TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Tierzucht

Identification and Characterization of Candidate Genes for Complex Traits in Cattle

Xiaolong Wang

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.

Vorsitzender: Univ.-Prof. Dr. W. M. Windisch

Prüfer der Dissertation:

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

Die Dissertation wurde am 30.05.2012 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum für Ernährung, Landnutzung und Umwelt am 09.11.2012 angenommen.

Table of Contents

Abbreviations

°C	Degrees celsius
ACOP	Ambilateral circumocular pigmentation
ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta
ART4	ADP-ribosyltransferase 4
ASIP	Agouti
ATF7IP	Activating transcription factor 7 interacting protein
BOSCC	Bovine ocular squamous cell carcinoma
bp	Base pair
BTA	*Bos taurus* autosome
CNV	Copy number variation
Ct	Cycle threshold
DERA	Deoxyribose-phosphate aldolase (putative)
DGAT1	Diacylglycerol O-acyltransferase 1
DMSO	Dimethyl sulfid
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotides
EBVs	Estimated breeding values
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EPS8	Epidermal growth factor receptor pathway substrate 8
ERP27	Endoplasmic reticulum protein 27
FP	Fat percentage
FV	German Fleckvieh
GEBV	Genetic estimated breeding value
GHR	Growth hormone receptor
GPAT4	1-acylglycerol-3-phosphate O-acyltransferase 6
GRIN2B	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B
GRM	Genomic relationship matrix
GSDMC	Gasdermin C
GUCY2C	Guanylate cyclase 2C
GWA	Genome-wide association
GWAS	Genome-wide association studies
H2AFJ	H2A histone family, member J
HF	German Holstein-Friesian
indel	Insertion-deletion polymorphism
KDR	Kinase insert domain receptor
KIT	Tyrosine-protein kinase Kit

LD	Linkage disequilibrium
LMO3	LIM domain only protein 3
MAF	Minor allele frequency
MAS	Marker assisted selection
Mb	Mega base pairs
MC1R	Melanocortin 1 receptor
MGP	Matrix Gla protein
MGST1	Microsomal glutathione S-transferase 1
miRNA	micro-RNA
MITF	Microphthalmia-associated transcription factor
mRNA	Messenger ribonucleic acid
mtDNA	Mitochondrial DNA
NCBI	National center of biotechnology information
NEFA	Nonesterified fatty acids
ng	Nanogram
PAX3	Paired box 3
PCR	Polymerase chain reaction
PDE6H	Phosphodiesterase 6H, cGMP-specific, cone, gamma
PTPRO	Protein tyrosine phosphatase, receptor type, O
QTL	Quantitative trait loci
QTG	Quantitative trait genes
QTN	Quantitative trait nucleotides
RACE	Rapid amplification of cDNA ends
RERG	RAS-like, estrogen-regulated, growth inhibitor
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RT-PCR	Real-time polymerase chain reaction
SNP	Single nucleotide polymorphism
SNP_id	In-house single nucleotide polymorphism identification code
STRAP	Serine/threonine kinase receptor associated protein
TBE	Tris-borate-EDTA
TE	Tris EDTA buffer
UTR	Untranslated region
UV	Ultraviolet
VNTR	Variable number tandem repeat
µL	Micro litre

List of tables

List of figures

List of appendices

Preface

This PhD thesis concludes my three years research training within the field of animal genetics in the Chair of Animal Breeding, Technische Universität München. This dissertation is based on research results from two independent studies, all trying to identify QTN of two prototypical complex traits (milk fat content and ambilateral circumocular pigmentation) in cattle (*Bos taurus*). In order to achieve this, candidate genes were selected from QTL regions mapped by previous GWAS. Subsequently, candidate gene re-sequencing and mutation analyses were performed in order to refine these QTL regions.

Chapter 1 provides an introduction to this thesis. It starts with a review of the conventional strategies to dissect candidate genes underlying complex traits, and mainly focuses on the methodology of GWAS, which is used to map the QTL regions in this dissertation. The procedures used to map the candidate genes for milk fat traits and coat colours in cattle are subsequently reviewed.

Chapter 2 investigates the findings of a GWAS on milk fat content in 2435 German Holstein-Friesian bulls. The genetic associations between the EBVs for milk fat percentage of these bulls were analysed, and two novel QTL regions on BTA5 and BTA27 were discovered. The most significant QTL on BTA5 was chosen for further analysis. Sixteen positional candidate genes (*LMO3, MGST1, DERA, STRAP, EPS8, PTPRO, RERG, ARHGDIB, PDE6H, ERP27, H2AFJ, MGP, ART4, GUCY2C, ATF7IP* and *GRIN2B*) within the interval of this locus were re-sequenced and analysed in order to locate the causative variant. A promoter SNP that resides in an evolutionarily conserved site and 100 bp upstream of the *EPS8* gene is highly associated with the EBVs of milk fat percentage.

Chapter 3 refers to the results from a GWAS of ambilateral circumocular pigmentation in 3579 German Fleckvieh bulls. Five positional and functional genes (*PAX3*, *KIT*, *KDR*, *GSDMC* and *MITF*) underlying four QTL regions were chosen for re-sequencing and analysis. Expression of the *KIT* gene between the follicles from pigmented and unpigmented hair was also investigated.

The general discussion in **Chapter 4** starts with a strategy of selecting candidate genes

derived from previous GWAS. This is followed by a discussion of genomic organization and characterization. Subsequently, the approaches of SNP detection, QTN validation and analysis of non-coding variants are discussed separately, and future perspectives for identification of QTN in farm animals conclude this thesis.

Chapter 1
General introduction

Genetic evaluation using genome-wide information is an important milestone in the long-awaited application of molecular genetics technology to animal improvement. Certain genetic markers could improve estimates of the genetic potentials of animals as they underlie many important production traits and scientists have been working for many years to map the genes responsible for phenotypic variation. Many of the monogenic traits regulated by single genes can be detected by performing linkage studies, whereas many phenotypes, which are quantitative in nature and have multiple interactions between genetic and environmental factors, remain difficult to pinpoint.

1. The methodology of mapping genetic variations

Practical improvements in animal breeding programs through the use of DNA markers to predict the performance of animals have been developed. The process was called marker assisted selection (MAS), this facilitates the exploitation of existing genetic diversity in breeding populations and can be used to improve a whole range of desirable traits (Beuzen *et al*. 2000). Some of the genetic markers commonly used in MAS include restriction fragment length polymorphisms (RFLP), variable number tandem repeats (VNTR), microsatellites and single nucleotide polymorphisms (SNP) (Williams 2005). Based upon these markers, linkage and association studies have been widely used to evaluate whether or not the genetic variations correspond to desired traits (Beuzen *et al*. 2000).

1.1 Linkage studies

Linkage studies, which are old-fashioned approaches, have been widely used to dissect genetic loci or genes by observing related individuals, for example, in a family. On the basis of pedigree information, the genetic markers identified can be employed to narrow down the genetic regions that are inherited by family members. Linkage studies have been successfully used to identify susceptibility genes in major genetic disorders (parametric linkage) and complex diseases (model-free or non-parametric linkage) (Dawn Teare and Barrett 2005). However, due to the complexity of common diseases, linkage studies have mainly been used to detect genes inherited in a Mendelian fashion. Furthermore, the application of genome-wide linkage analysis in order to identify genetic variants underlying complex traits has only been partially successful. This is mainly due to the low heritability or

small effect of a huge number of genes; however, they still contribute to complex traits (Pritchard and Cox. 2002).

1.2 Association studies

Association studies either use a population-based (case/control or quantitative trait models) or a family-based (transmission disequilibrium test) approach to gaining more power in mapping candidate genes. In comparison with linkage approaches, association studies localise genetic signals in smaller regions of chromosomes since they can capture all of the meiotic recombination events in a population rather than gaining information from pedigrees.

1.2.1 Genome-wide association studies (GWAS)

With the available genome sequences for many organisms, combined with comprehensive SNP data and advanced genotyping technologies, mapping genetic associations through entire genomes has become possible and has shed light on genetic studies of complex traits. In 2005, *Science* reported the first GWAS in which a variation associated with age-related macular degeneration was successfully identified by testing ~100,000 SNPs (Klein *et al*. 2005). The year after witnessed a surge of large-scale and high-density GWAS which identified hundreds of associations of complex traits both in humans and model organisms, and further improved our understanding of the genetic architecture of many complex traits to a great extent.

Although numerous susceptibility loci/genes have been so far identified by GWAS, a small fraction of the heritable variations in these traits were explained. The mystery of the "missing heritability" or "dark matter" is due to the limitations in the design of early GWAS and current GWAS are still underpowered for detecting most common variants (Manolio *et al*. 2009; Lango Allen *et al*. 2010). Moreover, the significant SNPs defined by GWAS are largely proxies for the true causative variants existing nearby. Fine-mapping of candidate quantitative trait loci (QTL) regions and validating their true causal variants are key challenges in dissecting complex traits (Freedman *et al*. 2011).

GWAS have been proven to be successful in human genetic studies, and can be conducted in the same manner in animal genetics on the basis of released higher-density SNP arrays and on-going re-sequencing projects. On the one hand, the results of GWAS can be applied to MAS, known as genomic selection in livestock in order to improve a number of important production traits (Hayes *et al*. 2009). On the other hand, the ample markers covering the entire genome can be used for mapping candidate genes and for eventually identifying of causal variants of economically important agronomic traits (Goddard 2008).

2. Identifying genes that underlie milk fat traits in dairy cows

The application of genetic and genomic tools to livestock selection at present is mainly focused on mapping QTL, which control agriculturally important traits. Once the QTL have been mapped, the far more difficult step is to identify their candidate quantitative trait genes (QTG), validate the relevant nucleotide polymorphisms (quantitative trait nucleotides, or QTN) and elucidate how they contribute to the quantitative traits (Lyman *et al*. 1999; Flint and Mott 2001). Ron and Weller (2007) proposed that candidate genes underlying QTL should meet at least one of the four following criteria: 1) have a known physiological role in relation to the phenotype; 2) the existence of studies of knock-outs, mutations or transgenics in other models; 3) be preferentially expressed in related tissues; and 4) be preferentially expressed during developmental stages relevant to the phenotype. Subsequently, sequencing of the most likely candidate genes is employed to search for nucleotide polymorphisms which may include the potential QTN. In order to validate a candidate polymorphism as an actual QTN from multiple SNPs in livestock, a schematic strategy including both statistical and physiological methods should be considered (Ron and Weller 2007). Cohen-Zinder *et al*. (2005) summarised four statistical points for determining a potential QTN: 1) the effect of the putative QTN accounts for the entire effect; 2) no other polymorphisms in linkage disequilibrium (LD) with the QTL have significant effects in models that also include the effect of the putative QTN; 3) the QTN is segregating in diverse breeds; and 4) changes in the allelic frequencies of the QTN correspond to the changes expected because of selection in the population. The functional analyses aim to demonstrate whether the alternative allele of QTN produces unequal phenotypes, or alters the function of proteins (Ron and Weller 2007).

Fat is the most variable component in milk because it contributes unique characteristics to the flavour, texture, appearance and satiability of dairy products, provides a source of fat soluble vitamins, essential fatty acids and other constitutional compounds (Anon 2006). Variations in fat composition will largely alter the value of milk. The composition of milk fat is continuously influenced by multiple environmental factors such as dietary supply, gestation, management of the cow, health status, lactation stage and season (Olori *et al*. 1997; Warnick *et al*. 2001; Kendall *et al*. 2009). Therefore, understanding the genetic and genomic basis of milk fat will facilitate dairy cattle improvements in certain herds or in the entire population.

Over the years, research has yielded dozens of QTL responsible for fat traits by performing linkage and association studies. Since the first QTL affecting milk production traits was detected in dairy cattle by exploiting "progeny testing" at the genome-scale (Georges *et al*. 1995), many other QTL associated with milk production traits have been reported **(Table 1-1)**. With the completion of the bovine genome sequence, it might be possible to link previously defined QTL for milk and lactation to the bovine genome to provide valuable insight into the molecular evolution of milk and lactation (Lemay *et al*. 2009). Lemay *et al.* (2009) used milk proteome data, thousands of milk protein and mammary genes and 238 milk production QTL, as reported in the literature, to produce a QTL distribution map throughout the whole bovine genome **(Figure 1-1)**. According to the Cattle QTL database (**Cattle QTLdb,** http://www.animalgenome.org/cgi-bin/QTLdb/BT/index), 175 QTL regions for milk yield, 97 for milk fat yield and 87 for milk fat content have been reported (Hu *et al*. 2010). However, the QTL interval on the bovine chromosomes is relatively large and therefore it is difficult to identify causative mutations.

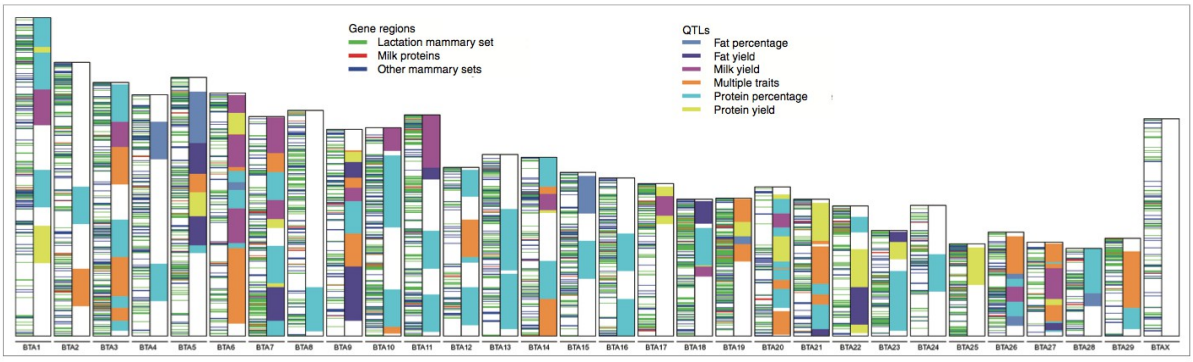

Figure 1-1 Distribution of milk and mammary genes across all bovine chromosomes (Lemay *et al*. 2009).

Although multiple QTL associated with milk fat traits (fat content and fat yield) have been described in cattle, relatively very few QTN have been identified. **Table 1-1** reviews the candidate genes that affect milk fat related traits. Among them, the *DGAT1* gene roles of a key enzyme in triglyceride synthesis (Cases *et al*. 1998) that has a considerable effect on fat content and other milk production traits (Grisart *et al*. 2002; Winter *et al*. 2002). A single coding variant underlying the K232A-substitution, accounts for 31% of the variation in milk fat content in the Holstein-Friesian population (Grisart *et al*. 2002), whereas the others listed in **Table 1-1** have shown minor effects on the milk fat traits.

Table 1-1 Candidate genes reported to affect milk fat traits in cattle.

BTA	Gene	Function	Molecular basis	Reference
5	*OLR1*	Oxidized low-density lipoprotein receptor 1		Khatib *et al*. 2006
6	*ABCG2*	ATP-binding cassette G2	Y581S	Cohen-Zinder *et al*. 2005
6	*SPP1 (or OPN)*	Secreted phosphoprotein-1 (Osteopontin)		Leonard *et al*. 2005
6	*PPARGC 1A*	Peroxysome proliferator-activated receptor-gamma coactivator-1alpha		Weikard *et al*. 2005
14	*DGAT1*	Diacylglycerol O-acyltransferase 1	K232A	Winter *et al*. 2002 Grisart *et al*. 2002
19	*STAT5A*	Signal transducer and activator of transcription 5A		Brym *et al*. 2004
19	*FASN*	Fatty acid synthase		Roy *et al*. 2006
20	*GHR*	Growth hormone receptor	F279Y	Blott *et al.* 2003 Viitala *et al*. 2006
20	PRLR	Prolactin receptor	S18N	Viitala *et al*. 2006
23	*PRL*	Prolactin		Brym *et al*. 2005
26	*SCD*	Stearoyl-CoA desaturase		Moioli *et al*. 2007

3. Dissecting the genes that underlie coat colour in cattle

The diversity of coat colour in mammals has provided abundant materials for scientific research over the last few decades. Genetic studies on mammalian coat colour have not only enriched our understanding of mammalian migration, development, and the regulation of melanocytes, but they have also provided fundamental insights into human evolution, development and diseases. The ability of most wild animals to adapt their coat colour helps to camouflage them in order to avoid predators, to catch their prey, to mate and even to regulate body temperature. Therefore, the evolutionary success of wild animals has

historically allowed them to keep the same coat colours. In contrast to wild animals, the coat colour of domestic animals is highly variable and has been widely adopted as a unique phenotype in the morphological selection that resulted in the constitution of modern breeds. A recent study described the *MC1R* gene as a perfect example for explaining the differences between domestic pigs and wild boars where intriguingly, a number of *MC1R* mutations that appeared in domestic populations were silent among wild counterparts (Fang *et al*. 2009). Fang *et al* suggested that these differences may be caused by either pre-existing variations in their wild ancestor population or by the result of positive selection arising from new mutations with major effects after domestication.

3.1 The study of coat colour in cattle

Hundreds of genes involved in coat colours and patterns so far have been discovered in mice. The most frequently recognized loci are Agouti (A), Brown (B), Albino (C), Dilution (D), Extension (E) and White spotting (W) (Jackson 1994). Farm animals share most of the same loci in that many genes are homologous among mammalian species. Only a few coat colour characteristics in mammals are considered to be monogenic traits, whereas many others are thought to be complex traits due to epistatic effects, epigenetic inheritance, and are influenced by genetic and environmental factors.

Human-mediated selection and domestication in cattle started a long time before animal breeding became a science. Studies on bovine mtDNA phylogeny indicated that the domestication of Zebu cattle, a European taurine breed originating from Indian and Africa, occurred more than 10,000 years ago (Loftus *et al*. 1994; Bradley *et al*. 1996). Early farmers took advantage of the meat, milk, skins and beasts of burden from the wild ancestors of modern cattle in order to fulfil agricultural, economic, cultural and even religious roles (Bradley *et al*. 1996). The process of long-term selection leads to changes in hormone chemistry, physiology, pathology and development in cattle (Slominski *et al*. 2004). These changes alter coat colouration, which becomes a kind of symbol of each breed.

Red and solid black are the two main coat colours observed in cattle. Other coat colours are basically produced by modifications of the three fundamental colours of red, brown-black and black. The cause of most genetic variations is through lightening or the removal of

pigmentation among these three colours, such as in the Limousin, Jersey, Zebu and Brown Swiss breeds. Other mutant genes participate in controlling the dilution of entire body pigmentation, such as in Charolais and Simmental breeds (Olson 1999). Several genes take part in the process of creating complex coat colours and patterns in domestic cattle **(Table 1-2)**.

Table 1-2 The major candidate genes reported to affect coat colour in cattle.

Locus	Gene (BTA)	Phenotype	Breed	Reference
Albinism	*TYR* (29)	Albinism	Braunvieh	Schmutz *et al*. 2004
Brown	*TYRP1* (8)	Dun brown coat colour	Dexter cattle	Berryere *et al*. 2003
Extension	*MC1R* (or *MSHR*) (18)	Red vs. Black dominant	Holstein	Joerg *et al*. 1996
Roan	*MGF* (or *KITLG*) (5)	Roan	Belgian Blue, Shorthorn	Seitz *et al*. 1998
White Spotting	*KIT* (6)	Spotting	German Simmental, German Holstein	Reinsch *et al*. 1999
Agouti	*ASIP* (13)	Agouti	Normande cattle	Girardot *et al*. 2006
Dilution	*PMEL* (or *SILV*) (5)	Coat colour dilution	Charolais	Kühn and Weikard 2007

3.2 Ambilateral circumocular pigmentation in cattle

Ambilateral circumocular pigmentation (ACOP) occurs in some large domestic animals and is characterized by different dome shapes around the eyes. In cattle, this pattern commonly appears in the Simmental and Hereford breeds (Olson 1999). It has been shown that ACOP in cattle can greatly reduce susceptibility to eye lesions that lead to the “cancer eye” (or bovine ocular squamous cell carcinoma, BOSCC), the incidence of BOSCC is increased by the ultraviolet (UV) component of sunlight (Heeney and Valli 1985; Anderson 1991). Anderson (1991) suggested that increasing eyelid pigmentation has resulted in a lower incidence of lesion development in the Hereford breed. Another disease called pink-eye (or Infectious bovine keratoconjunctivitis) is also related to the proportion of eye margin pigmentation. Non-pigmented animals are more susceptible to pink-eye than animals with a higher proportion of eye margin pigmentation (Frisch 1975). A plausible explanation for this is that cattle with a non-pigmented eye margin will be exposed to more UV radiation, especially during the summer months (Ward and Nielson 1979).

German Fleckvieh, a dual purpose breed (dairy and beef), originated as a hybrid of Swiss Simmental and the local cattle breed in southern Germany. This breed features ACOP in different patterns: ambilateral circumocular pigmented on both sides or a single side (left/right) **(Figure 1-2)**. In practice, pigmented skin around the eyes is an important characteristic in Fleckvieh breed selection. Breeders tend to select cows with a higher proportion of pigmented skin around the eyes due to their suitability for importing to tropical or subtropical regions such as southern Africa (www.bavarian-fleckvieh-genetics.de). However, the genetic basis of ACOP in cattle is largely unstudied and additional genetic studies are needed in order to decipher its complexity.

Figure 1-2 Ambilateral circumocular pigmentation in the German Fleckvieh breed (Photo: Dr. Johannes Buitkamp).

Chapter 2

Study of candidate genes in a milk fat content QTL region on BTA5 in the German Holstein-Friesian population

This chapter is based on the following publications:

Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population

Xiaolong Wang[1*], Christine Wurmser[1*], Hubert Pausch[1], Simone Jung[1], Friedrich Reinhardt[2], Jens Tetens[3], Georg Thaller[3], Ruedi Fries[1]

[1] Lehrstuhl für Tierzucht, Technische Universität München, Freising-Weihenstephan, Germany
[2] Vereinigte Informationssysteme Tierhaltung w.V., Verden/Aller, Germany
[3] Institut für Tierzucht und Tierhaltung, Christian-Albrechts-Universität, Kiel, Germany
*These authors contributed equally to this work

Polymorphism analysis of candidate genes in a fat content QTL region on BTA5 in the German Holstein-Friesian population

Xiaolong Wang, Hubert Pausch, Simone Jung, Ruedi Fries

Lehrstuhl für Tierzucht, Technische Universität München, Freising-Weihenstephan, Germany

Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population

Xiaolong Wang[1*], Christine Wurmser[1*], Hubert Pausch[1], Simone Jung[1], Friedrich Reinhardt[2], Jens Tetens[3], Georg Thaller[3], Ruedi Fries[1]

[1] Lehrstuhl für Tierzucht, Technische Universität München, Freising-Weihenstephan, Germany

[2] Vereinigte Informationssysteme Tierhaltung w.V., Verden/Aller, Germany

[3] Institut für Tierzucht und Tierhaltung, Christian-Albrechts-Universität, Kiel, Germany

*These authors contributed equally to this work

Abstract

Milk composition traits exhibit a complex genetic architecture with a small number of major quantitative trait loci (QTL) explaining a large fraction of the genetic variation and numerous QTL with minor effects. In order to identify QTL for milk fat percentage (FP) in the German Holstein-Friesian (HF) population, a genome-wide association study (GWAS) was performed. The study population consisted of 2327 progeny-tested bulls. Genotypes were available for 44,280 SNPs. Phenotypes in the form of estimated breeding values (EBVs) for FP were used as highly heritable traits. A variance components-based approach was used to account for population stratification. The GWAS identified four major QTL regions explaining 46.18% of the FP EBV variance. Besides two previously known FP QTL on BTA14 (P=8.91 x 10^{-198}) and BTA20 (P=7.03 x 10^{-12}) within *DGAT1* and *GHR*, respectively, we uncovered two additional QTL regions on BTA5 (P=2.00 x 10^{-13}) and BTA27 (P=9.83 x 10^{-5}) encompassing *EPS8* and *GPAT4*, respectively. *EPS8* and *GPAT4* are involved in lipid metabolism in mammals. Re-sequencing of *EPS8* and *GPAT4* revealed 50 polymorphisms. Genotypes for five of them were inferred for the entire study population. Two polymorphisms affecting potential transcription factor binding sites of *EPS8* (P=1.40 x 10^{-12}) and *GPAT4* (P=5.18 x 10^{-5}), respectively, were highly significantly associated with the FP EBV. Our results provide evidence that alteration of regulatory sites is an important aspect of genetic variation of complex traits in cattle.

Introduction

Improvement of milk yield and composition is a major objective of dairy cattle breeding programs and highly reliable breeding values are estimated to this end. Milk composition traits such as protein and fat content are not only important production traits but also permit insights into the metabolic constitution of lactating cows (Buttchereit *et al*. 2010).

Milk fat content, indicated as fat percentage (FP), is a prototypical complex quantitative trait determined by numerous loci with small effects and only few loci with major effects (Hayes *et al*. 2010). Family-based linkage studies and genome-wide association studies (GWAS) have already identified several genomic regions contributing to the genetic variation of FP in cattle (*e.g.* Schennink *et al*. 2009; Cohen-Zinder *et al*. 2005; Jiang *et al*. 2010; Hayes *et al*. 2010). Among them, most prominently, a K232A-substitution within the acyl-CoA:diacylglycerol acyltransferase encoding gene *DGAT1* (Grisart *et al*. 2002; Winter *et al*. 2002) and a F297Y-substitution within the growth hormone receptor encoding gene *GHR* (Viitala *et al*. 2006; Blott *et al*. 2003) have been well characterized. These two polymorphisms account for a major fraction of the genetic variation of FP in various cattle breeds (*e.g.* Thaller *et al*. 2003; Schennink *et al*. 2007; Signorelli *et al*. 2009).

The aim of the present study was to identify major quantitative trait loci (QTL) for FP in the German HF population by genome-wide association analysis and to pinpoint the causal variants. The study revealed four QTL, two of them novel. We report candidate genes and putative causal variants for the newly identified QTL.

Material and methods

Animals and phenotypes

The study population consisted of 2401 progeny-tested Holstein Friesian bulls. The animals descend from 376 different sires and 423 maternal grand-sires. The paternal half-sib families and maternal grandsire families encompass up to 83 members with an average of six members. Phenotypes in the form of estimated breeding values (EBVs) for milk FP were obtained from vit w.V. Verden (www.vit.de, April 2010 version). Breeding value estimation for FP was carried out using best linear unbiased prediction (BLUP).

Genotypes and quality control

The HF bulls were genotyped with the Illumina Bovine SNP 50K BeadChip® comprising 54,001 single nucleotide polymorphisms (SNPs). Of 2401 genotyped animals, 62 were removed from the data set because genotyping failed for more than 10% of the SNPs. The chromosomal positions of 53,452 SNPs were according to the University of Maryland UMD3.1 assembly of the bovine genome sequence (Zimin *et al*. 2009). 549 SNPs with unknown chromosomal position and 7951 SNPs with minor allele frequencies < 0.01 were excluded. 732 SNPs that were missing in more than 10% of the animals and 966 SNPs that deviated significantly from the Hardy-Weinberg equilibrium (P< 0.001) were omitted from subsequent analyses. The genomic relationship of each pair of animals was obtained as proposed by VanRaden (2008) and was compared with the corresponding pedigree relationship. Twelve animals showed major differences between pedigree and genomic relationship and were excluded from further analysis. The final data set comprised 2327 animals and 44,280 SNPs. Sporadically missing genotypes were imputed using default parameters of *Beagle* (version 3.2.1) (Browning and Browning 2009).

Genome-wide association study

To account for population stratification, *EMMAX* (Kang *et al*. 2010) was used to fit the model $y = Xb + u + \epsilon$, where y is a vector of EBVs for FP, b is the SNP effect, X is a design matrix of SNP genotypes, u is the random polygenic effect with $(0, \sigma_a^2 G)$, where σ_a^2 is the additive genetic variance, G is the genomic relationship matrix (GRM) among the 2327 animals (see above) and ϵ is the non heritable component of the random variation.

Chromosomal partitioning of the genetic variance

In order to estimate the proportion of EBV variance attributable to a particular chromosome and QTL, a GRM was built (see above) for each of the 30 chromosomes and the four QTL regions separately. A QTL was defined by the SNPs within a 5 Mb interval centered on the most significantly associated SNP. Chromosome-specific variance was estimated based on GRMs that included all SNPs on a chromosome except those within the 5 Mb QTL interval.

We used *GCTA* (Yang *et al*. 2011) to fit the model $y = \sum_{i=1}^{34} g_i + e$, where y is a vector of

EBVs for FP, g is a vector of genetic effects attributed to the ith chromosome/QTL, and e is a vector of random residual deviates. g_i is assumed to be normally distributed with $N(0, G_i\sigma^2_{gi})$, where G_i is the GRM built based on SNPs of the ith chromosome/QTL. The proportion of variance attributable to the ith chromosome/QTL was calculated as $\sigma^2_i/(\sum_{j=1}^{34}\sigma^2_j+\sigma^2_e)$.

Candidate gene annotation

The *GenomeThreader* software tool (Gremme *et al*. 2005) was used to predict the genomic structure and localization of the genes based on the University of Maryland UMD3.1 assembly of the bovine genome sequence (Zimin *et al*. 2009) and the Dana-Farber Cancer Institute bovine gene index release 12.0 (Quackenbush *et al*. 2001) together with the annotated RNA sequences of the UMD3.1 assembly. The *GenomeThreader* output was viewed and edited using the *Apollo Genome Annotation and Curation Tool* (Lee *et al*. 2009). The exon-intron organization of the annotated genes is summarized in **Table S1**. Transcription factor binding sites were predicted with *MatInspector* (Cartharius *et al*. 2005), *JASPAR* (Portales-Casamar *et al*. 2010) and *TESS* (Schug 2008). Prediction of microRNAs was carried out using *MicroCosm Targets* (http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/) and *TargetScanHuman 6.0* (http://www.targetscan.org/).

Re-sequencing of candidate genes

Candidate genes for the BTA5 and BTA27 FP QTL were re-sequenced in 24 and 12 HF animals, respectively. The animals were selected based on their genotypes for the most significantly associated SNP of the particular QTL. PCR primers were designed for the promoter regions (3000 bp and 1500 bp upstream of the transcription start site for *EPS8* and *GPAT4*, respectively), for all exons and intron-exon boundaries as well as for the 5' and 3' untranslated regions (UTR) (**Appendix 2**). The PCR products were sequenced using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) on the ABI 3130xl Genetic Analyzer (Applied Biosystems). The *Phred/Phrap/Polyphred* software suite (Nickerson *et al*. 1997) was used for base calling, sequence alignment and polymorphism identification, and *consed* (Gordon et al. 1998) was used for viewing.

Genotyping of selected polymorphisms

Three previously proposed quantitative trait nucleotides (QTN) for FP, namely the ones responsible for the K232A-substitution within *DGAT1* (Grisart *et al*. 2002, Winter *et al*. 2002), the F297Y-substitution within *GHR* (Blott *et al.* 2003) and the S18N-substitution within *PRLR* (Viitala *et al.* 2006) were obtained by TaqMan® genotyping analysis (Applied Biosystems Applera, Darmstadt, Germany). Potentially functional polymorphisms for the BTA5 and BTA27 QTL were genotyped in 2327 animals of our study population as well **(Table S3)**. Sporadically missing genotypes were imputed using default parameters of *Beagle* (version 3.2.1) (Browning & Browning 2009).

Results

Association study

The genome-wide association study based on 44,280 SNPs and 2327 progeny-tested bulls identified four QTL for FP on BTA5, BTA14, BTA20 and BTA27 in the German HF population (**Figure 2-1**).

