTL associated to blood creatine level. In addition to pathway enrichment analysis, QTL enrichment therefore suggested additional information to which traits are likely to be influenced by polymorphisms in miRNA genes. Similarly, these polymorphisms could be good candidates for improvement of meat quality traits in cattle.

Table 10 – QTL enriched in multiple miRNA variant targets

Trait related to QTLs	SNP type	unique miRNA	common miRNA	unique breeds: qval mean	common breeds: qval
Spleen percentage	Gain	miR-2425-3p miR-2284e	miR-2883	LM JE HO RO AN: 0.0; HO JE AN LM: 0.0;	AN, HO, RO: 0.0;
Blood creatinine level	Gain	miR-29e miR-2419-5p	miR-29d	AN HO LM FL JE RO: 0.0058; JE: 0.0099;	HO: 0.0058;
Meat colour L*	Gain	miR-2425-3p miR-2284e	miR-2883	LM JE HO RO AN: 0.0; HO JE AN LM: 0.0;	AN, HO, RO: 0.0;

3.3.7 Genotyping of seven miRNA polymorphisms showed difference in allele frequencies across 18 cattle breeds

The first seven SNPs identified in miRNA described previously have been integrated into a broader genotyping analysis. At this time in the study it has been hypothesis that miRNA polymorphisms, which differentially target genes, could be good marker for genomic selection. To test this hypothesis, the first step was to conduct a genotyping analysis, to investigate the distribution of the identified miRNA polymorphisms in a wider range of cattle breeds with a higher number of animals. The genotyping of the SNPs located in the miRNA identified on the first set from dbSNP (id:130) showed that these SNPs were present in all breeds but with different frequencies. By testing Hardy-Weinberg equilibrium (HWE), all the SNPs frequencies were identified as significantly different of expected frequencies if the population was following the equilibrium for the pooled breeds ($p\text{-value} < 0.01$) suggesting that the SNPs could have been positively selected (**Figure 26**).

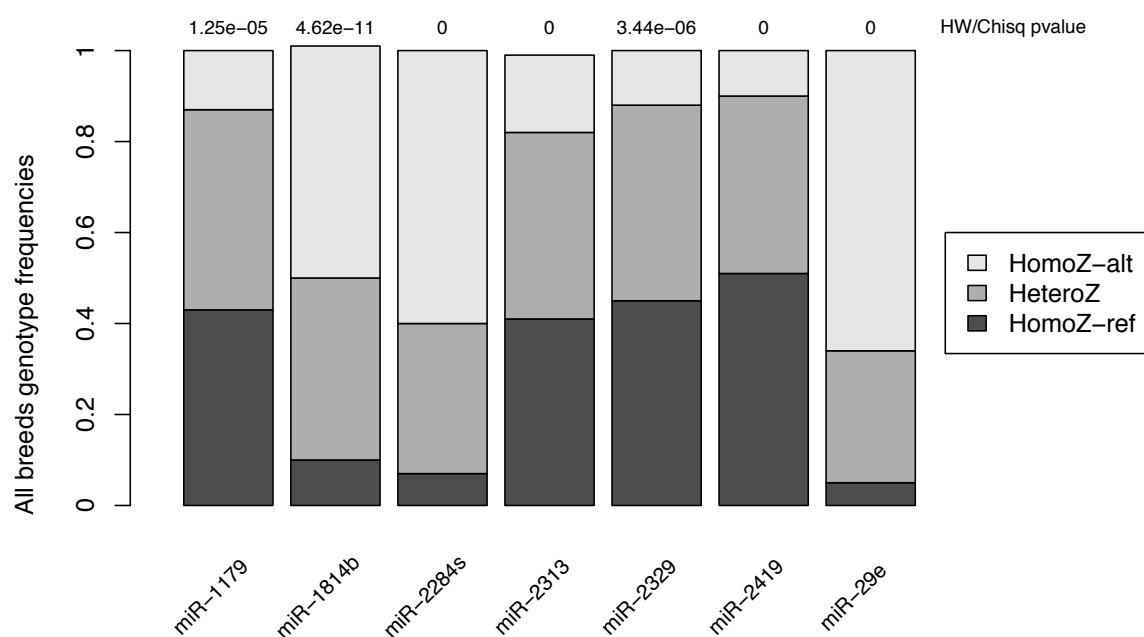


Figure 26 – Allele frequencies of miRNA SNPs in cattle breeds. Representation of the allele frequency for each miRNAs variant across 18 breeds, the value on the top of the bar is the $p\text{-value}$ from the χ^2 testing the Hardy-Weinberg equilibrium. HomoZ-alt: homozygous for the alternate allele, HeteroZ: heterozygous allele, HonoZ-ref: homozygous for the reference allele.

Out of the 18 breeds genotyped, nine breeds had a number of animals ranging from 1 to 12, therefore this made it difficult to assess these breeds independently using the Hardy-Weinberg equilibrium. Therefore, only the breeds having a sufficient number of animals (Aberdeen Angus: AAN, Belgian Blue: BBL, Charolais: CHA, Chillingham: CHL, Hereford: HER, Holstein: HOL, Limousin: LIM, Parthenaise: PAR, Simental: SIM) will be discussed here. It has been identified that alleles of miR-2329 do not show a significant difference between the observed and the estimated frequencies. In addition, the equilibrium was not respected in miR-29e for the Charolais (CHA) and Angus (AAN), in miR-1179 for the Parthenaise (PAR), in 1814b for Simmental (SIM), in miR-2284s for Holstein (HOL), in miR-2313 for Charolais and Holstein and in miR-2419 for Simmental (**Supplementary table 11**). These results show that, even if most of the miRNA SNPs are present in overall breeds, some specificities were identified for certain alleles, suggesting that these SNPs could have had an influence on selected traits.

3.3.8 Association analysis of seven miRNA SNPs in Holstein-Friesian breed

The seven SNPs identified from dbSNP were investigated for association with milk traits. The association analysis was performed on 914 Holstein-Friesian animals for which 42 traits were quantified. Only five miRNA SNPs went through the design stage, discarding *mir-29e* and *mir-2329* variants from the analysis as they were not complying with the criteria of the genotyping platform. On the 42 traits analysed, five traits showed a *p-value* smaller or equal to 0.05. However, the correction of these *p-values* did not show any significant Q-value for the five SNPs considered. Nevertheless, one SNP in mir-2419-3p and one SNP in mir-1814b showed significant *p-value* before correction (equal to 0.01) for respectively rear udder height and carcass conformation (**Table 11**). Even if the corrected *p-value* were not significant, the two SNPs in mir-2419-3p and mir-1814b may have some effect on the related traits.

Table 11 – Suggestive traits (p -value < 0.05) associated to the miRNA SNPs investigated

miRNA ID	dependent	solution	level	SE	trait	n	p-value	Q-value
mir-2419-3p	ruh	0.27	1	0.1	ruh	1295	0.01	0.63
mir-1814b	cconf	-0.05	1	0.02	cconf	1318	0.01	0.63
mir-2284s	fl	-0.37	1	0.17	fl	1380	0.02	1
mir-2284s	bd	0.29	1	0.14	bd	1370	0.05	1
mir-1179	civ	-0.5	1	0.25	civ	1395	0.05	1

3.3.9 Analysis of QTLs and Pathway enrichment for mir-2419-3p and mir-1814b

Despite the fact that the corrected test for association to traits did not show any significant results, our next analysis focused on miR-2419-3p and mirR1814b, which show significant p -value for rear udder height and carcass conformation. This analysis was similar to the one carried out previously on the SNPs identified using the USDA data and aimed to browse available data to identify function and traits, which could support weak associations.

It was found that two miRNAs were not located into the same locus of another coding gene, which could give additional indication to their potential function. Indeed miRNA located in intron of genes tend to have function related to the considered genes (Flynt, 2010; Najafi-Shoushtari, 2010). Consequently, to identify potential links between the miRNA polymorphisms and the traits identified, QTLs containing the polymorphic miRNA as well as pathways and QTL enrichment for the targets of these two miRNAs were investigated. For mir-2419-3p and mir-1814b, respectively 11 and 7 QTLs overlapping their locus were identified (**Table 12**).

Table 12 – mir-1814b and mir-2419 were overlapping QTLs loci

bta-miR-2419-3p	
Trait	QTL size (nt)
Lignoceric acid content	965383
Calving ease (maternal)	7011056
Longissimus muscle area	11309757
Marbling score	11309757
Body weight (slaughter)	13341760
Milking speed	16714497
Milk protein yield	16714497
Social separation--Standing alert	56832639
Social separation--Vocalization	56832639
Milk fat percentage	59857522
Marbling score	102158362
bta-miR-1814b	
Trait	QTL size (nt)
Pelvic area	127710
Structural soundness (legs, feet, penis, and prepuce)	7200115
Carcass weight	9026374
Semen volume	11844999
Teat placement	15972325
Teat placement	15972325
Milk protein percentage	15972325

The miRNA *mir-1814b* is located into a region which is close to the peak of a QTL associated with carcass weight trait (**Figure 27**), which could link to the carcass conformation traits suggested association, identified for *mir-1814b* SNP. *Mir-2419-3p* was located into three QTLs associated to milk traits (milk fat percentage, protein yield and milking speed) which suggested a link to udder traits weak association to the SNP in *miR-2419-3p*. In addition, QTL are enriched in genes targeted by *miR-2419-3p* and *miR-1814b*. 24 QTLs were significantly enriched in targets of *miR-2419-3p* ($p\text{-value} < 0.01$) and eight of them were still significant after $p\text{-value}$ correction. Among these eight QTLs, seven are associated with milk-related traits. In addition, nine QTLs were significantly enriched in targets of *miR-1814b* but no $p\text{-value}$ remained significant after correction.

By investigating the targets of *miR-1814b*, it has been found that this miRNA is targeting two important genes, namely *GATM* and *MSTN*. *GATM* is coding for creatine synthesis enzyme, playing a role in the creatine blood level (Stöckler, 1996). Creatine is a precursor of creatinine, which has been positively correlated to muscle mass characteristics as carcass weight, dressing percentage and proportion of lean meat, which can be related to the carcass conformation traits. Moreover, *MSTN* coding for the myostatin protein involved in the muscle development control (Grobet, 1997) is also a target of *miR-1814b*.

In addition, six out of the eight QTLs enriched in targets for *miR-2419-3p* were related to milk production traits. According to several studies, a correlation between udder-related traits and milk production have been suggested, supporting the link between the milk QTLs enriched in *miR-2419-3p* targets and the weak association with the SNP in *miR-2419-3p* and rear-udder height.

Although the association found for *mir-1814b* and *mir-2419-3p* is weak, the analysis of the location of their SNP in the genome and the function and traits associated with their targets could supports a participation of their polymorphism with respectively

carcass conformation and rear udder height traits identified in the candidate gene association study.

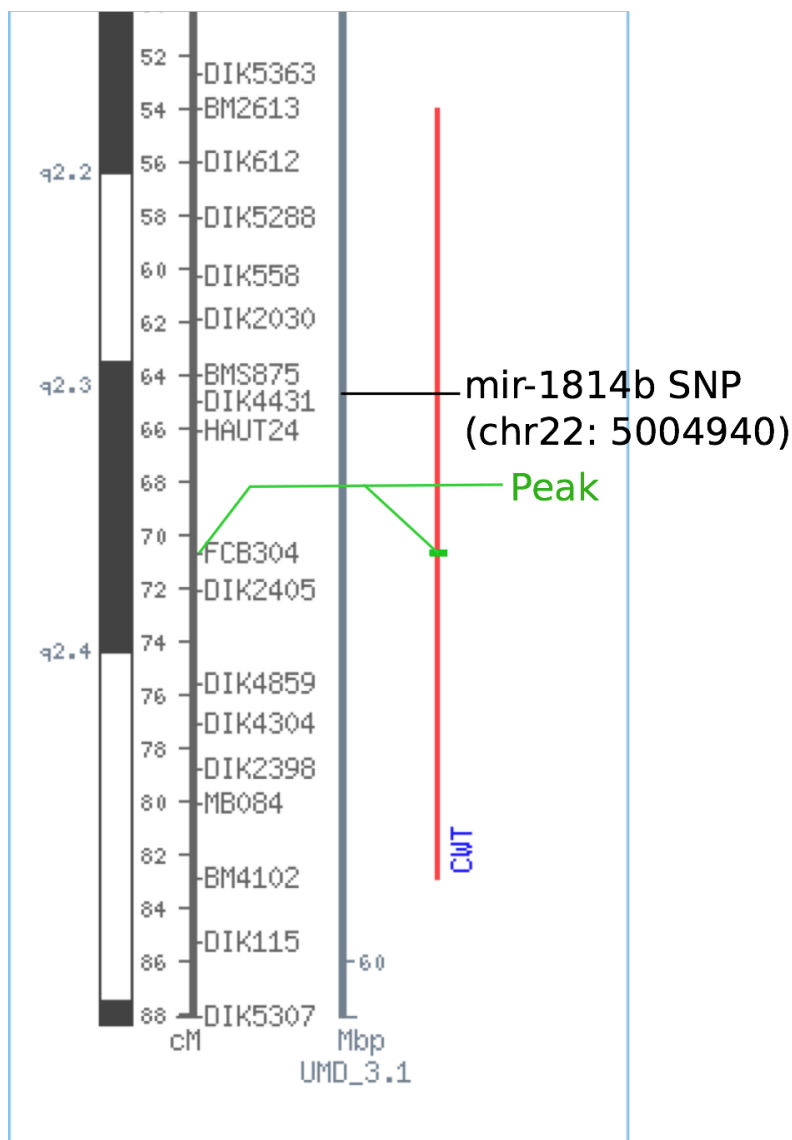


Figure 27 – Representation of carcass weight trait related QTL overlapping *mir-1814b* gene. The SNPs of *mir-1814b* is located near to the peak (green line) of a QTL associated with carcass weight (Figure extracted from AnimalQTLdb)

3.4 Conclusions

The role of miRNA in cattle is being increasingly studied as their ability to regulated multiple targets simultaneously can lead to broad regulation of genes, which can influence production traits in cattle. In this study, the high rate of polymorphisms in cattle were confirmed as described by Zorc et al. (2015) in their last version of mir Sniper, a software identifying SNPs miRNA (Zorc, 2015). Furthermore, variability in miRNAs throughout different breeds, confirmed through the genotyping of 18 breeds of cattle has been identified. Moreover, the importance of the localisation of the SNPs and the potential difference in targets hence in pathway regulation, highlighted links between productive pathways and miRNA variations. Some miRNA polymorphisms were observed to lead to a potentially stronger dis-regulation of many targets and pathways. Among all the miRNA analysed *mir-2419*, *mir-29e*, *mir-449b* could be retained for further in depth functional analysis.

3.4.1 *mir-2419* presented two different SNPs which could have an influence on both dairy and beef cattle.

In the case of *mir-2419* both mature miRNA (5p and 3p) were annotated as expressed and SNPs have been identified in both of them. The first SNP identified in *mir-2419* is located in the mature region of miR-2419-3p. This SNP has been identified in Holstein, Limousin and Romagnola breeds, but the genotyping showed that it is present in all the breeds with different allele ratios. The two alleles identified are T (UMD3.1 reference) and C (dbSNP variant) located in the mature miRNA at position 87083333 in chromosome 16. The variant allele is homozygous in 10% of the animals analysed but mainly present in Hereford breed (38%) followed by the double muscle phenotype breed Belgian Blue (16%) both being beef breed. However, the heterozygous form has been identified in half the overall cattle population studied (**Supplementary table 11**). This SNP can modify the predicted hairpin loop structure creating two small internal loops when only one is present in the reference one. The association study on Holstein animals shown that miR-2419-3p variant could be associated with rear udder height, which can be linked to milk production. The targets of miR-2419-3p were enriched in pathways and QTLs related to milk production. This suggested that the variation in the hairpin loop could impair the proper processing of the miRNA and the regulation of the

related targets. Although the variant identified from USDA in Holstein for *mir-2419-3p* indicated, based on the shift of targeted genes, a link with milk production, the genotyping shown that this variant is mainly present in non-milking breed with an allele ratio similar, higher or lower to Holstein allele ratio.

The second SNP in *mir-2419*, identified in Jersey breed, is located in the seed region of miR-2419-5p. The hairpin structure including this SNP is similar to the reference one except that the internal loop is 1 nucleotide longer. The analysis of the difference of targets, due to this polymorphism, showed that the variant has 10 times more target sites than the reference one. Moreover, the targets of miR-2419-5p variant are specifically enriched in 'Inhibition of Angiogenesis by TSP1' and 'S-adenosyl-L-methionine Biosynthesis' pathways in which none of the genes are targeted by the reference allele. Furthermore, miR-2419-5p had targets enriched in meat related QTL. Finally, miR-2419-5p variant has been identified only in the Jersey dairy breed from the USDA SNP data. These different enrichments could link this allele to a dairy rather than beef oriented selection.

It is interesting to notice that the gene *mir-2419* expressed two mature miRNAs, which both seemed to have an impact on production traits in cattle and present different allele forms, potentially modifying the processing of the pre-miRNA and the targeting function.

3.4.2 Mir29e and mir-449b

The two miRNAs *mir-29e* and *mir-449b* contained SNPs in their seed region. Mir-29e had one SNP identified in the seed region of multiple breeds and *mir-449b* had two SNPs, one in the seed region and one in the mature sequence, both only identified in the Jersey breed. The variants of *mir-29e* and *mir-449b* were linked by some of their targets, which are located in the same pathways. These pathways indicated a potential link with dairy cattle as most of these pathways were related to mammary development and milk production. The main difference between these two miRNA variants was the number of targets. When the reference *mir-29e* and *mir-449b* targeted respectively 627 and 518 genes, the variant form targeted 731 and 223. However, the majority of the pathway significantly enriched by *mir-29e* targets are the one targeted by the reference one. Few pathways enriched in targets of the variant form of *mir-29e* have been identified, even though these targets were more abundant than the reference one. The identification of allele frequencies in *mir-29e* showed that the reference allele

represented only 19.5% of the whole population and 5% of the cattle possessed the homozygous allele. Holstein breed was the one having the highest percentage of the reference homozygous allele (9%) and that this allele is absent in Charolais, Parthenaise and Chillingham. This led to consider polymorphisms in both *mir-449b* and *mir-29e* as potential candidates for dairy breeding.

3.4.3 Jersey specific SNPs in miRNA and population structure

It was interesting to notice that two of the best polymorphic miRNA candidate identified, *mir-2419* and *mir-449b* have been identified in Jersey breed. Unfortunately, the Jersey was not a breed taking part of the first version of the IDB programme, which could have given us the actual frequency of these SNP on a larger population. Furthermore, the SNPs specifically identified in Jersey were identified from the USDA data, after the first genotyping. Studies on variation in cattle breed gave information about the structure of these breeds and in particular the Jersey breed. A study on SNPs in taurine and indicine revealed that Jersey breed have a particular structure, clustering clearly in comparison with other European breed, showing little admixture. It also showed, based on the size of the region in LD, that this breed arise from a small original population (Consortium, 2009). On other study based on microsatellite showed that the genetic variability inside the native Jersey population was not significantly lower than other continental breed and that the inbreeding level was limited (Chikhi, 2004). However, a genotyping study on large group of animals from three breed from US (Holstein, Jersey and Swiss Brown) showed that Jersey breed was the most homozygous (VanRaden, 2011). Overall, that could explain the presence of these Jersey specific SNPs, related to milk production, that has been identified in miRNA sequences.

This study gave an insight into the importance of miRNA in the regulation of genes and pathways in cattle production traits and the potential of miRNA gene polymorphisms. These polymorphisms whether they were located in the hairpin-loop, the mature sequence or the seed region can all have a potential of disruption of the primary function of the miRNA. Further comparative analysis between cattle breed would be necessary to identify whether these polymorphisms are really unique to one breed or to groups of breeds, as *mir-449b* and *mir-2419* seemed specific to Jersey breed.

4 3' UTR variants in cattle breeds modify microRNA binding sites and display association with milk production traits

4.1 Introduction

As mentioned in the previous chapters, miRNAs were small non-coding RNAs regulating the expression of protein coding genes by binding into their 3' UTR. The untranslated regions of mRNA have a role in the regulation of the translation into protein. The expression of genes is regulated by two sets of mechanisms: transcription control, through locus control regions promoting, enhancing or silencing the transcription; and post-transcriptional control located in 5' and 3' untranslated regions presenting motifs or structure related to protein or RNA binding which regulate the translation and the stability of the messenger RNA (Pesole, 2001). Among the different mechanisms of regulation, polymorphisms in the 3' UTR sequence can modify the binding affinity of miRNA and hence modify the translation of the related gene.

Polymorphisms in cattle are used as markers for improving selection of animals for important traits in beef and dairy breeds. The sequencing of the cattle genome, the HapMap project and 1000 bull genome project helped to increase the identification of variants to improve genetic gain (Consortium, 2009; Daetwyler, 2014). Access to these data allowed the development of new approaches using molecular technologies to improve the selection of cattle breeds based on genetic information, and especially to reduce the impact of genetic diseases which can be related to single gene defects, and for creating genomically enhanced breeding value (Spelman, 2013). In Ireland a wide project of genotyping of dairy and beef breeds was launched in 2014 as part of the effort to improve the cattle herd using the Irish Cattle Breeding Federation (ICBF) database for genetic evaluation, and the economic indexes developed with Teagasc (Irish Agriculture and Food Development Authority) to help select superior beef and dairy cattle for farmers (Cromie, 2014).

The potential effects of polymorphisms in 3' UTR related to miRNA binding sites have been predicted computationally using human SNP genotype data and showed a negative selection on SNPs modifying miRNA-binding sites based on a weak allele frequency for these polymorphisms (Chen, 2006). These polymorphisms in human have been related to different diseases such as Tourette's syndrome, which was linked to a modification in a binding site for miR-189 in *SLITRK1* (Abelson, 2005). Other studies

showed links between variation in miRNA-binding sites and mental illness: a variation in *FGF20* 3' UTR was linked to Parkinson Disease and impaired the binding site of miR-433 (Wang, 2008); a variation in the 3' UTR of the Dopamine D2 Receptor gene (*DRD2*), linked to schizophrenia, impaired the binding site of miR-326 and modified the repression of *DRD2* (Shi, 2014). Furthermore, polymorphisms in 3' UTR related to miRNA-binding sites of genes involved in cancer have been identified in human (Landi, 2008; Ryan, 2010), for example *KRAS* 3' UTR presented a polymorphism which impaired the binding of let-7 and was linked to lung cancer risk (Chin, 2008). These findings in human showed that SNPs in miRNA-binding sites of a single gene can cause drastic change in expressed phenotypes.

The analysis of miRNA SNPs in livestock has also been studied, although to a lesser extent than in human. A study on sheep identified a polymorphism in the 3' UTR region of the *MSTN* gene, corrupting a binding site for miR-1 and miR-26 leading to the over-expression of the *MSTN* gene and to the double-muscle phenotype associated with the Texel sheep breed (Clop, 2006). A study on Japanese black cattle showed an association between a polymorphism in the 3' UTR of *GTF2F2* and the number of calves produced at four years. This could be linked to miRNA binding sites as the expression differs between the *GTF2F2* isoforms (Sasaki, 2013). Another study in the Holstein breed identified the presence of a SNP in the 3' UTR of *PRKAG1/AMPK γ 1* gene (Mahmoudi, 2015). This SNP has been linked to higher protein and milk yield and is suggested to impair a miRNA-binding site.

These studies on specific gene SNPs showed the potential of polymorphisms in 3' UTR related to miRNA-binding sites in the modification of phenotypes. Genome wide identification of these SNPs, and the related genes and pathways which can be modified, could give important indications to improve the selection of animals in genomic selection breeding programs.

This study aimed to investigate the hypothesis stating that genetic variants between cattle breeds in their 3' UTRs modified miRNA binding sites and are involved in the differential expression of important economic traits, thus could serve as useful candidates for genomic selection.

4.2 Materials and Methods

4.2.1 Identification of Single Nucleotide Polymorphism in cattle genes 3' UTR

To identify polymorphism in 3' UTR the same set of SNPs from six European cattle breeds identified in collaboration with USDA mentioned in the chapter 3 were used. In a similar way than to the chapter 3 for SNPs in miRNA genes, a custom Python script was used to identify the SNPs located in 3' UTR of cattle genes. The 3' UTR annotation and sequences have been extracted from ENSEMBL via Biomart tool, containing the sequence of 3' UTR identified in cattle from the UMD3.1 annotation. On this set of sequences, only the longest 3' UTR identified for a given gene have been selected.

4.2.2 Construction of 3' UTRs alignment

To identify miRNA binding sites, using TargetScan, an alignment of 3' UTR sequences from the different breed was needed. Using the reference 3' UTR sequences and variant ones, an alignment was generated. The cattle 3' UTRs reference sequences and their coordinates were retrieved from Ensembl through Biomart (*Bos taurus* UMD3.1) (Hubbard, 2002; Kinsella, 2011). The reference genome corresponded to the Hereford breed used for the sequencing and annotation of cattle genome for the UMD3.1 assembly (Zimin, 2009).

