TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Tierzucht

Genomic Characterisation and Polymorphism Analysis of Candidate Genes for Milk Production Traits and Association Studies in Three Cattle Breeds

Franz Reinhold Seefried

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende: Univ.-Prof. A. Schnieke, Ph.D. (Univ. of Edinburgh/UK)
Prüfer der Dissertation:

1. Univ.-Prof. Dr. Dr. H.-R. Fries
2. Univ.-Prof. Dr. G. A. Thaller (Christian-Albrechts-Universität zu Kiel)

Die Dissertation wurde am 10.12.2007 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum für Ernährung, Landnutzung und Umwelt am 11.06.2008 angenommen

List of figures

List of tables

Abbreviations

ABCG2	ATP-binding cassette sub-family G member 2 gene
ABCG2	ATP-binding cassette sub-family G member 2 protein
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bp	Base pair
BLAST	Basic local alignment search tool
BLUP	Best Linear Unbiased Prediction
BLUP-EBV	Breeding value estimated by Best Linear Unbiased Prediction
BTA3	Bovine chromosome 3
BTA6	Bovine chromosome 6
BTA8	Bovine chromosome 8
BV	German Brown
cDNA	coding DNA
CENTD1	Centaurin delta 1 gene
cM	centi Morgan
CSN1S1	Casein alpha s1 gene
CSN1S1	Casein alpha s1 protein
CSN1S2	Casein alpha s2 gene
CSN2	Casein beta gene
CSN3	Casein kappa gene
DD	Daughter design
DGAT1	Diacylglycerol O-acyltransferase 1 gene
DMSO	Dimethyl sulfid
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotides
DTT	Dithiothreitol
DYD	Daughter yield deviation
EST	Expressed sequence tag
FBN	Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere, 18196 Dummerstorf, Germany
FC1	Fat content lactation 1
FC2	Fat content lactation 2
FC3	Fat content lactation 3
FDR	False discovery rate
FV	German Fleckvieh
FY1	Fat yield lactation 1
FY2	Fat yield lactation 2
FY3	Fat yield lactation 3
GGA	Gallus gallus
GDD	Grand daughter design
GDF8	Growth differentiation factor 8 gene
HA	Haplotype analysis
HF	German Holstein
HSA4	Human chromosome 4
HWE	Hardy-Weinberg-Equilibrium
INDEL	Insertion-Deletion polymorphism
kb	Kilo base pairs
KLF3	Kruppel-like factor 3 gene

KLHL5	Kelch-like 5 gene
LD	Linkage disequilibrium
MA-BLUP	Marker Assisted Best Linear Unbiased Prediction
MAF	Minor allele frequency
MALDI-TOF MS	Matrix Assisted Laser Desorption Ionisation Time of Flight Mass Spectrometry
MAS	Marker assisted selection
Mb	Mega base pairs
MGS	Maternal grand sire
MGGS	Maternal grand grand sire
MMA	Multi marker analysis
MMU	Mus musculus
miRNA	micro-RNA
MW	Total net merit index for milk production in German Brown and German Fleckvieh
MY1	Milk yield lactation 1
MY2	Milk yield lactation 2
MY3	Milk yield lactation 3
NCBI	National Center of Biotechnology Information
OPN	Osteopontin gene
OPN	Osteopontin protein
PC1	Protein content lactation 1
PC2	Protein content lactation 2
PC3	Protein content lactation 3
PCDH7	Protocadherin 7 gene
PCR	Polymerase chain reaction
PGM1	Phosphoglucomutase 1 gene
PGM2	Phosphoglucomutase 2 gene
PGM3	Phosphoglucomutase 3 gene
PGM5	Phosphoglucomutase 5 gene
PPARGC1A	Peroxisome proliferator-activated receptor gamma coactivator 1-alpha gene
PRC	Peroxisome proliferator-activated receptor gamma, coactivator-related 1 gene
PS	Paternal sire
PTTG2	Pituitary tumor-transforming 2 gene
PY1	Protein yield lactation 1
PY2	Protein yield lactation 2
PY3	Protein yield lactation 3
QTL	Quantitative trait loci
QTN	Causative variant for a quantitative trait
RNA	Ribonucleic acid
RZM	Total net merit index for milk production in German Holstein
SAP	Shrimp Alkaline Phosphatase
sd	Standard deviation
SDS	Sodium dodecyl sulphate
SLC2A9	Facilitated glucose transporter member 9 gene
SNP	Single nucleotide polymorphism
SNP_id	In-house single nucleotide polymorphism identification code
SMA	Single marker analysis
TBE	Tris Borate EDTA buffer

TE	Tris EDTA buffer
UDP	Udenosin diphosphate
UTR	Untranslated region
VIT	Vereinigte Informationssysteme Tierhaltung w.V., 27283 Verden, Germany

1. Introduction and goals

In past decades, sophisticated statistical methods for breeding value estimation were established and enabled impressive genetic gain. However, when selection relies on statistical methods, genetic gain is limited in traits with low heritability, in traits with phenotypes evolved late in animals' life, or in traits that are not expressed in the gender of interest, e.g. lactation traits.

DNA is in fact available at any age and from both genders, so that efforts to decipher the molecular architecture of quantitative traits have been initiated in order to alleviate some of these limitations. Genome scans have identified markers that affect quantitative traits, and the bovine chromosome 6 (BTA6) turned out to harbour more quantitative trait loci (QTL) for milk traits than other chromosomes.

Milk synthesis is an excellent quantitative trait for molecular dissection. It is characterised by a great deal of phenotypic variation and the amount of existing phenotype data is immense and of high quality. For analysing the molecular architecture, milk synthesis has to be partitioned into the synthesis of its components, such as milk fat, milk protein and lactose. Milk components are physiologically correlated due to interconnected metabolic pathways and common precursors for the synthesis of different milk components, such as glucose, which is required for fat- and lactose synthesis. Apart from this, correlation among milk components relies on genetic sources and is caused by pleiotropic effects and temporarily also by linkage between loci.

Livestock species in particular are known for a high level of non-random association between different genetic loci (linkage disequilibrium), so that a molecular dissection of quantitative traits is complicated. Under the prevailing circumstances, it might not be opportune to analyse genes in isolation, so this thesis investigates six candidate genes together in an association study in three breeds.

The specific goals of this thesis were

1) Characterisation of candidate genes for milk synthesis on bovine chromosome 6 by sequence and structure analysis
2) Systematic screening for polymorphisms
3) Identification of associated and causative polymorphisms for milk synthesis by an association study in three different breeds.

2. Literature review

2.1. *Analysis of quantitative traits*

Milk production traits are quantitative traits showing a continuous distribution of phenotypic values. An infinitesimal number of genes with minute effects specify in their sum the continuous distribution of the heritable traits (Fisher 1930). Fisher's model had to be modified when studies of Quantitative Trait Loci (QTL) started (Geldermann 1976). QTLs are individual genes with quantifiable measurable effects and are traceable by DNA markers. These results are consistent with Robertson's theorem (Robertson 1969) and Orr's model (Orr 1998) that the distribution of allelic effects might be exponential and the quantitative variation is determined by few QTLs of large effect and many genes of small effect (see Figure 2.1) (HAYES and GODDARD 2001).

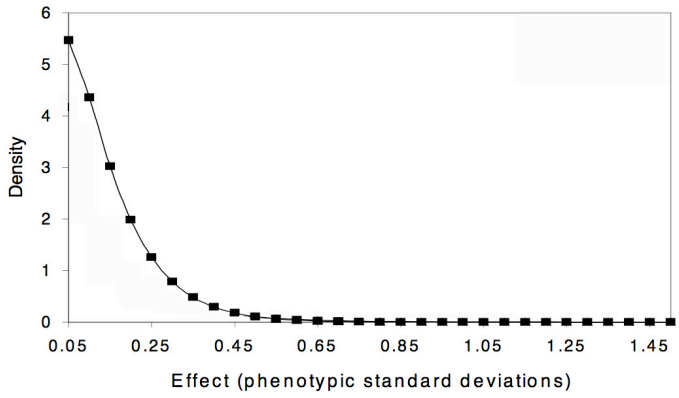


Figure 2.1 Distribution of QTL in dairy cattle (HAYES and GODDARD 2001)

Milk from cattle contains on average 87.3% water, 4.2% fat, 3.3% protein and 4.6% lactose but variation is seen between and within breeds. Since it is of central economic interest in dairy cattle breeding, QTL studies have been performed in past decades. Searching for a QTL starts with the association between a trait and a chromosomal position, represented by at least one polymorphic DNA marker. The marker alleles must be co-inherited with the alleles of the target gene. Markers that were available in the last decades were satellite-markers, mini-satellites or variable number of tandem repeat markers (WYMAN and WHITE 1980) and micro-satellites (WEBER and MAY 1989).

Mapping studies in cattle were first implemented with daughter designs (DD) (WELLER et al. 1990), but DDs were replaced in the following by granddaughter designs (GDD). GDDs are easily available in cattle populations, but more importantly, it has proved more effective

to use GDDs than DDs. The power of mapping studies in increased with a larger number of sons per grandsire also when using GDDs. For this reason, QTL mapping studies were commonly carried out with Holstein populations.

Later, and principally based on information about the genomic localization arising from QTL studies, candidate gene analyses were set up for a molecular dissection of the quantitative trait. During candidate gene analyses, the genomic region of QTLs is screened for functional candidate genes, genes that show a physiological context to the trait the QTL was identified for. In the context of association studies with candidate genes, polymorphisms of the selected gene are analysed for a putative effect on the quantitative trait to identify the causative mutation(s), so called QTN(s). However, the molecular situation remains complex, since several allelic effects of one gene may be coupled to operate as 'super-alleles' and show in their optimal ensemble the effect of the QTL (STAM and LAURIE 1996).

2.2. *QTL mapping approaches on cattle chromosome 6 (BTA6)*

A cattle QTL database gathering QTLs that were published during the past decade is available at http://www.animalgenome.org/cattle/. A consistent finding across all studies on BTA6 suggests a primary QTL for milk yield close to marker BM143 (Khatkar *et al.* 2004) (Olsen *et al.* 2002) (Ron *et al.* 2001). The position of marker BM143 corresponds to a chromosomal position on BTA6 at 40.57 Mb based on the bovine draft sequence *Btau 3.1* (http://www.ncbi.nlm.nih.gov/), which became available very recently. Apart from this, a QTL for milk yield was mapped to the casein gene cluster (Velmala *et al.* 1999) (Lien *et al.* 1995) (Velmala *et al.* 1995) (BOVENHUIS and WELLER 1994), which is located in *Btau 3.1* at about 75 Mb on BTA6.

For fat yield, QTLs were identified on BTA6 at markers BM143 (Ron *et al.* 2001) (Olsen *et al.* 2004), TGLA37A (Kuhn *et al.* 1999). The corresponding region in *Btau 3.1* is the region on BTA6 from 39.85 Mb to 46.13 Mb, in which also a QTL for protein yield as assumed (Kuhn *et al.* 1999) (Ron *et al.* 2001) (Zhang *et al.* 1998) (Olsen *et al.* 2004).

Most mapping studies identified QTLs for content traits. A QTL for fat content is assumed to be located close to the fat- and protein yield QTL at about 40 Mb on BTA6 in *Btau 3.1* (Zhang *et al.* 1998) (Ron *et al.* 2001) (Olsen *et al.* 2004). QTLs for protein percentage were discussed in two consensus regions on BTA6, one in the centre of BTA6 at several markers but approximately at 42 Mb (*Btau 3.1*) (Georges *et al.* 1995) (Zhang *et al.* 1998) (Velmala *et al.* 1999) (Spelman *et al.* 1996) (Ron *et al.* 2001) (Olsen *et al.* 2002) (Olsen *et al.* 2004) and

the second one more distal at the casein gene cluster (Velmala *et al*. 1999) (Velmala *et al*. 1995) (Boichard *et al*. 2003).

Along with single QTL studies, multi QTL models were also estimated. Two separate approaches suggested more than one segregating QTL. The first multi QTL model postulated the first QTL close to marker BM143 at 40.57 Mb and the second QTL close to marker BM415 at 67.26 Mb (Ron *et al*. 2001). The second multi QTL existence was reported for a much smaller chromosomal segment between markers BMS2508 at 39.8 Mb and BMS690 at 42.99 Mb (Olsen *et al*. 2004).

Due to inconsistencies in traits, chromosomal position and number of identified QTLs, numerous QTL mapping studies haven't been followed up or confirmed. However in summary for BTA6, it is most likely that three QTLs for milk traits reside in two distinct chromosomal regions. The proximal region between approximately 30 Mb and 45 Mb is most likely to harbour a QTL affecting milk yield and content traits (Olsen *et al*. 2005) besides a QTL with effects on fat and protein yield (Kuhn *et al*. 1999; Olsen *et al*. 2004). The third QTL is most likely to be located at the casein gene cluster and affects milk yield and content traits (Velmala *et al*. 1999).

2.3. *Previous association studies with candidate genes on cattle chromosome 6 (BTA6)*

The second step during a molecular dissection of quantitative traits is the search for the causative mutation(s) (QTN) and its gene(s) underlying the QTL effect. Candidate gene identification is often carried out using comparative gene maps. These are based on the conservation of synteny between species. Hence well-characterized model organisms and their chromosomal gene order were used to predict the chromosomal position of genes in non-model organisms. Whole genome comparative maps for humans and cattle (Hayes 1995) (Solinas-Toldo *et al*. 1995) and later radiation hybrid maps (Band *et al*. 2000) were available and were used for the identification of genes in the target species.

When selecting candidate genes, it is important to study the physiology influencing the trait of interest. The synthesis of milk components is a highly complex physiological process and for closer consideration, milk synthesis has to be divided into the synthesis of its main components, fat, protein and lactose. Lactose and caseins are mammary specific products and are synthesised de novo in the mammary epithelial cells after the precursors are blood-transported to the mammary system. Milk fat and its short and middle chain fatty acids are

synthesised de novo, with acetate and β-hydroxybutyrate as the main origins, but long chain fatty acids derive from body fat (see Figure 2.2).

De novo synthesised milk proteins need blood-transported amino acids as precursors. Caseins represent about 80% of the milk proteins and can be fractioned into α-S1, α-S2, β and κ-casein. Therefore it was logical that the first work on candidate gene analyses was done on genes of the casein cluster. In HF, no significant effects of casein variants were identified on milk-, fat- and protein yield (Sabour *et al.* 1996), but an association between a *CSN1S1*-variant and protein content was found (Prinzenberg *et al.* 2003). The casein genes were also investigated in cattle breeds where a high milk protein content is a breed characteristic. In Norwegian Red, significant associations were identified for milk and protein yield (Lien *et al.* 1995). In Brown Swiss, however, two separate studies have not shown consistent results (Braunschweig *et al.* 2000) (Boettcher *et al.* 2004).

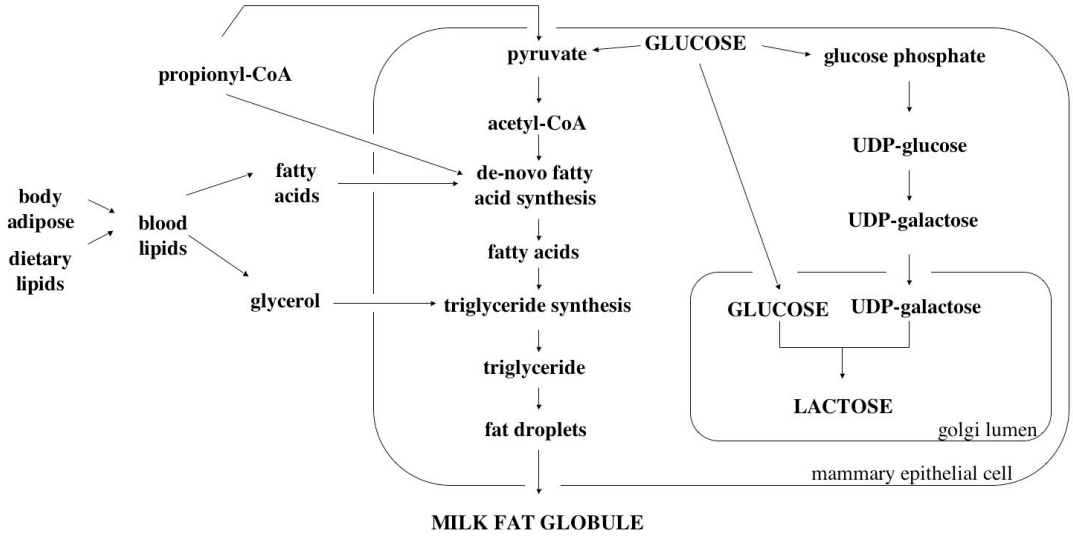


Figure 2.2 Synthesis of milk fat and lactose (http://classes.ansci.uiuc.edu/ansc438/)

Glucose is the essential precursor for the synthesis of lactose since lactose is a disaccharide composed of β1-4 glycosidic-linked galactose and glucose. Galactose occurs from the conversion of glucose to UDP-galactose via UDP glucose (see Figure 2.2) and the mammary gland cannot synthesize glucose from other precursors because of the lack of glucose-6-phosphatase (THREADGOLD and KUHN 1979) (Scott *et al.* 1976).

The lactose content in milk is relatively insensitive to environmental influences but it plays an important role in milk chemistry. Lactose is synthesised in the mammary epithelial cells' Golgi lumen. Due to a non-permeable character, diffusion from the Golgi lumen to the cell cytosol is not possible. Instead, water diffuses into the Golgi lumen and increases milk yield

to maintain the osmotic equilibrium. Since lactose is the main osmole in milk, lactose synthesis is the major factor influencing the volume and yield of milk (Cant *et al.* 2002) (Holt 1983). Considering a constant fat and protein yield synthesis, an increase in lactose synthesis ends in reduced milk fat and protein content due to an increased milk yield (Olsen *et al.* 2002). The negative correlation between milk yield and fat or protein yield is explained and it is shown that the several milk sub-traits depend on each other. It also becomes evident that glucose and the entire energy metabolism are crucial for milk synthesis, especially in ruminants where 40 – 60% of the blood glucose has to be synthesized via gluconeogenesis, from propionate and gucogenic amino acids in the liver (Kirchgeßner 1997). Therefore several candidate genes, which are known to be involved in different energy metabolic pathways were analysed in order to identify QTNs on BTA6.

Because of its chromosomal position and the key function in energy metabolism, *PPARGC1A* was discussed as a positional and functional candidate gene for a fat yield QTL on BTA6 (Kuhn *et al.* 1999). Statistical analyses revealed a significant association between a single nucleotide polymorphism (SNP) in intron 9 and fat yield (Weikard *et al.* 2005). Two other genes, *OPN* and *ABCG2*, were proposed to harbour a QTN for the proximal milk content QTL on BTA6 (Ron *et al.* 2001). *OPN* encodes a glycoprotein, which is involved in calcium metabolism, and *ABCG2* is a member of the ATP-binding cassette subfamily. Both were postulated by separate studies. For *OPN* and its insertion – deletion variant (INDEL) upstream of the transcription start site, a significant effect on fat and protein content was reported (Schnabel *et al.* 2005). The observed insertion deletion variant was described as a poly-T tract producing alleles of either nine (T9) or ten (T10) thymine nucleotides where the T9 variant decreased fat and protein content significantly (Schnabel *et al.* 2005). In the second study, eight genes within the chromosomal region of *ABCG2* were analysed for gene expression in the mammary gland. *OPN* and *ABCG2* showed significantly increased expression in the mammary gland from parturition through lactation (Ron *et al.* 2007) and were screened together with other genes in this region for variants in the bovine genomic sequence. Only one identified SNP showed co-segregation with the QTL marker. It was an SNP in exon 14 of *ABCG2*, encoding a tyrosine to serine mutation. The tyrosine variant has a frequency in the investigated population of 0.805 and was associated with an increase in content traits and an antagonistic decrease in milk yield (Cohen-Zinder *et al.* 2005). Both, *OPN* and *ABCG2* are anchored in the same chromosomal region, but only the approach in Israeli Holstein (Cohen-Zinder *et al.* 2005) involved DNA-variation of both genes. However the INDEL upstream of the *OPN* transcription start site was not included.

Finally, the molecular architecture of milk synthesis is complex so that previous association studies on BTA6 come to controversial conclusions since genes and DNA polymorphisms were analysed in isolation in most cases. A causative mutation however should be confirmed in other populations and breeds, as it was reported for the association of *DGAT1* (Winter *et al*. 2002) (Thaller *et al*. 2003). Hence, it is expected that the simultaneous observation of several genes located on BTA6 in more than one dairy cattle breed contribute to a molecular dissection of BTA6 in the context of milk synthesis.

3. Animals, Materials and Methods

3.1. *Animal panels for SNP screening and association studies*

There were two separate animal panels. The first was used for SNP screening and the second for the association studies. Both animal panels included sires in artificial insemination of three dairy cattle breeds, German Brown (BV), German Fleckvieh (FV) and German Holstein (HF). The panel for SNP screening contained 12 sires, 4 of BV, FV and HF, respectively. Prepared DNA (25 ng/μl) of the SNP screening panel was available from an earlier master thesis (see Table 3.1) (Giessler 2006).

Table 3.1 Animal panel for polymorphism screening

Identification code	Breed	Name	Year of birth	PS [a] / MGS [b] / MGGS [c]
276000934216432	BV	Vinstor	2000	Vinaut / Stor / Simon
276000934856703	BV	Humor	2000	Hussli / Pete-Rose / Videlus
276000934442734	BV	James	2001	Jemstone / Monopoly / Emory
276000935377641	BV	Efast	2001	Even / Strifast / Vigate
276000936199323	FV	Eibsee	2003	Einser / Horb / Romit
276000937578788	FV	Honduras	2002	Hippo / Horramor /Balist
276000937638989	FV	Pomal	2003	Poldi / Malhax / Halling
276000936995477	FV	Rügen	2003	Rumba / Samurai / Morten
276000937712020	HF	Renegade	2004	Reno / Picasso / - [d]
276000662206477	HF	Lionel	2003	Lheros / Bestow / - [d]
276000938688368	HF	Tartan	2003	Talent / Durham / Formation
250001450822003	HF	Koldy	2003	Courier / Manfred / Celsius

[a] Paternal sire
[b] Maternal grand sire
[c] Maternal grand grand sire
[d] Sire unknown

The panel that was used for the association studies contained 3124 animals in all, of which 332 sires (62 BS, 197 FV, 73 HF) were available at the Chair of Animal Breeding (Winter *et al*. 2003) as prepared DNA (25 ng/μl). Additionally, semen straws of BV and FV sires were obtained from German artificial insemination stations and genomic DNA was prepared. Genomic DNA (20 ng/μl) from HF sires was obtained from The Research Institute for the Biology of Farm Animals (FBN), Dummerstorf, so that the final animal panel contained 742 BV sires born between 1997 and 2001, 964 FV sires born between 1997 and 2000 and 1418 HF sires, born between 1979 and 2000. BV and FV samples are each organized into 141 different paternal half-sib groups. HF comprise 107 different groups. The mean family size is 5.26 in BV, 6.79 in FV and 13.25 in HF.

3.2. *Materials and Methods*

3.2.1. In-silico analyses

3.2.1.1. Selection of candidate genes

Recent QTL-mapping studies indicate that BTA6 plays an important role with regard to milk production traits. Six functional and positional candidate genes on BTA6 were selected for association studies with milk production traits. Four of the genes of interest, *ABCG2*, *OPN*, *PPARGC1A* and *CSN1S1* have been analysed in previous studies. Two functional and positional candidate genes, not yet investigated for milk production traits in dairy cattle, were added to this list. These genes were identified by *geneScore* (Bininda-Emonds *et al.* 2004), an in-house tool for candidate gene identification. *geneScore* search is based on physiological keywords and screens the *pubmed* database (http://www.ncbi.nlm.nih.gov/) for genes that contain the keyword in their gene ontology description. The search was confined to the part of the human chromosome 4 (HSA4), which is known to be homologue to BTA6. *geneScore* output is a list that orders the identified genes by their identification frequency in *pubmed*.

3.2.1.2. Gene identification in the bovine genomic sequence

In June 2005, the second version of the bovine draft sequence (*Btau 2.0*) became open source. The bovine draft sequence is a preliminary assembly of the cow (Bos taurus), and *Btau 2.0*, an updated version of *Btau 1.0*, which became available in September 2004 (Havlak *et al.* 2004) (http://www.hgsc.bcm.tmc.edu/projects/bovine/). *Btau 2.0* is organised in contigs, which are contiguous lengths of genomic sequence in which the order of bases is known at a high confidence level. Contigs show overlapping regions and can be organised by creating an assembly (scaffold). Gaps can occur where reads from the two sequenced ends of at least one fragment overlap with other reads in two different contigs (http://genome.jgi-psf.org/help/scaffolds.html). 50% of the bovine contigs span more than 19.9 kb and more than half of the created bovine scaffolds are bigger than 434.7 kb. *Btau 2.0* consists of 4409 chromosomally allocated and 98058 chromosomally unanchored scaffolds.

Btau 3.1 is the third assembly of the Bos taurus genome. It was produced with a combination of Whole Genome Shotgun Reads (WGS) from *Btau 2.0* and BAC sequences. *Btau 3.1* is a set of contigs and scaffolds, which include sequence contigs oriented with respect to each other. This reciprocal scaffold orientation leads to chromosomal stringed scaffolds. The total length of all contigs is 48.7 kb and 50% of the assembled genome lies in blocks of more than

997.5 kb. *Btau 3.1* is like *Btau 2.0*, still a draft version of the bovine genomic sequence (http://www.hgsc.bcm.tmc.edu/projects/bovine/) and may contain errors, such as miss-assemblies of repeated sequences, collapses of repeated regions and unmerged overlaps.

Both, *Btau 2.0* and *Btau 3.1* were used for the identification of the candidate genes in the bovine genome. The files Btau_20050310.contigs.fa.gz for *Btau 2.0* and Btau20060815.contigs.fa.gz for *Btau 3.1* were downloaded from the ftp server to a local server from the Human Genome Sequencing Centre at the Baylor College of Medicine (http://www.hgsc.bcm.tmc.edu/projects/bovine/).

Applying a *python* script (*python* v.2.4.1), the human coding sequences of the candidate genes were downloaded as fasta-files from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/).

Basic Local Alignment Search Tool (BLAST) was used for the in-silico identification of the candidate genes in *Btau 2.0* and *Btau 3.1*. BLAST is an algorithm for the comparison of biological sequence information (Altschul *et al*. 1990). BLAST offers the identification of a sub-sequence in the library sequence that is similar to a sub-sequence of the target sequence. Since coding sequence shows a high sequence similarity between species (Andersson *et al*. 2001), BLAST enabled an in-silico identification of the human reference sequence in the bovine draft sequences. BLAST was run on a local server.

3.2.1.3. Gene annotation

The *GenomeThreader* software (http://www.GenomeThreader.org/) was applied (Gremme *et al*. 2005) for a semi-automated annotation of the selected candidate genes in the bovine draft sequence. *GenomeThreader* applies an algorithm to compute gene structure predictions by using a similarity-based approach. It requires a query sequence and aligns this with a database sequence, which contains cDNA and EST sequences from humans, cattle, mice, rats, pigs and chickens. In the present study, subsets of *Btau 2.0* and *Btau 3.1*, identified previously by BLAST (see 3.2.1.2 Gene identification in the bovine genomic sequence) were taken as query sequences in *GenomeThreader*. For an identification of repetitive sequence, *GenomeThreader* runs *RepeatMasker* (http://www.repeatmasker.org/). The *GenomeThreader* output file was converted to an xml-file and was read into the *Apollo* Annotation and Curation Tool v.6.1 (Lewis *et al*. 2002) (http://www.fruitfly.org/annot/apollo/) for visualisation and editing.

3.2.1.4. Genomic characterisation of candidate genes

A genomic characterisation was done for the selected candidate genes. The conservation of the peptide sequence between species was observed with *ClustalW* v.1.83 (Thompson *et al.* 1997) and the bovine peptide sequence was aligned with the peptide sequence of human, mouse and chicken. The visualisation of the protein sequence alignment was carried out on the *BoxShade* server, available at http://www.ch.embnet.org/software/BOX_form.html. To calculate the peptide similarity, alignments were repeated in pairs using the alignment tool from the European Bioinformatics Institute (http://www.ebi.ac.uk/emboss/align/index.html) (Larsen *et al.* 1992).

The transcription start site in the bovine sequence was predicted with an online prediction tool (http://www.fruitfly.org/seq_tools/promoter.html) (Reese 2001). 500 bp upstream of the 3' end of the first exon were used without changing the standard settings of the online tool. A second prediction result was indirectly obtained from *GenomeThreader* annotation and results were compared.

The 5'-end sequence, the region upstream of the transcription start site, harbours regulatory elements for gene expression. The prediction of putative CpG-islands was run with the program *cpgplot* (http://www.ebi.ac.uk/emboss/cpgplot/) (Larsen *et al.* 1992) from the European Bioinformatics Institute. To analyse TATA-box enriched promoter regions, *Consite* tool (http://mordor.cgb.ki.se/cgi-bin/CONSITE/consite) was used and the bovine sequence was screened for a subset of transcription factor binding sites, namely NF-Y, TBP, SP1, NF-kappaB, MYF, HNF, GATA and cEBP. Transcription factor binding sites were only predicted when the similarity between human and bovine 5'-end sequences was at least 0.80. All in-silico identifications of regulatory elements in the 5'-end were conducted within a sequence fragment of 1 kb, upstream of the 3' splice site of exon 1.

The 3' untranslated region (UTR) is known to contain post-transcriptional regulation elements, such as miRNA binding sites (Clop *et al.* 2006). Putative binding sites were predicted with the web-tool available at http://microrna.sanger.ac.uk/sequences/ and 1 kb genomic sequence downstream of the stop codon triplet was screened.

The effect of SNPs in translated sequence on the primary structure of the peptide was observed with the online tool *Transeq* from the European Bioinformatics Institute (http://www.ebi.ac.uk/emboss/transeq/).

3.2.2. Primer design

Primers were designed using the Primer3 software available at http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi. Estimated melting temperatures of primers were set between 57 and 63°C with the optimum at 60°C. Optimal primer size was set at 20 bases and the optimal content of GC bases was at 50%. All primers were designed based on genomic sequence from *Btau 2.0* and *Btau 3.1*, respectively.

3.2.3. Preparation of genomic DNA

The genomic DNA was prepared (Buitkamp 1999) from the semen of the animals from the panel used for association studies. The content of the semen straws, containing a defined volume of the ejaculate but also glycerol and other additives, was transferred to 1.5 ml tubes. For washing, 1,000 μl of 1x PBS (8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, pH 7.4) was added and centrifuged at 16000 G for 5 minutes. The pellet was re-suspended after removal of the supernatant. Cell lysis was carried out by adding 750 μl 1x PKS (20 mM Tris HCl, 4 mM EDTA, 100 mM NaCl, 1% SDS, pH 7.4), 50 μl Proteinase K (20mg/ml), 20 μl DTT (1M) and by incubation at 56 °C for 12 hours including moderate shaking. After transferring the lysate to Vaccutainer Blood Collection Tubes (BD-368510; Becton, Dickinson and Company, Franklin Lakes, NJ), 2000 μl phenol/chlorophorm/isoamlyalcohol (25:24:1) was added and gently mixed while inverting the tubes. Afterwards this was centrifuged for 10 minutes at 600 G. 1000 μl chlorophorm/isoamylalcohol (24:1) was added and centrifugation at 600 G for 10 minutes was repeated. The supernatant containing the DNA was transferred to a 15 ml Falcon tube and was mixed with 3000 μl ethanol (99%) for DNA precipitation while inverting the tube. The precipitated DNA was transferred to a 1.5 ml tube, and centrifuged for 3 minutes at 16000 G. After eliminating the supernatant, the DNA-pellet was stored in an incubator at 60°C for 30 minutes to evaporate residual ethanol and 100 μl TE buffer was added.

To measure the concentration of the prepared DNA, a 1:5 dilution was made. The measurement was realised in a fluorometer (DyNA Quant 200, Hoefer Pharmacia Biotech Inc.; San Francisco, CA, USA) and with Bisbenzimide H33258 dye (Sigma-Aldrich; Steinheim; GER). 20 samples were added, together with lambda DNA (SD0011, MBI Fermentas, St. Leon-Rot, Germany), to a 0.8% agarose gel containing TBE buffer to control the measurement result. Afterwards the 1:5 dilutions were adjusted with TE buffer to a

concentration of 25 ng/μl and transferred to 96-well plates for storage. Each plate included two positive and two blank controls.

The 25 ng/μl working solution was diluted with TE buffer to achieve the required DNA-solution of 1 ng/μl for genotyping with matrix-assisted laser-desorption / ionization time-of-flight mass spectrometry (MALDI-TOF). For storage, 8 μl were transferred to 384-well plates by using a Tecan robotic pipetting device (Genesis RSP 150 Work Station, Tecan Trading AG, Switzerland). DNA was dried over night, covered with a plastic film and stored.

3.2.4. Polymerase chain reaction (PCR)

3.2.4.1. Standard PCR

PCR systems amplified 50 ng of genomic DNA in a Temperature Gradient Thermocycler (Biometra, Göttingen, Germany). The PCR under standard conditions was done in a reaction volume of 20 μl, containing 0.5 units Quiagen Taq DNA Polymerase (201203; Quiagen, Hilden, Germany), 1x Quiagen PCR buffer, 1.5mM $MgCl_2$, 200 μM of each nucleotide and 0.125 μM for both forward and reverse primer under the following conditions: initial denaturation at 95°C for three minutes, followed by 30 cycles consisting of 40 seconds at 94°C for denaturation, 50 seconds at 60°C for primer annealing and 90 seconds at 72°C for extension. This was followed by a final extension step for seven minutes.

To control PCR amplification, the product was applied to a 2% agarose gel, together with a DNA size marker (Genruler 100bp DNA Ladder, SM0241; MBI Fermentas, St. Leon-Rot; Germany).

3.2.4.2. PCR optimisation steps

If necessary the standard PCR system was modified by adding additive reagents. The reaction was completed by 1x Quiagen Q-Solution and 5% DMSO (dimethyl sulfoxid), respectively. Next, a temperature gradient (52.5°C – 67.5°C) was applied during primer annealing of the standard PCR reaction. For some PCR systems it was necessary to optimise the content of $MgCl_2$ and 2.5mM, 3.5mM or 4.5mM $MgCl_2$ was tested in separate assays. Some PCR systems amplified with HotStarTaq DNA Polymerase (203205; Quiagen, Hilden, Germany) and an initial denaturation at 95°C for 15 min. The PCR system supplying the best results was used for further experiments.

3.2.5. DNA sequencing

DNA sequencing was performed according to the Sanger method (Sanger *et al.* 1977) by base specific dye-labelled didesoxynucleotides. The mixture of didesoxynucleotides and unmodified nucleotides generates randomly distributed termination positions over the given template sequence. After denaturation, electrophoresis fractionation and dye detection, the distribution of the termination positions is shown in the newly synthesised DNA. Sequencing was performed on an ABI 377 automated sequencer (Applied Biosystems, Foster City, CA, USA).

3.2.5.1. PCR cleaning

The PCR products were purified of unincorporated nucleotides and reagents before using them as templates during sequencing reaction. 50 μl double distilled water was loaded together with the complete PCR product to MultiScreen$_{HTS}$ PCR$_{96}$ Filter Plates (MANU03050; Millipore S.A., Moisheim, France) and vacuum suction cleaning (14 inch Hg, 5 min.) followed on a Millipore vacuum manifold (MAVM0960R; Millipore S.A., Moisheim, France). After vacuum cleaning, PCR samples were re-suspended in 35 μl of Tris buffer (10mM, pH 8.0) and separated, together with a size standard (Genruler 100bp DNA Ladder, SM0241; MBI Fermentas, St. Leon-Rot; Germany) on a 2.0% agarose gel.

3.2.5.2. Sequencing reaction

Sequencing reaction was performed in a volume of 10 μl, containing 0.5 μM each of forward and reverse primer and 2 μl of BigDye® Terminator v1.1 Sequencing Kit (4337449; Perkin Elmer Applied Biosystems Division, Foster City, CA, USA). 10 – 20 ng of purified PCR product was supplemented for the sequencing reaction applying the following temperature and time profile: initial denaturation at 95°C for 15 seconds, followed by 35 cycles consisting of 10 seconds at 95°C for denaturation, 5 seconds at 51°C for primer annealing and 6 minutes at 60°C for extension.

3.2.5.3. Sequencing on an ABI377 automated sequencer

Sequencing reaction was purified of unincorporated nucleotides with Sephadex columns. MultiScreen HV Plates (MAHVN4510; Millipore S.A., Moisheim, France) were loaded with 25 mg Sephadex G-50 (Sigma-Aldrich Chemie GmbH, Deisenhofen, Germany) and 300 μl

double distilled water. After incubation for 2 hours at room temperature, columns were formed by centrifugation at 920 G for 5 min. Sequencing reaction was added to the sephadex columns, purified while centrifuging (920 G, 5 min) and afterwards vacuum-dried by centrifugation at 0.3 inch Hg in a vacuum centrifuge (Speed vac plus, SC 110A; Thermo Savant, Holbrook, NY, USA). The purified sequencing product was re-suspended in 2.5 μl of ABI loading buffer (formamid coloured with dextran blue, 47670 and 31393; Fluka, Buchs, Switzerland), denaturated at 96°C for 2 minutes, loaded to a 5% Polyacrylamid/7M Urea gel and separated by electrophoresis using an ABI Prism® 377 Sequencer (Applied Biosystems, CA, USA). DNA Sequencing Analysis Software v3.4.1 (Applied Biosystems, CA, USA) was used to process the obtained data.

3.2.5.4. Detection of DNA variants

DNA variants were identified by comparative sequencing of the animal panel for SNP screening (see 3.1 Animal panels for SNP screening and association studies). The obtained sequence was analysed to detect genomic variation using Phred/Phrap/Polyphred/Consed software suite (EWING and GREEN 1998; EWING *et al*. 1998; GORDON *et al*. 1998; NICKERSON *et al*. 1997). Each detected SNP was archived in a local database and assigned an in-house identification code (SNP_id). The databases contained information as follows: nucleotide exchange, forward and reverse primer, locus name and the upstream and downstream sequence as obtained from comparative sequencing.

3.2.6. Selection of SNPs for association studies

Among all identified DNA variants a subset was selected for genotyping and association studies. To execute a breed-comparative association, the set of selected DNA variants was genotyped across breeds. Due to the small number of animals used during SNP-screening, DNA variants were selected without considering SNP-tagging software. However, functional features, as well as technical features with respect to the genotyping method were considered.

3.2.7. Genotyping with iPLEX Assay for Matrix-assisted Laser-Desorption / Ionisation Time-of-Flight Mass spectrometry (MALDI-TOF)

The selected DNA variants were genotyped with iPLEX chemistry on a MALDI-TOF Mass Spectrometer (Sequenom Inc., San Diego, CA, USA). iPLEX-method requires first a PCR-

amplification of the genomic region of the DNA variant. A treatment with shrimp alakaline phosphatase (SAP) purifies the PCR product from unincorporated nucleotides to avoid that they are built in the extension products and falsify the results. In general, iPLEX-reaction follows the principles of Primer Extension Reaction (see Figure 3.1). The extension primers anneal adjacent to the SNP and the addition of a DNA polymerase along with a mixture of nucleotides and a terminator nucleotide allows extension through the polymorphic site and generates allele specific extension products having different masses depending on the incorporated nucleotide at the SNP.

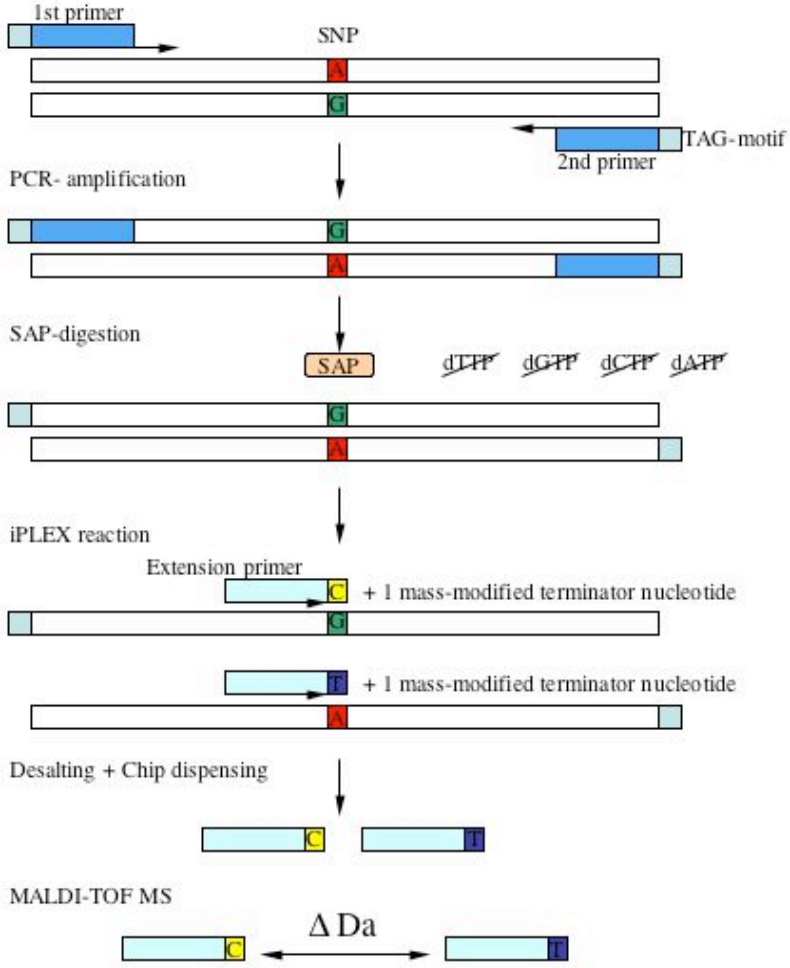


Figure 3.1 Principle of the iPLEX method (Juling 2007)
After PCR-amplification with tagged primers the unbound nucleotides are dephosphorylated by SAP-digestion. The iPLEX reaction contains chain terminating nucleotides. Extension primers are elongated by one base depending on the SNP-allele. The two extension products are discriminated by mass differences of the incorporated bases in MALDI-TOF MS.

In fact, mass differences between the terminator nucleotides are small, 9-40 Daltons, so iPLEX Assay uses mass modified terminators to increase the mass difference (http://www.sequenom.com). For Matrix Assisted Laser Desorption / Ionisation (MALDI)

iPLEX products are transferred to a matrix, which is a low molecular mass organic acid. After this, samples are exposed under vacuum conditions to laser impulses (KARAS and HILLENKAMP 1988). The matrix is required for the absorption of the laser energy and sample ionisation. During energy transfer to the matrix fixed sample, ions become free and change to vapour (Hillenkamp *et al*. 1991). Ions are extracted in an electric field and are separated by their mass charge-ratio before they arrive on the detector. Ions with a higher ratio need more time to run through the duct than ions with low ratio and the time-of-flight (TOF) (SAUER and GUT 2002) is specific for each product. MALDI-TOF mass array system is highly sensitive and is applicable for SNP analysis (Storm *et al*. 2003).

3.2.7.1. Primer design for iPLEX Assay

Every SNP was stored on an in-house database including the upstream and downstream sequence as obtained by comparative sequencing (see 3.2.5.4 Detection of DNA variants). Modification steps were necessary. Upstream and downstream sequence of the selected SNPs, was aligned with repetitive sequence from RepBase Update 20060129 using RepeatMasker v.3.1.3. Identified repetitive sequence was replaced by *N*s'. Polymorphic sites in the up- or downstream sequence from the SNP of interest were replaced by *N* to inhibit primer design in polymorphic region. Both steps were carried out with a *python* script (v.2.4.1).

Primer design and assay design were carried out applying MassARRAY® Typer 3.4 Software (Sequenom, San Diego, CA, USA). Two separate primer assays were made for each SNP: primers for the initial amplification and primers for iPLEX reaction. For a uniform amplification, the initial PCR product for each SNP should contain a comparable number of nucleotides (100 bp) (http://www.sequenom.com) for each SNP. The primers for the initial amplification contain, next to the sequence specific part, an additional sequence motif (Tag-motif) of 10 bp to increase the primer mass (see Figure 3.1) in order that the primer mass does not correspond to the mass of any extinction product later. Specific mass differences among the extinction primers of all SNPs and iPLEX chemistry allowed an amplification of more than one single base extension product in one assay (multi-plex). MassARRAY® Typer 3.4 Software (Sequenom, San Diego, CA, USA) was used for the assay design and plexing of extension products.

3.2.7.2. PCR for iPLEX Assay

8 ng of genomic DNA was amplified in the initial PCR reaction. PCR was done in a reaction volume of 5 μl, containing per μl master mix 0.1 U of Quiagen HotStarTaq DNA Polymerase (203205; Quiagen, Hilden, Germany), 1.25x Quiagen PCR buffer, 3.5mM $MgCl_2$, 0.5mM of each nucleotide and 0.1 μM each for the forward and reverse primer under the following conditions: initial denaturation at 94°C for 15 minutes, followed by 44 cycles consisting of 20 seconds at 94°C for denaturation, 30 seconds at 56°C for primer annealing and 60 seconds at 72°C for extension. This was followed by a final extension step at 72°C for 3 minutes.

The master mix was added to the cured genomic DNA by using a Tecan robotic pipetting device (Genesis RSP 150 Work Station, Tecan Trading AG, Switzerland). The amplification products of 24 samples, including negative controls, of each 382-well plate were applied to a 3% agarose gel, together with a DNA size marker (Genruler 100bp DNA Ladder, SM0241; MBI Fermentas, St. Leon-Rot; Germany).

3.2.7.3. Treatment with shrimp alakaline phosphatase (SAP)

Unincorporated dNTPs can increase contaminant peaks during data analysis; therefore it is necessary to dephosphorylate the PCR product of all unincorporated nucleotides by a treatment with shrimp alkaline phosphatase (SAP). The SAP treatment is divided into two steps, the SAP digestion and the SAP-enzyme deactivation. SAP assay had a volume of 2 μl, added, with the aid of a Tecan robotic pipetting device (Genesis RSP 150 Work Station, Tecan Trading AG, Switzerland), to the PCR volume of 5 μl. SAP assay contained per μl 0.3 U of SAP – Enzyme (#10002, Sequenom, San Diego, CA, USA) and 0.85x MassEXTEND™ Buffer (#10055, Sequenom, San Diego, CA, USA) and was performed under the following conditions: SAP-digestion at 37°C for 20 minutes, followed by 5 minutes at 85°C for enzyme deactivation.

3.2.7.4. iPLEX reaction

iPLEX reaction was carried out in a volume of 2 μl, added to the 7 μl assay after SAP treatment after preparing the extension primer mix. Extension primer dependent concentration adjustments for primers joined in one assay are crucial. MALDI-TOF Mass Spectrometry is characterised by the inverse relationship between analyte signal-to-noise ratio and increased mass of the analyte. Therefore larger extension products would result in smaller signal–to-

noise ratios without concentration adjustments (http://www.sequenom.com). To compensate this effect, it is suggested that all oligos of one multiplex assay are divided into four almost equal parts, depending on the mass of the extension products. The quarter of oligos characterised by the largest mass portion among all oligos in one multiplex-assay is recommended for loading 2 x the concentration of the quarter characterised by the lowest mass proportion. Similarly, the third and second quarter of oligos, respectively are recommended for loading 1,51 x and 1,24 x. Thus the final concentration ratios in the total assay for extension reaction for each quarter were adjusted as follows: 0.625μM : 0.780 μM : 0.96 μM : 1,25 μM. The assay for iPLEX reaction contained: 0.222x pPLEX buffer (Sequenom, San Diego, CA, USA), 1X iPLEX extension mix (Sequenom, San Diego, CA, USA) and 0.03 U/μl iPLEX enzyme (Sequenom, San Diego, CA, USA). Primer extension reaction was amplified using the temperature time profile shown in Table 3.2.

Table 3.2 Time and temperature profile for iPLEX reaction

iPLEX step	Temperature (°C)	Period	
1. Initial denaturation	94	30 sec	
2. Denaturation	94	5 sec	go 4-times to step 3
3. Oligo annealing	52	5 sec	
4. Strand extension	80	5 sec	go 39-times to step 2
5. Final extension	72	3 min	
6. Sample storing	4	∞	

3.2.7.5. Desalting the iPLEX reaction

Desalting of the iPLEX reaction products is necessary to optimize mass spectrometric analysis. If samples would not have been desalted properly, sodium (22 Da) and potassium (38 Da) adducts could complicate the heterozygote allele discrimination for A/C (24 Da) and C/G (40 Da) polymorphisms. 6 mg of MassEXTEND Clean Resin kit (#10053, Sequenom, San Diego, CA, USA) was dispensed on 6 mg dimple microplates (#11235, Sequenom, San Diego, CA, USA) and, after incubation at room temperature for 30 minutes, was added to the samples. 16 μl of nanopure water was added and samples were gently mixed on a rotator for 10 minutes.

3.2.7.6. Chip dispensing and MALDI-TOF Analysis

The desalted samples were spotted to SpectroCHIPs (#00601, Sequenom, San Diego, CA, USA), which contain the matrix for laser absorption. Using a nanoplotter (SpectroJET, Sequenom, San Diego, CA, USA), 15 nl of each sample was dispensed on the SpectroCHIP™

by piezoelectric pipetting. The SpectroCHIP is a silicon chip that incorporates a high-density photo resistant array of mass spectrometry analysis sites. Samples for calibrating the system are required on each chip: 3-Oligo Calibrant Mix (#00335, Sequenom, San Diego, CA, USA) contains a mixture of oligonucleotides of known molar mass. After charging the chip with samples and calibrant mixture, the chip is delivered to a carrier and introduced into the vacuum sluice of MassARRAY mass spectrometer.

3.2.8. Data processing

3.2.8.1. Phenotypic data

Phenotypic data of BV and FV were supplied by the Bayerischen Landesanstalt für Landwirtschaft, Institute of Animal Breeding (http://www.lfl.bayern.de). The Vereinigte Informationssysteme Tierhaltung w.V. (http://www.vit.de) supplied HF-data. Daughter Yield Deviations (DYD) (MRODE and SWANSON 2004) (Liu *et al.* 2004) served as phenotypic data. Apart from DYD, information about the number of daughters per sire, separately for each lactation, and genetic parameters, namely the heritability and phenotypic variance of a trait for all breeds and each lactation (see Table 3.3) were obtained and stored on an in-house database system. DYD for the following milk production traits were received separately for each of the first three lactations: milk yield (MY1, MY2, MY3), fat yield (FY1, FY2, FY3) and protein yield (PY1, PY2, PY3). DYD are average phenotypes of daughters corrected for fixed effects, herd, season and calving interval and corrected for the daughters' dams' genetic contributions.

Table 3.3 Genetic parameters for the weighted least squares analysis
Traits represent first, second and third lactation for milk yield (MY1, MY2, MY3), fat yield (FY1, FY2, FY3), protein yield (PY1, PY2, PY3), fat content (FC1, FC2, FC3) and protein content (PC1, PC2, PC3)

Breed	Genetic parameter	MY1	MY2	MY3	FY1	FY2	FY3	PY1	PY2	PY3
BV	h^2	0.39	0.34	0.32	0.35	0.32	0.27	0.30	0.31	0.29
	σ_p^2	585928	921114	1034988	921	1528	1690	644	1053	1139
FV	h^2	0.36	0.32	0.33	0.31	0.30	0.27	0.27	0.27	0.27
	σ_p^2	680549	1010029	1104146	1065	1747	1924	701	1124	770
HF	h^2	0.53	0.35	0.34	0.52	0.36	0.36	0.51	0.38	0.38
	σ_p^2	661091	918266	1037747	1010	1503	1803	578	771	882

The minimum number of daughters per lactation for DYD estimation was set at 10. Table 3.4 shows means and standard deviations for the DYD of each breed. For milk content traits, fat content (FC1, FC2, FC3) and protein content (PC1, PC2, PC3), no direct DYD estimates could be obtained due to the indirect estimation in random regression test day models.

Consequently, the DYD for milk content traits were calculated similarly to the procedure applied during estimations of breeding values (VIT 2004) (BLT 2007). The formula includes estimated DYD for milk, fat or protein yield as well as the population means for milk yield and respective fat or protein yield trait. Population means for yield traits were obtained from breeding evaluation centres.

Exemplary demonstration for protein content:

$$DYD_{proteincontent} = \frac{DYD_{fatyield} * 100 - \left(DYD_{milkyield} * PM_{proteincontent}\right)}{DYD_{milkyield} + PM_{milkyield}}$$

Linear regression models were carried out weighting each observation. The weight was proportional to one over the residual variance as shown in the following chapter 3.2.9.4 Linear regression models. For HF, the estimates from marker assisted estimation of breeding values were delivered by VIT and used for the derivation of *DGAT1* genotype (Winter *et al.* 2002).

Table 3.4 Phenotypic data

Total number of sons (n), mean and standard deviation (sd) of daughter yield deviations (DYD) of sons for milk production traits representing first, second and third lactation for milk yield (MY1, MY2, MY3), fat yield (FY1, FY2, FY3), protein yield (PY1, PY2, PY3), fat content (FC1, FC2, FC3) and protein content (PC1, PC2, PC3) for BV, FV and HF. DYD for content traits were calculated as described.

Trait	BV		FV		HF	
	n	Mean (± sd)	n	Mean (± sd)	n	Mean (± sd)
MY1	409	116 (± 194)	896	166 (± 256)	1304	441 (± 535)
MY2	363	141 (± 251)	779	187 (± 286)	1304	518 (± 591)
MY3	269	121 (± 252)	446	162 (± 305)	1209	547 (± 615)
FY1	402	5.64 (± 7.38)	917	5.84 (± 9.08)	1304	11.94 (± 19.06)
FY2	363	6.55 (± 9.79)	779	6.16 (± 10.93)	1304	14.75 (± 22.84)
FY3	269	5.79 (± 9.18)	446	8.21 (± 11.54)	1209	16.85 (± 24.79)
PY1	402	4.17 (± 5.53)	912	5.05 (± 7.19)	1304	13.17 (± 14.25)
PY2	363	5.24 (± 7.52)	779	5.45 (± 8.71)	1304	15.66 (± 16.80)
PY3	269	4.25 (± 7.54)	446	4.89 (± 8.69)	1209	15.34 (± 17.90)
FC1	402	0.016 (± 0,104)	896	-0.002 (± 0.136)	1304	-0.098 (± 0.290)
FC2	363	0.011 (± 0.110)	779	-0.021 (± 0.140)	1304	-0.096 (± 0.302)
FC3	269	0.010 (± 0.105)	446	-0.024 (± 0.153)	1209	-0.062 (± 0.309)
PC1	402	0.004 (± 0.064)	896	-0.015 (± 0,076)	1304	-0.024 (± 0.115)
PC2	363	0.004 (± 0.071)	779	-0.015 (± 0,077)	1304	-0.027 (± 0.121)
PC3	269	0.001 (± 0.073)	446	-0.008 (± 0.084)	1209	-0.038 (± 0.125)

3.2.8.2. Genotypic data

Genotypes from MALDI-TOF MS were transferred to an in-house database.

3.2.9. Statistical analyses

3.2.9.1. Testing for Hardy-Weinberg Equilibrium

Within each breed, genotypes were observed for deviations from the Hardy-Weinberg Equilibrium (FALCONER and MACKAY 1996). Tests were carried out using Fisher-Exact test as implemented in R (www.r-project.org). Results were adjusted for multiple testing by Bonferroni correction.

3.2.9.2. Analysing Linkage Disequilibrium between candidate genes

D' and r^2 were used to analyse Linkage Disequilibrium (LD) (FALCONER and MACKAY 1996). *Haploview* program v.4.0 was used for the visualization of linkage disequilibrium (LD) (Barrett *et al*. 2005).

For a distance dependent estimation of LD, D' was fitted in the Malecot model (Malécot 1948):

$$D'_i = (1 - L)M \exp(-\kappa d_i) + L + \varepsilon_i$$

D'_i is the observed LD for the ith pair of loci distance d_i apart, L ($0 \leq L \leq 1$) is the residual association at large distances, M ($0 \leq M \leq 1$) is the proportion of alleles transmitted from founders, κ (> 0) is the exponential decay rate of D' with physical distance d, and ε_i is the random error deviation for the ith pair of loci. The marker distances were calculated based on the assembly of the bovine genomic sequence *Btau 3.1*, which is a draft version of the bovine genomic sequence. Therefore genomic positions of markers have approximate character.

The Malecot model (Malécot 1948) was modified, henceforth termed modified Malecot model, in order to ascertain, taking into account the physical distance, whether LD for marker pairs located in the same gene is higher than for marker pairs of different genes:

$$D'_i = (1 - L)M \exp(-\kappa_1 d_i + (-k_2 d_i) * z_i) + L + \varepsilon_i$$

where D'_i is the observed LD for the ith pair of syntenic loci distance d_i apart, L ($0 \leq L \leq 1$) is the residual association at large distances, M ($0 \leq M \leq 1$) is the proportion of alleles

transmitted from founders, κ_1 (> 0) is the exponential decay rate of D' with physical distance d for markers within a gene, κ_2 (> 0) is the exponential decay rate of D' with physical distance d for markers of different genes, z_i is an indicator variable that takes the value 0 if both markers are in the same gene and 1 if both markers are in a different gene and ε_i is the random error deviation for the ith pair of loci. Both Malecot models were fitted in a non-linear regression model using the *nls* function as obtained by *R-{stats}* package (www.r-project.org).

3.2.9.3. Haplotype construction

Haplotypes were constructed separately for breeds applying two software packages. *SimWalk2* v.2.91 (SOBEL and LANGE 1996) constructs the haplotypes with a Markov chain Monte Carlo algorithm including parental genotypes. Parental genotypes were not available, but grandsires' genotypes were inferred by *SimWALK2* for grandsires with ≥ 12 genotyped sons. *fastPHASE* does haplotype construction based on linkage disequilibrium and allele frequencies (SCHEET and STEPHENS 2006) and was applied for animals in parental half sib groups counting < 12 bulls.

3.2.9.4. Linear regression models

3.2.9.4.1. Linear regression models in single marker analysis (SMA)

R environment (www.r-project.org) was used for data analyses. Models for single marker analyses (SMA) were applied as linear regression models on the number of alleles at each marker on DYD of the trait under consideration. Models included the sire as fixed effect. Paternal half-sib families counting less than 2 members were pooled. In case of HF, models included the *DGAT1*-genotype since *DGAT1*-variation is known to explain phenotypic variation in milk production traits of HF (Winter *et al.* 2002). Linear regression models fitting DYD derived from breeding value estimation do not allow estimation of dominance effects, and regression coefficients represent half of the gene substitution effect ($\alpha/2$) (FALCONER and MACKAY 1996). Allele substitution effects were estimated with the following model:

$$y_{ijk} = \mu + b * SNP + gs_i + DGAT1_j + e_{ijk}$$

where y_{ijk} is the DYD of son k within sire i, *DGAT1* genotype j. μ is the overall mean, gs_i is the fixed effect of sire i. *SNP is* the copy-number of allele 1 (0, 1, or 2) at each SNP and b is

the regression coefficient representing half of the gene substitution effect ($\alpha/2$) (FALCONER and MACKAY 1996). $DGAT1_j$ is the fixed effect of the *DGAT1*-genotype and e_{ijk} is the random residual effect. In SMA each yield trait was weighted proportional to one over $var(e_{ijk})$:

$$\mathrm{var}(e_{ijk}) = \left[\frac{1}{4} h^2 + \frac{\left(1 - \frac{1}{4} h^2\right)}{n_{ij}} \right] \sigma^2{}_p$$

where h^2 is the heritability, σ^2_p the phenotypic variance of the trait under consideration and n_{ij} is the number of daughters of sire *j* within sire *i* included for the calculation of the DYD. Due to the indirect estimation of milk content traits in random regression test day models, the calculation of $var(e_{ijk})$ was not possible and SMA was performed without weighting the phenotypic value.