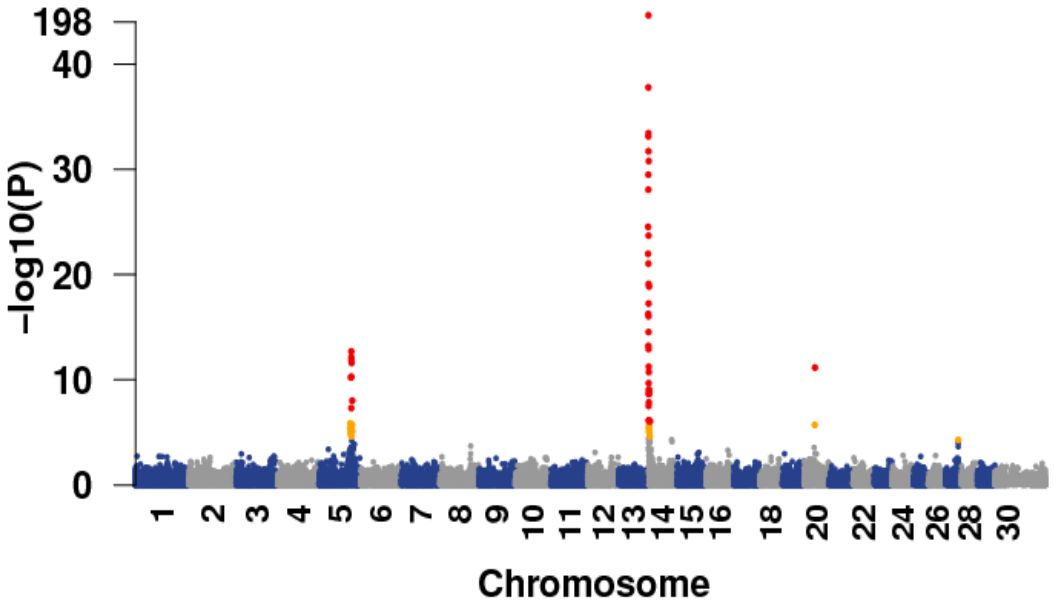

Figure 2-1. Association of 44,280 SNPs with the estimated breeding values for milk fat percentage in 2327 animals of the German Holstein-Friesian population. Red symbols represent SNPs with $P < 1.3 \times 10^{-6}$ (Bonferroni-corrected significance level), orange symbols indicate chromosome-wide significance.

Forty-eight significantly associated SNPs encompass the DGAT1 encoding gene (1.46 Mb – 7.31 Mb) and define the QTL region on BTA14. The most significantly associated SNP resides 1149 bp upstream of the postulated *DGAT1*-QTN (Grisart *et al*. 2002; Winter *et al*. 2002). Genotyping of the K232A-substitution within *DGAT1* showed close to complete linkage disequilibrium with the SNP ARS-BFGL-NGS-4939 (r^2=0.998). The genotypes of these two SNPs differ only in two animals, most likely due to imperfect genotype imputation, resulting in marginally different *P*-values **(Figure 2-2a)**.

On BTA20, a single SNP (BTA-84181-no-rs) located 357,004 bp downstream of the postulated causal F297Y-substitution within the GHR encoding gene (Blott *et al*. 2003) was significantly associated (P =1.95 x 10^{-6}). The association signal for the F297Y-substitution was more prominent than for any other SNP on BTA20 (P=7.03 x 10^{-12}) **(Figure 2-2b)**. BTA-84181-no-rs SNP is only moderately linked with the F297Y-substitution (r^2=0.2). We additionally obtained genotypes for the S18N-substitution within the PRLR encoding gene (Viitala *et al*. 2006), another presumed causal variant. The S18N-substitution is 7.56 Mb distant from the significantly associated BTA-84181-no-rs SNP. No association of the *PRLR*-variant (P=0.319) with the FP EBV was observed.

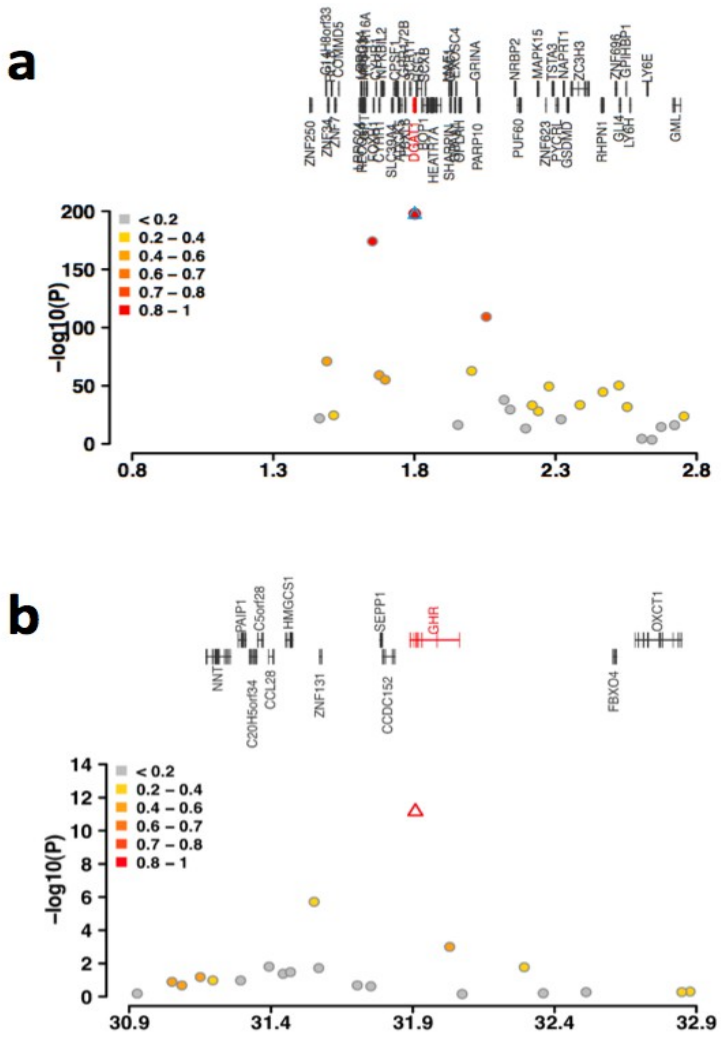

Figure 2-2. Detailed view of two genomic regions within known QTL for milk fat content in cattle. QTL regions on BTA14 **(a)** and BTA20 **(b)** encompassing *DGAT1* and *GHR*, respectively. Open symbols (blue and red triangles) represent the K232A- and the F297Y-substitution, respectively. Different colors indicate the extent of linkage disequilibrium (r^2) between the postulated QTN and all other SNPs.

Seven genome-wide significantly associated SNPs located between 91.2 Mb and 97.1 Mb delineate the QTL region on BTA5. The most significantly associated SNP Hapmap49734-BTA-74577 (P=2.00 x 10^{-13}) is located in the second intron of the epidermal growth factor receptor pathway substrate 8 encoding gene (*EPS8)* (**Table 2-1**) **(Figure 2-3)**.

Table 2-1. Characteristics of the most significantly associated 50K Illumina BeadChip SNPs and additional polymorphisms of four major QTL for milk fat percentage in the German Holstein-Friesian population.

SNP	Chr.	Physical position [a]	Minor allele (MAF)	*P* [b]	Neighboring gene
Hapmap49734-BTA-74577	5	94,570,828	A (0.09)	2.00×10^{-13}	*EPS8*
ARS-BFGL-NGS-4939	14	1,801,116	G (0.31)	1.57×10^{-198}	*DGAT1*
BTA-84181-no-rs	20	31,552,475	G (0.37)	1.95×10^{-6}	*GHR*
ARS-BFGL-NGS-57448	27	36,155,097	A (0.36)	9.83×10^{-5}	*GPAT4*
ss319604831	5	94,551,792	G (0.21)	4.92×10^{-6}	*EPS8*
ss319604833	5	94,553,580	T (0.09)	1.40×10^{-12}	*EPS8*
ss319604845	5	94,726,848	T (0.22)	2.40×10^{-5}	*EPS8*
K232A	14	1,802,265	A (0.31)	8.91×10^{-198}	*DGAT1*
F297Y	20	31,909,479	A (0.17)	7.03×10^{-12}	*GHR*
S18N	20	39,115,344	G (0.14)	3.19×10^{-1}	*PRLR*
ss410759404	27	36,211,257	GA (0.38)	5.18×10^{-5}	*GPAT4*
ss410758894	27	36,228,939	A (0.40)	2.27×10^{-4}	*GPAT4*

[a] The SNPs are ordered according to their position on the UMD3.1 assembly of the bovine genome sequence.

[b] The *P*-values are obtained after regression analysis and by using a variance components based approach to account for population stratification.

On BTA27, SNP ARS-BFGL-NGS-57448 is associated with $P = 9.83 \times 10^{-5}$ and thus with chromosome-wide significance only. It is located about 56-kb upstream of the glycerol-3-phosphate acyltransferase 4 encoding gene (*GPAT4)* (**Table2-1) (Figure 2-4)**.

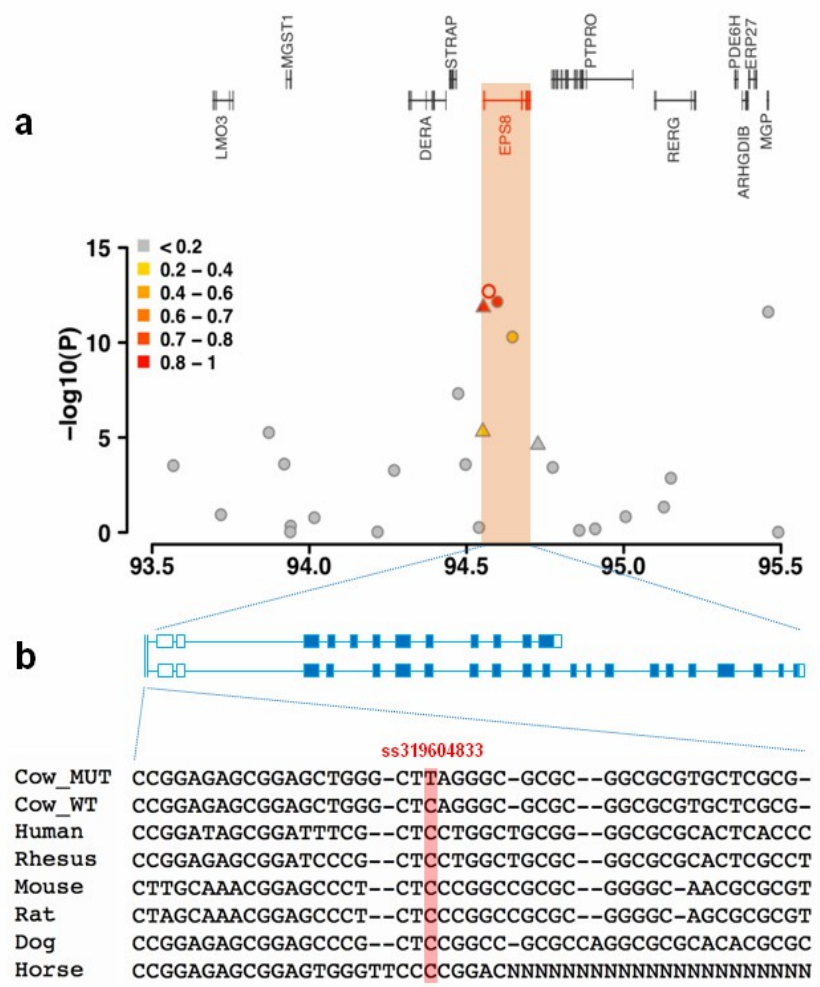

Figure 2-3. Schematic view of the BTA5 QTL region encompassing *EPS8*. **(a)** The open symbol represents the most significantly associated SNP. Different colors indicate the extent of linkage disequilibrium (r^2) between the most significantly associated SNP and all other SNPs. Triangles and circles indicate SNPs resulting from re-sequencing and genotyping with the 50K BeadChip, respectively. **(b)** Two alternative transcripts of *EPS8* are present in cattle. The multispecies sequence alignment of a segment in the promoter region of *EPS8* encompassing the highly significantly associated ss319604833 polymorphism (red background) illustrates the high conservation among species.

Assessing the impact of the four QTL

Alleles increasing the FP EBVs were identified for the most significantly associated SNP for each of the four identified QTL regions. The frequency distribution of animals with an increasing number of alleles is displayed in **Figure 2-5.** The EBVs of animals with one and seven FP increasing alleles differ by more than three standard deviations.

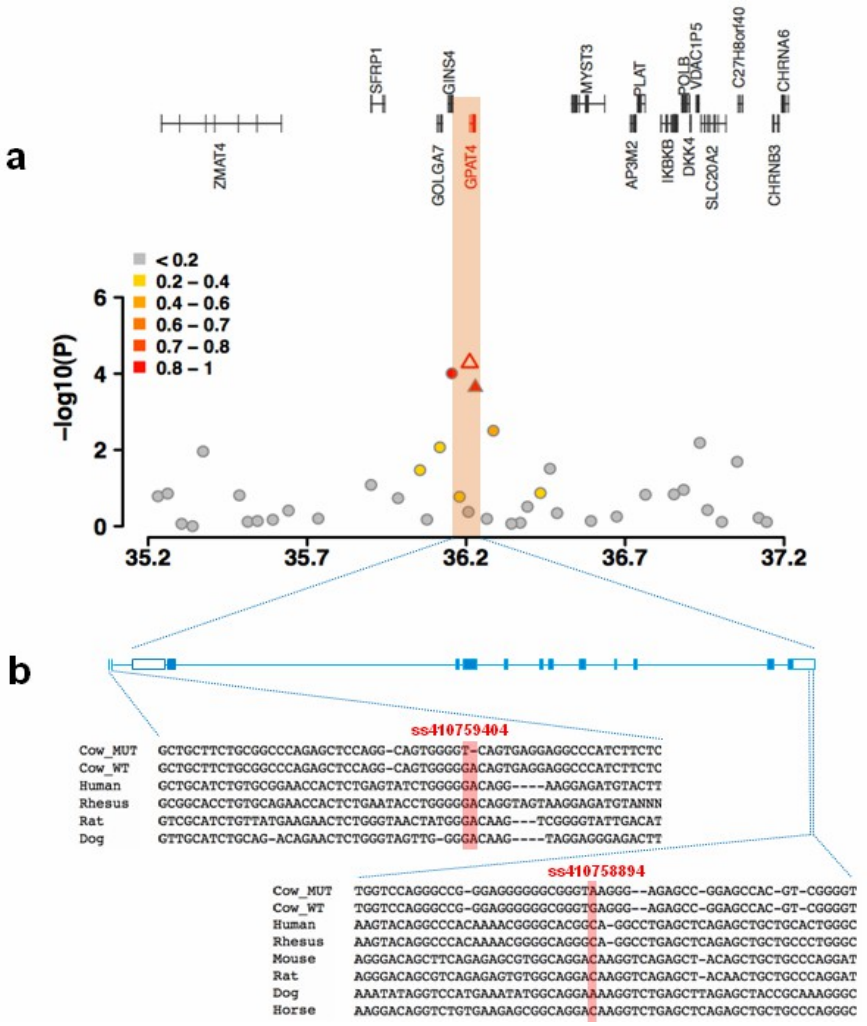

Figure 2-4. Schematic view of the BTA27 QTL region encompassing *GPAT4*. **(a)** The open symbol represents the most significantly associated SNP. Different colors indicate the extent of linkage disequilibrium (r^2) between the most significantly associated SNP and all other SNPs. The triangles indicate SNPs resulting from re-sequencing, circles indicate SNPs from genotyping with the 50K BeadChip. **(b)** Gene structure of *GPAT4* and multispecies sequence alignment of the promoter encompassing the highly significantly associated ss410759404 and of the 3'UTR encompassing the highly significantly associated ss410758894, respectively.

Chromosomal partitioning of the EBV variance

The proportion of the EBV variance attributed to a particular chromosome/QTL was estimated with the effects of all chromosomes/QTL fitted simultaneously. Totally, the 44,280 SNPs account for 85.97% of the EBV variance. The contribution of particular chromosomes

varies strongly (**Figure 2-6**). A major fraction of the EBV variance is attributable to BTA14 (33.60%), BTA5 (12.08%) and BTA20 (7.01%). BTA27 accounts for a minor fraction (1.19%) of the EBV variance only. Totally, the four identified QTL explain 46.18% of the FP EBV variance. The estimates of the EBV variance attributable to the four QTL are 8.35% (BTA5), 31.04% (BTA14), 5.91% (BTA20) and 0.88% (BTA27).

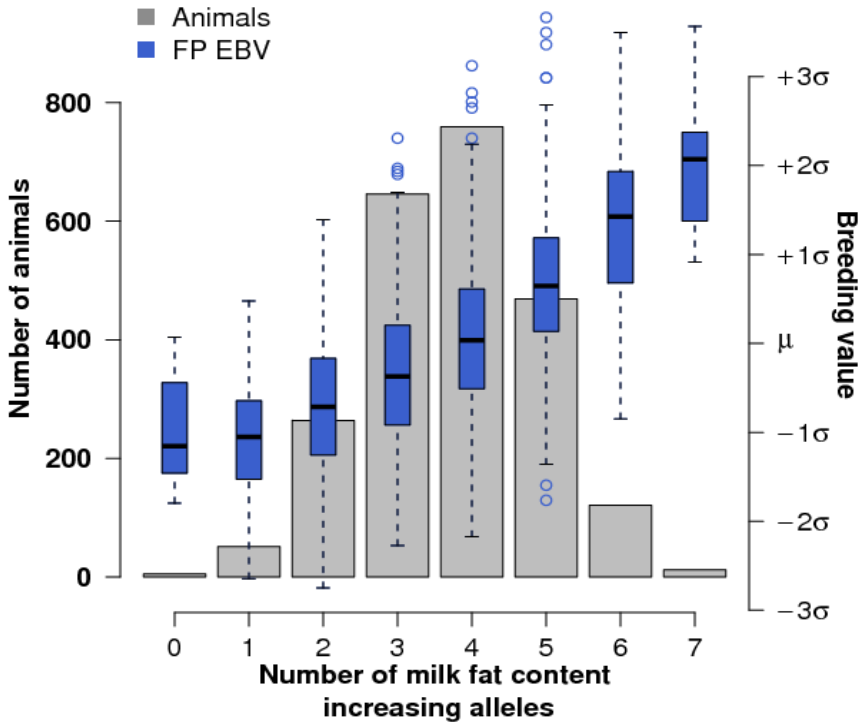

Figure 2-5. The combined impact of the four identified QTL on BTA5, 14, 20 and 27 on the estimated breeding value for milk fat percentage in the German Holstein-Friesian population. 2327 Holstein-Friesian animals are grouped according to the number of alleles that increase the milk FP EBV. The grey bars indicate the number of animals with an increasing number of FP increasing alleles. The box plots represent the FP EBVs for each group.

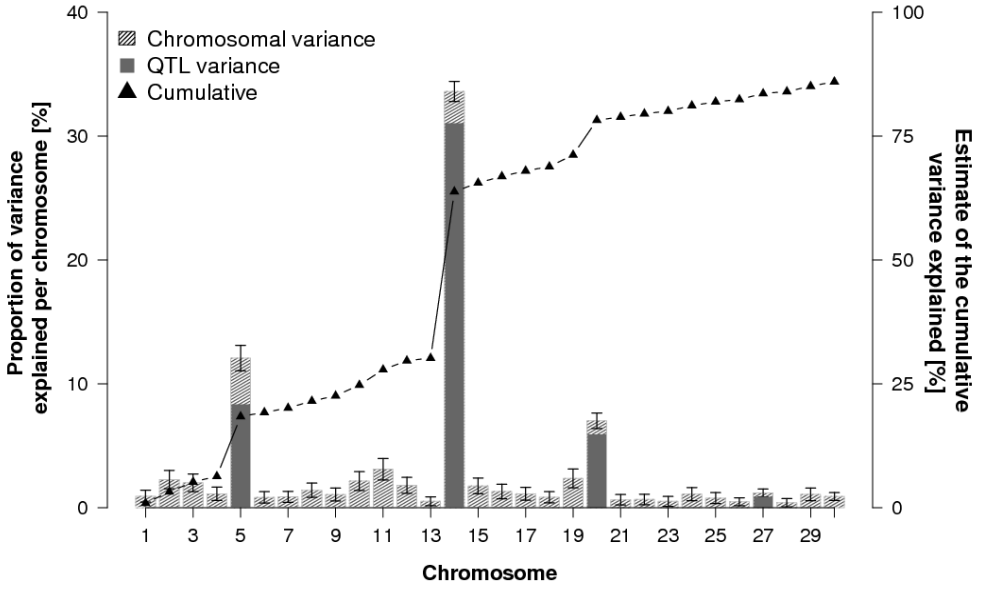

Figure 2-6. Partitioning of the genetic variance onto 30 chromosomes and four identified QTL regions on BTA5, 14, 20 and 27. The grey shaded bars indicate the fraction of EBV variance attributed to a particular chromosome and the corresponding standard error. The dark grey bars represent the fraction of EBV variance attributed to each of the four identified QTL regions. The black triangles represent the cumulative proportion of EBV variance explained.

Molecular-genetic analysis of the BTA5 QTL

EPS8 was considered as a positional candidate gene for the FP QTL region on BTA5 as the two most significantly associated SNPs were located within its second intron. *EPS8* plays a role in the fat metabolism of mammals as it is a substrate for the EGFR-kinase (Fazioli *et al*. 1993). The EGFR-kinase activates transcriptional regulators of fatty acid synthesis and thereby increases the concentration of intracellular fatty acids (Guo *et al*. 2009). Re-sequencing of *EPS8* in 24 animals resulted in the identification of 20 polymorphisms (**Table S4**). We genotyped two promoter polymorphisms (ss319604831 and ss319604833) in high LD (r^2>0.8) with SNP Hapmap49734-BTA-74577 and for one non-synonymous mutation (M599T, ss319604845) located in a highly conserved region of *EPS8* (**Table 1, Figure 3**). However, only ss319604833, located 100 bp upstream of the transcription start of *EPS8*, was highly associated with the FP EBV (P=1.40 x 10^{-12}) (**Table 1**, **Figure 2-3).** The significance level of the putative causal variant ss319604833 was marginally lower compared to the SNP

Hapmap49734-BTA-74577 (P=2.00 x 10^{-13}). This could be due to imperfect genotype imputation.

ss319604833 is located in a region of *cis*-acting regulatory elements. Five transcription factors whose binding may be affected by the polymorphism were predicted. The promoter variant with the more frequent C allele, associated with a lower FP EBV, contains a potential MEF3-element. In contrast, TBF1-, Ptx1-, MafB- and TFAP2A-elements were predicted for the sequence with the rare T allele, associated with an increased FP EBV **(Figure 2-7a)**. The expression of transcription factor TFAP2A is significantly correlated with the concentration of nonesterified fatty acids (NEFA) and liver triacylglycerol (Loor *et al*. 2005).

Molecular-genetic analysis of the BTA27 QTL

The FP QTL region on BTA27 encompasses the *GPAT4* gene. *GPAT4* encodes the rate-limiting enzyme glycerol-3-phosphate acyltransferase in the triacylglycerol biosynthesis pathway and plays a key role in milk fat biosynthesis (Bionaz & Loor 2008). Re-sequencing of the exons and regulatory flanking regions in a panel of twelve animals revealed 30 polymorphisms. Genotypes for two polymorphisms (ss410759404 and ss410758894) located within the 5'flanking region and within the 3'UTR, respectively, were obtained for the entire study population (**Table 2-1**) **(Figure 2-4)**. ss410759404 (GA>T), was more strongly associated with the FP EBV (P =5.18 x 10^{-5}) than any other SNP on BTA27.

ss410759404 is located 1378 bp upstream of the translation start of *GPAT4*. Transcription factors SREB, CREB, RXR-α and RAR-β were predicted for the sequence with the frequent T allele, which is associated with a lower FP EBV. The PAX6-element was predicted for the sequence with the minor GA allele, that was associated with an increased FP EBV. **(Figure 2-7b)**. RXR-α up-regulates genes involved in fatty acid and lipid metabolism during the process of adipogenesis (Hamza et al. 2009), CREB takes part in the regulation of gluconeogenesis (Chakravarty *et al*. 2005) and SREB is considered to be one of the central regulation factors in milk fat synthesis (Bionaz and Loor 2008).

The prediction of putative miRNA binding sites revealed that the binding of mmu-miR-712 might be affected by the ss410758894 polymorphism, with the minor A allele showing less

similarity to the miRNA-binding site **(Figure 2-7c)**. However, no conserved sites for miRNA families could be identified due to low conservation between species. Also, the *P*-value for ss410758894 (P =2.27 x 10^{-4}) did not meet the significance level.

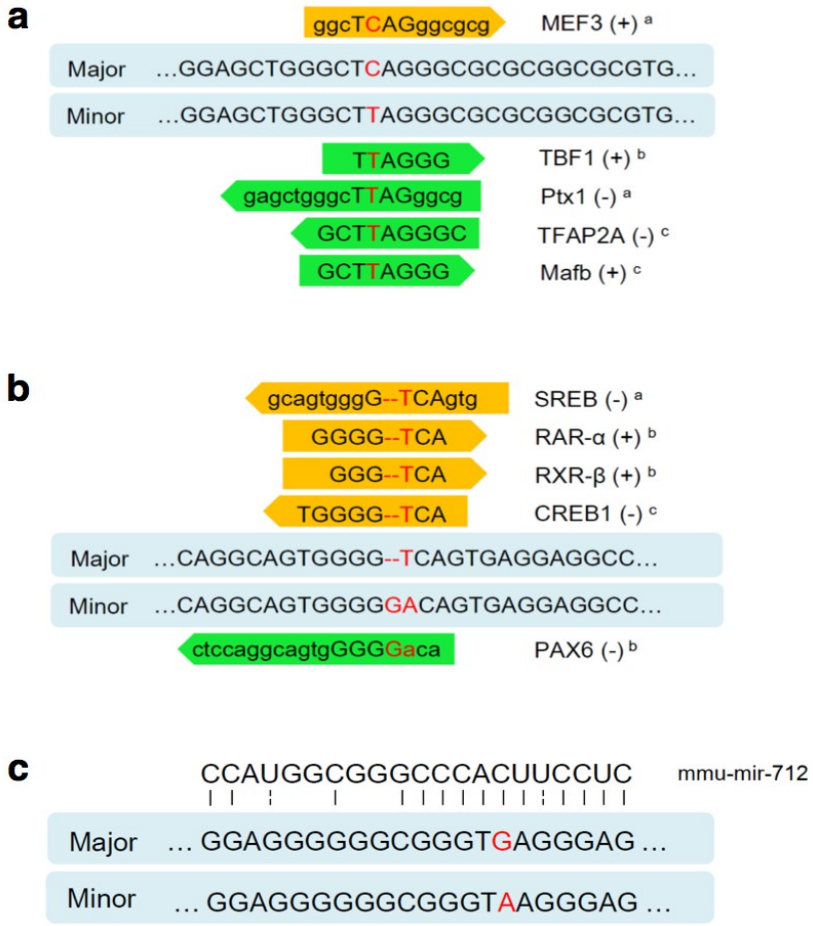

Figure 2-7. Prediction of regulatory sites for the polymorphisms in *EPS8* and *GPAT4*. Prediction of transcription factor binding sites within the promoter sequences of *EPS8* and *GPAT4* encompassing the SNP ss319604833 **(a)** and the SNP ss410759404 **(b)**. (+) and (-) indicate forward and reverse direction. *MatInspector*[a], *TESS*[b] and *JASPAR*[c] were used for the prediction of transcription factor binding sites. **(c)** Prediction of a miRNA binding site for mmu-mir-712 (free energy: -27.93 kcal/mol) within the 3'UTR of *GPAT4* encompassing the polymorphism ss410758894.

Discussion

Our genome-wide association study was based on a medium-sized sample of the German HF population and on a dense SNP map. It revealed four major QTL for FP. The four identified QTL regions account for a large part of the EBV variance (46.18%). However, a large

fraction of the EBV variance is attributable to chromosomes with no identified QTL, *e.g.* BTA11 and BTA19. The QTL region on BTA14 encompassing *DGAT1* accounts for 31.04% of the EBV variance which agrees with previous findings (Thaller *et al*. 2003). The QTL region on BTA20 encompassing the F297Y-substitution within the GHR encoding gene on BTA20 (Blott *et al*. 2003) accounts for 5.91% of the EBV variance. There was no evidence for a second FP QTL on BTA20 resulting from the S18N-substitution within the *PRLR* encoding gene (Viitala *et al*. 2006). We were able to identify two additional QTL regions for FP in the German HF population on BTA5 and BTA27 that together account for 9.23% of the EBV variance. Our findings support the proposed genetic architecture of FP with numerous loci with small effects and only few loci with larger effects (Hayes *et al*. 2010).

The presence of a FP QTL on BTA5 agrees with findings in the Australian HF population (Hayes *et al*. 2010). We identified a highly significantly associated polymorphism in the promoter region of *EPS8* which is supposed to mediate the binding of TFAP2A and concomitantly the transcription rate of *EPS8*. EPS8 physically interacts with the epidermal growth factor receptor (Fazioli *et al*. 1993). Recently it has been shown that sterol regulatory element-binding proteins (SREBPs) are regulated by the epidermal growth factor (Chatterjee *et al*. 2009). SREBPs control the expression of genes required for the uptake and synthesis of cholesterol, fatty acid and triglycerides. Thus, it seems likely that an enhanced transcription rate of *EPS8*, conferred by binding of TFAP2A, results in an increased milk fat biosynthesis in the lactating mammary gland. Therefore, a contribution of ss319604833 to the genetic variation of milk fat synthesis seems plausible.

Recently, Bouwman and colleagues (Bouwman *et al*. 2011) reported a QTL contributing to the genetic variation of milk fatty acid composition in the Dutch HF population nearby *GPAT4,* supporting our findings of a FP QTL in the German HF population on BTA27. *GPAT4* plays a crucial role in lipid biosynthesis in mammals (Bionaz and Loor 2008). The transcription rate of *GPAT4* is highly correlated with the concentration of milk diacylglycerols and triacylglycerols (Beigneux *et al*. 2006; Vergnes et al. 2006). Prediction of transcription factor binding sites for the highly significantly associated SNP ss410759404 suggests that the binding of transcription factors involved in fat metabolism might be affected by this variant (Bionaz and Loor 2008; Chakravarty *et al*. 2005; Hamza *et al*. 2009).

Hence, a contribution of ss410759404 to the milk fat biosynthesis capacity in lactating cows seems likely. Association analysis uncovered a second SNP (ss410758894) affecting a potential miRNA binding site of *GPAT4*. Although ss410758894 did not meet the criteria for significant association, it is possible that both variants contribute to the genetic variation of the BTA27 FP QTL in the German HF population.

Acknowledgements

This study is part of the projects FUGATO-plus GenoTrack and FUGATO-plus MeGA-M and was financially supported by the German Ministry of Education and Research, BMBF (grants 0315134A, 0315134D and 0315131A), the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven. X. Wang is supported by a grant from China Scholarship Council (CSC) for his Ph.D study.

Author contributions

Conceived and designed the experiments: XW CW HP RF. Performed the experiments: XW CW HP. Analyzed the data: HP XW CW. Contributed reagents/materials/analysis tools: SJ FR JT GT. Wrote and revised the paper: XW CW HP RF.

Polymorphism analysis of candidate genes for a fat content QTL region on BTA5 in the German Holstein-Friesian population

Xiaolong Wang, Hubert Pausch, Simone Jung, Ruedi Fries

Lehrstuhl für Tierzucht, Technische Universität München, Freising-Weihenstephan, Germany

Abstract

We have previously carried out a genome-wide association study (GWAS) to elucidate candidate genes underlying the genetic associations in milk fat percentage (FP) in the German Holstein-Friesian (HF) population. The most highly significant QTL on chromosome 5 accounted for >8% of the trait variation. In order to identify the potential causative variant, 16 positional candidate genes (*LMO3, MGST1, DERA, STRAP, EPS8, PTPRO, RERG, ARHGDIB, PDE6H, ERP27, H2AFJ, MGP, ART4, GUCY2C, ATF7IP* and *GRIN2B*) within this QTL region were characterised and further re-sequenced in twelve selected animals representing three breeds (Holstein-Friesian, Fleckvieh and Braunvieh). In total, 300-kb sequences were screened and resulted in the detection of 338 polymorphisms, including eleven non-synonymous mutations. We decided to determine the genotypes of three putative functional variants (P47Q in *MGST1*, M599T in *EPS8*, E6A in *GRIN2B*) in 750 animals, and used genotype imputation to determine the genotypes of 2327 animals. However, none of the three non-synonymous SNPs could be associated with FP at the genome-wide significance level in the HF breed.