The variant 3' UTR sequences were reconstructed, based on the reference sequences and the SNP data from the six cattle breeds described in Chapter 3 (USDA data: Holstein (HO), Jersey (JE), Angus (AN), Romagnola (RO), Limousin (LM) and Fleckvieh (FL)). Each SNP identified was mapped on the reference 3' UTR. Using the corresponding coordinates, the reference nucleotides were replaced by the variants one creating the new variant 3' UTRs for each breed. All cattle breed variants and their reference 3' UTR pairs were used as a basis for identifying differential miRNA binding sites. The 3' UTRs reference and breeds variant of each gene were aligned using clustalW (Larkin, 2007) and a unique identifying number (the breed tag) was given for each breed. The purpose of the breed tag was to distinguish the breed for each miRNA binding sites identified in the output file produce by TargetScan.

4.2.3 Identification and analysis of polymorphic miRNA binding sites

The analysis conducted to identify miRNA binding sites and their potential impact on pathways and hence traits in cattle breeds is identical to the one described in Chapter 2 (**Figure 10**) and Chapter 3. Only slight modifications have been made in the different scripts for parsing the results to consider the targets of the miRNAs in multiple breeds. In the

analysis of SNPs in miRNA binding sites, the ranking considered the targeted genes and not the miRNA and the pathway enrichment was performed considering the total sets of genes having polymorphic binding sites for each breed.

4.2.4 Genotyping for gene association

The gene association analysis has been conducted in association with Teagasc during the update of the International Beef and Dairy SNPchip (IDB v2.0) allowing the integration of some of the top SSr-ranked gene having SNPs in 3' UTR and presenting function related to important traits. The analysis of the genotyping data was identical to the method described in Chapter 3. In this second version of IDB, 25 SNPs in total have been selected, of which 18 were filtered in and tested on more than 30000 animals from 25 different breeds

4.2.5 Association study

The association study has been conducted with Teagasc and followed the same protocol as described in Chapter 3. An increased number of animal of the Holstein-Friesian breed have been used for the association study, bringing this number to 10707 animals genotyped. The 20 traits for which the association have been tested are fat percentage, fat weight, protein percentage, muscle weight, calving interval, carcass conformation, carcass fat, carcass weight, survivability, gestation time, somatic cell count, angularity, stature, body condition score, chest width, direct calving difficulty, maternal calving difficulty, calf carcass weight.

4.2.6 Analysis of SNPs from 83 candidate genes for miRNA binding discovery.

This set SNPs has been analysed in collaboration with Teagasc following a study on polymorphisms discovery from DNA sequencing, aiming to identify SNPs and related allele frequencies to be used as fertility marker in dairy cattle (Mullen, 2012). The SNPs were identified from 83 candidate genes of the somatotrophic axis from 150 Holstein-Friesian bull. These animals were separated in two groups diverging on genetic merit for fertility. A set of 697 SNPs has been identified to be located in these 83 candidate genes (Mullen, 2012). These SNPs were used to construct the variant 3' UTR related to these candidate genes and have been analysed for miRNA binding sites and ranked with the SSr method described previously.

4.3 Results and Discussion

4.3.1 Identification of 3' UTR SNPs in six cattle breeds

To identify polymorphisms in the 3' UTR of cattle, the same set of SNPs as in chapter 3 were used. This set corresponds to 24,032,577 SNPs obtained in collaboration with USDA from six European cattle breeds as described previously. A total of 11,761 annotated 3' UTR loci were retrieved from ENSEMBL, using the Biomart tool. Among these 11761 3' UTRs, 1680 contained a total of 3,253 SNPs. This represented an average of 0.46 SNP per kb of the annotated 3' UTRs. All cattle chromosomes contained genes having SNPs located in their 3' UTR (**Figure 28**).

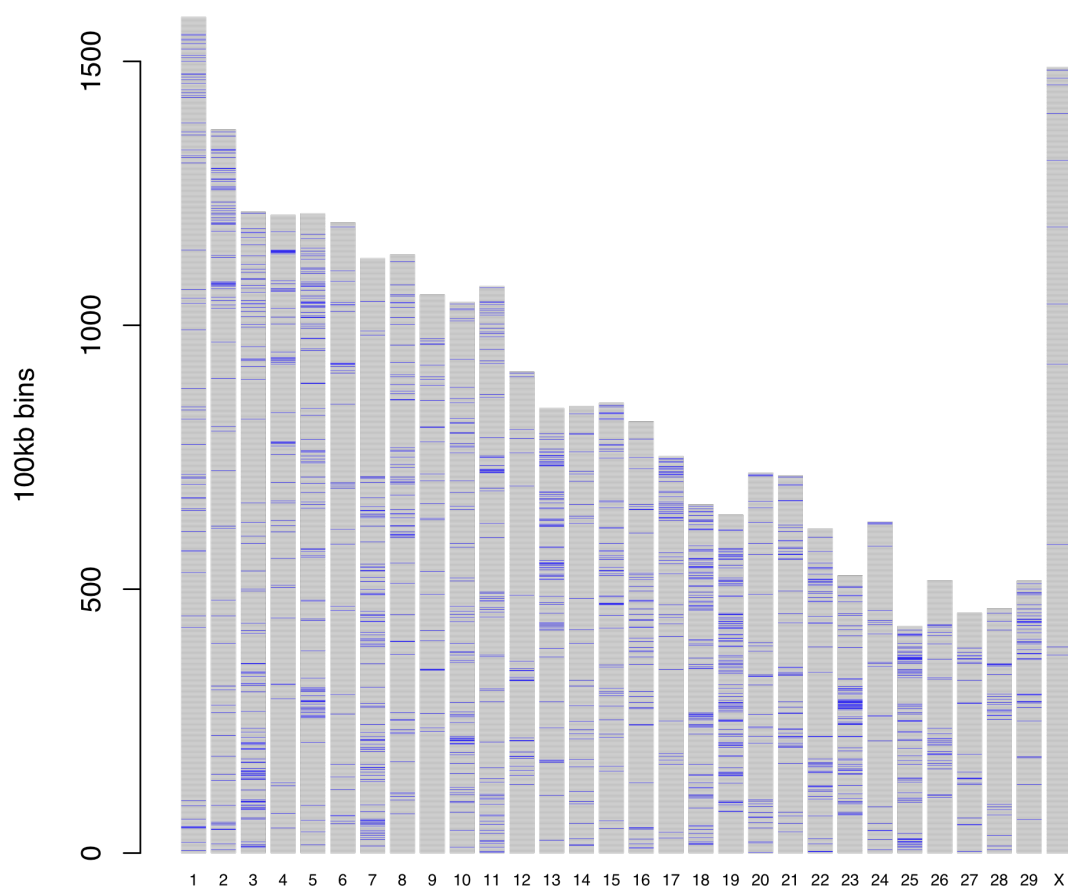


Figure 28 – Distribution of SNPs located in 3' UTR throughout cattle genome. The bar represented each 29 autosomes and chromosome X, divided into 100 kilo-bases bins. Each bins highlighted in blue is containing at least one SNPs in the 3' UTR of a gene.

However, there is a higher number of SNPs in 3' UTRs located in chromosome 4, corresponding to an average of 1.57 SNPs per kilo-base in the 3' UTRs with a standard deviation of 0.24. The gene containing the most SNPs in its 3' UTR (67 SNPs) was an uncharacterized gene located in chromosome 4 (ENSBTAG00000039928), coding for a short protein related to the GTPase IMAP family. In addition, 27 genes out of the 1680 having SNPs in their 3' UTRs, have been identified to have 10 or more SNPs in their 3' UTR. Out of these 27 genes, a third (9/26) were also located in chromosome 4 (Table 13).

Table 13 – List of top genes having 10 or more SNPs in their 3' UTR.

Gene ID	Number of SNPs							
	Chm	HO	JE	AN	LM	RO	FL	Total
ENSBTAG00000039928	4	24	17	34	34	39	13	67
ENSBTAG00000040331	4	19	1	28	38	36	8	55
ENSBTAG00000046257	4	3	2	5	24	10	13	39
ENSBTAG00000039588	4	11	0	8	5	20	9	34
ENSBTAG00000046433	4	12	8	14	14	1	0	27
ENSBTAG00000025664	10	23	10	20	25	4	0	26
ENSBTAG00000030940	4	4	3	3	8	17	3	25
ENSBTAG00000039093	4	21	11	6	15	0	0	24
ENSBTAG00000005759	2	18	13	4	14	0	0	20
ENSBTAG00000045588	7	10	5	8	17	1	0	20
ENSBTAG00000039028	13	11	2	1	10	0	0	16
ENSBTAG00000039037	24	1	12	0	5	0	0	15
ENSBTAG00000002859	4	8	12	12	11	0	0	15
ENSBTAG00000006738	21	2	11	1	13	0	0	14
ENSBTAG00000002786	21	7	6	9	5	0	0	13
ENSBTAG00000000653	13	4	7	6	4	0	0	13
ENSBTAG00000004894	4	0	0	2	4	5	1	12
ENSBTAG00000009363	18	8	8	4	2	0	0	11
ENSBTAG00000040586	24	6	2	0	7	0	0	10
ENSBTAG00000040580	21	5	2	3	4	0	0	10
ENSBTAG00000025494	23	5	2	2	5	0	0	10
ENSBTAG00000017056	3	6	3	2	10	0	0	10
ENSBTAG00000002414	3	0	0	10	1	0	0	10
ENSBTAG00000007374	10	0	9	5	5	0	0	10
ENSBTAG00000021939	19	6	5	1	7	0	0	10
ENSBTAG00000008747	19	3	7	0	7	0	0	10

4.3.2 Polymorphism in 3' UTRs create or impair miRNA-binding sites

To identify which SNPs in 3' UTRs are located in miRNA-binding sites and potentially modify their functions, the polymorphic 3' UTR sequences and their corresponding reference sequences have been analysed with TargetScan using annotated miRNA sequences from miRBase. The analysis of miRNA binding sites showed that of the 1680 polymorphic 3' UTR, 726 have polymorphic binding sites counting for 1097 different SNPs. Among these 726 polymorphic 3' UTRs, 288 have been identified only in dairy breeds (Jersey and Holstein), 215 only in beef breeds (Angus, Romagnola and Limousin) and two unique to the Fleckvieh breed (**Figure 29**)

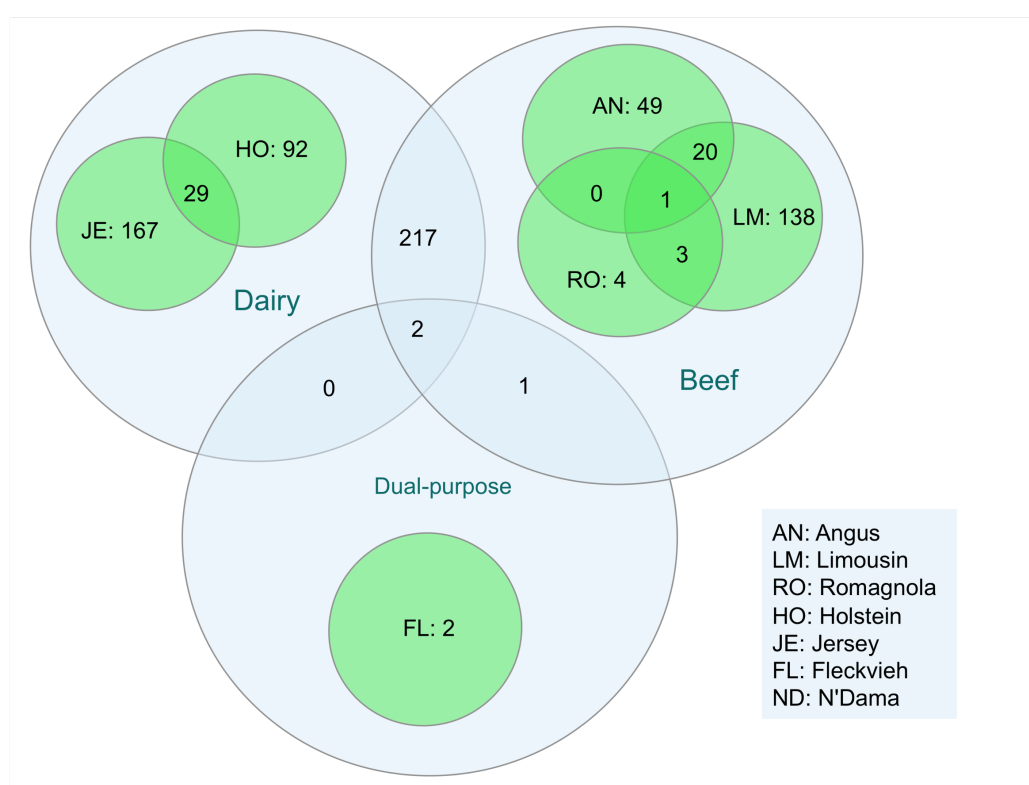


Figure 29 – Gene distribution across the different groups and breed. The Venn diagram represents the distribution of genes having polymorphisms in their 3' UTR identified in certain breeds and also regrouped following production interest.

All the polymorphic 3' UTRs for each breed have been ranked using the scoring system described in chapter 2 (**Supplementary table 12**). Among the highest ranked genes with polymorphic miRNA-binding sites, the *GIMAP1* Limousin variant is the top ranked one. A total of 55 SNPs have been identified in the 3' UTR of *GIMAP1*. Out of these 55 SNPs, 37 were modifying a total of 34 different target sites changing the binding affinity of 50 miRNAs. Among these 37 SNPs, 10 have only been identified in Limousin, four in Angus, four in Romagnola and one in Fleckvieh (**Figure 30**). The 10 SNPs identified in Limousin created seven unique miRNA binding sites and the loss of two miRNA binding sites. Moreover, the gene identified as ENSBTAG00000039928, under the Ensembl annotation and associated as a member of the GIMAP family and containing 11 members in cattle, have been identified as overrepresented in copy number variation analysis (Bickhart, 2012). As mentioned previously, this gene has the highest number of SNPs in the 3' UTR and was ranked 11th for its miRNA binding site score. Among the 67 SNPs identified in the 3' UTR of ENSBTAG00000039928, 24 are located in miRNA-binding sites. Two of these SNPs led to the formation of one unique miRNA-binding site in Angus, one SNPs created one unique miRNA-binding site in Jersey, one SNPs formed a new miRNA binding sites in Holstein and two SNPs formed two miRNA-binding sites in Fleckvieh. The GTPase of Immunity Associated Protein (GIMAP) proteins are a family of proteins involved in the immune system for which the precise functional mechanism is still unclear. The expression of GIMAP members have been found to vary during the T-helper cell differentiation and could be involved in asthma, allergies and auto-immune diabetes in human (Krücken, 2004; Riaz, 2015). A study on mouse knock-down of *GIMAP1* showed a role of this gene in mature B and T cell development and survival (Saunders, 2010). Moreover, *GIMAP1* has been shown to be over-expressed in peripheral blood mononuclear cells during bovine tuberculosis infection in the presence of tuberculin (Meade, 2008). The two polymorphic GIMAP family members which have been identified could have a strong potential for immunity improvement in cattle selection.

In the top thirty ranked genes (**Table 14**), apart from *GIMAP1* and ENSBTAG00000039928; *GUCA2B*, *LAMTOR2*, *PRKAG1*, *BST2*, *CTC1*, *TARBP1*, *CYP3A5*, *MED23*, and *CYP17A1* suggested a potential for cattle production trait variations.

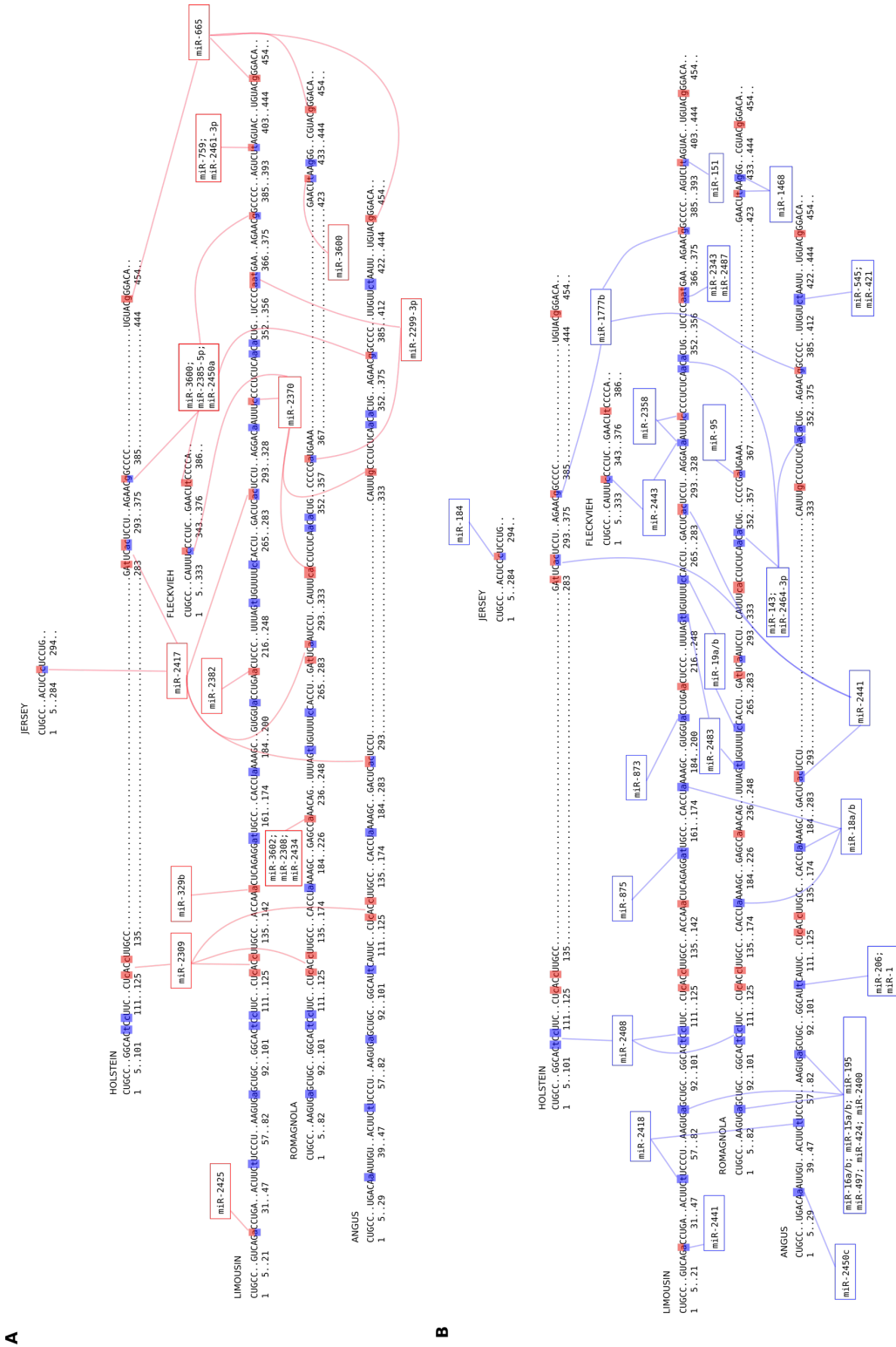


Figure 30 – Single nucleotide polymorphisms in *GIMAP1* 3' UTR through six cattle breeds creates lost and gain of miRNA binding sites. The 27 SNPs identified in miRNA binding sites created changes in predicted binding of miRNAs to the 3' UTR of *GIMAP1*. A) 20 SNPs created a disruption (red) of 18 miRNA-binding sites for 16 miRNAs and B) 27 SNPs (blue) created 24 new miRNA-binding sites for 34 miRNAs

Table 14 – Top 30 ranked genes having the highest SSr score

Gene ID	Gene name	number	number	SSr	Breeds
		polymorphic site	common site		
					LM RO AN HO
ENSBTAG00000040331_BTLM	<i>GIMAP1</i>	39	21	-13.0629	FL JE
ENSBTAG00000007194_BTTHO	<i>GUCA2B</i>	5	1	-5.2000	HO LM
ENSBTAG00000004517_BTJE	<i>TARBP1</i>	14	14	-2.2700	JE
ENSBTAG00000005450_BTLM	<i>CYP3A5</i>	5	3	-2.2283	LM
ENSBTAG00000012721_BTTHO	<i>HOGA1</i>	13	17	-2.0953	HO
ENSBTAG00000039028_BTLM	<i>PI3</i>	10	7	-1.9743	LM HO
					FL AN RO LM
ENSBTAG00000039928_BTFL	<i>GIMAP-like</i>	16	31	-1.8348	HO JE
ENSBTAG00000011754_BTLM	<i>LAMTOR2</i>	5	5	-1.5240	LM
ENSBTAG00000007117_BTAN	<i>CTC1</i>	11	19	-1.4966	AN LM HO
ENSBTAG00000012418_BTJE	<i>MED23</i>	2	1	-1.3560	JE
ENSBTAG00000014335_BTJE	<i>CYP17A1</i>	8	13	-1.2843	JE AN HO LM
ENSBTAG00000031725_BTLM	<i>LOC517833</i>	12	35	-1.0893	LM
ENSBTAG00000010548_BTJE	<i>LRRC6</i>	11	15	-1.0333	JE
ENSBTAG00000005759_BTJE	<i>PLA2G2A</i>	9	22	-0.8628	JE HO LM AN
ENSBTAG00000012149_BTTHO	<i>HOXC8</i>	2	1	-0.7860	HO
ENSBTAG00000035012_BTRO	<i>TSPY-like</i>	6	12	-0.7420	RO AN LM HO
ENSBTAG00000003898_BTAN	<i>HMGCS2</i>	12	34	-0.6868	AN JE LM HO
ENSBTAG00000010225_BTLM	<i>POLR2D</i>	11	31	-0.6472	LM HO
ENSBTAG00000004595_BTJE	<i>GML</i>	4	7	-0.5737	JE
ENSBTAG00000014426_BTAN	<i>PRKAG1</i>	9	34	-0.5519	AN
ENSBTAG00000025663_BTLM	<i>RNASE1</i>	10	27	-0.5456	LM
					LM AN HO JE
ENSBTAG00000045588_BTLM	<i>BST2B</i>	23	140	-0.5341	RO
ENSBTAG000000047141_BTJE	<i>PAG7</i>	3	6	-0.5235	JE
ENSBTAG00000038706_BTJE	<i>MT1E</i>	7	15	-0.4363	JE HO
ENSBTAG00000024772_BTLM	<i>TRIP6</i>	5	14	-0.4336	LM
ENSBTAG00000015004_BTJE	<i>TBCB</i>	4	5	-0.4248	JE
ENSBTAG00000039813_BTLM	<i>GZMB</i>	3	4	-0.4140	LM HO JE
ENSBTAG00000011559_BTLM	<i>RPL7A</i>	2	1	-0.3920	LM
ENSBTAG00000046433_BTJE	<i>TRB-like</i>	7	15	-0.3892	JE
ENSBTAG00000021939_BTTHO	<i>ELAC2</i>	14	70	-0.3778	HO LM JE AN

The list is presenting the 30 genes which had polymorphism in miRNA binding site causing the most disruption base on their score. The gene ID represent the genes with the highest ranked breed specific polymorphism, the breed column is showing which breeds also harbour polymorphisms in this given gene but with a lower score.

In the 3' UTR of *GUCA2B*, one SNP, only identified in Holstein and Limousin, is forming one miRNA binding site. *GUCA2B* codes for an endocrine hormone, the guanylate cyclase activator 2B or uroguanylin, which is expressed in intestine and inhibits sodium ion absorption and stimulates electrolyte excretion in kidney (Sindić, 2006). Moreover it has been shown in mice that differences of expression of *GUCA2B* had a role in pregnancy and fertility (McConaha, 2011) and intron variations in *GUCA2B* have been related to hypertension in human (Yoshikawa, 2007). It also has been found that down-regulation of *GUCA2B* among other genes in the guanylate cyclase-C pathway is related to gastrointestinal inflammation (Brenna, 2015). It has also been shown that uroguanylin is involved in satiation in mice (Valentino, 2011). It could be suggested that variation of expression in *GUCA2B*, which played a role in gastrointestinal function, could be involved in feed efficiency in cattle.

In the 3' UTR of *LAMTOR2*, one SNP, present in a miRNA-binding site, has been identified only in Limousin. *LAMTOR2* coded for a protein called p14, which was part of the Ragulator/LAMTOR complex, regulating mTOR (mammalian target of Rapamycin). It has been shown to be crucial for dendritic cell homeostasis, playing an important role in phagocyte for the endolysosomal transport during *Salmonella* infection (Taub, 2012). Furthermore, a study on human immunodeficiency identified p14 variant as responsible of many phenotype including B-cell and cytotoxic T-cell deficiency. They also identified a new role for p14 in biogenesis of endosome. They used a MAS to genotyped parents and their 12 siblings of which four were affected. By coupling the marker analysis with a microarray analysis, they narrowed down the causal marker to a point mutation in the 3' UTR of *LAMTOR2*. The microarray analysis showed that in their candidate genes, only *LAMTOR2* was underexpressed. They suggested that the mutation was destabilising the messenger RNA causing the immunodeficiency syndrome (Bohn, 2007). Therefore, *LAMTOR2* is an important factor in the adaptive immune response and the latter study suggested a mutation which could be related to miRNA binding site.

