Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science

Validation of the role of a SNP in the VPS35 gene and its effect on non-coagulating milk in Swedish Red Cattle

Validering av rollen av en SNP i VPS35 genen och dess effekt för icke koagulerande mjölk hos Svensk Rödbrokig Boskap

Gro Hylén

Degree project • 30 credits
Agronom programmet inriktning husdjur/ Animal Science MSc
Department of Animal Breeding and Genetics
Examensarbete / SLU, Institutionen för husdjursgenetik no 540
Uppsala 2018

Validation of the role of a SNP in the VPS35 gene and its effect on non-coagulating milk in Swedish Red Cattle

Validering av rollen av en SNP i VPS35 genen och dess effekt för icke koagulerande mjölk hos Svensk Rödbrokig Boskap

Gro Hylén

Supervisor:	Sandrine Duchemin, Department of Animal Breeding and Genetics, SLU
Assitant supervisor:	Kim Jäderkvist Fegraeus, Department of Animal Breeding and Genetics, SLU
Examinator:	Dirk-Jan De Koning, Department of Animal Breeding and Genetics, SLU
Credits:	30 credits
Level:	A2E
Course title:	Degree Project in Animal Science
Course code:	EX0556
Program:	Agricultural Science program – Animal Science
Place of publication:	Uppsala
Year of publication:	2018
Serietitel:	Examensarbete / SLU, Institutionen för husdjursgenetik
Delnummer i serien:	540
Elektronisk publisering:	https://stud.epsilon.slu.se

Keywords: VPS35, genotype, SNP, SR, Cow, NC-milk, MCP, Genomic selection
Nykelord: VPS35, genotyp, SNP, SRB, ko, icke-koagulerande, mjölk, genomisk selektion

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinary Medicine and Animal Science Department of Animal Breeding and Genetics

Abstract

From earlier studies, it is clear that nearly 20 % of the Swedish red (SR) dairy herd possess milk that doesn't coagulate within 40 minutes after rennet addition. This is a problem when processing milk into cheese because lower cheese yield is the result of non-coagulating (NC) milk. An estimate of 17.5 M SEK could be gained if cheese yield increased with one percent. The current payment strategies for milk from Swedish dairies favours a high total protein content and high yield. However, there is a moderate unfavourable genetic correlation between total protein content and NC milk in SR cows, which could possibly lead to an increase of NC milk within the SR cattle population. Much of the literature in the subject suggests that the largest contribution to differences in milk quality has a genetic background, which allows breeding cows with better milk quality. A polymorphism in the VPS35 gene has previously been suggested to be in strong association with NC milk in SR cows. The aim of this thesis was to validate a SNP in the VPS35 gene that has been found to influence NC milk. This validation included genotyping of the SNP in 1046 SR cows, estimation of the heritability for NC and the effect of this SNP on the proportion of NC milk. If the SNP is proven to be significant for NC milk, it can be relevant to include it in breeding programmes for SR cows. Phenotype records were available on the following animals: 382 animals sampled between 2010 and 2011 and 603 animals sampled between 2016 and 2017. We compared the real genotypes obtained from this study and genotypes from previous imputed data that was available for 382 animals. Based on this comparison, we validated and re-estimated the effect of the VPS35 SNP with the 603 genotyped animals, that haven't yet undergone a genome-wide association study (GWAS). The results showed high correlation between the imputed and the real data and high heritability for NC milk. However, for the 603 animals that haven't yet undergone a GWAS, the effect of the VPS35 SNP was low as well as the heritability.

Keywords: VPS35, genotype, SNP, SR, Cow, NC-milk, MCP, Genomic selection

Sammanfattning på Svenska

Från tidigare studier står det klart att nästan 20 % av rasen SRB (svensk rödbrokig boskap) mjölkar mjölk som inte koagulerar inom 40 minuter efter att löpe har tillsats. Detta är ett problem när mjölken ska användas till ostproduktion, då ostutbytet blir mindre om mjölken inte koagulerar. Ett uppskattat värde av 17,5 miljoner kronor skulle kunna sparas om ostutbytet ökade med en procent. Betalningssystemet från mejerier till mjölkproducenter favoriserar högt totalt proteininnehåll och hög mjölkmängd. Det är en måttlig ogynnsam korrelation mellan proteininnehåll och icke koagulerande mjölk hos SRB, vilket skulle kunna försämra koagulering egenskaper ytterligare. Mycket av litteraturen i ämnet föreslår att den största anledningen till skillnader i mjölkkvalitet beror på genetiska skillnader mellan djur. Vetskapen om detta gör det möjligt att avla kor med bättre mjölkkvalitet. Polymorfism i VPS35 genen har påvisat stark association med icke koagulerande mjölk hos SRB. Syftet med den här studien var att validera en ’single nucleotide polymorphism’ (SNP) i VPS35 genen som påverkar icke koagulerande mjölk. I valideringen ingår genotypning av SNP:et och skattning av arvbarhet och effekten av SNP:et för icke koagulerande mjölk. Om SNP:et påvisas vara signifikant för icke koagulerande mjölk kan det vara relevant att använda det i avelsprogram för SRB. I den här studien blev 1046 SRB kor genotypade för polymorfism i ett SNP i VPS35 genen. Fenotyp observationer fanns tillgängliga för följande djur; 382 djur provtagna mellan år 2010 och 2011 och 603 djur provtagna mellan år 2016 och 2017. I studien jämförde vi resultaten mellan de riktiga genotypvärdena från denna studie med de imputerade gentypvärdena från tidigare studie som fanns tillgängliga för 382 djur. Baserat på denna jämförelse kunde vi validera och uppskatta effekten av VPS35 SNP:et för de 603 genotypade djuren. Resultaten visar på hög korrelation mellan den imputerade versus den riktiga datan och på hög arvbarhet för NC mjölk för 382 djur. Men för de 603 djur som inte ännu genomgått en genome-wide association study (GWAS) var effekten av gentyp i SNP:et lågt såväl som arvbarheten.