Keywords: Fat percentage, re-sequencing, SNP, association

Introduction

Milk fat is an important economic trait for dairy cattle, because it contributes unique characteristics to the flavour and nutritional properties of milk (Haug *et al*. 2007). The biosynthesis and regulation of milk fat in the bovine mammary gland are complicated processes in which a network of genes participates (Bionaz and Loor 2008). Previous observations have investigated several genes contributing to milk fat content including the

K232A-substitution within the *DGAT1* gene, which explains 31% of the variation in fat yield and composition in Holstein-Friesian (HF) cattle (Grisart *et al*. 2002; Winter *et al*. 2002; Thaller *et al*. 2003). Moreover, a number of SNPs in other candidate genes (*e.g. GHR*, *ABCG2*, *FASN*, *OLR1*, *PPARGC1A*, *PRL* and *STAT5A*) conferring changes in fat percentage and/or fat yield have been reported in different cattle populations (Blott *et al.* 2003; Brym *et al*. 2004, 2005; Cohen-Zinder *et al*. 2005; Dybus *et al*. 2005; Weikard *et al*. 2005; Khatib *et al*. 2006; Roy *et al*. 2006; Viitala *et al*. 2006; Morris *et al*. 2007).

Three recent genome-wide association studies (GWAS) have independently mapped a highly significant QTL affecting milk fat content on bovine chromosome (BTA) 5 in the HF breed (Hayes *et al*. 2010; Pimentel *et al.* 2011; Wang *et al.* 2012). However, candidate genes within this QTL region and their molecular basis remain unknown. The objectives of this study were to elucidate the genetic basis of candidate genes, to detect polymorphisms and to evaluate associations between the identified variants and milk fat content in the HF population.

Materials and Methods

Animals and phenotypes

Phenotypes in the form of estimated breeding values (EBVs) for fat content, indicated as fat percentage (FP), were obtained from 2401 progeny-tested HF bulls by vit w.V. Verden (www.vit.de, April 2010 version). Breeding value estimation for FP was carried out using best linear unbiased prediction (BLUP).

Candidate gene annotation

The *GenomeThreader* software tool (Gremme *et al*. 2005) was used to predict the genomic structure and localization of the candidate genes based on the University of Maryland UMD3.1 assembly of the bovine genome sequence (Zimin *et al*. 2009) and the Dana-Farber Cancer Institute bovine gene index release 12.0 (Quackenbush *et al*. 2001), together with the annotated RNA sequences of the UMD3.1 assembly. The *GenomeThreader* output was viewed and edited using the *Apollo* sequence annotation editor (Lee *et al*. 2009).

Re-sequencing of candidate genes

A DNA panel of twelve animals representing three breeds (Holstein-Friesian, Fleckvieh and Braunvieh) was used for polymorphism detection. PCR primers were designed for the promoter region (3000 bp upstream of the transcription start site), for all exons and intro-exon boundary regions as well as 5′ and 3′ gene flanking regions (primer sequences are summarized in **Appendix 2**). The PCR products were purified, and sequenced using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI 3130xl Genetic Analyzer (Applied Biosystems) following the protocol shown in the **Appendix 5**. The *Phred/Phrap/Polyphred* software suite (Nickerson *et al*. 1997) was used for sequence dissection, and the sequences were viewed with *consed* (Gordon *et al*. 1998). *Haploview* software was used to discover linkage disequilibrium (LD) regions and to determine tag SNPs (Barrett *et al*. 2005). Functional significance of the coding non-synonymous variants was predicted by *PANTHER* (Thomas and Kejariwal 2004) and *PolyPhen-2* (Adzhubei *et al*. 2001) methods.

Genotyping of selected polymorphisms

Genotypes of three potentially functional SNPs **(**P47Q in *MGST1*, M599T in *EPS8* and E6A in *GRIN2B*) on BTA5 were determined in 750 animals of our study population by TaqMan® genotyping assays (Applied Biosystems Applera, Darmstadt, Germany). *Beagle* 3.0.1 (Browning and Browning 2009) was used to impute the genotypes in a total of 2327 animals.

Results

Genomic structure

Inspection of the bovine-human comparative gene maps revealed 16 protein-coding genes within a 3.0 Mb genomic region (BTA5:93.5-96.5 Mb, bovine genome assembly UMD3.1) surrounding the most significantly associated SNP Hapmap49734-BTA-74577 derived from Wang *et al*. (2012). These genes were: *LMO3, MGST1, DERA, STRAP, EPS8, PTPRO, RERG, ARHGDIB, PDE6H, ERP27, H2AFJ, MGP, ART4, GUCY2C, ATF7IP* and *GRIN2B*. Genomic size and annotation information are listed in **Appendix 1**.

Polymorphism screening in the QTL region on BTA5

In order to detect the polymorphisms, approximately 300-kb genomic sequences, including

all exons, intron-exon boundaries, alternative promoter regions and 5' and 3' UTRs of these 16 positional candidate genes, were re-sequenced in a panel of 12 animals from the three breeds (Holstein-Friesian, Fleckvieh and Braunvieh). In total, 338 variants were identified comprising 30 insertion/deletion polymorphisms (indels) and 308 SNPs **(Table 2-2)**. The polymorphism data have been submitted to the NCBI Single Nucleotide Polymorphism Database (dbSNP) and are included in **Appendix 3**. Of the 338 variants, 42 SNPs were located in the coding regions, eleven SNPs lead to non-synonymous amino acid exchanges **(Table 2-3)**.

Table 2-2 Localization of the identified sequence polymorphisms.

Gene	Polymorphisms in coding regions		Polymorphisms in non-coding regions			Total
	Synonymous	Non-synonymous	5' end	Intronic region	3' end	
LMO3	0	0	3	5	0	8
MGST1	3	1	18	5	1	28
DERA	2	0	8	9	1	20
STRAP	0	0	11	5	11	27
EPS8	1	1	7	17	1	27
PTPRO	2	4	6	16	1	29
RERG	3	0	8	1	16	28
ARHGDIB	0	0	2	10	2	14
PDE6H	0	0	8	0	2	10
ERP27	0	1	3	13	3	20
H2AFJ	0	0	2	0	1	3
MGP	0	0	8	0	2	10
ART4	1	0	5	0	0	6
GUCY2C	4	2	15	25	1	47
ATF7IP	1	0	3	11	2	17
GRIN2B	13	2	9	20	0	44
Total	30	11	116	137	44	338

The 5’ end includes 5’ UTR and its upstream region, and the 3’ end contains 3’ UTR and its downstream sequence.

Association study of selected polymorphisms

Of these eleven non-synonymous SNPs, three of which responsible for the P47Q-substitution within *MGST1*, the M599T-substitution within *EPS8* and the E6A-substitution within *GRIN2B* were determined in 2327 animals according to their high multispecies conservation

(Figure 2-8a-c) and/or the function of the genes. MGST1, as a member of the MAPEG superfamily, contains a fatty acid/phospholipid substrate binding site (Busenlehner *et al*. 2007) and is capable of catalysing a glutathione-dependent reduction of certain lipid hydroperoxides such as fatty acid hydroperoxides (Morgenstern and DePierre 1983; Mosialou *et al*. 1995). The product of *MGST1* is predominantly localised to the outer mitochondrial membrane and the endoplasmic reticulum, the site of lipid synthesis, and is presumed to protect these membranes from oxidative stress (Siritantikorn *et al*. 2007; Johansson *et al*. 2010). It is likely that this protective function of MGST1 affects the lipid synthesis capacity in the mammary gland. EPS8 plays a role in the fat metabolism of mammals as it is a substrate for EGFR-kinase (Fazioli *et al*. 1993). The EGFR-kinase activates transcriptional regulators of fatty acid synthesis and thereby enhances the concentration of intracellular fatty acids (Guo *et al*. 2009). However, no significant correlations were found between these three non-synonymous mutations and FP EBVs (**Table 2-4**).

Table 2-3 Amino acid substitutions caused by eleven non-synonymous SNPs found in the re-sequenced genes.

Gene	Non-synonymous SNPs	Protein residue	Location	Amino acid property change	Evolutionary analysis	
					PANTHER	*PolyPhen-2*
MGST1	c.140A>C	P47Q	Exon 2	Nonpolar/polar	0.4905	0.434, benign
EPS8	c.1796C>T	M599T	Exon 18	No change	0.8529	0.000, benign
PTPRO	c.874C>G	E292Q	Exon 5	Polar, negatively charged/uncharged	0.8384	0.701, possibly damaging
	c.896G>T	L299W	Exon 5	No change	0.7807	0.000, benign
	c.964A>G	E322K	Exon 5	Polar, negatively charged/positively charged	0.9323	0.520, possibly damaging
	c.2581A>G	I861V	Exon 16	No change	0.8564	0.001, benign
ERP27	c.398A>G	R133H	Exon 4	No change	0.8812	0.024, benign
GUCY2C	c.47A>G	R16Q	Exon 12	Polar, positively charged/uncharged	-	0.000, benign
	c.59C>T	P20L	Exon 14	No change	-	0.000, benign
GRIN2B	c.17A>G	E6A	Exon 1	Polar, negatively charged/nonpolar, uncharged	0.2406	0.000, benign
	c.3694A>G	T1232A	Exon 12	Polar/nonpolar	-	0.000, benign

The output of *PANTHER,* $P_{deleterious}$, refers to the probability that a given coding variant will cause a deleterious effect on protein function. $P_{deleterious}$ is continuous values from 0 (neutral) to 1 (most likely to be deleterious). - means the position does not align exactly with a reference location. The *PolyPhen-2* software reports a score ranging from 0 (neutral) to 1 (damaging).

Figure 2-8 Multispecies protein sequence alignment for the three genotyped sequence variants. **(a)** P47Q (ss319604543) in the MGST1 protein. **(b)** M599T (ss319604845) in the EPS8 protein. **(c)** E6A (ss319604793) in the GRIN2B protein.

Table 2-4 Association of the three non-synonymous SNPs with milk FP EBV in the HF population.

SNP (Amino acid exchange)	Physical position [a]	Minor allele (MAF)	P value	Gene
ss319604543 (P47Q)	93,939,231	A (0.13)	9.37×10^{-1}	*MGST1*
ss319604845 (M599T)	94,726,848	T (0.22)	2.40×10^{-5}	*EPS8*
ss319604793 (E6A)	96,408,820	C (0.25)	2.70×10^{-4}	*GRIN2B*

[a] The SNPs are ordered according to their position on the UMD3.1-assembly of the bovine genome sequence.

Discussion

Although various studies have proposed milk fat related QTL regions on BTA5 (Kolbehdari *et al*. 2009; Hayes *et al*. 2010; Awad *et al.* 2010, 2011; Jiang *et al*. 2010; Mai *et al*. 2010; Pimentel *et al.* 2011; Wang *et al.* 2012), and three of them (Hayes *et al*. 2010; Pimentel *et al*. 2011; Wang *et al*. 2012) have independently mapped the same locus (Hapmap49734-BTA-

74577 at position 94,570,828 bp, UMD3.1 assembly) as the second most significant QTL for milk fat content in the Australian and German HF populations, none of the candidate genes in this QTL have been investigated so far. In this study, 16 positional candidate genes derived from this QTL region were re-sequenced in an attempt to investigate the genetic basis for its impact on the milk fat content. The candidate gene analysis revealed 338 variants by re-sequencing 300-kb of putative functional genomic regions (promoters, coding regions, intron-exon boundaries and UTR). However, there was no significant association of the three selected non-synonymous SNPs with the FP EBV in the HF breed.

The genomic structures of 16 selected candidates were computationally predicted by using comparative maps of human-bovine sequence similarity. However, the existing genomic sequences are incomplete, making the genomic structures uncertain, and are not helpful in identifying bovine candidate genes accurately, especially in the non-coding regions. Moreover, it is difficult to define the position of polymorphisms because homology-based annotation yields multiple alternatively spliced transcripts. Thus, inappropriate annotation of relevant genes may mislead the identification of genomic variants.

To assess the variants in an efficient and cost effective manner, three non-synonymous mutations were prioritized for genotyping because they may positively or negatively affect the functions of protein products. However, none of the three selected non-synonymous SNPs could be associated with milk fat percentage at a genome-wide level. The inability to detect a putative functional variant might be because of the limited number of animals used in the variation detection phase to capture all potential polymorphisms. Moreover, with the growing evidence for genetic effects by non-coding variants, the functional variants responsible for fat content may be located in non-protein-coding regions (5' UTR including promoters, 3' UTRs, and introns or intragenic regions) or in intergenic regions. For instance, the candidate polymorphisms in the 5'-flanking regions or introns of genes may influence transcription rate and eventually affect the expression of protein products; intronic variants may result in truncated protein products by splice variants; regulatory variants in 3' UTRs may also regulate expression by modulating microRNAs (Ibeagha-Awemu *et al*. 2008). After excluding two non-synonymous SNPs from the two most likely candidate genes (*MGST1* and *EPS8*), we assume that the causative variant is rather located in the non-coding regulatory regions of one of these two genes.

In summary, this study was unable to detect a genetic variant that is likely to be responsible for milk fat content in the HF population. The QTN that may be included in the list of identified polymorphisms of present study, further studies are therefore required to determine the actual functional variant.

Chapter 3
Study of ambilateral circumocular pigmentation related genes in the German Fleckvieh population

This chapter is based on the following publications:

Polymorphism analysis of candidate genes associated with ambilateral circumocular pigmentation in the German Fleckvieh population

Xiaolong Wang, Hubert Pausch, Michal Wysocki, Simone Jung, Ruedi Fries

Chair of Animal Breeding, Technische Universität München, Freising-Weihenstephan, Germany

Differential expression of the *KIT* gene in pigmented and non-pigmented bovine hair follicles

Xiaolong Wang, Daniel Brugger, Hubert Pausch, Ruedi Fries, Michal Wysocki

Chair of Animal Breeding, Technische Universität München, Freising-Weihenstephan, Germany

Polymorphism analysis of candidate genes associated with ambilateral circumocular pigmentation in the German Fleckvieh population

Xiaolong Wang, Hubert Pausch, Michal Wysocki, Simone Jung, Ruedi Fries

Chair of Animal Breeding, Technische Universität München, Freising-Weihenstephan, Germany

Abstract

Ambilateral circumocular pigmentation (ACOP) is characterized by different dome shapes around the eyes. ACOP occurs in some large domestic animals such as the cattle breeds Simmental and Hereford. Previous investigations showed that ACOP can reduce susceptibility to some carcinogenic agents, which are induced, to a certain degree, by the ultraviolet light component of sunlight. We previously localised twelve ACOP-associated QTL regions using a genome-wide scan in 3579 German Fleckvieh (FV) bulls (Pausch *et al*. 2012). In the present study, five positional and functional genes (*PAX3, KIT, KDR, GSDMC* and *MITF*) from four loci, which are related to melanocyte development or eye development, were prioritized for further investigation. 180 polymorphisms were identified by re-sequencing 76.5-kb genomic sequences from 16 selected FV bulls. A single nucleotide change in exon 8 is capable of encoding a substitution, T242M in PAX3, strongly affecting the progeny-derived ACOP phenotype ($P = 3.14 \times 10^{-5}$). This missense variant, which encodes part of the transactivation domain and is located in a highly conserved region, most likely represents the causal mutation of the proportion of daughters with ACOP in the FV population.

Keywords: ambilateral circumocular pigmentation, Fleckvieh, coat colour, PAX3

Introduction

Studies of coat colours and patterns in both natural populations and domesticated animals have provided valuable insights into molecular genetics and considerably enriched our understanding of biology, evolution and diseases, in humans and animals. Ambilateral circumocular pigmentation (ACOP) features different dome shapes surrounding the eyes in

some large domestic animals. In cattle, it appears frequently in the Simmental and Hereford breeds (Olson 1999). The German Fleckvieh (FV) breed, a dual purpose breed originating from the Simmental breed in Switzerland, exhibits a typical ACOP phenotype, which shown in **Figure 1-2**.

Previous findings indicated that the amount of ACOP is correlated with the incidence of ocular squamous cell tumours in cattle (Bailey *et al*. 1990). ACOP can reduce susceptibility to eye lesions, which lead to "cancer eye", also known as bovine ocular squamous cell carcinoma (BOSCC), the incidence of which is increased by ultraviolet (UV) radiation in sunlight (Anderson 1991; Tsujita and Plummer 2010). It has been reported that approximately 82% of all bovine tumours at slaughter are caused by "cancer eye" (Russell *et al*. 1956). A plausible explanation for this is that ambilateral circumocular unpigmented cattle are exposed to a greater amount of UV radiation (Ward and Nielson 1979), which makes them more susceptible to carcinogenesis (Tran *et al*. 2008).

In order to understand the genetic architecture involved in forming ACOP in the FV population, a genome-wide association analysis was previously performed using high-density SNP panels (>650,000 SNPs) for 3579 FV bulls (Pausch *et al*. 2012). Twelve QTL regions were mapped to distinct regions and their positional candidate genes were also presented. To further characterize these loci, five candidate genes (*PAX3*, *KIT*, *KDR*, *GSDMC* and *MITF*) from four loci were selected in an attempt to identify further mutations that could account for each QTL region by using a candidate gene re-sequencing approach.

Materials and Methods

Study samples

The proportion of daughters with ACOP were provided by the Bavarian State Research Center for Agriculture (www.lfl.bayern.de) for 320,186 FV cows, sired by 3579 progeny tested bulls. Genomic DNA panels comprising 16 FV bulls (eight animals with the highest proportion of ACOP from their daughters and eight with the lowest) were constructed for DNA re-sequencing (**Appendix 6**).

Gene annotation

Five positional and functional genes, *PAX3* (BTA2), *KIT* (BTA5), *KDR* (BTA5), *GSDMC* (BTA14) and *MITF* (BTA22), as well as a 5.6-kb conserved region upstream of the *KDR* gene, were chosen for DNA re-sequencing. *GenomeThreader* software (Gremme *et al*. 2005) was used to determine the genomic structure and localization of the candidate genes based on the University of Maryland UMD3.1 assembly of the bovine genome sequence (Zimin *et al*. 2009) and the Dana-Farber Cancer Institute bovine gene index release 12.0 (Quackenbush *et al*. 2001) together with the annotated RNA sequences of the UMD3.1 assembly. The *GenomeThreader* output was viewed and edited using the *Apollo* sequence annotation editor (Lee *et al*. 2009).

Gene re-sequencing and variation screening

All of the alternative transcripts of the five genes (*PAX3*, *KIT*, *KDR*, *GSDMC* and *MITF*) including exons, intro-exon boundary regions, 3000 bp upstream of the 5'UTR, along with 1000 bp downstream of the 3'UTR, were re-sequenced using selected DNA samples. The primers were designed using Primer 3 software (http://frodo.wi.mit.edu/primer3/) and are summarized in **Appendix 2**. The PCR products spanning regions of interest were purified, and sequenced using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI 3130xl Genetic Analyzer (Applied Biosystems) following the protocol shown in the **Appendix 5**. The *Phred/Phrap/Polyphred* software suite (Nickerson *et al*. 1997) was used for sequence dissection, and sequences were viewed with *consed* (Gordon *et al*. 1998). *Haploview* software was used to discover LD regions and to determine tag SNPs (Barrett *et al*. 2005). Functional consequences of the non-synonymous SNPs were predicted by *PANTHER* (Thomas and Kejariwal 2004) and *PolyPhen-2* (Adzhubei *et al*. 2010) methods.

Single SNP genotyping

The SNP encoding a T424M-substitution in *PAX3* was genotyped in 1100 animals by TaqMan genotyping assay (Applied Biosystems Applera) applied on an ABI7500 system (Applied Biosystems). Genotype imputation was performed in a total of 3675 animals using *Beagle* 3.0.1 (Browning and Browning. 2009).

Results

Annotation of candidate genes

Candidate genes correlated with melanocyte development and its pathway, or eye development, were prioritized for re-sequencing. Five positional and functional genes (*PAX3, KIT, KDR, GSDMC* and *MITF),* including the three (*KIT, KDR* and *MITF*) accounting for the two major QTL regions, were selected for the re-sequencing study.

Previous investigations showed that these five genes are related to coat colour in cattle or other mammals. The *KIT* gene is essential for proliferation and migration of melanoblasts during embrogenesis, and subsequently for the development and maintenance of pigmentation (Hou *et al*. 2000). Several mutations in *KIT* have been reported to increase the incidence of different pigmentation patterns in animals and have shown distinct pleiotropic effects. In cattle, the spotting locus was mapped on BTA6 in the region comprising the *KIT* gene (Reinsch *et al*. 1999), and the alleles of *KIT* were found to vary in different cattle breeds (Olson 1999; Fontanesi *et al*. 2010a). It is known that the *s* (spotted) allele of spotting loci is related to recessive spotting in the Simmental breed but the causative mutation remains unknown (Olson 1999). The *KDR* (*VEGFR-2*) gene is a candidate gene with critical roles in controlling eye development. The inhibition of *VEGFR-1* and *-2* prevents retinal neovascularization and, therefore, *KDR* and its receptors are crucial for normal neural development in the retinal (Robinson *et al*. 2001). Mutations in mice *MITF* were found to cause coat colour dilution, white spotting and a completely white coat due to the inadequacy of melanocytes (review by Steingrímsson *et al*. 2004). In cattle, the white spotting locus and proportion of black coat colour were recently mapped to a region on BTA22 near *MITF* (Liu *et al*. 2009; Hayes *et al*. 2010). The *PAX3* gene is correlated with melanocytic proliferation, apoptosis resistance, migration, lineage specificity and differentiation (Kubic *et al*. 2008). The *GSDMC* gene, belongs to the gasdermin (Gsdm) family, and was initially isolated from B16 melanoma sublines in humans and thought to be a marker of melanoma progression, as well as roles as a tumour suppressor (Watabe *et al*. 2001; Saeki *et al*. 2009). The genes in the Gsdm family are expressed in a tissue-specific manner, such as in skin epithelium and the gastrointestinal tract (Tamura *et al*. 2007), and thus *GSDMC* may play important roles in skin development. However, the exact function of *GSDMC* in coat pigmentation remains unclear.

Gene annotation with *GenomeThreader* software revealed the genetic structures of the selected bovine genes: *PAX3*, with three transcripts, harbours 4, 8 and 9 exons, and encodes 215, 479 and 484 amino acids respectively; *KIT* consists of 15 exons and encodes 977 amino acids, *KDR* possesses 30 exons and encodes 1356 amino acids; *GSDMC* harbours 15 exons and encodes 487 amino acids; while *MITF*, with three transcripts, harbours 9, 10, 10 exons and encodes 410, 510 and 526 amino acids, respectively. Genomic size and annotation information are listed in **Appendix 1**.

Genomic sequencing of the candidate genes

To search for polymorphisms associated with the proportion of daughters with ACOP, approximately 76.5-kb of the genomic sequences were re-sequenced using established DNA panels. In total, 180 polymorphisms (150 SNPs and 30 insertions and deletions) were identified (**Table 3-1**, **Appendix 3)**. Among them, 17 SNPs were located in the coding regions, 71 in the promoter regions, five in the 5' untranslated region (UTR) and 12 in the 3'UTR. Ten SNPs indicated non-synonymous amino-acid exchanges (**Table 3-2)**, and their functional effect were predicted using *PANTHER* and *PolyPhen-2*. Two variants, the T424M within PAX3 and the L27P within GSDMC, may strongly influence their protein function due to a higher deleterious effect. However, the substitution of the L27P in GSDMC has not change the property of amino acid.

Table 3-1 Localization of the identified sequence polymorphisms.

Gene	Polymorphisms in coding regions		Polymorphisms in non-coding regions			Total
	Synonymous	Non-synonymous	5' end	Intronic region	3' end	
PAX3	1	1	10	5	5	22
KIT	0	0	0	2	0	2
KDR	5	3	23	28	15	74
GSDMC	0	6	26	17	3	52
MITF	1	0	17	5	7	30
Total	7	10	76	57	30	180

The 5’ end includes 5’ UTR and its upstream region, and the 3’ end contains 3’ UTR and its downstream sequence.

Table 3-2 Amino acid substitutions caused by ten non-synonymous SNPs found in the re-sequenced genes.

Gene	Non-synonymous SNPs	Protein residue	Location	Amino acid property change	Evolutionary analysis	
					PANTHER	*PolyPhen-2*
PAX3	c.1271C>T	T424M	Exon 8	Polar/non-polar	0.8556	0.726, possibly damaging
KDR	c.1075C>T	P359S	Exon 8	Nonpolar/polar	0.6153	0.003, benign
	c.1870A>G	I624V	Exon 13	No change	-	0.001, benign
	c.1882A>G	E628K	Exon 13	No change	-	0.267, benign
GSDMC	c.80C>T	L27P	Exon 2	No change	0.9892	0.967, probably damaging
	c.490A>G	V164I	Exon 4	No change	0.4013	0.032, benign
	c.599A>T	V200E	Exon 5	Nonpolar/polar, negatively charged	0.3178	0.015, benign
	c.788A>G	K263R	Exon 8	No change	0.3198	0.006, benign
	c.930A>G	M310I	Exon 12	No change	0.7315	0.410, benign
	c.1121A>G	Y374C	Exon 14	No change	0.8049	0.083, benign

The output of *PANTHER,* $P_{deleterious}$, refers to the probability that a given coding variant will cause a deleterious effect on protein function. $P_{deleterious}$ is continuous values from 0 (neutral) to 1 (most likely to be deleterious). - means the position does not align exactly with a reference location. The *PolyPhen-2* software reports a score ranging from 0 (neutral) to 1 (damaging).

Subsequently, *Haploview* software was used to select the tag SNPs which in high LD (r^2>0.9) with the most significantly associated SNP of each QTL region. This analysis revealing negative evidences among the polymorphisms in *KIT*, *KDR* and its 5' flanking region, *GSDMC* and *MITF*, respectively, implying that the causative mutations did not existed in the regions studied, or not highly linked with the most significantly associated SNP. However, two SNPs (ss475875628 in the promoter region and ss475875642 in the exon 8) in *PAX3,* were found to be in perfect LD (r^2=1) with the most significant SNP on BTA2 (UA-IFASA-5029) in the 16 animals **(Figure 3-1).**

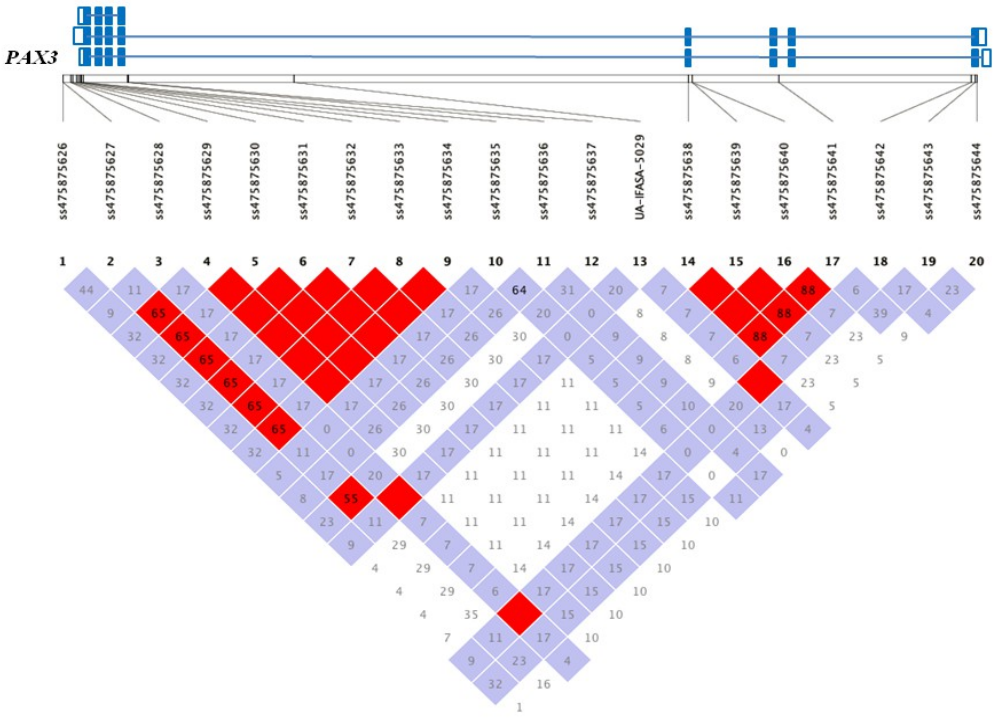

Figure 3-1 Schematic of the exonic structure and linkage disequilibrium plot of *PAX3*. The rectangles show the approximate size of the exons of *PAX3*; the coding regions are coloured blue. The SNP ss475875628 and the SNP ss475875642 (T424M) are highly linked with the most significant SNP (UA-IFASA-5029). Each diamond indicates the level of LD (r^2), the plot was generated using *Haploview* (Barrett *et al.* 2005). The intensity of the colours corresponds to the levels of r^2.

Analysis of the missense variant T424M in *PAX3*

We have totally identified 22 variants in the *PAX3* gene using 16 selected individuals (**Table 3-3)**. The non-synonymous SNP, c.1271C>T, occurs in exon 8 of *PAX3* at position 111,129,964 bp (UMD3.1 assembly) and leads to a missense polar threonine to a non-polar methionine encoding Thr424Met. The functional effect of this variant was evaluated using *PANTHER* and *PolyPhen-2,* all showing a relatively higher score which indicating that the amino-acid substitution alters the structure of protein. The rare T allele (TT and CT genotypes) of this low-frequency variant (MAF=0.04) is responsible for a higher proportion of ACOP from progenies than C allele (CC genotype) **(Figure 3-2)**.

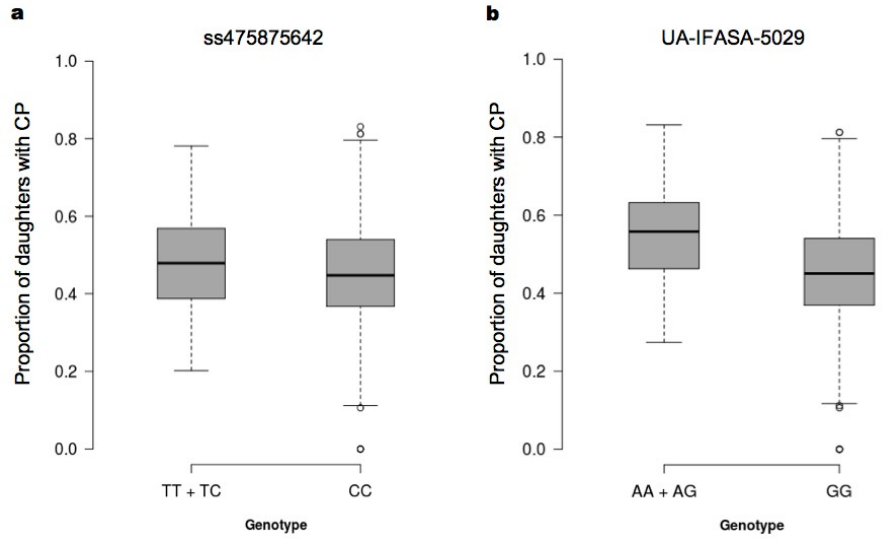

Figure 3-2 The proportion of daughter with ACOP by genotypes for ss475875642 (p.T424M) **(a)** and UA-IFASA-5029 (the most significantly associated SNP on BTA2) **(b)**.

Additionally, this missense mutation is located in the COOH-terminal transactivation domain of the PAX3 protein and in a region of high interspecies conservation **(Figure 3-3)**. The transactivation domain is generally thought to activate transcription by contacting transcription factors. It has been reported that the COOH-terminal transactivation domain in PAX3 affects DNA recognition specificity by regulating the activity of another domain – the homeodomain (Cao and Wang 2000). PAX3 is a crucial transcription factor for transactivating the promoter of the *KIT* gene (Guo *et al.* 2010), as well as for promoting and inhibiting melanogenesis through the transcriptional regulation of MITF, DCT, and TYRP1 (Kubic *et al*. 2008). We regarded this variant (T424M) as being of great importance since it may influence the transactivation of its target gene, and tested the association between T424M and the proportion of daughters with ACOP phenotype.

The missense mutation T424M was located in the transactivation domain and is marked in red background. Genotyping of p.T424M was performed in 1100 FV bulls and imputed in a total of 3645 animals. A significant association was observed ($P = 3.14 \times 10^{-5}$), although not as significant as the most SNP UA-IFASA-5029 ($P = 2.49 \times 10^{-9}$) **(Figure 3-4)**.