In the 3' UTR of *CYP3A5*, one SNP has been identified only in Limousin which is present in a miRNA binding site. The cytochrome P450s (CYP) is an enzyme family mainly involved in the metabolism of drugs and more generally xenobiotics as plant secondary metabolites or pollutant and endogenous compounds as steroid or bile acids, helping to degrade this wide range of chemicals. In total, 57 members of the CYP family have been identified in humans and had many orthologs in cattle (Nebert, 2002; Zancanella, 2010). *CYP3A* is the major cytochrome P450 subfamily involved in drug metabolism, metabolising

more than half of the drugs in human but also in cattle (Thummel, 1998; Gellner, 2001; Cantiello, 2009). It has been shown that *CYP3A5* and other CYP family members presented numerous variabilities in their sequence in different populations of human and rhesus macaques, inducing a modification of the drug metabolism activity (Kuehl, 2001; Uno, 2010; Onizuka, 2011). In cattle, it has been shown that *CYP3A5*, a member of the CYP3A family, is mainly expressed in liver and that differences of expression has been identified among different cattle breeds (Zancanella, 2014). However, it also has been shown that the nomenclature for *CYP3A5* in cattle as orthologue of human *CYP3A5s* is not accurate and direct correlation should be made carefully (Zancanella, 2010). In addition *CYP3A5* expression has been correlated to miRNA and polymorphisms as SNPs in the 3' UTR of *CYP3A5* genes have been associated with miRNA-binding sites in human (Wei, 2012; Ramamoorthy, 2013).

In the 3' UTR of *CYP17A1*, three SNPs created three miRNA binding sites and one SNP impaired one miRNA binding site. *CYP17A1* is also a member of the cytochrome P450 family. It have been identified as a catalyser involved in the synthesis of cortisol, androgen and oestrogen and mutation in *CYP17A1* in human have been linked to deficiency in gonadal steroids production inducing hypertension and hypokalaemia (Benetti-Pinto, 2007; E. S. Lee, 2013). *CYP17A1* is mainly expressed in theca cells and its expression varies during maturation (Voss, 1993; Bao, 1998; Nimz, 2009). Moreover, a study on post-partum dairy cow undergoing severe negative energy balance showed that *CYP17A1* was upregulated in liver of cows under severe negative energy balance and was linked to DHEA synthesis which could have a role in hepatic signalling pathways (McCabe, 2012).

In the 3' UTR of *PRKAG1*, one SNP only identified in Angus, created a novel miRNA-binding site. *PRKAG1* coded for a subunit of the adenosine monophosphate-activated protein kinase (AMPK), involved in the control of the energy status of the cell, reacting to the increase of AMP/ATP ratio and have also been implicated in the metabolism of fatty acid and glycogenesis. It also has been shown that cattle oocyte meiosis is inhibited by AMPK. The SNP in the 3' UTR, which has been identified in the present study, has been previously identified in Holstein cattle and is associated with dairy production traits. It has been shown that the C/C allele have a frequency of 65.8% in high milk yield animal while the T/T allele represent 80.8% in the low milk yield animals. In the same study these SNPs have also been linked to miRNA binding sites. However, the miRNAs identified in that Holstein study differ to those which have been identified here as they used an inaccurate method solely based on BLAST for identification. The method they used consisted of the

miRNA search tool of miRBase using their variant sequence. This tool is not a miRNA binding site search but a BLAST version for small sequence homology search. The result they provided, showed clearly that the match they identified for miRNA-423-5p was actually a spurious one as they showed a homology alignment and not a complementary alignment. Furthermore, the alignment they provided showed that the necessary perfect seed region match of the miRNA (nucleotide 2 to 8) was not preserved, their matches showed to be from the nucleotide 5 to 18 of the miRNA.

In the 3' UTR of *BST2B*, 10 SNPs created 10 miRNA binding sites and eight SNPs impaired eight miRNA binding sites. Bone marrow stroma antigen 2 (*BST2*) has been recently characterised in cattle, presenting three isoforms, which are known as *BTS2A1*, *BST2A2* and *BST2B*. *BST2* has been reported to have antiviral activity, however the *BST2B* form, corresponding to the one identified in the present study, has a weaker antiviral activity due to the lack of a GPI anchor (Takeda, 2012). This lower antiviral activity has also been found in sheep. However, another study in sheep showed that *BST2B*, located in the Golgi apparatus, has the ability to prevent the trafficking of viral protein envelope by sequestration in the Golgi, hence reducing the infectiousness.

In the 3' UTR of trans-activation response RNA-binding protein 1 (*TARBP1*), one SNP only identified in Jersey, created one miRNA-binding site and impaired another one. *TARBP1* is one of the key components of the RISC complex, with DICER and ARGONAUTE 2 (AGO2). TARBP is a double-strand RNA binding protein, loading small RNA into the RISC complex (Daniels, 2009; Chakravarthy, 2010).

In the 3' UTR of *CTCI*, one SNP formed one miRNA-binding site and another SNP impaired one miRNA-binding site. This two SNPs were in Limousin, Angus and Holstein breed. CTS telomere maintenance complex component 1 (*CTCI*) is part of the CTS complex involved in the protection of telomere degradation. A study on coping in piglets, which regrouped all behavioural and physiological responses to challenging situation, found a cis-eQTL associating polymorphisms and level of expression of *CTCI* in the hypothalamus (Ponsuksili, 2015). Polymorphisms in *CTCI* have also been associated to cerebral abnormalities in human (Polvi, 2012). The polymorphisms in miRNA binding sites of *CTCI* could modify the level of expression perturbing its function. The results on variation of expression of *CTCI* in piglets could suggest that these polymorphisms could also be linked to differences of behaviour in cattle. The difference in coping can impact stress level, which itself can have effect on diseases vulnerability in animals (Koolhaas, 1999).

In the 3' UTR of *MED23*, one SNP, only identified in Jersey, formed one miRNA-binding site. *MED23* is coding for a subunit of the mediator complex, which is involved in the transcription mechanism, promoting the communication between transcription factors, polymerase II and initiation factors (Conaway, 2005). The subunit coded by *MED23*, in association with *ELK1*, has been shown to be an important regulator in adipogenesis as the knockdown of *MED23* in embryonic fibroblast cell inhibited the adipocyte differentiation (Wang, 2009). It has also been shown that the specialisation of multipotent mesenchymal stem cells into smooth muscle cell (SMC) or adipocytes was dependent of the absence or the presence of *MED23*, acting as a regulator of sets of genes involved in the development of SMC and adipocytes (Yin, 2012).

This approach of variation in miRNA-binding sites amongst different breeds of cattle showed that the presence of SNPs in miRNA-binding sites was not an uncommon phenomenon. Moreover, the most extensive variations in 3' UTRs were related to genes which could have an impact on important traits for breeding improvement.

4.3.3 Pathways enriched for polymorphic 3' UTR are related to immune response

In order to have a broader insight of the function potentially affected by the variation in miRNA-binding sites, an analysis on the molecular pathways was conducted. To identify pathways most affected by miRNA-binding site variants, Ingenuity Pathway Analysis (IPA) software was used. The significantly enriched pathways have been corrected with Benjamini-Hochberg method (adjusted *p-value* < 0.01). All the genes having SNPs in their 3' UTR, modifying miRNA-binding sites were separated according to breed and whether a miRNA-binding site was gained or lost. As showed previously, more genes having polymorphic binding sites have been identified specifically in dairy (288) than in beef breeds (215). Although the number of genes with variant miRNA-binding sites was similar between Holstein and Jersey, only the Holstein group presented significantly enriched pathways (adjusted *p-value* < 0.05). Similarly, Limousin breed was the beef breed which presented the majority of significantly enriched pathways for variant binding sites when the other beef breed present, Romagnola, had only one pathway significantly enriched.

Overall, 23 significantly enriched pathways have been identified for a total of 33 candidate genes. Out of these 24 pathways, the most significant ones included Antioxidant Action of Vitamin C, Endothelin-1 Signalling, Synaptic Long Term Depression and VEGF Family Ligand-Receptor Interactions. Based on the scoring developed, it has been found that the highest scored, significantly enriched pathway, is fMLP Signalling in Neutrophils (**Table 15**). The polypeptide N-formyl-methionyl-leucyl-phenylalanine (fMLP) is a chemoattractant and activator of granulocytes as neutrophils, leading to the release of reactive oxygen species (ROS) involved in the immune response (Panaro, 1999). Vitamin C had a role of protection against oxidative compounds. The candidate genes identified in this pathway were related to the recycling of vitamin C (*TXNRD1*) (Mein, 2012) or were activated by ROS in the case of phospholipase A2 (*PLA2G2A*, *PLA2G2F*, *PLA2G12A*) (Clark, 1995; Andersen, 2003; Shelat, 2008) and NF- κ B (*IKBKB*, *NFKBIE*) (Schenk, 1994; Schmidt, 1995) mediating an inflammatory response. Endothelin-1 was a small peptide having a role of vasoconstriction and have been identified to play a role in many diseases as cancer (Grant, 2003) cardiopathy (Sharma, 2000) or lung diseases (Fagan, 2001).

It has also been noticed that the majority (22/24) of the pathways enriched concerned miRNA targets variation, impairing miRNA-binding sites. The two pathways, which were enriched in gained miRNA-binding sites, were Ephrin B Signalling (Limousin variants) and Retinoate Biosynthesis I (Romagnola variant). However, Ephrin B Signalling pathway was also enriched for impaired binding sites. Finally, four pathways were enriched in targets presenting miRNA binding sites impairing four variants common between Limousin and Holstein.

Table 15 – Pathways significantly enriched with gene having polymorphic miRNA-binding sites

Pathway name	SNP type	Overlap type	Breeds	q-value mean	score mean	number binding	number miRNA	number targets
fMLP Signalling in Neutrophils	Lost	unique_single	BTHO	0.048	-0.130	394	5	252
Virus Entry via Endocytic Pathways	Lost	unique_single	BTHO	0.038	-0.130	376	5	249
Ephrin B Signalling	Gain	common	BTLM	0.021	-0.129	613	6	329
NF-κB Activation by Viruses	Lost	unique_multi	BTHO BTLM	0.047	0.038	-0.129	350 464	4 5 234 292
Ephrin B Signalling	Lost	common	BTHO	0.048	-0.128	372	4	233
Colorectal Cancer Metastasis	Lost	unique_single	BTHO	0.048	-0.127	712	7	378
LPS-stimulated MAPK Signalling	Lost	unique_single	BTHO	0.048	-0.127	356	4	232
Apoptosis Signalling	Lost	unique_single	BTLM	0.019	-0.125	747	6	365
Phospholipase C Signalling	Lost	unique_single	BTHO	0.017	-0.125	1112	9	472
Macropinocytosis Signalling	Lost	unique_single	BTHO	0.048	-0.124	343	4	229
IL-8 Signalling	Lost	unique_single	BTHO	0.017	-0.123	737	8	369
Antioxidant Action of Vitamin C	Lost	unique_single	BTLM	0.011	-0.120	790	7	384
Role of MAPK Signalling in the Pathogenesis of Influenza	Lost	unique_single	BTHO	0.048	-0.120	565	4	333
Androgen Signalling	Lost	unique_single	BTHO	0.048	-0.120	568	5	341
Retinoate Biosynthesis I	Gain	unique_single	BTRO	0.022	-0.119	159	1	141
Endothelin-1 Signalling	Lost	unique_single	BTLM	0.011	-0.119	1078	9	432
Synaptic Long Term Depression	Lost	unique_single	BTLM	0.011	-0.117	1029	8	424
IL-1 Signalling	Lost	unique_single	BTHO	0.038	-0.117	624	5	350
CCR3 Signalling in Eosinophils	Lost	unique_multi	BTHO BTLM	0.025	0.047	-0.116	617 752	6 6 325 382
Fc Epsilon RI Signalling	Lost	unique_single	BTHO	0.048	-0.115	579	5	318
VEGF Family Ligand-Receptor Interactions	Lost	unique_multi	BTHO BTLM	0.025	0.010	-0.115	586 757	5 6 322 385
Sperm Motility	Lost	unique_single	BTHO	0.048	-0.111	754	5	394
MIF-mediated Glucocorticoid Regulation	Lost	unique_multi	BTHO BTLM	0.047	0.019	-0.109	307 480	3 4 199 295
MIF Regulation of Innate Immunity	Lost	unique_single	BTLM	0.038	-0.108	480	4	295

The number of non-polymorphic potential miRNA-binding sites and their related miRNAs identified in the candidate genes involved in the enriched pathways was very high (from 307 to 1112 binding sites for 141 to 472 miRNAs) compared to the number of polymorphic binding sites in the same genes. The difference between the polymorphic and non-polymorphic binding sites was highlighted through the score system used to rank the different genes and high ranking genes were the ones not only having the higher target score for binding sites but also the highest number of polymorphic binding sites in comparison with non-polymorphic ones for the same gene. In relation with this polymorphic sites number, the highest ranked gene in the enriched pathways was *PLA2G2A* (rank 25 for the Holstein variant). This gene was present in 14 of the enriched pathways and was the third most represented gene in the enriched pathways after *MRAS* (23 pathways) and *PRKCQ* (18 pathways). However, *MRAS* and *PRKCQ* were lowly ranked (rank 955 and under) therefore presenting a low TargetScan score and a high number of non-polymorphic binding sites. Furthermore, as mentioned previously, there were only 33 different genes involved and some of these genes were present in almost all the pathways identified which indicated an overlapping of the candidate genes between the pathways. It is not surprising as the majority of these pathways were related to defence and immune system. The four most significantly enriched pathways mentioned previously are all related to Limousin variants.

4.3.4 QTL enrichment in polymorphic 3' UTR

Similar to Chapter 3, to identify potential traits linked to polymorphisms in miRNA targets, these polymorphic targets were tested for enrichment in QTL. The polymorphic targets were grouped depending of the breeds in which the SNPs have been identified. On the seven breeds in which polymorphisms in miRNA-binding sites have been identified, four showed significant enrichment of miRNA polymorphic target genes in a total of 14 QTL-associated traits (**Table 16**). Among these 14 traits, nine were enriched for the reference targets and five were enriched for the variant targets. In contrary to the QTL analysis in the Chapter 3, there were no common QTL-related traits, which were enriched for both variant targets and reference targets. Furthermore, the Limousin breed showed enrichment in traits related to QTL only for the variants impairing miRNA binding sites and the majority of the traits (4/6) were related to milk production. It can also be noted that Limousin variant targets lost enrichment in QTL for ovulation rate.

Table 16 – Traits associated to QTLs significantly enriched with genes having polymorphic miRNA binding sites

Breed	Type	Traits	p-value	Chromosomes	Number QTL IDS	Maximum targets	Number maximum
BTJE	Lost	Body weight (weaning)	0.006	Chr26 Chr27 Chr25	39	10681 10671 10647 15731	7
				Chr22 Chr23 Chr21			
				Chr29 Chr7 Chr6			
				Chr4 Chr3 Chr2 Chr1			
				Chr13 Chr11 Chr10			
				Chr17 Chr16 Chr15			
BTFL	Gain	Calving ease (maternal)	0.010	Chr4	1	10721	4
BTFL	Gain	Marbling score	0.013	Chr4	1	10013	4
BTFL	Lost	Body weight (yearling)	0.014	Chr26 Chr4	2	10720	3
BTLM	Lost	Milk conjugated linoleic acid	0.017	Chr19	1	10031	27
BTLM	Lost	Milk oleic acid percentage	0.018	Chr19	1	10028	21
BTLM	Lost	Milk trans-vaccenic acid percentage	0.018	Chr19	1	10029	25
BTRO	Gain	Parasites mean of natural logarithm	0.021	Chr4	1	10515	3
BTFL	Lost	Milk fat percentage	0.029	Chr26 Chr4	3	5055	3
BTRO	Gain	Milk protein yield (EBV)	0.031	Chr13 Chr4 Chr8	3	6061	3
BTLM	Lost	Ovulation rate	0.038	Chr7 Chr5 Chr19 Chr14	4	10573	13
BTLM	Lost	Milk stearic acid percentage	0.040	Chr19	1	10027	25
BTRO	Gain	Somatic Cell Count	0.042	Chr18 Chr4 Chr8	4	2492	3
BTLM	Lost	Social separation-- Walking/running	0.042	Chr19 Chr10	2	7143	17

The table represented a ranked list of traits related to Quantitative Traits Loci, which are significantly enriched in genes having SNPs in their miRNA binding sites. Each trait can be related to more than one QTL and the maximum ids is listing the QTL id which contain the higher number of candidate genes.

In the contrary, the Romagnola breed had significant enrichment only for the variant targets related to parasite load, milk production and somatic cell count traits. Only one trait, body weight at weaning, had been enriched for the Jersey breeds for the reference targets and it was spread across 39 different QTL located on 22 chromosomes. It was interesting to notice that the difference in targeting of genes between the reference and variant for this traits could be linked to the small stature and growth rate identified in Jersey breed (Albertí, 2008). Finally, four traits associated to QTL have been significantly enriched for the Fleckvieh breed. Two of these, calving ease and marbling score, were related to QTL enriched in variant targets and the two other, body weight at yearling stage and the percentage of fat in milk, were related to QTL enriched in reference targets.

4.3.5 Genotyping of 25 3' UTR SNPs across 25 cattle breeds

Genotyping of Irish beef and dairy cattle is an important process in the national genomic selection breeding program implemented to improve genetic gain in Irish cattle herds. Following the first International Dairy and Beef SNPs identification, a second set of SNPs was similarly designed to be integrated into the IDB 2.0. On this genotyping platform, 1,873 SNPs were introduced for research purposes, including 25 of the highest SSr-ranked SNPs in miRNA binding sites identified previously and related to dairy cattle or dairy production for the following association study. This analysis provided an overview of the frequency of the different alleles in Irish dairy and beef breed.

Out of the 25 breeds genotyped, 11 had a number of animals ranging from one to nine and therefore are not discussed here, as the number was too small to assess any Hardy-Weinberg equilibrium. All the breed allele frequencies are summarised in supplementary material (**Supplementary table 13**). Considering the 14 other breeds, the number of animals ranged from 42 (Piedmontese) to 10562 (Holstein-Friesian). Out of the 25 SNPs present in the 3' UTR of the top ranked potentially dysregulated genes selected for the genotyping, only 18 went through the process of integration on to the SNPchip. Out of these 18 SNPs, 10 presented a majority of homozygous allele, for the reference or the variant allele, with a frequency ranging from 99.5 to 100%. Among these 10 SNPs, six were homozygous for the variant allele in *PPP1R16B*, *TRGC-like*, *PLA2G2A*, *GIMAP1*, *SRPRB* and *GIMAP4* (locus 113,867,507) with a frequency ranging from 99.5 to 100% and four were homozygous for the reference allele in *CTCI*, *ERLIN1*, *GIMAP4* (locus 113,867,369) and *PI3-like*, with a frequency ranging from 97.6% to 100%.

Furthermore, by testing the Hardy-Weinberg equilibrium, allele frequencies showed significant differences with the estimated one for *CYP3A5*, *GUCA2B*, *LAMTOR2*, *HOXC8*, *RHOBTB2*, *PI3* and *BST2B*, suggesting a potential positive selection of these SNPs (**Figure 31**). By examining the allele frequency breed by breed the SNPs which significantly diverging to Hardy-Weinberg ($p\text{-value} < 0.01$) were present only in the Aubrac breed in *HOXC8*, for Charolais, Angus and Holstein-Friesian in *CYP3A5*, for all the breeds in *BST2B*, for Salers, Limousin, Hereford, Charolais, Belgian Blue and Holstein-Friesian in *PI3*, for Limousin, Charolais and Angus in *RHOBTB2*. These results showed a difference of variability in SNPs located in miRNA binding sites when some were almost totally absent of the observed population some others presented a frequency which suggested that positive selection could have influenced specific traits.

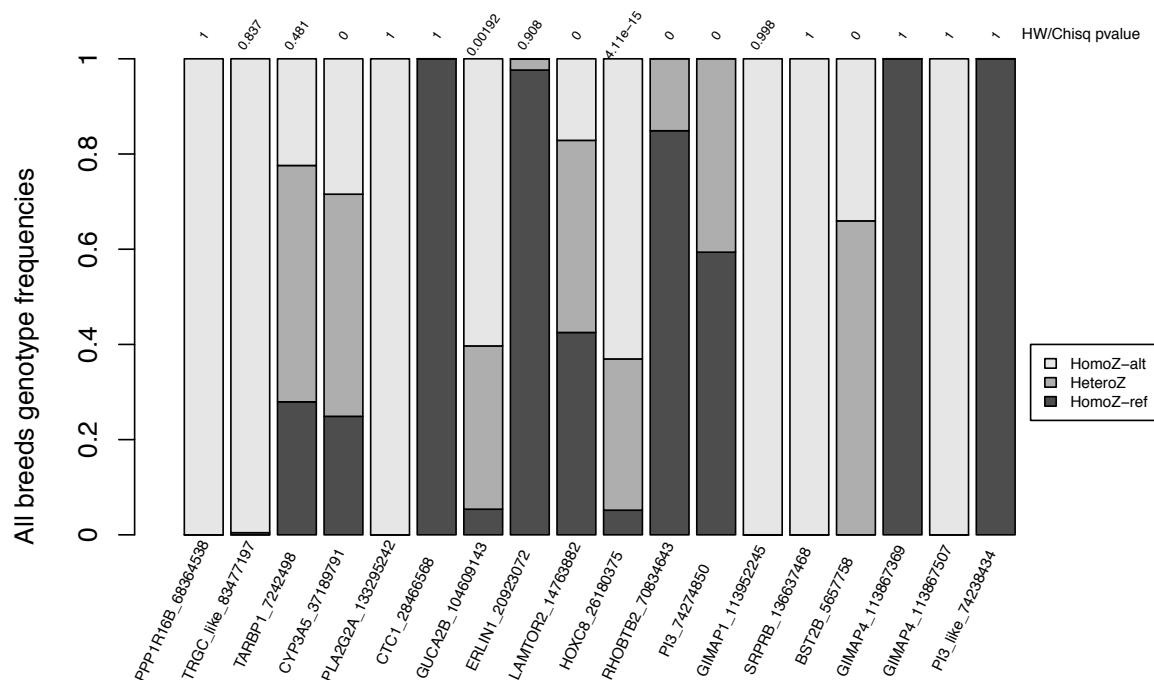


Figure 31 – Genotyping frequency of SNPs in miRNA binding sites across cattle breeds

Representation of the allele proportion for each selected SNPs located in 3' UTR miRNA binding sites investigated in whole individual across 25 breeds

4.3.6 Association analysis of genotyped SNPs in Holstein-Friesian breed

Following the genotyping, ten SNPs were selected to be integrated into a broader candidate gene association study. If the allele was detected in the genotyping, a second cutoff was applied based on a minimum frequency threshold and Hardy-Weinberg equilibrium. This resulted in ten SNPs, located in *LAMTOR2*, *TRGC-like*, *TARBPI*, *CYP3A5*, *GUCA2B*, *ERLIN1*, *HOXC8*, *RHOBTB2*, *PI3*, *BST2B* and *PI3-like* UTRs, to be included in the association study. The results showed that 10 traits were significantly enriched (nominal *p-value* <0.05) for 6 SNPs. After adjustment of the *p-value*, three traits were significantly associated with three of the candidate SNPs (adjusted *p-value* <0.05). The SNP located in the 3' UTR of *TARBPI* was associated with milk protein percentage (protper), the SNP located in the 3' UTR of *GUCA2B* was associated with milk fat percentage (fatper) and the SNP located in the 3' UTR of *CYP3A5* was associated with milk fat mass (fatkg). These three traits were all linked to milk production and could suggest a role of these three SNPs toward milk quality. The allele frequencies analysed previously showed that only *CYP3A5* presented a significant deviation of the Hardy-Weinberg equilibrium for the Holstein-Friesian set and was the most significant one overall the other breeds for this SNP. The different alleles were present in similar proportion for *CYP3A5* and *TARBPI*, however the variant *GUCA2B* was present in 79% of the animal of the Holstein-Friesian breed. The role of *GUCA2B* in gastro intestinal inflammation and satiety has been discussed previously and could suggest a modification of absorption of nutrient, modifying the milk quality. However, it is not clear how the polymorphisms in *CYP3A5* and *TARBPI* could affect respectively the fat mass and the protein percentage in milk (**Table 17**).

Overall, these traits associated with polymorphic miRNA binding sites showed the importance of regulatory sites in cattle breeding and that SNPs in 3' UTR were a component of traits modification, even if miRNAs binding sites were considered as highly conserved.