Multiple testing was accounted for by estimation of the false discovery rate (FDR) for each trait *x* SNP combination (STOREY and TIBSHIRANI 2003). The false discovery rate is estimated as follows:

$$\hat{FDR}(t) = \frac{\hat{\pi}_0 m * t}{S(t)}$$

where *m* represents the number of tests and *t* the threshold calling features significant with $0 < t \leq 1$. *S(t)* is the number of features with *p*-values below *t*. $\hat{\pi}_0$ is the estimated proportion of null hypotheses and quantified as follows:

$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda; i = 1,\ldots,m\}}{m(1-\lambda)}$$

where λ is the parameter derived from the *p*-value distribution from *m* tests.

3.2.9.4.2. Linear regression models in haplotype analysis (HA)

Constructed haplotypes were analysed, fitting the number of haplotypes of each animal in linear regression models. Models included the grandsire as a fixed effect with half-sib groups counting less than 2 members being pooled. Models for HF also included the *DGAT1*-genotype as a fixed effect. Each haplotype was fitted separately, regressing the trait value on the number of haplotypes while pooling the remaining haplotypes. Substitution effects for a certain haplotype were estimated while pooling all other haplotypes applying the following model:

$$y_{ijk} = \mu + b * HAT + gs_i + DGAT1_j + e_{ijk}$$

where y_{ijk} is the DYD of son k within sire i, *DGAT1* genotype j. μ is the overall mean, gs_i is the fixed effect of sire i and *HAT* is the copy-number (0, 1, or 2) of each haplotype-allele (*HAT*), b is the regression coefficient representing half of the gene substitution effect ($\alpha/2$) (FALCONER and MACKAY 1996), $DGAT1_j$ is the fixed effect of the *DGAT1*-genotype and e_{ijk} is the random residual effect. Each yield trait was weighted proportional to one over $var(e_{ijk})$ in HA, as shown in SMA (see 3.2.9.4.1 Linear regression models in single marker analysis (SMA)).

False Discovery Rate (*FDR*) (STOREY and TIBSHIRANI 2003) for each trait x haplotype combination was applied to account for multiple testing, as shown in 3.2.9.4.1 Single marker analysis (SMA).

Each significantly associated haplotype was inspected for SNPs that discriminate the significantly associated haplotype from all others. Discriminating SNPs capture information of significantly associated haplotypes, so that it is sufficient to genotype the discriminating SNP only. In cases where no SNP allowing the discrimination of the significantly associated haplotype from all others was identified, haplotypes were grouped according to the effect direction in the trait of interest. Subsequently haplotype groups were screened for discriminating SNPs.

3.2.9.4.3. Linear regression models in multi marker analysis (MMA)

Multi marker analysis (MMA) was carried out applying linear regression models fitting the number of alleles of several markers. Models included the sire as a fixed effect, while paternal half sib groups counting less than 2 animals were grouped. The *DGAT1* genotype was fitted as fixed effect in models with data from HF:

$$y_{ijk} = \mu + b_1 * SNP1 + \ldots + b_n * SNPn + gs_i + DGAT1_j + e_{ijk}$$

where y_{ijk} is the DYD of son *k* within sire *i, DGAT1* genotype *j*. μ is the overall mean, gs_i is the fixed effect of sire *i*, *SNP1* the copy-number of allele 1 (0, 1 or 2) at SNP *1*, *SNPn* the copy-number of allele 1 (0, 1 or 2) at SNP *n*. b_1 is the regression coefficient representing half of the gene substitution effect (α/2) of SNP *1*, b_n is the regression coefficient representing half of the gene substitution effect (α/2) of SNP n (FALCONER and MACKAY 1996), $DGAT1_j$ is the fixed effect of the *DGAT1*-genotype and e_{ijk} is the random residual effect. Each yield trait was weighted proportional to one over $var(e_{ijk})$ (see 3.2.9.4.1 Linear regression models in single marker analysis (SMA)).

MMA was carried out fitting markers of one gene and fitting markers of several genes to detect correlated effects due to linkage disequilibrium among significantly associated markers in SMA and HA. The initial multi marker model includes markers of one gene that are significantly associated in SMA and that are identified as discriminating significantly associated haplotypes. To detect correlated effects among the fitted markers, the initial multi marker model is stepwise reduced following the model optimization criteria of *AIC*-criterion (Akaike 1974). Following these steps separately for gene *x* trait combinations with more than one significantly associated marker after SMA and HA, the initial multi marker model resulted in a model, fitting a set of markers, explaining the data in the best possible way. After the initial multi marker model was optimized, model optimization according to *AIC*-criteria was repeated including two-way marker interactions. Models were compared to the models without interaction and interaction terms were dropped if *p* values for the F-tests were ≥ 0.05.

The final marker set was created after completing MMA separately for all breeds including markers that were significantly associated in SMA, that discriminated significantly associated haplotypes and that were identified in MMA to be associated with marker specific effects. Haplotypes were constructed with these markers for each breed as described in 3.2.9.3 Haplotype construction.

3.2.9.4.4. Haplotype analysis with final marker set

Haplotypes constructed with markers of the final marker set were used to calculate the proportion of explained additive genetic variance. This is a quotient of the additive genetic variance explained by the haplotypes, divided by the total additive genetic variance. The total additive genetic variance was calculated for yield traits using heritability and phenotypic variance as obtained from breeding evaluation centres (see 3.2.8.1 Phenotypic data).

Multiple alleles occur in all breeds and were considered during estimation of the substitution effects. The additive genetic variance of loci with multiple alleles under non-random mating is calculated as (LYNCH and WALSH 1998):

$$\sigma^2_A = 2\sum_{i=1}^{n} p_i \alpha^2_i$$

where p is the frequency of haplotype i, α is the allele substitution effect of haplotype i, summarized from haplotype 1 to haplotype n with an estimation of α according to:

$$\alpha_i = \sum_{j=1}^{n} p_j G_{ij} - \mu_G$$

where p_j is the frequency of the haplotype allele j, G is the genotypic value of diplotype ij of haplotype j and μ_G is the population mean of the trait of interest (LYNCH and WALSH 1998). Least square means estimation of G_{ij} considered the grandsire, in HF also the *DGAT1*-genotype, as fixed effects. μ_G was calculated using G_{ij} and p_j.

4. Results

4.1. *Selection of candidate genes*

Six candidate genes were selected on BTA6 for the association studies. Four genes were selected by literature review: *ABCG2* (Cohen-Zinder *et al.* 2005), *OPN* (Schnabel *et al.* 2005), *PPARGC1A* (Weikard *et al.* 2005) and *CSN1S1* (Prinzenberg *et al.* 2003). The identification of additional candidate genes was realised with *geneScore*, an in-house tool for candidate gene identification. In total, 14 genes were identified on HSA4 to have 'glucose', the keyword that was used in *geneScore* search, in their gene ontology description. Two out of these 14 genes, *PGM2* and *SLC2A9*, were selected based on their chromosomal position. *PGM2* is located within the region of a previously identified QTL (Kuhn *et al.* 1999). *SLC2A9* is located at a distance of approximately 18 Mb in distal orientation of the casein gene cluster. However a QTL has not been mapped to this part of BTA6 so far, *SLC2A9* is selected as a functional candidate gene for milk synthesis. Table 4.1 lists the selected candidate genes.

Table 4.1 Selected candidate genes

Gene	Gene name	Accession number [a]	Genomic localisation [b]
ABCG2	ATP-binding cassette subfamily G member 2	NM_004827	4q22
CSN1S1	Casein alpha-S1	NM_001890	4q21.1
OPN	Osteopontin	NM_000582	4q21-q25
PGM2	Phosphoglucomutase-2	NM_018290	4p14
PPARGC1A	Peroxisome proliferate activated receptor alpha	NM_013261	4p15.1
SLC2A9	Solute carrier family 2 member 9	NM_001001290	4p16-p15.3

[a] Human mRNA reference sequence number (Acc. from NCBI)
[b] Genomic localisation in humans

4.2. In-silico *gene structure derivation*

4.2.1. Positioning of candidate genes in bovine draft sequences

Complete bovine annotations of the selected genes were not available. Therefore, local BLAST searches used the reference sequences, as downloaded from the National Center of Biotechnology Information (http://www.ncbi.nlm.nih.gov) to localise the candidate genes in the bovine genome. The first BLAST search was run against *Btau 2.0*. Results are shown in Table 4.2. All exons of *PPARGC1A* and *CSN1S1* were identified in the bovine sequence, applying a significance level of 0.10 for the sequence alignment. In the case of *ABCG2*, *OPN*,

PGM2 and *SLC2A9*, respectively, one exon remained unidentified. Human SLC2A9 exon counts 102 bp. Scaffold 6.200 includes only 70 bp from the 5' part of SLC2A9 exon 10 due to unknown sequence. However, the gap could be closed by comparative sequencing during SNP-screening: Acc. EU295941 (http://www.ncbi.nlm.nih.gov/).

Table 4.2 BLAST result against *Btau 2.0*

For each BLAST template, the scaffold number in *Btau 2.0*, the number of exons of the gene in humans and the corresponding bovine exons as identified by BLAST is displayed.

Gene	BLAST template [a]	Number of exons (HSA)	Scaffold	Identified exons in *Btau 2.0*
ABCG2	NM_004827	16	ChrUn.7178	2 to 8, 10 to 13
			Chr6.46	14 to 16
			ChrUn.82793	9
OPN	NM_000582	6	Chr6.46	1, 2, 4 to 6
PPARGC1A	NM_013261	13	Chr6.67	1 to 12
			ChrUn.84773	13
PGM2	NM_018290	14	Chr6.82	1 to 4, 6 to 14
CSN1S1	NM_181029 [b]	19	ChrUn.7178	1 to 19
SLC2A9	NM_001001290	13	ChrUn.2195	12,13
			Chr6.200	2 to 6, 9 to 11 [c]
			Chr6.174	7, 8

[a] Human mRNA reference sequence number (Acc. from NCBI)
[b] The bovine reference sequence was taken for the positioning in *Btau 2.0* (Acc. from NCBI).
[c] Chr6.200 harbours Exon 10 only partially, *Btau 2.0* updated with sequence from comparative sequencing during SNP-screening, EU295941 (Acc. from NCBI).

Gene positioning was repeated when *Btau 3.1* became available. Due to a linear chromosomal anchorage of the scaffolds in *Btau 3.1*, the BLAST results provide information about the chromosomal position and BLAST searches map the gene in-silico. All exons of *PPARGC1A*, *PGM2* and *CSN1S1* were identified at a significance level of 0.10 for the sequence alignment (see Table 4.3.) The first exon of *ABCG2*, the third exon of *OPN* and exon 1, exon 9 and exon 10 of *SLC2A9* remained undetected. The missing exons of *ABCG2* and *OPN*, respectively, could be identified by manual annotation in *Btau 3.1*. *SLC2A9* exon 11 and exon 12 were identified in chromosomally unanchored scaffold sequence in *Btau 3.1*, but exon 2 to exon 8 anchor in scaffold sequence allocated to BTA6. Surprisingly, the BLAST output for exon 2 to exon 8 showed inconsistent sequence orientation in Btau 3.1. The second, third and fourth exon were identified in 5' – 3' orientation, exon 5 to exon 8 in reverse complement orientation. SLC2A9 exon 9 and exon 10 have not been identified in *Btau 3.1*.

BLAST searches identified *ABCG2* and *OPN* in the chromosomal region at 33 Mb. *PPARGC1A* is positioned in *Btau 3.1* on BTA6 at 41 Mb, *CSN1S1* at 79 Mb and *SLC2A9* at 97 Mb. As it was detected during the chromosomal positioning of *SLC2A9*, the corresponding BTA6-region in *Btau 3.1* is inaccurate. *Btau 3.1* was only taken for the rough localisation of

SLC2A9 on BTA6 while Btau 2.0 sequence was taken for the genomic annotation. A position on the minus strand is assumed, and the 5' splice donor of exon 3 is assumed at 97157422 bp on BTA6 for further analyses.

Table 4.3 BLAST result against *Btau 3.1*

For each BLAST template, the number of exons of the gene in humans and the corresponding location in *Btau 3.1* and strand is displayed.

Gene	BLAST template [a]	Number of exons (HSA)	Identified exons in *Btau 3.1*	Localisation in *Btau 3.1* (bp)	Strand in *Btau 3.1*
ABCG2	NM_004827	16	2 to 16	BTA6 (33575490 to 33620247)	+
OPN	NM_000582	7	1, 2, 4 to 7	BTA6 (33710591 to 33717339)	-
PPARGC1A	NM_013261	13	1 to 13	BTA6 (41139556 to 41242941)	-
PGM2	NM_018290	14	1 to 14	Chr.Un.003.335 [c] (77487 to 113728)	-
CSN1S1	NM_181029 [b]	19	1 to 19	BTA6 (79752087 to 79769150)	-
SLC2A9	NM_001001290	13	2 to 8	BTA6 (97178174 to 102566754)	+ / - [d]
			11 to 12	Chr.Un003.893 [c] (2708 to 28864)	+

[a] Human mRNA reference sequence number (Acc. from NCBI)
[b] The bovine reference sequence was taken for the positioning in *Btau 2.0* (Acc. from NCBI).
[c] Chromosomal unanchored scaffold
[d] Orientation of coding sequence is inconsistent in *Btau 3.1*

The coding sequence of *PGM2* was identified in *Btau 3.1* on a chromosomally unanchored scaffold, Chr.Un.003.335. In contrast, the annotation with *Btau 2.0* identified *PGM2* on BTA6. Since it is unlikely that the chromosomal position of a single gene differs between *Btau 2.0* and *Btau 3.1*, five *PGM2*-neighboured genes in the human genome assembly were taken for additional BLAST searches against *Btau 3.1*. Two adjacent genes were localized in the human genome proximally of *PGM2* and three distally (see Table 4.4.).

BLAST searches anchored *PCDH7*, *CENTD1*, *KLF3* and *KLHL5* in *Btau 3.1* to BTA6 (see Table 4.4). *PGM2* and its closest neighbour in the human genome assembly, *PTTG2* were identified on chromosomally unanchored sequence in *Btau 3.1*. *PCDH7* and *CENTD1*, the two genes located proximally to *PGM2* in the human genome assembly were identified in Btau 3.1 on BTA6 at 47 Mb and 51 Mb, respectively. *KLF3* and *KLHL5*, localized distally to *PGM2* in the human genomic sequence, were mapped by BLAST in *Btau 3.1* on BTA6 to 53 Mb and 54 Mb, respectively. Hence, comparing the BLAST result of *PCDH7* and *CENTD1* with the result of *KLF3* and *KLHL5* indicates a gap of unknown sequence from approximately 51.7 Mb to 53.8 Mb on BTA6 in *Btau 3.1*.

Table 4.4 BLAST result against *Btau 3.1* for genes adjacent of *PGM2* in humans
For each BLAST template, the chromosomal position and strand in humans, the corresponding location in *Btau 3.1* and strand is displayed

Gene	BLAST template [a]	Location on HSA4 in bp (strand)	Localisation on BTA6 (*Btau 3.1*) in bp (strand)
PCDH7	NM_032457.1	30278000 to 30807000 (+)	46617089 to 47101758 (-)
CENTD1	NM_015230.2	35722000 to 35945000 (-)	51261150 to 51730680 (+)
PGM2	NM_013261	37544200 to 37500200 (+)	Chr.Un.003.335 [b]
PTTG2	NM_006607.1	37638379 to 37639098 (+)	Chr.Un.003.101 [b] Chr.Un.003.127 [b]
KLF3	NM_016531.3	38337900 to 38381100 (+)	53831438 to 53841603 (+)
KLHL5	NM_001007075.1	38713000 to 38810000 (+)	54195293 to 54264293 (+)

[a] Human mRNA reference sequence number (Acc. from NCBI)
[b] Chromosomal unanchored scaffold, unknown orientation in the bovine genomic sequence

However, a high conservation level of gene order between human and cattle (Solinas-Toldo *et al.* 1995; Weikard *et al.* 2007) supports a chromosomal localisation of *PGM2* in the bovine genomic sequence on BTA6 at approximately 52 Mb. The annotation in the reverse complement sequence of Chr.Un.003.335 was taken for further analyses, while the first nucleotide of the start codon was set to a position of 52180182 on BTA6. In fact, the genomic orientation of *PGM2* relative to the centromer could not be resolved by these methods. The neighbouring genes of *PGM2* do not show similar orientation between the strand in humans and cattle (see Table 4.4), but the 5' – 3' orientation is assumed in cattle, since *PGM2* is located on the plus strand in the human genomic sequence.

4.2.2. Annotation of candidate genes

The annotation of the candidate genes was done semi-automatically using *GenomeThreader* (Gremme *et al.* 2005) after results from manual positioning by BLAST were available. The *GenomeThreader* output file was converted to an xml-file before it was visualised by the Apollo Annotation and Curation Tool v.6.1 (Lewis *et al.* 2002). The subset of bovine sequence identified previously by BLAST was taken as query sequence for the *GenomeThreader* alignment (see 4.2.1 Positioning of candidate genes in bovine draft sequences). Since *ABCG2* and *OPN* are positioned closely in the bovine genome assembly, both genes were annotated in one *GenomeThreader* run using the chromosomal region of BTA6 from 33.50 Mb to 33.63 Mb (see Table 4.5). For the annotation of *PPARGC1A*, the BTA6-region from 40.8 Mb to 41.4 Mb was used. *PGM2* was identified on a chromosomally

unanchored scaffold, Un.003.335, whereof the reverse complement sequence between 60.0 kb and 130.0 kb was run in *GenomeThreader*. *CSN1S1* was annotated with *GenomeThreader* together in one run with the other four genes of the casein cluster and within the BTA6-region from 79.4 Mb to 79.9 Mb. As shown in 4.2.1 Positioning of candidate genes with bovine draft sequences, in-silico identification of *SLC2A9* by BLAST resulted in a consistent annotation only with *Btau 2.0* sequence. Hence, *Btau 2.0* sequence was taken for the *GenomeThreader* annotation of *SLC2A9*. As shown in Table 4.2, exon 2 to exon 6, and exon 9 to exon 11 were identified in Btau 2.0 on scaffold 6.200, but exon 7 and exon 8 were encompassed by scaffold 6.174. Hence, scaffold 6.200 was cut into two parts, to insert the sequence of scaffold 6.174. Additionally, exon 10 was included incompletely in scaffold 6.200 due to missing sequence within this region. By comparative sequencing of this region in cattle, the unknown sequence could be replaced and the scaffold sequence was updated manually within this region. To complete the annotation of SLC2A9 with *GenomeThreader*, Chr.Un.2195 was attached to the rearranged sequence of scaffold 6.200 and 6.147. The *GenomeThreader* annotations resulted, apart from *SLC2A9*, in complete gene annotations for the candidate genes. Compared to the manual annotations by BLAST, where the first exon of *ABCG2* and exon 3 of *OPN* did not result in a BLAST hit, the *GenomeThreader* annotation revealed complete gene annotations. In contrast, the detection of the first exon of *SLC2A9* also failed when using *GenomeThreader*. For details of the gene annotations, see chapter below (4.3 Characterisation of candidate genes).

Table 4.5 Results from semi-automated gene annotation with *GenomeThreader* (Gremme *et al.* 2005)

Gene	Number of exons in (HSA)	Localisation in *Btau 3.1* and sequence (Mb) selected for annotation		Identified exons by *GenomeThreader*
ABCG2	16	BTA6 [a]	(33.50 to 33.63)	1 to 16
OPN	7	BTA6 [a]	(33.50 to 33.63)	1 to 7
PPARGC1A	13	BTA6 [a]	(40.80 to 41.40)	1 to 13
PGM2	14	Chr.Un.003.335 [b]	(0.06 to 0.13)	1 to 14
CSN1S1	19	BTA6 [a]	(79.40 to 79.90)	1 to 19
SLC2A9	13	ChrUn.2195 [c]	[d]	12, 13
		Chr6.200 [c]	[d]	2 to 6, 9 to 11 [e]
		Chr6.174 [c]	[d]	7, 8

[a] Scaffold anchors in *Btau 3.1* to BTA6, as identified by BLAST
[b] Chromosomal unanchored scaffold in *Btau 3.1*
[c] Sequence from *Btau 2.0*
[d] Entire sequence of *Btau 2.0* scaffold was taken for annotation
[e] Chr6.200 harbours from Exon 10 only 70 bp (see 4.2.1 Positioning of candidate genes in bovine draft sequences), *Btau 2.0* updated with sequence from comparative sequencing during SNP-screening (EU295941, Acc. from NCBI)

4.3. *Characterisation of candidate genes*

4.3.1. ATP-binding cassette sub-family G member 2 gene (*ABCG2*)

ABCG2 is a member of the ATP binding cassette (ABC) subfamily, which has five genes: *ABCG1*, *ABCG2*, *ABCG4*, *ABCG5* and *ABCG8* (KUSUHARA and SUGIYAMA 2007). The proposed topology model of human ABCG2 protein represents an N-terminal nucleotide binding domain and the C-terminal transmembrane domain with six transmembrane segments. Between the fifth and sixth transmembrane domain is a large extracellular loop, which includes three cysteines of which one is involved in the formation of an intramolecular disulfide bond. The other two cysteines form an intermolecular disulfide bond (Henriksen *et al.* 2005). ABCG2 accepts a broad ligand spectrum of drugs (Breedveld *et al.* 2004) (Hirano *et al.* 2005) (Jonker *et al.* 2005) (Merino *et al.* 2005) (Merino *et al.* 2006), among them drugs which are involved in ATP-dependent processes (Litman *et al.* 2000). *ABCG2* is expressed in different tissues whereas the highest expression among 61 investigated organs was detected in the mammary gland (RON *et al.* 2007; SU *et al.* 2002). In mice as representatives for mammals, ABCG2 was recently reported to be the first active efflux transporter, mediating riboflavin concentration in milk (van Herwaarden *et al.* 2007). This study raises the possibility that nutrients can be transported into milk by ABC transporters. Therefore it is possible that *ABCG2* has an impact on milk synthesis and milk composition.

Using BLAST searches, *ABCG2* was mapped in silico on BTA6 at 33.5 Mb. It is interesting to note that in Israeli Holstein, a QTL affecting protein content was mapped to marker BM143 (Weller *et al.* 2002) which corresponds in Btau 3.1 to the chromosomal position at 40.5 Mb and explains the role of *ABCG2* as positional candidate gene.

Based on the *GenomeThreader* annotation, the bovine *ABCG2*, as well as the human, is partitioned into 14 exons (see Figure 4.1). The coding sequence of *ABCG2* encompasses in humans 68.8 kb and in cattle 70.52 kb and a large intron 1, counting about 25.5 kb in humans and cattle (see Table 4.6), is characteristic for this gene.

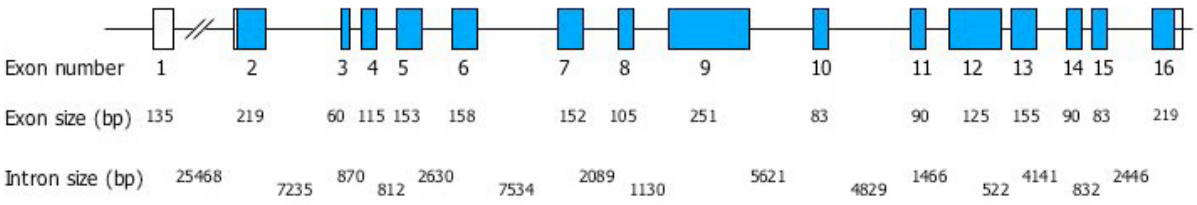


Figure 4.1 Graphical overview of the bovine *ABCG2*

The blue bars represent translated sequence. White bars represent untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in *Btau 3.1* is given below the bars.

Table 4.6 Intron / exon organisation of *ABCG2* in *Btau 3.1*

No.	Exon Size (bp)	Position in cDNA	Localisation on BTA6 (bp)	3' splice acceptor ... 5' splice donor [a]	No.	Intron Size (bp)
1	135	1 - 135	34049793 to 34049927	agt**ct**AAAGC … CGCCG**gt**gag	1	25468
2 [b]	219	136 - 355	34075396 to 34075614	gga**ag**GCGGA … ATCAA**gt**atg	2	7235
3	60	356 - 416	34082850 to 34082909	tac**ag**TGGAG … TCTTC**gt**gag	3	870
4	115	417 - 532	34083780 to 34083894	tat**ag**GTTGT … TACAA**gt**aag	4	812
5	153	533 - 686	34084707 to 34084859	tta**ag**GATGA … CCAAG**gt**aat	5	2630
6	158	687 - 845	34087489 to 34087646	ttt**ag**GTTGG … AAGAG**gt**aaa	6	7543
7	152	846 - 998	34095190 to 34095341	ttt**ag**GATGT … CATAG**gt**atg	7	2089
8	105	999 - 1104	34097430 to 34097534	tgc**ag**GTTTC … TGAAG**gt**aag	8	1130
9	251	1105 - 1356	34098665 to 34098915	cac**ag**CTAAC … CTCAG**gt**aac	9	5621
10	83	1357 - 1440	34104537 to 34104619	aaa**ag**CTAAT … AACAG**gt**gag	10	4829
11	90	1441 - 1531	34109449 to 34109538	cac**ag**AGCCG … TTTAT**gt**gag	11	1466
12	125	1532 - 1657	34111005 to 34112129	tct**ag**ACATG … GTTAG**gt**aag	12	522
13	155	1658 - 1813	34112652 to 34112506	cta**ag**GACTG … TGATG**gt**aag	13	4141
14	90	1814 - 1904	34116648 to 34116737	tgt**ag**ATATT … ATGCG**gt**atg	14	832
15	83	1905 - 1988	34117568 to 34117650	tat**ag**GCTTT … GCCAT**gt**aag	15	2446
16 [c]	219	1989 - 2208	34120097 to 34120316	ttc**ag**ATGTA … AAGTT**tt**ttt		

[a] Intron sequence is indicated by lowercase letters, exon sequence by capital letters, the splice site sequence is in bold print. Information follows *GenomeThreader* annotation (see 4.2.2 Annotation of candidate genes).
[b] Exon contains start codon.
[c] Exon contains stop codon.

The translated sequence consists of 1965 bp in the bovine genome, corresponding to a protein sequence of 655 residues. In bovine and human, the translation start site is in exon 2, the stop codon in exon 16. The organisation of the protein motifs in cattle is comparable to the organisation in humans. Walker A motif was found in exon 3, while exon 6 contains the nucleotide binding domain ABC signature motif, as well as the Walker B motif. Together, Walker A and Walker B motif form the ATP-binding cassette and are responsible for ATP binding (Langmann *et al*. 2000). The transmembrane regions of the ABCG2 are coded by exons 10, 13, 14, 15 and 16 (Bailey-Dell *et al*. 2001).

ClustalW (v.1.83) (Thompson *et al*. 1997) was used to examine the similarity of the amino acid sequence between human, bovine, murine and galline species. Figure 4.2 shows the alignment displayed by *BoxShade* v.3.21 (http://www.ch.embnet.org/software/BOX_form.html). The similarity between the human and the bovine peptide sequence is 90.1%. The similarity between the murine and human sequence is 71.9% and 76.2% between chicken and human.

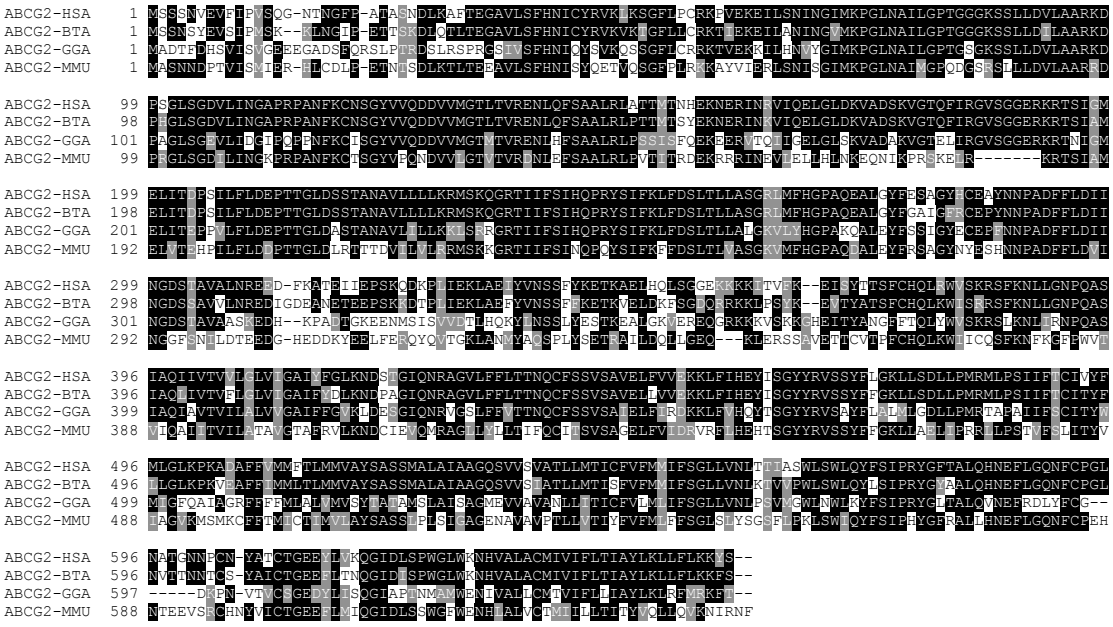


Figure 4.2 Alignment for *ABCG2* amino acid sequences of different species
Sequence alignment of *ABCG2* from Homo sapiens (ABCG2-HSA) (Acc. NP_004818.2, NCBI), Bos taurus (ABCG2-BTA) as translated from predicted mRNA sequence, Gallus gallus (ABCG2-GGA) (Acc. XP_421638.2, NCBI) and Mus musculus (ABCG2-MMU) (Acc. NP_036050.1). Numbers represent protein residues. Black and grey marked fonts show identical and conservative amino acids, respectively.

Transcription start site prediction failed using the online tool *Promoter Prediction* (http://www.fruitfly.org/seq_tools/promoter.html). The predicted transcription start site in *GenomeThreader* annotation (Gremme *et al.* 2005) did not correspond to the human transcription start site.

Therefore, a sequence fragment of 1 kb upstream of the 3' end of exon 1 was screened for CpG islands with the program *cpgplot* from the Emboss web-service (http://www.ebi.ac.uk/em
boss/cpgplot/) (Larsen *et al.* 1992). A CpG island was identified from 202 bp to 491 bp upstream of the 3' splice site of exon 1 (see Figure 4.3). This prediction was confirmed in the human sequence, where a CpG-island with a length of 729 nucleotides was detected.
Screening for putative transcription factor binding sites at a minimum conservation cut-off level of 0.80 failed between human and cattle sequence failed. A miRNA binding site has not been identified in 3' UTR sequence.

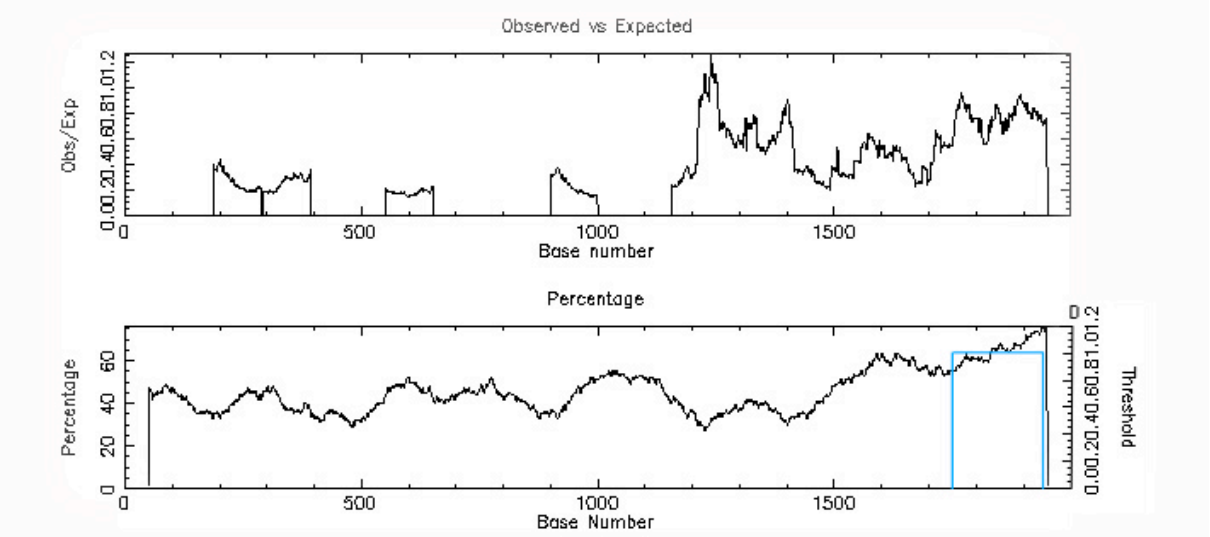


Figure 4.3 CpG island prediction for *ABCG2* in cattle
Upper box represents ratios of observed versus expected CpG content. The lower box shows the relative C+G content and the predicted CpG-island by the blue box. Horizontal axes show nucleotide numbers in relation to the 3' end of first exon.

In conclusion, *ABCG2* can be described as a gene that belongs to the class of genes characterised by CpG-rich promoter regions (Carninci *et al*. 2006). As was observed during the genomic characterisation, the sequence similarity in the 5' region is low. No transcription factor binding sites were detected, but the CpG-island prediction resulted in a putative CpG-island counting 289 bp so the promoter region of *ABCG2* can be described as a region having a high GC content but lacking canonical TATA and CAAT boxes. Genes that are regulated by CpG-islands are known to show low sequence similarity and their expression is known to be relatively tissue unspecific (Carninci *et al*. 2006).

4.3.2. Osteopontin gene (*OPN*)

OPN encodes a highly acidic phosphoprotein. It is involved in cell to cell communication but also in interactions between cell and extracellular matrix components (DENHARDT and GUO 1993). Potentially relevant against the background of milk synthesis are the effects of OPN on gene expression, Ca^{2+}-regulation, production of nitric oxides and effects on an involvement in cell migration (DENHARDT and GUO 1993). *OPN* is expressed in the mammary gland, whereas expression in the murine mammary gland depends on postnatal development (RITTLING and NOVICK 1997). Nemir et. al (2000) have shown, that OPN is essential for the development of the mammary gland and for lactation. They generated transgenic mice that expressed OPN antisense-RNA in the mammary gland. These mice showed a lack of mammary gland structure, lactation deficiency and a reduced synthesis of β-casein (Nemir *et al*. 2000).

Based on the result of a local BLAST search of human *OPN* cDNA (NM_000582) against the bovine genomic sequence *Btau 3.1* the position of *OPN* on BTA6 is close to 33,7 Mb. As already shown in 2.2 QTL mapping approaches on cattle chromosome 6 (BTA6), several QTL mapping studies for milk production traits have focused upon the QTL on BTA6 near *BM143* for milk yield and milk content traits. Comparing the chromosomal position of *OPN* with the *BM143*-position in *Btau 3.1*, *OPN* is localised proximally to *BM143* at a distance of approximately 6.8 Mb. Knowledge about the function and position on BTA6 of *OPN* underlines the role as a functional and positional candidate gene for milk synthesis.

The bovine gene structure of *OPN* was obtained with *GenomeThreader*. (Gremme *et al.* 2005). *OPN* shows a similar intron / exon structure in cattle and in humans. The coding sequence of *OPN* encompasses 7.9 kb in humans and 7.0 kb in cattle. The bovine translated sequence consists of 1965 bp, corresponding to a protein sequence of 655 residues. The bovine *OPN*, as well as the human *OPN* include the translation start site in exon 2 and the stop codon in exon 7 (see Figure 4.4 and Table 4.7).

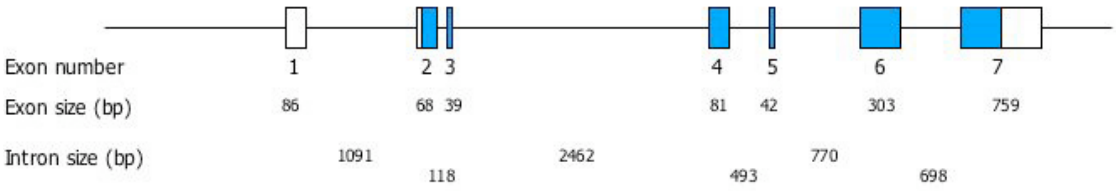


Figure 4.4 Graphical overview of the bovine *OPN*
The blue bars exhibit translated sequence. White bars exhibit untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in *Btau 3.1* is given below the bars.

Table 4.7 Intron / exon organisation of *OPN* in *Btau 3.1*

No.	Exon Size (bp)	Position in cDNA	Position on BTA6 (bp)	3' splice acceptor ... 5 'splice donor [a]	No.	Intron Size (bp)
1	86	1 - 86	34217297 to 34217212	tca**gc**AGCAG … CCAAG**gt**aag	1	1091
2 [b]	68	87 - 155	34216120 to 34216053	tgc**ag**GAAAA … TTCCA**gt**gag	2	118
3	39	156 - 195	34215934 to 34215896	caa**ag**GTTAA … AGCAG**gt**aag	3	2462
4	81	196 - 277	34213433 to 34213353	tgt**ag**CTTAA … CACAG**gt**att	4	493
5	42	278 - 320	34212859 to 34212818	ttc**ag**AATTC … AAAAT**gt**gag	5	770
6	303	621 - 724	34212047 to 34211745	tct**ag**ACCCT … TTCAG**gt**aaa	6	698
7 [c]	759	625 - 1384	34211046 to 34210288	ttc**ag**AGTCC … TCTGG**tg**cta		

[a] Intron sequence is indicated by lowercase letters, exon sequence by capital letters, the splice site sequence is in bold print. Information follows *GenomeThreader* annotation (see 4.2.2 Annotation of candidate genes).
[b] Exon contains start codon.
[c] Exon contains stop codon.

The conservation level of peptide sequence between the human, bovine, murine and galline species was analysed using the software *ClustalW* v.1.83 (Thompson *et al.* 1997) (see Figure

4.5). Similarities of 69.1% and 71.2% were detected between humans and cattle and between humans and mice. The galline peptide sequence indicated a lower similarity to its mammalian orthologs. Comparing the peptide sequence between humans and chicken lead to a similarity of 42.5%.

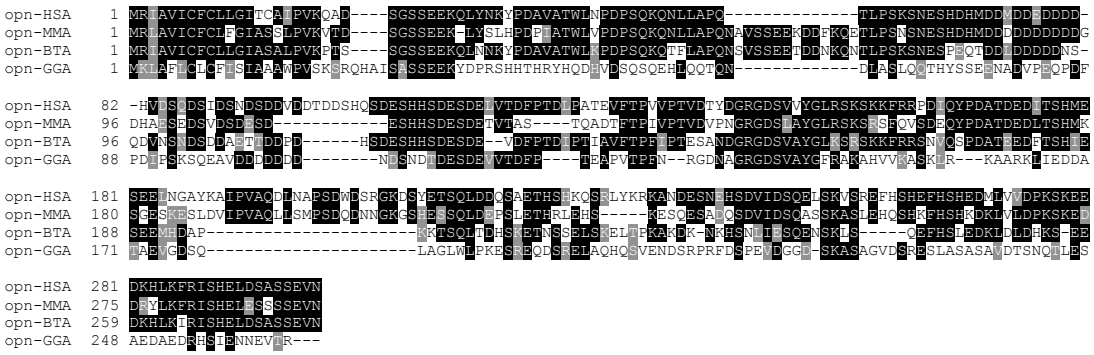


Figure 4.5 Alignment for *OPN* amino acid sequences of different species
Sequence alignment of *OPN* from Homo sapiens (OPN-HSA) (Acc. NP_000573.1, NCBI), Bos taurus (OPN-BTA) as translated from predicted mRNA sequence, Gallus gallus (OPN-GGA) (ENSGALP00000017754, www.ensembl.org) and Mus musculus (OPN-MMU) (Acc. NP_033289.1, NCBI). Numbers represent protein residues. Black and grey marked fonts show identical and conservative amino acids, respectively.

The putative bovine transcription start was predicted applying the open source software *Promoter Prediction* (http://www.fruitfly.org/seq_tools/promoter.html). The larger font in Figure 4.6 shows the putative transcription start, which was identical to the predicted transcription start by *GenomeThreader* (Gremme *et al.* 2005). The transcription start in humans is almost identical with the observed result in cattle, suggesting a high sequence similarity between humans and cattle within this region.

Start	End	Score	Promoter Sequence
161	211	0.97	

TGGAGCAGCCTTTAAATTCTGGGAGATCCTGGTTGTCAGC**A**GCAGGGAGA

Figure 4.6 Transcription start site prediction of bovine *OPN*
The putative transcription start site is shown by heightened font. Start and end position is relative to the 3' end of exon 1. The score value indicates the prediction accuracy with values ranging from minimal 0 to maximal 1.

The identification of CpG-islands using *cpgplot* at (http://www.ebi.ac.uk/emboss/cpgplot/) (Larsen *et al.* 1992) failed in cattle as well as in human.

The analyses for putative transcription factor binding sites within the bovine and human 5' region demonstrated partially a sequence similarity higher than 80%. Three putative transcription factor binding sites were postulated by the *Consite* tool (see Figure 4.7), along with two SP1 binding sites and one CAAT enhancer binding protein (cEBP). One SP1

binding site starts at 179 bp, the other at 198 bp upstream of the 3' splice site of exon 1. The CAAT enhancer binding protein (cEBP) is located 306 bp upstream of the 3' splice site of exon 1 (see Figure 4.7).

The observation of the 3' UTR in humans and cattle with regard to miRNA binding sites did not lead to a successful identification.

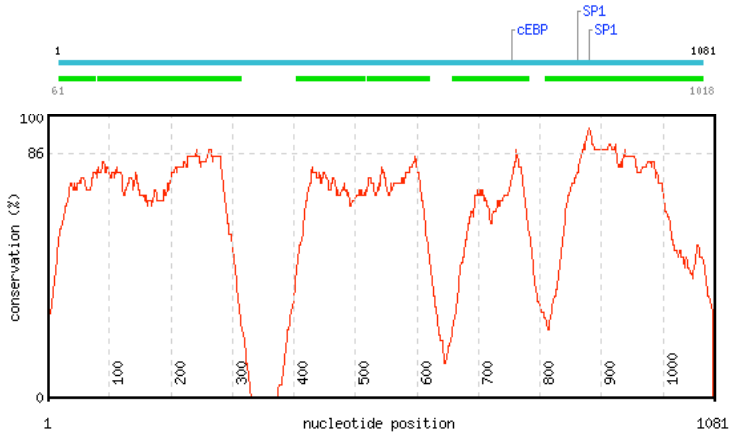


Figure 4.7 Conservation profile of the aligned bovine and human 5'-region of *OPN*
The blue line represents bovine sequence; the green, human sequence. The red line shows sequence identity between human (blue line) and bovine sequence (green line). Transcription factor binding sites located in conserved regions (≥ 86% identity) are displayed. SP1 represents the SP1-binding site, cEBP represents the CAAT enhancer binding protein site.

Based on the results from transcription factor site prediction, *OPN* can be assigned to the group of genes regulated by transcription factor binding sites. Genes belonging to this category usually show high tissue specific gene expression (Carninci *et al.* 2006).

4.3.3. Peroxisome proliferator-activated receptor gamma coactivator 1-alpha gene (*PPARGC1A*)

PPARGC1A encodes a member of the nuclear receptor superfamily PGC-1 (Muller *et al.* 2003). All three members of this family, *PPARGC1A*, *PRC* and *PGC-1β* have a related modular structure. *PPARGC1A* is a transcription coactivator interacting with a broad range of transcription factors that are involved in a wide variety of biological responses including mitochondrial biogenesis, glucose and fatty acid metabolism as well as in adaptive thermogenesis (LIANG and WARD 2006). *PPARGC1A* is expressed in tissues of high metabolic activity, such as liver, heart and muscle fibre (Franks *et al.* 2007) (HANDSCHIN and SPIEGELMAN 2006). Recently strong evidence for a mitochondrial dysregulation in type 2 diabetes has been found, and the identified influencing genes are transcriptional targets of

PPARGC1A and *PGC-1β* (Song *et al.* 2004) (Zhang *et al.* 2004) (HANDSCHIN and SPIEGELMAN 2006). Lactation requires a very complex regulation of glucose metabolism, so that PPARGC1A lends itself as a functional candidate gene for lactation related traits.

The local BLAST search using *Btau 3.1* as subject sequence positioned *PPARGC1A* on BTA6 at about 41.0 Mb. A QTL with effects on fat- and protein yield was previously mapped within a chromosomal interval bracketed by markers TGLA37 and FBN13 (Kuhn *et al.* 1999). The identification of these markers in *Btau 3.1* mapped TGLA37 on BTA6 at 46.1 Mb. Finally, *PPARGC1A* is 5.1 Mb proximally to TGLA37 supporting the role of *PPARGC1A* as a positional candidate gene.

The candidate gene hypothesis has already been tested and a previous association study detected a significant association between an SNP in intron 9 and milk fat yield (Weikard *et al.* 2005).

Both annotations, the manual annotation by BLAST and the semi-automated annotation with *GenomeThreader* (Gremme *et al.* 2005) detected all human exons in the bovine genomic sequence. The whole coding sequence of *PPARGC1A* counts 103.1 kb in cattle and 98.26 kb in humans. *PPARGC1A* is composed of 13 exons in both species (see Figure 4.8 and Table 4.8). The translated sequence in cattle includes 2388 bp, corresponding to a protein sequence of 796 residues. Characteristic for the intron / exon structure of *PPARGC1A* is a large exon 2, which encompasses 53.0 kb in humans and 59.1 kb in cattle. There is also a remarkably big difference in the exon size. The shortest exon is exon comprising 46 nucleotides in humans and cattle. Exon 8 encompasses 919 nucleotides and human assembly 916 nucleotides. Start codon of *PPARGC1A* was identified in the first exon, stop codon triplet in exon 13.

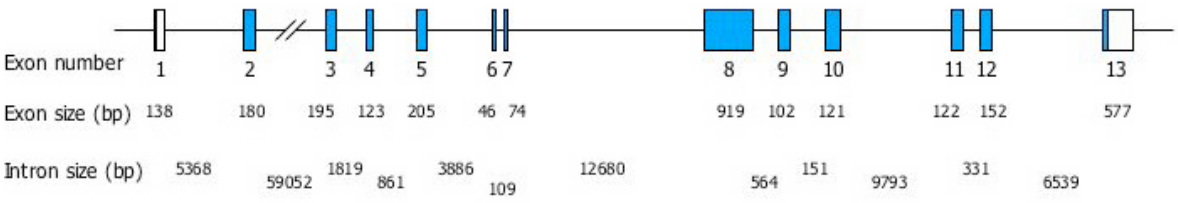


Figure 4.8 Graphical overview of the bovine *PPARGC1A*

The blue bars exhibit translated sequence. White bars exhibit untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in *Btau 3.1* is given below the bars.

Table 4.8 Intron / exon organisation of *PPARGC1A* in *Btau 3.1*

No.	Exon Size (bp)	Position in cDNA	Position on BTA6 (bp)	3' splice acceptor ... 5' splice donor [a]	No.	Intron Size (bp)
1 [b]	138	1 - 137	41242911 to 41242774	ctc**tc**AGTAA … TCGAG**gt**gag	1	5362
2	180	138 - 318	41237412 to 41237233	tgc**ag**TGTGC … TTGAG**gt**aag	2	59052
3	195	319 - 514	41178181 to 41177987	tgc**ag**AAGAT … CTCTA**gt**aag	3	1819
4	123	515 - 638	41176168 to 41176046	tct**ag**CTTAA … TTAAG**gt**aca	4	861
5	205	639 - 844	41175185 to 41174981	aga**ac**GAGAA … ACAAG**gt**agg	5	3886
6	46	845 - 891	41171094 to 41171049	ttt**ag**CCAAA … CCAAA**gt**aag	6	109
7	74	892 - 966	41170939 to 41171866	tcc**ag**TGACC … TGCAG**gt**aag	7	12680
8	919	967 - 1886	41159185 to 41158267	ctt**ag**GCCTA … TCAAG**gt**aaa	8	564
9	102	1887 - 1989	41157702 to 41157601	ctc**ag**ATCTT … CCCAG**gt**aat	9	151
10	121	1990 - 2111	41157449 to 41157329	ctc**ag**GTATG … CAATT**gt**aag	10	9793
11	122	2112 - 2234	41147535 to 41147414	tcc**ag**GAAGA … GATGG**gt**gag	11	331
12	152	2235 - 2387	41147082 to 41146931	tgc**ag**AGACA … CCTAG**gt**atg	12	6539
13 [c]	577	2388 - 2965	41140391 to 41139825	ttc**ag**ATTCA … AAGGA**aa**aaa		

[a] Intron sequence is indicated by lowercase letters, exon sequence by capital letters, the splice site sequence is in bold print. Information follows *GenomeThreader* annotation (see 4.2.2 Annotation of candidate genes).
[b] Exon contains start codon.
[c] Exon contains stop codon.

Figure 4.9 shows the bovine *PPARGC1A* amino acid sequence aligned with the peptide sequence of humans, cattle, mice and chickens. The peptide sequences show substantial similarity between each of the four observed species. Each observed species showed a similarity higher than 90% with the human ortholog.

However, the promoter prediction with the online tool did not result in an unambiguous identification of a transcription start site http://www.fruitfly.org/seq_tools/promoter.html. In contrast, the semi-automated gene annotation identified a transcription start site, which corresponds with the human transcription start site (Gremme *et al.* 2005).

5' end sequence of 1 kb length was analysed for CpG islands with the programme *cpgplot* from the Emboss web-service (http://www.ebi.ac.uk/emboss/cpgplot/) (Larsen *et al.* 1992). Human and bovine 5' end sequences were analysed. A successful prediction was not achieved in humans or cattle.

The analyses for transcription factor binding sites marked four predicted sites. The bovine region illustrates a substantial sequence similarity to the human orthologue and subsets of the regarded region showed almost complete identity. Therefore the conservation cut-off was set at 99% during the prediction of transcription factor binding sites. Three sites were predicted to be SP1 binding sites; the fourth was identified as a TATA-box binding site (TBP) (see Figure 4.10). One SP1 binding site was marked in *Consite* within the 5' UTR region, at nucleotide position 5 of the first exon. The other two lay in 5' end sequence 168 nucleotides and 202 nucleotides, respectively upstream of the 3' splice site of exon 1. The TATA-box lay 191 nucleotides upstream of the 3' splice site of exon 1.

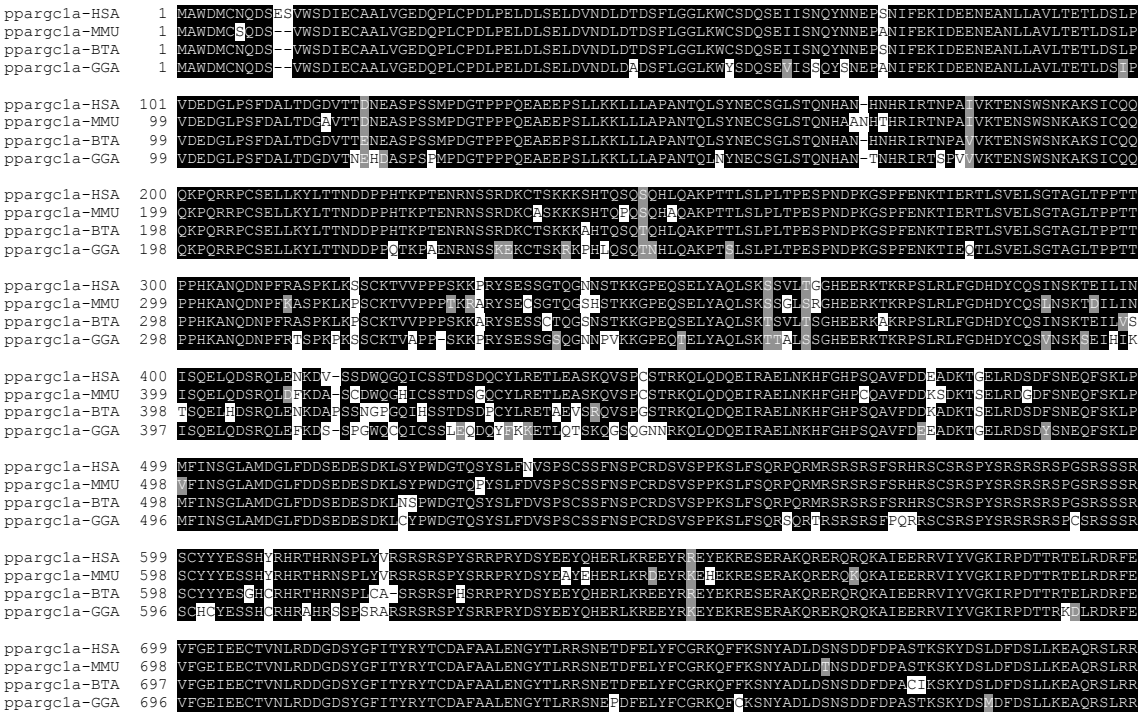


Figure 4.9 Alignment for *PPARGC1A* amino acid sequences of different species

Sequence alignment of *PPARGC1A* from Homo sapiens (PPARGC1A-HSA) (Acc. NP_037393.1, NCBI), Bos taurus (PPARGC1A -BTA) as translated from predicted mRNA sequence, Gallus gallus (PPARGC1A -GGA) (Acc. NP_001006457.1, NCBI) and Mus musculus (PPARGC1A -MMU) (Acc. NP_032930.1, NCBI). Numbers represent protein residues. Black and grey marked fonts show identical and conservative amino acids, respectively.

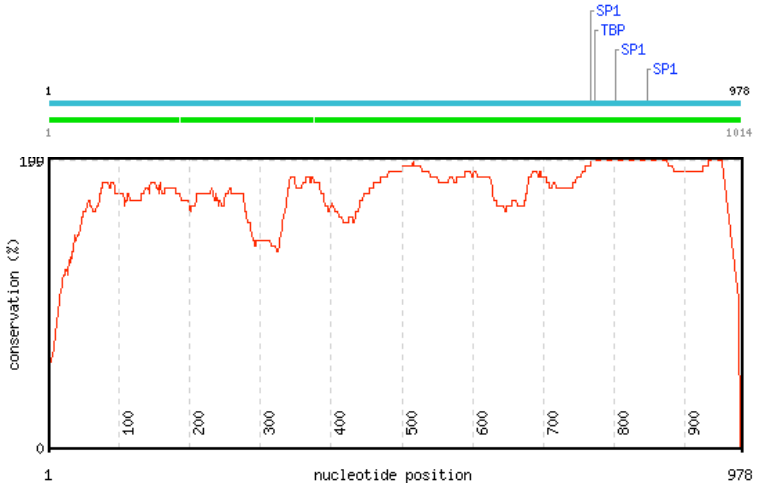


Figure 4.10 Conservation profile of the aligned bovine and human 5'-region of *PPARGC1A*

The blue line represents bovine sequence, the green human sequence. The red line shows sequence identity between human (blue line) and bovine sequence (green line). Transcription factor binding sites located in conserved regions (≥ 99% identity) are displayed. SP1 represents the SP1-binding site, TBP represents the TATA box binding protein site.

A miRNA binding site has not been detected in the human and bovine 3' sequence.
In conclusion, *PPARGC1A* belongs to the group of genes with highly conserved promotor sequence and transcription factor binding site regulated gene expression (Carninci *et al.* 2006).

4.3.4. Phosphoglucomutase-2 gene (*PGM2*)

In humans, four members represent the family of phosphoglucomutase (PGM) enzymes: *PGM1*, *PGM2*, *PGM3* and *PGM5* (Whitehouse *et al.* 1998) (Edwards *et al.* 1995). The PGM enzymes initially were known in humans for their enzymatic activity while a genomic isolation by molecular and immunological approaches failed for *PGM2* and PGM3. However, *PGM2* and *PGM3* were identified later by a phylogenetic approach using 47 highly divergent prokaryotic and eukaryotic PGM-like database sequences (Whitehouse *et al.* 1998). PGMs have key functions in glycogen metabolism and gluconeogenesis. Lactose is a disaccharide that consists of β-D-galactose and β-D-glucose molecules bonded through a β1-4 glycosidic linkage. PGM enzymes are involved in lactose synthesis by the PGM2-catalysed interconversion of glucose 1-phosphate to glucose 6-phosphate (Levin *et al.* 1999). *PGM2* is expressed in all organisms from bacteria to plants and animals (Whitehouse *et al.* 1998).
BLAST searches against *Btau 3.1* mapped *PGM1* on BTA3, *PGM3* on BTA9 and *PGM5* on BTA8 but *PGM2* harbours in sequence that is chromosomally unanchored. However, in sequence of *Btau 2.0, PGM2* was identified to anchor on BTA6. As already mentioned in 4.2.1 Positioning of candidate genes with bovine draft sequences, the most likely position of PGM2 is presumed to be at approximately 52.0 Mb on BTA6. QTL mapping studies identified a QTL for fat- and protein yield bracketed by markers TGLA37 and FBN13 (Kuhn *et al.* 1999), which corresponds in *Btau 3.1* to the BTA6-localisation at 46.1 Mb. Accepting the *PGM2*-position on BTA6 at 52.0 Mb, it is located distally to this QTL.
With their fundamental role in metabolic pathways of energy metabolism, the genes of the PGM family in general can be considered as functional candidate genes for milk synthesis. Among all *PGMs*, *PGM2* is the most promising candidate gene since it is located on BTA6 within a region that has been identified as harbouring a QTL for milk production traits.

PGM2 was annotated manually by local BLAST searches, but also semi-automatically by using the *GenomeThreader* algorithm (Gremme *et al.* 2005). The entire gene spans 36.6 kb in cattle and 36.4 kb in humans. In both species, *PGM2* counts 14 exons (see Figure 4.11 and

Table 4.9). The translated sequence in cattle includes 1812 bp, corresponding to a protein sequence of 604 residues. Exon 1 contains the translation start site; translation stop codon was identified within exon 14.

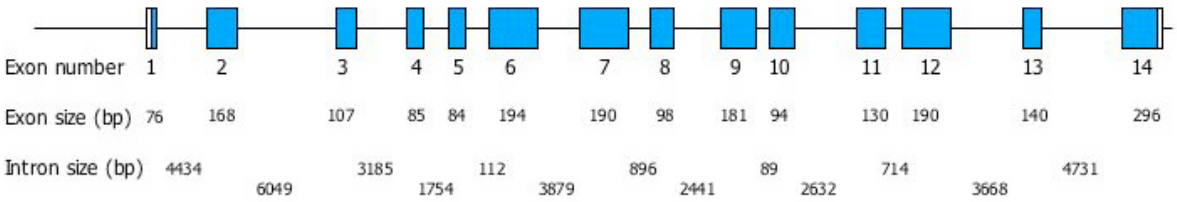


Figure 4.11 Graphical overview of the bovine *PGM2*
The blue bars represent translated sequence. White bars represent untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in *Btau 3.1* is given below the bars.

Table 4.9 Intron / exon organisation of *PGM2* in *Btau 3.1*

No.	Exon Size (bp)	Position in cDNA	Position on Chr.Un.003.335 [d] (bp)	3' splice acceptor ... 5' splice donor [a]	No.	Intron Size (bp)
1 [b]	76	1 - 76	113804 to 113728	acg**ag**GGCTC … ACAAG**gt**gag	1	4434
2	168	77 - 245	109292 to 109126	tgc**ag**AATCC … CACAG**gt**acc	2	6049
3	107	246 - 353	103076 to 102966	ttt**ag**GGATT … AGAAG**gt**acg	3	3185
4	85	354 - 439	99784 to 99700	tcc**ag**ATTTG … TTGTG**gt**aag	4	1754
5	84	440 - 524	97945 to 97862	tac**ag**CCCTA … ATAAG**gt**att	5	112
6	194	525 - 719	97749 to 97556	tcc**ag**GTCTA … CACAG**gt**aaa	6	3879
7	190	720 - 910	93676 to 94487	tct**ag**GGCTG … TCTTG**gt**aac	7	896
8	98	911 - 1009	92591 to 92494	ttc**ag**ACTTT … GACAG**gt**aca	8	2441
9	181	1010 - 1191	90053 to 89873	ttc**ag**TGGCA … TCGAG**gt**ata	9	89
10	94	1192 - 1286	89784 to 89690	ttt**ag**GAAAC … TATTG**gt**aag	10	2632
11	130	1287 - 1417	87057 to 86928	tgc**ag**GATAT … GTTGA**gt**aag	11	714
12	190	1418 - 1608	86213 to 86020	gat**ag**GTATG … AAGCT**gt**aag	12	3668
13	140	1609 - 1749	82355 to 82216	cct**ag**ATTCT … TATGG**at**ggc	13	4731
14 [c]	296	1750 - 2046	77484 to 77188	tgc**ag**TGATC … TCACA**ac**tgg		

[a] Intron sequence is indicated by lowercase letters, exon sequence by capital letters, the splice site sequence is in bold print. Information follows *GenomeThreader* annotation (see 4.2.2 Annotation of candidate genes).
[b] Exon contains start codon.
[c] Exon contains stop codon.
[d] Scaffold chromosomally unanchored in *Btau 3.1*.