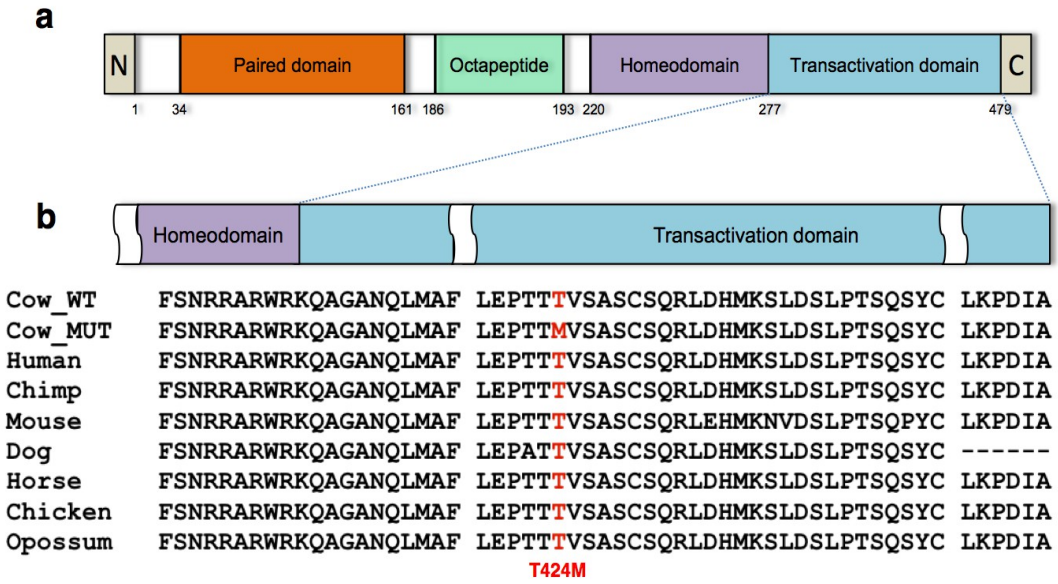

Figure 3-3 Domain structure and multispecies sequence alignment of the PAX3 protein. **(a)** The domain structure of the PAX3 protein. PAX3 possesses four domains: the paired domain, the octapeptide motif, the homeodomain and the transactivation domain (Kubic *et al.* 2008). Each domain is shown by a coloured box with the domain name indicated. The NH_2-terminus (N) and the COOH-terminus (C) of the protein sequences are also indicated. **(b)** Multiple protein sequence alignment of PAX3.

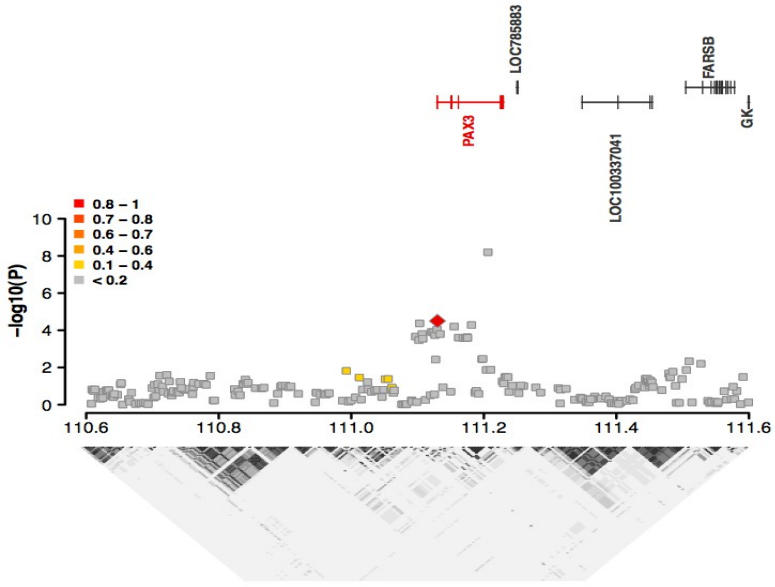

Figure 3-4 Evidence for the association of the p.T424M variant in the *PAX3* locus. The red diamond indicates the non-synonymous mutation p.T424M (ss475875642). Bonferroni-corrected threshold for the genome-wide significance ($P < 7.59 \times 10^{-8}$).

Discussion

In search of functional variants, the five genes selected (*PAX3, KIT, KDR,GSDMC* and *MITF*) were re-sequenced, including the two most significant QTL regions on BTA6 and BTA22, which explained 18.29% and 12.53% of the phenotypic variation, respectively (Pausch *et al*. 2012). Although the molecular basis of these genes in relation to certain coat colours in animals has been determined, re-sequencing of the candidate genes only revealed a missense variant in *PAX3* for explaining the likely causality of the ACOP phenotype derived from daughters.

KIT and *KDR* Among the genetic associations for ACOP, the QTL region on BTA6, including *KIT* and *KDR,* reached the most significant level ($P = 3.83 \times 10^{-54}$) (Pausch *et al*. 2012). However, both *KIT* and *KDR* are functional candidates, making it difficult to determine which gene contain the actual QTN. The *KIT* gene is responsible for many coat colour traits, especially the spotting locus in farm animals such as horses, dogs, pigs and cattle (Hagen *et al*. 2004; Brooks and Bailey. 2005; Brooks *et al*. 2007; Fontanesi *et al*. 2010a; Fontanesi *et al*. 2010b). Surprisingly, only two intronic SNPs were identified by re-sequencing the *KIT* gene. Another study discovered 111 *KIT* polymorphisms throughout the entire coding region in three cattle breeds (Angus, Hereford and Holstein) (Fontanesi *et al*. 2010a), suggesting that the investigated regions in the current study are highly homozygous in the FV breed. Moreover, it has been reported recently that the colour sidedness phenotype in the Belgian blue and brown Swiss cattle is caused by translocated segments between BTA6 (including *KIT*) and BTA29 (Durkin *et al*. 2012), implying that in regulatory variants even structural variations may contribute to various coat colour phenotypes in cattle.

Previous studies indicated that the *KDR* gene correlates with vascular and ocular development (Robinson *et al*. 2001; Gogat *et al*. 2004), and its functional role in melanocyte development still elusive (Aoki *et al*. 2009). Of the 74 polymorphisms identified in *KDR*, most were found in a high LD block but none was linked with the most significant SNP. Re-sequencing of two additional highly conserved DNA elements located in the 5' flanking region of *KDR* also failed to reveal any potential SNPs for causality. Since the importance of *KIT* in coat colour pigmentation has been shown in various species (reviewed by Hofreiter and Schöneberg 2010), and the differential expression of *KIT* in the follicles of pigmented

and non-pigmented hair, we speculated that the QTN for this locus might exist in the regulatory region of *KIT* rather than *KDR*, or that it might be located in their intergenic regions.

GSDMC The *GSDMC* (or *MLZE*) gene is the most likely candidate for the locus on BTA14 (Pausch *et al*. 2012). As a tumour suppressor gene (Saeki *et al*. 2009), *GSDMC* is up-regulated in growing metastatic melanomas and is thought to be liable for melanoma progression (Watabe *et al*. 2001). We suggest that instead of being involved in the development of eye pigmentation directly, *GSDMC* might predispose cattle to the "cancer eye". In this study, a total of 52 polymorphisms were identified and most of them were linked each other, and no polymorphisms were highly linked with the most associated SNP.

MITF *MITF* has been proved to be a master regulator for the development of melanin-containing pigment cells. This study investigated the polymorphisms within three alternative *MITF* transcripts. None of the polymorphisms showed high LD (r^2 >0.8) with the most significant SNP. We postulate that the causal variant does not exist in the region evaluated, or it is situated in a region that was not properly annotated. Moreover, an intronic SNP in the bovine *MITF-M* isoform that indicated significant differences between two spotted breeds (Italian Holstein and Italian Simmental) and two solid coloured breeds (Italian Brown and Reggiana) (Fontanesi *et al*. 2012), suggesting that intronic variants are important contributors to complex traits, especially for some large effect QTL (Keane *et al*. 2011).

Notably, microRNA (*e.g.* miR-25; miR-137, miR-148 and miR-340) in *MITF* also play important roles in modulating melanoma cells or skin melanocytes (Bemis *et al*. 2008; Goswami *et al*. 2010; Haflidadóttir *et al*. 2010; Zhu *et al*. 2010). Zhu *et al* (2010) concluded that miR-25 regulates brown coat colour versus white in alpacas (*Lama pacos*) by influencing the expression of microRNAs, suggesting that variants in 3'UTR or intronic regions that affect miRNA binding should also be considered. However, we carefully re-sequenced the 3'UTR of *MITF* and identified seven polymorphisms, but none of them showed high LD values with the most associated SNP.

PAX3 PAX3, a key player in melanogenesis, is a transcription factor that binds a proximal

region of the *MITF* promoter (Bondurand *et al*. 2000). It has been shown that PAX3, together with MITF, regulates Waardenburg syndrome, which is a hereditary disorder with sensorineural deafness and pigmentation defects in the hair, skin and iris (Watanabe *et al*. 1998). However, there is also evidence showing that PAX3 interacts with another important receptor tyrosine kinase (KIT), and transactivates the promoter of the *KIT* gene in an indirect manner (Guo *et al*. 2010).

The unique *PAX3* missense mutation, T424M, encodes part of the transactivation domain and the functional connections between PAX3 and KIT or MITF. We speculated that the transcription factor PAX3 regulates the transcription of KIT or MITF by binding to a specific DNA element. However, genotyping of this low-frequency variant (MAF=0.04) in 1100 animals and follow-up genotype imputation in a total of 3645 animals demonstrated a less significant P value (3.14×10^{-5}) in comparison with the most associated SNP ($P = 2.49 \times 10^{-9}$). This could be due to the genotype imputation bias of low-frequency variants, because the majority of homozygous genotypes for rare SNPs make it difficult to impute them (Asimit and Zeggini 2010). Bickel *et al* (2011) reported an abdominal pigmentation study in *Drosophila melanogaster*, in which many small-effect polymorphisms in three distinct functional regions of a locus contributed to a large phenotypic effect. It is plausible there might be other small-effect variants (*e.g*. in the regulatory regions of *PAX3*) together with this missense variant T424M that are responsible for cumulative effect of the QTL underlying the ACOP phenotype.

In summary, the re-sequencing of five candidate genes with a panel of animals identified a number of candidate polymorphisms that might include functional variants underlying these loci. Additionally, a low-frequency missense variant (p.T424M) in *PAX3* was found to be strongly associated with progeny-derived ACOP. This variant encodes part of the transactivation domain and resides in a highly conserved region, making it likely to be responsible for the proportion of daughters with ACOP in the FV breed.

Differential expression of the *KIT* gene in pigmented and non-pigmented bovine hair follicles

Xiaolong Wang, Daniel Brugger, Hubert Pausch, Ruedi Fries, Michal Wysocki

Chair of Animal Breeding, Technische Universität München, Freising-Weihenstephan, Germany

Summary

c-kit is a receptor for the mast/stem cell growth factor and plays a role in melanogenesis, erythropoiesis, spermatogenesis and T-cell differentiation. Variants of its gene, *KIT*, define the *spotting* locus in cattle. To investigate the role of *KIT* for melanogenesis in the bovine hair follicles, we optimised a protocol for total RNA extraction from hair follicles, and analysed the expression of *KIT* by qPCR. We show that *KIT* is expressed in the follicles of pigmented hair but is undetectable in the follicles of unpigmented counterparts. The absence of expression is most likely due to the absence of melanogenic melanocytes.

Keywords: *KIT,* hair follicles, coat colour, cattle, quantitative PCR

Text:

The coat of animals appears in a wide range of colours and patterns which allow for the differentiation of different breeds in domesticated species. More than 100 genes are known to affect skin and hair pigmentation in mammals (Slominski *et al*. 2004). One of the key factors is *KIT*, encoding the mast/stem cell growth factor receptor c-kit. It is essential for proliferation and migration of melanoblasts during embrogenesis, and subsequently for the development and maintenance of pigmentation (Hou *et al*. 2000). Absence of mast cell growth factor/c-kit is accompanied by loss of melanocytes or obstruction of melanin synthesis and therefore results in hair depigmentation (Botchkareva *et al.* 2001).

The *KIT* gene represents the white spotting locus in mice (Geissler *et al.* 1988; Chabot *et al.* 1988), horses (Haase *et al*. 2007; Haase *et al*. 2009), pigs (Marklund *et al*. 1998), cats (Cooper *et al*. 2006). In cattle, researchers have mapped the spotting locus to a region on

BTA6 that includes *KIT* in different populations (Grosz and MacNeil 1999; Reinsch *et al*. 1999; Liu *et al*. 2009; Hayes *et al*. 2010). However, the exact function of *KIT* for the development of the spotting phenotype in cattle remains largely unknown. In order to further investigate the role of *KIT* in hair pigmentation of cattle, a qPCR assay was set up to examine the expression of *KIT* in the follicles of pigmented (red) and unpigmented (white) in the Fleckvieh breed.

The hair roots, including follicles, are a reliable source of RNA for expression studies (Kim *et al*. 2006). However, the retrieval of high quality RNA from this material is rather challenging. Here we present an efficient protocol for collecting and extracting total RNA from bovine hair follicles. A total of 20 samples of pigmented (red) and unpigmented (white) hair were collected from the body and the head of five Fleckvieh animals. 100 µL RLT lysis buffer (Qiagen) was used for hair transportation. The hair follicles were first centrifuged for 10 seconds at 14000 rpm to make sure all follicles are contained within the RLT buffer and are subsequently homogenized by vortexing for 1 min. The obtained lysate was immediately used for RNA extraction using the RNeasy Micro Kit (Qiagen) according to the manufacturer's instructions but with a reduced buffer RW1 amount of 350 µL. Nanodrop ND1000 (Peqlab) and Bioanalyzer 2100 (Agilent) were used to evaluate the quantity and quality of the extracted RNA. The results shown in **Table 3-3** indicate sufficient amount of high quality RNA which can be used for downstream qPCR analysis.

Table 3-3 Quantification of RNA from the bovine hair follicles.

	RNA average quantity (ng/µL)	RNA average A^{260}/A^{280}	RNA Integrity Number (RIN)
Nanodrop1000	40.70 (±49.28)	1.94 (±0.16)	-
Bioanalyzer2100	67.35 (±78.21)	-	8.50 (±1.00)

The displayed values represent the average of 20 samples. For RNA, an A^{260}/A^{280} ratio of 1.8 - 2.0 and a RIN > 8.0 are optimal for qPCR measurement.

Synthesis of cDNA was performed using the QuantiTect Reverse Transcription Kit (Qiagen) with 50 ng of total RNA per sample. qPCR amplification reactions were carried out with the Fast SYBR Green MasterMix (Applied Biosystems) and the ABI Prism 7500 Fast Sequence Detection System (Applied Biosystems). In order to avoid amplification of residual genomic

DNA, primers used to amplify the *KIT* gene were located in exon2 and 3 (F: 5'-CCACCCAGCAAAATCAGAGT, R: 5'-TGTCAAATCCTTGGGGAGAG, fragment size: 367 bp). Primers for the *GAPDH* gene (F: 5'-GGCGTGAACCACGAGAAGTATAA, R: 5'-CCCTCCACGATGCCAAAGT, fragment size:173 bp) were previously described by Leutenegger *et al*. (2000). The cycling conditions comprised 10 min polymerase activation at 95°C and 40 cycles at 95°C for 5 s and 60°C for 35 s. A melting curve analysis was performed to determine the specificity of the PCR products after the experimental run. All samples were measured in duplicate. All values of the threshold cycles (Ct) were averaged across technical replicates, for *KIT* and *GAPDH* for each sample. The Ct value of *KIT*-expreesion was divided by the Ct value of the *GAPDH*-expreesion to determine the relative expression level of *KIT*.

The *KIT* gene was expressed in the follicles of pigmented hair, whereas in unpigmented counterparts an expression signal was detected in one sample presumably due to contamination **(Fig. 3-5)**. These findings are in line with a *KIT* expression study using human hair follicles, with a significantly lower expression in the follicles of unpigmented hair (grey) than in the follicles of pigmented (Hachiya *et al*. 2009). Most likely a few melanocytes are left in the grey hair follicles while completely absent in the white hair follicles (Commo *et al*. 2004). Botchkareva *et al* (2001) showed that *KIT* expression is only associated with melanogenic melanocytes in skin and hair follicles, our findings further suggest that there are no melanocytes present in the follicles of unpigmented hair.

KIT is suggested as the main candidate for the degree of spotting in cattle (Grosz and MacNeil 1999; Reinsch *et al*. 1999; Liu *et al*. 2009; Hayes *et al*. 2010), but the underlying variants have not been identified yet. It seems more likely that regulatory mutations modulating *KIT* expression are responsible for the spotting patterns in cattle. We assume that the lack of *KIT* expression in the follicles of unpigmented hair is due to lack of melanogenic melanocytes, which results from a disturbed migration/proliferation of the melanoblasts during embryogenesis (Mackenzie *et al*. 1997).

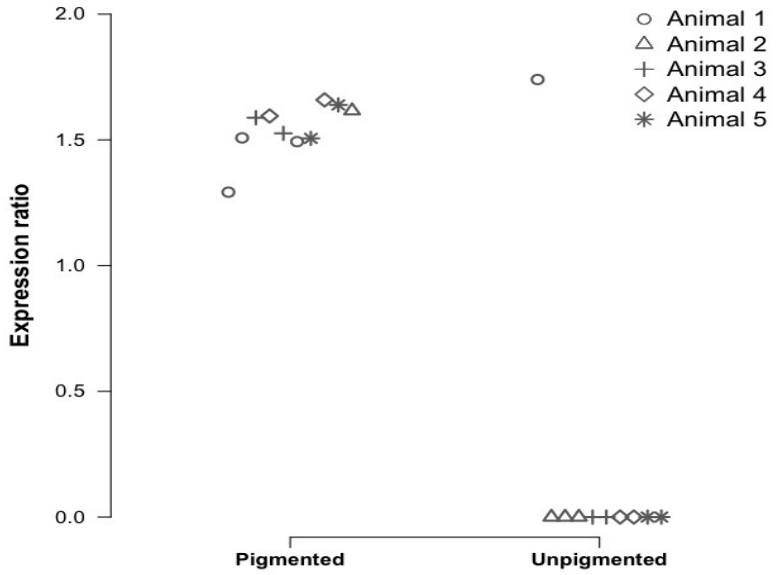

Figure 3-5 Differential expression of *KIT* in bovine hair follicles. Quantification of *KIT* expression in the follicles of pigmented (n=10) and unpigmented (n=10) hair from five animals was measured by qPCR. Each symbol indicates the expression ratio of the Ct of the target gene (*KIT*) divided by the Ct of the reference gene (*GAPDH*). Distinct symbols indicate the five different animals. The unpigmented sample from Animal 1 shows an expressed signal, this was thought to be due to contamination.

Acknowledgements

This research is financially supported by the German Ministry of Education and Research, BMBF (grants 0315134A and 0315134D), the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven. We acknowledge Dr. Johannes Buitkamp at Institut für Tierzucht, Bayerische Landesanstalt für Landwirtschaft for providing Fleckvieh hair samples. X. Wang is supported by a grant from China Scholarship Council (CSC) for his Ph.D study.

Chapter 4
General discussion

GWAS using high-density SNP arrays have been widely used to examine associations between genetic variants and phenotypes in both humans and domestic animals. However, genetic refinement of association signals identified by GWAS is rather challenging, because the GWAS associated SNPs are most likely to represent tagging markers which are in LD with the causal variants rather than be the causal variants themselves. In this dissertation, the genetic basis of two model complex traits (milk fat content and ambilateral circumocular pigmentation) in cattle were investigated by re-sequencing candidate genes derived from recent GWAS.

Identification of candidate genes mapped by GWAS

The criteria for selecting candidate genes from previous studies included: 1) a gene has a known function; 2) positional involvement in a developmental pathway; 3) affects the phenotypes in question in other species; 4) is expressed in different developmental stages or in certain organs; 5) is closed to reported QTL regions (reviewed by Ibeagha-Awemu *et al*. 2008). Position-based genetic mapping approaches (*e.g.* linkage analysis or GWAS) provide physical information in the search for candidate genes according to the positions of the most associated signals. In comparison with the conventional candidate gene approach, GWAS can provide unbiased information on variation in all genes across an entire genome, and can potentially narrow down the candidate loci to small genomic regions. Rather than focusing on candidate genes with known biological function, genes mapped by GWAS can overcome the difficulties and obstacles imposed by incomplete understanding of the biological and biochemical roles of desirable traits (Kitsios and Zintzaras 2009).

In this thesis, the positional approach has produced credible candidate genes underlying complex traits in cattle. In the fat content study in the HF breed, the GWAS utilized 44,280 available SNPs to map the QTL region to a 3 Mb interval on BTA5 (Wang *et al*. 2012). The two most highly associated SNPs were located in the second intron of the *EPS8* gene. A subsequent fine-mapping assessment confirmed that *EPS8* is the underlying gene responsible for fat content in the HF population. Through imputed genotypes at 12 million polymorphic sites in 3668 animals based on whole-genome re-sequencing of 43 key ancestors of the FV population, a highly significant FP QTL region on BTA5 was identified. The 17 most significant SNPs were located within a short interval (2532 bp) of a putative regulatory

region of the *MGST1* gene (Unpublished data by Hubert Pausch). Therefore, association analysis of denser SNPs and imputed data can increase the power of GWAS and bring more accurate location information to localise candidate genes. It can be concluded that the positional candidate genes, especially those with available functional information, should be prioritized for further assessment.

Genomic structure characterization

Gene discovery and annotation are critical steps toward the follow-up genetic analysis. Despite abundant resources for annotation and discovery of bovine genes (Childers *et al*. 2011), the sites of gene transcription initiation and termination, as well as differential splicing information, remain incomplete. Thus, the genomic structure of bovine genes is referred to the prediction and annotation of homologs in other well-annotated organisms such as human and mouse. Nevertheless, the non-protein-coding regions and regulatory regions often provide inaccurate or incomplete information, due to the complexity of genome annotation in humans (Brent 2005).

An inability to identify potential QTN may due to incomplete annotation of the candidate genes. Considering that only three *MITF* transcripts were predicted in the present study, while nine *MITF* transcripts with varied 5' specificities have been reported in humans (Levy *et al*. 2006), and multiple *MITF* transcripts have also been described in mice (Bharti *et al*. 2008; Bauer *et al*. 2009). Thus, there might be other bovine *MITF* transcripts that were not studied here. Moreover, the genomic structures of candidate genes may alter according to species-specific expression patterns, tissue specific patterns and even variations in different developmental stages.

Additionally, experimental validation is required to confirm or refine the genomic structures predicted by *in silico* methods. Rapid amplification of cDNA ends (RACE), a PCR-based method, has often been used to empirically annotate the transcription start/end sites for a single gene, as well as for large-scale structural transcript annotation (Salehi-Ashtiani *et al*. 2009).

SNP detection approach

Screening as many SNPs as possible for the most likely candidate genes is the basis for identification of the actual QTN. In this study, genomic DNA from a small number of individuals were used as templates in order to detect sequence variants. Typically, causative polymorphisms are expected to exist in different populations since segregation events occurred historically. Thus, a comparative re-sequencing approach among three breeds (Holstein-Friesian, Fleckvieh and Braunvieh) was used for SNP detection in the initial polymorphism screening phase. However, this method can be less effective in selecting candidate polymorphisms for further dissection, due to the limited sample sizes of each breed, and the unknown functions of non-coding variants. Three non-synonymous SNPs leading to amino acid exchanges (P47Q in *MGST1*, M599T in *EPS8* and E6A in *GRIN2B*) were chosen for genotyping according to their locations and the functions of the genes. Unfortunately, none of these three SNPs were able to explain the causality of the phenotypes. As the hypothesis stated that the causal variants should generally be in high LD with the most significant SNPs, another DNA panel that included 24 HF bulls was selected according to the genotypes of the most significant SNP to further refine the region of association. A promoter SNP of *EPS8* is in high LD with the most significantly associated SNP, and showed strong association with the FP EBV.

In order to screen potentially functional variants that might account for ACOP in the FV breed, DNA panels from 16 bulls with progeny-derived phenotypes were prepared. Sequencing of candidate genes in these panels identified SNPs in high LD with the most significantly associated SNPs. By using this approach, however, polymorphisms with low allele frequencies could have been lost from these DNA panels due to the limited sample sizes.

Although the coding and potential regulatory regions of the investigated candidate genes were preferentially sequenced in the current study, QTN that are located in uninvestigated regions (*e.g.* introns, intergenic regions) cannot be discovered easily. Therefore, re-sequencing the entire genomic region that shared the haplotypes among the multiple breeds will facilitate the detection of SNPs and identification of the causative mutations (Boyko 2011). Alternatively, imputation of nearly complete genotypes can be achieved by using the

whole-genome sequence information from a few key ancestors and the genotypes of the bulls analysed with high-density SNP arrays (Barris *et al*. 2012).

QTN validation

Since the QTN is thought to segregate in diverse breeds, we also genotyped the SNP ss319604833 in the 1021 FV bulls (data not shown). The results showed that this variant segregates in the FV population (MAF=0.17). However, it does not reach a genome-wide significance level. This might result from the limited sample size used from the FV population, or there might be several QTN on BTA5 accounting for fat content in the HF population (Kolbehdari *et al*. 2009; Awad *et al.* 2011; Jiang *et al*. 2010; Mai *et al*. 2010). Moreover, it is not always the case that putative QTN segregate across diverse populations and have the same effects. For instance, the effect of a QTN (G to A transition) in the ovine *GDF8* gene was only observed in specific cross breeds (Clop *et al*. 2006).

We have also conducted a large scale GWAS on milk FP in 3645 FV bulls using imputed whole-genome sequencing data from 43 key ancestors (unpublished data). Surprisingly, the most significantly associated SNP on BTA5 does not reside within the *EPS8* region but is located in the putative regulatory region of its neighbouring gene - *MGST1* (~ 0.5 Mb distance away) **(Figure 4-1)**. *MGST1* was also considered a candidate gene for FP in the present study. However, genotyping of a missense variant (P47Q) of *MGST1* provided no evidence of its causality for the FP QTL in the HF population. It is more plausible that two distinct QTL exist in the FV and HF breeds.

The functional characterization of QTN aims to investigate whether an alternative allele of QTN produces unequal phenotypes, or affects the function of proteins (Ron and Weller 2007). In the milk FP study, the rare T allele (TT/CT genotypes) of ss319604833 was found to be responsible for higher milk FP EBVs than the C allele (CC genotype) (**Figure 4-2**). A similar result was observed with the most significantly associated SNP (Hapmap49734-BTA-74577). In order to validate the functional effect of the promoter SNP ss319604833, a variety of different molecular techniques can be implemented. For instance, qPCR can be used to examine whether ss319604833 affects the expression of *EPS8* in relevant organs such as the liver and mammary gland. Additionally, *in vivo* (*e.g.* Luciferase reporter assay) and *in vitro* (*e.g.* EMSA, RNA-seq) methods can be applied for transcriptional activity studies and

transcription factor (*e.g.* TFAP2A) binding affinity changes. Since we did not provide evidence to verify whether ss319604833 is biologically related to fat biosynthesis or metabolism, further investigations are therefore needed to determine the mechanistic connections between this locus and milk fat content or to identify the real causal variant, which was not observed in the current study.

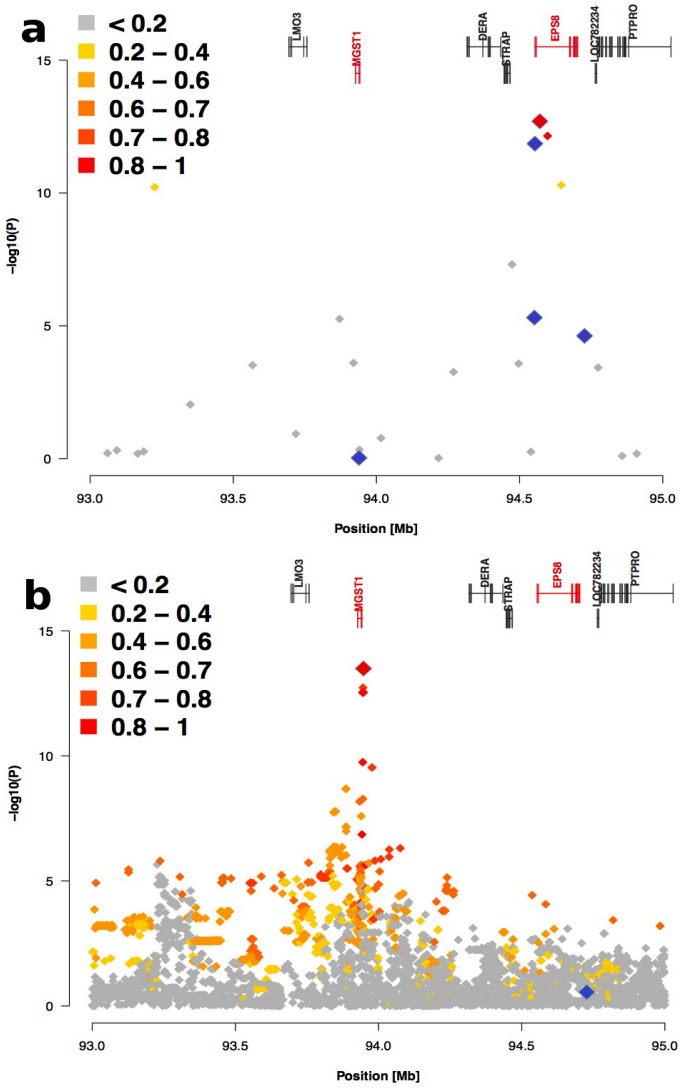

Figure 4-1 Detailed evidence of the FP QTL region on BTA5 in the HF and FV breeds, respectively. Different colours indicate the extent of LD (r^2) between the most significantly associated SNP and all other SNPs. Blue symbols represent the SNPs genotyped by TaqMan assay. **(a)** The QTL region in the HF population. SNPs were genotyped with 50K BeadChip arrays. **(b)** The QTL region in the FV population. SNPs were genotyped with 50K and 777K (high-desnity) BeadChip arrays and imputed on the basis of whole-genome sequencing data from 43 key ancestors (from unpublished data by Hubert Pausch).

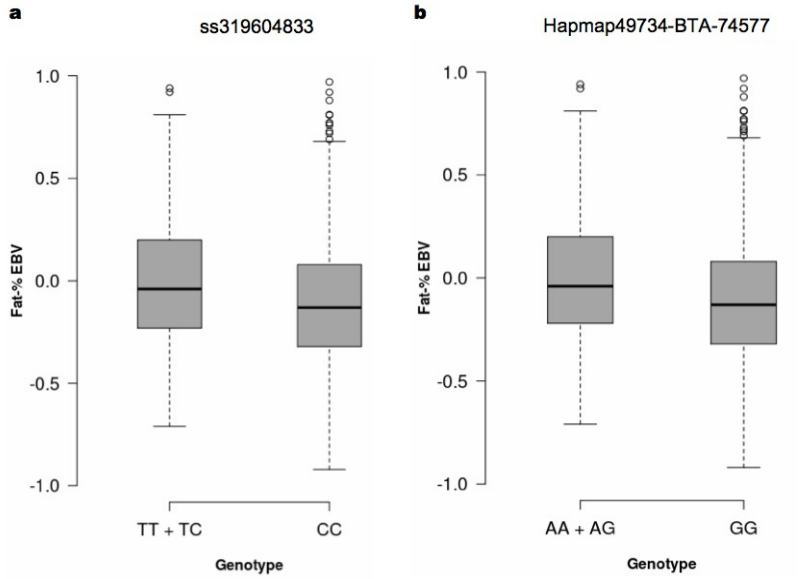

Figure 4-2 FP EBVs by genotypes for ss319604833 **(a)** and Hapmap49734-BTA-74577 **(b)**.