Nyckelord: VPS35, genotyp, SNP, SRB, ko, icke-koagulerande mjölk, koaguleringsegenskaper, Genomisk selektion

Preface

It has been an interesting journey to work with this thesis. I have met really nice people and I have got an insight in the research world. I must admit I didn't quite know what I was getting myself into before I started. It was my previous experience of handcrafted cheesemaking that got me interested in the subject of milk quality. To be able to combine milk quality traits with genomic data sounded fascinating and important! I have learnt so much during the process of this thesis. There is so much more I would like to add in this thesis. But for the purpose and the aim of the thesis I hope I have manage to include the most essential parts. For me, the thesis has also served as a perfect summary of my time as a student at SLU. In terms of what I have learnt and practical application of the knowledge. Getting into the genetic field of animal science started year one with a group project about the evolution; Darwins finches, and Mendel. It has from then on continued with a bachelor about gene editing and it is now finished with this thesis about SNP in dairy cows. I hope that my work has contributed to the bigger project it has had a small part in, and I hope the reader can find the literature review interesting to read.

Thank you!

Table of contents

List of tables

List of Figures

Abbreviations

CF	Curd firmness
CN	Casein, occurring in four types; α_{s1}, α_{s2}, β and κ
GWAS	Genome-wide association study
MBP	Mega base pair
MCP	Milk coagulating properties
MFG	Milk fat globule
NC	Non-coagulating milk
QTL	Quantitative trait loci
RCT	Rennet coagulation time
SH	Swedish Holstein, usually abbreviated SLB in Sweden
SJ	Swedish Jersey, usually abbreviated SJB in Sweden
SNP	Single nucleotide polymorphism
SP	Swedish polled, usually abbreviated SKB in Sweden
SR	Swedish red dairy cow, Usually abbreviated SRB in Sweden
VPS35	Vacuolar protein sorting gene
α-LA	Alfa lactalbumin, whey protein
β-LG	Beta lactoglobulin, whey protein

1 Introduction

In Sweden there are three main dairy cow breeds. One of the breeds is the Swedish red cow (SR). The SR population has a total of 93,252 animals and it is the second largest cow breed in Sweden (Swedish Dairy Association, 2016). It is a high producing robust cow (*SRB-föreningen*) yielding an average of 9,747 kg energy corrected milk (ECM) per cow and year. SR produce more protein and fat content in their milk than the most popular dairy breed, Swedish Holstein (SH) (Swedish Dairy Association, 2016). Within the SR population in Sweden, there is a high (18%) prevalence of non-coagulating (NC) milk (Gustavsson *et al.*, 2014a). Milk that doesn't coagulate is a problem for the dairy industry as it means lower cheese yield. Of the total milk that are bought and processed by Swedish dairies 24.4 % is being used for cheese production. If the cheese yield increased with only one percent, it would result in around 17.5 M SEK annual economic gain for the industry. The payment strategy from the dairies today is based on kilo ECM, which also accounts for the protein content. Gustavsson *et al.* (2014b) suggest that there is a moderate (0.38) unfavourable genetic correlation between protein content and NC- milk, hence there is reason to include MCP in breeding programmes. Inclusion of MCP in breeding programmes could prevent further increase of NC milk. Milk that coagulate well is usually characterized by a high relative content of κ-casein (Wedholm *et al.*, 2006). But there is genetic variation within the κ-casein (Fox & McSweeney, 2003) that can affect the coagulation properties (Gustavsson *et al.*, 2014c). In today's breeding programs for livestock animals, especially dairy, for which genomic selection is widely used, it is possible to select animals upon a favourable genetic variant. But before you select an animal based on their genotype, regions in the genome that are affecting the trait must be located. This has been done for the trait NC milk in the SR population by Duchemin *et al.* (2016). A genome wide association study (GWAS) by Duchemin *et al.* (2016) identified a region on *Bos Taurus* chromosome 18 (BTA 18). More specifically, they found an intronic SNP in the VPS35 gene, that is strongly associated to NC milk in SR cows. The discovery of this region was different from other studies where most of the findings were dependent of variation

in the casein cluster of genes located on BTA 6 (Glantz *et al.*, 2015; Gregersen *et al.*, 2015). The purpose of this study was to re-estimate the effect of the SNP found in the GWAS study and evaluate if the finding is reasonable to use as a molecular marker in genomic selection. This was done by genotyping SR cows for the polymorphism in the VPS35 gene. The effect of each genotype in the VPS35 gene on NC milk are going to be estimated by statistical analyses in SAS and DMU.

The hypothesis is that there is an additive or dominant relationship between the genotypes and NC milk, and that one or two of the genotypes will show a significant relationship to NC milk. The SNP can then be used as a marker for selecting SR cows with better milk quality.

2 Background

2.1 Milk

2.1.1 Synthesis of milk

The milk synthesis starts in the udder. The udder of a cow consists of four separate compartments called mammary glands. The mammary gland possesses circular structures called alveoli. The alveolar function is to extract nutrients from the blood, transform the nutrients to the milk and secrete the milk to the alveolar lumen. The epithelial cells located at the base membrane of the alveoli are responsible for the synthesis of milk. The process is activated by hormonal changes at the time of parturition. The precursors for milk are: water, mineral & vitamins, immunoglobulins, amino acids, glucose, and fatty acids. Vitamins & minerals are transported through the epithelial cells directly into the milk. Immunoglobulins are transported through the mammary cells by a process called transcytosis. Glucose, amino acids, and fatty acids undergo transformation in the epithelial cells. Depending on which nutrient is going to be synthesized, different pathways are activated. Glucose is converted to lactose by the enzyme lactose synthetase. Fatty acids are synthesised to triglycerides in the rough endoplasmic reticulum. Amino acids are synthesised to milk proteins in the rough endoplasmic reticulum of the alveolar epithelial cells. The milk proteins are thereafter phosphorylated or glycosylated in the Golgi apparatus before they are transported to the secretory vesicles and the alveolar lumen. The synthesis of milk components is under genetic control. This is important to remember when looking into variation in milk composition (Sjaastad *et al.*, 2010; Wattiaux).

