Comparative bioinformatics analyses of transcriptome and epigenome data using bovine model systems

Naveed Jhamat
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
Uppsala

Doctoral thesis
Swedish University of Agricultural Sciences
Uppsala 2019

Acta Universitatis agriculturae Sueciae
2019:15

ISSN 1652-6880
ISBN (print version) 978-91-7760-348-1
ISBN (electronic version) 978-91-7760-349-8

Print: SLU Service/Repro, Uppsala 2019

Comparative bioinformatics analyses of transcriptome and epigenome data using bovine model systems

Abstract

Bioinformatics methodology is used to understand complex biological processes in molecular biology, genetics and epigenetics. This thesis performed analysis of the complex biological processes controlling reproduction in mammals including dairy cattle. Successful fertility in dairy cattle is required for farm sustainability as it supports milk productivity and longevity. Disorder in metabolism due to higher levels of insulin during pregnancy and disturbance of immune response due to bacterial infections after calving are detrimental in reproduction. Determinants controlling these molecular mechanisms are poorly understood. To investigate these mechanisms, we performed bioinformatics analyses on data obtained from two experimental studies:

Insulin is a regulator of metabolism and conditions such as hyperinsulinemia is known to impair fertility especially during embryonic development. This study investigates the potential effect of insulin treatment on gene expression and DNA methylation patterns of bovine embryos during *in vitro* oocyte maturation by using the EmbryoGENE DNA Methylation Array. The results of *Paper I* revealed that the identified differentially methylated regions (DMRs) were correlated with differentially expressed genes involved in metabolic regulation.

Uterine diseases impair fertility in dairy cattle and Lipopolysaccharide (LPS) from gram-negative Escherichia coli, is a major source of uterine diseases by activating pro-inflammatory pathways. This study investigates *in vitro* the effects of infection by LPS on bovine endometrial epithelial cells (bEEC) mimicking *in vivo* processes. In *Paper II*, RNA Sequencing analysis revealed that LPS has significantly affected the transcriptome of bEEC, identifying more than 2000 differentially expressed genes, involved in immune response, proliferation, cell adhesion, and implantation. In *Paper III*, Bisulfite sequencing revealed that LPS has profoundly affected the DNA methylation pattern of bEEC. 1291 DMRs were found and their associated genes were involved in molecular processes related to proliferation, apoptosis and embryo development. In *Paper IV*, the enrichment of motifs within these DMRs revealed transcription factor binding sites for immunologically important transcription factors. Thus, the transcriptomics, epigenomics and bioinformatics results obtained from these analyses revealed the complexity of the regulatory transcriptional network activated during inflammation.

Keywords: Insulin, Metabolic imbalance, RNA-Seq, LPS, Endometritis, Implantation, Epigenetics, RRBS, Differential methylation, TFBS.

Author's address: Naveed Jhamat, SLU, Department of Animal Breeding and Genetics, P.O. Box 7023, SE-750 07, Uppsala, Sweden

Dedication

To all my teachers especially Dr. Jehangir Tamimi (Late) and my father Ghulam Qadir Jhamat who has always been a source of inspiration for me.

Education is what remains after one has forgotten what one has learned in school.
Albert Einstein

Contents

List of publications

I Denise Laskowski, Patrice Humblot, Marc-Andre Sirard, Ylva Sjunnesson, **Naveed Jhamat**, Renee Båge, Göran Andersson (2018). DNA methylation pattern of bovine blastocysts associated to hyperinsulinemia *in vitro*. *Molecular Reproduction & Development,* DOI: 10.1002/mrd.22995.

II Yongzhi Guo, Tom Van Schaik, **Naveed Jhamat**, Adnan Niazi, Metasu Chanrot, Gilles Gharpigny, Jean-Francois Valarcher, Erik Bongcam-Rudloff, Göran Andersson, Patrice Humblot. Differential gene expression in bovine endometrial epithelial cells after challenge with LPS; specific implications for embryo maternal interactions. (submitted)

III **Naveed Jhamat**, Adnan Niazi, Yongzhi Guo, Shumaila Sayyab, Gavin Kelsey, Eva Ivanova, Patrice Humblot, Erik Bongcam-Rudloff, Göran Andersson. LPS-treatment of bovine endometrial epithelial cells causes differential DNA methylation of genes associated with endometrial function and inflammation. (manuscript)

IV **Naveed Jhamat**, Adnan Niazi, Yongzhi Guo, Shumaila Sayyab, Patrice Humblot, Göran Andersson, Erik Bongcam-Rudloff. Genome-wide identification of transcription factor binding sites in LPS-associated differentially methylated regions in bovine endometrial epithelial cells. (manuscript)

The contribution of Naveed Jhamat to the papers included in this thesis was as follows:

I. Partly contributed in bioinformatics analysis and compiling results.

II. Partly contributed in bioinformatics analysis, compiling results and drafting manuscript.

III. Developed bioinformatics work flow, performed analysis, annotation, compiled the results and wrote manuscript.

IV. Performed bioinformatics analysis, compiled the results and wrote manuscript.

List of figures

1 Introduction

1.1 Bioinformatics

Bioinformatics is an interdisciplinary field that applies the knowledge of computer science, applied mathematics and statistics to decode, analyse and organise the information associated with molecular genetics and genomics. Broadly, the three major aims of bioinformatics are to; *i)* organise biological data for efficient retrieval and to store new results *ii)* develop tools and resources that facilitate data analysis and *iii)* use tools for data analysis and interpretation of the results for the biologists (Luscombe et al., 2001).

Although it is difficult to pinpoint about the inception of bioinformatics, however this term was reported in 1977 by Paulien Hogeweg when she pronounced bioinformatics as her main field of research and established bioinformatics group at University of Utrecht (Hogeweg, 1978; Hogeweg and Hesper, 1978). Dr. Margret Oakley Dayhoff is also considered as one of the pioneers in bioinformatics who dedicated herself in applying the evolving computer technologies for research in medicine and biology. In '60s, she developed protein and nucleic acid databases and computer methods for comparison of protein sequences to derive evolutionary history of biological kingdom. Her database was considered as the first *Atlas of Protein Sequence and Structure* (Hunt, 1983).

The emergence of DNA sequencing technologies was a revolution in producing biological data at a phenomenal rate which made the use of computers inevitable for biological research. Many publicly open databases were developed to feed

the need and Attwood and colleagues narrated in their paper (Attwood et al., 2011) the emergence and history of theses biological databases such as Protein Data Bank (PDB) was established in 1971 and a computer-based archival file for three-dimensional structural data of macromolecules (Bernstein et al., 1977); The EMBL Nucleotide Database was founded in 1980 followed by the EMBL Biocomputing programme in 1984. In 1987, four years before the worldwide web was created, EMBL began serving biological data over the internet, following up in 1988 with a fileserver for bioinformatics software. EMBL Nucleic archive founded was the first internationally supported central resource for nucleic acid sequence data (Hamm and Cameron, 1986); GenBank was founded in 1982 at NCBI and it works in close collaboration with the EMBL and DNA Data Bank of Japan (Burks et al., 1985). In this era of data explosion, new set of tools started to emerge for biological data mining and annotation. Currently bioinformatics has become a necessary and integral part of biological research, as it includes the analysis of genomic sequence data, genome assembly and annotation, comparative genomics, epigenomics, protein structure prediction, development of programs and databases (Stein, 2001; Visel et al., 2009).

1.1.1 DNA sequencing technologies

The era of genetic sequencing was started by Walter Gilbert through chemical sequencing method (Maxam and Gilbert, 1977) and Frederick Sanger through chain termination method (Sanger et al., 1977), both shared the Nobel Prize in 1980 for their contributions regarding the determination of base sequences in nucleic acids. The molecular biology community rapidly adopted the new DNA sequencing technologies in the late 70's and early 80's and Larhammar and colleagues were among the very first to use these techniques to completely sequence a human MHC class II gene (Larhammar et al., 1983). The Sanger sequencing was known as first generation sequencing technology that allowed only a few hundred base pairs of DNA fragments to be sequenced per reaction. This technique turned out to be parallelised and automated with Human Genome Project in 1990´s whose first draft was completed in 2001 (Lander et al., 2001; Venter et al., 2001).

Sanger sequencing was quite expensive with cost of $0.05 per kilo-base (Shendure and Ji, 2008). Other limitations of Sanger sequencing, such as scalability, throughput and speed, stimulated the emergence of a new technology based on combination of Sanger sequencing and advanced fluorescent detection

methods, which was known as next-generation sequencing (NGS) and considered as second-generation sequencing or high-throughput sequencing (HTS). NGS brought a revolution in genomics, transcriptomics, and epigenomics and made comparative genomics possible by sequencing related organisms and resequencing many individuals from the same and between different population and subsequently comparing their sequences to comprehend how genetic variations affect phenotypic features.

The first NGS platform was the 454 technology which was introduced in 2005, and this was shortly followed by the Solexa/Illumina genome sequencing platform and SOLiD by Applied Biosystems (van Dijk et al., 2014). The NGS platforms have distinct features such as variable read lengths, type of sequencing, run times and throughput capacity (Goodwin et al., 2016). At the first release, these technologies generated short reads of 35bp (SOLiD/Illumina) to 110bp (454 Roche) but later on these produced around 20M reads (454 Roche), and 30M reads (SOLiD) at comparatively very low cost and in lesser time than Sanger sequencing (Mardis, 2008; van Dijk et al., 2014). This development opened a new era of research projects using DNA sequencing that enhanced the need and importance of Bioinformatics. The Illumina has multiple platforms such as HiSeq, MiSeq, and NextSeq, with variable read lengths and run times.

Since then, many new technologies have been developed which increased their output, read length, and quality along with decreasing its cost. Ion-torrent (from Life technologies) was the first semiconductor-based platform that can generate up to 1Gb of data, with a longer read lengths up to 400 bases. The newer versions, Ion proton and Ion S5, can even produce up to 15Gb of data. The latest NGS platforms from Pacific Bio (Rhoads and Au, 2015) and Oxford Nanopore have been introduced to produce longer sequences of more than 200 kb (Jain et al., 2016; Tyson et al., 2018). Illumina HiSeq 2000 was able to produce around 600Gb data per run which was sufficient data for 6 human genomes in 11 days (Mardis, 2013). The era of third generation sequencing has been started with single molecule sequencing which allows to sequence the samples without amplifying the included DNA (Chaisson et al., 2015). This reduces the preparation time and cost and allows to sequence the unknown microbes but error rates are high (Mardis, 2013; Shapiro et al., 2013).

1.2 Genetics and transcriptomics

Genetics is the field of biology which, study genes, genetic variations and heredity in living organisms and has strong link with information systems to decode the genetic information into biological information. All living organisms are made up of cells that contain genetic information on DNA in the form of nucleotide bases (the different bases are Adenine, Thymine, Cytosine and Guanine that are abbreviated A, T, C, and G, respectively). These nucleotide bases are arranged in a certain DNA sequence throughout a genome. A particular pattern, combination or permutation of the subset of DNA sequence provides a template for genes that carry a piece of genetic instructions. These instructions are used for making proteins and regulating the chemical reactions for development and survival of the organism. These genes can be transferred from parents to offspring through inheritance. With some exception, all the cells of an individual organism have the same DNA sequence but particular genes are expressed in different cell types, tissues and organs at different levels. Some genes are turned on and off in context of environmental changes (diet, exercise, medications, metabolism, and infection) that leads to phenotype differences or disease. During transcription, template strand of DNA is copied to produce many different types of RNAs (see below) including messenger RNAs (mRNAs) which are translated into proteins (Alberts et al., 2014; Lodish et al., 2007).

The transcriptome is the set of all RNA molecules in one cell or a population of cells whereas transcriptomics technologies are used to study an organism's transcriptome. DNA stores the genetic information, which is expressed through transcription. The definition of transcription is DNA-dependent, RNA polymerase-mediated synthesis of RNA. As mentioned above, mRNA acts as a transient intermediary molecule before protein synthesis while noncoding RNAs perform additional functions. A non-coding RNA (ncRNA) is transcribed from DNA but not translated into protein and performs different functions within the cell by taking part in transcription and translation. The DNA sequence used for transcription of ncRNA is referred as RNA gene. There are different forms of ncRNAs:

- The two larger subtypes of ribosomal RNAs (18S and 28S rRNAs) are transcribed by RNA polymerase I. These subunits along with 5S rRNA and associated proteins form ribosomes which is a cellular machinery to translate mRNAs into proteins by reading mRNAs and linking amino acids accordingly to produce polypeptide chain (Simsek et al., 2017).
- Transfer RNAs (tRNAs) are transcribed by RNA polymerase III. It is an adaptor molecule that links codons of mRNA to corresponding

amino acids by carrying the amino acids to the ribosomes (Sharp et al., 1985).

- Small nuclear RNAs (snRNAs) are transcribed by either RNA polymerase II or III and forms spliceosome in the nucleus with the help of associated proteins called snRNPs. Spliceosome is a molecular machinery that removes introns from the pre-mRNA to generate mature-mRNA and the process is referred as splicing. snRNAs also support in the regulation of transcription factors or RNA polymerase II as well as maintaining telomeres which are regions of repetitive nucleotide sequences at the end of chromatid which protect the ends of chromosome from deterioration during replication (Matera et al., 2007).
- Small nucleolar RNAs (snoRNAs) are mostly located on introns of the genes which transcribed by RNA polymerase II and guide chemical modifications of rRNAs, tRNAs and snRNAs primarily through methylation or pseudouridylation (Mannoor et al., 2012).
- Micro RNAs (miRNAs) are part of RNA interference (RNAi) which are usually transcribed by RNA polymerase II; and are post transcriptional regulators of gene expression through base pairing with target complementary messenger RNAs (mRNAs) that results in gene silencing through translation repression or target degradation (Ambros, 2004; Zhang et al., 2018)
- Small interfering RNA or short interfering RNA (siRNA) is also part of RNAi and the Dicer enzyme catalyzes its production from double-stranded RNA or pre-microRNA. It functions in the same way like miRNA to perform post transcriptional gene silencing (PTGS) as a result of mRNA degradation and preventing translation (Carthew and Sontheimer, 2009).
- Long non-coding RNAs (lnc RNAs) are larger than 200 nucleotides and often transcribed by RNA polymerase II; and play a role in chromatin remodeling, transcriptional and post transcriptional regulation such as X-inactive specific transcript (Xist) has role in X-chromosome inactivation process in female mammals (Ransohoff et al., 2018).

A transcriptome represents the total sets of transcripts present in a cell at a particular time point. Measuring the quantity and quality of mRNA expression in different tissues, conditions or time points contributes to an increased knowledge of how gene expression is regulated. Furthermore, defining the transcriptome is valuable to improve the genome annotation of a particular species. The possibility to study the whole transcriptome started in the early 1990s and technological advances have made transcriptomics a widespread

discipline (Lowe et al., 2017). There are two key techniques / approaches of transcriptomics to define differential gene expression:

1.2.1 Microarray

DNA microarray or (DNA Chip) or (Bio Chip) is a collection of microscopic DNA spots attached to a solid surface. Synthesized oligonucleotide probes complementary to the corresponding mRNAs are placed on an array. Then a sample of mRNAs are allowed to hybridize to the probes. Scientists use DNA microarrays to measure the expression levels of large numbers of genes (Bumgarner, 2013).

1.2.2 RNA-Sequencing (RNA-Seq)

With the advent of NGS and advances in technology, RNA-Seq has become technology of choice for gene expression profiling. RNA-Seq provides the sequence and frequency of RNA molecules that are present at any particular time point in a specific cell type. Counting the number of mRNAs that are encoded by individual genes provides an indicator of protein coding potential, a main contributor to phenotype development.

RNA-Seq refers to the blend of different high-throughput sequencing methodologies with computational methods to capture and quantify transcripts present in an RNA extract. The nucleotide sequences generated are typically around 100 bp in length, but their read length can vary from 30 bp to over 10,000 bp, depending on the sequencing method used. RNA-Seq may be used to identify the expressed genes within a genome or identify which genes are active at a particular point in time, and read counts can be used to accurately model the relative gene expression level. The transcriptomics field has been emerged swiftly with the initiation of next-generation sequencing technologies. RNA-seq has now largely displaced microarrays and taken the place as preferred method for gene expression profiling (Lowe et al., 2017; McGettigan, 2013). However, the cost-efficiency and efficacy of using microarrays are still used as a valuable resource to define differential gene expression (*Paper I*).

Direct sequencing using Oxford Nanopore MinIon and similar technologies will give the complete RNA sequence from nucleotide 1 to the last nucleotide of all the mRNAs and all the splice variants. This combined with Illumina-based

RNA-seq, which will allow quantification is the future in RNA studies. Then when all the mRNAs have been identified there will be microarrays with probes for all the different RNAs that are known to be expressed.

1.3 Epigenetics, DNA methylation, RRBS

Conrad Waddington coined the term *epigenetics* in the early 1940s (Waddington, 1942, 2012) and defined epigenetics as ‘‘the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being’’(Waddington, 1968). Waddington introduced the concept of epigenetic landscape to illustrate embryonic development that a cell can adopt various developmental pathways towards differentiation. The cellular differentiation occurs multiple times during the development of a multicellular organism and these changes are governed by epigenetic modifications rather than alteration in DNA sequence. Critical epigenetic reprogramming arises throughout germ cell development and early embryogenesis in mammals (Dupont et al., 2009). Except B lymphocytes and T lymphocytes that rearrange their specific Ig and TCR genes, respectively, all the cells in the body contain the same DNA sequence but the different genes are expressed in different cell types with different expression levels that causes phenotype differences (Pontén et al., 2008; Thul et al., 2017). These phenotype differences are due to epigenetic phenomena that control changes in phenotype without a change in genotype. Epigenetic change is a regular and natural phenomenon but can also be influenced by many factors such as age, environment, lifestyle and disease state (Goldberg et al., 2007). Epigenetic changes due to external / environmental factors, lead to changes in chromatin structure and thus affect /modulate / regulate /control the gene expression.

The word epigenome is derived from epigenetics and refers to the complete description of chemical changes to the DNA and histone proteins of an organism that ultimately modulate chromatin structure and genome function; and these changes can, if they influence the germline, be transferred to the next generation (Bernstein et al., 2007). The epigenome may be dynamically changed by environmental factors and plays a vital role in regulation of gene expression, development, cell / tissue differentiation, and inhibition of transposable elements (Conley and Jordan, 2012). Epigenomics is the study of epigenome *i.e.* the complete set of epigenetic modifications (Russell, 2009). These modifications are reversible and epigenome maintenance is a continuous dynamic process,

ensuring the stability of eukaryotic genomes *e.g.* DNA repair (Alabert and Groth, 2012).