Table 17 – List of traits and SNPs tested for association with significant nominal *p*-value

SNP	trait	effect	SE	t	p-value	Q-value
ENSBTAG00000007194_104609143	fatper	0.0188	0.00422	4.45	8.77E-06	1.33E-03
ENSBTAG00000004517_7242498	protper	0.00634	0.00162	3.92	8.97E-05	6.82E-03
ENSBTAG00000005450_37189791	fkg	-0.814054	0.24024	-3.39	7.04E-04	3.57E-02
ENSBTAG000000031916_70834643	protper	0.00828	0.0029	2.86	4.25E-03	1.62E-01
ENSBTAG00000005450_37189791	fatper	-0.00996	0.00362	-2.75	5.98E-03	1.82E-01
ENSBTAG00000004517_7242498	fatper	0.00907	0.00347	2.62	8.82E-03	1.92E-01
ENSBTAG000000045588_5657758	ccf	-0.715461	0.26950	-2.65	8.07E-03	1.92E-01
ENSBTAG000000031916_70834643	fatper	0.0153	0.00619	2.46	1.39E-02	2.12E-01
ENSBTAG00000004517_7242498	su	-1.55391	0.62088	-2.5	1.25E-02	2.12E-01
ENSBTAG000000039028_74274850	scc	0.0724	0.0291	2.49	1.28E-02	2.12E-01
ENSBTAG00000005450_37189791	mcd	0.345115	0.14636	2.36	1.83E-02	2.53E-01
ENSBTAG00000004517_7242498	gest	-0.776255	0.33699	-2.3	2.15E-02	2.72E-01
ENSBTAG00000005450_37189791	ccf	0.387246	0.17567	2.2	2.79E-02	3.26E-01
ENSBTAG00000007194_104609143	cwt	-11.6236	5.50506	-2.11	3.49E-02	3.79E-01
ENSBTAG00000007194_104609143	protper	0.004	0.00197	2.02	4.34E-02	3.98E-01
ENSBTAG000000045588_5657758	fatper	0.0109	0.00541	2.01	4.45E-02	3.98E-01
ENSBTAG00000007194_104609143	mkg	-15.0961	7.45898	-2.02	4.34E-02	3.98E-01

4.3.7 Identification of polymorphism in 3' UTR related to miRNA binding sites in candidate genes.

A study on Holstein bull fertility identified a set of 697 SNPs in the 3' UTR of 83 genes (Mullen, 2012). Out of these 697 SNPs, a total of 251 SNPs, related to 62 genes were located in miRNA binding sites. Each polymorphic sites of each genes was compared to all the reference sites for these particular genes then ranked with the SSr method. The top ten ranked sites showed the best candidates that could undergo a modification of regulation by miRNAs (**Table 18**). The polymorphism in *SIRT2* created a new site that could be targeted by nine new miRNAs compared to the reference. Moreover, three occurrences of *MAP3K11*, corresponding to three polymorphisms creating new binding sites for a total of seven new miRNAs has also been identified.

Table 18 – List of top 10 polymorphic miRNA binding sites among a subset of 83 genes

Gene Ensembl ID	Gene name	Polymorphic sites	Common sites	SSR	Event: number of miRNA
ENSBTAG00000001776_48860657	<i>SIRT2</i>	9	86	-0.121	Gain:9
ENSBTAG00000009190_27622484	<i>SLC2A4</i>	7	131	-0.107	Gain:7
ENSBTAG00000020935_74141827	<i>HIF1A</i>	11	198	-0.077	Gain:11
ENSBTAG00000004114_44443481	<i>MAP3K11</i>	6	161	-0.040	Increase:4 Gain:2
ENSBTAG00000017220_48766012	<i>GH1</i>	7	212	-0.033	Increase:1 Gain:6
ENSBTAG00000004114_44443923	<i>MAP3K11</i>	5	161	-0.025	Increase:4 Gain:1
ENSBTAG00000004114_44443965	<i>MAP3K11</i>	4	161	-0.018	Increase:4
ENSBTAG00000000730_34410623	<i>GCG</i>	4	208	-0.014	Gain:4
ENSBTAG00000007825_42868329	<i>MARK2</i>	5	245	-0.012	Increase:3 Gain:2
ENSBTAG00000013003_50040081	<i>INS</i>	4	237	-0.011	Increase:4

SIRT2 is part of sirtuin protein family. This gene family have been shown to be tissue specifically expressed and involved in glucose metabolism (Ghinis-Hozumi, 2011). The modification of the regulation of this enzyme due to polymorphic miRNA binding sites could be of high importance in different cattle breed development. *MAP3K11* produce a kinase which have been shown to bind to the CRISP2 protein. This protein is expressed in testis and could have a role in spermatozoid mobility (Gibbs, 2007). The regulation of genes having an impact on gametes had a good potential for selection for reproduction improvement. In relation with the study made on Holstein bull fertility, *MAP3K11* SNPs in miRNA binding sites proved to be the best candidate.

4.4 Conclusions

This study confirmed the role of variant in miRNA binding sites as an active member of phenotypic variation in livestock as it has been observed in Texel sheep for myostatin variant and double-muscled trait. Although considered as highly conserved in animals, SNPs in 3' UTRs, related to miRNA binding sites have been shown not to be an uncommon event. Moreover, some of these SNPs have been experimentally proved to be associated with production traits in Holstein breed.

The identification of polymorphism in 3' UTR of cattle showed variability in the allelic frequencies between breeds which can be linked to artificial selection. Polymorphisms in miRNA binding sites could, if not totally impairing sites, modify the regulatory effect of miRNAs, therefore the expression of genes, pathway and traits related. This made these variants great candidates for selection as they do not knock down the gene function but could modulate its expression producing more controllable phenotypes. This regulation of expression could give moderate traits change, hence creating a tool for more precise selection.

Out of the 25 SNPs selected from the top SSr-ranked list as candidate for genotyping and association studies, only ten proved to be valid for the association. Out of these ten, three SNPs located in three genes, *TARBPI*, *GUCA2B*, *CYP3A5*, have been found to be significantly associated with milk production traits in Holstein cattle. Therefore, these genes had the potential to be interesting targets for selection and should be investigated further to decipher their putative role related to milk production. Furthermore, the addition of these SNPs in the International Dairy and Beef SNPchip contributed to the improvement of genomic selection cattle breeding in Ireland. It can also be noticed the preponderance of traits related to milk production in the QTL enrichment analysis as well as the high ranked gene *PRKAG1*, which had a polymorphism in 3' UTR which have been associated with milk production in another study. The role of *CYP17A1* in post-partum dairy cow in negative energy balance also presented an interest for the dairy cattle production to favour recovery after calving and a better use of feed resources.

Overall the genes and pathways enrichment identified a predominance of function related to immune system and diseases. It can be noticed that the vitamin C activator pathway was regrouping many candidate genes involved in immune response as well as the fMLP signalling in neutrophils. It can also be noticed the high ranking of *BTS2B*, *GIMAP1* and *LAMTOR2* also involved in immune response.

The variants in miRNA binding sites showed to be interesting candidate to improve cattle selection and further analysis on the association of the other SNPs identified in this study could bring new insight in the regulation of other important genes for beef and dairy production.

5 General discussion

Ireland is well placed as a global producer of beef (BordBia, 2015). Ireland currently exports 90% of national beef output, making Ireland the largest net beef exporter in the European Union (EU), and the fourth largest net beef exporter globally. Beef exports were worth approximately €1.9 billion to the Irish economy in 2012 (Renwick, 2013). However, it has been estimated that the global population will increase by 32% by 2050 (United Nation, 2015). This growing world populous is fuelling an increased demand for meat (Foley, 2011). This represented an ideal opportunity for Ireland to further benefit financially from beef exports. Similarly, Irish milk and dairy production are globally important. Ireland is one of the top countries having surpluses in milk production (FAO, 2008). In 2015, Irish milk production exports increased by 4%, representing €3.24 billion. Furthermore the end of European quota in 2015 on milk presented an opportunity to increase the production with an expected growth of 50% by 2020 (Fareilly, 2014). This places Ireland as a major contributor to address the global milk market demand. In order for Ireland to fully capitalise on this increased global demand for beef and milk, their production must be increased in a sustainable manner, while also enabling farmers to maximise their profits.

Traditionally, if there was an increased demand for food, farmers would increase total area farmed and herd size. This is known as intensification of farming (Garnett, 2009). This is no longer a feasible method to increase production because of the impact farming has on climate change (Garnett, 2009). In Ireland, agriculture is responsible for a third of the national greenhouse gas (GHG) emissions (Fitzsimons, 2013) thus an increase in the size of the national herd would further increase GHG emissions and would negatively impact Ireland's aim of reducing emissions by 40% before 2030 (Europea, 2014). Thus, there is a requirement to improve the economically important traits in beef and dairy cattle in a sustainable manner. This can be facilitated through the discovery and application of accurate, reliable and robust DNA based markers for these traits into the genomic selection and breeding programme, which will enhance the efficiency beef and dairy production.

Since the foundation of ICBF in 1998 and the first annual reports for cattle breeding was released in 2002, the breeding value of Irish cattle kept increasing. In

2009, a national genomic selection breeding programme has been set in Ireland to drive genetic improvement of the dairy herd. This has been achieved using the International Dairy and Beef (IDB) custom SNP chip (Mullen, 2013), which is now in its third iteration. Versions one, two, and three of the IDB chip were being used to improve the genetic merit of dairy cattle in Ireland and contained SNPs for traits such as milk production (Mullen, 2013). Similar progress is underway for beef through the launch of the national beef genomics scheme by the DAFM.

An important regulatory role of miRNA in beef and dairy cattle production has been recognized (H. Li, 2011; Muroya, 2013). This thesis examined polymorphisms in miRNA genes and associated targets in *Bos taurus* breeds and assessed their association with economically important beef and dairy cattle production traits. Candidate SNPs found to be associated with these traits of interest could then be added to the IDB v3 SNPchip and applied in national genomic evaluations in the breeding programme.

The cattle species, *Bos taurus*, in its European form originated from the Middle East, more than 10,000 years ago, from the domestication of *Bos. primigenius* relatives. The multiple generation of breeding and selection of different cattle for specific traits modified the appearance of the ancestral cattle from which the cattle breeds known today emerged. The improvement of genetics and genomics in the last decade produced a tremendous amount of data, which contributed to the identification of markers, improving the knowledge of the origination of *Bos Taurus* and the tools for cattle breeding.

The rise of marker assisted selection studies increased the production of genomic data and the identification of QTL. This information, together with the availability of full genome sequencing data have increased the opportunity for cattle genomic selection, including the accurate identification of genetic markers related to specific traits. These considerations led to the design of this project, which aimed to identify polymorphisms in miRNA and miRNA binding sites as candidate for cattle selection improvement. The miRNAs are small non-coding genes, which regulated protein-coding genes by binding to their 3' UTRs. These microRNAs have been extensively studied in cattle showing many traits have variation based on differential expression and in a few cases, genetic polymorphisms. The role of polymorphic miRNA binding sites has been shown to be responsible for double muscle traits in Texel sheep, which led to considering miRNA having a role in domestication and in cattle selection.

For the purpose of the identification of miRNA binding sites, TargetScan software has been used with custom file for the polymorphic sequences, based on SNP data collected from different sources.

The first analysis of miRNA and miRNA binding site in this thesis was conducted in association with the whole nuclear genome sequencing of *B. primigenius* study. The aim was to investigate the hypothesis that variants in miRNA and miRNA binding sites between *Bos taurus* and *Bos primigenius* were responsible for modification of the regulation of genes involved in the domestication process. The aim was to identify SNPs in miRNA and miRNA binding sites and identify the best candidate. For that purpose, a workflow based on TargetScan was developed and a ranking score, SSr was created. The polymorphisms were assessed by comparing the reference sequence of *B. taurus* to the *B. primigenius* genome assembly. The use of complete genome of other breeds, helped to assess if the SNPs were fixed in cattle or if the *B. primigenius* variants was still present in modern cattle. Only one miRNA gene, *mir-2893*, expressed in ovaries, has been identified having a polymorphic seed region. The shift of targets of *mir-2893* showed that regulation modifications could have occurred in neurodevelopment and food metabolism pathways. Furthermore, the analysis of polymorphisms in the 3' UTRs also led to a potential shift of miRNA targeting between *B. primigenius* and *B. taurus*. These changes also resulted in enrichment in immunological and fertility pathways. Overall these alterations were in line with domestication events, as the influence of early selection by human tended to converge in livestock with the pathways identified. These changes also gave an indication of major sites that could be further investigated for cattle breeding purposes.

Following this first analysis and the access to extensive breed-related polymorphism data, the analysis of SNPs in miRNA genes of cattle has been conducted to investigate the hypothesis that genetic variants in miRNA sequences between beef and dairy cattle breeds modified the regulation of genes involved in important economic traits and thus could be used as candidate for genomic selection. The SNPs from available data were mapped on all annotated miRNA structure (pre-miRNA, mature miRNA and seed region) adapting the workflow and the SSr score developed previously. Some of the SNPs in the pre-miRNA changed their MFE and their predicted secondary structure, which could destabilise them and impair their processing. Among the SNPs identified in mature miRNA and seed region, three major polymorphic miRNAs were identified: *mir-2419*, *mir-449b* and *mir-29e*. The SNPs potentially

changed their targets and the analysis of these targets found that these three miRNAs were involved in mammary development and maintenance through modification of the regulation of IL-17 or prolactin pathways, which could have an impact for milk production.

Finally, in the last chapter, the SNPs in the target sites were investigated. The aim was to investigate the hypothesis that genetic variants in 3' UTR between cattle breeds modified miRNA binding sites in these genes and be involved in the expression of important economic traits and thus could serve as useful candidates for genomic selection. It has been shown that polymorphisms in the 3' UTR of genes can impair miRNA binding-sites. A total of 726 genes having SNPs corresponding to miRNA binding sites were identified using the same workflow and SSr score. Among these polymorphic genes, three in particular were investigated: *TARBPI*, *GUCA2B*, *CYP3A5*. These three genes have been associated with milk production traits in a candidate gene association study on Holstein breed cattle. Furthermore, *PRKAG1* have been shown to have a SNP in a miRNA binding site. This SNP has also been previously related to milk production traits. The pathway enrichment of polymorphic miRNA targets showed mainly a link with immune response of which *BTS2B*, *GIMAP1* and *LAMTOR2* are the main candidates in the study.

The analysis of miRNA and miRNA binding sites polymorphism, associated with the SSr ranking system developed here, allowed the selection of candidate genes. These genes have been shown to have a function of importance for breeding and some of the SNPs in these genes were associated with milk production traits. Only a small fraction of the identified polymorphisms has been tested for genotyping and association when many others have also potential in beef and dairy breeds. The identification of these polymorphic miRNA and miRNA binding sites, for cattle selection could potentially explain polygenic phenotypes and variation in traits due to the wide range of function and genes that were shown to be differentially targeted and potentially differentially regulated by few variations in miRNA sequence. The data produced in this thesis constituted a base for further candidate gene association studies and can be used for the improvement of cattle selection.

The first step to improve the identification of candidate polymorphism in miRNA could be to infer the co-expression of the miRNA gene and its targets. It is potentially achievable using RNA-seq or microarrays combining miRNA genes and target gene profiling (Yin, 2008; Tesfaye, 2009; Fatima, 2014). However the temporal

and spatial features of miRNA expression hinder the design of a comprehensive analysis (Ziats, 2014). The SNPs identified in miRNA related genes can be grouped in three types: SNPs influencing the secondary structure (pre-miRNA), SNPs influencing the seed region (mature miRNA) and SNPs influencing the targeted genes.

Polymorphisms have been identified in the hairpin loop structure of miRNA. These polymorphisms have been shown to have an impact on the predicted secondary structure of the miRNA transcript. It could be interesting to evaluate *in vitro* if the hairpin loop variants modifications that were predicted had an impact on the loading of the pre-miRNA and therefore on their repression potential. For this purpose it will be necessary to assess the level of processing of these variant loop by measuring and comparing the level of mature miRNA in the reference and the variant forms using, for example, northern blot and quantitative RT-PCR (Chen, 2005; Duan, 2007).

The selection of candidate genes has been centred on the miRNA/miRNA target site recognition prediction. To integrate more variants in association studies it would be necessary to assess the potential of binding modification *in vitro*. It is possible to create a miRNA construct having the polymorphic sequence identified in the present set of data. The miRNA construct could be artificially expressed through a plasmid in different type of cattle cells possibly from different breeds. Then, it would be possible to measure the effect of the variant miRNA in comparison with the reference one using gene expression profiling method. This could be done by extracting the RNA product for each variant and reference sample then sequencing them or using microarrays to measure then compare the level of expression throughout the whole genome to identify variation of expression for the predicted targets (Willenbrock, 2009). In a similar way, the polymorphic targets can be tested, by creating a construct of the 3' UTR containing the SNPs identified inserted in a plasmid. It would be then possible to measure the level of degradation or translational repression using reporters (Kuhn, 2008; Thomson, 2011).

The association study conducted on Holstein dairy for which the candidate SNPs were selected, was focused on dairy production. Hence it would be interesting to include the identified SNPs into candidate gene association studies on other breeds like Angus or Limousine beef breeds or animals from a pool of breeds for more general traits like stature, carcass weight or muscle mass. Increasing the number of animals would also help to explain complex traits (Gibson, 2010).

The genotyping results from IDB project showed that the allele ratio corresponding to the variant identified vary among breeds and could be interesting candidate for selection. These SNPs are now included in the last version of the IDB SNPchip (v3) for which the number of animal tested currently increased in addition with phenotypic information on dairy and beef. This will permit in the future to conduct association studies on a larger scale for the selected SNPs allowing more accurate association on more diverse traits.

6 Conclusion

This study aimed to test the hypothesis that genetic variants between *B. taurus* and *B. primigenius* located in miRNA and miRNA binding sites influenced the domestication process by affecting important traits. Subsequently, it was hypothesised that genetic variants in miRNA and in miRNA binding sites between cattle breeds could influence traits of economic importance and serve as DNA based markers for future genomic selection breeding programmes. To address these hypotheses, a workflow and the SSr scoring system were developed, aiming to identify SNPs between aurochs and cattle miRNA and miRNA-binding sites and also in miRNAs and 3' UTR miRNA targets between modern cattle breeds.

One polymorphism in a miRNA seed region (*mir-2893*) and 1606 polymorphic miRNA target genes have been identified between aurochs and modern cattle. Genetic variation of these targets were shown to potentially affect major functions such as neurodevelopment, immunity, metabolism or pigmentation, which represents important traits in domestication, thus identifying these SNPs as interesting markers of domestication events in the bovine species.

The identification of SNPs in miRNA between different cattle breeds revealed 89 miRNA variants. The analysis of target shifts and their related function showed that three miRNAs, *mir-2419*, *mir-449b* and *mir-29e*, were strong candidates for modifying important phenotypes in cattle such as immune response to mastitis. These SNPs were incorporated on to the IDB SNPchip, which was used for genomic evaluations in the Irish national genomic selection breeding programme. An association study was conducted by genotyping using the IDB SNPchip which suggested an association between *mir-2419-3p* with rear udder height and *mir-1814b* with carcass conformation.

The SNPs discovered between cattle breed, which were mapped onto 3' UTR led to the identification of 726 genes containing 1097 different SNPs in their miRNA-binding sites. The functional analysis of these genes revealed important functions potentially playing important roles in traits of economic importance in cattle production, such as milk production and immune response. Some of the SNPs were included in a candidate SNP association study, which revealed that *TARBPI*, *GUCA2B* and *CYP3A5* binding site variants were associated with milk production traits.

Further investigation of the SNPs identified in this study is necessary and will involve larger association studies using beef and dairy cattle in the national herd. This is actually an on-going process as the SNPs identified in miRNA and miRNA binding sites are now incorporated onto the latest version of IDB SNPchip (version 3), and will be used for genotyping many hundreds of thousands of dairy and beef cattle across Ireland as part of the BGDGP project (BGDP, 2015). This will facilitate very powerful SNP association studies to be conducted in the future, particularly when comprehensive phenotypic information will be collected as part of the BDGP project. SNPs showing significant associations with traits of economic importance will be maintained on future versions of the chip and employed in future dairy and beef genomic selection breeding programme.

7 List of Appendices (Volume II)

Supplementary table 1 – List of all genes predicted to be affected by miR-2893 polymorphism between *B. primigenius* and *B. taurus*, ranked with SSr score

Supplementary table 2 – SSr-ranked genes with 3' UTR polymorphic miRNA binding sites between *B. taurus* and *B. primigenius*

Supplementary table 3 Polymorphic miRNA seed regions for TargetScan

Supplementary table 4 – Polymorphic mature miRNAs for TargetScan

Supplementary table 5 – List of all the breeds used for genotyping the 7 miRNAs SNPs

Supplementary table 6 – List of polymorphism in miRNA genes and their location for hairpin, mature and seed region, in six cattle breeds

Supplementary table 7 – Minimum free energy of reference and variant pre-miRNA structure of all polymorphic miRNAs

Supplementary table 8 – List of all breed-related polymorphic miRNAs and their targets, ranked by score

Supplementary table 9 – Ranked list of all significantly enriched pathways in miRNA targets for the different pathway groups

Supplementary table 10 – List of QTL terms enriched in miRNA variant targets

Supplementary table 11 – List of all alleles frequencies for the 7 miRNAs genotyped

Supplementary table 12 – List of polymorphic 3' UTR ranked by their miRNA binding score

Supplementary table 13 – Haplotype frequencies from genotyping of SNPs in miRNA binding sites on 25 cattle breeds

8 Publication

8.1 Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle

Stephen D E Park, David A. Magee, Paul A. McGettigan, Matthew D. Teasdale, Ceiridwen J. Edwards, Amanda J. Lohan, Alison Murphy, **Martin Braud**, Mark T. Donoghue, Yuan Liu, Andrew T. Chamberlain, Kévin Rue-Albrecht, Steven Schroeder, Charles Spillane, Shuaishuai Tai, Daniel G. Bradley, Tad S. Sonstegard, Brendan J. Loftus and David E. MacHugh

Genome Biology 2015 16:234 DOI: 10.1186/s13059-015-0790-2

Published 26th October 2015

8.2 Genome-Wide MicroRNA Binding Site Variation between Extinct Wild Aurochs and Modern Cattle Identifies Candidate microRNA-Regulated Domestication Genes

Martin Braud, David A. Magee, Stephen D. E. Park, Tad S. Sonstegard, Sinead M. Waters, David E. MacHugh, and Charles Spillane

Frontiers in Genetic 2017; 8: 3. DOI: 10.3389/fgene.2017.00003

Published online 31st January 2017

Bibliography

- Abd El Naby, W. S.Hagos, T. H.Hossain, M. M.Salilew-Wondim, D.Gad, A. Y.Rings, F.Cinar, M. U.Tholen, E.Loof, C.and Schellander, K. (2013) 'Expression analysis of regulatory microRNAs in bovine cumulus oocyte complex and preimplantation embryos', *Zygote*, 21(1), pp. 31–51.
- Abelson, J. F.Kwan, K. Y.O'Roak, B. J.Baek, D. Y.Stillman, A. A.Morgan, T. M.Mathews, C. A.Pauls, D. L.Rašin, M.-R.and Gunel, M. (2005) 'Sequence variants in SLITRK1 are associated with Tourette's syndrome', *Science*, 310(5746), pp. 317–320.
- Accorsi, P. A.Pacioni, B.Pezzi, C.Forni, M.Flint, D. J.and Seren, E. (2002) 'Role of prolactin, growth hormone and insulin-like growth factor 1 in mammary gland involution in the dairy cow', *Journal of dairy science*, 85(3), pp. 507–513.
- Achilli, A.Bonfiglio, S.Olivieri, A.Malusa, A.Pala, M.Kashani, B. H.Perego, U. A.Ajmone-Marsan, P.Liotta, L.and Semino, O. (2009) 'The multifaceted origin of taurine cattle reflected by the mitochondrial genome', *PLoS One*, 4(6), p. e5753.
- Achilli, A.Olivieri, A.Pellecchia, M.Uboldi, C.Colli, L.Al-Zahery, N.Accetturo, M.Pala, M.Kashani, B. H.and Perego, U. A. (2008) 'Mitochondrial genomes of extinct aurochs survive in domestic cattle', *Current Biology*, 18(4), pp. R157–R158.
- Agarwal, V.Bell, G. W.Nam, J.-W.and Bartel, D. P. (2015) 'Predicting effective microRNA target sites in mammalian mRNAs', *Elife*, 4, p. e05005.
- Ajmone-Marsan, P.Garcia, J. F.and Lenstra, J. A. (2010) 'On the origin of cattle: how aurochs became cattle and colonized the world', *Evolutionary Anthropology: Issues, News, and Reviews*, 19(4), pp. 148–157.
- Alb, J. G.Phillips, S. E.Rostand, K.Cui, X.Pinxteren, J.Cotlin, L.Manning, T.Guo, S.York, J. D.and Sontheimer, H. (2002) 'Genetic ablation of phosphatidylinositol transfer protein function in murine embryonic stem cells', *Molecular biology of the cell*, 13(3), pp. 739–754.
- Albertí, P.Panea, B.Sañudo, C.Olleta, J. L.Ripoll, G.Ertbjerg, P.Christensen, M.Gigli, S.Failla, S.and Concetti, S. (2008) 'Live weight, body size and carcass characteristics of young bulls of fifteen European breeds', *Livestock Science*, 114(1), pp. 19–30.
- Altuvia, Y.Landgraf, P.Lithwick, G.Elefant, N.Pfeffer, S.Aravin, A.Brownstein, M. J.Tuschl, T.and Margalit, H. (2005) 'Clustering and conservation patterns of human microRNAs', *Nucleic acids research*. Oxford Univ Press, 33(8), pp. 2697–2706.