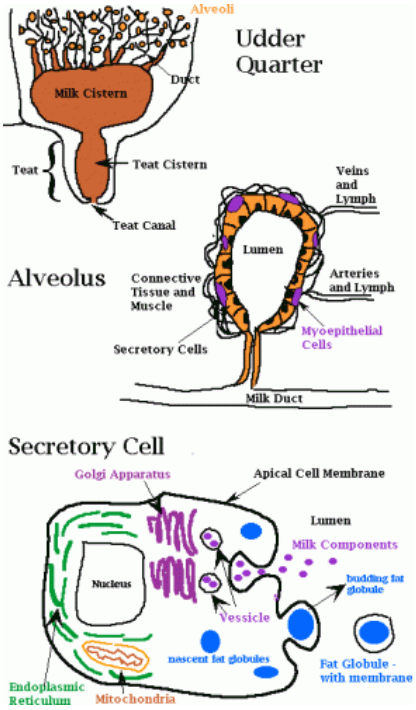


Figure 1. Schematic picture of the milk biosynthesis in the mammary gland (*Milk Biosynthesis, Food Science, University of Guelph*).

2.1.2 Composition of milk

Milk consists of water, lactose, fat, proteins, minerals and vitamins in different concentrations. Lactose is present in 4.5-5.0 % in bovine milk. It is a disaccharide composed of glucose and galactose. The lactose content influences the milk yield. This is due to the osmotic pressure in the mammary gland during milk synthesis, i.e. more water is drawn into the milk if the concentration of lactose is greater. Lactose is a highly fermentable medium. Lactic acid bacteria can hydrolyse the lactose in milk to lactic acid. This reduces the pH of milk and allowing the milk to coagulate. This is an important step in cheese and yogurt processing. (Griffiths, 2010).

The fat content in bovine milk can range between 3-5 %, sometimes even more. It is present in milk in the form of milk fat globules (MFG). The MFG are composed of different types of triglycerides. Different length of the fatty acids chains included in the triglyceride or the amount of saturation gives dairy products much of its texture and flavour. The protein content in bovine milk varies between 3-4 %. Milk proteins mostly consists of caseins (around 80%) and whey proteins (around 20%). Albumin, lactoferrin and lactoperoxidase are also present but in smaller amounts (e.g., Lindmark-Mansson *et al.*, 2003). Casein molecules are clustered together in

complex structures called micelles. Caseins precipitate at pH 4.6 and form a gel network in the milk. This unique property of caseins is important in the process of cheese. Whey proteins are subdivided in lactalbumin (α-LA) and lactoglobulin (β-LG). Whey protein remains soluble in the aqueous part of the milk when pH decreases. Finally, milk contains many minerals and one mineral worth mentioning is called colloidal calcium phosphate and is associated with the casein micelles. They play an important part in micelle stability and cheese process abilities (Griffiths, 2010).

Milk specific protein - Casein

The Swedish chemist J.J. Berzelius published the first paper in the subject 'milk protein' in 1814. Twenty-four years later the Dutch chemist J.G. Mulder described a method for extracting the protein from milk by acid precipitation. The name casein was universally adopted in the early 1900´s, and is defined by a protein that precipitates from milk at pH 4,6. (Fox & McSweeney, 2003). The casein micelle has been widely studied throughout the years after its discovery and it can be described as a spherical aggregate of individual casein molecules. Four types of caseins have been identified: α_{s1}-, α_{s2}-, β- and κ-caseins. κ-caseins are known to be the most critical fraction in micelle stability. The different types of caseins appear in different genetic variant that can alter the processability of milk.

2.1.3 Reasons for variation in milk composition

There are many reasons for variation in milk composition. Milk composition is most variable between species. Thereafter comes breed differences, but variation between individuals within the same breed also occurs (e.g., Poulsen *et al.*, 2013; Stocco *et al.*, 2017). Physiological reason such as stage of lactation, udder health and parity influences the composition (e.g., Lindmark-Mansson *et al.*, 2003; Sjaastad *et al.*, 2010; Bittante *et al.*, 2012). Environmental factors such as feeding strategy and season also affect the composition of milk (Lindmark-Mansson *et al.*, 2003; Griffiths, 2010; Bittante *et al.*, 2012). Feeding more concentrate to lactating dairy cows increases milk yield, fat and protein content significantly (Tyrisevä *et al.*, 2004). Dairy cows are lactating approximately 300 days. Milk yield per day during the lactation is at its peak five to ten weeks after parturition. At peak lactation the concentration of fat and protein are the lowest (Griffiths, 2010). With increasing days in milk (DIM) the content of fat and protein increases (Poulsen *et al.*, 2013). Table 1 shows the average fat, protein and yield for the major four Swedish dairy breeds.

Table 1. *Breed differences in major nutrients and yield in the major four dairy breeds in Sweden.*

Breed	Fat %	Protein%	Average yield ECM/year
SR	4.4	3.6	9 747
SH	4.1	3.4	10 452
SJ	5.9	4.1	8 984
SP	4.2	3.6	5 716

SR = Swedish Red, SH = Swedish Holstein, SJ = Swedish Jersey, SP = Swedish Polled (Swedish Cattle statistics 2017).