The development of high-throughput technologies has made it possible to perform whole genome sequence studies of epigenetic modifications of different cell types from different organisms. Broadly, two epigenetic modifications are considered as being the most important; *i)* histone modification and *ii)* DNA methylation.

1.3.1 Histone modification

Chromosomes are compressed in the form of chromatin and chromatin consists of DNA and proteins, primarily four different types of histone. DNA is wrapped around the core eight histones (two of each H2A, H2B, H3, and H4) forming the nucleosome which is a repeating unit and smallest functional unit of chromatin. Histones are responsible for maintaining the shape and structure of chromatin. Any change in chromatin is considered as the key regulator of genomic functions (Fischle et al., 2003; Downs et al., 2000). Epigenetic modifications are covalent post-translational modifications that occur at the amino terminal tails of the histones which alter chromatin organisation and ultimately availability of genes in DNA to be activated (Lennartsson and Ekwall, 2009), such as histone acetylation and histone methylation. Acetylation of histone plays vital role in gene regulation and is controlled by the balance in activity of two enzymes, *i)* Histone Acetyltransferase (HAT) and *ii)* Histone Deacetylase (HDAC). Acetylation of histones by HAT triggering the uncoiling of DNA and opening the chromatin structure which causes the genes to be accessible for transcription factors and allow the transcription. When transcription is no longer required, then deacetylation of histone by HDAC results in wrapping the DNA and closes the chromatin structure. Histone methyltransferase (HMT) is responsible for histone methylation by transferring methyl group on the target residues of histones, which causes to increase or decrease gene expression depending on which residue of histone is methylated with how many methyl groups, *e.g.* it is reported in mouse study, if methylation occurs on lysine residue 3 (L3) of histone 3, it unwraps the DNA but if it occurs on L9 of histone 3, it condenses the DNA, such as in X-inactivation. Although, histone modifications are very significant in epigenetic studies but we have not explored this aspect as yet in our current research.

1.3.2 DNA methylation

DNA methylation is considered as the first and major epigenetic modification (Wyatt, 1951) which is essential for normal development and plays a decisive role in many biological processes, such as cell differentiation, gene expression, genomic imprinting and embryogenesis (Yuan et al., 2016). DNA methylation preferably occurs at 5′-CpG-3′ dinucleotides (Gardiner-Garden and Frommer, 1987) and if it occurs in the promoter region or first exon of a gene then normally it leads to suppression of gene expression (Su et al., 2014a). DNA methylation is performed by two classes of DNA methyltransferases (DNMTs), called *de novo* DNMTs and maintenance DNMTs, which ensure that the daughter cells after cell division are properly methylated on both strands (Okano et al., 1999).

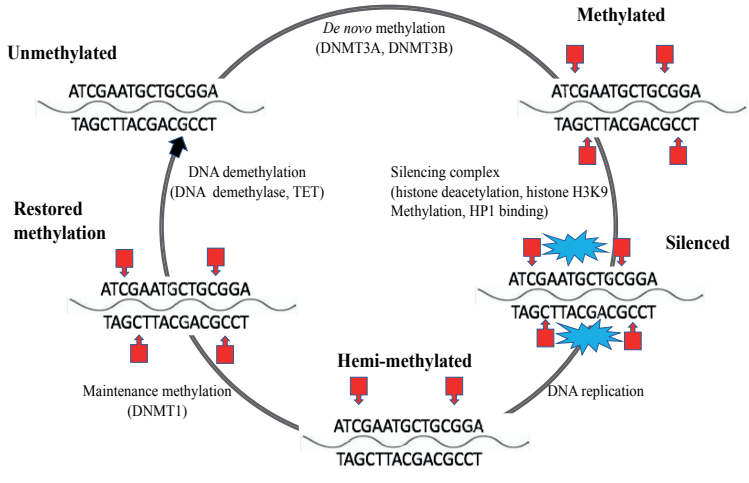

Figure 1. Methylation and Demethylation cycle. Adapted from *Nature Reviews Cancer* (Issa, 2004).

Passive DNA demethylation may occur during cell division when DNMT1 levels are suboptimal or when this enzyme is inactive. For example this may occur with increasing age. However, active DNA demethylation is dependent on DNA demethylase enzymes in stem cells and early embryos. TET family member proteins, are reported to convert 5-methylcytosine to 5-

hydroxymethylcytosine in human and mouse studies (Carey et al., 2011; Feng et al., 2010; Ito et al., 2010).

Aberrant DNA methylation patterns of promoter regions are associated with heritable losses of gene function and formation of transcriptionally repressive chromatin that might be major cause of cancer (Baylin et al., 2001; Baylin and Herman, 2000; Jones and Laird, 1999). The alteration of LPS-induced immune responses in bovine dermal fibroblasts cells is epigenetically regulated (Green and Kerr, 2014; Walker et al., 2015). During early pregnancy in cattle, DNA methylation is correlated with gene expression in the endometrium (Walker et al., 2013) and may regulate the uterine response during implantation and aberrant DNA methylation may result in pregnancy loss (Walker and Mitchell, 2013). CpG islands are regions in the genome of DNA with high frequency of CpG sites having *i)* length of sequence at least 200 bp *ii)* GC content > 50% *iii)* ratio of observed to expected CpGs > 0.6. The global DNA methylation profile of embryos with different kinetics of development revealed that more hypermethylated regions were distributed on introns, exons and promoters in fast growing embryos as compared to slow growing embryos in which more hypermethylated regions were found in CpG islands (Ispada et al., 2018). Genome-wide DNA methylation analysis of bovine embryos derived *in vivo*, subjected to *in vitro* culture before or during or after the time of embryonic genome activation (EGA), has reported increased number of hypomethylated genomic loci in blastocysts during (EGA) (Salilew-Wondim et al., 2018). The analysis of WGBS from mammalian placentas revealed lower global methylation levels as compared to their somatic tissues and higher gene body methylation pattern was the conserved feature among all mammalian placentas (Schroeder et al., 2015). In our current study *Paper I*, DNA methylation profiling of bovine blastocysts with high concentrations of insulin *in vitro* identified differentially methylated regions which were correlated with genes expression related to metabolic regulation. In mammals, DNA methylation occurs mostly in CpG context but in rare cases non-CpG context (CHG, CHH) is also reported in embryonic stem cells and brain cells, however non-CpG context methylation is reported most abundant in plants (Law and Jacobsen, 2010; Stroud et al., 2014).

However, DNA methylation is complicated to analyse experimentally as it does not change DNA sequence and is not preserved during cycles of polymerase chain reaction (PCR) because DNA polymerase is unable to discriminate between the methylated and unmethylated cytosines (Kristensen and Hansen, 2009). The advent of next-generation DNA sequencing technology has

empowered to investigate the DNA methylation dynamics of key biological functions (Guo et al., 2014; Smith et al., 2014). The major sequencing technologies to investigate genome-wide DNA methylation are: *i)* methylated DNA binding domain sequencing (Aberg et al., 2012), *ii)* methylated DNA immunoprecipitation sequencing (Taiwo et al., 2012a), *iii)* whole genome bisulfite sequencing (WGBS) (Lister et al., 2009) and *iv)* reduced representation bisulfite sequencing (RRBS) (Nagarajan et al., 2014). The first two technologies employ the enrichment of methylated DNA to obtain utmost resolution up to 150 bp (Harris et al., 2010) whereas the last two attain single-base resolution using bisulfite conversion, and bisulfite conversion methods are more precise than enrichment methods, furthermore RRBS is cost effective and reduces the amount of sequencing required as compared to WGBS (Bock et al., 2010; Harris et al., 2010). Some other sequencing technologies are also being used to detect DNA methylation; such as PacBio sequencing, which directly detect DNA methylation without bisulfite conversion, through single-molecule, real-time (SMRT) sequencing (Flusberg et al., 2010), Nanopore sequencing detect imprinted DNA methylation through long-read sequencing (Gigante et al., 2018) and whole-genome methylation profiling of single cells through single-cell combinatorial indexing for methylation analysis (sci-MET) (Mulqueen et al., 2018).

Although a few studies of genome-wide DNA methylation pattern were reported in pigs, sheep, horses and cattle with low resolution and limited tissue types. Furthermore, the reported studies in cattle on muscle and placental tissues using methylated DNA immunoprecipitation sequencing (MeDIP-seq) had less resolution. (Cao et al., 2015; Couldrey et al., 2014; Gao et al., 2014; Huang et al., 2014; Lee et al., 2014; Su et al., 2014b), but still our understanding of DNA methylation profiles in cattle is limited as compared to humans and rodents (Zhou et al., 2016a).

In livestock, the use of the RRBS technique has been reported for the detection and quantification of DNA methylation at single base resolution (Doherty and Couldrey, 2014). RRBS approach enriches for CpG regions by digesting genomic DNA with single restriction enzyme *Msp*I targeting at 5′ -CCGG- 3′ to ensure that each fragment covers minimum one CpG site (Gu et al., 2011, 2010). During bisulfite conversion, methylated cytosines remain unaffected while unmethylated cytosines are converted into uracil (Frommer et al., 1992). Although, RRBS has been widely applied for DNA methylome research in human (Pei et al., 2012) and other model organisms (Chatterjee et al., 2013; Hartung et al., 2012; Meissner et al., 2008), but seldom in cattle and specifically

the understanding of methylation profiles in bovine endometrial epithelial cells (bEEC) remains unknown.

In our current research, we used RRBS because it is an effective and representative method to describe the DNA methylation profiling on a genome-wide level due to targeting specifically CpG rich regions. Other reasons are the relative low cost and high coverage as compared to WGBS and that the method can provide high resolution along with the information of all three methylation contexts (CG, CHG and CHH) as compared to MeDIP-seq and methyl-binding domain sequencing (MBD-seq) (Choi et al., 2015; Meissner et al., 2005; Zhou et al., 2016). Our current study is the first to investigate DNA methylation profiles in bEEC challenged with lipopolysaccharide (LPS).

1.4 Transcription Factors and Transcription Factor Binding Site mining

Transcription factors (TFs) are modular regulatory proteins that bind to a specific DNA sequence to control the rate of gene expression. TFs activate, repress or suppress gene transcription by supporting or blocking RNA polymerase binding to DNA (Latchman, 1993; Roeder, 1996). TFs occur in families and the members of these families share similar types of binding domains. During evolution, these families expanded and diverged by a combination of gene duplication and mutation and ultimately fully redundant, partially redundant and completely nonredundant TFs were emerged. This expansion usually correlates with organismal complexity (Grove et al., 2009) *e.g.* basic helix loop helix (bHLH) family contains 42 members in the nematode *Caenorhabditis elegans* (Reece-Hoyes et al., 2005) and more than 100 members in humans (Simionato et al., 2007). This increased complexity in mammals is reflected by more than 2000 genes encoding TFs (Venter et al., 2001; International Human Genome Sequencing Consortium, 2001) .

TFs are modular proteins consisting of at least one DNA-binding domain (DBD), containing specific motif that can recognize short specific DNA sequences, *i.e.* referred to as Transcription Factor Binding Sites (TFBSs) (Mitchell and Tjian, 1989). Furthermore, they have activation/repression domains and dimerization domains. The binding of TFs to their specific TFBS regulatory *cis*-acting DNA sequences is the critical feature in controlling transcription. TFBSs are short and often well-conserved sequences, called motifs that are located in regulatory regions such as promoters, enhancers and silencer

elements in non-coding DNA sequences (Wei and Yu, 2007). Transcription factors act as molecular switches in turning genes on and off and the identification of TFBSs and the TFs that bind to them provides basic and critical understanding of gene regulatory networks (Selvaraj and Natarajan, 2011).

TFs establish protein-protein interactions with activators or repressors and the assembly of pre-initiation complex (PIC) is required for regulated gene transcription. This process involves the binding of at least six general TFs (GTFs), TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH, to the core promoter, which helps RNA polymerase II to establish a stable PIC and effectively initiate transcription. The elements present in most mammalian core promoters are the TATA-box located at position -26 to -31 relative to the start site of transcription (TSS), Initiator sequences (INR) overlapping the TSS, TFIIB recognition element (BRE) immediately 5' of the TATA-box at position -37 to -32 relative the TSS and finally the downstream promoter element (DPE) located at position +28 to +32 relative the TSS. The binding of TATA binding protein (TBP), which is a subunit of TFIID to the TATA-box of core promoter leads to the assembly of other GTFs and RNA polymerase II. It should be noted that some genes lack bona fide TATA-boxes and are called TATA-less promoters. In such promoters, transcription start sites are not as strictly localized. Furthermore, other regulatory TFs bind to proximal promoter, silencers and enhancers to regulate transcription (Levine and Tjian, 2003).

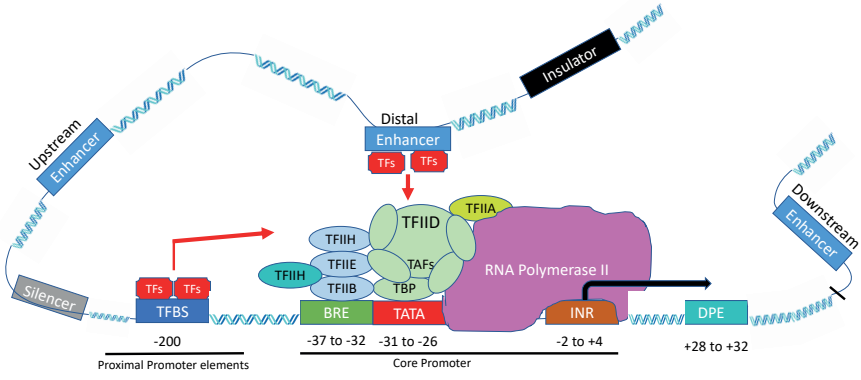

Figure 2. Complex metazoan transcriptional control modules showing the assembly of pre-initiation complex at the core promoter in combination with GTFs and RNA polymerase II along with proximal and distal regions. The figure is adapted from (Levine and Tjian, 2003).

In the 1960s, Jacob and Monod presented a model for gene regulation in response to cellular environmental conditions; depicted an operon as model to control the transcription of a set of structural genes that code proteins used for metabolic functions. Mark Ptashne considered this model as advent of new era in molecular biology and highlighted its historical perspective. Lac operon is an inducible set of genes that breakdown the lactose into sugar which is used for cellular metabolism. A repressor inhibits gene expression by binding upstream of the controlling region (operator) of lac operon. Lactose is the inducer molecule for the lac operon and when lactose enters into cell environment, it binds to the repressor that releases the operator and allowing the RNA polymerase to start transcription of operon in *E. coli*. Mainly two proteins are produced after translation of polycistronic mRNA by ribosomes; *i)* Beta-Galactosidase, which breakdown the cellular lactose into sugar, glucose and galactose *ii)* Permease, which binds to the cell membrane that enhances the rate of importing lactose from outside the cell. This process continues until the concentration of lactose becomes low and the lactose bound to repressor are released. This allows the repressor to bind again the operator of lac operon and stop the gene expression. This genetic regulatory mechanism continues for synthesis of required proteins responsible for lactose metabolism when the lactose concentration is high in the cell and glucose level is low (Jacob and Monod, 1961; Lewis, 2011).

The prediction of TFBSs or regulatory regions on a whole genome level is the challenge for genomics and epigenomics studies. It is evident that these TFBSs are highly conserved in related species and their conservation is required to predict a non-coding DNA sequence as TFBS (Pritsker et al., 2004). As described above, high throughput technology of NGS has enabled researchers to perform whole genome DNA sequencing of multiple species at low cost and in short time. Now, whole genome DNA sequence datasets of different species in the form of databases are available for mapping to predict potential regulatory regions (McGuire et al., 2000). Resources such as TRANSFAC, JASPAR, ENCODE Project, modENCODE and Mouse ENCODE are available and allow to determine all these functional elements in the genomes of human and model organisms.

We used oPOSSUM application programming interface (API) to access oPOSSUM database for the prediction of potential TFBS located in differentially methylated regions in response to LPS-treatment of bEEC (*Paper III*). To obtain optimal results from oPOSSUM, it should be ensured that *i)* Average length of the sequences in target dataset must match to the average lengths of sequences in background *ii)* GC composition of the target must closely match to the GC composition of the background dataset (otherwise a bias will be introduced) *iii)* Number of background sequences must be >= number of target sequences. In addition, oPOSSUM-3 database provides cluster analysis of TFBSs as TFs are classified into families / classes on the basis of DNA binding domains. Consensus similarity of domain varies in classes, some have almost identical and some have variation in their domains. oPOSSUM-3 provides the result having subsets of profiles that are almost identical by dividing the structural classes into clusters based on profile similarity whereas TRANSFAC uses MATCH algorithm to construct position weight matrix (PWM) for prediction of potential TFBSs (Kwon et al., 2012; Wingender et al., 1996).

1.5 Bovine model system in reproduction

Reproductive performance in dairy cows is a complex phenotype and the underlying molecular mechanisms in dairy cows are only poorly understood. Major efforts are made to increase our understanding of the factors influencing successful fertility and this is required for the farm sustainability as it supports milk productivity and longevity. Massive genetic selection has increased their milk yield but there is steady decline in reproductive performance (Garnsworthy et al., 2008; Miglior et al., 2005). Disorder of metabolism by hyperinsulinemia

and disturbance of immune response by microbial infections are considered very crucial in reproduction (Pasquali et al., 2007; Park et al., 2016). To investigate the underlying factors influencing these molecular mechanisms, we performed bioinformatics analysis on data obtained from two experimental studies:

Study 1: Insulin is a critical regulator of metabolism and conditions such as hyperinsulinemia is known to impair fertility especially during embryonic development. For normal conditions during pregnancy correct levels of insulin play a crucial role to obtain appropriate regulation of metabolism. The molecular mechanisms underlying hyperinsulinemia are only partly understood. The study was designed to investigate the dose dependent effect of insulin treatment during bovine *in vitro* oocyte maturation at the epigenetic and transcriptomics level. Differential DNA methylation and changes in gene expression patterns, was examined using two different insulin concentrations, 0.1 μg/ml (INS0.1) or 10 μg/ml (INS10) on oocytes. Samples were drawn at week 8 from blastocysts originated from control and insulin treated oocytes. The experiments in *Paper I* were performed to find the potential effect of insulin treatment on gene expression and DNA methylation patterns of embryos during *in vitro* oocyte maturation by using the EmbryoGENE DNA Methylation Array (EDMA).

