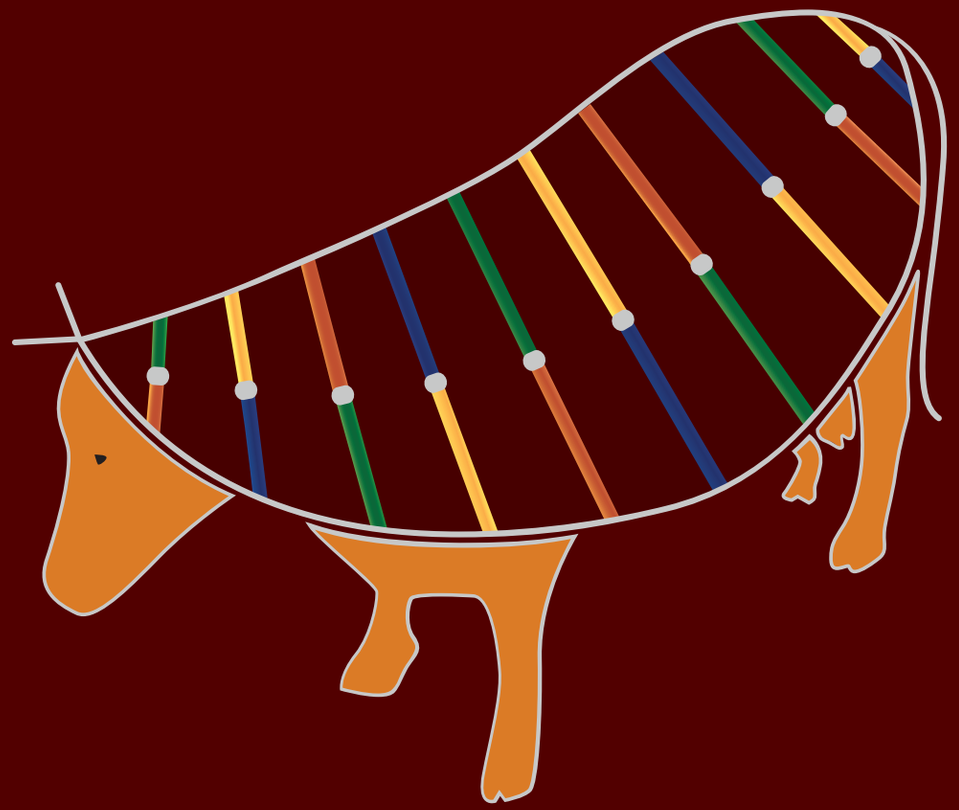
Genomic regulation of oestrous behaviour in dairy cows Arun Kommadath 2012

Genomic regulation of oestrous behaviour in dairy cows



Arun Kommadath

Propositions

1. Quantitative trait associated gene expression analysis answers different questions than the classical differential expression analysis in microarray based studies (this thesis).

2. Communication between brain areas orchestrates sexual behaviour (this thesis).

3. In bioinformatics analyses, biological interpretation is more important than statistical significance.

4. The trend that technological advances in science outpace legal regulatory frameworks governing them should be reversed.

5. It is from our children that we fully realize what we mean to our parents.

6. In science as in life, what seemed the absolute truth yesterday can be refuted today, and what seems farfetched today can be conventional tomorrow.

Propositions belonging to the thesis entitled, 'Genomic regulation of oestrous behaviour in dairy cows'.

Arun Kommadath
Wageningen, 24th February 2012

Genomic regulation of oestrous behaviour in dairy cows

Thesis committee

Thesis supervisor
Prof. dr. Mari A. Smits
Personal chair at Animal Breeding and Genomics Centre
Wageningen University

Prof. dr. Martien A.M. Groenen
Personal chair at Animal Breeding and Genomics Centre
Wageningen University

Thesis co-supervisors
Dr. Marinus F.W. te Pas
Senior Researcher at Animal Breeding and Genomics Centre
Wageningen University

Other members
Prof. dr. Bas Kemp, Wageningen University
Prof. dr. Jaap Keijer, Wageningen University
Prof. dr. Dirk-Jan de Koning, Swedish University of Agricultural Sciences, Uppsala, Sweden
Prof. dr. Eckhard Wolf, Ludwig-Maximilians University, Munich, Germany

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

Genomic regulation of oestrous behaviour in dairy cows

Arun Kommadath

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof.dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday February 24 2012
at 1.30 p.m. in the Aula.

Kommadath, A.
Genomic regulation of oestrous behaviour in dairy cows.

PhD thesis, Wageningen University, The Netherlands (2012)
With references, with summaries in English and Dutch

ISBN 978-94-6173-193-7

Dedicated to my beloved parents, wife and son

Abstract

Kommadath, A. (2012). Genomic regulation of oestrous behaviour in dairy cows. PhD thesis, Wageningen University, the Netherlands

Concurrent to the impressive improvement achieved over the last few decades for the trait of milk production in dairy cows was a steady decline in several fertility traits including oestrous behaviour (OB). An understanding of the genomic regulation of OB, which is currently lacking in dairy cows, will present new opportunities for managing this trait to help improve fertility. The research described in this thesis therefore aimed to achieve this understanding by studying gene expression in the bovine anterior pituitary (AP) and four brain areas (amygdala, hippocampus, dorsal hypothalamus and ventral hypothalamus) that are involved in regulating OB. A series of different analyses were performed that included model based association of gene expression with OB scores, gene co-expression and differential expression. In the association analyses, the identified genes included those previously not known to be related to OB associated processes as well as several genes expressed in mid-cycle that may have a function in the proper expression of OB at the next oestrus. Expected genes known to be involved in OB associated processes like socio-sexual behaviour (e.g. *OXT, AVP, GABRA6, HTR2A, DRD2*), anxiety, stress and feeding motivation (e.g. *POMC*, *MCHR1, TTR*) were found along with genes associated with nervous system related processes (e.g. *CHRM1*, *CHRM3, CHRNA5, CTLA4, IL1RL1, MARCO*), suggesting a link between neuronal plasticity and OB. In the co-expression analyses, biological terms found common to several OB correlated consensus modules included general cellular processes like oxidative phosphorylation, ribosome and biosynthetic processes, indicating increased transcription and protein synthesis. These processes are primary events in the activation of neuronal cells and pathways involved in female reproductive behaviour and they precede the oestrogen driven expansion of dendrites and synapses. Hub genes within the OB correlated modules (e.g. *NEFL*, *NDRG2*, *GAP43*, *THY1*, *TCF7L2 etc.*) are strong candidates among genes regulating OB expression. Further, we showed the phenomenon of chromosomal regional regulation of transcription to exist in the bovine genome. To conclude, this study has revealed important new aspects of the genomic regulation of OB in dairy cows with the key findings presented within the framework of the GAPPS modules. The new knowledge could be used to optimize fertility of dairy cows by aiding to improve existing or to devise novel reproductive management tools like diagnostic tools to determine the reproductive health, energy and fertility status of the cow, oestrus detection tools and so on.

Contents

1

General introduction

1.1 Dairy cow fertility

Dairy farming is one of the basic agricultural based occupations providing direct and indirect livelihoods to scores of people around the world. Milk, the main product of dairy farming, is a highly valued and nutritious food source which is either consumed as such or forms the raw material from which several dairy based products are processed. Today, dairy processing is a high end industry, which along with dairy farming, contributes considerably to the economy of several nations. The European Union (EU) holds a dominant position in the world dairy market. The dairy exports from the EU-25 amounted to 21 billion Euros in the year 2009, which was over 70% of the total international trade in dairy products (Tacken *et al.*, 2009). To keep the dairy industry profitable is therefore important to the economy of the EU as well as several major dairy producing nations in the world like India, US, Canada, Australia and New Zealand.

The average milk yield potential of dairy cows worldwide has more than doubled in major dairy breeds over the past half century thanks to breeding programmes that intensively selected animals with a high breeding value for milk production traits. During the last 25 years (1985-2010), average 305-day milk production of pedigree Holstein-Friesian cows in the Netherlands has increased from 5765 kg to 8804 kg (CRV, 2010). Starting at about the same period as that when rapid increases in milk yield potential were being achieved, a declining trend was noticed for several fertility traits in the dairy cow (Royal *et al.*, 2000b, Roxstrom *et al.*, 2001, Veerkamp *et al.*, 2003). This led to the belief that the two events were linked. It later became evident that a negative genetic correlation exists between production traits and fertility traits (Pryce and Veerkamp, 1999, Royal *et al.*, 2000a, Roxstrom *et al.*, 2001, Berry *et al.*, 2003). However, genetics only contributes to approximately half of the progress in milk yield while the rest is attributed to improved environmental factors like better nutrition, housing, health and management (Pryce and Veerkamp, 2001).

Similar to milk production, the factors affecting fertility traits are multiple and interacting. In addition to genetics, the declining trend in fertility traits is partly due to the type of dairy management and production system followed and partly to nutritional demands and physiological adaptations of the cow in response to high milk production (Butler, 2000, Lucy, 2001, Butler, 2003, Lucy, 2003). Darwash *et al.* (1997) broadly defined fertility in dairy cows as the accomplishment of pregnancy at the desired time which encompasses “the inherent capacity of the cow to establish ovarian function postpartum, to show overt oestrus, and to conceive and maintain pregnancy when inseminated at the appropriate time”. Complementary

to this, Royal *et al.* (2000a) defined subfertility as "any condition leading to a failure to establish pregnancy following completion of uterine involution". Here, involution refers to the process by which the uterus returns to its normal pre-pregnant state at 40–50 days after parturition. Reproductive success of the dairy cow is linked to its body energy reserves (Roche, 2006), which may be compromised especially during early and peak lactation periods. The metabolic responses in the cow to low energy reserves involve signalling molecules and hormones which are central not only to the somatotrophic axis that regulates the partitioning of energy/nutrients but also to the hypothalamic-pituitary-ovarian axis that regulates reproductive function (Chagas *et al.*, 2007, Garnsworthy *et al.*, 2008) (Figure 1.1).

It is encouraging to note that dairy breeding programmes in the Scandinavian countries successfully managed to maintain fertility levels in dairy cows along with significant increases in their milk yield potential (Philipsson and Lindhé, 2003, Pryce *et al.*, 2004). This success, in spite of the low heritability of fertility traits, typically below 0.05 (Veerkamp and Beerda, 2007), and their negative correlation with production traits, was achieved by means of breeding programmes employing multi-trait selection indices that included both fertility and milk production traits (Pryce *et al.*, 2004). Similar breeding programmes initiated in other countries over the past decade have successfully arrested and subsequently begun improving fertility traits in dairy cows, although the rate of progress has been slow. Strategies employing genomics techniques can help improve the rate of genetic progress by allowing selection at an earlier stage, even as early as foetal life. Marker assisted selection is applicable in commercial dairy breeding and is particularly useful for traits that are difficult to measure, exhibit low heritability, and/or are expressed late in development (Spelman and Garrick, 1997). The most abundant of these markers, called SNPs (Single Nucleotide Polymorphisms), hold great promise to improve breeding strategies especially with the recent advances in genomic selection (Hayes *et al.*, 2009). Although marker based selection techniques do not require information on the actual genes that regulate the trait under selection, it would be advantageous to know what these genes are and their functions in order to understand how the trait is regulated. Research directed at understanding the regulation of fertility traits could lead to the development of better management practices and tools to optimize fertility of dairy herds (Veerkamp and Beerda, 2007, Beerda *et al.*, 2008), for example, by aiding development of diagnostic tools to determine the reproductive health, energy and/or fertility status of the cow, or of tools to detect oestrus and so on.

1.2 Hormonal regulation of reproductive cycle in female mammals

The hormonal regulation of the reproductive or oestrous cycle is well studied in mammalian species and several hormone-driven mechanisms in the brain that influence mammalian oestrous behaviour or female sexual behaviour are known (Pfaff, 2005). Key hormones regulating the oestrous cycle in mammals are: gonadotropin-releasing hormone (GnRH) produced in the hypothalamus; follicle-stimulating hormone (FSH) and luteinizing hormone (LH) produced in the anterior pituitary; oestrogen (E2) and progesterone (P4) produced in the ovaries; and prostaglandin produced in the uterus (Downey, 1980). Some of these hormones also influence the expression of oestrous behaviour by targeting specific brain areas, directly or indirectly. For example, E2 dependent gene transcription in the hypothalamus and midbrain establishes the neural circuitry required for lordosis behaviour in female rats (Pfaff, 2005), which is characterised by a vertebral dorsiflexion to permit mating and fertilization.

Organs other than the brain, including liver, adipose, and muscle, link reproduction and metabolism through the production of several metabolites and hormones, with the growth hormone (GH) being the principal coordinator of these activities (Lucy, 2003). GH stimulates these organs to release metabolites such as glucose and non-esterified fatty acids and also the pancreas and gut, to promote the release of

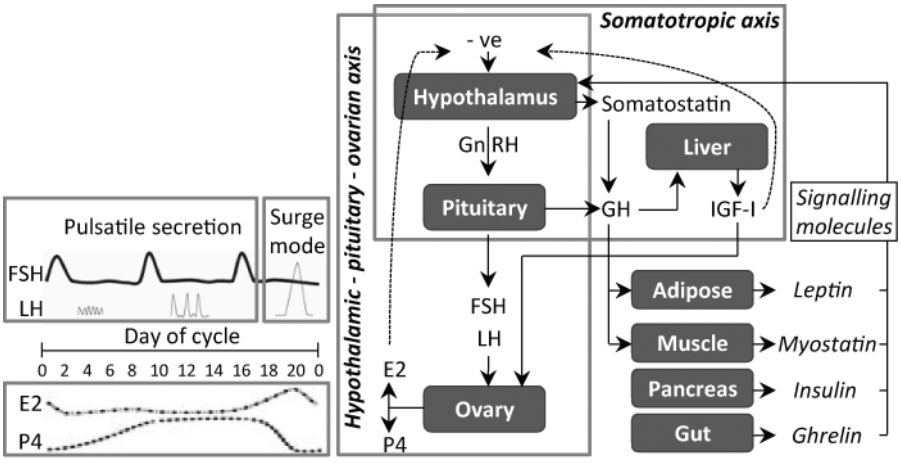


Figure 1.1 Interactions between regulatory mechanisms that control the hypothalamic-pituitary-ovarian axis and the somatotropic axis (adapted from Chagas *et al.* (2007) and Garnsworthy *et al.* (2008)). The negative (-ve) feedback by E2 and IGF-I on the release of GnRH by the hypothalamus is indicated. The hormonal profiles for FSH, LH, E2 and P4 across the oestrous cycle are also depicted.

several hormones and metabolites like insulin, insulin-like growth factor 1 (IGF-I), ghrelin, and leptin that signal back to the centres in the brain that control appetite and reproduction (Lucy, 2003). Figure 1.1 represents the interaction between regulatory mechanisms that control the hypothalamic-pituitary-ovarian axis and the somatotropic axis thereby linking reproductive process with nutritional and metabolic inputs (adapted from Chagas *et al.*(2007) and Garnsworthy *et al.*(2008)).
The oestrous cycle has two phases - follicular phase which ends with a surge release of gonadotrophins (LH and FSH) from the anterior pituitary accompanied by ovulation; and a luteal phase in which P4 is secreted from the corpus luteum. The oestrous cycle is driven primarily by GnRH, which is secreted into the hypophyseal portal system in a pulsatile fashion from the terminals of GnRH neurons originating around the preoptic area and mediobasal hypothalamus at the base of the brain (Garnsworthy *et al.*, 2008). These neurons represent the final output component of a neuronal network that integrates multiple environmental and internal factors to regulate reproductive hormone secretion (Herbison, 1998). The secretion of GnRH from the hypothalamus into the pituitary portal vessels is governed by two distinct control mechanisms: the pulsatile and surge mode centres. The former, called the GnRH pulse generator, plays a role as the master regulator of the reproductive function and its activity is regulated by neuropeptides like neuropeptide Y, cholecystokinin-octapeptide and melanocortins (Okamura and Ohkura, 2007). Whereas FSH induces follicular development, both FSH and LH together stimulate the production of hormones E2 and P4 in the ovary. The hormones E2 and P4, via a negative feedback action on the hypothalamic-pituitary-ovarian axis, help maintain low levels of FSH and LH through most of the oestrous cycle. However, towards proestrus, the rising E2 levels have a positive feedback on LH secretion, resulting in the LH surge that causes ovulation. This pre-ovulatory rise in E2 is responsible for the hormonally based components of oestrous behaviour (Houpt, 2011). The role of GnRH is to synchronize reproductive behaviour with the ovulatory surge of LH (Mong and Pfaff, 2003). E2 is one of the principal determinants of GnRH neuron functioning and, acting as a classic homeostatic feedback molecule between gonad and brain, is critical in enabling GnRH neurons to exhibit fluctuating patterns of biosynthetic and secretory activity (Herbison, 1998). Gore and Roberts (1995) reported that an increase in GnRH primary transcript expression occurred early in proestrus, and Petersen *et al.* (1996) found that this increased transcription was initiated by an E2 dependent signal. GnRH neurons are also affected by signals from neurotransmitters like glutamate that act via the *N*-methyl *D*-aspartate receptor (NMDAR), which in itself is regulated by E2 (Gore, 2001).

1.3 Oestrous behaviour

Female sexual behaviour or oestrous behaviour (OB) is an active process comprising copulation as well as critical non-copulatory behaviours that precede and facilitate mating. In many mammals, sexual behaviour only occurs at a specific stage of a female's oestrous cycle, in synchrony with the timing of ovulation. Because of the temporal and adaptive links between female mating and ovulation, it is not surprising that these two processes are usually regulated by the same factors, E2 and P4 (Kauffman *et al.*, 2010). In cows, during the period of oestrus (phase of oestrous cycle when the female exhibits OB and is sexually receptive to the male), the following behavioural signs are exhibited: flehmen, restlessness, sniffing the vulva of another cow, resting with chin on the back of another cow, mounting other cows and standing heat. Scoring based on observation of these signs has been used as a way to detect and quantify the intensity of oestrus (van Eerdenburg *et al.*, 1996, Roelofs *et al.*, 2005). The proper detection of oestrus is important to ascertain the optimum time for insemination of the cow. Intensive oestrus signalling relates to high reproductive performance (Garcia *et al.*). However, the expression of OB has decreased both in duration and intensity over generations of cows coinciding with the period in which rapid increases in milk yield potential was achieved (Lopez *et al.*, 2004). The duration of oestrus, defined as the time between the first and last recorded standing event, has reduced from an average of 14.9 h in Friesian cattle in 1976 (Esslemont and Bryant, 1976) to about 7 h in the modern Holstein cows (Lopez *et al.*, 2004). Short heat periods and the absence of clear behavioural signs of heat make farmers fail to detect heat or misjudge the optimum time of insemination of their cows, resulting in financial losses due to reduced conception rates and increased calving intervals. Therefore, better techniques to detect or improve the expression of OB can help improve fertility in dairy cows.

1.4 Brain regions regulating oestrous behaviour

Underlying all socio-sexual behaviour is the fundamental arousal of brain and behaviour (Pfaff, 2005). Studies in rodents show that female reproductive behaviour in mammals centres on E2 actions in the hypothalamus (Pfaff, 2005) with a significant role for the ventromedial hypothalamus (VMH) followed by others like the arcuate nucleus, pre-optic area, paraventricular nucleus and suprachiasmic area nuclei (Etgen, 1987, Richmond and Clemens, 1988, Matuszewich *et al.*, 2000, Richards *et al.*, 2010). E2 was shown to up-regulate expression of proteins implicated in stimulating neuronal plasticity and altering signal transduction in the VMN in ovariectomized rats, which could be involved in

priming female sexual behaviour (Mo *et al.*, 2008). The brain regions used by E2 to activate GnRH neurons are the antero-ventral periventricular nucleus in rats and the mediobasal hypothalamus in mice, monkeys and ewes (Herbison, 1998). Based on extensive literature review, Newman (1999) postulated the existence of a social behaviour network that included six limbic system areas that are reciprocally interconnected anatomically: lateral septum, medial preoptic area, anterior hypothalamus, VMH, midbrain and central amygdaloid nuclei of the stria terminalis. Limbic regions, such as the hippocampus and amygdala, that have connections to the hypothalamic–pituitary–adrenal axis, may be targets for E2's effects on anxiety and depression (Walf and Frye, 2006) and have functions related to sexual behaviour and associated emotional responses (Gallagher and Chiba, 1996, Salamon *et al.*, 2005). The importance of the amygdala in regulating sexual behaviour is conclusive from several studies. The medial nucleus of amygdala is one of the major E2 concentrating sites in the brain (Stumpf, 1970) and has direct input into the core of the VMH (Krettek and Price, 1977) thus suggesting that it can influence lordosis by directly acting on the VMH. Metabolic conversion of P4 to 3α,5α-tetrahydroprogesterone, which is important for P4 facilitated sexual receptivity of rats and hamsters takes place in the ventral tegmental area of the brain (Frye and Vongher, 2001, Frye *et al.*, 2008). Thus the limbic structures of the brain along with the pituitary gland that forms part of the hypothalamic–pituitary–ovarian axis are among the main centres that regulate OB.

1.5 Neuroendocrine and genomic regulation of oestrous behaviour

A recent review (Boer *et al.*, 2010) discusses the current knowledge on the hormonal and genomic regulation of OB. Most of our current understanding of the genomic regulation of OB comes from studies in rodents. E2 is among the most important endocrine regulators of reproduction and are ligands for either E2 receptor alpha (ERα) or beta (ERβ). The molecular mechanisms by which E2 affects the response of target tissues may be based on genomic or non-genomic actions.

Genomic actions of E2 are those wherein its binding to either of the ER isoforms results in regulation of transcription of genes. In the central nervous system, several genes are also regulated in this manner including certain norepinephrine receptors to promote lordosis behaviour (Mong and Pfaff, 2004). Norepinephrine signalling to VMH neurons presents one way for linking generalized arousal pathways, through hormone sensitive signal transduction channels and ionic currents, to hormone-sensitive reproductive behaviours (Lee and Pfaff, 2008). Pfaff *et al.* (2008) provided an overview of genes whose transcript levels are affected by

E2 acting in the hypothalamic or preoptic neurons through its receptors. These E2 responsive genes whose products foster female sex behaviours include *GnRH* and its receptor (*GnRHR*), P4 receptor (*PGR*), adrenergic and muscarinic receptors, oxytocin (*OXT*) and its receptor (*OXTR*), encephalin and opioid receptors and nitric oxide synthase (Pfaff *et al.*, 2008, Vasudevan and Pfaff, 2008). It was reported that serotonin has a role in mediating hypothalamic control of sexual behaviour in E2-primed ovariectomized rats and that LH-releasing hormone has opposing effects on the control of lordosis behaviour (Foreman and Moss, 1978). Similarly, dopamine has an inhibitory role in the control of sexual receptivity (Everitt and Fuxe, 1977). Suppression of Prostaglandin D synthetase by E2 in the preoptic area may account for increased generalized arousal and movement evident in E2-treated females (Mong *et al.*, 2003). Salamon *et al.* (2005) demonstrated that amygdalar regulation of the female sexual cycle is mediated by E2 related signalling molecules which exert their influences on ovulation and sexual behaviour via coupled nitric oxide release.

The non-genomic actions of E2 are rapid effects, generally initiated at the plasma membrane by the natural ligand, 17β-oestradiol, resulting in the activation of signal transduction pathways within target cells (Vasudevan and Pfaff, 2008). Reproductive behaviour in female rodents is dependent both on gene expression as well as kinase activation and changes in neuronal excitability by E2, thus supporting the hypothesis that rapid signalling cascades initiated at the membrane by E2 can lead to changes in transcription from E2 response element-containing genes (Vasudevan and Pfaff, 2008).

Based on information on gene inductions by E2 acting in the brain, effect on downstream genes and their physiological routes of action, Mong and co-workers (Mong *et al.*, 2003, Mong and Pfaff, 2004) suggested a set of modules which help to account for the causal relations between sex hormones and female mating behaviours. These modules abbreviated as GAPPS are described below:

1. **Growth** (G) response, dependent on hormones E2 and P4 that allows for a greater range of neuronal input and output connections for the hormone-facilitated, behaviour-directing hypothalamic neurons.
2. **Amplification** (A) of the E2 effect by P4, in part through the downstream genes like *ERα*, *ERβ*, *OXT* and *OXTR*.
3. **Preparation** (P) of the female for mating behaviour through indirect behavioural means – the reduction of anxiety and a partial analgesia effected through *OXT* and *OXTR* as well as the genes for encephalin and its receptors.
4. **Permission** (P) phase which refers to the permissive actions by hypothalamic neurons for the mating behaviour to occur. Here, neurotransmitter receptor

induction by E2 permits the neural circuit for lordosis behaviour to be activated. Key players here are the α1-adrenoceptor and muscarinic acetylcholine receptors in the ventromedial nucleus (VMN) of the hypothalamus.

5. **Synchronisation** (S) of mating behaviour with ovulation. Here E2 inducts the genes for GnRH, GnRH receptor and LH releasing hormone under positive feedback to synchronize OB with the ovulatory surge of LH which in turn triggers ovulation.

Though the GAPPS modules were based on studies mainly in rodents, some of the processes within these modules could very likely overlap with processes involved in the regulation of OB in dairy cows. However, the extent of this overlap has not yet been studied.

1.6 Gene expression and its regulation

The process of gene expression is universal to all living organisms and refers to all steps involved in the creation of a functional gene product, either a protein via messenger RNA (mRNA) or other functional RNA like ribosomal RNA, transfer RNA or microRNA (miRNA). In eukaryotes, protein-coding gene expression begins with gene transcription to produce a primary transcript called pre-mRNA, which is then subjected to various post transcriptional modifications like 5' capping, 3' cleavage, polyadenylation and splicing to create the mature mRNA. This mRNA is then transported out of the nucleus to the cytoplasm where it is translated to protein. The biosynthesis of proteins ends with post-translational modifications. The entire process is complex and requires fine regulation at every step.

The best-studied level of gene regulation is at the DNA sequence level which involves two types of elements: cis-acting and trans-acting elements. Promoters, enhancers and silencers belong to the former whereas DNA-binding transcription factors (TFs), cofactors, miRNAs, chromatin-remodelling systems and RNA polymerases are part of the latter. The roles and mechanism of action of all these gene regulatory factors are well-studied and therefore not reviewed here. The most numerous of all regulatory factors are the trans-acting TFs and miRNAs. The miRNAs are a recently discovered class of non-protein-coding transcripts about 19-23 nucleotides long that act by binding to complementary sequences on target mRNAs. Several reviews elaborate the similarities and differences between TFs and miRNAs and their gene regulatory mechanisms (Chen and Rajewsky, 2007, Hobert, 2008). While TFs act at gene transcriptional level to positively or negatively regulate transcription, miRNAs act at post-transcriptional level, acting mostly through repression of the mRNAs (Pillai *et al.*, 2007, Vasudevan *et al.*, 2007). Individual TFs and miRNAs can control several tens to hundreds of target genes and

conversely, individual genes may be controlled by a combination of TFs and miRNAs.

Other levels of gene regulation in addition to sequence level are at the chromatin level and the nuclear level (van Driel *et al.*, 2003). Gene regulation at the chromatin level involves changes in chromatin structure that are controlled by factors such as histone modification and DNA methylation. This level of regulation acts usually at the level of gene clusters, allowing switches between different functional states. Finally, gene regulation at the nuclear level involves the dynamic spatial organization of the genome inside the nucleus. Evolutionary constraints have shaped the genome organisation at all three levels to allow for efficient control of gene expression (Castillo-Davis *et al.*, 2002, Eisenberg and Levanon, 2003, Vinogradov, 2004, Janga *et al.*, 2008, Nie *et al.*, 2010a, Nie *et al.*, 2010b). A manifestation of this is the finding that genes are also clustered on chromosomes based on their activity. For example, it has been shown that there exist regions on the chromosomes with highly expressed genes called RIDGEs (Regions of Increased Gene Expression) and others with lowly expressed genes called anti-RIDGEs (Caron *et al.*, 2001, Versteeg *et al.*, 2003, Gierman *et al.*, 2007) compared to the average expression level across the genome. Interestingly, certain genomic features observed for RIDGEs were in striking contrast with those for anti-RIDGEs (Versteeg *et al.*, 2003). RIDGEs were found to be gene dense, GC rich and Short Interspersed Element (SINE) repeat rich and mostly harboured genes with shorter than average intron sizes. In contrast, anti-RIDGES showed low gene density, were AT rich and Long Interspersed Element (LINE) repeat rich and had comparatively higher intron sizes. Based on their findings, Versteeg *et al.* (2003) postulated that RIDGEs globally govern the expression levels of their embedded genes and that this higher level regulation of gene transcription was dependent on factors that act on chromosomal domains like chromatin conformation and position in the nucleus. The studies mentioned above have shown that the phenomenon of chromosomal regional regulation of transcription occurs in several mammalian species and also chicken. It is therefore likely to occur in bovines too, though not yet proven.

1.7 Gene expression measurement and analysis

Gene expression is the most fundamental level at which an organism's genotype dictates its phenotype. A measure of gene expression is thus quite valuable and finds multiple applications in science and medicine. Technologies for gene expression measurement have been evolving rapidly over the past decade with conventional techniques like Northern blotting and Reverse transcription polymerase chain reaction (RT-PCR) giving way to high throughput genome wide

approaches like Serial analysis of gene expression (SAGE), DNA microarrays and more recently Next-generation sequencing technologies like RNA-seq.

Conventionally, DNA microarray technology was used to identify differentially expressed genes in cells/tissues/organs between groups of individuals belonging to contrasting classes based on, for example, physiological stages, disease status, stages in cell cycle and such qualitative attributes. However, when the trait of interest is quantitative (e.g. OB), the grouping of individuals into qualitative classes, ignoring the continuous scale, is indistinct and dilutes the available information. In such cases, the association between gene expression and phenotypic trait is better analysed using the individual quantitative measurements (Reiner-Benaim *et al.*, 2007). Recent studies on microarray gene expression data have successfully linked genes to quantitative traits of interest by correlation, linear regression or complex regression models (Blalock *et al.*, 2004, Jia and Xu, 2005, Qu and Xu, 2006, Jia *et al.*, 2008). Microarray technology, though a genome-wide approach, can only measure transcripts for which a probe has been designed for on the array. Further, the probes need to be well annotated to genes for use in post analysis procedures like gene ontology and pathway analysis. While interpreting the results of these post analysis procedures, it is important to keep in mind that although useful, there is an inherent bias to be expected in these procedures. This is due to the unbalanced information available for gene functions. Since a lot of research has been done on cancer and immunology, there is a lot known about gene functions related to these. In comparison, not much is known about genes with functions related to the nervous system and behaviour. The recent whole genome assemblies available for several species including livestock have contributed to improved gene annotation. This has greatly enhanced the value of microarray analysis because a lot more is now known about genes for which probes have been designed on the microarrays. For the domestic cow (*Bos taurus*), considered a model for ruminants, genome assemblies were published by two different groups in April 2009: the Btau 4.0 assembly (Liu *et al.*, 2009) and the UMD 2.0 assembly (Zimin *et al.*, 2009). The bovine genome contains about 2,87 Gbp, of which over 90% has been mapped to 29 autosomes and the X chromosome. The coverage for the bovine Y chromosome is significantly lower and work in underway to improve this. Of an estimated 22,000 protein-coding genes on the bovine genome, a core set of 14,235 orthologs were shown to be shared among 7 mammalian species (The Bovine Genome Sequencing and Analysis Consortium *et al.*, 2009). Thus, gene expression studies benefit from comparative genomics where information on orthologous genes from a better studied species is used in species which are less well studied.

1.8 Outline of this thesis

Most of the current knowledge on genomic regulation of OB is based on research done in rodents. Not much is known about genomic regulation of OB in dairy cows and other farm animals. The major aim of this thesis was to improve the understanding of genomic regulation of OB in dairy cows by studying gene expression in the bovine anterior pituitary and four brain areas (amygdala, hippocampus, dorsal hypothalamus and ventral hypothalamus) that are involved in regulating OB. This knowledge will present new opportunities for managing the trait of OB to improve fertility in dairy cows. The questions we addressed were: which genes and which biological processes associated with variations in the expression of OB in dairy cows and were likely to be involved in OB regulation. To answer these questions, a large scale microarray experiment was designed and performed, followed by a series of different analyses that included model based association of gene expression with OB scores, gene co-expression and differential expression. Genes thus found were subsequently studied in detail to identify biological processes or pathways they are part of.

The experiment involved recording of OB in 28 healthy Holstein Friesian cows from 30 days in milk (DIM) onwards till their time of sacrifice which varied between 77 and 139 DIM. The OB was quantified as heat scores and the scores from multiple consecutive cycles per cow were averaged to obtain the average heat score per cow. Of the 28 cows, 14 were sacrificed at the start of their oestrous cycle (day0) and 14 around mid-cycle (day12). Following sacrifice, tissue samples were collected from the anterior pituitary and four brain areas (Figure 1.2).

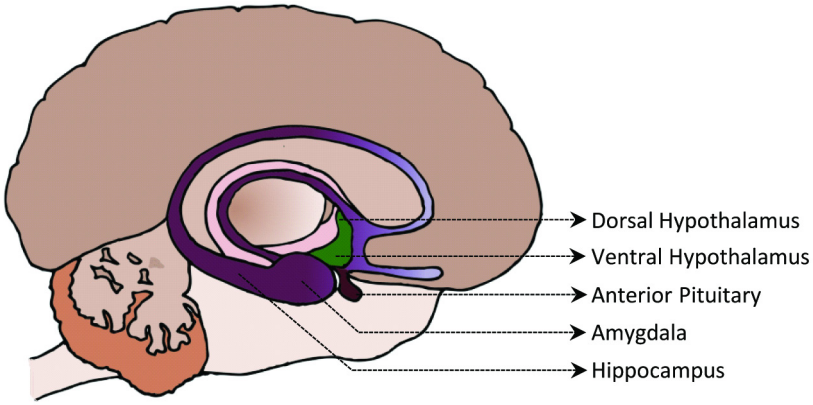


Figure 1.2 Schematic diagram of bovine brain showing selected brain areas and anterior pituitary.

Total cytoplasmic RNA extracted from individual tissue samples were hybridized on Bovine 24K oligonucleotide (70-mer) microarrays designed and produced by the Bovine Oligonucleotide Microarray Consortium (BOMC), USA (http://www.bovineoligo.org/). A total of 280 arrays (i.e. 28 cows x 5 tissues samples x 2 dye swaps) were prepared in a common reference design with the dye labels swapped between individual samples from each tissue and a reference sample consisting of equal proportions of pooled RNA from all tissues of all cows. Microarray data pre-processing and analysis was done using the LIMMA (Linear models for microarray data) package (Smyth, 2005) within Bioconductor project (Gentleman *et al.*, 2004) of R statistical programming language (http://www.r-project.org). Gene expression levels expressed as M-values (2 log differential-expression ratio of sample versus reference) were thus obtained for each tissue per cow. This data was used for all subsequent analyses described in the following chapters.

In Chapter 2, we aimed to identify and study those genes that were associated with OB among genes expressed in the bovine anterior pituitary either at the start of oestrous cycle or at the mid-cycle, or regardless of the phase of cycle. Gene expression was modelled as a function of the orthogonally transformed average heat score values using a Bayesian hierarchical mixed model on data from day 0 cows alone, day 12 cows alone and the combined data from day 0 and day 12 cows. A similar approach was followed for the analysis of the four bovine brain areas, the findings of which are reported in Chapter 3. Our objective in Chapter 4 was to identify genes and biological processes shared among the bovine anterior pituitary and four brain areas that act together to regulate OB. For this, we investigated networks of co-expressed genes between these tissues. In Chapter 5, we explored the existence of chromosomal regional regulation of transcription in the bovine genome and certain genomic features of genes within these regions and of the tissue-specific and housekeeping genes. Finally in Chapter 6, I discuss the main findings of all the previous chapters to arrive at a more complete picture of the genomic regulation of OB in dairy cows. I end the discussion with remarks on the steps needed to apply the new knowledge in practice to help optimize fertility in dairy cows.

References

Beerda B, Wyszynska-Koko J, te Pas MFW, de Wit AAC and Veerkamp RF 2008. Expression profiles of genes regulating dairy cow fertility: recent findings, ongoing activities and future possibilities. Animal 2, 1158-1167.

Berry DP, Buckley F, Dillon P, Evans RD, Rath M and Veerkamp RF 2003. Genetic parameters for body condition score, body weight, milk yield, and fertility estimated using random regression models. J Dairy Sci 86, 3704-3717.

Blalock EM, Geddes JW, Chen KC, Porter NM, Markesbery WR and Landfield PW 2004. Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. P Natl Acad Sci USA 101, 2173-2178.

Boer HMT, Veerkamp RF, Beerda B and Woelders H 2010. Estrous behavior in dairy cows: identification of underlying mechanisms and gene functions. Animal 4, 446-453.

Butler WR 2000. Nutritional interactions with reproductive performance in dairy cattle. Anim Reprod Sci 60-61, 449-457.

Butler WR 2003. Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. Livest Prod Sci 83, 211-218.

Caron H, van Schaik BDC, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus M-C, van Asperen R, Boon K, Voute PA, Heisterkamp S, van Kampen A and Versteeg R 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science 291, 1289-1292.

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV and Kondrashov FA 2002. Selection for short introns in highly expressed genes. Nat Genet 31, 415-418.

Chagas LM, Bass JJ, Blache D, Burke CR, Kay JK, Lindsay DR, Lucy MC, Martin GB, Meier S, Rhodes FM, Roche JR, Thatcher WW and Webb R 2007. Invited review: New perspectives on the roles of nutrition and metabolic priorities in the subfertility of high-producing dairy cows. J Dairy Sci 90, 4022-4032.

Chen K and Rajewsky N 2007. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet 8, 93-103.

CRV 2010. CRV Jaarstatistieken 2010. CRV, Arnhem, the Netherlands.

Darwash AO, Lamming GE and Woolliams JA 1997. Estimation of genetic variation in the interval from calving to postpartum ovulation of dairy cows. J Dairy Sci 80, 1227-1234.

Downey BR 1980. Regulation of the estrous cycle in domestic animals-- a review. Can Vet J. La revue veterinaire canadienne 21, 301-306.

Eisenberg E and Levanon EY 2003. Human housekeeping genes are compact. Trends Genet 19, 362-365.

Esslemont R and Bryant M 1976. Oestrous behaviour in a herd of dairy cows. Vet Rec 99, 472-475.

Etgen AM 1987. Inhibition of estrous behavior in rats by intrahypothalamic application of agents that disrupt nuclear binding of estrogen-receptor complexes. Horm Behav 21, 528-535.

Everitt BJ and Fuxe K 1977. Dopamine and sexual behaviour in female rats. Effects of dopamine receptor agonists and sulpiride. Neurosci Lett 4, 209-213.

Foreman MM and Moss RL 1978. Role of hypothalamic serotonergic receptors in the control of lordosis behavior in the female rat. Horm Behav 10, 97-106.

Frye CA and Vongher JM 2001. Ventral tegmental area infusions of inhibitors of the biosynthesis and metabolism of 3α,5α-thp attenuate lordosis of hormone-primed and behavioural oestrous rats and hamsters. J Neuroendocrinol 13, 1076-1086.

Frye CA, Paris JJ and Rhodes ME 2008. Exploratory, anti-anxiety, social, and sexual behaviors of rats in behavioral estrus is attenuated with inhibition of 3[alpha],5[alpha]-THP formation in the midbrain ventral tegmental area. Behav Brain Res 193, 269-276.

Gallagher M and Chiba AA 1996. The amygdala and emotion. Curr Opin Neurobiol 6, 221-227.

Garcia E, Hultgren J, Fällman P, Geust J, Algers B, Stilwell G, Gunnarsson S and Rodriguez-Martinez H Oestrous intensity is positively associated with reproductive outcome in high-producing dairy cows. Livest Sci 2011, 139(3):191-195.

Garnsworthy PC, Sinclair KD and Webb R 2008. Integration of physiological mechanisms that influence fertility in dairy cows. Animal 2, 1144-1152.

Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini A, Sawitzki G, Smith C, Smyth G, Tierney L, Yang J and Zhang J 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80.

Gierman HJ, Indemans MH, Koster J, Goetze S, Seppen J, Geerts D, van Driel R and Versteeg R 2007. Domain-wide regulation of gene expression in the human genome. Genome Res 17, 1286 - 1295.

Gore AC 2001. Gonadotropin-releasing hormone neurons, NMDA receptors, and their regulation by steroid hormones across the reproductive life cycle. Brain Res Rev 37, 235-248.

Gore AC and Roberts JL 1995. Regulation of gonadotropin-releasing hormone gene expression in the rat during the luteinizing hormone surge. Endocrinology 136, 889-896.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92, 433-443.

Herbison AE 1998. Multimodal influence of estrogen upon gonadotropin-releasing hormone neurons. Endocr Rev 19, 302-330.

Hobert O 2008. Gene regulation by transcription factors and microRNAs. Science 319, 1785-1786.

Houpt KA 2011. Domestic animal behaviour for veterinarians and animal scientists. Wiley-Blackwell, Oxford.

Janga SC, Collado-Vides J and Babu MM 2008. Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. P Natl Acad Sci USA 105, 15761-15766.

Jia Z and Xu S 2005. Clustering expressed genes on the basis of their association with a quantitative phenotype. Genet Res 86, 193-207.

Jia Z, Tang S, Mercola D and Xu S 2008. Detection of quantitative trait associated genes using cluster analysis. In Evolutionary computation, machine learning and data mining in bioinformatics (eds. E Marchiori and JH Moore), pp. 83-94, Springer-Verlag, Berlin Heidelberg.

Kauffman AS, Michael DB and Janice M 2010. Mammalian female sexual behavior and hormones. In Encyclopedia of Animal Behavior, pp. 355-369, Academic Press, Oxford.

Krettek JE and Price JL 1977. Projections from the amygdaloid complex to the cerebral cortex and thalamus in the rat and cat. J Comp Neurol 172, 687-722.

Lee AW and Pfaff DW 2008. Hormone effects on specific and global brain functions. J Physiol Sci 58, 213-220.

Liu Y, Qin X, Song X-Z, Jiang H, Shen Y, Durbin KJ, Lien S, Kent M, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley K, Weinstock G and Gibbs R 2009. *Bos taurus* genome assembly. BMC Genomics 10, 180.

Lopez H, Satter LD and Wiltbank MC 2004. Relationship between level of milk production and estrous behavior of lactating dairy cows. Anim Reprod Sci 81, 209-223.

Lucy MC 2001. Reproductive loss in high-producing dairy cattle: Where will it end? J Dairy Sci 84, 1277-1293.

Lucy MC 2003. Mechanisms linking nutrition and reproduction in postpartum cows. Reproduction (Cambridge, England) Supplement 61, 415-427.

Matuszewich L, Lorrain DS and Hull EM 2000. Dopamine release in the medial preoptic area of female rats in response to hormonal manipulation and sexual activity. Behav Neurosci 114, 772-782.

Mo B, Callegari E, Telefont M and Renner KJ 2008. Estrogen regulation of proteins in the rat ventromedial nucleus of the hypothalamus. J Proteome Res 7, 5040-5048.

Mong J, Easton A, Kow L-M and Pfaff D 2003. Neural, hormonal and genetic mechanisms for the activation of brain and behavior. Eur J Pharmacol 480, 229-231.

Mong JA and Pfaff DW 2003. Hormonal and genetic influences underlying arousal as it drives sex and aggression in animal and human brains. Neurobiol Aging 24, S83-S88.

Mong JA and Pfaff DW 2004. Hormonal symphony: steroid orchestration of gene modules for sociosexual behaviors. Mol Psychiatry 9, 550-556.

Newman SW 1999. The medial extended amygdala in male reproductive behavior a node in the mammalian social behavior network. Ann N Y Acad Sci 877, 242-257.

Nie H, Crooijmans R, Bastiaansen J, Megens H-J and Groenen M 2010a. Regional regulation of transcription in the chicken genome. BMC Genomics 11, 28.

Nie H, Crooijmans RPMA, Lammers A, van Schothorst EM, Keijer J, Neerincx PBT, Leunissen JAM, Megens H-J and Groenen MAM 2010b. Gene expression in chicken reveals correlation with structural genomic features and conserved patterns of transcription in the terrestrial vertebrates. PLoS ONE 5, e11990.

Okamura H and Ohkura S 2007. Neuroendocrine control of reproductive function in ruminants. Anim Sci J 78, 105-111.

Petersen S, Gardner E, Adelman J and McCrone S 1996. Examination of steroid-induced changes in LHRH gene transcription using 33P-and 35S-labeled probes specific for intron 2. Endocrinology 137, 234-239.

Pfaff D 2005. Hormone-driven mechanisms in the central nervous system facilitate the analysis of mammalian behaviours. J Endocrinol 184, 447-453.

Pfaff DW, Kow LM, Loose MD and Flanagan-Cato LM 2008. Reverse engineering the lordosis behavior circuit. Horm Behav 54, 347-354.

Philipsson J and Lindhé B 2003. Experiences of including reproduction and health traits in Scandinavian dairy cattle breeding programmes. Livest Prod Sci 83, 99-112.

Pillai RS, Bhattacharyya SN and Filipowicz W 2007. Repression of protein synthesis by miRNAs: how many mechanisms? Trends Cell Biol 17, 118-126.

Pryce JE and Veerkamp RF 1999. The incorporation of fertility indices in genetic improvement programmes. In Fertility in the high-producing dairy cow. - Galway, Ireland: British Society of Animal Science, 20-22 September 1999 - Occasional meeting, 29-30.

Pryce JE and Veerkamp RF 2001. The incorporation of fertility indices in genetic improvement programmes. BSAS Occasional Publication 26, 237-250.

Pryce JE, Royal MD, Garnsworthy PC and Mao IL 2004. Fertility in the high-producing dairy cow. Livest Prod Sci 86, 125-135.

Qu Y and Xu S 2006. Quantitative trait associated microarray gene expression data analysis. Mol Biol Evol 23, 1558-1573.

Reiner-Benaim A, Yekutieli D, Letwin NE, Elmer GI, Lee NH, Kafkafi N and Benjamini Y 2007. Associating quantitative behavioral traits with gene expression in the brain: searching for diamonds in the hay. Bioinformatics 23, 2239-2246.

Richards N, Wayman C and Allers KA 2010. Neuronal activity in the hypothalamic paraventricular nucleus varies across the estrous cycle in anesthetized female rats: effects of dopamine receptor agonism. J Sex Med 7, 1104-1115.

Richmond G and Clemens L 1988. Ventromedial hypothalamic lesions and cholinergic control of female sexual behavior. Physiol Behav 42, 179-182.

Roche JF 2006. The effect of nutritional management of the dairy cow on reproductive efficiency. Anim Reprod Sci 96, 282-296.

Roelofs JB, van Eerdenburg FJCM, Soede NM and Kemp B 2005. Various behavioral signs of estrous and their relationship with time of ovulation in dairy cattle. Theriogenology 63, 1366-1377.

Roxstrom A, Strandberg E, Berglund B, Emanuelson U and Philipsson J 2001. Genetic and environmental correlations among female fertility traits and milk production in different parities of Swedish Red and White dairy cattle. Acta Agr Scand, A 51, 7-14.

Royal MD, Mann GE and Flint APF 2000a. Strategies for Reversing the Trend Towards Subfertility in Dairy Cattle. Vet J 160, 53-60.

Royal MD, Darwash AO, Flint APF, Webb R, Woolliams JA and Lamming GE 2000b. Declining fertility in dairy cattle: Changes in traditional and endocrine parameters of fertility. Anim Sci 70, 487-501.

Salamon E, Esch T and Stefano GB 2005. Role of amygdala in mediating sexual and emotional behavior via coupled nitric oxide release. Acta Pharmacol Sin 26, 389-395.

Smyth GK 2005. Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor (eds. R Gentleman, V Carey, S Dudoit, R Irizarry and W Huber), pp. 397-420, Springer, New York.

Spelman R and Garrick D 1997. Utilisation of marker assisted selection in a commercial dairy cow population. Livest Prod Sci 47, 139-147.

Stumpf WE 1970. Estrogen-Neurons and estrogen-neuron systems in the periventricular brain. Am J Anat 129, 207-217.

Tacken GML, Banse M, Batowska A, Gardebroek C, Nesha Turi K, Wijnands JHM and Poppe KJ 2009. Competitiveness of the EU dairy industry. LEI Wageningen UR, Den Haag.

The Bovine Genome Sequencing and Analysis Consortium, Elsik CG, Tellam RL and Worley KC 2009. The genome sequence of taurine cattle: A window to ruminant biology and evolution. Science 324, 522-528.

van Driel R, Fransz PF and Verschure PJ 2003. The eukaryotic genome: a system regulated at different hierarchical levels. J Cell Sci 116, 4067-4075.

van Eerdenburg FJCM, Loeffler HSH and Van Vliet JH 1996. Detection of oestrus in dairy cows: A new approach to an old problem. Vet Q 18, 52-54.

Vasudevan N and Pfaff DW 2008. Non-genomic actions of estrogens and their interaction with genomic actions in the brain. Front Neuroendocrinol 29, 238-257.

Vasudevan S, Tong Y and Steitz JA 2007. Switching from repression to activation: MicroRNAs can up-regulate translation. Science 318, 1931-1934.

Veerkamp RF and Beerda B 2007. Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68, S266-S273.

Veerkamp RF, Beerda B and van der Lende T 2003. Effects of genetic selection for milk yield on energy balance, levels of hormones, and metabolites in lactating cattle, and possible links to reduced fertility. Livest Prod Sci 83, 257-275.

Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ and van Kampen AHC 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res 13, 1998-2004.

Vinogradov AE 2004. Compactness of human housekeeping genes: selection for economy or genomic design? Trends Genet 20, 248-253.

Walf AA and Frye CA 2006. A review and update of mechanisms of estrogen in the hippocampus and amygdala for anxiety and depression behavior. Neuropsychopharmacology 31, 1097-1111.

Zimin A, Delcher A, Florea L, Kelley D, Schatz M, Puiu D, Hanrahan F, Pertea G, Van Tassell C, Sonstegard T, Marcais G, Roberts M, Subramanian P, Yorke J and Salzberg S 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol 10, R42.

2

Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows

A. Kommadath[1], H.A. Mulder[1], A.A.C. de Wit[1], H. Woelders[1], M.A. Smits[1], B. Beerda[1#], R.F. Veerkamp[1], A.C.J. Frijters[2] and M.F.W. te Pas[1]

[1]Animal Breeding and Genomics Centre, Animal Sciences Group of Wageningen UR, P.O. Box 65, 8200 AB Lelystad, The Netherlands; [2]CRV, Research & Development, P.O. Box 454, 6800AL Arnhem, The Netherlands

[#] Present address: Adaptation Physiology Group, Department of Animal Sciences, Wageningen University, 6709PG Wageningen, The Netherlands

Animal (2010) 4: 1297-1307

Abstract

Intensive selection for high milk yield in dairy cows has raised production levels substantially but at the cost of reduced fertility, which manifests in different ways including reduced expression of oestrous behaviour. The genomic regulation of oestrous behaviour in bovines remains largely unknown. Here, we aimed to identify and study those genes that were associated with oestrous behaviour among genes expressed in the bovine anterior pituitary either at the start of oestrous cycle or at the mid-cycle (around day 12 of cycle), or regardless of the phase of cycle. Oestrous behaviour was recorded in each of 28 primiparous cows from 30 days in milk onwards till the day of their sacrifice (between 77 and 139 days in milk) and quantified as heat scores. An average heat score value was calculated for each cow from heat scores observed during consecutive oestrous cycles excluding the cycle on the day of sacrifice. A microarray experiment was designed to measure gene expression in the anterior pituitary of these cows, 14 of which were sacrificed at the start of oestrous cycle (day 0) and 14 around day 12 of cycle (day 12). Gene expression was modelled as a function of the orthogonally transformed average heat score values using a Bayesian hierarchical mixed model on data from day 0 cows alone (analysis 1), day 12 cows alone (analysis 2) and the combined data from day 0 and day 12 cows (analysis 3). Genes whose expression patterns showed significant linear or non-linear relationships with average heat scores were identified in all three analyses (177, 142 and 118 genes, respectively). Gene ontology terms enriched among genes identified in analysis 1 revealed processes associated with expression of oestrous behaviour whereas the terms enriched among genes identified in analysis 2 and 3 were general processes which may facilitate proper expression of oestrous behaviour at the subsequent oestrus. Studying these genes will help to improve our understanding of the genomic regulation of oestrous behaviour, ultimately leading to better management strategies and tools to improve or monitor reproductive performance in bovines.

Key words: oestrous behaviour, gene expression pattern, quantitative trait, anterior pituitary, dairy cow

2.1 Introduction

Several decades of intensive selection for high milk yield in dairy cows has raised production levels substantially, but at the cost of reduced fertility as the unfavourable genetic correlation between milk yield and fertility traits used to be ignored (Royal *et al*., 2000; Roxstrom *et al*., 2001). The expression of oestrous behaviour (heat), a key fertility trait that marks the fertile period in cows, has decreased both in duration and intensity over generations of cows selected for high milk yield (Lopez *et al*., 2004). Short heat periods and the absence of clear behavioural signs of heat make farmers fail to detect heat or misjudge the optimum time of insemination of their cows, resulting in financial losses because of prolonged interval from calving to first insemination, reduced conception rates and increased calving intervals.

Some key hormones that regulate oestrous cycle in mammals are gonadotropin-releasing hormone (GnRH) produced in the hypothalamus; follicle-stimulating hormone (FSH) and luteinizing hormone (LH) in the anterior pituitary; oestrogen and progesterone in the ovaries; and prostaglandin in the uterus. These hormones may influence the expression of oestrous behaviour by targeting specific brain areas, directly or indirectly. For example, oestrogen dependent gene transcription in the hypothalamus and midbrain establishes the neural circuitry required for lordosis behaviour in female rats (Pfaff, 2005). The hormonal regulation of oestrous cycle is well studied and several hormone-driven mechanisms in the brain that influence the mammalian sexual behaviour are known (Pfaff, 2005), but specific knowledge on the genomic regulation of oestrous behaviour in cows is lacking. Understanding the genomic regulation of oestrous behaviour may help to develop better management strategies and tools to improve or monitor reproductive performance in bovines (Veerkamp and Beerda, 2007).

Conventionally, DNA microarray technology is used to identify differentially expressed genes between groups of individuals belonging to contrasting classes of a phenotypic trait of interest. However, when the trait of interest is quantitative (e.g. oestrous behaviour quantified as heat score), the grouping of individuals into qualitative classes, ignoring the continuous scale, is indistinct and dilutes the available information. In such cases, the association between gene expression and phenotypic trait is better analysed using the individual quantitative measurements (Reiner-Benaim *et al*., 2007). Recent studies on microarray gene expression data have successfully linked genes to quantitative traits of interest by correlation, linear regression or complex regression models (Blalock *et al*., 2004; Jia and Xu, 2005; Qu and Xu, 2006; Jia *et al*., 2008). Some of these models have the added advantage

that they account for non-linear relationships between gene expression and phenotypic trait that occur because of complex interactions among genes in regulatory networks. The Bayesian hierarchical mixed model developed by Jia *et al*. (2008) fits linear as well as nonlinear associations between gene expression patterns and quantitative measures of a trait using orthogonal polynomials.
Here, the Bayesian hierarchical mixed model was used for the objective to identify and study those genes that were associated with oestrous behaviour among genes expressed in the bovine anterior pituitary, either at the start of oestrous cycle, or at the mid-cycle (around day 12 of cycle), or regardless of the phase of cycle.

2.2 Material and Methods

Phenotypic data recording and tissue isolation

The current study is a part of an experiment which was originally set up to investigate differential gene expression in brain tissue samples of Holstein Friesian (HF) cows, which were either of a low or high genetic merit for fertility and which belonged to two different stages of oestrous cycle. However, here we focus on identifying genes whose expression profiles were associated with oestrous behaviour at the start or mid of oestrous cycle, or regardless of the phase. The associated gene lists thus identified were compared for similarities or differences. No tests were performed to identify genes that significantly differ in their level of expression between the groups of animals in different phases.
Twenty-eight healthy HF heifers were selected for this study, of which 14 belonged to a lower fertility group with estimated breeding values (EBVs) for fertility ranging between 93 and 97 whereas the remaining 14 belonged to a higher fertility group with values ranging between 101 and 103. The EBVs are expressed on a standardised scale with 100 as the base value and with a standard deviation of 4 units for EBV with a reliability of 0.80. The base value of 100 corresponds to the average EBV of black and white HF cows born in the year 2000. The EBV for fertility was based on the traits: time to first insemination, percentage non-return within 56 days after first insemination and interval between calvings (NRS, 2009). The EBVs for fertility of the selected heifers were calculated using the EBVs for fertility of their sire (weighing factor 2) and their dam's sire (weighing factor 1).
At least 4 weeks before their expected calving date, the experimental heifers were moved to a free stall barn with slatted concrete floor at the Waiboerhoeve research farm at Lelystad in the Netherlands and reared under identical conditions of feeding and management. The age of the heifers at calving varied between 22 and 30 months, with calving dates in the period from September 2006 to December 2006. After calving, the cows were observed daily for the signs of oestrous

expression during two observation periods of 30 min each: one in the morning (after milking but before feeding) and one in the evening (after milking and feeding). Cows were kept in one group of animals of similar age, which over time ranged in size from 15 to 30 animals. At the start and end of the study, cows from the main herd were added to the experimental group so as to maintain a group size of at least 15 individuals. Oestrous behaviour was expressed as heat scores, with specific behaviours being weighted according to the scoring protocol described by Van Eerdenburg (2006): mucous vaginal discharge (three points), flehmen (three points), restlessness (five points), being mounted but not standing (ten points), sniffing the vulva of another cow (ten points), resting with chin on the back of another cow (fifteen points), mounting other cows, or attempting to do so (thirty-five points), mounting head side of other cows (forty-five points), standing heat (hundred points). Heat scores per 30 min observation bouts were calculated by adding up the recorded occurrence of each specific behaviour multiplied by its weight as per the scoring protocol described above. However, the recording of the occurrence of restlessness behaviour was limited to one per observation bout. From 30 days in milk (DIM) onwards, milk progesterone levels were assessed twice a week. Ovarian structure was evaluated by trans-rectal ultrasonography either performed on alternate days or daily during the week preceding expected oestrus. The time of oestrus was established on the basis of milk progesterone levels and trans-rectal ultrasonography to help determine the right moment of euthanizing cows and not miss those with possible silent heat. Cows in silent heat were awarded a heat score of 0. Cows were assumed to be in oestrus when the cumulative heat score from two consecutive observation periods exceeded 50 points. An average heat score value was calculated for each cow to quantify the degree to which it expressed oestrous behaviour (i.e. as a trait) based on its heat scores at oestrus observed during all consecutive oestrous cycles from 30 DIM onwards, excluding the heat score at oestrus on the day of sacrifice. All cows were euthanized in the period ranging between 77 and 139 DIM by intravenous injection of 20 ml of T61. They were sacrificed in a stress-free, quick and standardised way, between 8 and 11 am, in a room only a few metres away from their home section. Fourteen of the 28 cows were euthanized at the start of their oestrous cycle (hereafter indicated as day 0) and the remaining 14 were euthanized at mid-cycle around day 12 (hereafter indicated as day 12) when milk progesterone levels were high. Each group of 14 cows that were euthanized at a particular time point consisted of seven cows with high and seven with low-genetic merit for fertility. Within an hour of death, tissue samples from the anterior pituitary and brain areas, that is the hippocampus, amygdala, dorsal and ventral hypothalamus, were

collected, snap-frozen in liquid nitrogen and stored at −70°C until the RNA isolation was made.
The study was approved by the Animal Care and Ethics Committee of the Animal Sciences Group of Wageningen University and Research Centre, Lelystad.

RNA isolation and microarray hybridisation

The procedure used for RNA isolation from all the separate brain samples collected was similar to that described by Niewold *et al*. (2007) but without the sodium citrate/NaCl precipitation step. Bovine 24K oligonucleotide (70-mer) microarrays designed and produced by the Bovine Oligonucleotide Microarray Consortium (BOMC) (http://www.bovineoligo.org/) were used. A total of 56 arrays were prepared in a common reference design with the dye labels swapped between individual anterior pituitary samples and a reference sample consisting of equal proportions of RNA from the anterior pituitary, hippocampus, amygdala, dorsal hypothalamus and ventral hypothalamus of all 28 cows. Exactly 5 µg of RNA was used per labelling using the RNA MICROMAX TSA™ labelling and detection kit (Perkin-Elmer, Boston, MA, USA). Following hybridization, cover slips were removed after allowing a 5 min delay at room temperature followed by three washings of 15 min each in: 2 × SSC (saline sodium citrate) + 0.2% SDS (at 42°C), 2 × SSC (room temperature) and 0.2 × SSC (room temperature). The rest of the protocol conformed to that described by Niewold *et al*. (2007). Part of the processed slides were scanned using ScanArray 5000 (Packard Biosciences, Billerica, MA, USA), part using ScanArray Express (Perkin-Elmer, Boston, MA, USA) and the remaining using GenePix 4200A (Molecular Devices, Sunnyvale, CA, USA), with identical settings. Image processing was performed using GenePix Pro 6 software. All processed images were visually inspected for proper data recording and any bad spots that remained undetected by the software were manually flagged.

Pre-processing of microarray data

LIMMA (linear models for microarray data) package (Smyth, 2005) within Bioconductor project (Gentleman *et al*., 2004) of R statistical programming language (http://www.r-project.org/) was used for pre-processing the microarray data including microarray data quality checking. The quality of the array data was checked by means of several data visualisation plots such as image plots, MA-plots, density plots and box plots. The image plots of the background intensities for the two dyes showed that the background was inconsistent within and between arrays. Background correction was therefore required and was performed using the

'normexp + offset' method (Ritchie *et al*., 2007) available in LIMMA. This method always produces positive corrected values so that no spot information is lost. In addition, an offset of 50 was used to stabilise the variability of the log-ratios (M-values) as a function of intensity. Within array normalisation was performed using print tip loess method, which is a good method to correct for spatial effects and intensity dependent biases. To make the arrays comparable, between arrays normalisation was performed using quantile method, which equalizes the intensity distribution across all arrays. A comparison of the MA-plots and box plots on data before and after normalisation showed that the normalisation procedure followed had corrected intensity dependent biases and made the ranges of intensity distributions comparable across arrays. On the basis of observations of the above-mentioned plots, we concluded that the microarray data quality was good and that the background correction and array normalisation procedures followed were adequate. The M-values per probe of the dye-swap pairs for each individual were averaged (after reversing the sign of M-value for one of the dye channels) to obtain the gene expression data. As the gene expression data was generated using a two colour common reference design microarray experiment, the gene expression levels in the anterior pituitary were assessed relative to expression levels in reference tissue consisting of samples from the anterior pituitary and four brain areas as obtained during two phases of the oestrous cycle. This means that genes with negative M-values were lower expressed in the anterior pituitary than in the common reference while genes with positive M-values were higher expressed in the anterior pituitary than in the common reference.

Associating heat scores with gene expression data

The Bayesian hierarchical mixed model developed by Jia *et al*. (2008) was used to associate quantitative phenotypes to expressed genes using orthogonal polynomials. In this model, the expression level of gene *i* (Y_i) across N subjects, as a function of the phenotypic value of a quantitative trait, *Z*, is given by:

$$Y_i(Z) = \alpha_i + \beta_i(Z) + \varepsilon_i$$

where α_i is the gene specific intercept for gene *i*, $\beta_i(Z)$ is an arbitrary function describing the relationship between the gene expression of gene *i* and the phenotypic values and ε_i is the random error term with assumed normal distribution, N(0, σ_ε^2). Using orthogonal polynomials to describe the functional relationships between the model parameters and *Z*, followed by a linear contrasting scheme to remove the mean expression (μ) of each gene (Qu and Xu, 2006), the model becomes:

$$y_i = Y_i(Z) - \mu = \sum_{j=1}^{p} X_j \beta_{ij} + \varepsilon_i$$

where *p* is the order of the orthogonal polynomial (*j*) and *X* is an *N*x*p* matrix denoting *Z* after transformation to its orthogonal polynomials. In the mixed model, the gene specific regression coefficient (β_{ij}) is assumed to be sampled from a mixture of two normal distributions: one with a very small variance, *N(0,∂)* and one with a larger variance, N(0,σ_j^2),where $\partial = 10^{-4}$ (a small positive number) and σ_j^2 is an unknown variance assigned to the j^{th} polynomial. This approach used by Jia *et al*. (2008) is based on a procedure called stochastic search variable selection developed by George and McCulloch (1993). The variable η_{ij} = {0,1} is used to indicate whether β_{ij} is sampled from the distribution with the small variance in which case β_{ij} is approximating 0 or whether β_{ij} is sampled from the distribution with the large variance, in which case β_{ij} will have a non-trivial value and should be estimated from the data. Both variance components, σ_j^2 and $\sigma_\varepsilon^2$are estimated by borrowing information across all genes. Using the Markov Chain Monte Carlo simulation, the association status of gene *i* with the j^{th} polynomial is determined based on the posterior mean of η_{ij}. According to this association status, genes are clustered in 2^p clusters. More details can be found in Jia *et al*. (2008).

For the current study, we used the above mentioned algorithm of Jia *et al*. (2008) coded in SAS® language, kindly provided by the authors. Average heat score was selected as the quantitative phenotype to associate with gene expression, as it was the most representative for the trait oestrous behaviour. This was determined in a principal component analysis where oestrous behaviour related parameters as collected over time were correlated, with heat scores showing the strongest loadings (B. Beerda, unpublished results). The interpretation of the results obtained in the association methodology used in this study differs from that for differential gene expression analysis in LIMMA where contrasting groups of individuals are compared against each other for significant differences in gene expression levels without considering its association with a quantitative trait.

The clustering program was run on SAS® software, Version 9.1 of the SAS® System for Windows. Gene expression was modelled as a function of average heat score transformed into third order polynomials, thereby clustering genes into eight binary based categories. The categories were represented as 000, 100, 010, 001, 110, 101, 011 and 111. Genes with no association in all three orders of the polynomial belong to cluster 000, those with a linear association alone belong to cluster 100, those with a linear and quadratic association belong to cluster 110 and so on. The algorithm was run for 10,000 iterations with a burn-in period of 5000

Table 2.1 Description of the three analyses and their objectives.

Analysis	Data	Objective
1	Gene expression data of day 0 cows and their average heat scores from previous cycles	To identify genes of which the expression in the bovine anterior pituitary at the start of oestrus was associated with oestrous behaviour
2	Gene expression data of day 12 cows and their average heat scores from previous cycles	To identify genes of which the expression in the bovine anterior pituitary around mid-cycle (dioestrus) was associated with oestrous behaviour
3	Gene expression data of day 0 and day 12 cows and their average heat scores from previous cycles	To identify genes of which the expression in the bovine anterior pituitary was associated with oestrous behaviour regardless of the phase of oestrous cycle

iterations. Trace plots of the estimated residual variance in consecutive iterations showed that the parameter stabilised within a narrow range at approximately 3000 iterations and therefore the selected burn-in period of 5000 iterations was sufficient. After the burn-in period, results of one iteration in 20 were saved, resulting in 250 samples used for calculating posterior means of each variable. The cut-off value for $\overline{\eta_{ij}}$ for cluster assignment was set at 0.8 to limit the false discovery rate (FDR) of cluster assignment below 1% (Jia and Xu, 2007).

The clustering program was run in three separate analyses (Table 2.1). To identify genes of which the expression in the bovine anterior pituitary at the start of oestrus was associated with oestrous behaviour (objective 1), we analysed gene expression data of day 0 cows and their average heat scores (analysis 1). Similarly, to identify genes of which the expression in the bovine anterior pituitary around mid-cycle (dioestrus) was associated with oestrous behaviour (objective 2), we analysed gene expression data of day 12 cows and their average heat scores (analysis 2). Finally, to identify genes of which the expression in the bovine anterior pituitary at the start of oestrus and at around mid-cycle was associated with oestrous behaviour, that is regardless of the phase of oestrous cycle (objective 3), we analysed together, the gene expression data of day 0 and day 12 cows and their average heat scores (analysis 3).

Gene ontology (GO) based analysis of oestrous behaviour associated genes

For gaining insight into biological processes underlying associations between gene expression profiles and oestrous behaviour as identified by the described clustering

program, we performed statistical analysis for over-representation of GO terms in sets of oestrous behaviour associated genes (study sets) in the three analyses compared to the set of all genes represented on the array (population set). The GO term over-representation analysis was performed using the Ontologizer (Bauer *et al*., 2008), a web based software package, with the parent–child intersection method developed by Grossmann *et al*. (2007) that addresses the problem of dependencies between annotation of parent and child terms in the GO hierarchy. The Ontologizer software uses a modified Fisher's exact test to calculate the statistical significance of over-represented GO terms. In addition to the study sets and the population set, the Ontologizer software required as input, the OBO (Open Biomedical Ontologies) file and the GO association file, both of which were available for download from the website of the GO project (http://www.geneontology.org/). The OBO file consists of GO terms, their definition and structure, whereas the GO association file maps gene products (protein, gene and transcript, *etc*.) to GO terms. Here, the OBO and GO association files downloaded on 22 July 2008 were used. The microarray probe annotation file provided by the BOMC (http://www.bovineoligo.org/) for the bovine oligonucleotide array provides the Ensembl ID (http://www.ensembl.org/) of the best matching human homologous protein for 16,464 probes in the array. As human genes are better annotated than bovine genes, it was decided to use the human homologous protein information for the GO-based analysis. The human homologous protein Ensembl IDs were converted to the corresponding UniProt ID (http://www.uniprot.org/) wherever possible using the 'g:Convert' module of the web-based tool 'g:Profiler' (Reimand *et al*., 2007) as the majority of the GO terms in the GO association file were mapped to UniProt IDs followed by Ensembl and NCBI RefSeq IDs (http://www.ncbi.nlm.nih.gov/RefSeq/). After filtering out genes in the population set without any GO term associations, 7635 genes remained. The GO analysis was performed separately on the list of genes from each associated cluster in each analysis and also with the combined list of genes from all associated clusters per analysis because associated genes from different clusters could be involved in common biological processes. To reduce the stringency of FDR control on the small number of genes having GO annotation and thereby observe general trends in the data, the adjusted P-value cut-off of correction for multiple testing by Benjamini-Hochberg method was relaxed to 0.35 while still considering only GO terms with ordinary P-value below 0.01 as significant.

Further, the GO analysis was re-performed by defining a smaller refined set of genes as the population set. The refined set included only genes that were expressed in the anterior pituitary with variability across samples and excluded

genes, which were either not expressed or whose expression remained fairly constant across all samples. To determine this refined population set, we filtered out genes whose expression values had an inter-quartile range less than 0.5 across all the array samples, leaving a total of 9608 genes. After filtering out genes in this set without any GO term associations, 2461 genes remained.

2.3 Results

The trait of expressing oestrous behaviour was quantified as a cow's average heat score. Data from one of the day 0 cows was excluded from further analysis because of its high outlier heat score (1750). The average heat scores for the remaining 13 day 0 cows ranged from 0 to 405, with a mean value of 178.4 (standard deviation 125.7), and the average heat scores for the 14 day 12 cows ranged from 2 to 505, with a mean value of 244.7 (standard deviation 175.4). These scores were used with the corresponding gene expression data to run the three analyses as already described. The total number of heat score associated genes found per cluster in each analysis is provided in Table 2.2, including the number of genes for which UniProt/Ensembl annotated gene product information was available. All three analyses showed that over 99% of the genes had no association with heat scores. Genes that were associated with heat scores typically showed linear, quadratic or cubic relationships and only a few showed combinations of these relations. The combined numbers of associated genes in all clusters in the three analyses were 177, 142 and 118, respectively. Figure 2.1 presents a Venn diagram showing the

Table 2.2 Association status of gene expression patterns to average heat score in the three analyses.

Cluster	Association status	No. of genes in cluster (no. of genes with UniProt/Ensembl annotated gene product)		
		Analysis 1	Analysis 2	Analysis 3
1 (000)	No association	23319	23354	23378
2 (100)	Linear	45 (35)	65(48)	23(18)
3 (010)	Quadratic	85 (60)	25(16)	10(4)
4 (001)	Cubic	37 (33)	35(21)	76(57)
5 (110)	Linear + Quadratic	2 (2)	3(3)	1(1)
6 (101)	Linear + Cubic	1 (0)	12(11)	6(6)
7 (011)	Quadratic + Cubic	7 (6)	1(1)	1(1)
8 (111)	Linear + Quadratic + Cubic	0 (0)	1(1)	1(1)
Combined clusters 2 to 8 (all associated genes)		177 (135)	142(101)	118(88)

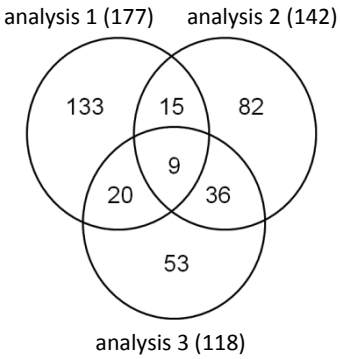


Figure 2.1 Venn diagram of number of genes identified as associated with heat score in the three analyses.

number of oestrous behaviour associated genes found per analysis and their overlap. The number of common genes that were associated with oestrous behaviour across analyses was lower when comparing the results from analyses 1 and 2 than comparing those from 1 and 3 or 2 and 3. The difference in the sets of genes identified in the three analyses indicates that the relationship between gene expression profiles and oestrous behaviour depend on the phase of oestrous cycle.

Some of the genes/probes had limited or no annotation in the microarray probe annotation file provided by BOMC, in which case, we reverted to the recent re-annotation of the bovine microarray probes (Version 2, Ensembl 50: 11 September 2008) by the European Animal Disease Genomics Network of Excellence for Animal Health and Food Safety (EADGENE) (http://www.eadgene.info). Supplementary Table A (Additional file 1) provides a list of all oestrous behaviour associated genes found in the three analyses along with their annotation from 2 sources – BOMC and EADGENE.

Of the nine genes associated with oestrous behaviour in all three analyses, interestingly, three genes were immunoglobulin related and one was follicle-stimulating hormone beta (*FSHB*) subunit, a key hormone in the regulation of oestrous cycle. Figure 2.2 shows the association of gene expression profile of FSHB with heat score for the three analyses – a linear relationship in analysis 1, cubic in analysis 2 and quadratic in analysis 3. Unfortunately, the association status of LH, another key hormone in the regulation of oestrous cycle could not be ascertained as probes for *LH* beta subunit were not present on the array. Several genes in the list of heat score associated genes are known from literature to be related to oestrous expression. These included prolactin (*PRL*) precursor, pituitary-specific positive transcription factor 1 (*PIT1*) and melanin-concentrating hormone (*MCH*) which associated with oestrous behaviour in analyses 1 and 2 separately but not in

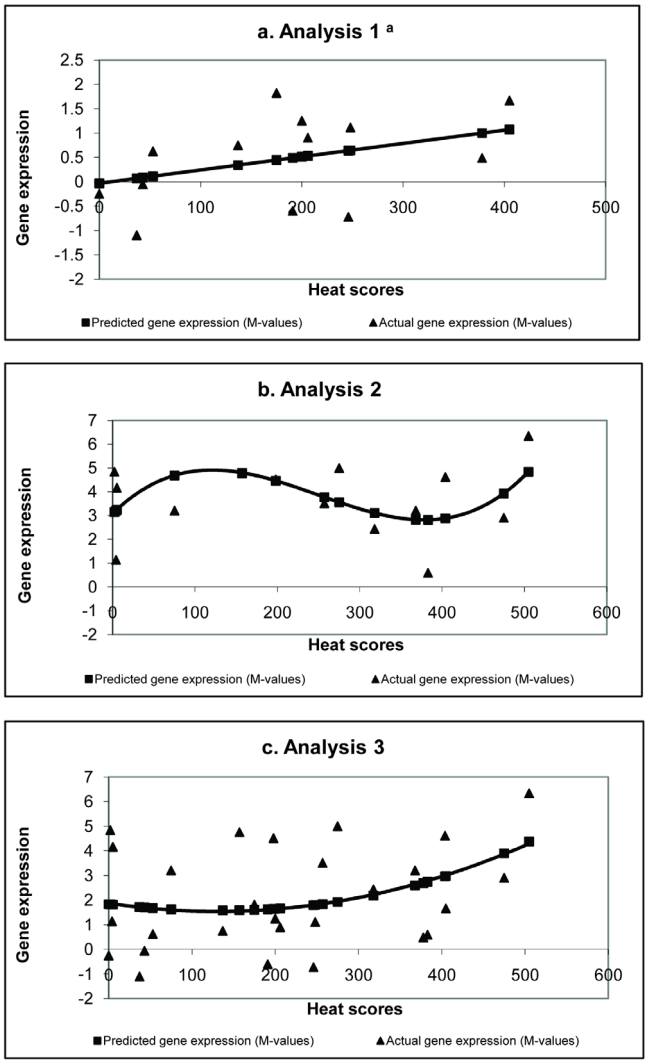


Figure 2.2 Association of gene expression profile of FSH (beta) gene with heat score in the three analyses. [a] Note difference in scale in Figure 2.2a.

analysis 3. Further, considering a few examples of associated genes with a nervous system related function, Homo sapiens neurotrimin precursor (*HNT*), associated with oestrous behaviour in analyses 1 and 3, neurokinin-B precursor (*NKB*) and neurogenic differentiation factor 2 (*NEUROD2*) associated with oestrous behaviour only in analysis 1.

The genes that showed associations with oestrous behaviour were further analysed

based on their GO annotations provided by BOMC. The significant GO terms in the clusters of heat score associated genes in analyses 1, 2 and 3 are given in Tables 2.3, 2.4 and 2.5, respectively. The study count and population count provided in the tables correspond to the number of gene products annotated to the particular over-represented GO term in the cluster being analysed and in the whole array, respectively. It was noticed that certain GO terms enriched in the combined cluster were different from the terms in the separate cluster analysis.

Considering only those GO terms that were supported by at least four genes, significant terms within the GO category 'biological processes' for analysis 1 were: 'behaviour'; 'secretion'; 'nervous system development' and 'antigen processing and presentation' and for analysis 2 was 'multicellular organismal process'. Within the GO category 'molecular function' the following terms resulted: from analysis 1 – 'major histocompatibility complex (MHC) class I receptor activity' and 'identical protein binding'; and from analysis 2 – 'hormone activity'. Within the GO category

Table 2.3 Gene ontology (GO) terms over-represented in clusters of heat score associated genes from analysis 1 (day0 cows) (adjusted P < 0.35).

GO ID	GO term	GO category [a]	P - value	Adj. P - value	Study Count	Population Count
Cluster 010 (quadratic)						
GO:0032393	MHC class I receptor activity	M	< 0.001	0.212	2	5
GO:0007610	Behaviour	B	0.002	0.323	5	169
GO:0046903	Secretion	B	0.003	0.323	4	139
GO:0031225	Anchored to membrane	C	0.003	0.323	3	58
GO:0042802	Identical protein binding	M	0.004	0.323	4	239
GO:0048732	Gland development	B	0.004	0.323	2	24
GO:0046870	Cadmium ion binding	M	0.005	0.323	1	1
GO:0007399	Nervous system development	B	0.005	0.323	8	350
GO:0048154	S100 beta binding	M	0.006	0.323	1	2
GO:0031984	Organelle sub compartment	C	0.006	0.323	1	6
GO:0042995	Cell projection	C	0.007	0.329	4	181
GO:0045202	Synapse	C	0.008	0.337	3	91
Clusters 2 - 8 combined						
GO:0032393	MHC class I receptor activity	M	< 0.001	< 0.001	4	5
GO:0042611	MHC protein complex	C	< 0.001	0.257	4	18
GO:0005576	Extracellular region	C	0.001	0.257	16	743
GO:0019882	Antigen processing and presentation	B	0.001	0.257	4	29

[a] B – Biological Process; M – Molecular Function; C – Cellular Component.