2.1.4 Milk coagulation properties

Milk coagulation properties (MCP) is an important trait to study since it affects the processability of milk, e.g. cheese. Cheese making usually begins with adding enzyme to milk. One enzyme that is widely used is Rennet. It is derived from calf stomach. Microbial alternative to the animal derived enzyme also exists and are frequently used. In the presence of Ca^{2+}, the enzyme promotes the coagulation of the caseins in the milk (Fox & McSweeney, 2003). MCP can be analysed by lactodynamography that monitor the viscosity of the milk after addition of rennet. There are three parameters that can be measured. The first one is rennet coagulation time (RCT), which is the time it takes from the addition of rennet to the time when the milk starts to coagulate. The second and third parameter measure curd firmness (CF). One parameter measures the time at a specific CF and the other measures CF at a specific time. (Reviewed by Bittante *et al.*, 2012). While some milk samples have a delayed coagulation time others fail to coagulate, and is commonly called non-coagulating (NC) milk. NC milk have been associated to late lactation stage in Holstein but not in Finnish Ayrshire cows (Tyrisevä *et al.*, 2004). Some regions of the genome and some known genes seem to be associated with NC milk. For example, a relative low content of κ-CN is associated with poorly coagulated milk (e.g., Wedholm *et al.*, 2006) whereas a relative high content of κ-CN has been associated with less casein loss in the whey, leading to higher cheese yield (Hallen *et al.*, 2010). The relative content of whey protein in the protein profile of milk is of importance if cheese making is the purpose of processing milk. Much of the variation in MCP is due to the effect of genotype, crude protein content, individual protein fractions, milk acidity and mineral content (Bittante *et al.*, 2012).

2.2 Genes and genetic variation

Genes are segments in the DNA that encode RNA and polypeptide molecules. The bovine genome consists of more than 22,000 protein-coding genes (Elsik *et al.*,

2009). The expression of the genes is different in the different types of cells. (e.g., Alberts *et al.*, 2010), leading to observed genetic variation between and within species. Between species there is more genetic variation than within species (Gibbs *et al.*, 2009). The key to evolution and modern animal breeding is that there is genetic variation both between and within breeds. Usually animals within the same breed share large proportion of their genomes, hence these animals belong to similar haplotype groups. The genetic diversity can be measured in terms of how many different alleles are occurring within a population. Inbred populations have less genetic diversity because they have more alleles in common. Different cattle breeds are genetically different and this is mostly because they originate from different geographical regions (Blott *et al.*, 1998).

In the bovine genome, 197 protein coding genes that are expressed in the mammary gland have been identified (Lemay *et al.*, 2009). These expressed genes are on average more conserved than other genes in the bovine genome. However, the large variation seen in milk composition may be due to transcriptional and translational regulation of genes that are expressed in the mammary gland and other organs that are involved in the synthesis of milk. The variation that is seen in expression of milk proteins are likely due to copy number variation or interactions with other genes expressed in the mammary gland. Further studies on non-coding regions within the genome, especially those with putative regulatory functions should be investigated as this could be the source of species specific variation in milk composition. (Lemay *et al.*, 2009). The most studied region related to MCP is the *Bos Taurus* casein gene complex located on chromosome 6 (BTA 6).

Single nucleotide polymorphism and its usage

Single nucleotide polymorphism (SNP) is an alteration in a single nucleotide in a specific position on the genome. It is a position where a point mutation has occurred, and it can be an indicator for a change in gene expression for a closely related gene. When the whole bovine genome was sequenced in 2009, the discoveries of thousands of DNA markers, in form of SNP were found (Hayes *et al.*, 2009). In animal breeding SNPs are useful for detecting variation between animals within populations.

Usually animals are genotyped for several thousand SNPs with commercially high-throughput technologies. Genetic variants within SNPs can be used to link a certain variant with a trait. This is done with genome-wide association studies (GWAS), and the success of GWAS are dependent on the large number of markers derived from the genotyping (Brouard *et al.*, 2017). This genetic information about the animal can then be used to estimate its genomic breeding value (GEBV), which is the

sum of the effects of dense genetic markers, or haplotypes of these markers. The GEBV has the potential to capture all the quantitative trait loci that are contributing to variation in a trait (e.g., Hayes *et al.*, 2009). The principle behind genomic selection is based on selecting animals upon its GEBV.

2.2.1 Variation in casein genes responsible for variation in MCP

Within the major milk proteins α_{s1}- α_{s2} -, β - κ-caseins, α-LA and β-LG up to 47 different genetic variants have been observed (Caroli *et al.*, 2009). Some of the variants are only found within a certain breed but there is also individual variation within each breed (Fox & McSweeney, 2003). The three most common dairy breeds in Scandinavia show large differences in MCP and Jersey shows superior MCP compared to the other breeds (Poulsen *et al.*, 2013). Although milk from jersey cows have higher protein and fat content contributing to better MCP, variation in the casein genes seems to be the predominant reason for differences in MCP (Poulsen *et al.*, 2013).

A study by Hallen *et al.* (2007) concluded that κ-CN with allele B was associated with better coagulation properties than the A and E-alleles. An explanation for this is that genotype AB in the κ-CN has been associated with higher concentration of κ-CN than the genotype AA (e.g. Wedholm *et al.*, 2006). The genotype A^2A^2 in the β-CN has been negatively correlated with MCP in SR and SH (Hallen *et al.*, 2007). Poulsen *et al.* (2013) found high prevalence (16%) of NC milk in SR and associated the trait with the variant A^2 in the β-CN gene.