Study 2: Uterus is an important organ that has pivotal role from conception to successful delivery. It provides pathway for the sperm to reach the ovum, implant the zygote and develop, protect and nourish the fetus until birth (Soffar, 2015). Uterine diseases significantly impair reproductive performance and are considered as the main sources of delayed breeding and infertility (Grimard et al., 2006), which lead to milk production losses and enhance frequency of slaughtering infected animals (Royal et al., 2000). Pathogens such as specific strains of gram-negative Escherichia coli (*E. coli*) and Bovine Herpes Virus 4 (BoHV-4) are among the major sources of uterine immune-mediated diseases. They dysregulate transcription of the host cells by influencing the epigenetic factors such as DNA methylation and histone modifications (Bierne et al., 2012; Haller et al., 2003; Lieberman, 2006; Takahashi et al., 2011), which leads to disease development and reduces conception rates and enhances pregnancy loss (Galvão et al., 2009; Opsomer et al., 2000). Lipopolysaccharide (LPS) expressed on the surface of Gram-negative bacteria (*E. coli*) is considered as strong inducer of immune responses, activates pro-inflammatory pathways, dys-regulates the function of endometrial cells and is a key player in the mechanisms involved in endometritis (Herath et al., 2009; Williams et al., 2007). How the epigenome is vulnerable to DNA methylation in bovine endometrial epithelial cells (bEEC) during inflammation; and epigenetic

factors that modulate gene regulation and molecular mechanisms that are involved in disease development and impair fertility are unclear. To investigate the effects of LPS on bEEC on *i)* transcriptome (gene expression), ii) epigenetic factors (DNA methylation patterns) that modulate gene regulation, and iii) transcription factors (TFs) that bind to a specific DNA sequence to control the rate of gene expression regarding reproduction, we have established an *in vitro* model for bovine endometrial inflammation caused by gram-negative bacteria, using primary bEEC.

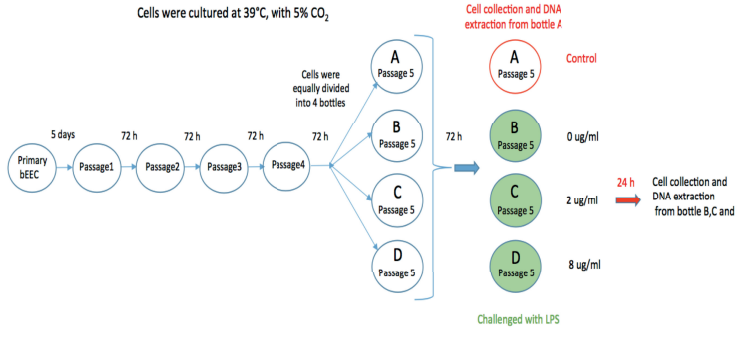

Figure 3. The cell culture protocol of bEEC. For control sample at time 0 hours, DNA was extracted from bottle A after passage 5. At this time bottles B, C, and D were treated with 0, 2, and 8 µg/mL of LPS, respectively. After 24 hours, DNA was extracted separately from bottles B, C, and D for treated samples.

Cells were obtained from three cows after slaughter and after isolation grown in tissue culture and treated with 0, 2, and 8 µg/mL LPS from *E. coli* (O111:B4) for 24 hours. These concentrations of LPS were chosen because *in vivo* these concentrations were reported in uterine fluid of cows during clinical endometritis and related to earlier experiments exposing the effect of LPS on cell survival and proliferation profiles (Herath et al., 2009; Williams et al., 2007). The time at LPS challenging was set as 0 hours. The samples of total RNAs were extracted at time 0 hours and after 24 hours of LPS treatment; considering control samples and treated samples, respectively. *Paper II* was performed using this experimental design to identify differentially expressed genes (DEGs) in bEEC after challenge with LPS and their implications for embryo maternal interactions by using RNA sequencing (RNA-seq). For *Paper III*, genomic DNA was extracted from same cells (bEECs) with same

concentration of LPS treatments at the same time points. To investigate whether LPS treatment had an effect on DNA methylation levels in the bEECs, we used reduced representation bisulfite sequencing (RRBS) for identification of differentially methylated regions (DMRs) and their association with gene expression in the previous study (*Paper II*). For *Paper IV*, the identification and classification of the genome-wide transcription factor binding sites (TFBSs) located within the identified DMRs in *Paper III* was performed by using oPOSSUM databases. Results obtained from studies of endometritis in cattle may also be used as a model for endometritis in women and thereby having a substantial comparative value.

1.6 Impact of insulin on oocyte maturation

Insulin is a hormone produced by beta cells in the islets of Langerhans of the pancreas that controls blood sugar levels within a normal range as part of metabolism and allows the cells to take up glucose. Disturbance of normal insulin levels leads to severe health problems such as: *i)* Type 1 diabetes; in which pancreas does not produce enough insulin and consequently the blood sugar level rises up, *ii)* Type 2 diabetes; in which the body does not use insulin properly and pancreas tries to make more insulin to control the blood sugar levels, thus becomes insulin resistance, and *iii)* Gestational diabetes; if high blood sugar levels and insulin resistance occur during pregnancy (Sonksen and Sonksen, 2000). Exact causes of these types of diabetes are not known, however researchers believe in some risk factors that may more likely to involve such as genetic, environmental, overweight or inactive, diabetes history in family, high blood pressure etc. (Antosik and Borowiec, 2016; Cryer et al., 2016; Dendup et al., 2018; Lin et al., 2018).

Hyperinsulinemia refers to conditions associated with excessive levels of insulin circulating in the blood whereas hyperglycemia (high blood sugar) refers to excessive amount of glucose in the blood plasma. Insulin being the critical regulator of metabolism play crucial role during successful pregnancy. Obesity and overfeeding elevate the risk of diabetes type 2 which lead to insulin resistance and hyperinsulinemia (Kahn et al., 2006) and prolonged hyperinsulinemia may also lead to dyslipidemia. Metabolic imbalance, caused by changes in the blood profile such as hyperinsulinemia, hyperglycemia and dyslipidemia, leads to female infertility (Pasquali et al., 2007). The adverse consequences of insulin on reproductive performance has been reported in bovine as well as in humans (Adamiak et al., 2005; Sakumoto et al., 2010).

Insulin can influence female fertility both positively and negatively depending upon the reproductive stage and nutritional condition of the female (Gong et al., 2002; Grazul-Bilska et al., 2012) and it is useful for continuation of cyclicity in lean cows but detrimental for already over-conditioned animals in damaging the oocyte quality (Freret et al., 2006).

Although, negative effects of obesity and hyperinsulinemia, on oocyte quality and development of embryos in both human and ruminant, have been reported (Ashworth et al., 2009; Grazul-Bilska et al., 2012; Robker, 2008) but there is need to further investigate the potential pathological mechanisms at the molecular level underlying hyperinsulinemia.

In our current study, we have investigated the potential effects of hyperinsulinemia on DNA methylation patterns of bovine embryos development.

1.7 Impact of LPS on bEEC

Uterine microbial diseases are common cause of decreased fertility among half of all dairy cattle after calving that negatively influence the ovarian functions, uterine and animal health (Lincke et al., 2007; Sheldon et al., 2009). Lipopolysaccharide (LPS), being the component of outer membrane of gram-negative bacteria *E. coli,* is detected by Toll Like Receptor 4 (TLR4) and this interaction leads to activation of signal transduction that subsequently causes the production of cytokines, chemokines and antimicrobial peptides to communicate with the immune system for combating pathogenic microbes (Sheldon IM et al., 2009).

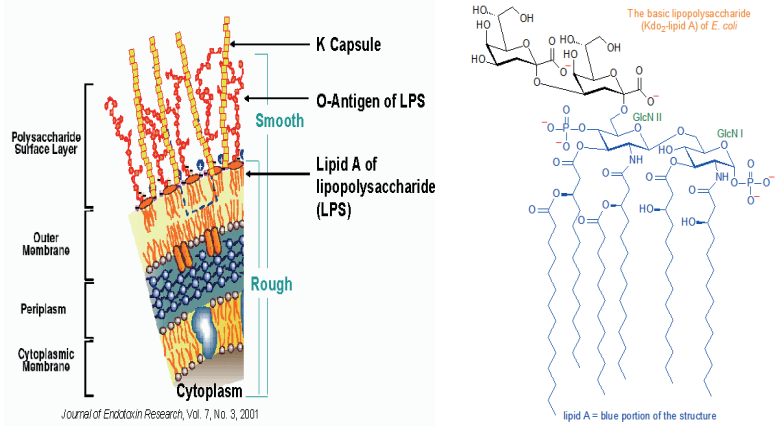

Figure 4. The major components of LPS; *a)* o antigen *b)* core *c)* lipid and the figure is adapted from *Journal of Endotoxin Research*, Vol. 7, No.3, 2001.

Specific strains of endometrial pathogenic *E. coli* (EnPEC) generate prostaglandin E2 and interleukin-8 (IL-8) in the endometrial epithelial cells and stromal cells of cattle and mice that accumulates neutrophils and macrophages in the endometrium causing endometritis (Sheldon et al., 2010). Infections of the endometrium by gram-negative bacteria induces the activation of pro-inflammatory pathways that have the potential to negatively affect fertility and more specifically the implantation process (Gilbert, 2011; Park et al., 2016). Gram negative bacteria are directly or indirectly involved in development of endometritis by inducing inflammation of the endometrium and uncontrolled inflammation is likely to be a major cause for implantation failure in cattle (Piras et al., 2017a) and human (Agrawal et al., 2013); and persistent inflammation may develop endometrial cancer (Modugno et al., 2005).

In our current study, we have investigated the effects of LPS on bEEC *in vitro* from whole transcriptomic information with focus on genes and pathways involved in the regulation of implantation by using high-throughput RNA sequencing (RNA-seq) and the potential effect of DNA methylation on gene expression by using Reduced Representation Bisulfite Sequencing (RRBS).

2 Objective of thesis

The overall aim of the studies described in this thesis was to investigate *in vitro,* the effect of DNA methylation on gene regulation of bovine embryo in response to high doses of insulin during oocyte maturation; the effect of mimic infection due to *E. coli* LPS treatment in bEEC on *i)* gene expression, *ii)* DNA methylation by identifying the regulatory regions correlated with DEGs, and *iii)* the classification of transcription factor binding sites (TFBSs).

The specific objectives were:

- To investigate the potential effect of hyperinsulinemia on gene expression and DNA methylation of bovine embryo during oocyte maturation (***Paper I***).

- To identify differentially expressed genes (DEGs) in bovine endometrial epithelial cells (bEEC) in response to LPS, specifically their implications in embryo maternal interactions (***Paper II***).

- To identify differentially methylated regions (DMRs) in the genome of bovine endometrial epithelial cells (bEEC) in response to LPS and to investigate regulatory regions (DMRs) that are correlated with DEGs (***Paper III***).

- To identify and classify the transcription factor binding sites (TFBSs) in the genome of bovine endometrial epithelial cells (bEEC) in response to LPS (***Paper IV***).

3 Summary of current research

3.1 Paper I

DNA methylation pattern of bovine blastocysts associated to hyperinsulinemia *in vitro*

3.1.1 Materials and methods

- Experimental design

The study was designed to investigate the dose-dependent effect of insulin treatment during bovine *in vitro* oocyte maturation at the epigenetic and transcriptomics level. Differential DNA methylation and changes in gene expression patterns, was examined using two different insulin concentrations, 0.1 µg/ml (INS0.1) or 10 µg/ml (INS10) on oocytes. Samples were drawn at Day 8 blastocysts originated from control and insulin treated oocytes using four replicates per treatment group. The gene expression data of the insulin-treated groups were compared to a control without insulin supplementation (INS0).

- Transcriptome and DNA methylation patterns Analysis

Gene expression studies were performed using the EmbryoGENE platform according to standardized methods previously described in general (Robert et al., 2011) and specifically for this study (Laskowski et al., 2017b) . The potential effect of insulin treatment on DNA methylation patterns of embryos during *in vitro* oocyte maturation was analysed using the bovine EmbryoGENE DNA Methylation Array (EDMA) (Salilew-Wondim et al., 2015; Shojaei Saadi et al., 2014). After Loess and quantile inter-array scale normalization, a linear model for microarray data (LIMMA) was fitted to identify differentially methylated probes between the control and the insulin treated groups considering significant with p-value <0.05 and the log2 fold-change ≥ 0.5 or absolute fold change ≥ 1.5.

3.1.2 Main findings

The comparison of insulin-treated samples INS0.1 and INS10 with controls INS0 revealed that 13,658 and 12,418 probes, respectively, were differentially methylated regions (DMRs) with p-value < 0.05 and the log2 fold-change ≥ 0.5 or absolute fold change ≥1.5. Out of these total DMRs, an overlap of 3,233 probes (DMRs) was found for both insulin groups and among these overlaps, 1,381 probes were hypomethylated and 1,852 probes were hypermethylated.

Overall, a relative hyper-methylated patterns of probes was commonly observed in both groups. The deviation of methylation patterns was also observed between both insulin groups regarding introns and non-CpG islands. For the comparison of 0.1 µg/ml (INS0.1) insulin treatment, more hypermethylation was seen on exonic regions and CpG islands and for treated samples with 10 µg/ml (INS10) insulin, more hypermethylation was seen in all gene regions including promoters.

For the comparison of epigenetic and transcriptomic analyses, the data were selected for genes related to metabolism, oxidative stress response, proliferation and mitochondrial functions. In addition, the genes related to epigenetic regulation such as DNA-methyl-transferases (*DNMT*) and Chromodomain Helicase DNA Binding Protein 4 (*CDH4*), were also scrutinized and significant differences were observed in the methylation patterns and/or mRNA-expression or both.

The differential CpG hypo-methylation pattern was complex at the *IGF2R* locus which upholds its imprinted status and these probes cannot distinguish between the two alleles and thus the real effect on differential DNA methylation may be undervalued. At the *IGF2R* locus, six probes were significantly hypo-methylated and four probes were significantly hyper-methylated. Furthermore, probes spanning at the 5' flanking region of *IGF2R* including the promoter were found close to be significant hypo-methylated for both insulin treatment groups that support the increased mRNA expression of *IGF2R* in response to insulin treatment. Whereas, probe 28, covering the imprint control element (ICE) was also observed differentially hypo-methylated.

Further analysis of differentially expressed candidate genes, related to metabolic functions, revealed that many of the previously described genes having a significant upregulation also had modifications in the CpG methylation pattern. Thirteen out of the 24 described genes were upregulated when the promoter

region was hypo-methylated which showed a classic methylation pattern in relation to the transcriptome. Furthermore, the genes involved in epigenetic regulation such as DNA-methyltransferases (*DNMTs*) and chromodomain helicase DNA binding protein 4 *(CHD4)* and histone variants *(HIST1H1C)* showed an aberrant DNA methylation and gene expression patterns.

3.2 Paper II

Differential gene expression in bovine endometrial epithelial cells after challenge with Lipopolysaccharide (LPS); specific implications for embryo maternal interactions

3.2.1 Materials and methods

- Isolation of bovine endometrial epithelial cells (bEEC)

The technique for isolation and cell culture was adapted from Charpigny et al (Charpigny et al., 1999). Briefly, the uterus of three Swedish Red Breed (SRB) cows from the local slaughter house were transferred to the laboratory on ice. The bEEC preparation was performed within one hour and uterine horns were cut longitudinally. The endometrial tissue from the middle part of each uterine horn were separated from the myometrium and cut into 1 mm^3 pieces and transferred to enzymatic dissociation solution. The microscopic observation revealed that no fibroblasts were found after passage 3 to passage 5 and flow-cytometry showed that more than 98% of cells expressed cytokeratin, reflecting high purity of the epithelial cell culture system (Nongbua et al., 2018; Piras et al., 2017b).

- LPS challenge and Isolation of total RNA

Following in vitro culture, on passage 5, bEEC samples from three cows were either mock-challenged (0) or challenged with either 2 or 8 µg/mL LPS from a virulent strains of *E. coli* (serotype O111:B4). The time of LPS challenge was taken as 0 hours in the experiment. At time 0 hours and 24 hours later; cells were detached, washed and total RNA was extracted. The extracted RNA samples at time 0 hours without LPS dosage and 24 hours without LPS (exposed to 0 µg/ml) were considered as "control samples" whereas extracted RNA samples after 24 hours of LPS challenge with 2 µg/ml and 8 µg/ml were considered as "treated samples".

- Library preparation and sequencing

RNA sequencing (RNA-Seq) libraries for 4 samples from each cow (on Time 0, and with 0, 2, and 8 µg/mL LPS 24h) were prepared by the sequencing platform at Science for Life Laboratory, Uppsala University. In total, there were 12 samples from three cows. RNA-Seq was performed using the Illumina HiSeq2500 system. Paired-end reads of 125bp length were generated having average sequence depth of 28.33 million reads per sample with SD of 2.03.

Table 1. *Number of raw reads in millions (10^6) generated from RNA sequencing using the Illumina HiSeq2500 system for each sample.*

Time points	Treatment	Cow 1	Cow 2	Cow 3
0 hours	Without LPS	28.43	29.99	26.11
24 hours	Without LPS	27.58	30.08	25.30
24 hours	2 µg/mL	28.16	29.19	24.97
24 hours	8 µg/mL	29.82	28.78	31.54

- Data quality control

We used FastQC to check the quality control of the RNA-Seq data generated from control (control time 0, N = 3 and control 24h, N = 3) and LPS-treated samples (2 µg/mL, N = 3 and 8 µg/mL, N = 3). We used Trimmomatic 0.32 to remove adapter sequences and filter low quality reads (quality score < 30) (Bolger et al., 2014).

- Read alignment to the *Bos Taurus* reference genome sequence

We used the STAR 2.4.0 software (Dobin et al., 2013) for alignment of our RNA sequence reads to the *Bos Taurus* reference genome sequence bosTau6. After trimming and adaptor removal, approximately 97% of the reads were mapped to the reference genome whereas around 90% were uniquely mapped with an average of 26.35 million and SD of 1.96 million per sample.

- Gene expression quantification

We used HTSeq 0.6.1 (Anders et al., 2015) for the htseq-count scripting to count for each gene to determine how many aligned reads overlap to specific exons. Uniquely mapped reads in SAM/BAM format and ENSEMBLv78 annotations file in GTF format were used as an input in the script. This script was designed specifically for differential expression analysis and it counted only those reads which were mapped unambiguously to a single gene whereas reads aligned to multiple positions or overlapping with more than one gene were discarded.

- Identification of differentially expressed genes (DEGs)

DESeq2 version 1.6.3 (Love et al., 2014) was used to the identify of DEGs using pairwise comparisons *i*) between control at time 0 and control at 24h, *ii*) between control at 24h *vs* 2 µg/mL LPS *iii*) between control at 24h *vs* 8 µg/mL LPS *iv*) between 2 µg/mL and 8 µg/mL LPS-treated samples at 24h, and *v*) between control samples at 24h and LPS-treated samples at 24h (2 and 8 µg/mL LPS taken together). The resulting *p*-values were adjusted for multiple testing using the Benjamini–Hochberg procedure. DEGs with adjusted p-values < 0.05 were regarded as statistically significant. MA-plots were used to visualize the differential gene expression between two samples.

- Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment were performed using DAVID 6.7 (Huang et al., 2009) with Ensembl genes as background. GO analysis used hypergeometric distribution to provide information related to three ontologies: biological process (BP), cellular component (CC) and molecular function (MF). Furthermore, classification analysis of genes that could possibly be involved in disease and physiological functions was performed using data available in the GeneCards database (http://www.genecards.org/).

- RT^2-qPCR validation and changes in gene expression with time

Technical as well as biological validations of the differential gene expression obtained by RNA-Seq data were performed by Real Time-quantitative PCR (RT^2-qPCR). 19 DEGs from the RNA sequencing data were selected for a commercial custom-made array (Control 24 hours and 24 hours after 2 µg/mL

LPS). Another set of samples was taken from the cell culture of 3 additional cows with 2 µg/mL LPS treatment at three different time points. Cells from these cows were exposed to 2 µg/mL LPS for 6, 24 and 48 hours and comparisons were performed with controls obtained at the same time. 19 DEGs (over- or under-expressed) from RNA-seq were tested for RT^2-qPCR validation and all the tested genes with adjusted p-value < 0.05 were confirmed as differentially expressed by the RT^2-qPCR validation.

3.2.2 Main findings

- Overall differential gene expression analysis

The overall RNA-Seq analysis revealed that a large number of genes has altered their gene expression with LPS exposure to bEEC. However, 2035 and 2073 differentially expressed genes (DEGs) were identified between control samples at time 24 hours and treated samples with 2 and 8 µg/mL LPS respectively. Moreover, 1748 DEGs were common between these two treatment groups. But when comparison was done between two treated samples with 2 µg/mL LPS and 8 µg/mL LPS, no DEG was found. The comparison of combined LPS treatment samples (both groups of 2 and 8 µg/mL LPS are taken together) and control samples (at time 24 hours) revealed 3032 DEGs. In almost all cases of response to LPS, genes showed similar variation of expression levels. The number over-expressed genes was higher as compared to under-expressed genes. Out of 2035 DEGs, 1066 (52.4%) and 969 (47.6%) were over-expressed and under-expressed, respectively. In general, higher fold changes were observed for the over-expressed genes as compared to that of under-expressed genes.

Due to large number of common DEGs detected among the treatment groups and high degree of similarity between transcriptomic profiles (r2 = 0.99) observed in 2 and 8 µg/mL LPS, the comprehensive analysis of this study was performed mainly on the comparison between control samples at time 24 hours *vs.* treated samples 2 µg/mL LPS.

The principal component analysis (PCA) of the top 500 genes revealed common response of LPS groups when they were compared to controls and displayed variability across samples. Although, cow 2 showed variability in comparison to cows 1 and 3 along the first principal component but the three cows were aligned along the second principal component.

- Gene ontology analysis of the DEGs

Gene ontology analysis of the DEGs were categorized into 484 functional groups with a corresponding adjusted p value < 0.001, out of which 274 were over-expressed and 210 were under-expressed groups. Of the over-expressed groups, 214, 15 and 45 were categorized into biological process (BP), cellular component (CC) and molecular function (MF), respectively. For the under-expressed groups, 134, 52 and 24 were categorized into BP, CC, and MF, respectively. Gene ontology analysis showed that four out of the top 10 significantly overrepresented pathways were associated to immune response, inflammatory response and antigen processing and presentation. External stimuli, transcription, catalytic activity, response to stress and glycolysis were also among the overrepresented. The top five underrepresented GO-terms are related to cell structures, cytoskeleton, cell membrane, binding and organelle.

- KEGG pathway analysis of the DEGs

KEGG pathway analysis revealed 22 significant enriched pathways (adjusted *p*-values < 0.05), out of which 11 were overrepresented and 11 were underrepresented pathways. As expected, the overrepresented pathways were related to inflammation (NOD)-like receptor signaling pathway (16 DEGs) enriched, Toll-like receptor signaling pathway (19 DEGs), cytokine-cytokine receptor interaction (25 DEGs), antigen processing and presentation (12 DEGs), chemokine signaling pathway (23 DEGs), and apoptosis (19 DEGs). Whereas, three underrepresented pathways were found to be related to focal adhesion, regulation of actin cytoskeleton and adherent junction.

- Downstream analysis of DEGs

According to GeneCards database (http://www.genecards.org), the identified DEGs encode protein involved in acute inflammation (410 DEGs), innate immunity (441 DEGs), immune tolerance in pregnancy (120 DEGs), allergy (153 DEGs), and cell adhesion (626 DEGs). Twenty-four common DEGs (*EDN1, TGFB2, TGFB3, FAS, C3, ICAM1, CXCL8, IL1RN, TP53, TNF, IL1A, CAT, F5, NFKB1, IL1B, PTGS1, CDKN2A, IL1R1, CCL2, ADA, IL6, CSF2, MMP1*, and *MMP9*) encode proteins involved in all of the above biological processes. In addition, a large number of genes were also classified for cell skeleton (466 DEGs), cell proliferation (880 DEGs), cell apoptosis (760 DEGs) and signal transduction (755 DEGs).

Nineteen DEGs have log2 fold change > 3 in the list of total over-expressed DEGs and out of these 19, seven DEGs (*CXCL6, BCL2A1, LGALS9, C3, BIRC3, CFB,* and *CD40*) were related with inflammation and 9 DEGs (*CXCL6, BCL2A1, LGALS9, C3, BIRC3, CFB, CD40, CTSC,* and *TCN1*) were associated with infection. Several BoLA (bovine leukocyte antigen) genes were shown to over-expressed in response to LPS treatment.

- Genes related to implantation and maternal response to the embryo

In our study, several DEGs encode protein with functions related to embryo maternal interactions and implantation in response to LPS treatment. These have been involved in cell structure, calcium metabolism and membrane properties, cell adhesion, and enzymes such as Matrix Metallo Peptidase (MMPs) regulating tissue remodelling. Four members of the cadherin superfamily (*ITGB6, CDH26, ITGAV,* and *CELSR1*) and MMPs family (*MMP1, MMP7, MMP9,* and *MMP13*) were over-expressed whereas their inhibitor TIMP3 was found to be under-expressed after LPS-treatment. Many genes of the mucin family (*MUC1, MUC4, MUC13, MUC16, MUC20,* and *F1MUC1*) were found to be over-expressed. Mucins are proteins that elicit gel-like secretions for lubrication and cell signalling. Their role in vertebrates bone formation is also reported and they bind to pathogens as part of an active immune system. Four transcripts of the integrin family (*ITGB3, ITGB4, ITGB5*, and *ITGA7*) and most of transcripts coding for cell adhesion molecules (*CTNNA3, CTNNAL1, CDH2, PCDH7, CT, PKP1,* and *PKP4*) were found to be under-expressed after LPS-treatment. A well-balanced combination of adhesion molecules (like cadherins, integrins and MMPs) is essential for regulation of tissue remodelling that plays an vital role in successful embryo implantation (Aplin, 1997; Lessey, 2002).

Several interferon-τ (IFNT)-induced genes (*IFIT1, IFIT2, IFIT3, IFIT5, IFITM2, IFITM3, PARP12, ZNFX1, HERC6, RNF213, CXCR7, DDX58, PLAC8, RSAD2,* and *STAT1*) were also found over-expressed after LPS-treatment. Among other significant genes, leukemia inhibitor factor (LIF) was found over-expressed that has crucial role in successful implantation and Gal-9 (LGALS-9) was over-expressed which is involved in infection response whereas Gal-1 (LGALS-1) and LGALS-3 were found to be under-expressed.

3.3 Paper III

LPS-treatment of bovine endometrial epithelial cells causes differential DNA methylation of genes associated with endometrial function and inflammation

3.3.1 Materials and methods

- Isolation of bovine endometrial epithelial cells (bEEC)

The uterine horns from three Swedish Red Breed (SRB) cows were collected from Lövsta slaughterhouse. Endometrial epithelial cells were purified according to procedures previously described in *Paper II* and in detail Charpigny et al (Charpigny et al., 1999; Piras et al., 2017a).

- LPS challenge and genomic DNA isolation

At passage 5 *in vitro* culture bEEC samples from three cows were challenged with 0 or 2 or 8 µg/ml LPS with the same time points as in our previous study *Paper II*. Genomic DNA was extracted from samples at time 0 hours and 24 hours without LPS for control samples whereas genomic DNA was extracted for treated samples with 2 µg/ml and 8 µg/ml LPS at time 24h.

- Library preparation and sequencing processes

Reduced Representation Bisulfite Sequencing (RRBS) method was used for the detection and quantification of DNA methylation at single base resolution. Following genomic DNA extraction of 4 samples from each cow (on Time 0 without treatment, and after 24h with 0, 2, and 8 µg/mL LPS), libraries for RRBS were prepared at The Babraham Institute, UK. In total, there were 12 samples from three cows. DNA libraries for sequencing were generated by *Msp*I digestion followed by end-repair/A-tailing and 5mC adaptor ligation, and bisulfite conversion plus PCR. RRBS was then performed by using Illumina HiSeq2500-RapidRun at The Babraham Institute, UK. Illumina sequencing platform generated an output of 17-21 million 50bp single-end raw reads per sample in the compressed fastq format containing phred quality scores.

- Data quality control

The quality of the RRBS data sets generated from control samples (control time 0, N = 3 and control 24h, N = 3) and LPS-treated samples (2 µg/mL, N = 3 and 8 µg/mL, N = 3) were analysed by using FastQC version 0.11.5. Then adapter contaminated sequences and low-quality reads (quality score < 30) were filtered by using TrimGalore version 0.4.1(Krueger F: Trim Galore! http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). TrimGalore used automatically Illumina standard adapters by default and removed 2 additional bases containing a cytosine which was artificially introduced in the end-repair step during the RRBS libraries preparations.

- Reads alignment to the *Bos Taurus* reference genome sequence

After removing adapter contamination and performing quality control, we used BS-Seeker2 (Guo et al., 2013) for the alignment of quality reads with the reference genome bosTau8 (Bos_taurus_UMD3.1) that produced an average of 42% uniquely mapped reads per sample.

- Calculation of differential DNA methylation and DMRs

R package methylKit v0.9.5 (Akalin et al., 2012) was used to define differential DNA methylation and to identify DMRs. The cytosines having the coverage of at least five reads were considered for analysis. Logistic regression test was used to determine which one regions are differentially methylated between the samples. The genome was tiled for 100bp and statistically significant DMRs of 100bp were calculated based on *i)* at least two CpG sites in the region, *ii)* methylation difference > 10%, *iii)* and adjusted p-values (q-value) < 0.05. DMRs were identified by taking comparisons between samples *i*) control on time 0 vs control 24h, *ii*) control 24h vs LPS-treated 2 µg/ml, *iii*) control 24h vs LPS-treated 8 µg/ml, *iv*) control 24h vs LPS-treated combined. For annotation of genic regions, we used Ensembl genome browser database whereas UCSC genome browser database was used for CpG islands and repetitive element annotations.

- Gene Ontology over-representation and pathways analyses

The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used for Gene Ontology (GO) analysis of DMR-associated genes with all

annotated genes in Bos taurus genome as background. KEGG database available within DAVID platform (Huang et al., 2009), and with WikiPathways database (https://www.wikipathways.org/) were used for pathway enrichment analysis.

3.3.2 Main findings

- Quality control and alignment

RRBS of bEEC generated 17 to 21 million raw reads per sample. After performing quality control and removing adapter contamination, 60-62% of the raw reads were successfully aligned to the reference genome sequence (bosTau8), whereas an average of 42% reads per sample were uniquely mapped.

- DNA methylation profiles and identification of DMRs

For DNA methylation profiles on uniquely mapped reads, we found 2.1 to 2.3 million CpG sites per sample, out of which 1.93 million CpG sites were common in all samples, that covered 7.1% of the total CpG sites. When genome was tilled for 100bp regions on the basis of at least one CpG site and read coverage >=5, we identified 700,323 regions. Among these regions, 157,202 regions were scrutinized, having at least two CpG sites for differential methylation analysis. Analysis of RRBS data revealed 822 significant DMRs (q-value < 0.05) between control samples at 0h *vs*. 24 h. The analysis further revealed 511 significant DMRs between control 24h and samples treated with LPS at low concentration (2 µg/ml), whereas 469 significant DMRs were identified in comparison between control 24h and high concentration (2 µg/ml samples. Only 69 DMRs were found common between 2 µg/ml and 8 µg/ml LPS groups that may be due to coverage threshold in these groups. Therefore, we performed a combined comparison of 2 and 8 µg/ml LPS groups with 24hr control samples that revealed 803 significant DMRs having 423 DMRs common between both LPS group. Then we merged 803 DMRs of combined group with those (488) DMRs of individual groups which were not found in combined group. Hence, an overall aggregate of 1291 significant DMRs was finalized for further analysis, and out of which 707(55%) and 584(45%) DMRs were hypo-methylated and hyper-methylated respectively.

- Genomic and CGI distribution of DMRs

In order to assess the functional significance of the methylation differences arising due to LPS treatment, the distribution of these DMRs in the genome was examined. To define the genomic distribution of the identified 1291 DMRs, their positions in the bovine reference genome sequence were investigated in relation to annotated bovine protein-coding genes, their established promoters and CpG islands. This revealed that almost 46% of the DMRs were overlapping with CpG islands (CGIs), 31% with CGI shores, and 23% with other genomic regions. Furthermore, approximately 47.5% (n=613) of the DMRs were found to be associated with gene body (exons and introns), 47.6% (n=615) were intergenic and 4.8% (n=63) mapped to the promoter regions (-2kb upstream of TSS). Furthermore, out of 600 differentially methylated genes (DMGs), 589 protein-coding, 7 miRNA, and 4 pseudogenes were also identified. In addition, the number of DMRs were found significantly correlated with the number of genes per chromosome (R=0.45, P = 0.011) but DMRs were not significantly correlated with chromosomal size (R=0.32, P = 0.084).

- Distribution of methylated region *vs.* telomeric region on chromosome

The DNA methylation pattern of all target regions (n = 157,202) was skewed towards chromosomal ends. When telomeric region was taken as 20 kbp, we found that there was no association between the distribution of targeted regions and telomeric regions. Whereas, Fisher's Exact test (*p*-value < 1.14e-05) revealed that DMRs were significantly enriched in 2Mbp regions (taken adjacent to telomeres) compared to non-telomeric regions.

- Comparison between DNA-methylation and RNA-Seq in bEEC: (DMRs *vs.* DEGs)

We performed overall comparison of differential DNA methylation obtained using RRBS analysis with the DEGs identified by RNA-Seq analysis on the same cell samples in our *Paper II*. We observed a significant negative association of gene expression with mean methylation of promoter regions (Spearman rho = −0.41; P < 2.2e-16) and gene body (Spearman rho = −0.22; P < 2.2e-16). However, 39 DEGs had inverse relationship between their gene expression and degree of methylation having |Δmethylation| > 5% and |Δexpression log2FC| >1. Their combined functional analysis disclosed that they were related to ion/calcium ion transport and signal transduction processes.

- Gene ontology and pathway analysis

GO analysis revealed that these genes were also distributed among important biological and molecular processes related to signal transduction, cell proliferation, apoptotic process, vasculogenesis and embryo development. Furthermore, for molecular functions, DMRs were observed enriched for calcium and zinc ion binding, ATP binding and transcription coactivator activity.

KEGG pathway analysis revealed significantly enriched pathways including "Calcium Signaling Pathway", "MAPK Signaling Pathway", and "Oxytocin Signaling Pathway". However, pathway analysis with WikiPathways database revealed enrichment of additional pathways including "Wnt Signaling", "Sudden Infant Death Syndrome (SIDS) susceptibility", "Iron metabolism in placenta", "Myometrial relaxation and contraction", and "IL-2 Signaling".

3.4 Paper IV

Genome-wide identification of transcription factor binding sites in LPS-associated differentially methylated regions in bovine endometrial epithelial cells

3.4.1 Materials and methods

- Isolation of bovine endometrial epithelial cells (bEEC)

The technique for isolation and cell culture was generally explained in our study *Paper III* and specifically in Charpigny et al (Charpigny et al., 1999; Piras et al., 2017a).

- LPS challenge and genomic DNA isolation

LPS treatment and extraction of genomic DNA from bEEC samples *in vitro* culture was performed with the same time points as explained in our previous study *Paper III.*

- Library preparation and sequencing processes

DNA libraries for RRBS were prepared at The Babraham Institute, UK for 12 samples from three cows as explained in our study *Paper III*.

- Data quality control

The quality of the RRBS data sets was analysed by using FastQC and maintained by using TrimGalore on the same samples as discussed in our *Paper III*.

- Read alignment to the *Bos Taurus* reference genome sequence

BS-Seeker2 was used for the alignment of quality reads with the reference genome bosTau8 (Bos_taurus_UMD3.1) as discussed in our *Paper III*.

- Identification of Transcription Factor Binding Sites

Our previous study *Paper III* revealed overall 1291 differentially methylated regions (DMRs) which were used for identification and classification of transcription factor binding sites by using JASPAR database (Khan et al., 2018) through oPOSSUM application programming interface (API) (Kwon et al., 2012).

3.4.2 Main findings

TFBS Enrichment analysis by using oPOSSUM revealed overall 73 significantly over-represented motifs within the 1291 DMRs identified in our previous study *Paper III*. We found that among the overrepresented motifs include TFBS for five members of the signal transducer and activator of transcription (STAT) family (STAT1, STAT2, STAT3, STAT5a/b, STAT6), three members of interferon regulatory factor family (IRF1, IRF2 and IRF7), early growth response 3 (EGR3), activating enhancer-binding protein 2 (TFAP2A and TFAP2C), aryl hydrocarbon receptor nuclear translocator (ARNT), and neuroblastoma MYC oncogene (MYCN). However, seven motifs were found significantly enriched in both hyper-methylated and hypo-methylated regions (AHR::ARNT, ARNT::HIF-1α, NRF1, HES1, HEY1, FOXB1, and TFE3).

On classification of over-represented motifs for transcription factors, we found that a large number of overrepresented motifs were the members of basic helix-

loop-helix factors (bHLH), basic leucine zipper factors (bZIP), and homeo-domain factors classes.

The enrichment analysis of TFBS in the promoter regions (1 kb upstream) of the differentially expressed genes (DEGs) previously reported in our *Paper II,* revealed that 145 motifs were enriched, which indicates that corresponding factors are relevant for the functional changes during infection.

Five significantly enriched transcription factors with most hits (hits >20% of total DMRs sequences) were selected for co-localization with other TFBS consensus sequence motifs. This analysis showed that motifs of hypermethylated regions Arnt::HIF-1α, PAX2, NRF1, HLTF and MEIS1 were found to be co-localized, and motifs of hypomethylated regions Ahr::Arnt, Arnt::HIF-1α, NRF1, TCFL5, and HES1 were found to be co-localized.