Table 2.4 Gene ontology (GO) terms over-represented in clusters of heat score associated genes from analysis 2 (day 12 cows) (adjusted P < 0.35).

GO ID	GO term	GO category [a]	P - value	Adj. P - value	Study Count	Population Count
Cluster 010 (quadratic)						
GO:0032501	Multicellular organismal process	B	< 0.001	0.068	8	1496
GO:0005516	Calmodulin binding	M	0.003	0.324	2	94
GO:0016829	Lyase activity	M	0.003	0.324	2	71
Cluster 001 (cubic)						
GO:0006091	Generation of precursor metabolites and energy	B	< 0.001	0.231	3	159
Cluster 101 (linear + cubic)						
GO:0030276	Clathrin binding	M	0.001	0.111	1	6
Clusters 2 - 8 combined						
GO:0005576	Extracellular region	C	< 0.001	< 0.001	19	743
GO:0040034	Regulation of development, heterochronic	B	< 0.001	0.056	2	3
GO:0005179	Hormone activity	M	< 0.001	0.078	5	42
GO:0032501	Multicellular organismal process	B	0.002	0.343	19	1496

[a] B – Biological Process; M – Molecular Function; C – Cellular Component.

Table 2.5 Gene ontology (GO) terms over-represented in clusters of heat score associated genes from analysis 3 (day 0 + day 12 cows) (adjusted P < 0.35).

GO ID	GO term	GO category [a]	P - value	Adj. P - value	Study Count	Population Count
Cluster 100 (linear)						
GO:0001871	Pattern binding	M	0.001	0.226	2	58
GO:0010463	Mesenchymal cell proliferation	B	0.002	0.265	1	1
Cluster 110 (linear + quadratic)						
GO:0048511	Rhythmic process	B	0.006	0.279	1	37

[a] B – Biological Process; M – Molecular Function; C – Cellular Component.

'cellular component', both analyses 1 and 2 generated the term 'extracellular region' and analysis 1 produced in addition 'cell projection' and 'MHC protein complex'. Genes identified in analysis 3 did not have any significant enriched GO term with at least four genes. The gene products in the over represented GO terms are presented in Supplementary Tables B, C and D (Additional file 2) for analyses 1, 2 and 3, respectively.

When the GO analysis was re-performed with a population set consisting of only the genes expressed with a certain variability across samples, the top listed over represented GO terms corresponded to those found in the earlier analysis but the order of ranking of the terms based on P-values differed and the P-values were less significant because of the smaller population size. However, to understand the general trends in the biological processes enriched in the lists of heat score associated genes, the choice of either population set did not matter in this case.

2.4 Discussion

Gene expression profiles in the anterior pituitary of dairy cows were associated with oestrous behaviour by applying a Bayesian hierarchical mixed model based method for clustering genes on the basis of their linear, quadratic or cubic relation with a quantitative phenotype of interest, that is heat score. Genes in a regulatory network may interact in complex ways (feedback mechanisms, cooperation or competition between genes) to result in non-linear associations between gene expression levels and phenotype. Therefore it is advantageous to use a model that accounts for both linear and non-linear associations, for example, through polynomial regression. In this study, we began the microarray data analysis with pre-processing steps performed in LIMMA. However, for the association analysis, we did not attempt to model polynomial regression in LIMMA but instead we chose the SAS program developed by Jia *et al*. (2008) in which polynomial regression was already implemented. Jia *et al*. (2008) reported that their method was a better and faster algorithm to detect quantitative trait associated genes in comparison to similar methods like the one described by Qu and Xu (2006) based on an expectation-maximization (EM) algorithm. The increased speed is achieved by logically fixing the number of clusters beforehand thereby obviating the need for extra model evaluations for determining optimal cluster number as required in the EM algorithm. In our study, we tested the association between gene expression values and phenotypic trait measurements transformed in three orders of orthogonal polynomials, and consequently, fixed the number of clusters to eight. By running the three analyses, as defined earlier, we were able to detect several genes associated with the expression of oestrous behaviour. Each analysis took less than 6 h to complete on a normal desktop computer. Analyses were performed to identify genes that were associated with heat score during different phases of the oestrous cycle but not intended to identify genes differentially expressed between these phases. Analysis 1 identified genes that may have a direct association with the expression of oestrous behaviour whereas analysis 2 identified genes that may be involved in facilitator processes that prepare the cow for later oestrous

behaviour. Analysis 3 identified genes whose expression during both phases of the cycle may be associated with the expression of oestrous behaviour. The increased power gained by the greater number of data points in analysis 3 revealed significant genes that did not show up in the other two analyses. The overlap in the results from the three analyses indicate that certain genes that were directly associated with the expression of oestrous behaviour at oestrus were also involved in oestrous behaviour related processes at other phases in the cycle.

An alternative approach for doing three separate analyses would be to add nested regressions to account for phase-specific associations in a single model. However, we were not able to do so within the current framework of the SAS program as the algorithm followed would not support the steps following the model fitting if we altered the model by adding nested regressions. This could be taken up as a useful feature to add to the existing algorithm in the future.

It is likely that some of the genes predicted by this program to be associated with heat scores are false positives. To test the robustness of the algorithm used, we performed multiple runs of the algorithm on data from analysis 1 as a test case. The overlap between the original lists of associated genes reported here and the three repeated test runs were 79%, 83% and 88%. This showed that the algorithm was reasonably robust. In addition, we also did a leave-one-out analysis, although we realise that this analysis would give drastic results given the size of the current dataset. Nevertheless, even in the worst-case scenario using the smallest dataset (analysis 1), the overlap varied from 44% to 76% depending on the position of the data point removed. As expected, the differences were not only in the number of associated genes found but also the pattern of association it was assigned to. Owing to the small number of data points available, leaving out even one data point was expected to result in a change in the pattern of the association and thereby changes in the association status. Therefore, the relatively low number of animals sampled in this study was a constraint on this methodology due to which, especially the higher order relationships captured by, this method seemed to be in some cases an artefact of the method and were less reliable. For datasets of the size as in this study, it would be better to initially focus on the more reliably predicted linearly associated genes and then carefully consider genes that are associated with higher order relationships. The results may be checked for genes already known from literature to be involved in regulating oestrous. Further, rather than focusing on individual genes found associated by this method, it would be more reliable to focus on those genes that were together involved in certain biological processes as detected in the GO term enrichment analyses.

Known oestrus regulating genes and new candidate genes associated with oestrous behaviour

On the basis of annotation of the list of genes identified to be associated to oestrous behaviour (Supplementary Table A in Additional file 1), the list included genes that encoded for hormones, transcription factors, signalling molecules or other gene products. Some of these genes could be identified to have a function related to oestrous regulation and/or behaviour, examples of which are discussed below.

Among the key hormones known to regulate the oestrous cycle, probes for *GnRH* and *LH* were, unfortunately, not represented on the DNA microarray while the beta subunit of *FSH* was. *FSHB* was found to be associated with heat score either linearly or non-linearly in all the three analyses (Figure 2.2). Transient rises in FSH drive the emergence and growth of small antral follicles, with FSH concentrations reaching peak values around 28 h after the onset of a new oestrous cycle (Mihm and Austin, 2002). Relatively high expression levels of the FSHB gene may mirror a general good functioning of pituitary gonadotropes, promoting gonadotropin-induced steroidogenesis by ovarian follicles with the resulting oestrogens facilitating oestrous behaviour (Pfaff, 2005). *PRL* was associated with the oestrous behaviour in analysis 1. PRL, usually associated with lactation, is a multifunctional hormone that has been reported to have a negative effect on sexual arousal (Bancroft, 2005). Oestrous behaviour may, therefore, be enhanced at lower PRL level. The findings indicate, to some degree, the opposite functioning of biological processes underlying lactation and reproduction. Some of the genes that were found to be associated with oestrous behaviour have known functions related to behaviour or nervous development, making them likely candidates regulating the oestrous behaviour expression. These include *HNT* and synapsin-2 as found in analyses 1 and 3; neurexophilin-2, *PIT1* and *MCH* in analyses 1 and 2; tachykinin-3 precursor (containing NKB and neuromedin-K), *NEUROD2* and early growth response protein 1 (*EGR1*) found in analysis 1. *EGR1* regulates *LH* (beta subunit) gene expression in the pituitary gland (Lee *et al*., 1996). Relatively high *LH* activity, like that of *FSH*, may reflect appropriate functioning of pituitary gonadotropes and responsiveness to the oestrogens that synchronise oestrous behaviour and LH surge-induced ovulation. Similarly, *PIT1* may be an important regulator of oestrous behaviour due to its function as a transcription factor that activates expression of growth hormone and *PRL* genes. MCH is a neuropeptide whose administration in female rats stimulates sexual behaviour (Gonzalez *et al*., 1996) and proestrus FSH and LH releases, which resembles the effect produced by GnRH (Chiocchio *et al*., 2001).

Possibly, it plays a similar role in regulating bovine oestrous behaviour. The predominant expression site for *MCH* is not the pituitary but the hypothalamus, hence the relatively negative expression values noticed for this gene in all cows at both phases.

The gene product sodium/potassium-transporting ATPase subunit alpha-3 (*ATP1A3*) was found to be heat score associated in analysis 1. *ATP1A3* has been implicated in rapid-onset dystonia parkinsonism, characterised by sudden onset of neurological symptoms over hours to a few days (de Carvalho Aguiar *et al*., 2004), suggesting a role in the sudden onset of behavioural changes like during oestrus.

Several genes annotated to have immunological function were found to be associated with oestrous behaviour in all three analyses. Genes with immunological function were also identified in a related study by Beerda *et al*. (2008) on the same experimental cows as in this study but using tissue from the ventral tegmental area of the brain. In their study, an analysis using LIMMA for differential gene expression between groups of day 0 and day 12 cows demonstrated up-regulation of multiple immunoglobulin superfamily proteins in day 0 cows. Immunoglobulin superfamily proteins may play important roles in brain developmental processes and the functioning of neuronal networks in adults because they provide the ideal structure for protein–protein interactions and, thus, cell-cell interactions (Rougon and Hobert, 2003). Immunoglobulins may facilitate remodelling of synaptic networks, which occurs during oestrogen promoted female sexual behaviour (Pfaff, 2005).

The genes discussed above can be linked to oestrous behaviour on the basis of earlier reports on the functioning of their products, which identifies them as candidates for regulating oestrous behaviour in dairy cows. The heat score associated genes found in the different analyses may help us to postulate hypotheses on the genomic regulation of oestrous behaviour. There were several oestrous behaviour associated genes that are not currently annotated or whose function in the brain is still unknown and these genes may be of particular interest to target in future research.

Biological processes associated with oestrous behaviour

Gene ontology analysis of genes found in the different clusters of the three analyses revealed over-represented GO terms. The highest number of over-represented GO terms was found in analysis 1, confirming that relationships between gene expression profiles and oestrous behaviour are most strong around the time of oestrus. As expected, the biological processes, ‘behaviour’, ‘secretion’, and ‘nervous system development’ emerged from analysis 1. Also, the GO term, ‘antigen processing and presentation’, enriched in analysis 1 was in line with

expectations, given the role of immunoglobulins in remodelling of synaptic networks, which occurs during oestrogen promoted female sexual behaviour. Some of the over-represented GO terms, particularly in analyses 2 and 3, had no clear relationship with oestrous behaviour or fertility, and may represent more general processes that facilitate oestrous expression at a later phase. It would be interesting to study further the genes associated with these processes for useful new insights.

Some of the GO terms that emerged from the three analyses were supported by only a few genes in the study set. In part, this resulted from the limited number of genes for that GO term appearing on the whole array. For example, the term 'cadmium ion binding' in analysis 1 and 'mesenchymal cell proliferation' in analysis 3 have only one gene each in the study set and in the whole array associated to it. The GO terms that are based upon only few genes need to be evaluated critically regarding their importance in the regulation of oestrous behaviour, as it could be that these terms appear enriched just by chance.

2.5 Conclusions

The Bayesian hierarchical mixed model based clustering method used in this study was successful in detecting the oestrous behaviour associated genes based on the pattern of the relationship of the expression values of these genes with the quantitative phenotype i.e. heat score. Although most of the oestrous behaviour associated genes and the biological processes they controlled were activated around the time of oestrus, there were also genes expressed in mid-cycle that associate with oestrous behaviour, indicating that these genes may play a role in facilitating the next oestrus. Studying these genes and the processes they control will help improve our understanding of the genomic regulation of oestrous behaviour expression, ultimately leading to better management strategies or tools to improve or monitor reproductive performance. The list of oestrous behaviour associated genes identified may be useful for studying gene networks and also for inferring possible functions for non-annotated genes. On the basis of success of this method, a similar study may be repeated on other brain areas already sampled in this experiment. The results could then be integrated to get an overall view of the gene expression patterns of oestrous behaviour associated genes in different brain areas and the likely genetic cross-talk between them and how they contribute to the expression of oestrous behaviour.

Acknowledgements

Regarding the animal experiment, the support of Jan Bloemert and co-workers of the research farm Waiboerhoeve is highly appreciated and we are indebted to Henk Sulkers and Diana Baumann for their assistance. We thank Norbert Stockhofe for helping to set up the study and Ad Korevaar for his crucial role in the tissue collection. Zhenyu Jia and Shizhong Xu are acknowledged for kindly providing the SAS code for detection of quantitative trait associated genes and helpful initial instructions. The study was financially supported by the international cattle breeding co-operative CRV and has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2006-016250 (SABRE) and FOOD-CT-2004-506416 (EADGENE). The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

Supplementary Materials

Additional file 1.

Supplementary Table A. Annotation from 2 sources for the oestrous behaviour associated probes/genes identified in the 3 analyses – original annotation provided by BOMC and recent re-annotation by EADGENE.

http://journals.cambridge.org/downloadsup.php?file=/S1751731110000303sup001.xls&code=89e49314d05c5ef3bcc11dafcdd37ae9&mime=application/vnd.ms-excel

Additional file 2.

Supplementary Table B. Gene products within over-represented GO terms from analysis 1 (day0 cows).

Supplementary Table C. Gene products within over-represented GO terms from analysis 2 (day12 cows).

Supplementary Table D. Gene products within over-represented GO terms from analysis 3 (day0+day12 cows).

http://journals.cambridge.org/downloadsup.php?file=/S1751731110000303sup002.doc&code=89e49314d05c5ef3bcc11dafcdd37ae9&mime=application/msword

References

Bancroft, J 2005. The endocrinology of sexual arousal. J Endocrinol 186, 411–427.

Bauer, S, Grossmann, S, Vingron, M, Robinson, PN 2008. Ontologizer 2.0 – a multifunctional tool for GO term enrichment analysis and data exploration. Bioinformatics 24, 1650–1651.

Beerda, B, Wyszynska-Koko, J, te Pas, MFW, de Wit, AAC, Veerkamp, RF 2008. Expression profiles of genes regulating dairy cow fertility: recent findings, ongoing activities and future possibilities. Animal 2, 1158–1167.

Blalock, EM, Geddes, JW, Chen, KC, Porter, NM, Markesbery, WR, Landfield, PW 2004. Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. Proc Natl Acad Sci USA 101, 2173–2178.

Chiocchio, SR, Gallardo, MGP, Louzan, P, Gutnisky, V, Tramezzani, JH 2001. Melanin-concentrating hormone stimulates the release of luteinizing hormone-releasing hormone and gonadotropins in the female rat acting at both median eminence and pituitary levels. Biol Reprod 64, 1466–1472.

de Carvalho Aguiar, P, Sweadner, KJ, Penniston, JT, Zaremba, J, Liu, L, Caton, M, Linazasoro, G, Borg, M, Tijssen, MAJ, Bressman, SB, Dobyns, WB, Brashear, A, Ozelius, LJ 2004. Mutations in the Na+/K+-ATPase [alpha]3 gene ATP1A3 are associated with rapid-onset dystonia Parkinsonism. Neuron 43, 169–175.

Gentleman, R, Carey, V, Bates, D, Bolstad, B, Dettling, M, Dudoit, S, Ellis, B, Gautier, L, Ge, Y, Gentry, J, Hornik, K, Hothorn, T, Huber, W, Iacus, S, Irizarry, R, Leisch, F, Li, C, Maechler, M, Rossini, A, Sawitzki, G, Smith, C, Smyth, G, Tierney, L, Yang, J, Zhang, J 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5, R80.

George, EI, McCulloch, RE 1993. Variable selection via Gibbs sampling. J Am Stat Assoc 88, 881–889.

Gonzalez, MI, Vaziri, S, Wilson, CA 1996. Behavioral effects of [alpha]-MSH and MCH after central administration in the female rat. Peptides 17, 171–177.

Grossmann, S, Bauer, S, Robinson, PN, Vingron, M 2007. Improved detection of overrepresentation of Gene-Ontology annotations with parent child analysis. Bioinformatics 23, 3024–3031.

Jia, Z, Tang, S, Mercola, D, Xu, S 2008. Detection of quantitative trait associated genes using cluster analysis. In: Evolutionary computation, machine learning and data mining in bioinformatics (ed. E Marchiori and JH Moore), pp. 83–94. Springer-Verlag, Berlin Heidelberg.

Jia, Z, Xu, S 2005. Clustering expressed genes on the basis of their association with a quantitative phenotype. Genet Res 86, 193–207.

Jia, Z, Xu, S 2007. Mapping quantitative trait loci for expression abundance. Genetics 176, 611–623.

Lee, SL, Sadovsky, Y, Swirnoff, AH, Polish, JA, Goda, P, Gavrilina, G, Milbrandt, J 1996. Luteinizing hormone deficiency and female infertility in mice lacking the transcription factor NGFI-A (Egr-1). Science 273, 1219–1221.

Lopez, H, Satter, LD, Wiltbank, MC 2004. Relationship between level of milk production and estrous behavior of lactating dairy cows. Anim Reprod Sci 81, 209–223.

Mihm, M, Austin, EJ 2002. The final stages of dominant follicle selection in cattle. Domest Anim Endocrinol 23, 155–166.

Niewold, TA, Veldhuizen, EJA, van der Meulen, J, Haagsman, HP, de Wit, AAC, Smits, MA, Tersteeg, MHG, Hulst, MM 2007. The early transcriptional response of pig small intestinal mucosa to invasion by *Salmonella enterica* serovar *typhimurium* DT104. Mol Immunol 44, 1316–1322.

NRS 2009. The Royal Dutch Cattle Syndicate (NRS) Handbook. Retrieved from: https://www.cr-delta.nl/nl/fokwaarden/pdf/E17.pdf.

Pfaff, D 2005. Hormone-driven mechanisms in the central nervous system facilitate the analysis of mammalian behaviours. J Endocrinol 184, 447–453.

Qu, Y, Xu, S 2006. Quantitative trait associated microarray gene expression data analysis. Mol Biol Evol 23, 1558–1573.

Reimand, J, Kull, M, Peterson, H, Hansen, J, Vilo, J 2007. g:Profiler – a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res 35, W193–W200.

Reiner-Benaim, A, Yekutieli, D, Letwin, NE, Elmer, GI, Lee, NH, Kafkafi, N, Benjamini, Y 2007. Associating quantitative behavioral traits with gene expression in the brain: searching for diamonds in the hay. Bioinformatics 23, 2239–2246.

Ritchie, ME, Silver, J, Oshlack, A, Holmes, M, Diyagama, D, Holloway, A, Smyth, GK 2007. A comparison of background correction methods for two-colour microarrays. Bioinformatics 23, 2700–2707.

Rougon, G, Hobert, O 2003. New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. Annu Rev Neurosci 26, 207–238.

Roxstrom, A, Strandberg, E, Berglund, B, Emanuelson, U, Philipsson, J 2001. Genetic and environmental correlations among female fertility traits and milk production in different parities of Swedish Red and White dairy cattle. Acta Agric Scand, A 51, 7–14.

Royal, MD, Darwash, AO, Flint, APF, Webb, R, Woolliams, JA, Lamming, GE 2000. Declining fertility in dairy cattle: changes in traditional and endocrine parameters of fertility. Anim Sci 70, 487–501.

Smyth, GK 2005. Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and bioconductor (ed. R Gentleman, V Carey, S Dudoit, R Irizarry and W Huber), pp. 397–420. Springer, New York, NY, USA.

Van Eerdenburg, FJCM 2006. Estrus detection in dairy cattle: how to beat the bull. Vlaams Diergeneeskundig Tijdschrift 75, 61–69.

Veerkamp, RF, Beerda, B 2007. Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68, S266–S273.

3

Gene expression patterns in four brain areas associate with quantitative measure of estrous behavior in dairy cows

Arun Kommadath[1] , Henri Woelders[1] , Bonne Beerda[1,2] , Herman A Mulder[1] , Agnes AC de Wit[1] , Roel F Veerkamp[1] , Marinus FW te Pas[1] and Mari A Smits[1]

[1]Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 65, 8200 AB Lelystad, The Netherlands; [2]Adaptation Physiology Group, Department of Animal Sciences, Wageningen University, 6709 PG Wageningen, The Netherlands

BMC Genomics (2011) 12:200

Abstract

Background

The decline noticed in several fertility traits of dairy cattle over the past few decades is of major concern. Understanding of the genomic factors underlying fertility, which could have potential applications to improve fertility, is very limited. Here, we aimed to identify and study those genes that associated with a key fertility trait namely estrous behavior, among genes expressed in four bovine brain areas (hippocampus, amygdala, dorsal hypothalamus and ventral hypothalamus), either at the start of estrous cycle, or at mid cycle, or regardless of the phase of cycle.

Results

An average heat score was calculated for each of 28 primiparous cows in which estrous behavior was recorded for at least two consecutive estrous cycles starting from 30 days post-partum. Gene expression was then measured in brain tissue samples collected from these cows, 14 of which were sacrificed at the start of estrus and 14 around mid-cycle. For each brain area, gene expression was modeled as a function of the orthogonally transformed average heat score values using a Bayesian hierarchical mixed model. Genes whose expression patterns showed significant linear or quadratic relationships with heat scores were identified. These included genes expected to be related to estrous behavior as they influence states like socio-sexual behavior, anxiety, stress and feeding motivation (*OXT, AVP, POMC, MCHR1*), but also genes whose association with estrous behavior is novel and warrants further investigation.

Conclusions

Several genes were identified whose expression levels in the bovine brain associated with the level of expression of estrous behavior. The genes *OXT* and *AVP* play major roles in regulating estrous behavior in dairy cows. Genes related to neurotransmission and neuronal plasticity are also involved in estrous regulation, with several genes and processes expressed in mid-cycle probably contributing to proper expression of estrous behavior in the next estrus. Studying these genes and the processes they control improves our understanding of the genomic regulation of estrous behavior expression.

Key words: estrous behavior, gene expression pattern, quantitative trait, brain, dairy cow

3.1 Background

Maintaining good fertility and thereby optimum reproductive performance in dairy cows is of great economic importance for the dairy industry. Knowledge on factors influencing fertility is already being applied to improve or regulate fertility. For example, the importance of limiting negative energy balance in early lactation cows for proper reproductive performance is well recognized [1,2]. Insight into the hormonal regulation of estrous cycle has found practical application to artificially regulate the cycle in farm animals and to manage or treat fertility related problems. However, current understanding of genomic factors underlying fertility is limited and this obstructs the development of novel genomic tools and managemental strategies for improving and optimizing reproductive performance, such as biomarkers to monitor the fertility status of cows. Studying the genomic factors underlying fertility may help to optimize nutritional or management systems that improve reproductive performance [3] and also to explain the genetic basis for the decline in several fertility traits of high producing dairy cows. Currently it is known that this decline may be partly attributed to physiological adaptations by the cow to high milk production [4].

Among the fertility traits, the expression of estrous behavior (heat), a key fertility trait that marks the fertile period in cows, has decreased both in duration and intensity over generations of cows selected for high milk yield [5]. Short heat periods and the absence of clear behavioral signs of heat cause farmers to fail to detect heat or to misjudge the optimum time of insemination of their cows, resulting in financial losses due to prolonged interval from calving to first insemination, reduced conception rates and increased calving intervals.

In an effort to understand the genomic regulation of estrous behavior expression in dairy cows, a microarray experiment was set up to study gene expression levels in 4 different brain areas and the anterior pituitary of cows sacrificed at either the start of estrus (day0 of estrous cycle) or at mid-cycle (day12). Differential gene expression analysis between day0 and day12 cows for each of these tissues revealed a limited number of significant genes in the anterior pituitary alone and none in the brain areas (detailed results not reported here). When the trait of interest is quantitative, as in this case (estrous behavior quantified as heat score), the grouping of individuals into qualitative classes dilutes the available information. Therefore, the association between gene expression and phenotypic trait may be better analyzed using the individual quantitative measurements. Using this approach in an earlier study, we identified a set of a few hundred probes out of approximately 24,000 probes on a bovine microarray corresponding to genes

whose level of expression in the anterior pituitary of experimental cows associated with the degree to which the individual cows expressed estrous behavior [6]. Among these probes were genes encoding hormones like FSH and prolactin, whose roles in estrous regulation are well-known. Further, biological processes relevant to estrous behavior were over-represented in this set of genes. These results give confidence in the association analysis methodology followed, though experimental validation is needed to determine to what extent the associated genes regulate estrous behavior. In addition to the anterior pituitary, it is likely that a number of areas of the cow brain also have genes whose levels of expression associate with the degree to which cows express estrous behavior. Studies in rodents have revealed that estrogen dependent female reproductive behavior happens via well-orchestrated genomic responses in the forebrain with the hypothalamus playing a major role [7,8]. Areas in the limbic region of the forebrain like the amygdala and hippocampus were found to have functions related to sexual behavior and associated emotional responses [9,10]. As yet, there have been no studies in cows linking gene expression in the brain to estrous behavior. Identifying and studying genes whose level of expression in different brain areas of cows associate with the degree to which these cows express estrous behavior will help improve our understanding of genomic factors underlying fertility.

The objective here is to identify and study those genes that associated with estrous behavior, among genes expressed in four bovine brain areas (hippocampus, amygdala, dorsal hypothalamus and ventral hypothalamus), either at the start of estrous cycle, or at mid cycle, or regardless of the phase of cycle.

3.2 Results

The degree of estrous behavior expression was quantified as a cow's average heat score using heat scores recorded from at least two consecutive cycles (Table 3.1). The data from one of the day0 cows was excluded from further analysis because of its high outlier value of 1750 for average heat score, which we attributed to that cow's several consecutive attempts to mount other cows during one observation period. The average heat scores for the remaining 13 day0 cows ranged from 0 to 405, with a mean value of 178.4 (SD 125.7), and the average heat scores for the 14 day12 cows ranged from 2 to 505, with a mean value of 244.7 (SD 175.4). The average heat scores were used with the corresponding gene expression data to run the three analyses per brain area as summarized in Table 3.2.

Additional file 1 lists the associated genes found in each analysis and describes the pattern of association between gene expression and heat score for each of the genes in the list. The patterns noticed were: linear (positive or negative slopes) or

Table 3.1 Average heat scores of the experimental cows sacrificed at day0 or day12 of estrous cycle.

day0 Cow Nr	Heat score	day12 Cow Nr	Heat score
d0_5006	0	d12_1528	2
d0_9284	37	d12_9303	4
d0_3739	43	d12_8860	5
d0_8855	53	d12_1773	75
d0_1194	137	d12_8873	157
d0_1821	175	d12_7724	198
d0_8870	191	d12_1520	257
d0_5507	200	d12_7942	275
d0_1786	206	d12_1607	318
d0_3472	246	d12_1822	368
d0_6487	248	d12_1638	383
d0_3747	378	d12_6956	404
d0_7008	405	d12_8857	475
d0_5125 *	1750	d12_2540	505

* Data from this cow was not used in the analysis due to outlier heat score value.

Table 3.2 Description of the three analyses and their objectives.

Analysis	Data	Objective
day0	Gene expression data from the tested brain area of day0 cows and their average heat scores	To identify genes of which the expression in the tested brain area at the start of estrus was associated with estrous behavior
day12	Gene expression data from the tested brain area of day12 cows and their average heat scores	To identify genes of which the expression in the tested brain area around mid-cycle (diestrus) was associated with estrous behavior
day0 +day12	Gene expression data from the tested brain area of day0 and day12 cows and their average heat scores	To identify genes of which the expression in the tested brain area was associated with estrous behavior regardless of the phase of estrous cycle

quadratic (positive/convex shaped or negative/concave shaped curves). The total number of heat score associated probes found common to all 5 Gibbs sampling chains per analysis in each brain area is provided in Figure 3.1. Additional file 2 depicts Venn diagrams that show the number of overlapping probes between

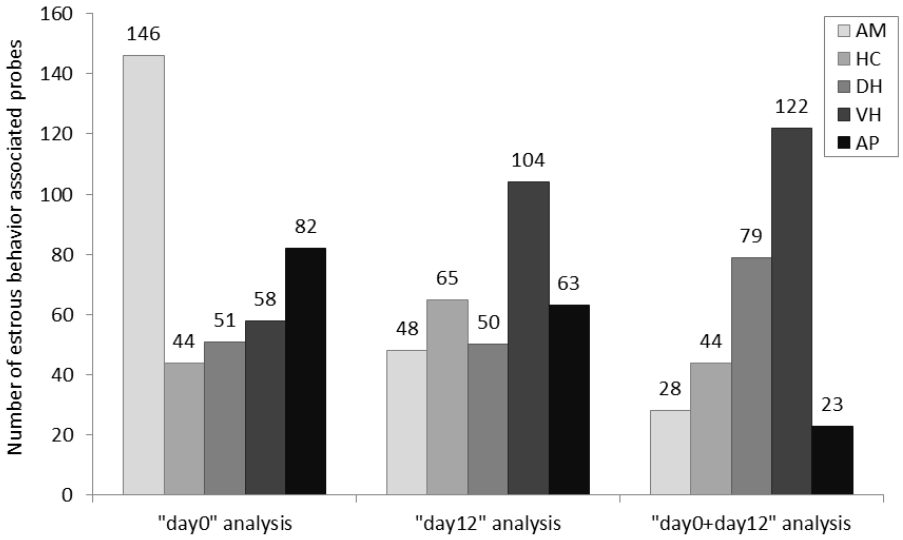


Figure 3.1 Number of estrous behavior associated probes found in 4 brain areas and anterior pituitary (AM-Amygdala; HC-Hippocampus; DH-Dorsal Hypothalamus; VH-Ventral Hypothalamus; AP-Anterior Pituitary).

the different brain areas per analysis. The overlap was highest between the DH and VH and then between the AM and HC. The figure 3.1 and the additional file 2 includes the results of the re-analysis on the AP as well, for which the number of associated probes found for the three analyses (82, 63 and 23 for day0, day12 and day0+day12 respectively) were now considerably lower than the numbers (177, 142 and 118 for day0, day12 and day0+day12 respectively) from the association analysis with third order polynomials and one Gibbs sampling chain per analysis as done earlier [6]. For the different brain areas, the percentage of associated probes found common to all chains per analysis varied from 50-80%. For day0, a relatively high number of heat score associated probes were detected for AM and AP (146 and 82 respectively) whereas for day12 this was true for VH (104). For DH, the numbers were approximately equal at day0 and day12 (51 and 50 respectively). Figure 3.2 provides the association patterns for three genes whose expression values were found to be associated with heat score at day0. For *AVP* and *MCHR1*, a linear trend was observed in HC and AM respectively whereas for *OXT*, a quadratic trend was observed in DH.

Of the total 640 unique probes found associated with heat score in all analyses and brain areas, 372 had gene annotations. Sets of genes associated with functional categories which we group as: "transcription and regulation of gene expression",

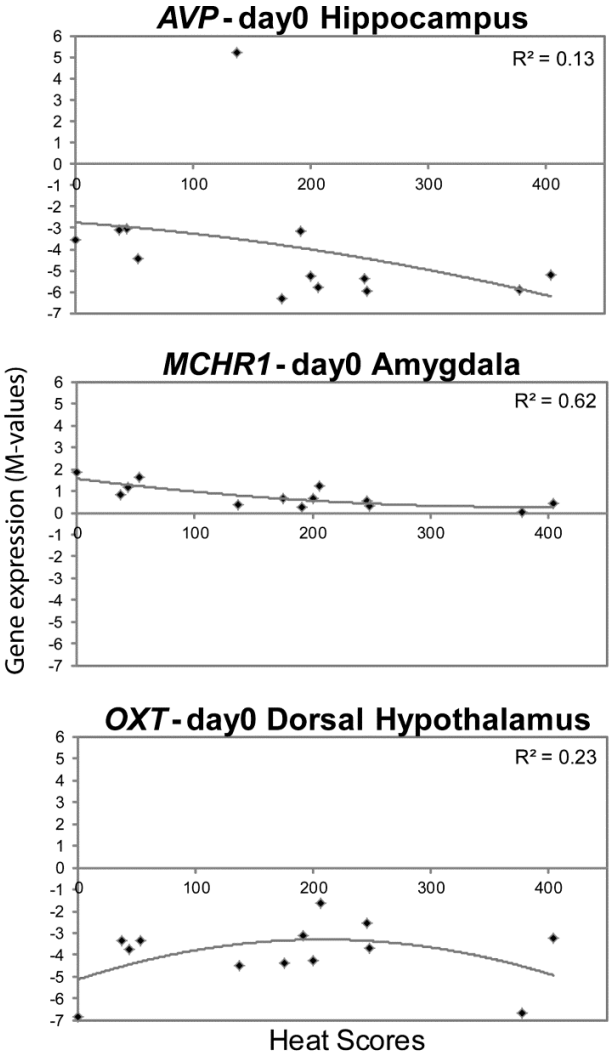


Figure 3.2 Association patterns of gene expression profile with heat scores for three estrous behavior associated genes at day0.

"detection of and responses to stimuli" and "signaling pathways" each made up about 15% of the genes that were identified. In the category "detection of and responses to stimuli" we included the genes involved in neurotransmission through encoding for neurotransmitters/hormones (*OXT, POMC, AVP, CCK, CGA*) or neurotransmitter receptors (*GABRA6, HTR2A, MCHR1, CHRM1, CHRM3, DRD2, CHRNA5*) and metabolizing enzymes (*PTGDS, PTGIS, PTGR1, ACHE, SULT4A1*). We

Table 3.3 Significantly enriched Gene Ontology Biological Process terms in the estrous behavior associated gene lists*

Gene Ontology Biological Process term	day0	day12	day0+day12
Behavior related terms			
regulation of female receptivity; reproductive behavior in a multicellular organism; female mating behavior; maternal aggressive behavior; maternal behavior; parental behavior	HC	DH	HC
grooming behavior			HC
locomotory behavior	AP^{+}		
behavior	AP^{+}		
Neurotransmission and signaling related term			
neurological system process	AP^{+}		
regulation of neurotransmitter levels	AP		
transmission of nerve impulse	AP^{+}		
regulation of synaptic transmission	DH		
cell surface receptor linked signal transduction		DH^{+}	
cell-cell signaling	HC^{+}		
second-messenger-mediated signaling			VH^{+}
response to cAMP	DH		
Wnt receptor signaling pathway	DH		
phosphoinositide-mediated signaling			VH
Ion regulation related terms			
ion transport	AP^{+}		
elevation of cytosolic calcium ion concentration			DH
negative regulation of ion transmembrane transporter activity	AM		
cytosolic calcium ion homeostasis	AM		

* Only terms with p < 0.10 and FDR < 0.20 are included. Terms typical of other organs (these terms appeared here due to genes with different functions in different organs) have been removed from this summary table. Where the term is supported by 5 or more genes in the gene list, the corresponding brain area is followed by a $^{+}$ sign whereas in cases where the term is supported by only 2-4 genes, the sign is absent.

could also identify 5 more sets of genes, each consisting of between 5 to 10% of the total genes. These included genes associated with: "transport and localization" (positioning of a substance or cellular entity, like sorting nexin family member genes), "transporter activity" (e.g. calcium channels and Na+/K+ transport), "metabolism" (e.g. glucose and amino acid metabolism or steroidogenesis), "cell

cycle" (here including processes linked to DNA folding and repair) and "multi-organism processes". In the last category we included immune system related genes like *CTLA4, IL1RL1, MARCO, FCRLA* and *IL33*.

A detailed list of all significant enriched GO and KEGG pathway terms ($p < 0.10$) found in the different brain areas and analyses is provided in Additional file 3. For illustration, a summary of only those GO biological processes that cleared a FDR cutoff of 20% is provided in Table 3.3. Several relevant processes related to behavior, neurotransmission and signaling, and ion regulation were found especially in day0. The presence of enriched processes related to behavior like 'grooming behavior', 'regulation of female receptivity' and 'female mating behavior' within the associated genes found in HC indicates the key role of the genes *AVP* and *OXT* that contribute to these processes. Terms related to neurotransmission and signaling and the associated ion regulation terms too have biological implications related to estrous behavior as evident from the discussion below on some of the genes that contribute to these processes.

3.3 Discussion

Variation in the behavioral trait to express estrous behavior was found to be associated with variation in the expression of genes in the cow's brain areas: DH, VH, HC and AM and also the AP, an endocrine gland pivotal in synchronizing estrous behavior with hormonal and ovarian events preceding ovulation. The choice of these tissues for this study was based on their reported involvement in regulating female sexual behavior. Although genes differ in their influence on specific traits, it is tempting to consider the AM to be of relatively higher importance for regulating estrous behavior as the largest number of associated genes was found here at day0 and because AM is known for its central role in regulating emotions. Key genes and biological processes as identified from the lists of heat score associated genes are discussed next and linked to estrous behavioral expression, though the links are not always as expected.

Genes and biological processes associated with estrous behavior in line with expectations

An association with heat scores was detected for the *CGA* gene, which encodes for the alpha subunit of glycoprotein hormones (FSH, LH, TSH), with associations being time and brain area specific. It seems that cows with clear expression of estrous behavior have relatively high expression of this gene in the hypothalamus around mid-cycle and low expression around start of estrus.

The *OXT* gene was found associated in the HC and DH whereas the *AVP* gene was found associated in the HC and AM. The known functional properties of these

genes contributed to several enriched GO terms in the DAVID analysis, especially those related to female mating behavior. Oxytocin is released within the brain where it acts on specific oxytocin receptors to elicit effects like female sexual receptivity, grooming behavior and partner bonding [11]. In the presence of estrogen, oxytocin exerts an anxiolytic effect thereby favoring courtship and mating [12,13]. Similar to oxytocin, *AVP* is associated with sexual behavior and bonding [14] and its expression is under control of the steroids progesterone and estrogen [15,16]. In mice, the absence of estrogen receptors (ERα and ERβ) impairs social recognition similar to the effect of *OXT* gene deletion [17]. Studies with rodents and humans [18] demonstrate that oxytocin and vasopressin modulate complex socio-sexual behavior, typically under the influence of reproductive steroid hormones. The present association of *OXT* and *AVP* in several brain areas is in line with the above mentioned findings and suggests a major influence of these genes on estrous behavior expression in dairy cows.

The genes *CCK, POMC, MCHR1, GABRA6, HTR2A* and *DRD2*, which associated with heat score in at least one brain area, are known to modulate emotional states like anxiety and satiety [19-21] or even sexual motivation. In sheep, dopamine-mediated D2 receptor (*DRD2*) signaling in the mediobasal hypothalamus affects female sexual motivation and receptivity [22]. Interactions between monoamines (dopamine, serotonin, noradrenaline) and steroid hormones play a major role in the integration of reproductive behavior and gonadal function [23]. The perception and awareness of male-related cues differs with the stage of estrous cycle, with releases of monoamines (linked to *HTR2A* and *DRD2*) and gamma-aminobutyric acid (GABA) (linked to *GABRA6*) in the mediobasal hypothalamus being triggered by such cues only when ewes are in estrus [24]. We found serotonin receptor 2A (*HTR2A*) associated with heat score at day12 in VH. Studies in female rats and hamsters have shown the inhibitory and facilitatory effects of serotonin receptor agonists and antagonists on the hypothalamic regulation of sexual receptivity [25,26], and this regulation is also mediated by GABAergic neurons interacting with serotonin containing neurons [27].