Pair-wise comparisons between humans and the other investigated species resulted in similarity levels of more than 90%. As shown in Figure 4.12, differences between species are almost exclusively located in the region of the peptides' N termini.

The prediction of the transcription start site with the online program failed but the result corresponds with the low similarity in the N termini of the peptide. In humans, two different transcription start sites are described (http://www.ensembl.org). One shows substantial similarity with the predicted transcription start site identified in the *GenomeThreader* annotation. Therefore, and based on the failed prediction with Promoter Prediction, one can assume that several transcription start sites for PGM2 exist in the bovine genome as well.

Human and bovine 5' sequences of *PGM2* were analysed for CpG islands with the programme *cpgplot* from the Emboss web-service (http://www.ebi.ac.uk/emboss/cpgplot/) (Larsen *et al.* 1992). The prediction resulted in the identification of a putative CpG-island in the bovine 5' end sequence (see Figure 4.13). The predicted CpG-island encompasses 231 nucleotides, and its genomic position starts 105 nucleotides upstream of the 3' splice site of exon 1. The prediction in the bovine genomic sequence found confirmation in the complementary region of humans, where a CpG-island with a length of 259 nucleotides was identified.

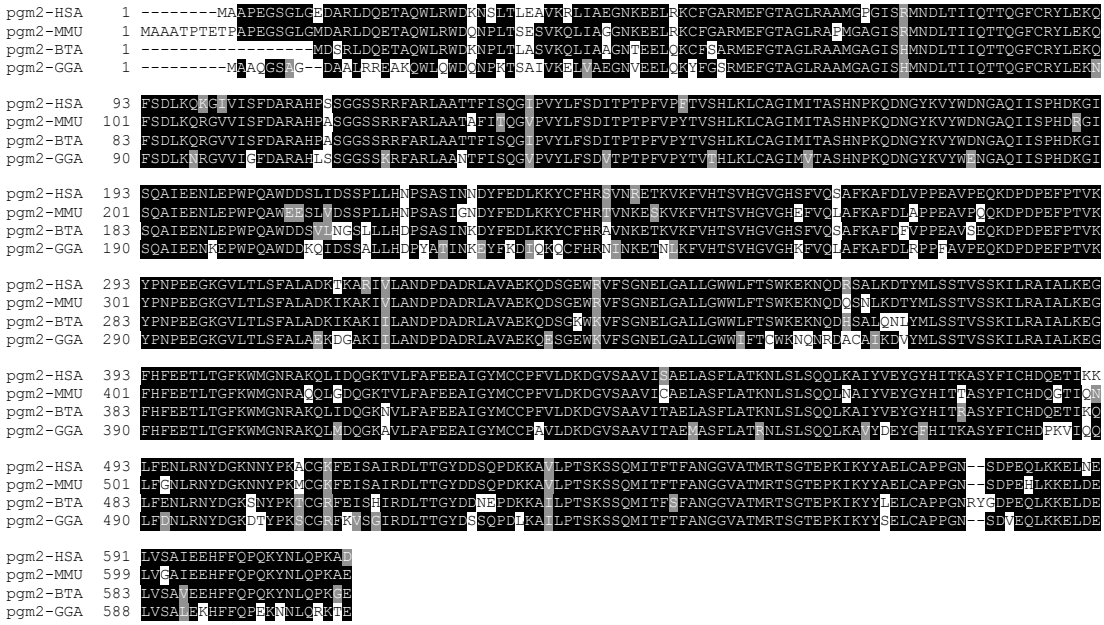


Figure 4.12 Alignment for *PGM2* amino acid sequences of different species
Sequence alignment of *PGM2* from Homo sapiens (PGM2-HSA) (Acc. NP_060760.2, NCBI), Bos taurus (PGM2-BTA) as translated from predicted mRNA sequence, Gallus gallus (PGM2-GGA) (Acc. XP_001236252.1, NCBI) and Mus musculus (PGM2-MMU) (Acc. NP_082408.2, NCBI). Numbers represent protein residues. Black and grey marked fonts show identical and conservative amino acids, respectively.

During the prediction of putative transcription factor sites, a low sequence similarity was observed between the *PGM2*-promoter region of humans and cattle. Only short sequence fragments reach a similarity of more than 80%. Consequently, it was not possible to assign putative transcription factor binding sites in the submitted sequence.

The search for miRNA binding sites (Clop *et al.* 2006) in the human and bovine 3' UTR region was without success.

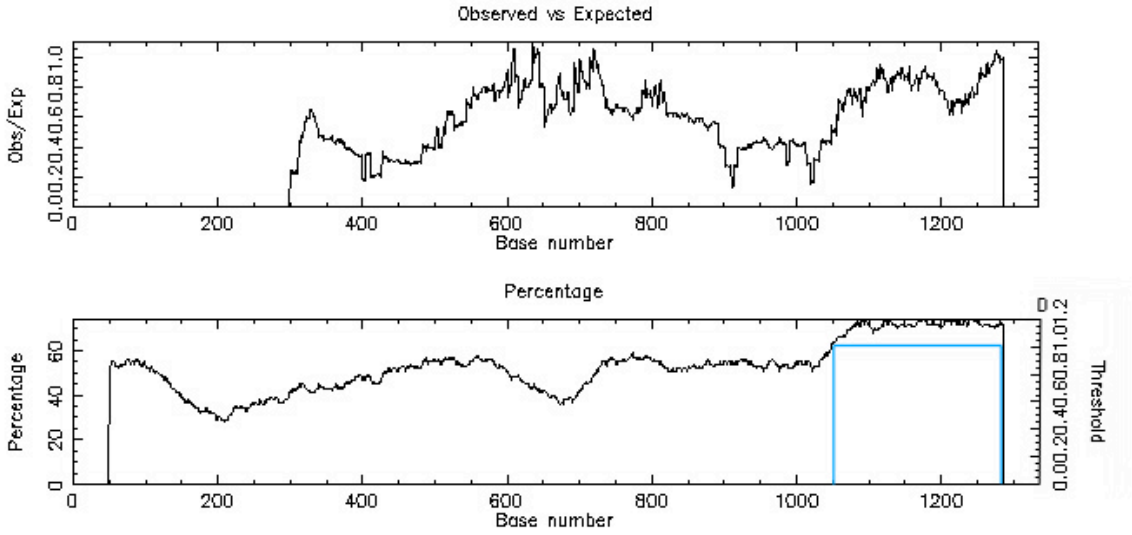


Figure 4.13 CpG island prediction for *PGM2* in cattle
Upper box represents ratios of observed versus expected CpG content. The lower box shows the relative C+G content and the predicted CpG-island by the blue box. Horizontal axes show nucleotide numbers in relation to the 3' end of first exon.

Conclusively all obtained results from the genomic characterisation underline *PGM2* as a gene that can be allocated in the class of CpG-rich genes (Carninci *et al.* 2006). The 5' sequence of *PGM2* showed low similarity between the bovine and human species. No cis-elements were predicted. In contrast, the CpG island prediction in the bovine genome was confirmed in the 5' end sequence of humans and mice. Therefore it is presumable, that *PGM2* does not show a tissue specific regulation of the gene expression (Carninci *et al.* 2006).

4.3.5. Alpha-s1-casein gene (*CSN1S1*)

Six proteins, which together comprise 95% of the total protein in bovine milk, have been classified according to their biochemical characteristics into caseins and whey proteins. Four different casein genes have been described: *CSN1S1* encodes αS1-casein and *CSN1S2* the αS2-casein. β-casein is encoded by *CSN2* and κ-casein by *CSN3* (THREADGILL and WOMACK 1990). As the caseins represent the main part of the bovine milk protein fraction, their common basic function is the supply of the offspring with amino acids. All casein peptides are characterised by low solubility at pH 4.6 and calcium dependent precipitation. Caseins are organised by clusters of protein chains, called micelles and show similar sites of phosphorylation (Ramunno *et al.* 2004). This information supports the hypothesis of a common evolutionary root (Rijnkels 2002) and a development by gene duplication. All milk proteins, apart from colostral immunoglobulins, are synthesised de novo in the mammalian

epithelial tissue, which explains the expression of all casein genes in the mammary gland (Kuss *et al*. 2005). Milk casein consists of 35 to 45% αS1-casein, so that *CSN1S1* encodes the major part of milk casein.

Due to a strong interest in production traits in dairy cattle, the four casein genes were the subject of early molecular genetic approaches. In situ hybridisations mapped the casein genes to BTA6 at q31 – 33 (THREADGILL and WOMACK 1990) and the four casein genes were reported to reside on less than 300 kb on BTA6 (THREADGILL and WOMACK 1990). BLAST searches and *Btau 3.1* identified *CSN1S1* on BTA6 at 79 Mb approximately. With *GenomeThreader* (Gremme *et al*. 2005), each of the casein genes was annotated within the chromosomal fragment of BTA6 bracketed between 79.4 Mb and 79.9 Mb. Hence, the results of Threadgill et. al (1990) were confirmed, while the region on BTA6 where the caseins are encoded could be accurately measured at precisely 242.9 kb. Of all casein genes, *CSN3* occupies the most proximal position. In distal orientation, *CSN3* is followed by CSN1S2 at a distance of 110 kb and more distally *CSN2* follows at a distance of approximately 80 kb. The most distal position on BTA6 of all casein genes is taken by *CSN1S1*, which is located in a distal distance of approximately 40 kb to *CSN1S2*. Several QTLs were mapped to this chromosomal region and QTLs were mostly identified for protein content (ASHWELL and VAN TASSELL 1999; BOICHARD *et al*. 2003; MOSIG *et al*. 2001; VELMALA *et al*. 1999). However, QTL affecting yield traits near the casein region were also reported (Wiener *et al*. 2000) (Nadesalingam *et al*. 2001).

The gene annotations by BLAST and *GenomeThreader* (Gremme *et al*. 2005) showed similar gene structures for all casein genes with several small exons and a low intron / exon ratio. The entire bovine *CSN1S1* extends over a chromosomal distance of 17.5 kb. The human *CSN1S1* encompasses 15.7 kb. The bovine translated sequence of *CSN1S1* is assembled by 642 bp, which corresponds to a peptide sequence of 214 residues. The size in cattle is comparable to the human orthologue, where the coding sequence is assembled by 561 bp and the peptide consists 187 residues. In cattle, the coding sequence of *CSN1S1* is partitioned into 19 exons (see Figure 4.14 and Table 4.10) and in humans into 15 exons. In cattle as well as in humans, the first exon is untranslated sequence and the start codon has been identified in exon 2. Characteristic for *CSN1S1* in humans and in cattle is the structure of the stop codon: the final nucleotides TG of exon 17 are coupled with the first nucleotide A of exon 18.

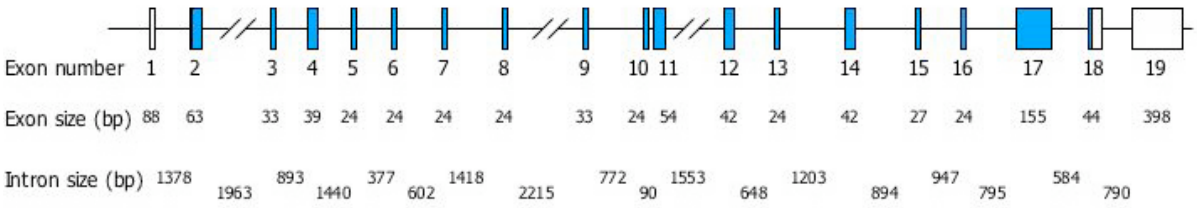


Figure 4.14 Graphical overview of the bovine *CSN1S1*
The blue bars represent translated sequence. White bars represent untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in Btau 3.1 is given below the bars.

Table 4.10 Intron /exon organisation of *CSN1S1* in *Btau 3.1*

No.	Exon Size (bp)	Position in cDNA	Position on BTA6 (bp)	3' splice acceptor ... 5' splice donor [a]	No.	Intron Size (bp)
1	88	1 - 88	79769237 to 79769150	taa**tc**AGTAG … TCAAG**gt**att	1	1378
2 [b]	63	89 - 152	79767771 to 79767709	cat**ag**ATCTT … GGCCT**gt**gag	2	1963
3	33	153 - 168	79765745 to 79765713	aac**ag**AAACA … CTCAA**gt**aag	3	893
4	39	169 - 226	79764819 to 79764781	tac**ag**GAAGT … TGGCA**gt**aag	4	1440
5	24	227 - 251	79763340 to 79764718	tgt**ag**CCTTT … GAAAG**gt**aag	5	377
6	24	252 - 276	79764340 to 79763652	tta**ag**GAGAA … GCAAG**gt**aag	6	602
7	24	277 - 301	79763049 to 79763026	ggc**ag**GATAT … CTGAG**gt**aag	7	1418
8	24	302 - 326	79762233 to 79762310	cta**ag**GATCA … TTAAG**gt**aag	8	2215
9	33	327 - 360	79760810 to 79760778	ttt**ag**CAAAT … GTGAG**gt**ata	9	772
10	24	361 - 385	79760005 to 79759982	tta**ag**GAAAT … TTGAG**gt**gag	10	90
11	54	386 - 440	79759891 to 79759838	ttt**ag**CAGAA … ATCTG**gt**aaa	11	1533
12	42	440 - 483	79758305 to 79758264	aat**ag**GAACA … AGCTG**gt**aat	12	648
13	24	484 - 508	79757615 to 79757592	tta**ag**GAAAT … CTGAG**gt**gag	13	1203
14	42	509 - 551	79756388 to 79756347	ttt**ag**GAACG … AACAG**gt**aat	14	894
15	27	552 - 579	79755452 to 79755426	tcc**ag**AAAGA … ATCAG**gt**aag	15	947
16	24	580 - 604	79754479 to 79754456	tta**ag**GAACT … CTGAG**gt**gaa	16	795
17 [c]	155	605 - 760	79753660 to 79753505	tct**ag**CTTTT … TGGTG**gt**aag	17	584
18	44	761 - 805	79752920 to 79752877	ttc**ag**AAGAG … CTTTG**gt**aag	18	790
19	398	806 - 1204	79752086 to 79751689	tac**ag**ATGGT … GCACT**at**tcc		

[a] Intron sequence is indicated by lowercase letters, exon sequence by capital letters, the splice site sequence is bolded. Information follows *GenomeThreader* annotation (see 4.2.2 Annotation of candidate genes).
[b] Exon contains start codon.
[c] Exon contains stop codon.

The similarity analysis of the amino acid sequence between species was carried out without the galline species since chicken are avian. The detected level of interspecies peptide sequence conservation was surprisingly small. Only the first residues of the peptide revealed a high level of similarity (see Figure 4.15). Pair-wise species comparisons detected levels of similarity of around 40%. The lowest similarity level, 39.5%, was observed between cattle and mice. The alignment between the human and bovine amino acid sequence was similar at 44.0% and the alignment between humans and mice was 41.3%.

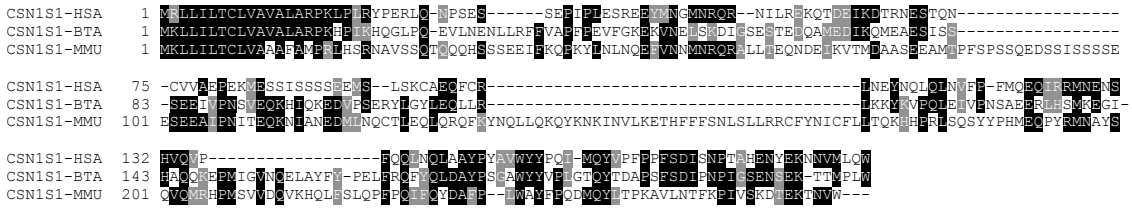


Figure 4.15 Alignment for *CSN1S1* amino acid sequences of different species
Sequence alignment of *CSN1S1* from Homo sapiens (CSN1S1-HSA) NP_001881.1, Bos taurus (CSN1S1-BTA) as translated from predicted mRNA sequence and Mus musculus (CSN1S1-MMU) NP_031810.1. Numbers represent protein residues. Black and grey marked fonts show identical and conservative amino acids, respectively.

The web-based tool *Promoter Prediction* http://www.fruitfly.org/seq_tools/promoter.html was used to identify the putative transcription start site. The genomic sequence of a length of 1 kb upstream of 3' splice site of the first exon was examined. Larger font in Figure 4.16 shows the putative transcription start at position 59. The result from transcription start site prediction with *GenomeThreader* (Gremme *et al.* 2005) was identified within the same region, but was not identical. However, the *GenomeThreader* transcription start site corresponds to the human site (http://www.ensembl.org).

Start	End	Score	Promoter Sequence
18	68	0.72	

TCAGTAGGTTTAAATAGCTTGGAAGCAAAAGTCTGCCATC**A**CCTTGATCA

Figure 4.16 Transcription start site prediction of bovine *CSN1S1*
The putative transcription start site is shown by heightened font. Start and end position are relative to the 3' end of exon 1. The score value indicates the accuracy of the prediction with values ranging from minimum 0 to maximum 1.

CSN1S1 5' end sequence of cattle, humans and mice were screened for the presence of a CpG-island with *cpgplot*, but no CpG island was identified.

Afterwards, 1 kb upstream of the 3' splice site of exon 1 was screened for transcription factor binding sites. A low sequence similarity between species, which was already seen in the amino acid alignment, was also observed for the genomic sequence of the 5' region. However, it was possible to identify a TATA-box (TBP) (see Figure 4.17) in the bovine 5' end sequence. It is located at a distance of 128 nucleotides upstream of the 3' splice site of exon 1. In addition, the bovine 5' end showed a 16-bp Milk Box motif that was previously described in goats. It is positioned with reference to the 3' splice site of exon 1 at a distance of 194 nucleotides (Ramunno *et al.* 2004) (Schmitt-Ney *et al.* 1991).

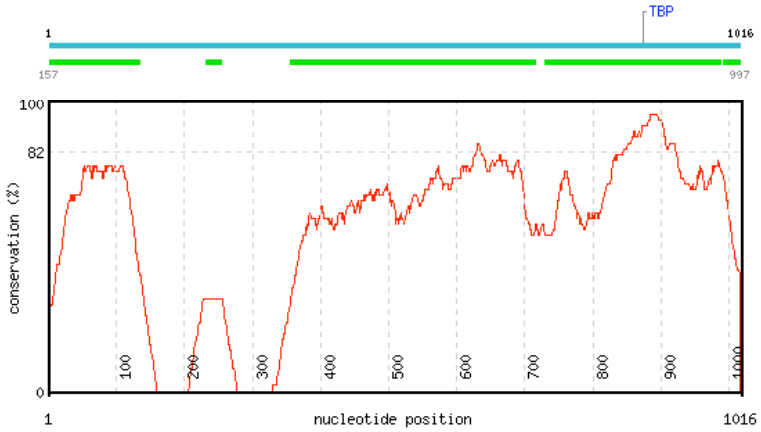


Figure 4.17 Conservation profile of the aligned bovine and human 5'-region of *CSN1S1*
The blue line represents bovine sequence, the green, human sequence. The red line shows sequence identity between human (blue line) and bovine sequence (green line). Transcription factor binding sites located in conserved regions (≥ 82% identity) are displayed. TBP represents the site of the TATA-box binding protein

The 3' UTR of humans and cattle were scanned for miRNA binding sites (Clop *et al.* 2006), but no site was detected.

However the amino acid similarity between species was observed to be low, based on the results from transcription factor binding site prediction, *CSN1S1* can be assigned to the class of genes regulated by TATA-rich promoter regions. Genes regulated by TATA-enriched promoter regions are usually known for a high tissue specific regulation of the gene expression (Carninci *et al.* 2006). This seems to be obvious due to the mammary specific gene expression (Schmitt-Ney *et al.* 1991). The interspecies differences that were detected by the peptide alignment lead to the expectation that the amino acid constitution of CSN1S1 is species specific.

4.3.6. Facilitated glucose transporter member 9 gene (*SLC2A9*)

SLC2A9 is a member of the facilitated transporter protein family 2. Proteins of this family have 12 putative transmembrane domains and carry two highly conserved motifs between domain 2 and 3 and between domain 8 and 9 (Phay *et al.* 2000). Glucose, the crucial precursor for lactose, crosses biologic membranes in mammals through two families of glucose transporters. It is transported either by active transporters, which establish a sodium co-transport induced concentration gradient (HEDIGER and RHOADS 1994) or passively (Mueckler 1994) (Thorens 1996). SLC2A9 belongs to the passive transporters (ZHAO and KEATING 2007). *SLC2A9* was shown to be expressed in various tissues, like liver, kidney and placenta (Augustin *et al.* 2004) (Richardson *et al.* 2003). In induced diabetic mice, the

expression of *SLC2A9* in kidney and liver tissue increased significantly compared to non-diabetic animals (Keembiyehetty *et al.* 2006).

Local BLAST searches with *Btau 3.1* did not result in a precise localisation on BTA6, but the region distal to 95 Mb is most probable (see 4.2.1 Positioning of candidate genes in bovine draft sequences), so that the closest QTLs for milk production traits are those mapped to markers BP7 (Wiener *et al.* 2000), BM415 (Mosig *et al.* 2001), INRAK (Boichard *et al.* 2003) and to the casein gene cluster (Velmala *et al.* 1999). These markers correspond in *Btau 3.1* to a genomic sequence on BTA6 from 67.3 Mb for marker BM415 to 84.5 Mb for marker BP7, so that *SLC2A9* is a functional candidate gene for milk related traits.

A two-sided gene annotation, manually by BLAST searches and semi-automated with *GenomeThreader* (Gremme *et al.* 2005) was applied. As already shown in 4.2.1 Positioning of candidate genes in bovine draft sequences, the annotation in *Btau 2.0* sequence was more complete than in *Btau 3.1*. However, exon 1 was identified neither by manual BLAST searches nor in the *GenomeThreader* annotation. Human exon 1 contains untranslated sequence, so that all translated sequence was detected in the bovine genome assembly. In cattle, translated sequence encompasses 1566 nucleotides, corresponding to a peptide sequence of 522 residues. Translated sequence of human *SLC2A9* counts 1582 nucleotides, and the peptide sequence 527 residues. A correct size - determination of all bovine introns was not possible due to the fact that exonic sequence of *SLC2A9* was identified on more than one Btau 2.0 scaffold. The bovine peptide is encoded by exons 2 to 13 (see Figure 4.18 and Table 4.11).

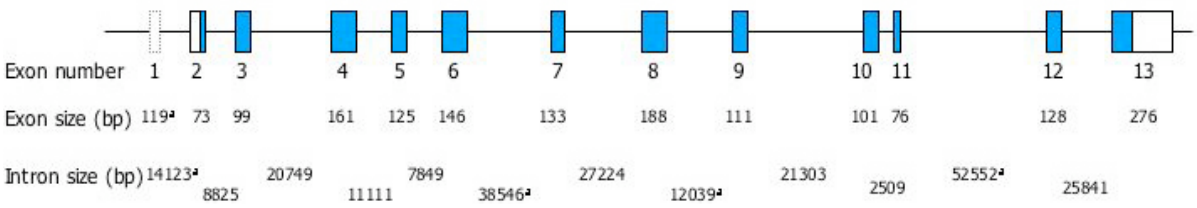


Figure 4.18 Graphical overview of the bovine *SLC2A9*

The blue bars represent translated sequence. White bars represent untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in *Btau 2.0* is given below the bars.

[a] The size information of intron 1, 6, 8 and 11 is from humans.

Sequence alignments of the amino acid sequence with human, bovine, murine and galline species resulted in conservation levels between 65.0% and 83.1%. The similarity between the species is shown by the *BoxShade* v.3.21 output in Figure 4.19. Among all investigated species, the bovine peptide shows the highest sequence similarity to the human peptide: 83.1%. Murine and galline peptide indicate lower similarities: 65.0% and 66.7%.

Table 4.11 Intron / exon organisation of *SLC2A9* in *Btau 2.0*

No.	Exon Size (bp)	Position in cDNA	Scaffold	Scaffold position (*Btau 2.0*)	3' splice acceptor ... 5' splice donor [a]	No.	Intron Size (bp)
1			Not identified			1	-
2 [b]	73	1 - 73	6.200	205408 to 205480	ggc**tg**ACAAG ... AACTG**gt**cag	2	8825
3	99	74 - 173	6.200	214306 to 214404	ccc**ag**GACTG ... CTCCG**gt**gat	3	20749
4	161	174 - 335	6.200	235154 to 235314	ttc**ag**TACAT ... GGGAG**gt**cag	4	11111
5	125	336 - 458	6.200	246426 to 246550	tcc**ag**GAAAA ... TGGAG**gt**gag	5	7849
6	146	459 - 605	6.174	254400 to 254546	cac**ag**GCATC ... GAAAG**gt**aag	6	[f]
7	133	606 - 739	6.174	27823 to 27968	ttc**ag**GAGAG ... GAAAG**gt**agg	7	27224
8	188	740 - 927	6.200	55193 to 55326	ggt**ag**CCTTC ... ATGCG**gt**gag	8	[f]
9	111	928 - 1039	6.200	376383 to 376493	tat**ag**ATTTG ... TCTCT**gt**aat	9	21303
10	102 [d]	1040 - 1141	6.200	397797 to 397866 [e]	ccc**ag**GGCTT ... TGCAG**gt**gag	10	2509
11	76	1142 - 1218	6.200	401408 to 401483	cac**ag**GATCA ... GCCAG**gt**aag	11	[f]
12	128	1219 - 1347	Un.2195	2712 to 2838	tgc**ag**GTGGC ... TTCAG**gt**gag	12	25841
13 [c]	276	1348 - 1624	Un.2195	28679 to 28955	cac**ag**CAAAG ... TTTTA**at**acc		

[a] Intron sequence is indicated by lowercase letters, exon sequence by capital letters, the splice site sequence is in bold print. Information follows *GenomeThreader* annotation (see 4.2.2 Annotation of candidate genes).
[b] Exon contains start codon.
[c] Exon contains stop codon.
[d] Length determination by resequencing due to incomplete anchorage in *Btau 2.0* (see 4.2.2 Annotation of candidate genes): Acc. EU295941, NCBI.
[e] *Btau 2.0* contains only 70 bp from the 5' part of exon 10.
[f] Size determination not possible due to adjacent exons anchored on different scaffolds.

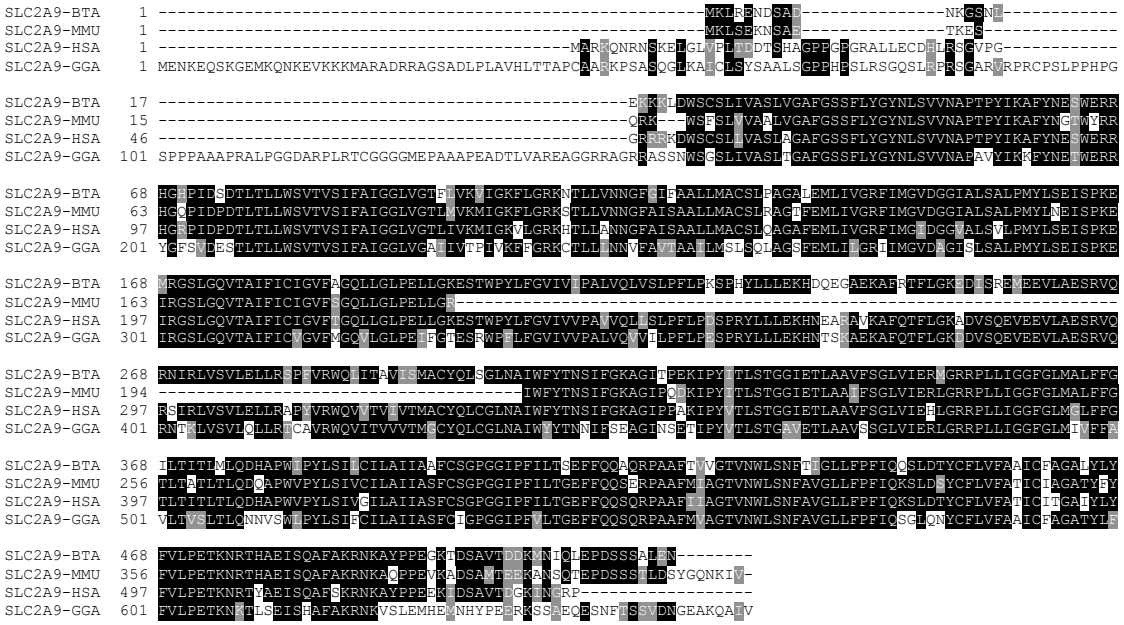


Figure 4.19 Alignment for *SLC2A9* amino acid sequences of different species
Sequence alignment of *SLC2A9* from Homo sapiens (SLC2A9-HSA) (Acc. NP_001001290.1, NCBI), Bos taurus (SLC2A9-BTA) as translated from predicted mRNA sequence, Gallus gallus (SLC2A9-GGA) (Acc. XP_420789.2, NCBI) and Mus musculus (SLC2A9-MMU) (Acc. NP_663534.1, NCBI). Numbers represent protein residues. Black and grey marked fonts show identical and conservative amino acids, respectively.

Since the first exon was not identified in the bovine genomic sequence, no genomic characterisation for the bovine 5' region was done.

The human and bovine 3' regions were scanned for miRNA binding sites (Clop *et al*. 2006). The web tool available at http://microrna.sanger.ac.uk/sequences/ was used for miRNA binding analyses, but searches in each observed species were without success.

4.3.7. Summary of candidate gene characterisation

ABCG2, *OPN*, *PPARGC1A*, *PGM2* and *SLC2A9* are involved in metabolic pathways that are related to milk synthesis (see Table 4.12). *CSN1S1* encodes a milk component, so that all selected genes are functional candidate genes for milk synthesis. Apart from *SLC2A9*, all selected genes are located within confidence intervals of QTLs for milk production traits, thus *ABCG2*, *OPN*, *PPARGC1A*, *PGM2* and *CSN1S1* are both functional and positional candidate genes for milk synthesis.

Varying data was obtained concerning the intron / exon ratio. In relation to the exonic sequence, *OPN* has the shortest introns. *PGM2* and *PPARGC1A* are characterized by relatively long intronic sequence.

Table 4.12 Functional and positional gene characterisation in terms of milk synthesis

Gene	Functional relevance		Positional relevance	
	Peptide function	Gene expression	Localisation on BTA6[a] (Mb)	QTL marker bracket and localisation on BTA6 [a] (Mb)
ABCG2	Transporter for ATP-related ligands	Mammary gland	33.5	BM1329 / BM143 [c] (26.6 to 40.5)
OPN	Ca^{2+} regulation, Mammary gland differentiation	Mammary gland	33.7	BM1329 / BM143 [c] (26.6 to 40.5)
PPARGC1A	Regulation of transcription factors	Liver, kidney, heart, muscle	41.0	BM143 / BM4528 [d] (46.1 to 58.5)
PGM2	Interconversion of glucose	Muscle	52.0 [b]	BM143 / BM4528 [d] (46.1 to 58.5)
CSN1S1	Milk protein component	Mammary gland	79.7	BM415 / AFR227 [d] (67.2 to 86.2)
SLC2A9	Glucose transport	Liver, kidney	97.0 [b]	[f]

[a] Localisation as identified by BLAST against *Btau 3.1*
[b] Approximate position on BTA6, see 4.2.1 Positioning of candidate genes in bovine draft sequences
[c] (Viitala *et al*. 2003)
[d] (Kuhn *et al*. 1999)
[e] (Viitala *et al*. 2003)
[f] Not located in a confidence interval of a QTL for milk production traits.

A large variation was obtained for the conservation of the peptide sequence between species (see Table 4.13). Three genes indicate a fundamental metabolic importance due to a high

peptide sequence similarity between species: *ABCG2*, *PPARGC1A* and *PGM2* (see Figure 4.2, Figure 4.9, Figure 4.12). The expression of *OPN*, *PPARGC1A* and *CSN1S1* is expected to be regulated by transcription factor binding sites and to be highly tissue specific (Carninci *et al*. 2006). The latter was reported for *CSN1S1* (VONDERHAAR and ZISKA 1989; YOSHIMURA and OKA 1990). miRNA binding sites have not been discovered.

Table 4.13 Summary of genomic characterisation of the candidate genes

Gene	No. of exons	Intron / exon ratio	Size of bovine peptide (No. of amino acids)	Peptide sequence similarity between HSA and BTA (%)	Regulatory elements in 5'- region
ABCG2	14	30.8	655	90.1	CpG island
OPN	7	4.1	278	69.1	TFBs [a]
PPARGC1A	13	34.5	796	96.6	TFBs [a]
PGM2	14	17.0	604	94.5	CpG island
CSN1S1	19	15.6	214	44.0	TFBs [a]
SLC2A9	13	[a]	522	83.1	[b]

[a] Transcription factor binding sites
[b] Prediction not possible, due to unidentified genomic region in bovine draft sequences

4.4. *Polymorphism analysis*

4.4.1. Identified polymorphisms

In all, 206 DNA variants, 191 SNPs and 15 INDELs were discovered. Identified DNA variants mostly have intronic position (n = 134) or are located upstream of the transcription start site (n = 40) (see Table 4.14). Translated regions of all six candidate genes contain 23 SNPs whereof *SLC2A9* harboured almost half of these (n = 11). No SNP was detected in translated sequence of *OPN*.

Table 4.14 Localisation of identified sequence variation

Gene	Upstream		5' UTR [a]		Translated region		Intronic region		3' UTR [a]		Down-stream		Σ
	SNP	INDEL	SNP	INDEL	SNP	INDEL	SNP	INDEL	SNP	INDEL	SNP	INDEL	
ABCG2	5	1	1	-	4	-	22	-	-	-	[d]		33
OPN	3	1	2	-	-	-	5	-	-	-	[d]		11
PPARGC1A	-	-	-	-	2	-	16	2	-	-	[d]		20
PGM2	10	2	-	-	5	-	34	4	1	-	[d]		56
CSN1S1	17	1	2	-	1	-	28	2	-	1	[d]		52
SLC2A9	[b]		[c]	[c]	11	-	20	1	-	-	2	-	34
Σ	35	5	5	-	23	-	125	9	1	1	2	-	206

[a] Untranslated region
[b] Exon 1 of *SLC2A9* could not be annotated
[c] 5' UTR of Exon 1 not screened for DNA variants
[d] Region not screened for DNA variants

In total 69.2 kb genomic sequence was re-sequenced during screening for DNA variants. A fragment with the size of 30 nucleotides from the tenth exon of *SLC2A9* was re-sequenced de novo, because *Btau 2.0* contained a short gap-region of unknown sequence, depicted by 220 *N*s'. The main focus during polymorphism screen was on the re-sequencing of sequence that was annotated in 4.2.2 Annotation of candidate genes as coding. Apart from *SLC2A9*, approximately 1.5 kb of sequence upstream of the transcription start site were also re-sequenced in each gene. 3' UTR region was screened partially. Downstream sequence of 3' UTR was screened exclusively in *SLC2A9*. Primers for the SNP screen of the coding regions were designed in intronic sequence, so that parts of intronic sequence were screened for DNA variants automatically. The entire intron was re-sequenced in case of short intronic sequence (see Table 4.15) since exons framing the intron were amplified in a single PCR product. The number of re-sequenced nucleotides per gene ranges from 7.0 kb for *SLC2A9* to 13.9 kb for *PGM2*. 2.97 DNA variants per 1 kb genomic sequence were identified on average across genes. Comparing the DNA-variant frequency between the genes divides the candidate genes into two categories. In *ABCG2*, *OPN* and *PPARGC1A* less than 3 DNA variants per 1 kb sequence were discovered but more than 3 SNPs per 1 kb were discovered in *PGM2*, *CSN1S1* and *SLC2A9*.

Table 4.15 Frequency of polymorphisms as calculated from information in Table 4.14

Gene	No. of exons	Introns completely re-sequenced	Re-sequenced nucleotides (kb)		Identified DNA variants per kb	
			Genomic sequence	Translated sequence	Genomic sequence	Translated sequence
ABCG2	16	-	13.8	1.98	2.39	2.02
OPN	7	1, 2, 4 to 6	8.6	0.91	1.28	0.00
PPARGC1A	13	6, 9	12.9	2.39	1.55	0.84
PGM2	14	5, 9	13.9	1.80	4.03	2.78
CSN1S1	19	5, 10	13.0	0.64	4.00	1.56
SLC2A9	13	-	7.0	1.56	4.86	7.05
Σ	82	-	69.2	9.28		
Average					2.97	2.47

Confining the result of SNP-screening to translated sequence, 2.34 DNA variants were found per 1 kb over all six candidate genes (see Table 4.15). Most of the genes, like *ABCG2*, *PPARGC1A*, *PGM2* and *CSN1S1*, contain within translated sequence between one and 2.5 DNA variants per 1 kb. However, the differences between the candidate genes were immense and ranged from zero in *OPN* to 7.05 in *SLC2A9*, so that *SLC2A9* is characterised by a disproportionately high figure of DNA variants in translated regions. *SLC2A9* was the only

gene showing a higher frequency of DNA variants in translated sequence than in non-translated sequence.

Ten of the SNPs from translated regions were non-synonymous (see Table 4.16). Non-synonymous SNPs were seen most frequently in *SLC2A9* and three of these are located in exon 8. *ABCG2* and *PGM2* contain two non-synonymous SNPs each. Both exchanges in *ABCG2* have conservative character both in *PGM2* are non-conservative. The previously suggested causative mutation for milk production traits in exon 14 (Cohen-Zinder *et al*. 2005) was not identified by comparative sequencing of the animal panel used for SNP-screening.

Table 4.16 Identified non-synonymous SNPs
SNPs are labelled by the in-house identification code (SNP_id).

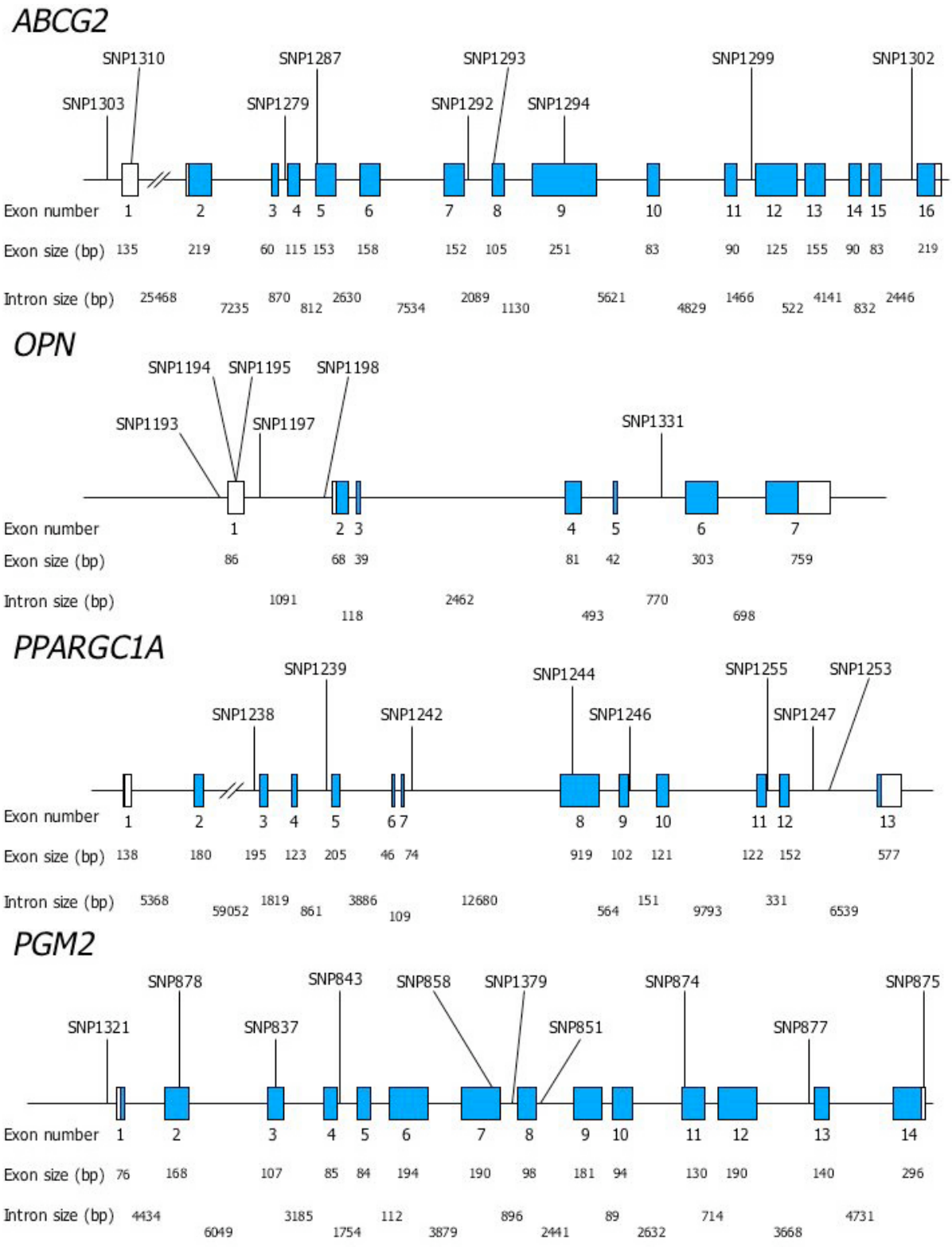
SNP	Gene	Exon	Nucleotide exchange in translated sequence [a]	Amino acid exchange	Exchange character
1293	*ABCG2*	8	AGGTTTCCC(A/G)CTGTGAGCC	IGF(H/R)CEP	Conservative
1294	*ABCG2*	9	TTCCATCCT(A/G)CAAGGAGGT	LPS(Y/C)KEV	Conservative
878	*PGM2*	2	GCAGGTAAT(A/G)CAGAAGAAC	AGN(T/A)EEL	Nonconservative
858	*PGM2*	7	GAGGCTGTT(C/T)CCGAACAGA	FVP(P/S)EAV	Nonconservative
1191	*CSN1S1*	17	AGAACAGTG(A/G)AAAGACTAC	ENS(E/G)KTT	Nonconservative
1213	*SLC2A9*	8	TTGGGCAAA(C/G)AGGACNTCT	LGK(E/Q)DIS	Nonconservative
1214	*SLC2A9*	8	AAANAGGAC(A/G)TCTCCCGNG	KED(I/V)SRE	Conservative
1215	*SLC2A9*	8	CGNGAGATG(C/G)AGGAGGTCC	REM(Q/E)EVL	Nonconservative
1226	*SLC2A9*	11	AGGATCATG(C/T)GCCCTGGAC	QDH(A/V)PWI	Conservative
1230	*SLC2A9*	13	TTTGCTGCA(A/G)TCTGCTTCG	FAA(I/V)CFA	Conservative

[a] Adjacent SNPs of the SNP of interest are replaced by *N* in genomic sequence

4.4.2. Polymorphisms selected for association studies

65 polymorphisms were selected for the association studies according to the following priority order: non-synonymous SNP – SNP from untranslated sequence – SNP from the region upstream of the transcription start site– SNP with synonymous character – SNP with intronic position. The SNPs were submitted to the MassARRAY® Typer 3.4 Software (Sequenom, San Diego, CA, USA) for primer design. This software identified six SNPs whose genotyping with MALDI-TOF mass spectrometry is not possible. Among them was SNP1214, a non-synonymous SNP of *SLC2A9*. Three polymorphisms are located in untranslated sequence, one in *CSN1S1* (SNP1150) and two in *OPN* (SNP1314, SNP1315). One SNP is in intronic sequence of *ABCG2* (SNP1289) and one in *CSN1S1* (SNP1164). The 59 remaining SNPs were combined in four separate iPLEX assays for genotyping, the fourth assay containing one SNP only. This polymorphism, SNP1320, has intronic position in *CSN1S1* and was not suitable for one of the other three assays so that it was not genotyped.

Primer design for the final set of 58 polymorphisms, 57 SNPs and one INDEL, was carried out. For detailed information about the position of each polymorphism, see Figure 4.20.



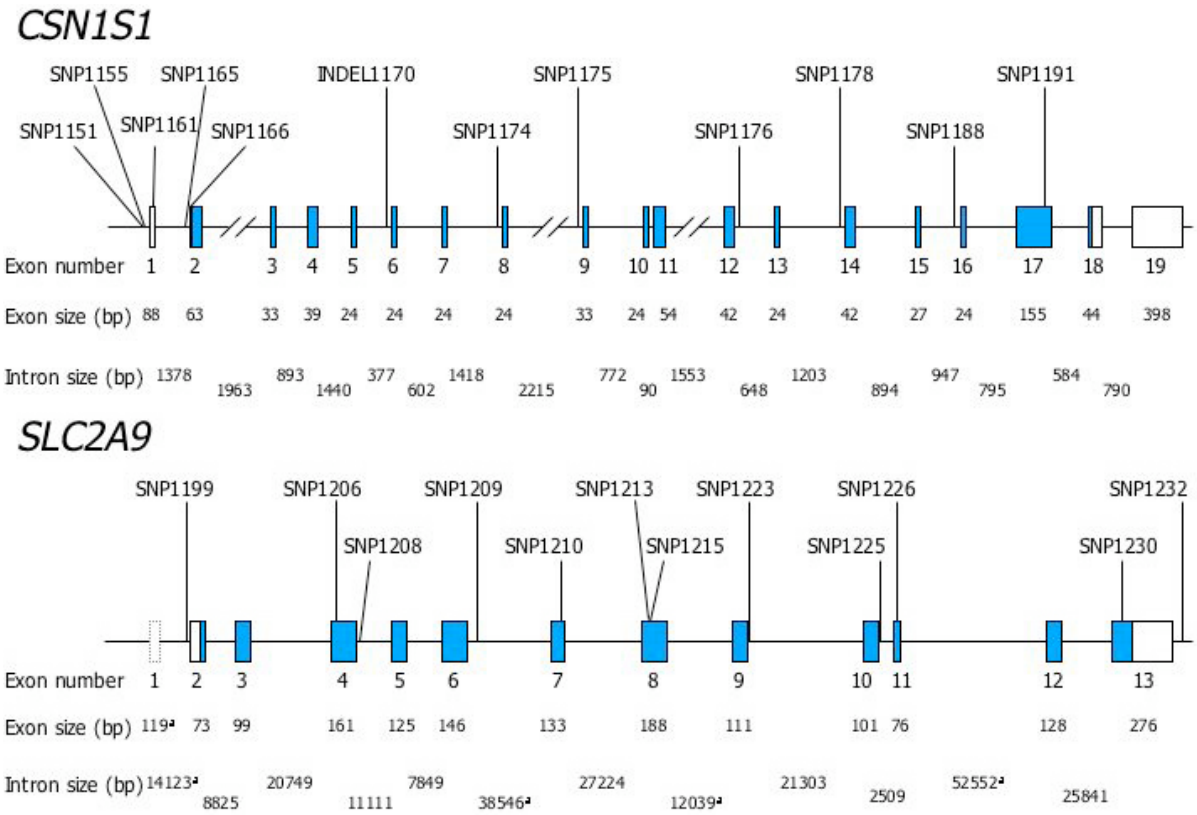


Figure 4.20 Position of the selected polymorphisms in the candidate genes
The blue bars represent translated sequence. White bars represent untranslated sequence. Intronic sequence is represented by the black line. Size information based on annotation in *Btau 3.1* is given below the bars. SNPs are labelled with the in-house identification code.
[a] Size information from humans.

4.5. *Genotyping result and test for Hardy-Weinberg Equilibrium*

The final set of DNA variants for the association studies contains 58 DNA variants, 57 SNPs and one insertion – deletion polymorphism (INDEL). 50 of these variants were successfully genotyped with MALDI-TOF MS. Table 4.17 lists all SNPs for which a genotyping result could not be obtained.

Table 4.17 SNPs with failed genotyping

SNP	Gene	SNP position	Reason for failed genotyping
1198	*OPN*	Intron 1	Clusters not well differentiated
1253	*PPARGC1A*	Intron 12	Primer amplification failed
1255	*PPARGC1A*	Intron 11	Primer amplification failed
1151	*CSN1S1*	Upstream	Primer amplification failed
1155	*CSN1S1*	Upstream	Primer amplification failed
1178	*CSN1S1*	Intron 13	Primer amplification failed
1230	*SLC2A9*	Exon 13	Primer amplification failed

The number of genotyped animals per marker is on average 2852.9 and ranges from 2620 (SNP1302) to 2920 (SNP1247). Accepting the common rule of a minor allele frequency

(MAF) for SNPs of more than 1%, BV and FV are monomorphic at five markers, respectively. In BV, the MAF is lower than 0.01 at SNP1293, SNP1294, SNP1331, SNP1239 and SNP1209 and lower than 0.01 in FV at SNP1293, SNP1294, SNP1223, SNP1209 and SNP1206 (see Table 4.18). HF is monomorphic at nine SNPs: SNP1287, SNP1302, SNP1191, SNP1188, SNP1175, SNP1165, SNP1223, SNP1213 and SNP1206.

Table 4.18 Results from genotyping

Polymorphisms are described by SNP_id, the in-house SNP-identification code. In addition, the gene, the position in the gene and the nucleotide exchange is listed. Columns five, six and seven comprehend for each breed the minor allele with frequency. The SNP position on BTA6 (*Btau 3.1*) is given by nucleotide position (bp).

SNP	SNP position	Nucleotide exchange	Minor allele (frequency)			Localisation on BTA6 (bp) in *Btau 3.1*
			BV	FV	HF	
ABCG2						
1303	Upstream	A > G	G (0.474)	G (0.278)	A (0.437)	33549254
1310	5' UTR	A > T	A (0.113) [c]	A (0.395)	A (0.074)	33549828
1279	Intron 3	A > T	T (0.030)	T (0.190)	A (0.441)	33583613
1287	Exon 5	C > T	C (0.474)	C (0.183)	C (0.001)	33584710
1292	Intron 7	A > G	A (0.117)	A (0.449)	A (0.472)	33595516
1293	Exon 8	A > G	A (0.000)	A (0.001)	A (0.243)	33597436
1294	Exon 9	A > G	G (0.001)	G (0.000)	G (0.153)	33598812
1299	Intron 11	G > T	G (0.030)	G (0.040)	G (0.158)	33610928
1302	Intron 15	A > G	A (0.455)	G (0.287)	G (0.006)	33620004
MAF average			0.188	0.203	0.221	
OPN						
1331	Intron 5	C > T	C (0.001)	C (0.073)	C (0.477)	33712386
1197	Intron 1	A > G	A (0.067)	A (0.406)	A (0.277)	33716486
1195	5' UTR	A > G	G (0.067)	G (0.484)	A (0.076)	33717226
1194	5' UTR	C > T	T (0.495)	T (0.279)	T (0.035)	33717228
1193	5' end	C > T	T (0.066)	T (0.409)	T (0.278)	33717256
MAF average			0.139	0.330	0.237	
PPARGC1A						
1247	Intron 12	A > C	C (0.324)	A (0.420)	C (0.107)	41143032
1246	Intron 9	C > T	T (0.060)	T (0.132)	T (0.219)	41157582
1244	Exon 8	C > T	T (0.294)	T (0.392)	C (0.170)	41158848
1242	Intron 7	A > G	A (0.481)	A (0.231)	A (0.483)	41170736
1239	Intron 4	G > T	T (0.000)	T (0.015)	T (0.071)	41175394
1238	Intron 2	A > G	G (0.494)	A (0.490)	G (0.371)	41178281
MAF average			0.276	0.280	0.237	
PGM2						
1321	Upstream	A > G	A (0.211)	A (0.079)	A (0.340)	52179988 [a]
878	Exon 2	A > G	G (0.185)	G (0.287)	G (0.096) [c]	52184812 [a]
837	Exon 3	C > T	C (0.213)	C (0.089)	C (0.409)	52190478 [a]
843	Intron 4	A > G	A (0.214)	A (0.089)	A (0.408)	52194174 [a]
858	Exon 7	C > T	T (0.213)	T (0.090)	T (0.407)	52199253 [a]
1379	Intron 7	C > G	C (0.213)	C (0.088)	C (0.407)	52200022 [a]
851	Intron 8	A > G	A (0.211)	A (0.089)	A (0.408)	52200411 [a]
874	Exon 11	C > T	C (0.214)	C (0.089)	C (0.407)	52205780 [a]
877	Intron 12	C > T	T (0.156)	T (0.470)	T (0.433)	52210430 [a]
875	3' UTR	A > G	A (0.216)	A (0.088)	A (0.408)	52215454 [a]
MAF average			0.205	0.146	0.372	

CSN1S1						
1191	Exon 17	A > G	G (0.068)	G (0.076)	G (0.003)	79753530
1188	Intron 15	A > G	G (0.050)	G (0.076)	G (0.003)	79754369
1176	Intron 12	C > T	T (0.333)	T (0.270)	T (0.454)	79758171
1175	Intron 8	C > T	C (0.050)	C (0.076)	C (0.003)	79760940
1174	Intron 7	A > T	T (0.194)	T (0.162)	T (0.308)	79762464
1170	Intron 5	D > T	T (0.357)	D (0.416)	T (0.456)	79763772
1166	5' UTR	C > T	T (0.050)	T (0.090)	T (0.050)	79767761
1165	Intron 1	A > C	A (0.047)	A (0.089)	A (0.002)	79767876
1161	5' UTR	C > T	C (0.050)	C (0.092)	C (0.050)	79769175
MAF average			0.133	0.150	0.148	
SLC2A9						
1232	3' UTR	A > C	A (0.470)	C (0.373)	C (0.369)	96990000 [b]
1226	Exon 11	C > T	T (0.038)	T (0.074)	T (0.053)	97048880 [b]
1225	Intron 10	A > G	G (0.116)	G (0.328)	G (0.358)	97051880 [b]
1223	Intron 9	A > G	G (0.216)	G (0.003)	G (0.000)	97068880 [b]
1215	Exon 8	C > G	G (0.399)	G (0.486)	G (0.361)	97081380 [b]
1213	Exon 8	C > G	C (0.221)	C (0.197)	C (0.002)	97081400 [b]
1210	Exon 7	A > G	A (0.087)	A (0.092)	A (0.220)	97102400 [b]
1209	Intron 6	G > T	T (0.007) [c]	T (0.003) [c]	T (0.017) [c]	97141229 [b]
1208	Intron 4	A > G	A (0.340)	G (0.404)	A (0.116)	97157129 [b]
1206	Exon 4	C > T	T (0.047)	T (0.007)	T (0.003)	97157399 [b]
1199	Intron 1	G > T	G (0.252)	G (0.263)	G (0.267)	97181400 [b]
MAF average			0.199	0.203	0.161	
MAF average across all genes			0.190	0.204	0.227	

[a] Chromosomal positions of *PGM2* polymorphisms are calculated using the annotation in Chr.Un.003.335 (*Btau 3.1*) and assuming a chromosomal anchorage of the first nucleotide of the start codon at 52180182 bp (see 4.2.1 Positioning of candidate genes in bovine draft sequences)

[b] Chromosomal positions of *SLC2A9* polymorphisms are calculated using the annotation in *Btau 2.0* (see 4.2.1 Positioning of candidate genes in bovine draft sequences) and assuming a chromosomal anchorage of the 5' splice donor of exon 3 at 97157422 bp on BTA6 in *Btau 3.1* (see 4.2.1 Positioning of candidate genes in bovine draft sequences)

[c] Significant ($p < 0.001$) departure from Hardy-Weinberg Equilibrium

The MAF average within breeds calculated with all markers is comparable across breeds. BV has a total MAF average of 0.190, ranging from fixation (SNP1293, SNP1239) to 0.495 (SNP1194). FV has a total MAF average of 0.204, with MAFs between zero (SNP1294) and 0.490 (SNP1238). HF shows the highest total MAF mean, 0.227 (see Table 4.18) and MAFs range from fixation (SNP1223) to 0.483 (SNP1242). A subset of 40 among all 50 genotyped markers shows identical minor alleles across breeds whereas *PGM2* presents for each marker x breed combination the same minor allele. *CSN1S1* is characterised by the lowest MAF average across breeds (0.143) and exhibits the highest level of allelic fixation. *PPARGC1A* has a mean MAF of 0.264 across breeds showing the lowest level of allelic fixation among all observed genes.

However, when regarding genes separately, there are differences in the allele frequencies between breeds. BV has the lowest intra-gene MAF average in *CSN1S1* (0.133) and the highest in *PPARGC1A* (0.276). FV shows the minimal MAF average in *PGM2* (0.146) and the maximal average in *OPN* (0.330). *CSN1S1* is the gene with the lowest MAF average in

HF (0.148), the highest mean MAF in HF appears in *PGM2* (0.372), so that differences in the MAF averages per gene were obtained between breeds. The allele frequency of neutral variants is pushed either upwards or downwards by selection on a linked beneficial mutation (Kim 2006). An excess of high frequency derived alleles can be a signature of single selective sweeps. Selective sweeps generate a local reduction of sequence variation (KIM and STEPHAN 2002) and DNA affected by selective sweeps contains fewer segregating alleles of intermediate frequency (Braverman *et al.* 1995). *CSN1S1* has a low MAF average across breeds. This fact might indicate that the casein gene cluster has been under intense selective pressure in all breeds.

Markers were tested within breeds for a departure from the Hardy-Weinberg Equilibrium (HWE). Results were adjusted to multiple testing according to the Bonferroni correction method and markers that depart from HWE expectation are marked in Table 4.18. SNP1209 does not fit HWE expectation in all breeds, showing a deficit in heterozygous animals in BV, FV and HF. Additionally in both BV and HF, one marker does not comply with the HWE expectation. SNP1310 deviates in BV with a heterozygous deficit. SNP878 deviates in HF showing excess in homozygous GG animals.

The test for HWE is a useful indicator for genotype frequencies within populations and whether the frequencies are based on a valid definition of alleles. Calculations during association studies presume that the observed population is in HWE. HWE assumes stable populations under random mating and without selective pressure, migration or genetic drift (FALCONER and MACKAY 1996). Therefore, departure from the HWE expectation can be an indication that one of these assumptions is not true. One the other hand, a departure from HWE might indicate technical problems during genotyping. Since SNP1209 departs from HWE expectation in all breeds, problems during genotyping, such as an undetected polymorphism in the genomic region of primer annealing is likely. Both the other polymorphisms appear breed-specific and might occur for breed-specific, population genetical reasons, rather than from technical problems during genotyping.

4.6. *Linkage Disequilibrium between candidate genes*

Two measures are commonly used to evaluate LD: D' and r^2. Neither is completely independent of allele frequency (Lewontin 1988). However, the D' coefficient was used since the estimation of its variation does not depend on allele frequency (HEDRICK 1987; HEDRICK 1988; LEWONTIN 1988). D' averages for marker pairs within genes are > 0.80 across all gene *x*

breed combinations, apart from *SLC2A9* in FV and HF and from *ABCG2* in HF (see Table 4.19).

Table 4.19 Breed average of *D'* for gene *x* gene combinations
Diagonal values represent *D'* averages for markers if both markers are located in the same gene, off-diagonal values are *D'* averages for maker pairs of different genes.