Non-coding variation analysis

In some cases, variants can be easily found that cause non-synonymous changes in the coding region of a gene. However, there is increasing evidence that non-coding regulatory variants can also have profound biological implications. Hindorff *et al* (2009) showed that 88% of trait-associated SNPs derived from GWAS were intronic (45%) or intergenic (43%). In farm animals, the QTN validated by previous studies existed in various genomic regions. These included an exon (K232A within the bovine *DGAT1* gene, Grisart *et al*. 2002; Winter *et al.* 2002), an intron (g. G3027A within the porcine *IGF2* gene, Van Laere *et al*. 2003) and the 3'UTR (G to A transition within the ovine *GDF8* gene, Clop *et al*. 2006).

Although hundreds of polymorphisms were identified by re-sequencing relevant candidate genes in the present study, most of them were located in the non-coding regions. The causal variants for these genes, in general, may exist in the regulatory elements of non-coding regions (Keane *et al.* 2011). The most common regulatory elements are enhancers; other regulatory sequences such as promoters, insulators and silencers may also be involved (Maston *et al*. 2006). For instance, the putative functional SNP ss319604833 identified in the promoter of *EPS8* may influence the binding of a predicted transcription factor TFAP2A, which is significantly correlated with triacylglycerol synthesis. Considering the differential expression of follicles in the pigmented and unpigmented hair, and the existence of only two intronic SNPs found in *KIT*, we assume that more likely the regulatory mutations which

modulate the *KIT* expression are responsible for the spotting patterns in cattle.

Besides the genetic consequences of variants in coding and non-coding regions, epigenetic mechanisms including DNA methylation and histone modifications seem to allow an organism to respond to the environment through the control of gene expression. In the porcine *IGF2* gene, the QTN (g. G3027A) underlying the QTL for muscle growth and back fat thickness is within a CpG island and is flanked at its 5′ site by a putative differentially methylated region (Van Laere *et al*. 2003).

Computational methods based on the evolutionary analysis of nucleotide sequences are widely used to predict the deleteriousness or potential function of non-coding variants (Cooper and Shendure 2011). Specifically, predicting the potential impact of regulatory variants on transcription factor binding sites, miRNAs, miRNA binding sites, intronic splice sites, exonic splicing enhancers and silencers (Bansal *et al*. 2010). However, due to the incomplete annotation of non-coding elements, the predictive power of regulatory elements is weak compared to coding sequences. Experimental analyses are needed to verify the regulatory effects of non-coding variants predicted by *in silico* methods. For instance, high-throughput assessments such as chromatin immunoprecipitation followed by sequencing (ChIP–seq) were used to investigate transcription factor binding and RNA-seq for transcriptional activity analysis at genome-wide levels.

Future perspectives

Tremendous progress is being made in mapping complex and mendelian traits underlying production and diseases in cattle. Recently, progress in gene-mapping studies in cattle was accelerated by next generation sequencing technologies (*e.g.* deep sequencing and RNA-seq) and high-density SNP arrays. Given the declining cost of DNA sequencing, future directions relying upon denser SNP arrays in large populations and ongoing whole genome re-sequencing projects will continue to capture additional genetic variations, and elucidate the causality between genotypes and phenotypes through functional investigations. A great understanding of functional and biological consequences following the identification of newly mapped genes or causal variants, especially those that account for major loci, will unravel additional insights into the understanding of complex phenotypes, and thus will be beneficial for both farm animals and human beings.

Summary

Most economically relevant objects are complex traits as they are controlled by numerous genes. However, the contribution of the underlying polymorphisms to the trait variation differs considerably. GWAS using high-density SNP arrays have been successfully applied to the identification of QTL regions explaining a large fraction of the genetic variation of important traits in cattle. However, the fine-mapping of the identified QTL regions is difficult as the association signals often expand over several million base pairs because of the high linkage disequilibrium in livestock species. This thesis describes the molecular-genetic analysis of significantly associated genomic regions identified by recent GWAS, and in an attempt to identify QTN underlying the most important QTL for two complex traits (milk fat content and ambilateral circumocular pigmentation) in cattle.

A genome-wide association study for milk fat percentage in the German HF population identified four major loci explaining a large fraction of the EBV variation (46.19%). Besides DGAT1, the most significantly associated QTL on BTA5 accounts for 8.35% of the EBV variation. Re-sequencing 16 positional candidate genes (*LMO3, MGST1, DERA, STRAP, EPS8, PTPRO, RERG, ARHGDIB, PDE6H, ERP27, H2AFJ, MGP, ART4, GUCY2C, ATF7IP* and *GRIN2B*) identified 338 polymorphisms. Among them a promoter SNP (ss319604833) of *EPS8* showed a similar level of significance as the most significantly associated SNP of the GWAS ($P = 1.40 \times 10^{-12}$ *vs.* $P = 2.00 \times 10^{-13}$). This potentially regulatory site resides in a highly conserved region. The polymorphism is likely to affect the binding of transcription factor TFAP2A, which is correlated with triacylglycerol synthesis. It seems likely that the binding of TFAP2A enhances the transcription rate of *EPS8* and thereby increases milk fat biosynthesis in the lactating mammary gland. These findings suggest that this polymorphism might be the underlying QTN for the BTA5 milk fat content QTL; further biological assessments are required to verify its regulatory role.

Studies of coat colour in domestic animals have led to crucial insights regarding their biology, evolution and diseases. The genetic architecture of ambilateral circumocular pigmentation (ACOP) in the FV population was investigated by GWAS and twelve major loci were identified (Pausch *et al*. 2012). Five candidate genes (*PAX3*, *KIT*, *KDR*, *GSDMC* and *MITF*) within four QTL regions were selected for fine-mapping. In total, 180

polymorphisms were detected by re-sequencing these genes in a panel of selected animals. In-depth analyses showed that a missense mutation, c.1271C>T (p.T242M) in *PAX3* largely affects the progeny-derived ACOP phenotype ($P = 3.14 \times 10^{-5}$). This low-frequency variant encodes part of the transactivation domain and resides in a highly conserved region. Gene expression analysis revealed that *KIT* is expressed in the follicles of pigmented hair and is undetectable in unpigmented counterparts. This provides evidence that regulatory sites modulating the expression of *KIT* are very likely to cause spotting patterns in cattle.

In conclusion, this thesis demonstrates a strategy for fine-mapping QTL regions and provides evidence that functional variants are very likely to contribute to the genetic variation of complex traits in cattle. These findings highlight the important role of molecular-genetic dissection of QTL regions to unravel the underlying mechanisms causing genetic variation of complex traits and furthermore provide new insights into biological processes underlying genetic variation in cattle.

Zusammenfassung

Die meisten ökonomisch wichtigen Merkmale weisen eine komplexe genetische Architektur mit einer Vielzahl zugrundeliegender Gene auf. Der Beitrag einzelner Polymorphismen zur genetischen Varianz dieser Merkmale unterscheidet sich aber erheblich. Mit genomweiten Assoziationsstudien konnten QTL identifiziert werden, die einen großen Anteil der genetischen Variation wichtiger Merkmale beim Rind erklären. Die Feinkartierung dieser QTL-Regionen erweist sich als schwierig, da hohes Kopplungsungleichgewicht in Nutztierpopulationen oftmals mehrere Millionen Basenpaare umfassende signifikante Regionen hervorruft. Diese Dissertation beschreibt die molekulargenetische Analyse zweier signifikant assoziierten QTL-Regionen zur Identifizierung von QTN für zwei komplexe Merkmale (Milch-Fett-Gehalt und beidseitige Augenpigmentierung) beim Rind.

Eine genomweite Assoziationsstudie (GWAS) für des Merkmal Milch-Fett-Gehalt in der Deutschen Holstein-Friesian Population identifizierte vier QTL, die einen großen Teil der Varianz des Zuchtwertes erklären (46,19 %). Der neben DGAT1 signifikanteste QTL befindet sich auf BTA5 ($P = 2.00 \times 10^{-13}$) und erklärt 8,35 % der Zuchtwertvarianz. Die Re-Sequenzierung von 16 positionellen Kandidatengenen (*LMO3, MGST1, DERA, STRAP, EPS8, PTPRO, RERG, ARHGDIB, PDE6H, ERP27, H2AFJ, MGP, ART4, GUCY2C, ATF7IP* und *GRIN2B*) lieferte insgesamt 338 Polymorphismen. Ein Promotor-Polymorphismus des *EPS8*-Gens (ss319604833) erzielte ein ähnliches Signifikanzniveau wie der signifikanteste assoziierte SNP aus der GWAS ($P = 1.40 \times 10^{-12}$ *vs.* $P = 2.00 \times 10^{-13}$). Diese Variante befindet sich in einer hoch konservierten Region mit regulatorischem Potential. Es ist möglich, dass diese Variante die Bindung des Transkriptionsfaktors TFAP2A beeinträchtigt, der mit der Triacylglycerol-Synthese korreliert ist. Es scheint wahrscheinlich, dass eine verstärkte Transkriptionsrate von *EPS8* aufgrund der Bindung von TFAP2A, zu einer erhöhten Biosynthese von Milchfett in der Milchdrüse laktierender Kühe führt. Diese Beobachtungen deuten darauf hin, dass die gefundene Variante möglicherweise der QTN für den QTL für Milch-Fett-Gehalt auf Chromosom 5 ist. Jedoch sind weitere biologische Untersuchungen notwendig, um diese regulatorische Funktion zu verifizieren.

Studien zur Fellfarbe in domestizierten Tieren führten zu relevanten Einblicken ihrer Biologie, Evolution und Krankheiten. Die genetische Architektur der beidseitigen

Augenpigmentierung (ACOP) in der FV Population wurde mittels GWAS untersucht und identifizierte zwölf einflussreiche genomische Regionen (Pausch *et al*. 2012). Insgesamt wurden fünf Kandidatengene (*PAX3*, *KIT*, *KDR*, *GSDMC* und *MITF*) für vier verschiedene QTL-Regionen zur Feinkartierung ausgewählt. Durch die Re-Sequenzierung dieser Gene in selektierten Tieren konnten 180 Polymorphismen detektiert werden. Eine genauere Untersuchung zeigte, dass eine Missense-Mutation, c.1271C>T (p.T242M) in *PAX3* den nachkommenbasierten Phänotyp für ACOP maßgeblich beeinflusst ($P = 3.14 \times 10^{-5}$). Diese Variante mit niedriger Allelfrequenz kodiert einen Teil des Transaktivierungsbereichs und befindet sich in einer hoch konservierten Region. Zusätzlich konnten Expressionsstudien zeigen, dass *KIT* in den Follikeln pigmentierter Haare exprimiert wird, nicht aber in Follikeln unpigmentierter Haare. Diese Ergebnisse deuten darauf hin, dass regulatorische Mutationen, die die Expression von *KIT* beeinflussen für die Scheckung beim Rind verantwortlich sind.

Zusammenfassend demonstriert diese Dissertation eine Strategie zur Feinkartierung von QTL-Regionen und zeigt dass regulatorische Varianten sehr wahrscheinlich zur genetischen Variation komplexer Merkmale beim Rind beitragen. Diese Ergebnisse heben die Wichtigkeit der molekulargenetischen Analyse von QTL-Regionen zur Identifizierung kausaler Varianten für komplexe Merkmale hervor und liefern außerdem neue Erkenntnisse über biologische Prozesse, die der genetischen Variation komplexe Merkmale beim Rind zugrunde liegt.

Bibliography

Adzhubei, I.A. et al., 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), pp.248–249.

Anderson, D.E., 1991. Genetic Study of Eye Cancer in Cattle. *Journal of Heredity*, 82(1), pp.21 -26.

Anon, 2006. The Importance of Milk and Milk Products in the Diet. In *Handbook of Dairy Foods and Nutrition, Third Edition*. CRC Press, pp. 1-53.

Aoki, H. et al., 2009. Two Distinct Types of Mouse Melanocyte: Differential Signaling Requirement for the Maintenance of Non-Cutaneous and Dermal Versus Epidermal Melanocytes. *Development*, 136(15), pp.2511–2521.

Asimit, J. & Zeggini, E., 2010. Rare variant association analysis methods for complex traits. *Annual Review of Genetics*, 44, pp.293–308.

Awad, A. et al., 2010. Confirmation and refinement of a QTL on BTA5 affecting milk production traits in the Fleckvieh dual purpose cattle breed. *Animal Genetics*, 41(1), pp.1-11.

Awad, A. et al., 2011. Mapping of a milk production quantitative trait locus to a 1.056 Mb region on bovine chromosome 5 in the Fleckvieh dual purpose cattle breed. *Genetics, Selection, Evolution: GSE*, 43, p.8.

Bailey, C.M., Hanks, D.R. & Hanks, M.A., 1990. Circumocular pigmentation and incidence of ocular squamous cell tumors in Bos taurus and Bos indicus X Bos taurus cattle. *Journal of the American Veterinary Medical Association*, 196(10), pp.1605-1608.

Bansal, V. et al., 2010. Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics*, 11(11), pp.773–785.

Barrett, J.C. et al., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), pp.263 -265.

Barris, W. et al., 2012. Next generation sequencing of African and Indicine cattle to identify single nucleotide polymorphisms. *Anim. Prod. Sci.*, 52(3), pp.133–142.

Bauer, G.L. et al., 2009. The Role of MITF Phosphorylation Sites During Coat Color and Eye Development in Mice Analyzed by Bacterial Artificial Chromosome Transgene Rescue. *Genetics*, 183(2), pp.581 -594.

Bemis, L.T. et al., 2008. MicroRNA-137 Targets Microphthalmia-Associated Transcription Factor in Melanoma Cell Lines. *Cancer Research*, 68(5), pp.1362 -1368.

Berryere, T.G. et al., 2003. TYRP1 is associated with dun coat colour in Dexter cattle or how now brown cow? *Animal Genetics*, 34(3), pp.169-175.

Beuzen, N.D., Stear, M.J. & Chang, K.C., 2000. Molecular markers and their use in animal breeding. *Veterinary Journal (London, England: 1997)*, 160(1), pp.42-52.

Bharti, K. et al., 2008. Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF. *Development*, 135(6), pp.1169 -1178.

Bickel, R.D., Kopp, A. & Nuzhdin, S.V., 2011. Composite Effects of Polymorphisms near Multiple Regulatory Elements Create a Major-Effect QTL. *PLoS Genet*, 7(1), p.e1001275.

Bionaz, M. & Loor, J.J., 2008. ACSL1, AGPAT6, FABP3, LPIN1, and SLC27A6 Are the Most Abundant Isoforms in Bovine Mammary Tissue and Their Expression Is Affected by Stage of Lactation. *The Journal of Nutrition*, 138(6), pp.1019–1024.

Blott, S. et al., 2003. Molecular Dissection of a Quantitative Trait Locus: A Phenylalanine-to-Tyrosine Substitution in the Transmembrane Domain of the Bovine Growth Hormone Receptor Is Associated With a Major Effect on Milk Yield and Composition. *Genetics*, 163(1), pp.253-266.

Bondurand, N. et al., 2000. Interaction among SOX10, PAX3 and MITF, three genes altered in Waardenburg syndrome. *Human Molecular Genetics*, 9(13), pp.1907 -1917.

Botchkareva N.V. et al., 2001. SCF/c-kit signaling is required for cyclic regeneration of the hair pigmentation unit. *The FASEB Journal*, 15(3), pp.645 -658.

Bouwman, A.C. et al., 2011. Genome-wide association of milk fatty acids in Dutch dairy cattle. *BMC Genetics*, 12(1), p.43.

Boyko, A.R., 2011. The domestic dog: man's best friend in the genomic era. *Genome Biology*, 12(2), p.216.

Bradley, D.G. et al., 1996. Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 93(10), pp.5131 -5135.

Brent, M.R., 2005. Genome annotation past, present, and future: how to define an ORF at each locus. *Genome Research*, 15(12), pp.1777–1786.

Brooks, S.A. et al., 2007. A chromosome inversion near the *KIT* gene and the Tobiano spotting pattern in horses. *Cytogenetic and Genome Research*, 119(3-4), pp.225-230.

Brooks, Samantha A. & Bailey, Ernest, 2005. Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mammalian Genome*, 16(11), pp.893-902.

Browning, B.L. & Browning, S.R., 2009. A Unified Approach to Genotype Imputation and Haplotype-Phase Inference for Large Data Sets of Trios and Unrelated Individuals. *The American Journal of Human Genetics*, 84(2), pp.210-223.

Brym, P., Kamiński, Stanislaw & Ruść, A., 2004. New SSCP polymorphism within bovine STAT5A gene and its associations with milk performance traits in Black-and-White and Jersey cattle. *Journal of Applied Genetics*, 45(4), pp.445-452.

Brym, P., Kamiński, Stanisław & Wójcik, E., 2005. Nucleotide sequence polymorphism within exon 4 of the bovine prolactin gene and its associations with milk performance traits. *Journal of Applied Genetics*, 46(2), pp.179-185.

Busenlehner, L.S. et al., 2007. Location of Substrate Binding Sites within the Integral Membrane Protein Microsomal Glutathione Transferase-1. *Biochemistry*, 46(10), pp.2812–2822.

Buttchereit, N. et al., 2010. Evaluation of five lactation curve models fitted for fat:protein ratio of milk and daily energy balance. *Journal of Dairy Science*, 93(4), pp.1702-1712.

Cao, Y. & Wang, C., 2000. The COOH-terminal transactivation domain plays a key role in regulating the in vitro and in vivo function of Pax3 homeodomain. *The Journal of Biological Chemistry*, 275(13), pp.9854-9862.

Cartharius, K. et al., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics*, 21(13), pp.2933 -2942.

Cases, S. et al., 1998. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 95(22), pp.13018-13023.

Chakravarty, K. et al., 2005. Factors that control the tissue-specific transcription of the gene for phosphoenolpyruvate carboxykinase-C. *Critical Reviews in Biochemistry and Molecular Biology*, 40(3), pp.129–154.

Chatterjee, S. et al., 2009. Identification of Novel Genes and Pathways Regulating SREBP

Transcriptional Activity. *PLoS ONE*, 4(4), p.e5197.

Childers, C.P. et al., 2010. Bovine Genome Database: integrated tools for genome annotation and discovery. *Nucleic Acids Research*, 39(Database), p.D830–D834.

Clop, A. et al., 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics*, 38(7), pp.813–818.

Cohen-Zinder, M. 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 Research*, 15(7), pp.936 -944.

Commo, S., et al., 2004. Human hair greying is linked to a specific depletion of hair follicle melanocytes affecting both the bulb and the outer root sheath. *The British Journal of Dermatology*, 150(3), pp.435-443.

Cooper, G.M. & Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet*, 12(9), pp.628-640.

Cooper, M.P. et al., 2006. White spotting in the domestic cat (Felis catus) maps near KIT on feline chromosome B1. *Animal Genetics*, 37(2), pp.163–165.

Dawn Teare, M. & Barrett, J.H., 2005. Genetic linkage studies. *The Lancet*, 366(9490), pp.1036-1044.

Durkin, K. et al., 2012. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*, 482(7383), pp.81–84.

Dybus, A. et al., 2005. Association of genetic variants of bovine prolactin with milk production traits of Black-and-White and Jersey cattle. *Arch Tierz Dummerstorf*, 48(2), pp.149–156.

Fang, M. et al., 2009. Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs. *PLoS Genet*, 5(1), p.e1000341.

Fazioli, F. et al., 1993. eps15, a novel tyrosine kinase substrate, exhibits transforming activity. *Molecular and Cellular Biology*, 13(9), pp.5814-5828.

Flint, J. & Mott, R., 2001. Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews. Genetics*, 2(6), pp.437–445.

Fontanesi, L. et al., 2010. Genetic heterogeneity at the bovine KIT gene in cattle breeds carrying different putative alleles at the spotting locus. *Animal Genetics*, 41(3), pp.295-303.

Fontanesi, L., D'Alessandro, E., et al., 2010. Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Animal Genetics*, 41(5), pp.478-492.

Fontanesi, L., Scotti, E. & Russo, V., 2012. Haplotype variability in the bovine MITF gene and association with piebaldism in Holstein and Simmental cattle breeds. *Animal Genetics*, 43(3), pp.250–256.

Freedman, M.L. et al., 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*, 43(6), pp.513-518.

Frisch, J.E., 1975. The Relative Incidence and Effect of Bovine Infectious Keratoconjunctivitis in Bos Indicus and Bos Taurus Cattle. *Animal Science*, 21(03), pp.265-274.

Georges, M. et al., 1995. Mapping Quantitative Trait Loci Controlling Milk Production in Dairy Cattle by Exploiting Progeny Testing. *Genetics*, 139(2), pp.907-920.

Girardot, M. et al., 2006. The insertion of a full-length Bos taurus LINE element is responsible for a transcriptional deregulation of the Normande Agouti gene. *Pigment Cell Research / Sponsored by the European Society for Pigment Cell Research and the International Pigment Cell Society*, 19(4), pp.346-355.

Goddard, M E, 2008. The use of high density genotyping in animal health. *Developments in Biologicals*, 132, pp.383-389.

Gogat, K. et al., 2004. VEGF and KDR Gene Expression During Human Embryonic and Fetal Eye Development. *Investigative Ophthalmology & Visual Science*, 45(1), pp.7–14.

Gordon, D., Abajian, C. & Green, P., 1998. Consed: a graphical tool for sequence finishing. *Genome Research*, 8(3), pp.195-202.

Goswami, S. et al., 2010. MicroRNA-340-mediated degradation of microphthalmia-associated transcription factor mRNA is inhibited by the coding region determinant-binding protein. *The Journal of Biological Chemistry*, 285(27), pp.20532-20540.

Gremme, G. et al., 2005. Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology*, 47(15), pp.965-978.

Grisart, Bernard et al., 2002. Positional Candidate Cloning of a QTL in Dairy Cattle: Identification of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on Milk Yield and Composition. *Genome Research*, 12(2), pp.222 -231.

Grosz, M.D. & MacNeil, M.D., 1999. The "spotted" locus maps to bovine chromosome 6 in a Hereford-Cross population. *The Journal of Heredity*, 90(1), pp.233-236.

Guo, D. et al., 2009. EGFR Signaling Through an Akt-SREBP-1-Dependent, Rapamycin-Resistant Pathway Sensitizes Glioblastomas to Antilipogenic Therapy. *Sci. Signal.*, 2(101), p.ra82.

Guo, X.-L. et al., 2010. Identification of a novel nonsense mutation on the Pax3 gene in ENU-derived white belly spotting mice and its genetic interaction with c-Kit. *Pigment Cell & Melanoma Research*, 23(2), pp.252-262.

Haase, B et al., 2009. Seven novel KIT mutations in horses with white coat colour phenotypes. *Animal Genetics*, 40(5), pp.623–629.

Haase, Bianca et al., 2007. Allelic Heterogeneity at the Equine KIT Locus in Dominant White (W) Horses. *PLoS Genet*, 3(11), p.e195.

Hachiya A, et al., 2009. Stem cell factor-KIT signalling plays a pivotal role in regulating pigmentation in mammalian hair. *J. Pathol* 218: 30-39.

Haflidadóttir, B.S. et al., 2010. miR-148 regulates Mitf in melanoma cells. *PloS One*, 5(7), p.e11574.

Hagen, M.A.E. et al., 2004. Analysis of the inheritance of white spotting and the evaluation of KIT and EDNRB as spotting loci in Dutch boxer dogs. *The Journal of Heredity*, 95(6), pp.526-531.

Hamza, M.S. et al., 2009. De-novo identification of PPARgamma/RXR binding sites and direct targets during adipogenesis. *PloS One*, 4(3), p.e4907.

Haug, A., Høstmark, A.T. & Harstad, O.M., 2007. Bovine milk in human nutrition--a review. *Lipids in Health and Disease*, 6, p.25.

Hayes, B.J. et al., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), pp.433-443.

Hayes, Ben J. et al., 2010. Genetic Architecture of Complex Traits and Accuracy of Genomic Prediction: Coat Colour, Milk-fat content, and Type in Holstein Cattle as Contrasting Model Traits. *PLoS Genet*, 6(9), p.e1001139.

Heeney, J.L. & Valli, V.E., 1985. Bovine ocular squamous cell carcinoma: an epidemiological perspective. *Canadian Journal of Comparative Medicine*, 49(1), pp.21-26.

Hindorff, L.A. et al., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America*, 106(23), pp.9362–9367.

Hofreiter, M. & Schöneberg, T., 2010. The genetic and evolutionary basis of colour variation in vertebrates. *Cellular and Molecular Life Sciences*, 67(15), pp.2591–2603.

Hou, L., Panthier, J.J. & Arnheiter, H., 2000. Signaling and transcriptional regulation in the neural crest-derived melanocyte lineage: interactions between KIT and MITF. *Development*, 127(24), pp.5379-5389.

Hu Z. et al., 2010. QTLdb: A Comprehensive Database Tool Building Bridges between Genotypes and Phenotypes. *The 9th World Congress on Genetics Applied to Livestock Production*. Leipzig, Germany August 1-6, 2010.

Ibeagha-Awemu, E.M., Kgwatalala, P. & Zhao, X., 2008. A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 19(9), pp.591–617.

Jackson, I.J., 1994. Molecular and Developmental Genetics of Mouse Coat Color. *Annual Review of Genetics*, 28(1), pp.189–217.

Jiang, L. et al., 2010. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. *PLoS ONE*, 5(10), p.e13661.

Joerg, H. et al., 1996. Red coat color in Holstein cattle is associated with a deletion in the MSHR gene. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 7(4), pp.317-318.

Johansson, K. et al., 2010. Multiple roles of microsomal glutathione transferase 1 in cellular protection: a mechanistic study. *Free Radical Biology & Medicine*, 49(11), pp.1638–1645.

Kang, H.M. et al., 2010. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*, 42(4), pp.348-354.

Keane, T.M. et al., 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, 477(7364), pp.289-294.

Kendall, C. et al., 2009. Intake and milk production of cows fed diets that differed in dietary neutral detergent fiber and neutral detergent fiber digestibility. *Journal of Dairy Science*, 92(1), pp.313-323.

Khatib, H. et al., 2006. Association of the OLR1 gene with milk composition in Holstein dairy cattle. *Journal of Dairy Science*, 89(5), pp.1753–1760.

Kim, S.J. et al., 2006. Gene Expression in Head Hair Follicles Plucked from Men and Women. *Annals of Clinical & Laboratory Science*, 36(2), pp.115–126.

Kitsios, G.D. & Zintzaras, E., 2009. Genome-wide association studies: hypothesis-"free" or "engaged"? *Translational Research*, 154(4), pp.161–164.

Klein, R.J. et al., 2005. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science (New York, N.Y.)*, 308(5720), pp.385-389.

Kolbehdari, D. et al., 2009. A whole genome scan to map QTL for milk production traits and somatic cell score in Canadian Holstein bulls. *Journal of Animal Breeding and Genetics*, 126(3), pp.216-227.

Kubic, J.D. et al., 2008. Pigmentation PAX-ways: the role of Pax3 in melanogenesis, melanocyte stem cell maintenance, and disease. *Pigment Cell & Melanoma Research*, 21(6), pp.627-645.

Kühn, Ch & Weikard, R., 2007. An investigation into the genetic background of coat colour dilution in a Charolais x German Holstein F2 resource population. *Animal Genetics*, 38(2), pp.109-113.

Lango Allen, H. et al., 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 467(7317), pp.832-838.

Lee, E. et al., 2009. Apollo: a community resource for genome annotation editing. *Bioinformatics*, 25(14), pp.1836 -1837.

Lemay, D.G. et al., 2009. The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biology*, 10(4), p.R43.

Leonard, S. et al., 2005. Effects of the Osteopontin Gene Variants on Milk Production Traits in Dairy Cattle. *Journal of Dairy Science*, 88(11), pp.4083-4086.

Leutenegger, C.M. et al., 2000. Quantitation of bovine cytokine mRNA in milk cells of healthy cattle by real-time TaqMan® polymerase chain reaction. *Veterinary Immunology and Immunopathology*, 77(3–4), pp.275–287.

Levy, C., Khaled, M. & Fisher, D.E., 2006. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends in Molecular Medicine*, 12(9), pp.406-414.

Liu, L. et al., 2009. Genome scan for the degree of white spotting in dairy cattle. *Animal Genetics*, 40(6), pp.975-977.

Loftus, R.T. et al., 1994. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 91(7), pp.2757 -2761.

Loor, J.J. et al., 2005. Temporal gene expression profiling of liver from periparturient dairy cows reveals complex adaptive mechanisms in hepatic function. *Physiological Genomics*, 23(2), pp.217 -226.

Lyman, R.F., Lai, C. & MacKay, T.F., 1999. Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in Drosophila melanogaster. *Genetical Research*, 74(3), pp.303–311.

Mackenzie, M.A. et al., 1997. Activation of the receptor tyrosine kinase Kit is required for the proliferation of melanoblasts in the mouse embryo. *Developmental Biology*, 192(1), pp.99-107.

Mai, M.D. et al., 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *J. Anim Sci.*, 88(11), pp.3522-3528.

Manolio, T.A. et al., 2009. Finding the missing heritability of complex diseases. *Nature*, 461(7265), pp.747-753.

Marklund, S. et al., 1998. Molecular Basis for the Dominant White Phenotype in the Domestic Pig. *Genome Research*, 8(8), pp.826–833.

Maston, G.A., Evans, S.K. & Green, M.R., 2006. Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics*, 7, pp.29-59.

Moioli, B. et al., 2007. Short Communication: Effect of Stearoyl-Coenzyme A Desaturase Polymorphism on Fatty Acid Composition of Milk. *Journal of Dairy Science*, 90(7), pp.3553-3558.

Morgenstern, R. & DePierre, J.W., 1983. Microsomal glutathione transferase. Purification in unactivated form and further characterization of the activation process, substrate specificity and amino acid composition. *European Journal of Biochemistry / FEBS*, 134(3), pp.591-597.

Moriguchi, T. et al., 2006. MafB Is Essential for Renal Development and F4/80 Expression in Macrophages. *Molecular and Cellular Biology*, 26(15), pp.5715-5727.

Morris, C.A. et al., 2007. Fatty acid synthase effects on bovine adipose fat and milk fat. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 18(1), pp.64–74.

Mosialou, E. et al., 1995. Microsomal glutathione transferase: lipid-derived substrates and lipid dependence. *Archives of Biochemistry and Biophysics*, 320(2), pp.210-216.

Nickerson, D.A., Tobe, V.O. & Taylor, S.L., 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Research*, 25(14), pp.2745-2751.

Olori, V.E. et al., 1997. Effect of gestation stage on milk yield and composition in Holstein Friesian dairy cattle. *Livestock Production Science*, 52(2), pp.167-176.

Olson, T.A. 1999. Genetics of color variation. In The Genetics of Cattle, ed. Fries R, Ruvinsky A. CABI, Wallingford, England.

Pausch et al., 2012. Progeny phenotyping and genome-wide association analysis of sires uncovers the genomic architecture of UV-protective eye area pigmentation in cattle. *PLoS ONE*, (In press).

Pimentel, E. da C.G. et al., 2011. Genome partitioning of genetic variation for milk production and composition traits in Holstein cattle. *Frontiers in Livestock Genomics*, 2, p.19.

Portales-Casamar, E. et al., 2010. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 38(Database issue), pp.D105-110.

Pritchard, J.K. & Cox, N.J., 2002. The allelic architecture of human disease genes: common disease–common variant…or not? *Human Molecular Genetics*, 11(20), pp.2417 -2423.

Pryce, J.E. et al., 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *Journal of Dairy Science*, 93(7), pp.3331-3345.

Quackenbush, J. et al., 2001. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic Acids Research*, 29(1), pp.159 -164.

Reinsch, N. et al., 1999. A QTL for the degree of spotting in cattle shows synteny with the KIT locus on chromosome 6. *Journal of Heredity*, 90(6), pp.629 -634.

Robinson, G.S. et al., 2001. Nonvascular role for VEGF: VEGFR-1, 2 activity is critical for neural retinal development. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 15(7), pp.1215-1217.

Ron, M. & Weller, J.I., 2007. From QTL to QTN identification in livestock – winning by points rather than knock–out: a review. *Animal Genetics*, 38(5), pp.429–439.

Roy, R. et al., 2006. Association of polymorphisms in the bovine FASN gene with milk-fat content. *Animal Genetics*, 37(3), pp.215-218.

Russell, W.O. et al., 1956. Studies on bovine ocular squamous carcinoma ("cancer eye"). I. Pathological anatomy and historical review. *Cancer*, 9(1), pp.1-52.