Noteworthy, are the heat score associated genes (*PTGDS, PTGIS, PTGFR*) that regulate prostaglandin functioning. In the central nervous system, prostaglandins are involved in functions like thermoregulation and influencing neuronal morphology. Prostaglandins are also known to be under the influence of estradiol [28] and are capable of directly affecting neurons that synthesize and secrete gonadotropin-releasing hormone [29].

Another heat score associated gene, *TTR*, a carrier of thyroid hormone, is known to influence anxiety [30], behavioral activity [31] and mental functions [32]. The

melanin-concentrating hormone receptor, *MCHR1*, plays a role in metabolic rate and feed intake [33]. Changes in anxiety behavior and feeding motivation are likely to facilitate mating.

The present association between heat scores and the expression of *ACHE* and several cholinergic receptors (*CHRM1, CHRM3* and *CHRNA5*) may be explained by the effect of the neurotransmitter, acetylcholine on arousal, plasticity and reward. The products of the muscarinic cholinergic receptor genes, *CHRM1* and *CHRM3*, are G_q-protein coupled receptors whose activation releases intracellular Ca^{2+} via the phospholipase C - inositol 1,4,5-trisphosphate signaling pathway [34]. The genes for phospholipase C and inositol triphosphate kinase (*PLCB2*, *ITPKA*) and several protein kinases were also found associated to heat score. These findings may be explained based on the hypothesis put forward by Kow and Pfaff [35] that the membrane actions of estrogen can modulate the genomic actions of estrogen and that this transcriptional potentiation was mediated via signaling pathways requiring the activation of certain protein kinases and increased intracellular Ca^{2+}.

Genes and biological processes unexpectedly associated with estrous behavior

The finding of expected processes related to estrus as described in the previous section supports assumed neurophysiological mechanisms underlying female sexual behavior in dairy cows. Here, examples are given of more novel candidate genes and mechanisms. The *TAC3* gene encodes the protein tachykinin (or neurokinin B), which in humans has been considered a critical regulator of gonadotropins (LH, FSH) via regulating GnRH secretion [36]. The present findings encourage further investigation towards the importance of tachykinin associated mechanisms in dairy cow fertility. Behavioral changes during estrus represent changes in central perception and processing of information, i.e. cognition, and some of the genes that associated with heat scores have been linked to cognition, e.g. *PEBP1, MOBP, LTA4H* and *KCNN2*. *PEBP1* has been suggested to be involved in chronic stress-induced memory impairment [37], *MOBP* has been linked to mood disorders [38], *LTA4H* to depression in women [39], and *KCNN2* to anxiety and stress responses [40].

The gene *LIPN1* in mice seems to establish a cross-talk between reproduction and metabolic events [41] and was here associated with heat scores. Also the *POU1F1* gene, which encodes transcription factors involved in activation of growth hormone and prolactin, was found associated with heat score in the AP on day0. This gene may contribute in part to the generally observed reduction in estrous behavior in high producing cows. The heat score associated gene *GARNL1* is noted

here because it has been linked to egg productivity in laying hens [42] and, thus, may have a fertility related function.

A number of heat score associated genes have been linked to the immune system, e.g. *CTLA4, IL1RL1, MARCO, FCRLA, IL33, CCL26* and *CXCL10*, indicating the importance of cell-cell interactions. It has been shown that immunoglobulin superfamily proteins may play important roles in brain developmental processes and the functioning of neuronal networks in adults because they provide the ideal structure for protein-protein interactions and, thus, cell-cell interactions [43]. Remodeling of synaptic networks, which occurs during estrogen promoted female sexual behavior [7], may also be facilitated by immunoglobulins. Together with the many associated genes that are known to regulate cell fate, activity and morphology, this seems to underline the importance of neural tissue plasticity in the appropriate expression of estrous behavior.

The gene product Sodium/potassium-transporting ATPase subunit alpha-3 (*ATP1A3*), which was found to be associated to heat score at estrus in AP in our earlier study was also found in VH. *ATP1A3* has been implicated in rapid-onset dystonia parkinsonism (RDP), characterized by sudden onset of neurological symptoms over hours to a few days [44], suggesting a role in the sudden onset of behavioral changes like during estrus.

Estrous behavior associated genes expressed at estrus and mid-cycle

Including data from both day0 and day12 cows in a combined analysis not only revealed genes that were associated with estrous behavior regardless of phase of cycle but also resulted in greater statistical power and helped reveal some associated genes which could not be found in the separate analysis. For example, *DRD2* (dopamine receptor D2) gene has been found associated in DH. However, care needs to be taken to interpret the results of the combined analysis where an associated gene was also found in one or both of the separate analyses. There could be cases where the day0+day12 combined analysis found an associated gene because of the effect of a strong association found in one of the separate analyses. An example of this case is *HTR2A* which did not associate with heat scores on day0, but did on day12 (R^2 = 0.27) and also weakly (R^2 = 0.1) in the day0+day12 combined analysis, probably as a carryover effect.

Interesting observations were also made by investigating associated genes that appeared in several brain areas and analyses. For example, the gene *SLC17A7* found in VH at day0 and in the combined analysis and also in the AP, is known to mediate the uptake of glutamate into synaptic vesicles at presynaptic nerve terminals of excitatory neural cells and may also mediate the transport of inorganic

phosphate [45]. The gene could be a contributor to neurotransmission associated with estrous behavior expression and was one of the genes contributing to several ion transport related GO terms found enriched in the DAVID analysis. The gene *NKD1*, found in several brain areas across all analyses, was seen to contribute to several GO terms related to binding and signal transduction. It has been reported to have a role as an antagonist of *Wnt* signaling pathway which may influence the development of neurons in dorsal midbrain [46], suggesting again a link between neuronal plasticity and estrous behavior. The gene *ANO8*, also found associated in several brain areas and analyses, is a member of the anoctamin family which has been implicated in calcium ion-activated chloride channels that perform several important functions including neuronal excitability.

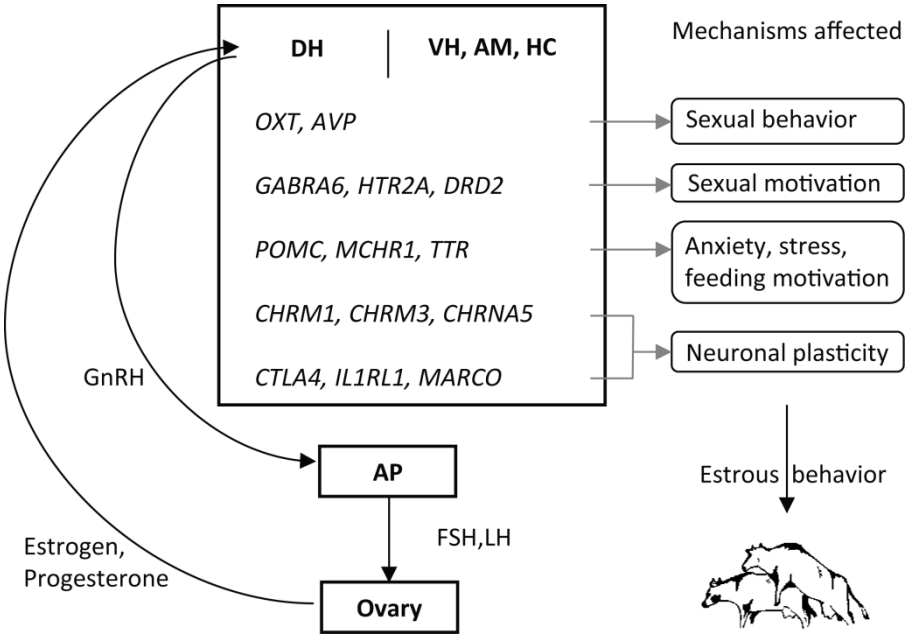


Figure 3.3 Schematic representation of key genes found associated with estrous behavior and their relationship with estrus. Some of the key genes whose expressions in the brain areas were found associated with estrous behavior are depicted here along with mechanisms affected during estrus that are important for estrous behavior. GnRH released from the DH stimulates release of FSH and LH from the AP which in turn influences the ovary to release estrogen and progesterone (Hypothalamic-pituitary-gonadal axis). In the brain areas, estrogen influences expression of *OXT* and *AVP*, which play major roles in the regulation of sexual behavior. The neurotransmitter receptors, *GABRA6*, *HTR2A* and *DRD2* affect sexual motivation, whereas *CHRM1*, *CHRM3* and *CHRNA5* along with "immune related" genes like *CTLA4*, *IL1RL1* and *MARCO* affect neuronal plasticity. The genes *POMC*, *MCHR1* and *TTR* are involved in altering anxiety, stress and feeding motivation.

To summarize, several genes were found here whose expression levels in the brain areas were associated with the degree to which cows express estrous behavior. For some of these genes, there is a known function relating them to processes regulating estrous behavior, while for others, the current association suggests such a relation. We propose that the genes *OXT* and *AVP* play major roles in regulating estrous behavior along with genes affecting neurotransmission and neuronal plasticity. Genes, whose expression in mid-cycle associated with estrous behavior, may contribute to preparing the cow for the next estrus. Figure 3.3 depicts the key findings. This study may assist in the search for biomarkers for estrus detection or in screening for the most likely genes within QTLs associated with fertility. Further research with more animals sampled at multiple time points in the estrous cycle may improve our understanding of the dynamic regulation of estrous behavior over the cycle. Nevertheless, this first study provides an understanding of some of the genes and processes related to expression of estrous behavior in dairy cows.

3.4 Conclusions

The study predicted estrous behavior associated genes in the 4 brain areas at two time points of the estrous cycle. These included genes expected to be related to estrous behavior as they influence states like socio-sexual behavior, anxiety, stress and feeding motivation (*OXT, AVP, POMC, MCHR1*), but also genes whose association with estrous behavior is novel and warrants further investigation. Studying these genes and the processes they control improves our understanding of the genomic regulation of estrous behavior expression, ultimately leading to better management strategies or tools to improve or monitor reproductive performance.

3.5 Methods

Data recording, sample isolation and microarray hybridization

Estrous behavior was recorded in 28 healthy Holstein Friesian cows from 30 days in milk (DIM) onwards till their time of sacrifice which varied between 77 and 139 DIM. The estrous behavior recorded in these cows were quantified as heat scores and the scores from multiple consecutive cycles per cow were averaged to obtain the average heat score per cow [6]. Of the 28 cows, 14 were sacrificed at the start of their estrous cycle (day0) and 14 around mid-cycle (day12). Following sacrifice, brain tissue samples were collected from 4 brain areas: amygdala (AM), hippocampus (HC), dorsal hypothalamus (DH) and ventral hypothalamus (VH). Since the brain areas are not clearly demarcated, a standardized system was used for positioning and dissecting the brain which ensured that the collected samples

representing each brain area in all cows were identical. The procedure followed for collecting brain samples is described in detail in Additional file 4. The study was approved by the Animal Care and Ethics Committee of the Animal Sciences Group of Wageningen University and Research Centre, Lelystad.

RNA extracted from brain tissue samples were hybridized on Bovine 24K oligonucleotide (70-mer) microarrays designed and produced by the Bovine Oligonucleotide Microarray Consortium (BOMC), USA (http://www.bovineoligo.org/). This array was amongst the few whole genome bovine spotted microarrays available at the time this experiment was performed in the year 2007-08. The choice of microarray technology over RNA sequencing, which was not a well-established technology then, was appropriate for fulfilling the objectives of this study in a precise and cost effective manner. The procedures followed for RNA preparation and microarray hybridization were as described in our earlier study [6]. The purity of the total RNA was assessed using the A260/280 ratios given by NanoDrop spectrophotometer and found to have a ratio above 1.8 indicating good quality RNA. In addition, we also performed agarose gel electrophoresis for a visual inspection of the RNA integrity. There was no indication of degraded RNA as deduced from the intensities of the 28S and 18S bands. Based on the combination of results of NanoDrop and gel electrophoresis, we were satisfied with the quality of RNA extracted by our extraction procedure. Five µg RNA was used per labelling using the RNA MICROMAX TSA labelling and detection kit (Perkin-Elmer). A total of 224 arrays (i.e. 28 cows × 4 brain areas × 2 dye swaps) were prepared in a common reference design with the dye labels swapped between individual samples from each brain area and a reference sample consisting of equal proportions of RNA from all four brain areas as well as AP of all cows. Processed slides were scanned using GenePix 4200A (Molecular Devices), with identical settings and the images processed using GenePix Pro 6 software (Molecular Devices). Regarding microarray quality and validation, a recent re-annotation of the probes in this array following the drafting of the bovine genome ensured that only good quality probes were considered in further analysis (details in the second following section). All array hybridizations and processing were done in a standard lab with experienced personnel using a standard microarray platform. Moreover, the experimental design used 14 biological replicates per time point which is relatively large and sufficient for a sound statistical analysis and the use of dye swap technical replication also improves the analysis. Given the above facts, we believe that a PCR validation of the individual probes was not required, the more so because we analysed clusters of genes instead of individual genes.

The array experiment data is publicly accessible on ArrayExpress (accession: E-TABM-916 on http://www.ebi.ac.uk/arrayexpress).

Microarray data analysis and identification of estrous behavior associated genes

Microarray data pre-processing and analysis was done using the LIMMA (Linear models for microarray data) package [47] within Bioconductor project [48] of R statistical programming language (http://www.r-project.org), identical to the procedures described and used in our earlier study [6]. Gene expression levels expressed as M-values (log differential-expression ratio of sample versus reference) were thus obtained for each brain area per cow. For each gene, we then modeled its expression level across all cows as a function of the cows' average heat score. For this we used the Bayesian hierarchical mixed model developed by Jia *et al.* [49] that employs orthogonal polynomials to quantify linear and non-linear associations between quantitative phenotypes and gene expression data. We used the algorithm of Jia *et al.* [49] coded in SAS® language which the authors kindly provided. The program was run on SAS® software, Version 9.1 of the SAS® System for Windows. Similar to the approach in our previous study on the AP [6], here we used the SAS program for three separate analyses on data from each of the 4 brain areas to achieve the objectives of this study (summarized in Table 3.2). However, in contrast to the procedure followed in our earlier study, where gene expression was modeled as a function of average heat score with third order polynomials, here we used second order polynomials. The Bayesian hierarchical mixed model used here remains the same as was explained in our earlier study [6]. However, we limited our analyses to two orders, linear and quadratic, because the resulting relationships for third order polynomials were observed to be less reliable considering the low number of animals used in this study. For each gene, posterior probabilities were obtained for the two regression coefficients to be different from zero. Genes were considered to have a significant association when the posterior probability of at least one of the regression coefficients was larger than 0.80, thereby limiting the FDR to below 1% [50]. We performed the Gibbs sampling chains multiple times (arbitrarily set to 5 chains with each chain running 10,000 iterations of which 5000 were burn-in) for each analysis of a particular brain area and only those genes that appeared significant in all chains were considered in the functional analysis (see next section). By doing so, we limited the variations due to Gibbs sampling.

Functional analysis of estrous behavior associated genes

The original annotation of the bovine 24 K oligonucleotide microarray provided by

BOMC dates back to June 2007. For our analysis, we used the bovine oligonucleotide array probe re-annotation (Version 5) based on Ensembl (http://www.ensembl.org) release 56 (October 2009) provided on the EADGENE website by the authors of the oligo-set re-annotation pipeline, sigReannot [51]. For the re-annotation, out of the 23,496 probes (all control probes removed) on the bovine oligonucleotide array, only 16,620 probes that were assigned a quality score between 1 and 4 based on their specificity to hits on the bovine genome were considered. Probes with quality scores between 5 and 7 had either no hits or multiple hits and were not annotated as they were not specific.

For gaining insight into biological processes underlying estrous behavior, we performed functional analyses of the sets of estrous behavior associated genes (study sets) identified in the 3 analyses for the 4 brain areas and the anterior pituitary. To increase the accuracy of functional analysis, we used the re-annotated probes information. The probe re-annotation also provides the Ensembl gene ID of the orthologous human genes for 14,585 probes, which we used in the functional analysis instead of bovine genes so as to benefit from the greater gene annotation information available for human species.

Functional analysis was done using DAVID bioinformatics resources [52,53], a freely available web-based tool (http://david.abcc.ncifcrf.gov/) that integrates biological data from several sources including GO [54] and biological pathway databases. In order to get an indication of the major processes over-represented among the annotated probes (genes) in the study sets (all linear/quadratic associated genes in each of the 3 analyses for the 4 brain areas), we interpreted the significance of the DAVID results based on EASE score (p value derived from a modified Fisher's exact test) [55] threshold set at 0.10. We did not consider multiple testing correction as it was too conservative for our purpose of discovery of biological themes in our study sets. Genes in the study sets were tested for enriched GO terms and KEGG [56] pathway terms using DAVID functional annotation tool. The population set against which the study set genes were tested consisted of the Ensembl IDs of the orthologous human genes, of which 11,589 remained after removing duplicates.

Additional material

Additional file 1. Association patterns between gene expression and heat score of estrous behavior associated genes identified in the four brain areas and anterior pituitary.

http://www.biomedcentral.com/content/supplementary/1471-2164-12-200-s1.pdf

Additional file 2. Venn diagrams showing the number of overlapping probes between the different brain areas per analysis.

http://www.biomedcentral.com/content/supplementary/1471-2164-12-200-s2.pdf
The figures in brackets represent the total number of estrous behavior associated probes found in each brain area per analysis. Here, AM - Amygdala; HC - Hippocampus; DH - Dorsal Hypothalamus; VH - Ventral Hypothalamus; AP - Anterior Pituitary.

Additional file 3. Significantly enriched Gene Ontology and KEGG pathway terms in the estrous behavior associated gene lists.
http://www.biomedcentral.com/content/supplementary/1471-2164-12-200-s3.pdf
Table shows all terms with $p < 0.10$ and includes gene ontology terms in the 3 categories: Biological Process, Molecular function and Cellular component.

Additional file 4. Collection of brain samples and pituitary from the experimental cows.
http://www.biomedcentral.com/content/supplementary/1471-2164-12-200-s4.pdf
Procedure followed for collection of anterior pituitary and brain samples: amygdala, hippocampus, dorsal hypothalamus and ventral hypothalamus (with pictures).

List of abbreviations

AM: Amygdala; AP: Anterior Pituitary; BOMC: Bovine Oligonucleotide Microarray Consortium; DAVID: Database for Annotation, Visualization and Integrated Discovery; DH: Dorsal Hypothalamus, EADGENE: European Animal Disease Genomics Network of Excellence for Animal Health and Food Safety; FDR: False Discovery Rate; FSH: Follicle stimulating hormone; GnRH: Gonadotropin-releasing hormone; GO: Gene Ontology; HC: Hippocampus; KEGG: Kyoto Encyclopedia of Genes and Genomes; LH: Luteinizing hormone; QTL: Quantitative trait loci; SD: Standard Deviation; TSH: Thyroid stimulating hormone; VH: Ventral Hypothalamus.

Acknowledgements

Regarding the animal experiment, the support of Jan Bloemert and co-workers of the research farm Waiboerhoeve is highly appreciated and we are indebted to Henk Sulkers and Diana Baumann for their assistance. We thank Norbert Stockhofe for helping to set up the study and Ad Korevaar for his crucial role in tissue collection. Zhenyu Jia and Shizhong Xu are acknowledged for kindly providing the SAS code for quantitative trait associated gene expression analysis and helpful initial instructions. Prof. Dr. Martien Groenen is thanked for valuable suggestions while writing the manuscript. The work was co-financed by the international cattle breeding co-operative, CRV and by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2006-016250 (SABRE) and FOOD-

CT-2004-506416 (EADGENE). The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

Authors' contributions

AACW performed the microarray experiments. AK performed the data analysis and drafted the manuscript. HAM and RFV helped with the statistical analysis. BB, HW, MFWP and MAS helped with the biological interpretation of the results. All authors read and helped in improving the manuscript.

References

1. Butler WR: Nutritional interactions with reproductive performance in dairy cattle. Anim Reprod Sci 2000, 60-61:449-457
2. Butler WR: Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. Livest Prod Sci 2003, 83(2-3):211-218.
3. Veerkamp RF, Beerda B: Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 2007, 68(Supplement 1):S266-S273.
4. Lucy MC: Reproductive loss in high-producing dairy cattle: Where will it end? J Dairy Sci 2001, 84(6):1277-1293.
5. Lopez H, Satter LD, Wiltbank MC: Relationship between level of milk production and estrous behavior of lactating dairy cows. Anim Reprod Sci 2004, 81(3-4):209-223.
6. Kommadath A, Mulder HA, de Wit AAC, Woelders H, Smits MA, Beerda B, Veerkamp RF, Frijters ACJ, te Pas MFW: Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows. Animal 2010, 4(08):1297-1307.
7. Pfaff D: Hormone-driven mechanisms in the central nervous system facilitate the analysis of mammalian behaviours. J Endocrinol 2005, 184(3):447-453.
8. Mong JA, Pfaff DW: Hormonal symphony: steroid orchestration of gene modules for sociosexual behaviors. Mol Psychiatry 2004, 9(6):550-556.
9. Gallagher M, Chiba AA: The amygdala and emotion. Curr Opin Neurobiol 1996, 6(2):221-227.
10. Salamon E, Esch T, Stefano GB: Role of amygdala in mediating sexual and emotional behavior via coupled nitric oxide release. Acta Pharmacol Sin 2005, 26(4):389-395.
11. Leng G, Meddle SL, Douglas AJ: Oxytocin and the maternal brain. Curr Opin Pharmacol 2008, 8(6):731-734.

12. McCarthy MM, McDonald CH, Brooks PJ, Goldman D: An anxiolytic action of oxytocin is enhanced by estrogen in the mouse. Physiol Behav 1997, 60(5):1209-1215.
13. Mong JA, Pfaff DW: Hormonal symphony: steroid orchestration of gene modules for sociosexual behaviors. Mol Psychiatry 9(6):550-556.
14. Curley JP, Keverne EB: Genes, brains and mammalian social bonds. Trends Ecol Evol 2005, 20(10):561-567.
15. Kalamatianos T, Kalló I, Goubillon ML, Coen CW: Cellular Expression of V1a Vasopressin Receptor mRNA in the Female Rat Preoptic Area: Effects of Oestrogen. J Neuroendocrinol 2004, 16(6):525-533.
16. Patisaul HB, Scordalakes EM, Young LJ, Rissman EF: Oxytocin, but not oxytocin receptor, is regulated by oestrogen receptor β; in the female mouse hypothalamus. J Neuroendocrinol 2003, 15(8):787-793.
17. Choleris E, Gustafsson JA, Korach KS, Muglia LJ, Pfaff DW, Ogawa S: An estrogen-dependent four-gene micronet regulating social recognition: A study with oxytocin and estrogen receptor-α and -β knockout mice. Proc Natl Acad Sci USA 2003, 100(10):6192-6197.
18. Donaldson ZR, Young LJ: Oxytocin, vasopressin, and the neurogenetics of sociality. Science 2008, 322(5903):900-904.
19. Uhart M, McCaul ME, Oswald LM, Choi L, Wand GS: GABRA6 gene polymorphism and an attenuated stress response. Mol Psychiatry 2004, 9(11):998-1006.
20. Rex A, Marsden CA, Fink H: Cortical 5-HT-CCK interactions and anxiety-related behaviour of guinea-pigs: a microdialysis study. Neurosci Lett 1997, 228(2):79-82.
21. Millington G: The role of proopiomelanocortin (POMC) neurones in feeding behaviour. Nutr Metab 2007, 4(1):18.
22. Fabre-Nys C, Chesneau D, De La Riva C, Hinton MR, Locatelli A, Ohkura S, Kendrick KM: Biphasic role of dopamine on female sexual behaviour via D2 receptors in the mediobasal hypothalamus. Neuropharmacology 2003, 44:354-366.
23. Fabre-Nys C: Steroid control of monoamines in relation to sexual behaviour. Rev Reprod 1998, 3(1):31-41.
24. Fabre-Nys C, Ohkura S, Kendrick KM: Male faces and odours evoke differential patterns of neurochemical release in the mediobasal hypothalamus of the ewe during oestrus: An insight into sexual motivation. Eur J Neurosci 1997, 9:1666-1677.

25. Uphouse L: Female gonadal hormones, serotonin, and sexual receptivity. Brain Res Rev 2000, 33(2-3):242-257.
26. Caldwell HK, Albers HE: The effects of serotonin agonists on the hypothalamic regulation of sexual receptivity in Syrian hamsters. Horm Behav 2002, 42(1):78-84.
27. Luine VN, Wu V, Hoffman CS, Renner KJ: GABAergic Regulation of Lordosis: Influence of Gonadal Hormones on Turnover of GABA and Interaction of GABA with 5-HT. Neuroendocrinology 1999, 69(6):438-445.
28. Amateau SK, McCarthy MM: A novel mechanism of dendritic spine plasticity involving estradiol induction of prostaglandin-E2. J Neurosci 2002, 22(19):8586-8596.
29. Jasoni CL, Todman MG, Han SK, Herbison AE: Expression of mRNAs encoding receptors that mediate stress signals in gonadotropin-releasing hormone neurons of the mouse. Neuroendocrinology 2005, 82(5-6):320-328.
30. Sousa JC, Grandela C, Fernández-Ruiz J, Miguel Rd, Sousa Ld, Magalhães AI, Saraiva MJ, Sousa N, Palha JA: Transthyretin is involved in depression-like behaviour and exploratory activity. J Neurochem 2004, 88(5):1052-1058.
31. Venero C, Guadano-Ferraz A, Herrero AI, Nordstrom K, Manzano J, de Escobar GM, Bernal J, Vennstrom B: Anxiety, memory impairment, and locomotor dysfunction caused by a mutant thyroid hormone receptor α1 can be ameliorated by T3 treatment. Genes Dev 2005, 19(18):2152-2163.
32. Nunez J, Celi FS, Ng L, Forrest D: Multigenic control of thyroid hormone functions in the nervous system. Mol Cell Endocrinol 2008, 287(1-2):1-12.
33. Marsh DJ, Weingarth DT, Novi DE, Chen HY, Trumbauer ME, Chen AS, Guan XM, Jiang MM, Feng Y, Camacho RE, *et al.*: Melanin-concentrating hormone 1 receptor-deficient mice are lean, hyperactive, and hyperphagic and have altered metabolism. Proc Natl Acad Sci USA 2002, 99(5):3240-3245.
34. Billups D, Billups B, Challiss RAJ, Nahorski SR: Modulation of G_q-protein-coupled inositol trisphosphate and Ca^{2+} signaling by the membrane potential. J Neurosci 2006, 26(39):9983-9995.
35. Kow LM, Pfaff DW: The membrane actions of estrogens can potentiate their lordosis behavior-facilitating genomic actions. Proc Natl Acad Sci USA 2004, 101(33):12354-12357.
36. Topaloglu AK, Reimann F, Guclu M, Yalin AS, Kotan LD, Porter KM, Serin A, Mungan NO, Cook JR, Ozbek MN, *et al.*: TAC3 and TACR3 mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. Nat Genet 2009, 41(3):354-358.

37. Feldmann RE, Maurer MH, Hunzinger C, Lewicka S, Buergers HF, Kalenka A, Hinkelbein J, Broemme JO, Seidler GH, Martin E, *et al.*: Reduction in rat phosphatidylethanolamine binding protein-1 (PEBP1) after chronic corticosterone treatment may be paralleled by cognitive impairment: A first study. Stress 2008, 11(2):134-147.
38. Sokolov BP: Oligodendroglial abnormalities in schizophrenia, mood disorders and substance abuse. Comorbidity, shared traits, or molecular phenocopies? Int J Neuropsychopharmacol 2007, 10(04):547-555.
39. Zhao J, Quyyumi AA, Patel R, Zafari AM, Veledar E, Onufrak S, Shallenberger LH, Jones L, Vaccarino V: Sex-specific association of depression and a haplotype in leukotriene A4 hydrolase gene. Psychosom Med 2009, 71(7):691-696.
40. Mitra R, Ferguson D, Sapolsky RM: SK2 potassium channel overexpression in basolateral amygdala reduces anxiety, stress-induced corticosterone secretion and dendritic arborization. Mol Psychiatry 2009, 14(9):847-855.
41. Nadra K, Charles A-SdP, Medard JJ, Hendriks WT, Han GS, Gres S, Carman GM, Saulnier-Blache JS, Verheijen MHG, Chrast R: Phosphatidic acid mediates demyelination in Lpin1 mutant mice. Genes Dev 2008, 22(12):1647-1661.
42. Chih-Feng C, Yow-Ling S, Cheng-Ju Y, Pin-Chi T, Hui-Chiu C, Yen-Pai L: Laying traits and underlying transcripts, expressed in the hypothalamus and pituitary gland, that were associated with egg production variability in chickens. Theriogenology 2007, 68(9):1305-1315.
43. Rougon G, Hobert O: New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. Annu Rev Neurosci 2003, 26(1):207-238.
44. de Carvalho Aguiar P, Sweadner KJ, Penniston JT, Zaremba J, Liu L, Caton M, Linazasoro G, Borg M, Tijssen MAJ, Bressman SB, *et al.*: Mutations in the Na+/K+-ATPase [alpha]3 gene ATP1A3 are associated with rapid-onset Dystonia Parkinsonism. Neuron 2004, 43(2):169-175.
45. Aihara Y, Mashima H, Onda H, Hisano S, Kasuya H, Hori T, Yamada S, Tomura H, Yamada Y, Inoue I, *et al.*: Molecular cloning of a novel brain-type Na+-dependent inorganic phosphate cotransporter. J Neurochem 2000, 74(6):2622-2625.
46. Chittka A, Volff J, Wizenmann A: Identification of genes differentially expressed in dorsal and ventral chick midbrain during early development. BMC Dev Biol 2009, 9(1):29.

47. Smyth GK: Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor. Edited by Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W. New York: Springer; 2005:397-420.
48. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, *et al.*: Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004, 5(10):R80.
49. Jia Z, Tang S, Mercola D, Xu S: Detection of quantitative trait associated genes using cluster analysis. In Evolutionary computation, machine learning and data mining in bioinformatics. Volume 4973/2008. Edited by Marchiori E, Moore JH. Berlin Heidelberg: Springer-Verlag; 2008:83-94.
50. Jia Z, Xu S: Mapping quantitative trait loci for expression abundance. Genetics 2007, 176(1):611-623.
51. Casel P, Moreews F, Lagarrigue S, Klopp C: sigReannot: an oligo-set re-annotation pipeline based on similarities with the Ensembl transcripts and Unigene clusters. BMC Proc 2009, 3(Suppl 4):S3.
52. Dennis G, Sherman B, Hosack D, Yang J, Gao W, Lane HC, Lempicki R: DAVID: Database for annotation, visualization, and integrated discovery. Genome Biol 2003, 4(5):P3.
53. Huang DW, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protoc 2009, 4(1):44-57.
54. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.*: Gene Ontology: tool for the unification of biology. Nat Genet 2000, 25(1):25-29.
55. Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA: Identifying biological themes within lists of genes with EASE. Genome Biol 2003, 4(10):R70
56. Kanehisa M, Goto S: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000, 28:27-30.

4

Gene co-expression network analysis identifies genes and biological processes shared among anterior pituitary and brain areas that affect estrous behavior in dairy cows

Arun Kommadath[1], Marinus F.W. te Pas[1] and Mari A. Smits[1]

[1]Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, P.O. Box 65, 8200 AB Lelystad, The Netherlands

Ready for submission

Abstract

Background

The expression of EB, a key fertility trait in dairy cows, has been declining over the past few decades both in intensity and duration. Knowledge of the genomic factors underlying EB, which is currently lacking, may lead to novel applications to improve fertility. Our objective was to identify genes and biological processes shared among the bovine anterior pituitary (AP) and four brain areas that act together to regulate EB by investigating networks of co-expressed genes between these tissues.

Results

We used a systems biology approach called Weighted gene co-expression network analysis (WGCNA) for defining gene co-expression networks using gene expression data from the following tissues collected from 14 cows at estrus: AP, dorsal hypothalamus (DH), ventral hypothalamus (VH), amygdala (AM) and hippocampus (HC). Consensus modules of co-expressed genes were identified between the networks for the AM-DH, HC-DH, VH-DH, AP-DH and AM-HC tissue pairs. The correlation between the module's eigengene (weighted average gene expression profile) and levels of EB exhibited by the experimental cows were tested. EB correlated modules (absolute correlation > 0.3 and $p < 0.2$) were found enriched for gene ontology terms like glial cell development and regulation of neural projection development as well as for KEGG pathway terms related to brain degenerative diseases. General cellular processes like oxidative phosphorylation, ribosome and biosynthetic processes were found enriched in several correlated modules, indicating increased transcription and protein synthesis. Stimulation of ribosomal RNA synthesis is known from rodent studies to be a primary event in the activation of neuronal cells and pathways involved in female reproductive behavior and this precedes the estrogen driven expansion of dendrites and synapses. Similar processes therefore operate in cows too to affect EB. Hub genes within EB correlated modules (e.g. *NEFL*, *NDRG2*, *GAP43*, *THY1*, *TCF7L2 etc.*) are strong candidates among genes regulating EB expression.

Conclusions

The study improved our understanding of the genomic regulation of EB in dairy cows by providing new insights into genes and biological processes shared among the bovine AP and brain areas acting together to regulate EB. The new knowledge could lead to novel management strategies to monitor and improve reproductive performance in dairy cows.

Key words: gene co-expression, estrous behavior, WGCNA, brain, dairy cow

4.1 Background

Dairy cow fertility is a trait of great economic importance to dairy farmers and dairy industry, therefore losses due to sub-fertility or infertility are of major concern [1]. We now have a basic understanding of the physiological mechanisms that influence the multi-factorial trait of fertility [2], especially the hormones involved and the link with nutrition and energy balance [3]. At the genomic level, gene expression studies on tissues and organs involved in female reproductive processes, embryo development and implantation and maintenance of pregnancy have helped to identify candidate genes and biological processes related to fertility [4-7]. This has contributed to our understanding of the genomic regulation of certain aspects of fertility and may lead to novel applications to improve fertility. However, there is a lack in understanding of genomic factors underlying fertility traits that are regulated by the brain and pituitary, for example, estrous behavior (EB). The expression of EB, a key fertility trait, has been declining both in intensity and duration over the past few decades along with the decline in several other fertility traits [8]. The proper expression of EB is important because it helps the farmer to identify the fertile period in the cows' estrous cycles when artificial insemination has the greatest chance to result in successful fertilization.

In an effort to understand the genomic regulation of EB, we performed an experiment to study gene expression in the anterior pituitary (AP) and four brain areas: dorsal hypothalamus (DH), ventral hypothalamus (VH), amygdala (AM) and hippocampus (HC) at two time points of the cows' estrous cycle: start of estrus (day0) and mid-cycle (day12). The choice of these tissues was motivated by the fact that the anterior pituitary and hypothalamus are part of the hypothalamic-pituitary-ovarian axis and act as important centers controlling the female reproductive cycle and behavior [9, 10] whereas the amygdala and hippocampus have functions related to controlling sexual behavior and associated emotional responses [11, 12]. In earlier studies [13, 14], we focused on the association between gene expression profiles and the trait of showing EB (quantified as heat scores) and successfully identified sets of genes in all five tissues under study that associated with the degree to which the individual cows expressed EB. In these studies, we looked separately per tissue into EB associated gene expression. However, in real life scenario, the expression of a complex trait like behavior may be the result of intricate interactions between the different tissues in response to hormonal, visual, physiological and other signals. Therefore, studying gene co-expression patterns preserved among the AP and brain areas may shed light on biological processes shared among these tissues acting together in regulating EB

expression. It has been shown that gene expression profiles across microarray samples can be highly correlated and genes with similar expression patterns may form complexes, pathways, or participate in common regulatory and signaling circuits [15-18]. Specifically, in a study of brain gene expression data, it was shown that the transcriptomes of human brain regions were robustly organized into modules of co-expressed genes that reflect their underlying cellular composition [19]. Here, modules refer to clusters of highly interconnected genes within a gene co-expression network. Other studies used co-expression networks to screen for genes underlying complex traits, for example, body weight [20]. Therefore, gene co-expression networks constructed from gene expression data can effectively capture the relationships between transcripts which group as modules that reveal biologically meaningful higher-order organization of the transcriptome [21].
The objective of this study was to identify genes and biological processes shared among the bovine AP and four brain areas that act together to regulate EB by investigating networks of co-expressed genes between these tissues.

4.2 Results

Estrous behavior scores and gene expression data

For each cow, the level of EB expression was quantified as its average heat score using EB scores recorded from at least 2 consecutive cycles (Table 4.1). For identifying modules of co-expressed genes correlated to EB, we adjusted the heat score of one of the day0 cows to the level of the next highest cow (405) because of its high outlier value of 1750, which we attributed to that cow's several consecutive attempts to mount other cows during one observation period. The average heat scores for the remaining 13 day0 cows ranged from 0 to 405, with a mean value of 178.4 (SD 125.7).
Following pre-processing of the microarray data, gene expression levels expressed as M-values (log differential-expression ratio between sample tissue and a common reference which is the pool of all tissues) were obtained for each tissue per cow. Of the 16,620 probes on the microarray that were found to be of good quality based on re-annotation, gene expression measures for 13,234 unique genes remained after averaging the expression levels of genes represented by multiple probes (see Table 4.2 for a summary of the steps taken in the analysis and their results).

Consensus modules identified from co-expression networks between tissue pairs

We first checked the microarray data to ensure that the genes and arrays that were selected for detailed analysis did not have excessive numbers of missing values. These checks revealed that weren't any such issues for any of the tissue pairs under

Table 4.1 Estrous behavior expressed as heat scores.

day0 Cow Nr	Average heat score
d0_5006	0
d0_9284	37
d0_3739	43
d0_8855	53
d0_1194	137
d0_1821	175
d0_8870	191
d0_5507	200
d0_1786	206
d0_3472	246
d0_6487	248
d0_3747	378
d0_7008	405
d0_5125 †	1750

† For the analysis done here, heat score of this cow was replaced by that of the next highest scoring cow (i.e. score of 405) due to its unrealistically high outlier value.

Table 4.2 Summary of steps taken in the analysis and their results.

Steps	Results
Process data of 14 pairs of dye-swapped microarrays per tissue collected from 14 cows at day0 of estrous cycle	M-values of 23,496 probes per array
Select good quality probes based on probe re-annotation and average the M-values of probes representing the same gene	16,620 good quality probes per array representing 13,234 genes
Select the top 50% most variable genes per tissue and identify genes shared by each tissue within a pair	Approximately 4000-5000 genes shared per tissue pair - AM-DH, HC-DH, VH-DH, AM-HC and AP-DH
Perform WGCNA on the gene expression data of shared genes within each tissue pair and identify consensus modules	Gene co-expression networks constructed for tissues within each pair and consensus modules identified
Identify consensus modules within tissue pairs whose module eigengenes correlate with estrous behavior (EB) scores	Consensus modules that correlated with EB - AM-DH: 1 of 3, HC-DH: 5 of 10, VH-DH: 0 of 2, AM-HC: 3 of 8, AP-DH: 10 of 23
Test for enriched GOBP and KEGG pathway terms within EB correlated consensus modules	A summary of the significant terms per module are reported in Figure 4.1
Identify hub genes within the EB correlated consensus modules	The top 3 hub genes per module are reported in Table 4.3

study: AM-DH, HC-DH, VH-DH, AM-HC and AP-DH. The clustering of arrays per tissue based on Euclidian distance did not point to any outlier arrays. For network construction, we chose one half of the total 13,234 unique genes on the array that had the most variable expression levels across cows showing different levels of EB. This choice followed by selecting only those genes shared between the tissues in each pair resulted in between 4000 and 5000 genes per analysis, thereby also making the analyses less computationally demanding.