	ABCG2	*OPN*	*PPARGC1A*	*PGM2*	*CSN1S1*	*SLC2A9*
BV						
ABCG2	0,801	0,496	0,443	0,446	0,320	0,353
OPN		0,817	0,410	0,219	0,201	0,292
PPARGC1A			0,893	0,262	0,366	0,227
PGM2				0,970	0,249	0,311
CSN1S1					0,972	0,587
SLC2A9						0,808
FV						
ABCG2	0,826	0,846	0,346	0,224	0,300	0,323
OPN		0,999	0,254	0,153	0,100	0,262
PPARGC1A			0,843	0,279	0,135	0,268
PGM2				0,936	0,086	0,306
CSN1S1					0,971	0,310
SLC2A9						0,611
HF						
ABCG2	0,772	0,570	0,427	0,240	0,350	0,254
OPN		0,998	0,543	0,171	0,213	0,219
PPARGC1A			0,860	0,209	0,276	0,214
PGM2				0,893	0,364	0,167
CSN1S1					0,919	0,360
SLC2A9						0,649

Across breeds, *SLC2A9* has the lowest *D'* average while *CSN1S1* and *PGM2* show the highest level of LD among all observed genes. In BV, *ABCG2* and *CSN1S1* show the lowest and highest intra-gene LD respectively. *SLC2A9* and *OPN* show the lowest and highest LD within genes (see Table 4.19) in FV and HF. Only in FV, between ABCG2 and OPN, does LD between genes reach a level that otherwise appears for marker-pairs within a gene (*D'* = 0.846). The average D' for marker-pairs with < 5 cM distance is 0.816 in BV, 0.832 in FV and 0.777 in HF, which is approximately equivalent to the *D'* average for markers within genes (see 4.1 Selection of candidate genes). Markers that are located > 20 cM apart show, on average, *D'* values of 0.296 in BV, 0.228 in FV and 0.261 in HF. *Haploview* (Barrett *et al.* 2005) was used to visualize the breed-specific pattern of LD. The MAF threshold for SNPs to be included in the *Haploview* analysis was set at 0.01 to eliminate monomorphic markers. HWE-expectation was tested and markers were excluded when $p < 0.001$. The higher number of marker-pairs between *CSN1S1* and *SLC2A9* showing a $D' > 0.80$ in BV (see Figure 4.21), as well as the higher number of marker-pairs of *ABCG2* and *PPARGC1A* showing a $D' > 0.80$

in HF (see Figure 4.23) are examples for the breed-specific pattern of LD. The LD plot for FV is shown in Figure 4.22.

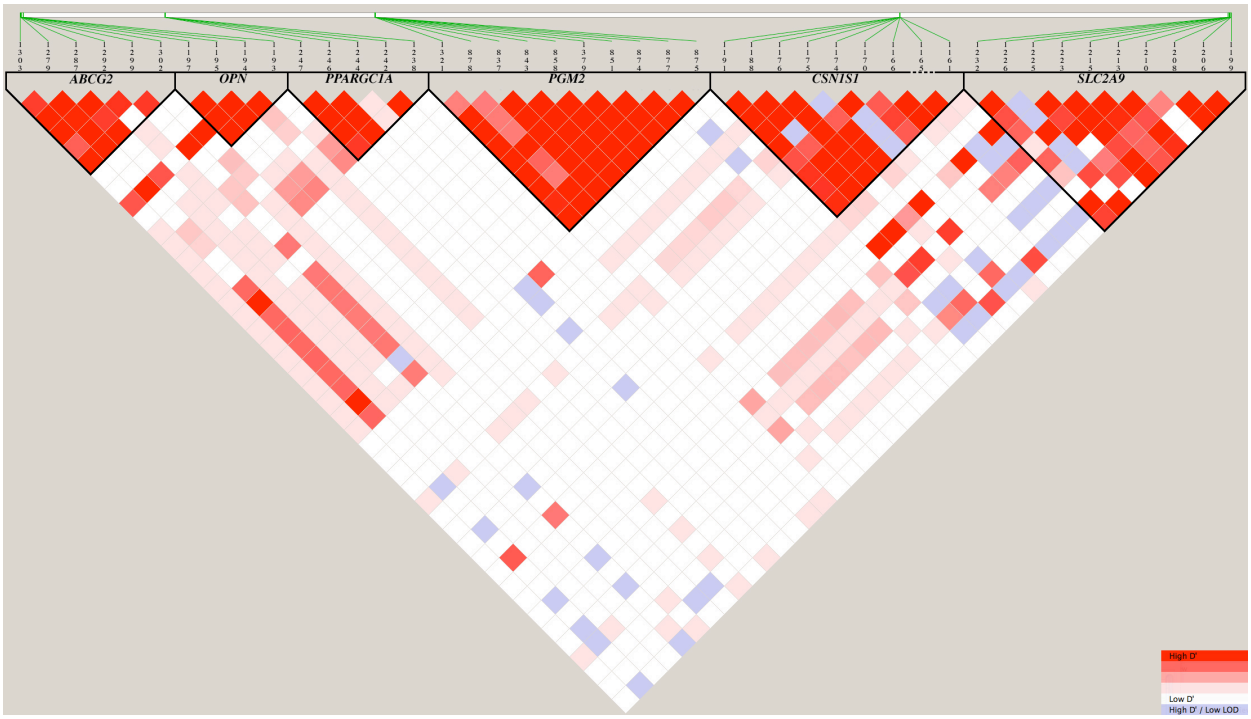


Figure 4.21 LD plot for BV across the chromosome segments
Each diamond contains the level of LD, measured in *D'*, between the markers specified, visualized using *Haploview* (Barrett *et al.* 2005). The intensity of the colours corresponds to levels of *D'*. Markers belonging to the same gene are framed by a black triangle. The SNPs are identified by the in-house database code (SNP_id), represented above the plot in vertical orientation.

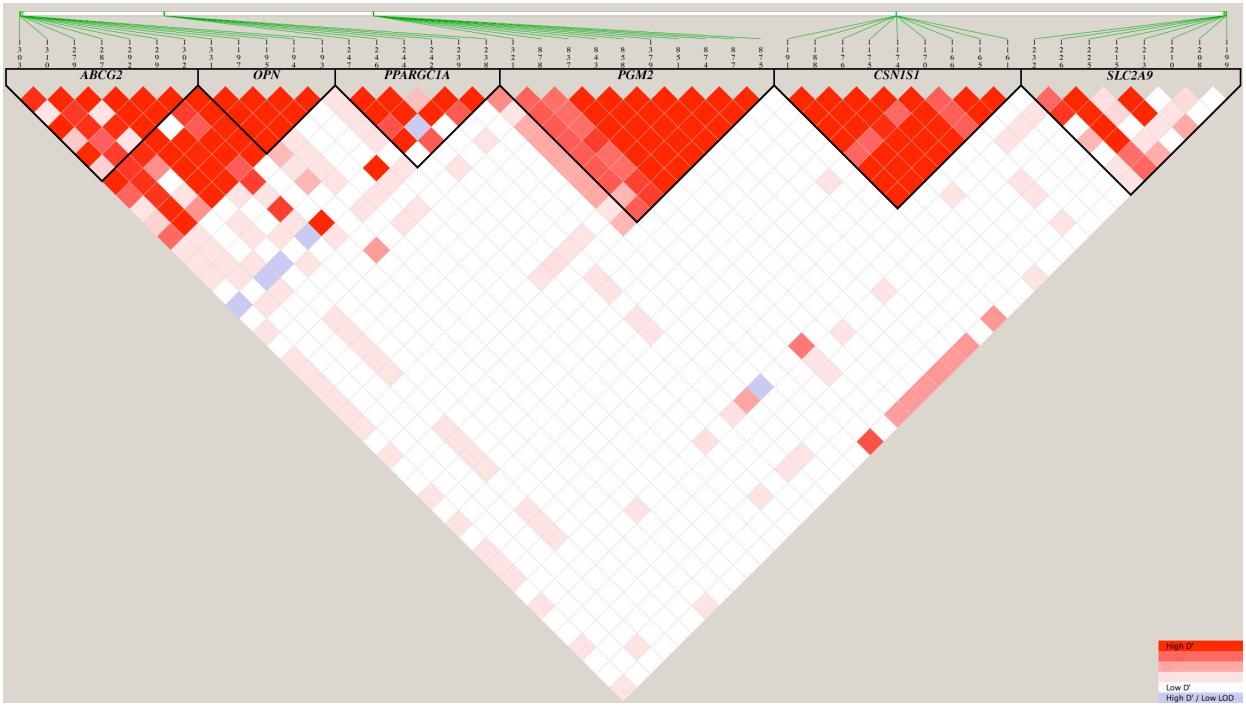


Figure 4.22 LD plot for FV across the chromosome segments
Each diamond contains the level of LD, measured in *D'*, between the markers specified, visualized using *Haploview* (Barrett *et al.* 2005). The intensity of the colours corresponds to levels of *D'*. Markers belonging to the same gene are framed by a black triangle. The SNPs are identified by the in-house database code (SNP_id), represented above the plot in vertical orientation.

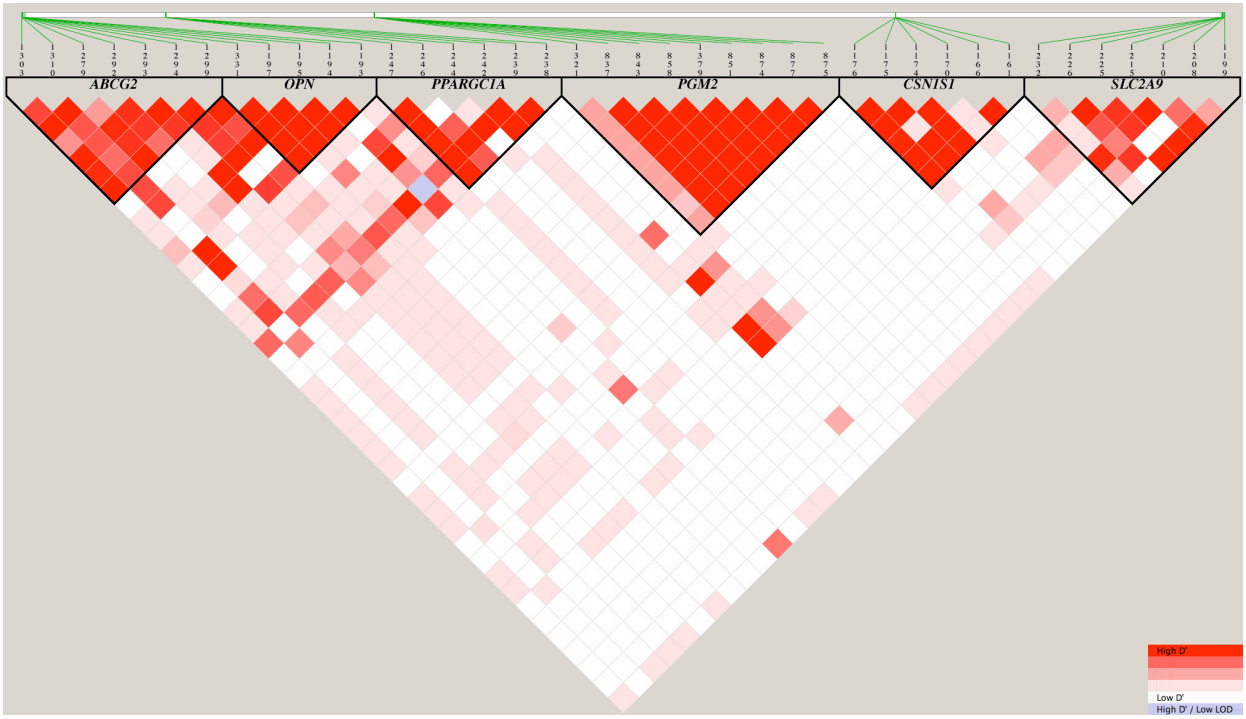


Figure 4.23 LD plot for HF across the chromosome segments
Each diamond contains the level of LD, measured in *D*', between the markers specified, visualized using *Haploview* (Barrett *et al*. 2005). The intensity of the colours corresponds to levels of *D*'. Markers belonging to the same gene are framed by a black triangle. The SNPs are identified by the in-house database code (SNP_id), represented Figure the plot in vertical orientation.

D' was analysed by means of a Malecot model to study the pattern of LD in the breeds relative to the marker distance. The decay of D' with increasing distance (see Figure 4.24) is highly significant ($p < 0.001$) across breeds (see Table 4.20). The parameter *k* represents the expected decay of *D*' when the marker distance increases by the value of 1 Mb (see Table 4.20).

Table 4.20 Parameter estimates from Malecot model

Breed	Parameter	Estimate ± Standard error	*p*-value
BV	*L*	0.319 ± 0.009	$p < 0.001$
	M	0.731 ± 0.020	$p < 0.001$
	k	0.158 ± 0.017	$p < 0.001$
FV	*L*	0.233 ± 0.007	$p < 0.001$
	M	0.793 ± 0.016	$p < 0.001$
	k	0.281 ± 0.038	$p < 0.001$
HF	*L*	0.254 ± 0.008	$p < 0.001$
	M	0.710 ± 0.017	$p < 0.001$
	k	0.171 ± 0.018	$p < 0.001$

The background *D'* for loosely linked markers (> 20 Mb apart) is about 0.30 across breeds, while the level of background *D'* is marginally higher in BV than in FV and HF. The decay of *D'* with marker distance is most rapid in FV and least rapid in BV (see Figure 4.24).

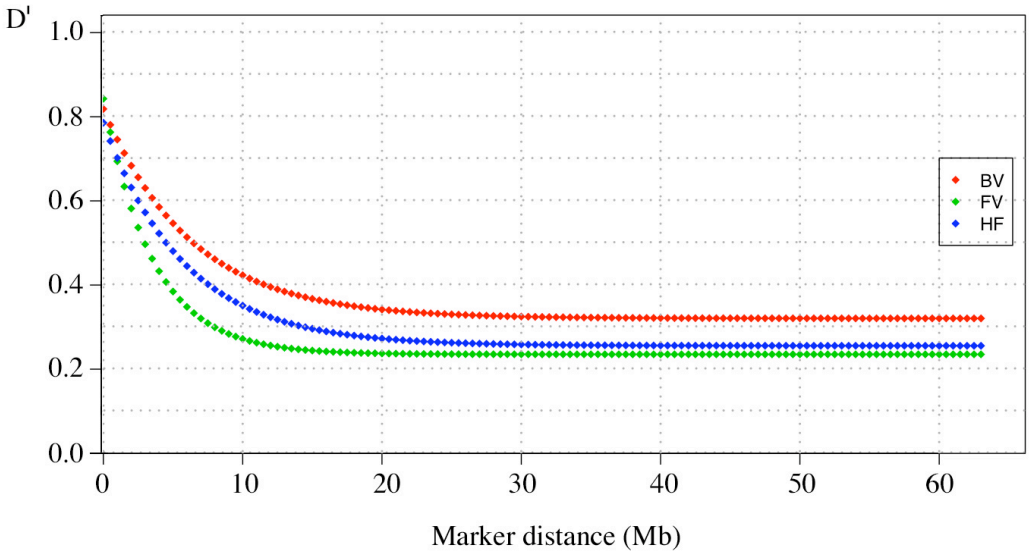


Figure 4.24 Malecot model estimated dependence of *D'* on physical marker distance

The modified Malecot model was applied in order to ascertain, taking physical distance into account, whether LD within genes is more extensive than LD between genes. LD is significantly higher within genes than between genes in BV and HF, but not in FV, as parameterised by k_2 (see Table 4.21).

Table 4.21 Parameter estimates from the modified Malecot model

Breed	Parameter	Estimate ± Standard error	*p*-value
BV	*L*	0.270 ± 0.030	$p < 0.001$
	M	0.838 ± 0.020	$p < 0.001$
	k_1	0.038 ± 0.015	$p < 0.05$
	k_2	1.018 ± 0.104	$p < 0.001$
FV	*L*	0.234 ± 0.007	$p < 0.001$
	M	0.786 ± 0.017	$p < 0.001$
	k_1	0.291 ± 0.040	$p < 0.001$
	k_2	0.058 ± 0.053	$p > 0.05$
HF	*L*	0.250 ± 0.009	$p < 0.001$
	M	0.764 ± 0.018	$p < 0.001$
	k_1	0.123 ± 0.018	$p < 0.001$
	k_2	0.510 ± 0.089	$p < 0.001$

4.7. *Constructed haplotypes*

SimWalk2 v2.91 was used for haplotype construction including parental information. Only sires of paternal half-sib families counting of ≥ 10 sons were included in *SimWalk2* haplotype estimation. In case of smaller half-sib family groups, haplotypes were derived with *fastPHASE*, which does not use pedigree information. Gabriel et. al (2002) proposed obtaining haplotypes with marker pairs showing a $D' > 0.90$ (Gabriel *et al.* 2002). In most cases, results from LD analysis show $D' > 0.85$ for markers within genes (see 4.6 Linkage Disequilibrium between candidate genes). *SLC2A9* showed $D' < 0.70$ in FV and HF, but a different block definition than that incorporating all *SLC2A9*-markers does not appear more obvious (see Figure 4.22). Thus, haplotypes were constructed separately for each gene by incorporating all markers, except for *ABCG2* and *OPN* in FV. *ABCG2* and *OPN* are almost in complete LD in FV, so haplotypes were constructed that incorporate markers of both genes. Only haplotypes that occurred more frequently than 2.5% were considered.

In total, 62 haplotypes were constructed (see Table 11.3), most of them in *SLC2A9* as a consequence of relatively low LD. As a consequence of extensive LD, *CSN1S1* and *PGM2* are the loci that have the highest amount of haplotypes that are shared between breeds (see Table 4.22).

Table 4.22 Haplotype sharing between breeds

	Number of constructed haplotypes		
Gene	Total	Shared between all breeds	Shared between two breeds
ABCG2	10	[a]	2
OPN	5	[a]	3
PPARGC1A	8	5	0
PGM2	6	3	1
CSN1S1	6	2	3
SLC2A9	18	5	2

[a] Haplotypes in FV were constructed with markers of *ABCG2* and *OPN*

4.8. *Association studies with milk production traits*

4.8.1. Single marker analysis (SMA)

4.8.1.1. Markers of ABCG2 *in single marker analysis*

ABCG2 markers show associations with milk production traits only in HF, with two markers (SNP1293, SNP1294) being significantly associated (see Figure 4.25). Both are non-synonymous polymorphisms causing conservative amino acid exchanges (see 4.4.1 Identified

polymorphisms). SNP1293 is associated significantly with FY3 and PY2, whereas SNP1294 exhibits a significant association with PY2. The *p*-values for SNP1293 on FY2 and PY3 and for SNP1294 on FY3 are marginally above the threshold of significance ($p < 0.01$, $FDR > 10\%$) (see Figure 4.25). Suggestive effects ($p < 0.05$, $FDR > 10\%$) on MY2 and MY3 were identified for both markers. The significant effects ($\alpha/2$) for the G allele of SNP1293 range from +2.58 kg in PY2 to +3.56 kg in FY3 (see Table 11.8).

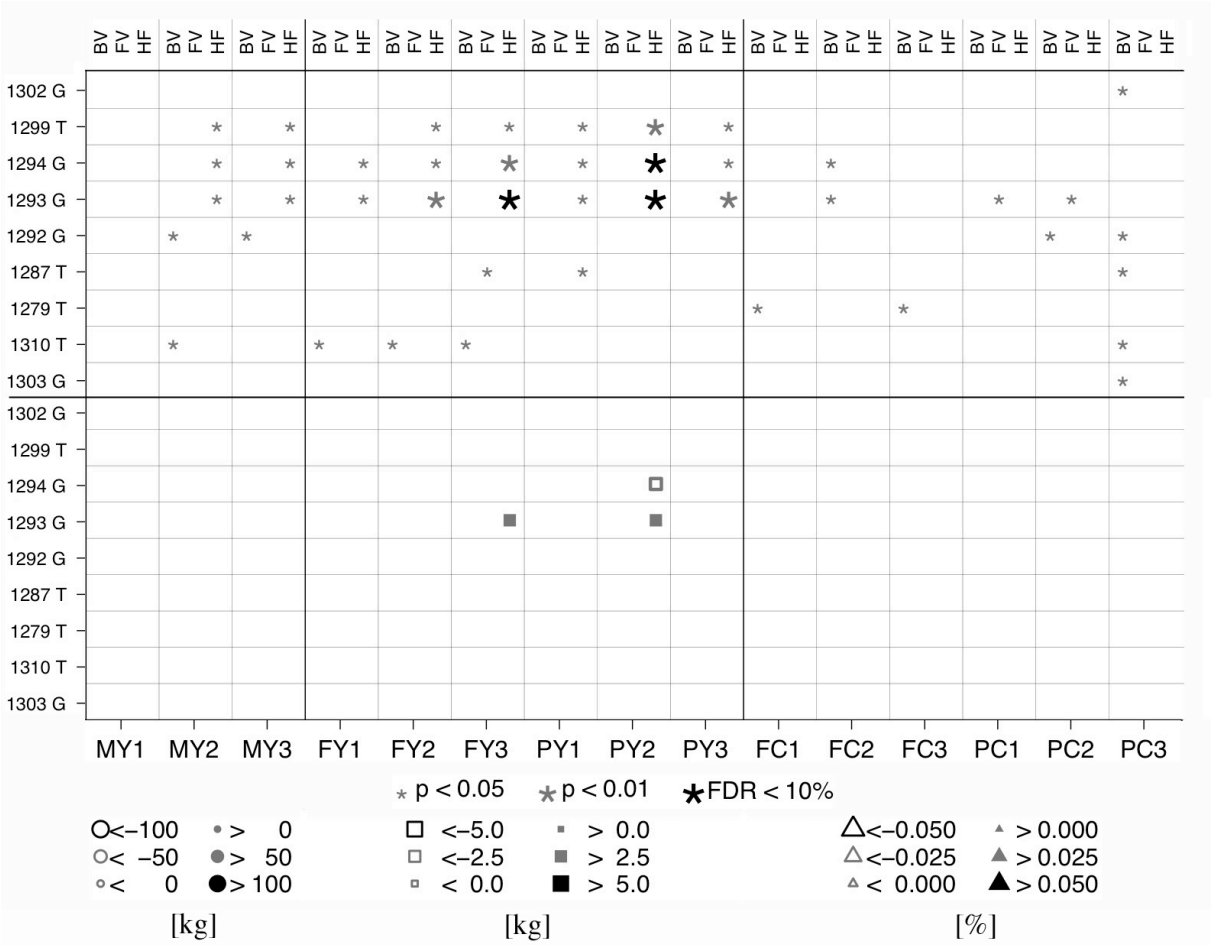


Figure 4.25 Significance level and effects ($\alpha/2$) of *ABCG2*-markers with *FDR* < 10%

4.8.1.2. Markers of OPN in single marker analysis

OPN, namely SNP1331, exhibits significant effects on milk production traits exclusively in HF. SNP1331 is located in the fifth intron and significant associations were discovered on FY3, PY2 and PY3. The estimated effects ($\alpha/2$) for the T variant of SNP1331 vary between -2.41 kg in PY2 and -3.83 kg in FY3 (see Table 11.8). Suggestive associations ($p < 0.05$, $FDR > 10\%$) were observed with MY3, FY1, FY2 and PY1 (see Figure 4.26).

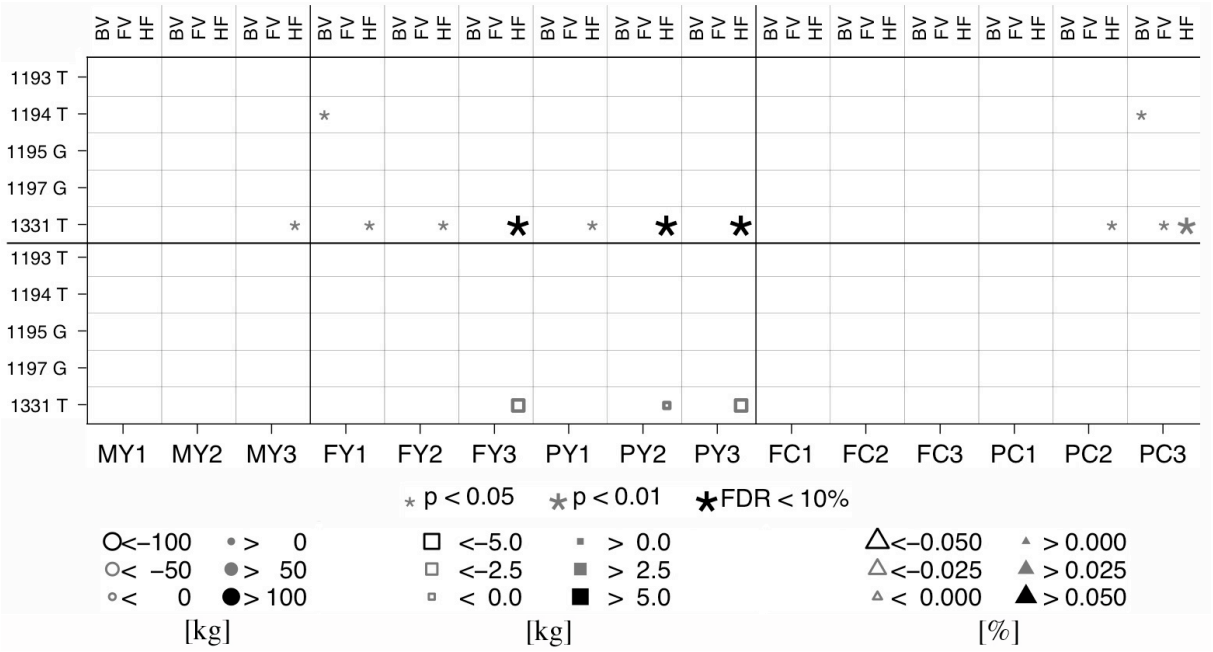


Figure 4.26 Significance level and effects (α/2) of *OPN*-markers with *FDR* < 10%

4.8.1.3. Markers of PPARGC1A in single marker analysis

PPARGC1A markers show significant effects on milk production traits in each breed and at three intronic polymorphisms: SNP1238, SNP1239 and SNP1242. Significant associations were detected in BV for SNP1238 and SNP1242 with MY2 and FY2 (see Table 11.5), in FV for SNP1238 with FC1 (see Table 11.7) and in HF for SNP1239 with FY3, PY2 and PY3 (see Table 11.8). None of the associated markers shows significant effects through all three lactations (see Figure 4.27). In HF the effects of SNP1239 on MY3, FY1, FY2, and PY1 are suggestive ($p < 0.05$, $FDR > 10\%$). In FV the p-value of SNP1238 is marginally above the threshold of significance ($p < 0.01$, $FDR > 10\%$) for FC2 and suggestive for FC3 ($p < 0.05$, $FDR > 10\%$). In BV, only MY2 and FY2 are significantly associated with SNP1238 and the effects on MY1 and FY1 are suggestive ($p < 0.05$, $FDR > 10\%$). The effects (α/2) for the T allele of SNP1239 range from +4.38 kg in PY2 and +7.17 kg in FY3 in HF (see Table 11.8) to +1.01 kg on FY2 for SNP1238 in BV. In FV, the G variant of SNP1238 is significantly associated with an effect (α/2) of -0.023% in FC1.

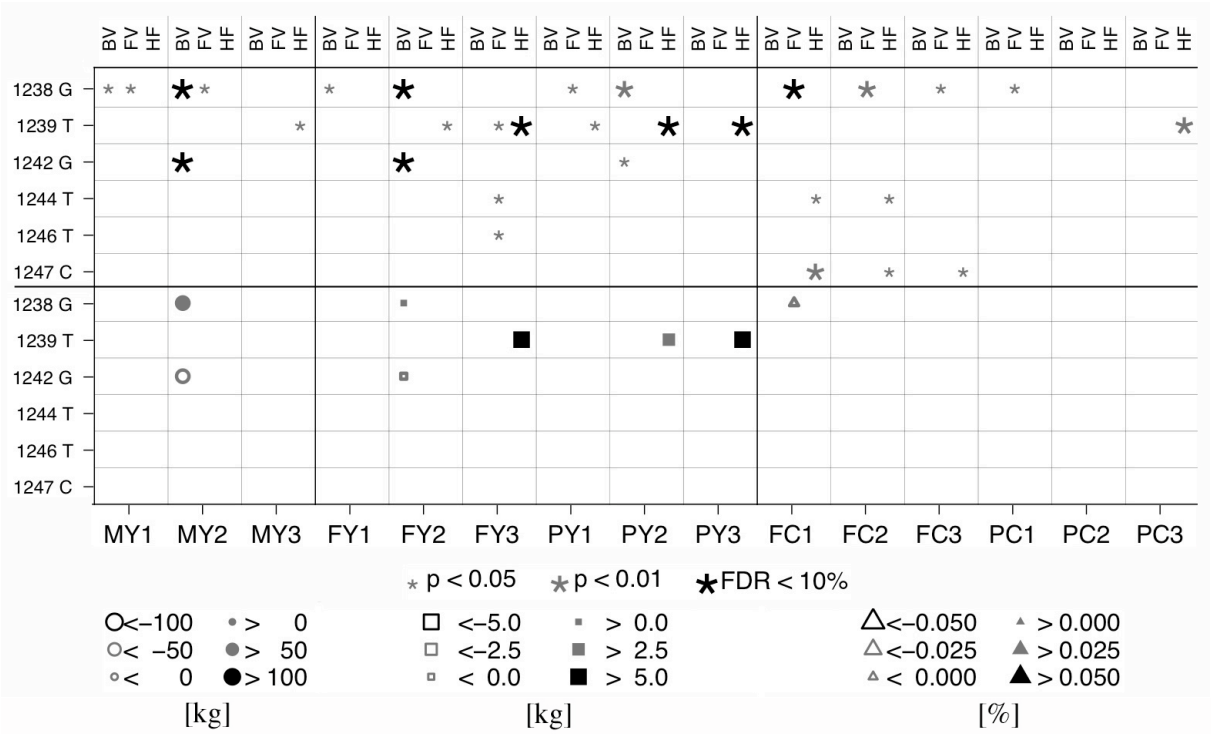


Figure 4.27 Significance level and effects (α/2) of *PPARGC1A*-markers with *FDR* < 10%

4.8.1.4. Markers of PGM2 in single marker analysis

Among the markers of *PGM2*, only SNP877 is significantly associated in FV. This polymorphism, located in intron 12, exhibits a significant effect on FC3 (see Figure 4.28) and the estimated effect (α/2) for its T variant is –0.029% (see Table 11.7). SNP877 does not show any association with FC1 and FC2 (see Table 11.7).

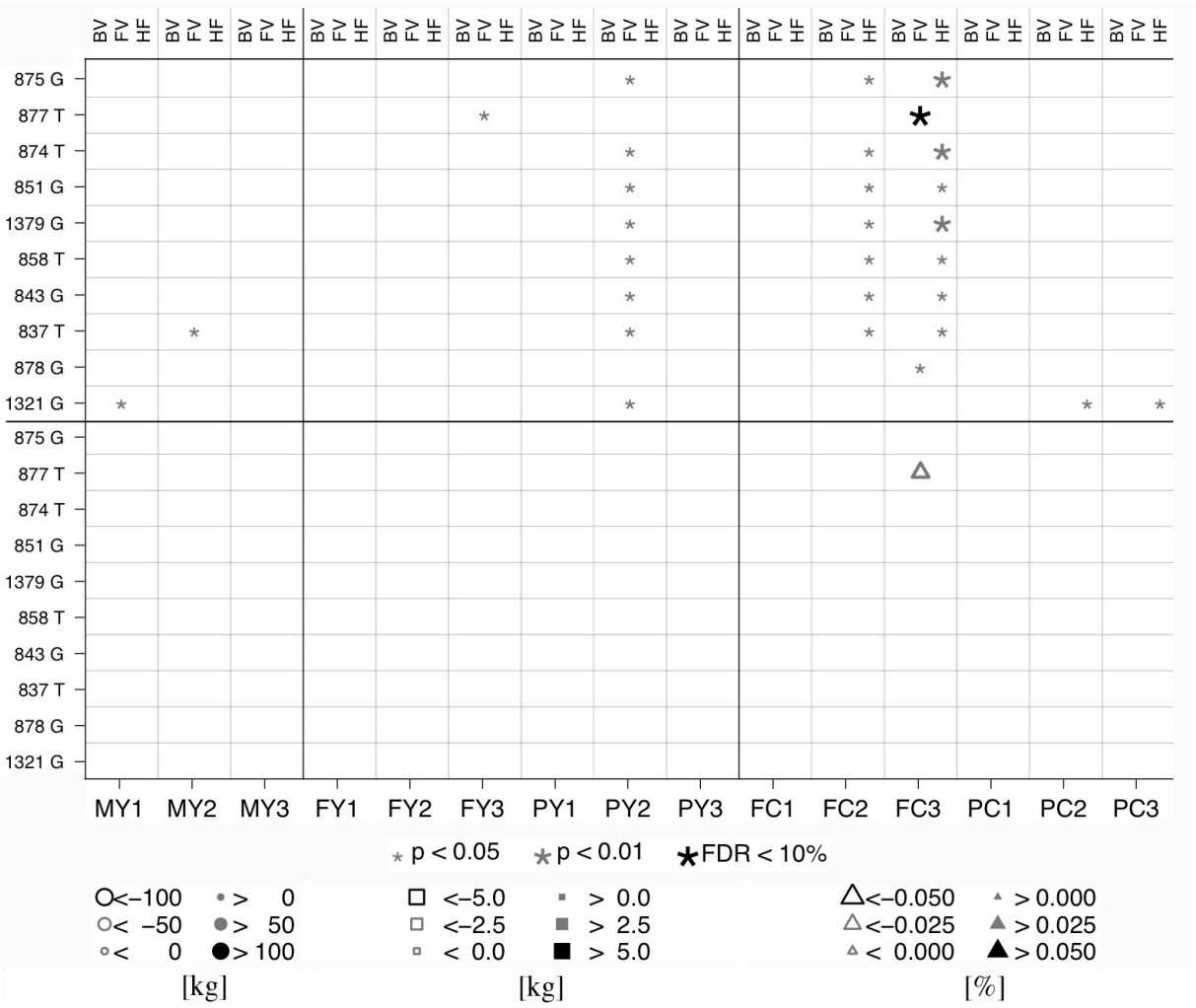


Figure 4.28 Significance level and effects (α/2) of *PGM2*-markers with *FDR* < 10%

4.8.1.5. Markers of CSN1S1 in single marker analysis

CSN1S1 has nine significantly associated markers across all breeds and traits and associations have mainly been detected with protein content. In BV and FV however, *CSN1S1* markers also show significant effects on fat content and milk yield (see Figure 4.29). Among the significantly associated markers is SNP1191, the non-synonymous polymorphism in exon 17. SNP1191 exhibits significant effects on MY1, FC2, PC1 and PC2 in BV and on PC1, PC2 and PC3 in FV. The associations of CSN1S1 markers with protein content are, for the most part, significant in all lactations. If not, the *p*-values for the non-significant lactations are just above the threshold of significance ($p < 0.01$, $FDR > 10\%$) (see Figure 4.29). The effect directions, increasing or decreasing, are comparable across breeds at each SNP, although effect sizes differ between breeds. In BV the effects (α/2) range from +0.033% for the G allele of SNP1191 in FC1 to +0.064% for the G allele of SNP1188 in FC1 (see Table 11.4). FV effects vary from +0.019% for the T variant of SNP1176 on FC1 and FC2 to +0.39% for

the T variant of INDEL1170 in FC3 (see Table 11.6), whereas the HF effects on protein content are between +0.13% for the T allele of INDEL1170 in PC1 and +0.19% for the T allele of SNP1174 in PC2 (see Table 11.8). SMA in HF revealed suggestive associations ($p < 0.05$, $FDR > 10\%$) on MY1 and MY2 for SNP1174. Antagonistic effects between milk yield and content traits were observed across all CSN1S1 markers and breeds, showing that significant effects on fat- and protein yield are minor.

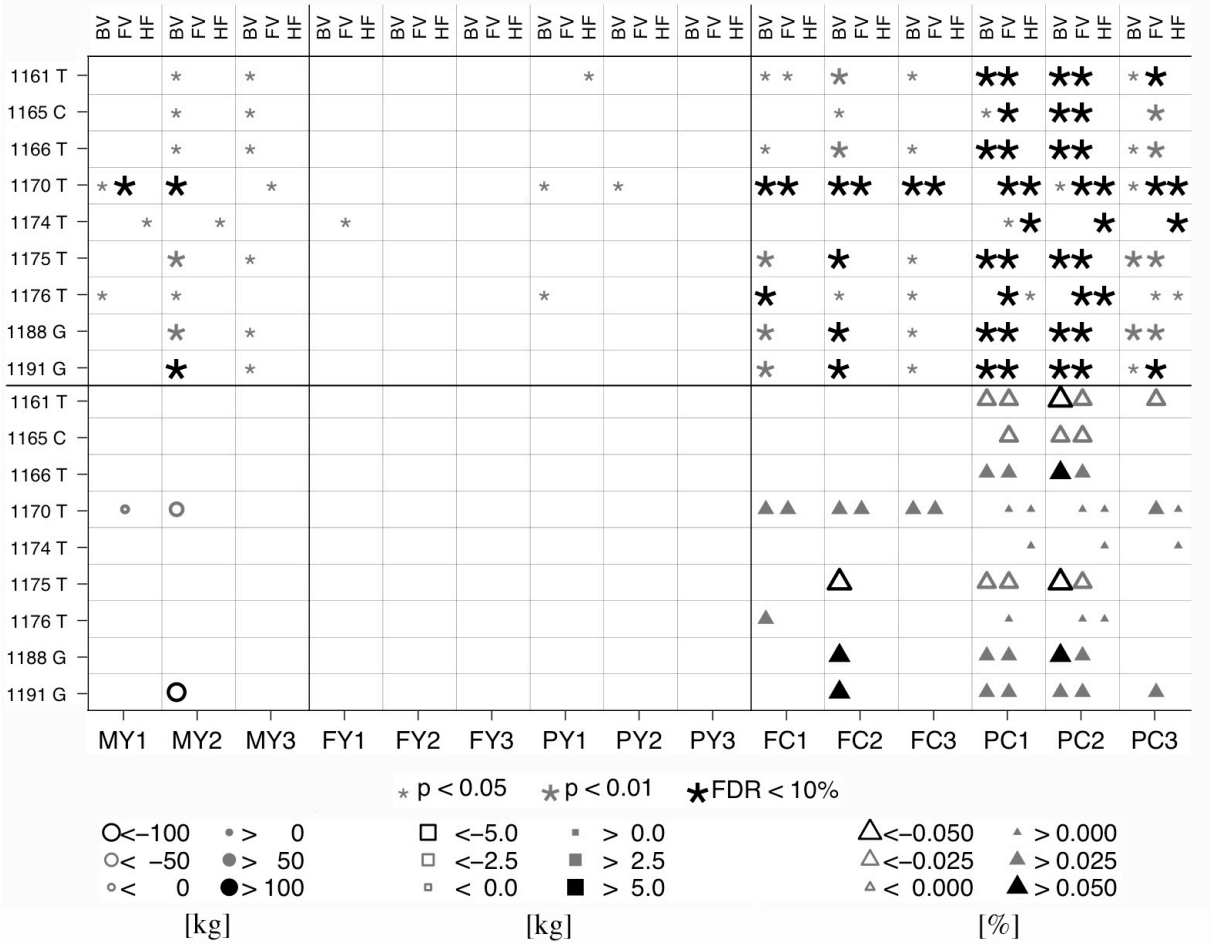


Figure 4.29 Significance level and effects ($\alpha/2$) of *CSN1S1*-markers with *FDR* < 10%

4.8.1.6. Markers of SLC2A9 in single marker analysis

Six markers of *SLC2A9* are significantly associated in BV. Three of the significantly associated markers are located in a translated region (see Figure 4.20). SNP1213 and SNP1215 are non-synonymous polymorphisms and SNP1206 is a synonymous polymorphism. SNP1223, SNP1208 and SNP1199 are significantly associated and have intronic position. SNP1206 exhibits significant effects on MY1, FC2, PC1 and PC2 (see Table 11.5) and suggestive effects ($p < 0.05$, $FDR > 10\%$) on MY2, FY1, PY1, PC1 and PC3

(see Figure 4.30). The T variant has an estimated effect ($\alpha/2$) of +132.2 kg, –0.036% and –0.060% on MY1, PC1 and FC2, respectively. The effects ($\alpha/2$) for the T allele of SNP1215 range from +1.59 kg in PY1 to +3.23 kg in FY2 and for the C allele of SNP1213 from +2.33 kg in FY1 to +3.88 kg in FY2 (see Table 11.5). The *p*-values for SNP1213 on MY2, FY3, PY1 and PY3 are marginally above the threshold of significance ($p < 0.05$, $FDR > 10\%$) and a suggestive association ($p < 0.05$, $FDR > 10\%$) was found for SNP1213 in MY1. SNP1215 is significantly associated with fat- and protein yield across all lactations.

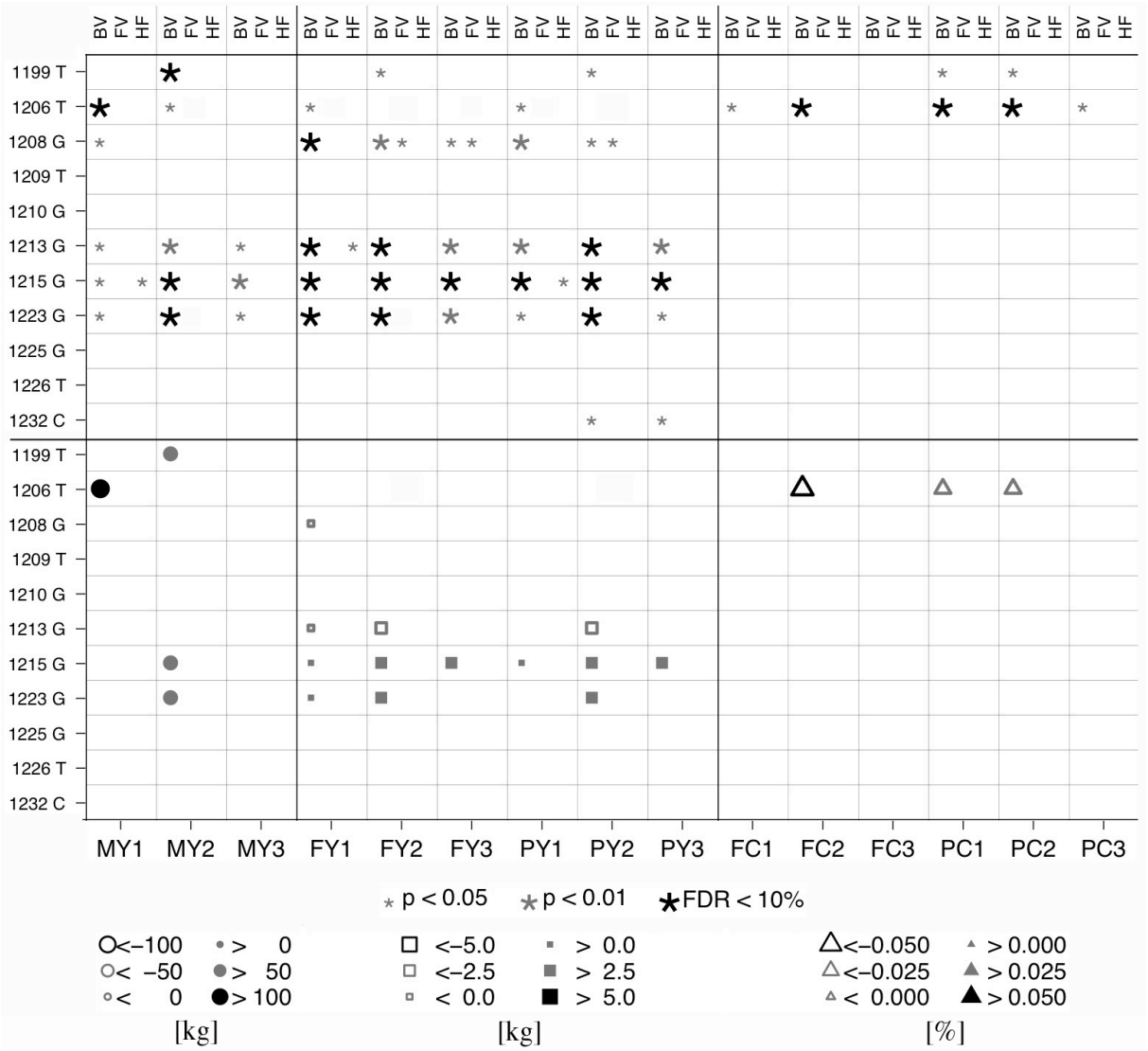


Figure 4.30 Significance level and effects ($\alpha/2$) of *SLC2A9*-markers with *FDR* < 10%

4.8.1.7. Summary of single marker analysis

Across breeds and traits 22 polymorphisms show significant associations and usually more than one marker of a gene is associated. Depending on breed and trait, two markers of *ABCG2* and *PPARGC1A*, respectively, one *OPN*-marker, six *SLC2A9*-markers and all

observed *CSN1S1*-markers are significantly associated. Significant associations of *ABCG2*, *PPARGC1A* and *SLC2A9* are seen mainly with fat- and protein yield. *CSN1S1* markers are significantly associated with protein content in each breed as well as significant effects on milk yield and fat content in BV and FV. The *PPARGC1A*-region is significantly associated across breeds, but significant trait *x* marker combinations are breed-specific. In BV and FV, no *ABCG2*- / *OPN*-marker shows an association, but BV is monomorphic at each marker significantly associated in HF. FV is monomorphic at *ABCG2*-markers significantly associated in HF. Fewer *CSN1S1*-markers are significantly associated in HF than in BV or FV, but HF is monomorphic at four *CSN1S1*-markers (SNP1191, SNP1188, SNP1175, SNP1165) significantly associated in BV and FV.

CSN1S1 and *PGM2* show the most extensive LD among all six genes but the two differ in the frequency of significantly associated SNPs (see Table 4.23): one *PGM2*-marker versus all *CSN1S1*-markers. Thus it is more likely that a QTN will be located in *CSN1S1* than in PGM2.

Table 4.23 Linkage disequilibrium (*D'* average) for gene *x* breed combinations and number of significantly associated markers (*FDR* < 10%) in single marker analysis

	ABCG2			*OPN*			*PPARGC1A*			*PGM2*			*CSN1S1*			*SLC2A9*		
	BV	FV	HF	BV	FV	HF	BV	FV	HF	BV	FV	HF	BV	FV	HF	BV	FV	HF
D'	0,80	0,83	0,77	0,82	0,99	0,99	0,89	0,84	0,86	0,97	0,94	0,89	0,97	0,97	0,92	0,81	0,61	0,65
Trait																		
MY1	-	-	-	-	-	-	2	-	-	-	-	-	-	1	-	1	-	-
MY2	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	3	-	-
MY3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
FY1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-
FY2	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	3	-	-
FY3	-	-	1	-	-	1	-	-	1	-	-	-	-	-	-	1	-	-
PY1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PY2	-	-	2	-	-	1	-	-	1	-	-	-	-	-	-	1	-	-
PY3	-	-	-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-
FC1	-	-	-	-	-	-	-	1	-	-	-	-	4	1	-	1	-	-
FC2	-	-	-	-	-	-	-	-	-	-	-	-	4	1	-	1	-	-
FC3	-	-	-	-	-	-	-	-	-	-	1	-	1	1	-	-	-	-
PC1	-	-	-	-	-	-	-	-	-	-	-	-	5	8	2	-	-	-
PC2	-	-	-	-	-	-	-	-	-	-	-	-	6	8	3	-	-	-
PC3	-	-	-	-	-	-	-	-	-	-	-	-	-	3	2	-	-	-

SNPs with significant effects on fat- and protein content have effects on milk yield in opposite effect sign, so that effects on fat- and protein yield are nor significant.

Associated markers in one breed were compared in their effect directions with other breeds in case of a MAF above 1%. SNPs with significant effects in a given trait in more than one breed exhibited similar effect directions across all breeds. In most cases this was also observed for SNPs that are significantly associated in one breed.

No significantly associated SNP is located in a promoter region, but seven significantly associated SNPs are located within the translated sequence. Five significantly associated SNPs are non-synonymous (SNP1293, SNP1294, SNP1191, SNP1215, SNP1213).
Finally, significantly associated markers were identified across all breeds. In most cases several markers of a gene are significantly associated. Therefore haplotypes were constructed and analysed in linear regression models.

4.8.2. Haplotype analysis (HA)

Haplotype analysis (HA) was carried out as described (see 3.2.9.4.2 Linear regression models in haplotype analysis (HA)) and each significantly associated haplotype was inspected for SNPs that allow discrimination of the significantly associated haplotype from all others. In cases where no SNP was identified allowing the discrimination of the significantly associated haplotype from all others, haplotypes were grouped according to the effect direction in the trait of interest and haplotype groups were screened for discriminating SNPs. SNPs that are significantly associated in SMA but do not discriminate a significantly associated haplotype can be excluded as QTNs, while discriminating SNPs or at least markers in LD with a discriminating SNP could represent a QTN. Furthermore, discriminating SNPs could be used in selection schemes. In the context of selection it would be sufficient to genotype the discriminating SNP only.

4.8.2.1. Haplotypes of ABCG2 in haplotype analysis

Haplotypes constructed with *ABCG2* markers were analysed in BV and HF. Four breed specific significantly associated haplotypes were discovered. ABC_HT3 shows in BV a significant effect ($\alpha/2$) of –154.61 kg on MY3 (see Table 11.9 and Figure 4.31) and non-significant antagonistic effects on content traits, but no specific nucleotide was identified.

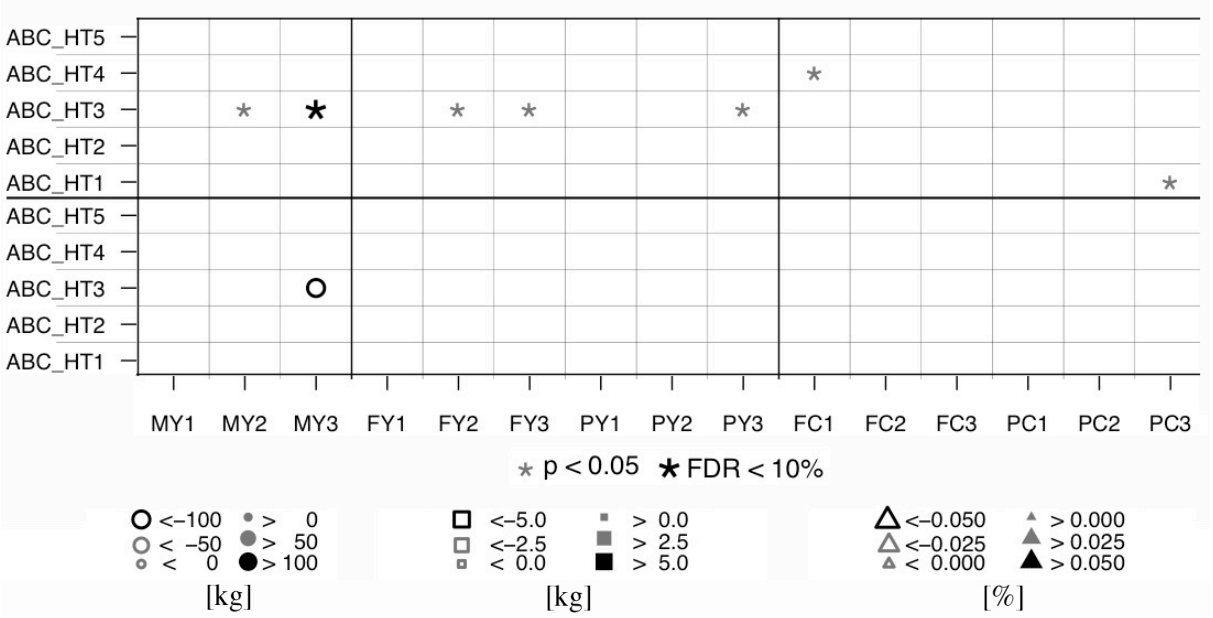


Figure 4.31 Significance level and effects (α/2) in BV of *ABCG2* haplotypes with *FDR* < 10%

Three *ABCG2*-haplotypes are significantly associated in HF. ABC_HT2 is significantly associated with content traits (FC1, FC2, FC3, PC1, PC2, PC3) (see Figure 4.32) and exhibits a non-significant decrease in MY1, MY2 and MY3. However, no specific nucleotide could be detected in ABC_HT2. This is in accordance with SMA, which failed to detect an association for any *ABCG2* marker on any milk content trait (see 4.8.1.1 Markers of *ABCG2* in single marker analysis). ABC_HT6 and ABC_HT8 are associated with protein yield and the effects (α/2), -2.89 kg on PY2 for ABC_HT6 and +2.74 kg on PY3 for ABC_HT8, are significant (see Table 11.17). ABC_HT6 shows suggestive associations ($p < 0.05$, $FDR > 10\%$) with MY3, FY2, FY3 and PY2, and for ABC_HT8 suggestive associations ($p < 0.05$, $FDR > 10\%$) were observed with MY2, FY2, FY3, PY1 and PY3 (see Figure 4.32). Both haplotypes are frequent in HF, but were not detected in BV.

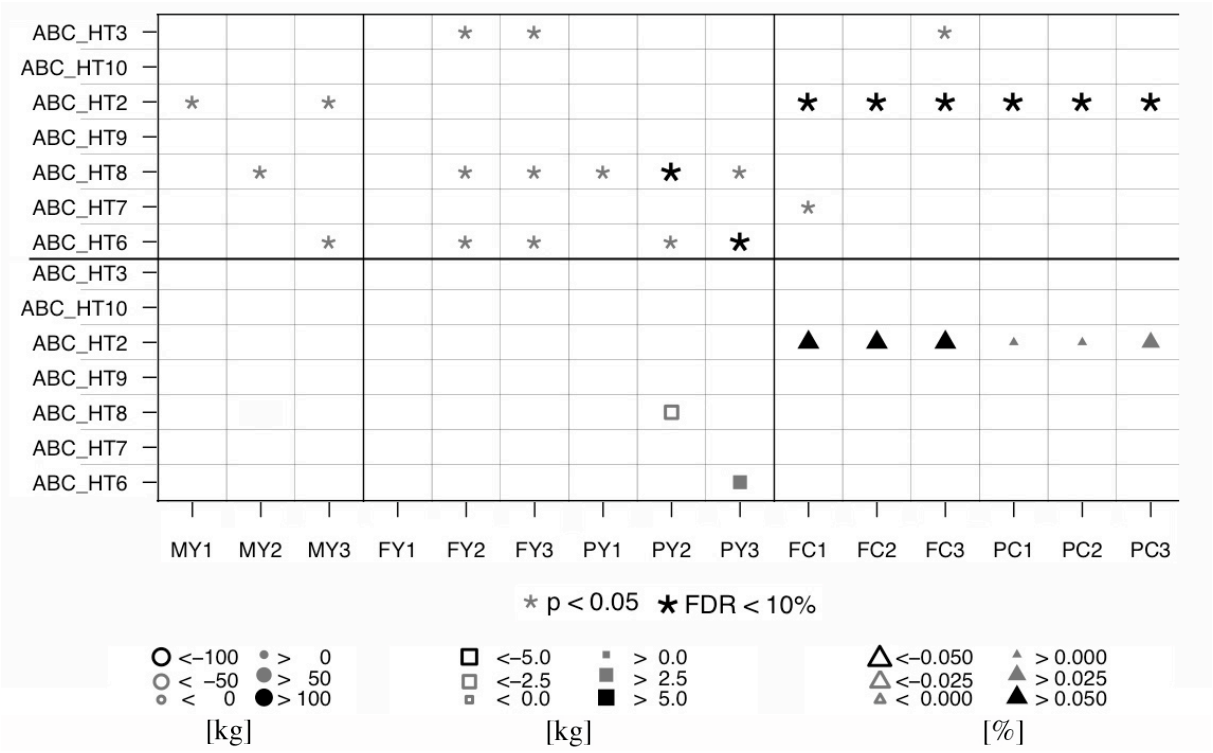


Figure 4.32 Significance level and effects (α/2) in HF of *ABCG2* haplotypes with *FDR* < 10%

SNP1294 tags ABC_HT8 since ABC_HT8 is the only haplotype that carries the G variant of SNP1294. This is the rare variant of SNP1294, which is associated significantly in SMA with a decline of 2.91 in PY2. ABC_HT6 does not carry any specific nucleotide itself, but together with other haplotypes whose effects (α/2) on PY1, PY2 and PY3 have a positive effect sign, the tagging role of SNP1293 becomes evident as Tag-position. Haplotypes whose effects (α/2) on PY1, PY2 and PY3 are trait increasing carry the G variant, haplotypes that carry the A variant of SNP1293 are trait decreasing. This corresponds well to the result from SMA, where a significant increase was detected for the G variant of SNP1293 in PY2. Consequently, both non-synonymous SNPs of *ABCG2*, which were significantly associated in SMA, were identified tagging significantly associated haplotypes.

4.8.2.2. Haplotypes of OPN in haplotype analysis

Haplotype OP_HT4 and haplotype OP_HT5 are significantly associated in HF. OP_HT5 shows significant associations with PC2 (α/2 = –0.017%) and PC3 (α/2 = –0.016). The decline in MY1, MY2 and MY3 is not significant. Compared with trait-increasing OPN-haplotypes in PC2 and PC3, OP_HT5 is not specific at any marker. OP_HT4 is the most

frequent haplotype in HF and significantly associated with FY2, PY2 and PY3 (see Figure 4.33). The effects ($\alpha/2$) range from +1.95 kg in PY2 to +3.18 kg in FY3 (see Table 11.18). OPN-haplotypes with negative effects on FY3, PY2, and PY3 differ from OP_HT5 in the T allele at SNP1331. In SMA, the C allele of SNP1331 is significantly associated with an increase in FY3, PY2 and PY3.

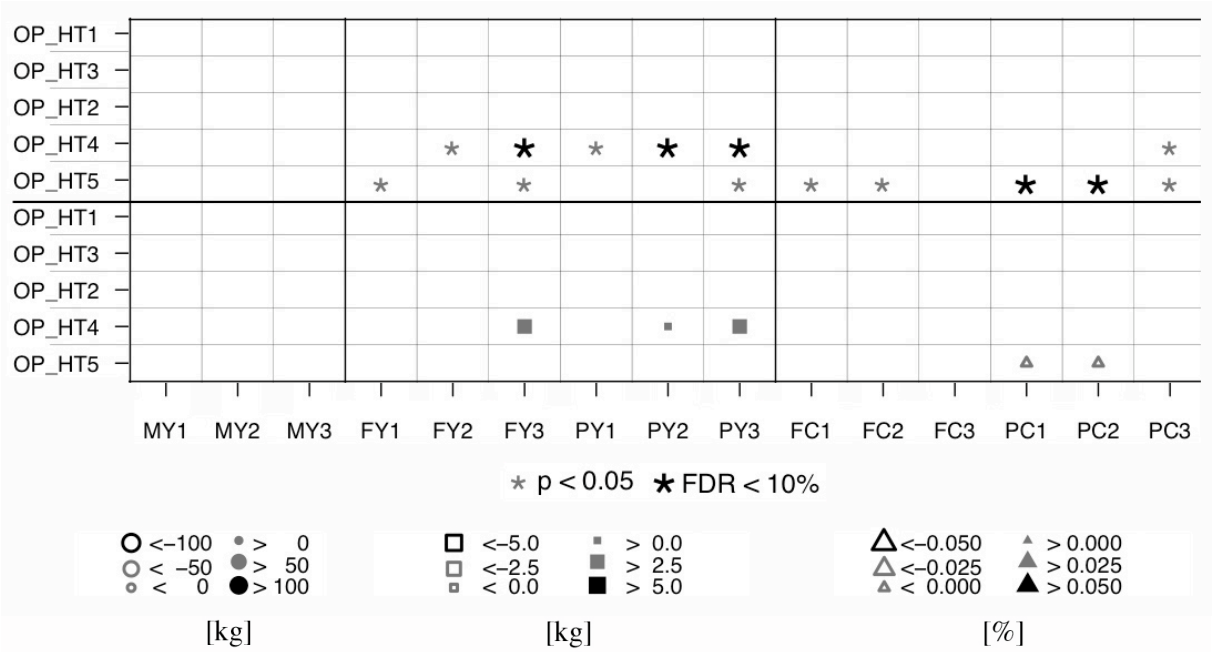


Figure 4.33 Significance level and effects ($\alpha/2$) in HF of *OPN* haplotypes with *FDR* < 10%

4.8.2.3. Haplotypes constructed with ABCG2- and OPN-polymorphisms in haplotype analysis

Haplotypes constructed with markers of *ABCG2* and *OPN* were analysed exclusively in FV (see 4.7 Constructed haplotypes). Haplotype ABOP_HT1 exhibits a significant decrease in PC1 and PC2 of –0.038% and –0.040%, respectively (see Table 11.13). Suggestive associations ($p < 0.05$, $FDR > 10\%$) on MY1, MY2 and MY3 are antagonistic and diminish effects on fat- or protein yield (see Figure 4.34). ABOP_HT1 is the only haplotype in FV with a decline in PC1, PC2 and PC3, but no SNP was identified that tags ABOP_HT1. SMA in FV yielded no significant association on protein content for any *ABCG2*- or *OPN*-marker.