Saeki, N. et al., 2009. Distinctive expression and function of four GSDM family genes (GSDMA-D) in normal and malignant upper gastrointestinal epithelium. *Genes, Chromosomes & Cancer*, 48(3), pp.261-271.

Salehi-Ashtiani, K. et al., 2009. Large-Scale RACE Approach for Proactive Experimental Definition of C. Elegans ORFeome. *Genome Research*, 19(12), pp.2334–2342.

Schennink, A. et al., 2007. DGAT1 underlies large genetic variation in milk–fat composition of dairy cows. *Animal Genetics*, 38(5), pp.467-473.

Schennink, A. et al., 2009. Effect of polymorphisms in the FASN, OLR1, PPARGC1A, PRL and STAT5A genes on bovine milk-fat composition. *Animal Genetics*, 40(6), pp.909-916.

Schmutz, Sheila M. et al., 2004. A form of albinism in cattle is caused by a tyrosinase frameshift mutation. *Mammalian Genome*, 15(1), pp.62-67.

Schug, J., 2008. Using TESS to predict transcription factor binding sites in DNA sequence. *Current Protocols in Bioinformatics*, Chapter 2, p.Unit 2.6.

Seitz, J.J. et al., 1999. A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mammalian Genome*, 10(7), pp.710-712.

Signorelli, F. et al., 2009. Exploring polymorphisms and effects on milk traits of the DGAT1, SCD1 and GHR genes in four cattle breeds. *Livestock Science*, 125(1), pp.74-79.

Siritantikorn, A. et al., 2007. Protection of cells from oxidative stress by microsomal glutathione transferase 1. *Biochemical and Biophysical Research Communications*, 355(2), pp.592–596.

Slominski, A. et al., 2004. Melanin Pigmentation in Mammalian Skin and Its Hormonal Regulation. *Physiological Reviews*, 84(4), pp.1155 -1228.

Steingrímsson, E., Copeland, N.G. & Jenkins, N.A., 2004. Melanocytes and the microphthalmia transcription factor network. *Annual Review of Genetics*, 38, pp.365-411.

Tamura, M. et al., 2007. Members of a novel gene family, Gsdm, are expressed exclusively in the epithelium of the skin and gastrointestinal tract in a highly tissue-specific manner. *Genomics*, 89(5), pp.618-629.

Thaller, G. et al., 2003. Effects of DGAT1 variants on milk production traits in German cattle breeds. *Journal of Animal Science*, 81(8), pp.1911-1918.

Thomas, P.D. & Kejariwal, A., 2004. Coding Single-Nucleotide Polymorphisms Associated with Complex Vs. Mendelian Disease: Evolutionary Evidence for Differences in Molecular Effects. *Proceedings of the National Academy of Sciences of the United States of America*, 101(43), pp.15398–15403.

Tran, T.T., Schulman, J. & Fisher, D.E., 2008. UV and pigmentation: molecular mechanisms and social controversies. *Pigment Cell & Melanoma Research*, 21(5), pp.509-516.

Tsujita, H. & Plummer, C.E., 2010. Bovine Ocular Squamous Cell Carcinoma. *Veterinary Clinics of North America: Food Animal Practice*, 26(3), pp.511-529.

Van Laere, A.-S. et al., 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature*, 425(6960), pp.832–836.

VanRaden, P.M., 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91(11), pp.4414-4423.

Viitala, S. et al., 2006. The Role of the Bovine Growth Hormone Receptor and Prolactin Receptor Genes in Milk, Fat and Protein Production in Finnish Ayrshire Dairy Cattle. *Genetics*, 173(4), pp.2151 -2164.

Wang et al., 2012. Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. *PLoS ONE* (Submitted)

Ward, J.K. & Nielson, M.K., 1979. Pinkeye (Bovine Infectious Keratoconjunctivitis) in Beef Cattle. *J. Anim Sci.*, 49(2), pp.361-366.

Warnick, L.D. et al., 2001. The effect of lameness on milk production in dairy cows. *Journal of Dairy Science*, 84(9), pp.1988-1997.

Watabe, K. et al., 2001. Structure, expression and chromosome mapping of MLZE, a novel gene which is preferentially expressed in metastatic melanoma cells. *Japanese Journal of Cancer Research: Gann*, 92(2), pp.140-151.

Watanabe, A. et al., 1998. Epistatic relationship between Waardenburg Syndrome genes MITF and PAX3. *Nat Genet*, 18(3), pp.283-286.

Weikard, R. et al., 2005. The bovine PPARGC1A gene: molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiological Genomics*, 21(1), pp.1–13.

Williams, J.L., 2005. The use of marker-assisted selection in animal breeding and biotechnology. *Revue Scientifique Et Technique (International Office of Epizootics)*, 24(1), pp.379-391.

Winter, A. 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. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14), pp.9300 -9305.

Yang, J. et al., 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet*, 43(6), pp.519-525.

Zhu, Z. et al., 2010. MicroRNA-25 functions in regulation of pigmentation by targeting the transcription factor MITF in alpaca (*Lama pacos*) skin melanocytes. *Domestic Animal Endocrinology*, 38(3), pp.200-209.

Zimin, A.V. et al., 2009. A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology*, 10(4), p.R42.

Acknowledgements

I have been truly blessed to work with one of the best supervisors in the field of animal genetics and breeding: Prof. Dr. Ruedi Fries. I wish to express my sincerest gratitude to him for his continuous involvement, supervision, encouragement, patience and advice, and for giving me the opportunity to conduct research and work in his program with scientific freedom. Not only did he provide professional guidance but he also extended my horizon to the future of animal genetics and breeding in research.

I am thankful to the German Ministry of Education and Research (BMBF), the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven for financially supporting my project.

Special thanks to Prof. Dr. W. Windisch for chairing the examination board and Prof. Dr. G. Thaller for acting as a second referee of this thesis.

Sincere appreciation is extended to my lovely colleagues in Lehrstuhl für Tierzucht. They are: Hubert Pausch, Dr. Michal Wysocki, Dr. Krzysztof Flisikowski, Kai Li, Simone Jung, Christine Wurmser, Andrea Kaiser, Bernhard Aigner, Sandra Jansen, Christoph Gschöderer and Daniel Brugger. I thank them for their tremendous help and generous support over the past three years, as well as their friendship, jokes, the pleasant working atmosphere and for the parties. Special thanks go to Hubert Pausch for his wonderful assistance with the statistical analysis and Birgit Hoffmann for her kind help in administrative matters. My special thanks are also extended to Bettina Hayn, Anne keller, Rudolf Antes, Theresia Böhm and Christine Braig for their great laboratory technical assistance.

I gratefully acknowledge the Chinese Scholarship Council (CSC) and Munich Consult for their financial support. Special thanks to my Chinese friends in Freising and Munich, who helped me and shared their time and heart with me in our daily lives.

Lastly, I would like to express special appreciation and thanks to my family and relatives in China, and my girlfriend in England for their constant encouragement, support and love over the years.

No matter where I go or what I do, I'll always remember the times I spent in Freising, and in Germany.

Appendix

Appendix 1. Exon/intron organization of the investigated bovine genes

Gene	Exon	Length (bp)	Start position (bp)	End position (bp)
LMO3	1	212	93697406	93697617
	2	427	93698009	93698435
	3	214	93703878	93704091
	4	126	93746276	93746401
	5	2896	93757057	93759952
	CDS	435	93703886	93757162
MGST1	1	141	93942196	93942056
	2	95	93939245	93939151
	3	763	93927397	93926635
	CDS	465	93942196	93926635
DERA	1	122	94434516	94434395
	2	98	94397913	94397768
	3	148	94396005	94395910
	4	96	94394028	94393881
	5	135	94390292	94390184
	6	129	94372005	94371891
	7	113	94324147	94324017
	8	150	94319453	94319360
	9	700	94315879	94316336
	CDS	954	94434486	94316462
STRAP	1	438	94467480	94467043
	2	136	94466314	94466179
	3	82	94458500	94458419
	4	73	94457607	94457535
	5	97	94455360	94455264
	6	138	94454700	94454563
	7	137	94450990	94450854
	8	150	94448877	94448728
	9	66	94448341	94448276
	10	561	94447573	94447013
	CDS	1050	94467154	94447634
EPS8_transcript I	1	227	94553680	94553906
	2	74	94558442	94558515
	3	80	94674462	94674541
	4	74	94677617	94677690
	5	51	94688540	94688590
	6	68	94689953	94690020
	7	162	94691356	94691517
	8	150	94693298	94693447
	9	83	94696745	94696827
	10	137	94697410	94697546

	11	74	94700936	94701009
	12	533	94702355	94702887
	CDS	1035	94674483	94702534
EPS8_transcript II	1	227	94553680	94553906
	2	74	94558442	94558515
	3	80	94674462	94674541
	4	74	94677617	94677690
	5	68	94689953	94690020
	6	162	94691356	94691517
	7	150	94693298	94693447
	8	83	94696745	94696867
	9	137	94697410	94697546
	10	74	94700936	94701009
	11	127	94702355	94702481
	12	89	94704771	94704859
	13	75	94705230	94705304
	14	149	94712914	94713062
	15	184	94716329	94716512
	16	134	94721752	94721885
	17	100	94725123	94725222
	18	144	94726718	94726861
	19	223	94736955	94737177
	20	181	94745018	94745198
	21	130	94746416	94746545
	22	296	94748798	94749093
	CDS	2457	94674483	94748911
PTPRO	1	249	95029001	95028753
	2	274	94881737	94881464
	3	159	94871611	94871453
	4	153	94869687	94869535
	5	444	94867035	94866592
	6	162	94864763	94864602
	7	197	94860714	94860518
	8	121	94852092	94851972
	9	194	94851629	94851436
	10	112	94850138	94850027
	11	152	94844962	94844811
	12	121	94843391	94843271
	13	140	94823829	94823690
	14	133	94820326	94820194
	15	121	94818254	94818134
	16	69	94814420	94814352
	17	84	94811817	94811734
	18	36	94804585	94804550
	19	82	94801107	94801026
	20	91	94791761	94791671

	21	77	94788440	94788364
	22	135	94787538	94787404
	23	123	94786926	94786804
	24	155	94781123	94780969
	25	136	94779787	94779652
	26	120	94774166	94774047
	CDS	3648	95028827	94774062
RERG	1	70	95102110	95102179
	2	164	95102298	95102461
	3	57	95250484	95250540
	4	74	95260194	95260267
	5	1297	95262016	95263312
	CDS	597	95102401	95262423
PED6H	1	24	95357597	95357573
	2	173	95357307	95357134
	3	40	95356127	95356087
	4	465	95353276	95352811
	CDS	249	95357267	95353200
ARHGDIB	1	187	95374966	95375152
	2	137	95375410	95375546
	3	57	95379465	95379521
	4	184	95385868	95386051
	5	84	95386881	95386964
	6	77	95388897	95388973
	7	64	95390544	95390607
	8	669	95393345	95394013
	CDS	600	95385874	95393544
ERP27	1	177	95395868	95396044
	2	98	95396613	95396710
	3	138	95398929	95399066
	4	117	95412207	95412323
	5	126	95416716	95416841
	6	198	95418407	95418604
	7	716	95419496	95420211
	CDS	816	95395951	95419543
MGP	1	87	95456475	95456561
	2	33	95458045	95458077
	3	76	95458937	95459012
	4	377	95459610	95459986
	CDS	309	95456501	95459751
H2AFJ	1	675	95568292	95567617
	CDS	387	95568181	95567792
ART4	1	499	95499801	95500299
	2	694	95504232	95504925
	3	97	95513385	95513481
	CDS	954	95500129	95513476

GUCY2C	1	285	95605363	95605647
	2	113	95609801	95609913
	3	65	95611559	95611623
	4	213	95614717	95614929
	5	122	95616388	95616509
	6	97	95619350	95619446
	7	118	95622494	95622611
	8	136	95624623	95624758
	9	86	95629112	95629197
	10	112	95630527	95630638
	11	82	95636728	95636809
	12	106	95637839	95637944
	13	63	95641580	95641642
	14	72	95642246	95642317
	15	105	95642779	95642883
	16	87	95645524	95645610
	17	133	95646741	95646873
	18	138	95649779	95649916
	19	89	95651193	95651281
	20	92	95655313	95655404
	21	159	95659602	95659760
	22	193	95661454	95661646
	23	175	95663548	95663722
	24	99	95666009	95666107
	25	95	95666857	95666951
	26	77	95668475	95668551
	27	663	95670087	95670749
	28	119	95681765	95681883
	CDS	3216	95605431	95670261
ATF7IP	1	145	95859493	95859348
	2	905	95858256	95857351
	3	580	95857071	95856491
	4	83	95848326	95848243
	5	145	95846875	95846730
	6	137	95845355	95845218
	7	65	95838006	95837941
	8	73	95829269	95829196
	9	88	95828613	95828525
	10	638	95826421	95825783
	11	64	95819643	95819579
	12	78	95810977	95810899
	13	155	95809335	95809180
	14	182	95806747	95806565
	15	112	95804626	95804514
	16	589	95800726	95800137
	CDS	3729	95858249	95800307

GRIN2B	1	411	96408804	96409214
	2	599	96530804	96531402
	3	115	96648521	96648635
	4	203	96716090	96716292
	5	172	96717007	96717178
	6	154	96718215	96718368
	7	126	96723432	96723557
	8	230	96726176	96726405
	9	161	96769017	96769177
	10	188	96770969	96771156
	11	239	96773957	96774195
	12	2347	96777686	96780032
	CDS	4491	96408804	96779581
PAX3_transcript I	1	454	111229728	111229275
	2	236	111227967	111227732
	3	130	111226399	111226270
	4	412	111225102	111224691
	CDS	645	111229728	111224691
PAX3_transcript II	1	492	111229766	111229275
	2	236	111227967	111227732
	3	130	111226399	111226270
	4	135	111225102	111224968
	5	206	111161746	111161541
	6	166	111151864	111151699
	7	215	111150534	111150320
	8	850	111130061	111129212
	CDS	1437	111229766	111129212
PAX3_transcript III	1	365	111229639	111229275
	2	236	111227967	111227732
	3	130	111226399	111226270
	4	135	111225102	111224968
	5	206	111161746	111161541
	6	166	111151864	111151699
	7	215	111150534	111150320
	8	247	111130061	111129815
	9	429	111129320	111128892
	CDS	1452	111229359	111129286
KIT	1	169	71796318	71796486
	2	270	71868544	71868813
	3	282	71871914	71872195
	4	137	71873269	71873405
	5	172	71877586	71877757
	6	190	71882316	71882505
	7	116	71884641	71884756
	8	115	71901100	71901214
	9	194	71902899	71903092

	10	107	71904293	71904399
	11	127	71904490	71904616
	12	105	71904890	71904994
	13	111	71905079	71905189
	14	151	71906521	71906671
	15	92	71908606	71908697
	16	128	71909153	71909280
	17	123	71910297	71910419
	18	112	71913325	71913436
	19	100	71913550	71913649
	20	106	71913996	71914101
	21	2240	71915195	71917434
	CDS	2931	71796420	71915323
KDR	1	359	72277958	72277600
	2	94	72273946	72273853
	3	197	72271948	72271752
	4	131	72268884	72268754
	5	169	72268485	72268317
	6	140	72267603	72267464
	7	178	72266851	72266674
	8	115	72265293	72265179
	9	164	72265091	72264928
	10	157	72263063	72262907
	11	124	72261810	72261687
	12	109	72261170	72261062
	13	342	72260251	72259910
	14	147	72257871	72257725
	15	132	72256613	72256482
	16	107	72254216	72254110
	17	136	72253680	72253545
	18	105	72252964	72252860
	19	114	72251326	72251213
	20	89	72250639	72250551
	21	154	72249965	72249812
	22	98	72247897	72247800
	23	123	72245345	72245223
	24	112	72245070	72244959
	25	100	72244695	72244596
	26	106	72244185	72244080
	27	152	72242930	72242779
	28	100	72237636	72237537
	29	86	72237039	72236954
	30	2117	72234965	72232849
	CDS	4068	72277958	72232849
GSDMC	1	245	11907547	11907303
	2	221	11905410	11905190

	3	184	11904201	11904018
	4	169	11900383	11900215
	5	112	11898349	11898238
	6	48	11896964	11896917
	7	30	11896539	11896510
	8	48	11896416	11896369
	9	27	11896044	11896018
	10	30	11895849	11895820
	11	27	11895529	11895503
	12	140	11894658	11894519
	13	118	11894196	11894079
	14	74	11893413	11893340
	15	879	11892719	11891841
	CDS	1461	11907547	11891841
MITF_transcript I	1	168	31769469	31769301
	2	227	31768246	31768019
	3	83	31766967	31766884
	4	95	31764753	31764658
	5	117	31754358	31754241
	6	56	31752271	31752215
	7	75	31746514	31746439
	8	147	31742504	31742357
	9	995	31736618	31735623
	CDS	1230	31769469	31735623
MITF_transcript II	1	235	31961854	31961619
	2	249	31826664	31826415
	3	227	31768246	31768019
	4	83	31766967	31766884
	5	95	31764753	31764658
	6	117	31754358	31754241
	7	74	31752289	31752215
	8	75	31746514	31746439
	9	147	31742504	31742357
	10	836	31736618	31735782
	CDS	1530	31961854	31735782
MITF_transcript III	1	55	31839314	31839259
	2	249	31826664	31826415
	3	227	31768246	31768019
	4	83	31766967	31766884
	5	95	31764753	31764658
	6	117	31754358	31754241
	7	74	31752289	31752215
	8	75	31746514	31746439
	9	147	31742504	31742357
	10	836	31736618	31735782
	CDS	1578	31839314	31735782

Appendix 2. Primers used in candidate gene re-sequencing for sequence variant screening

Gene	Forward primer sequence		Reward primer sequence		Region
LMO3	8004_F	TGGAGGAACATGAAAGTGGA	8005_R	TTTGAACCTGGCATCATTCA	PROM
	7726_F	TAATTGCAGAGGGAGGGAAA	7727_R	CTAGGATTGGGGCTGACAAG	PROM
	7728_F	CCTCATCACACCCTTTGTCC	7729_R	GATCCCCAACCTCCAATTTT	EX1
	7940_F	GGAGCTGCTTCTCTGTGGTC	7941_R	GCCCGCATCTATTTCATGTC	EX2
	7732_F	GGCAGGTATTTTGCCATCTG	7733_R	CAGCCTGACCCTGTCACTTT	EX3
	7734_F	ATGGGTTTCTTTTCCAACGA	7735_R	TGATGTGTTTCTGGCCTTGA	EX4
	7736_F	AAGCCAAACTTTCAGACTCAGG	7737_R	GCGGCAATAATAGTGGATGTT	EX5
MGST1	7976_F	CCAGTGTGGCTGCTCAATATC	7977_R	CAGCCCCACTACTACACCTTC	PROM
	8329_F	GTAAGAGACGCGGGTTTGAT	8330_R	ATTTTCCTCCTGCCCTCAGT	PROM
	7974_F	TGATGCTGGGACAGACTGAG	7975_R	TCCCTGTGCTGTAGAGTAGGC	PROM
	6928_F	GGGTGAGATGAATTGGGAGA	6929_R	TTGCTTTCCAGAATCCCAAC	EX1
	6930_F	CAGGGATCAAACCTGCATTT	6931_R	CCCCTAACAAAGTGCACAGAA	EX2
	6932_F	AAAGCAAATATGGCGAAACG	6933_R	AAAGCAAATATGGCGAAACG	EX3
DERA	7345_F	TCAGTCGCTCAGTTGTGTCC	7346_R	CACCACCCTTATGGCAGAAA	PROM
	7347_F	CCCCACAAGAAAGTCTTCCA	7348_R	ATGAGTGCAATTGTGCGGTA	PROM
	7349_F	AAAGGCAATGCCAAAGAATG	7350_R	TGGAGCCCAGAAAAACAAAG	PROM
	7351_F	AAGATCATGGCATCCTGTCC	7352_R	CAAATGGGTCCAATTCCCTA	PROM
	7353_F	CGACTGAGCGACTGATCTGA	7354_R	AGAAACAGGGGGCTTAGGAA	PROM
	7114_F	AGGCATTTATGGCCAGTGTC	7115_R	GTTGAATCTCCTGGGTTGCT	PROM
	7116_F	TCCCACGTGCTATGAGATGA	7117_R	CAACAATGGACCCATCAGGT	EX2
	7118_F	TCTCCAACCAGGGATTGAAC	7119_R	TGGCTAGGGAATCAGAATCA	EX3
	7238_F	GTAGGCCGTTATGCCTCCTT	7239_R	TAGTCCATGAGGTCGCGAAG	EX4
	7122_F	CTGCCAGCAGCTTAATTTCA	7123_R	TCTCCTCCATTCACCCTTTG	EX5
	7124_F	AAAAGCTCTGTGGCTTGGAA	7125_R	GGACTCTCCAGGCAAGAACA	EX6
	7240_F	GGGTTAGATCCCTGGGTCAG	7241_R	CACACCTACCAACTCCCTCA	EX7
	7242_F	TGGCCTCCTGATAGAAATGG	7243_R	GTTGCAAGAAGTCGGACACA	EX8
	7130_F	TGCCTTCTTCCAAGCAAACT	7131_R	GAGAGACTTAGAAAGGCATGGA	EX9
STRAP	7339_F	GCTTCCCTACACACAATGAGC	7340_R	GCTCCGTATGGCTTTGTCAT	PROM
	7437_F	CAAAGCCATACGGAGCTCAT	7438_R	CGGCAATCTTGATTCCAACT	PROM
	7343_F	TCACACCATGCAGGGATAAA	7344_R	CAACACCACACTTCAAAAGCA	PROM
	7232_F	TTTTATTTTGAGGGGCTCCA	7233_R	CCCAGCCACGTAACATTTCT	
	7234_F	AATGGGAATGGGAACTCAAG	7235_R	AAGCTGCATAGAGGCTCGAA	PROM
	7096_F	TTTGAACCAGAGCCAAAAGC	7097_R	CTACTCCTCTCCCCATCAGC	EX1
	7098_F	GCTTCCCAGAGAGAATGTGC	7099_R	TTCTGCTGGTGGAATCAACA	EX2
	7100_F	TGGCTACAGGCTAAAAGGAAA	7101_R	CCAACTACCGCCTGTGAAAT	EX3
	7102_F	TGGTTGGTGGTGTGACATTT	7103_R	ACTTTCCTGGTGGTCCAGTG	EX4
	7439_F	ATGAGAAACTGCAGGGGATG	7440_R	CAGATGCAGCCACAGTTGTAA	EX5
	7338_F	TGTGGCTTTGGCATAAAACA	7107_R	AATGTCCCAAACACCTGGAA	EX6
	7236_F	TCTTTTCGACCCTGTGGACT	7237_R	ACCCAGGGATAGAACCCAAC	EX7
	7110_F	GGACCATCATGGAAGTCTCC	7111_R	CCTCAATTTTCCATCCTCCA	EX8,EX9
	7112_F	TTCTCTTTTGCTTCCCCTCA	7113_R	AATCCCCGGTCAGGAAACTA	EX10
	8086_F	GTGGCACTTGGGATCATTTT	8087_R	AAAGGGCTTCTGGACTGTCA	3'UTR
	8088_F	CACGCAAAATCCCAGTAACC	8089_R	GGTTGCCATTTCCAGAAGAA	3'UTR
EPS8	8527_F	GATATTTCATGTCCTGACTAGACCA	8528_R	CATATTCTCATCAAACCTGTAAGAAA	PROM

	8529_F	TTTTTCAACTCCTGGTCCAAA	8530_R	CATGGACAGTGAGCCTAGCA	PROM
	8525_F	GTATTTTTCCCCCTCCCGTA	8526_R	CTTTGTGAGCCCATGAAATG	PROM
	8375_F	ATTGGGGACAGGAGGAGAAG	8522_R	TGCCTTGAATGTGCTTTGAG	PROM
	8819_F	CGACTGGGTGACTGAACTGA	8560_R	AAATGAGATGGGCCTCGTTA	PROM
	8379_F	TAACGAGGCCCATCTCATTT	8380_R	TTCCGTCATCCTAACCAACC	PROM
	8381_F	ATTCCTGGGAGTCCTGACCT	8382_R	AAGTCGGACCGAAGTTCTGA	PROM
	8383_F	CTCTTGCTCTCCTCCCCTTC	8384_R	CTCCCGACACCAGTTGAAAG	EX1
	8385_F	ATTCTGGCCTGGAGAAAACC	8386_R	GGACACCTCAGTAGCAGCAA	EX2
	8523_F	ACTCACCTGGGGTTTGATCC	8524_R	GTCATTGGGGTTGATTGCTT	EX3
	7026_F	TGGGTCAGGATCCGATAGTC	7027_R	TCAGCCCATGATCAACAGAA	EX4
	8387_F	AACCGCTGCCATTTCTTTTA	8388_R	CGGTTTCACAATACCCATGA	EX5
	7028_F	TGACCGTATTTGCAGTTCCA	7029_R	TGTGCCTCAAGCTTCATTCT	EX5
	7030_F	GTTGCCATTTCCTTCTCCAA	7031_R	TTCCTCACTTCTTCCACCTCA	EX6
	7079_F	GGCATCCTCTCTTTCCCATA	7080_R	GACACGACTGAGCGACTTCA	EX7
	7034_F	GCGCATAATCCCTCTTTCTG	7035_R	TGCCTTGGGGTAGAAAAACA	EX8
	7036_F	CCTTGTTTTTCTACCCCAAGG	7037_R	TCAAACTCCCTCTTTGAGTGC	EX9
	7038_F	TGGAAGGCTGGTACTTCAGC	7039_R	TTGGCATAAAGGCAAAAATTG	EX10
	7040_F	AGCCACAGGTATGCTAGTTGAA	7041_R	CAGATGGCCACTATTGAATGA	EX11
	7042_F	CTTGGAACTTTGCCCTCATC	7043_R	CACAACAGACGCCAAAAGAA	EX12,EX13
	8373_F	TGCATATAATCATCATCTTCTTGC	8374_R	TGCATATAATCATCATCTTCTTGC	EX14
	7387_F	TGGAAAGAGGGCAGACTAGG	7388_R	GACGTCTCTGGTGGCTCAGT	EX15
	7389_F	TACAGTCCATGGGGTCACAA	7390_R	GGCATTTGCGCTGTTAGAAT	EX16
	7391_F	CCCTGGCCATAGTTTGTTCA	7392_R	TTCCCCACAAAACAAACACA	EX17
	7393_F	ATATGGACCTGGGGTGGAAT	7394_R	GGAAAAACCCGAATGAACCT	EX18
	7395_F	TTTGCTGTGGATCTGCAAGT	7396_R	GCTTCCCTAGTGGCTCAGTG	EX19
	7397_F	GGGGTCTTGCCAAATGAATA	7398_R	GGGTGCCAACAAAAGACAGT	EX20
	7399_F	GTTTACCCCAGGATGCTGAA	7400_R	TTCATTTTGCAGGCACACTT	EX21
	8885_F	TCTTGCTTGACCTAAAGAGCA	8886_R	CCCAAGGTCAAAACTTCTGC	EX22
	7401_F	TGATCTGCAGCCTCACAGAC	7402_R	GCAACAGAGCCGTGACAAAT	EX22
	8302_F	TCCAGGGTGTAGCTTGCTTT	8303_R	CCTTTTGCCAATGAGACACA	3'UTR
PTPRO	8371_F	CTCCACCCAGTTTCCTTCAA	8372_R	GCTCAGCTGGTCAAGAATCC	PROM
	8020_F	AGGACAGCCCTTGCAGTTTA	8021_R	CATGGACAGAGGACCCTAGC	PROM
	7956_F	GGGTGCCATTGTACGTGAG	7957_R	TGACCACTAGCGCCACCT	PROM
	7738_F	GAACACTGGGGTGGGTTG	7739_R	CCAGCAGCCGAGATACTCAC	PROM
	7740_F	AAGAACTCTTGCGTCCCAGA	7741_R	AGAGGCAAGGACAACAAAGC	EX1
	7742_F	CTACATGGCACAATCCGTGA	7743_R	AGTTGATCTTGCCGTTCTGG	EX2
	7744_F	GCAAAAGGCAATTGGAAAAA	7745_R	CCTACACCAGACATGCAGGA	EX3
	7746_F	TCTGTGGCATTTAGAAAACCAA	7747_R	GGTGATGGGTATTTAGCATGTG	EX4
	7748_F	GACTGATGATACGATAATGTGTGC	7749_R	TTGCACCATTTAAGTCCTCTGA	EX5
	7750_F	CAACTCGACATAGGGAAAGAGG	7751_R	GCTAATTTTTGGAGAGACCACCT	EX6
	7750_F	CAACTCGACATAGGGAAAGAGG	7751_R	GCTAATTTTTGGAGAGACCACCT	EX6
	7752_F	GAAACCACACTGCCACACAC	7753_R	ATGCACCAGAGAGACCTTGG	EX7
	7754_F	TGAGTCACACGCTGAAAACC	7755_R	GCACGTTACCTGCATGTGTT	EX8,EX9
	7756_F	GATTTCCCAGGCAGCAATAC	7757_R	GAATCGGACATGACTGAGCA	EX10
	7758_F	CAGGCTGCAGTCCATAGAGTC	7759_R	AGGCTTCTCAGGTGGTGCTA	EX11
	7760_F	TTTGCTGTGAATGCAAGAGG	7761_R	TCCCGACTCTGTGGTTCTTC	EX12