While applying the Weighted gene co-expression network analysis (WGCNA) to define gene co-expression networks using gene expression data per tissue pair, we chose a soft-thresholding power (beta = 12) based on the criterion of approximate scale-free topology. This was because biological networks generally follow a power law and tend to be approximately scale-free. The genes in the network were clustered into modules based on their connectivity strengths with all other genes in the network. The hierarchical clustering dendrograms produced for the consensus module identification in each pair of tissues are given in supplementary Figure S1 (posted at http://edepot.wur.nl/183833). In the applied WGCNA software tool, the consensus modules found in the co-expression networks between tissue pairs were labeled by colors. After excluding the grey module, that is used to hold all genes that do not clearly belong to any other module, the number of modules found in each tissue pair were as follows: AM-DH: 3, HC-DH: 10, VH-DH: 2, AM-HC: 8, AP-DH: 23. The highest number of genes within the consensus modules was observed for the AP-DH pair (n=1904). This was followed by the HC-DH pair at 846, AM-HC pair at 843, AM-DH pair at 225 and VH-DH pair at 197. A comparison of the genes within the consensus modules in the tissue pairs that had the DH in common (AP-DH, AM-DH, HC-DH and VH-DH) revealed that the highest number of genes were shared between the AP-DH and HC-DH pairs and that 59 genes were shared among all four pairs (Venn diagram provided in supplementary Figure S2 posted at http://edepot.wur.nl/183833). The major biological processes enriched within these 59 shared genes were ribosome and translation.

Consensus modules correlated with estrous behavior

The consensus modules in the co-expression networks between tissue pairs that correlated with EB were detected based on the correlation between the module eigengene (weighted average expression profile of all genes within the module) and the heat scores. An absolute correlation above 0.3 was considered significant when the corresponding p-value was below 0.2. The consensus modules that correlated with EB are presented in Table 4.3 along with the corresponding correlations and p-values per tissue. A module may be found significant in one

Table 4.3 Consensus modules whose module eigengenes correlated with estrous behavior.

Nr	Modules	n	Correlation (p-value) with estrous behavior		Top 3 hub genes
	AM-DH		**AM**	**DH**	
1	brown	40	0.14 (0.65)	0.4 (0.16) *	*OLFM1, NEFL, LY6E*
	HC-DH		**HC**	**DH**	
1	magenta	31	0.47 (0.093) *	−0.3 (0.31)	*SLC25A47, TP53I3, CRYAA*
2	red	51	0.62 (0.016) *	0.014 (0.96)	*MATN1, MCM9, MTR*
3	black	45	0.49 (0.078) *	0.026 (0.93)	*CENPI, EXOSC1, DEPDC7*
4	blue	152	0.47 (0.089) *	0.32 (0.28)	*RPL13A, RPS20, RPL18*
5	green	59	0.55 (0.039) *	0.26 (0.37)	*CEP63, KRTAP13-1, AQP3*
	AM-HC		**AM**	**HC**	
1	red	52	−0.052 (0.86)	0.48 (0.081) *	*NUP160, CCL21, NCAPG2*
2	pink	34	−0.33 (0.25)	0.47 (0.091) *	*MED22, GBA, TP53I3*
3	green	77	0.32 (0.28)	0.42 (0.14) *	*RPL18, HSPA8, RPS17*
	AP-DH		**AP**	**DH**	
1	greenyellow	74	0.42 (0.14) *	0.0098 (0.97)	*NGEF, FAM5C, C1QTNF5*
2	midnightblue	58	0.4 (0.15) *	0.37 (0.2)	*NEFL, VSNL1, NDRG2*
3	salmon	71	0.4 (0.16) *	0.35 (0.23)	*RNF128, PODNL1, HS3ST6*
4	black	88	0.4 (0.16) *	0.26 (0.37)	*ENTPD6, MGC134087, PHF13*
5	purple	77	0.39 (0.17) *	0.26 (0.37)	*MRPL40, HSCB, POLR2B*
6	turquoise	343	−0.38 (0.18) *	0.22 (0.45)	*SPG11, FRS2, GPC3*
7	magenta	84	−0.42 (0.14) *	0.17 (0.57)	*KBTBD10, FNDC3A, IBSP*
8	brown	114	−0.39 (0.18) *	−0.13 (0.65)	*CHEK1, CBLL1, FANK1*
9	pink	86	−0.34 (0.25)	−0.39 (0.18) *	*NPY2R, A5PK81,TMEM107*
10	lightcyan	54	−0.38 (0.19) *	0.09 (0.77)	*MAN2C1, TRDN,SLC16A11*

The consensus modules (labeled by colors) in the co-expression networks between tissue pairs whose module eigengenes correlated with estrous behavior scores are presented here. Modules with an absolute correlation above 0.3 and having a p-value below 0.2 (in brackets) are indicated by * symbol. The top 3 hub genes within these modules based on their intra-modular connectivity measures (kME) are also presented. The VH-DH pair did not have any estrous behavior correlated consensus modules.

tissue of the pair but not in the other. This difference in the correlation between the EB scores and module eigengenes for the same module between the individual tissues of a pair is due to the fact that the gene expression profiles per gene are not exactly the same between the tissues. The highest number of significant modules were found in the AP-DH pair i.e. 10 whereas the VH-DH pair had none. The AM-DH pair had only one significant module, the HC-DH pair had five ($p < 0.1$) and the AM-HC had four ($p < 0.1$). Also presented in Table 4.3 are the top 3 hub genes within each EB correlated module i.e. the most highly connected genes in each module based on their module eigengene based intra-modular connectivity measures (kME). The complete list of all genes in all the EB correlated consensus modules, ordered by their kME measures, is presented in supplementary Table S1 (posted at http://edepot.wur.nl/183833). Also presented there are the corresponding gene significance (GS) values which are a measure of the correlation of that gene's expression with EB across all samples per tissue pair.

Tests for enriched Gene Ontology Biological Process (GOBP) terms and KEGG pathway terms in the EB correlated modules revealed significant terms in several modules in all the tissue pairs except the VH-DH. The results are summarized in Figure 4.1 whereas the full list can be found in supplementary Table S2 (posted at http://edepot.wur.nl/183833). Some of the key GOBP terms found in both AM-DH and HC-DH pairs were translation, biosynthetic process and glial cell development. The terms translation and biosynthetic process were also found in the AM-HC pair but in addition, there were terms like reproduction, intracellular signaling cascade and production of molecular mediator of immune response. In all 3 pairs, the KEGG terms included one or more of the following: ribosome, oxidative phosphorylation and Alzheimer's disease. The terms enriched in the AP-DH pair shared less with the other tissue pairs and included GOBP terms like response to caffeine, release of calcium ion and regulation of neuron projection development, regulation of axon diameter and neuron differentiation. The top KEGG terms found here were amyotrophic lateral sclerosis, biosynthesis of alkaloids and Glycolysis/Gluconeogenesis. The 'midnight blue' module in the AP-DH pair was especially interesting as it showed several nervous system related processes. From the biological processes found enriched in the different consensus modules in the five tissue pairs, it seems that processes specifically related to nervous system functioning are few whereas the majority relates to general cellular processes. There were also several modules that did not reveal any significant GOBP or KEGG pathway terms.

As an example, we visualized as a network the 'midnight blue' module found in the AP-DH (Figure 4.2). The nodes in this network correspond to genes in the module

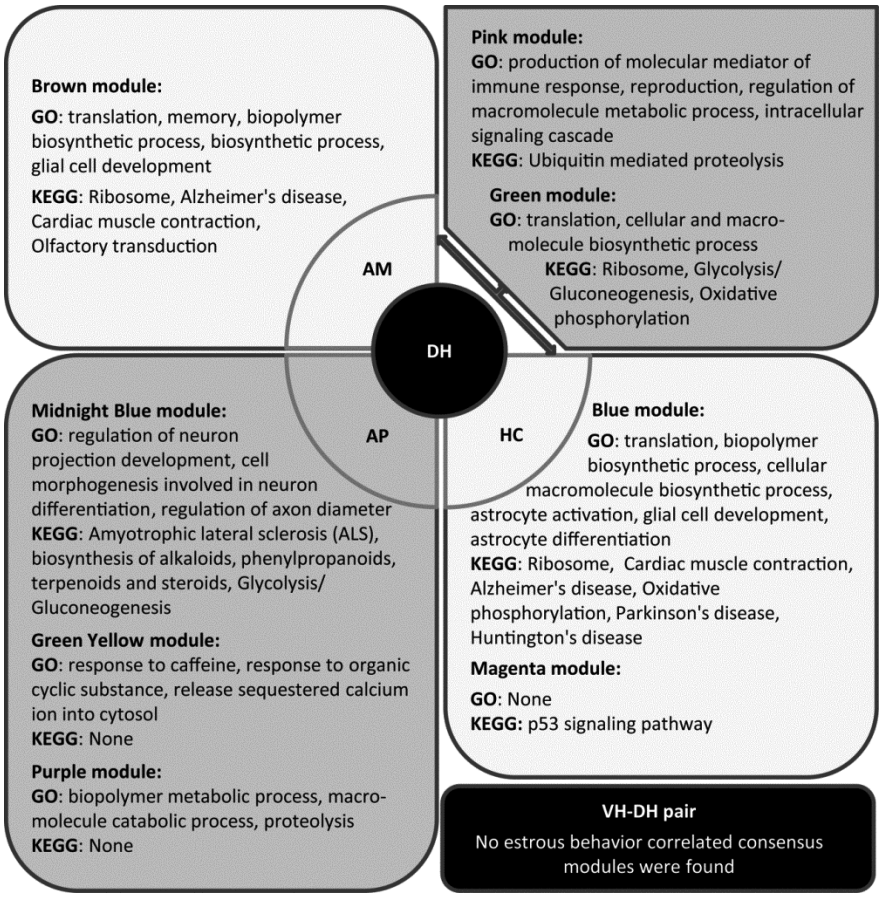


Figure 4.1 Biological processes within consensus modules between tissue pairs that correlated with estrous behavior. The genes in the consensus modules found in co-expression networks between tissue pairs that significantly correlated with estrous behavior scores (absolute correlation > 0.3 and $p < 0.2$) were tested for over-represented GOBP and KEGG pathway terms. Those modules which showed significant GOBP and KEGG terms ($p < 0.1$) are presented here within boxes corresponding to each tissue pair: AM-DH, HC-DH, VH-DH, AP-DH and AM-HC.

and the edges correspond to the inter-genic connection strengths. It may be noted that most of the strongest connections are between genes that are also well-connected with other genes i.e. the hub genes. The kME measures may be used to prioritize candidate genes for further in-depth studies and validation. The top 10 hub genes in the 'midnight blue' network (colored blue, Figure 4.2) were: *NEFL*, *VSNL1*, *NDRG2*, *AVP*, *GAP43*, *BCAS1*, *HAPLN2*, *THY1*, *OXT* and *TCF7L2*. The known functions of these hub genes indicated that they could be strong candidates among genes regulating EB. These genes are discussed in detail in the following section.

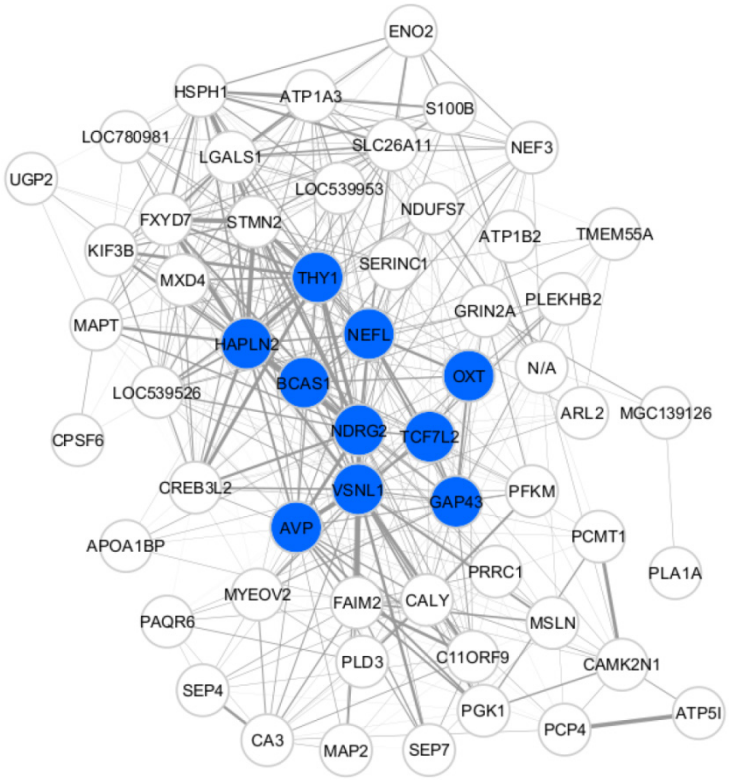


Figure 4.2 Network visualization of the 'midnight blue' consensus module in the AP-DH pair. Nodes represent the genes in the module and edges represent the inter-genic connection strengths. The 10 nodes with the highest intra-modular connectivity measures (kME) are colored blue. In general, nodes with higher kME values are placed towards the center of the network and have higher inter-genic connection strengths.

4.3 Discussion

The objective of this study was to identify genes and biological processes shared among the bovine AP and four brain areas that act together to regulate EB by investigating networks of co-expressed genes between these tissues. For this, we used a systems biology approach called WGCNA. The gene sets or modules thus identified were based on unsupervised clustering and not *a priori* defined. The genes in a module may have a common function which can be ascertained based on pathway or gene ontology enrichment tests. Further, the most connected genes within a module, i.e. the hub genes, may indicate the most important genes that drive the process regulated by the genes in the module. WGCNA has been used successfully in a wide variety of studies to identify gene co-expression modules and genes within them that relate either to a trait or disease condition of interest [22-

25] or that reflect underlying cellular composition or function of a tissue [19, 26] or that are conserved between species to help understand aspects of tissue evolution [27]. One of the strengths of this method is that the problem of multiple testing corrections is eliminated. This is because here the tested correlation is between the trait and the module eigengene and not to individual gene expression levels. Further, studying gene co-expression promises new insights by suggesting that the co-expressed genes may be responding to common transcriptional regulatory factors or signals even in different tissues.

The highest number of consensus modules and genes within them that correlated with EB were found in the AP-DH pair. This may be a reflection of the fact that the AP and DH play roles as centers in the hypothalamic-pituitary-ovarian axis that regulate reproduction and fertility via the sex hormones. Also, the high number of consensus modules may indicate multiple biological processes that are either associated with the physiological state of estrus or that co-ordinate the regulation of the female reproductive cycle via a high level of communication between these areas and the gonads and via responding to common signals especially estrogen.

The fact that a number of modules correlate with EB may indicate that several genes and biological processes across the AP and brain areas work together in the regulatory mechanism for EB expression. Knowledge of the components in this regulatory mechanism will help improve our understanding of EB regulation. Over-represented GOBP terms in modules significantly correlated with EB included general brain processes like glial cell development, memory, astrocyte activation, regulation of neuron projection development, axon diameter and neurogenesis, or neurodegenerative diseases like Alzheimer's and Parkinson's diseases. Several of these processes were also found enriched among genes found associated with EB in our earlier studies at individual tissue level [13, 14]. From the current co-expression study, it becomes evident that these are important processes shared among the AP and brain areas and hints at a coordinated regulation via regulatory factors or signals across these tissues. We also found genes involved in general cellular processes like oxidative phosphorylation, glycolysis/gluconeogenesis and ribosome. These general processes have been classified as general constitutive or as having housekeeping functions in several tissues [28] and were recognized as principal pathways in the brain [29, 30] to meet the high energy demands. However, these processes are also enhanced during stress and vary with the stage of estrous cycle. It has been shown that progesterone and estrogen increase the oxidative capacity of whole-brain mitochondria and the activity of cytochrome c oxidase, a major mitochondrial enzyme responsible for oxidative respiration [31]. Further, a study of the cerebral cortex in female rats at different stages of the

estrous cycle indicated a significant increase of the oxidative metabolism of the posterior cortex in the estrus and proestrus phases compared to the diestrus phase [32]. The increased synthesis of ribosomal RNA in the ventromedial hypothalamus has been reported to be one of the early effects of estrogen administered sub-cutaneously in ovariectomised rats and represents a primary event in the activation of neuronal cells and neuronal pathways involved in female reproductive behavior [33]. Estrogen driven expansion of dendrites and synapses follows from the stimulation of ribosomal RNA synthesis [9, 10]. Similar processes as identified in rodent studies therefore operate in cows too to affect EB.

The genes with the highest intra-modular connectivity per module i.e hub genes, are generally the most important in the network and are supposed to be key regulators. Studying these genes can reveal the biological process driven by modules in the network. We found that the hub genes in some modules (e.g. midnight blue) had a common function related to behavior regulation, nervous system processes or energy metabolism, while in other modules, a common function could not be ascertained for the hub genes. In the latter case, it may be an indication that these genes are involved in some common hitherto unassigned function in the mechanisms regulating EB. We take here the example of the hub genes in the 'midnight blue' module found in AP-DH pair. Several of the top 10 hub genes in this module are known to have functions that could explain their role in EB regulation. For example, *NEFL* is known to encode the light chain neurofilament protein comprising the axoskeleton and thereby functionally maintains the neuronal caliber and also plays a role in intracellular transport to axons and dendrites. The role of *NEFL* in protein phosphatase-1 targeting to neuronal membranes and cytoskeleton and its potential for regulating the dephosphorylation of phosphoproteins implicated in synaptic plasticity has been shown [34]. Other hub genes involved in neuronal growth or plasticity include *NDRG2* [35], *THY1* [36] and *GAP43* [37]. Another hub gene, *TCF7L2*, encodes a high mobility group (HMG) box-containing transcription factor [38] and is an important effector in the *Wnt* signaling pathway [39], that is active in the development of the blood brain barrier. The gene *VSNL1* is a neuronal calcium sensor that has recently been identified as part of a signaling complex associated with purinergic receptors [40]. These functions may link to findings of our earlier studies [13, 14] where we concluded that synaptic plasticity is of importance for proper expression of EB. Further, hub genes like *OXT* and *AVP*, that encode the pituitary hormones oxytocin and arginine vasopressin, were also found in our earlier studies [13, 14] as key genes whose expressions within individual tissues were correlated to EB and are

known from literature to have important regulatory functions related to socio-sexual behavior [41-43].

The relatively low number of animals in a network based study affects the statistical significance of the relationships found here. Despite this, the moderate correlation levels and the fact that many of the genes found here have biological relevance to the trait under study, indicates that even with a relatively high p-value cut-off of 0.2 chosen to identify consensus modules that correlated to EB, the genes within these modules may be considered as candidates regulating EB. It should be noted that the products of some of the genes within the co-expressed modules may fall in different cellular compartments and may have no possibility for physical interaction *in vivo*. Nevertheless, their belonging together in a module may indicate that the co-expressed genes across tissues may be responding to common regulatory factors or signals. Thereby new hypotheses may be formulated on how the AP and brain areas function together in regulating the trait of EB. A recent study showed that though transcriptome organization was generally poorly preserved between brain and blood, there were certain genes in the consensus modules, especially those involved in basic cellular processes, that exhibited a strong preservation [44]. Therefore in future, the correlation between the level of a protein or metabolite secreted in the milk or blood and the gene expression profile of genes found correlated to EB here may be studied. Highly correlated cases may be potential biomarkers which could be considered for development of quick assays to determine the estrus status of the cow. Biomarkers for estrus may also be developed from EB correlated gene products from the brain that cross the blood brain barrier or those from the AP which can potentially be measured in the blood.

4.4 Conclusions

To conclude, the study improved our understanding of the genomic regulation of EB in dairy cows by providing new insights into genes and associated biological processes shared among the bovine AP and brain areas acting together to regulate EB. The new knowledge could lead to development of novel management strategies to monitor and improve reproductive performance in dairy cows, for example, biomarkers for estrus detection.

4.5 Methods

Phenotypic data recording and gene expression measurements

The microarray experiment was carried out as part of a study aimed at identifying and studying genes that contribute to differences in EB expression and fertility levels of dairy cows. The procedures followed for EB recording, tissue collection,

RNA extraction and microarray hybridization were as described in our earlier studies [13, 14]. Briefly, EB was recorded in 28 healthy Holstein Friesian cows from 30 days in milk (DIM) onwards till their time of sacrifice which varied between 77 and 139 DIM i.e. after at least 2 estrous cycles. The EB recorded per cow were quantified as heat scores and the scores from multiple consecutive cycles per cow were averaged to obtain the average heat score per cow. Samples from 4 brain areas (DH, VH, AM and HC) and the AP were collected from these cows, 14 of which were sacrificed at start of estrus (day0) and 14 at mid of estrous cycle (day12). The cows were euthanized in a stress-free, quick and standardized way and all efforts were made to minimize suffering. The study was approved by the Animal Care and Ethics Committee of the Animal Sciences Group of Wageningen University and Research Centre, Lelystad (Approval ID 2006087a). RNA extracted from tissue samples were hybridized on Bovine 24K oligonucleotide (70-mer) microarrays designed and produced by the Bovine Oligonucleotide Microarray Consortium (BOMC), USA (http://www.bovineoligo.org/). A total of 280 arrays (i.e. 28 cows x 5 tissues x 2 for dye swaps) were prepared in a common reference design with the dye labels swapped between individual samples from each tissue and a reference sample consisting of equal proportions of RNA from all tissues of all cows. Microarray data pre-processing and analysis was done using the LIMMA (Linear models for microarray data) package [45] within Bioconductor project [46] of R statistical programming language (http://www.r-project.org). Gene expression levels expressed as M-values (log differential-expression ratio of sample versus reference) were thus obtained for all probes per array.

The array design (accession number: A-MEXP-1781) and the array experiment data (accession number: E-TABM-916) are publicly accessible on ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/). The original annotation of the bovine 24K oligonucleotide microarray provided by BOMC dates back to June 2007. For our analysis, we used the bovine oligonucleotide array probe re-annotation (Version 5) based on Ensembl (http://www.ensembl.org) release 56 (October 2009) provided on the EADGENE website (http://www.eadgene.info/ToolsResources/EADGENEOligoSetsAnnotationFiles/tabid/324/Default.aspx) by the authors of the oligo-set re-annotation pipeline, sigReannot [47]. For the re-annotation, out of the 23,496 probes (excluding control probes) on the bovine oligonucleotide array, only 16,620 probes that were assigned a quality score between 1 and 4 for their specificity to hits on the bovine genome were considered. Probes with quality scores between 5 and 7 had either no hits or multiple hits and were not annotated as they were not specific.

For the analysis done here, we used only the data from the day0 cows. The M-values of genes represented by multiple probes were averaged, leaving gene expression data of 13,234 unique genes per array. The various steps taken in the analysis are summarized in Table 4.2.

Construction of gene co-expression networks and identification of consensus modules between tissue pairs

For this analysis, we chose to study gene co-expression patterns in tissues taken as pairs with the DH in common i.e. AP-DH, VH-DH, AM-DH and HC-DH pairs. This was because the hypothalamus is the master regulatory center within the hypothalamic-pituitary-ovarian axis that listens to hormonal signals and stimulates the AP to release hormones by secreting releasing factors and signaling by neural pathways. It is likely that the genes influencing EB expression in other tissues would have an expression profile similar to that in the DH when the cows are in estrus and under the influence of estrogen. The genes with a similar expression profile may be detected as preserved modules of genes between the different tissues. In addition, we also analyzed the AM-HC pair, as these tissues are physically and functionally close to each other with both involved in emotional responses and likely to be inter-communicating while affecting EB.

To begin with, genes and arrays having an excessive number of missing values were tested for and removed if any. The arrays were then clustered to identify any outliers. From the 13,234 unique good quality genes in the array, only the top 50 per cent genes with the most variable expression per tissue were selected. Among these, only those genes shared between each pair of tissues under study were used for network construction. The selection was done because the most interesting genes would be the ones whose varying expression contributes to the value of the trait expressed. Genes whose expression remains the same in all cases would generally not be associated with the varying trait. Further, analyzing a subset of the most interesting genes instead of taking all genes together makes the procedures involved in network construction and analysis less computationally demanding.

For defining gene co-expression networks and identifying consensus modules of co-expressed genes among tissue pairs, we used a systems biology approach called Weighted Gene Co-expression Network Analysis [48, 49] implemented in R statistical programming language (http://www.r-project.org) as a freely available software package called WGCNA. Briefly, WGCNA is a network based gene screening method that integrates information on gene significance (correlation between gene expression and a trait of interest) and module membership to identify candidate genes affecting a trait. Detailed tutorials with examples for the

use of this method can be found at http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/index.html. Pearson correlations were first calculated for all pairs of genes across all microarray samples, followed by a noise reduction step where these correlations were weighted by raising their absolute values to a certain power, beta. This soft-thresholding power, beta, was determined using the scale free topology criterion, a property of most biological networks. For the resulting weighted network, gene co-expression modules were defined as branches of a hierarchical clustering tree. Specifically, we used average linkage hierarchical clustering with the topological overlap similarity measure to define a cluster tree. The topological overlap is a robust measure of interconnectedness, which keeps track of shared patterns of connection strengths. For branch cutting (module detection) we used the dynamic branch-cutting algorithm [50] implemented in WGCNA to detect modules larger than 20 genes. The modules that are shared between the tissues in each pair (consensus modules) were then identified and assigned unique color labels.

Characterization of consensus modules that correlate with estrous behavior

The consensus modules whose module eigengenes correlated with EB (absolute correlation>0.3 and p<0.2) were identified using a function within WGCNA package that correlates module eigengenes with trait values. The module eigengene is the most representative gene expression in the module and corresponds to the first principal component of that module. The correlation is tested by taking the module eigengenes as covariates of a multivariate regression model that regresses the trait of EB expressed as heat scores on the eigengenes transcriptome [21].

For gaining insight into biological processes underlying EB, we performed functional analyses of the genes within the EB correlated modules (study sets) using R package GOstats [51]. Genes in the study sets were tested for enriched Gene Ontology Biological Process [52] terms and KEGG [53] pathway terms. Bovine Ensembl IDs were used for the study sets whereas the population set consisted of Entrez gene IDs as required by the GOstats package. Of the 13,234 unique good quality genes on the array, 12,655 remained after conversion to Entrez gene IDs.

The genes in the consensus modules were ordered by their module eigengene based intra-modular connectivity measures (kME) to identify genes with the highest connectivity. The most important genes in the network are usually the most well-connected with other genes and are called hub genes. As an example, we looked in detail at the known functions of the top hub genes in one of the EB correlated consensus modules for their likely relation with EB regulation. Further, we visualized this module as a network with the nodes representing the genes and

the edges representing the inter-genic connection strengths. We used a function in WGCNA to export the genes from the weighted network data of these modules in a form that was compatible for visualization using the 'Cytoscape' software [54]. We chose a cut-off threshold of 0.2 to filter out low strength connections between genes in the network.

List of abbreviations

AM: Amygdala; AP: Anterior Pituitary; BOMC: Bovine Oligonucleotide Microarray Consortium; DH: Dorsal Hypothalamus; EB: Estrous behavior; GH: gene significance; GOBP: Gene Ontology Biological Process; HC: Hippocampus; KEGG: Kyoto Encyclopedia of Genes and Genomes; kME: module eigengene based intra-modular connectivity measures; SD: Standard Deviation; VH: Ventral Hypothalamus; WGCNA: Weighted gene co-expression network analysis

Acknowledgements

We thank Roel Veerkamp and Martien Groenen for helpful discussions and suggestions to improve the manuscript. This work was co-financed by the European Commission funded projects SABRE (www.sabre-eu.eu/) (contract No. FOOD-CT-2006-016250) and EADGENE (www.eadgene.info/) (contract No. FOOD-CT-2004-506416), the Dutch Ministry of Agriculture, Nature and Food Quality (www.minlnv.nl) (KB-04-004-012) and the International cattle breeding co-operative - CRV BV (www.crv4all.com/). The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

Authors' contributions

AK performed the data analysis and drafted the manuscript. MFWP and MAS helped with the biological interpretation of the results and in improving the manuscript.

Supplementary material

Posted at http://edepot.wur.nl/183833

Figure S1. Consensus gene dendrograms and modules of co-expressed genes identified by weighted gene co-expression network analysis for the tissue pairs.
Presented here are the gene dendrograms produced by clustering the dissimilarity based on consensus topological overlap for each tissue pair: AM-DH, HC-DH, VH-DH, AP-DH and AM-HC. The corresponding modules of co-expressed genes are represented by colors in the color row below the dendrograms.

Figure S2. Venn diagram comparing genes in all consensus modules per tissue pair.
The genes in all consensus modules of the AP-DH, AM-DH, HC-DH and VH-DH tissue pairs are compared here to determine the overlap of genes among them.

Table S1. Genes within consensus modules among tissue pairs that significantly correlated with estrous behavior.
The Ensembl IDs and gene symbols of the genes within each of the estrous behavior correlated consensus modules are presented for AM-DH, HC-DH, AP-DH and AM-HC tissue pairs. No significantly correlated modules were found in the VH-DH pair. The genes within each consensus module are ordered by their intra-modular connectivity measures (kME). The corresponding gene significance values (correlation of gene expression with estrous behavior) are also reported.

Table S2. Gene Ontology Biological Process and KEGG pathway terms enriched in consensus modules among tissue pairs that significantly correlated with estrous behavior.
The consensus modules (labeled by colors) in the co-expression networks between tissue pairs whose module eigengenes correlated with estrous behavior scores with an absolute correlation above 0.3 and having a p-value below 0.2 in at least one of the tissues in each pair were tested for enriched GOBP and KEGG pathway terms. Those terms that were supported by at least two genes and had a Benjamini-Hochberg (BH) corrected p-value below 0.1 are reported. Here, GOBPID stands for gene ontology biological process ID; ExpCount stands for expected number of genes in the module that would relate to the GO/KEGG term; Count stands for actual number of genes in the module that relate to the term; and Size stands for the number of genes in the population set (total genes on the microarray considered in the study) that relate to the term.

References

1. Stott AW, Veerkamp RF, Wassell TR: The economics of fertility in the dairy herd. Anim Sci 68, 1999, 49-57 1999.
2. Garnsworthy PC, Sinclair KD, Webb R: Integration of physiological mechanisms that influence fertility in dairy cows. Animal 2008, 2(08):1144-1152.
3. Butler WR: Nutritional interactions with reproductive performance in dairy cattle. Anim Reprod Sci 2000, 60-61:449-457.
4. Bauersachs S, Ulbrich SE, Gross K, Schmidt SEM, Meyer HHD, Einspanier R, Wenigerkind H, Vermehren M, Blum H, Sinowatz F, Wolf E: Gene expression profiling of bovine endometrium during the oestrous cycle: detection of

molecular pathways involved in functional changes. J Mol Endocrinol 2005, 34(3):889-908.

5. Casey OM, Morris DG, Powell R, Sreenan JM, Fitzpatrick R: Analysis of gene expression in non-regressed and regressed bovine corpus luteum tissue using a customized ovarian cDNA array. Theriogenology 2005, 64(9):1963-1976.
6. Ushizawa K, Herath C, Kaneyama K, Shiojima S, Hirasawa A, Takahashi T, Imai K, Ochiai K, Tokunaga T, Tsunoda Y, Tsujimoto G, Hashizume K: cDNA microarray analysis of bovine embryo gene expression profiles during the pre-implantation period. Reprod Biol Endocrinol 2004, 2(1):77.
7. Salilew-Wondim D, Hölker M, Rings F, Ghanem N, Ulas-Cinar M, Peippo J, Tholen E, Looft C, Schellander K, Tesfaye D: Bovine pretransfer endometrium and embryo transcriptome fingerprints as predictors of pregnancy success after embryo transfer. Physiol Genomics 2010, 42(2):201-218.
8. Lopez H, Satter LD, Wiltbank MC: Relationship between level of milk production and estrous behavior of lactating dairy cows. Anim Reprod Sci 2004, 81(3-4):209-223.
9. Mong JA, Pfaff DW: Hormonal symphony: steroid orchestration of gene modules for sociosexual behaviors. Mol Psychiatry 2004, 9(6):550-556.
10. Pfaff D: Hormone-driven mechanisms in the central nervous system facilitate the analysis of mammalian behaviours. J Endocrinol 2005, 184(3):447-453.
11. Gallagher M, Chiba AA: The amygdala and emotion. Curr Opin Neurobiol 1996, 6(2):221-227.
12. Salamon E, Esch T, Stefano GB: Role of amygdala in mediating sexual and emotional behavior via coupled nitric oxide release. Acta Pharmacol Sin 2005, 26(4):389-395.
13. Kommadath A, Mulder HA, de Wit AAC, Woelders H, Smits MA, Beerda B, Veerkamp RF, Frijters ACJ, te Pas MFW: Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows. Animal 2010, 4(08):1297-1307.
14. Kommadath A, Woelders H, Beerda B, Mulder H, de Wit A, Veerkamp R, te Pas M, Smits M: Gene expression patterns in four brain areas associate with quantitative measure of estrous behavior in dairy cows. BMC Genomics 2011, 12(1):200.
15. Horvath S, Dong J: Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput Biol 2008, 4(8):e1000117.
16. Ideker T, Ozier O, Schwikowski B, Siegel AF: Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 2002, 18(suppl 1):S233-S240.

17. Huang Y, Li H, Hu H, Yan X, Waterman MS, Huang H, Zhou XJ: Systematic discovery of functional modules and context-specific functional annotation of human genome. Bioinformatics 2007, 23(13):i222-i229.
18. Eisen MB, Spellman PT, Brown PO, Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998, 95(25):14863-14868.
19. Oldham M, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind D: Functional organization of the transcriptome in human brain. Nat Neurosci 2008, 11(11):1271 - 1282.
20. Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt E, Drake T, Lusis A: Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet 2006, 2(8):e130.
21. Langfelder P, Horvath S: Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol 2007, 1(1):54.
22. Horvath S, Zhang B, Carlson M, Lu K, Zhu S, Felciano R, Laurance M, Zhao W, Shu Q, Lee Y et al: Analysis of Oncogenic Signaling Networks in Glioblastoma Identifies ASPM as a Novel Molecular Target. Proc Natl Acad Sci USA 2006, 103(46):17402 - 17407.
23. Ghazalpour A, Doss S, Zhang B, Plaisier C, Wang S, Schadt E, Thomas A, Drake T, Lusis A, Horvath S: Integrating Genetics and Network Analysis to Characterize Genes Related to Mouse Weight. PloS Genetics 2006, 2(8):e130.
24. Fuller T, Ghazalpour A, Aten J, Drake T, Lusis A, Horvath S: Weighted Gene Co-expression Network Analysis Strategies Applied to Mouse Weight. Mamm Genome 2007, 6(18):463 - 472.
25. Miller J, Oldham M, Geschwind D: A Systems Level Analysis of Transcriptional Changes in Alzheimer's Disease and Normal Aging. J Neurosci 2008, 28(6):1410 - 1420.
26. de Jong S, Fuller T, Janson E, Strengman E, Horvath S, Kas M, Ophoff R: Gene expression profiling in C57BL/6J and A/J mouse inbred strains reveals gene networks specific for brain regions independent of genetic background. BMC Genomics 2010, 11(1):20.
27. Oldham M, Horvath S, Geschwind D: Conservation and Evolution of Gene Co-expression Networks in Human and Chimpanzee Brains. Proc Natl Acad Sci USA 2006, 103(47):17973 - 17978.
28. Prieto C, Risueño A, Fontanillo C, De Las Rivas J: Human Gene Coexpression Landscape: Confident Network Derived from Tissue Transcriptomic Profiles. PLoS ONE 2008, 3(12):e3911.

29. Gan L, Liu X, Xiang Z, He N: Microarray-based gene expression profiles of silkworm brains. BMC Neurosci 2011, 12(1):8.
30. Marques F, Sousa J, Coppola G, Gao F, Puga R, Brentani H, Geschwind D, Sousa N, Correia-Neves M, Palha J: Transcriptome signature of the adult mouse choroid plexus. Fluids Barriers CNS 2011, 8(1):10.
31. Irwin RW, Yao J, Hamilton RT, Cadenas E, Brinton RD, Nilsen J: Progesterone and Estrogen Regulate Oxidative Metabolism in Brain Mitochondria. Endocrinology 2008, 149(6):3167-3175.
32. Menéndez-Patterson A, Florez-Lozano JA, Marin B: Oxidative and glicolytic metabolism of the frontal cortex (latero-frontal) and of the posterior cortex (latero-occipital) in relation with the sexual activity of the rat. Reproduction 1976, 3(3-4):279-285.
33. Jones KJ, Harrington CA, Chikaraishi DM, Pfaff DW: Steroid hormone regulation of ribosomal RNA in rat hypothalamus: early detection using in situ hybridization and precursor-product ribosomal DNA probes. J Neurosci 1990, 10(5):1513-1521.
34. Terry-Lorenzo RT, Inoue M, Connor JH, Haystead TAJ, Armbruster BN, Gupta RP, Oliver CJ, Shenolikar S: Neurofilament-L Is a Protein Phosphatase-1-binding Protein Associated with Neuronal Plasma Membrane and Post-synaptic Density. J Biol Chem 2000, 275(4):2439-2446.
35. Takahashi K, Yamada M, Ohata H, Honda K, Yamada M: Ndrg2 promotes neurite outgrowth of NGF-differentiated PC12 cells. Neurosci Lett 2005, 388(3):157-162.
36. Rege TA, Hagood JS: Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. FASEB J 2006, 20(8):1045-1054.
37. Gispen W, Nielander H, De Graan P, Oestreicher A, Schrama L, Schotman P: Role of the growth-associated protein B-50/GAP-43 in neuronal plasticity. Mol Neurobiol 1991, 5(2):61-85.
38. Castrop J, van Norren K, Clevers H: A gene family of HMG-box transcription factors with homology to TCF-1. Nucleic Acids Res 1992, 20(3):611.
39. Jin T, Liu L: Minireview: The Wnt Signaling Pathway Effector TCF7L2 and Type 2 Diabetes Mellitus. Mol Endocrinol 2008, 22(11):2383-2392.
40. Chaumont S, Compan V, Toulme E, Richler E, Housley GD, Rassendren F, Khakh BS: Regulation of P2X2 Receptors by the Neuronal Calcium Sensor VILIP1. Sci Signal 2008, 1(41):ra8.
41. Leng G, Meddle SL, Douglas AJ: Oxytocin and the maternal brain. Curr Opin Pharmacol 2008, 8(6):731-734.

42. Curley JP, Keverne EB: Genes, brains and mammalian social bonds. Trends Ecol Evol 2005, 20(10):561-567.
43. Donaldson ZR, Young LJ: Oxytocin, Vasopressin, and the Neurogenetics of Sociality. Science 2008, 322(5903):900-904.
44. Cai C, Langfelder P, Fuller T, Oldham M, Luo R, van den Berg L, Ophoff R, Horvath S: Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 2010, 11(1):589.
45. Smyth GK: Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor. Edited by Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W. New York: Springer; 2005:397-420.
46. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J et al: Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004, 5(10):R80.
47. Casel P, Moreews F, Lagarrigue S, Klopp C: sigReannot: an oligo-set re-annotation pipeline based on similarities with the Ensembl transcripts and Unigene clusters. BMC Proc 2009, 3(Suppl 4):S3.
48. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008, 9(1):559.
49. Zhang B, Horvath S: A General Framework for Weighted Gene Co-expression Network Analysis. Stat Appl Genet Mol Biol 2005, 4:Article 17.
50. Langfelder P, Zhang B, Horvath S: Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 2008, 24(5):719 - 720.
51. Falcon S, Gentleman R: Using GOstats to test gene lists for GO term association. Bioinformatics 2007, 23(2):257-258.
52. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al: Gene Ontology: tool for the unification of biology. Nat Genet 2000, 25(1):25-29.
53. Kanehisa M, Goto S: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000, 28:27-30.
54. Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res 2003, 13(11):2498-2504.