- Amaral, A. J.Ferretti, L.Megens, H.-J.Crooijmans, R. P. M. A.Nie, H.Ramos-Onsins, S. E.Perez-Enciso, M.Schook, L. B.and Groenen, M. A. M. (2011) 'Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA', *PloS one*. Public Library of Science, 6(4), p. e14782.
- Ambros, V. (1989) 'A hierarchy of regulatory genes controls a larva-to-adult developmental switch in *C. elegans*', *Cell*. Elsevier, 57(1), pp. 49–57.
- Ambros, V. (2004) 'The functions of animal microRNAs', *Nature*, 431(7006), pp. 350–355.
- Ambros, V. and Horvitz, H. R. (1984) 'Heterochronic mutants of the nematode *Caenorhabditis elegans*', *Science*, 226(4673), pp. 409–416.
- Ambros, V. and Horvitz, H. R. (1987) 'The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events', *Genes & Development*, 1(4), pp. 398–414.
- An, X.Song, Y.Bu, S.Ma, H.Gao, K.Hou, J.Wang, S.Lei, Z.and Cao, B. (2016) 'Association of polymorphisms at the microRNA binding site of the caprine *KITLG* 3'-UTR with litter size', *Scientific reports*. Nature Publishing Group, 6.
- Andersen, J. M.Myhre, O.and Fonnum, F. (2003) 'Discussion of the role of the extracellular signal-regulated kinase-phospholipase A2 pathway in production of reactive oxygen species in Alzheimer's disease', *Neurochemical research*, 28(2), pp. 319–326.
- Andersson, L. and Georges, M. (2004) 'Domestic-animal genomics: deciphering the genetics of complex traits', *Nature Reviews Genetics*, 5(3), pp. 202–212.
- Arrillaga, C. G.Henning, W. L.and Miller, R. C. (1952) 'The effects of environmental temperature and relative humidity on the acclimation of cattle to the tropics', *Journal of Animal Science*. The American Society of Animal Science, 11(1), pp. 50–60.
- Auchtung, T. L.Rius, A. G.Kendall, P. E.McFadden, T. B.and Dahl, G. E. (2005) 'Effects of photoperiod during the dry period on prolactin, prolactin receptor, and milk production of dairy cows', *Journal of dairy science*, 88(1), pp. 121–127.
- Baek, D.Villén, J.Shin, C.Camargo, F. D.Gygi, S. P.and Bartel, D. P. (2008) 'The impact of microRNAs on protein output', *Nature*, 455(7209), pp. 64–71.
- Bao, B. and Garverick, H. A. (1998) 'Expression of steroidogenic enzyme and gonadotropin receptor genes in bovine follicles during ovarian follicular waves: a review', *Journal of animal science*, 76(7), pp. 1903–1921.
- Baranov, K. O.Volkova, O. Y.Mechetina, L. VChikaev, N. A.Reshetnikova, E. S.Nikulina, G. M.Taranin, A. Vand Najakshin, A. M. (2012) 'Expression of human B-Cell specific receptor FCRL1 in healthy individuals and in patients with autoimmune diseases',

- Molecular Biology*, 46(3), pp. 450–456.
- Benetti-Pinto, C. L.Vale, D.Garmes, H.and Bedone, A. (2007) ‘17-Hydroxyprogesterone deficiency as a cause of sexual infantilism and arterial hypertension: Laboratory and molecular diagnosis-a case report’, *Gynecological endocrinology*, 23(2), pp. 94–98.
- Bermingham, M. L.Bishop, S. C.Woolliams, J. A.Pong-Wong, R.Allen, A. R.McBride, S. H.Ryder, J. J.Wright, D. M.Skuce, R. A.and McDowell, S. W. J. (2014) ‘Genome-wide association study identifies novel loci associated with resistance to bovine tuberculosis’, *Heredity*.
- Berry, D. P.Bastiaansen, J. W. M.Veerkamp, R. F.Wijga, S.Wall, E.Berglund, B.and Calus, M. P. L. (2012) ‘Genome-wide associations for fertility traits in Holstein-Friesian dairy cows using data from experimental research herds in four European countries’, *animal*, 6(8), pp. 1206–1215.
- Berry, D. P.Kearney, F.and Harris, B. L. (2009) ‘Genomic selection in Ireland’, *Interbull Bulletin*, (39), p. 29.
- Berry, D. P.Meade, K. G.Mullen, M. P.Butler, S.Diskin, M. G.Morris, D.and Creevey, C. J. (2011) ‘The integration of ,’omic’ disciplines and systems biology in cattle breeding’, *animal*, 5(4), pp. 493–505.
- Berry, D. P.Shalloo, L.Cromie, A. R.Veerkamp, R. F.Dillion, P.Amer, P. R.Kearney, J. F.Evans, R. D.and Wickham, B. (2007) ‘The economic breeding index: a generation on’, *ICB Federation*, pp. 29–34.
- BGDP (2015) *Beef Genomic and Data Programme*.
- Bickhart, D. M.Hou, Y.Schroeder, S. G.Alkan, C.Cardone, M. F.Matukumalli, L. K.Song, J.Schnabel, R. D.Ventura, M.Taylor, J. F.Garcia, J. F.Van Tassell, C. P.Sonstegard, T. S.Eichler, E. E.and Liu, G. E. (2012) ‘Copy number variation of individual cattle genomes using next-generation sequencing’, *Genome Res*.
- Bohn, G.Allroth, A.Brandes, G.Thiel, J.Glocker, E.Schäffer, A. A.Rathinam, C.Taub, N.Teis, D.and Zeidler, C. (2007) ‘A novel human primary immunodeficiency syndrome caused by deficiency of the endosomal adaptor protein p14’, *Nature medicine*, 13(1), pp. 38–45.
- Boichard, D.Chung, H.Dassonneville, R.David, X.Eggen, A.Fritz, S.Gietzen, K. J.Hayes, B. J.Lawley, C. T.and Sonstegard, T. S. (2012) ‘Design of a bovine low-density SNP array optimized for imputation’, *PLoS One*, 7(3), p. e34130.
- Bolormaa, S.Pryce, J. E.Kemper, K.Savin, K.Hayes, B. J.Barendse, W.Zhang, Y.Reich, C. M.Mason, B. A.and Bunch, R. J. (2013) ‘Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in, and composite beef

- cattle', *Journal of animal science*. American Society of Animal Science, 91(7), pp. 3088–3104.
- Bonfiglio, S.Achilli, A.Olivieri, A.Negrini, R.Colli, L.Liotta, L.Ajmone-Marsan, P.Torroni, A.and Ferretti, L. (2010) 'The enigmatic origin of bovine mtDNA haplogroup R: sporadic interbreeding or an independent event of *Bos primigenius* domestication in Italy?', *PLoS One*, 5(12), p. e15760.
- Bonfiglio, S.Ginja, C.De Gaetano, A.Achilli, A.Olivieri, A.Colli, L.Tesfaye, K.Agha, S. H.Gama, L. T.and Cattonaro, F. (2012) 'Origin and spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1', *PLoS One*, 7(6), p. e38601.
- Bonnet, E.Wuyts, J.Rouzé, P.and Van de Peer, Y. (2004) 'Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences', *Bioinformatics*. Oxford Univ Press, 20(17), pp. 2911–2917.
- BordBia (2015) 'Meat and livestock review and outlook 2014/15'.
- Bradley, D. G. and Magee, D. A. (2006) 'Genetics and the origins of domestic cattle', *Documenting domestication: new genetic and archaeological paradigms*, pp. 317–328.
- Braud, M.Magee, D. A.Park, S. D. E.Sonstegard, T. S.Waters, S. M.MacHugh, D. E.and Spillane, C. (2017) 'Genome-wide microRNA binding site variation between extinct wild aurochs and modern cattle identifies candidate microRNA-regulated domestication genes', *Frontiers in Genetics*. Frontiers Media SA, 8.
- Brenna, Ø.Bruland, T.Furnes, M. W.Granlund, A. van B.Drozdo, I.Emgård, J.Brønstad, G.Kidd, M.Sandvik, A. K.and Gustafsson, B. I. (2015) 'The guanylate cyclase-C signaling pathway is down-regulated in inflammatory bowel disease', *Scandinavian journal of gastroenterology*, (0), pp. 1–12.
- Briggs, A. W.Stenzel, U.Johnson, P. L. F.Green, R. E.Kelso, J.Prüfer, K.Meyer, M.Krause, J.Ronan, M. T.and Lachmann, M. (2007) 'Patterns of damage in genomic DNA sequences from a Neandertal', *Proceedings of the National Academy of Sciences*. National Acad Sciences, 104(37), pp. 14616–14621.
- Bruno, D. R.Rossitto, P. V.Bruno, R. G. S.Blanchard, M. T.Sitt, T.Yeargan, B. VSmith, W. L.Cullor, J. S.and Stott, J. L. (2010) 'Differential levels of mRNA transcripts encoding immunologic mediators in mammary gland secretions from dairy cows with subclinical environmental *Streptococci* infections', *Veterinary immunology and immunopathology*, 138(1), pp. 15–24.
- Burden, S. and Yarden, Y. (1997) 'Neuregulins and their receptors: a versatile signaling module in organogenesis and oncogenesis', *Neuron*, 18(6), pp. 847–855.

- Canavez, F. C. Luche, D. D. Stothard, P. Leite, K. R. Sousa-Canavez, J. M. Plastow, G. Meidanis, J. Souza, M. A. Feijao, P. Moore, S. S. and Camara-Lopes, L. H. (2012) 'Genome Sequence and Assembly of *Bos indicus*', *J Hered.*
- Cantiello, M. Giantin, M. Carletti, M. Lopparelli, R. M. Capolongo, F. Lasserre, F. Bollo, E. Nebbia, C. Martin, P. G. P. and Pineau, T. (2009) 'Effects of dexamethasone, administered for growth promoting purposes, upon the hepatic cytochrome P450 3A expression in the veal calf', *Biochemical pharmacology*, 77(3), pp. 451–463.
- Carneiro, M. Rubin, C.-J. Di Palma, F. Albert, F. W. Alföldi, J. Barrio, A. M. Pielberg, G. Rafati, N. Sayyab, S. and Turner-Maier, J. (2014) 'Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication', *Science*, 345(6200), pp. 1074–1079.
- Carrington, J. C. and Ambros, V. (2003) 'Role of microRNAs in plant and animal development', *Science*, 301(5631), pp. 336–338.
- Carthew, R. W. and Sontheimer, E. J. (2009) 'Origins and mechanisms of miRNAs and siRNAs', *Cell*, 136(4), pp. 642–655.
- Chabory, E. Damon, C. Lenoir, A. Henry-Berger, J. Vernet, P. Cadet, R. Saez, F. and Drevet, J. R. (2010) 'Mammalian glutathione peroxidases control acquisition and maintenance of spermatozoa integrity', *Journal of animal science*, 88(4), pp. 1321–1331.
- Chakravarthy, S. Sternberg, S. H. Kellenberger, C. A. and Doudna, J. A. (2010) 'Substrate-specific kinetics of Dicer-catalyzed RNA processing', *Journal of molecular biology*, 404(3), pp. 392–402.
- Chalfie, M. Horvitz, H. R. and Sulston, J. E. (1981) 'Mutations that lead to reiterations in the cell lineages of *C. elegans*', *Cell*. Elsevier, 24(1), pp. 59–69.
- Chambaz, A. Scheeder, M. R. L. Kreuzer, M. and Dufey, P.-A. (2003) 'Meat quality of Angus, Simmental, Charolais and Limousin steers compared at the same intramuscular fat content', *Meat Science*. Elsevier, 63(4), pp. 491–500.
- Charlier, C. Coppieters, W. Rollin, F. Desmecht, D. Agerholm, J. S. Cambisano, N. Carta, E. Dardano, S. Dive, M. and Fasquelle, C. (2008) 'Highly effective SNP-based association mapping and management of recessive defects in livestock', *Nat Genet*, 40(4), pp. 449–454.
- Chekulaeva, M. and Filipowicz, W. (2009) 'Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells', *Current opinion in cell biology*, 21(3), pp. 452–460.
- Chen, C. Ridzon, D. A. Broomer, A. J. Zhou, Z. Lee, D. H. Nguyen, J. T. Barbisin, M. Xu, N. L. Mahuvakar, V. R. and Andersen, M. R. (2005) 'Real-time quantification of microRNAs

- by stem-loop RT-PCR', *Nucleic acids research*. Oxford Univ Press, 33(20), pp. e179–e179.
- Chen, K. and Rajewsky, N. (2006) 'Natural selection on human microRNA binding sites inferred from SNP data', *Nat Genet*, 38(12), pp. 1452–1456.
- Chen, L.Zeng, Y.Yang, H.Lee, T. D.French, S. W.Corrales, F. J.García-Trevijano, E. R.Avila, M. A.Mato, J. M.and Lu, S. C. (2004) 'Impaired liver regeneration in mice lacking methionine adenosyltransferase 1A', *The FASEB Journal*, 18(7), pp. 914–916.
- Chen, S.Lin, B.-Z.Baig, M.Mitra, B.Lopes, R. J.Santos, A. M.Magee, D. A.Azevedo, M.Tarroso, P.and Sasazaki, S. (2010) 'Zebu cattle are an exclusive legacy of the South Asia Neolithic', *Molecular biology and evolution*, 27(1), pp. 1–6.
- Chen, X.Paranjape, T.Stahlhut, C.McVeigh, T.Keane, F.Nallur, S.Miller, N.Kerin, M.Deng, Y.and Yao, X. (2015) 'Targeted resequencing of the microRNAome and 3' UTRome reveals functional germline DNA variants with altered prevalence in epithelial ovarian cancer', *Oncogene*. Nature Publishing Group, 34(16), pp. 2125–2137.
- Chikhi, L.Goossens, B.Treanor, A.and Bruford, M. W. (2004) 'Population genetic structure of and inbreeding in an insular cattle breed, the Jersey, and its implications for genetic resource management', *Heredity*. Nature Publishing Group, 92(5), pp. 396–401.
- Chin, L. J.Ratner, E.Leng, S.Zhai, R.Nallur, S.Babar, I.Muller, R.-U.Straka, E.Su, L.and Burki, E. A. (2008) 'A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk', *Cancer research*, 68(20), pp. 8535–8540.
- Clark, J. D.Schievella, A. R.Nalefski, E. A.and Lin, L.-L. (1995) 'Cytosolic phospholipase A₂', *Journal of lipid mediators and cell signalling*, 12(2), pp. 83–117.
- Cleaveland, S.Laurenson, M. K.and Taylor, L. H. (2001) 'Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1411), pp. 991–999.
- Clop, A.Marcq, F.Takeda, H.Pirottin, D.Tordoir, X.Bibe, B.Bouix, J.Caiment, F.Elsen, J. M.Eychenne, F.Larzul, C.Laville, E.Meish, F.Milenkovic, D.Tobin, J.Charlier, C.and Georges, M. (2006) 'A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep', *Nat Genet*, 38(7), pp. 813–818.
- Cohen-Zinder, M.Seroussi, E.Larkin, D. M.Loor, J. J.Everts-van der Wind, A.Lee, J.-H.Drackley, J. K.Band, M. R.Hernandez, A. G.and Shani, M. (2005) 'Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on

- chromosome 6 affecting milk yield and composition in Holstein cattle', *Genome Res*, 15(7), pp. 936–944.
- Comerford, J. W. Bertrand, J. K. Benyshek, L. L. and Johnson, M. H. (1987) 'Reproductive rates, birth weight, calving ease and 24-h calf survival in a four-breed diallel among Simmental, Limousin, Polled Hereford and Brahman beef cattle', *Journal of animal science*. The American Society of Animal Science, 64(1), pp. 65–76.
- Comin, A. Cassandro, M. Chessa, S. Ojala, M. Dal Zotto, R. De Marchi, M. Carnier, P. Gallo, L. Pagnacco, G. and Bittante, G. (2008) 'Effects of Composite β - and κ -Casein Genotypes on Milk Coagulation, Quality, and Yield Traits in Italian Holstein Cows', *Journal of dairy science*, 91(10), pp. 4022–4027.
- Conaway, R. C. Sato, S. Tomomori-Sato, C. Yao, T. and Conaway, J. W. (2005) 'The mammalian Mediator complex and its role in transcriptional regulation', *Trends in biochemical sciences*, 30(5), pp. 250–255.
- Consortium, B. H. (2009) 'Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds', *Science*, 324(5926), pp. 528–532.
- Cortés, O. Tupac-Yupanqui, I. Dunner, S. García-Atance, M. A. García, D. Fernández, J. and Cañón, J. (2008) 'Ancestral matrilineages and mitochondrial DNA diversity of the Lidia cattle breed', *Animal genetics*, 39(6), pp. 649–654.
- Cromie, A. Kearney, F. Evans, R. and Berry, D. P. (2014) 'Genomics for Pedigree and Cross-bred Beef Cattle Populations; Some experiences from Ireland', in *10th World Congress on Genetics Applied to Livestock Production*. Asas.
- Daetwyler, H. D. Capitan, A. Pausch, H. Stothard, P. Van Binsbergen, R. Brøndum, R. F. Liao, X. Djari, A. Rodriguez, S. C. and Grohs, C. (2014) 'Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle', *Nat Genet*, 46(8), pp. 858–865.
- Daetwyler, H. D. Villanueva, B. Bijma, P. and Woolliams, J. A. (2007) 'Inbreeding in genome-wide selection', *Journal of Animal Breeding and Genetics*. Wiley Online Library, 124(6), pp. 369–376.
- Daetwyler, H. D. Wiggans, G. R. Hayes, B. J. Woolliams, J. A. and Goddard, M. E. (2011) 'Imputation of missing genotypes from sparse to high density using long-range phasing', *Genetics*, 189(1), pp. 317–327.
- Daniels, S. M. Melendez-Peña, C. E. Scarborough, R. J. Daher, A. Christensen, H. S. El Far, M. Purcell, D. F. J. Lainé, S. and Gatignol, A. (2009) 'Characterization of the TRBP domain

- required for dicer interaction and function in RNA interference', *BMC molecular biology*, 10(1), p. 38.
- Dann, S. G.Selvaraj, A.and Thomas, G. (2007) 'mTOR Complex1–S6K1 signaling: at the crossroads of obesity, diabetes and cancer', *Trends in molecular medicine*, 13(6), pp. 252–259.
- Dillon, K. (2008) *Holstein Friesians in Ireland*, *World Holstein Friesian Federation*.
- Druet, T. and Georges, M. (2010) 'A hidden Markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping', *Genetics*, 184(3), pp. 789–798.
- Duan, R.Pak, C.and Jin, P. (2007) 'Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA', *Human molecular genetics*, 16(9), pp. 1124–1131.
- Edwards, C. J.Bollongino, R.Scheu, A.Chamberlain, A.Tresset, A.Vigne, J.-D.Baird, J. F.Larson, G.Ho, S. Y. W.and Heupink, T. H. (2007) 'Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs', *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1616), pp. 1377–1385.
- Edwards, C. J.Magee, D. A.Park, S. D. E.McGettigan, P. A.Lohan, A. J.Murphy, A.Finlay, E. K.Shapiro, B.Chamberlain, A. T.and Richards, M. B. (2010) 'A complete mitochondrial genome sequence from a mesolithic wild aurochs (*Bos primigenius*)', *PLoS One*, 5(2), p. e9255.
- 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.Guigo, 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.Camara, 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. VEyras, 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.Lau, L. P.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.Iivanainen, A.Lahmers, K. K.Bennett, A. K.Dickens, C. M.Gilbert, J. G.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.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. S.Qin, X.Ren, Y.Shen, Y.Song, H.Bell, S. N.Davis, C.Johnson, A. J.Lee, S.Nazareth, L. VPatel, B. M.Pu, L. L.Vattathil, S.Williams Jr., 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.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.de Oliveira, C. J.Deobald, H.Devinoy, E.Donohue, K. E.Dovc, 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.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.More, D. D.Muntean, C. T.Nandakumar, H. P.Nogueira, M. F.Olsaker, I.Pant, S. D.Panzitta, F.Pastor, R. C.Poli, M. A.Poslusny, N.Rachagani, S.Ranganathan, S.Razpet, A.Riggs, P. K.Rincon, G.Rodriguez-Ororio, 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. VTammen, I.Telugu, B. P.Urbanski, J. M.Utsunomiya, Y. T.Verschoor, C. P.Waardenberg, A. J.Wang, Z.Ward, R.Weikard, R.Welsh Jr., T. H.White, S. N.Wilming, L. G.Wunderlich, K. R.Yang, J.and Zhao, F. Q. (2009) 'The genome sequence of taurine cattle: a window to ruminant biology and evolution', *Science*, 324(5926), pp. 522–528.

- Enright, A. J. John, B. Gaul, U. Tuschl, T. Sander, C. and Marks, D. S. (2004) 'MicroRNA targets in *Drosophila*', *Genome Biology*, 5(1), pp. R1–R1.
- Esmailzadeh, A. K. Morris, C. A. Cullen, N. G. Kruk, Z. A. Lines, D. S. Hickey, S. M. Dobbie, P. M. Bottema, C. D. K. and Pitchford, W. S. (2011) 'Genetic mapping of quantitative trait loci for meat quality and muscle metabolic traits in cattle', *Animal genetics*, 42(6), pp. 592–599.
- Europea, C. (2014) 'A policy framework for climate and energy in the period from 2020 to 2030', *COM (2014)*, 15.
- Fagan, K. A. McMurtry, I. F. and Rodman, D. M. (2001) 'Role of endothelin-1 in lung disease', *Respiratory research*, 2(2), p. 90.
- FAO (2008) 'Global Dairy Sector: Status and Trends'.
- Farely, P. and Crosse, S. (2014) 'The Department of Agriculture, Food and the Marine'.
- Fatima, A. and Morris, D. G. (2013) 'MicroRNAs in domestic livestock', *Physiological genomics*, 45(16), pp. 685–696.
- Fatima, A. Waters, S. O'Boyle, P. Seoighe, C. and Morris, D. G. (2014) 'Alterations in hepatic miRNA expression during negative energy balance in postpartum dairy cattle', *BMC genomics*. BioMed Central, 15(1), p. 28.
- Felius, M. Koolmees, P. A. Theunissen, B. Lenstra, J. A. and Consortium, E. C. G. D. (2011) 'On the breeds of cattle—historic and current classifications', *Diversity*. Molecular Diversity Preservation International, 3(4), pp. 660–692.
- Feuermann, Y. Mabeesh, S. J. and Shamay, A. (2004) 'Leptin affects prolactin action on milk protein and fat synthesis in the bovine mammary gland', *Journal of dairy science*, 87(9), pp. 2941–2946.
- Finlay, E. K. Berry, D. P. Wickham, B. Gormley, E. P. and Bradley, D. G. (2012) 'A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle', *PLoS One*, 7(2), p. e30545.
- Fitzsimons, C. Kenny, D. A. Deighton, M. H. Fahey, A. G. and McGee, M. (2013) 'Methane emissions, body composition, and rumen fermentation traits of beef heifers differing in residual feed intake', *Journal of animal science*. American Society of Animal Science, 91(12), pp. 5789–5800.
- Fleckvieh Society of Australia* (2011) www.fleckvieh.org.au.
- Flynt, A. S. Greimann, J. C. Chung, W.-J. Lima, C. D. and Lai, E. C. (2010) 'MicroRNA Biogenesis via Splicing and Exosome-Mediated Trimming in *Drosophila*', *Mol Cell*, 38(6), pp. 900–907.