4 General discussion

Bioinformatics is the application of computer science, applied mathematics and statistics to answer questions regarding complex biological systems using different omics-based technologies in molecular biology, genomics, and epigenetics. Biological processes are essential for living organisms to survive such as metabolism, cell differentiation, reproduction etc. In particular, the biological processes controlling reproduction in mammals including modern dairy cattle are complicated. Disorder of metabolism due to higher levels of insulin during pregnancy and disturbance of immune response due to bacterial infections after calving are considered very detrimental in reproduction. The determinants involved in controlling these molecular mechanisms are not fully understood. The results reported in the current thesis provide additional information regarding these complex biological systems.

4.1 Paper I

Previous studies from our group have shown the detrimental effect of elevated insulin concentrations on development of blastocysts during oocyte maturation such as decreased blastocyst rates (D. Laskowski et al., 2016), morphological changes (Laskowski et al., 2017a), and disturbed gene expression especially in genes related to lipid metabolism, proliferation, mitochondrial functions, and oxidative stress (Laskowski et al., 2017b).

In the present study, we used the EDMA platform to define the potential molecular effects of insulin treatment on embryonic development during *in vitro* oocyte maturation. Differential gene expression and DNA methylation patterns were analysed using the EmbryoGENE DNA Methylation Array (EDMA). A stronger differential DNA methylation effect was observed when the higher concentration of insulin was employed. When using the highly concentrated INS10 group, a higher number of hyper-methylated probes were observed that advocates the difference of dose-dependent methylation patterns. However, an overlap of DMRs exists that showed the methylation pattern in the same direction in both insulin treatments. Thus, the two treatments can serve as biological replicates.

We observed different relationships between transcriptome and DNA methylation pattern such as a classic one (downregulation *vs* hypermethylation and vice versa) and a less typical one (upregulation vs hypermethylation). Conservation odds of methylation were found to be higher in exons, promoters,

enhancers, and CpG islands. An inverse correlation between methylation pattern and transcriptome is well established for promoter regions of genes (Weber et al., 2007) and DNA methylation in the first exon of protein-coding genes has been reported to be correlated with silencing of transcription but might have different effect in introns (Brenet et al., 2011; Jones, 2012).

Insulin can be mitogenic through its own receptors and its mitogenic functions are reflected in our study for the activation of genes associated to growth and proliferation such as Insulin-like growth factor 2 receptor (*IGF2R*), DEAH-Box Helicase 38 (*DHX38*), lamin A (*LMNA*), mitogen-activated protein kinase kinase 2 (*MAP2K2*), protein tyrosine phosphatase non-receptor type 5 (*PTPN5*), and tubulin beta-3 chain (*TUBB3*). Previous studies have reported that the expression level of *IGF2R* in the embryo varies due to maternal metabolic conditions (Arias-Álvarez et al., 2013). Insulin binds to IGF2R receptor and its maternal expression eradicates increased amount of insulin (O'Dell and Day, 1998). In our study, the upregulation of *IGF2R* mRNA expression in the insulin-treated groups advocates a mechanism to remove the excessive amount of insulin molecules.

The maternal expression of imprinted gene I*GF2R* is controlled by an imprint control element (ICE) known as Region2 that overlaps the promoter of the noncoding AirRNA gene (*AIRn*) (Zwart et al., 2001). The hypomethylation at the promoter region of *IGF2R* validates our finding of increased *IGF2R* mRNA expression whereas the observed complex pattern of DNA methylation, at the imprinted *IGF2R* locus in response to insulin treatment, was expected in the array-based method used because of the fact that the probes cannot distinguish between two alleles. To further define the exact results of insulin-treated bovine embryos at the maternal or paternal alleles, we need further differential RNA sequencing studies that can scrutinize the paternal and maternal genetic variants along with whole genome bisulfite sequencing (WGBS) for epigenetic studies.

DNA methyltransferases (DNMTs) are highly conserved and considered as writers of the epigenome and responsible for maintenance of CpG methylation patterns through mitosis by binding to hemi-methylated DNA at CpG sites or *de novo* methylation of DNA, which occurs after embryo implantation (Okano et al., 1999; Uysal et al., 2015). Significantly higher expression level of *DNMT3B* was observed that influenced the epigenetic modifications which were reported for embryonic development, X-chromosome inactivation and imprinting through *de novo* methylation (Walton et al., 2014).

4.2 Paper II

The observed over-expression of genes encoding cytokines and chemokines in LPS-treated bEEC *in vitro* model supports the notion that such effects are occurring *in vivo* during conditions such as inflammation and could cause migration of immune effector cells to the inflamed tissues and ultimately the persistent presence of immune cells in the endometrium that possibly impairs implantation (Dekel et al., 2010). A prominent example that we identified as being over-expressed in response to LPS is Tumor Necrosis Factor alpha (*TNF-α*), which has well-known pleiotropic effects on inflammation, innate immunity and cell growth in the endometrium and its high concentrations has been reported to be the source of implantation failures and pregnancy loss (Torchinsky et al., 2005). Other examples that we observed are the chemokines CXCL1 and CXCL6 that both lead to the recruitment of neutrophils and are linked to inflammation and apoptosis pathways (Chittur et al., 2011). From this *in vitro* model, although cells were exposed to low dosages of LPS for a short period of time, we observed significant changes in gene expression of genes encoding cytokines and chemokines.

Another prominent example of DEGs was the finding of several over-expressed BoLA (bovine leukocyte antigen) genes. This was an expected finding as BoLA genes encode Major Histocompatibility Complex (MHC) antigen presenting molecules that have essential functions as immune response activators for the acquired immune system and bind antigens derived from pathogens and present these antigens to T cell receptors expressed on T lymphocytes. During a normal pregnancy expression of both MHC class I and II genes needs to be suppressed preventing premature abortion of the fetus. A further example is that four members of the cadherin superfamily are among the over-expressed DEGs which encode proteins that are transmembrane glycoproteins that mediate cell-cell interactions through calcium binding and has role in cell adhesion to bind cell with each other and possible changes could impair implantation (Lessey, 2002).

MMPs enzymes are involved in immune response and fusion of muscle cells during embryo development by remodelling of the extracellular matrix (ECM) (Tency et al., 2012), and down-regulation of *MMP2* and *TIMP3* in the endometrium is associated with implantation failure (Konac et al., 2009), our obtained data supports that down-regulation of *TIMP3* may be detrimental for implantation.

The interferon-τ (IFNT)-induced genes encode proteins that belong to a group of signalling proteins with crucial functions in response to pathogen challenges. Developing embryo during pregnancy activates STAT signal transduction pathway through secretion of IFNT having crucial role in implantation (Binelli et al., 2001; Maj and Chelmonska-Soyta, 2007). The up-regulation of IFNT-induced genes increases our understanding for biological consequences during endometrial inflammation *in vitro* LPS-bEEC model. The activation of leukemia inhibitor factor (LIF)-dependent STAT signal transduction pathway is crucial for embryo implantation and its over-expression is reported in mouse uterus during receptivity phase (Stewart et al., 1992) whereas its under-expression causes implantation failure (Salleh and Giribabu, 2014). In our current study, exposure of LPS to bEEC for 6 hours increased the expression levels of both *LIF* and *STAT1* but their differential expressions were gradually decreasing after time 48 hours. The duration of these changes demands further investigation for understanding of response to a living embryo at early pregnancy stage.

LPS-treatment also significantly de-regulated the galectin family genes. Strictly controlled expression levels of galectin family genes are crucial during pregnancy and especially for early embryo–maternal interactions. During normal pregnancy, over-expression of *Gal-1* leads towards immune tolerance and is vital in human preimplantation embryos and its expression becomes significantly increased during implantation (Popovici et al., 2005). Low expression of *Gal-1* in the endometrium has been connected with early pregnancy loss and miscarriages (Barrientos et al., 2014). Studies reported the presence of *Gal-1* in the bovine endometrium (in lamina propria) (Froehlich et al., 2012), but its exact role to regulate BoLA gene expression during pregnancy is unknown in cattle. Due to the unknown implication and significance of these changes for the successful implantation / pregnancy in cattle, specific functional studies would be beneficial.

4.3 Paper III

We used RRBS for identification of differentially methylated regions (DMRs) in bovine endometrial epithelial cells (bEECs) in response to LPS. Furthermore, the obtained DMRs were scrutinized in relation to DEGs from our study *Paper II*. Previously it has been reported that the endometrium is dynamic tissue whose methylome changes throughout the reproductive life cycle such as menstrual cycle (Houshdaran et al., 2014; Saare et al., 2016), implantation and even its extensive remodelling is required after calving (Sheldon et al., 2008). To our

knowledge, no study has been performed using bEEC to investigate the DNA methylation changes taking place while uterine infection is being established. We found a global DNA hypomethylation trend on comparison of control *vs* LPS treated samples, which indicates that LPS induces hypomethylation in bEECs. DNA methylation changes were observed enriched in sub-telomeric regions, which approves that mammalian telomeres are rich in repetitive regions but lack of CpG islands, whereas adjacent regions to telomere are rich in CpG islands (Blasco, 2007; Gadalla et al., 2012). More hypomethylated genes were observed in X chromosome as compared to other chromosomes. Enrichment of DMRs in the X chromosome of uterine leiomyoma has been reported (Maekawa et al., 2011).

A large number of DMRs were mapped to genes involved in the controlling the endometrial functions or were related to endometrial dysfunction and infertility. Furthermore, gene ontology and pathway analyses of these genes revealed several pathways such as proliferation and differentiation, cell migration, cell adhesion and extracellular matrix remodelling and immune responses.

Class II histone deacetylases (HDACs) are signal transducers often acting as co-repressors of transcription by removing histone acetylation (Li and Yang, 2016). The function of Class II histone deacetylases (HDACs) is influencing histone modification and thereby chromatin structure. Two hypomethylated and one hypermethylated DMRs were found in HDAC4 intron. Due to its multiple targets, other functions may be changed in the endometrium. The above result associated to the proliferative phenotype previously observed in our model (Chanrot et al., 2017) indicates the specific studies to decipher the implications of the multiple changes in HDAC4 methylation for epithelial cell proliferation.

Wnt signaling pathway is among the most important pathways involved in cell proliferation and differentiation in the endometrium. Wnt7A has been reported to encode a protein that acts as a main actor for the controlling of β-catenin expression. In our LPS model, the Wnt7A gene has two hypo-methylated DMRs that may be consistent with the proliferative phenotype previously observed (Chanrot et al., 2017) in cow epithelial cells.

A strong hypomethylation of the hypoxia inducible factor 1-alpha (*HIF-1A*) was observed and has negative correlation with its strong mRNA over-expression in our study *Paper II*. *HIF-1A* is reported to be involved in physiological processes in the endometrium such as tissue repair (Maybin et al., 2018, 2011), implantation (Gou et al., 2017), peri-implantation blastocyst development (Song

et al., 2008), and its over-expression is associated with miscarriages in the human (Chen et al., 2016). Due to its vital role in the development of endometrial vascularization, pregnancy and in uterine pathology, the disturbed methylation and transcriptomic profiles of this gene under LPS would demand further studies to explain its potential role in tissue repair and persistence of inflammation following infection.

4.4 Paper IV

We performed an *in silico* annotation of the transcription factor binding site (TFBS) that were present in differentially methylated regions (DMRs) identified in our previous study *Paper III*. After enrichment analysis, a detailed *in silico* annotation of the identified consensus sequence motifs using the Jaspar database of transcription factor binding profiles was performed (Khan et al., 2018; Mathelier et al., 2016). Enrichment analysis revealed 73 enriched TFBS consensus sequence motifs that were sub-classified for their presence in hyper-methylated and hypo-methylated DMRs. Majority of these motifs were located in promoters, followed by enhancers. We next analysed the enrichment of TFBS in the promoter regions (1 kb upstream) of differentially expressed genes (DEGs) identified by RNA-Seq of the same cells in our *Paper II* . We found 145 enriched motifs and most of these were present in the promoter regions of DEGs.

Co-localization between the five most significantly enriched transcription factors and other TFBS revealed that binding motifs of transcription factors: Arnt::HIF-1α, PAX2, NRF1, HLTF and MEIS1 were found to be co-localized in hyper-methylated regions, and motifs of Ahr::Arnt, Arnt::HIF-1α, NRF1, TCFL5, and HES1 were found to be co-localized in hypo-methylated regions. Several of these transcription factors are well-known to have their transcriptional role for the genes involved in inflammatory and hypoxia-related pathways.

Further comparison of the top 10 TFBS and their relative mean expression level revealed that all the top 10 TFBS were significantly hypo-methylated and the mRNA expression of two of the genes, *HIF-1A* and *STAT1* was found to be highly over-expressed in response to LPS-treatment. Significant enrichment of STAT1, STAT2 and STAT3 consensus sequence motifs were observed in hyper-methylated DMRs. STAT proteins have been reported to perform significant biological functions during early pregnancy (Maj and Chelmonska-Soyta, 2007). STAT family members are known to be involved in LPS-induced pathways and has crucial role in embryo implantation and fertility processes. This indicates the critical functional role of STAT1 during endometrial inflammations. The gene

expression patterns altered by hypoxic conditions are mostly controlled at the transcriptional level and hypoxia-inducible transcription factor 1 alpha (HIF-1α) is among the main transcription factors responding to hypoxia. We observed that mRNA expression of *HIF-1A* was highly over-expressed whereas HIF-1α/Arnt consensus binding motifs were observed highly enriched in the hypomethylated DMRs.

A functional association between inflammation and hypoxia is reported in rheumatoid arthritis (Konisti et al., 2012). In our study, both hypoxia-inducible and inflammatory transcription factor were found critically activated and the enrichment of their consensus sequence binding motifs strongly advocates a connection between hypoxia and inflammation in endometritis.

5 Conclusions

Paper I

Insulin treatments have affected the epigenome and transcriptome of bovine embryos during oocytes maturation by inducing changes in overall DNA methylation patterns as well as on the individual gene level of candidate genes. Overall 13,658 and 12,418 probes (DMRs) were identified for insulin-treated samples INS0.1 and INS10 respectively on comparison with controls samples INS0. An overlap of 3,233 DMRs was found for both insulin groups with 1,381 hypomethylated and 1,852 hypermethylated. A relative hyper-methylated pattern was observed in the INS0.1 and comparatively much higher number of hyper-methylated probes was observed in INS10 showing the dose-dependent methylation patterns. In comparison of transcriptome with epigenome, 13 out of 24 candidate genes were upregulated while their promoter region was hypo-methylated and one gene was downregulated and its corresponding methylation pattern was hyper-methylated. Differentially expressed transcripts (DET) of genes, related to lipid metabolism, glucose metabolism and insulin-dependent signaling and resistance, were observed in the vicinity of DMRs. Mitogenic function of insulin due to DNA methylation have been observed in activation of genes related to growth and proliferation such as Insulin-like growth factor 2 receptor (*IGF2R*) and its increased mRNA expression corresponding to hypomethylation at promoter region describes a mechanism to eradicate the excessive quantity of insulin. Futhermore, expression levels of *DNMT3B* was significantly increased, which influenced the epigenetic modifications which were critical in embryonic development, X-chromosome inactivation and imprinting.

Whether or not the above methylome changes persist later during embryonic development is not known and, we have not investigated in our current study about their effects on the health of offspring later in life.

Paper II

LPS *in vitro* treatment of bEECs at biologically relevant concentrations was shown to profoundly affect gene expression. More than 2000 protein-coding genes (approx. 10% of all genes) were found to be differentially expressed. A large number of biological pathways was influenced suggesting that LPS has a major impact on multiple functions in the endometrial cells. We observed that DEGs involved in pathways relating to immune response, proliferation, glycolysis, oxidative stress and general metabolism were detected. Many of

these have also been reported in literature before but the specific molecules involved in implantation and immuno-tolerance mechanisms in response to LPS are documented here for the first time. This shows that our *in vitro* model both confirms previous knowledge concerning endometritis but, importantly it also identified novel DEGs in such pathways. Our results also pinpoint that LPS-induced cellular disturbances responsible for crucial functions in the endometrium at implantation and subsequently during embryonic development may impair successful fetal development. These results provide new insights in the molecular mechanisms linking LPS and impairment of bovine reproductive function. These findings also open the way for further functional studies to link the results from this in vitro model to the in vivo situation and find reliable markers of persistent inflammation in the bovine endometrium that may damage fertility even after infection has disappeared.

Paper III

LPS treatment has affected DNA methylation patterns of bEEC and overall a hypo methylation pattern was observed. In aggregate, 1291 DMRs were identified, out of which 707(55%) and 584(45%) DMRs were hypo-methylated and hyper-methylated respectively. The genomic distribution of DMRs revealed that 46% of the DMRs were overlapping with CpG islands (CGIs), 31% with CGI shores, and 23% with other genomic regions. Furthermore, 47.5% were on the gene body, 47.6% were intergenic and 4.8% mapped to the promoter regions. Furthermore, 600 differentially methylated genes (DMGs) were found comprising 589 protein-coding, 7 miRNA, and 4 pseudogenes. DMRs were found significantly enriched in 2Mbp regions, taken adjacent to telomeres as compared to non-telomeric regions. A negative correlation was found between 39 DEGs ($|\Delta$expression log2FC$| > 1$) and DMRs ($|\Delta$methylation$| > 5\%$) and related to ion/calcium ion transport and signal transduction processes. GO analysis of methylated genes revealed that apoptotic process, signal transduction, cell proliferation, vasculogenesis and embryo development were enriched. In addition, significantly enriched pathways including Calcium Signaling, MAPK Signaling, Oxytocin Signaling, Wnt Signaling, Sudden Infant Death Syndrome (SIDS) susceptibility, and IL-2 Signaling were found.

These analyses allow the identification of regions harbouring candidates for key regulatory elements of endometrial function, thus contributing to the understanding of LPS-induced deregulation that may impact implantation.

Paper IV

Our previous study *Paper III*, revealed 1291 differentially methylated regions (DMRs) using RRBS data in the genome of bEEC, challenged with LPS treatment. In this study, we used these DMRs to explore their regulatory role in gene transcription by identifying consensus sequence motifs within these DMRs and corresponding potential transcription factor binding sites (TFBS). The most enriched motifs were further scrutinized for TFBS and their known transcription factors. TFBS were found related to some immunologically important transcription factors.

Furthermore, mRNA expressions of *HIF-1A* and *STAT1* were highly over-expressed and the enrichment of their consensus sequence binding motifs strongly indicates a link between hypoxia and inflammation in endometritis. Identification of additional transcription factors that regulate gene expression under normal and hypoxic conditions in bEEC may lead to useful clinical applications to treat patients with endometritis as well as other inflammatory diseases. Results obtained from studies of endometritis in cattle may also be used as a model for endometritis in women and thereby having a substantial comparative value.