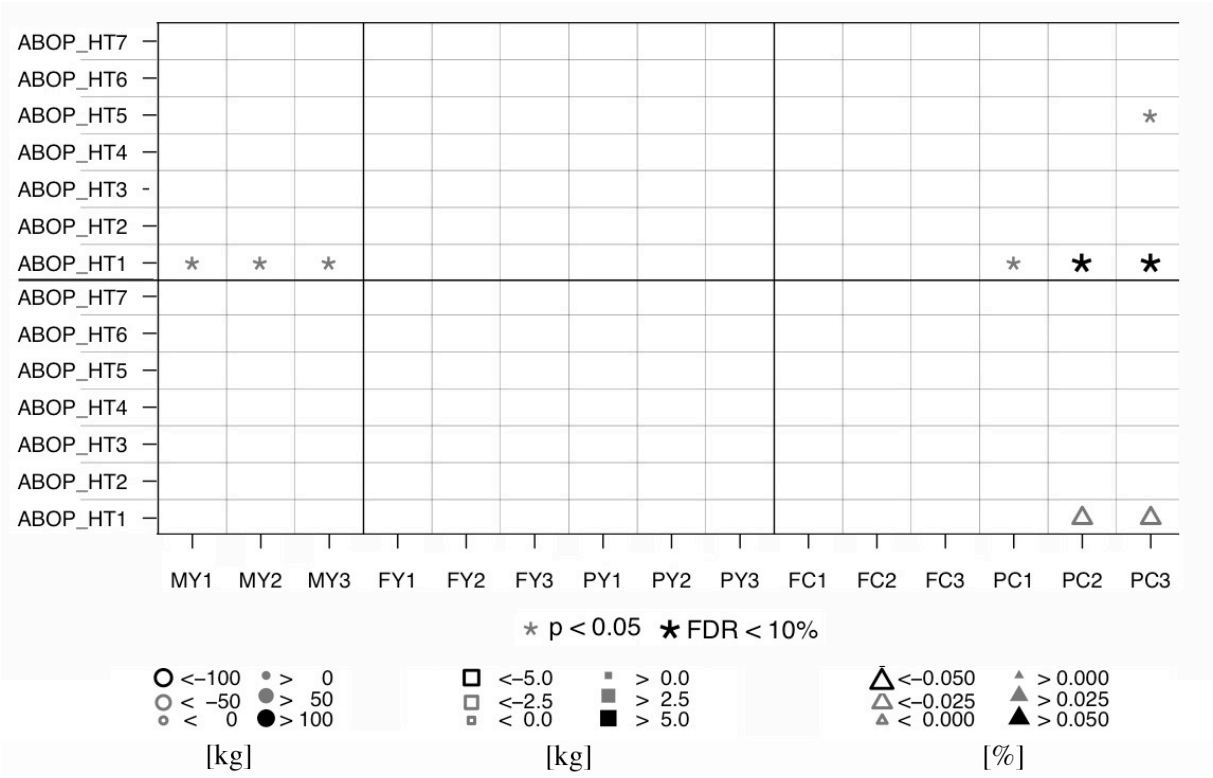


Figure 4.34 Significance level and effects (α/2) in FV of haplotypes incorporating *ABCG2* and *OPN* markers with *FDR* < 10%

4.8.2.4. Haplotypes of PPARGC1A in haplotype analysis

Eight haplotypes were constructed with *PPARGC1A*-markers of which six are significantly associated (see Table 11.10, Table 11.14 and Table 11.19). PPARG_HT4 exhibits significant effects in HF on FC1 (α/2 = +0.039%) (see Table 11.19 and Figure 4.37) and in FV on FC1 (α/2 = +0.027%), FC2 (α/2 = +0.028%) and FC3 (α/2 = +0.036%) (see Table 11.14 and Figure 4.36). Antagonistic effects on MY1, MY2 and MY3 were observed in both breeds but were only significant on MY1 in FV. Identical results from HA in two breeds however, did not result in an identical SNP being discriminatory for PPARG_HT4 in both breeds. SNP1238 tags PPARG_HT4 in FV, while SNP1247 tags PPARG_HT4 in HF. The A allele at SNP1238 is specific for PPARG_HT4 in FV, the C allele at SNP1247 is specific for PPARG_HT4 in HF. Both polymorphisms showed comparable effects on FC1, FC2 and FC3 in SMA, but only SNP1238 was significantly associated.

Haplotype PPARG_HT2 was identified with a significant effect of +2.23 kg (α/2) in FY2 in BV (see Table 11.10 and Figure 4.35). The identification of an SNP tagging PPARG_HT2 was ambiguous. Regarding only the frequent haplotypes, PPARG_HT2 was the only

haplotype in BV carrying the A variant at SNP1242 and the G variant at SNP1238 (see 11.3 Haplotypes). Both polymorphisms were identified significantly in SMA of BV with an effect of +1.23 kg on FY2 for the G allele of SNP1238 and +1.01 kg for the G allele of SNP1242.

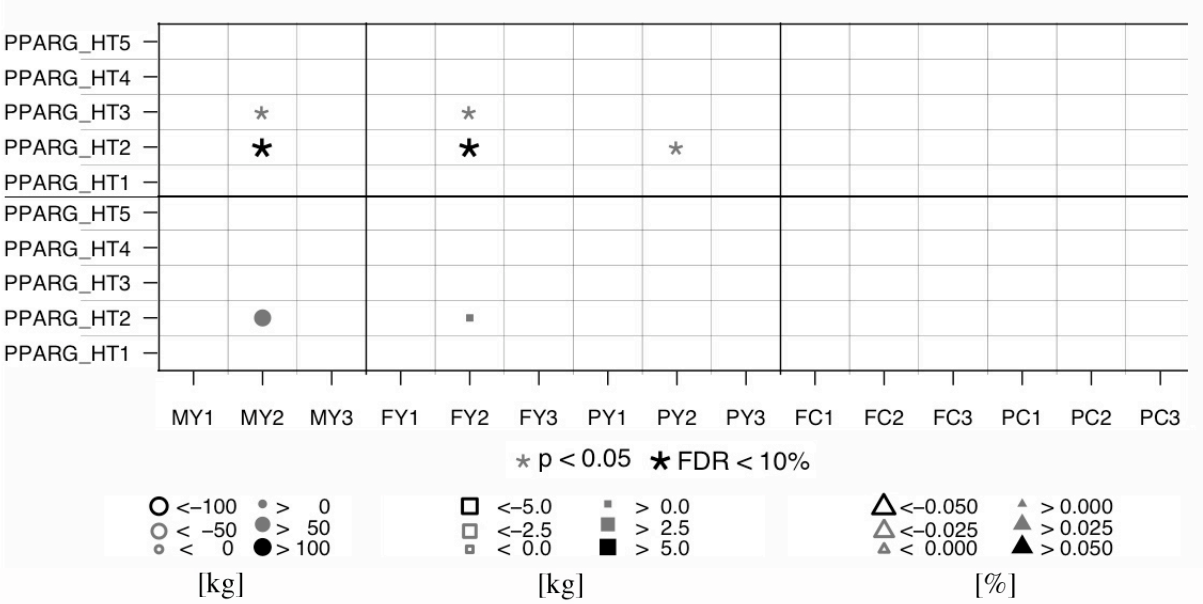


Figure 4.35 Significance level and effects ($\alpha/2$) in BV of *PPARGC1A* haplotypes with *FDR* < 10%

PPARG_HT6 shows significant effects ($\alpha/2$) on FC2 (–0.024%) and on FC3 (–0.021%) (see Table 11.14 and Figure 4.36) in FV. No nucleotide is specific for PPARG_HT6, but PPARG_HT6 carries the G variant at SNP1238 and has the largest effect deviation in FC2 and FC3 from the PPARG_HT4-effect among all haplotypes with the G allele at SNP1238. This explains the significant association of PPARG_HT6.

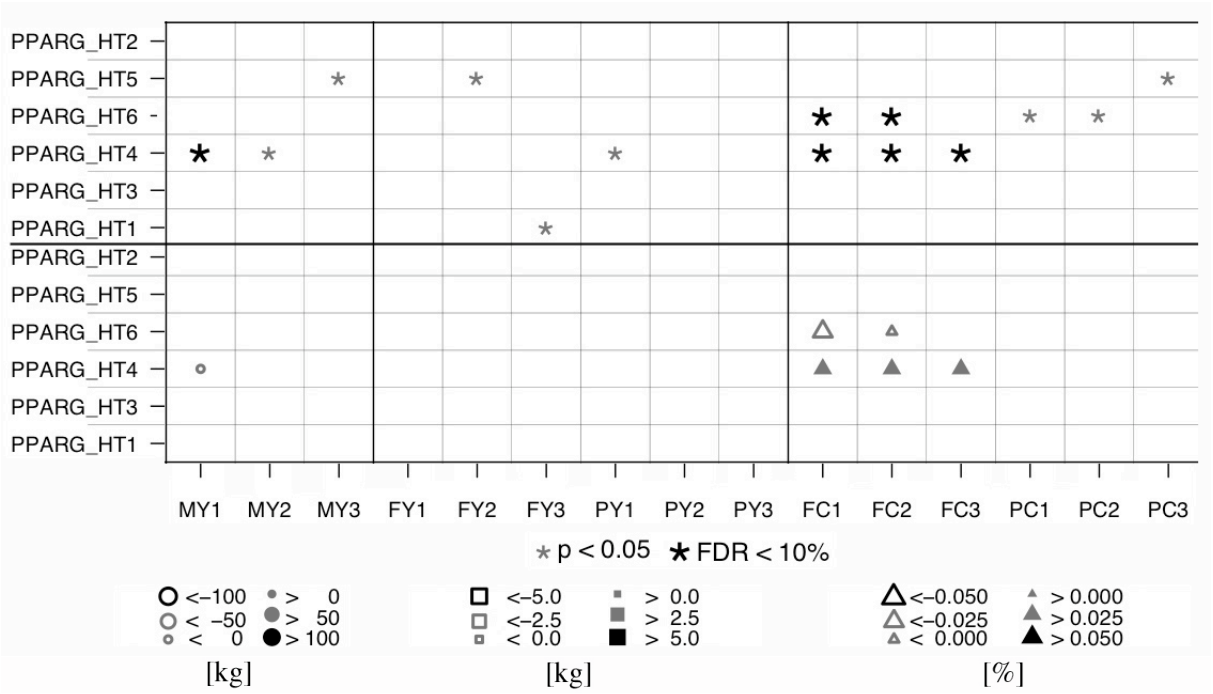


Figure 4.36 Significance level and effects (α/2) in FV of *PPARGC1A* haplotypes with *FDR* < 10%

In HF, haplotype PPARG_HT7 shows significant effects (α/2) on FY3 (+7.25 kg), PY2 (+4.32 kg), PY3 (+6.27 kg) and PC3 (+0.031%). Associations with MY3, FY2 and PY1 are suggestive ($p < 0.05$, $FDR > 10\%$) (see Figure 4.37 and Table 11.19). The T nucleotide at SNP1239 is haplotype-discriminating, and was also significantly associated in SMA with significantly increased FY2, PY2 and PY3. The significant effect of PPARG_HT8 on FC1 and FC2 in HF is explained by the largest effect deviation in FC1 and FC2 from the effects of PPARG_HT4. PPARG_HT8 carries the A allele and PPARG_HT4 the C allele at SNP1247.

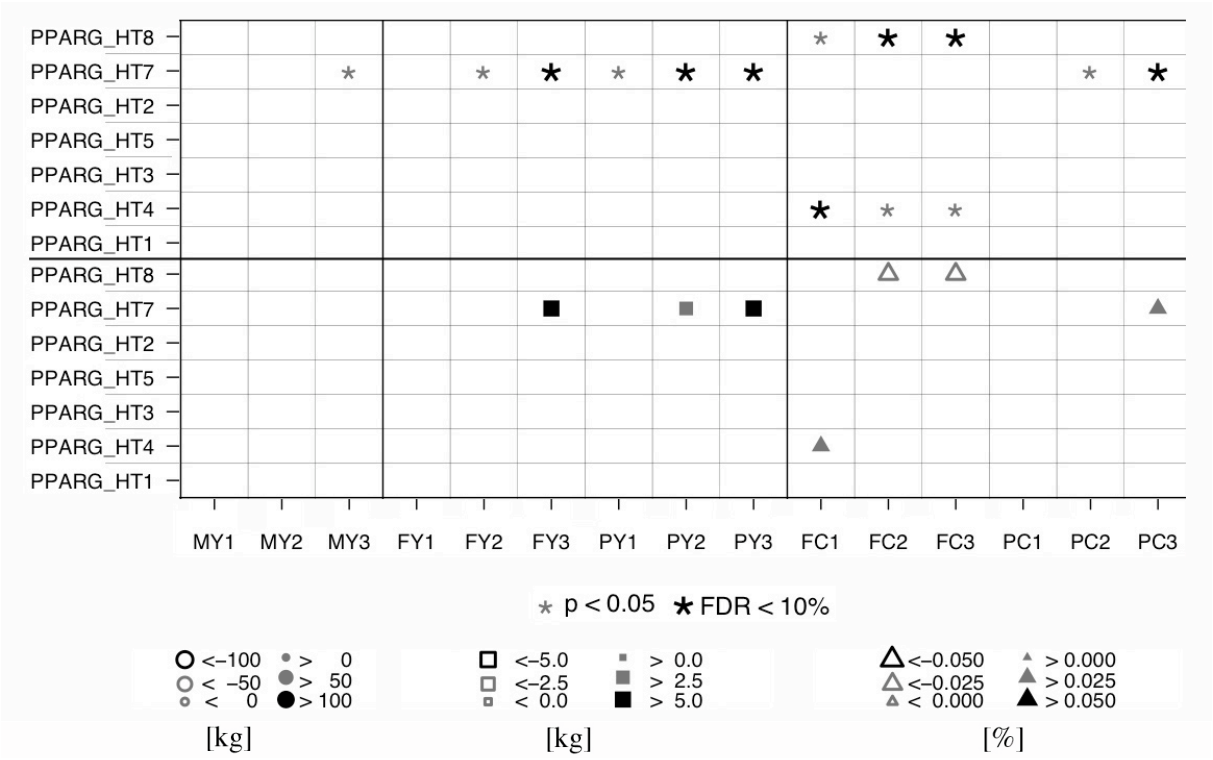


Figure 4.37 Significance level and effects ($\alpha/2$) in HF of *PPARGC1A* haplotypes with *FDR* < 10%

4.8.2.5. Haplotypes of PGM2 in haplotype analysis

In HF, HA found PGM2_HT6 to be significantly associated with a decreased PC3 ($\alpha/2$ = –0.028%) (see Table 11.20). PGM2_HT6 and PGM2_HT2 carry the A allele at SNP1321, but *PGM2* haplotypes with increasing effects on PC1, PC2, and PC3 carry the G allele. SNP1321 is an A-G polymorphism located in a previously identified CpG-island in the 5' end region of PGM2. SNP1321 is not significantly associated in SMA, but associations with PC1, PC2 and PC3 are suggestive ($p < 0.05$, $FDR > 10\%$) (see Figure 4.38).

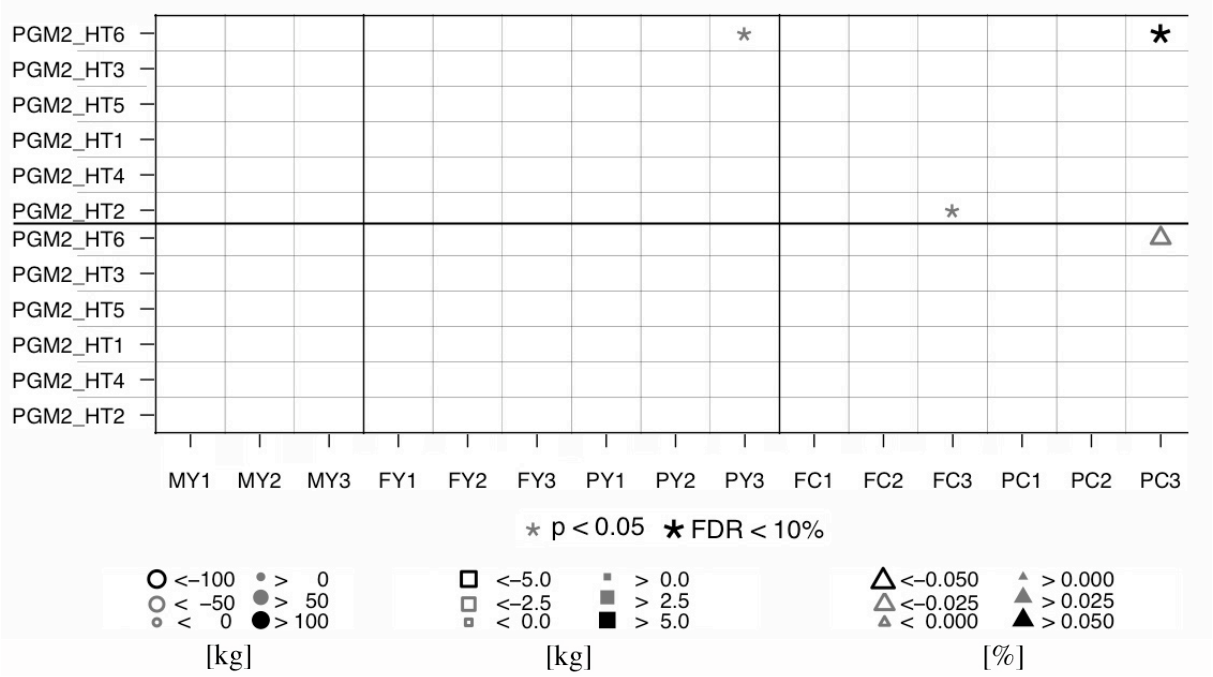


Figure 4.38 Significance level and effects ($\alpha/2$) in HF of *PGM2* haplotypes with *FDR* < 10%

4.8.2.6. Haplotypes of CSN1S1 in haplotype analysis

Across all breeds and traits, five among six *CSN1S1*-haplotypes exhibit significant effects. Three haplotypes show significant associations in more than one breed: CSN1S_HT2 in all breeds and CSN1S_HT4 and CSN1S_HT5 in BV and FV, respectively.

CSN1S_HT2 indicates significant associations on MY1, MY2, FC1, FC2 and FC3 in BV (see Figure 4.39), in FV on all content traits (FC1, FC2, FC3, PC1, PC2, PC3) (see Figure 4.40) and in HF on PC1, PC2 and PC3 (see Figure 4.41). The effects on fat percentage vary in BV from -0.036% in FC3 to -0,040% in FC1 (see Table 11.11) and from -0,024% in FC2 to -0.035% in FC3 in FV. The FV-effects on protein content range from -0.023% in PC1 to –0.026% in PC3 (see Table 11.15), and HF-effects range from –0,013% in PC1 to –0.016% in PC2 (see Table 11.21). Antagonistic effects on MY1, MY2 and MY3 are observed across all breeds, but are significant (MY1, MY2) in BV, only. Across all breeds, INDEL1170 tags CSN1S_HT2. The deletion variant of SNP1170 is significantly associated in SMA with a decrease across all content traits and breeds (see Figure 4.29).

CSN1S_HT4 exhibits significant effects in BV on MY2, FC2, PC1 and PC2 and in FV on PC1, PC2, and PC3, (see Figure 4.39 and Figure 4.40). CSN1S_HT4 is rare in BV and FV and in HF the frequency is below 2.5% (see 11.3 Haplotypes). The effects range in FV from +0.030% in PC2 to +0.037% in PC3 (see Table 11.15) and in BV from +0.035% in PC1 to +0.056% in FC2 (see Table 11.11). Antagonistic effects on milk yield are significant

exclusively in MY2 (-134,36 kg) in BV. CSN1S_HT4 is tagged by each of the following six *CSN1S1* markers: SNP1191, SNP1188, SNP1175, SNP1166, SNP1165, SNP1161.

The significant effects of CSN1S_HT5 on FC1, FC2 in FV and on FC1, FC2, FC3, PC2, PC3 in BV do not correspond with any specific nucleotide.

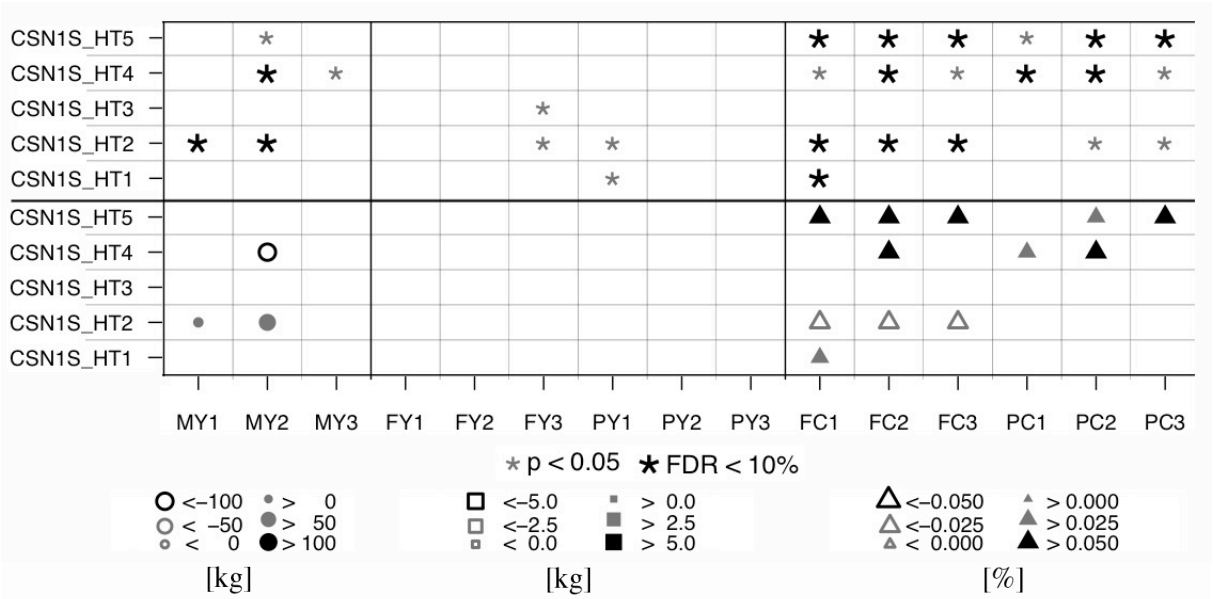


Figure 4.39 Significance level and effects ($\alpha/2$) in BV of *CSN1S1* haplotypes with *FDR* < 10%

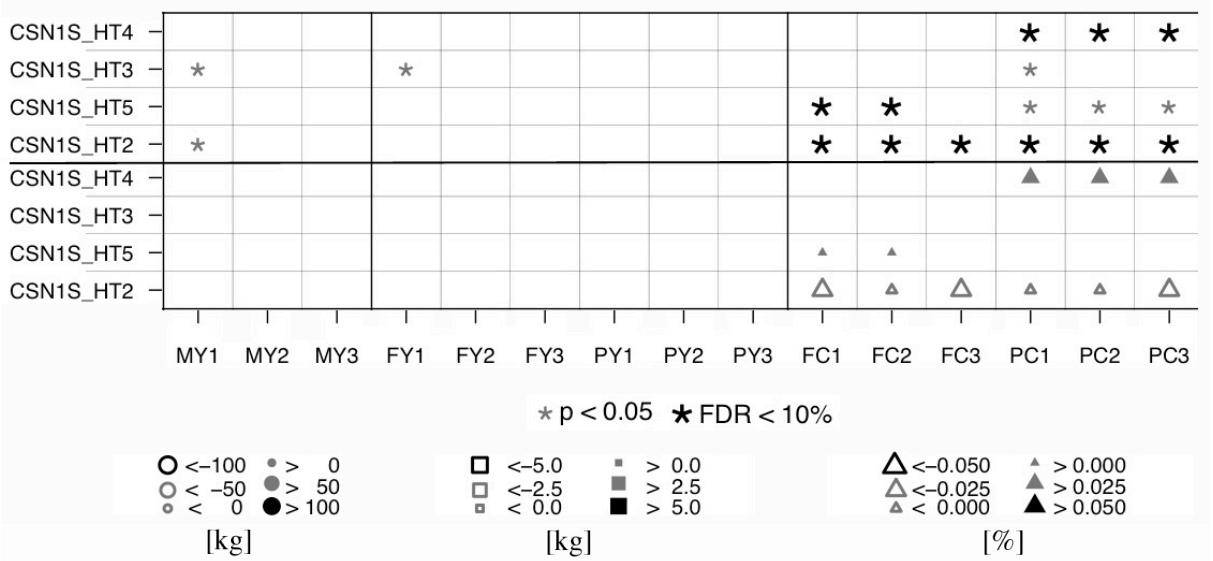


Figure 4.40 Significance level and effects ($\alpha/2$) in FV of *CSN1S1* haplotypes with *FDR* < 10%

CSN1S_HT3 appears across all breeds, but shows significant associations in HF exclusively and with the traits PC1, PC2 and PC3 (see Figure 4.41). Effects range from +0.018% on PC1 to +0.020% on PC2 (see Table 11.21) and CSN1S_HT3 is tagged by the T allele of SNP1174. SMA revealed that the T variant of SNP1174 increases PC1, PC2 and PC3 significantly in HF. Both significantly associated *CSN1S1* haplotypes in HF are associated with PC1, PC2 and PC3.

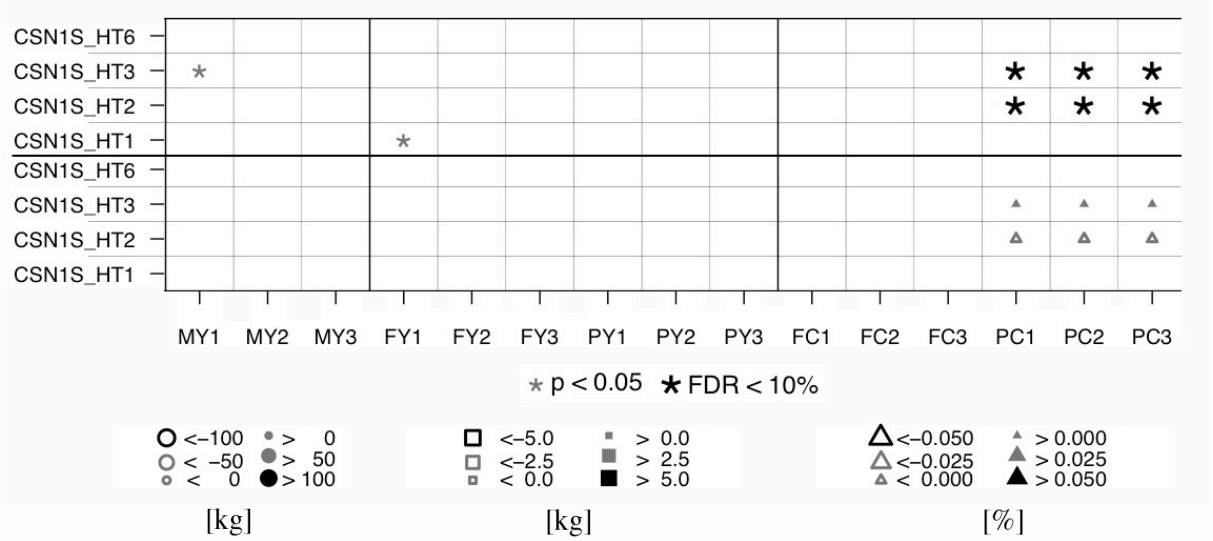


Figure 4.41 Significance level and effects (α/2) in HF of *CSN1S1* haplotypes with FDR < 10%

4.8.2.7. Haplotypes of SLC2A9 in haplotype analysis

SLC2A9-markers yielded 18 inferred haplotypes, of which eight show significant associations. SLC2_HT8 exhibits significant effects in BV on PC1 (+0.013%) and in FV on PC1 (+0.013%) and PC2 (+0.014%). The effects on milk yield are antagonistic but not significant in both breeds (see Table 11.12 and Table 11.16). In neither of these breeds was it possible to identify markers that tag SLC2_HT8.

HA in BV showed significant effects of the BV-specific SLC2_HT1 on MY2, FY1, FY2 and PY2 (see Figure 4.42) with the following effects: +2.37 kg on FY1, +4.29 kg on FY2, +3.09 kg on PY2 and +92.27 kg on MY2 (see Table 11.12). SNP1213 (C allele) and SNP1223 (G allele) tag SLC2_HT1. The C allele of SNP1213 significantly increased FY1, FY2 and PY2 and the G variant of SNP1223 significantly increased MY2, FY1, FY2 and PY2 in SMA (see 4.8.1.6 Markers of SLC2A9 in single marker analysis). SLC2_HT7 exhibits significant effects in BV on MY1 (+ 118.83 kg) and on FC2 (–0.058%) (see Figure 4.42). SLC2A_HT7 appears in BV at a frequency of 4%, and is tagged by SNP1206. An association of SNP1206 was

observed in SMA on MY1, FC1, PC1 and PC2. Haplotype SLC2_HT2 exhibits a significantly declining effect on MY2. The effects of SLC2_HT2 and SLC2_HT1 on MY2 differ most among all haplotypes: both carry opposite alleles at SNP1223 and SNP1223. The situation for SLC2_HT8 is similar, causing increased PC1. SLC2_HT8 carries the T allele at SNP1206 and its PC1-effect differs most from the effect of SLC2_HT7. Both carry opposite nucleotides at SNP1206.

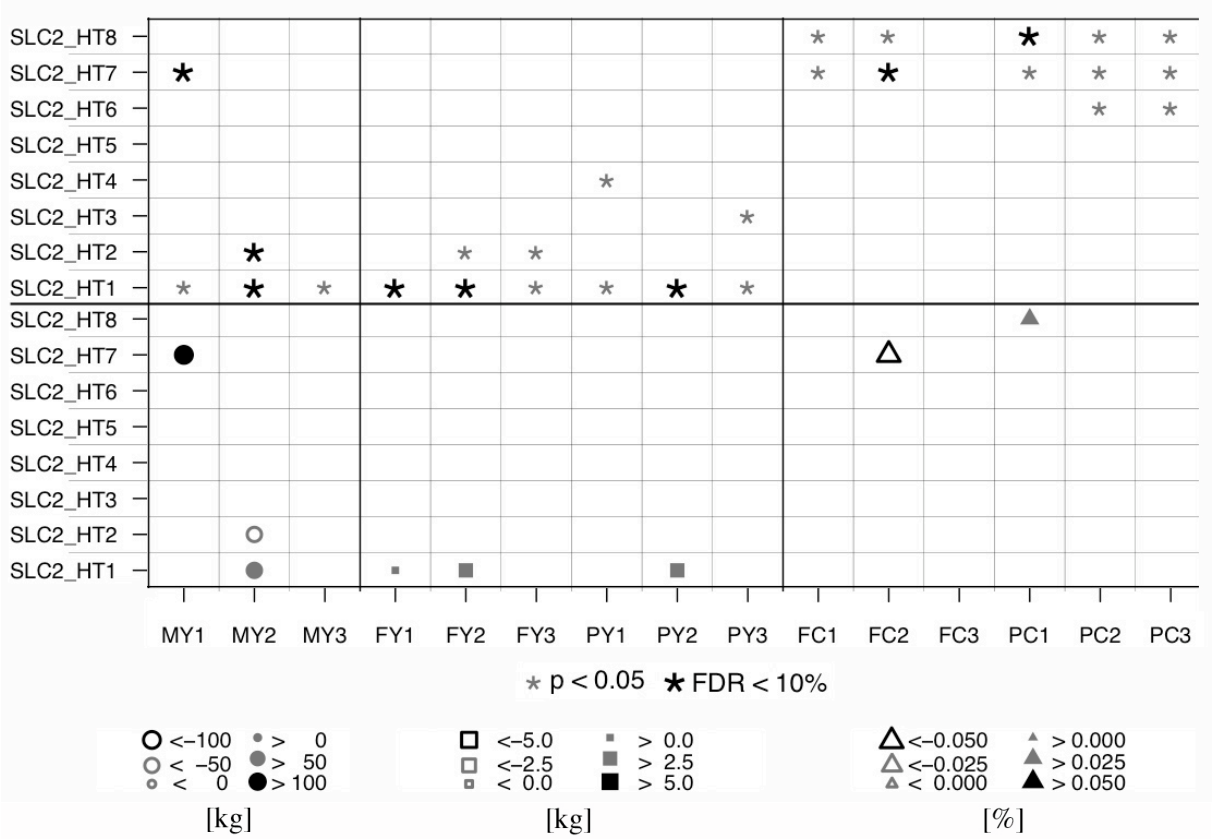


Figure 4.42 Significance level and effects ($\alpha/2$) in BV of *SLC2A9* haplotypes with *FDR* < 10%

Four SLC2A9-haplotyoes are significantly associated in FV (SLC2_HT8, SLC2_HT3, SLC2_HT12, SLC2_HT16) (see Figure 4.43) but no SNP was identified as tagging one of these haplotypes.

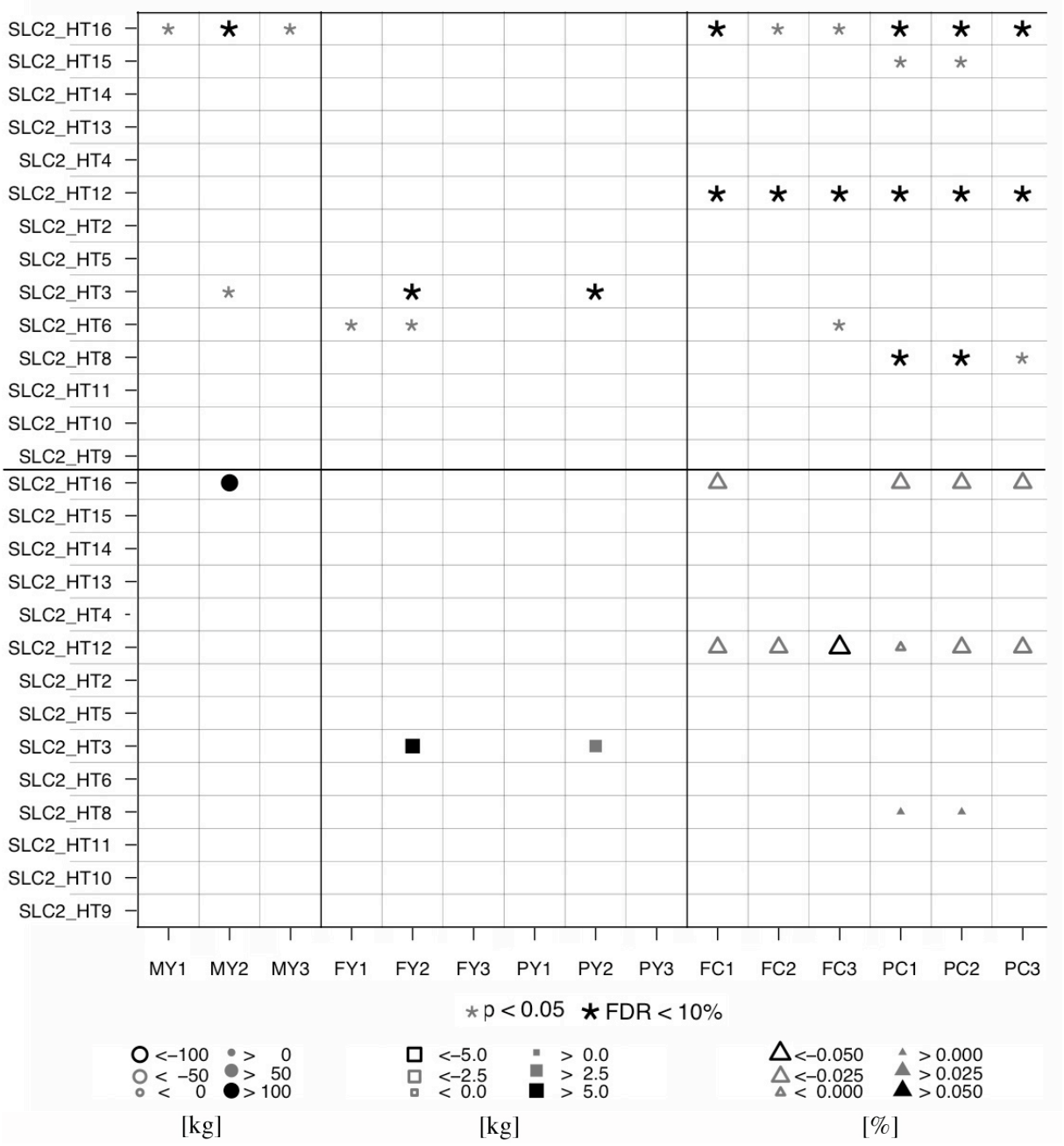


Figure 4.43 Significance level and effects ($\alpha/2$) in FV of *SLC2A9* haplotypes with *FDR* < 10%

In HF, SLC2_HT2 is associated significantly with FY3 (see Table 11.22 and Figure 4.44) but a haplotype discriminating SNP was not identified.

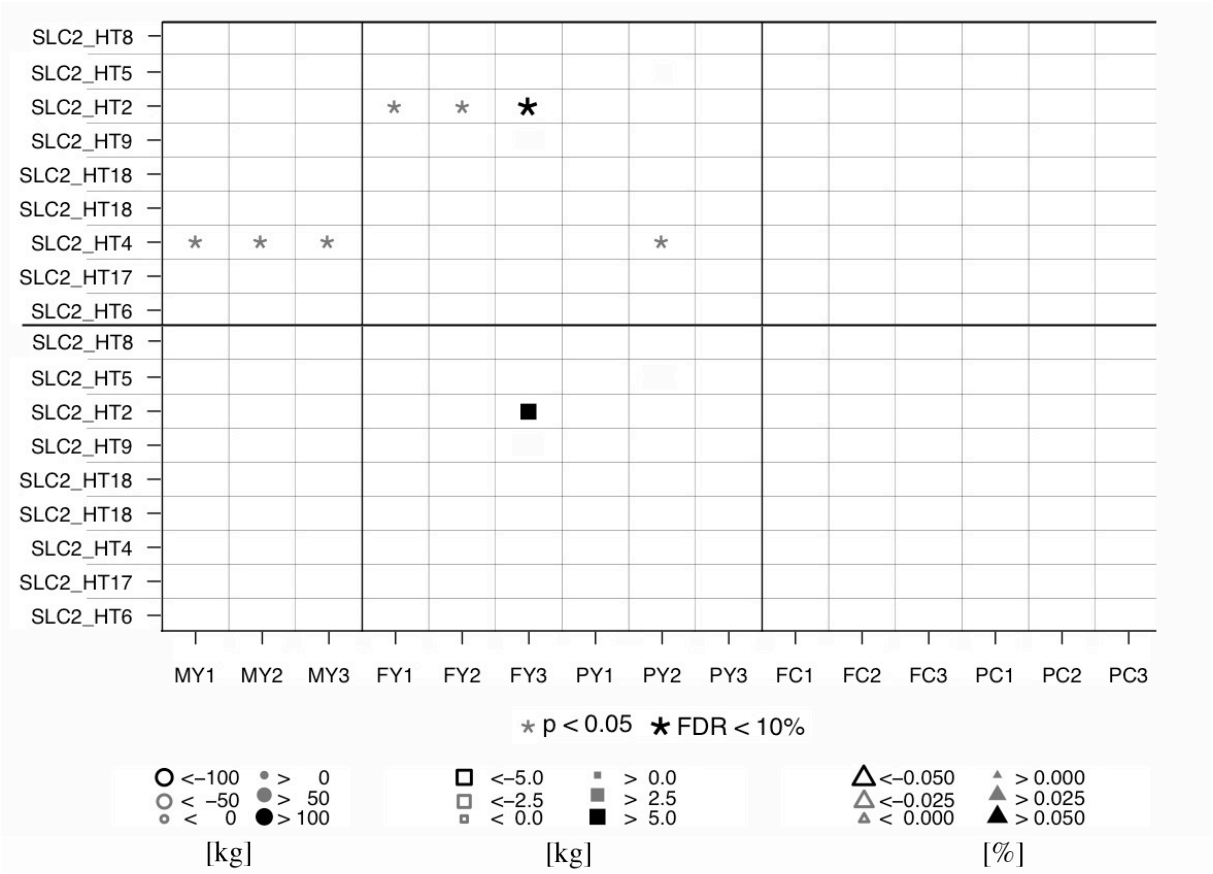


Figure 4.44 Significance level and effects (α/2) in HF of *SLC2A9* haplotypes with *FDR* < 10%

4.8.2.8. Summary of haplotype analysis

Across all breeds and traits 25 haplotypes are significantly associated, of which 18 significantly associated haplotypes can be tagged by haplotype discriminating SNPs. Similar effects of the same haplotype in different breeds provide evidence for a causative mutation in the gene. In contrast, if a haplotype is significantly associated only in one breed it is more likely that the effects result from closely linked genes.

CSN1_HT2, which is discriminated from other CSN1S1-haplotypes with the deletion variant of INDEL1170, is significantly associated with content traits across all breeds. CSN1S_HT4 is associated with protein percentage in BV and FV and is tagged in both breeds by one of the following SNPs: SNP1191, SNP1188, SNP1175, SNP1166, SNP1165, SNP1161. PPARG_HT4 is associated with fat content in FV and HF, but haplotype discriminating markers are breed-specific. CSN1S_HT5 is significantly associated with fat content in BV and FV, and SLC2A_HT8 has significant effects on protein content in BV and FV but CSN1S_HT5 and SLC2A_HT8 are not tagged by a discriminating marker. Associations of other haplotypes were discovered in a single breed only.

Effects of alleles that tag a significantly associated haplotype were compared with allele effects from SMA with regard to magnitudes and directions. Across breeds and traits, 17 markers significantly associated in SMA are discriminatory for a significantly associated haplotype (see Table 4.24).

Table 4.24 Markers with significant association in single marker analysis (SMA) while being discriminatory for significantly associated haplotypes in haplotype analysis (HA)

Breed	SNP	Associated traits (FDR < 10%) in SMA	Discriminated haplotype	Associated traits (FDR < 10%) in HA
BV	1242	MY2, FY2	PPARG_HT2	MY2, FY2
	1238	MY2, FY2	PPARG_HT2	
	1191	MY2, FC2, PC1, PC2	CSN1S_HT4	
	1188	FC1, FC2, PC1, PC2	CSN1S_HT4	
	1175	FC1, FC2, PC1, PC2	CSN1S_HT4	MY2, FC2, PC1, PC2
	1166	PC1, PC2	CSN1S_HT4	
	1165	PC2	CSN1S_HT4	
	1161	PC1, PC2	CSN1S_HT4	
	1170	MY2, FC1, FC2, FC3	CSN1S_HT2	MY1, MY2, FC1, FC2, FC3
	1223	MY2, FY1, FY2, PY2	SLC2_HT1	MY2, FY1, FY2, PY2
	1213	FY1, FY2, PY2	SLC2_HT1	
	1206	MY1, FC2, PC1, PC2	SLC2_HT7	MY1, FC2
FV	1238	FC1	PPARG_HT4	MY1, FC1, FC2, FC3
	1191	PC1, PC2, PC3	CSN1S_HT4	
	1188	PC1, PC2	CSN1S_HT4	
	1175	PC1, PC2	CSN1S_HT4	PC1, PC2, PC3
	1166	PC1, PC2	CSN1S_HT4	
	1165	PC1, PC2	CSN1S_HT4	
	1161	PC1, PC2, PC3	CSN1S_HT4	
	1170	MY1, FC1, FC2, FC3, PC1, PC2, PC3	CSN1S_HT2	FC1, FC2, FC3, PC1, PC2, PC3
HF	1293	FY3, PY2	ABC_HT6	PY3
	1294	PY2	ABC_HT8	PY2
	1331	FY3, PY2, PY3	OP_HT4	FY3, PY2, PY3
	1239	FY3, PY2, PY3	PPARG_HT7	FY3, PY2, PY3, PC3
	1170	PC1, PC2, PC3	CSN1S_HT2	PC1, PC2, PC3
	1174	PC1, PC2, PC3	CSN1S_HT3	PC1, PC2, PC3

Six markers significantly associated in SMA do not tag a significantly associated haplotype (see Table 4.25). Associations of these SNPs due to incomplete LD with (a) causative mutation(s) are more probable than a causative function of the marker itself.

Table 4.25 Markers with significant association in single marker analysis (SMA) while being non-discriminatory for significantly associated haplotypes

Breed	Gene	SNP	Associated traits (FDR < 10%) in SMA
BV	*CSN1S1*	1176	FC1
	SLC2A9	1215	MY2, FY1, FY2, FY3, PY1, PY2, PY3
	SLC2A9	1208	FY1
	SLC2A9	1199	MY2
FV	*PGM2*	877	FC3
	CSN1S1	1176	FC1, FC2
HF	*CSN1S1*	1176	PC2

HA in HF also identified markers, SNP1247 and SNP1321, that discriminate a significantly associated haplotype but that are without significant association in SMA (see Table 4.26). However, *p*-values in SMA are either marginally above the threshold of significance ($p < 0.01$, $FDR > 10\%$) or suggestive ($p < 0.05$, $FDR > 10\%$) (see Figure 4.27 and Figure 4.28).

Table 4.26 Discriminatory markers of significantly associated haplotypes in haplotype analysis (HA) without significant association in single marker analysis

Breed	SNP	Discriminated haplotype	Significantly associated trait in HA
HF	1247	PPARG_HT4	FC1
	1321	PGM2_HT6	PC3

In conclusion, nine haplotypes showed significant associations but the identification of specific markers failed (see Table 4.27). Haplotypes capture information about the entire chromosomal region spanned by the incorporated markers. Therefore significantly associated haplotypes, indicate a causative mutation in the region although not being discriminated by a single polymorphism. Such a causative mutation is expected to be in strong LD with the haplotype, but LD to each single marker is not extensive enough to appear in SMA at one of the analysed markers. ABC_HT2 in HF and SLC_HT12 in FV are significantly associated with all content traits without being tagged by a single marker. None of the *ABCG2*- or *SLC2A9*-markers shows significant association with a content trait in SMA either in HF or FV. Extensive LD was seen within genes (see 4.6 Linkage Disequilibrium between candidate genes). Due to this the causative mutation is more likely to be located in an adjacent gene than in the gene that is investigated.

Table 4.27 Associated haplotypes (FDR < 10%) in haplotype analysis (HA) without discriminating polymorphism

Breed	Haplotype	Significantly associated traits in HA
BV	ABC_HT3	MY3
	CSN1S_HT1	FC1
	CSN1S_HT5	FC1, FC2, FC3, PC2, PC3
FV	ABOP_HT1	PC2, PC3
	CSN1S_HT5	FC1, FC2
	SLC2_HT3	$FY2_{FV}$, $PY2_{FV}$
	SLC2_HT12	$FC1_{FV}$, $FC2_{FV}$, $FC3_{FV}$, $PC1_{FV}$, $PC2_{FV}$, $PC3_{FV}$
	SLC2_HT16	$FC1_{FV}$, $PC1_{FV}$, $PC2_{FV}$, $PC3_{FV}$
HF	ABC_HT2	FC1, FC2, FC3, PC1, PC2, PC3
	OP_HT5	PC1, PC2

More than one marker for *ABCG2*, *PPARGC1A*, *CSN1S1* and *SLC2A9* discriminate a significantly associated haplotype and are significantly associated in SMA. Thus it is uncertain whether markers within the same gene exhibit separate effects or represent correlated effects due to LD. To identify correlated effects between markers of the same gene, MMA was carried out.

4.8.3. Multi marker analysis (MMA)

Multi Marker Analysis (MMA) and *AIC*-optimisation was applied to identify whether markers are associated due to correlated effects caused by extensive LD. MMA was used to descry distinct effects of markers that are significantly associated in SMA and discriminate a significantly associated haplotype in HA. Therefore the threshold of significance was set at $p < 0.05$.

In the first step and in case of several markers of a same gene that are significant associated in SMA and have a discriminating feature of a significantly associated haplotype, MMA was carried out within genes by fitting markers of the same gene. MMA within genes was performed in traits and breeds in which the markers were obtained in SMA to be significantly associated. Table 4.28 lists investigated breed *x* marker *x* trait combinations.

Table 4.28 Polymorphisms fitted in multi marker models within genes

Gene	SNPs	Breed(s)	Trait
ABCG2	1293, 1294	HF	FY, PY
PPARGC1A	1242, 1238	BV	MY, FY
CSN1S1	1191, 1188, 1170, 1166, 1165, 1161	BV, FV	MY, FC, PC
	1174, 1170	HF	PC
SLC2A9	1213, 1206	BV	MY, FY, PY, FC, PC

Secondly, MMA was performed across genes in order to detect correlated effects due to LD between genes. Models in MMA across genes included markers that were identified in MMA within genes to be associated with distinct effects. MMA across genes was performed in breeds for milk production traits and for the total merit index for milk production, so-called MW in BV and FV (BLT 2007), and RZM in HF (VIT 2004). In the following in this work, the total merit index for milk production is called MW across breeds.

4.8.3.1. Markers of ABCG2 in multi marker analysis

SNP1293 and SNP1294 were fitted with HF-data on fat- (FY1, FY2, FY3) and protein yield (PY1, PY2, PY3). SNP1294 dropped out as non-significant ($p > 0.05$) (see Figure 4.45) in model selection, suggesting that the effect of SNP1294 is correlated to the effect of SNP1293 due to LD being suggested by $r^2 = 0.54$. SNP1293 is used in further analysis of HF.

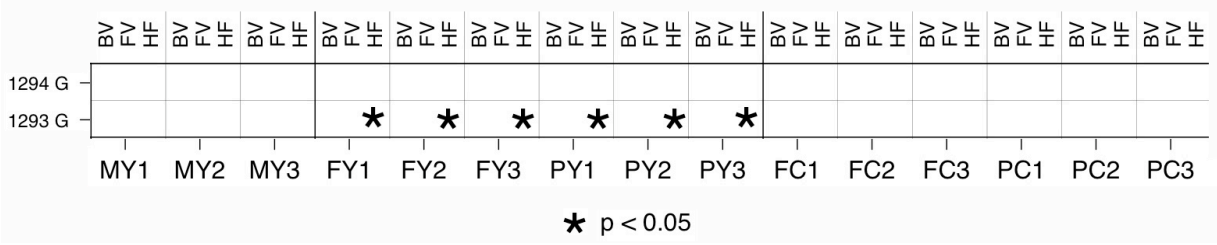


Figure 4.45 Significantly associated markers of *ABCG2* in multi marker analysis after backward model selection

4.8.3.2. Markers of PPARGC1A in multi marker analysis

SNP1242 and SNP1238 were fitted with BV-data on milk- (MY1, MY2, MY3) and fat yield (FY1, FY3, FY3). In model selection, SNP1242 dropped out as non-significant ($p > 0.05$) (see Figure 4.46). This suggests that the effects of SNP1242 and SNP1238 are correlated due to LD. The $r^2 = 0.92$ between both markers. SNP1238 is taken for further analysis in BV.

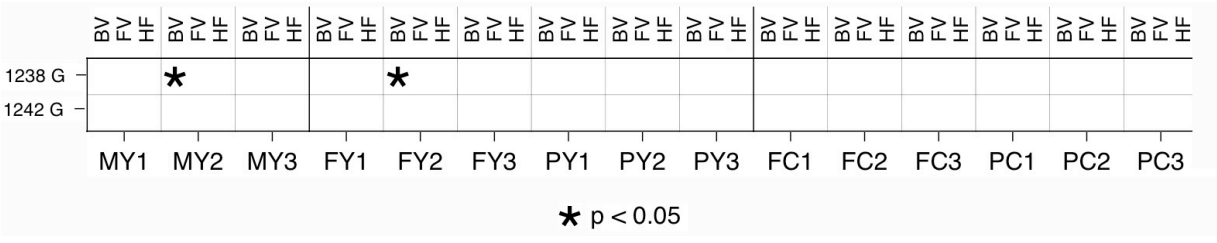


Figure 4.46 Significantly associated markers of *PPARGC1A* in multi marker analysis after model optimisation

4.8.3.3. Markers of CSN1S1 in multi marker analysis

In BV and FV, each of SNP1191, SNP1188, SNP1175, SNP1166, SNP1165 and SNP1161 discriminates itself the significantly associated haplotype CSN1S_HT4. INDEL1170 tags CSN1S_HT2. Multi marker analyses with these markers were carried out in BV and FV on milk yield (MY1, MY2, MY3) and all content traits. 5 of the six SNPs that discriminate CSN1S_HT4 dropped out in model selection while the sixth SNP is significantly associated (see Figure 4.47). This suggests that the 6 SNPs each of which discriminates CSN1S_HT4, represent one effect and markers are significantly associated in SMA due to extensive LD (see Table 4.29). A reliable selection of one marker was not possible across traits and breeds because markers are almost in complete LD (see Table 4.29, Figure 4.21, Figure 4.22 and Table 4.29), and the marker that remained in optimised models differed between traits. However, SNP1191 is chosen to represent the effect of CSN1S_HT4 in further analyses of BV and FV.

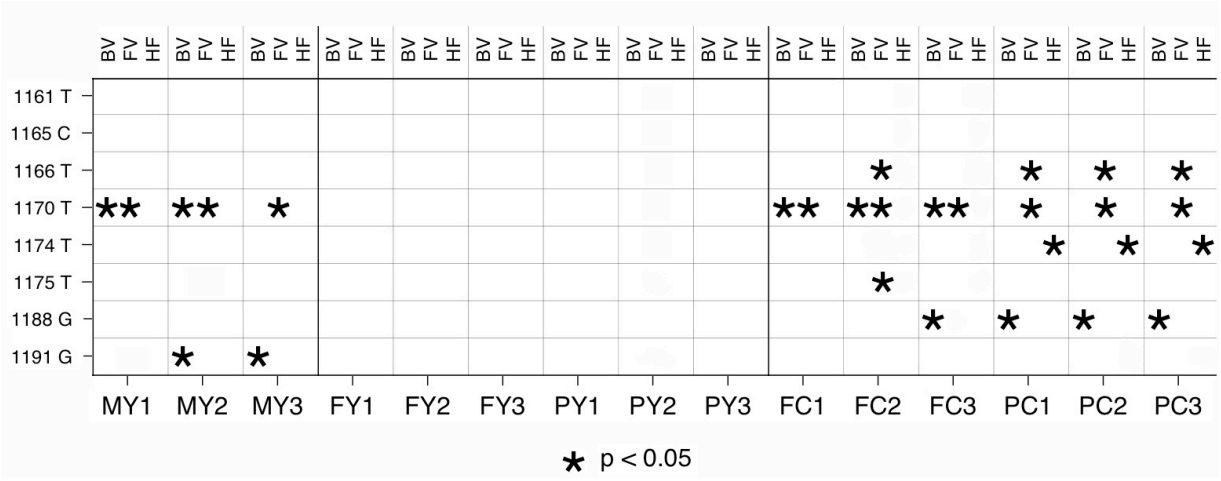


Figure 4.47 Significantly associated markers of *CSN1S1* in multi marker analysis within genes after model optimisation

Table 4.29 Linkage disequilibrium (r^2) in BV and FV for *CSN1S1*-markers fitted in multi marker analysis within genes

		1188	1175	1170	1166	1165	1161
BV (FV)	1191	0.79 (0.99)	0.79 (0.99)	0.01 (0.03)	0.06 (0.82)	0.74 (0.81)	0.76 (0.82)
BV (FV)	1188		0.99 (0.99)	0.01 (0.03)	0.04 (0.82)	0.94 (0.81)	0.95 (0.81)
BV (FV)	1175			0.04 (0.03)	0.94 (0,82)	0.93 (0,82)	0,95 (0,81)
BV (FV)	1170				0.44 (0.04)	0.01 (0.04)	0.01 (0.04)
BV (FV)	1166					0.04 (0.99)	0.04 (0.99)
BV (FV)	1165						0.99 (0.99)

However, models for MY2 and FC3 in BV and models for FC2, PC1 and PC2 in FV included, along with one of the six CSN1S_HT4 discriminating markers, INDEL1170. A

separate effect of INDEL1170 can be assumed and is supported by low LD between INDEL1170 and markers that discriminate CSN1S_HT4 (see Table 4.29).

In HF, SNP1174 and INDEL1170 were fitted in MMA on all protein content traits. INDEL1170 dropped out during model selection as non-significant ($p > 0.05$), suggesting that INDEL1170 represents in HF a correlated effect of SNP1174. Both markers show extensive LD in HF: $r^2 = 0.55$ and SNP1174 is chosen for further analysis in HF.

4.8.3.4. Markers of SLC2A9 in multi marker analysis

The significant effects of SNP1223, SNP1213 and SNP1206 in SMA in BV were confirmed because SNP1206 discriminates SLC2_HT7 and each of SNP1223 and SNP1213 discriminates SLC2_HT1 in HA. Multi marker models with SNP1223, SNP1213 and SNP1206 were computed for all traits. In content traits, only SNP1206 remained after model selection, while for MY1, MY2, FY1 and PY2 two markers were selected: SNP1206 together with SNP1223 or SNP1213 (see Figure 4.48). SNP1223 and SNP1213, however, are in extensive LD ($r^2 = 0.96$) and no reduced model included both, showing that their effects are correlated. Thus, based on results of MMA it can be concluded that *SLC2A9* encompasses two markers with separate effects. One effect is represented by SNP1206, the other either by SNP1213 or by SNP1223. A definite selection of either SNP1223 or SNP1213 was not possible. SNP1213 is taken in further analysis of BV to represent the effect of SLC2_HT1 because SNP1213 was included in optimised models of the first lactation traits in which most phenotypic data is available (see Table 3.4).

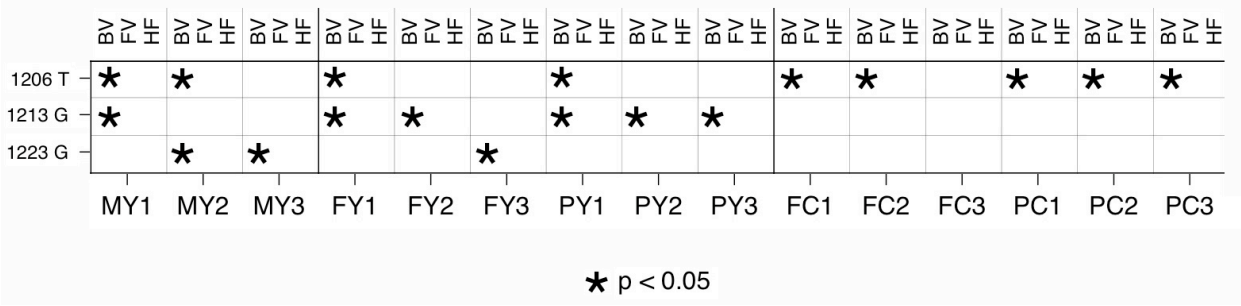


Figure 4.48 Significantly associated markers of *SLC2A9* in multi marker analysis within genes after model optimisation

4.8.3.5. Multi marker analysis across genes

MMA across genes analysed markers that were identified in MMA within genes to be associated with distinct effects (see Table 4.30). Models were optimised according to the *AIC*-criteria and included the sire, and in the case of HF also the *DGAT1*-genotype, as fixed effects.