5

Regional Regulation of Transcription in the Bovine Genome

Arun Kommadath[1#], Haisheng Nie[2#¤], Martien A. M. Groenen[2], Marinus F. W. te Pas[1], Roel F. Veerkamp[1], Mari A. Smits[1]

[1]Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, The Netherlands; [2]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, The Netherlands

[#] These authors contributed equally to this work.
[¤] Current address: Illumina, Beijing, People's Republic of China

PLoS ONE (2011) 6(6): e20413

Abstract

Eukaryotic genes are distributed along chromosomes as clusters of highly expressed genes termed RIDGEs (Regions of IncreaseD Gene Expression) and lowly expressed genes termed anti-RIDGEs, interspersed among genes expressed at intermediate levels or not expressed. Previous studies based on this observation suggested a dual mechanism of gene regulation, where, in addition to transcription factors, the chromosomal domain influences the expression level of their embedded genes. The objectives here were to provide evidence for the existence of chromosomal regional regulation of transcription in the bovine genome, to analyse the genomic features of genes located within RIDGEs versus anti-RIDGEs and tissue-specific genes versus housekeeping and to examine the genomic distribution of genes subject to positive selection in bovines. Gene expression analysis of four brain tissues and the anterior pituitary of 28 cows identified 70 RIDGEs and 41 anti-RIDGEs (harbouring 3735 and 1793 bovine genes respectively) across the bovine genome which are significantly higher than expected by chance. Housekeeping genes (defined here as genes expressed in all five tissues) were over-represented within RIDGEs but tissue-specific genes (genes expressed in only one of the five tissues) were not. Housekeeping genes and genes within RIDGEs had, in general, higher expression levels and GC content but shorter gene lengths and intron lengths than tissue-specific genes and genes within anti-RIDGES. Our findings suggest the existence of chromosomal regional regulation of transcription in the bovine genome. The genomic features observed for genes within RIDGEs and housekeeping genes in bovines agree with previous studies in several other species further strengthening the hypothesis of selective pressure to keep the highly and widely expressed genes short and compact for transcriptional efficiency. Further, positively selected genes were found non-randomly distributed on the genome with a preference for RIDGEs and regions of intermediate gene expression compared to anti-RIDGEs.

Key words: regulation of transcription, RIDGE, bovine, brain, positive selection

5.1 Introduction

Eukaryotic genes are not randomly distributed along chromosomes but organised as clusters of highly expressed genes, termed RIDGEs (Regions of IncreaseD Gene Expression), and lowly expressed genes, termed anti-RIDGEs [1], interspersed among genes expressed at intermediate levels or not expressed. Certain genomic features observed for RIDGEs were in striking contrast with those for anti-RIDGEs [2]. RIDGEs were found to be gene dense, GC rich and SINE repeat rich and mostly harboured genes with shorter than average intron sizes. In contrast, anti-RIDGES showed low gene density, were AT rich and LINE repeat rich. Surprisingly, the gene expression patterns of highly and weakly expressed chromosomal regions were roughly similar in all 12 human tissues analysed [1,2]. Based on their findings, Versteeg *et al*. [2] postulated that RIDGEs globally govern the expression levels of their embedded genes and that this higher level regulation of gene transcription was dependent on factors that act on chromosomal domains like chromatin conformation and position in the nucleus. The existence of this novel domain wide transcription regulatory mechanism, in addition to the well-known regulatory mechanism at the individual gene level that involve transcription factors and regulatory sequences on the gene, was proven in a later study by Gierman *et al*. [3]. They showed that the expression levels of identical green fluorescent protein (GFP) reporter constructs integrated at several different chromosomal positions corresponded to the overall expression level of genes within the domains of integration. To explain this, a dual mechanism of gene regulation was suggested wherein transcription factors determine a basal level of transcription for a gene, whereas the chromosomal domain in which the gene was located determined its ultimate expression level. They also established a range for the sizes of such chromosomal domains by showing that the correlation between GFP expression and domain activity was highest for window sizes of roughly 19–79 genes around the integration sites.

The observation of clusters of highly expressed genes was reported to be a consequence of the clustering of housekeeping (HK) genes, which in turn, was probably the outcome of selective pressure to position widely expressed genes in genomic regions where the higher-order chromatin structure allows better accessibility to the transcription machinery [4,5]. However, RIDGEs are not restricted to HK genes, but encode tissue-specific (TS) genes too [2]. Moreover, not all genes on a RIDGE are highly expressed but the average expression level of all genes per cluster taken together would be higher than the average gene expression across the genome. Nevertheless, RIDGEs consist of a higher proportion of HK

genes than would be expected by chance and the HK genes share certain genomic features with genes on RIDGEs such as higher expression levels and shorter lengths for introns, genes and coding sequences [6,7]. It has been proposed that since transcription is an energy expensive process, there is an evolutionary selection pressure for economy of transcription to keep widely and highly expressed genes short and compact [6,8,9,10].

The phenomenon of domain wide regulation of transcription has been shown to exist in several species: mammals (human, mouse) [1,2,3,11,12], flies (drosophila) [13,14] and recently in birds (chicken) [6]. Though believed to be a feature for all eukaryotes, this has been experimentally verified only in a few species. The recently published bovine genome assembly [15] allowed us for the first time to investigate the existence of chromosomal regional regulation of transcription in bovines using a brain transcriptome dataset of closely related tissues: Amygdala (AM), Hippocampus (HC), Dorsal Hypothalamus (DH), Ventral Hypothalamus (VH) and Anterior Pituitary (AP). Prior to this study, it was not known whether the phenomenon of chromosomal regional regulation of transcription existed in bovine genome and whether the genomic features for RIDGEs and anti-RIDGEs reported in earlier studies in other species could be expected in bovines as well. It was also not known whether the above mentioned findings could be expected when closely related tissues are analysed as was done here, which is in contrast to previous studies that used gene expression data from a wide variety of tissues.

A study on expression levels of genes subject to positive selection [16] showed that positively selected genes, in general, had reduced expression levels and were expressed in a TS manner in several human tissues. Targets of positive selection have been linked to complex trait phenotypes in humans and primates where adaptation to local environmental changes has been the driving force. Likewise, in laboratory, farm and companion animals, human intervention in the form of domestication and intensive artificial selection may have been additional drivers [16,17,18,19,20]. Therefore the quest for genes controlling complex traits of interest in domesticated animals may be supported by finding signatures of positive selection and regional patterns of gene expression that help narrow down the regions on the genome to look for these genes.

The primary objectives of this study were: 1. to provide evidence for the existence of chromosomal regional regulation of transcription in the bovine genome, 2. to analyse the genomic features of genes located within RIDGEs versus anti-RIDGEs and TS genes versus HK, and 3. to examine the genomic distribution of genes subject to positive selection in bovines.

Here, HK refers to genes expressed in all five tissues studied here and TS refers to genes expressed in only one of the five tissues.

5.2 Results

Gene expression data

The normalised transformed gene expression data matrix analysed here consisted of average expression per tissue for 13,234 bovine Ensembl genes (Table S1). Prior to averaging, a hierarchical cluster analysis of the normalised gene intensities per individual per tissue showed clustering by tissue type. Among the brain areas, the AM and HC clustered closer to each other and so did DH and VH, whereas the AP stood out as a separate cluster (Figure S1). The chromosome lengths and chromosome wise distribution of all bovine Ensembl genes and of those bovine genes on the BOMC array represented by good quality probes (see Materials and Methods section) are shown in Figure 5.1. The number of Ensembl genes represented on the re-annotated BOMC array was roughly 50% of the total known Ensembl genes on each chromosome. The total bovine genome size analysed in this study was 2,608,296,415 bp which was approximately 91% of the latest bovine

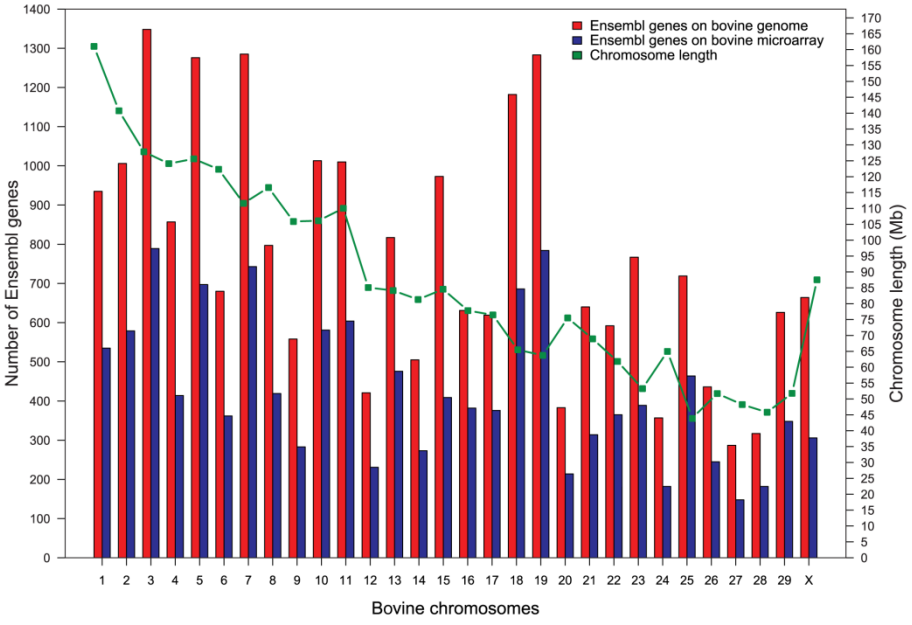


Figure 5.1 Chromosome wise distribution of Ensembl genes on bovine genome and microarray. The number of Ensembl genes on each chromosome of the bovine genome and those represented on the microarray are depicted by bar plots. The chromosome length is overlaid on this plot.

genome assembly version Btau_4.0 [15,21] having an estimated genome size of 2.87 Gb. Chromosomes 3, 5, 7, 18 and 19 have the most genes, with over 1100 each. The bovine chromosomes 18 and 19 (BTA18 and BTA19), though ranked 21st and 23rd respectively by length, are ranked fifth and third respectively by the number of genes harboured. We used the 'Synteny Tracker' tool [22] to identify syntenic regions between the bovine and human genomes. The high gene density of these bovine chromosomes corresponds to the high gene density of their syntenic human chromosomes. BTA18 shared a large syntenic region with HSA19 which has the highest gene density amongst human chromosomes [23]and also with HSA16 which is of moderate gene density [24]. Likewise, BTA19 was almost entirely syntenic with HSA17 which has the second highest gene density amongst human chromosomes [25].

Genomic regions of high and low gene expression identified

RIDGEs and anti-RIDGEs were so selected that each covers approximately 10% of the bovine genome (see Materials and Methods section). For our dataset, this criterion was satisfied by taking the expression thresholds as 1.25 times larger than the genomic median expression value in the case of RIDGEs and 1.45 times lower than the genomic median expression value for anti-RIDGEs. With the chosen window size of 39 genes and genome coverage threshold of 10%, a reasonable

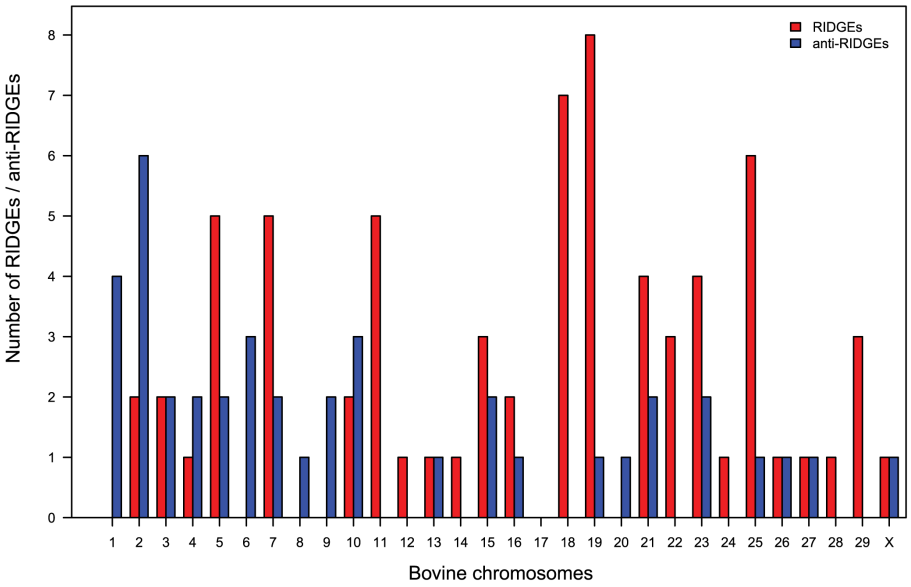


Figure 5.2 Chromosome wise distribution of RIDGEs and anti-RIDGEs. The number of RIDGEs and anti-RIDGEs found per chromosome (based on median gene expression with a window size of 39 genes) is depicted here.

number of RIDGEs and anti-RIDGEs could be identified: 70 RIDGEs harbouring 3735 bovine Ensembl genes and 41 anti-RIDGEs harbouring 1793 bovine Ensembl genes. The chromosome wise distribution of these RIDGEs and anti-RIDGEs is shown in Figure 5.2 and their genomic locations on the bovine genome are provided in Table S2. The chromosome wise transcriptome maps based on median expression depicting the identified RIDGEs and anti-RIDGEs are given in Figure S2.

Higher number of RIDGEs and anti-RIDGEs found than expected by chance

A permutation test repeated 10000 times using the same window size (39 genes) and threshold for RIDGE identification (1.25 times larger than the genomic median expression value) as used in our analysis showed that there was only about a 10% chance for obtaining an equal or higher number of RIDGEs (mean = 60.13, s.d. = 7.07) than what was found in our analysis (n = 70). A similar test for anti-RIDGEs revealed that there was less than 1% chance for obtaining an equal or higher number of anti-RIDGEs (mean = 24.44, s.d. = 4.92) than what was found in our analysis (n = 41).

Transcriptome maps in four brain areas and anterior pituitary are highly correlated

Spearman rank correlation test on pair-wise comparisons of transcriptome maps of all tissues showed high correlations (Figure 5.3), with an average correlation of 0.91 (all p-values below 2.2e-16). The highest correlations are between DH and VH and between AM and HC. All brain areas are more correlated with each other than with the anterior pituitary. The similarity in the transcriptome maps across tissues was clearly visible taking chromosome 2 as an example (Figure 5.4).

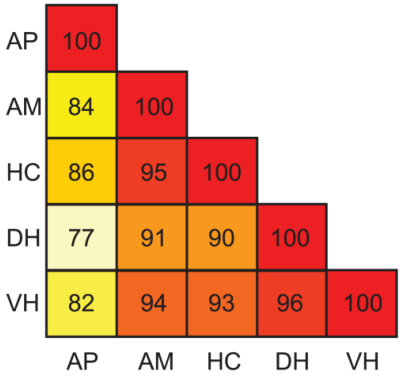


Figure 5.3 Pair wise correlations between tissues based on median expression values of all genes per tissue. AP-Anterior Pituitary, AM-Amygdala, HC-Hippocampus, DH-Dorsal Hypothalamus, VH-Ventral Hypothalamus.

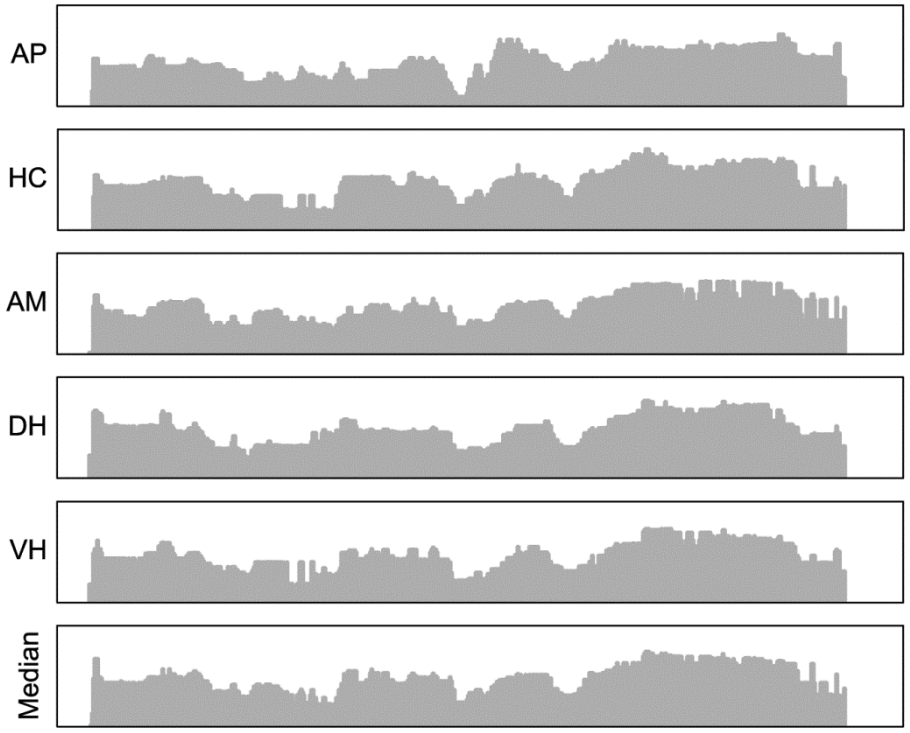


Figure 5.4 Tissue wise transcriptome maps for chromosome 2. Transcriptome maps for AP, HC, AM, DH, VH and one based on the average of the median gene expressions across all 5 tissues. The y-axis shows the median gene expression levels (log transformed normalised gene intensities ranging from 9 to 14) for the 579 Ensembl genes on chromosome 2 represented on the array (AP-Anterior Pituitary, AM-Amygdala, HC-Hippocampus, DH- Dorsal Hypothalamus, VH-Ventral Hypothalamus).

'Housekeeping' and 'Tissue-specific' genes identified

In order to define the threshold for expression, the normalized expression intensities of all genes and negative controls across all arrays and tissues was determined (Figure 5.5). The threshold for expression defined as the 99.9% quantile value of the log transformed expression levels of negative controls across all arrays, was 11.94. Figure 5.5 also shows the number of tissues in which the expressed genes are distributed, where '1' represents the TS genes expressed only in one tissue (total of 1035 for the five tissues), '2' represents genes expressed in 2 tissues and so on. Number '5' represents genes expressed in all 5 tissues which we defined here as HK genes (total of 3651). The distribution of the 1035 TS genes across individual tissues is also shown (see Table S3 for list of HK and TS genes). The

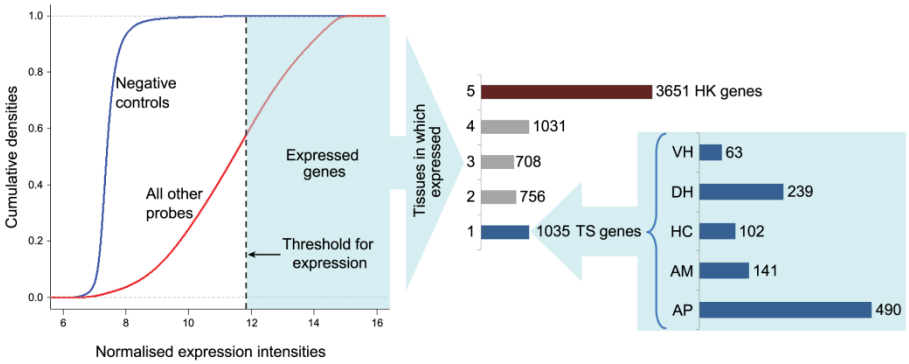


Figure 5.5. Determination of expressed genes and their tissue wise distribution. The blue line in the graph on the left represents expression levels of negative controls on the microarray and the red line represents that of all other probes. The 99.9% quantile value of expression levels of all negative controls across all arrays was 11.94, which was taken as the threshold above which a gene was considered as expressed. The graph on the right depicts the number of tissues in which the expressed genes are distributed and their tissue wise distributions. Here, '1' represents the 'tissue-specific' genes i.e. expressed only in one tissue (total of 1035), '2' represents genes expressed in 2 tissues and so on. Number '5' represents genes expressed in all five tissues which we defined here as 'housekeeping' genes (total of 3651). (AP-Anterior Pituitary, AM-Amygdala, HC-Hippocampus, DH- Dorsal Hypothalamus, VH-Ventral Hypothalamus).

maximum number of TS genes was seen for the AP (490) which is an endocrine gland. The brain areas, being all of neurological tissue type, share a lot of functions in common and therefore genes. Hence the number of unique TS genes found were smaller (141 for AM, 102 for HC, 239 for DH, 63 for VH). Functional analysis of the orthologous human genes of the bovine genes (human orthologues used for better annotation information available) in R package 'GOstats' (see Materials and Methods) revealed enriched gene ontology and KEGG pathway terms for the HK and TS genes (Table S4). In most cases, the terms were not significant at a Benjamini-Hochberg p-value threshold of 0.1, hence the top 10-15 terms were considered to get an indication of the most important functions. Briefly, functional analysis of TS genes per tissue revealed processes like 'negative regulation of synaptic transmission, glutamatergic', 'negative regulation of transmission of nerve impulse' and 'regulation of behaviour' for the anterior pituitary, and processes like 'androgen receptor signalling pathway', 'neurotransmitter catabolic process', 'behavioural response to pain' for the brain areas. For HK genes, general processes related to ATP synthesis, mitotic activity and translation, typical of most tissues, were found to be enriched.

Housekeeping genes are over-represented on RIDGEs but tissue-specific genes are not

Based on 10,000 random samples, the mean percentage overlap between the randomly sampled genes (equivalent in number to the total HK genes) and the RIDGE genes was found to be 17.55% (s.d. = 0.53). However, the overlap between actual HK genes found in our analysis and the RIDGE genes was 23.28% (850 out of 3651 HK genes) which was clearly above expectation ($p<0.01$). Similarly, the mean percentage overlap between the randomly sampled genes (equivalent in number to the total TS genes) and the RIDGE genes, was found to be 17.52% (s.d. = 1.14). However, the overlap between actual TS genes found in our analysis with RIDGE genes was found to be 13.14% (136 out of 1035 TS genes) which was clearly below expectation ($p<0.01$).

Differences observed in several genomic features of RIDGE versus anti-RIDGE genes and 'Housekeeping' versus 'Tissue-specific' genes

The gene length and average intron length were significantly lower in HK genes compared to TS genes and also significantly lower in genes located on RIDGEs compared to anti-RIDGEs. In contrast, the GC content was significantly higher in HK genes compared to TS genes and also significantly higher in genes located on RIDGEs compared to anti-RIDGEs (Figure 5.6). Further, transcript length and exon count were significantly lower in HK genes compared to TS genes whereas these were not significantly different between genes on RIDGEs compared to anti-RIDGEs. Differences in exon length were non-significant in both cases (Figure S3). The results of these comparisons are summarized in Table 5.1.

Table 5.1 Comparisons of genomic features of housekeeping vs. tissue-specific genes and of genes on RIDGEs vs. anti-RIDGEs.

	Genomic features	Housekeeping vs. Tissue-specific genes	RIDGE vs. anti-RIDGE genes
1	Gene length	Lower *	Lower *
2	Intron length	Lower *	Lower *
3	GC content	Higher *	Higher *
4	Transcript length	Lower *	Non-significant
5	Exon count	Lower *	Non-significant
6	Exon length	Non-significant	Non-significant

* $p<0.01$

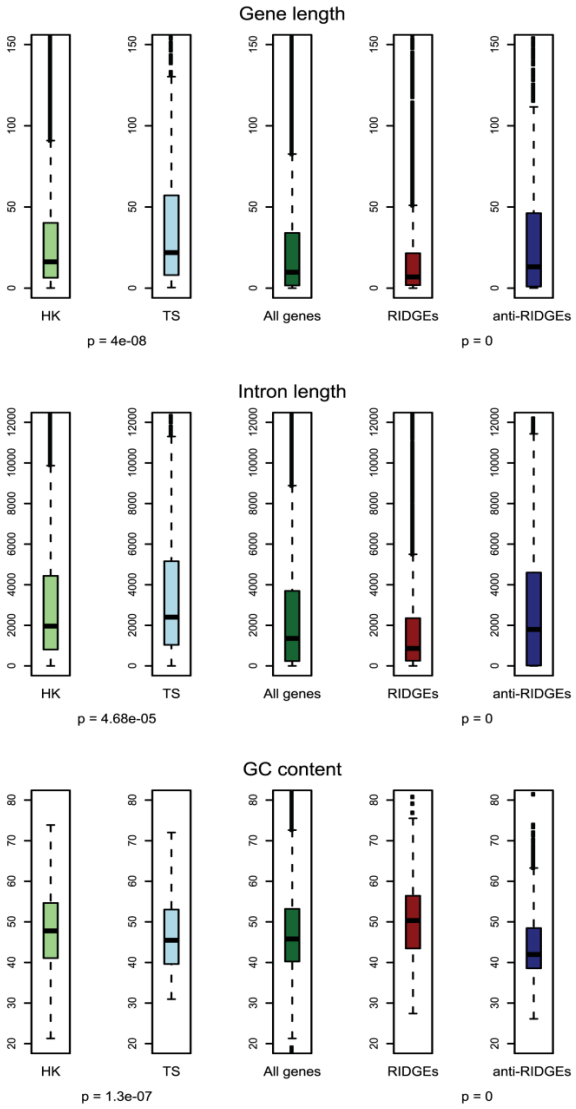


Figure 5.6 Genomic features of housekeeping vs. tissue-specific genes and of genes present on the RIDGEs vs. anti-RIDGEs. The following genomic features are represented here: Gene length, Intron length and GC content. The p-values of the significance of difference between the genomic feature comparisons are given below each pair of box plots separated by a box plot depicting the feature for all genes together. The bottom and top of the box are represents the 25th and 75th percentile (the lower and upper quartiles, respectively), and the band near the middle of the box represents the 50th percentile (the median). The ends represent the lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Any data not included between the ends are plotted as an outlier with a dot (HK - housekeeping genes, TS - tissue-specific genes).

Genes subject to positive selection are not randomly distributed on the genome

Out of the 71 genes reported to be subject to positive selection [21], 54 could be mapped to bovine Ensembl genes with a known genomic location. Of these 54 genes, 12 were found to be located within RIDGEs whereas only one was within an anti-RIDGE (Table S5). This suggests that genes under positive selection are not randomly distributed on the genome ($p<0.01$) but are less likely to be within anti-RIDGEs i.e. more likely to be found within RIDGEs and regions of intermediate gene expression.

5.2 Discussion

Hierarchical clustering of the arrays showed that arrays clustered based on tissue type but not on days (day0 and day12). The AP being an endocrine gland is clearly different from the brain tissues and this was evident in the clustering results. Among the brain areas, the AM and HC tended to cluster together which may be explained by the close physical and functional proximity between these tissues in the brain. Similarly, the DH and VH are closer to each other physically and functionally as they share many nuclei.

A number of regions of high and low gene expression along the chromosomes were identified. The probability to find as many regions of high and low gene expression as found in our analysis purely by chance is below 10% for RIDGEs and below 1% for anti-RIDGEs, suggesting that RIDGEs and anti-RIDGEs are not a random feature but have an underlying purpose. Given the fact that our data is based on only about 13,000 genes corresponding to good quality probes on the array, we expect more significant results if we had greater number of genes under study. Therefore, even the relatively high 10% chance for RIDGEs may be considered to reflect the trend that RIDGEs are not a random feature in bovine genome. Moreover, the high correlations among the transcriptome maps from the five tissues studied here suggest that regional differences in transcription are a general trend in the bovine genome. High correlations in the global transcriptome profile between diverse tissues have been reported earlier in other species [8].

HK and TS genes were also identified and the fact that HK genes were significantly over-represented on RIDGEs indicates that most of these genes are expressed at a relatively high level. The observations on genomic features of RIDGE versus anti-RIDGE genes and of HK versus TS genes are in agreement with previous studies on human, chicken and other vertebrates. Hence this study extends these general observations also to bovines. The high degree of correlation between gene expression levels and genomic features may indicate an evolutionary adaptation

for energy efficiency of transcription by keeping genes which are highly and widely expressed, short and compact [6,8,9,10].

Analysis of the TS genes for enriched gene ontology terms revealed neurological processes as expected for TS genes. For the HK genes, even though our definition was based on only 5 tissues of which 4 were of neurological type, processes enriched among those genes were general processes related to energy metabolism, cell division, transcription etc. expected of most tissues and cells as they are HK functions. However, it is not possible to conclusively prove that all these genes as truly HK in the absence of information from other diverse tissues and physiological conditions of expression. Hence the HK genes reported here, as defined earlier, only represent genes expressed in all brain tissues represented here and the anterior pituitary.

RIDGEs were suggested to be preferred sites for formation of chromosomal aberrations, development of gene deregulation as in tumours and evolutionary breakpoints in mammalian evolution [26]. Here, we found that positively selected genes were preferentially distributed on RIDGEs and regions of intermediate gene expression compared to anti-RIDGEs. This finding seems to contradict that in an earlier study [16], where it was shown that positively selected genes had reduced expression levels and were expressed in a TS manner in several human tissues. However, it may be noted that selection pressures acting on bovines could be quite different than those in humans, particularly because of the effects of domestication and intensive artificial selection in bovines for traits preferred by humans. For example, positively selected genes in bovines include a number of immune related genes which probably arose in response to the substantial rumen microbial load or due to keeping cattle in herds where rapid disease transmission is a persistent threat [21]. Other points to note while comparing our results with those of other studies which may be based on a large number of diverse tissues is that our definition of HK and TS genes are based on only five closely related tissues and that HK genes as defined in our study could overlap with brain-specific genes of other studies.

To conclude, our findings suggest the existence of chromosomal regional regulation of transcription in the bovine genome. The HK and TS genes reported here represent a useful resource for further studying bovine brain expressed genes. The genomic features observed for genes within RIDGEs and HK genes in bovines agree with previous studies in several other species further strengthening the hypothesis of selective pressure to keep the highly and widely expressed genes short and compact for transcriptional efficiency. Another striking observation was that positively selected genes were non-randomly distributed on the genome with a

preference for RIDGEs and regions of intermediate gene expression compared to anti-RIDGEs.

5.2 Materials and Methods

Ethics Statement

The study was approved by the Animal Care and Ethics Committee of the Animal Sciences Group of Wageningen University and Research Centre, Lelystad (Approval ID 2006087a).

Microarray experiment description and analysis

The microarray experiment was carried out as part of a study aimed at identifying and studying genes that contribute to differences in oestrous behaviour expression and fertility levels of dairy cows [27]. Briefly, oestrous behaviour was recorded in 28 healthy Holstein Friesian cows from 30 days in milk (DIM) onwards till their time of sacrifice which varied between 77 and 139 DIM i.e. after at least 2 oestrous cycles. Samples from 4 brain areas (dorsal hypothalamus, ventral hypothalamus, amygdala and hippocampus) and the anterior pituitary were collected from these cows, 14 of which were sacrificed at start of oestrus and 14 at mid of oestrous cycle. The cows were euthanized in a stress-free, quick and standardized way and all efforts were made to minimize suffering.

RNA extracted from brain tissue samples were hybridized on Bovine 24K oligonucleotide (70-mer) microarrays designed and produced by the Bovine Oligonucleotide Microarray Consortium (BOMC), USA (http://www.bovineoligo.org/). The procedures followed for tissue collection, RNA isolation and microarray hybridization were as described in our earlier study [27]. Briefly, a total of 280 arrays (i.e. 14 cows×2 phases×5 tissues×2 for dye swaps) were prepared in a common reference design with the dye labels swapped between individual samples from each brain area and a reference sample consisting of equal proportions of RNA from all tissues of all cows. Microarray data pre-processing and analysis was done using the LIMMA (Linear models for microarray data) package [28] within Bioconductor project [29] of R statistical programming language (http://www.r-project.org). Briefly, background correction was performed using LIMMA's 'normexp' method (with an offset of 50) followed by within-array 'print tip loess' normalisation and between-arrays quantile normalisation ('Aquantile' method). We then transformed the normalized data, converting M-values back to normalised intensity values so that we could perform an intensity based analysis rather than ratio based. The final gene expression data matrix analysed here was obtained by averaging the intensities of the red and green labelled samples per

individual per tissue, then averaging the median gene expression across individuals per phase, and finally averaging the expression values for genes represented by multiple probes on the array.

All microarray experiment data is MIAME compliant and has been deposited in ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/) (accession number: E-TABM-916). The original annotation provided by BOMC for the bovine 24K oligonucleotide microarray dates back to June 2007. For our analysis, we used the bovine oligonucleotide array probe re-annotation (Version 5) based on Ensembl (http://www.ensembl.org) release 56 (October 2009) provided on the EADGENE website (http://www.eadgene.info/ToolsResources/EADGENEOligoSetsAnnotationFiles/tabid/324/Default.aspx) by the authors of the oligo-set re-annotation pipeline, sigReannot [30]. For the re-annotation, out of the 23,496 probes (excluding control probes) on the bovine oligonucleotide array, only 16,620 probes that were assigned a quality score between 1 and 4 for their specificity to hits on the bovine genome were considered. Probes with quality scores between 5 and 7 had either no hits or multiple hits and were not annotated as they were not specific.

Identifying genomic regions of high and low gene expression

Similar to the protocol followed by Nie *et al*. [6] for identifying regions of high and low gene expression levels, we first calculated the gene expression levels along the chromosome at the position of every gene on it using moving medians based on a window size of 39 genes i.e. the median of the expression level of that gene and that of 19 genes flanking it on either side. The Robust Scatter Plot Smoothing technique (function 'runmed' in R package 'stats') was used to calculate the moving median gene expression. The chosen window size of 39 genes falls within the range of sizes established for chromosomal domains [3] and is similar to the size used in earlier studies [1,2,6]. We then defined RIDGEs and anti-RIDGEs as those regions on the chromosome that contain contiguous stretches of at least 10 gene positions where the moving median gene expression level at each gene position is a certain fold threshold higher or lower than the overall genomic median expression for the bovine transcriptome [1,2,3,6]. These higher and lower fold thresholds were identified by testing a range of thresholds to select the ones that resulted in the identification of RIDGEs and anti-RIDGEs to cover about 10% of the genome each. As a consequence, the gene expression fold thresholds are not symmetric. The choice of 10% coverage is arbitrary and we could have chosen other realistic genome coverage thresholds above or below this level too, provided enough genes are selected from RIDGEs and anti-RIDGEs to support the study of the genomic

features for genes in these regions. Here, we chose 10% genome coverage threshold to maintain uniformity with previous studies [6,26] and to provide a reasonable number of genes for further study. The rationale for using a genome coverage threshold, in addition to expression fold threshold, was to ensure that the total region studied is constant for both RIDGEs and anti-RIDGEs. In the absence of any prior knowledge on the presence and extent of coverage of RIDGEs and anti-RIDGEs on bovine genome and given the fact the there are several criteria that could affect the size of the RIDGEs and anti-RIDGEs, we set the genome coverage threshold to a constant value in order to make the study consistent. In addition, the criteria of genomic coverage thresholds will make the methodology used in this study scalable to any microarray platform or species.

Testing the probability of finding RIDGEs and anti-RIDGEs by chance

We know from studies in other species that RIDGEs and anti-RIDGEs represent a higher order structure in the genome and there is a non-random distribution or clustering of genes based on their expression levels [1]. Here, after identifying RIDGEs and anti-RIDGEs with the criteria we defined, we tested whether the numbers of RIDGEs and anti-RIDGEs identified in bovines were indeed greater than that expected by random chance. The number of regions identified would be a function of the region size which depends on criteria we defined i.e. a combination of thresholds for gene expression level, genome coverage, window size and contiguous stretch of gene positions satisfying the gene expression threshold. Nie *et al*. [6] showed that with window sizes of 29, 39 and 59, the number of RIDGEs identified reduced proportionately but it was still significantly higher in the actual genome than on a randomised genome based on permutations of gene positions. Here too, we do the random permutation test to test whether, with similar settings, there would be an over-representation of regions with overall high or low gene expression on the actual genome compared to the randomised genome. In brief, the test involved permuting the genomic locations of Ensembl genes on the genome and repeating the RIDGE/anti-RIDGE analysis 10,000 times to create 10,000 random transcriptome maps and then comparing the number of RIDGEs/anti-RIDGEs thus identified to the number identified in the actual analysis.

Testing the correlation between tissue-specific transcriptome maps

To test for pair wise correlations among the transcriptome maps on all the chromosomes (applied to the running median expression values with window size of 39 genes) across tissues, the non-parametric Spearman correlation test was used on the ranks of the paired transcriptome maps.

Identifying 'Housekeeping' and 'Tissue-specific' genes and their functional analysis

We defined a threshold level for expression as the 99.9% quantile value of the expression levels of all negative controls on the array. Genes expressed above this level specifically in one tissue alone were defined as TS genes and those expressed in all five tissues were defined as HK genes. In order to test whether the identified TS genes and HK genes contained genes with specific functions expected of these categories, a functional analysis was done using R package 'GOstats' [31]. To take advantage of the better annotation available for human genes, we first converted the cattle genes to their orthologous human genes. Genes in the study sets were tested for enriched Gene Ontology (GO) [32] terms and KEGG [33] pathway terms. The population set against which the study set genes were tested consisted of the Ensembl IDs of the orthologous human genes, of which 11,589 remained after removing duplicates.

Testing RIDGEs for over-representation of housekeeping or tissue-specific genes

For testing if HK genes are over-represented on RIDGEs, we first randomly sampled the same number of genes from the total Ensembl genes on the array as the number of HK genes found in our analysis. We repeated this sampling procedure 10000 times and for each random sample, we calculated the percentage overlap of the sampled genes with the genes found to be present on the RIDGEs. We performed a similar procedure to test if TS genes are over-represented on RIDGEs. Based on these, we derived the distributions of expected percentages of HK genes and of TS genes on RIDGEs under the null hypothesis. We compared this with the actual percentage of HK genes and TS genes found on RIDGEs in our analysis.

Genomic features of RIDGE versus anti-RIDGE genes and 'Housekeeping' versus 'Tissue-specific' genes

Similar to the protocol followed by Nie *et al*. [6], genomic features of RIDGEs versus anti-RIDGEs and HK versus TS genes were determined and the significance of the differences between each pair for all features were statistically tested using Wilcoxon rank-sum test (function '*Wilcox.test*' in R package '*stats'*). The features compared were: gene length (genomic length), intron length (averaged intron length of all transcripts per gene), GC content, transcript length, exon count and exon length. The information required for these calculations was retrieved from the Ensembl genome database using BioMart [34] on 1st September 2010.

Genomic distribution of genes subject to positive selection in bovines

To analyse the relation between chromosomal regions of differing gene expression levels and genes under positive selection, we examined the genomic distribution of 71 bovine genes previously identified as being subject to positive selection [21]. Using Fisher's Exact Test, we tested whether the positively selected genes have a non-random pattern of distribution on the genome with respect to RIDGEs, anti-RIDGEs or regions of intermediate gene expression levels.

Supporting Information

Figure S1. Hierarchical clustering of microarrays.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s001
Hierarchical cluster analysis of the gene expression intensities for all individuals showed that arrays from the same brain area tended to cluster together but the effect of day was not evident. (AP-Anterior Pituitary, AM-Amygdala, HC-Hippocampus, DH-Dorsal Hypothalamus, VH-Ventral Hypothalamus).

Figure S2. Chromosome wise transcriptome maps depicting the identified RIDGEs and anti-RIDGEs.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s002
The y-axis shows the median gene expression levels (log transformed normalised gene intensities ranging from 9 to 14) for the Ensembl genes on each chromosome represented on the array. The solid red and blue lines depict the RIDGEs and anti-RIDGEs respectively whereas the dotted red and blue lines represent the expression threshold to qualify as RIDGE or anti-RIDGE respectively.

Figure S3. Genomic features of housekeeping vs. tissue-specific genes and of genes present on the RIDGEs vs. anti-RIDGEs.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s003
The following genomic features are represented here: Transcript length, exon count and exon length. The p-value of the significance of difference between the genomic feature comparisons is given below each pair of boxplots separated by a boxplot depicting the feature for all genes together. The bottom and top of the box are represents the 25th and 75th percentile (the lower and upper quartiles, respectively), and the band near the middle of the box represents the 50th percentile (the median). The ends represent the lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile.

Any data not included between the ends are plotted as an outlier with a dot (HK - housekeeping genes, TS - tissue-specific genes).

Table S1. Expression data of 13,234 genes in five bovine tissues.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s004

Table S2. Genomic locations of RIDGEs and anti-RIDGEs on the bovine genome.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s005

Table S3. List of housekeeping and tissue-specific genes.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s006

Table S4. Enriched Gene Ontology and KEGG pathway terms in the lists of housekeeping and tissue-specific genes.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s007

Table S5. Genomic locations of genes subject to positive selection and their overlap with RIDGEs/ anti-RIDGEs.
http://www.plosone.org/article/fetchSingleRepresentation.action?uri=info:doi/10.1371/journal.pone.0020413.s008

Acknowledgements

We thank Herman Mulder (for help with the statistics), Henri Woelders (for helpful discussions) and Agnes de Wit (for performing the microarray work), all of whom work at the Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Lelystad, The Netherlands. The study was financially supported by the international cattle breeding co-operative CRV and has been co-financed by the European Commission, within the 6th Framework Programme, contract No. FOOD-CT-2006-016250 (SABRE) and FOOD-CT-2004-506416 (EADGENE). The text represents the authors' views and does not necessarily represent a position of the European Commission, which will not be liable for the use made of such information.