6 Future perspectives and implications

In *Paper I*, it is clear from our *in vitro* model that insulin has a major functional effect. The fact that insulin treatment during oocyte maturation has a massive effect on both the transcriptome and epigenome suggests that standard *in vitro* fertilisation regimes needs to be further investigated. There is a concern based on our findings that the effects of persistent DNA methylome and mRNA expression changes during early development of *in vitro* fertilised embryos may also have an impact on the health of individuals born after such conditions.

In *Papers II, III* and *IV* we have performed an in depth analyses of differential gene expression and methylation in our *in vitro* model for bacterial-induced inflammation in the bovine endometrium. The future perspectives and implications of our results are largely based on an increased knowledge of the changed transcriptome and the identification of the regulatory networks of *cis*-acting and *trans*-acting factors that are critical for these biological processes. Our obtained results may lead to the development of future strategies to treat and prevent major consequences of uterine inflammation in mammals. There is a need for further functional studies aiming to relate the results from LPS treated bEEC *in vitro* model to uterine diseases where cow endometrium is exposed to higher LPS concentrations for a longer time and to study them *in vivo*.

Due to the limitation of RRBS technology, whole genome-based approaches should be conducted because we might have missed to identify some regulatory regions that could play vital role in regulation of genes regarding reproduction, hence WGBS is recommended for further / deep investigation. Our in depth Bioinformatics study regarding enriched and co-localized TFBS located within the identified DMRs opens up obvious experiments to analyse the molecular mechanisms for how such transcription factors regulate transcription during an ongoing inflammation. We have identified novel TFs in combination with well-known TFs involved in inflammatory responses. Our *in vitro* model for bovine endometritis and their *in silico* annotation will guide us for further functional studies aiming at characterization of the functional factors which could be involved during bacterially-induced inflammatory diseases in cattle and other mammals including human.

7 References

Aberg, K.A., McClay, J.L., Nerella, S., Xie, L.Y., Clark, S.L., Hudson, A.D., Bukszár, J., Adkins, D., Swedish Schizophrenia Consortium, Hultman, C.M., Sullivan, P.F., Magnusson, P.K.E., van den Oord, E.J.C.G., 2012. MBD-seq as a cost-effective approach for methylome-wide association studies: demonstration in 1500 case--control samples. Epigenomics 4, 605–621. https://doi.org/10.2217/epi.12.59

Adamiak, S.J., Mackie, K., Watt, R.G., Webb, R., Sinclair, K.D., 2005. Impact of Nutrition on Oocyte Quality: Cumulative Effects of Body Composition and Diet Leading to Hyperinsulinemia in Cattle. Biology of Reproduction 73, 918–926. https://doi.org/10.1095/biolreprod.105.041483

Agrawal, V., Jaiswal, M.K., Kumar, Y., 2013. LPS-Induced Implantation Failure: One of the Causes of Female Infertility 4.

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E., 2012. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biology 13, R87. https://doi.org/10.1186/gb-2012-13-10-r87

Alabert, C., Groth, A., 2012. Chromatin replication and epigenome maintenance. Nature Reviews Molecular Cell Biology 13, 153–167. https://doi.org/10.1038/nrm3288

Alberts, B., Johnson, A.D., Lewis, J., Morgan, D., Raff, M., Roberts, K., Walter, P., 2014. Molecular Biology of the Cell, Sixth edition. ed. W. W. Norton & Company, New York, NY.

Ambros, V., 2004. The functions of animal microRNAs. Nature 431, 350–355. https://doi.org/10.1038/nature02871

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Antosik, K., Borowiec, M., 2016. Genetic Factors of Diabetes. Arch. Immunol. Ther. Exp. (Warsz.) 64, 157–160. https://doi.org/10.1007/s00005-016-0432-8

Aplin, J.D., 1997. Adhesion molecules in implantation. Rev. Reprod. 2, 84–93.

Arias-Álvarez, M., García-García, R.M., Lorenzo, P.L., Gutiérrez-Adán, A., Sakr, O.G., González-Bulnes, A., Rebollar, P.G., 2013. Embryo gene expression in response to maternal supplementation with glycogenic precursors in the rabbit. Anim. Reprod. Sci. 142, 173–182.

Ashworth, C.J., Toma, L.M., Hunter, M.G., 2009. Nutritional effects on oocyte and embryo development in mammals: implications for reproductive efficiency and environmental sustainability. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 364, 3351–3361. https://doi.org/10.1098/rstb.2009.0184

Attwood, T.K., Gisel, A., Bongcam-Rudloff, N.-E.E. and E., 2011. Concepts, Historical Milestones and the Central Place of Bioinformatics in

Modern Biology: A European Perspective. Bioinformatics - Trends and Methodologies. https://doi.org/10.5772/23535
Barrientos, G., Freitag, N., Tirado-González, I., Unverdorben, L., Jeschke, U., Thijssen, V.L.J.L., Blois, S.M., 2014. Involvement of galectin-1 in reproduction: past, present and future. Hum. Reprod. Update 20, 175–193. https://doi.org/10.1093/humupd/dmt040
Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., Herman, J.G., 2001. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum Mol Genet 10, 687–692. https://doi.org/10.1093/hmg/10.7.687
Baylin, S.B., Herman, J.G., 2000. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends in Genetics 16, 168–174. https://doi.org/10.1016/S0168-9525(99)01971-X
Bernstein, B.E., Meissner, A., Lander, E.S., 2007. The Mammalian Epigenome. Cell 128, 669–681. https://doi.org/10.1016/j.cell.2007.01.033
Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M., 1977. The protein data bank: A computer-based archival file for macromolecular structures. Journal of Molecular Biology 112, 535–542. https://doi.org/10.1016/S0022-2836(77)80200-3
Bierne, H., Hamon, M., Cossart, P., 2012. Epigenetics and Bacterial Infections. Cold Spring Harb Perspect Med 2, a010272. https://doi.org/10.1101/cshperspect.a010272
Binelli, M., Subramaniam, P., Diaz, T., Johnson, G.A., Hansen, T.R., Badinga, L., Thatcher, W.W., 2001. Bovine interferon-tau stimulates the Janus kinase-signal transducer and activator of transcription pathway in bovine endometrial epithelial cells. Biol. Reprod. 64, 654–665.
Blasco, M.A., 2007. The epigenetic regulation of mammalian telomeres. Nat. Rev. Genet. 8, 299–309. https://doi.org/10.1038/nrg2047
Bock, C., Tomazou, E.M., Brinkman, A.B., Müller, F., Simmer, F., Gu, H., Jäger, N., Gnirke, A., Stunnenberg, H.G., Meissner, A., 2010. Quantitative comparison of genome-wide DNA methylation mapping technologies. Nature Biotechnology 28, 1106. https://doi.org/10.1038/nbt.1681
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170
Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A.J., Socci, N.D., Scandura, J.M., 2011. DNA methylation of the first exon is tightly linked to transcriptional silencing. PLoS ONE 6, e14524. https://doi.org/10.1371/journal.pone.0014524
Bumgarner, R., 2013. Overview of DNA microarrays: types, applications, and their future. Curr Protoc Mol Biol Chapter 22, Unit 22.1. https://doi.org/10.1002/0471142727.mb2201s101

Burks, C., Fickett, J.W., Goad, W.B., Kanehisa, M., Lewitter, F.I., Rindone, W.P., Swindell, C.D., Tung, C.S., Bilofsky, H.S., 1985. The GenBank nucleic acid sequence database. Comput. Appl. Biosci. 1, 225–233.

Cao, J., Wei, C., Liu, D., Wang, H., Wu, M., Xie, Z., Capellini, T.D., Zhang, L., Zhao, F., Li, L., Zhong, T., Wang, L., Lu, J., Liu, R., Zhang, S., Du, Y., Zhang, H., Du, L., 2015. DNA methylation Landscape of body size variation in sheep., DNA methylation Landscape of body size variation in sheep. Sci Rep 5, 5, 13950–13950. https://doi.org/10.1038/srep13950, 10.1038/srep13950

Carey, N., Marques, C.J., Reik, W., 2011. DNA demethylases: a new epigenetic frontier in drug discovery. Drug Discovery Today 16, 683–690. https://doi.org/10.1016/j.drudis.2011.05.004

Carthew, R.W., Sontheimer, E.J., 2009. Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642–655. https://doi.org/10.1016/j.cell.2009.01.035

Chaisson, M.J.P., Huddleston, J., Dennis, M.Y., Sudmant, P.H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J.M., Stamatoyannopoulos, J.A., Hunkapiller, M.W., Korlach, J., Eichler, E.E., 2015. Resolving the complexity of the human genome using single-molecule sequencing. Nature 517, 608–611. https://doi.org/10.1038/nature13907

Chanrot, M., Guo, Y., Dalin, A.M., Persson, E., Båge, R., Svensson, A., Gustafsson, H., Humblot, P., 2017. Dose related effects of LPS on endometrial epithelial cell populations from dioestrus cows. Anim. Reprod. Sci. 177, 12–24. https://doi.org/10.1016/j.anireprosci.2016.12.002

Charpigny, G., Reinaud, P., Creminon, C., Tamby, J.P., 1999. Correlation of increased concentration of ovine endometrial cyclooxygenase 2 with the increase in PGE2 and PGD2 in the late luteal phase. Journal of reproduction and fertility 117, 315–324.

Chatterjee, A., Ozaki, Y., Stockwell, P.A., Horsfield, J.A., Morison, I.M., Nakagawa, S., 2013. Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing. Epigenetics 8, 979–989. https://doi.org/10.4161/epi.25797

Chen, X., Jiang, L., Wang, C.C., Huang, J., Li, T.C., 2016. Hypoxia inducible factor and microvessels in peri-implantation endometrium of women with recurrent miscarriage. Fertil. Steril. 105, 1496-1502.e4. https://doi.org/10.1016/j.fertnstert.2016.02.032

Chittur, S., Parr, B., Marcovici, G., 2011. Inhibition of inflammatory gene expression in keratinocytes using a composition containing carnitine, thioctic Acid and saw palmetto extract. Evid Based Complement Alternat Med 2011, 985345. https://doi.org/10.1093/ecam/nep102

Choi, M., Lee, J., Le, M.T., Nguyen, D.T., Park, S., Soundrarajan, N., Schachtschneider, K.M., Kim, J., Park, J.-K., Kim, J.-H., Park, C., 2015. Genome-wide analysis of DNA methylation in pigs using

reduced representation bisulfite sequencing. DNA Res 22, 343–355. https://doi.org/10.1093/dnares/dsv017
Conley, A.B., Jordan, I.K., 2012. Endogenous Retroviruses and the Epigenome, in: Witzany, G. (Ed.), Viruses: Essential Agents of Life. Springer Netherlands, Dordrecht, pp. 309–323. https://doi.org/10.1007/978-94-007-4899-6_16
Couldrey, C., Brauning, R., Bracegirdle, J., Maclean, P., Henderson, H.V., McEwan, J.C., 2014. Genome-Wide DNA Methylation Patterns and Transcription Analysis in Sheep Muscle. PLOS ONE 9, e101853. https://doi.org/10.1371/journal.pone.0101853
Cryer, M.J., Horani, T., DiPette, D.J., 2016. Diabetes and Hypertension: A Comparative Review of Current Guidelines. J Clin Hypertens (Greenwich) 18, 95–100. https://doi.org/10.1111/jch.12638
Dekel, N., Gnainsky, Y., Granot, I., Mor, G., 2010. Inflammation and implantation. Am. J. Reprod. Immunol. 63, 17–21. https://doi.org/10.1111/j.1600-0897.2009.00792.x
Dendup, T., Feng, X., Clingan, S., Astell-Burt, T., 2018. Environmental Risk Factors for Developing Type 2 Diabetes Mellitus: A Systematic Review. Int J Environ Res Public Health 15. https://doi.org/10.3390/ijerph15010078
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635
Doherty, R., Couldrey, C., 2014. Exploring genome wide bisulfite sequencing for DNA methylation analysis in livestock: a technical assessment. Front. Genet. 5. https://doi.org/10.3389/fgene.2014.00126
Downs, J.A., Lowndes, N.F., Jackson, S.P., 2000. A role for *Saccharomyces cerevisiae* histone H2A in DNA repair. Nature 408, 1001–1004. https://doi.org/10.1038/35050000
Dupont, C., Armant, D.R., Brenner, C.A., 2009. Epigenetics: Definition, Mechanisms and Clinical Perspective. Semin Reprod Med 27, 351–357. https://doi.org/10.1055/s-0029-1237423
Feng, S., Jacobsen, S.E., Reik, W., 2010. Epigenetic reprogramming in plant and animal development. Science 330, 622–627. https://doi.org/10.1126/science.1190614
Fischle, W., Wang, Y., Allis, C.D., 2003. Histone and chromatin cross-talk. Current Opinion in Cell Biology 15, 172–183. https://doi.org/10.1016/S0955-0674(03)00013-9
Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., Turner, S.W., 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nature Methods 7, 461–465. https://doi.org/10.1038/nmeth.1459
Freret, S., Grimard, B., Ponter, A.A., Joly, C., Ponsart, C., Humblot, P., 2006. Reduction of body-weight gain enhances in vitro embryo production

in overfed superovulated dairy heifers. Reproduction 131, 783–794. https://doi.org/10.1530/rep.1.00689
Froehlich, R., Hambruch, N., Haeger, J.D., Dilly, M., Kaltner, H., Gabius, H.J., Pfarrer, C., 2012. Galectin fingerprinting detects differences in expression profiles between bovine endometrium and placentomes as well as early and late gestational stages. Placenta 33, 195–201. https://doi.org/10.1016/j.placenta.2011.12.016
Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., Paul, C.L., 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U.S.A. 89, 1827–1831.
Gadalla, S.M., Katki, H.A., Shebl, F.M., Giri, N., Alter, B.P., Savage, S.A., 2012. The relationship between DNA methylation and telomere length in dyskeratosis congenita. Aging Cell 11, 24–28. https://doi.org/10.1111/j.1474-9726.2011.00755.x
Galvão, K.N., Greco, L.F., Vilela, J.M., Sá Filho, M.F., Santos, J.E.P., 2009. Effect of intrauterine infusion of ceftiofur on uterine health and fertility in dairy cows. Journal of Dairy Science 92, 1532–1542. https://doi.org/10.3168/jds.2008-1615
Gao, F., Zhang, J., Jiang, P., Gong, D., Wang, J.-W., Xia, Y., Østergaard, M.V., Wang, J., Sangild, P.T., 2014. Marked methylation changes in intestinal genes during the perinatal period of preterm neonates. BMC Genomics 15, 716. https://doi.org/10.1186/1471-2164-15-716
Gardiner-Garden, M., Frommer, M., 1987. CpG islands in vertebrate genomes. J. Mol. Biol. 196, 261–282.
Garnsworthy, P.C., Sinclair, K.D., Webb, R., 2008. Integration of physiological mechanisms that influence fertility in dairy cows*. animal 2, 1144–1152. https://doi.org/10.1017/S1751731108002358
Gigante, S., Gouil, Q., Lucattini, A., Keniry, A., Beck, T., Tinning, M., Gordon, L., Woodruff, C., Speed, T.P., Blewitt, M., Ritchie, M., 2018. Using long-read sequencing to detect imprinted DNA methylation. bioRxiv 445924. https://doi.org/10.1101/445924
Gilbert, R.O., 2011. The effects of endometritis on the establishment of pregnancy in cattle. Reprod. Fertil. Dev. 24, 252–257. https://doi.org/10.1071/RD11915
Goldberg, A.D., Allis, C.D., Bernstein, E., 2007. Epigenetics: A Landscape Takes Shape. Cell 128, 635–638. https://doi.org/10.1016/j.cell.2007.02.006
Gong, J.G., Lee, W.J., Garnsworthy, P.C., Webb, R., 2002. Effect of dietary-induced increases in circulating insulin concentrations during the early postpartum period on reproductive function in dairy cows. Reproduction 123, 419–427.
Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17, 333–351. https://doi.org/10.1038/nrg.2016.49

Gou, J., Jia, J., Feng, J., Zhao, X., Yi, T., Cui, T., Li, Z., 2017. Stathmin 1 plays a role in endometrial decidualisation by regulating hypoxia inducible factor-1α and vascular endothelial growth factor during embryo implantation. Reprod. Fertil. Dev. 29, 1530–1537. https://doi.org/10.1071/RD15539

Grazul-Bilska, A.T., Borowczyk, E., Bilski, J.J., Reynolds, L.P., Redmer, D.A., Caton, J.S., Vonnahme, K.A., 2012. Overfeeding and underfeeding have detrimental effects on oocyte quality measured by in vitro fertilization and early embryonic development in sheep. Domest. Anim. Endocrinol. 43, 289–298. https://doi.org/10.1016/j.domaniend.2012.05.001

Green, B.B., Kerr, D.E., 2014. Epigenetic contribution to individual variation in response to lipopolysaccharide in bovine dermal fibroblasts. Vet. Immunol. Immunopathol. 157, 49–58. https://doi.org/10.1016/j.vetimm.2013.10.015

Grimard, B., Freret, S., Chevallier, A., Pinto, A., Ponsart, C., Humblot, P., 2006. Genetic and environmental factors influencing first service conception rate and late embryonic/foetal mortality in low fertility dairy herds. Animal Reproduction Science 91, 31–44. https://doi.org/10.1016/j.anireprosci.2005.03.003

Grove, C.A., De Masi, F., Barrasa, M.I., Newburger, D.E., Alkema, M.J., Bulyk, M.L., Walhout, A.J.M., 2009. A Multiparameter Network Reveals Extensive Divergence between C. elegans bHLH Transcription Factors. Cell 138, 314–327. https://doi.org/10.1016/j.cell.2009.04.058

Gu, H., Bock, C., Mikkelsen, T.S., Jäger, N., Smith, Z.D., Tomazou, E., Gnirke, A., Lander, E.S., Meissner, A., 2010. Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution. Nat Methods 7, 133–136. https://doi.org/10.1038/nmeth.1414

Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A., Meissner, A., 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 6, 468–481. https://doi.org/10.1038/nprot.2010.190

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., Fan, X., Yong, J., Wen, L., Xie, S.X., Tang, F., Qiao, J., 2014. The DNA methylation landscape of human early embryos. Nature 511, 606–610. https://doi.org/10.1038/nature13544

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.-Y., Pellegrini, M., 2013. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics 14, 774. https://doi.org/10.1186/1471-2164-14-774

Haller, D., Holt, L., Kim, S.C., Schwabe, R.F., Sartor, R.B., Jobin, C., 2003. Transforming Growth Factor-β1 Inhibits Non-pathogenic

Gramnegative Bacteria-induced NF-κB Recruitment to the Interleukin-6 Gene Promoter in Intestinal Epithelial Cells through Modulation of Histone Acetylation. J. Biol. Chem. 278, 23851–23860. https://doi.org/10.1074/jbc.M300075200

Hamm, G.H., Cameron, G.N., 1986. The EMBL data library. Nucleic Acids Res. 14, 5–9.