Table 4.30 Markers fitted in multi marker analysis across genes

Breed	*ABCG2*	*OPN*	*PPARGC1A*	*PGM2*	*CSN1S1*	*SLC2A9*
BV	-	-	1238	-	1191, 1170	1213, 1206
FV	-	-	1238	-	1191, 1170	-
HF	1293	1331	1239	-	1174	-

The effect of the *DGAT1*-genotype on MW is not significant. Results for markers are shown in Figure 4.49. All markers listed in Table 4.30 were identified as exhibiting distinct effect.

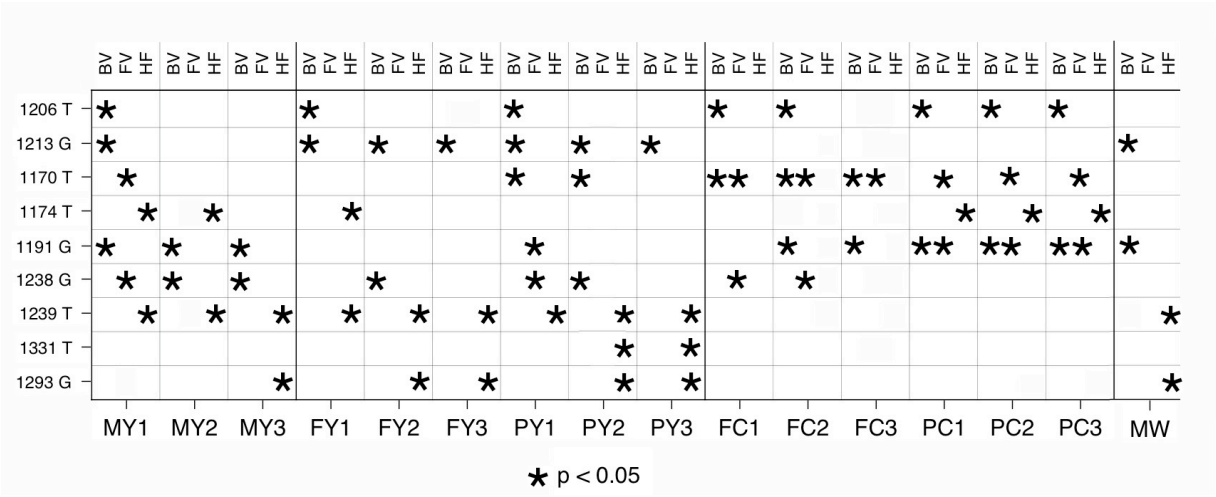


Figure 4.49 Significantly associated markers in multi marker analysis across genes
The significant ($p < 0.05$) interaction in BV between INDEL1170 and SNP1206 in FC1 and FC2 is not displayed.

Extensive (long-range) LD between markers from Table 4.30 was not detected (see Table 4.31), which indicates the absence of correlated effects between markers from different genes. In BV, interaction between INDEL1170 and SNP1206 on FC1 and FC2 was significant ($p < 0.05$) but the interaction is based on less than 30 observations and therefore ignored in Figure 4.49 and in further analysis.

Table 4.31 Linkage disequilibrium (r^2) between markers included in multi marker analysis across genes

Breed	Gene	SNP	1331	1239	1238	1191	1174	1170	1213	1206
BV	*PPARGC1A*	1238	[a]	[a]	[a]	0.01	[a]	0.00	0.00	0.00
	CSN1S1	1191	[a]	[a]	[a]	[a]	[a]	0.06	0.01	0.00
		1170	[a]	[a]	[a]	[a]	[a]	[a]	0.01	0.02
	SLC2A9	1213	[a]	[a]	[a]	[a]	[a]	[a]	[a]	0.01
FV	*PPARGC1A*	1238	[a]	[a]	[a]	0.00	[a]	0.00	[a]	[a]
	CSN1S1	1191	[a]	[a]	[a]	[a]	[a]	0.03	[a]	[a]
HF	*ABCG2*	1293	0.03	0.01	[a]	[a]	0.00	[a]	[a]	[a]
	OPN	1331	[a]	0.06	[a]	[a]	0.00	[a]	[a]	[a]
	PPARGC1A	1239	[a]	[a]	[a]	[a]	0.03	[a]	[a]	[a]

[a] Marker combination no fitted in multi marker analysis in this breed

Significant effects on MW were identified for SNP1191 ($p < 0.01$) and SNP1213 ($p < 0.05$) in BV (see Figure 4.49). SNP1293 ($p < 0.01$) and SNP1239 ($p < 0.01$) are significantly associated with the total merit index for milk production in HF (see Figure 4.49). Effects ($\alpha/2$), shown in Table 4.32, are estimated in single marker models.

Table 4.32 Regression coefficients on the number of copies of allele '*x*' representing half of the allele substitution effects ($\alpha/2$) and standard error for the total merit index of milk production (MW) of significantly associated polymorphisms ($p < 0.05$)

	BV		HF	
	CSN1S1	*SLC2A9*	*ABCG2*	*PPARGC1A*
Marker	1191	1213	1293	1239
Allele '*x*'	G	G	G	T
MW	-2.422 ± 1.196	-1.760 ± 0.649	1.741 ± 0.560	2.928 ± 0.986

4.8.4. Summary of association studies

SMA identified 22 trait- and breed-specific significantly associated SNPs (FDR < 10%), of which 18 discriminate a significantly associated haplotype. Among these, MMA allowed in most cases the identification of those markers that exhibit distinct effects (see Table 4.30). Haplotypes where constructed with the markers listed in Table 4.30, except SNP1238 in BV and SNP1331 in HF. SNP1238 is significantly associated in SMA of BV only in second lactation of milk- and fat yield, but not in first lactation for which most phenotypic data is available (see Table 3.4). SNP1331 is significantly associated in HF with several traits in SMA and was also confirmed in HA. However, MMA across genes included SNP1331 in the final marker set only in four traits, while SNP1293 and SNP1239 explain the major part of the SNP1331-effect in fat- and protein yield, since SNP1293 and SNP1239 are included in most cases in models for fat-and protein yield after optimisation.

The proportion of genetic variance, which is explained by the haplotypes, was calculated after haplotype construction and substitution effect estimation. Results from effect estimation are shown in Table 11.23, Table 11.24 and Table 11.25. Haplotypes were constructed as described in 3.2.9.3 Haplotype construction and substitution effects were estimated as described in 3.2.9.4.4 Haplotype analysis with haplotypes constructed with markers from the final marker set. Results from haplotype construction are shown in Table 4.33.

Table 4.33 Haplotypes constructed for calculation of explained additive genetic variance
Haplotypes were constructed separately for each breed. Only haplotypes with a frequency > 0.025 are included. Numbers in third column indicate in vertical orientation the SNP by the in-house identification code SNP_id.

Breed	Haplotype name	Haplotype [a]	Frequency
BV		SNP 1 1 1 1 1 1 2 2 9 7 1 0 1 0 3 6	
	HT1_BV	A T C C	0.073
	HT2_BV	A D C C	0.120
	HT3_BV	A D G C	0.499
	HT4_BV	A T G C	0.227
	HT5_BV	G T G C	0.045
	HT6_BV	A D G T	0.036
FV		SNP 1 1 1 2 1 1 3 9 7 8 1 0	
	HT1_FV	G A D	0.218
	HT2_FV	A A D	0.193
	HT3_FV	G A T	0.262
	HT4_FV	A A T	0.272
	HT5_FV	A G T	0.027
	HT6_FV	G G T	0.028
HF		SNP 1 1 1 2 2 1 9 3 7 3 9 4	
	HT1_HF	G G A	0.497
	HT2_HF	A G A	0.194
	HT3_HF	A G T	0.052
	HT4_HF	G G T	0.219
	HT5_HF	G T T	0.038

[a] 'D' represents the deletion variant of INDEL1170

The proportion of explained genetic variation depends on effect size and the frequency of haplotypes (see 3.2.9.4.4 Haplotype analysis with final marker set) (LYNCH and WALSH 1998). Effects with larger magnitude explain more genetic variation, but a medial allele

frequency distribution is crucial for the proportion of explained genetic variation. The proportion additive genetic variance explained by the haplotype listed in Table 4.33 varies from 0.1% in fat and protein yield traits of FV to 3.3% in MY2 and PY2 of HF, and MY2 of BV (see Table 4.34). The associated markers in FV only had significant effects in milk yield and fat- and protein percentage. Antagonistic effect directions resulted in non-significant effects on fat- and protein yield. Thus the proportion of explained genetic variance in fat- and protein yield is low in FV.

Table 4.34 Proportion (%) of genetic variance explained by haplotypes constructed with markers that are associated with distinct effects

Trait	BV	FV	HF
MY1	2.5	2.0	2.5
MY2	3.3	1.1	3.3
MY3	1.4	2.3	2.4
FY1	1.1	0.1	1.8
FY2	1.4	0.1	2.5
FY3	1.8	1.7	3.2
PY1	2.3	0.1	1.6
PY2	2.2	0.1	3.3
PY3	1.1	0.1	3.0

5. Discussion

5.1. *General approach*

Whole genome scans identified several markers on bovine chromosome 6 (BTA6) that affect milk production traits, so that this chromosome has been the focus of studies searching for causative polymorphisms for milk synthesis. However, previous studies analysed candidate genes in isolation, while in this work six genes were characterised and screened for polymorphisms. DNA-variants were analysed in association studies in various cattle breeds in order to assess the possibility of a causative site among associated polymorphisms by the breed-comparative approach.

5.2. *Breed-comparative linkage disequilibrium analysis*

Several studies reported extensive linkage disequilibrium (LD) in cattle (FARNIR *et al.* 2000; KHATKAR *et al.* 2006a; KHATKAR *et al.* 2006b; KHATKAR *et al.* 2007; TENESA *et al.* 2003). Across studies, which calculated D' from microsatellite data, the *D'* was approximately 0.50 between markers < 5 cM apart (FARNIR *et al.* 2000; KHATKAR *et al.* 2006b; TENESA *et al.* 2003). This work focuses mainly on association studies, although LD was analysed since it is relevant in terms of association studies. *D'* is, across breeds, about 0.80 for markers separated by < 5 cM, which shows that *D'* from SNP-data in this study is higher than has been concluded in previous studies with microsatellite data. In general, *D'* is upwards biased when analysing multi-allelic loci (ARDLIE *et al.* 2002; NSENGIMANA and BARET 2004), so that our result is surprising. Otherwise, in the present work, *D'* for markers within a distance of 5 cM is, apart from *ABCG2* and *OPN*, equivalent to the calculated *D'* within genes. Accounting for the physical distance, the modified Malecot model pointed out that *D'* is significantly higher within than between genes in BV and HF, which is in line with recent results from the HapMap project (http://www.hapmap.org) (Frazer *et al.* 2007) and with results from a haplotype analysis in goats (Hayes *et al.* 2006). Hence, the higher *D'* within genes might explain the *D'*-bias between markers < 5 cM apart in the present work.

The Malecot model quantifies the exponential decay of *D'* with increasing marker distance and detected, across breeds, a *D'*-decrease to a level of about 0.30 for loosely linked loci (> 20 Mb). This is in agreement with results from previous studies (Khatkar *et al.* 2006a; Khatkar *et al.* 2006b) with regard to the marker distance. The magnitude *D'* between loosely linked loci,

however, is elevated in this study as compared to Khatkar et. al (2006). LD is governed by numerous factors, such as inbreeding (NSENGIMANA and BARET 2004) or selection (HARTL and CLARK 2007; PARSCH *et al*. 2001), but this contrast is most likely to originate from differences in the SNP-coverage on BTA6 between both studies. Khatkar et. al (2006) investigated 220 evenly spaced SNPs, while in this work the SNPs are located in a small subset of selected candidate genes.

The decay rate of *D'* with marker distance and the distribution of LD differed between breeds. The decay proceeded less rapidly in FV and most rapidly in BV while the latter is hardly explainable. Studies analysing inbreeding were not subject in the present work. However it is known that HF has lower estimates for the effective population size than other breeds (HAGGER 2005; MC PARLAND *et al*. 2007; SORENSEN *et al*. 2005; ZENGER *et al*. 2007), so that a delayed decay rate was expected in HF. The approach in three breeds exhibited a breed-specific pattern of LD. This is consistent with a previous breed-comparative approach in cattle (Barendse *et al*. 2007), and with a study that analysed several dog breeds (Sutter *et al*. 2004).

5.3. *Is one marker causative?*

Several QTLs for milk production traits have been mapped on BTA6, but QTL-confidence intervals span tens of map units and encompass hundreds of genes. Methods for an identification of a causative mutation (QTN) such as inbred lines, transgenic organisms or knock-out organisms are applicable in model organisms but are not readily applicable in livestock species, so that the identification of causative mutations in livestock populations more closely resembles a collection of multiple evidence than a proof (Mackay 2001). As a result of these limitations, only few postulated QTNs in livestock species, e.g. *DGAT1* in cattle (Winter *et al*. 2002), *GDF8* in sheep (Clop *et al*. 2006) and *IGF2* in pigs (Van Laere *et al*. 2003) come close to the status of a verified QTN (RON and WELLER 2007).

Two distinct regions on BTA6 are most likely to harbour three QTLs for milk production traits (Khatkar *et al*. 2004). The proximal region is located between 30 Mb and 45 Mb and the distal region is located at about 70 Mb and it is probable that two distinct QTLs exist in the proximal region. One QTL affects milk yield and content traits with opposite effect signs (Olsen *et al*. 2005) and the second QTL was detected for fat- and protein yield (Kuhn *et al*. 1999). Thus both are differentiated by traits rather than by chromosomal position since confidence intervals overlap and different mapping studies have used different animal panels,

populations and marker sets. However, *ABCG2* and *OPN* are located closer to the QTL with effects on milk yield and content traits (Olsen *et al.* 2005) and *PPARGC1A* and *PGM2* are located within the confidence interval for the fat- and protein yield QTL (Kuhn *et al.* 1999). The third QTL on BTA6 has effects on milk yield and content traits in antagonistic direction and is mapped to the region of the casein gene cluster (Velmala *et al.* 1999). Consequently *ABCG2*, *OPN*, *PPARGC1A*, *PGM2* and *CSN1S1* are functional and positional candidate genes.

Association studies were carried out in a three-stage procedure: single marker analysis, haplotype analysis and multi marker analysis. Multi marker models included markers that were significantly associated in single marker analysis while being discriminatory for significantly associated haplotypes and were fitted by backwards selection using the *AIC*-criterion to identify whether the effects of different markers are correlated due to extensive LD. One may conclude that markers with distinct effects present strong evidence for being causal (see Table 4.30).

Apart from statistical support, a strong QTN-candidate should be characterised by functional features. The present work is an association study where functional studies such as expression analyses or studies in knockouts of model organisms are not included. Therefore one might argue that a phenotype is a priori affected by non-synonymous exchanges (RON and WELLER 2007), such as SNP1293, SNP1213 and SNP1191. However, this is not always the case. It is also possible that intronic markers are causative since it is known that introns can encompass regulative elements (GREENWOOD and KELSOE 2003; LE HIR *et al.* 2003).

LD governs the identification of associated markers since only a subset of DNA-variants is genotyped. LD within genes was extensive and comparable between breeds, but LD between genes differed in level and distribution between breeds. The level was higher in BV and HF, and both breeds showed a larger number of markers associated with distinct effects than FV. HF had higher LD between *ABCG2* and *PPARGC1A* than other breeds and BV had more *CSN1S1*- and *SLC2A9*-markers in $D' > 0.80$ than other breeds. Both breeds had the majority of markers with distinct effects in regions with breed-specific enhanced LD, so that the breed-specific pattern of LD was reflected in the number of significantly associated markers. In addition, *PGM2*, which contains across breeds the smallest number of associated markers and haplotypes, is isolated from other genes in the context of LD, since very few *PGM2*-markers are in extensive LD ($D' > 0.80$) with markers from other genes. Therefore, results from this study support the relevance of LD in identifying associated markers, while also underlining how LD can complicate the question of whether an associated marker is causative. However,

causative polymorphisms are expected to be confirmed in different populations and breeds, so that the breed-comparative approach in this study yields information that can be used to discuss associated SNPs as QTN-candidates, although functional studies are missing. Markers associated with distinct effects (see Table 4.30) are strong QTN-candidates if the conditions listed below are fulfilled:

1) The SNP segregates with significant association in all breeds and shows comparable effect magnitudes and effect signs across breeds.
2) The SNP is significantly associated in a breed across all lactations of a trait.
3) The SNP was identified with significant effects in a previous association study.
4) The significantly associated traits of the SNP correspond to the traits affected by the QTL in the corresponding region.

5.3.1. Are markers significantly associated with the same trait across breeds?

No marker is significantly associated with distinct effects in all breeds but allelic fixation has to be accounted for. SNP1293 is associated in HF and SNP1206 is associated in BV, but both SNPs are monomorphic in breeds without association. However, BV and FV carry at SNP1293 the allele associated in HF with a yield increasing effect. This is unexpected on the one hand, since BV and FV exhibit lower yield traits as compared to HF, and on the other hand since Red-Holstein sires have been introduced into FV in past decades to increase yield traits. Due to fixation in FV, it has to be assumed that the introduced Red-Holstein sires have been monomorphic for the preferred allele of SNP1293. This is unexpected since Red-Holstein sires in the analysed data and the entire HF sample are similar in terms of the allele frequency at SNP1293 (results not shown). FV and HF carry at SNP1206 the allele that is associated in BV with increased content traits. This is not in line with performance differences between breeds, since BV has higher milk content than FV and HF.
SNP1331, SNP1238 and SNP1213 are significantly associated in one breed only, although they are polymorphic in at least one additional breed.

5.3.2. Are markers significantly associated across all lactations of a trait?

Across breeds, no SNP is significantly associated with all lactations of all milk production traits. *CSN1S1*-markers, SNP1191, SNP1174 and INDEL1170 are in most cases significantly associated in all lactations of content traits (see Figure 4.29). SNP1293, SNP1239, SNP1331,

SNP1238 and SNP1213 are associated with some lactations of significantly associated traits. This may be due to effect magnitudes changing from lactation to lactation. However, this argument is less applicable in HF, because the HF-effects in distinct lactations are comparable in magnitude and in the number of animals providing phenotypic data. In contrast, the number of animals for which phenotypic data was available decreases from first to third lactation in BV and FV, so that results such as SNP1238 in BV, which is not significantly associated with the first lactation of a trait, have to be interpreted with scepticism.

5.3.3. Do significantly associated markers agree with previous association studies?

Proximal QTL-region of BTA6

ABCG2, *OPN*, *PPARGC1A* and *PGM2* are located in the proximal QTL-region of BTA6 and previous studies reported markers of *ABCG2*, *OPN* and *PPARGC1A* to be associated with milk traits. DNA variants of *ABCG2* and *OPN* were analysed in Israeli (Cohen-Zinder *et al.* 2005) and U.S. HF (Leonard *et al.* 2005; Schnabel *et al.* 2005) and Norwegian Red (Olsen *et al.* 2007). In these studies, markers showed the strongest effects on milk yield and content traits. Studies that included markers from both genes (Cohen-Zinder *et al.* 2005; Olsen *et al.* 2007) postulated a missense mutation in exon 14 of *ABCG2* (SNPex14) (Cohen-Zinder *et al.* 2005) as a strong QTN-candidate.

In the present study, *ABCG2*- and *OPN*-markers are significantly associated in HF, but the effects are significant on fat- and protein yield without exhibiting significant effects on milk yield and content traits. In addition, no *ABCG2*- or *OPN*-marker is significantly associated with a milk production trait in BV and FV. The present study does not confirm the findings of previous studies, which were however conducted without accounting for *DGAT1* (COHEN-ZINDER *et al.* 2005; LEONARD *et al.* 2005; OLSEN *et al.* 2007; SCHNABEL *et al.* 2005). Ron et. al (2006) report that SNPex14 segregates in German FV, which was not confirmed in the present SNP-screen. On the other hand, markers of *ABCG2* and *OPN* are almost in complete LD in FV, so that, although not genotyped directly, one would expect the effect of SNPex14 to be identified in FV among the analysed *ABCG2*- or *OPN*-markers. Previous studies with *PPARGC1A* identified SNP1246 and a SNP in UTR of exon 13 as being significantly associated with fat-yield in German HF (Weikard *et al.* 2005), while in American HF the UTR-SNP in exon 13 was significantly associated with protein percentage (Khatib *et al.* 2007). In the present study, markers of *PPARGC1A* are significantly associated with fat- and protein yield in HF (SNP1239) and BV (SNP1238), and with fat content in FV (SNP1238).

Hence the HF-result from this study agrees with the previous study in German HF with regard to the associated trait, but both studies are not in line in terms of the associated marker (Weikard *et al.* 2005), because SNP1246 is without any significant association despite being polymorphic in all three breeds.

Distal QTL-region on BTA6

CSN1S1 and *SLC2A9* are located in the distal part of BTA6, while *SLC2A9* is not anchored within a QTL confidence interval and has not been analysed in previous association studies. *CSN1S1* is a member of the casein gene cluster and numerous casein gene variants have been analysed with regard to milk production traits. *CSN1S1*-variants were identified in previous studies as associated with milk production traits in Norwegian Red (Lien *et al.* 1995) and HF (Sabour *et al.* 1996), while in Finish Ayrshire (Ikonen *et al.* 1999) and Brown Swiss (Boettcher *et al.* 2004) no significant associations were identified. In HF, a four-allelic INDEL in the promoter region of *CSN1S1* showed a significant association with protein percentage (Prinzenberg *et al.* 2003). In a second study in HF, the *CSN1S1*-B-allele, which is analogous to the A variant of SNP1191 (Farrell *et al.* 2004) had significant effects on milk yield (Lin *et al.* 1986). In Norwegian Red however significant effects on protein yield were observed (Lien *et al.* 1995), suggesting that previous studies investigating genes from the casein cluster are not in line. *CSN1S1*-markers in the present work are significantly associated with protein percentage across breeds, showing opposite effect signs between milk yield and content traits, so that significant effects on fat- and protein yield were not detected. In HF, less *CSNS1*-markers are significantly associated than in BV or FV. This is explained by an enhanced level of allelic fixation in HF, while at each marker the milk yield increasing allele is predominant, which is in line with performance differences between HF and other breeds. Hence results from this study confirm previous studies that reported effects on milk yield and content traits, but do not agree with the *CSN1S1*-effect on protein yield in Norwegian Red (Lien *et al.* 1995).

5.3.4. Are marker-associated traits in correspondence with QTL- affected traits?

PPARGC1A-markers, in particular SNP1239 in HF and SNP1238 in BV and *CSN1S1*-markers (SNP1191, SNP1170, INDEL1170), are associated with traits that correspond with the QTL-affected traits in the region of relevance (KHATKAR *et al.* 2004; KUHN *et al.* 1999).

SNP1293 in *ABCG2* and SNP1331 in *OPN* are associated with fat- and protein yield but a QTL for milk yield and content traits is more likely in this region (Olsen *et al.* 2005).

5.4 Appraisal of the breed-comparative approach

The molecular dissection of milk production with four selected genes located in a region harbouring most likely two QTLs is complicated since milk traits are correlated. However, previous studies claimed QTN-identification. As discussed in the previous chapter 5.3 Is one marker causative?, it is more likely that neither the present work, nor any of the previous studies have identified a polymorphism with causative function, or that the molecular background in the analysed regions is too complex to be deciphered by investigating selected genes.

Nevertheless, our results are compatible with two segregating QTLs between 30 and 45 Mb on BTA6. The fat- and protein yield causing variant (Kuhn *et al.* 1999) is probably in LD with SNP1293, SNP1331 and SNP1239. This marker-trio affects identical traits. The traits affected by SNP1293 are not in correspondence with those affected by the QTL (Olsen *et al.* 2005) and the effects of SNP1331 in all lactations on fat yield and in first lactation on protein yield are correlated with the effects of SNP1293 and SNP1239 due to LD between markers. Therefore, it appears more likely that each of this marker-trio is in incomplete LD with one causative site than with different sites. This is supported by an extensive level of LD ($D' >$ 0.80) between several marker pairs of *ABCG2* and *PPARGC1A* in HF. Single *ABCG2*- or *OPN*-markers that affect milk yield and content traits, the traits of the second QTL in this region (Olsen *et al.* 2005) are missing in this study. However, haplotype ABC_HT2 is associated with these traits. Haplotypes capture information about the entire chromosomal region that is spanned by incorporated markers, so beneficial information is expected when haplotypes are investigated in addition to analysing SNPs in isolation. Therefore, it is likely that the QTL for milk yield and content traits (Olsen *et al.* 2005) is in LD with ABC_HT2.

Marker selection after multi marker analysis was indefinite in both genes selected from the distal part of BTA6: *CSN1S1* and *SLC2A9*. Markers in final multi marker models changed from lactation to lactation within a trait. Thus, it is not justified to prefer one of these markers as a strong QTN-candidate. Recently, it has been shown in goats that the expression of *CSN1S2* is influenced by *CSN1S1*-variants (Leroux *et al.* 2003), so that epistatic effects are likely to complicate a QTN-identification in this region. As a consequence of the cluster-organisation and due to extensive LD, it is innately complicated to postulate a casein polymorphism as causative as long as functional analyses are missing. However, *SLC2A9*-

markers show significant effects on fat- and protein-yield, which are not affected by *CSN1S1*-markers. This suggests, apart from one or several causative sites in the casein cluster, an additional causative polymorphism although no QTL has been mapped distally to the casein cluster so far.

LD is one main factor that complicates a QTN-call. However, combining different breeds facilitates a QTN-call since it increases the number of potentially informative recombination elements and shortens linkage disequilibrium distances (Parker *et al.* 2007). The statement 'Markers from this study are most probably not causative' depends solely on results from the breed-comparative approach. Therefore this work illustrates the advantage of combining cattle breeds in studies that aim to identify causative sites. This is most important since previous studies that investigated BTA6-regions claimed a QTN-identification without considering different breeds. Breed-comparative approaches exist rarely in cattle, but are state of the art in other species, for example dogs (Karlsson *et al.* 2007) (Lindblad-Toh *et al.* 2005). Therefore, based on this and base on a genome-wide breed-comparative study in cattle (Barendse *et al.* 2007), it is recommended that breed-comparative approaches in cattle are followed up in the future. This is of particular importance since present results from BTA6 lead to the conclusion that the genetic regulation of milk synthesis is more complex than was assumed in previous studies on this chromosome.

5.5. *Including identified markers in breeding programmes*

It is most likely that the identified markers are not causative but markers that represent distinct effects could be used in breeding programmes. In general, three types of genetic loci are distinguished in terms of a usage in marker-assisted selection (MAS): direct markers, LD-markers and LE-markers (Dekkers 2004). Direct markers are single loci that influence the associated trait directly as causative mutation (QTN). LD-markers have no causative function themselves, but are single markers that are in extensive population-wide LD with a causative site. LE-markers are loci that are in population-wide linkage equilibrium with the functional mutation. The three marker types differ in application in MAS, in the work required for genotyping and in the amount of extra genetic gain. Linkage phases differ between families, so that the effects of LE-markers have to be assessed on a within-family basis. QTNs, and to a lesser degree also LD-markers, allow a direct selection on the genotype due to a consistent population-wide association between genotype and phenotype (Dekkers 2004). Genotyping efforts are much higher for LE-markers than for direct or LD-markers, because LE-markers

require large families to be genotyped. For QTNs and LD-markers, it is sufficient to genotype only selection candidates. Genotypic data can be incorporated as fixed genotype effects into marker-assisted BLUP (MA-BLUP) (Dekkers 2004). Prospects for extra genetic gain are higher for QTNs and LD-markers than for LE-markers because of the limited information and uncertainty in terms of the inheritance of the QTL-loci.

Markers obtained in this study are family-independent. Milk production traits are quantitative genetic traits, so benefits from marker-assisted BLUP are defined by the proportion of additive genetic variation that is explained by marker information (Meuwissen *et al*. 2001). This proportion depends on allele frequency and effect size (LYNCH and WALSH 1998). Depending on lactation and breed, markers from this study explain between 1.1% and 3.0% of additive genetic variation in milk-, fat- and protein yield. Hence markers from this study are less important than DGAT1, which explains proportions of additive genetic variation in yield traits between 5% in protein yield and 12% in milk yield (Bennewitz *et al*. 2004). However, it would be beneficial to integrate markers from this study in MA-BLUP (Villanueva *et al*. 2004).

MA-BLUP is most useful for animals that have inaccurately estimated breeding values from BLUP (BLUP-EBV), e.g. young sires before entering progeny testing and bull dams. MA-BLUP allows pre-selection of selection candidates, which is especially beneficial among full sibs with identical BLUP-EBVs.

6. Conclusion and outlook

The following major conclusion can be drawn from the results of this study:

- The breed-comparative approach was helpful for collecting evidence about a causative function of associated polymorphisms.

It is very likely that markers identified in this study are associated due to extensive LD with causative sites. This follows from the results of the breed-comparative approach used in this study. The study demonstrated the advantage that can be obtained by combining several breeds in studies aiming to identify causative polymorphisms. Despite the fact that markers identified by this study are most likely non-causative, the associations noted confirm anew the anchorage of causative polymorphisms for milk traits on BTA6. Even though the present research analysed a few tens of markers located in several candidate genes, the achieved chromosomal coverage appears insufficient to uncover the molecular background for milk synthesis anchoring in the regions analysed. Only very recently, bovine SNP-sets containing several thousands of genome-wide spread markers were made available. In spite of this fact, the expectation that genome-wide association studies will decipher the molecular architecture of quantitative traits has to be lowered since dominance and epistatic effects, recently exemplified for milk traits in cattle (Kuehn *et al.* 2007) and goats (Leroux *et al.* 2003), complicate molecular dissection. Nevertheless, these upcoming studies will achieve an enhanced SNP-coverage in order to refine positional indication of chromosomal regions anchoring causative variants, and could lead to elementary changes in livestock selection schemes, such as genomic selection (Schaeffer 2006).

7. Summary

Various mapping experiments in cattle resulted in the detection of several markers affecting milk production traits on chromosome 6 (BTA6). The aim of this study was to identify association and causality of polymorphisms for milk production traits. Six candidate genes were characterised, screened for polymorphisms and analysed in association studies.

Some of the analysed genes; namely *ABCG2*, *OPN*, *PPARGC1A* and *CSN1S1* were analysed in previous association studies, and causal sites were declared for each of them. However, previous studies led to distinct conclusions concerning the candidate genes, as the genes were analysed in isolation and in a single breed only. In contrast, this study employed a breed-comparative approach analysing the four genes mentioned above together with *PGM2* and *SLC2A9* in order to find out the relative importance of each of these genes for milk production traits. In total, 206 DNA-variants were discovered, of which 50 were genotyped for association study in 742 sires of German Brown, 964 sires of German Fleckvieh and 1418 sires of German Holstein. Association studies were carried out separately for each breed. The traits milk-, fat- and protein yield, as well as fat- and protein content were investigated for each of the first three lactations using a three-stage procedure – single marker analysis, haplotype analysis and multi marker analysis.

A total of 22 polymorphisms were found to be significantly associated with milk production traits in single marker analysis, whereas across breeds most genes harboured more than one associated marker due to extensive linkage disequilibrium. Furthermore, significant associations with various milk production traits were found for 25 haplotypes. In addition, 18 polymorphisms were found to be significantly associated in single marker analysis while being discriminatory for significantly associated haplotypes. Multi marker models were applied to investigate whether these markers are associated with correlated effects. Overall, five markers represent distinct effects in German Brown, that is: one *PPARGC1A*-marker and two markers of each of the two genes *CSN1S1* and *SLC2A9*, respectively. Apart from missing *SLC2A9*-effects, the result of German Fleckvieh is similar to that of German Brown. In German Holstein, one marker each of *ABCG2*-, *OPN*, *PPARGC1A* and *CSN1S1*, respectively turned out to be associated with distinct effects. Markers associated with milk traits can be grouped into two categories: (a) Markers, significantly associated with milk yield and content traits showing opposite effect signs, exhibit non-significant effects on fat and protein yield. (b) Markers have significant effects on fat and protein yield without being significantly associated with milk yield and content traits. *PPARGC1A* and *SLC2A9* encompass markers of

both categories. Markers of *ABCG2* and *OPN* are exclusively associated with fat and protein yield. *CSN1S1*-markers belong to the category of antagonistic markers. Depending on trait, lactation and the breed under consideration, markers of this study explain up to 3.3% of genetic variation in milk, fat and protein yield. Marker associations vary depending on breed and trait. Hence, it is likely that markers identified in this study are not causative. Nonetheless, the identified markers can be used in selection schemes of the breeds analysed in this research.

Linkage disequilibrium in livestock species hinders molecular dissection of quantitative traits. The breed-comparative approach proved to be advantageous with regard to finding evidence about a causative function of associated polymorphisms and is suggested for further studies. It can also be concluded from this study that the molecular background of QTLs on BTA6 for milk traits is too complex to be clarified by investigating a few selected genes. Therefore, it is recommended that future studies on BTA6 consider an increased marker density and evenly spaced polymorphisms to decipher the molecular architecture of QTLs for milk related traits.

8. Zusammenfassung

In den vergangenen Jahren wurden mittels QTL-Studien Genomregionen auf dem Rinderchromosom 6 (BTA6) identifiziert, die Milchleistungsmerkmale beeinflussen. Ziel der hier vorliegenden Arbeit war es, für diese Effekte kausale DNA-Polymorphismen zu finden. Dafür wurden zunächst sechs Kandidatengene auf BTA6 ausgewählt und charakterisiert, sowie durch Resequenzierung auf DNA-Polymorphismen untersucht. Anschließend wurden die Polymorphismen in drei Rassen genotypisiert und die Ergebnisse in einer vergleichenden Assoziationsstudie analysiert.

Vier der sechs ausgewählten Gene (*ABCG2, OPN, PPARGC1A* und *CSN1S1*) wurden bereits in früheren Studien untersucht. Die Autoren dieser Studien erhoben den Anspruch kausale Varianten entdeckt zu haben, obwohl die Gene meist getrennt voneinander und nur in einer Rasse untersucht wurden. Die vorliegende Arbeit ist eine Assoziationsstudie, in der diese Gene neben zwei weiteren Genen, PGM2 und SLC2A9, gleichzeitig analysiert wurden, um so die relative Bedeutung der einzelnen Gene in Bezug auf Milchleistungsmerkmale genauer ermitteln zu können. Von den insgesamt 206 identifizierten DNA Variationen wurden 50 ausgewählt und in 742 Braunvieh-Bullen, 942 Fleckvieh-Bullen und 1414 Holstein-Bullen genotypisiert. Die Assoziationsstudie wurde innerhalb der Rassen getrennt für die ersten drei Laktationen der Milchleistungsmerkmale Milch-, Fett- und Eiweißmenge, sowie Fett- und Eiweißgehalt durchgeführt und umfasste Einzelmarker-, Haplotypen- und Mehrmarkeranalysen.

Im Rahmen der Einzelmarkeranalyse konnten 22 signifikant assoziierte Marker nachgewiesen werden. Hohes Kopplungsungleichgewicht innerhalb der Gene führte in den meisten Fällen zu mehr als einem assoziierten Marker pro Gen. Die Haplotypenanalyse identifizierte 25 signifikant assoziierte Haplotypen. Ein Vergleich zwischen der Einzelmarkeranalyse und der Haplotypenanalyse ergab 18 Marker, die neben einer signifikanten Assoziation als Marker auch spezifisch für einen assoziierten Haplotypen sind. Mit diesen wurde die Mehrmarkeranalyse durchgeführt, um korrelierte Effekte der verschiedenen Marker zu finden. Beim Braunvieh konnten insgesamt fünf getrennte Effekte nachgewiesen werden, ein Effekt im *PPARGC1A*, sowie jeweils zwei in *CSN1S1* und *SLC2A9*. Mit Ausnahme der SLC2A9-Effekte wurde bei der Rasse Fleckvieh die gleichen Effekte festgestellt. Bei der Rasse Holstein wiesen *ABCG2*, *OPN*, *PPARGCA* und *CSN1S1* jeweils einen Marker auf, dessen Effekte mit den Effekten anderer Marker nicht korrelieren.

Die assoziierten Marker können unabhängig von Rasse und Gen in zwei Kategorien eingeteilt werden. (a) Marker mit signifikanten Effekten in den Merkmalen Milchmenge, Milchfett- bzw. Milcheiweißgehalt zeigten in diesen Merkmalen antagonistische Effektrichtungen, so dass keine Assoziation zu den Merkmalen Fett- und Eiweißmenge nachgewiesen werden konnte. (b) Marker, deren Effekte in den Merkmalen Fett- und Eiweißmenge signifikant sind, zeigen keine signifikante Assoziation zur Milchmenge, sowie zu den Milchgehalten. Die Gene *PPARGC1A* und *SLC2A9* enthalten Marker beider Kategorien, dagegen sind Marker von *ABCG2* und *OPN* ausschließlich mit den Merkmalen Fett- und Eiweißmenge assoziiert. Unter den assoziierten *CSN1S1*-Markern wurden nur Marker gefunden, die der Kategorie der Marker mit antagonistischen Effektrichtungen zuzuordnen sind. Je nach Merkmal, Laktation und Rasse erklären die Marker dieser Studie in den Merkmalen Milch, Fett- und Eiweißmenge bis zu 3,3% der additiv genetischen Varianz. Als Folge der je nach Rasse, Marker und Merkmal unterschiedlich identifizierten Assoziationen ist es unwahrscheinlich, dass einer der identifizierten Marker eine kausale Variante für Milchleistungsmerkmale darstellt. Die Marker können dennoch im Rahmen von Selektionsentscheidungen der betreffenden Rasse verwendet werden, da hierfür der Nachweis der Kausalität nicht erforderlich ist.

Kopplungsungleichgewicht in Nutztierpopulationen erschwert wesentlich die Aufklärung molekulargenetischer Zusammenhänge quantitativer Merkmale. Diesbezüglich stellte sich der Rasse vergleichende Ansatz der vorliegenden Studie als vorteilhaft heraus und sollte daher in Folgestudien berücksichtigt werden. Diese Studie zeigt auch, dass der molekulargenetische Hintergrund von QTLs in den betrachteten Regionen auf BTA6 zu komplex ist, um ihn durch eine simultane Betrachtung weniger Gene aufklären zu können. Zukünftige Studien erfordern daher auf BTA6 neben einer deutlich höheren Markerdichte eine gleichmäßigere Markerverteilung, um zur Klärung der molekulargenetischen Zusammenhänge der QTLs für Milchleistungsmerkmale beizutragen.

9. Acknowledgements

First, I am most thankful to my supervisor Prof. Ruedi Fries for giving me the opportunity to work on this exciting subject and for providing me with outstanding facilities.

For financial support, I acknowledge the Federal Ministry of Education and Research (Förderkennzeichen AZ0313381) and the Förderverein Biotechnologieforschung e.V., Bonn. I also thank the artificial insemination stations and cattle breeding associations for supplying semen samples. For supplying Holstein-DNA, I thank the FBN-Dummerstorf, namely Dr. Christa Kühn. I want to offer special thanks to the Bayerische Landesanstalt für Landwirtschaft (LfL), Institut für Tierzucht and to the VIT - Vereinigte Informationssysteme Tierhaltung w.V., namely Dr. Reiner Emmerling and Friedrich Reinhardt, for delivering phenotypic data.

I owe my gratitude to the following persons from the Institute of Human Genetics of GSF National Research Center for Environment and Health, Neuherberg: Prof. Thomas Meitinger for providing access to the genotyping facility, Dr. Peter Lichtner for advice in the genotyping setup and assay design and Bianca Schmick and Jelena Golic for their technical help.

With regards to the laboratory work at the Chair of Animal Breeding, I send many thanks to Theresia Böhm and Anne Keller, but also to Amra Kharahasanovic, Nadine Sobotzki, Steffi Simon and Louisa Mühlberger, for attending this project during student training.

I thank Rita Popken and Birgit Hoffmann, who helped me in administrative matters.

Further, I want to express my deepest gratitude to Dr. Hermann Schwarzenbacher for numerous and fruitful discussions, for support in statistics and for any proposal to finish this work.

Special thanks to all my colleagues at the Chair of Animal Breeding not mentioned by name for their help and friendship during these years. In particular, I want to thank by name Katrin Juling for office sharing and creating an enjoyable and profitable office atmosphere.

Best thanks to the Schwayer family, Astrid, Matthias, the "bicycle-friends" and the friends not mentioned by name for listening to my problems and spending time together to relax.

Finally, I thank my parents, my sister and my family. You always support me wherever I go and whatever I do.

Franz
Freising, December 2007

10. Bibliography

AKAIKE, H., 1974 A new look at the statistical model identification. IEEE Control Systems Society **19:** 716-723.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J Mol Biol **215:** 403-410.

ANDERSSON, O., A. STENQVIST, A. ATTERSAND and G. VON EULER, 2001 Nucleotide sequence, genomic organization, and chromosomal localization of genes encoding the human NMDA receptor subunits NR3A and NR3B. Genomics **78:** 178-184.

ARDLIE, K. G., L. KRUGLYAK and M. SEIELSTAD, 2002 Patterns of linkage disequilibrium in the human genome. Nat Rev Genet **3:** 299-309.

ASHWELL, M. S., and C. P. VAN TASSELL, 1999 Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan. J Dairy Sci **82:** 2497-2502.

AUGUSTIN, R., M. O. CARAYANNOPOULOS, L. O. DOWD, J. E. PHAY, J. F. MOLEY *et al.*, 2004 Identification and characterization of human glucose transporter-like protein-9 (GLUT9): alternative splicing alters trafficking. J Biol Chem **279:** 16229-16236.

BAILEY-DELL, K. J., B. HASSEL, L. A. DOYLE and D. D. ROSS, 2001 Promoter characterization and genomic organization of the human breast cancer resistance protein (ATP-binding cassette transporter G2) gene. Biochim Biophys Acta **1520:** 234-241.

BAND, M. R., J. H. LARSON, M. REBEIZ, C. A. GREEN, D. W. HEYEN *et al.*, 2000 An ordered comparative map of the cattle and human genomes. Genome Res **10:** 1359-1368.

BARENDSE, W., A. REVERTER, R. J. BUNCH, B. E. HARRISON, W. BARRIS *et al.*, 2007 A validated whole genome association study of efficient food conversion in cattle. Genetics.

BARRETT, J. C., B. FRY, J. MALLER and M. J. DALY, 2005 Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics **21:** 263-265.

BENNEWITZ, J., N. REINSCH, S. PAUL, C. LOOFT, B. KAUPE *et al.*, 2004 The DGAT1 K232A mutation is not solely responsible for the milk production quantitative trait locus on the bovine chromosome 14. J Dairy Sci **87:** 431-442.

BININDA-EMONDS, O. R. P., B. HAUBOLD and R. FRIES, 2004 GeneScore: identifying candidate genes using bioinformatics., pp. in *NGFN evaluation meeting*. Bininda-Emonds, O.R.P.,, Berlin.

BLT, 2007 Informationen zur Zuchtwertschätzung beim Rind.

BOETTCHER, P. J., A. CAROLI, A. STELLA, S. CHESSA, E. BUDELLI *et al.*, 2004 Effects of casein haplotypes on milk production traits in Italian Holstein and Brown Swiss cattle. J Dairy Sci **87:** 4311-4317.

BOICHARD, D., C. GROHS, F. BOURGEOIS, F. CERQUEIRA, R. FAUGERAS *et al.*, 2003 Detection of genes influencing economic traits in three French dairy cattle breeds. Genet Sel Evol **35:** 77-101.

BOVENHUIS, H., and J. I. WELLER, 1994 Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Genetics **137:** 267-280.

BRAUNSCHWEIG, M., C. HAGGER, G. STRANZINGER and Z. PUHAN, 2000 Associations between casein haplotypes and milk production traits of Swiss Brown cattle. J Dairy Sci **83:** 1387-1395.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783-796.

BREEDVELD, P., N. ZELCER, D. PLUIM, O. SONMEZER, M. M. TIBBEN *et al.*, 2004 Mechanism of the pharmacokinetic interaction between methotrexate and benzimidazoles: potential role for breast cancer resistance protein in clinical drug-drug interactions. Cancer Res **64:** 5804-5811.

BUITKAMP, J., ANTES, R., WAGNER, V., 1999 *DNA profiling in veterinary medicine, A Laboratory Guide to DNA Fingerprinting/Profiling*. Birkhäuser, Basel.

CANT, J. P., D. R. TROUT, F. QIAO and N. G. PURDIE, 2002 Milk synthetic response of the bovine mammary gland to an increase in the local concentration of arterial glucose. J Dairy Sci **85:** 494-503.

CARNINCI, P., A. SANDELIN, B. LENHARD, S. KATAYAMA, K. SHIMOKAWA *et al.*, 2006 Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet **38:** 626-635.

CLOP, A., F. MARCQ, H. TAKEDA, D. PIROTTIN, X. TORDOIR *et al.*, 2006 A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet **38:** 813-818.

COHEN-ZINDER, M., E. SEROUSSI, D. M. LARKIN, J. J. LOOR, A. EVERTS-VAN DER WIND *et al.*, 2005 Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. Genome Res **15:** 936-944.

DEKKERS, J. C., 2004 Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci **82 E-Suppl:** E313-328.

DENHARDT, D. T., and X. GUO, 1993 Osteopontin: a protein with diverse functions. Faseb J **7:** 1475-1482.

EDWARDS, Y. H., W. PUTT, M. FOX and J. H. IVES, 1995 A novel human phosphoglucomutase (PGM5) maps to the centromeric region of chromosome 9. Genomics **30:** 350-353.

EWING, B., and P. GREEN, 1998 Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res **8:** 186-194.

EWING, B., L. HILLIER, M. C. WENDL and P. GREEN, 1998 Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res **8:** 175-185.

FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics*. Pretntice Hall.

FARNIR, F., W. COPPIETERS, J. J. ARRANZ, P. BERZI, N. CAMBISANO *et al.*, 2000 Extensive genome-wide linkage disequilibrium in cattle. Genome Res **10:** 220-227.

FARRELL, H. M., JR., R. JIMENEZ-FLORES, G. T. BLECK, E. M. BROWN, J. E. BUTLER *et al.*, 2004 Nomenclature of the proteins of cows' milk--sixth revision. J Dairy Sci **87:** 1641-1674.

FISHER, R. A., 1930 *The genetical theory of natural selection*. Oxford University Press, Oxford.

FRANKS, P. W., U. EKELUND, S. BRAGE, J. LUAN, A. J. SCHAFER *et al.*, 2007 PPARGC1A coding variation may initiate impaired NEFA clearance during glucose challenge. Diabetologia **50:** 569-573.

FRAZER, K. A. P. I., D. G. BALLINGER, D. R. COX, D. A. HINDS, L. L. STUVE *et al.*, 2007 A second generation human haplotype map of over 3.1 million SNPs. Nature **449:** 851-861.

GABRIEL, S. B., S. F. SCHAFFNER, H. NGUYEN, J. M. MOORE, J. ROY *et al.*, 2002 The structure of haplotype blocks in the human genome. Science **296:** 2225-2229.

GELDERMANN, H., 1976 Investigations on Inheritance of Quantitative Characters in Animals by Gene Markers

II. Expected Effects. Theoretical and Applied Genetics **47**.

GEORGES, M., D. NIELSEN, M. MACKINNON, A. MISHRA, R. OKIMOTO *et al.*, 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics **139:** 907-920.

GIESSLER, A., 2006 Genomic characterisation of bovine PRKAA2 gene and association studies with lactation and beef traits in German Brown, pp. 75 in *Department für Tierwissenschaften*. TU München-Weihenstephan, Freising.

GORDON, D., C. ABAJIAN and P. GREEN, 1998 Consed: a graphical tool for sequence finishing. Genome Res **8:** 195-202.

GREENWOOD, T. A., and J. R. KELSOE, 2003 Promoter and intronic variants affect the transcriptional regulation of the human dopamine transporter gene. Genomics **82:** 511-520.

GREMME, G., V. BRENDEL, M. E. SPARKS and S. KURTZ, 2005 Engineering a Software Tool for Gene Prediction in Higher Organisms. Information and Software Technology **47:** 965-978.

HAGGER, C., 2005 Estimates of genetic diversity in the brown cattle population of Switzerland obtained from pedigree information. J Anim Breed Genet **122:** 405-413.

HANDSCHIN, C., and B. M. SPIEGELMAN, 2006 Peroxisome proliferator-activated receptor gamma coactivator 1 coactivators, energy homeostasis, and metabolism. Endocr Rev **27:** 728-735.

HARTL, D. L., and A. G. CLARK, 2007 *Principles of Population Genetics*. Sinauer Associates, Inc.

HAVLAK, P., R. CHEN, K. J. DURBIN, A. EGAN, Y. REN *et al.*, 2004 The Atlas genome assembly system. Genome Res **14:** 721-732.

HAYES, B., and M. E. GODDARD, 2001 The distribution of the effects of genes affecting quantitative traits in livestock. Genet Sel Evol **33:** 209-229.

HAYES, B., N. HAGESAETHER, T. ADNOY, G. PELLERUD, P. R. BERG *et al.*, 2006 Effects on production traits of haplotypes among casein genes in Norwegian goats and evidence for a site of preferential recombination. Genetics **174:** 455-464.

HAYES, H., 1995 Chromosome painting with human chromosome-specific DNA libraries reveals the extent and distribution of conserved segments in bovine chromosomes. Cytogenet Cell Genet **71:** 168-174.

HEDIGER, M. A., and D. B. RHOADS, 1994 Molecular physiology of sodium-glucose cotransporters. Physiol Rev **74:** 993-1026.

HEDRICK, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331-341.

HEDRICK, P. W., 1988 Inference of recombinational hotspots using gametic disequilibrium values. Heredity **60 (Pt 3):** 435-438.

HENRIKSEN, U., J. U. FOG, T. LITMAN and U. GETHER, 2005 Identification of intra- and intermolecular disulfide bridges in the multidrug resistance transporter ABCG2. J Biol Chem **280:** 36926-36934.

HILLENKAMP, F., M. KARAS, R. C. BEAVIS and B. T. CHAIT, 1991 Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem **63:** 1193A-1203A.

HIRANO, M., K. MAEDA, S. MATSUSHIMA, Y. NOZAKI, H. KUSUHARA *et al.*, 2005 Involvement of BCRP (ABCG2) in the biliary excretion of pitavastatin. Mol Pharmacol **68:** 800-807.

HOLT, C., 1983 Swelling of Golgi vesicles in mammary secretory cells and its relation to the yield and quantitative composition of milk. J Theor Biol **101:** 247-261.

IKONEN, T., M. OJALA and O. RUOTTINEN, 1999 Associations between milk protein polymorphism and first lactation milk production traits in Finnish Ayrshire cows. J Dairy Sci **82:** 1026-1033.

JONKER, J. W., G. MERINO, S. MUSTERS, A. E. VAN HERWAARDEN, E. BOLSCHER *et al.*, 2005 The breast cancer resistance protein BCRP (ABCG2) concentrates drugs and carcinogenic xenotoxins into milk. Nat Med **11:** 127-129.

JULING, K., 2007 Genomic Characterisation, Polymorphism Analysis and Association Studies of Candidate genes for BSE Susceptibility, pp. 158 in *Department für Tierwissenschaften*. TU München-Weihenstephan, Freising.

KARAS, M., and F. HILLENKAMP, 1988 Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem **60:** 2299-2301.

KARLSSON, E. K., I. BARANOWSKA, C. M. WADE, N. H. SALMON HILLBERTZ, M. C. ZODY *et al.*, 2007 Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet.

KEEMBIYEHETTY, C., R. AUGUSTIN, M. O. CARAYANNOPOULOS, S. STEER, A. MANOLESCU *et al.*, 2006 Mouse glucose transporter 9 splice variants are expressed in adult liver and kidney and are up-regulated in diabetes. Mol Endocrinol **20:** 686-697.

KHATIB, H., I. ZAITOUN, J. WIEBELHAUS-FINGER, Y. M. CHANG and G. J. ROSA, 2007 The association of bovine PPARGC1A and OPN genes with milk composition in two independent Holstein cattle populations. J Dairy Sci **90:** 2966-2970.

KHATKAR, M. S., A. COLLINS, J. A. CAVANAGH, R. J. HAWKEN, M. HOBBS *et al.*, 2006a A first-generation metric linkage disequilibrium map of bovine chromosome 6. Genetics **174:** 79-85.

KHATKAR, M. S., P. C. THOMSON, I. TAMMEN, J. A. CAVANAGH, F. W. NICHOLAS *et al.*, 2006b Linkage disequilibrium on chromosome 6 in Australian Holstein-Friesian cattle. Genet Sel Evol **38:** 463-477.

KHATKAR, M. S., P. C. THOMSON, I. TAMMEN and H. W. RAADSMA, 2004 Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet Sel Evol **36:** 163-190.

KHATKAR, M. S., K. R. ZENGER, M. HOBBS, R. J. HAWKEN, J. A. CAVANAGH *et al.*, 2007 A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in holstein-friesian cattle. Genetics **176:** 763-772.

KIM, Y., 2006 Allele frequency distribution under recurrent selective sweeps. Genetics **172:** 1967-1978.

KIM, Y., and W. STEPHAN, 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160:** 765-777.

KIRCHGEßNER, M., 1997 *Tierernährung*. DLG Verlag Frankfurt (Main).

KUEHN, C., C. EDEL, R. WEIKARD and G. THALLER, 2007 Dominance and parent-of-origin effects of coding and non-coding alleles at the acylCoA-diacylglycerol-acyltransferase (DGAT1) gene on milk production traits in German Holstein cows. BMC Genet **8:** 62.

KUHN, C., G. FREYER, R. WEIKARD, T. GOLDAMMER and M. SCHWERIN, 1999 Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim Genet **30:** 333-340.

KUSS, A. W., J. GOGOL, H. BARTENSCHLAGER and H. GELDERMANN, 2005 Polymorphic AP-1 binding site in bovine CSN1S1 shows quantitative differences in protein binding associated with milk protein expression. J Dairy Sci **88:** 2246-2252.

KUSUHARA, H., and Y. SUGIYAMA, 2007 ATP-binding cassette, subfamily G (ABCG family). Pflugers Arch **453:** 735-744.

LANGMANN, T., M. PORSCH-OZCURUMEZ, U. UNKELBACH, J. KLUCKEN and G. SCHMITZ, 2000 Genomic organization and characterization of the promoter of the human ATP-binding cassette transporter-G1 (ABCG1) gene. Biochim Biophys Acta **1494:** 175-180.

LARSEN, F., G. GUNDERSEN, R. LOPEZ and H. PRYDZ, 1992 CpG islands as gene markers in the human genome. Genomics **13:** 1095-1107.

LE HIR, H., A. NOTT and M. J. MOORE, 2003 How introns influence and enhance eukaryotic gene expression. Trends Biochem Sci **28:** 215-220.

LEONARD, S., H. KHATIB, V. SCHUTZKUS, Y. M. CHANG and C. MALTECCA, 2005 Effects of the osteopontin gene variants on milk production traits in dairy cattle. J Dairy Sci **88:** 4083-4086.

LEROUX, C., F. LE PROVOST, E. PETIT, L. BERNARD, Y. CHILLIARD *et al.*, 2003 Real-time RT-PCR and cDNA macroarray to study the impact of the genetic polymorphism at the alphas1-casein locus on the expression of genes in the goat mammary gland during lactation. Reprod Nutr Dev **43:** 459-469.

LEVIN, S., S. C. ALMO and B. H. SATIR, 1999 Functional diversity of the phosphoglucomutase superfamily: structural implications. Protein Eng **12:** 737-746.

LEWIS, S. E., S. M. SEARLE, N. HARRIS, M. GIBSON, V. LYER *et al.*, 2002 Apollo: a sequence annotation editor. Genome Biol **3:** RESEARCH0082.

LEWONTIN, R. C., 1988 On measures of gametic disequilibrium. Genetics **120:** 849-852.

LIANG, H., and W. F. WARD, 2006 PGC-1alpha: a key regulator of energy metabolism. Adv Physiol Educ **30:** 145-151.

LIEN, S., L. GOMEZ-RAYA, T. STEINE, E. FIMLAND and S. ROGNE, 1995 Associations between casein haplotypes and milk yield traits. J Dairy Sci **78:** 2047-2056.

LIN, C. Y., A. J. MCALLISTER, K. F. NG-KWAI-HANG and J. F. HAYES, 1986 Effects of milk protein loci on first lactation production in dairy cattle. J Dairy Sci **69:** 704-712.

LINDBLAD-TOH, K., C. M. WADE, T. S. MIKKELSEN, E. K. KARLSSON, D. B. JAFFE *et al.*, 2005 Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature **438:** 803-819.

LITMAN, T., M. BRANGI, E. HUDSON, P. FETSCH, A. ABATI *et al.*, 2000 The multidrug-resistant phenotype associated with overexpression of the new ABC half-transporter, MXR (ABCG2). J Cell Sci **113 (Pt 11):** 2011-2021.

LIU, Z., F. REINHARDT, A. BUNGER and R. REENTS, 2004 Derivation and calculation of approximate reliabilities and daughter yield-deviations of a random regression test-day model for genetic evaluation of dairy cattle. J Dairy Sci **87:** 1896-1907.

LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland, Massachusetts, USA.

MACKAY, T. F., 2001 The genetic architecture of quantitative traits. Annu Rev Genet **35:** 303-339.

MALÉCOT, G., 1948 *Les mathématiques de l'hérédité*.

MC PARLAND, S., J. F. KEARNEY, M. RATH and D. P. BERRY, 2007 Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. J Anim Sci **85:** 322-331.

MERINO, G., A. I. ALVAREZ, M. M. PULIDO, A. J. MOLINA, A. H. SCHINKEL *et al.*, 2006 Breast cancer resistance protein (BCRP/ABCG2) transports fluoroquinolone antibiotics and affects their oral availability, pharmacokinetics, and milk secretion. Drug Metab Dispos **34:** 690-695.

MERINO, G., J. W. JONKER, E. WAGENAAR, A. E. VAN HERWAARDEN and A. H. SCHINKEL, 2005 The breast cancer resistance protein (BCRP/ABCG2) affects pharmacokinetics, hepatobiliary excretion, and milk secretion of the antibiotic nitrofurantoin. Mol Pharmacol **67:** 1758-1764.

MEUWISSEN, T. H., B. J. HAYES and M. E. GODDARD, 2001 Prediction of total genetic value using genome-wide dense marker maps. Genetics **157:** 1819-1829.

MOSIG, M. O., E. LIPKIN, G. KHUTORESKAYA, E. TCHOURZYNA, M. SOLLER *et al.*, 2001 A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. Genetics **157:** 1683-1698.

MRODE, R. A., and G. J. T. SWANSON, 2004 Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. Livest. Prod. Sci. **86:** 253-260.

MUECKLER, M., 1994 Facilitative glucose transporters. Eur J Biochem **219:** 713-725.

MULLER, Y. L., C. BOGARDUS, O. PEDERSEN and L. BAIER, 2003 A Gly482Ser missense mutation in the peroxisome proliferator-activated receptor gamma coactivator-1 is associated with altered lipid oxidation and early insulin secretion in Pima Indians. Diabetes **52:** 895-898.

NADESALINGAM, J., Y. PLANTE and J. P. GIBSON, 2001 Detection of QTL for milk production on Chromosomes 1 and 6 of Holstein cattle. Mamm Genome **12:** 27-31.

NEMIR, M., D. BHATTACHARYYA, X. LI, K. SINGH, A. B. MUKHERJEE *et al.*, 2000 Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency. J Biol Chem **275:** 969-976.

NICKERSON, D. A., V. O. TOBE and S. L. TAYLOR, 1997 PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res **25:** 2745-2751.

NSENGIMANA, J., and P. V. BARET, 2004 Linkage disequilibrium and the genetic distance in livestock populations: the impact of inbreeding. Genet Sel Evol **36:** 281-296.

OLSEN, H. G., L. GOMEZ-RAYA, D. I. VAGE, I. OLSAKER, H. KLUNGLAND *et al.*, 2002 A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. J Dairy Sci **85:** 3124-3130.

OLSEN, H. G., S. LIEN, M. GAUTIER, H. NILSEN, A. ROSETH *et al.*, 2005 Mapping of a milk production quantitative trait locus to a 420-kb region on bovine chromosome 6. Genetics **169:** 275-283.