Author contributions

Conceived and designed the experiments: AK HN MG MP RV MS; Performed the experiments: AK HN; Analysed the data: AK HN; Contributed reagents/materials/ analysis tools; AK HN MG MP RV MS; Wrote the manuscript: AK; Edited the manuscript: HN MG MP RV MS.

References

1. Caron H, van Schaik BDC, van der Mee M, Baas F, Riggins G, *et al*. (2001) The Human Transcriptome Map: Clustering of Highly Expressed Genes in Chromosomal Domains. Science 291: 1289–1292.
2. Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, *et al*. (2003) The Human Transcriptome Map Reveals Extremes in Gene Density, Intron Length, GC Content, and Repeat Pattern for Domains of Highly and Weakly Expressed Genes. Genome Res 13: 1998–2004.
3. Gierman H, Indemans M, Koster J, Goetze S, Seppen J, *et al*. (2007) Domain-wide regulation of gene expression in the human genome. Genome Res 17: 1286–1295.
4. Lercher MJ, Urrutia AO, Hurst LD (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nature Genet 31: 180–183.
5. Lercher MJ, Urrutia AO, Pavlicek A, Hurst LD (2003) A unification of mosaic structures in the human genome. Human Mol Genet 12: 2411–2415.
6. Nie H, Crooijmans R, Bastiaansen J, Megens H-J, Groenen M (2010) Regional regulation of transcription in the chicken genome. BMC Genomics 11: 28.
7. Nie H, Crooijmans RPMA, Lammers A, van Schothorst EM, Keijer J, *et al*. (2010) Gene Expression in Chicken Reveals Correlation with Structural Genomic Features and Conserved Patterns of Transcription in the Terrestrial Vertebrates. PLoS ONE 5: e11990.
8. Eisenberg E, Levanon EY (2003) Human housekeeping genes are compact. Trends Genet 19: 362–365.
9. Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV, Kondrashov FA (2002) Selection for short introns in highly expressed genes. Nature Genet 31: 415–418.
10. Vinogradov AE (2004) Compactness of human housekeeping genes: selection for economy or genomic design? Trends Genet 20: 248–253.
11. Singer GAC, Lloyd AT, Huminiecki LB, Wolfe KH (2005) Clusters of Co-expressed Genes in Mammalian Genomes Are Conserved by Natural Selection. Mol Biol Evol 22: 767–775.
12. Mijalski T, Harder A, Halder T, Kersten M, Horsch M, *et al*. (2005) Identification of coexpressed gene clusters in a comparative analysis of transcriptome and proteome in mouse tissues. Proc Natl Acad Sci USA 102: 8621–8626.
13. Spellman P, Rubin G (2002) Evidence for large domains of similarly expressed genes in the Drosophila genome. J Biol 1: 5.

14. Boutanaev A, Kalmykova A, Shevelyov Y, Nurminsky D (2002) Large clusters of co-expressed genes in the Drosophila genome. Nature 420: 666–669.
15. Liu Y, Qin X, Song X-Z, Jiang H, Shen Y, *et al*. (2009) Bos taurus genome assembly. BMC Genomics 10: 180.
16. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, *et al*. (2008) Patterns of Positive Selection in Six Mammalian Genomes. PLoS Genet 4: e1000144.
17. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4: e72.
18. Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L (2009) Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. PLoS Genet 5: e1000341.
19. Flori L, Fritz S, Jaffrezic F, Boussaha M, Gut I, *et al*. (2009) The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. PLoS ONE 4: e6595.
20. Biswas S, Akey JM (2006) Genomic insights into positive selection. Trends Genet 22: 437–446.
21. Elsik CG, Tellam RL, Worley KC, The Bovine Genome Sequencing Analysis Consortium (2009) The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. Science 324: 522–528.
22. Donthu R, Lewin H, Larkin D (2009) SyntenyTracker: a tool for defining homologous synteny blocks using radiation hybrid maps and whole-genome sequence. BMC Res Notes 2: 148.
23. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, *et al*. (2004) The DNA sequence and biology of human chromosome 19. Nature 428: 529–535.
24. Martin J, Han C, Gordon LA, Terry A, Prabhakar S, *et al*. (2004) The sequence and analysis of duplication-rich human chromosome 16. Nature 432: 988–994.
25. Zody MC, Garber M, Adams DJ, Sharpe T, Harrow J, *et al*. (2006) DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. Nature 440: 1045–1049.
26. Folle GA, Liddle P, Lafon-Hughes L, Di Tomaso MV (2010) Close Encounters: RIDGEs, Hyperacetylated Chromatin, Radiation Breakpoints and Genes Differentially Expressed in Tumors Cluster at Specific Human Chromosome Regions. Cytogenet Genome Res 128: 17–27.
27. Kommadath A, Mulder HA, de Wit AAC, Woelders H, Smits MA, *et al*. (2010) Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows. Animal 4: 1297–1307.

28. Smyth GK (2005) Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer. pp. 397–420.
29. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, *et al*. (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80.
30. Casel P, Moreews F, Lagarrigue S, Klopp C (2009) sigReannot: an oligo-set re-annotation pipeline based on similarities with the Ensembl transcripts and Unigene clusters. BMC Proc 3: S3.
31. Falcon S, Gentleman R (2007) Using GOstats to test gene lists for GO term association. Bioinformatics 23: 257–258.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, *et al*. (2000) Gene Ontology: tool for the unification of biology. Nature Genet 25: 25–29.
33. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28: 27–30.
34. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, *et al*. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21: 3439–3440.

6

General discussion

6.1 Background

Oestrous behaviour (OB) is considered one of the key fertility traits in dairy cows. Modern dairy farming systems practise artificial insemination (AI) of cows. Proper timing of AI is crucial to maximize conception rates. The timing of AI is determined based on the observation of behavioural signs of oestrus which is the phase in oestrous cycle when the cow is sexually receptive or "in heat". Therefore it is important that cows express OB well enough to enable proper detection of oestrus. Several management tools now exist to aid oestrus detection. Most of these are based on recording of increased physical activity or evidence of the cow having been mounted by herd mates or teaser animals or on the milk/blood progesterone profile. The expression of OB has however been declining both in intensity and duration over the past few decades along with the decline in several other fertility traits (Lopez *et al.*, 2004). Poor oestrus detection leads to missing the appropriate timing for AI which in turn results in poor conception rates, extended inter-calving intervals and associated financial losses. An understanding of the genomic regulation of OB in dairy cows will present new opportunities for managing this trait to help improve fertility of dairy cows. Considerable work has been done to understand the genomic mechanisms regulating OB using model animals, mainly rodents (Frohlich *et al.*, 1999, Lee and Pfaff, 2008, Pfaff *et al.*, 2008, Vasudevan and Pfaff, 2008, Pfaff *et al.*, 2011). However, little or no research has been conducted to study the same in dairy cows or other farm animals. The research described in this thesis therefore aimed to improve the understanding of genomic regulation of OB in dairy cows by studying gene expression in the bovine anterior pituitary (AP) and four brain areas (amygdala (AM), hippocampus (HC), dorsal hypothalamus (DH) and ventral hypothalamus (VH)) that are involved in regulating OB.

In chapters 2 and 3, we specifically addressed the analyses of the gene expression data from the AP and brain areas to identify genes whose expression patterns showed significant linear or non-linear associations with OB scores. In chapter 4, we investigated gene co-expression networks to identify genes and biological processes shared among the AP and four brain areas that act together to regulate OB. In chapter 5, we looked into the phenomenon of chromosomal regional regulation of transcription and certain genomic properties of genes belonging to these regions as well as of tissue-specific and housekeeping genes in the AP and brain areas. The main findings of the chapters 2-5 are discussed in the following sections to arrive at a more complete picture of the genomic regulation of OB in dairy cows. Genes differentially expressed in the AP between day0 and day12 are then discussed in the context of OB related processes. This is followed by a brief

discussion on regulation of gene expression. The advantages and limitations of the analytical methods used in this thesis are discussed next followed by some points that would be useful to consider in future studies on genomic regulation of OB. I end this chapter with a discussion on the steps needed to apply the new knowledge in practice to optimize fertility in dairy cows.

6.2 Genomic regulation of oestrous behaviour in dairy cows

The genomic regulation of OB in dairy cows is discussed here using the framework of the GAPPS modules (Mong *et al.*, 2003, Mong and Pfaff, 2004). As introduced in chapter 1, GAPPS represents a set of five modules, based mainly on studies in rodents, to help explain the causal relations between sex hormones and female mating behaviour. A summary of the main findings is provided in Table 6.1.

Table 6.1 Oestrous behaviour associated genes and processes in dairy cows known to correspond to processes within the GAPPS modules.

GAPPS modules	Characteristics	Corresponding genes and processes in cows
Growth	Increase in the input / output connections for behaviour-directing hypothalamic neurons	Synaptic plasticity • Immune related genes - *CTLA4, IL1RL1, MARCO* • Neurotransmitter receptors - *CHRM1*, *CHRM3*, *CHRNA5* • Ribosomal genes - *RPL14, RPL18, RPL24, RPS11, RPS18* • Others - *NEFL, NDRG2, THY1*
Amplification	Amplification of oestrogen effect by progesterone mediated by progesterone receptor	*PGR* gene up-regulated in the AP at day0
Preparation	Preparation for mating	Female sexual receptivity • *OXT, AVP, HTR2A, DRD2, GABRA6* Anxiolytic effect • *OXT, TTR, KCNN2* Altered feeding motivation and mood • *POMC, MCHR1, MOBP, LTA4H*
Permission	Permissive actions by hypothalamic neurons for the mating behaviour to occur	Arousal, activation of protein kinases and release of **C**a^{2+} • *CHRM1, CHRM3, CHRNA5, PLCB2, ITPKA*
Synchronisation	Synchronise mating behaviour with ovulation	Prostaglandin regulators • *PTGDS, PTGIS, PTGFR*

The first of these modules is the hormone dependent **Growth** (G) response, driven by oestrogen and progesterone that triggers an increase in the input and output connections for behaviour-directing hypothalamic neurons. This may be considered equivalent to processes related to synaptic plasticity found in our analyses. The genes responsible for synaptic plasticity were found here not only in the hypothalamic neurons but also other brain areas and the AP during the time of oestrus. Several "immune-related" genes found in the association studies (chapters 2-3) like *CTLA4*, *IL1RL1*, *MARCO*, *FCRLA*, *IL33*, *CCL26* and *CXCL10* may facilitate remodelling of synaptic networks. It has been shown that immunoglobulin superfamily proteins play important roles in brain developmental processes and the functioning of neuronal networks in adults because they provide the ideal structure for protein–protein interactions and, thus, cell-cell interactions (Rougon and Hobert, 2003). The co-expression study (chapter 4) also revealed certain hub genes within gene co-expression networks between brain areas and AP that have functions related to neuronal growth or plasticity - *NEFL* (Terry-Lorenzo *et al.*, 2000), *NDRG2* (Takahashi *et al.*, 2005), *THY1* (Rege and Hagood, 2006) and *GAP43* (Gispen *et al.*, 1991). The gene *NEFL*, for example, regulates the dephosphorylation of phosphoproteins implicated in synaptic plasticity. Further, several ribosomal genes associated with OB in the association studies (*RPL24*) or the co-expression studies (*RPL14*, *RPL18*, *RPS11*, *RPS18* etc.) across several brain areas. The increased synthesis of ribosomal RNA in the ventromedial hypothalamus has been reported to be one of the early effects of oestrogen administered sub-cutaneously in ovariectomised rats and represents a primary event in the activation of neuronal cells and neuronal pathways involved in female reproductive behaviour (Jones *et al.*, 1990). Oestrogen driven expansion of dendrites and synapses follows from the stimulation of synthesis of ribosomal RNA (Mong and Pfaff, 2004, Pfaff, 2005).

The second module in GAPPS is the **Amplification** (A) of the oestrogen effect by progesterone mediated by nuclear progesterone receptor (PGR). In our study, the expression of *PGR* was not found associated with the level of OB in any of the tissues studied. However, a differential expression analysis between day0 and day12 of the oestrous cycle (described in section 6.3) revealed that *PGR* was up-regulated in the AP at day0. Therefore the increased expression of *PGR* during day0 supports the amplification of the oestrogen effect in cows which is a requisite for OB to occur. However, the level of OB displayed did not depend on the *PGR* expression level.

Next, the **Preparation** (P) module in GAPPS refers to preparation of the female for mating which involves analgesia, anxiety reduction, social recognition and aggression. Among the genes found associated with OB in several brain areas in our

analyses were oxytocin (*OXT*) and arginine vasopressin (*AVP*). Oxytocin, produced by the supraoptic and paraventricular nuclei of the hypothalamus is released within the brain where it acts on specific oxytocin receptors to elicit effects like female sexual receptivity, grooming behaviour and partner bonding (Leng *et al.*, 2008). In the presence of oestrogen, oxytocin exerts an anxiolytic effect, mediated by increases in oxytocin binding density in the lateral septum, thereby favouring courtship and mating (McCarthy *et al.*, 1997, Mong and Pfaff, 2004). Similar to oxytocin, vasopressin is associated with sexual behaviour and bonding and its expression is under the control of oestrogen and progesterone (Patisaul *et al.*, 2003, Kalamatianos *et al.*, 2004, Curley and Keverne, 2005, Donaldson and Young, 2008). The genes *POMC*, *MCHR1*, *CCK*, *DRD2*, *HTR2A* and *GABRA6*, which associated with OB score in at least one of the brain areas studied here (details in chapter 3), are known to modulate emotional states like anxiety and satiety (Rex *et al.*, 1997, Marsh *et al.*, 2002, Uhart *et al.*, 2004, Millington, 2007). The link between fertility and appetite is evident from the finding of genes *POMC* and *MCHR1*, both of which play roles in feeding behaviour, metabolic rate and feed intake (Marsh *et al.*, 2002, Millington, 2007). It is known that interactions between monoamines (dopamine, serotonin, noradrenaline) and steroid hormones play a major role in the integration of reproductive behaviour and gonadal function (Fabre-Nys, 1998). In ewes, dopamine-mediated D2 receptor (*DRD2*) signalling in the mediobasal hypothalamus is known to affect female sexual motivation and receptivity (Fabre-Nys *et al.*, 2003). Further, the perception and awareness of male-related cues differs with the stage of oestrous cycle, with releases of monoamines (linked to *HTR2A* and *DRD2*) and gamma-aminobutyric acid (*GABA*) (linked to *GABRA6*) in the mediobasal hypothalamus being triggered by such cues only when ewes are in oestrus (Fabre-Nys *et al.*, 1997). Studies in female rats and hamsters have shown the inhibitory and facilitator effects of serotonin receptor agonists and antagonists on the hypothalamic regulation of sexual receptivity (Uphouse, 2000, Caldwell and Albers, 2002). This regulation is also mediated by GABAergic neurons interacting with serotonin containing neurons. Many other genes that we found in our analyses are known from studies in other species to play a role in emotional responses, e.g. *TTR* has been linked to anxiety (Sousa *et al.*, 2004), *MOBP* to mood disorders (Sokolov, 2007), *LTA4H* to depression (Zhao *et al.*, 2009) and *KCNN2* to anxiety and stress responses (Mitra *et al.*, 2009).

The next GAPPS module called **Permission** (P) refers to the permissive actions by hypothalamic neurons for the mating behaviour to occur. Involved in this process are noradrenergic (NA) α-1b receptors as well as muscarinic receptors which are induced by oestrogen in the ventromedial hypothalamic neurons. Our finding of

the association between OB scores and the expression of *ACHE* and several cholinergic receptors (*CHRM1*, *CHRM3* and *CHRNA5*) can be explained by the effect of the neurotransmitter acetylcholine on arousal, plasticity and reward. The products of the muscarinic cholinergic receptor genes, *CHRM1* and *CHRM3*, are G_q-protein coupled receptors whose activation releases intracellular calcium (Ca^{2+}) via the phospholipase C - inositol 1,4,5-trisphosphate signalling pathway (Billups *et al.*, 2006). The genes for phospholipase C and inositol triphosphate kinase (*PLCB2*, *ITPKA*) and several protein kinases were also found associated to OB scores. These findings can be explained based on the hypothesis put forward by Kow and Pfaff (2004) that the membrane actions of oestrogen can modulate the genomic actions of oestrogen and that this transcriptional potentiation is mediated via signalling pathways requiring the activation of certain protein kinases and increased intracellular Ca^{2+}.

Finally, the OB score associated genes (*PTGDS*, *PTGIS*, *PTGFR*) that regulate prostaglandin functioning may be involved in the last GAPPS module, i.e. **Synchronisation** (S) of mating behaviour with ovulation where oestrogen inducts GnRH. Prostaglandins are known to be under the influence of oestrogen (Amateau and McCarthy, 2002) and are capable of directly affecting neurons that synthesize and secrete gonadotropin-releasing hormone (Jasoni *et al.*, 2005, Clasadonte *et al.*, 2011).

It was thus possible to place many of the genes and processes that we identified as associated with bovine OB regulation within the framework of the GAPPS modules. This provided new insight into the similarity of processes involved in the genomic regulation of OB in dairy cows compared to rodent models, though certain individual genes may vary between the two species. There were also several OB associated genes which were not well-annotated or whose current annotation does not reveal its role in OB regulation. These genes will be especially interesting to target in future research to establish their potential role in OB regulation.

6.3 Genes differentially expressed between oestrus and mid-cycle

To identify genes whose expression related to the phase of the oestrous cycle, we performed differential expression analyses of the AP and the brain areas between day0 and day12 cows (unpublished study). This analysis revealed 29 differentially expressed probes (27 up regulated and 2 down regulated) in the AP and none in the brain areas that showed at least a 2-fold change in expression at a p-value of below 0.01. The failure to identify any differentially expressed genes in the brain areas between the two phases may be because of one or more of the following

reasons: phase specific genes were expressed in the brain at levels too low to be detected or at levels that did not differ significantly between the phases or certain probes corresponding to phase specific genes were missing on the microarray.

Regarding the differentially expressed probes in the AP, only 10 probes had specific genes annotated to them based on the microarray probe re-annotation (Casel *et al.*, 2009) that we followed. These included the following genes, all of which were up-regulated at day0: *NPPC, SULF2, PGR, PDLIM3, LARP1, C17ORF48, KIAA1755, FXC1, CLEC14* and *KAZN*. This small number of genes did not reveal any significant biological processes to explain their functioning together using gene ontology (GO) (Ashburner *et al.*, 2000) and KEGG pathway (Kanehisa and Goto, 2000) enrichment tests. However, the functions of some of the differentially expressed genes fitted with key biological processes associated with OB regulation namely anxiolytic activities and synaptic plasticity in the brain (Frohlich *et al.*, 1999, Lee and Pfaff, 2008, Pfaff *et al.*, 2008, Vasudevan and Pfaff, 2008, Pfaff *et al.*, 2011). For example, the product of the gene *NPPC* called C-type natriuretic peptide (CNP) was suggested to have a role in mediating anxiety in rats via a mechanism that depends on corticotrophin-releasing hormones (Bíró *et al.*, 1996, Montkowski *et al.*, 1998). Anxiolytic effects mediated by hormones like oxytocin are known to favour courtship and mating during oestrus (McCarthy *et al.*, 1997, Mong and Pfaff, 2004), and *NPPC* could also be playing a part in this process. Further, extensive localisation of CNP transcripts have been reported in several brain areas like the olfactory nuclei of the forebrain, anteroventral periventricular nuclei and arcuate nuclei of the hypothalamus, limbic cortices and hippocampal subfields CA1-3 (Langub *et al.*, 1995), some of which are known to be involved in mediating emotional responses. Another gene, *SULF2* was found to be associated with neurite outgrowth, synaptic plasticity and behaviour in mice (Kalus *et al.*, 2009), which are functions shared by several other genes found in the association analyses described in the previous section. The gene *PGR* was an obvious candidate because progesterone is a major hormone regulating female reproductive cycle and it is known that reproductive behaviours and progesterone-regulated neural responses are mediated by progesterone receptors (Brinton *et al.*, 2008).

The differential expression analysis therefore revealed oestrus phase up-regulated genes in the AP some of which are known to be involved in processes related to OB. However, not much is currently known about the other genes in the list of differentially expressed genes to link them to processes associated with the phase of oestrous cycle or OB. Therefore it would be interesting to investigate these genes for novel alternative functions that have to do with processes related to oestrus.

6.4 Regulation of gene expression

Now that several genes and biological processes associated with OB regulation have been identified, a valuable next step would be to identify the regulatory factors like transcription factors (TFs) or microRNAs (miRNAs) that control the expression of these genes. Identifying these factors is important to help reconstruct gene regulatory networks. However, identifying genes encoding TFs in microarray experiments is difficult because of their generally low expression levels and because their activity is mainly regulated at the post-transcriptional level (Kel *et al.*, 2006). The fact that gene expression is controlled not only by individual TFs but also by combinatorial interaction of TFs and their binding motifs on the genome further complicates the reconstruction of gene regulatory networks. An indirect way to determine the TFs involved is to analyse the promoters of co-expressed genes for common TF binding sites. However, this approach has several limitations: combinatorial regulation by TFs is not addressed in this approach; relevant TFs may not be identified even when over-represented binding sites are found; and poor statistical significance of the results (Kato *et al.*, 2004). Genome-wide analysis of TF binding sites by chromatin immunoprecipitation (ChIP) based assays like ChIP-chip (Ren *et al.*, 2000, Lieb, 2003) or ChIP-Seq (Robertson *et al.*, 2007) using bovine AP and brain tissue samples will help reveal TFs specific to these tissues. This information combined with the OB associated genes and co-expressed genes found in our analysis will help identify the TFs involved in regulating genes associated with OB.

At the level of post-transcriptional gene regulation, one of the most important players is the miRNA. A relatively large number of the known miRNAs are expressed in the mammalian brain (Sempere *et al.*, 2004, Nelson *et al.*, 2006), and at comparatively high levels (Fiore *et al.*, 2008). Brain expressed miRNAs have been shown to be involved in neuronal development and differentiation (Makeyev *et al.*, 2007, Visvanathan *et al.*, 2007), neuronal morphogenesis (Wayman *et al.*, 2008), modulation of the circadian clock (Cheng *et al.*, 2007) and in synapse functioning (Schratt, 2009). Further, there is now evidence to link miRNAs with neurodegenerative disorders like Alzheimer's disease, Parkinson's disease, Schizophrenia and others (reviewed in Gerhard, 2009, Saba and Schratt, 2010) where behavioural changes are evident. The above mentioned functions of miRNAs hint to their likely involvement in the regulation of genes associated with OB, several of which have been found here to be involved in biological processes like synaptic plasticity or with processes affected in neurodegenerative disorders. Stable extra-cellular miRNAs have even been identified in various body fluids, thus

generating great interest in miRNAs as potential biomarkers and therapeutic targets in neurological disorders (Etheridge *et al.*, Siegel *et al.*, 2011). Therefore studying the role of miRNAs in the brain in relation to OB is especially interesting to identify miRNAs associated with OB regulating genes and to explore the possibility of their use as biomarkers. Microarrays and next generation sequencing technologies have been developed to detect miRNAs. The current release (Release 17: April 2011) of the online repository (http://www.mirbase.org) for published microRNA sequences and their associated annotation (Griffiths-Jones *et al.*, 2005) enlists 662 bovine miRNAs. However, much research is still needed to fully understand the roles of these miRNAs, especially those expressed in the brain.

Another aspect associated with regulation of gene expression is the clustering of genes on chromosomes based on their expression levels. Chromosomal regions of increased or decreased gene expression (RIDGEs and anti-RIDGEs) compared to the average gene expression across the genome have been shown in several species (Caron *et al.*, 2001, Versteeg *et al.*, 2003, Gierman *et al.*, 2007). In chapter 5, we showed that this phenomenon of chromosomal regional regulation of transcription exists in the bovine genome too. Further, the genes we identified as expressed in the AP and four brain areas (which we referred to as housekeeping genes) were over-represented within RIDGEs but genes found expressed in only one of the five tissues studied here (referred to as tissue-specific genes) were not. Housekeeping genes and genes within RIDGEs had, in general, higher expression levels and GC content but shorter gene lengths and intron lengths than tissue-specific genes and genes within anti-RIDGES. The genomic features observed for genes within RIDGEs and housekeeping genes in bovines agree with previous studies in several other species further strengthening the hypothesis of selective pressure to keep highly and widely expressed genes short and compact for transcriptional efficiency (Castillo-Davis *et al.*, 2002, Lercher *et al.*, 2002, Lercher *et al.*, 2003, Vinogradov, 2004). Further, positively selected genes in bovines were found non-randomly distributed on the genome with a preference for RIDGEs and regions of intermediate gene expression compared to anti-RIDGEs. Targets of positive selection have been linked to complex trait phenotypes. Therefore the quest for genes controlling complex traits may be supported by finding signatures of positive selection and regional patterns of gene expression that help narrow down the regions on the genome to look for these genes.

6.5 Analytical methods used to identify genes associated with oestrous behaviour

In chapters 2 and 3, we used a Bayesian hierarchical model based method

(Jia *et al.*, 2008) to identify genes whose expression patterns showed significant linear or non-linear relationships with OB scores. Since genes in a regulatory network may interact in complex ways (feedback mechanisms, cooperation or competition between genes) to result in non-linear associations between gene expression levels and phenotype, it was advantageous to use a model that accounts for both linear and non-linear associations, for example, through polynomial regression. Use of a Bayesian approach over frequentist also had several advantages for the regression model applied to our data. Firstly, inferences are based on the posterior distribution, which provides an easily interpretable alternative to p-values. Secondly, Bayesian interval estimates obtained from Markov chain Monte Carlo (MCMC) procedures are more appropriate in small samples than confidence intervals from ordinary regression (Dunson, 2001). However, as we found out in our analyses, given the relatively small number of animals used in our experiment (14 at day0 and 14 at day12), the most reliable associations were the linearly associated gene expression patterns. Second order associations were less reliable whereas third order or higher associations were untrustworthy as it was not possible to distinguish between true and spurious associations. We tested the reliability of using the Bayesian hierarchical model based method on our dataset by repeating the analyses three times and checking for the consistency of results obtained as well as by performing a leave-one-out analyses (details in chapter 2). The overlap in the results varied between 79-88% for the repeated runs and between 44-76% for the leave-one-out analyses. As expected, the differences were not only in the number of associated genes found but also the pattern of association it was assigned to. Therefore, the relatively low number of animals sampled in this study was a constraint on this methodology due to which, especially the higher order relationships seemed to be in some cases an artefact of the method. We therefore focussed on the more reliably predicted linearly associated genes and then carefully considered genes with higher order associations.

In the association based study mentioned above, we looked separately per brain area into OB associated gene expression. However, in real life scenario, the expression of a complex trait like behaviour may be the result of intricate interactions between different brain areas and the AP in response to hormonal, visual, physiological and other signals. Therefore, in chapter 4, we used a method called 'Weighted Gene Co-expression Network Analysis'(WGCNA) (Zhang and Horvath, 2005, Langfelder and Horvath, 2008) to study gene co-expression patterns preserved among relevant brain areas and AP. The study revealed additional OB associated genes that were not found when studying the brain areas separately in

the association study. Valuable information on hub genes (e.g. *NEFL*, *NDRG2*, *THY1* etc.) and biological processes (e.g. glial cell development, regulation of neural projection development, ribosome and biosynthetic processes etc.) shared between brain areas acting together in regulating OB expression were obtained (details in chapter 4). It should be noted that the products of some of the genes within the co-expressed modules may fall in different cellular compartments and may have no possibility for physical interaction *in vivo*. Nevertheless, their belonging together in a module may indicate that the co-expressed genes across the AP and four brain areas may be responding to common regulatory factors or signals.

6.6 Considerations in future research

I now discuss some points to consider in follow on studies on genomic regulation of OB in dairy cows. In our experiment, to ensure that there were at least a few cows coming into oestrus at the same time, a group size between 15 and 30 was maintained throughout the experiment by adding additional cows from the main herd to replace the experimental cows that were sacrificed. In spite of this, the number of cows that came into oestrus on a particular day still varied. This might have contributed to variations in the intensity of OB displayed as the number of cows in oestrus at the same time in a herd can positively influence the behaviour displayed (Hurnik *et al.*, 1975). Further, to avoid the effects of circadian rhythm on gene expression related to OB, the euthanizing of the cows and sample collection were done only during the morning within a fixed time slot. This might have resulted in slight variations in the progression of oestrus phase in the different cows at the time of their sacrifice. Both the above mentioned factors could have been better controlled by synchronizing the oestrous cycle of experimental cows using external hormonal treatments (Ryan *et al.*, 1999). Many of the studies in rodents and other experimental animals that aimed to study OB, in fact, used external hormone treatments to study consequent changes in gene expression in the brain. However, we did not follow this approach as the external hormone treatment could alter the normal hormonal balance of the cows. The disturbed hormonal balance may also affect the normal expression profile of genes involved in the regulation of OB. Further, the extent to which the normal OB may deviate due to the influence of external hormones is not clear. We therefore opted to study genomic regulation of OB within the normal undisturbed oestrous cycle of the experimental cows. With the current research findings as a benchmark, future research may be conducted after oestrous synchronisation too. An improvement in future studies could be to increase the observation periods for cows in oestrus and

use of oestrus detection aids. Technologies like video recording of the cow herd or pedometers or electronic devices like "Heat watch" (http://www.cowchips.net/) would ensure that cows showing enhanced OB outside of the observation times is accounted for.

In the current experiment, whole brain areas were used as samples for the microarray. Using samples of specific nuclei in the brain which are known to influence sexual behaviour could result in a more specific and sensitive gene expression study. For example, the ventromedial nuclei (Etgen, 1987, Richmond and Clemens, 1988), the medial pre-optic nuclei (Matuszewich *et al.*, 2000), and para-ventricular nuclei (Richards *et al.*, 2010) of the hypothalamus are known to control sexual receptivity and behaviour in female rats. Methods like laser-capture micro-dissection (LCM)(Espina *et al.*, 2006) can be used to specifically collect these nuclei or subpopulations of tissue cells under direct microscopic visualization and use them for the gene expression study, with several methods optimised for brain tissue (Bonaventure *et al.*, 2002, Segal *et al.*, 2005). Bernard *et al.* (2011) recently described a region-specific in situ hybridization-guided LCM method on post-mortem human brain tissue coupled with gene expression quantification.

Another point to consider is to have gene expression measures from several relevant time points across the oestrous cycle. Microarrays only provide a snapshot of the gene expression at a particular point of time when the sample is collected. For a broader perspective on the dynamics of gene expression across the cycle for key genes associated with OB regulation, it would be advantageous to use tissues collected from not only day0 and day12 but also other important time points in the oestrous cycle that correspond to major hormonal changes. For example, microarrays of samples collected on days just preceding oestrus would capture the effects of the GnRH and LH surges on gene expression.

Regarding the choice of microarray platform, while one-colour or two-colour microarray platforms are essentially equivalent in quality of performance (Patterson *et al.*, 2006), for the association and co-expression analyses done here, absolute expression levels (from one-colour arrays) would make the interpretation of results more straightforward than ratio-based expression levels (from two-colour arrays). New technologies like RNA-seq which is based on next generation sequencing not only give absolute expression levels but also offer several advantages over microarrays. Unlike microarrays, RNA-seq does not depend of pre-designed probes but on sequencing, and therefore it can read the entire transcriptome including currently unknown transcripts and splice variants in a relatively unbiased way. RNA-seq is also more sensitive and has a broader dynamic range than microarrays to measure differences in transcript levels. Though RNA-seq

is currently an expensive technology, its costs are rapidly falling. The advantages of RNA-seq over microarrays make it worthwhile to consider using RNA-seq technology in future research on genomic regulation of OB.

6.7 Concluding remarks

This study has successfully identified genes and their associated biological processes to provide a better understanding of the genomic regulation of OB in dairy cows. To make this knowledge practically useful in dairy farming, it is important to first validate the most promising candidate genes identified here, especially the genes also known from literature to have functions related to OB in other species and the hub genes from the co-expression study. How the validation can be achieved would however be a challenge, given the difficulty to collect brain tissue samples from cow during their lifetime. Use of laboratory animals like rodents will allow the use of selective knock out or knock down mutants to study the influence of genes identified in our study. There is also the possibility to use advanced mathematical models to assist in the validation of the roles of these genes. The use of mathematical models for oestrous cycle in cows with inputs from physiological parameters has been reported (Boer *et al.*, 2011). It would be valuable to extend this model to study how changes in expression of key genes identified in our study affect OB expression.

Following validation, the genes may be screened for biomarker development. One possibility could be to correlate the level of specific proteins or metabolites in cow milk or blood with the gene expression profile of genes found here to be correlated to OB. Highly correlated cases may be considered as biomarkers for development of quick assays to determine the oestrous status of the cow. In a recent study (Cai *et al.*, 2010) that identified consensus modules preserved between blood and brain in a gene co-expression based analysis, it was shown that transcriptome organization was generally poorly preserved between these two tissues. However, there were certain genes in the consensus modules, especially those involved in basic cellular processes, which exhibited a strong preservation. Some of the genes identified in our study could be involved in the synthesis of particular metabolites that cross the blood brain barrier or are released into the portal blood and reach general circulation. These metabolites could potentially be measured and should be explored for biomarker development.

Genome-wide association studies (GWAS) based on high throughput SNP genotyping technologies is now a preferred and efficient method to identify SNPs associated with quantitative traits. This method requires phenotypic information on several hundreds to thousands of individuals. In dairy cattle, GWAS has

successfully identified SNPs associated with several fertility traits and production traits (Daetwyler *et al.*, 2008, Lillehammer *et al.*, 2009, Mai *et al.*, 2010, Bouwman *et al.*, 2011). GWAS for OB traits, which has yet to be done, has the potential to reveal OB associated SNPs. Combining the information on genes that associated with OB in our analysis with positions of SNP that may be associated with OB can potentially help identify the most relevant genes affecting OB.

To conclude, this study has revealed important new aspects of the genomic regulation of OB in dairy cows. Key findings are presented within the framework of the GAPPS modules (Table 6.1). The findings here would serve as a valuable reference to provide leads on genes to follow on in future research on understanding OB regulation. Further, it could help geneticists to prioritize the regions on the genome, the QTLs and the genes in the vicinity of SNPs that are associated with fertility traits. The new knowledge could also be used to optimize fertility of dairy cows by aiding to improve existing or to devise novel reproductive management tools like diagnostic tools to determine the reproductive health, energy and/or fertility status of the cow, oestrus detection tools and so on.

References

Amateau SK and McCarthy MM 2002. A novel mechanism of dendritic spine plasticity involving estradiol induction of prostaglandin-E2. J Neurosci 22, 8586-8596.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25-29.

Bernard R, Burke S and Kerman IA 2011. Region-specific in situ hybridization-guided laser-capture microdissection on postmortem human brain tissue coupled with gene expression quantification. Methods Mol Biol 755, 345-361.

Billups D, Billups B, Challiss RAJ and Nahorski SR 2006. Modulation of G_q-protein-coupled inositol trisphosphate and Ca^{2+} signaling by the membrane potential. J Neurosci 26, 9983-9995.

Bíró É, Tóth G and Telegdy G 1996. Effect of receptor blockers on brain natriuretic peptide and C-type natriuretic peptide caused anxiolytic state in rats. Neuropeptides 30, 59-65.

Boer HMT, Stötzel C, Röblitz S, Deuflhard P, Veerkamp RF and Woelders H 2011. A simple mathematical model of the bovine estrous cycle: Follicle development and endocrine interactions. J Theor Biol 278, 20-31.

Bonaventure P, Guo H, Tian B, Liu X, Bittner A, Roland B, Salunga R, Ma X, Kamme F, Meurers B, Bakker M, Jurzak M, Leysen J and Erlander M 2002. Nuclei and subnuclei gene expression profiling in mammalian brain. Brain Res. 943, 38-47.

Bouwman A, Bovenhuis H, Visker M and van Arendonk J 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. BMC Genet 12, 43.

Brinton RD, Thompson RF, Foy MR, Baudry M, Wang J, Finch CE, Morgan TE, Pike CJ, Mack WJ, Stanczyk FZ and Nilsen J 2008. Progesterone receptors: Form and function in brain. Front Neuroendocrinol 29, 313-339.

Cai C, Langfelder P, Fuller T, Oldham M, Luo R, van den Berg L, Ophoff R and Horvath S 2010. Is human blood a good surrogate for brain tissue in transcriptional studies? BMC Genomics 11, 589.

Caldwell HK and Albers HE 2002. The effects of serotonin agonists on the hypothalamic regulation of sexual receptivity in Syrian hamsters. Horm Behav 42, 78-84.

Caron H, van Schaik BDC, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus M-C, van Asperen R, Boon K, Voute PA, Heisterkamp S, van Kampen A and Versteeg R 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. Science 291, 1289-1292.

Casel P, Moreews F, Lagarrigue S and Klopp C 2009. sigReannot: an oligo-set re-annotation pipeline based on similarities with the Ensembl transcripts and Unigene clusters. BMC Proc 3, S3.

Castillo-Davis CI, Mekhedov SL, Hartl DL, Koonin EV and Kondrashov FA 2002. Selection for short introns in highly expressed genes. Nat Genet 31, 415-418.

Cheng H-YM, Papp JW, Varlamova O, Dziema H, Russell B, Curfman JP, Nakazawa T, Shimizu K, Okamura H, Impey S and Obrietan K 2007. MicroRNA modulation of circadian-clock period and entrainment. Neuron 54, 813-829.

Clasadonte J, Poulain P, Hanchate NK, Corfas G, Ojeda SR and Prevot V 2011. Prostaglandin E2 release from astrocytes triggers gonadotropin-releasing hormone (GnRH) neuron firing via EP2 receptor activation. Proc Natl Acad Sci USA 108, 16104-16109.

Curley JP and Keverne EB 2005. Genes, brains and mammalian social bonds. Trends Ecol Evol 20, 561-567.

Daetwyler H, Schenkel F, Sargolzaei M and Robinson J 2008. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. J Dairy Sci 91, 3225 - 3236.

Donaldson ZR and Young LJ 2008. Oxytocin, vasopressin, and the neurogenetics of sociality. Science 322, 900-904.

Dunson DB 2001. Commentary: Practical advantages of bayesian analysis of epidemiologic data. Am J Epidemiol 153, 1222-1226.

Espina V, Wulfkuhle JD, Calvert VS, VanMeter A, Zhou W, Coukos G, Geho DH, Petricoin EF and Liotta LA 2006. Laser-capture microdissection. Nat Protoc 1, 586-603.

Etgen AM 1987. Inhibition of estrous behavior in rats by intrahypothalamic application of agents that disrupt nuclear binding of estrogen-receptor complexes. Horm Behav 21, 528-535.

Etheridge A, Lee I, Hood L, Galas D and Wang K 2011. Extracellular microRNA: A new source of biomarkers. Mutat Res Fundam Mol Mech Mutag 717, 85-90.

Fabre-Nys C 1998. Steroid control of monoamines in relation to sexual behaviour. Rev Reprod 3, 31-41.

Fabre-Nys C, Ohkura S and Kendrick KM 1997. Male faces and odours evoke differential patterns of neurochemical release in the mediobasal hypothalamus of the ewe during oestrus: An insight into sexual motivation. Eur J Neurosci 9, 1666-1677.

Fabre-Nys C, Chesneau D, De La Riva C, Hinton MR, Locatelli A, Ohkura S and Kendrick KM 2003. Biphasic role of dopamine on female sexual behaviour via D2 receptors in the mediobasal hypothalamus. Neuropharmacology 44, 354-366.