	7762_F	TTTGCCTTCCAAGTTAGACGA	7763_R	ACAGACAACCGCCTCAATTT	EX13
	7764_F	TGGGAAGTACAGGAAGAAGGAA	7765_R	ACCACTGAAGAGGGTGATGG	EX14
	7766_F	ACCAGCAAGAAGCCATCACT	7767_R	TCTTTACCACTGAGCCACCA	EX15
	7768_F	TCCCTGTCCCTCATCATCTC	7769_R	GAATTGAGAGCCCCAAACTG	EX16
	7800_F	AGGCCGCATGTGAATAAAAC	7801_R	CCCCCAAAACAGAGAACAGT	EX17
	7770_F	GGCAATGCTGATAGCACAAA	7771_R	CACACCAGCACATAACAGCA	EX18
	7772_F	AACGTCTGTTTTCTGCAAGGA	7773_R	GTCCAATTCATCCACCATCC	EX19
	7774_F	GAGTGGGATTGCTGGATCAT	7775_R	GCCTTTTTCACACCTTGCAT	EX20
	7776_F	TGGGATTGCTGGATCATGTA	7777_R	GGGGAGAAACCTGGTAAAGG	EX21
	7778_F	TCTTGCCCAGAGAATTCCAC	7779_R	AGTCTCCCGTGAAGGAGGAT	EX22,EX23
	7802_F	GATGCAAACCCTGATCTGCT	7803_R	TCATGATGGTTTTTGCCGTA	EX24
	7804_F	CAGGCAGACGCTTTAACCTC	7805_R	TAAAGCACACGTGCAACACA	EX25
	7784_F	CCCTAACCAGGGATTGAACC	7785_R	ACTCTTTTGTGACCCCATGC	EX26
	8090_F	AGGATGGAGTCAGGAAGTGA	8091_R	AGGGCCAAGGAAGAAACACT	3'UTR
RERG	8010_F	TCGGAGGAAGCTCTGACACT	8011_R	TTAGGAATGACGCCAAGGTC	PROM
	8012_F	CTGTGATGGCTGGAGGATTT	8013_R	TTTCCCTTTGGGCTGACTAA	PROM
	8014_F	AAACCTTGAAAGGCAAATTCA	8015_R	GCAAAAGCGGACATAAAACC	EX1
	7862_F	GCATGGCAACCCAGTCTAAT	7863_R	GGCTCTTTGCGTGATTTGAT	EX2
	7864_F	AGCAGCTCCTGTCTGTCTGC	7865_R	TTTCAAAACCCTTGGAGCAC	EX3
	7866_F	TCCAAACTGAGAGGGTCAGC	7867_R	CACTCTGCAACTTCCCAACA	EX4
	8078_F	GCCCTTTTTCTTCCTCCAAA	8079_R	TGAAACTCCCTGTGATAATGGA	3'UTR
	8080_F	CCATAGGGTCACAAAGAGTCG	8081_R	TCAGGCATAGTGAACTCTAACCA	3'UTR
	8082_F	CTGTTGAGGACCCCAGGTT	8083_R	TGTTTTCCCTTTTGGCTCAG	3'UTR
ARHGDIB	7990_F	TCAGGTCCTTCTTGTTCCAAA	7991_R	CTGCAGTCTATGGGGTCACA	PROM
	8325_F	CCCATGTCTGGGAGTCTGTT	8326_R	AGCACCCAAACCACATTTTC	PROM
	6912_F	TTCACAAGCCATTTCCTCAA	6913_R	TCTGCCTCTCAAGCCTGACT	EX1
	6914_F	CCAGGGTTTCCTCTTTGAGT	6915_R	TAGCCATGCGTTCTGCTATG	EX2
	6916_F	AGCGAGGAGAAGAAGGGAAC	6917_R	GAACCCTCACACCCATGTTT	EX3
	6918_F	AGAGTGGGCTTTCCTGATGA	6919_R	TGCAGCTCACTCCCTTAACC	EX4
	7008_F	GGTGAATCTGCCACTTCCAT	7009_R	TCCCAAGCCAAAGATCAAAC	EX5
	6922_F	GGGGCAGAGTCTTTAAGCAT	6923_R	ACTGCTCCACAGATCATCCA	EX6
	6924_F	AACAATTGGAACCAGCTTCG	6925_R	TCTAACCGAGCCAGGTATCG	EX7
	6926_F	CACAGTGAGCTCTGCTCCAA	6927_R	AATGGGGACCTTCCAGTTTC	EX8
PED6H	7994_F	GGAGGGTCTCCGACCTAGAG	7995_R	CCTTGAACCAACCTCTTCCA	PROM
	7996_F	CATCGTCAGGGAGAAGGAAG	7997_R	CTTGGCAGCACTGAAATGTT	PROM
	6842_F	AGGCCATAAGAATGGGTCAA	6843_R	GAAACAGAGGGTGAGGTGGA	EX1
	6844_F	GGGAAACTCTCTTCTGGGACA	6845_R	GGGGTTCTCAGGCAAGAATA	EX2
	6846_F	CCATTCTTCCCCATCATCTG	6847_R	CGGGGCTTACTTCCTGTGTA	EX3
ERP27	7998_F	CTCGGCCAGAGGAATATGAG	7999_R	CTGCCCACCTATTCTGAGGA	PROM
	8000_F	AGAGCTGACTGAGACCCTCCT	8001_R	CCAGAAAGATCCCACATGCT	PROM
	8002_F	TCTTCATTGTGGTGCTCAGG	8003_R	ATGTCTGACTCCAGGGAGGA	PROM
	6826_F	TGTGGGGAAGATCCAAAGAG	6827_R	TTGGGGAACAAATACCCAGA	EX1
	6828_F	GAGTCCCCCTTCCACAAGTT	6829_R	CCCTGCCTTGGAAATTAAAA	EX2
	6830_F	ATGAGCAAAGTGCCAACTCC	6831_R	TCAGCCGGTAAAGAATCTGC	EX3
	6832_F	CAGCCTGGAAAATTCAACCA	6833_R	CATATATCCCCTCCCTCCTGA	EX4
	6834_F	TTGTCACCTTCTGTGGGATG	6835_R	TGGGAAGATCCACTGGAGAA	EX5

	6836_F	GTGGGGATGAGGGGTCTACT	6837_R	AGGCCATAGAGTCACCAGGA	EX6
	6838_F	TTTAAATTGTGTGGCACAGACC	6839_R	TGACATGACTGAGCGACATTC	EX7
	6840_F	CGATGGGAAACGAGTCAAAT	6841_R	GTTAGGAACGCGAAACAAGC	EX8
H2AFJ	8016_F	CTAGCCCTTTCCCTCCTTTG	8017_R	CCTGGGATTCTCTAGGCAAG	PROM
	8018_F	GGGTTTTGCTTCTCAAACGA	8019_R	TCTGGTTGAGTATGCCTGAAA	PROM
	7664_F	GGCAAAATAGGGGTGTGTGT	7665_R	ACAACTCCAGCGAGAAGGAA	EX1
	7666_F	GTGCTAATTGGGAGCAGAGC	7667_R	CCAGACGATTTTCACCGACT	EX1
	8076_F	GCCCGCAGTGTTTTTGAATA	8077_R	TCTGACACCCCAAACAAGAA	3'UTR
MGP	7954_F	GAAGCTTAAATCGGGCTTCC	7955_R	ATCAGATGTGTCCCTGTTGC	PROM
	7654_F	GCTCTTCCACATCTGCTGTTC	7655_R	TGGGTCTTTCAGGTTTGGTC	PROM
	7656_F	AAGTCCTTTCTCCCCCAGAA	7657_R	TTGATGGAGCAGATCAGTGC	EX1
	7658_F	CCTTGATTTCAGGTCCCAAA	7659_R	GTGTGGTCTCCCAGCTGACT	EX2
	7660_F	AGCCTGTGTGCCAAGTCTCT	7661_R	AGTCTGAGCCACCAGGGATT	EX3
	7662_F	TGGGGTCACAAAGAGACACA	7663_R	AAGCAGCCATTCACTGATCC	EX4
ART4	7978_F	TTCAGATGATGGCTTATGGGTA	7979_R	CCCACCACCAGTTAAGAAAGG	PROM
	8298_F	GTGGAGCTAGGAGCCCATTA	8299_R	GCGTCAGATCCTCAATCCAC	PROM
	7448_F	GCACCTTGCAACCTCTTCTT	7449_R	AGGCATGCCCTGTATGATGT	PROM
	7450_F	CTTCTGGGTGGGGTCTATCA	7451_R	GCCCCTATCCCAAGGAATTA	EX1
	7452_F	TCAGGTCAACAACACTCACCA	7453_R	TTTATACTGCTGCGGAGACC	EX2
	7454_F	AAGCATTACCGCAAGGTCTG	7455_R	TAACCCGATGGATTGTAGCC	EX2
	7456_F	GCGAACCAAAGGAGAAATGA	7457_R	CTGGAGAATTCCGTGGACAG	EX3
GUCY2C	8006_F	CCCGTCAGAAAAGTGTCCAT	8007_R	GGGTAAGCCAGGGAGTTTTC	PROM
	8008_F	TGGACTCTGGGTTCATCCAT	8009_R	GCTTCAGCATCAGTCCTTCC	PROM
	7808_F	GTCCTTTGGACAGCAAGGAG	7809_R	GGGAGATTACCTGTGCCTCA	PROM
	7810_F	TACCGCCCTCTTCCTCTTTT	7811_R	GGGCTCAAATCCTTCTTTCTG	EX1
	7812_F	CGTGGCCCCAAAATAAAAAT	7813_R	ATGGGGAGAGAGTCAGAGCA	EX2
	7814_F	CTCTCCCAGCTCAGACATCC	7815_R	GCTTCTCCAGGCAAGAACAC	EX3
	7816_F	TTTTGTCATTTCCACCCACA	7817_R	TCAATCACACTGAGGCAAGG	EX4
	7818_F	TGAGCAAACGACATGGAAAG	7819_R	TTCTCTGGTTGTGGTGCAAG	EX5
	7820_F	AACCTGTGTCCCCTGAGTTG	7821_R	TCCTGTGCCACATGGTAGAA	EX6
	7822_F	CGGTAGGAGTTCCATTTCCA	7823_R	TGGGTCTTCATTGCTGTGTG	EX7
	7942_F	TCGCATCACGCTACTCAAAG	7943_R	ACGGTTGACCCATCTATTGC	EX8
	7826_F	CTGCTTTATGCCCTGTGTCA	7827_R	CGGCTATTTTTATGTGCCTGA	EX9
	7828_F	GAATGTCTATTGGGCAAGCA	7829_R	TGGGCAGAGAAGTAAAGTCCA	EX10
	7944_F	TATCCTTCCCTCCCACCTTT	7945_R	TTTCGATGCAGAATGCCTTT	EX11
	7832_F	GTCTCCCATGTTTCCTGCAT	7833_R	CAAGTGTGGGCAAAGGATTC	EX12
	7834_F	ATGAGAGGGGACACAGGACA	7835_R	GATGACCGCTGTATGCATGT	EX13
	7836_F	TGCATACAGCGGTCATCCTA	7837_R	GGCCCATACATTTTTCCTGA	EX14, 15
	7838_F	TGGTTGGGCAACTAAGATCC	7839_R	ATGGGGCCCTAGTTACTGGT	EX16
	7840_F	ACTCGACGGTGATTCTGTCC	7841_R	TTTGGGCAATTCTCTTAGCC	EX17
	7842_F	CCATGTTTCCTCACCCAGAT	7843_R	AGGATGTATGCGGTCCTGTT	EX18
	7844_F	AGCATCTCGGTCAGAGGAAA	7845_R	ATGGGGTCACAAAGAACTGG	EX19
	7846_F	TGTCCTCTAAGCCCGTGAGT	7847_R	CCTCCAAACATGGGAAATCT	EX20
	7848_F	ACATTGATTGCCATTGTGCT	7849_R	TCCCTGCTCTGGGAACTAGA	EX21
	7850_F	CATGTGCTCTGTGTGTGTGC	7851_R	TGAATTTGCCACAACTCCAG	EX22
	7946_F	GCTTTCTAGGTGGCGAAGTG	7947_R	GCAGGCGGATTCAAGAGTAA	EX23

	7948_F	ATTGCAGGCGGATTCTTTAC	7949_R	CCAGGCTCACAGTCAGTCAA	EX24
	7856_F	CTCTCCCCGTGTCTGGTAAT	7857_R	CCGAAAGGCAGTTGCTCTAA	EX25
	7858_F	GTGGCAAAAACAGCCAGAAT	7859_R	CCTTGGATTCAATGCCTCAT	EX26
	7860_F	TGCAGCACAGAGTAGGCATT	7861_R	ACATCCACAGATCCAGGAAA	EX27
ATF7IP	8022_F	TGGATTGTAGGTGGGGAAAA	8023_R	TTGCTTGGAGAAGGGAAAAA	PROM
	8024_F	TTTATGTGTTGCTGGGTTGC	8025_R	AACGCTGTGTGGTAACTGGA	PROM
	7806_F	GCCATGCCCTTACTGATGAA	7807_R	TCTCACACACACCAGGATGG	EX1
	7622_F	TTAGGTGGCTTTGGTTTCAA	7623_R	CTGGATCATCAGAGGCCAGT	EX2
	7624_F	CCTCTGAATCACTGGCCTCT	7625_R	TGGTCAATCTCCACACTGCT	EX2
	7626_F	TGCAGAAGCTCTCGAAACAG	7627_R	TGAAGTCGCTCAGTGGTGTC	EX2
	7628_F	CCACCAGTTCATGCTCCTTT	7629_R	TTGTTTTCCCCTTCCCATAA	EX3
	7630_F	CCTGAGGATGGAAAATAAGCA	7631_R	CCTTCCAGAGCCTTGTTCAT	EX4
	7632_F	TGGTCCTTCACCTTGTTTCC	7633_R	TTACAGCCTGATGGCAGCTT	EX5
	7634_F	GCTGGTTTGTGAGCAAGGAG	7635_R	GGTCTGGAACTGAACCCACA	EX6
	8335_F	TAAGCTGGGAGAAGGCACTG	8336_R	CAGCAAAGCAAAAAGAAGCA	EX7
	8337_F	CCGCATTCACTCCCAGTTTA	8338_R	TTCCCAGGCAAGAATAGTGG	EX8
	7638_F	TGCCAGTGTTCTAAGGCATTT	7639_R	AGGAACGTTGTTTGGTTTCG	EX9
	7640_F	CTGTATCCTCCCAGCCTCAG	7641_R	TCCCACTTTAACATCATTCCTC	EX9
	7642_F	GGAAAATCATCTGTGATTGAGC	7643_R	TGTCAAACAACACAGAGAAGCTG	EX10
	7644_F	GCCTTTGGTCTTCCTGTGAA	7645_R	TGGGCATTACAGGGTTATTG	EX11
	7646_F	TTGGTTTTCACATTGGCAGA	7647_R	GAACAGATGGCAAACAAGCA	EX12
	7648_F	GACAGACGAGAGTGGGAAGG	7649_R	TCCATCTGGGCTAAGTCACA	EX13
	7650_F	TGGTAGCTTGCATATGTGAGGA	7651_R	AATGGAAAAGAATCTGAAGCTG	EX14
	7652_F	TGCCAGATGTGTGATATTTGC	7653_R	TCCAGCCTCAAAACTAAAGCA	EX15
	8304_F	TGAACAATTCAGGCTTCAAGG	8305_R	TTGAAGGGAAATTGGCTCAT	3'UTR
GRIN2B	7988_F	CGGGTGTCTAGGAAGACCAA	7989_R	GGATGGTGAAGCAAAGCCTA	PROM
	6934_F	GGAAGATCTGGGTGCATGTT	6935_R	CCTTTTCCTGATCCTCAAACC	PROM
	6936_F	GTCCTTGCCTGTCTGCTGAT	6937_R	GGACTATCCCCTCACCCAGT	EX1
	6938_F	AAACAAACCTCTGCGTCCAG	6939_R	GCCAACCTTCAACCTCAGAC	EX2
	6940_F	TGTGAAAACTTGGATGGCAGT	6941_R	CACAGGGAAAATCATCTCTTGA	EX3
	6942_F	AGCACACGGTCTAAGGAGGA	6943_R	TGGCAATTCTTCCAGGTCAT	EX4
	6944_F	ATTGTGCTGAGCTGTGAAGG	6945_R	CCTAGAGGAGTGGGATGCAG	EX5
	6946_F	GGGGACACAGGGATGACTAA	6947_R	TTGCTACTGGTCCTGCATTG	EX6
	6948_F	ATTTCCACGTGCCTATCAGC	6949_R	GTATCTCCTGCCAGCTCTGC	EX7
	6950_F	GCCCCAGAAGAAGTGAGACA	6951_R	GCCGACCTTTCACAAATGAG	EX8
	6952_F	TCCTGAGCAAGGGCTAATATG	6953_R	AACACATGAACTTGAGCCTTGA	EX9
	6954_F	CCACCTGGGAGTTAGGCTTT	6955_R	GAAGGAGAAGGAAGCCTGGT	EX10
	6956_F	CACACATTCATACAGGGCACA	6957_R	CTACCCCCTGGGGAAATGTA	EX11
	6958_F	ACAAGGCAGGCAAGCTTCTA	6959_R	GCTTCTTCAGGCTGTCCTTG	EX12
	7020_F	GTCACCTACGGCAACATCG	7021_R	CTTTCTGCAGGTCCACGAAC	EX12
	6962_F	CGTCTAAGTACCCGCAGAGC	6963_R	CTGGAAAGGGAGCAGAGATG	3'UTR
PAX3	9247_F	GGGTGCAGAAGAGTCGAGTT	9248_R	TCAGTTTGTCAGGGGGAAAC	PROM
	9249_F	ACTTGGAAAAAGCCCAGCTC	9250_R	CTTGGTAAGTGCCAGCGAAC	PROM
	9181_F	CCCGATTCTCATCCTTTTGA	9182_R	TACCTGTGGCTCTCCCTCTC	PROM
	9183_F	GAGGAACCCCATAGGGAAGA	9184_R	GGAAAAGTTTGGTGCGAGTC	PROM
	9185_F	GTCACGACGGGAGGAGACT	9186_R	TCCAAAACAACAGGGGAAAG	EX1

	9187_F	TTCATCTTTTGGGCTGCTCT	9188_R	GTGAGCACGTACAGCTTGGA	EX2
	9189_F	AGAGTTGGGGTCCTTTTGCT	9190_R	CCCTTCGGTTAGGGTTACCA	EX3
	9191_F	GACAAGGATGCCCAAAGATG	9192_R	CCCTCAACCAAGGAAGCTCT	EX4
	9193_F	TCTTTCCCTCCATGTTTTGG	9194_R	CCCACAGACATACCTTGAGGA	EX5
	9195_F	TACAAGCCAGAAGGCAGAGG	9196_R	CCCTCAATTCCTACCCCAGT	EX6
	9197_F	TGTTCTGGTTAAGTGTTCCTTGG	9198_R	GGGTGGAGAGAAAGGAAACC	EX7
	9199_F	ATCACTAGGGGCATGGTGTC	9200_R	CAGCTTTCACGTCTCAGCAA	EX8
	9201_F	TGGACAAAGTAAGCCTTGGA	9202_R	GCTTTTGTTCGACACGTTCA	EX9
	9203_F	TGAAAAGCTGCGTGTGTTTC	9204_R	TCCACCCAGGTCGTATATTGA	3'UTR
KIT	8421_F	TGTGGTCTGATGGTCCTGAG	8422_R	GGAAGTGCTTTACCCAAGGA	PROM
	8423_F	GTGTGATTTTGGGCCACTTT	8424_R	TGCTCCCCAGACAATAAAGG	PROM
	8425_F	TTTGGTCCTCCTCGAACAAC	8426_R	AATCTTCCCGAGGCTGAAAC	EX1
	8427_F	CATGGCATCCAGCAAGTCT	8428_R	CATACCCGAAGCCACTATGC	EX2
	8429_F	GGTTTAAGTGCCACCTTTTCC	8430_R	ATTTTGAGGCTGGGAGAACC	EX3
	8431_F	TTTACTTGGGGGACGCATAG	8432_R	TCTCAACGCACTCCAGACAG	EX4
	8433_F	TTGATAGGATGCAGGTATGCTG	8434_R	TTTTCCACGTATTCATTCATTCA	EX5
	8435_F	GGCGGATTCTTTACCAGCTT	8436_R	TGAGACAACTGAGGCATGGA	EX6
	8437_F	AAGCCTTTCCCCACTCTGTT	8438_R	TCCCCATCTCAGCAATCAAT	EX7
	8439_F	TCACATTCTCCCCTTTGAGC	8440_R	ACAGTCACATTTCCCACACA	EX8
	8441_F	GAGCCAGCCGTCTTACTGAA	8442_R	GCCAGTGATGGAATGGACTT	EX9
	8443_F	GTGATGGGCGGTAGGTAGTG	8444_R	GGGTTTTCCCTGTAGGAAGG	EX10,11
	8445_F	TTTCTTCAGTTCCACCACCAC	8446_R	TCAAGAGGTCAAAGCTGGTGT	EX12,13
	8447_F	GCGTTCTGCTAAGCCATGTT	8448_R	CGACACACAGGAAAGAACCA	EX14
	8449_F	ATAGCCTGCCTCTCACATGC	8450_R	AAATACTCCTTGGGGGTGGT	EX15,16
	8451_F	ACCATTCAGAGCGCCATCTA	8452_R	TTCTCCTGCTGTGACCTTCA	EX17
	8453_F	TACCCTATTCCACCCCACAA	8454_R	GCGACCGAAATAACATTTGC	EX18,19
	8455_F	TCAAGGAAGGTTTCCGAATG	8456_R	CATCTCATATCCCTTGGCTCA	EX20
	8457_F	CCCTTTGTTCCTTGTGGGTA	8458_R	GAGGCCATCTCTCCAATTCA	EX21
	8459_F	TGAATATCTGGGCTCACGAA	8460_R	TGCAGGTGGATTGCAAATAA	3'UTR
	8461_F	GATTCAGCCTCCATCACAGG	8462_R	TGAAGTGCCCCTGAAGTACC	3'UTR
	8463_F	CTATGCTCTCGCACCTTTCC	8464_R	TGCTCAGTCGCAAGGTTATG	3'UTR
KDR	8553_F	AGGAAACGCAACCATACACC	8554_R	TCTGGAAGCTGTGCTCTTGA	5FL
	8535_F	GCGTACCTTGTTCGTGTGTG	8536_R	CTGGAGAATCCCATGGACTG	5FL
	8887_F	TCCAAACTCCTCCTCCTTGA	8888_R	TCCAAACTCCTCCTCCTTGA	5FL
	8889_F	GGCTACTGAGTGTGGGCTTC	8890_R	GCCAGAGTGCATCCTGTTTT	5FL
	8539_F	TCCTCCTTCCAAAGGGTAGC	8540_R	CAAAAACCCTTGGCAAAAGA	5FL
	8541_F	TGCAAGTGTGTTCAGCTTCC	8542_R	GTGCAAGTTCCCATCCATCT	5FL
	8543_F	TGGTTCCATCACCACCAGTA	8544_R	CCTGGCTTTGTGGTGAAAAT	5FL
	8545_F	GCCAGTACACATCCTATTCAAGC	8546_R	AAGGACTTGCACCACCAAAT	5FL
	8547_F	TTGAAAAATCCTGCCACCAT	8548_R	CCTGAATTGGTGCCCTCTAA	5FL
	8549_F	TGTTTGGCATTCATGCTTTC	8550_R	TTGGGTTTTGGTTTGATGGT	5FL
	8551_F	TGCCTGGCTTCCAAGATAAG	8552_R	AGATCCCCTGGAGAAGGAAA	5FL
	8563_F	ACAGTAAGGGCTTCCCTGGT	8564_R	TCTCAAGAGCACAGCTTCCA	PROM
	8565_F	AGTCGAGAGGTTTGGCACAG	8566_R	CCCACTCCATCTCTGCATCT	PROM
	8567_F	AATATCTCGGGGGTTTGGAG	8568_R	TGGGCTAGTTCAGTTCCCTCT	EX1
	8569_F	TCCCAAAGGAGTGAGGTCAT	8570_R	AACTCTTGGGGTACATCCTTGA	EX2

	8571_F	AGTTTGGTTTGTTCGCTGCT	8572_R	AGGAGGGCTTAGGAGTGAGG	EX3
	8820_F	CATGCGCTCTTTCACTTTGA	8821_R	CTGAAGCCTTTCTGGCTGTC	EX4
	8822_F	CATCTTCTTCCATCATCTCTTACC	8823_R	TAGCAGCAGCAGCACATAGG	EX5
	8575_F	TGGGCTACAACACATTGGAA	8576_R	GGACTTTGGAGAATGGAGGA	EX6
	8577_F	GCAAGAAGTGGCATGAGACA	8578_R	CCCTGAGGTCCTAACAAGAGC	EX7
	8579_F	TGGAATGGGACCTGTAGGAG	8580_R	CCCTGCTGGCACTTTTATTC	EX8,9
	8581_F	GGCACAAAGTCACCCAACTT	8582_R	CCACCAACAGAGGAGTGGAT	EX10
	8583_F	TGCTTCCCTCCTGTATCCTG	8584_R	GCACATGCACTCAGACCAAA	EX11
	8585_F	TGGTTCCAAAGCTCATTTTGT	8586_R	TCAACAGCCAGCCCATAAAT	EX12
	8587_F	AATGCCAGGGCTTCACTTTT	8588_R	ATCCCCTGAAGACAGCAGTG	EX13
	8589_F	TCTGTTTGATAGGAGGGAAGTCA	8590_R	TCACTATGTTGTCCACCTGAAA	EX14
	8591_F	ATGGGGCTGCAAAATAATCA	8592_R	CTACCTGCCGGGCTAAGAAT	EX15
	8593_F	GCTCCATCCAAGTTAGAGATTG	8594_R	GCCTGGTAAACCCAATAGCA	EX16,17
	8595_F	CGGGCTACTTATCCATCGTC	8596_R	GCCACAAAGCCACTGATACA	EX18
	8597_F	AGGGAAAGTTGCACAGATGC	8598_R	GGCCACAGGTGAGGTAGAAA	EX19
	8599_F	TGGTGATTGTGGAGTTCTGC	8600_R	TTCTTTGTCACATCCCGTCA	EX20,21
	8601_F	TTTGCCTCCTGGATTTGTTC	8602_R	TTTGGGGATGATTCAAGAGC	EX22
	8891_F	TTCAGAGTCCGAGGCTGTTT	8892_R	AGCATGGATGGAGGAACAAG	EX23
	8893_F	GATGCACCCTGCGATCTATT	8894_R	TGGTGTAATCAGGGGCTCTC	EX24,25
	8605_F	GAAAGAAAGGGGGAAAAAGG	8606_R	TTCAGTGGCTGAGAGGACAG	EX26
	8607_F	CCACGTTCCCCATAAACAAG	8608_R	CGCACAGAACTCTTCCTTCC	EX27
	8609_F	CCAGAGGATGTTTTGCAGGT	8610_R	GCGCTTATGCTCAAAGTCCT	EX28,29
	8611_F	GGTAGGGAATGGGGTTAGGA	8612_R	ATGAGAAGAGCACGCAACCT	EX30
	8613_F	TGGAAAGGAAGACGCTGACT	8614_R	ATTGCTCCCAGCCTCTCATA	3'UTR
	8615_F	CTCCCAGCCAGTCAGTATTC	8616_R	GGGAAGCCCCATAAGGATAG	3'UTR
GSDMC	9251_F	TGGTCTTGGCAATTTGACTG	9252_R	TTCCTCCCAAACAGTTGGAC	PROM
	9253_F	TGGATTAAGGGAGGAGCACA	9254_R	TGGCGAAGTAAAGAAGATGGA	PROM
	9205_F	GATGGAGGGACTGTGAATCC	9206_R	AACACACCTGGCTCACATTTT	PROM
	9207_F	GTGTTTTGCAATGCCCCTAC	9208_R	GCGGATTCCTTACCATCTGA	EX1
	9255_F	CCTGGGCACCATCATACTCT	9256_R	CAGGGAGCCGGTTTTACC	EX2
	9300_F	ACTTCATCCCTTCTTGGAATCA	9301_R	GCATGATATGGGAGGTCATTTT	EX3
	9213_F	CCAGTACCTTCCCACCGTTA	9214_R	TGGACATGAGTTGGCGACTA	EX4
	9215_F	CGCCCAAATACATGCAATAA	9216_R	GGAAATGGCCTTACATCTGC	EX5
	9217_F	GCCTCTCTGCACCATTCAG	9218_R	ATGGGGAAGAAGGCAGGTAA	EX6,7
	9219_F	AAGGGAGAAAAGGCACCTGA	9220_R	TCCCTGAACTTCTGCCCTTA	EX8,9
	9221_F	GATGGGTGAGTGGATGACCT	9222_R	CAGCAAGGGTTTGTTGTGTG	EX10,11
	9223_F	TCCTCCCAGCCTTCTACTCC	9224_R	GAGGAGACCCGTATTAGCAGAA	EX12
	9225_F	CCTTCCCCTCCTACCTGAGT	9226_R	GAGCCTGGAATCATCCACAC	EX13
	9227_F	CAAGAGCAGGAGGGAACTTG	9228_R	GGTCAGGGTGAGCTTGGATA	EX14
	9229_F	TATCACCATTTGGCACCTGA	9230_R	CATGCAATGAAATGCTCAGA	EX15
	9231_F	CCACTCCAGTTTGGGTGAAC	9232_R	CATGGCAACCCACTTCAGTA	3'UTR
MITF	8391_F	CCATTTGAGCTACAGGGAAGTC	8392_R	TGGCTCATTTGAACTTCCAG	PROM
	8393_F	TCCTCCCTGATACCTCCAAA	8394_R	TCAAGGGTACCGACTTGACC	PROM
	8472_F	TGGCTTTTGTCAGAGGTGTATG	8473_R	CCCCAGTTTCTTCACCTGAG	PROM
	8474_F	CAGCTTGCTTCCCTAAACTG	8475_R	GCGAGGAATGGGAGAAGAGT	PROM
	8476_F	CCACTGCTTCAGTCTCATGG	8477_R	CATCTTTAGCCCATACTTCTGC	PROM

8478_F	TGTTCTGACCTTTGAATTGTGTG	8479_R	GATAAGGCCCTGGAAACTCC	PROM
8480_F	CTAGTAGGGGGCGAGGGTTT	8481_R	AGTCTGCTGGGCATGAGAAC	PROM
8482_F	TCCTAAAACCCATCCCTGTG	8483_R	TAATTGACTGCAGCCCTGTG	PROM
8484_F	CAAGTTCAGGTTGCAGGCTA	8485_R	CTGAGAGTCTGGGGTGAAGC	PROM
8486_F	CTTCATTAGCCTGGGCATTT	8487_R	TCAAAGTCAAACTCGCTGTCA	PROM
8395_F	GTGAGCTTGGAGGAGGAGTG	8396_R	CGGAGACCCTCATTGGTG	EX1
8397_F	GTGCCTTGGTCCTTGTGAAG	8398_R	CCGGCTTCAGACTACCTCAG	EX2
8399_F	TATTTTGCACGGGCTCATTT	8400_R	TTTCTCACGGGACATCAACA	EX3
8401_F	TGTTCTCCAAAGAGGGCATC	8402_R	GGAGAATACAGCCATGCACA	EX4
8403_F	GCATGAGAAGAAGGGCACA	8404_R	CCTGGAGCCATAACCCAATA	EX5
8405_F	CCTTTTACCCAAACCAACCA	8406_R	TGTCCAAGTTGCTAACCCATAA	EX6
8407_F	CATCCTCCTCATCACCCATT	8408_R	TATTTCCCTCGGCAGCAGTA	EX7
8409_F	TCCCTATGTCTGCTGCCTTT	8410_R	CAACTGTGGGTCTGAACGTG	EX8
8470_F	GGGAGCTTCTTGGCTCAGTT	8471_R	CCAGTCCTCAAGATACTGAAAACA	EX9
8413_F	TGTCCCAACCAAGCAGAGAT	8414_R	TCCCCTTGATTGGTAAAACA	EX10
8415_F	TTGACTTTGTCATCAACAGACCT	8416_R	TCGGAATAAACACACCAGCTAA	EX11
8417_F	TGTCCTAAAATCCCCTGTGC	8418_R	GCCAAAATCCTCTTGCTCTG	EX12
8419_F	ACCTGAAAGGGGTTTTCTTGA	8420_R	CAGAGCACAGAGTGAAGAGCA	EX13
8533_F	TCGATTCATTGGTACTGCCTTA	8534_R	AACAGGCATATTAACGCTATGG	3'UTR
9115_F	GAAAACCAAACTGGGCACAT	9116_R	TGCCACTTGCTATTCAATCGT	3'UTR
9117_F	GGTGTTAAGAGGCAGGCTGT	9118_R	TGACTTTTGCACCCAGCTC	3'UTR
9119_F	ACTCAAGGGGGTAGGGAAAA	9120_R	GGCAAAGTCAGAAGGTTTGG	3'UTR
9121_F	TAAGCCAAGAGGTGGTGGTT	9122_R	CATTAAGTCCGAAGAACTTGTGC	3'UTR
9123_F	GCTGTTGGATGCAGCAATAA	9124_R	GGGAAGTTTCTTGCTCATTTTC	3'UTR

5FL: 5' flanking region.