OLSEN, H. G., S. LIEN, M. SVENDSEN, H. NILSEN, A. ROSETH *et al.*, 2004 Fine mapping of milk production QTL on BTA6 by combined linkage and linkage disequilibrium analysis. J Dairy Sci **87:** 690-698.

OLSEN, H. G., H. NILSEN, B. HAYES, P. R. BERG, M. SVENDSEN *et al.*, 2007 Genetic support for a quantitative trait nucleotide in the ABCG2 gene affecting milk composition of dairy cattle. BMC Genet **8:** 32.

ORR, H. A., 1998 *The population genetics of adaption: the distribution of factors fixed during adaptive evolution*. Evolution.

PARKER, H. G., A. V. KUKEKOVA, D. T. AKEY, O. GOLDSTEIN, E. F. KIRKNESS *et al.*, 2007 Breed relationships facilitate fine-mapping studies: A 7.8-kb deletion cosegregates with Collie eye anomaly across multiple dog breeds. Genome Res.

PARSCH, J., C. D. MEIKLEJOHN and D. L. HARTL, 2001 Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of Drosophila simulans. Genetics **159:** 647-657.

PHAY, J. E., H. B. HUSSAIN and J. F. MOLEY, 2000 Cloning and expression analysis of a novel member of the facilitative glucose transporter family, SLC2A9 (GLUT9). Genomics **66:** 217-220.

PRINZENBERG, E. M., C. WEIMANN, H. BRANDT, J. BENNEWITZ, E. KALM *et al.*, 2003 Polymorphism of the bovine CSN1S1 promoter: linkage mapping, intragenic haplotypes, and effects on milk production traits. J Dairy Sci **86:** 2696-2705.

RAMUNNO, L., G. COSENZA, A. RANDO, R. ILLARIO, D. GALLO *et al.*, 2004 The goat alphas1-casein gene: gene structure and promoter analysis. Gene **334:** 105-111.

REESE, M. G., 2001 Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem **26:** 51-56.

RICHARDSON, S., G. NEAMA, T. PHILLIPS, S. BELL, S. D. CARTER *et al.*, 2003 Molecular characterization and partial cDNA cloning of facilitative glucose transporters expressed in human articular chondrocytes; stimulation of 2-deoxyglucose uptake by

IGF-I and elevated MMP-2 secretion by glucose deprivation. Osteoarthritis Cartilage **11:** 92-101.

RIJNKELS, M., 2002 Multispecies comparison of the casein gene loci and evolution of casein gene family. J Mammary Gland Biol Neoplasia **7:** 327-345.

RITTLING, S. R., and K. E. NOVICK, 1997 Osteopontin expression in mammary gland development and tumorigenesis. Cell Growth Differ **8:** 1061-1069.

ROBERTSON, A., 1969 *The nature of quantitative genetic variation*. University of Wisconsin Press, Madison, WI, (USA)

RON, M., G. ISRAELI, E. SEROUSSI, J. I. WELLER, J. P. GREGG *et al.*, 2007 Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. BMC Genomics **8:** 183.

RON, M., D. KLIGER, E. FELDMESSER, E. SEROUSSI, E. EZRA *et al.*, 2001 Multiple quantitative trait locus analysis of bovine chromosome 6 in the Israeli Holstein population by a daughter design. Genetics **159:** 727-735.

RON, M., and J. I. WELLER, 2007 From QTL to QTN identification in livestock - winning by points rather than knock-out: a review. Anim Genet.

SABOUR, M. P., C. Y. LIN, A. J. LEE and A. J. MCALLISTER, 1996 Association between milk protein genetic variants and genetic values of Canadian Holstein bulls for milk yield traits. J Dairy Sci **79:** 1050-1056.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:** 5463-5467.

SAUER, S., and I. G. GUT, 2002 Genotyping single-nucleotide polymorphisms by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci **782:** 73-87.

SCHAEFFER, L. R., 2006 Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet **123:** 218-223.

SCHEET, P., and M. STEPHENS, 2006 A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet **78:** 629-644.

SCHMITT-NEY, M., W. DOPPLER, R. K. BALL and B. GRONER, 1991 Beta-casein gene promoter activity is regulated by the hormone-mediated relief of transcriptional repression and a mammary-gland-specific nuclear factor. Mol Cell Biol **11:** 3745-3755.

SCHNABEL, R. D., J. J. KIM, M. S. ASHWELL, T. S. SONSTEGARD, C. P. VAN TASSELL *et al.*, 2005 Fine-mapping milk production quantitative trait loci on BTA6: analysis of the bovine osteopontin gene. Proc Natl Acad Sci U S A **102:** 6896-6901.

SCOTT, R. A., D. E. BEUMAN and J. H. CLARK, 1976 Cellular gluconeogenesis by lactating bovine mammary tissue. J Dairy Sci **59:** 50-56.

SOBEL, E., and K. LANGE, 1996 Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet **58:** 1323-1337.

SOLINAS-TOLDO, S., C. LENGAUER and R. FRIES, 1995 Comparative genome map of human and cattle. Genomics **27:** 489-496.

SONG, S., Y. ZHANG, K. MA, L. JACKSON-HAYES, E. N. LAVRENTYEV *et al.*, 2004 Peroxisomal proliferator activated receptor gamma coactivator (PGC-1alpha) stimulates carnitine palmitoyltransferase I (CPT-Ialpha) through the first intron. Biochim Biophys Acta **1679:** 164-173.

SORENSEN, A. C., M. K. SORENSEN and P. BERG, 2005 Inbreeding in Danish dairy cattle breeds. J Dairy Sci **88:** 1865-1872.

SPELMAN, R. J., W. COPPIETERS, L. KARIM, J. A. VAN ARENDONK and H. BOVENHUIS, 1996 Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics **144:** 1799-1808.

STAM, L. F., and C. C. LAURIE, 1996 Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in Drosophila melanogaster. Genetics **144:** 1559-1564.

STOREY, J. D., and R. TIBSHIRANI, 2003 Statistical significance for genomewide studies. Proc Natl Acad Sci U S A **100:** 9440-9445.

STORM, N., B. DARNHOFER-PATEL, D. VAN DEN BOOM and C. P. RODI, 2003 MALDI-TOF mass spectrometry-based SNP genotyping. Methods Mol Biol **212:** 241-262.

SU, A. I., M. P. COOKE, K. A. CHING, Y. HAKAK, J. R. WALKER *et al.*, 2002 Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A **99:** 4465-4470.

SUTTER, N. B., M. A. EBERLE, H. G. PARKER, B. J. PULLAR, E. F. KIRKNESS *et al.*, 2004 Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res **14:** 2388-2396.

TENESA, A., S. A. KNOTT, D. WARD, D. SMITH, J. L. WILLIAMS *et al.*, 2003 Estimation of linkage disequilibrium in a sample of the United Kingdom dairy cattle population using unphased genotypes. J Anim Sci **81:** 617-623.

THALLER, G., W. KRAMER, A. WINTER, B. KAUPE, G. ERHARDT *et al.*, 2003 Effects of DGAT1 variants on milk production traits in German cattle breeds. J Anim Sci **81:** 1911-1918.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN and D. G. HIGGINS, 1997 The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res **25:** 4876-4882.

THORENS, B., 1996 Glucose transporters in the regulation of intestinal, renal, and liver glucose fluxes. Am J Physiol **270:** G541-553.

THREADGILL, D. W., and J. E. WOMACK, 1990 Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res **18:** 6935-6942.

THREADGOLD, L. C., and N. J. KUHN, 1979 Glucose-6-phosphate hydrolysis by lactating rat mammary gland. Int J Biochem **10:** 683-685.

VAN HERWAARDEN, A. E., E. WAGENAAR, G. MERINO, J. W. JONKER, H. ROSING *et al.*, 2007 Multidrug Transporter ABCG2/Breast Cancer Resistance Protein Secretes Riboflavin (Vitamin B2) into Milk. Mol Cell Biol **27:** 1247-1253.

VAN LAERE, A. S., M. NGUYEN, M. BRAUNSCHWEIG, C. NEZER, C. COLLETTE *et al.*, 2003 A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature **425:** 832-836.

VELMALA, R., J. VILKKI, K. ELO and A. MAKI-TANILA, 1995 Casein haplotypes and their association with milk production traits in the Finnish Ayrshire cattle. Anim Genet **26:** 419-425.

VELMALA, R. J., H. J. VILKKI, K. T. ELO, D. J. DE KONING and A. V. MAKI-TANILA, 1999 A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Anim Genet **30:** 136-143.

VIITALA, S. M., N. F. SCHULMAN, D. J. DE KONING, K. ELO, R. KINOS *et al.*, 2003 Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. J Dairy Sci **86:** 1828-1836.

VILLANUEVA, B., J. C. DEKKERS, J. A. WOOLLIAMS and P. SETTAR, 2004 Maximizing genetic gain over multiple generations with quantitative trait locus selection and control of inbreeding. J Anim Sci **82:** 1305-1314.

VIT, 2004 Beschreibung der Zuchtwertschätzung für Milchleistungsmerkmale, Zellzahl, Exterieurmerkmale, Nutzungsdauer und Zuchtleistungsmerkmale.

VONDERHAAR, B. K., and S. E. ZISKA, 1989 Hormonal regulation of milk protein gene expression. Annu Rev Physiol **51:** 641-652.

WEBER, J. L., and P. E. MAY, 1989 Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet **44:** 388-396.

WEIKARD, R., T. GOLDAMMER, J. E. WOMACK and C. KUEHN, 2007 Forward to a detailed sequence map of bovine chromosome 6. Cytogenet Genome Res **116:** 72-79.

WEIKARD, R., C. KUHN, T. GOLDAMMER, G. FREYER and M. SCHWERIN, 2005 The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. Physiol Genomics **21:** 1-13.

WELLER, J. I., H. WELLER, D. KLIGER and M. RON, 2002 Estimation of quantitative trait locus allele frequency via a modified granddaughter design. Genetics **162:** 841-849.

WHITEHOUSE, D. B., J. TOMKINS, J. U. LOVEGROVE, D. A. HOPKINSON and W. O. MCMILLAN, 1998 A phylogenetic approach to the identification of phosphoglucomutase genes. Mol Biol Evol **15:** 456-462.

WIENER, P., I. MACLEAN, J. L. WILLIAMS and J. A. WOOLLIAMS, 2000 Testing for the presence of previously identified QTL for milk production traits in new populations. Anim Genet **31:** 385-395.

WINTER, A., W. KRAMER, F. A. WERNER, S. KOLLERS, S. KATA *et al.*, 2002 Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. Proc Natl Acad Sci U S A **99:** 9300-9305.

WINTER, A., M. VAN ECKEVELD, O. R. BININDA-EMONDS, F. A. HABERMANN and R. FRIES, 2003 Genomic organization of the DGAT2/MOGAT gene family in cattle (Bos taurus) and other mammals. Cytogenet Genome Res **102:** 42-47.

WYMAN, A. R., and R. WHITE, 1980 A highly polymorphic locus in human DNA. Proc Natl Acad Sci U S A **77:** 6754-6758.

YOSHIMURA, M., and T. OKA, 1990 Hormonal induction of beta-casein gene expression: requirement of ongoing protein synthesis for transcription. Endocrinology **126:** 427-433.

ZENGER, K. R., M. S. KHATKAR, J. A. CAVANAGH, R. J. HAWKEN and H. W. RAADSMA, 2007 Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. Anim Genet **38:** 7-14.

ZHANG, Q., D. BOICHARD, I. HOESCHELE, C. ERNST, A. EGGEN *et al.*, 1998 Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. Genetics **149:** 1959-1973.

ZHANG, Y., L. W. CASTELLANI, C. J. SINAL, F. J. GONZALEZ and P. A. EDWARDS, 2004 Peroxisome proliferator-activated receptor-gamma coactivator 1alpha (PGC-1alpha) regulates triglyceride metabolism by activation of the nuclear receptor FXR. Genes Dev **18:** 157-169.

ZHAO, F. Q., and A. F. KEATING, 2007 Expression and regulation of glucose transporters in the bovine mammary gland. J Dairy Sci **90 Suppl 1:** E76-86.

11. Appendices

11.1. *Buffers*

TE (10mM Tris-Cl, 1mM EDTA, pH 8.0)

TBE 10x (900mM Tris-Cl, 900mM Borate, 20mM EDTA, pH 8.3)

11.2. *Primers*

Table 11.1 Primers for SNP-screening

Gene	Primer	Pair	Lab-Name	Sequence	Region
ABCG2	4532	4533	ABCG2_4532up	CAGGGCATTTGAAAATAGTTGG	Upstream
	4533	4532	ABCG2_4533dn	ACGCTACAAATCCTGCTTCC	Upstream
	4530	4531	ABCG2_4530up	CCCTCCGAACACCCTTAGAT	Exon 1
	4531	4530	ABCG2_4531dn	CTCTGGCTCCTTTCCTCCTT	Exon 1
	3898	3899	ABCG2_3898up	CTGGAAGTTTGTACCAAAGCA	Exon 2
	3899	3898	ABCG2_3899dn	TGATCCAACCCTTTCACTGC	Exon 2
	3900	3901	ABCG2_3900up	TCAACAATTCAGTTTCTCCTTCC	Exon 3
	3901	3900	ABCG2_3901dn	GACGGCCTGACTTATGATGG	Exon 3
	4025	4026	ABCG2_4025up	CCCAAGGTCAGACAACCAAG	Exon 4
	4026	4025	ABCG2_4026dn	GCACACTCAACTATCAGCCAAG	Exon 4
	4029	4030	ABCG2_4029up	ACCTCGACCTGCCAATTTTA	Intron 4
	4030	4029	ABCG2_4030dn	TGGGCTGCTGAGAACTGTAA	Intron 4
	4027	4028	ABCG2_4027up	CTTGGGTATCTTATTTTTGTGGA	Exon 5
	4028	4027	ABCG2_4028dn	GAACACATTGCCACACTTGC	Exon 5
	4033	4034	ABCG2_4033up	AACTCTTTTAGGGCTGGAACA	Exon 6
	4034	4033	ABCG2_4034dn	CCACGCAGTAGATCAGTGAAA	Exon 6
	4035	4036	ABCG2_4035up	GGAATAAATGATCTAAAGGCAGGA	Exon 7
	4036	4035	ABCG2_4036dn	CACGGGCATGTACTCAAACA	Exon 7
	4037	4038	ABCG2_4037up	TGTCTTCCCTAGCAGCACCT	Exon 8
	4038	4037	ABCG2_4038dn	GGCACTTCTTTTAAAATGGGAAC	Exon 8
	3912	3913	ABCG2_3912up	AGACCGAAGAGCCTTCCAA	Exon 9
	3913	3912	ABCG2_3913dn	GAGCTATAGAAGCCTGGGGATT	Exon 9
	3914	3915	ABCG2_3914up	ACTGGCAGTGCCCGCTAA	Exon 10
	3915	3914	ABCG2_3915dn	GCTTTCAGTTCACTGACTTGTCC	Exon 10
	3916	3917	ABCG2_3916up	GCTCATGCTTCCCTCTCTGT	Exon 11
	3917	3916	ABCG2_3917dn	GACGGTGCCCTAATCTTGAA	Exon 11
	4039	4040	ABCG2_4039up	CCTCTGGTAGCAGTGGAACTC	Exon 12
	4040	4039	ABCG2_4040dn	CCCTCCCATCCAAGACTAAA	Exon 12
	3920	3921	ABCG2_3920up	AGTCCCTGATAGACAGTGTTGG	Exon 13
	3921	3920	ABCG2_3921dn	CAGCCTCCTTAAAGCAGAGTC	Exon 13
	3922	3923	ABCG2_3922up	ACAGCCTCAGCTCCAGAGAG	Exon 14
	3923	3922	ABCG2_3923dn	ATCTCCCAATCCTTCAGCTC	Exon 14
	4041	4042	ABCG2_4041up	TGTATCAGAGCCCCAAATCC	Exon 15
	4042	4041	ABCG2_4042dn	AAAGGATTCCTCAGCCCAGT	Exon 15
	3926	3927	ABCG2_3926up	CATGCACATACACACACCACA	Exon 16
	3927	3926	ABCG2_3927dn	GGCAAGTGAAAAGAAGACAACC	Exon 16
OPN	4358	4359	OPN_4358up	GGTGCTGCCAAAGAAAAGTT	Upstream
	4359	4358	OPN_4359dn	AAAATCCCCAGTCACGTTCA	Upstream
	4356	4357	OPN_4356up	CCCTGCTCGGTCATAAACTG	Upstream
	4357	4356	OPN_4357dn	AAGAGTCCAGTCCCCTGTGA	Upstream
	3862	3863	OPN_3862up	GGCTGGGTAGTGGCAAAAT	Exon 1
	3863	3862	OPN_3863dn	AAAATAACCGGCCACCTTTC	Exon 1
	4600	4601	OPN_4600up	TGCCTCATTTCATTGGGAAG	Intron 1

	4601	4600	OPN_4601dn	TTTTGTGTGGAAACACTCCA	Intron 1
	3864	3865	OPN_3864up	TTCCTCTCTCCCTTGCCTAA	Exon 2+3
	3865	3864	OPN_3865dn	GAAGCATAGCAGGCACACAA	Exon 2+3
	4007	4008	OPN_4007up	TCACTTAGAGACCCCTGTTTCTTT	Exon 4
	4008	4007	OPN_4008dn	ACAGGCAGGGAAGATCCATT	Exon 4
	4602	4603	OPN_4602up	GGCTAAAGCCTGACCCATCT	Intron 4
	4603	4602	OPN_4603dn	TTGTTGTCATCAGTTTCCTCAGA	Intron 4
	3868	3869	OPN_3868up	TCTTCCGGCTTAATCCTTCC	Exon 5
	3869	3868	OPN_3869dn	CACCTCCCTATTAGCCTGGA	Exon 5
	4691	4692	OPN_4691up	CCTCTGAGGAAACTGATGACAA	Intron 5
	4692	4691	OPN_4692dn	TCAGGGCTTTCATTGGACTT	Intron 5
	3870	3871	OPN_3870up	GCCATTAAGTGCTTTGTTGTGA	Exon 6
	3871	3870	OPN_3871dn	ATGGACTTGAGCGCAAAAAC	Exon 6
	4604	4605	OPN_4604up	TCGCAGTTTTCACTCCGTTT	Intron 6
	4605	4604	OPN_4605dn	TGTGTGATGTGAAGTCCTCCTC	Intron 6
	3872	3873	OPN_3872up	GAGATGACATAATGAAAGAACTTGG	Exon 7
	3873	3872	OPN_3873dn	GGTGTTACCATGAAGCCACA	Exon 7
PPARGC1A	4596	4597	PPARGC1A_4596up	TTGCGCTTTCAAACACTCC	Upstream
	4597	4596	PPARGC1A_4597dn	TCCATACAGAGTCCTGGTTGC	Upstream
	4009	4010	PPARGC1A_4009up	TGACGTCACAAGTTTGAGCA	Exon 1
	4010	4009	PPARGC1A_4010dn	CCAGGCCACTAAACACCACT	Exon 1
	4011	4012	PPARGC1A_4011up	AAACCAGCTTTGGTGGTGTC	Exon 2
	4012	4011	PPARGC1A_4012dn	TGGTGACAGCAAAGCAAGAA	Exon 2
	4013	4014	PPARGC1A_4013up	GACCAAAAGGGAGTTTGGAA	Exon 3
	4014	4013	PPARGC1A_4014dn	TTGGGGATTATTTTGGCATC	Exon 3
	4146	4147	PPARGC1A_4146up	ATTGAGGGGAAAATGCACAC	Exon 4
	4147	4146	PPARGC1A_4147dn	TGGCACTGCAGCATCTTAAA	Exon 4
	4148	4149	PPARGC1A_4148up	GGCGCTGTAGGGTCTTGTTA	Exon 5
	4149	4148	PPARGC1A_4149dn	TCAGGAGCCTTTTGGTGATT	Exon 5
	4150	4151	PPARGC1A_4150up	TTCGTGTCTAAACAAACCGTGT	Exon 6 + 7
	4151	4150	PPARGC1A_4151dn	TGTCTTAAGTGGGAGGAGGTTT	Exon 6 + 7
	4021	4022	PPARGC1A_4021up	AACCATTGACTGAAATCTTAGGG	Exon 8
	4022	4021	PPARGC1A_4022dn	GCGGTCTCTCTCAGGTAGCA	Exon 8
	3887	3888	PPARGC1A_3887up	AGGAGCTCCATGACTCCAGA	Exon 8
	3888	3887	PPARGC1A_3888dn	GCTGGTGGGTTTTAAATGCT	Exon 8
	3889	3890	PPARGC1A_3889up	TTCCAGGGGCTACTCAGTCA	Exon 9+10
	3890	3889	PPARGC1A_3890dn	CCCCCAAAATGAGAGAGAGA	Exon 9+10
	3891	3892	PPARGC1A_3891up	TGGGGGTTCTTGTTATGCAA	Exon 11
	3892	3891	PPARGC1A_3892dn	TGGGACAGGAAATCTCAACTG	Exon 11
	4023	4024	PPARGC1A_4023up	CGATAGCATTGGCATCATTG	Intron 11
	4024	4023	PPARGC1A_4024dn	ACGACCCAGAGGGATGGTAT	Intron 11
	4688	4754	PPARGC1A_4688up	TTTTCCTCTTTGTCTGTTGCAG	Exon 12
	4754	4688	PPARGC1A_4754dn	TCCATTTTGAACTCAACATTATGG	Exon 12
	4378	4379	PPARGC1A_4378up	TCCATTTGAAAGAGCTTGATGA	Intron 12
	4379	4378	PPARGC1A_4379dn	ACCAGCTTTCATGGGCATAC	Intron 12
	4380	4381	PPARGC1A_4380up	TCCCAGATCCCAACACAAAT	Intron 12
	4381	4380	PPARGC1A_4381dn	GCCCATCTTTCATCACCAAA	Intron 12
	4384	4385	PPARGC1A_4384up	TCATCACTCCTTACCCTCTGC	Exon 13
	4385	4384	PPARGC1A_4385dn	TGTCTCTTGCCTCTTCAGCA	Exon 13
PGM2	4296	4297	PGM2_4296up	GCTGCCTCTCTTCCTTCTCA	Upstream
	4297	4296	PGM2_4297dn	CACTGATGCCATGACACTGAT	Upstream
	4614	4615	PGM2_4614up	AGCGCCACCTCTGAAATAAA	Upstream
	4615	4614	PGM2_4615dn	AAGCCATTCATTCAGGTCCTA	Upstream
	4288	4289	PGM2_4288up	AACACACACACACACGTCCA	Upstream
	4289	4288	PGM2_4289dn	CTTGTCCCATCTCAGCCACT	Upstream
	4689	4690	PGM2_4689up	TGCCTGGAGCTCAGTAGGAC	Upstream
	4690	4689	PGM2_4690dn	GTCTGAGCCGTCTCCTGGT	Upstream
	3504	3505	PGM2_3504up	TAAGGCGAGCGAGTCATTTT	Exon 1

	3505	3504	PGM2_3505dn	GAGGGAAGGAAGCACTTTCA	Exon 1
	3590	3591	PGM2_3590up	ATTGTGGGTTGGTTGGTTGT	Exon 2
	3591	3590	PGM2_3591dn	GCCTGATCCTGAAAGAGAGC	Exon 2
	3323	3324	3323_PGM2_e3f	TCAATTTTTGGTTTGGTGTGTT	Exon 3
	3324	3323	3324_PGM2_e3r	GTGGGCCTTACAAAGCACTG	Exon 3
	3399	3400	3399_PGM2_i3f	GCATCATCCTTGAGGGTTTT	Intron 3
	3400	3399	3400_PGM2_i4r	CTGATCCTGGGACAGGAAAT	Intron 3
	3518	3519	PGM2_3518up	AAGCCAGATGCTGTGGAAAT	Exon 4
	3519	3518	PGM2_3519dn	ACTAGGGTGTGGGAGGAACA	Exon 4
	3401	3403	3401_PGM2_i4f	TTCTCAGCTACCGCAATTAGG	Intron 4
	3403	3401	3403_PGM2_i4r	CATTGTTTGTACCATTTTCCCTTT	Intron 4
	3325	3326	3325_PGM2_e5r	TGAAACTTTGTGCTGGAATCA	Exon 5 + 6
	3326	3325	3326_PGM2_e6r	GAAGCGCTTGGGTCATGTAG	Exon 5 + 6
	3588	3589	PGM2_3588up	GTGTTGTGGGAGCGTGTATG	Exon 7
	3589	3588	PGM2_3589dn	GGCACAATGAAAACACACAAA	Exon 7
	3349	3350	3349_PGM2_e8f_p	ATCTTTTGCTTTGGCTGACA	Exon 8
	3350	3349	3350_PGM2_e8r_p	TGTCTTGCTTTTCTGCCACA	Exon 8
	3405	3408	3405_PGM2_i7f	TGTGCTCAGTTCTTGGTTGG	Intron 8
	3408	3405	3408_PGM2_e9r	CCCATCCACTTAAAGCCAGT	Intron 8
	3508	3509	PGM2_3508up	TGCCGCTTACAGTTAAGCAC	Exon 9+10
	3509	3508	PGM2_3509dn	CTTTGCCCGTTTATTTGCAT	Exon 9+10
	3510	3511	PGM2_3510up	CTCATGCAATAGGCAGGATG	Exon 11
	3511	3510	PGM2_3511dn	CGGGTGCACAAGAAGAGAGT	Exon 11
	3522	3523	PGM2_3522up	TTGGACCTCTTTATTTAATGCTGA	Exon 12
	3523	3522	PGM2_3523dn	GCACAGTGAAAAAGTAAATCGAGA	Exon 12
	3524	3525	PGM2_3524up	CCGAGTACATTTTCTAACTTCCTCA	Exon 13
	3525	3524	PGM2_3525dn	ACATACTGGACAGATGAGCTATTTT	Exon 13
	3516	3517	PGM2_3516up	CATGCTCCAGAATAATCCACTG	Exon 14
	3517	3516	PGM2_3517dn	TCAAACAGGAGATGGCAAGA	Exon 14
CSN1S1	4336	4337	CSN1S1_4336up	GGATTTTTGTTCTCAGCTCCTT	Upstream
	4337	4336	CSN1S1_4337dn	GGACAAAATGGGGTCTTCAA	Upstream
	4334	4335	CSN1S1_4334up	CAACTAGTACACCCAAAATGAACAA	Upstream
	4335	4334	CSN1S1_4335dn	GGTGATGGCAGACTTTTGCT	Upstream
	4396	4397	CSN1S1_4396up	TGGGAGACGAACTGAACAGA	Exon 1
	4397	4396	CSN1S1_4397dn	CACAAAATAAAAATGGCCTTGA	Exon 1
	4160	4161	CSN1S1_4160up	CTGAAAGAGCAGGCTAAAGGA	Exon 2
	4161	4160	CSN1S1_4161dn	TGATCAAAATGTGAAGGATTACCA	Exon 2
	4162	4163	CSN1S1_4162up	CAGGGGAACTTGGTGTCAAA	Exon 3
	4163	4162	CSN1S1_4163dn	ACACCTCAGCGAATGTGAAA	Exon 3
	4164	4165	CSN1S1_4164up	CCTTCATTTCAATTGTATCCAGA	Exon 4
	4165	4164	CSN1S1_4165dn	TGCACAAAACAGCGTAAGTG	Exon 4
	4166	4167	CSN1S1_4166up	TGATAGGCAACCCAATTTAGC	Exon 5 + 6
	4167	4166	CSN1S1_4167dn	TCCCTCAAAAATTCTCTTTGTCA	Exon 5 + 6
	4168	4169	CSN1S1_4168up	TGCAACCATGAGGAGACAAG	Exon 7
	4169	4168	CSN1S1_4169dn	TTGTCCTTAGAGAATTGTTCATTCA	Exon 7
	4063	4064	CSN1S1_4063up	GGTCATGGAAACTGGACACA	Exon 8
	4064	4063	CSN1S1_4064dn	GGGTATGATAGCACTGCTTTAGG	Exon 8
	4170	4171	CSN1S1_4170up	TTGGGAACAGAAATCATATCCA	Exon 9
	4171	4170	CSN1S1_4171dn	TGGACACCACAGATATACAGATAGG	Exon 9
	4172	4173	CSN1S1_4172up	CCCAGGAATTTGTGGCTAAA	Exon 10+11
	4173	4172	CSN1S1_4173dn	CCACATTTACCGAGGGCTTA	Exon 10+11
	4069	4070	CSN1S1_4069up	GATGGTATCCACGAAATTGACA	Exon 12
	4070	4069	CSN1S1_4070dn	GGGATTAGGGGAAGGAGAGA	Exon 12
	4174	4175	CSN1S1_4174up	TGCTATGTTCATGAGACCTTTGA	Exon 13
	4175	4174	CSN1S1_4175dn	CATTATTGACTTCCTCCTTCCTACA	Exon 13
	4176	4177	CSN1S1_4176up	GTGGATTATTTTCCCCCAAA	Exon 14
	4177	4176	CSN1S1_4177dn	CAGCTCCCATTCCAAAGAAG	Exon 14
	4178	4179	CSN1S1_4178up	CAGCCTTAAAAGATCACCCCTA	Exon 15

	4179	4178	CSN1S1_4179dn	GCTCTCTTCACTCAGGGATGG	Exon 15
	4180	4181	CSN1S1_4180up	TCATGGGCCAAAGTAAGAGAA	Exon 16
	4181	4180	CSN1S1_4181dn	GGAGAGTCCTCGTCCCAGA	Exon 16
	4182	4183	CSN1S1_4182up	TTGATTTCTCTGTTTTCCTCACA	Exon 17
	4183	4182	CSN1S1_4183dn	TTAGCAAAAGTGCAAGGAGGA	Exon 17
	4184	4185	CSN1S1_4184up	TGCTGGGAAAATTAGTGCTCA	Exon 18
	4185	4184	CSN1S1_4185dn	GAGGACCAGCAAAACAACTTT	Exon 18
	4186	4187	CSN1S1_4186up	TCAGGAGCAGTGGGTATGTG	Exon 19
	4187	4186	CSN1S1_4187dn	GGATCAGCTAGAAGTGAACATTTG	Exon 19
SLC2A9	4047	4048	SLC2A9_4047up	ACACAGCAGTGGTCCAACCT	Exon 2
	4048	4047	SLC2A9_4048dn	GTGGTCATAGCCCAGCATTC	Exon 2
	4156	4157	SLC2A9_4156up	CTGAGCTCCCTGGACTGG	Exon 3
	4157	4156	SLC2A9_4157dn	GAGGCACTCACCGGAGTC	Exon 3
	3930	3931	SLC2A9_3930up	GCTTAGCGAAATGCCTGGTA	Exon 4
	3931	3930	SLC2A9_3931dn	GCTCCTCCTTGTTTGGACAG	Exon 4
	3932	3933	SLC2A9_3932up	ACACCAGGGAAATCCACAGA	Exon 5
	3933	3932	SLC2A9_3933dn	CCCAGTCATGTGACAACCAG	Exon 5
	3934	3935	SLC2A9_3934up	ACAGAAAGCGTTCCCTGAAG	Exon 6
	3935	3934	SLC2A9_3935dn	AAACGAAAGGCTTGAGTGCT	Exon 6
	3942	3943	SLC2A9_3942up	AACAAATCCAGTGTCCAGTCC	Exon 7
	3943	3942	SLC2A9_3943dn	ACAGATGGGAAGACGAAACG	Exon 7
	3944	3945	SLC2A9_3944up	GTCTGGATGGTGGGACAGAC	Exon 8
	3945	3944	SLC2A9_3945dn	TCTGGGCTAGCTGCTTCATT	Exon 8
	3936	3937	SLC2A9_3936up	CAAATCATCACCTTGGCTCA	Exon 9
	3937	3936	SLC2A9_3937dn	AAGAAGACAACAGGGAAACAAGA	Exon 9
	3938	3939	SLC2A9_3938up	GAGCCATCCTTTAGCCATGA	Exon 10
	3939	3938	SLC2A9_3939dn	CCAGTCCAAGGACCCATTAC	Exon 10
	3940	3941	SLC2A9_3940up	GCATCATGCCACTTCACTGT	Exon 11
	3941	3940	SLC2A9_3941dn	CGACCATCAGTTGGATGAAA	Exon 11
	3946	3947	SLC2A9_3946up	GGACCAGGTCAGAGAAATGG	Exon 12
	3947	3946	SLC2A9_3947dn	AAGGGCAACTCCCAAAAGAT	Exon 12
	3948	3949	SLC2A9_3948up	GTGCACCTCATGTGAGAACG	Exon 13
	3949	3948	SLC2A9_3949dn	GGGGATTCAGCTGTGAAGAA	Exon 13

Table 11.2 Primers for genotyping

Well	Gene_SNP	Sequence		
		Forward PCR-Primer	Reverse PCR-Primer	Extension Primer
1 [a]	abcg2_1303	ACGTTGGATGTCAAAAAGTTTTGTTTCCC	ACGTTGGATGCCCAGGGCCAATAAAATCAG	TTTGTTTCCCCCTCCG
	slc2a9_1213	ACGTTGGATGTAGCCTTCCGGACCTTCTTG	ACGTTGGATGATACCAGCCGGATGTTCCTC	GACCTTCTTGGGCAAA
	pgm2_1321	ACGTTGGATGACTTCTATCCCGCTGCCCC	ACGTTGGATGCCCGGAGGAAAATGACTCG	AGGAGGCGTGCTCTTA
	csn1s1_1161	ACGTTGGATGATACCTTGAACCCAAGACTG	ACGTTGGATGGGAAGCAAAAGTCTGCCATC	AAGACTGGGAAGAAGC
	abcg2_1287	ACGTTGGATGACTGTCAGAGTTCCCATCAC	ACGTTGGATGCTACAGAGTGATGGTATTAG	GTTCCCATCACAACATC
	slc2a9_1206	ACGTTGGATGTATGGGATGTCCGTGCCTTC	ACGTTGGATGAGCTGGGAGGAATAACATGG	CCTTCTCTCCCACGACTC
	pgm2_858	ACGTTGGATGTGTTGGGAACTCAGGATCAG	ACGTTGGATGAGTCGGCTTTCAAGGCATTC	gAGGGTCCTTCTGTTCGG
	pgm2_837	ACGTTGGATGTGACCTGAAGCAAAGAGGTG	ACGTTGGATGTTACGTACCTTCTGCTGCTG	AAAGAGGTGTTGTGATCAG
	ppargc1a_1244	ACGTTGGATGACTCGTTAGTACATCACAGG	ACGTTGGATGTGGAAGAGTGTATTTGCCCC	AGTACATCACAGGAGCTCCA
	ppargc1a_1246	ACGTTGGATGGTTCAAGATCGCCCCATAGC	ACGTTGGATGTCTACTCGAGGGATAAGAGG	cGGTAATGATGCACGTTCGC
	slc2a9_1225	ACGTTGGATGTCACCATTACGCTGATGCTG	ACGTTGGATGAATGGTCCCAGAGCATGAGG	CCAGGGAGTTCAGGATGGGC
	slc2a9_1208	ACGTTGGATGGGGATTCAATTTTCCTTGGC	ACGTTGGATGAGAATTCCACTCAGTCCAGC	cTTTTCCTTGGCTTTTCCTTA
	csn1s1_1191	ACGTTGGATGTTACCACCACAGTGGCATAG	ACGTTGGATGATGCCCCATCATTCTCTGAC	aCACAGTGGCATAGTAGTCTTT
	csn1s1_1170	ACGTTGGATGGAGGAATTGGTAAAGTGGAG	ACGTTGGATGTGGCGTTTTCATCATCAAGC	cAGAAAGCTGTGCAGTTCTGTCTT
	csn1s1_1178	ACGTTGGATGGTATAAACGGTATTCTCTGCC	ACGTTGGATGGTAAGTATAACATTGAGTTG	cTAAACGGTATTCTCTGCCAAAATT
	ppargc1a_1253	ACGTTGGATGCTCCTTACCCTCTGCAAATA	ACGTTGGATGTGGTAGATGGTATGATATA	TGCAAATATATACAAGTGGTATAAA
	abcg2_1302	ACGTTGGATGCGCCAGTACATCTGAAATTC	ACGTTGGATGGGTTTCTTTAAGGAACAGTG	TTCAACAGAAAAAGAATATAACTTCA
	csn1s1_1155	ACGTTGGATGGGTGGTATAATTAAAATGC	ACGTTGGATGGTGGTCTAATCTTTTTCCTGC	cGTGGTATAATTAAAATGCCACCAAA
	opn_1331	ACGTTGGATGGACACTGTTTTTCCTTGTTC	ACGTTGGATGGTCTCACAATTATGTATCTG	tCTAATTTACCAAATTTCAATGATCCA
2 [b]	opn_1195	ACGTTGGATGAGTCTGTAGCAAACTGCAGG	ACGTTGGATGAGCACTGCATCAGCATCACA	TTACCTTGGTCTGCGG
	slc2a9_1209	ACGTTGGATGTCAGTTCATGTCACCATCCC	ACGTTGGATGGACAAGGTTTTGGAGGAAGG	ACGAAGGGCAGTCAGA
	pgm2_874	ACGTTGGATGCCCAATCCTCAGGATATGTG	ACGTTGGATGAGCGCTGACTCCATCTTTG	cTATATGTGCTGCCCTTT
	abcg2_1294	ACGTTGGATGGCCTGGAAATCCATTTGAGC	ACGTTGGATGGGATCAGAGAAGGAAGAAGC	GCATAAGTGACCTCCTTG
	pgm2_877	ACGTTGGATGGTAAAAAACAACTTACGTTGC	ACGTTGGATGACTAGTCGGGAGAATCTAGG	ACAACTTACGTTGCTACTTA
	pgm2_843	ACGTTGGATGTCTCAGCTACCGCAATTAGG	ACGTTGGATGAGTTGGTACACCAATTTGTC	gCTGAATTCACCTCAAACCAT
	pgm2_851	ACGTTGGATGGACAGAGGTTCAGTGAGATG	ACGTTGGATGAGATGTTTCTGGAGAGACCG	AGATGATCATTTGCAAGATAA
	csn1s1_1176	ACGTTGGATGCTCAGAACTGAATTTGGGAA	ACGTTGGATGCTTTGACTAAATTCTACATC	aGAATCATTTGGGAGGCTCTA
	csn1s1_1166	ACGTTGGATGGACCTAACTATTTTATTTTC	ACGTTGGATGAAGAGCAACAGCCACAAGAC	TTCTTACATAGATCTTGACAAC
	csn1s1_1174	ACGTTGGATGTGGGTGGCAATAGTGAAGTG	ACGTTGGATGTCAGTTAGCCTGGTAGGTAG	cAGTGAAGTGGTAAGGAAAATT
	ppargc1a_1238	ACGTTGGATGACACTGGGAGATGAGTAATC	ACGTTGGATGGTGTTATGGCATAAATGTTCC	CTGAATATGTTTAACCAAATGAG

	abcg2_1293	ACGTTGGATGATGTCCAGGAAGAAGTCTGC	ACGTTGGATGTGTGTATCTGAGTTACCTTC	CAGGGTTATTATAGGGCTCACAG
	csn1s1_1175	ACGTTGGATGATTTTATTGGCCCTGAATC	ACGTTGGATGCACACTTTCATACTGGGTTG	cTTTATTGGCCCTGAATCTTTTTA
	slc2a9_1230	ACGTTGGATGTGGACACCTACTGTTTCCTG	ACGTTGGATGTGTTTTTGGTCTCAGGCAGG	tcCTGTTTCCTGGTTTTTGCTGCA
	csn1s1_1188	ACGTTGGATGATGGAAAGGTGAGGATAGTG	ACGTTGGATGTCTTCACTCATGGGCCAAAG	ggGAGGATAGTGTACTCTTTAGAA
	abcg2_1292	ACGTTGGATGACGGGCATGTACTCAAACAG	ACGTTGGATGCATCTTCGGTTAGTTTGGTG	cAAAGAGACATCAATAACAAACTGG
	slc2a9_1232	ACGTTGGATGAGAGTCCCCTGCAAGTTTAG	ACGTTGGATGAGAAATCCCAAATGGGTTGC	cccATATTTAAAAATAAGCCTTTGCT
	ppargc1a_1255	ACGTTGGATGTAGCAGGTTTATTTTCTTC	ACGTTGGATGCCTACAAATTATACTCAGAGA	TCTTCTTTTGTTTTAAAGAAAGTATA
	opn_1197	ACGTTGGATGGAAGAAAAACATGAAGAGAG	ACGTTGGATGCACATTTGAAATGCATGTTGG	cGATAAAAATAAGAAAGAGAAACTTG
	abcg2_1299	ACGTTGGATGTAAGACCAGGAGATTAAACC	ACGTTGGATGTTGCAGGCGGATTCTTTACC	cccTTAAACCATTTGCTATACAGATGT
3 [c]	slc2a9_1215	ACGTTGGATGATACCAGCCGGATGTTCCTC	ACGTTGGATGTAGCCTTCCGGACCTTCTTG	TCCGCCAGGACCTCCT
	abcg2_1310	ACGTTGGATGCTAAAGCTTGCGAAGTGAGG	ACGTTGGATGTTACCGTCACGTCGGAGAGA	TGAAGTCGGTGCTGCC
	slc2a9_1199	ACGTTGGATGGAAACAGGCCAGAGAGGAAC	ACGTTGGATGATAAGGAACCCAGGGGAGGA	AGAGAGGAACAGAGGC
	opn_1193	ACGTTGGATGTGACAACCAGGATCTCCCAG	ACGTTGGATGAATGCCCCATGACACATCTC	gGCTCTCCACCTACACAG
	abcg2_1279	ACGTTGGATGGGTAAGGGAGCTTTACTTTC	ACGTTGGATGATATTGTAAATCTTACCTTC	ACTTTCATATTTTGGGGC
	slc2a9_1226	ACGTTGGATGATGATGGCCAGAATGCACAG	ACGTTGGATGATGACTGCTTCCTCATCTGC	gGGTAGGGGATCCAGGGC
	slc2a9_1210	ACGTTGGATGTCGTGCTTCTCAAGTAGCAG	ACGTTGGATGTGATTGTCATCCCTGCCTTG	TCTCAAGTAGCAGGTAGTG
	slc2a9_1223	ACGTTGGATGAACCCAGGGTTATTCCTGAG	ACGTTGGATGAAAGGCATGCTGGTTTGTCC	AGAACATCATGGGGTTTAA
	opn_1194	ACGTTGGATGAGCACTGCATCAGCATCACA	ACGTTGGATGAGTCTGTAGCAAACTGCAGG	ccctGGGACTGGACTCTTCT
	pgm2_1379	ACGTTGGATGAGAGGTCAGATAACGAGAGC	ACGTTGGATGTTGGAAATAACCAGGCCAGC	ggggAACGAGAGCATCAACG
	pgm2_878	ACGTTGGATGTGGCGTCAGTGAAACAACTC	ACGTTGGATGGTCCCAAATTCCATTCGAGC	gccaTCATTGCAGCAGGTAAT
	pparg1a_1239	ACGTTGGATGGTTGCTGAGTAGGAACATTG	ACGTTGGATGGCAGAACAAGGAAAGAAAA	gAGGAACATTGAAATGTAGGA
	ppargc1a_1242	ACGTTGGATGCTAAGAGTATAAACAACATTG	ACGTTGGATGGCCCAAGTCTGACAAACAGT	GTATAAACAACATTGAACAAGAA
	ppargc1a_1247	ACGTTGGATGGAAGGATGGATCAGGATTAAG	ACGTTGGATGCCTTTTTCTTGCTTAAGTAT	AGAATATGATCATTTAAATCAGGA
	pgm2_875	ACGTTGGATGTTAGTTGTAGGTAAATGTA	ACGTTGGATGGCCACAAAAGTATAACCTCC	TTAGTTGTAGGTAAATGTAATTATG
	opn_1198	ACGTTGGATGATGGAAAATAGAGGTGGCAG	ACGTTGGATGGTGTGGAAACACTCCACTAA	GCAGTACAAATATTAACTCAAAAGAT
	csn1s1_1165	ACGTTGGATGTGAAGAGCAAAATTAAAAAC	ACGTTGGATGCCAACAAAGGAGTATATTATG	AATGAATAAAAATATTACTTTCAAGTG

[a] Well contains in addition primers of 9 SNPs that are not subject of this thesis.
[b] Well contains in addition primers of 5 SNPs that are not subject of this thesis.
[c] Well contains in addition primers of 1 SNP that is not subject of this thesis.

11.3. *Haplotypes*

Table 11.3 Constructed haplotypes

Haplotypes were constructed separately for each breed and separately for each gene. Only haplotypes with a frequency > 0.025 are included. Numbers in the third column indicate in vertical orientation the SNP by the in-house identification code SNP_id.

Gene	Haplotype name	Haplotype	Frequency		
ABCG2		SNP	BV	FV [b]	HF
		1 1 1 1 1 1 1 1 1 3 3 2 2 2 2 2 2 3 0 1 7 8 9 9 9 9 0 3 0 9 7 2 3 4 9 2			
	ABC_HT1	G T A C G G A T G	0.45		[a]
	ABC_HT2	A T A T G G A T A	0.49		0.07
	ABC_HT3	A A A T A G A T A	0.03		0.07
	ABC_HT4	A T T T A G A G A	0.03		[a]
	ABC_HT5	A A A T A G A T G	0.05		[a]
	ABC_HT6	G T T T G G A T A	[a]		0.22
	ABC_HT7	G T T T A G A T A	[a]		0.09
	ABC_HT8	G T T T G A G G A	[a]		0.14
	ABC_HT9	G T T T G A A T A	[a]		0.09
	ABC_HT10	A T A T A G A T A	[a]		0.29
OPN		SNP			
		1 1 1 1 1 3 1 1 1 1 3 9 9 9 9 1 7 5 4 3			
	OP_HT1	T G A C C	0.49		0.04
	OP_HT2	T A G C T	0.44		0.27
	OP_HT3	T G A T C	0.07		0.04
	OP_HT4	C G G C C	[a]		0.48
	OP_HT5	T G G C C	[a]		0.17
ABCG2/OPN		SNP			
		1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 2 2 2 2 2 2 3 3 1 1 1 1 0 1 7 8 9 9 9 9 0 3 9 9 9 9 3 0 9 7 2 3 4 9 2 1 7 5 4 3			
	ABOP_HT1	A T A T A G A T A T G A C C		0.03	
	ABOP_HT2	G T A C G G A T G T G A T C		0.16	
	ABOP_HT3	A T A T G G A T A T G A C C		0.21	
	ABOP_HT4	A A A T A G A T A T A G C T		0.35	
	ABOP_HT5	G T T T A G A G A C G G C C		0.08	
	ABOP_HT6	A T T T G G A T G T G A T C		0.03	
	ABOP_HT7	G T T T G G A T A C G G C C		0.03	

PPARGC1A	SNP			
	1 1 1 1 1 1 2 2 2 2 2 2 4 4 4 4 3 3 7 6 4 2 9 8			
PPARG_HT1	A T T A G G	0.06	0.12	0.19
PPARG_HT2	A C C A G G	0.38	0.03	0.04
PPARG_HT3	A C T G G A	0.19	0.19	0.38
PPARG_HT4	C C C G G A	0.30	0.28	0.10
PPARG_HT5	A C T A G G	0.03	0.03	0.07
PPARG_HT6	C C C G G G	[a]	0.28	[a]
PPARG_HT7	A C T A T G	[a]	[a]	0.07
PPARG_HT8	A C T A G A	[a]	[a]	0.11

PGM2	SNP			
	1 1 3 8 8 8 8 3 8 8 8 8 2 7 3 4 5 7 5 7 7 7 1 8 7 3 8 9 1 4 7 5			
PGM2_HT1	G A T G C G G T C G	0.61	0.43	0.15
PGM2_HT2	A A C A T C A C C A	0.20	0.06	0.27
PGM2_HT3	G G T G C G G T T G	0.15	0.26	0.07
PGM2_HT4	G A T G C G G T T G	[a]	0.20	0.31
PGM2_HT5	G A C A T C A C C A	[a]	[a]	0.12
PGM2_HT6	A A T G C G G T T G	[a]	[a]	0.06

CSN1S1	SNP [c]			
	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 9 8 7 7 7 7 6 6 6 1 8 6 5 4 0 6 5 1			
CSN1S_HT1	A A T T A T C C T	0.08	[a]	0.10
CSN1S_HT2	A A C T A D C C T	0.64	0.41	0.53
CSN1S_HT3	A A T T T T C C T	0.19	0.15	0.29
CSN1S_HT4	G G T C A T T A C	0.05	0.06	[a]
CSN1S_HT5	A A C T A T C C T	0.03	0.33	[a]
CSN1S_HT6	A A T C A T T C C	[a]	[a]	0.04

SLC2A9	SNP													
	1 2 3 2	1 2 2 6	1 2 2 5	1 2 2 3	1 2 1 5	1 2 1 3	1 2 1 0	1 2 0 9	1 2 0 8	1 2 0 6	1 1 1 9			
SLC2_HT1	C	C	A	G	G	C	G	G	A	C	T	0.18	[a]	[a]
SLC2_HT2	C	C	A	A	C	G	G	G	G	C	G	0.22	0.04	0.04
SLC2_HT3	C	C	A	A	G	G	A	G	G	C	T	0.06	0.03	[a]
SLC2_HT4	A	C	A	A	C	G	G	G	G	C	T	0.29	0.07	0.19
SLC2_HT5	A	T	A	A	C	G	G	G	A	C	T	0.03	0.06	0.04
SLC2_HT6	C	C	A	A	C	G	G	G	G	C	T	0.03	0.04	0.09
SLC2_HT7	A	C	G	A	G	G	G	G	A	T	T	0.04	[a]	[a]
SLC2_HT8	A	C	G	A	G	G	G	G	A	C	T	0.05	0.18	0.03
SLC2_HT9	A	C	A	A	C	G	G	G	G	C	G	[a]	0.05	0.19
SLC2_HT10	A	C	A	A	G	G	A	G	G	C	T	[a]	0.06	[a]
SLC2_HT11	A	C	A	A	C	G	G	G	A	C	T	[a]	0.06	[a]
SLC2_HT12	C	C	A	A	C	G	G	G	A	C	T	[a]	0.08	[a]
SLC2_HT13	A	C	G	A	C	G	G	G	G	C	T	[a]	0.06	[a]
SLC2_HT14	C	C	A	A	G	C	G	G	G	C	T	[a]	0.04	[a]
SLC2_HT15	C	C	A	A	G	C	G	G	A	C	G	[a]	0.09	[a]
SLC2_HT16	A	C	A	A	G	G	G	G	G	C	T	[a]	0.03	[a]
SLC2_HT17	C	C	G	A	G	G	A	G	G	C	T	[a]	[a]	0.19
SLC2_HT18	A	C	G	A	G	G	G	G	G	C	T	[a]	[a]	0.08

[a] Frequency < 0.025

[b] FV-haplotypes were constructed across ABCG2 and OPN

[c] 'D' represents the deletion variant of INDEL1170

11.4. *Substitution effects*

Table 11.4 Regression coefficients on the number of copies of allele '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of significantly associated (*FDR* < 10%) *CSN1S1*-polymorphisms in BV

Gene	CSN1S1							
SNP	1191	1188	1176	1175	1170	1166	1165	1161
Allele '*x*'	G	G	T	T	T	T	C	T
Trait								
MY1 (kg)	-65.27 ± 33.46	-68.24 ± 37.60	-32.00 ± 15.47	68.24 ± 37.59	-38.37 ± 15.40	-59.17 ± 37.17	54.35 ± 42.20	57.53 ± 37.28
MY2 (kg)	-140.48* ± 46.19	-143.22 ± 50.95	-47.65 ± 21.76	143.212 ± 50.95	-66.23* ± 21.99	-129.61 ± 50.40	127.32 ± 58.80	127.47 ± 50.64
MY3 (kg)	-123.20 ± 47.54	-131.58 ± 52.86	-11.06 ± 25.30	131.58 ± 52.86	-22.65 ± 25.31	-114.26 ± 52.17	130.72 ± 58.95	111.70 ± 52.51
FY1 (kg)	0.11 ± 1.31	0.55 ± 1.48	0.20 ± 0.61	-0.54 ± 1.48	0.46 ± 0.61	0.53 ± 1.46	0.16 ± 1.72	-0.56 ± 1.46
FY2 (kg)	-1.81 ± 1.81	-1.56 ± 2.00	-0.49 ± 0.85	1.56 ± 2.00	-0.55 ± 0.86	-1.42 ± 1.98	1.97 ± 2.35	1.38 ± 1.97
FY3 (kg)	-1.13 ± 1.85	-1.44 ± 2.07	1.36 ± 0.97	1.44 ± 2.07	1.54 ± 0.97	-1.25 ± 2.04	1.87 ± 2.31	1.22 ± 2.03
PY1 (kg)	-0.38 ± 1.00	0.34 ± 1.13	-0.99 ± 0.46	-0.34 ± 1.13	-0.95 ± 0.46	0.58 ± 1.11	0.11 ± 1.28	-0.61 ± 1.11
PY2 (kg)	-1.58 ± 1.46	-0.60 ± 1.61	-1.27 ± 0.68	0.60 ± 1.61	-1.44 ± 0.68	-0.24 ± 1.59	1.06 ± 1.85	0.20 ± 1.59
PY3 (kg)	-1.86 ± 1.56	-1.45 ± 1.75	0.09 ± 0.83	1.46 ± 1.75	0.55 ± 0.82	-0.96 ± 1.73	2.03 ± 1.95	0.91 ± 1.724
FC1 (%)	0.049 ± 0.018	0.058* ± 0.020	0.027* ± 0.008	-0.058* ± 0.020	0.036* ± 0.008	0.051 ± 0.020	-0.038 ± 0.022	-0.050 ± 0.020
FC2 (%)	0.060* ± 0.019	0.064* ± 0.021	0.023 ± 0.009	-0.064* ± 0.021	0.034* ± 0.009	0.057 ± 0.021	-0.049* ± 0.024	-0.057 ± 0.021
FC3 (%)	0.053 ± 0.021	0.054 ± 0.023	0.023 ± 0.011	-0.054 ± 0.022	0.033* ± 0.011	0.046 ± 0.023	-0.048 ± 0.025	-0.045 ± 0.023
PC1 (%)	0.033* ± 0.011	0.047* ± 0.013	0.004 ± 0.005	-0.047* ± 0.013	0.008 ± 0.006	0.045* ± 0.012	-0.032 ± 0.014	-0.044* ± 0.012
PC2 (%)	0.047* ± 0.013	0.063* ± 0.014	0.006 ± 0.006	-0.063* ± 0.014	0.014 ± 0.006	0.061* ± 0.014	-0.049 ± 0.016	-0.061* ± 0.014
PC3 (%)	0.032 ± 0.014	0.042 ± 0.016	0.006 ± 0.008	-0.042 ± 0.016	0.017 ± 0.008	0.041 ± 0.016	-0.0343 ± 0.0177	-0.040 ± 0.0157

* FDR < 10%

Table 11.5 Regression coefficients on the number of copies of allele '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of significantly associated (*FDR* < 10%) *PPARGC1A*- and *SLC2A9*-polymorphisms in BV

Gene	PPARGC1A		SLC2A9					
Marker	1242	1238	1223	1215	1213	1208	1206	1199
Allele '*x*'	G	G	G	T	G	G	T	T
Trait								
MY1 (kg)	-23.31 ± 14.20	29,12 ± 13,82	37,80 ± 18,13	39,39 ± 15,32	-40,30 ± 17,81	-38,60 ± 15,07	132,20* ± 32,13	32,65 ± 19,21
MY2 (kg)	-60,19* ± 19,37	63,00* ± 18,80	72,07* ± 24,70	64,13* ± 21,11	-67,34 ± 24,21	-39,15 ± 20,82	97,67 ± 45,24	76,67* ± 25,75
MY3 (kg)	-38,68 ± 22,69	38,73 ± 21,53	59,27 ± 26,51	63,58 ± 22,50	-55,64 ± 26,23	-30,43 ± 22,80	91,55 ± 54,69	40,63 ± 30,17
FY1 (kg)	-1,01 ± 0,54	1,15 ± 0,53	2,17* ± 0,69	2,06* ± 0,59	-2,33* ± 0,68	-2,10* ± 0,58	2,58 ± 1,26	0,83 ± 0,75
FY2 (kg)	-2,26* ± 0,75	2,33* ± 0,73	3,91* ± 0,94	3,23* ± 0,80	-3,88* ± 0,92	-2,16 ± 0,80	-0,37 ± 1,75	2,34 ± 1,00
FY3 (kg)	-1,23 ± 0,86	1,01 ± 0,83	2,81 ± 1,00	2,83* ± 0,85	-2,68 ± 0,99	-1,79 ± 0,86	1,61 ± 2,09	1,84 ± 1,15
PY1 (kg)	-0,55 ± 0,42	0,77 ± 0,41	1,19 ± 0,53	1,59* ± 0,45	-1,39 ± 0,52	-1,25 ± 0,44	2,25 ± 0,96	0,35 ± 0,57
PY2 (kg)	-1,40 ± 0,60	1,65 ± 0,59	2,68* ± 0,75	2,57* ± 0,64	-2,63* ± 0,74	-1,44 ± 0,64	0,71 ± 1,41	1,73 ± 0,80
PY3 (kg)	-0,79 ± 0,74	0,94 ± 0,70	2,21 ± 0,85	2,71* ± 0,72	-2,31 ± 0,84	-1,27 ± 0,73	0,22 ± 1,78	0,90 ± 0,98
FC1 (%)	-0,001 ± 0,008	-0,002 ± 0,008	0,007 ± 0,010	0,004 ± 0,008	-0,008 ± 0,010	-0,007 ± 0,008	-0,045 ± 0,018	-0,010 ± 0,010
FC2 (%)	0,004 ± 0,008	-0,005 ± 0,008	0,012 ± 0,011	0,008 ± 0,009	-0,013 ± 0,011	-0,007 ± 0,009	-0,060* ± 0,019	-0,012 ± 0,011
FC3 (%)	0,004 ± 0,010	-0,006 ± 0,010	0,002 ± 0,012	0,007 ± 0,010	-0,002 ± 0,012	-0,005 ± 0,010	-0,030 ± 0,024	-0,003 ± 0,013
PC1 (%)	0,005 ± 0,005	-0,005 ± 0,005	-0,004 ± 0,006	0,001 ± 0,005	0,003 ± 0,006	0,003 ± 0,005	-0,036* ± 0,011	-0,013 ± 0,007
PC2 (%)	0,010 ± 0,006	-0,008 ± 0,005	0,001 ± 0,007	0,004 ± 0,006	-0,002 ± 0,007	0,002 ± 0,006	-0,038* ± 0,013	-0,015 ± 0,007
PC3 (%)	0,007 ± 0,007	-0,005 ± 0,007	0,001 ± 0,008	0,006 ± 0,007	-0,003 ± 0,008	-0,002 ± 0,007	-0,040 ± 0,017	-0,009 ± 0,009

* FDR < 10%

Table 11.6 Regression coefficients on the number of copies of allele '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of significantly associated (*FDR* < 10%) *CSN1S1*-polymorphisms in FV