Harris, R.A., Wang, T., Coarfa, C., Nagarajan, R.P., Hong, C., Downey, S.L., Johnson, B.E., Fouse, S.D., Delaney, A., Zhao, Y., Olshen, A., Ballinger, T., Zhou, X., Forsberg, K.J., Gu, J., Echipare, L., O'Geen, H., Lister, R., Pelizzola, M., Xi, Y., Epstein, C.B., Bernstein, B.E., Hawkins, R.D., Ren, B., Chung, W.-Y., Gu, H., Bock, C., Gnirke, A., Zhang, M.Q., Haussler, D., Ecker, J.R., Li, W., Farnham, P.J., Waterland, R.A., Meissner, A., Marra, M.A., Hirst, M., Milosavljevic, A., Costello, J.F., 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. Nat. Biotechnol. 28, 1097–1105. https://doi.org/10.1038/nbt.1682

Hartung, T., Zhang, L., Kanwar, R., Khrebtukova, I., Reinhardt, M., Wang, C., Therneau, T.M., Banck, M.S., Schroth, G.P., Beutler, A.S., 2012. Diametrically opposite methylome-transcriptome relationships in high- and low-CpG promoter genes in postmitotic neural rat tissue. Epigenetics 7, 421–428. https://doi.org/10.4161/epi.19565

Herath, S., Lilly, S.T., Fischer, D.P., Williams, E.J., Dobson, H., Bryant, C.E., Sheldon, I.M., 2009. Bacterial lipopolysaccharide induces an endocrine switch from prostaglandin F2alpha to prostaglandin E2 in bovine endometrium. Endocrinology 150, 1912–1920. https://doi.org/10.1210/en.2008-1379

Hogeweg, P., 1978. Simulating the growth of cellular forms. SIMULATION 31, 90–96. https://doi.org/10.1177/003754977803100305

Hogeweg, P., Hesper, B., 1978. Interactive instruction on population interactions. Computers in Biology and Medicine 8, 319–327. https://doi.org/10.1016/0010-4825(78)90032-X

Houshdaran, S., Zelenko, Z., Irwin, J.C., Giudice, L.C., 2014. Human endometrial DNA methylome is cycle-dependent and is associated with gene expression regulation. Mol. Endocrinol. 28, 1118–1135. https://doi.org/10.1210/me.2013-1340

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44–57. https://doi.org/10.1038/nprot.2008.211

Huang, Y.Z., Sun, J.J., Zhang, L.Z., Li, C.J., Womack, J.E., Li, Z.J., Lan, X.Y., Lei, C.Z., Zhang, C.L., Zhao, X., Chen, H., 2014. Genome-wide DNA methylation profiles and their relationships with mRNA and the microRNA transcriptome in bovine muscle tissue (Bos taurine)., Genome-wide DNA Methylation Profiles and Their Relationships with mRNA and the microRNA Transcriptome in Bovine Muscle Tissue

(Bos taurine). Sci Rep 4, 4, 6546–6546. https://doi.org/10.1038/srep06546, 10.1038/srep06546
Hunt, L.T., 1983. Margaret O. Dayhoff 1925-1983. DNA 2, 97–98.
International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062
Ispada, J., de Lima, C.B., Sirard, M.-A., Fontes, P.K., Nogueira, M.F.G., Annes, K., Milazzotto, M.P., 2018. Genome-wide screening of DNA methylation in bovine blastocysts with different kinetics of development. Epigenetics Chromatin 11, 1. https://doi.org/10.1186/s13072-017-0171-z
Issa, J.-P., 2004. CpG island methylator phenotype in cancer. Nat. Rev. Cancer 4, 988–993. https://doi.org/10.1038/nrc1507
Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., Zhang, Y., 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature 466, 1129–1133. https://doi.org/10.1038/nature09303
Jacob, F., Monod, J., 1961. Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology 3, 318–356. https://doi.org/10.1016/S0022-2836(61)80072-7
Jain, M., Olsen, H.E., Paten, B., Akeson, M., 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biology 17, 239. https://doi.org/10.1186/s13059-016-1103-0
Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492. https://doi.org/10.1038/nrg3230
Jones, P.A., Laird, P.W., 1999. Cancer-epigenetics comes of age. Nature Genetics 21, 163–167. https://doi.org/10.1038/5947
Kahn, S.E., Hull, R.L., Utzschneider, K.M., 2006. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature 444, 840–846. https://doi.org/10.1038/nature05482
Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., Baranasic, D., Arenillas, D.J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W.W., Parcy, F., Mathelier, A., 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 46, D260–D266. https://doi.org/10.1093/nar/gkx1126
Konac, E., Alp, E., Onen, H.I., Korucuoglu, U., Biri, A.A., Menevse, S., 2009. Endometrial mRNA expression of matrix metalloproteinases, their tissue inhibitors and cell adhesion molecules in unexplained infertility and implantation failure patients. Reprod. Biomed. Online 19, 391–397.

Konisti, S., Kiriakidis, S., Paleolog, E.M., 2012. Hypoxia—a key regulator of angiogenesis and inflammation in rheumatoid arthritis. Nature Reviews Rheumatology 8, 153–162. https://doi.org/10.1038/nrrheum.2011.205

Kristensen, L.S., Hansen, L.L., 2009. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. Clin. Chem. 55, 1471–1483. https://doi.org/10.1373/clinchem.2008.121962

Kwon, A.T., Arenillas, D.J., Worsley Hunt, R., Wasserman, W.W., 2012. oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets. G3 (Bethesda) 2, 987–1002. https://doi.org/10.1534/g3.112.003202

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, Y., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou,

M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., Szustakowki, J., International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Larhammar, D., Hyldig-Nielsen, J.J., Servenius, B., Andersson, G., Rask, L., Peterson, P.A., 1983. Exon-intron organization and complete nucleotide sequence of a human major histocompatibility antigen DC beta gene. PNAS 80, 7313–7317. https://doi.org/10.1073/pnas.80.23.7313

Laskowski, D., Båge, R., Humblot, P., Andersson, G., Sirard, M.-A., Sjunnesson, Y., 2017a. Insulin during in vitro oocyte maturation has an impact on development, mitochondria, and cytoskeleton in bovine day 8 blastocysts. Theriogenology 101, 15–25. https://doi.org/10.1016/j.theriogenology.2017.06.002

Laskowski, D., Sjunnesson, Y., Humblot, P., Andersson, G., Gustafsson, H., Båge, R., 2016. The functional role of insulin in fertility and embryonic development-What can we learn from the bovine model? Theriogenology 86, 457–464. https://doi.org/10.1016/j.theriogenology.2016.04.062

Laskowski, D., Sjunnesson, Y., Humblot, P., Sirard, M.-A., Andersson, G., Gustafsson, H., Båge, R., 2017b. Insulin exposure during in vitro bovine oocyte maturation changes blastocyst gene expression and developmental potential. Reprod. Fertil. Dev. 29, 876–889. https://doi.org/10.1071/RD15315

Latchman, D.S., 1993. Transcription factors: an overview. Int J Exp Pathol 74, 417–422.

Law, J.A., Jacobsen, S.E., 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11, 204–220. https://doi.org/10.1038/nrg2719

Lee, J.-R., Hong, C.P., Moon, J.-W., Jung, Y.-D., Kim, D.-S., Kim, T.-H., Gim, J.-A., Bae, J.-H., Choi, Y., Eo, J., Kwon, Y.-J., Song, S., Ko, J., Yang, Y.M., Lee, H.-K., Park, K.-D., Ahn, K., Do, K.-T., Ha, H.-S., Han, K., Yi, J.M., Cha, H.-J., Cho, B.-W., Bhak, J., Kim, H.-S., 2014. Genome-wide analysis of DNA methylation patterns in horse. BMC Genomics 15, 598. https://doi.org/10.1186/1471-2164-15-598

Lennartsson, A., Ekwall, K., 2009. Histone modification patterns and epigenetic codes. Biochimica et Biophysica Acta (BBA) - General Subjects, Epigenetic Control of Gene Expression 1790, 863–868. https://doi.org/10.1016/j.bbagen.2008.12.006

Lessey, B.A., 2002. Adhesion molecules and implantation. J. Reprod. Immunol. 55, 101–112.

Levine, M., Tjian, R., 2003. Transcription regulation and animal diversity. Nature 424, 147–151. https://doi.org/10.1038/nature01763

Lewis, M., 2011. A tale of two repressors – a historical perspective. J Mol Biol 409, 14–27. https://doi.org/10.1016/j.jmb.2011.02.023

Li, L., Yang, X.-J., 2016. Molecular and Functional Characterization of Histone Deacetylase 4 (HDAC4). Methods Mol. Biol. 1436, 31–45. https://doi.org/10.1007/978-1-4939-3667-0_4

Lieberman, P.M., 2006. Chromatin regulation of virus infection. Trends in Microbiology 14, 132–140. https://doi.org/10.1016/j.tim.2006.01.001

Lin, J., Marcum, C.S., Myers, M.F., Koehly, L.M., 2018. Racial differences in family health history knowledge of type 2 diabetes: exploring the role of interpersonal mechanisms. Transl Behav Med 8, 540–549. https://doi.org/10.1093/tbm/ibx062

Lincke, A., Drillich, M., Heuwieser, W., 2007. [Subclinical endometritis in dairy cattle and its effect on fertility--a review of recent publications]. Berl. Munch. Tierarztl. Wochenschr. 120, 245–250.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315. https://doi.org/10.1038/nature08514

Lodish, H., Berk, A., Kaiser, C.A., Krieger, M., Scott, M.P., Bretscher, A., Ploegh, H., Matsudaira, P., 2007. Molecular Cell Biology, 6th edition. ed. W. H. Freeman, New York.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., Shafee, T., 2017. Transcriptomics technologies. PLoS Comput. Biol. 13, e1005457. https://doi.org/10.1371/journal.pcbi.1005457

Luscombe, N.M., Greenbaum, D., Gerstein, M., 2001. What is Bioinformatics? A Proposed Definition and Overview of the Field. Methods of

Information in Medicine 40, 346–358. https://doi.org/10.1055/s-0038-1634431
Maekawa, R., Yagi, S., Ohgane, J., Yamagata, Y., Asada, H., Tamura, I., Sugino, N., Shiota, K., 2011. Disease-dependent differently methylated regions (D-DMRs) of DNA are enriched on the X chromosome in uterine leiomyoma. J. Reprod. Dev. 57, 604–612.
Maj, T., Chelmonska-Soyta, A., 2007. Pleiotropy and redundancy of STAT proteins in early pregnancy. Reprod. Domest. Anim. 42, 343–353. https://doi.org/10.1111/j.1439-0531.2006.00787.x
Mannoor, K., Liao, J., Jiang, F., 2012. Small nucleolar RNAs in cancer. Biochim Biophys Acta 1826. https://doi.org/10.1016/j.bbcan.2012.03.005
Mardis, E.R., 2013. Next-generation sequencing platforms. Annu Rev Anal Chem (Palo Alto Calif) 6, 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628
Mardis, E.R., 2008. The impact of next-generation sequencing technology on genetics. Trends Genet. 24, 133–141. https://doi.org/10.1016/j.tig.2007.12.007
Matera, A.G., Terns, R.M., Terns, M.P., 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. Nature Reviews Molecular Cell Biology 8, 209–220. https://doi.org/10.1038/nrm2124
Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.-Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A., Wasserman, W.W., 2016. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 44, D110-115. https://doi.org/10.1093/nar/gkv1176
Maxam, A.M., Gilbert, W., 1977. A new method for sequencing DNA. Proceedings of the National Academy of Sciences 74, 560–564. https://doi.org/10.1073/pnas.74.2.560
Maybin, J.A., Hirani, N., Jabbour, H.N., Critchley, H.O.D., 2011. Novel Roles for Hypoxia and Prostaglandin E2 in the Regulation of IL-8 During Endometrial Repair. The American Journal of Pathology 178, 1245–1256. https://doi.org/10.1016/j.ajpath.2010.11.070
Maybin, J.A., Murray, A.A., Saunders, P.T.K., Hirani, N., Carmeliet, P., Critchley, H.O.D., 2018. Hypoxia and hypoxia inducible factor-1α are required for normal endometrial repair during menstruation. Nature Communications 9, 295. https://doi.org/10.1038/s41467-017-02375-6
McGettigan, P.A., 2013. Transcriptomics in the RNA-seq era. Current Opinion in Chemical Biology, Omics 17, 4–11. https://doi.org/10.1016/j.cbpa.2012.12.008
McGuire, A.M., Hughes, J.D., Church, G.M., 2000. Conservation of DNA Regulatory Motifs and Discovery of New Motifs in Microbial Genomes. Genome Res. 10, 744–757. https://doi.org/10.1101/gr.10.6.744

Meissner, A., Gnirke, A., Bell, G., Ramsahoye, B., Lander, E., Jaenisch, R., 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33. https://doi.org/10.1093/nar/gki901

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., Gnirke, A., Jaenisch, R., Lander, E.S., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766. https://doi.org/10.1038/nature07107

Miglior, F., Muir, B.L., Van Doormaal, B.J., 2005. Selection Indices in Holstein Cattle of Various Countries. Journal of Dairy Science 88, 1255–1263. https://doi.org/10.3168/jds.S0022-0302(05)72792-2

Mitchell, P.J., Tjian, R., 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245, 371–378. https://doi.org/10.1126/science.2667136

Modugno, F., Ness, R.B., Chen, C., Weiss, N.S., 2005. Inflammation and Endometrial Cancer: A Hypothesis. Cancer Epidemiol Biomarkers Prev 14, 2840–2847. https://doi.org/10.1158/1055-9965.EPI-05-0493

Mulqueen, R.M., Pokholok, D., Norberg, S.J., Torkenczy, K.A., Fields, A.J., Sun, D., Sinnamon, J.R., Shendure, J., Trapnell, C., O'Roak, B.J., Xia, Z., Steemers, F.J., Adey, A.C., 2018. Highly scalable generation of DNA methylation profiles in single cells. Nature Biotechnology 36, 428–431. https://doi.org/10.1038/nbt.4112

Nagarajan, A., Roden, C., Wajapeyee, N., 2014. Reduced representation bisulfite sequencing to identify global alteration of DNA methylation. Methods Mol. Biol. 1176, 23–31. https://doi.org/10.1007/978-1-4939-0992-6_3

Nongbua, T., Guo, Y., Edman, A., Humblot, P., Morrell, J.M., 2018. Effect of bovine seminal plasma on bovine endometrial epithelial cells in culture. Reproduction in Domestic Animals 53, 85–92.

O'Dell, S.D., Day, I.N.M., 1998. Molecules in focus Insulin-like growth factor II (IGF-II). The International Journal of Biochemistry & Cell Biology 30, 767–771. https://doi.org/10.1016/S1357-2725(98)00048-X

Okano, M., Bell, D.W., Haber, D.A., Li, E., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247–257.

Opsomer, G., Gröhn, Y.T., Hertl, J., Coryn, M., Deluyker, H., de Kruif, A., 2000. Risk factors for post partum ovarian dysfunction in high producing dairy cows in Belgium: A field study. Theriogenology 53, 841–857. https://doi.org/10.1016/S0093-691X(00)00234-X

Park, H.J., Kim, Y.S., Yoon, T.K., Lee, W.S., 2016. Chronic endometritis and infertility. Clin Exp Reprod Med 43, 185–192. https://doi.org/10.5653/cerm.2016.43.4.185

Pasquali, R., Patton, L., Gambineri, A., 2007. Obesity and infertility. Current Opinion in Endocrinology, Diabetes and Obesity 14, 482. https://doi.org/10.1097/MED.0b013e3282f1d6cb
Pei, L., Choi, J.-H., Liu, J., Lee, E.-J., McCarthy, B., Wilson, J.M., Speir, E., Awan, F., Tae, H., Arthur, G., Schnabel, J.L., Taylor, K.H., Wang, X., Xu, D., Ding, H.-F., Munn, D.H., Caldwell, C., Shi, H., 2012. Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. Epigenetics 7, 567–578. https://doi.org/10.4161/epi.20237
Piras, C., Guo, Y., Soggiu, A., Chanrot, M., Greco, V., Urbani, A., Charpigny, G., Bonizzi, L., Roncada, P., Humblot, P., 2017a. Changes in protein expression profiles in bovine endometrial epithelial cells exposed to E. coli LPS challenge. Mol. BioSyst. 13, 392–405. https://doi.org/10.1039/C6MB00723F
Piras, C., Guo, Y., Soggiu, A., Chanrot, M., Greco, V., Urbani, A., Charpigny, G., Bonizzi, L., Roncada, P., Humblot, P., 2017b. Changes in protein expression profiles in bovine endometrial epithelial cells exposed to E. coli LPS challenge. Molecular BioSystems 13, 392–405.
Pontén, F., Jirström, K., Uhlen, M., 2008. The Human Protein Atlas—a tool for pathology. The Journal of Pathology 216, 387–393. https://doi.org/10.1002/path.2440
Popovici, R.M., Krause, M.S., Germeyer, A., Strowitzki, T., von Wolff, M., 2005. Galectin-9: a new endometrial epithelial marker for the mid- and late-secretory and decidual phases in humans. J. Clin. Endocrinol. Metab. 90, 6170–6176. https://doi.org/10.1210/jc.2004-2529
Pritsker, M., Liu, Y.-C., Beer, M.A., Tavazoie, S., 2004. Whole-Genome Discovery of Transcription Factor Binding Sites by Network-Level Conservation. Genome Res 14, 99–108. https://doi.org/10.1101/gr.1739204
Ransohoff, J.D., Wei, Y., Khavari, P.A., 2018. The functions and unique features of long intergenic non-coding RNA. Nature Reviews Molecular Cell Biology 19, 143–157. https://doi.org/10.1038/nrm.2017.104
Reece-Hoyes, J.S., Deplancke, B., Shingles, J., Grove, C.A., Hope, I.A., Walhout, A.J., 2005. A compendium of Caenhorabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. Genome Biol 6, R110. https://doi.org/10.1186/gb-2005-6-13-r110
Rhoads, A., Au, K.F., 2015. PacBio Sequencing and Its Applications. Genomics, Proteomics & Bioinformatics, SI: Metagenomics of Marine Environments 13, 278–289. https://doi.org/10.1016/j.gpb.2015.08.002
Robert, C., Nieminen, J., Dufort, I., Gagné, D., Grant, J.R., Cagnone, G., Plourde, D., Nivet, A.-L., Fournier, É., Paquet, É., Blazejczyk, M., Rigault, P., Juge, N., Sirard, M.-A., 2011. Combining resources to obtain a comprehensive survey of the bovine embryo transcriptome