Gustavsson *et al*. (2014a) evaluated the effect of different combinations of genotypes in the α_{s1}-β-κ-casein cluster (CN-cluster) with regards to protein profile in milk of SR cows. Ten different CN-cluster genotypes were evaluated. The most common CN-cluster genotype for SR cows were BB/A^1A^2/AE, followed by BB/A^1A^2/AA which was present in 18 % and 17 % respectively in the sampled population. The authors suggested that the clustered genotype BB/A^1A^2/AB could have a positive effect on milk processabilities due to increased relative concentration of κ- and β-casein. This genotype was only present in 9 % of the SR sampled population (Gustavsson *et al.*, 2014a). The effect of the CN-cluster in relation to the concentration of whey protein in milk had no significant effect on the concentration of α-LA. However, the CN-cluster BB/A^1A^2/AB and BB/A^1A^1/EE has significantly higher concentration of β-LG in SR (Gustavsson *et al.*, 2014c). The relative concentration of whey protein will affect the cheese yield, i.e. more whey protein gives less cheese yield. But higher concentrations of β-LG can be desired in other production lines, such as milk powder (Anema, 2009).

Heritability

The heritability of NC milk in SR cows has been estimated at 0.45, which can be considered to be relatively high (Gustavsson *et al.*, 2014b). The average heritability for MCP measured in RCT from many studies and from different breeds is 0.26 with SD 0.06 (Bittante *et al.*, 2012). There is an moderate unfavourable correlation with NC milk in respect to protein content and milk yield (Gustavsson *et al.*, 2014b). The current payment strategies from Swedish dairies favours total fat and protein content and higher yield. If breeding programmes do not include milk quality traits other than total fat and protein content, the unfavourable correlation could increase NC milk in the SR population.

2.2.2 Other chromosomal regions affecting MCP

Several regions in the bovine genome are known to influence MCP. For example, in Finnish Ayrshire, the ribosome biogenesis protein (BMS1126) on chromosome 2 and BMS1355 on chromosome 18 were associated with NC milk (Tyrisevä *et al.*, 2008). In SR and SH, *Bos Taurus* leptin (LEP) and leptin receptor (LEPR) has been shown to be associated with technological properties of the milk (Glantz *et al.*, 2011). In Brown Swiss, caspase recruitment domain 15 protein (CARD15) and lipin 1 (LPIN1) have been shown to impact on protein and casein content (Cecchinato *et al.*, 2014).

Duchemin *et al.* (2016) discovered a region on chromosome 18 distributed over seven mega base pairs (MBP) including 14 SNPs that explained 7 to 11 % of the phenotypic variation of NC milk in SR. The region was better characterized with region-wide association study and haplotype analyses. One of the strongest associations was found for an intronic SNP in the Vacuolar protein sorting-associated protein 35 (VPS35) gene (Duchemin *et al.*, 2016). The SNPs exact position is chromosome 18:15,046826 (*Reference SNP (refSNP) Cluster Report: rs379827811*). VPS35 is a hydrophobic membrane protein and its function is to deliver organelle specific protein within the cell. The gene is well preserved between species. This indicates that it has an important function. It is highly expressed in brain, heart, testis, ovary, small intestines, spleen and placenta in humans (Zhang *et al.*, 2000). Mao *et al.* (2014) identified expressed protein of the VPS35 gene in human while studying breast cancer. This study is the only one linking the expression of the VPS35 gene to the mammary gland apart from Duchemin *et al.*(2016). Its function is not yet fully understood in cattle. However in human genetic research, mutation in the VPS35 gene have been associated with Parkinson and Alzheimer's disease (Vilarino-Gueell *et al.*, 2011; Wen *et al.*, 2011; Zimprich *et al.*, 2011; Zavodszky *et al.*, 2014). The limitation of the study by Duchemin *et al.* (2016) is that the results

need to be validated to verify if the identified marker is associated with NC milk in SR.

2.2.3 The aim of the thesis

The aim of this thesis is to validate the SNP in the VPS35 gene. In this study, we extracted DNA from the blood samples of SR cows and genotyped a SNP in the VPS35 gene. The validation considers a comparison between three data sets. The first data set contains: a) MCP records from 2010 and 2011 available on SR cows, and includes the genotypes of the casein cluster, and b) the genotyped VPS35 SNP. The second data set is the same as the first data set, except for the inclusion of the imputed VPS35 SNP by Duchemin et al. (2016). The third data set contains: a) MCP records from 2016 and 2017 available on SR cows, and b) the genotyped VPS35 SNP. The evaluation of MCP has just started on this third data set, and therefore, it will be used as an independent data set for this validation.

Secondary aims of this study include: a) estimating of the effect of the SNP in the VPS35 gene, and b) estimating the heritability for NC milk. If the SNP is proven to be significant for NC milk, it may be relevant to include it in breeding programmes for SR cows.

3 Materials and Methods

3.1.1 Animals and sample collection

A total of 1046 SR cows divided in two groups were used. The first sample set included 382 animals from 21 conventional herds in the south of Sweden. The cows were fed according to standard practice and milked 2 or 3 times a day. Lactation stage ranged between 2.5 to 61 weeks. The cows were in parity 1 to 3. They were chosen to be as unrelated as possible, being daughters to 160 sires. The milk and blood samples from those cows were collected in April through May 2010, and in September 2010 through April 2011. The second sample set included 603 animals from 31 conventional herds in the south of Sweden. The cows were fed according to standard practice and milked 2 or 3 times a day. Lactation stage ranged between 0 to 49 weeks. The cows were in parity 1 to 8. They were chosen to be as unrelated as possible. The milk and blood samples from those cows were collected in the winter season of 2016 and 2017.

3.1.2 Phenotypes/NC milk

The phenotype records of the milk were recorded by researchers at Lund university, who kindly allowed me to use their phenotypic records. After collection of individual milk samples at farms, the milk were cooled and transported to Lund University (Lund, Sweden). The milk samples were defatted by centrifugation and stored at 4°C for no longer than 3 days. A rheological test was performed to evaluate MCP. This was done by adding rennet (0.44mL/L Chy-Max Plus, 205 international milk clotting units (IMCU)/mL, Chr. Hansen A/S Hørsholm, Denmark) to fresh milk samples that had been preheated to 32°C for 30 min. RCT and yield strength (YS) was measured with a stresstech rheometer (Reological Instruments AB Lund, Sweden) with low amplitude oscillation. Forty minutes after rennet addition a YS test was performed to evaluate the coagulation properties of individual milk samples.