Gene	CSN1S1							
Marker	1191	1188	1176	1175	1170	1166	1165	1161
Allele '*x*'	G	G	T	T	T	T	C	T
Trait								
MY1 (kg)	-17,62 ± 23,87	-18,05 ± 23,94	-25,12 ± 14,91	20,38 ± 24,13	-37,45* ± 12,83	-22,41 ± 22,52	18,07 ± 22,97	19,93 ± 22,34
MY2 (kg)	-17,02 ± 29,18	-15,25 ± 29,28	-4,66 ± 18,24	15,68 ± 29,58	-28,67 ± 15,76	-21,34 ± 27,16	18,83 ± 27,64	22,20 ± 26,86
MY3 (kg)	-54,41 ± 39,73	-48,97 ± 40,09	-26,37 ± 24,73	53,65 ± 40,87	-50,80 ± 21,69	-41,21 ± 39,25	47,73 ± 39,47	41,51 ± 38,34
FY1 (kg)	0,37 ± 0,85	0,32 ± 0,86	-0,43 ± 0,53	-0,16 ± 0,86	0,41 ± 0,47	0,33 ± 0,80	-0,34 ± 0,82	-0,51 ± 0,80
FY2 (kg)	0,11 ± 1,18	0,17 ± 1,18	0,35 ± 0,74	-0,03 ± 1,19	0,69 ± 0,65	0,32 ± 1,10	-0,28 ± 1,12	-0,38 ± 1,09
FY3 (kg)	-1,32 ± 1,66	-1,27 ± 1,67	0,29 ± 1,03	1,72 ± 1,70	0,53 ± 0,91	-0,63 ± 1,63	1,00 ± 1,64	0,29 ± 1,59
PY1 (kg)	1,17 ± 0,69	1,11 ± 0,69	0,09 ± 0,43	-1,02 ± 0,69	0,14 ± 0,37	0,87 ± 0,64	-0,96 ± 0,66	-1,01 ± 0,64
PY2 (kg)	1,46 ± 0,94	1,46 ± 0,94	1,07 ± 0,59	-1,42 ± 0,95	0,62 ± 0,51	1,28 ± 0,87	-1,27 ± 0,89	-1,32 ± 0,86
PY3 (kg)	0,28 ± 1,20	0,29 ± 1,20	0,29 ± 0,74	-0,11 ± 1,23	0,31 ± 0,66	0,61 ± 1,18	-0,39 ± 1,18	-0,75 ± 1,15
FC1 (%)	0,020 ± 0,013	0,019 ± 0,013	0,012 ± 0,008	-0,018 ± 0,013	0,033* ± 0,007	0,022 ± 0,012	-0,019 ± 0,012	-0,023 ± 0,012
FC2 (%)	0,014 ± 0,014	0,014 ± 0,014	0,009 ± 0,009	-0,012 ± 0,014	0,028* ± 0,007	0,019 ± 0,013	-0,017 ± 0,013	-0,020 ± 0,013
FC3 (%)	0,018 ± 0,019	0,015 ± 0,019	0,020 ± 0,012	-0,012 ± 0,020	0,039* ± 0,010	0,019 ± 0,019	-0,017 ± 0,019	-0,023 ± 0,019
PC1 (%)	0,031* ± 0,007	0,030* ± 0,007	0,019* ± 0,004	-0,030* ± 0,007	0,025* ± 0,004	0,029* ± 0,007	-0,027* ± 0,007	-0,030* ± 0,007
PC2 (%)	0,030* ± 0,008	0,029* ± 0,008	0,019* ± 0,005	-0,029* ± 0,008	0,025* ± 0,004	0,029* ± 0,007	-0,028* ± 0,007	-0,030* ± 0,007
PC3 (%)	0,032* ± 0,011	0,029 ± 0,011	0,017 ± 0,007	-0,029 ± 0,011	0,030* ± 0,006	0,030 ± 0,011	-0,030 ± 0,011	-0,032* ± 0,010

* FDR < 10%

Table 11.7 Regression coefficients on the number of copies of allele '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of significantly associated (*FDR* < 10%) *PPARGC1A*-, *PGM2*- and *SLC2A9*-polymoprhisms in FV

Gene	PPARGC1A	PGM2
Marker	1238	877
Allele '*x*'	G	T
Trait		
MY1 (kg)	33,23 ± 13,26	13,94 ± 12,41
MY2 (kg)	31,96 ± 16,26	14,56 ± 15,05
MY3 (kg)	38,68 ± 21,88	9,14 ± 21,12
FY1 (kg)	0,05 ± 0,48	-0,02 ± 0,45
FY2 (kg)	0,06 ± 0,66	0,31 ± 0,61
FY3 (kg)	0,07 ± 0,91	-1,84 ± 0,87
PY1 (kg)	0,76 ± 0,38	0,18 ± 0,36
PY2 (kg)	0,72 ± 0,52	0,42 ± 0,48
PY3 (kg)	0,72 ± 0,66	-0,30 ± 0,63
FC1 (%)	-0,023* ± 0,007	-0,008 ± 0,007
FC2 (%)	-0,020 ± 0,008	-0,005 ± 0,007
FC3 (%)	-0,023 ± 0,011	-0,029* ± 0,010
PC1 (%)	-0,008 ± 0,004	-0,004 ± 0,004
PC2 (%)	-0,007 ± 0,004	-0,002 ± 0,004
PC3 (%)	-0,011 ± 0,006	-0,008 ± 0,006

* FDR < 10%

Table 11.8 Regression coefficients on the number of copies of allele 'x' representing half of the allele substitution effects (α/2) and standard error for milk production traits of significantly associated (*FDR* < 10%) *ABCG2*-, *OPN*-, *PPARGC1A*- and *CSN1S1*-polymoprhisms in HF

Gene	ABCG2		OPN	PPARGC1A	CSN1S1		
SNP	1293	1294	1331	1239	1176	1174	1170
Allele 'x'	G	G	T	T	T	T	T
Trait							
MY1 (kg)	46,33 ± 24,15	-54,19 ± 27,87	-27,81 ± 22,50	58,39 ± 42,78	-6,22 ± 20,59	-55,51 ± 22,74	-14,64 ± 20,38
MY2 (kg)	66,86 ± 26,54	-77,83 ± 30,58	-45,17 ± 24,63	83,46 ± 46,85	-10,04 ± 22,64	-50,87 ± 25,03	-17,43 ± 22,33
MY3 (kg)	63,94 ± 28,67	-65,18 ± 33,18	-61,72 ± 26,83	124,16 ± 50,96	1,05 ± 24,66	-42,11 ± 27,87	-13,14 ± 24,43
FY1 (kg)	2,00 ± 0,91	-2,09 ± 1,06	-1,67 ± 0,85	2,95 ± 1,62	0,22 ± 0,78	-1,62 ± 0,86	-0,10 ± 0,77
FY2 (kg)	2,98 ± 1,06	-3,07 ± 1,22	-2,53 ± 0,98	4,71 ± 1,87	0,34 ± 0,91	-1,03 ± 1,00	0,08 ± 0,90
FY3 (kg)	3,56* ± 1,21	-3,78 ± 1,40	-3,82* ± 1,13	7,17* ± 2,14	0,64 ± 1,04	-0,19 ± 1,18	0,40 ± 1,05
PY1 (kg)	1,65 ± 0,69	-1,93 ± 0,80	-1,58 ± 0,64	2,83 ± 1,22	0,62 ± 0,59	-0,48 ± 0,65	0,49 ± 0,58
PY2 (kg)	2,58* ± 0,82	-2,91* ± 0,94	-2,41* ± 0,76	4,38* ± 1,44	0,90 ± 0,70	-0,04 ± 0,78	0,86 ± 0,69
PY3 (kg)	2,44 ± 0,92	-2,54 ± 1,06	-3,28* ± 0,86	6,39* ± 1,63	1,04 ± 0,79	0,24 ± 0,90	0,89 ± 0,79
FC1 (%)	0,002 ± 0,011	0,001 ± 0,012	-0,007 ± 0,010	0,007 ± 0,019	0,008 ± 0,009	0,009 ± 0,010	0,007 ± 0,009
FC2 (%)	0,005 ± 0,011	0,001 ± 0,012	-0,008 ± 0,010	0,016 ± 0,019	0,010 ± 0,009	0,013 ± 0,010	0,009 ± 0,009
FC3 (%)	0,011 ± 0,011	-0,012 ± 0,013	-0,015 ± 0,011	0,026 ± 0,021	0,009 ± 0,010	0,019 ± 0,011	0,012 ± 0,010
PC1 (%)	0,002 ± 0,005	-0,002 ± 0,006	-0,009 ± 0,005	0,011 ± 0,009	0,011 ± 0,004	0,018* ± 0,005	0,013* ± 0,004
PC2 (%)	0,004 ± 0,005	-0,004 ± 0,006	-0,010 ± 0,005	0,018 ± 0,009	0,014* ± 0,004	0,019* ± 0,005	0,017* ± 0,004
PC3 (%)	0,004 ± 0,006	-0,005 ± 0,007	-0,014 ± 0,005	0,026 ± 0,010	0,012 ± 0,005	0,018* ± 0,006	0,015* ± 0,005

* FDR < 10%

Table 11.9 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *ABCG2*-haplotypes in BV

Haplotype name	ABC_HT1	ABC_HT2	ABC_HT3	ABC_HT4	ABC_HT5
Haplotype '*x*'	GTACGGATG	ATATGGATA	AAATAGATA	ATTTAGAGA	AAATAGATG
Trait					
MY1 (kg)	9,80 ± 14,34	15,73 ± 15,56	-49,34 ± 31,92	-8,78 ± 39,51	5,82 ± 35,21
MY2 (kg)	15,47 ± 20,20	41,04 ± 22,05	-106,41 ± 45,59	-15,19 ± 52,99	-33,44 ± 50,61
MY3 (kg)	40,85 ± 21,85	15,91 ± 24,05	-154,61* ± 55,03	-48,65 ± 55,31	12,62 ± 52,80
FY1 (kg)	-0,08 ± 0,55	0,93 ± 0,59	-2,01 ± 1,22	2,38 ± 1,49	-1,07 ± 1,35
FY2 (kg)	0,13 ± 0,77	1,65 ± 0,84	-3,81 ± 1,75	1,87 ± 2,02	-1,59 ± 1,94
FY3 (kg)	1,03 ± 0,83	0,78 ± 0,91	-5,33 ± 2,09	1,49 ± 2,10	-0,35 ± 2,00
PY1 (kg)	-0,22 ± 0,42	0,73 ± 0,45	-1,37 ± 0,93	0,88 ± 1,14	0,33 ± 1,03
PY2 (kg)	0,27 ± 0,63	1,25 ± 0,68	-2,51 ± 1,42	0,92 ± 1,64	-0,09 ± 1,57
PY3 (kg)	0,25 ± 0,71	0,68 ± 0,78	-3,62 ± 1,79	-0,04 ± 1,79	1,19 ± 1,70
FC1 (%)	-0,009 ± 0,008	0,004 ± 0,009	0,000 ± 0,017	0,045 ± 0,021	-0,020 ± 0,019
FC2 (%)	-0,009 ± 0,009	0,000 ± 0,010	0,011 ± 0,020	0,035 ± 0,022	-0,005 ± 0,022
FC3 (%)	-0,010 ± 0,010	0,002 ± 0,011	0,016 ± 0,024	0,043 ± 0,024	-0,013 ± 0,023
PC1 (%)	-0,009 ± 0,005	0,002 ± 0,005	0,006 ± 0,011	0,019 ± 0,013	0,004 ± 0,012
PC2 (%)	-0,006 ± 0,006	-0,003 ± 0,006	0,020 ± 0,013	0,020 ± 0,015	0,015 ± 0,014
PC3 (%)	-0,016 ± 0,007	0,002 ± 0,007	0,023 ± 0,017	0,020 ± 0,017	0,008 ± 0,016

* FDR < 10%

Table 11.10 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *PPARGC1A*-haplotypes in BV

Haplotype name	PPARG_HT1	PPARG_HT2	PPARG_HT3	PPARG_HT4	PPARG_HT5
Haplotype '*x*'	ATTAGG	ACCAGG	ACTGGA	CCCGGA	ACTAGG
Trait					
MY1 (kg)	2,33 ± 30,95	26,38 ± 15,03	-21,03 ± 17,71	-26,78 ± 15,77	4,18 ± 40,34
MY2 (kg)	18,38 ± 41,35	59,97* ± 20,45	-49,04 ± 24,47	-41,58 ± 22,19	34,67 ± 56,78
MY3 (kg)	23,04 ± 41,29	33,61 ± 23,32	-28,59 ± 26,52	-21,20 ± 24,95	25,36 ± 78,78
FY1 (kg)	0,81 ± 1,17	1,12 ± 0,57	-0,88 ± 0,68	-0,93 ± 0,61	-0,25 ± 1,52
FY2 (kg)	0,79 ± 1,58	2,23* ± 0,78	-2,05 ± 0,94	-1,31 ± 0,85	2,15 ± 2,17
FY3 (kg)	1,79 ± 1,56	1,03 ± 0,88	-0,73 ± 1,01	-0,57 ± 0,94	-0,20 ± 2,97
PY1 (kg)	0,36 ± 0,89	0,66 ± 0,44	-0,64 ± 0,52	-0,41 ± 0,46	-0,31 ± 1,16
PY2 (kg)	0,73 ± 1,28	1,44 ± 0,64	-1,20 ± 0,76	-0,69 ± 0,69	0,88 ± 1,75
PY3 (kg)	0,51 ± 1,33	0,60 ± 0,76	-0,80 ± 0,86	-0,36 ± 0,81	1,22 ± 2,54
FC1 (%)	0,013 ± 0,017	0,001 ± 0,008	-0,001 ± 0,010	0,005 ± 0,009	-0,006 ± 0,022
FC2 (%)	0,001 ± 0,018	-0,004 ± 0,009	0,001 ± 0,011	0,006 ± 0,010	0,010 ± 0,025
FC3 (%)	0,008 ± 0,018	-0,003 ± 0,010	0,005 ± 0,012	0,002 ± 0,011	-0,019 ± 0,035
PC1 (%)	0,006 ± 0,010	-0,005 ± 0,005	0,001 ± 0,006	0,011 ± 0,005	-0,006 ± 0,014
PC2 (%)	0,001 ± 0,012	-0,009 ± 0,006	0,007 ± 0,007	0,011 ± 0,006	-0,003 ± 0,016
PC3 (%)	-0,005 ± 0,013	-0,006 ± 0,007	0,002 ± 0,008	0,005 ± 0,008	0,002 ± 0,025

* FDR < 10%

Table 11.11 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *CSN1S1*-haplotypes in BV

Haplotype name	CSN1S_HT1	CSN1S_HT2	CSN1S_HT3	CSN1S_HT4	CSN1S_HT5
Haplotype '*x*'	AATTATCCT	AACTADCCT	AATTTTCCT	GGTCATTAC	AACTATCCT
Trait					
MY1 (kg)	-41,39 ± 25,76	41,43* ± 15,06	-5,88 ± 17,93	-42,01 ± 37,39	-65,62 ± 34,97
MY2 (kg)	-32,33 ± 36,08	60,37* ± 21,06	-5,05 ± 24,73	-134,63* ± 50,59	-108,08 ± 48,13
MY3 (kg)	-21,58 ± 40,16	19,66 ± 23,93	38,30 ± 27,41	-120,99 ± 54,26	-81,68 ± 51,44
FY1 (kg)	0,64 ± 0,98	-0,56 ± 0,59	0,15 ± 0,68	1,19 ± 1,44	0,75 ± 1,37
FY2 (kg)	0,21 ± 1,38	-0,01 ± 0,82	0,21 ± 0,95	-1,75 ± 1,96	0,10 ± 1,86
FY3 (kg)	0,67 ± 1,52	-1,91 ± 0,90	2,17 ± 1,03	-1,34 ± 2,08	0,66 ± 1,96
PY1 (kg)	-1,47 ± 0,74	0,94 ± 0,45	-0,46 ± 0,52	0,58 ± 1,10	-0,37 ± 1,04
PY2 (kg)	-1,59 ± 1,11	1,12 ± 0,66	-0,21 ± 0,76	-1,00 ± 1,58	-1,13 ± 1,50
PY3 (kg)	-0,78 ± 1,30	-0,58 ± 0,77	1,20 ± 0,88	-1,54 ± 1,77	0,98 ± 1,67
FC1 (%)	0,040* ± 0,014	-0,040* ± 0,008	0,007 ± 0,010	0,050 ± 0,020	0,058* ± 0,019
FC2 (%)	0,024 ± 0,015	-0,038* ± 0,009	0,006 ± 0,011	0,056* ± 0,021	0,069* ± 0,020
FC3 (%)	0,023 ± 0,018	-0,036* ± 0,010	0,005 ± 0,012	0,049 ± 0,023	0,059* ± 0,022
PC1 (%)	0,001 ± 0,009	-0,010 ± 0,005	-0,003 ± 0,006	0,035* ± 0,013	0,031 ± 0,012
PC2 (%)	-0,007 ± 0,010	-0,015 ± 0,006	-0,001 ± 0,007	0,055* ± 0,014	0,041* ± 0,014
PC3 (%)	0,001 ± 0,012	-0,016 ± 0,007	-0,003 ± 0,009	0,036 ± 0,016	0,051* ± 0,016

* FDR < 10%

Table 11.12 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *SLC2A9*-haplotypes in BV

Haplotype name	SLC2_HT1	SLC2_HT2	SLC2_HT3	SLC2_HT4	SLC2_HT5	SLC2_HT6	SLC2_HT7	SLC2_HT8
Haplotype '*x*'	CCAGGCGGACT	CCAACGGGGCG	CCAAGGAGGCT	ACAACGGGGCT	ATAACGGGACT	CCAACGGGGCT	ACGAGGGGATT	ACGAGGGGACT
Trait								
MY1 (kg)	47,27 ± 20,02	-28,43 ± 19,04	14,26 ± 31,76	-18,96 ± 15,24	17,61 ± 36,09	-25,49 ± 40,35	118,83* ± 37,42	-43,33 ± 31,36
MY2 (kg)	92,72* ± 27,61	-67,05* ± 25,64	56,06 ± 43,14	-7,99 ± 21,02	-22,97 ± 49,06	-9,03 ± 54,79	83,34 ± 52,18	-35,96 ± 44,06
MY3 (kg)	72,92 ± 30,66	-42,06 ± 29,53	78,19 ± 43,43	-22,39 ± 22,34	-32,40 ± 56,17	-69,70 ± 60,49	58,16 ± 61,09	-53,97 ± 51,57
FY1 (kg)	2,37* ± 0,76	-0,91 ± 0,73	0,38 ± 1,22	-1,07 ± 0,58	1,25 ± 1,36	-1,00 ± 1,52	1,68 ± 1,45	0,50 ± 1,19
FY2 (kg)	4,29* ± 1,05	-2,30 ± 0,98	1,57 ± 1,65	-0,66 ± 0,80	-0,80 ± 1,88	-0,44 ± 2,10	-0,81 ± 2,00	1,84 ± 1,68
FY3 (kg)	2,96 ± 1,16	-2,20 ± 1,11	2,48 ± 1,65	-0,88 ± 0,85	-0,64 ± 2,13	-1,80 ± 2,29	-0,10 ± 2,31	0,44 ± 1,96
PY1 (kg)	1,33 ± 0,58	-0,30 ± 0,56	0,82 ± 0,93	-1,10 ± 0,44	0,15 ± 1,04	0,23 ± 1,16	2,14 ± 1,10	0,38 ± 0,91
PY2 (kg)	3,09* ± 0,85	-1,44 ± 0,80	1,59 ± 1,34	-0,85 ± 0,65	-0,98 ± 1,52	1,82 ± 1,69	0,63 ± 1,62	0,91 ± 1,36
PY3 (kg)	2,39 ± 0,99	-1,43 ± 0,95	2,81 ± 1,40	-1,27 ± 0,72	-2,35 ± 1,81	0,67 ± 1,96	-0,97 ± 1,97	0,74 ± 1,67
FC1 (%)	0,003 ± 0,011	0,004 ± 0,010	-0,009 ± 0,017	-0,002 ± 0,008	0,008 ± 0,019	0,004 ± 0,022	-0,049 ± 0,020	0,036 ± 0,017
FC2 (%)	0,005 ± 0,012	0,006 ± 0,011	-0,009 ± 0,018	-0,004 ± 0,009	0,001 ± 0,021	0,002 ± 0,024	-0,058* ± 0,022	0,045 ± 0,019
FC3 (%)	-0,004 ± 0,014	-0,002 ± 0,013	-0,009 ± 0,019	0,001 ± 0,010	0,007 ± 0,025	0,018 ± 0,027	-0,033 ± 0,027	0,035 ± 0,022
PC1 (%)	-0,008 ± 0,007	0,011 ± 0,007	0,001 ± 0,011	-0,005 ± 0,005	-0,008 ± 0,012	0,023 ± 0,014	-0,030 ± 0,013	0,031* ± 0,011
PC2 (%)	-0,005 ± 0,008	0,013 ± 0,007	-0,004 ± 0,012	-0,008 ± 0,006	-0,003 ± 0,014	0,033 ± 0,015	-0,032 ± 0,015	0,031 ± 0,012
PC3 (%)	-0,004 ± 0,010	0,002 ± 0,009	0,003 ± 0,014	-0,006 ± 0,007	-0,018 ± 0,017	0,044 ± 0,018	-0,040 ± 0,019	0,035 ± 0,016

* FDR < 10%

Table 11.13 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of haplotypes including *ABCG2*- and *OPN*-polymorphisms in FV

Haplotype name	ABOP_HT1	ABOP_HT2	ABOP_HT3	ABOP_HT4	ABOP_HT5	ABOP_HT6	ABOP_HT7
Haplotype '*x*'	ATATAGATATGACC	GTACGGATGTGATC	ATATGGATATGACC	AAATAGATATAGCT	GTTTAGAGACGGCC	ATTTGGATGTGATC	GTTTGGATACGGCC
Trait							
MY1 (kg)	67,64 ± 34,42	13,48 ± 16,25	-10,09 ± 14,44	-8,61 ± 12,04	-5,72 ± 21,31	-10,63 ± 36,36	-32,62 ± 35,74
MY2 (kg)	90,14 ± 42,14	-0,40 ± 19,07	-23,59 ± 17,17	0,26 ± 14,40	-5,71 ± 24,95	-2,43 ± 44,09	8,36 ± 42,98
MY3 (kg)	115,20 ± 53,88	-23,86 ± 24,85	15,85 ± 24,78	-13,98 ± 20,37	4,44 ± 34,79	-55,81 ± 52,25	-4,98 ± 80,96
FY1 (kg)	0,74 ± 1,24	-0,07 ± 0,58	-0,14 ± 0,51	-0,04 ± 0,43	-0,10 ± 0,75	-0,30 ± 1,27	-1,07 ± 1,28
FY2 (kg)	1,81 ± 1,69	-1,03 ± 0,76	-0,55 ± 0,69	0,52 ± 0,58	-0,24 ± 1,00	-0,17 ± 1,76	0,52 ± 1,72
FY3 (kg)	1,37 ± 2,19	-1,82 ± 1,00	0,88 ± 1	0,82 ± 0,82	0,63 ± 1,41	-1,40 ± 2,12	0,21 ± 3,27
PY1 (kg)	0,78 ± 1,00	0,46 ± 0,47	-0,18 ± 0,41	-0,19 ± 0,35	-0,03 ± 0,60	0,87 ± 1,02	-1,09 ± 1,02
PY2 (kg)	0,61 ± 1,35	-0,18 ± 0,61	-0,40 ± 0,55	0,25 ± 0,46	0,05 ± 0,80	1,40 ± 1,41	0,25 ± 1,37
PY3 (kg)	1,10 ± 1,59	-0,65 ± 0,73	0,73 ± 0,73	-0,48 ± 0,60	0,52 ± 1,02	0,99 ± 1,54	-0,18 ± 2,38
FC1 (%)	-0,030 ± 0,018	-0,011 ± 0,008	0,005 ± 0,008	0,008 ± 0,006	-0,002 ± 0,011	-0,003 ± 0,019	0,008 ± 0,019
FC2 (%)	-0,028 ± 0,020	-0,016 ± 0,009	0,007 ± 0,008	0,007 ± 0,007	0,000 ± 0,012	-0,003 ± 0,020	0,002 ± 0,020
FC3 (%)	-0,045 ± 0,026	-0,013 ± 0,012	0,009 ± 0,012	0,016 ± 0,010	0,007 ± 0,017	0,007 ± 0,025	0,006 ± 0,038
PC1 (%)	-0,025 ± 0,010	0,000 ± 0,005	0,004 ± 0,004	0,002 ± 0,004	0,000 ± 0,006	0,017 ± 0,011	0,001 ± 0,011
PC2 (%)	-0,038* ± 0,011	-0,002 ± 0,005	0,006 ± 0,005	0,003 ± 0,004	0,003 ± 0,007	0,021 ± 0,012	-0,001 ± 0,011
PC3 (%)	-0,040* ± 0,015	0,002 ± 0,007	0,004 ± 0,007	0,000 ± 0,006	0,005 ± 0,010	0,036 ± 0,014	-0,002 ± 0,022

* FDR < 10%

Table 11.14 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *PPARGC1A*-haplotypes in FV

Haplotype name	PPARG_HT1	PPARG_HT3	PPARG_HT4	PPARG_HT6	PPARG_HT5	PPARG_HT2
Haplotype '*x*'	ATTAGG	ACTGGA	CCCGGA	CCCGGG	ACTAGG	ACCAGG
Trait						
MY1 (kg)	-0,81 ± 18,23	0,54 ± 14,90	-38,00* ± 13,74	27,11 ± 14,73	54,61 ± 36,29	-23,51 ± 36,21
MY2 (kg)	-4,24 ± 21,59	11,13 ± 18,41	-38,81 ± 16,72	19,63 ± 17,46	77,91 ± 42,84	-46,56 ± 45,54
MY3 (kg)	-55,23 ± 28,48	-10,34 ± 26,26	-39,82 ± 24,17	41,93 ± 24,53	176,25 ± 69,22	103,77 ± 75,10
FY1 (kg)	-0,12 ± 0,65	-0,13 ± 0,53	0,10 ± 0,49	-0,24 ± 0,53	2,19 ± 1,29	-0,95 ± 1,28
FY2 (kg)	-0,29 ± 0,86	-0,15 ± 0,74	0,23 ± 0,67	-0,74 ± 0,70	3,67 ± 1,71	-1,04 ± 1,82
FY3 (kg)	-2,87 ± 1,15	-1,10 ± 1,06	0,89 ± 0,98	0,38 ± 0,99	4,38 ± 2,82	3,41 ± 3,04
PY1 (kg)	0,55 ± 0,52	0,05 ± 0,43	-0,94 ± 0,39	0,40 ± 0,42	0,92 ± 1,05	-0,53 ± 1,03
PY2 (kg)	0,25 ± 0,69	0,23 ± 0,59	-0,86 ± 0,53	0,03 ± 0,56	2,02 ± 1,37	-0,40 ± 1,45
PY3 (kg)	-1,11 ± 0,84	-0,43 ± 0,77	-0,51 ± 0,71	0,64 ± 0,72	3,47 ± 2,05	2,62 ± 2,21
FC1 (%)	-0,002 ± 0,009	-0,001 ± 0,008	0,027* ± 0,007	-0,025 ± 0,008	0,008 ± 0,019	-0,005 ± 0,019
FC2 (%)	-0,002 ± 0,010	-0,009 ± 0,009	0,028* ± 0,008	-0,024* ± 0,008	0,006 ± 0,020	0,013 ± 0,021
FC3 (%)	-0,009 ± 0,014	-0,008 ± 0,013	0,036* ± 0,011	-0,021* ± 0,012	-0,041 ± 0,033	-0,013 ± 0,036
PC1 (%)	0,008 ± 0,005	0,001 ± 0,004	0,007 ± 0,004	-0,011 ± 0,004	-0,015 ± 0,011	0,003 ± 0,011
PC2 (%)	0,005 ± 0,006	-0,002 ± 0,005	0,009 ± 0,004	-0,011 ± 0,005	-0,011 ± 0,011	0,018 ± 0,012
PC3 (%)	0,011 ± 0,008	0,001 ± 0,007	0,013 ± 0,007	-0,013 ± 0,007	-0,038 ± 0,019	-0,012 ± 0,021

* FDR < 10%

Table 11.15 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (*a*/2) and standard error for milk production traits of *CSN1S1*-haplotypes in FV

Haplotype name	CSN1S_HT2	CSN1S_HT5	CSN1S_HT3	CSN1S_HT4
Haplotype '*x*'	AACTADCCT	AACTATCCT	AATTTTCCT	GGTCATTAC
Trait				
MY1 (kg)	30,52 ± 12,30	-15,32 ± 13,53	-38,78 ± 16,97	-20,16 ± 24,95
MY2 (kg)	19,61 ± 14,81	-17,92 ± 16,08	-13,38 ± 20,39	-20,44 ± 30,46
MY3 (kg)	36,63 ± 20,73	-21,57 ± 22,62	-41,88 ± 27,97	-60,32 ± 41,72
FY1 (kg)	-0,39 ± 0,44	0,76 ± 0,48	-1,33 ± 0,60	0,46 ± 0,88
FY2 (kg)	-0,78 ± 0,59	0,58 ± 0,64	-0,29 ± 0,82	0,14 ± 1,22
FY3 (kg)	-0,84 ± 0,84	0,46 ± 0,92	-0,26 ± 1,13	-0,68 ± 1,69
PY1 (kg)	-0,27 ± 0,35	0,16 ± 0,39	-0,83 ± 0,48	1,11 ± 0,71
PY2 (kg)	-0,84 ± 0,47	0,02 ± 0,51	0,14 ± 0,65	1,31 ± 0,97
PY3 (kg)	-0,53 ± 0,61	0,23 ± 0,67	-0,41 ± 0,82	0,39 ± 1,23
FC1 (%)	-0,027* ± 0,006	0,022* ± 0,007	0,007 ± 0,009	0,023 ± 0,013
FC2 (%)	-0,024* ± 0,007	0,020* ± 0,008	0,004 ± 0,010	0,017 ± 0,014
FC3 (%)	-0,035* ± 0,010	0,021 ± 0,011	0,020 ± 0,013	0,030 ± 0,020
PC1 (%)	-0,023* ± 0,004	0,010 ± 0,004	0,012 ± 0,005	0,032* ± 0,007
PC2 (%)	-0,023* ± 0,004	0,010 ± 0,004	0,010 ± 0,005	0,030* ± 0,008
PC3 (%)	-0,026* ± 0,006	0,015 ± 0,006	0,014 ± 0,008	0,037* ± 0,011

* FDR < 10%

Table 11.16 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *SLC2A9*-haplotypes in FV

Haplotype name	SLC2_HT9	SLC2_HT10	SLC2_HT11	SLC2_HT8	SLC2_HT6	SLC2_HT3	SLC2_HT5
Haplotype '*x*'	ACAACGGGGCG	ACAAGGAGGCT	ACAACGGGACT	ACGAGGGGACT	CCAACGGGGCT	CCAAGGAGGCT	ATAACGGGACT
Trait							
MY1 (kg)	-20,64 ± 28,01	-23,37 ± 37,76	-24,02 ± 26,79	-6,90 ± 15,39	33,70 ± 33,00	52,46 ± 34,68	-32,21 ± 29,91
MY2 (kg)	-8,54 ± 31,79	-2,76 ± 45,06	28,03 ± 34,74	-27,51 ± 18,66	50,28 ± 37,41	89,45 ± 42,15	-13,73 ± 36,95
MY3 (kg)	20,01 ± 43,36	-32,75 ± 65,77	-25,53 ± 56,99	-23,40 ± 25,30	2,25 ± 51,66	-19,82 ± 58,17	-0,84 ± 48,15
FY1 (kg)	-0,11 ± 0,99	0,44 ± 1,34	-1,37 ± 0,94	0,35 ± 0,55	2,49 ± 1,19	1,83 ± 1,23	-0,50 ± 1,07
FY2 (kg)	0,46 ± 1,27	-0,12 ± 1,80	0,29 ± 1,39	-0,72 ± 0,75	3,07 ± 1,49	5,45* ± 1,68	-1,49 ± 1,48
FY3 (kg)	-0,67 ± 1,75	-0,54 ± 2,66	-0,92 ± 2,30	-0,14 ± 1,02	3,84 ± 2,07	2,04 ± 2,35	0,37 ± 1,95
PY1 (kg)	-0,59 ± 0,80	-0,72 ± 1,10	-0,44 ± 0,76	0,49 ± 0,44	1,21 ± 0,95	1,81 ± 1,00	-0,80 ± 0,86
PY2 (kg)	-0,09 ± 1,01	-0,95 ± 1,43	0,82 ± 1,11	-0,09 ± 0,60	1,99 ± 1,19	4,34* ± 1,34	-1,00 ± 1,18
PY3 (kg)	0,29 ± 1,27	-0,11 ± 1,93	-1,67 ± 1,67	0,06 ± 0,75	0,03 ± 1,52	0,49 ± 1,71	0,33 ± 1,42
FC1 (%)	0,012 ± 0,015	0,018 ± 0,020	-0,008 ± 0,014	0,013 ± 0,008	0,022 ± 0,017	0,003 ± 0,018	0,013 ± 0,016
FC2 (%)	0,009 ± 0,015	0,001 ± 0,021	-0,013 ± 0,016	0,008 ± 0,009	0,018 ± 0,018	0,024 ± 0,020	-0,010 ± 0,017
FC3 (%)	-0,023 ± 0,021	0,011 ± 0,031	0,004 ± 0,027	0,017 ± 0,012	0,059 ± 0,025	0,040 ± 0,027	0,008 ± 0,023
PC1 (%)	0,002 ± 0,008	0,001 ± 0,011	0,007 ± 0,008	0,013* ± 0,005	0,002 ± 0,010	0,001 ± 0,010	0,005 ± 0,009
PC2 (%)	0,003 ± 0,008	-0,009 ± 0,012	-0,001 ± 0,009	0,014* ± 0,005	0,003 ± 0,010	0,016 ± 0,011	-0,006 ± 0,010
PC3 (%)	-0,008 ± 0,012	0,013 ± 0,018	-0,007 ± 0,016	0,015 ± 0,007	0,003 ± 0,015	0,016 ± 0,016	0,003 ± 0,013

* FDR < 10%

Haplotype name	SLC2_HT2	SLC2_HT12	SLC2_HT4	SLC2_HT13	SLC2_HT14	SLC2_HT15	SLC2_HT16
Haplotype '*x*'	CCAACGGGGCG	CCAACGGGACT	ACAACGGGGCT	ACGACGGGGCT	CCAAGCGGGCT	CCAAGCGGACG	ACAAGGGGGCT
Trait							
MY1 (kg)	-23,99 ± 27,63	35,77 ± 22,65	-11,17 ± 25,46	-16,51 ± 25,22	36,54 ± 32,01	-12,28 ± 21,56	79,98 ± 32,60
MY2 (kg)	-3,40 ± 33,07	39,81 ± 27,60	-15,79 ± 29,43	-23,46 ± 30,22	3,91 ± 37,51	-35,89 ± 25,45	105,41* ± 39,15
MY3 (kg)	60,61 ± 45,14	65,14 ± 37,07	-7,91 ± 37,95	-49,47 ± 47,68	-12,53 ± 52,45	-46,04 ± 38,56	121,13 ± 49,39
FY1 (kg)	-0,75 ± 1,00	-1,29 ± 0,81	-0,70 ± 0,90	-0,08 ± 0,89	1,65 ± 1,14	0,78 ± 0,76	0,28 ± 1,18
FY2 (kg)	-0,48 ± 1,32	-0,45 ± 1,11	-0,40 ± 1,18	-0,21 ± 1,21	1,15 ± 1,50	-0,25 ± 1,02	1,48 ± 1,57
FY3 (kg)	0,97 ± 1,83	-0,78 ± 1,50	1,19 ± 1,53	-1,24 ± 1,93	2,30 ± 2,11	-1,12 ± 1,56	1,22 ± 2,01
PY1 (kg)	-1,18 ± 0,80	-0,29 ± 0,65	-0,09 ± 0,72	-0,22 ± 0,72	0,70 ± 0,92	0,53 ± 0,62	0,96 ± 0,94
PY2 (kg)	0,11 ± 1,05	-0,66 ± 0,88	0,04 ± 0,94	-0,16 ± 0,96	0,30 ± 1,20	-0,18 ± 0,81	1,31 ± 1,25
PY3 (kg)	1,58 ± 1,33	-0,19 ± 1,10	1,04 ± 1,11	-1,10 ± 1,41	0,47 ± 1,54	-0,29 ± 1,14	1,41 ± 1,46
FC1 (%)	0,004 ± 0,014	-0,044* ± 0,012	-0,007 ± 0,013	0,008 ± 0,013	0,005 ± 0,017	0,016 ± 0,011	-0,047* ± 0,017
FC2 (%)	-0,006 ± 0,015	-0,034* ± 0,013	0,004 ± 0,014	0,012 ± 0,014	0,015 ± 0,018	0,018 ± 0,012	-0,041 ± 0,018
FC3 (%)	-0,023 ± 0,021	-0,051* ± 0,017	0,020 ± 0,018	0,015 ± 0,022	0,036 ± 0,025	0,012 ± 0,018	-0,052 ± 0,024
PC1 (%)	-0,006 ± 0,008	-0,025* ± 0,007	0,007 ± 0,008	0,007 ± 0,008	-0,007 ± 0,010	0,013 ± 0,006	-0,029* ± 0,010
PC2 (%)	0,002 ± 0,009	-0,032* ± 0,007	0,010 ± 0,008	0,010 ± 0,008	0,003 ± 0,010	0,015 ± 0,007	-0,035* ± 0,010
PC3 (%)	-0,008 ± 0,012	-0,035* ± 0,010	0,019 ± 0,011	0,009 ± 0,013	0,012 ± 0,015	0,018 ± 0,011	-0,039* ± 0,014

* FDR < 10%

Table 11.17 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *ABCG2*-haplotypes in HF

Haplotype name	ABC_HT2	ABC_HT3	ABC_HT6	ABC_HT7	ABC_HT8	ABC_HT9	ABC_HT10
Haplotype '*x*'	ATATGGATA	AAATAGATA	GTTTGGATA	GTTTAGATA	GTTTGAGGA	GTTTGAATA	ATATAGATA
Trait							
MY1 (kg)	-88,24±40,09	29,42±22,46	33,59±25,82	27,96±36,27	-57,10±29,25	-3,32±36,70	29,42±22,46
MY2 (kg)	-52,62±43,82	23,27±24,62	49,57±28,24	13,61±39,76	-82,12±32,01	-11,35±40,19	23,27±24,62
MY3 (kg)	-105,78±50,70	20,31±26,62	76,11±30,95	18,93±43,03	-60,89±34,97	-31,36±43,52	20,31±26,62
FY1 (kg)	0,92±1,52	0,69±0,85	1,61±0,97	-1,48±1,37	-1,97±1,10	-1,31±1,39	0,69±0,85
FY2 (kg)	2,51±1,75	0,30±0,98	2,31±1,12	-2,32±1,58	-2,95±1,28	-1,93±1,60	0,30±0,98
FY3 (kg)	2,13±2,14	0,28±1,12	2,94±1,30	-1,61±1,81	-3,32±1,47	-2,61±1,83	0,28±1,12
PY1 (kg)	-1,11±1,15	0,28±0,64	1,35±0,74	1,16±1,03	-1,93±0,83	-0,27±1,05	0,28±0,64
PY2 (kg)	0,27±1,35	0,00±0,76	2,19±0,87	0,84±1,22	-2,89*±0,98	-0,76±1,24	0,00±0,76
PY3 (kg)	-0,48±1,6248	-0,19±0,8508	2,74*±0,9884	0,83±1,3741	-2,32±1,1156	-1,30±1,39	-0,19±0,8508
FC1 (%)	0,054*±0,018	-0,006±0,010	0,003±0,011	-0,032±0,016	0,004±0,013	-0,014±0,016	-0,006±0,010
FC2 (%)	0,052*±0,018	-0,007±0,010	0,003±0,012	-0,030±0,016	0,003±0,013	-0,018±0,016	-0,007±0,010
FC3 (%)	0,067*±0,021	-0,006±0,011	0,000±0,013	-0,025±0,017	-0,009±0,014	-0,015±0,017	-0,006±0,011
PC1 (%)	0,024*±0,008	-0,009±0,005	0,003±0,005	0,003±0,008	-0,001±0,006	-0,002±0,008	-0,009±0,005
PC2 (%)	0,023*±0,009	-0,009±0,005	0,006±0,006	0,005±0,008	-0,002±0,006	-0,005±0,008	-0,009±0,005
PC3 (%)	0,031*±0,010	-0,010±0,005	0,004±0,006	0,004±0,008	-0,004±0,007	-0,003±0,009	-0,010±0,005

* FDR < 10%

Table 11.18 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *OPN*-haplotypes in HF

Haplotype name	OP_HT1	OP_HT2	OP_HT3	OP_HT4	OP_HT5
Haplotype '*x*'	TGACC	TAGCT	TGATC	CGGCC	TGGCC
Trait					
MY1 (kg)	-7,63±51,01	-14,33±25,66	-1,07±53,63	18,63±21,60	-6,98±27,91
MY2 (kg)	-7,66±55,60	-34,71±28,11	-23,67±58,72	35,33±23,64	-7,81±30,58
MY3 (kg)	2,14±61,86	-34,20±30,40	-17,56±62,97	50,31±25,74	-36,20±33,14
FY1 (kg)	0,03±1,93	-0,01±0,97	0,83±2,02	1,22±0,82	-2,15±1,05
FY2 (kg)	-0,24±2,22	-0,73±1,12	-0,29±2,34	2,05±0,94	-2,38±1,22
FY3 (kg)	1,73±2,60	-1,72±1,28	-1,10±2,65	3,18*±1,08	-3,36±1,39
PY1 (kg)	-0,79±1,45	-0,23±0,73	0,30±1,53	1,23±0,62	-1,42±0,80
PY2 (kg)	-1,11±1,71	-0,94±0,87	-0,05±1,81	1,95*±0,73	-1,72±0,94
PY3 (kg)	-0,88±1,98	-1,37±0,97	0,00±2,012	2,76*±0,82	-2,58±1,06
FC1 (%)	0,007±0,022	0,007±0,011	0,013±0,024	0,006±0,010	-0,025±0,012
FC2 (%)	0,004±0,023	0,008±0,011	0,009±0,024	0,007±0,010	-0,025±0,012
FC3 (%)	0,017±0,025	-0,003±0,012	-0,003±0,025	0,013±0,010	-0,022±0,013
PC1 (%)	-0,006±0,011	0,003±0,005	0,005±0,011	0,008±0,005	-0,016*±0,006
PC2 (%)	-0,010±0,011	0,003±0,006	0,009±0,012	0,009±0,005	-0,017*±0,006
PC3 (%)	-0,014±0,012	-0,001±0,006	0,008±0,012	0,012±0,005	-0,016±0,007

* FDR < 10%

Table 11.19 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *PPARGC1A*-haplotypes in HF

Haplotype name	PPARG_HT1	PPARG_HT4	PPARG_HT3	PPARG_HT5	PPARG_HT2	PPARG_HT7	PPARG_HT8
Haplotype '*x*'	ATTAGG	CCCGGA	ACTGGA	ACTAGG	ACCAGG	ACTATG	ACTAGA
Trait							
MY1 (kg)	-7,33 ± 27,49	-25,38 ± 34,04	2,86 ± 21,45	-4,60 ± 39,64	-29,98 ± 59,61	42,30 ± 43,12	32,09 ± 32,89
MY2 (kg)	-21,50 ± 30,06	-25,09 ± 37,23	-8,53 ± 23,45	32,65 ± 43,47	-5,16 ± 65,26	70,59 ± 47,19	25,80 ± 36,03
MY3 (kg)	-37,12 ± 32,60	-35,81 ± 40,99	-13,90 ± 25,15	-6,92 ± 47,12	23,19 ± 70,11	105,84 ± 51,16	72,76 ± 39,02
FY1 (kg)	-0,40 ± 1,04	2,04 ± 1,28	-0,15 ± 0,81	-1,29 ± 1,50	0,01 ± 2,25	2,81 ± 1,63	-1,16 ± 1,24
FY2 (kg)	-0,34 ± 1,20	2,18 ± 1,48	-0,49 ± 0,93	-0,71 ± 1,73	1,13 ± 2,60	4,78 ± 1,88	-2,27 ± 1,43
FY3 (kg)	-0,99 ± 1,37	1,70 ± 1,72	-0,78 ± 1,06	-2,19 ± 1,98	1,53 ± 2,95	7,25* ± 2,14	-0,78 ± 1,64
PY1 (kg)	-0,51 ± 0,78	-0,15 ± 0,97	-0,11 ± 0,61	-0,43 ± 1,13	0,11 ± 1,70	2,68 ± 1,23	0,52 ± 0,94
PY2 (kg)	-1,22 ± 0,92	-0,23 ± 1,15	-0,39 ± 0,72	0,56 ± 1,34	1,21 ± 2,01	4,32* ± 1,45	0,20 ± 1,11
PY3 (kg)	-1,48 ± 1,04	-0,55 ± 1,31	-0,90 ± 0,80	-0,61 ± 1,50	1,50 ± 2,24	6,27* ± 1,63	1,89 ± 1,25
FC1 (%)	-0,002 ± 0,012	0,039* ± 0,015	-0,003 ± 0,009	-0,015 ± 0,017	0,020 ± 0,026	0,014 ± 0,019	-0,031 ± 0,014
FC2 (%)	0,006 ± 0,012	0,037 ± 0,015	-0,002 ± 0,010	-0,024 ± 0,018	0,018 ± 0,027	0,023 ± 0,019	-0,038* ± 0,015
FC3 (%)	0,006 ± 0,013	0,035 ± 0,017	-0,002 ± 0,010	-0,022 ± 0,019	0,012 ± 0,028	0,035 ± 0,021	-0,042* ± 0,016
PC1 (%)	-0,004 ± 0,006	0,009 ± 0,007	-0,002 ± 0,005	-0,004 ± 0,008	0,015 ± 0,013	0,015 ± 0,009	-0,007 ± 0,007
PC2 (%)	-0,006 ± 0,006	0,008 ± 0,007	-0,001 ± 0,005	-0,006 ± 0,009	0,016 ± 0,013	0,022 ± 0,009	-0,008 ± 0,007
PC3 (%)	-0,004 ± 0,006	0,006 ± 0,008	-0,004 ± 0,005	-0,004 ± 0,009	0,009 ± 0,014	0,031* ± 0,010	-0,006 ± 0,008

* FDR < 10%

Table 11.20 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *PGM2*-haplotypes in HF

Haplotype name	PGM2_HT2	PGM2_HT4	PGM2_HT1	PGM2_HT5	PGM2_HT3	PGM2_HT6
Haplotype '*x*'	AACATCACCA	GATGCGGTTG	GATGCGGTCG	GACATCACCA	GGTGCGGTTG	AATGCGGTTG
Trait						
MY1 (kg)	-16,39 ± 22,89	15,82 ± 21,64	-14,75 ± 28,51	30,16 ± 30,67	-27,51 ± 41,41	-30,53 ± 45,68
MY2 (kg)	-4,50 ± 25,07	-10,37 ± 23,70	4,11 ± 31,20	51,53 ± 33,59	-38,33 ± 45,33	-20,39 ± 50,06
MY3 (kg)	29,37 ± 27,01	-32,13 ± 25,69	10,64 ± 33,79	37,70 ± 36,23	-54,70 ± 50,45	-32,99 ± 54,43
FY1 (kg)	-1,64 ± 0,86	0,70 ± 0,82	0,27 ± 1,08	0,22 ± 1,16	0,98 ± 1,56	0,92 ± 1,72
FY2 (kg)	-1,25 ± 1,00	-0,12 ± 0,94	1,23 ± 1,24	0,37 ± 1,34	0,68 ± 1,81	1,80 ± 1,99
FY3 (kg)	-0,70 ± 1,14	-0,34 ± 1,08	1,97 ± 1,42	0,20 ± 1,52	-0,42 ± 2,12	0,25 ± 2,29
PY1 (kg)	-0,78 ± 0,65	0,99 ± 0,62	-0,32 ± 0,81	1,23 ± 0,87	-1,00 ± 1,18	-2,11 ± 1,30
PY2 (kg)	-0,56 ± 0,77	0,18 ± 0,73	0,38 ± 0,96	1,93 ± 1,03	-0,62 ± 1,40	-2,28 ± 1,54
PY3 (kg)	0,54 ± 0,86	-0,33 ± 0,82	0,65 ± 1,08	1,60 ± 1,16	-1,63 ± 1,61	-3,61 ± 1,74
FC1 (%)	-0,012 ± 0,010	0,001 ± 0,010	0,011 ± 0,013	-0,012 ± 0,013	0,026 ± 0,018	0,024 ± 0,020
FC2 (%)	-0,013 ± 0,010	0,003 ± 0,010	0,013 ± 0,013	-0,018 ± 0,014	0,024 ± 0,018	0,027 ± 0,020
FC3 (%)	-0,021 ± 0,011	0,010 ± 0,010	0,018 ± 0,014	-0,014 ± 0,015	0,022 ± 0,020	0,015 ± 0,022
PC1 (%)	-0,003 ± 0,005	0,006 ± 0,005	0,002 ± 0,006	0,003 ± 0,007	0,000 ± 0,009	-0,015 ± 0,010
PC2 (%)	-0,005 ± 0,005	0,006 ± 0,005	0,003 ± 0,006	0,002 ± 0,007	0,009 ± 0,009	-0,019 ± 0,010
PC3 (%)	-0,005 ± 0,005	0,008 ± 0,005	0,004 ± 0,007	0,004 ± 0,007	0,005 ± 0,010	-0,028* ± 0,011

* FDR < 10%

Table 11.21 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *CSN1S1*-haplotypes in HF

Haplotype name	CSN1S_HT1	CSN1S_HT2	CSN1S_HT3	CSN1S_HT6
Haplotype '*x*'	AATTATCCT	AACTADCCT	AATTTTCCT	AATCATTCC
Trait				
MY1 (kg)	45,75 ± 34,56	16,75 ± 19,86	-46,04 ± 22,34	75,73 ± 53,48
MY2 (kg)	64,09 ± 37,73	18,61 ± 21,73	-45,07 ± 24,44	24,87 ± 58,72
MY3 (kg)	62,76 ± 41,85	11,49 ± 23,57	-32,88 ± 27,19	24,80 ± 65,54
FY1 (kg)	2,60 ± 1,30	0,20 ± 0,75	-1,23 ± 0,84	0,55 ± 2,02
FY2 (kg)	2,94 ± 1,50	0,04 ± 0,87	-0,73 ± 0,97	-1,25 ± 2,34
FY3 (kg)	1,98 ± 1,76	-0,20 ± 0,99	0,21 ± 1,14	-1,03 ± 2,75
PY1 (kg)	1,11 ± 0,99	-0,39 ± 0,57	-0,15 ± 0,64	2,71 ± 1,52
PY2 (kg)	1,81 ± 1,16	-0,71 ± 0,67	0,24 ± 0,75	1,71 ± 1,81
PY3 (kg)	1,80 ± 1,34	-0,75 ± 0,75	0,64 ± 0,87	0,86 ± 2,09
FC1 (%)	0,010 ± 0,015	-0,008 ± 0,009	0,008 ± 0,010	-0,032 ± 0,023
FC2 (%)	0,005 ± 0,015	-0,010 ± 0,009	0,013 ± 0,010	-0,029 ± 0,024
FC3 (%)	-0,005 ± 0,017	-0,010 ± 0,010	0,019 ± 0,011	-0,027 ± 0,026
PC1 (%)	-0,005 ± 0,007	-0,013* ± 0,004	0,018* ± 0,005	0,001 ± 0,011
PC2 (%)	-0,003 ± 0,008	-0,016* ± 0,004	0,020* ± 0,005	0,008 ± 0,012
PC3 (%)	-0,003 ± 0,008	-0,014* ± 0,005	0,019* ± 0,005	-0,001 ± 0,013

* FDR < 10%

Table 11.22 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of *SLC2A9*-haplotypes in HF

Haplotype name	SLC2_HT6	SLC2_HT17	SLC2_HT4	SLC2_HT18	SLC2_HT9	SLC2_HT2	SLC2_HT5	SLC2_HT8
Haplotype '*x*'	CCAACGGGGCT	CCGAGGAGGCT	ACAACGGGGCT	ACGAGGGGGCT	ACAACGGGGCG	CCAACGGGGCG	ATAACGGGACT	ATAACGGGACT
Trait								
MY1 (kg)	-6,56 ± 34,66	-2,92 ± 26,06	36,21 ± 24,78	-85,00 ± 39,03	-18,00 ± 25,35	32,19 ± 49,85	61,03 ± 53,68	-36,17 ± 57,25
MY2 (kg)	0,41 ± 37,98	28,43 ± 28,54	42,53 ± 27,12	-99,96 ± 42,68	-24,63 ± 27,75	35,85 ± 54,58	-12,11 ± 58,77	-91,39 ± 62,64
MY3 (kg)	-6,44 ± 40,88	26,33 ± 31,05	47,11 ± 29,55	-100,49 ± 46,50	-32,86 ± 30,48	53,25 ± 59,20	-50,90 ± 63,37	-96,50 ± 67,76
FY1 (kg)	-0,30 ± 1,31	-0,51 ± 0,98	0,71 ± 0,94	-1,85 ± 1,48	-1,08 ± 0,96	4,11 ± 1,88	2,73 ± 2,03	-1,96 ± 2,16
FY2 (kg)	-0,13 ± 1,51	0,54 ± 1,14	0,85 ± 1,08	-1,90 ± 1,70	-1,29 ± 1,11	4,40 ± 2,17	0,61 ± 2,34	-4,05 ± 2,50
FY3 (kg)	-0,17 ± 1,72	0,01 ± 1,31	1,62 ± 1,24	-1,09 ± 1,96	-1,92 ± 1,28	6,71* ± 2,48	-0,22 ± 2,67	-4,94 ± 2,85
PY1 (kg)	-0,09 ± 0,99	-0,37 ± 0,74	1,11 ± 0,71	-2,15 ± 1,11	-0,35 ± 0,72	0,70 ± 1,42	2,36 ± 1,53	-1,55 ± 1,63
PY2 (kg)	0,36 ± 1,17	0,72 ± 0,88	1,41 ± 0,83	-2,61 ± 1,31	-0,97 ± 0,85	0,39 ± 1,68	0,76 ± 1,81	-2,84 ± 1,93
PY3 (kg)	-0,25 ± 1,31	0,45 ± 0,99	1,68 ± 0,94	-2,18 ± 1,49	-1,26 ± 0,97	1,30 ± 1,89	-0,20 ± 2,03	-3,40 ± 2,16
FC1 (%)	0,001 ± 0,015	-0,004 ± 0,011	-0,009 ± 0,011	0,022 ± 0,017	-0,003 ± 0,011	0,035 ± 0,022	0,002 ± 0,024	-0,011 ± 0,025
FC2 (%)	-0,001 ± 0,015	-0,007 ± 0,012	-0,010 ± 0,011	0,027 ± 0,017	-0,001 ± 0,011	0,033 ± 0,022	0,012 ± 0,024	-0,007 ± 0,026
FC3 (%)	0,003 ± 0,016	-0,010 ± 0,012	-0,004 ± 0,012	0,034 ± 0,019	-0,005 ± 0,012	0,046 ± 0,024	0,021 ± 0,026	-0,018 ± 0,027
PC1 (%)	0,002 ± 0,007	-0,003 ± 0,006	-0,002 ± 0,005	0,011 ± 0,008	0,004 ± 0,005	-0,005 ± 0,011	0,002 ± 0,011	-0,006 ± 0,012
PC2 (%)	0,004 ± 0,008	-0,003 ± 0,006	0,000 ± 0,005	0,011 ± 0,009	-0,001 ± 0,006	-0,009 ± 0,011	0,012 ± 0,012	0,001 ± 0,012
PC3 (%)	0,000 ± 0,008	-0,005 ± 0,006	0,001 ± 0,006	0,014 ± 0,009	-0,001 ± 0,006	-0,006 ± 0,012	0,015 ± 0,013	-0,003 ± 0,013

* FDR < 10%

Table 11.23 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of haplotypes including polymorphisms associated with distinct effects in BV

Haplotype name	HT1_BV	HT2_BV	HT3_BV	HT4_BV	HT5_BV	HT6_BV
Haplotype '*x*'	ATCC	ADCC	ADGC	ATGC	GTGC	ADGT
Trait						
MY1 (kg)	-73,86	56,62	-11,54	-73,26	-102,72	86,15
MY2 (kg)	-100,60	108,53	-12,78	-73,25	-191,10	14,74
MY3 (kg)	-29,51	56,93	-35,87	-28,07	-141,27	55,04
FY1 (kg)	1,21	0,10	-1,46	-1,60	-0,15	0,08
FY2 (kg)	0,81	2,14	-1,57	-1,50	-1,83	-4,91
FY3 (kg)	2,38	-1,47	-2,39	1,61	0,29	-1,44
PY1 (kg)	-1,56	1,73	-0,28	-2,48	-2,27	0,00
PY2 (kg)	-2,45	3,03	-0,60	-2,15	-3,51	-2,98
PY3 (kg)	0,84	1,22	-1,40	0,18	-2,58	-2,90

Table 11.24 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of haplotypes including polymorphisms associated with distinct effects in FV

Haplotype name	HT1_FV	HT2_FV	HT3_FV	HT4_FV	HT5_FV	HT6_FV
Haplotype '*x*'	GAD	AAD	GAT	AAT	AGT	GGT
Trait						
MY1 (kg)	51,37	5,13	6,46	-54,34	-169,08	-102,58
MY2 (kg)	46,61	-17,93	-2,54	-31,49	-156,50	-109,70
MY3 (kg)	44,30	40,28	8,10	-55,56	-305,08	-56,95
FY1 (kg)	-0,91	-0,53	0,22	0,39	-2,40	-4,11
FY2 (kg)	-1,13	-1,18	0,21	1,21	-1,10	-5,44
FY3 (kg)	-2,47	0,92	1,38	0,01	-8,11	-4,80
PY1 (kg)	0,30	-1,41	0,58	-0,62	-2,12	-0,90
PY2 (kg)	-0,09	-2,39	0,48	0,29	-0,56	-1,30
PY3 (kg)	-0,69	-0,55	0,96	-0,41	-5,26	1,41

Table 11.25 Regression coefficients on the copy number of haplotype '*x*' representing half of the allele substitution effects (α/2) and standard error for milk production traits of haplotypes including polymorphisms associated with distinct effects in HF

Haplotype name	HT1_HF	HT2_HF	HT3_HF	HT4_HF	HT5_HF
Haplotype '*x*'	GGA	AGA	AGT	GGT	GTT
Trait					
MY1 (kg)	51,93	-21,35	-200,51	-55,53	-69,80
MY2 (kg)	56,94	-46,15	-228,42	-39,19	-86,20
MY3 (kg)	52,19	-56,34	-175,87	-51,43	-10,40
FY1 (kg)	1,56	-1,93	-6,04	-1,64	3,17
FY2 (kg)	1,67	-3,75	-5,99	-0,71	3,87
FY3 (kg)	1,14	-5,03	-5,86	0,55	8,63
PY1 (kg)	0,67	-1,74	-5,11	0,19	2,02
PY2 (kg)	0,77	-3,32	-5,79	1,23	2,74
PY3 (kg)	0,59	-3,65	-3,75	0,92	5,78

Curriculum vitae

Name:	Franz Reinhold Seefried
Date of birth:	23.04.1978
Place of birth:	Kempten / Allgäu
Marital status:	single
Nationality:	German
Parents:	Dipl. Ing. (FH) Reinhold Seefried, Maria Seefried

Education

Sep. 1984 – Sep. 1988	Primary School, Konrad Adenauer Grundschule Kempten /Allgäu
Sep. 1988 – Jul. 1997	Secondary School, Hildegardis Gymnasium Kempten / Allgäu

Studies

Sep. 1998 – Jul. 2004	Study of Agricultural Science with focus on Animal Production at the Life Science Center of the Technical University of Munich
Jul. 2004	Diploma degree, finals with distinction, alumnus award Diploma topic: Untersuchungen zur Trinkschwäche neugeborener Braunviehkälber (Chair of Animal Breeding, TUM)

Postgraduate education

Since Aug. 2004	Ph.D. studies at the Chair of Animal Breeding, Technical University of Munich, Topic: Characterisation and polymorphism analysis of candidate genes for milk production traits and association studies with milk production traits in three cattle breeds

Trainings

Aug. 1998 – Sep. 1998	Agricultural farm training, Armin Schönmetzler GbR, Betzigau,
Nov. 2000 – Mar. 2001	Student assistant at the Chair of Animal Breeding, Technical University of Munich
Apr. 2001 – Dec. 2001	Agricultural farm training, Schwayer – Hofer GbR, Lauchdorf
Jan. 2002 – Mar. 2002	Training at the German Brown Cattle Association AHG Kempten and at RBG Memmingen