Appendix 3. Sequence variants identified in the investigated bovine genes

Gene	Intern SNP_ID	dbSNP_ID (ss)	Polymorphism	Localization	Amino acid
LMO3	3737	319604511	C>T	PROM	
	3738	319604512	TTAGT_indel	PROM	
	3739	319604513	A>T	PROM	
	3575	319604514	A>G	INT3	
	3576	319604515	C>G	INT3	
	3577	319604516	C>T	INT3	
	3578	319604517	A>G	INT3	
	3579	319604518	A_indel	INT3	
MGST1	3797	319604519	G>T	PROM	
	3798	319604520	G>T	PROM	
	3799	319604521	C>T	PROM	
	3800	319604522	A>G	PROM	
	3801	319604523	A>G	PROM	
	3802	319604524	A>T	PROM	
	3803	319604525	G>T	PROM	
	3741	319604526	C>G	PROM	
	3742	319604527	C>G	PROM	
	3743	319604528	A>C	PROM	
	3744	319604529	A>G	PROM	
	3745	319604530	C>T	PROM	
	3746	319604531	A>G	PROM	
	3747	319604532	A>G	PROM	
	3164	319604533	C>T	PROM	
	3159	319604534	A>C	PROM	
	3160	319604535	G>T	PROM	
	3161	319604536	A>G	PROM	
	3162	319604537	A>G	EX1	S5
	3163	319604538	A>G	EX1	T21
	3138	319604539	A>C	INT1	
	3139	319604540	A>G	INT1	
	3140	319604541	C_indel	INT1	
	3141	319604542	C_indel	INT1	
	3142	319604543	A>C	EX2	P47Q
	3143	319604544	A>G	INT2	
	3099	319604545	C>T	EX3	R130
	3100	319604546	A>C	3'UTR	
DERA	3274	319604547	C>T	PROM	
	3275	319604548	A>G	PROM	
	3276	319604549	C>T	PROM	
	3277	319604550	C>T	PROM	
	3278	319604551	C>T	PROM	
	3279	319604552	C>G	PROM	
	3196	319604553	A>C	PROM	
	3197	319604554	C_indel	PROM	
	3198	319604555	A>G	INT1	

	3200	319604556	T_indel	INT2	
	3187	319604557	A>G	INT3	
	3205	319604558	C>T	EX4	R105
	3206	319604559	A>G	INT4	
	3207	319604560	GG_indel	INT4	
	3199	319604561	G>T	INT4	
	3188	319604562	A>G	INT6	
	3216	319604563	C>T	INT6	
	3217	319604564	A>G	EX7	T221
	3218	319604565	C>T	INT7	
	3202	319604566	C_indel	3'UTR	
STRAP	3261	319604567	C>G	PROM	
	3262	319604568	A>G	PROM	
	3259	319604569	C>T	PROM	
	3209	319604570	C>G	PROM	
	3210	319604571	C>T	PROM	
	3211	319604572	TTGAAGGCAGG_indel	PROM	
	3212	319604573	C>T	PROM	
	3213	319604574	A>G	PROM	
	3214	319604575	C>T	PROM	
	3215	319604576	A_indel	PROM	
	3191	319604577	C>G	PROM	
	3192	319604578	G>T	INT1	
	3201	319604579	C>T	INT1	
	3208	319604580	A>G	INT2	
	3193	319604581	A>G	INT3	
	3372	319604582	C>G	INT4	
	3373	319604583	A>T	INT4	
	3194	319604584	A>G	INT6	
	3195	319604585	A>G	INT6	
	3189	319604586	A>C	INT8	
	3190	319604587	A>G	INT8	
	3203	319604588	TTTTTTT_indel	INT8	
	3204	319604589	A>G	3'UTR	
	3814	319604590	A>G	3'UTR	
	3815	319604591	C>T	3'UTR	
	3816	319604592	C>T	3'UTR	
	3817	319604593	C>G	3'UTR	
EPS8	3856*	319604828	C>T	PROM	
	3857*	319604829	A_indel	PROM	
	3871*	319604830	A>G	PROM	
	3956*	319604831	TAA_indel	PROM	
	3858*	319604832	A>G	PROM	
	3826*	319604833	C>T	PROM	
	3827*	319604834	G>T	EX1 (5'UTR)	
	3853*	319604835	C>T	INT1	
	3258	319604594	C>T	INT5	

	3151*	319604836	GT_indel	INT8	
	3152*	319604837	C>T	INT8	
	3153*	319604838	A>G	INT8	
	3825*	319604839	C>T	INT9	
	3823*	319604840	C>T	INT10	
	3822*	319604841	C>T	INT11	
	3263*	319604842	A_indel	INT15	
	3264	319604595	A>G	INT15	
	3265	319604596	C>T	INT15	
	3266*	319604843	A>C	INT15	
	3267	319604597	A>C	INT15	
	3254*	319604844	A>G	INT16	
	3255	319604598	A>G	EX17	E545
	3256*	319604845	C>T	EX18	M599T
	3260*	319604846	A>C	INT19	
	3257	319604599	C>T	INT20	
	3824*	319604847	C>T	INT20	
	3716	319604600	C>T	3'UTR	
PTPRO	3748	319604601	A>G	PROM	
	3749	319604602	A>C	PROM	
	3750	319604603	A>G	PROM	
	3781	319604604	A>G	PROM	
	3674	319604605	AC_indel	PROM	
	3563	319604606	A>G	PROM	
	3564	319604607	A_indel	INT4	
	3565	319604608	C>G	EX5	E292Q
	3566	319604609	G>T	EX5	L299W
	3567	319604610	A>G	EX5	E322K
	3568	319604611	A>G	INT6	
	3545	319604612	A>G	INT6	
	3546	319604613	C>T	INT7	
	3552	319604614	C>T	EX8	L499
	3553	319604615	A>G	EX9	P530
	3547	319604616	G>T	INT11	
	3548	319604617	A>G	INT12	
	3569	319604618	A>G	EX16	I861V
	3551	319604619	A>G	INT16	
	3549	319604620	C>T	INT18	
	3550	319604621	A>C	INT18	
	3542	319604622	G>T	INT21	
	3543	319604623	C_indel	INT21	
	3544	319604624	C>T	INT21	
	3581	319604625	T>A	INT23	
	3672	319604626	C>T	INT24	
	3673	319604627	G_indel	INT24	
	3541	319604628	A>C	INT25	
	3813	319604629	C>G	3'UTR	

RERG	3765	319604630	A>G	PROM	
	3736	319604631	A>C	PROM	
	3660	319604632	C>T	PROM	
	3661	319604633	CA>TC	PROM	
	3662	319604634	C>T	PROM	
	3663	319604635	A>T	PROM	
	3664	319604636	A>T	PROM	
	3665	319604637	C>T	PROM	
	3645	319604638	A>G	INT3	
	3646	319604639	C>T	EX4	N120
	3647	319604640	A>G	EX4	A141
	3648	319604641	C>T	EX4	N155
	3649	319604642	C>G	3'UTR	
	3724	319604643	C>G	3'UTR	
	3725	319604644	A>G	3'UTR	
	3726	319604645	C>T	3'UTR	
	3727	319604646	G_indel	3'UTR	
	3728	319604647	G_indel	3'UTR	
	3729	319604648	A>G	3'UTR	
	3730	319604649	C>T	3'UTR	
	3732	319604650	A>T	3'UTR	
	3731	319604651	A>G	3'UTR	
	3733	319604652	A>G	3'UTR	
	3734	319604653	C>T	3'UTR	
	3735	319604654	A>G	3'UTR	
	3804	319604655	A>G	3'UTR	
	3805	319604656	C>T	3'UTR	
	3723	319604657	AT_indel	3'UTR	
ARHGDIB	3968	319604658	C>T	PROM	
	3049	319604659	C>T	INT2	
	3050	319604660	A>G	INT2	
	3051	319604661	A>T	INT2	
	3052	319604662	A>G	EX3 (5'UTR)	
	3053	319604663	A>G	INT3	
	3158	319604664	A>G	INT5	
	3054	319604665	A>G	INT5	
	3055	319604666	C>T	INT6	
	3056	319604667	C>T	INT6	
	3057	319604668	C>G	INT7	
	3058	319604669	C>T	INT7	
	3059	319604670	C>T	3'UTR	
	3060	319604671	C>T	3'UTR	
PDE6H	3757	319604672	G>T	PROM	
	3758	319604673	C>T	PROM	
	3759	319604674	C>T	PROM	
	3760	319604675	A>G	PROM	
	3761	319605013	C>T	PROM	

	3762	319604676	A>G	PROM	
	3763	319604677	A>C	PROM	
	3764	319604678	CCT_indel	PROM	
	3009	319604679	C>G	3'UTR	
	3010	319604680	G>T	3'UTR	
ERP27	3007	319604681	C>T	PROM	
	3819	319604682	A>G	PROM	
	3820	319604683	A>G	PROM	
	3008	319604684	C>G	INT1	
	3028	319604685	G>T	INT2	
	3029	319604686	C>G	INT2	
	3030	319604687	G>T	INT3	
	3026	319604688	A>G	INT4	
	3027	319604689	A>C	INT4	
	3031	319604690	C>T	INT4	
	3032	319604691	A>G	EX5	R163H
	3033	319604692	A>G	INT5	
	3034	319604693	C>T	INT5	
	3035	319604694	T_indel	INT5	
	3036	319604695	A>G	INT6	
	3025	319604696	A>T	INT7	
	3021	319604697	A>G	INT7	
	3022	319604698	AC_indel	3'UTR	
	3023	319604699	A>C	3'UTR	
	3024	319604700	A>G	3'UTR	
H2AFJ	3810	319604701	A>G	PROM	
	3766	319604702	G>T	PROM	
	3812	319604703	A>G	3'UTR	
MGP	3533	319604704	A>T	PROM	
	3534	319604705	AGGTGTGTGTTTGTGA_indel	PROM	
	3535	319604706	C>T	PROM	
	3536	319604707	C>T	PROM	
	3537	319604708	G>T	PROM	
	3538	319604709	A>C	PROM	
	3539	319604710	A>G	PROM	
	3540	319604711	G>T	PROM	
	3519	319604712	G>T	INT1	
	3520	319604713	C>T	INT1	
ART4	3374	319604714	C>T	PROM	
	3719	319604715	A>G	PROM	
	3720	319604716	A>G	PROM	
	3721	319604717	A>G	PROM	
	3722	319604718	C>T	PROM	
	3375	319604719	C>T	EX2	D138
GUCY2C	3767	319604720	C>T	PROM	
	3768	319604721	C>T	PROM	

	3769	319604722	C>T	PROM	
	3770	319604723	G>T	PROM	
	3771	319604724	A>G	PROM	
	3772	319604725	A>G	PROM	
	3773	319604726	C>T	PROM	
	3774	319604727	G>T	PROM	
	3775	319604728	A>G	PROM	
	3776	319604729	G>T	PROM	
	3651	319604730	A>G	PROM	
	3652	319604731	A>G	PROM	
	3653	319604732	A>G	PROM	
	3654	319604733	C>T	PROM	
	3655	319604734	C>T	PROM	
	3656	319604735	A>G	EX1	T3
	3657	319604736	A>G	EX1	R16Q
	3658	319604737	C>T	EX1	P20L
	3659	319604738	C>T	EX1	S38
	3683	319604739	A>G	INT1	
	3684	319604740	A>T	INT1	
	3685	319604741	C>T	INT2	
	3782	319604742	GTTG_indel	INT2	
	3640	319604743	C>T	INT9	
	3641	319604744	A>C	INT9	
	3639	319604745	C>T	INT13	
	3643	319604746	A>T	EX15	T549
	3631	319604747	C>T	INT16	
	3632	319604748	A>G	INT17	
	3633	319604749	G>T	INT17	
	3644	319604750	C>T	INT18	
	3634	319604751	A>C	INT20	
	3635	319604752	C>T	INT20	
	3636	319604753	C>T	EX21	D790
	3637	319604754	A>G	INT21	
	3642	319604755	A>G	INT21	
	3666	319604756	C>T	INT22	
	3667	319604757	A>G	INT22	
	3777	319604758	A>C	INT23	
	3778	319604759	C>T	INT23	
	3779	319604760	A>G	INT23	
	3780	319604761	C>T	INT23	
	3668	319604762	C>G	INT25	
	3669	319604763	A>G	INT25	
	3670	319604764	C>T	INT25	
	3671	319604765	C>T	INT25	
	3638	319604766	C>T	3'UTR	
ATF7IP	3818	319604767	A>C	PROM	
	3740	319604768	A>G	PROM	

	3580	319604769	G>T	EX1 (5'UTR)	
	3513	319604770	A>G	EX5	E551
	3573	319604771	C_indel	INT7	
	3574	319604772	C>T	INT7	
	3808	319604773	C>T	INT7	
	3809	319604774	C>T	INT8	
	3512	319604775	A>T	INT10	
	3514	319604776	A>T	INT10	
	3515	319604777	A>G	INT10	
	3516	319604778	C>T	INT10	
	3517	319604779	T_indel	INT10	
	3521	319604780	A>G	INT13	
	3518	319604781	A>T	INT15	
	3717	319604782	A>G	3'UTR	
	3718	319604783	C>T	3'UTR	
GRIN2B	3751	319604784	C>T	PROM	
	3752	319604785	C>T	PROM	
	3753	319604786	C>T	PROM	
	3754	319604787	A>G	PROM	
	3755	319604788	C>T	PROM	
	3101	319604789	C>T	PROM	
	3102	319604790	A>G	PROM	
	3129	319604791	A>G	PROM	
	3130	319604792	A>G	PROM	
	3131	319604793	A>G	EX1	E6A
	3132	319604794	A>G	EX1	V21
	3133	319604795	A>G	EX1	P32
	3154	319604796	C>T	EX1	D56
	3155	319604797	C>T	EX1	H64
	3134	319604798	A>G	EX1	P78
	3135	319604799	C>G	EX1	T121
	3136	319604800	A>T	INT1	
	3103	319604801	C>T	EX2	T275
	3104	319604802	A>G	INT3	
	3105	319604803	A>G	INT3	
	3106	319604804	C>T	INT3	
	3107	319604805	C>T	INT3	
	3108	319604806	C>T	EX4	Y388
	3109	319604807	C>T	EX4	I408
	3110	319604808	A_del	INT5	
	3111	319604809	C>T	EX6	A506
	3112	319604810	C>T	INT6	
	3113	319604811	C>T	INT6	
	3114	319604812	C>T	INT6	
	3115	319604813	A>G	INT7	
	3116	319604814	C>G	INT8	
	3117	319604815	C>T	EX9	V714

	3118	319604816	A>G	INT10	
	3119	319604817	C>T	INT10	
	3120	319604818	A>G	INT10	
	3121	319604819	A>G	INT10	
	3122	319604820	C>T	INT11	
	3123	319604821	C>G	INT11	
	3124	319604822	A>G	INT11	
	3125	319604823	A>G	INT11	
	3126	319604824	G>T	INT11	
	3127	319604825	C>T	EX12	P953
	3156	319604826	C>T	EX12	C1173
	3157	319604827	A>G	EX12	T1232A
PAX3	4083	475875626	C>T	PROM	
	4084	475875627	C>T	PROM	
	4085	475875628	C>T	PROM	
	4086	475875629	G>T	PROM	
	4087	475875630	G>T	PROM	
	4050	475875631	A>G	PROM	
	4051	475875632	C>G	PROM	
	4052	475875633	C>G	PROM	
	4053	475875634	G>T	PROM	
	4054	475875635	A>G	5'UTR	
	4055	475875636	G>T	3'UTR	
	4056	475875637	A>G	3'UTR	
	4057	475875638	C>T	EX5	R223
	4058	475875639	C>T	INT5	
	4059	475875640	C>T	INT5	
	4060	475875641	G>T	INT6	
	4061	475875642	C>T	EX8	T424M
	4062	475875643	A>C	INT8	
	4063	475875644	C>T	INT8	
	4089	475875645	C>G	3'UTR	
	4090	475875646	ATAAG_indel	3'UTR	
KIT	3840		G>T	INT3	
	3841		A>G	INT17	
KDR	3881		G>T	5'FL	
	3999		C>T	5'FL	
	3872		C>G	5'FL	
	3873		A>G	5'FL	
	3874		A>G	5'FL	
	3875		A>G	5'FL	
	3902		C>T	5'FL	
	3903		T_indel	5'FL	
	3904		A>G	5'FL	
	3905		C>T	5'FL	
	3906		A>G	5'FL	
	3907		CTTA_indel	5'FL	
	3908		A>G	5'FL	
	3909		C>G	5'FL	

3910	A>G	5'FL	
3911	A>T	5'FL	
4000	A>G	5'FL	
3882	C>T	5'FL	
4001	GTT_indel	5'FL	
3922	A>G	PROM	
3954	T_indel	PROM	
3955	AA_indel	PROM	
3923	A>C	PROM	
3924	A>T	INT2	
4002	A>C	INT4	
4003	A>C	EX5	V171
3925	A>G	INT6	
3926	C>T	INT6	
3894	A>C	INT6	
3895	A>G	INT6	
3896	C>T	EX8	P359S
3897	A>G	INT9	
3883	A>C	INT11	
3884	A>G	INT11	
3885	A>G	INT11	
3886	A>G	INT11	
3887	C>T	INT12	
3888	C>T	INT12	
3889	A>G	INT12	
3890	A>G	EX13	I624V
3891	A>G	EX13	E628K
3892	C>T	EX13	N631
3893	C>T	INT13	
3933	T_indel	INT13	
3927	C>T	INT13	
3928	C>T	INT13	
3929	G_indel	INT13	
3930	A>G	INT13	
3931	A>G	INT13	
3932	G>T	INT13	
3898	A>G	INT21	
3899	C>T	EX22	I1006
3934	A>G	EX26	T1152
3900	A>C	INT27	
3901	A>T	INT27	
3935	A>G	EX28	T1238
3936	A>C	INT28	
3937	A>G	INT28	
3938	C>T	INT28	
3939	A>G	3'UTR	
3940	C>T	3'UTR	
3941	A>G	3'UTR	
3942	A>G	3'UTR	
3943	G>T	3'UTR	
3944	C>T	3'UTR	
3945	C>T	3'UTR	

	3946	A>G	3'UTR	
	3947	A>G	3'UTR	
	3948	C>T	3'UTR	
	3949	C>T	3'UTR	
	3950	TATAAA_indel	3'UTR	
	3951	A>G	3'UTR	
	3952	A>G	3'UTR	
	3953	A>G	3'UTR	
GSDMC	4064	A>G	PROM	
	4065	C>T	PROM	
	4066	A_indel	PROM	
	4067	A>G	PROM	
	4068	C>T	PROM	
	4069	A>G	PROM	
	4070	C>T	PROM	
	4071	C>T	PROM	
	4072	C>T	PROM	
	4073	C>T	PROM	
	4074	G>T	PROM	
	4075	A>T	PROM	
	4076	C>T	PROM	
	4077	C>G	PROM	
	4078	A>G	PROM	
	4079	C>T	PROM	
	4080	A>G	PROM	
	4017	A>G	PROM	
	4018	A>T	PROM	
	4019	A>G	PROM	
	4020	C>G	PROM	
	4021	G>T	PROM	
	4022	A>G	PROM	
	4023	A>G	PROM	
	4024	C>T	5'UTR	
	4025	A>G	5'UTR	
	4081	C>T	EX2	L27P
	4026	C>T	INT3	
	4027	A>G	EX4	V164I
	4028	C>T	INT4	
	4029	A>G	INT4	
	4030	A>T	EX5	V200E
	4031	G>T	INT7	
	4032	A>G	EX8	K263R
	4033	A>G	INT8	
	4082	GG_indel	INT8	
	4034	C>T	INT11	
	4035	A>G	EX12	M310I
	4036	A>G	INT12	
	4037	A>C	INT12	
	4038	A>C	INT12	
	4039	C>T	INT12	
	4040	A>G	INT12	
	4041	A>G	EX13	Y374C

	4042	C>T	INT13	
	4043	A>T	INT13	
	4044	C>G	INT13	
	4045	C>T	INT14	
	4046	A>G	INT14	
	4047	C>G	3'UTR	
	4048	C>T	3'UTR	
	4049	G_indel	3'UTR	
MITF	3833	C>T	PROM	
	3834	A>G	PROM	
	3835	A>G	PROM	
	3836	A>G	PROM	
	3870	C>T	PROM	
	3871	A>G	PROM	
	3878	C>T	PROM	
	3862	A>C	PROM	
	3863	G>T	PROM	
	3864	C>T	PROM	
	3865	A>G	PROM	
	3879	C>G	PROM	
	3866	C>T	PROM	
	3867	A>G	PROM	
	3868	C>T	PROM	
	3869	C>T	PROM	
	3843	A>G	5'UTR	
	3842	C>T	INT3	
	3844	A>T	INT4	
	3837	C>T	INT5	
	3839	A>G	INT7	
	3850	T_indel	INT8	
	3838	C>T	EX12	H439
	3846	GA_indel	3'UTR	
	3847	C>T	3'UTR	
	4012	A>G	3'UTR	
	4013	A>C	3'UTR	
	4014	G>T	3'UTR	
	4015	A>C	3'UTR	
	4016	A>G	3'UTR	

* Polymorphisms were detected by using the 24 HF DNA panel. 5'FL: 5' flanking region.

Appendix 4. Primers and probes used for TaqMan genotyping

Lab ID	dbSNP ID	Gene	Primer	Probe
3131	ss319604793	*GRIN2B*	F: GTTGTAATTTAACCTCTGTGTGTGTGT	VIC: CAGAGCGGAGTGCTGT
			R: CCAGCACCAACCAGAACTTG	FAM: AGAGCGGCGTGCTGT
3142	ss319604543	*MGST1*	F: GGTTTGTTTTTTTCTTGGTATTTGGAATAGGT	VIC: TTGCCAACCAAGAAGA
			R: GATACTTCTTGGCATTTTCTCCTTTGC	FAM: TTGCCAACCCAGAAGA
3256	ss319604845	*EPS8*	F: GAATCTGGATTAGGACGTGCTGAT	VIC: CTGTATGGTATGCGTGTAAG
			R: GGGCAATGAATCTAAAAATTTGAAAAAGTT	FAM: TGTATGGTATGCATGTAAG
3826	ss319604833	*EPS8*	F: GGCTGCCGCCAAGAC	VIC: CGCCCTGAGCCCAG
			R: CGCACGCCCAGTTCCT	FAM: CGCCCTAAGCCCAG
3956	ss319604831	*EPS8*	F: GGTTCTAGAGATTTTGGAGCAGAAGA	VIC: CTGGTAGTTACTTAATAATTG
			R: TGACATCACAAGGCTTGAAATATTTGAAC	FAM: TCTGGTAGTTACTTAATTG
4061	ss475875642	*PAX3*	F: CGCTCTCGCCTCTCACT	VIC: CCGACACCGTGGTGGT
			R: TCCAAGCTCTTCATATGGTCTAGTCT	FAM: CCGACACCATGGTGGT

Appendix 5. Gene re-sequencing and SNP genotyping protocol

1) Buffers

TE	10mM Tris-HCl, 1mM EDTA, pH 8.0
TBE 10x	900mM Tris-HCl, 900mM Borate, 20mM EDTA, pH 8.3
EDTA	0.1M EDTA.Na_2, 1.0 M Tris-HCl
Dilution buffer	1M Tris-HCl, 25mM $MgCl_2$, pH 8.0

2) PCR and re-sequencing methods

Genomic DNA from semen was prepared using standard protocols.

The primers used in this study were designed using the Primer 3 online tool (http://frodo.wi.mit.edu/primer3/) and ordered from Sigma-Aldrich Chemie GmbH, Steinheim, Germany. The primer sequences are shown in **Appendix 2** according to their primer ids. A standard PCR reaction at a total volume of 20 µL was prepared according to the following protocol.

Genomic DNA (20 – 30 ng/µL)	1.5-2 µL
10x Coraload Buffer	2 µL
dNTP	2 µL
Forward primer (5 pmol/µL)	1 µL
Reverse primer (5 pmol/µL)	1 µL
Taq-Polymerase (0.025 units/µL) (Qiagen, Hilden, Germany)	0.2 µL
H_2O	due
Total	20 µL

After pre-incubation at 94 °C for 3 min, the PCR mixture underwent 30-33 cycles of denaturation at 94 °C for 30 s, annealing for 60 s and extension at 72 °C for 60 s. A final extension step at 72 °C for 3 min followed. If the PCR failed under standard conditions, the reaction was optimized by adding Q-Solution (Qiagen, Hilden, Germany) or 5% DMSO or by adjusting the annealing temperature to meet the requirements of the primers. To check the success of the PCR products, 4 µL of the PCR products was loaded onto an 1.5% agarose gel with 0.5 TBE buffer and stained with ethidium bromide. Electrophoresis of PCR products was carried out at 100 V for 30 min.

The PCR product was cleaned up in order to remove any remaining primers and dNTPs before being run in the sequencing reaction. A standard PCR reaction at a total volume of 8 µL was prepared according to the following table and thermal cycling for each reaction was initiated at 37 °C for 30 min and 80 °C for 15 min.

PCR product	2-3 µL
FastAP Thermosensitive Alkaline Phosphatase (1 u/µL) (Fermentas, St. Leon-Rot, Germany)	0.4 µL
Exonuclease I (20 u/µL) (Fermentas, St. Leon-Rot, Germany)	0.1 µL
H_2O	due
Total	8 µL

A total volume of 10 µL of the sequencing reaction mixture was prepared according to following table and the thermal cycling for each reaction was initiated by denaturation at 95 °C for 10 s, followed by 32 cycles of 10 s at 95 °C for denaturation, 5 s at 51 °C for primer annealing and 4 min at 60 °C for extension.

PCR cleaning product	1.5- 2 µL
Dilution buffer	4 µL
Forward or Reverse primer	0.5 µL
BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, USA)	0.5 µL
H_2O	due
Total	10 µL

The sequencing reaction products were cleaned through gel filtration using a MultiScreen Sephadex plate (Millipore, Schwalbach, Germany) before being loaded into the capillary sequencer. A 45 g volume of Sephadex G-50 (Sigma-Aldrich Chemie GmbH, Steinheim, Germany) was incubated with 300 µL water in a Sephadex plate at room temperature for 2 h before being centrifuged (Centrifuge 5804, Eppendorf, Hamburg, Germany) at 950 g for 8 min to remove the water. The sequencing reaction products were added to 30 µL EDTA water and transferred into Sephadex columns, covered a fresh plate (MicroAmp 96 well reaction plate, Applied Biosystems, Foster City, USA) and centrifuged (Centrifuge 5804, Eppendorf, Hamburg, Germany) at 950 g for 4 min. The fresh plate was eventually loaded onto an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, USA).

3) Allelic discrimination with custom TaqMan genotyping assays

Genotyping with custom TaqMan assays is based on the annealing of two allele specific primers labeled with different fluorescent markers. The TaqMan SNP assays were designed via the online system of Applied Biosystem and ordered from Applied Biosystems, Foster City, USA. The PCR reaction mixture at a total volume of 8 µL for each well, was prepared according to the following table.

Genomic DNA	1 µL
TaqMan Genotyping Master Mix (Applied Biosystems, Foster City, USA)	3.5 µL
TaqMan Genotyping Assay (Applied Biosystems, Foster City, USA)	0.15 µL
H_2O	3.35 µL
Total	8 µL

The thermal cycling for the genotyping reaction was initiated at 95 °C for 10 min and was followed by 40 cycles of denaturation for 15 s at 92 °C, and annealing for 60 s at 60 °C. The reaction was performed on an ABI7500 System (Applied Biosystems, Foster City, USA).

Appendix 6. DNA panels used for polymorphism screening

1). Genomic DNA panel used for the polymorphism screening of the QTL region on BTA5 for the FP study.

Identification code	Lab_id	Name	Breed
276000934216432	4542	Vinstor	Braunvieh
276000934856703	4543	Humor	Braunvieh
276000934442734	4544	James	Braunvieh
276000935377641	4545	Efast	Braunvieh
276000936199323	4551	Eibsee	Fleckvieh
276000937578788	4552	Honduras	Fleckvieh
276000937638989	4553	Pomal	Fleckvieh
276000936995477	4554	Rügen	Fleckvieh
276000937712020	4547	Renegade	Holstein-Friesian
276000662206477	4548	Lionel	Holstein-Friesian
276000938688368	4649	Tartan	Holstein-Friesian
250001450822003	4550	Koldy	Holstein-Friesian

2). Genomic DNA panel (24 animals) used for the polymorphism screening of the *EPS8* gene.

Identification code	Lab_id	Name
528000240536505	KI-GT1842	Tempel
528000295614887	KI-GT1850	Dukan
276001401263581	KI-GT1845	Turnier
276000347515029	KI-GT0479	Dogan
276000347440967	KI-GT0481	Ruwillo
276000768137078	KI-GT0482	Maiball
276000113829256	KI-GT0483	Nog
276000346491684	KI-GT0484	Belmfort
276000345655132	KI-GT0433	Jockardi
528000337849019	KI-GT0458	Laslo
276000346335385	KI-GT0475	Fabello
276000579651209	KI-GT0478	Carnesto
276000578989280	KI-GT1602	Elatos
276001502341005	KI-GT1604	Jango
276001502037822	KI-GT1606	Mouster
250002831233721	KI-GT1609	Joplin
276000344666668	KI-GT1200	Banny
276001273114089	KI-GT1205	Maxim
276000342353778	KI-GT1272	Nog
276000578905876	KI-GT0753	Pirot
276000344252802	KI-GT0556	Nikito
276000578726718	KI-GT0619	Apaluso
276000113186559	KI-GT0678	C.outback
276000347906735	KI-GT0752	Jargas

3). Genomic DNA panel (16 FV bulls) used for the polymorphism screening of the *KIT*, *KDR* and *MITF* gene.

Identification code	Lab_id	Name
276000936661144	7499	Island
40000333051547	7514	Waldess
276000935486867	7420	Romer
276000935388458	7057	Ironman
276000931294024	7442	Lonur
276000932807446	7389	Winkel
276000935384192	6893	Bombard
276000932170406	7388	Wiege
276000937364841	7547	Zuidberg
276000934794886	3862	Perino
276000931435786	4017	Sauerbruch
276000931377648	4157	Herminator
276000935247757	4049	Viagra
276000933672905	4409	Harly
276000932035371	3464	Prowen
276000533288296	4012	Hardware

4). Genomic DNA panel (16 FV bulls) used for the polymorphism screening of the *PAX3* and *GSDMC* gene.

Identification code	Lab_id	Name
276000912178945	8564	Profi
276000931928543	8171	Spoliant
276000931380536	8436	Spalt
276000932878716	8173	Stegnatz
276000918951820	8565	Wagner
276000533288270	7122	Sonsbeck
276000933590279	6871	Zahlo
276000934794886	8398	Perino
276000936552424	7345	Romos
276000932322228	3877	Rendan
276000934742640	8077	Rolling
276000914910182	8566	Renoir
276000934897637	7136	Ridach
276000932734085	8176	Weinboy
276000932322216	4279	Honzug
276000938721535	8135	Weltadel

Curriculum Vitae

XIAOLONG WANG

Nationality:	Chinese
Date of Birth:	6th, Dec. 1983
Place of Birth:	Shaanxi, P.R.China

HIGHER EDUCATION

2008.10-2011.10	**Ph.D study**, Animal Genetics Chair of Animal Breeding Technische Universität München, Germany
2005.09-2008.07	**Master of Science**, Molecular Animal Nutrition College of Animal Science and Technology Northwest A&F University, P.R.China
2001.09-2005.07	**Bachelor of Science**, Animal Science College of Animal Science and Technology Northwest A&F University, P.R.China **Bachelor of Engineering**, Computer Science College of Information Technology and Engineering Northwest A&F University, P.R.China

PUBLICATIONS

Xiaolong Wang*, Christine Wurmser*, Hubert Pausch, Simone Jung, Friedrich Reinhardt, Jens Tetens, Georg Thaller, Ruedi Fries., 2012. Identification and characterization of four major QTL regions for milk fat content in the German-Holstein-Friesian population. *PLoS ONE* 7(7): e40711. (*co-first author)

Hubert Pausch, **Xiaolong Wang**, Simone Jung, Dieter Krogmeier, Christian Edel, Christian Edel, Reiner Emmerling, Kay-Uwe Götz, Ruedi Fries., 2012. Progeny phenotyping and genome-wide association analysis of sires uncovers the genomic architecture of UV-protective eye area pigmentation in cattle. *PLoS ONE* 7(5): e36346.

Xiaolong Wang, Daniel Brugger, Hubert Pausch, Ruedi Fries, Michal Wysocki., 2012. Differential expression of *KIT* in pigmented and non-pigmented bovine hair follicles. *Animal Genetics*, (Submitted).

Xiaolong Wang *et al.,* Polymorphism analysis of candidate genes for a fat content QTL region on BTA5 in the German Holstein-Friesian population. (In preparation).

Xiaolong Wang *et al.,* Polymorphism analysis of candidate genes associated with ambilateral circumocular pigmentation in the German Fleckvieh population. (In preparation).

J.Wang, Y.L.Chen, **X.L.Wang**, Z.X.Yang., 2007. The genetic diversity of seven indigenous Chinese goat breeds. Small Ruminant Research.

CONFERENCE ABSTRACT

Polymorphism analysis of genes in a 3.2-Mb region on bovine chromosome 5 associated with milk fat percentage. 32nd Conference of the International Society of Animal Genetics (ISAG) 2010, Edinburgh, UK.

Generation of genotypes for population genetics on two genotyping platforms. Synbreed Midterm Evaluation Meeting, 2012, Freising, Germany.

Fine-mapping of a QTL for milk fat percentage on bovine chromosome 5 based on imputed sequencing data. 33rd Conference of the International Society of Animal Genetics (ISAG) 2012, Cairns, Australia.