Samples unable to coagulate within those 40 min were considered NC milk (Poulsen *et al.*, 2013; Gustavsson *et al.*, 2014a). NC milk was scored as a binary trait: a) NC with the value 1 represent NC milk (coagulation time >40 min), and b) NC with a value 0 represents well coagulated milk (coagulation time <40 min).

3.1.3 Genotypes

Isolation of DNA

The practical part in this study was done by me, and this included extraction of DNA from 1046 blood samples belonging to the cows with phenotypic records. The blood samples were received from south of Sweden and were stored in a freezer (-20°C) at SLU. DNA from the blood samples were extracted using Qiasymphony SP® instrument (Qiagen, Hilden, Germany) together with a complementary reagent kit customized for blood samples. The instrument utilizes magnetic-particle technology and comprises four steps: lyse, bind, wash. It elutes the extracted DNA on a new well plate. Manual steps include pipetting 350 µl of each blood sample in tubes from Sarstedt® to fit the instrument. The instrument could handle a maximum of 96 samples at the same time; thus 12 batches were prepared in total. The extracted DNA was labelled and stored in a normal freezer (-20°C) until further processing in this study and for future research projects.

Determine genotypes in the VPS35 gene

The genotyping of the samples was done using StepOnePlus Real-Time PCR system (Life Technologies) together with a custom designed TaqMan SNP Genotyping Assays (Applied Biosystems). The samples were genotyped for the mutation in in the VPS35 gene located on chromosome 18:15046826. Allele frequencies and discrimination plots were obtained from all DNA samples.

3.1.4 Description of the data

The observed genotype for each animal was merged with the phenotypic record. Only animals with both phenotypic and genotypic data could be used to analyse the effect of the VPS35 SNP. The genotype file, which represented all genotyped animals, consisted of 1046 records in total. The NC phenotype file used by Gustavsson et al. (2014) and Duchemin et al. (2016) consisted of 382 records. After animals with missing data were removed, 375 records remained in the first sample set. The phenotype file for the second sample set included 603 records. After animals with missing data were removed, 582 animals remained in the second sample set. A third sample set was also used in the analyses. The third sample set included the same

animals as in the first sample set but the genotype record was the data from Duchemin et al. (2016) imputed data-set, seven additional animals were available for analyses in this data set. The data sets are referred to as Real375, Imputed382 and Real582 throughout the rest of this report.

G-matrix and additional information
The G-matrix for Imputed382 and Real375 was built using the Illumina BovineHD (high density) BeadChip (Illumina Inc., San Diego, CA) containing 777,963 SNPs. The G-matrix for Real582 was built using the Infinium BovineLD (low density) BeadChip (Illumina Inc., San Diego, CA) containing 6,909 SNPs.

The cows sampled between 2010 and 2011 (Imputed382 and Real375) had also been genotyped for genetic variants of αs1-, β-, and κ-caseins using TaqMan SNP genotyping assay. The data of the combined genotypes of caseins were used in this study.

Lactation stage was recorded as weeks in milk (WIM), and the wilmink curve ($e^{-0.05\text{-wim}}$) was added to correct for the dynamics of the lactation curve (Wilmink, 1987).

3.1.5 Statistical analyses

Statistical analyses were performed by using the Statistical Analysis Software SAS (v. 9.4, SAS Institute Inc, Cary, NC). Descriptive statistics were obtained using the statements PROC MEANS and PROC FREQ. One-way ANOVA using the statements PROC GLM was used to determined differences in the effect of the genotype in the VPS35 SNP. Fischer exact test was applied to determine the frequencies of the genotypes in NC- versus WC-milk. The effect of each genotype on NC milk for each animal was estimated using a linear mixed model in DMU software (Madsen & Jensen, 2007). The phenotypic and VPS35 SNP records was fitted in the software together with a G-matrix. The estimated genomic breeding value (GEBV) for each animal was used to calculate correlation between the Real375 and the Imputed382.

Two polygenic animal models were used. Model one is described as follows:

$$Y_{ijkl} = \mu + herd_i + parity_j + b_1 * wim + b_2 * e^{-0.05\text{-wim}} + SNP_k + a_l + e_{ijkl}$$

Where Y_{ijkl} is the phenotype (NC-milk); μ is the overall mean, herd *i* is the fixed effect of a cow belonging to a specific herd; parity *j* is the fixed effect of number of parities per cow*; wim* is the fixed regression effect of weeks in milk; $e^{-0.05\text{-wim}}$ is fixed regression effect of the weeks in milk described in a Wilmink curve; SNP_k is the

fixed effect of genotype in the VPS35 SNP; a_l is the random effect of animal *l* and is assumed to be distributed as $N \sim (0, G\sigma_a^2)$, where G is the genomic relationship matrix based on 375, 382 or 582 animals and σ_a^2 is the additive genetic variance; and e_{ijkl} is the random residual effect and is assumed to be distributed as $N \sim (0, I\sigma_e^2)$ where I is the identity matrix and σ_e^2 is the residual variance.

Model two includes the combined genotype of the CN cluster, which is the only difference with model one. Model two is described as follows:

$$Y_{ijklm} = \mu + herd_i + parity_j + b_1 * wim + b_2 * e^{-0.05\text{-wim}} + CNcluster_k + SNP_l + a_m + e_{ijklm}$$

CNcluster *k* is the fixed effect of the combined genotypes of the casein cluster. This model approach was only done with Real375 and Imputed382. For Real582, records of CN cluster were unavailable.

Heritability estimates for NC milk were calculated as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

Where: h^2 is the heritability, σ_a^2 is the additive genetic variance, and σ_e^2 is the residual variance. The variance components were obtained from the results in DMU.