- Foley, J. A. Ramankutty, N. Brauman, K. A. Cassidy, E. S. Gerber, J. S. Johnston, M. Mueller, N. D. O'Connell, C. Ray, D. K. and West, P. C. (2011) 'Solutions for a cultivated planet', *Nature*. Nature Publishing Group, 478(7369), pp. 337–342.
- Food and Agriculture Organisation (2015) 'The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture', eds B.D. Scherf and D. Pilling, *Rome FAO*. eds B.D. Scherf and D. Pilling, Rome (Italy) FAO.
- Forsyth, I. A. (1986) 'Variation among species in the endocrine control of mammary growth and function: the roles of prolactin, growth hormone, and placental lactogen', *Journal of dairy science*, 69(3), pp. 886–903.
- Fric, J. Zelante, T. Wong, A. Y. W. Mertes, A. Yu, H.-B. and Ricciardi-Castagnoli, P. (2012) 'NFAT control of innate immunity', *Blood*, 120(7), pp. 1380–1389.
- Friedman, R. C. Farh, K. K. Burge, C. B. and Bartel, D. P. (2009) 'Most mammalian mRNAs are conserved targets of microRNAs', *Genome Res.* 2008/10/29, 19(1), pp. 92–105. doi: 10.1101/gr.082701.108.
- Garcia, D. M. Baek, D. Shin, C. Bell, G. W. Grimson, A. and Bartel, D. P. (2011) 'Weak seed-pairing stability and high target-site abundance decrease the proficiency of lsy-6 and other microRNAs', *Nat Struct Mol Biol.* 2011/09/13, 18(10), pp. 1139–1146. doi: 10.1038/nsmb.2115.
- Garnett, T. (2009) 'Livestock-related greenhouse gas emissions: impacts and options for policy makers', *environmental science & policy*. Elsevier, 12(4), pp. 491–503.
- Gellner, K. Eiselt, R. Hustert, E. Arnold, H. Koch, I. Haberl, M. Deglmann, C. J. Burk, O. Buntfuss, D. and Escher, S. (2001) 'Genomic organization of the human CYP3A locus: identification of a new, inducible CYP3A gene', *Pharmacogenetics and Genomics*, 11(2), pp. 111–121.
- Georges, M. and Andersson, L. (1996) 'Livestock genomics comes of age.', *Genome Research*. Cold Spring Harbor Lab, 6(10), pp. 907–921.
- Georges, M. Nielsen, D. Mackinnon, M. Mishra, A. Okimoto, R. Pasquino, A. T. Sargeant, L. S. Sorensen, A. Steele, M. R. and Zhao, X. (1995) 'Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing', *Genetics*, 139(2), pp. 907–920.
- Ghinis-Hozumi, Y. Gonzalez-Gallardo, A. Gonzalez-Dávalos, L. Antaramian, A. Villarroja, F. Shimada, A. Varela-Echavarría, A. and Mora, O. (2011) 'Bovine sirtuins: Initial characterization and expression of sirtuins 1 and 3 in liver, muscle, and adipose tissue', *Journal of animal science*, 89(8), pp. 2529–2536.

- Gibbs, G. M. Bianco, D. M. Jamsai, D. Herlihy, A. Ristevski, S. Aitken, R. J. de Kretser, D. M. and O'Bryan, M. K. (2007) 'Cysteine-rich secretory protein 2 binds to mitogen-activated protein kinase kinase kinase 11 in mouse sperm', *Biology of reproduction*, 77(1), pp. 108–114.
- Gibson, G. (2010) 'Hints of hidden heritability in GWAS', *Nat Genet*, 42(7), pp. 558–560.
- Gilmour, A. R. Gogel, B. J. Cullis, B. R. Thompson, R. and Butler, D. (2009) 'ASReml user guide release 3.0', *VSN International Ltd, Hemel Hempstead, UK*.
- Glazier, A. M. Nadeau, J. H. and Aitman, T. J. (2002) 'Finding genes that underlie complex traits', *Science*, 298(5602), pp. 2345–2349.
- Golby, P. Villarreal-Ramos, B. Dean, G. Jones, G. J. and Vordermeier, M. (2014) 'MicroRNA expression profiling of PPD-B stimulated PBMC from M. bovis-challenged unvaccinated and BCG vaccinated cattle', *Vaccine*, 32(44), pp. 5839–5844.
- Gong, J. Tong, Y. Zhang, H. Wang, K. Hu, T. Shan, G. Sun, J. and Guo, A. (2012) 'Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis', *Human mutation*. Wiley Online Library, 33(1), pp. 254–263.
- Govindaraju, A. Uzun, A. Robertson, L. Atli, M. O. Kaya, A. Topper, E. Crate, E. A. Padbury, J. Perkins, A. and Memili, E. (2012) 'Dynamics of microRNAs in bull spermatozoa', *Reprod Biol Endocrinol*, 10(82), pp. 1–10.
- Grant, K. Loizidou, M. and Taylor, I. (2003) 'Endothelin-1: a multifunctional molecule in cancer', *British journal of cancer*, 88(2), pp. 163–166.
- Griffiths-Jones, S. Grocock, R. J. Van Dongen, S. Bateman, A. and Enright, A. J. (2006) 'miRBase: microRNA sequences, targets and gene nomenclature', *Nucleic acids research*, 34(suppl 1), pp. D140–D144.
- Grigson, C. (1978) 'The craniology and relationships of four species of Bos: 4. The Relationship between Bos primigenius Boj. and B. taurus L. and its implications for the Phylogeny of the Domestic Breeds', *Journal of Archaeological Science*, 5(2), pp. 123–152.
- Grigson, C. (1980) 'The craniology and relationships of four species of Bos 5. Bos iudicus L', *Journal of Archaeological Science*, 7(1), pp. 3–32.
- Grimson, A. Farh, K. K. Johnston, W. K. Garrett-Engele, P. Lim, L. P. and Bartel, D. P. (2007) 'MicroRNA targeting specificity in mammals: determinants beyond seed pairing', *Mol Cell*. 2007/07/07, 27(1), pp. 91–105. doi: 10.1016/j.molcel.2007.06.017.
- Grobet, L. Martin, L. J. R. Poncelet, D. Pirottin, D. Brouwers, B. Riquet, J. Schoeberlein,

- A.Dunner, S.Ménissier, F.and Massabanda, J. (1997) ‘A deletion in the bovine myostatin gene causes the double-musled phenotype in cattle’, *Nat Genet*, 17(1), pp. 71–74.
- Groenen, M. A. M.Archibald, A. L.Uenishi, H.Tuggle, C. K.Takeuchi, Y.Rothschild, M. F.Rogel-Gaillard, C.Park, C.Milan, D.and Megens, H.-J. (2012) ‘Analyses of pig genomes provide insight into porcine demography and evolution’, *Nature*, 491(7424), pp. 393–398.
- Gu, S.Jin, L.Zhang, F.Huang, Y.Grimm, D.Rossi, J. J.and Kay, M. A. (2011) ‘Thermodynamic stability of small hairpin RNAs highly influences the loading process of different mammalian Argonautes’, *Proceedings of the National Academy of Sciences*, 108(22), pp. 9208–9213.
- Gu, S.Jin, L.Zhang, Y.Huang, Y.Zhang, F.Valdmanis, P. N.and Kay, M. A. (2012) ‘The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo’, *Cell*, 151(4), pp. 900–911.
- Gu, Y.Li, M.Wang, T.Liang, Y.Zhong, Z.Wang, X.Zhou, Q.Chen, L.Lang, Q.and He, Z. (2012) ‘Lactation-related microRNA expression profiles of porcine breast milk exosomes’, *PLoS One*, 7(8), p. e43691.
- Guerra-Assuncao, J. A. and Enright, A. J. (2010) ‘MapMi: automated mapping of microRNA loci’, *BMC Bioinformatics*. 2010/03/18, 11, p. 133. doi: 10.1186/1471-2105-11-133.
- Gutiérrez-Gil, B.Arranz, J. J.and Wiener, P. (2016) ‘An interpretive review of selective sweep studies in *Bos taurus* cattle populations: identification of unique and shared selection signals across breeds’, *Advances in Farm Animal Genomic Resources*. Frontiers Media SA, p. 231.
- Harris, B. and Johnson, D. (1998) ‘Approximate reliability of genetic evaluations under an animal model’, *Journal of dairy science*, 81(10), pp. 2723–2728.
- Hata, T.Murakami, K.Nakatani, H.Yamamoto, Y.Matsuda, T.and Aoki, N. (2010) ‘Isolation of bovine milk-derived microvesicles carrying mRNAs and microRNAs’, *Biochemical and biophysical research communications*, 396(2), pp. 528–533.
- Hayes, B. J.Bowman, P. J.Chamberlain, A. J.and Goddard, M. E. (2009) ‘Invited review: Genomic selection in dairy cattle: Progress and challenges’, *Journal of dairy science*. Elsevier, 92(2), pp. 433–443.
- Hayes, B. J.Lewin, H. A.and Goddard, M. E. (2013) ‘The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation’, *Trends in Genetics*, 29(4), pp. 206–214.
- Hill, W. G. and Kirkpatrick, M. (2010) ‘What animal breeding has taught us about evolution’,

- Annual Review of Ecology, Evolution, and Systematics*, 41, pp. 1–19.
- Hoch, R. V and Soriano, P. (2003) ‘Roles of PDGF in animal development’, *Development*, 130(20), pp. 4769–4784.
- Hofacker, I. L. (2003) ‘Vienna RNA secondary structure server’, *Nucleic acids research*, 31(13), pp. 3429–3431.
- Hofreiter, M.Jaenicke, V.Serre, D.von Haeseler, A.and Pääbo, S. (2001) ‘DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA’, *Nucleic acids research*. Oxford Univ Press, 29(23), pp. 4793–4799.
- Hossain, M. M.Ghanem, N.Hoelker, M.Rings, F.Phatsara, C.Tholen, E.Schellander, K.and Tesfaye, D. (2009) ‘Identification and characterization of miRNAs expressed in the bovine ovary’, *BMC Genomics*. 2009/09/22, 10, p. 443. doi: 10.1186/1471-2164-10-443.
- Hossain, M. M.Sohel, M. M. H.Schellander, K.and Tesfaye, D. (2012) ‘Characterization and importance of microRNAs in mammalian gonadal functions’, *Cell and tissue research*, 349(3), pp. 679–690.
- Hou, J.An, X.Song, Y.Gao, T.Lei, Y.and Cao, B. (2015) ‘Two Mutations in the Caprine MTHFR 3’UTR Regulated by MicroRNAs Are Associated with Milk Production Traits’, *PloS one*. Public Library of Science, 10(7), p. e0133015.
- Hou, Q.Huang, J.Ju, Z.Li, Q.Li, L.Wang, C.Sun, T.Wang, L.Hou, M.and Hang, S. (2012) ‘Identification of splice variants, targeted microRNAs and functional single nucleotide polymorphisms of the BOLA-DQA2 gene in dairy cattle’, *DNA and cell biology*, 31(5), pp. 739–744.
- Hou, Y.Bickhart, D. M.Hvinden, M. L.Li, C.Song, J.Boichard, D. A.Fritz, S.Eggen, A.DeNise, S.and Wiggans, G. R. (2012) ‘Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array’, *Bmc Genomics*, 13(1), p. 1.
- Hu, Z.-L.Fritz, E. R.and Reecy, J. M. (2007) ‘AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond’, *Nucleic acids research*, 35(suppl 1), pp. D604–D609.
- Hu, Z.-L.Park, C. A.and Reecy, J. M. (2016) ‘Developmental progress and current status of the Animal QTLdb’, *Nucleic acids research*, 44(D1), pp. D827–D833.
- Huang, T.Tucker, K.Lee, Y.Crott, J.Parnell, L.Shen, J.Smith, C.Ordo vas, J.Li, D.and Lai, C. (2012) ‘MAT1A variants modulate the effect of dietary fatty acids on plasma homocysteine concentrations’, *Nutrition, Metabolism and Cardiovascular Diseases*, 22(4), pp. 362–368.
- Hubbard, T.Barker, D.Birney, E.Cameron, G.Chen, Y.Clark, L.Cox, T.Cuff, J.Curwen,

- V.Down, T.Durbin, R.Eyras, E.Gilbert, J.Hammond, M.Huminiecki, L.Kasprzyk, A.Lehvaslaiho, H.Lijnzaad, P.Melsopp, C.Mongin, E.Pettett, R.Pocock, M.Potter, S.Rust, A.Schmidt, E.Searle, S.Slater, G.Smith, J.Spooner, W.Stabenau, A.Stalker, J.Stupka, E.Ureta-Vidal, A.Vastrik, I.and Clamp, M. (2002) ‘The Ensembl genome database project’, *Nucleic Acids Res.* 2001/12/26, 30(1), pp. 38–41.
- Hughes, A. E.Bradley, D. T.Campbell, M.Lechner, J.Dash, D. P.Simpson, D. A.and Willoughby, C. E. (2011) ‘Mutation altering the miR-184 seed region causes familial keratoconus with cataract’, *The American Journal of Human Genetics*, 89(5), pp. 628–633.
- Huntzinger, E. and Izaurralde, E. (2011) ‘Gene silencing by microRNAs: contributions of translational repression and mRNA decay’, *Nature Reviews Genetics*. Nature Publishing Group, 12(2), pp. 99–110.
- Hutvagner, G.McLachlan, J.Pasquinelli, A. E.Bálint, É.Tuschl, T.and Zamore, P. D. (2001) ‘A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA’, *Science*, 293(5531), pp. 834–838.
- Iorio, M. V and Croce, C. M. (2012) ‘MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review’, *EMBO molecular medicine*, 4(3), pp. 143–159.
- Irvin, M. R.Zhi, D.Aslibekyan, S.Claas, S. A.Absher, D. M.Ordoas, J. M.Tiwari, H. K.Watkins, S.and Arnett, D. K. (2014) ‘Genomics of Post-Prandial Lipidomic Phenotypes in the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) Study’.
- Iwai, N. and Naraba, H. (2005) ‘Polymorphisms in human pre-miRNAs’, *Biochemical and biophysical research communications*, 331(4), pp. 1439–1444.
- Izumi, H.Kosaka, N.Shimizu, T.Sekine, K.Ochiya, T.and Takase, M. (2012) ‘Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions’, *Journal of dairy science*, 95(9), pp. 4831–4841.
- Jennen, D. G. J.Vereijken, A. L. J.Bovenhuis, H.Crooijmans, R.Veenendaal, A.Van der Poel, J. J.and Groenen, M. A. M. (2004) ‘Detection and localization of quantitative trait loci affecting fatness in broilers’, *Poultry science*, 83(3), pp. 295–301.
- Jevsinek Skok, D.Godnic, I.Zorc, M.Horvat, S.Dovc, P.Kovac, M.and Kunej, T. (2013) ‘Genome-wide in silico screening for microRNA genetic variability in livestock species’, *Animal genetics*, 44(6), pp. 669–677.
- Jin, W.Dodson, M. V.Moore, S. S.and Basarab, J. A. (2010) ‘Characterization of microRNA expression in bovine adipose tissues: a potential regulatory mechanism of subcutaneous

- adipose tissue development', *BMC molecular biology*, 11(1), p. 29.
- Jorge, E. C.Melo, C. M. R.Rosário, M. F.Rossi, J. R. S.Ledur, M. C.Moura, A.and Coutinho, L. L. (2010) 'Chicken skeletal muscle-associated macroarray for gene discovery', *Genetics and Molecular Research*, 9(1), pp. 188–207.
- Kadzere, C. T.Murphy, M. R.Silanikove, N.and Maltz, E. (2002) 'Heat stress in lactating dairy cows: a review', *Livestock production science*. Elsevier, 77(1), pp. 59–91.
- Kemper, K. E. and Goddard, M. E. (2012) 'Understanding and predicting complex traits: knowledge from cattle', *Human molecular genetics*. Oxford Univ Press, p. dds332.
- Kerje, S.Carlborg, Ö.Jacobsson, L.Schütz, K.Hartmann, C.Jensen, P.and Andersson, L. (2003) 'The twofold difference in adult size between the red junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs', *Animal genetics*, 34(4), pp. 264–274.
- Khvorova, A.Reynolds, A.and Jayasena, S. D. (2003) 'Functional siRNAs and miRNAs exhibit strand bias', *Cell*, 115(2), pp. 209–216.
- Kim, V. N.Han, J.and Siomi, M. C. (2009) 'Biogenesis of small RNAs in animals', *Nature reviews Molecular cell biology*, 10(2), pp. 126–139.
- Kim, Y.Ryu, J.Woo, J.Kim, J. B.Kim, C. Y.and Lee, C. (2011) 'Genome-wide association study reveals five nucleotide sequence variants for carcass traits in beef cattle', *Animal genetics*, 42(4), pp. 361–365.
- Kinsella, R. J.Kahari, A.Haider, S.Zamora, J.Proctor, G.Spudich, G.Almeida-King, J.Staines, D.Derwent, P.Kerhornou, A.Kersey, P.and Flicek, P. (2011) 'Ensembl BioMart: a hub for data retrieval across taxonomic space', *Database (Oxford)*. 2011/07/26, 2011, p. bar030. doi: 10.1093.
- Kiriakidou, M.Nelson, P. T.Kouranov, A.Fitziev, P.Bouyioukos, C.Mourelatos, Z.and Hatzigeorgiou, A. (2004) 'A combined computational-experimental approach predicts human microRNA targets', *Genes & Development*, 18(10), pp. 1165–1178.
- Klos, K. S.Wyszomierski, S. L.Sun, M.Tan, M.Zhou, X.Li, P.Yang, W.Yin, G.Hittelman, W. N.and Yu, D. (2006) 'ErbB2 increases vascular endothelial growth factor protein synthesis via activation of mammalian target of rapamycin/p70S6K leading to increased angiogenesis and spontaneous metastasis of human breast cancer cells', *Cancer research*, 66(4), pp. 2028–2037.
- Knight, C. H.Calvert, D. T.and Flint, D. J. (1986) 'Inhibitory effects of bromocriptine on mammary development and function in lactating mice', *Journal of endocrinology*, 110(2), pp. 263–270.

- Kõks, S.Lilleoja, R.Reimann, E.Salumets, A.Reemann, P.and Jaakma, Ü. (2013) 'Sequencing and annotated analysis of the Holstein cow genome', *Mammalian genome*. Springer, 24(7–8), pp. 309–321.
- Kõks, S.Reimann, E.Lilleoja, R.Lättekivi, F.Salumets, A.Reemann, P.and Jaakma, Ü. (2014) 'Sequencing and annotated analysis of full genome of Holstein breed bull', *Mammalian genome*. Springer, 25(7–8), pp. 363–373.
- Koolhaas, J. M.Korte, S. M.De Boer, S. F.Van Der Vegt, B. J.Van Reenen, C. G.Hopster, H.De Jong, I. C.Ruis, M. A. W.and Blokhuis, H. J. (1999) 'Coping styles in animals: current status in behavior and stress-physiology', *Neuroscience & Biobehavioral Reviews*. Elsevier, 23(7), pp. 925–935.
- Kostiuk, M. A.Keller, B. O.and Berthiaume, L. G. (2010) 'Palmitoylation of ketogenic enzyme HMGCS2 enhances its interaction with PPAR α and transcription at the Hmgcs2 PPRE', *The FASEB Journal*, 24(6), pp. 1914–1924.
- Kozomara, A. and Griffiths-Jones, S. (2013) 'miRBase: annotating high confidence microRNAs using deep sequencing data', *Nucleic acids research*, 42(D1), pp. D68–D73.
- Krek, A.Gr^von, D.Poy, M. N.Wolf, R.Rosenberg, L.Epstein, E. J.MacMenamin, P.da Piedade, I.Gunsalus, K. C.and Stoffel, M. (2005) 'Combinatorial microRNA target predictions', *Nat Genet*, 37(5), pp. 495–500.
- Krol, J.Loedige, I.and Filipowicz, W. (2010) 'The widespread regulation of microRNA biogenesis, function and decay', *Nature Reviews Genetics*. Nature Publishing Group, 11(9), pp. 597–610.
- Krol, J.Sobczak, K.Wilczynska, U.Drath, M.Jasinska, A.Kaczynska, D.and Krzyzosiak, W. J. (2004) 'Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design', *Journal of Biological Chemistry*, 279(40), pp. 42230–42239.
- Krücken, J.Schroetel, R. M. U.Müller, I. U.Saïdani, N.Marinovski, P.Benten, W. P. M.Stamm, O.and Wunderlich, F. (2004) 'Comparative analysis of the human gimap gene cluster encoding a novel GTPase family', *Gene*, 341, pp. 291–304.
- Krüger, J. and Rehmsmeier, M. (2006) 'RNAhybrid: microRNA target prediction easy, fast and flexible', *Nucleic acids research*, 34(suppl 2), pp. W451–W454.
- Kuehl, P.Zhang, J.Lin, Y.Lamba, J.Assem, M.Schuetz, J.Watkins, P. B.Daly, A.Wrighton, S. A.and Hall, S. D. (2001) 'Sequence diversity in CYP3A promoters and characterization of the genetic basis of polymorphic CYP3A5 expression', *Nat Genet*, 27(4), pp. 383–391.
- Kuhn, D. E.Martin, M. M.Feldman, D. S.Terry, A. V.Nuovo, G. J.and Elton, T. S. (2008)

- ‘Experimental validation of miRNA targets’, *Methods*, 44(1), pp. 47–54.
- Lagos-Quintana, M.Rauhut, R.Lendeckel, W.and Tuschl, T. (2001) ‘Identification of novel genes coding for small expressed RNAs’, *Science*, 294(5543), pp. 853–858.
- Lagos-Quintana, M.Rauhut, R.Meyer, J.Borkhardt, A.and Tuschl, T. (2003) ‘New microRNAs from mouse and human’, *RNA*, 9(2), pp. 175–179.
- Lai, E. C.Tomancak, P.Williams, R. W.and Rubin, G. M. (2003) ‘Computational identification of Drosophila microRNA genes’, *Genome biology*. BioMed Central, 4(7), p. R42.
- Lai, S.-J.Liu, Y.-P.Liu, Y.-X.Li, X.-W.and Yao, Y.-G. (2006) ‘Genetic diversity and origin of Chinese cattle revealed by mtDNA D-loop sequence variation’, *Molecular phylogenetics and evolution*, 38(1), pp. 146–154.
- Lande, R. (1976) ‘Natural selection and random genetic drift in phenotypic evolution’, *Evolution*. JSTOR, pp. 314–334.
- Landi, D.Gemignani, F.Barale, R.and Landi, S. (2008) ‘A catalog of polymorphisms falling in microRNA-binding regions of cancer genes’, *DNA and cell biology*, 27(1), pp. 35–43.
- Larkin, D. M.Daetwyler, H. D.Hernandez, A. G.Wright, C. L.Hetrick, L. A.Boucek, L.Bachman, S. L.Band, M. R.Akraiko, T. Vand Cohen-Zinder, M. (2012) ‘Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle’, *Proceedings of the National Academy of Sciences*, 109(20), pp. 7693–7698.
- Larkin, M. A.Blackshields, G.Brown, N. P.Chenna, R.McGettigan, P. A.McWilliam, H.Valentin, F.Wallace, I. M.Wilm, A.and Lopez, R. (2007) ‘Clustal W and Clustal X version 2.0’, *Bioinformatics*, 23(21), pp. 2947–2948.
- Larson, G. and Fuller, D. Q. (2014) ‘The evolution of animal domestication’, *Annual Review of Ecology, Evolution, and Systematics*. Annual Reviews, 45, pp. 115–136.
- Larson, G.Piperno, D. R.Allaby, R. G.Purugganan, M. D.Andersson, L.Arroyo-Kalin, M.Barton, L.Vigueira, C. C.Denham, T.and Dobney, K. (2014) ‘Current perspectives and the future of domestication studies’, *Proceedings of the National Academy of Sciences*. National Acad Sciences, 111(17), pp. 6139–6146.
- Lee, E. S.Kim, M.Moon, S.Jekarl, D. W.Lee, S.Kim, Y.and Choi, G. Y. (2013) ‘A new compound heterozygous mutation in the CYP17A1 gene in a female with 17 α -hydroxylase/17, 20-lyase deficiency’, *Gynecological endocrinology*, 29(7), pp. 720–723.
- Lee, J.-S.Kim, J.-M.Lim, K.-S.Hong, J.-S.Hong, K.-C.and Lee, Y. S. (2013) ‘Effects of polymorphisms in the porcine microRNA MIR206/MIR133B cluster on muscle fiber and meat quality traits’, *Animal genetics*, 44(1), pp. 101–106.
- Lee, R. C. and Ambros, V. (2001) ‘An extensive class of small RNAs in Caenorhabditis