through deep sequencing and microarrays. Mol. Reprod. Dev. 78, 651–664. https://doi.org/10.1002/mrd.21364
Robker, R.L., 2008. Evidence that obesity alters the quality of oocytes and embryos. Pathophysiology 15, 115–121. https://doi.org/10.1016/j.pathophys.2008.04.004
Roeder, R.G., 1996. The role of general initiation factors in transcription by RNA polymerase II. Trends in Biochemical Sciences 21, 327–335. https://doi.org/10.1016/S0968-0004(96)10050-5
Royal, M.D., Darwash, A.O., Flint, A.P.F., Webb, R., Woolliams, J.A., Lamming, G.E., 2000. Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. Animal Science 70, 487–501. https://doi.org/10.1017/S1357729800051845
Russell, P.J., 2009. iGenetics: A Molecular Approach, 3 edition. ed. Pearson, San Francisco.
Saare, M., Modhukur, V., Suhorutshenko, M., Rajashekar, B., Rekker, K., Sõritsa, D., Karro, H., Soplepmann, P., Sõritsa, A., Lindgren, C.M., Rahmioglu, N., Drong, A., Becker, C.M., Zondervan, K.T., Salumets, A., Peters, M., 2016. The influence of menstrual cycle and endometriosis on endometrial methylome. Clin Epigenetics 8, 2. https://doi.org/10.1186/s13148-015-0168-z
Sakumoto, T., Tokunaga, Y., Tanaka, H., Nohara, M., Motegi, E., Shinkawa, T., Nakaza, A., Higashi, M., 2010. Insulin resistance/hyperinsulinemia and reproductive disorders in infertile women. Reprod Med Biol 9, 185–190. https://doi.org/10.1007/s12522-010-0062-5
Salilew-Wondim, D., Fournier, E., Hoelker, M., Saeed-Zidane, M., Tholen, E., Looft, C., Neuhoff, C., Besenfelder, U., Havlicek, V., Rings, F., Gagné, D., Sirard, M.-A., Robert, C., Shojaei Saadi, H.A., Gad, A., Schellander, K., Tesfaye, D., 2015. Genome-Wide DNA Methylation Patterns of Bovine Blastocysts Developed In Vivo from Embryos Completed Different Stages of Development In Vitro. PLoS ONE 10, e0140467. https://doi.org/10.1371/journal.pone.0140467
Salilew-Wondim, D., Saeed-Zidane, M., Hoelker, M., Gebremedhn, S., Poirier, M., Pandey, H.O., Tholen, E., Neuhoff, C., Held, E., Besenfelder, U., Havlicek, V., Rings, F., Fournier, E., Gagné, D., Sirard, M.-A., Robert, C., Gad, A., Schellander, K., Tesfaye, D., 2018. Genome-wide DNA methylation patterns of bovine blastocysts derived from in vivo embryos subjected to in vitro culture before, during or after embryonic genome activation. BMC Genomics 19, 424. https://doi.org/10.1186/s12864-018-4826-3
Salleh, N., Giribabu, N., 2014. Leukemia Inhibitory Factor: Roles in Embryo Implantation and in Nonhormonal Contraception. ScientificWorldJournal 2014. https://doi.org/10.1155/2014/201514
Schroeder, D.I., Jayashankar, K., Douglas, K.C., Thirkill, T.L., York, D., Dickinson, P.J., Williams, L.E., Samollow, P.B., Ross, P.J., Bannasch, D.L., Douglas, G.C., LaSalle, J.M., 2015. Early Developmental and

Evolutionary Origins of Gene Body DNA Methylation Patterns in Mammalian Placentas. PLOS Genetics 11, e1005442. https://doi.org/10.1371/journal.pgen.1005442
Selvaraj, S., Natarajan, J., 2011. Microarray Data Analysis and Mining Tools. Bioinformation 6, 95–99.
Shapiro, E., Biezuner, T., Linnarsson, S., 2013. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nature Reviews Genetics 14, 618–630. https://doi.org/10.1038/nrg3542
Sharp, S.J., Schaack, J., Cooley, L., Burke, D.J., Soil, D., 1985. Structure and Transcription of Eukaryotic tRNA Gene. Critical Reviews in Biochemistry 19, 107–144. https://doi.org/10.3109/10409238509082541
Sheldon, I.M., Cronin, J., Goetze, L., Donofrio, G., Schuberth, H.-J., 2009. Defining Postpartum Uterine Disease and the Mechanisms of Infection and Immunity in the Female Reproductive Tract in Cattle. Biol Reprod 81, 1025–1032. https://doi.org/10.1095/biolreprod.109.077370
Sheldon IM, Price SB, Cronin J, Gilbert RO, Gadsby JE, 2009. Mechanisms of Infertility Associated with Clinical and Subclinical Endometritis in High Producing Dairy Cattle. Reproduction in Domestic Animals 44, 1–9. https://doi.org/10.1111/j.1439-0531.2009.01465.x
Sheldon, I.M., Rycroft, A.N., Dogan, B., Craven, M., Bromfield, J.J., Chandler, A., Roberts, M.H., Price, S.B., Gilbert, R.O., Simpson, K.W., 2010. Specific Strains of Escherichia coli Are Pathogenic for the Endometrium of Cattle and Cause Pelvic Inflammatory Disease in Cattle and Mice. PLOS ONE 5, e9192. https://doi.org/10.1371/journal.pone.0009192
Sheldon, I.M., Williams, E.J., Miller, A.N.A., Nash, D.M., Herath, S., 2008. Uterine diseases in cattle after parturition. Vet. J. 176, 115–121. https://doi.org/10.1016/j.tvjl.2007.12.031
Shendure, J., Ji, H., 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135–1145. https://doi.org/10.1038/nbt1486
Shojaei Saadi, H.A., O'Doherty, A.M., Gagné, D., Fournier, É., Grant, J.R., Sirard, M.-A., Robert, C., 2014. An integrated platform for bovine DNA methylome analysis suitable for small samples. BMC Genomics 15, 451. https://doi.org/10.1186/1471-2164-15-451
Simionato, E., Ledent, V., Richards, G., Thomas-Chollier, M., Kerner, P., Coornaert, D., Degnan, B.M., Vervoort, M., 2007. Origin and diversification of the basic helix-loop-helix gene family in metazoans: insights from comparative genomics. BMC Evolutionary Biology 7, 33. https://doi.org/10.1186/1471-2148-7-33
Simsek, D., Tiu, G.C., Flynn, R.A., Byeon, G.W., Leppek, K., Xu, A.F., Chang, H.Y., Barna, M., 2017. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. Cell 169, 1051-1065.e18. https://doi.org/10.1016/j.cell.2017.05.022

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., Meissner, A., 2014. DNA methylation dynamics of the human preimplantation embryo. Nature 511, 611–615. https://doi.org/10.1038/nature13581

Soffar, H., 2015. The function of the uterus in the female reproductive system. Science online. URL https://www.online-sciences.com/the-living-organisms/the-function-of-the-uterus-in-the-female-reproductive-system/ (accessed 9.16.18).

Song, G., Kim, J., Bazer, F.W., Spencer, T.E., 2008. Progesterone and interferon tau regulate hypoxia-inducible factors in the endometrium of the ovine uterus. Endocrinology 149, 1926–1934. https://doi.org/10.1210/en.2007-1530

Sonksen, P., Sonksen, J., 2000. Insulin: understanding its action in health and disease. Br J Anaesth 85, 69–79.

Stein, L., 2001. Genome annotation: from sequence to biology. Nature Reviews Genetics 2, 493–503. https://doi.org/10.1038/35080529

Stewart, C.L., Kaspar, P., Brunet, L.J., Bhatt, H., Gadi, I., Köntgen, F., Abbondanzo, S.J., 1992. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. Nature 359, 76–79. https://doi.org/10.1038/359076a0

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., Jacobsen, S.E., 2014. The roles of non-CG methylation in Arabidopsis. Nat Struct Mol Biol 21, 64–72. https://doi.org/10.1038/nsmb.2735

Su, J., Wang, Y., Xing, X., Liu, J., Zhang, Y., 2014a. Genome-wide analysis of DNA methylation in bovine placentas. BMC Genomics 15, 12. https://doi.org/10.1186/1471-2164-15-12

Su, J., Wang, Y., Xing, X., Liu, J., Zhang, Y., 2014b. Genome-wide analysis of DNA methylation in bovine placentas. BMC Genomics 15, 12. https://doi.org/10.1186/1471-2164-15-12

Taiwo, O., Wilson, G.A., Morris, T., Seisenberger, S., Reik, W., Pearce, D., Beck, S., Butcher, L.M., 2012. Methylome analysis using MeDIP-seq with low DNA concentrations. Nat Protoc 7, 617–636. https://doi.org/10.1038/nprot.2012.012

Takahashi, K., Sugi, Y., Nakano, K., Tsuda, M., Kurihara, K., Hosono, A., Kaminogawa, S., 2011. Epigenetic control of host gene by commensal bacteria in large intestinal epithelial cells. J. Biol. Chem. jbc.M111.271007. https://doi.org/10.1074/jbc.M111.271007

Tency, I., Verstraelen, H., Kroes, I., Holtappels, G., Verhasselt, B., Vaneechoutte, M., Verhelst, R., Temmerman, M., 2012. Imbalances between matrix metalloproteinases (MMPs) and tissue inhibitor of metalloproteinases (TIMPs) in maternal serum during preterm labor. PLoS ONE 7, e49042. https://doi.org/10.1371/journal.pone.0049042

Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Blal, H.A., Alm, T., Asplund, A., Björk, L., Breckels, L.M., Bäckström, A.,

Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey, C.M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J.M., Sivertsson, Å., Sjöstedt, E., Skogs, M., Stadler, C., Sullivan, D.P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen, M., Mardinoglu, A., Pontén, F., Feilitzen, K. von, Lilley, K.S., Uhlén, M., Lundberg, E., 2017. A subcellular map of the human proteome. Science 356, eaal3321. https://doi.org/10.1126/science.aal3321

Torchinsky, A., Markert, U.R., Toder, V., 2005. TNF-alpha-mediated stress-induced early pregnancy loss: a possible role of leukemia inhibitory factor. Chem Immunol Allergy 89, 62–71. https://doi.org/10.1159/000087913

Tyson, J.R., O'Neil, N.J., Jain, M., Olsen, H.E., Hieter, P., Snutch, T.P., 2018. MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome Res. 28, 266–274. https://doi.org/10.1101/gr.221184.117

Uysal, F., Akkoyunlu, G., Ozturk, S., 2015. Dynamic expression of DNA methyltransferases (DNMTs) in oocytes and early embryos. Biochimie 116, 103–113. https://doi.org/10.1016/j.biochi.2015.06.019

van Dijk, E.L., Auger, H., Jaszczyszyn, Y., Thermes, C., 2014. Ten years of next-generation sequencing technology. Trends Genet. 30, 418–426. https://doi.org/10.1016/j.tig.2014.07.001

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., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Miklos, G.L.G., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V.D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.-R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z.Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center,

A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A.D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., Zhu, X., 2001. The Sequence of the Human Genome. Science 291, 1304–1351. https://doi.org/10.1126/science.1058040

Visel, A., Rubin, E.M., Pennacchio, L.A., 2009. Genomic views of distant-acting enhancers [WWW Document]. Nature. https://doi.org/10.1038/nature08451

Waddington, C.H., 2012. The epigenotype. 1942. Int J Epidemiol 41, 10–13. https://doi.org/10.1093/ije/dyr184

Waddington, C.H., 1968. Towards a Theoretical Biology. Nature 218, 525–527. https://doi.org/10.1038/218525a0

Waddington, C.H., 1942. The epigenotype. Endeavour 1, 18–20.

Walker, C.G., Littlejohn, M.D., Meier, S., Roche, J.R., Mitchell, M.D., 2013. DNA methylation is correlated with gene expression during early pregnancy in Bos taurus. Physiological Genomics 45, 276–286. https://doi.org/10.1152/physiolgenomics.00145.2012

Walker, C.G., Meier, S., Hussein, H., McDougall, S., Burke, C.R., Roche, J.R., Mitchell, M.D., 2015. Modulation of the immune system during postpartum uterine inflammation. Physiol. Genomics 47, 89–101. https://doi.org/10.1152/physiolgenomics.00098.2014

Walker, C.G., Mitchell, M.D., 2013. Reproductive technologies for the future: a role for epigenetics. Animal Production Science 53, 954–964. https://doi.org/10.1071/AN12267

Walton, E.L., Francastel, C., Velasco, G., 2014. Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. Biology (Basel) 3, 578–605. https://doi.org/10.3390/biology3030578

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., Schübeler, D., 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. 39, 457–466. https://doi.org/10.1038/ng1990

Wei, W., Yu, X.-D., 2007. Comparative Analysis of Regulatory Motif Discovery Tools for Transcription Factor Binding Sites. Genomics Proteomics Bioinformatics 5, 131–142. https://doi.org/10.1016/S1672-0229(07)60023-0

Williams, E.J., Fischer, D.P., Noakes, D.E., England, G.C.W., Rycroft, A., Dobson, H., Sheldon, I.M., 2007. The relationship between uterine pathogen growth density and ovarian function in the postpartum dairy cow. Theriogenology 68, 549–559. https://doi.org/10.1016/j.theriogenology.2007.04.056

Wingender, E., Dietze, P., Karas, H., Knüppel, R., 1996. TRANSFAC: A Database on Transcription Factors and Their DNA Binding Sites. Nucleic Acids Res 24, 238–241. https://doi.org/10.1093/nar/24.1.238

Wyatt, G.R., 1951. Recognition and estimation of 5-methylcytosine in nucleic acids. Biochem J 48, 581–584.

Yuan, X.-L., Gao, N., Xing, Y., Zhang, H.-B., Zhang, A.-L., Liu, J., He, J.-L., Xu, Y., Lin, W.-M., Chen, Z.-M., Zhang, H., Zhang, Z., Li, J.-Q., 2016. Profiling the genome-wide DNA methylation pattern of porcine ovaries using reduced representation bisulfite sequencing. Sci Rep 6. https://doi.org/10.1038/srep22138

Zhang, Z., Cao, Y., Zhai, Y., Ma, X., An, X., Zhang, S., Li, Z., 2018. MicroRNA-29b regulates DNA methylation by targeting Dnmt3a/3b and Tet1/2/3 in porcine early embryo development. Development, Growth & Differentiation 60, 197–204. https://doi.org/10.1111/dgd.12537

Zhou, Y., Xu, L., Bickhart, D.M., abdel Hay, E.H., Schroeder, S.G., Connor, E.E., Alexander, L.J., Sonstegard, T.S., Van Tassell, C.P., Chen, H., Liu, G.E., 2016. Reduced representation bisulphite sequencing of ten bovine somatic tissues reveals DNA methylation patterns and their impacts on gene expression. BMC Genomics 17, 779. https://doi.org/10.1186/s12864-016-3116-1

Zwart, R., Sleutels, F., Wutz, A., Schinkel, A.H., Barlow, D.P., 2001. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev. 15, 2361–2366. https://doi.org/10.1101/gad.206201

8 Acknowledgement

The research work in this thesis was carried out at SLU Global Bioinformatics Centre, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. A financial support as a scholarship was granted for four years by University of the Punjab, Lahore, Pakistan. I am thankful to both prestigious universities from the cores of my heart.

Furthermore, this work was partially performed with the financial support from the EU project "PROLIFIC" (Grant KBBE 311776-PR and PH). I am thankful to Gavin Kelsey at the Babraham institute, UK for providing their platform for generating bisulfite sequence data. I am grateful to SLU Global Bioinformatics Centre, SLU, Uppsala (http://sgbc.slu.se) and SLU Bioinformatics Infrastructure (SLUBI) for management and processing of the sequencing data. I would like to acknowledge UPPMAX, Uppsala University for computational resources and EU COST Action FA1201 (EPICONCEPT).

This is not the end of world rather not end of journey to explore the knowledge, it is just beginning to new horizons of learning. There are many events and people who came across to support me in this journey. I would like to express my sincere gratitude to all who shared this journey with me and made it memorable. I learnt a lot personally and professionally with their company and enhanced my exposure.

I would like to acknowledge my supervisor group: **Erik Bongcam-Rudloff, Göran Andersson, Shumaila Sayyab and Patrice Humblot-** without you, none of this work would have been realized. Your throughout support, guidance and inspiration always kept me in high spirits. Your valuable discussions and ideas made the work more meaningful and grow stronger. It has been an honour for me to work with such a great professionals.

Erik my main supervisor, thank you for providing me the opportunity to be part of your group and accepted me as a PhD student. Your doors always remained open for discussions, guidance and support. You will be my pride forever. **Göran**, you are a great person and matchless scientist. You know the art of inculcating knowledge and inspiring the students. I will always feel proud to be your student. **Patrice,** your valuable scientific discussions and guidance have great impact on my work, I am proud of you. **Shumaila**, your guidance will be remembered.

Special thanks to all my co-authors especially Denise Laskowski, Tom Van Schaik, Yongzhi Guo and Adnan Niazi. Guo has done all the lab work including sample collection, purification of cells and sequence library preparation. Adnan was always there for bioinformatics support whenever it was needed.

I would also like to thank Erik's Bioinformatics Group especially for weekly group meetings in which we could share our findings / progress and seek further guidance. Thanks to Juliette, Tomas, Oskar, Hadrian, Mojtaba, Aleksei and Adnan.

I would also like to thank DJ and Lotta, being the head of department, you always encouraged, guided and supported me. Special thanks to Helena Pettersson for administrative support and Cano Merkan for IT support.

Further, I would like to thank all my colleagues at SLU especially Fernando Lopespinto, Erling Strandberg, Hossein Jorjani, Valentina Palucci, Freddy Fikse, Sofia Nyman, Agnese Viluma, Suvi Makelainen, Jovana Marjanovic, Maulik Upadhyay and Susanne Eriksson.

Special thanks to Sweden and Pakistani friends living in Uppsala for their jovial company; Ahsan, Akhtar, Aqib, Asim, Durani, Ghafoor, Haseeb, Imran, Javeed, Salman, Shaban, Sohaib, Moneeb, Mudassar, especially Shahid Manzoor and Khuram for their guidance.

My colleagues at Punjab University Gujranwala Campus, Pakistan always supported me in their prayers. Special thanks to Dr. Kamran Mujahid, Dr. Ehsan Malik, Dr. Shahid Kamal, Dr. Ahmad Saeed Akhtar and Dr. Shahid Manzoor. I am grateful to all my teachers throughout in my career from my school level to university level, I am going to dedicate my research work to all my teachers.

Finally, I would like to thank my family, wife and children who suffered a lot during this period. I am proud of you.