4 Results and discussion

4.1.1 Descriptive statistics

Table 2 *Means and measures of variance for the traits included in the analyses.*

	Data sets								
	Real375			Imputed382			Real582		
Trait	Mean	SD	CV (%)	Mean	SD	CV (%)	Mean	SD	CV (%)
Wim[1]	26.61	10.99	41.28	26.70	10.98	41.14	26.14	11.76	44.99
Wim005[2]	0.31	0.17	54.18	0.30	0.17	54.25	0.32	0.20	61.34
Parity	1.74	0.73	42.08	1.74	0.73	42.08	2.18	1.22	56.13
NC	0.18	0.38	214.69	0.18	0.38	215.17	0.19	0.39	205.03

1. Wim = weeks in milk,
2. Wim005 = Wilmink term.

Data sets Real375 and Imputed382 include similar animals and phenotypic records and differ only by seven missing animals in the Real375. This resulted in slightly different mean values between those sets. In Real375, first parity was represented by 161 animals (43 %), second parity by 149 animals (40 %) and third parity by 65 animals (17 %). In Imputed382, first parity was represented by 165 animals (43 %), second parity by 152 animals (39 %) and third parity by 65 animals (17 %). Data set Real582 included cows between first and eight parity, where first parity was represented by 204 animals (35 %), second parity by 196 animals (34 %) and third and more parities by 182 animals (31 %). Average number of animals per herd in the Real375 and Imputed382 was 18, with a range from 11-24 cows. Average number of animals per herd in Real582 was 15, with a range from 1-55 cows. The effect of parity and herd on MCP is usually small (reviewed by; Bittante *et al.*, 2012). The effect of week in milk is affecting MCP, where MCP usually is better at the start of lactation, less in mid lactation and recover in the end of the lactation (reviewed by; Bittante *et al.*, 2012). The NC milk in the SR population have increased with one

percent between the cows sampled 2010 through 2011 and the cows sampled between 2016 through 2017. The observed differences in mean values and variation from the mean is not likely to be the cause of higher frequency of NC milk in Real582.

4.1.2 Genotype frequencies

A summary of genotype frequencies, number of animals of each genotype and allele frequencies can be seen in Table 3. The allele frequencies in the VPS35 SNP are similar between the three sets. Unfortunately, 89 genotyped animals had no phenotypic records, and these were excluded from the analyses. Three DNA samples were undetermined due to low amplification in the PCR and these were also excluded from the analyses. In Table 3, only animals that also had phenotypic records were included. The Imputed382 and the Real375 differed in genotypes by seven observations.

Table 3. *Genotype and allele frequencies in VPS35 SNP*.

VPS35 SNP	Real375	Real582	Imputed382
CC	0.5 (n=2)	0.3 (n=2)	0.3 (n=1)
GC	6.1 (n=23)	7.6 (n=44)	5.2 (n=20)
GG	93.3 (n=350)	92.1 (n=536)	92.7 (n=354)
F(G)	0.964	0.954	0.952
F(C)	0.036	0.041	0.029

4.1.3 Genotype frequencies in NC milk and WC milk

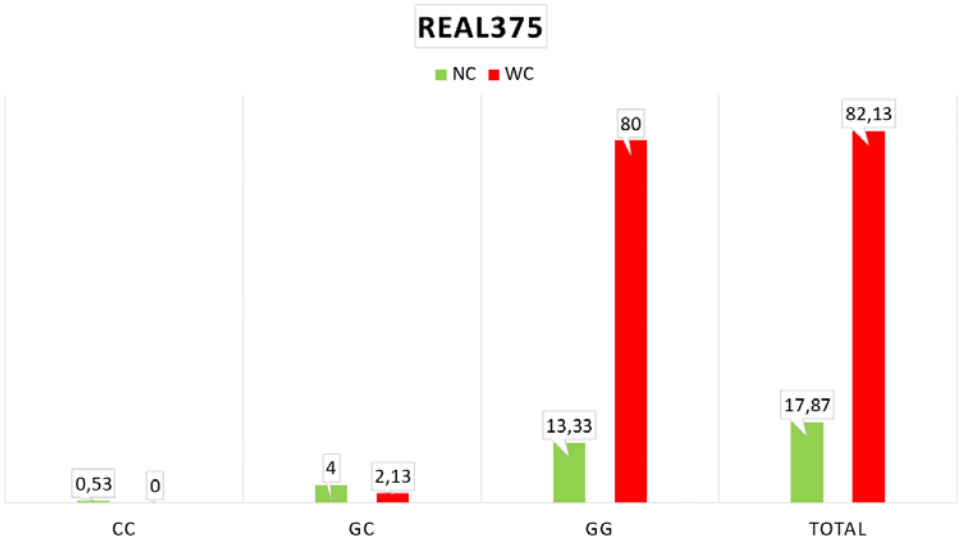


Figure 2. The frequencies (%) of NC and WC milk in the three genotypes in the Real375 (P<.0001).