- elegans*', *Science*, 294(5543), pp. 862–864.
- Lee, R. C. Feinbaum, R. L. and Ambros, V. (1993) 'The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*', *Cell*, 75(5), pp. 843–854.
- Lewandowska-Sabat, A. M. Günther, J. Seyfert, H. M. and Olsaker, I. (2012) 'Combining quantitative trait loci and heterogeneous microarray data analyses reveals putative candidate pathways affecting mastitis in cattle', *Animal genetics*, 43(6), pp. 793–799.
- Lewis, B. P. Burge, C. B. and Bartel, D. P. (2005) 'Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets', *Cell*, 2005/01/18, 120(1), pp. 15–20. doi: 10.1016/j.cell.2004.12.035.
- Lewis, B. P. Shih, I. Jones-Rhoades, M. W. Bartel, D. P. and Burge, C. B. (2003) 'Prediction of mammalian microRNA targets', *Cell*, 115(7), pp. 787–798.
- Lewontin, R. C. and Kojima, K. (1960) 'The evolutionary dynamics of complex polymorphisms', *Evolution*. JSTOR, pp. 458–472.
- Li, H. Zhang, Z. Zhou, X. Wang, Z. Wang, G. and Han, Z. (2011) 'Effects of microRNA-143 in the differentiation and proliferation of bovine intramuscular preadipocytes', *Molecular biology reports*, 38(7), pp. 4273–4280.
- Li, L. Huang, J. Zhang, X. Ju, Z. Qi, C. Zhang, Y. Li, Q. Wang, C. Miao, W. and Zhong, J. (2012) 'One SNP in the 3'UTR of HMGB1 gene affects the binding of target bta-miR-223 and is involved in mastitis in dairy cattle', *Immunogenetics*, 64(11), pp. 817–824.
- Li, Q. Wang, N. Du, Z. Hu, X. Chen, L. Fei, J. Wang, Y. and Li, N. (2012) 'Gastrocnemius transcriptome analysis reveals domestication induced gene expression changes between wild and domestic chickens', *Genomics*, 100(5), pp. 314–319.
- Li, T. Wu, R. Zhang, Y. and Zhu, D. (2011) 'A systematic analysis of the skeletal muscle miRNA transcriptome of chicken varieties with divergent skeletal muscle growth identifies novel miRNAs and differentially expressed miRNAs', *Bmc Genomics*, 12(1), p. 186.
- Li, Z. Wang, H. Chen, L. Wang, L. Liu, X. Ru, C. and Song, A. (2014) 'Identification and characterization of novel and differentially expressed microRNAs in peripheral blood from healthy and mastitis Holstein cattle by deep sequencing', *Animal genetics*, 45(1), pp. 20–27.
- Liang, Y. Buckley, T. R. Tu, L. Langdon, S. D. and Tedder, T. F. (2001) 'Structural organization of the human MS4A gene cluster on Chromosome 11q12', *Immunogenetics*, 53(5), pp. 357–368.
- Lindahl, G. Enfält, A.-C. von Seth, G. Josell, Å. Hedebrö-Velander, I. Andersen, H. J. Braunschweig, M. Andersson, L. and Lundström, K. (2004) 'A second mutant allele

- (V199I) at the PRKAG3 (RN) locus— I. Effect on technological meat quality of pork loin', *Meat Science*, 66(3), pp. 609–619.
- Linderholm, A. and Larson, G. (2013) 'The role of humans in facilitating and sustaining coat colour variation in domestic animals', in *Seminars in cell & developmental biology*. Elsevier.
- Litt, M. and Luty, J. A. (1989) 'A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene.', *American journal of human genetics*. Elsevier, 44(3), p. 397.
- Liu, F.-J.Jin, L.-J.Ma, X.-G.Zhang, Y.-L.Zhai, X.-W.Chen, J.-J.and Yang, X.-Y. (2014) 'Differentially expressed microRNAs and affected signaling pathways in placentae of transgenic cloned cattle', *Theriogenology*, 82(2), p. 338–346. e3.
- Liu, H.Hicks, J. A.Trakooljul, N.and Zhao, S. (2010) 'Current knowledge of microRNA characterization in agricultural animals', *Animal genetics*. Wiley Online Library, 41(3), pp. 225–231.
- Liu, Y.Qin, X.Song, X.-Z. H.Jiang, H.Shen, Y.Durbin, K. J.Lien, S.Kent, M. P.Sodeland, M.and Ren, Y. (2009) 'Bos taurus genome assembly', *Bmc Genomics*, 10(1), p. 180.
- Llave, C.Xie, Z.Kasschau, K. D.and Carrington, J. C. (2002) 'Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA', *Science*, 297(5589), pp. 2053–2056.
- Loftus, R. T.Ertugrul, O.Harba, A. H.El-Barody, M. A. A.MacHugh, D. E.Park, S. D. E.and Bradley, D. G. (1999) 'A microsatellite survey of cattle from a centre of origin: the Near East', *Molecular Ecology*. Wiley Online Library, 8(12), pp. 2015–2022.
- Loftus, R. T.MacHugh, D. E.Bradley, D. G.Sharp, P. M.and Cunningham, P. (1994) 'Evidence for two independent domestications of cattle', *Proceedings of the National Academy of Sciences*, 91(7), pp. 2757–2761.
- Lopez-Oceja, A.Muro-Verde, A.Gamarra, D.Cardoso, S.and de Pancorbo, M. M. (2015) 'New Q lineage found in bovine (*Bos taurus*) of Iberian Peninsula', *Mitochondrial DNA*, pp. 1–5.
- Ludwig, A.Pruvost, M.Reissmann, M.Benecke, N.Brockmann, G. A.Castaños, P.Cieslak, M.Lippold, S.Llorente, L.and Malaspinas, A.-S. (2009) 'Coat color variation at the beginning of horse domestication', *Science*. American Association for the Advancement of Science, 324(5926), p. 485.
- MacHugh, D. E.Larson, G.and Orlando, L. (2016) 'Taming the Past: Ancient DNA and the

- Study of Animal Domestication', *Annual Review of Animal Biosciences*. Annual Reviews
4139 El Camino Way, PO Box 10139, Palo Alto, California 94303-0139, USA, (0).
- MacHugh, D. E., Loftus, R. T., Bradley, D. G., Sharp, P. M. and Cunningham, P. (1994)
'Microsatellite DNA variation within and among European cattle breeds', *Proceedings of
the Royal Society of London B: Biological Sciences*. The Royal Society, 256(1345), pp.
25–31.
- Mackay, T. F. C. (2001) 'Quantitative trait loci in *Drosophila*', *Nature Reviews Genetics*.
Nature Publishing Group, 2(1), pp. 11–20.
- Madon, R. J., Ensor, D. M., Knight, C. H. and Flint, D. J. (1986) 'Effects of an antiserum to rat
growth hormone on lactation in the rat', *Journal of endocrinology*, 111(1), pp. 117–123.
- Magee, D. A., Berkowicz, E. W., Sikora, K. M., Berry, D. P., Park, S. D., E. Kelly, A. K., Sweeney,
T. Kenny, D. A., Evans, R. D. and Wickham, B. W. (2010) 'A catalogue of validated single
nucleotide polymorphisms in bovine orthologs of mammalian imprinted genes and
associations with beef production traits', *animal*, 4(12), pp. 1958–1970.
- Magee, D. A., MacHugh, D. E. and Edwards, C. J. (2014) 'Interrogation of modern and ancient
genomes reveals the complex domestic history of cattle', *Animal Frontiers*, 4(3), pp. 7–
22.
- Magee, D. A., Sikora, K. M., Berkowicz, E. W., Berry, D. P., Howard, D. J., Mullen, M. P., Evans, R.
D., Spillane, C. and MacHugh, D. E. (2010) 'DNA sequence polymorphisms in a panel of
eight candidate bovine imprinted genes and their association with performance traits in
Irish Holstein-Friesian cattle', *BMC genetics*, 11(1), p. 1.
- Mahmoudi, A., Zargaran, A., Amini, H.-R., Assadi, A., Hokmabad, R. V. and Eghbalsaied, S.
(2015) 'A SNP in the 3'-untranslated region of AMPK γ 1 may associate with serum ketone
body and milk production of Holstein dairy cows', *Gene*, 574(1), pp. 48–52.
- Maranville, J. C., Luca, F., Richards, A. L., Wen, X., Witonsky, D. B., Baxter, S., Stephens, M., Di
Rienzo, A. and Gibson, G. (2011) 'Interactions between glucocorticoid treatment and cis-
regulatory polymorphisms contribute to cellular response phenotypes', *PLoS Genet*, 7(7),
p. e1002162.
- Mato, J. M., Corrales, F. J., Lu, S. C. and Avila, M. A. (2002) 'S-Adenosylmethionine: a control
switch that regulates liver function', *FASEB J*. 2002/01/05, 16(1), pp. 15–26. doi:
10.1096/fj.01-0401rev.
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M.
P., O'Connell, J., Moore, S. S., Smith, T. P. L. and Sonstegard, T. S. (2009) 'Development
and characterization of a high density SNP genotyping assay for cattle', *PLoS One*, 4(4),

p. e5350.

- Mc Parland, S.Kearney, F.and Berry, D. P. (2009) ‘Purging of inbreeding depression within the Irish Holstein-Friesian population’, *Genetics Selection Evolution*. BioMed Central, 41(1), p. 16.
- Mc Parland, S.Kearney, J. F.Rath, M.and Berry, D. P. (2007) ‘Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians’, *Journal of Dairy Science*. Elsevier, 90(9), pp. 4411–4419.
- McCabe, M.Waters, S.Morris, D.Kenny, D.Lynn, D.and Creevey, C. (2012) ‘RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance’, *Bmc Genomics*, 13(1), p. 193.
- McClure, M.Sonstegard, T.Wiggans, G.and Van Tassell, C. P. (2012) ‘Imputation of microsatellite alleles from dense SNP genotypes for parental verification’, *Front. Genet*, 3(140), p. 10.3389.
- McConaha, M. E.Eckstrum, K.An, J.Steinle, J. J.and Bany, B. M. (2011) ‘Microarray assessment of the influence of the conceptus on gene expression in the mouse uterus during decidualization’, *Reproduction*, 141(4), pp. 511–527.
- McDaneld, T. G. (2009) ‘MicroRNA: mechanism of gene regulation and application to livestock’, *Journal of animal science*. American Society of Animal Science, 87(14_suppl), pp. E21–E28.
- Meade, K. G.Gormley, E.O’Farrelly, C.Park, S. D.Costello, E.Keane, J.Zhao, Y.and MacHugh, D. E. (2008) ‘Antigen stimulation of peripheral blood mononuclear cells from Mycobacterium bovis infected cattle yields evidence for a novel gene expression program’, *Bmc Genomics*, 9(1), p. 447.
- Mein, J. R.James, D. R.and Lakkanna, S. (2012) ‘Induction of phase 2 antioxidant enzymes by broccoli sulforaphane: perspectives in maintaining the antioxidant activity of vitamins A, C, and E’, *Frontiers in genetics*, 3, p. 7.
- Mencia, Á.Modamio-Høybjør, S.Redshaw, N.Morín, M.Mayo-Merino, F.Olavarrieta, L.Aguirre, L. A.del Castillo, I.Steel, K. P.Dalmay, T.Moreno, F.and Moreno-Pelayo, M. Á. (2009) ‘Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss’, *Nat Genet*, 41(5), pp. 609–613.
- Meuwissen, T. and Goddard, M. (2010) ‘Accurate prediction of genetic values for complex traits by whole-genome resequencing’, *Genetics*, 185(2), pp. 623–631.
- Meuwissen, T. H.Hayes, B. J.and Goddard, M. E. (2001) ‘Prediction of total genetic value using genome-wide dense marker maps.’, *Genetics*, 157(4), pp. 1819–1829.

- Milan, D.Jeon, J.-T.Looft, C.Amarger, V.Robic, A.Thelander, M.Rogel-Gaillard, C.Paul, S.Iannuccelli, N.and Rask, L. (2000) 'A mutation in PRKAG3 associated with excess glycogen content in pig skeletal muscle', *Science*, 288(5469), pp. 1248–1251.
- Miles, J. R.McDanel, T. G.Wiedmann, R. T.Cushman, R. A.Echternkamp, S. E.Vallet, J. L.and Smith, T. P. L. (2012) 'MicroRNA expression profile in bovine cumulus–oocyte complexes: Possible role of let-7 and miR-106a in the development of bovine oocytes', *Animal reproduction science*, 130(1), pp. 16–26.
- Miretti, S.Martignani, E.Taulli, R.Bersani, F.Accornero, P.and Baratta, M. (2011) 'Differential expression of microRNA-206 in skeletal muscle of female Piedmontese and Friesian cattle', *The Veterinary Journal*, 190(3), pp. 412–413.
- Mondou, E.Dufort, I.Gohin, M.Fournier, E.and Sirard, M. A. (2012) 'Analysis of microRNAs and their precursors in bovine early embryonic development', *Molecular human reproduction*, p. gas015.
- Morris, C. A.Bottema, C. D. K.Cullen, N. G.Hickey, S. M.Esmailzadeh, A. K.Siebert, B. D.and Pitchford, W. S. (2010) 'Quantitative trait loci for organ weights and adipose fat composition in Jersey and Limousin back–cross cattle finished on pasture or feedlot', *Animal genetics*, 41(6), pp. 589–596.
- Moseley, T. A.Haudenschield, D. R.Rose, L.and Reddi, A. H. (2003) 'Interleukin-17 family and IL-17 receptors', *Cytokine & growth factor reviews*. Elsevier, 14(2), pp. 155–174.
- Mullen, M. P.Creevey, C. J.Berry, D. P.McCabe, M. S.Magee, D. A.Howard, D. J.Killeen, A. P.Park, S. D.McGettigan, P. A.Lucy, M. C.Machugh, D. E.and Waters, S. M. (2012) 'Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples.', *BMC genomics*. BioMed Central, 13(1), p. 16. doi: 10.1186/1471-2164-13-16.
- Mullen, M. P.McClure, M. C.Kearney, J. F.Waters, S. M.Weld, R.Flynn, P.Creevey, C. J.Cromie, A. R.and Berry, D. P. (2013) 'Development of a custom SNP chip for dairy and beef cattle breeding, parentage and research', *Interbull Bulletin*, (47).
- Murakami, N.Bolton, D.and Hwang, Y.-W. (2009) 'Dyrk1A binds to multiple endocytic proteins required for formation of clathrin-coated vesicles', *Biochemistry*, 48(39), pp. 9297–9305.
- Murgiano, L.Jagannathan, V.Calderoni, V.Joechler, M.Gentile, A.and Drögemüller, C. (2014) 'Looking the cow in the eye: deletion in the NID1 gene is associated with recessive inherited cataract in Romagnola cattle', *PloS one*. Public Library of Science, 9(10), p. e110628.

- Murgiano, L.Sacchetto, R.Testoni, S.Dorotea, T.Mascarello, F.Liguori, R.Gentile, A.and Drögemüller, C. (2012) 'Pseudomyotonia in Romagnola cattle caused by novel ATP2A1 mutations', *BMC veterinary research*. BioMed Central, 8(1), p. 186.
- Muroya, S.Taniguchi, M.Shibata, M.Oe, M.Ojima, K.Nakajima, I.and Chikuni, K. (2013) 'Profiling of differentially expressed microRNA and the bioinformatic target gene analyses in bovine fast-and slow-type muscles by massively parallel sequencing', *Journal of animal science*, 91(1), pp. 90–103.
- Naeem, A.Zhong, K.Moisá, S. J.Drackley, J. K.Moyes, K. M.and Loor, J. J. (2012) 'Bioinformatics analysis of microRNA and putative target genes in bovine mammary tissue infected with *Streptococcus uberis*', *Journal of dairy science*, 95(11), pp. 6397–6408.
- Najafi-Shoushtari, S. H.Kristo, F.Li, Y.Shioda, T.Cohen, D. E.Gerszten, R. E.and Näär, A. M. (2010) 'MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis', *Science*, 328(5985), pp. 1566–1569.
- Nebert, D. W. and Russell, D. W. (2002) 'Clinical importance of the cytochromes P450', *The Lancet*, 360(9340), pp. 1155–1162.
- Neimann-Sorensen, A. and Robertson, A. (1961) 'The association between blood groups and several production characteristics in three Danish cattle breeds', *Acta Agriculturae Scandinavica*. Taylor & Francis, 11(2), pp. 163–196.
- Nimz, M.Spitschak, M.Schneider, F.Fürbass, R.and Vanselow, J. (2009) 'Down-regulation of genes encoding steroidogenic enzymes and hormone receptors in late preovulatory follicles of the cow coincides with an accumulation of intrafollicular steroids', *Domestic animal endocrinology*, 37(1), pp. 45–54.
- Nishimura, S.Watanabe, T.Mizoshita, K.Tatsuda, K.Fujita, T.Watanabe, N.Sugimoto, Y.and Takasuga, A. (2012) 'Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle', *BMC genetics*. BioMed Central, 13(1), p. 40.
- Noe-Nygaard, N.Price, T. D.and Hede, S. U. (2005) 'Diet of aurochs and early cattle in southern Scandinavia: evidence from 15N and 13C stable isotopes', *Journal of Archaeological Science*, 32(6), pp. 855–871.
- O'Connell, R. M.Rao, D. S.and Baltimore, D. (2012) 'microRNA regulation of inflammatory responses', *Annual review of immunology*, 30, pp. 295–312.
- Ogorevc, J.Kunej, T.Razpet, A.and Dovc, P. (2009) 'Database of cattle candidate genes and genetic markers for milk production and mastitis', *Animal genetics*, 40(6), pp. 832–851.

- Oklahoma State University (1997a) *Breed of Livestock: Angus Cattle*, www.ansi.okstate.edu/breeds/cattle/angus.
- Oklahoma State University (1997b) *Breed of Livestock: Jersey Cattle*, www.ansi.okstate.edu/breeds/cattle/jersey.
- Onizuka, M.Kunii, N.Toyosaki, M.Machida, S.Ohgiya, D.Ogawa, Y.Kawada, H.Inoko, H.and Ando, K. (2011) 'Cytochrome P450 genetic polymorphisms influence the serum concentration of calcineurin inhibitors in allogeneic hematopoietic SCT recipients', *Bone marrow transplantation*, 46(8), pp. 1113–1117.
- Otero, J. A.Miguel, V.González-Lobato, L.García-Villalba, R.Espín, J. C.Prieto, J. G.Merino, G.and Álvarez, A. I. (2015) 'Effect of bovine ABCG2 polymorphism Y581S SNP on secretion into milk of enterolactone, riboflavin and uric acid', *animal*, pp. 1–10.
- Panaro, M. A. and Mitolo, V. (1999) 'Cellular responses to FMLP challenging: a mini-review', *Immunopharmacology and immunotoxicology*, 21(3), pp. 397–419.
- Pant, P. V. K.Tao, H.Beilharz, E. J.Ballinger, D. G.Cox, D. R.and Frazer, K. A. (2006) 'Analysis of allelic differential expression in human white blood cells', *Genome Res*, 16(3), pp. 331–339.
- Park, S. D. E.Magee, D. A.McGettigan, P. A.Teasdale, M. D.Edwards, C. J.Lohan, A. J.Murphy, A.Braud, M.Donoghue, M. T.and Liu, Y. (2015) 'Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle', *Genome Biol*, 16(1), pp. 1–15.
- Patel, A. K.Bhatt, V. D.Tripathi, A. K.Sajnani, M. R.Jakhesara, S. J.Koringa, P. G.and Joshi, C. G. (2013) 'Identification of novel splice variants in horn cancer by RNA-Seq analysis in Zebu cattle', *Genomics*, 101(1), pp. 57–63.
- Pelosi, M.De Rossi, M.Barberi, L.and Musarò, A. (2014) 'IL-6 Impairs Myogenic Differentiation by Downmodulation of p90RSK/eEF2 and mTOR/p70S6K Axes, without Affecting AKT Activity', *BioMed Research International*, 2014.
- Pesole, G.Mignone, F.Gissi, C.Grillo, G.Licciulli, F.and Liuni, S. (2001) 'Structural and functional features of eukaryotic mRNA untranslated regions', *Gene*, 276(1), pp. 73–81.
- Polvi, A.Linnankivi, T.Kivelä, T.Herva, R.Keating, J. P.Mäkitie, O.Pareyson, D.Vainionpää, L.Lahtinen, J.and Hovatta, I. (2012) 'Mutations in CTC1, encoding the CTS telomere maintenance complex component 1, cause cerebroretinal microangiopathy with calcifications and cysts', *The American Journal of Human Genetics*, 90(3), pp. 540–549.
- Ponsuksili, S.Zebunke, M.Murani, E.Trakooljul, N.Krieter, J.Puppe, B.Schwerin, M.and Wimmers, K. (2015) 'Integrated Genome-wide association and hypothalamus eQTL

- studies indicate a link between the circadian rhythm-related gene PER1 and coping behavior', *Scientific reports*, 5, p. 16264.
- Prendiville, R.Lewis, E.Pierce, K. M.and Buckley, F. (2010) 'Comparative grazing behavior of lactating Holstein-Friesian, Jersey, and Jersey× Holstein-Friesian dairy cows and its association with intake capacity and production efficiency', *Journal of Dairy Science*. Elsevier, 93(2), pp. 764–774.
- Pryce, J. E.Bolormaa, S.Chamberlain, A. J.Bowman, P. J.Savin, K.Goddard, M. E.and Hayes, B. J. (2010) 'A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes', *Journal of dairy science*, 93(7), pp. 3331–3345.
- Pryce, J. E. and Veerkamp, R. F. (2001) 'The incorporation of fertility indices in genetic improvement programmes', *BSAS occasional publication*, pp. 237–250.
- Ramamoorthy, A.Liu, Y.Philips, S.Desta, Z.Lin, H.Goswami, C.Gaedigk, A.Li, L.Flockhart, D. A.and Skaar, T. C. (2013) 'Regulation of microRNA expression by rifampin in human hepatocytes', *Drug Metabolism and Disposition*, 41(10), pp. 1763–1768.
- Rehmsmeier, M.Steffen, P.Höchsmann, M.and Giegerich, R. (2004) 'Fast and effective prediction of microRNA/target duplexes', *RNA*, 10(10), pp. 1507–1517.
- Reinhart, B. J.Slack, F. J.Basson, M.Pasquinelli, A. E.Bettinger, J. C.Rougvie, A. E.Horvitz, H. R.and Ruvkun, G. (2000) 'The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*', *Nature*, 403(6772), pp. 901–906.
- Reinhart, B. J.Weinstein, E. G.Rhoades, M. W.Bartel, B.and Bartel, D. P. (2002) 'MicroRNAs in plants', *Genes & Development*, 16(13), pp. 1616–1626.
- Renwick, A. (2013) 'The Importance of the Cattle and Sheep Sectors to the Irish Economy', *The Irish Farmers' Association*.
- Riaz, T.Sollid, L. M.Olsen, I.and de Souza, G. A. (2015) 'Quantitative proteomics of gut derived Th1 and Th1/Th17 clones reveal the presence of CD28+ NKG2D-Th1 cytotoxic CD4+ T cells', *Molecular & Cellular Proteomics*, p. mcp. M115. 050138.
- Rincon, G.Weber, K. L.Van Eenennaam, A. L.Golden, B. L.and Medrano, J. F. (2011) 'Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys', *Journal of Dairy Science*. Elsevier, 94(12), pp. 6116–6121.
- Rischkowsky, B. and Pilling, D. (2007) *The state of the world's animal genetic resources for food and agriculture*. Food & Agriculture Org.
- Rokosz, M. (1995) 'History of the aurochs (*Bos taurus primigenius*) in Poland', *Animal Genetic Resources Information*, 16, pp. 5–12.