Fiore R, Siegel G and Schratt G 2008. MicroRNA function in neuronal development, plasticity and disease. Biochim Biophys Acta - Gene Regul Mech 1779, 471-478.

Frohlich J, Ogawa S, Morgan M, Burton L and Pfaff D 1999. Hormones, genes and the structure of sexual arousal. Behav Brain Res 105, 5-27.

Gerhard S 2009. Fine-tuning neural gene expression with microRNAs. Curr Opin Neurobiol 19, 213-219.

Gierman HJ, Indemans MH, Koster J, Goetze S, Seppen J, Geerts D, van Driel R and Versteeg R 2007. Domain-wide regulation of gene expression in the human genome. Genome Res 17, 1286 - 1295.

Gispen W, Nielander H, De Graan P, Oestreicher A, Schrama L and Schotman P 1991. Role of the growth-associated protein B-50/GAP-43 in neuronal plasticity. Mol Neurobiol 5, 61-85.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A and Enright AJ 2005. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 34, D140-D144.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci 92, 433-443.

Hurnik JF, King GJ and Robertson HA 1975. Estrous and related behaviour in postpartum Holstein cows. Appl Anim Ethol 2, 55-68.

Jasoni CL, Todman MG, Han SK and Herbison AE 2005. Expression of mRNAs encoding receptors that mediate stress signals in gonadotropin-releasing hormone neurons of the mouse. Neuroendocrinology 82, 320-328.

Jia Z, Tang S, Mercola D and Xu S 2008. Detection of quantitative trait associated genes using cluster analysis. In Evolutionary computation, machine learning and data mining in bioinformatics (eds. E Marchiori and JH Moore), pp. 83-94, Springer-Verlag, Berlin Heidelberg.

Jones K, Harrington C, Chikaraishi D and Pfaff D 1990. Steroid hormone regulation of ribosomal RNA in rat hypothalamus: early detection using in situ hybridization and precursor-product ribosomal DNA probes. J Neurosci 10, 1513-1521.

Kalamatianos T, Kalló I, Goubillon M-L and Coen CW 2004. Cellular expression of V1a vasopressin receptor mRNA in the female rat preoptic area: Effects of oestrogen. J Neuroendocrinol 16, 525-533.

Kalus I, Salmen B, Viebahn C, Figura Kv, Schmitz D, D'Hooge R and Dierks T 2009. Differential involvement of the extracellular 6-O-endosulfatases Sulf1 and Sulf2 in brain development and neuronal and behavioural plasticity. J Cell Mol Med 13, 4505-4521.

Kanehisa M and Goto S 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28, 27 - 30.

Kato M, Hata N, Banerjee N, Futcher B and Zhang M 2004. Identifying combinatorial regulation of transcription factors and binding motifs. Genome Biol 5, R56.

Kel A, Voss N, Jauregui R, Kel-Margoulis O and Wingender E 2006. Beyond microarrays: Finding key transcription factors controlling signal transduction pathways. BMC Bioinformatics 7, S13.

Kow L-M and Pfaff DW 2004. The membrane actions of estrogens can potentiate their lordosis behavior-facilitating genomic actions. Proc Natl Acad Sci USA 101, 12354-12357.

Langfelder P and Horvath S 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Langub MC, Watson RE and Herman JP 1995. Distribution of natriuretic peptide precursor mRNAs in the rat brain. J Comp Neurol 356, 183-199.

Lee AW and Pfaff DW 2008. Hormone Effects on Specific and Global Brain Functions. J Physiol Sci 58, 213-220.

Leng G, Meddle SL and Douglas AJ 2008. Oxytocin and the maternal brain. Curr Opin Pharmacol 8, 731-734.

Lercher MJ, Urrutia AO and Hurst LD 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 31, 180-183.

Lercher MJ, Urrutia AO, Pavlicek A and Hurst LD 2003. A unification of mosaic structures in the human genome. Hum Mol Genet 12, 2411-2415.

Lieb JD 2003. Genome-wide mapping of protein-DNA interactions by chromatin immunoprecipitation and DNA microarray hybridization. Methods Mol Biol 224, 99-109.

Lillehammer M, Hayes B, Meuwissen T and Goddard M 2009. Gene by environment interactions for production traits in Australian dairy cattle. J Dairy Sci 92, 4008 - 4017.

Lopez H, Satter LD and Wiltbank MC 2004. Relationship between level of milk production and estrous behavior of lactating dairy cows. Anim Reprod Sci 81, 209-223.

Mai M, Sahana G, Christiansen F and Guldbrandtsen B 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. J Anim Sci 88, 3522 - 3528.

Makeyev EV, Zhang J, Carrasco MA and Maniatis T 2007. The microRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. Mol Cell 27, 435-448.

Marsh DJ, Weingarth DT, Novi DE, Chen HY, Trumbauer ME, Chen AS, Guan X-M, Jiang MM, Feng Y, Camacho RE, Shen Z, Frazier EG, Yu H, Metzger JM, Kuca SJ, Shearman LP, Gopal-Truter S, MacNeil DJ, Strack AM, MacIntyre DE, Van der Ploeg LHT and Qian S 2002. Melanin-concentrating hormone 1 receptor-deficient mice are lean, hyperactive, and hyperphagic and have altered metabolism. Proc Natl Acad Sci USA 99, 3240-3245.

Matuszewich L, Lorrain DS and Hull EM 2000. Dopamine release in the medial preoptic area of female rats in response to hormonal manipulation and sexual activity. Behav Neurosci 114, 772-782.

McCarthy MM, McDonald CH, Brooks PJ and Goldman D 1997. An anxiolytic action of oxytocin is enhanced by estrogen in the mouse. Physiol Behav 60, 1209-1215.

Meuwissen THE, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819-1829.

Millington G 2007. The role of proopiomelanocortin (POMC) neurones in feeding behaviour. Nutr Metab (Lond) 4, 18.

Mitra R, Ferguson D and Sapolsky RM 2009. SK2 potassium channel overexpression in basolateral amygdala reduces anxiety, stress-induced corticosterone secretion and dendritic arborization. Mol Psychiatry 14, 847-855.

Mong J, Easton A, Kow L-M and Pfaff D 2003. Neural, hormonal and genetic mechanisms for the activation of brain and behavior. Eur J Pharmacol 480, 229-231.

Mong JA and Pfaff DW 2004. Hormonal symphony: steroid orchestration of gene modules for sociosexual behaviors. Mol Psychiatry 9, 550-556.

Montkowski A, Jahn H, Ströhle A, Poettig M, Holsboer F and Wiedemann K 1998. C-type natriuretic peptide exerts effects opposing those of atrial natriuretic peptide on anxiety-related behaviour in rats. Brain Res 792, 358-360.

Nelson PT, Baldwin DA, Kloosterman WP, Kauppinen S, Plasterk RHA and Mourelatos Z 2006. RAKE and LNA-ISH reveal microRNA expression and localization in archival human brain. RNA 12, 187-191.

Patisaul HB, Scordalakes EM, Young LJ and Rissman EF 2003. Oxytocin, but not oxytocin receptor, is regulated by oestrogen receptor β in the female mouse hypothalamus. J Neuroendocrinol 15, 787-793.

Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu T-M, Bao W, Fang H, Kawasaki ES, Hager J, Tikhonova IR, Walker SJ, Zhang L, Hurban P, de Longueville F, Fuscoe JC, Tong W, Shi L and Wolfinger RD 2006. Performance comparison of one-color and two-color platforms within the microarray quality control (MAQC) project. Nat Biotech 24, 1140-1150.

Pfaff D 2005. Hormone-driven mechanisms in the central nervous system facilitate the analysis of mammalian behaviours. J Endocrinol 184, 447-453.

Pfaff D, Waters E, Khan Q, Zhang X and Numan M 2011. Minireview: Estrogen receptor-initiated mechanisms causal to mammalian reproductive behaviors. Endocrinology 152, 1209-1217.

Pfaff DW, Kow LM, Loose MD and Flanagan-Cato LM 2008. Reverse engineering the lordosis behavior circuit. Horm Behav 54, 347-354.

Rege TA and Hagood JS 2006. Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis. FASEB J 20, 1045-1054.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP and Young RA 2000. Genome-wide location and function of DNA binding proteins. Science 290, 2306-2309.

Rex A, Marsden CA and Fink H 1997. Cortical 5-HT-CCK interactions and anxiety-related behaviour of guinea-pigs: a microdialysis study. Neurosci Lett 228, 79-82.

Richards N, Wayman C and Allers KA 2010. Neuronal activity in the hypothalamic paraventricular nucleus varies across the estrous cycle in anesthetized female rats: effects of dopamine receptor agonism. J Sex Med 7, 1104-1115.

Richmond G and Clemens L 1988. Ventromedial hypothalamic lesions and cholinergic control of female sexual behavior. Physiol Behav 42, 179-182.

Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M and Jones S 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Meth 4, 651-657.

Rougon G and Hobert O 2003. New insights into the diversity and function of neuronal immunoglobulin superfamily molecules. Annu Rev Neurosci 26, 207-238.

Ryan DP, Galvin JA and O'Farrell KJ 1999. Comparison of oestrous synchronization regimens for lactating dairy cows. Anim Reprod Sci 56, 153-168.

Saba R and Schratt GM 2010. MicroRNAs in neuronal development, function and dysfunction. Brain Res 1338, 3-13.

Schratt G 2009. MicroRNAs at the synapse. Nat Rev Neurosci 10, 842-849.

Segal JP, Stallings NR, Lee CE, Zhao L, Socci N, Viale A, Harris TM, Soares MB, Childs G, Elmquist JK, Parker KL and Friedman JM 2005. Use of laser-capture microdissection for the identification of marker genes for the ventromedial hypothalamic nucleus. J Neurosci 25, 4181-4188.

Sempere L, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E and Ambros V 2004. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol 5, R13.

Siegel G, Saba R and Schratt G 2011. microRNAs in neurons: manifold regulatory roles at the synapse. Curr Opin Genet Dev 21, 491-497.

Sokolov BP 2007. Oligodendroglial abnormalities in schizophrenia, mood disorders and substance abuse. Comorbidity, shared traits, or molecular phenocopies? Int J Neuropsychopharmacol 10, 547-555.

Sousa JC, Grandela C, Fernández-Ruiz J, Miguel Rd, Sousa Ld, Magalhães AI, Saraiva MJ, Sousa N and Palha JA 2004. Transthyretin is involved in depression-like behaviour and exploratory activity. J Neurochem 88, 1052-1058.

Takahashi K, Yamada M, Ohata H, Honda K and Yamada M 2005. Ndrg2 promotes neurite outgrowth of NGF-differentiated PC12 cells. Neurosci Lett 388, 157-162.

Terry-Lorenzo RT, Inoue M, Connor JH, Haystead TAJ, Armbruster BN, Gupta RP, Oliver CJ and Shenolikar S 2000. Neurofilament-L is a protein phosphatase-1-binding protein associated with neuronal plasma membrane and post-synaptic density. J Biol Chem 275, 2439-2446.

Uhart M, McCaul ME, Oswald LM, Choi L and Wand GS 2004. GABRA6 gene polymorphism and an attenuated stress response. Mol Psychiatry 9, 998-1006.

Uphouse L 2000. Female gonadal hormones, serotonin, and sexual receptivity. Brain Res Rev 33, 242-257.

Vasudevan N and Pfaff DW 2008. Non-genomic actions of estrogens and their interaction with genomic actions in the brain. Front Neuroendocrinol 29, 238-257.

Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ and van Kampen AHC 2003. The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. Genome Res 13, 1998-2004.

Vinogradov AE 2004. Compactness of human housekeeping genes: Selection for economy or genomic design? Trends Genet 20, 248-253.

Visvanathan J, Lee S, Lee B, Lee JW and Lee S-K 2007. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. Genes Dev 21, 744-749.

Wayman GA, Davare M, Ando H, Fortin D, Varlamova O, Cheng H-YM, Marks D, Obrietan K, Soderling TR, Goodman RH and Impey S 2008. An activity-regulated microRNA controls dendritic plasticity by down-regulating p250GAP. Proc Natl Acad Sci USA 105, 9093-9098.

Zhang B and Horvath S 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4, Article 17.

Zhao J, Quyyumi AA, Patel R, Zafari AM, Veledar E, Onufrak S, Shallenberger LH, Jones L and Vaccarino V 2009. Sex-specific association of depression and a haplotype in leukotriene a4 hydrolase gene. Psychosom Med 71, 691-696.

Summary

Concurrent to the impressive improvement achieved over the last few decades for the trait of milk production in dairy cows was a steady decline in several fertility traits. Oestrous behaviour (OB), a key fertility trait in dairy cows, has been declining both in intensity and duration, thereby making it difficult for farmers to identify the appropriate time for artificial insemination of their cows. An understanding of the genomic regulation of OB will present new opportunities for managing this trait to help improve fertility in dairy cows.

In Chapter 1, I begin with an introduction into the problems with fertility in high producing cows and the current state of knowledge of the physiological and genomic mechanisms regulating OB in female mammals. Although considerable work has been done to understand the genomic mechanisms regulating OB using model animals mainly rodents, little or no research has been conducted to study the same in dairy cows. The research described in this thesis therefore aimed to improve the understanding of genomic regulation of OB in dairy cows by studying gene expression in the bovine anterior pituitary (AP) and four brain areas (amygdala, hippocampus, dorsal hypothalamus and ventral hypothalamus) that are involved in regulating OB.

Towards this aim, a large scale microarray experiment was conducted followed by a series of different analyses that included model based association of gene expression with OB scores, gene co-expression and differential expression. In brief, OB was recorded in 28 healthy primiparous Holstein Friesian cows for at least two consecutive oestrous cycles, following which, 14 were sacrificed at the start of their oestrous cycle (day0) and 14 around mid-cycle (day12). Tissue samples were collected from the AP and four brain areas that are involved in OB regulation. A total of 280 microarrays (i.e. 28 cows x 5 tissues samples x 2 dye swaps) were prepared in a common reference design to generate the gene expression data. The data pre-processing and analysis was done using the LIMMA package within Bioconductor project of R statistical programming language. Gene expression levels expressed as M-values were thus obtained for each tissue per cow. This data was used for all subsequent analyses.

In Chapter 2, we identified and studied genes that associated with OB among genes expressed in the bovine AP either at the start of oestrous cycle or at the mid-cycle, or regardless of the phase of cycle. Gene expression levels were modelled as a function of the orthogonally transformed average OB scores per cow using a Bayesian hierarchical mixed model on data from day 0 cows alone (analysis 1), day 12 cows alone (analysis 2) and the combined data from day 0 and day 12 cows (analysis 3). Genes whose expression patterns showed significant linear or non-linear relationships with OB scores were identified in all three analyses. Gene

ontology terms enriched among genes identified in analysis 1 revealed processes associated with expression of OB (e.g. 'behaviour', 'secretion', and 'nervous system development') whereas the terms enriched among genes identified in analyses 2 and 3 were general processes (e.g. 'multicellular organismal process') which may facilitate proper expression of OB at the subsequent oestrus.

Similar to the analysis of the gene expression data of the bovine AP described in Chapter 2, the four bovine brain areas were analysed and their results reported in Chapter 3. Some additional steps to refine the analyses of the brain areas included limiting further analysis to genes whose expression patterns showed significant linear or quadratic relationships but not third order relationships and to consider only those genes as significant that showed up in 5 repeated runs of the Gibbs sampling based algorithm. The identified genes included those that are known to be involved in processes associated with OB like socio-sexual behaviour (e.g. *OXT, AVP, GABRA6, HTR2A, DRD2*), anxiety, stress and feeding motivation (e.g. *POMC*, *MCHR1, TTR*) as well as nervous system related processes (e.g. *CHRM1*, *CHRM3, CHRNA5, CTLA4, IL1RL1, MARCO*), suggesting a link between neuronal plasticity and OB. We also found genes whose association with OB is novel and these warrant further investigation. Further, as in Chapter 2, here too several genes expressed in mid-cycle were found associated with OB scores suggesting that they may have a function in the proper expression of OB at the next oestrus.

Next, in Chapter 4, we identified genes and biological processes shared among the bovine AP and four brain areas that act together to regulate OB. Here, gene co-expression networks were defined using the method called 'weighted gene co-expression network analysis' (WGCNA). Consensus modules of co-expressed genes shared between the networks were found for several tissue pairs. Modules whose representative expression profile correlated with the level of OB exhibited by the individual cows were then identified. OB correlated modules were found enriched for gene ontology terms like glial cell development and regulation of neural projection development as well as for KEGG pathway terms related to brain degenerative diseases. Biological terms found common to several OB correlated consensus modules included general cellular processes like oxidative phosphorylation, ribosome and biosynthetic processes, indicating increased transcription and protein synthesis. Stimulation of ribosomal RNA synthesis is known from rodent studies to be a primary event in the activation of neuronal cells and pathways involved in female reproductive behaviour and this precedes the oestrogen driven expansion of dendrites and synapses. Similar processes therefore operate in cows too to affect OB. Hub genes within the OB correlated modules (e.g. *NEFL*, *NDRG2*, *GAP43*, *THY1*, *TCF7L2 etc.*) are strong candidates among genes

regulating OB expression. Thus, new insights were obtained into genes and biological processes shared among the bovine AP and brain areas acting together to regulate OB.

In Chapter 5, we showed the phenomenon of chromosomal regional regulation of transcription to exist in the bovine genome. Housekeeping genes in the brain were found over-represented within chromosomal regions of increased gene expression (RIDGEs) but tissue-specific genes were not. Housekeeping genes and genes within RIDGEs had, in general, higher expression levels and GC content but shorter gene lengths and intron lengths than tissue-specific genes and genes within chromosomal regions of decreased gene expression (anti-RIDGES). The genomic features observed for genes within RIDGEs and housekeeping genes in bovines agree with previous studies in several other species further strengthening the hypothesis of selective pressure to keep the highly and widely expressed genes short and compact for transcriptional efficiency. Further, positively selected genes were found non-randomly distributed on the genome with a preference for RIDGEs and regions of intermediate gene expression compared to anti-RIDGEs.

Finally in Chapter 6, I discuss the main findings of the previous chapters to arrive at a more complete picture of the genomic regulation of OB in dairy cows. The GAPPS modules that account for the causal relations between sex hormones and female mating behaviours were used as a framework to discuss the OB associated genes and processes in dairy cows (summary table provided). Genes differentially expressed in the AP between day0 and day12 are then discussed in the context of OB related processes. This is followed by a brief discussion on regulation of gene expression. The advantages and limitations of the analytical methods used in this thesis are discussed next followed by some points that would be useful to consider in future studies on genomic regulation of OB. I end this chapter with a discussion on the steps needed to apply the new knowledge in practice to help optimize fertility in dairy cows.

Samenvatting

Gelijktijdig met de indrukwekkende verbetering die in de afgelopen decennia voor de melkproductie bij melkkoeien is bereikt was er een gestage daling van de vruchtbaarheid. Oestrus (Tochtigheid) gedrag (OG), een belangrijke eigenschap vruchtbaarheid van melkkoeien, is zowel in intensiteit als in duur afgenomen waardoor het voor de boeren moeilijk is om het juiste inseminatie moment van hun koeien te bepalen. Een goed begrip van de genomische regulering van OG zal nieuwe mogelijkheden bieden voor het beheer van deze eigenschap om te helpen de vruchtbaarheid van melkkoeien te verbeteren.

In hoofdstuk 1, geef ik een introductie van de problemen met vruchtbaarheid van hoogproductieve koeien en de huidige stand van kennis over de fysiologische en genetische mechanismen van de regulatie van OG bij vrouwelijke zoogdieren. Er is reeds veel werk gedaan aan de genomische regulatie mechanismen van OG in model dieren, vooral knaagdieren, maar er is weinig of geen onderzoek verricht bij melkkoeien. Het onderzoek beschreven in dit proefschrift is dan ook gericht op het begrijpen van de genetische regulatie van OG bij melkkoeien door het bestuderen van gen-expressie in de runder-hypofysevoorkwab (AP) en vier hersengebieden (amygdala, hippocampus, hypothalamus dorsale en ventrale hypothalamus) die betrokken zijn bij het reguleren van OG.

Hiervoor werd een groot microarray experiment uitgevoerd gevolgd door een reeks van verschillende analyses. De associatie van genexpressie met OG scores, gen co-expressie en differentiële expressie werden bestudeerd. In het kort werd OG bepaald bij 28 gezonde primipara Holstein Friese koeien voor ten minste twee opeenvolgende oestruscycli, waarna, 14 koeien werden geofferd aan het begin van hun oestrogeencyclus (dag0) en 14 koeien rond mid-cyclus (dag12). Weefsel monsters werden verzameld uit de AP en vier hersengebieden die betrokken zijn bij OG regulatie. Een totaal van 280 microarrays (dat wil zeggen 28 koeien x 5 weefsels samples x 2 duplo experimenten met omkering van de kleur) gehybridiseerd volgens de common-referentie methode om de genexpressie gegevens te genereren. De gegevens pre-processing en analyse werd gedaan met behulp van het Limma softwarepakket binnen het Bioconductor project in de R-statistische programmeertaal. Genexpressie niveaus uitgedrukt als M-waarden werden aldus verkregen voor elk weefsel per koe. Deze gegevens werden gebruikt voor alle volgende analyses.

In hoofdstuk 2 hebben we genen geïdentificeerd en bestudeerd van de genen die tot expressie komen in de koeien AP. De genexpressieniveaus werden geassocieerd met de OG, hetzij bij het begin van oestrogeencyclus of in het midden van de cyclus, of onafhankelijk van de fase van de cyclus. Genexpressie niveaus werden gemodelleerd als een functie van het orthogonaal getransformeerde gemiddelde

van de OG scores per koe met behulp van een Bayesiaanse hiërarchisch gemengd model van de gegevens van dag 0 koeien alleen (analyse 1), dag 12 koeien alleen (analyse 2) en de gecombineerde gegevens van dag 0 en dag 12 koeien (analyse 3). Genen waarvan de expressie patronen een significante lineaire of niet-lineaire relaties met OG scores vertoonden werden geïdentificeerd in alle drie de analyses. Gen ontologie termen van deze genen werden geïdentificeerd. Belangrijke groepen waren bijv. 'gedrag', 'afscheiding' en 'ontwikkeling van het zenuwstelsel'. In analyses 2 en 3 zijn juist veel algemene processen gevonden (bijvoorbeeld 'meercellige organisme proces'), die belangrijk kunnen zijn voor de volgende oestrus.

De vier runderen hersengebieden werden geanalyseerd en de resultaten gerapporteerd in hoofdstuk 3. Een aantal extra stappen om de analyses van de hersengebieden te verfijnen werden genomen, waaronder het beperken van verdere analyse van genen tot de expressie patronen die significant lineair of kwadratisch relaties vertonen, maar niet de derde orde relaties. Verder werden alleen die genen die in 5 herhaalde runs waargenomen werden op basis van een Gibbs sampling algoritme. Van de geïdentificeerde genen is bekend is dat ze betrokken zijn bij de processen die samenhangen met OG als sociaal-seksueel gedrag (bijv. *OXT, AVP, GABRA6, HTR2A, DRD2*), angst, stress en voeding motivatie (bijv. *POMC, MCHR1, TTR*) en als zenuwstelsel gerelateerde processen (bijv. *CHRM1, CHRM3, CHRNA5, CTLA4, IL1RL1, MARCO*). Dit suggereert een verband tussen neuronale plasticiteit en OG. We vonden ook genen waarvan de connectie met OG nieuw is en deze werden nader onderzocht. Verder werden net als in hoofdstuk 2, ook hier een aantal genen gevonden die tot expressie komen in het midden van de cyclus gevonden en associëren met OG scores wat suggereert dat zij een functie hebben bij het reguleren van de OG bij de volgende oestrus.

Vervolgens hebben we in hoofdstuk 4 de genen en biologische processen geïdentificeerd die steeds gelijktijdig in twee weefsels tot expressie komen, namelijk in de AP en vier hersengebieden die samenwerken om de OG te reguleren. Hier werden gen co-expressie netwerken volgens de methode 'gewogen gen co-expressie-netwerk analyse' (WGCNA) geïdentificeerd. Consensus modules van genen die in steeds 2 weefsels tot expressie komen werden gevonden voor verschillende weefsel-paren. Modules waarvan de “eigengenes” expressie profielen correleerde met het niveau van de OG van de individuele koeien werden geïdentificeerd. OG gecorreleerde modules werden verrijkt voor ontologie termen als gliale cel ontwikkeling en regulatie van neurale projectie ontwikkeling, alsmede voor de KEGG pathway termen die gerelateerd zijn aan de hersenen degeneratieve ziekten. Biologische termen voor verschillende OG gecorreleerde consensus

modules voor algemene cellulaire processen zoals oxidatieve fosforylering, ribosoom en biosynthetische processen werden gevonden, wat aangeeft dat er verhoogde transcriptie en eiwitsynthese plaat vindt. Stimulatie van ribosomaal RNA synthese is bekend van onderzoeken bij knaagdieren als een primaire gebeurtenis in de activatie van neuronale cellen betrokken bij vrouwelijk reproductief gedrag en dat vooraf gaat aan de oestrogeen-gedreven groei van dendrieten en synapsen. Soortgelijke processen in koeien zijn dus aangetoond die ook naar OG beïnvloeden. Hub genen binnen de OG gecorreleerd modules (bijv. *NEFL, NDRG2, GAP43, THY1, TCF7L2* etc.) zijn sterke kandidaat genen voor de regulatie van OG expressie. Zo werden nieuwe inzichten verkregen in de genen en biologische processen die plaatsvinden in de runderen AP en hersengebieden welke samenwerken om het OG te reguleren.

In hoofdstuk 5 hebben we het fenomeen van chromosomale regionale regulatie van transcriptie in de rundergenoom aangetoond. Housekeeping genen in de hersenen werden oververtegenwoordigd gevonden binnen de chromosomale regio's met verhoogde genexpressie (Ridges), maar weefselspecifieke genen niet. Housekeeping genen en genen in Ridges hadden, in het algemeen, een hogere expressie en hogere GC inhoud, maar kortere gen lengtes en intron lengtes dan de weefsel-specifieke genen en genen in chromosomale regio's van verminderde genexpressie (anti-Ridges). De genomische eigenschappen waargenomen voor genen binnen de Ridges en housekeeping genen in runderen zijn samen met eerdere studies in verschillende andere soorten een verdere versterking van de hypothese van selectieve druk voor de genen die hoog tot expressie komen om kort en compact te zijn t.b.v. transcriptionele efficiency. Verder werd geconstateerd dat positief geselecteerde genen niet-willekeurig verdeeld zijn over het genoom maar een voorkeur voor Ridges en regio's van intermediaire genexpressie hebben.

Tenslotte bespreek ik in hoofdstuk 6 de belangrijkste bevindingen van de voorgaande hoofdstukken om te komen tot een meer compleet beeld van de genomische regulatie van OG bij melkkoeien. De GAPPS modules die de causale relaties tussen geslachtshormonen en vrouwelijke paargedrag beschrijven werden gebruikt als een kader voor de OG geassocieerde genen en processen bij melkkoeien (inclusief samenvattende tabel). Genen die differentieel tot expressie komen in de AP tussen dag0 en dag12 worden vervolgens besproken in de context van de OG gerelateerde processen. Dit wordt gevolgd door een korte discussie over de regulatie van genexpressie. De voordelen en beperkingen van de analytische methoden die in dit proefschrift werden gebruikt worden besproken en gevolgd door een aantal punten die nuttig zouden zijn om te overwegen in toekomstige

studies van de genomische regulatie van OG. Ik eindig dit hoofdstuk met een bespreking van de stappen die nodig zijn om de nieuwe kennis in de praktijk te brengen t.b.v. het optimaliseren van de vruchtbaarheid bij melkkoeien

Acknowledgements

It was in the year 2004 that I first put my plan to shift career from a veterinarian to bioinformatics researcher into action. By keeping to animal genetics, the shift wasn't as dramatic as I imagined. My term at Wageningen University helped make this shift as smooth as possible. More than the place and the facilities, it was the people that made the greatest positive impact. I would like to express my profound gratitude to all those wonderful people who I had the opportunity to interact with and who stood by me over the last few challenging and very exciting years.

First of all, I would like to thank my PhD supervisors: Mari Smits, Marinus te Pas, Roel Veerkamp and Martien Groenen. They have all been a great source of inspiration and support throughout my PhD term. I greatly appreciate that they made time for me whenever I needed their help, even at short notice. By allowing me to work independently and at my own pace, Mari and Marinus helped me think about my research topic and define my own path in the PhD project. The confidence that Marinus placed on me and his warm and inspiring words are greatly appreciated. Roel, the good manager that he is, assessed me well quite early on and exerted just the right level of pressure to keep me on track and within the time schedule. I was amazed by his sharp reviews and the different view point he saw my work from, thus helping me improve my work. Equally amazing was the way in which Mari and Martien would suggest changes to my writing style that made it much more effective in publications.

Next, I would like to thank my colleagues at the Animal Breeding and Genomics Centre, Wageningen University Livestock Research who in general made life at work very warm and cordial. I am especially thankful to Han, Agnes, Henri, Bonne and Mario for professional help and scientific discussions that helped improve my publications; to Jack and Henk for being so friendly and helpful throughout; to Krista, Myrthe, Marike, Ina and Amir with whom I shared my office room at different points of time and who have been great company and to Dirk-Jan for all the above reasons. I really appreciate all those who took efforts to include me in their conversations, switching from Dutch to English at coffee or lunch breaks. Further I thank the great contacts outside the group, especially Alex and Freddy, who I hope to continue interacting with at my new job too.

I am very grateful to Haisheng who has not only been a great friend but also collaborated with me at work which even resulted in a nice publication. I thank Martin for his tips during the thesis preparation and formatting. I also thank Tette van der Lende, with whom I had several discussions about my work and publications.

I would like to express my sincere gratitude to my supervisors in my Master's theses, Hinri Kerstens and Pieter Neerincx, who initiated me into the wonderful world of programming within bioinformatics research.

I now thank my friends from Wageningen University who I have known from the time of my Master's degree to the more recent ones. I would like to especially mention Kishore who have been my best buddy here throughout as well as Shibu, Shahanas, Byjesh, Nagesh, Feana, Hema, Abhishek, Srinivas and Sachin. Also among my long-time friends here is Audrey, who along with Kishore, has agreed to be my paranymph at my thesis defence. I sincerely thank them for this. I greatly appreciate the company of my friends in and around our residence, especially Shanoj, Geetha, Rakshit, Monika, Hari, Bandhavi, Kiran, Bismi, Lisa, Brandt, Binu and Ferozan. They have been great company and great help to us at times of need.

I sincerely hope to continue my professional and personal relationship in future too with all the great people I met at Wageningen over the past few years, and I take this opportunity to thank them all again and also those whose names I may have missed mentioning above.

I now take this opportunity to thank my mentors, Prof. Dr. K.C. Raghavan and Dr. Vahida A.M. who along with my parents inspired me to take the decision to opt for higher studies.

With this, I now thank the most important people in my life, my parents and family. It is the limitless love, support and encouragement from my father, Viswanathan Kommadath and mother, Sasikala Velarambath throughout my life that made me what I am today. I am extremely grateful to them for this and for all the sacrifices they made for me. I thank my sister, Amitha, and brother, Akhil, for all the great times we shared. In 2007, I met Esha with whom I have been married for 4 years now. I thank her for her love and understanding and for being a wonderful partner. In 2010, our son, Agnith arrived and lighted up our world even more. One look at him would take away all the stresses of the day. I thank him for that and for the time I borrowed from his play time the last few months to complete my PhD work and thesis. I thank my parents and parents-in-law, Prakash and Prameela, for visiting us in the Netherlands and helping with caring for Agnith and for having good Indian food ready at the dining table when we got back from work. I also thank my brother-in-law, Kapil for helping with designing the cover page of this thesis. I thank the rest of my family too (grandmother, uncles, aunts, cousins and a cute nephew and niece) for their love and support.

Acknowledgements

Finally I thank God for the wonderful life that we have and for seeing us through the tough times and helping us make the right choices.

Arun Kommadath

January 2012, Wageningen

Curriculum vitae

About the author

Arun Kommadath was born in Kozhikode district of Kerala state in India on 5th June 1974. At the age of three, he and his family moved to Kuwait where he did most of his schooling. In September 1990, he along with his family returned to India and he completed his high school at Thrissur, Kerala, India. In July 1998, he graduated in Veterinary Sciences from the College of Veterinary and Animal Sciences, Mannuthy, Thrissur, Kerala, India. Two years later, he completed his M.Sc. in Animal Genetics and Breeding from the National Dairy Research Institute in Karnal, Haryana, India. After serving for four years in India, first as a Research Associate in the College of Veterinary and Animal Sciences, Mannuthy for a year and then as a Government Veterinary Surgeon at Palakkad, he decided to follow his interests in bioinformatics research. Between 2004 and 2006, he worked on his M.Sc. in Bioinformatics at the Wageningen University, The Netherlands, during the course of which he completed a major thesis in the group of Prof. Dr. Martien A.M. Groenen and a minor thesis in the group of Prof. Dr. Jack A.M. Leunissen. In July 2007, he started his PhD programme at the Animal Breeding and Genomics Centre, Wageningen University Livestock Research, Lelystad, which resulted in this thesis. From January 2012, he is working as a postdoctoral fellow at the Stothard Research Group, University of Alberta, Edmonton, Canada.

List of publications

Peer-reviewed publications

Kommadath A, Woelders H, Beerda B, Mulder HA, de Wit AAC, Veerkamp RF, te Pas MFW, Smits MA: Gene expression patterns in four brain areas associate with quantitative measure of estrous behavior in dairy cows. BMC Genomics 2011, 12(1):200.

Kommadath, A, Nie, H, Groenen MAM, te Pas MFW, Veerkamp RF, Smits MA: Regional Regulation of Transcription in the Bovine Genome. PLoS ONE 2011, 6(6): e20413.

Kommadath A, Mulder HA, de Wit AAC, Woelders H, Smits MA, Beerda B, Veerkamp RF, Frijters ACJ, te Pas MFW: Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows. Animal 2010, 4(08):1297-1307.

Kerstens HHD, Kollers S, Kommadath A, del Rosario M, Dibbits B, Kinders SM, Crooijmans RP, Groenen MAM: Mining for single nucleotide polymorphisms in pig genome sequence data. BMC Genomics 2009, 10(1):4.

Hedegaard J, Arce C, Bicciato S, Bonnet A, Buitenhuis B, Collado-Romero M, Conley LN, SanCristobal M, Ferrari F, Garrido JJ, Groenen MAM, Hornshoj H, Hulsegge I, Jiang L, Jimenez-Marin A, Kommadath A, Lagarrigue S, Leunissen JAM, Liaubet L, Neerincx PBT, Nie H, Poel J, Prickett D, Ramirez-Boo M, Rebel JMJ, Robert-Granie C, Skarman A, Smits MA, Sorensen P, Tosser-Klopp G, Watson M: Methods for interpreting lists of affected genes obtained in a DNA microarray experiment. BMC Proceedings 2009, 3(Suppl 4):S5.

Hulsegge I, Kommadath A, Smits MA: Globaltest and GOEAST: two different approaches for Gene Ontology analysis. BMC Proceedings 2009, 3(Suppl 4):S10.

Kommadath, A, Sharma A, Jakhar KK: Hepatotoxic, nephrotoxic and genotoxic effects in mice fed urea adulterated milk. Indian J. Dairy Sci. 2001, 54(6): 316-321

Publications in preparation

Kommadath, A, te Pas MFW, Smits MA: Gene co-expression network analysis identifies candidate genes and biological processes affecting estrous behavior in dairy cows.

Training and Supervision Plan

The Basic Package (3 ECTS)	
WIAS Introduction Course	2008
WGS Course: Ethics and Philosophy in Animal Sciences	2008
Scientific Exposure (16 ECTS)	
International conferences	
4th EADGENE conference: Animal Disease Genomics: Opportunities and Applications, Edinburgh, UK	2008
3rd SABRE Conference: Welfare and Quality Genomics, Foulum, Denmark	2008
60th Annual Meeting of the European Federation of Animal Science (Integrated EAAP-SABRE conference), Barcelona, Spain	2009
32nd Conference for the International Society for Animal Genetics (Integrated ISAG-SABRE conference), Edinburgh, UK	2010
11th International Conference on Systems Biology (ICSB), Edinburgh, UK	2010
Seminars and workshops	
4th Integrative Bioinformatics Workshop, Ghent, Belgium	2007
5th Annual Cytoscape Public Symposium, Amsterdam, NL	2007
WIAS science day 2008, Wageningen, NL	2008
EADGENE and SABRE Post-analyses Workshop, Lelystad, NL	2008
4th SABRE Consortium meeting, Schiphol, NL	2009
WIAS science day 2009, Wageningen, NL	2009
Systems Biology Day, Wageningen, NL	2009
WIAS science day 2010, Wageningen, NL	2010
Presentations (type of presentation - Oral/Poster)	
"GOEAST and Globaltest: Two different methods for Gene Ontology analysis", EADGENE and SABRE Post-analyses Workshop, Lelystad (Oral)	2008
"Bovine brain transcriptome analysis for gene pathways regulating oestrous behaviour in cattle", 3rd SABRE Conference: Welfare and Quality Genomics, Foulum (Poster)	2008
"Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows", 4th SABRE Consortium meeting, Schiphol (Oral)	2009
"Gene expression patterns in anterior pituitary associated with quantitative measure of oestrous behaviour in dairy cows", 60th Annual Meeting of the European Federation of Animal Science (Integrated EAAP-SABRE conference), Barcelona (Oral)	2009

"Genes expressed in bovine pituitary associated with oestrous behaviour regulation", WIAS science day 2009, Wageningen (Poster)	2009
"Gene expression patterns in four brain areas associate with quantitative measure of oestrous behaviour in dairy cows" - WIAS science day 2010, Wageningen (Oral)	2010
"Gene expression patterns in four brain areas associate with quantitative measure of estrous behavior in dairy cows" - 32nd Conference for the International Society for Animal Genetics (Integrated ISAG-SABRE conference), Edinburgh (Oral)	2010
"Bovine Brain Transcriptome Analysis for Genomic Features and Regional Regulation of Transcription" - 11th International Conference on Systems Biology (ICSB), Edinburgh (Poster)	2010

In-Depth Studies (9 ECTS)	
WIAS Course: Introduction to R for statistical analysis	2008
ARK Genomics Course: Microarray data analysis and meta-analysis	2008
WIAS Course: Statistics for the Life Sciences	2008
Advanced visualisation, integration and biological interpretation of -omics data	2009
Next generation sequencing: Technologies, Applications and Data Analysis	2011
Next generation sequencing: Data Analysis Pipelines	2011
NBIC course: Comparative Genomics: From Evolution to Function	2011

Professional Skills Support Courses (4 ECTS)	
WGS Course: Personal Efficacy	2007
WGS Course: Techniques for Writing and Presenting Scientific Papers	2009
WGS Course: Interpersonal Communication for PhD students	2010
WGS Course: Career Orientation	2011
Research Skills Training (6 ECTS)	
Preparing own PhD research proposal	2007
Didactic Skills Training (2 ECTS)	
Supervising practicals for MSc Course Genomics (ABG-30306)	2010

Education and Training Total: 39 ECTS

- One credit equals a study load of approximately 28 hours.

Colophon

Colophon

This research described in this thesis was co-financed by the European Commission funded projects SABRE (www.sabre-eu.eu/) (contract No. FOOD-CT-2006-016250) and EADGENE (www.eadgene.info/) (contract No. FOOD-CT-2004-506416), the Dutch Ministry of Agriculture, Nature and Food Quality (www.minlnv.nl) (KB-04-004-012) and the International cattle breeding co-operative - CRV BV (www.crv4all.com/). The text represents the authors' views and does not necessarily represent a position of the Commission who will not be liable for the use made of such information.

Financial support from the Animal Breeding and Genomics Centre, Wageningen University, The Netherlands for printing this thesis is gratefully acknowledged.

The cover design was conceived by Arun Kommadath and implemented by Kapil KT, Web designer, Kozhikode, Kerala, India.

The thesis was printed by GVO drukkers & vormgevers B.V. | Ponsen & Looijen, Ede, The Netherlands.