- Ron, M. and Weller, J. I. (2007) 'From QTL to QTN identification in livestock—winning by points rather than knock-out: a review', *Animal genetics*, 38(5), pp. 429–439.
- Rong, E.Zhang, Z.Qiao, S.Yang, H.Yan, X.Li, H.and Wang, N. (2015) 'Functional Characterization of a Single Nucleotide Polymorphism in the 3'Untranslated Region of Sheep DLX3 Gene', *PloS one*. Public Library of Science, 10(9), p. e0137135.
- Roughgarden, J. (1972) 'Evolution of niche width', *American Naturalist*, pp. 683–718.
- Roussel, P.Cunha, P.Porcherie, A.Petzl, W.Gilbert, F. B.Riollet, C.Zerbe, H.Rainard, P.and Germon, P. (2015) 'Investigating the contribution of IL-17A and IL-17F to the host response during Escherichia coli mastitis', *Veterinary research*. BioMed Central, 46(1), p. 56.
- Roux, M.Nizou, A.Forestier, L.Ouali, A.Levéziel, H.and Amarger, V. (2006) 'Characterization of the bovine PRKAG3 gene: structure, polymorphism, and alternative transcripts', *Mammalian genome*, 17(1), pp. 83–92.
- Royal, M. D.Darwash, A. O.Flint, A. P. F.Webb, R.Woolliams, J. A.and Lamming, G. E. (2000) 'Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility.', *Animal science*, 70(3), pp. 487–501.
- Royal, M. D.Flint, A. P. F.and Woolliams, Ja. (2002) 'Genetic and phenotypic relationships among endocrine and traditional fertility traits and production traits in Holstein-Friesian dairy cows', *Journal of dairy science*, 85(4), pp. 958–967.
- Ruvkun, G. and Giusto, J. (1989) 'The Caenorhabditis elegans heterochronic gene lin-14 encodes a nuclear protein that forms a temporal developmental switch', *Nature*. Nature Publishing Group, 338(6213), pp. 313–319.
- Ryan, B. M.Robles, A. I.and Harris, C. C. (2010) 'Genetic variation in microRNA networks: the implications for cancer research', *Nature Reviews Cancer*, 10(6), pp. 389–402.
- Sasaki, S.Ibi, T.Watanabe, T.Matsushashi, T.Ikeda, S.and Sugimoto, Y. (2013) 'Variants in the 3'UTR of General Transcription Factor IIF, polypeptide 2 affect female calving efficiency in Japanese Black cattle', *BMC genetics*, 14(1), p. 41.
- Saunders, A.Webb, L. M. C.Janas, M. L.Hutchings, A.Pascall, J.Carter, C.Pugh, N.Morgan, G.Turner, M.and Butcher, G. W. (2010) 'Putative GTPase GIMAP1 is critical for the development of mature B and T lymphocytes', *Blood*, 115(16), pp. 3249–3257.
- Saunders, M. A.Liang, H.and Li, W.-H. (2007) 'Human polymorphism at microRNAs and microRNA target sites', *Proceedings of the National Academy of Sciences*, 104(9), pp. 3300–3305.
- Schaeffer, L.Gohlke, H.Müller, M.Heid, I. M.Palmer, L. J.Kompauer, I.Demmelmair, H.Illig,

- T.Koletzko, B. and Heinrich, J. (2006) 'Common genetic variants of the FADS1 FADS2 gene cluster and their reconstructed haplotypes are associated with the fatty acid composition in phospholipids', *Human molecular genetics*, 15(11), pp. 1745–1756.
- Schenk, H.Klein, M.Erdbrügger, W.Dröge, W. and Schulze-Osthoff, K. (1994) 'Distinct effects of thioredoxin and antioxidants on the activation of transcription factors NF-kappa B and AP-1', *Proceedings of the National Academy of Sciences*, 91(5), pp. 1672–1676.
- Schmidt, K. N.Amstad, P.Cerutti, P. and Baeuerle, P. A. (1995) 'The roles of hydrogen peroxide and superoxide as messengers in the activation of transcription factor NF-κB', *Chemistry & biology*, 2(1), pp. 13–22.
- Schroeder, J. A. and Lee, D. C. (1998) 'Dynamic expression and activation of ERBB receptors in the developing mouse mammary gland', *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research*, 9(6), pp. 451–464.
- Schubert, M.Jónsson, H.Chang, D.Der Sarkissian, C.Ermini, L.Ginolhac, A.Albrechtsen, A.Dupanloup, I.Foucal, A. and Petersen, B. (2014) 'Prehistoric genomes reveal the genetic foundation and cost of horse domestication', *Proceedings of the National Academy of Sciences*, 111(52), pp. E5661–E5669.
- Schwarz, D. S.Hutvágner, G.Du, T.Xu, Z.Aronin, N. and Zamore, P. D. (2003) 'Asymmetry in the assembly of the RNAi enzyme complex', *Cell*, 115(2), pp. 199–208.
- Sethupathy, P. and Collins, F. S. (2008) 'MicroRNA target site polymorphisms and human disease', *Trends in Genetics*, 24(10), pp. 489–497.
- Sharma, I. and Singh, D. (2012) 'Conjugated linoleic acids attenuate FSH-and IGF1-stimulated cell proliferation; IGF1, GATA4, and aromatase expression; and estradiol-17β production in buffalo granulosa cells involving PPARγ, PTEN, and PI3K/Akt', *Reproduction*, 144(3), pp. 373–383.
- Sharma, R.Coats, A. J. S. and Anker, S. D. (2000) 'The role of inflammatory mediators in chronic heart failure: cytokines, nitric oxide, and endothelin-1', *International journal of cardiology*, 72(2), pp. 175–186.
- Sheffield, L. G. (1998) 'Hormonal Regulation of Epidermal Growth Factor Receptor Content and Signaling in Bovine Mammary Tissue 1', *Endocrinology*, 139(11), pp. 4568–4575.
- Shelat, P. B.Chalimoniuk, M.Wang, J.Strosznajder, J. B.Lee, J. C.Sun, A. Y.Simonyi, A. and Sun, G. Y. (2008) 'Amyloid beta peptide and NMDA induce ROS from NADPH oxidase and AA release from cytosolic phospholipase A2 in cortical neurons', *Journal of neurochemistry*, 106(1), pp. 45–55.
- Shi, S.Leites, C.He, D.Schwartz, D.Moy, W.Shi, J. and Duan, J. (2014) 'MicroRNA-9 and

- microRNA-326 regulate human dopamine D2 receptor expression, and the microRNA-mediated expression regulation is altered by a genetic variant', *Journal of Biological Chemistry*, 289(19), pp. 13434–13444.
- Shin, S.Wolgammott, L.Yu, Y.Blenis, J.and Yoon, S.-O. (2011) 'Glycogen synthase kinase (GSK)-3 promotes p70 ribosomal protein S6 kinase (p70S6K) activity and cell proliferation', *Proceedings of the National Academy of Sciences*, 108(47), pp. E1204–E1213.
- Sindić, A. and Schlatter, E. (2006) 'Cellular effects of guanylin and uroguanylin', *Journal of the American Society of Nephrology*, 17(3), pp. 607–616.
- Slatkin, M. (1970) 'Selection and polygenic characters', *Proceedings of the National Academy of Sciences*, 66(1), pp. 87–93.
- Slatkin, M. (2008) 'Linkage disequilibrium—understanding the evolutionary past and mapping the medical future', *Nature Reviews Genetics*. Nature Publishing Group, 9(6), pp. 477–485.
- Smith, L. A.Cassell, B. G.and Pearson, R. E. (1998) 'The effects of inbreeding on the lifetime performance of dairy cattle', *Journal of dairy science*. Elsevier, 81(10), pp. 2729–2737.
- Snelling, W. M.Cushman, R. A.Keele, J. W.Maltecca, C.Thomas, M. G.Fortes, M. R. S.and Reverter, A. (2013) 'Breeding and Genetics Symposium: networks and pathways to guide genomic selection', *Journal of Animal Science*. American Society of Animal Science, 91(2), pp. 537–552.
- Sorensen, D. and Gianola, D. (2002) *Likelihood, Bayesian and MCMC methods in quantitative genetics*. Springer.
- Soskis, M. J.Ho, H.-Y. H.Bloodgood, B. L.Robichaux, M. A.Malik, A. N.Ataman, B.Rubin, A. A.Zieg, J.Zhang, C.and Shokat, K. M. (2012) 'A chemical genetic approach reveals distinct EphB signaling mechanisms during brain development', *Nature neuroscience*.
- Spelman, R. J.Hayes, B. J.and Berry, D. P. (2013) 'Use of molecular technologies for the advancement of animal breeding: genomic selection in dairy cattle populations in Australia, Ireland and New Zealand', *Animal Production Science*, 53(9), pp. 869–875.
- Spitzer, E. and Grosse, R. (1987) 'EGF receptors on plasma membranes purified from bovine mammary gland of lactating and pregnant animals', *Biochemistry international*, 14(4), pp. 581–588.
- Srivastava, K. and Srivastava, A. (2012) 'Comprehensive review of genetic association studies and meta-analyses on miRNA polymorphisms and cancer risk', *PLoS One*, 7(11), p. e50966.

- Stöckler, S.Hanefeld, F.and Frahm, J. (1996) 'Creatine replacement therapy in guanidinoacetate methyltransferase deficiency, a novel inborn error of metabolism', *The Lancet*, 348(9030), pp. 789–790.
- Stormont, C. (1967) 'Contribution of blood typing to dairy science progress', *Journal of dairy science*. Elsevier, 50(2), pp. 253–260.
- Sun, G.Yan, J.Noltner, K.Feng, J.Li, H.Sarkis, D. A.Sommer, S. S.and Rossi, J. J. (2009) 'SNPs in human miRNA genes affect biogenesis and function', *RNA*, 15(9), pp. 1640–1651.
- Sun, J.Zhang, B.Lan, X.Zhang, C.Lei, C.and Chen, H. (2014) 'Comparative Transcriptome Analysis Reveals Significant Differences in MicroRNA Expression and Their Target Genes between Adipose and Muscular Tissues in Cattle'.
- Takeda, E.Nakagawa, S.Nakaya, Y.Tanaka, A.Miyazawa, T.and Yasuda, J. (2012) 'Identification and functional analysis of three isoforms of bovine BST-2'.
- Tao, W. and Mallard, B. (2007) 'Differentially expressed genes associated with *Staphylococcus aureus* mastitis of Canadian Holstein cows', *Veterinary immunology and immunopathology*, 120(3), pp. 201–211.
- Taub, N.Nairz, M.Hilber, D.Hess, M. W.Weiss, G.and Huber, L. A. (2012) 'The late endosomal adaptor p14 is a macrophage host-defense factor against Salmonella infection', *Journal of cell science*, 125(11), pp. 2698–2708.
- Tesfaye, D.Worku, D.Rings, F.Phatsara, C.Tholen, E.Schellander, K.and Hoelker, M. (2009) 'Identification and expression profiling of microRNAs during bovine oocyte maturation using heterologous approach', *Molecular reproduction and development*, 76(7), pp. 665–677.
- Testoni, S.Bartolone, E.Rossi, M.Patrignani, A.Bruggmann, R.Lichtner, P.Tetens, J.Gentile, A.and Drögemüller, C. (2012) 'KDM2B is implicated in bovine lethal multi-organic developmental dysplasia', *PloS one*. Public Library of Science, 7(9), p. e45634.
- Thompson, J. R.Everett, R. W.and Hammerschmidt, N. L. (2000) 'Effects of Inbreeding on Production and Survival in Holsteins¹', *Journal of Dairy Science*. Elsevier, 83(8), pp. 1856–1864.
- Thomson, D. W.Bracken, C. P.and Goodall, G. J. (2011) 'Experimental strategies for microRNA target identification', *Nucleic acids research*, 39(16), pp. 6845–6853.
- Thummel, K. E. and Wilkinson, G. R. (1998) 'In vitro and in vivo drug interactions involving human CYP3A', *Annual Review of Pharmacology and Toxicology*, 38(1), pp. 389–430.
- Trapezov, O. VTrapezova, L. I.and Sergeev, E. G. (2008) 'Effect of coat color mutations on

- behavioral polymorphism in farm populations of American minks (*Mustela vison* Schreber, 1777) and sables (*Martes zibellina* Linnaeus, 1758)', *Russian Journal of Genetics*. Springer, 44(4), pp. 444–450.
- Trendelenburg, A. U.Meyer, A.Rohner, D.Boyle, J.Hatakeyama, S.and Glass, D. J. (2009) 'Myostatin reduces Akt/TORC1/p70S6K signaling, inhibiting myoblast differentiation and myotube size', *American Journal of Physiology-Cell Physiology*, 296(6), pp. C1258–C1270.
- Tripurani, S. K.Xiao, C.Salem, M.and Yao, J. (2010) 'Cloning and analysis of fetal ovary microRNAs in cattle', *Animal reproduction science*, 120(1), pp. 16–22.
- Troy, C. S.MacHugh, D. E.Bailey, J. F.Magee, D. A.Loftus, R. T.Cunningham, P.Chamberlain, A. T.Sykes, B. C.and Bradley, D. G. (2001) 'Genetic evidence for Near-Eastern origins of European cattle', *Nature*, 410(6832), pp. 1088–1091.
- Trut, L.Oskina, I.and Kharlamova, A. (2009) 'Animal evolution during domestication: the domesticated fox as a model', *Bioessays*. Wiley Online Library, 31(3), pp. 349–360.
- Tscherner, A.Gilchrist, G.Smith, N.Blondin, P.Gillis, D.and LaMarre, J. (2014) 'MicroRNA-34 family expression in bovine gametes and preimplantation embryos', *Reproductive Biology and Endocrinology*, 12(1), p. 85.
- Ulitsky, I.Laurent, L. C.and Shamir, R. (2010) 'Towards computational prediction of microRNA function and activity', *Nucleic acids research*, 38(15), pp. e160–e160.
- UN (2015) 'World Population Prospects: The 2015 Revision, Key Findings and Advance Tables.', *Working Paper, No. ESA/P/WP. 241*. United Nations Department of Economic and Social Affairs New York, NY.
- Uno, Y.Matsushita, A.Osada, N.Uehara, S.Kohara, S.Nagata, R.Fukuzaki, K.Utoh, M.Murayamay, N.and Yamazaki, H. (2010) 'Genetic variants of CYP3A4 and CYP3A5 in cynomolgus and rhesus macaques', *Drug Metabolism and Disposition*, 38(2), pp. 209–214.
- Utsunomiya, Y. T.O'Brien, A. M. P.Sonstegard, T. S.Sölkner, J.and Garcia, J. F. (2015) 'Genomic data as the "hitchhiker's guide" to cattle adaptation: tracking the milestones of past selection in the bovine genome'.
- De Vaccaro, L. P. (1990) 'Survival of European dairy breeds and their crosses with zebus in the tropics.', in *Animal Breeding Abstracts*, pp. 475–494.
- Valentino, M. A.Lin, J. E.Snook, A. E.Li, P.Kim, G. W.Marszalowicz, G.Magee, M. S.Hyslop, T.Schulz, S.and Waldman, S. A. (2011) 'A uroguanylin-GUCY2C endocrine axis regulates feeding in mice', *The Journal of clinical investigation*, 121(9), p. 3578.

- Valera, A. Pelegrin, M. Asins, G. Fillat, C. Sabater, J. Pujol, A. Hegardt, F. G. and Bosch, F. (1994) 'Overexpression of mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase in transgenic mice causes hepatic hyperketogenesis', *Journal of Biological Chemistry*, 269(9), pp. 6267–6270.
- VanRaden, P. M. Olson, K. M. Wiggans, G. R. Cole, J. B. and Tooker, M. E. (2011) 'Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss', *Journal of Dairy Science*. Elsevier, 94(11), pp. 5673–5682.
- Veerkamp, R. F. Dillon, P. Kelly, E. Cromie, A. R. and Groen, A. F. (2002) 'Dairy cattle breeding objectives combining yield, survival and calving interval for pasture-based systems in Ireland under different milk quota scenarios', *Livestock Production Science*, 76(1), pp. 137–151.
- Venter, J. C. Adams, M. D. Myers, E. W. Li, P. W. Mural, R. J. Sutton, G. G. Smith, H. O. Yandell, M. Evans, C. A. and Holt, R. A. (2001) 'The sequence of the human genome', *Science*, 291(5507), pp. 1304–1351.
- Venuti, S. E. and Helmkamp Jr, G. M. (1988) 'Tissue distribution, purification and characterization of rat phosphatidylinositol transfer protein', *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 946(1), pp. 119–128.
- Vigne, J.-D. (2011) 'The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere', *Comptes rendus biologies*, 334(3), pp. 171–181.
- Visser, P. M. Brown, M. A. McCarthy, M. I. and Yang, J. (2012) 'Five years of GWAS discovery', *The American Journal of Human Genetics*, 90(1), pp. 7–24.
- Visser, P. M. Hill, W. G. and Wray, N. R. (2008) 'Heritability in the genomics era - concepts and misconceptions', *Nature Reviews Genetics*, 9(4), pp. 255–266.
- Voss, A. K. and Fortune, J. E. (1993) 'Levels of messenger ribonucleic acid for cytochrome P450 17 alpha-hydroxylase and P450 aromatase in preovulatory bovine follicles decrease after the luteinizing hormone surge', *Endocrinology*, 132(5), pp. 2239–2245.
- Van Vuure, C. (2005) *Retracing the aurochs: history, morphology and ecology of an extinct wild ox*. Pensoft Pub.
- Walsh, S. W. Fair, T. Browne, J. A. Evans, A. C. O. and McGettigan, P. A. (2012) 'Physiological status alters immunological regulation of bovine follicle differentiation in dairy cattle', *Journal of Reproductive Immunology*.
- Wang, G. van der Walt, J. M. Mayhew, G. Li, Y.-J. Züchner, S. Scott, W. K. Martin, E. R. and Vance, J. M. (2008) 'Variation in the miRNA-433 binding site of FGF20 confers risk for

- Parkinson disease by overexpression of α -synuclein', *The American Journal of Human Genetics*, 82(2), pp. 283–289.
- Wang, H.Xiao, S.Wang, M.Kim, N.-H.Li, H.and Wang, G. (2015) 'In silico identification of conserved microRNAs and their targets in bovine fat tissue', *Gene*, 559(2), pp. 119–128.
- Wang, M.Moisá, S.Khan, M. J.Wang, J.Bu, D.and Loo, J. J. (2012) 'MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation', *Journal of dairy science*, 95(11), pp. 6529–6535.
- Wang, W.Huang, L.Huang, Y.Yin, J.Berk, A. J.Friedman, J. M.and Wang, G. (2009) 'Mediator MED23 links insulin signaling to the adipogenesis transcription cascade', *Developmental cell*, 16(5), pp. 764–771.
- Wang, X.Gu, Z.and Jiang, H. (2013) 'MicroRNAs in farm animals', *Animal*. Cambridge Univ Press, 7(10), pp. 1567–1575.
- Wang, X.Zhang, J.Li, F.Gu, J.He, T.Zhang, X.and Li, Y. (2005) 'MicroRNA identification based on sequence and structure alignment', *Bioinformatics*, 21(18), pp. 3610–3614.
- Wassermann, L.Halwachs, S.Lindner, S.Honscha, K. U.and Honscha, W. (2013) 'Determination of functional ABCG2 activity and assessment of drug-ABCG2 interactions in dairy animals using a novel MDCKII in vitro model', *Journal of pharmaceutical sciences*, 102(2), pp. 772–784.
- Waters, S. M.McCabe, M. S.Howard, D. J.Giblin, L.Magee, D. A.MacHugh, D. E.and Berry, D. P. (2011) 'Associations between newly discovered polymorphisms in the Bos taurus growth hormone receptor gene and performance traits in Holstein-Friesian dairy cattle.', *Animal genetics*, 42(1), pp. 39–49. doi: 10.1111/j.1365-2052.2010.02087.x.
- Weber, J. L. and May, P. E. (1989) 'Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction.', *American journal of human genetics*. Elsevier, 44(3), p. 388.
- Wei, R.Yang, F.Urban, T. J.Li, L.Chalasani, N.Flockhart, D. A.and Liu, W. (2012) 'Impact of the interaction between 3'UTR SNPs and microRNA on the expression of human xenobiotic metabolism enzyme and transporter genes', *Frontiers in genetics*, 3.
- White, S. L.Bertrand, J. A.Wade, M. R.Washburn, S. P.Green, J. T.and Jenkins, T. C. (2001) 'Comparison of fatty acid content of milk from Jersey and Holstein cows consuming pasture or a total mixed ration', *Journal of Dairy Science*. Elsevier, 84(10), pp. 2295–2301.
- Wightman, B.Bürglin, T. R.Gatto, J.Arasu, P.and Ruvkun, G. (1991) 'Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch

- during *Caenorhabditis elegans* development.’, *Genes & development*. Cold Spring Harbor Lab, 5(10), pp. 1813–1824.
- Wightman, B.Ha, I.and Ruvkun, G. (1993) ‘Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*’, *Cell*, 75(5), pp. 855–862.
- Willenbrock, H.Salomon, J.Søkilde, R.Barken, K. B.Hansen, T. N.Nielsen, F. C.Møller, S.and Litman, T. (2009) ‘Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing’, *RNA*, 15(11), pp. 2028–2034.
- Winter, J.Jung, S.Keller, S.Gregory, R. I.and Diederichs, S. (2009) ‘Many roads to maturity: microRNA biogenesis pathways and their regulation’, *Nature cell biology*. Nature Publishing Group, 11(3), pp. 228–234.
- Womack, J. E. (2012) ‘First steps: bovine genomics in historical perspective’, *Animal genetics*. Wiley Online Library, 43(s1), pp. 2–8.
- Wuchty, S.Fontana, W.Hofacker, I. L.and Schuster, P. (1999) ‘Complete suboptimal folding of RNA and the stability of secondary structures’, *Biopolymers*, 49(2), pp. 145–165.
- ‘www.affymetrix.com’ (2016) *Affymetrix*.
- www.animalgenome.org/QTLdb/ (2016) ‘AnimalQTLdb’. NAGRP - Bioinformatics Coordination Program.
- www.bordbia.ie (2015) *No Title*.
- www.genomics.neogen.com (2016) *Neogen Genomics*.
- ‘www.illumina.com’ (2016) *Illumina*.
- Xiong, X.Kang, X.Zheng, Y.Yue, S.and Zhu, S. (2013) ‘Identification of loop nucleotide polymorphisms affecting microRNA processing and function’, *Molecules and cells*, 36(6), pp. 518–526.
- Xu, B.Feng, N.-H.Li, P.-C.Tao, J.Wu, D.Zhang, Z.-D.Tong, N.Wang, J.-F.Song, N.-H.and Zhang, W. (2010) ‘A functional polymorphism in Pre-miR-146a gene is associated with prostate cancer risk and mature miR-146a expression in vivo’, *The Prostate*, 70(5), pp. 467–472.
- Yang, J.-S.Phillips, M. D.Betel, D.Mu, P.Ventura, A.Siepel, A. C.Chen, K. C.and Lai, E. C. (2011) ‘Widespread regulatory activity of vertebrate microRNA* species’, *RNA*, 17(2), pp. 312–326.
- Yarden, Y. and Sliwkowski, M. X. (2001) ‘Untangling the ErbB signalling network’, *Nature reviews Molecular cell biology*, 2(2), pp. 127–137.
- Yin, J.Liang, Y.Park, J. Y.Chen, D.Yao, X.Xiao, Q.Liu, Z.Jiang, B.Fu, Y.and Bao, M. (2012)

- ‘Mediator MED23 plays opposing roles in directing smooth muscle cell and adipocyte differentiation’, *Genes & Development*, 26(19), pp. 2192–2205.
- Yin, J. Q.Zhao, R. C.and Morris, K. V (2008) ‘Profiling microRNA expression with microarrays’, *Trends in biotechnology*, 26(2), pp. 70–76.
- Yoshikawa, Y.Nakayama, T.Saito, K.Hui, P.Morita, A.Sato, N.Takahashi, T.Tamura, M.Sato, I.and Aoi, N. (2007) ‘Haplotype-based case-control study of the association between the guanylate cyclase activator 2B (GUCA2B, uroguanylin) gene and essential hypertension’, *Hypertension Research*, 30(9), pp. 789–796.
- Zancanella, V.Giantin, M.and Dacasto, M. (2014) ‘Absolute quantification and modulation of cytochrome P450 3A isoforms in cattle liver’, *The Veterinary Journal*, 202(1), pp. 106–111.
- Zancanella, V.Giantin, M.Lopparelli, R. M.Patarnello, T.Dacasto, M.and Negrisolo, E. (2010) ‘Proposed new nomenclature for Bos taurus cytochromes P450 involved in xenobiotic drug metabolism’, *Journal of veterinary pharmacology and therapeutics*, 33(6), pp. 528–536.
- Zeng, Y. and Cullen, B. R. (2003) ‘Sequence requirements for micro RNA processing and function in human cells’, *RNA*, 9(1), pp. 112–123.
- Zhang, W.Winder, T.Ning, Y.Pohl, A.Yang, D.Kahn, M.Lurje, G.Labonte, M. J.Wilson, P. M.and Gordon, M. A. (2011) ‘A let-7 microRNA-binding site polymorphism in 3’-untranslated region of KRAS gene predicts response in wild-type KRAS patients with metastatic colorectal cancer treated with cetuximab monotherapy’, *Annals of oncology*, 22(1), pp. 104–109.
- Zheng, Y.Chen, K.Zheng, X.Li, H.and Wang, G. (2014) ‘Identification and bioinformatics analysis of microRNAs associated with stress and immune response in serum of heat-stressed and normal Holstein cows’, *Cell Stress and Chaperones*, 19(6), pp. 973–981.
- Ziats, M. N. and Rennert, O. M. (2014) ‘Identification of differentially expressed microRNAs across the developing human brain’, *Molecular psychiatry*, 19(7), pp. 848–852.
- Zimin, A. VDelcher, A. L.Florea, L.Kelley, D. R.Schatz, M. C.Puiu, D.Hanrahan, F.Pertea, G.Van Tassell, C. P.Sonstegard, T. S.Marcais, G.Roberts, M.Subramanian, P.Yorke, J. A.and Salzberg, S. L. (2009) ‘A whole-genome assembly of the domestic cow, Bos taurus’, *Genome Biol.* 2009/04/28, 10(4), p. R42. doi: 10.1186/gb-2009-10-4-r42.
- Zorc, M.Obsteter, J.Dovc, P.and Kunej, T. (2015) ‘Genetic variability of microRNA genes in 15 animal species’, *Journal of genomics*, 3, p. 51.
- Zuccolo, J.Deng, L.Unruh, T. L.Sanyal, R.Bau, J. A.Storek, J.Demetrick, D. J.Luider, J.

- M.Auer-Grzesiak, I. A. and Mansoor, A. (2013) 'Expression of MS4A and TMEM176 genes in human B lymphocytes', *Frontiers in immunology*, 4.
- Zuker, M. and Stiegler, P. (1981) 'Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information', *Nucleic acids research*, 9(1), pp. 133–148.