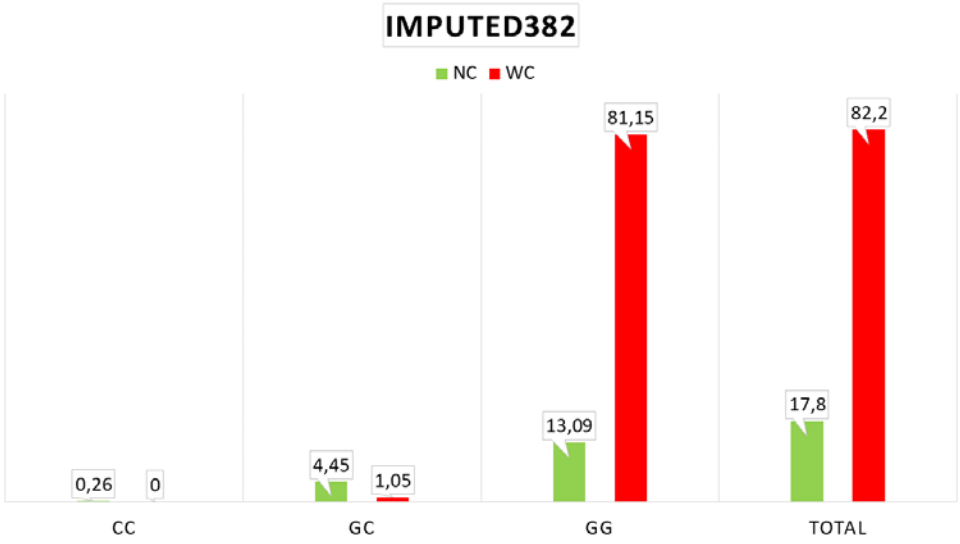


Figure 3. The frequencies (%) of NC and WC milk in the three genotypes in the Imputed382 (P<.0001).

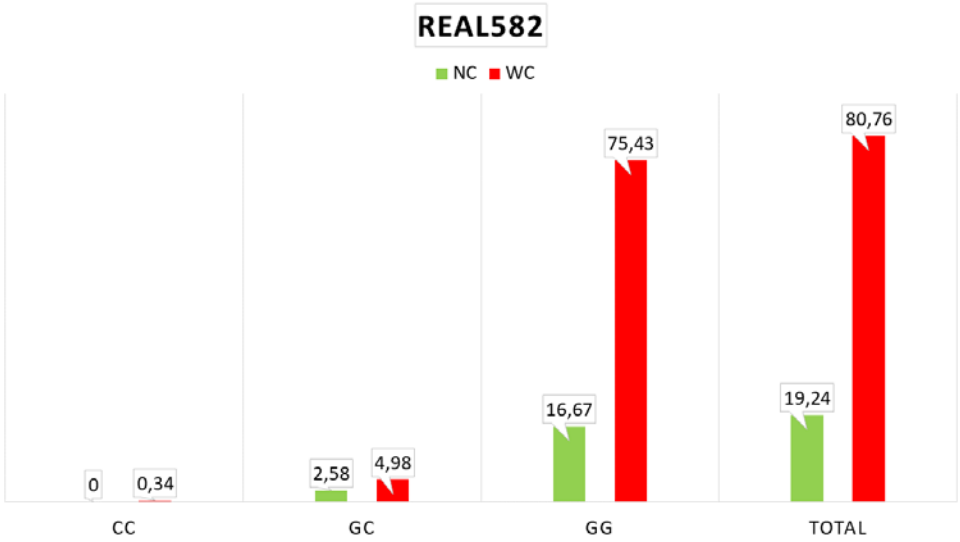


Figure 4. The frequencies (%) of NC and WC milk in the three genotypes in the Real582 (P=0.044).

The frequency of the GC genotype was significantly higher in animals with NC milk in the Real375 and the Imputed382 sets. The P-value from a Fisher's exacts test performed in SAS indicated the significant difference between NC and WC milk in each of the data sets.

4.1.4 Estimates of variance components in SAS

The effect of the VPS35 SNP on NC milk investigated with a F test in a one-way ANOVA in SAS were significant at 0.01 % for the Real375 and the Imputed382 (P=0.0001). The test hypothesis, for which H_0: was defined as: all the mean values between the genotypes are the same, was rejected. The same test showed lower significance (P=0.0276) for the Real582. The variance explained by the VPS35 SNP on NC milk was 3.48, 4.80 and 0.56 % in Real375, Imputed382 and Real582. The estimated difference between the three genotypes can be seen in Table 4. Keeping in mind that the CC was only represented by two animals in the Real375 and Real582 and only one animal in Imputed382. The estimated difference between the CC and the GG genotype might therefore not fully capture the relationship between the differences. However, I choose to include it in the model because of the already small sample size. The correlation between the genotypes in the Imputed382 and the Real375 was 0.77. This correlation reflects the following differences between the data sets: there are 7 missing genotypes and 7 mis-imputed genotypes in Imputed382 compared to Real375.

Table 4. *The estimated difference in mean between the genotypes and the probability that it equals zero (p-value). Standard error in brackets.*

	Real375		Imputed382		Real582	
Genotype	Estimate	P-value	Estimate	P-value	Estimate	P-value
CC	0.86 (0.25)	0.0009	0.86 (0.35)	0.0143	-0.18 (0.28)	0.5157
GC	0.51 (0.077)	<0.0001	0.67 (0.078)	<0.0001	0.16 (0.062)	0.0097
GG	0.00	-	0.00	-	0.00	-
Intercept	0.14	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

For comparison purposes, the model used was NC=VPS35, where VPS35 represent the class variable of the VPS35 genotypes (i.e., CC, GC and GG genotypes). The intercept is conditioning on the GG genotype. The estimates of GC and CC is the difference in mean from their value and the GG value.

4.1.5 Estimates of variance components in DMU

The results from the DMU can be seen in Table 5 and 6. The models that are used are polygenic animal models, for which the additive genetic relationships between animals are modelled using a G-matrix. Two models were run: model one, without the CN cluster and model two, with the CN cluster. Because the CN cluster is only available for Real375 and Imputed382, model two does not show results for Real582.

Model one:

$$Y_{ijkl} = \mu + herd_i + parity_j + b_1 * wim + b_2 * e^{-0.05\text{-}wim} + SNP_k + a_l + e_{ijkl}$$

Model two:

$$Y_{ijklm} = \mu + herd_i + parity_j + b_1 * wim + b_2 * e^{-0.05\text{-}wim} + CNcluster_k + SNP_l + a_m + e_{ijklm}$$

In model one (Table 5), the variance components are similar between Real375 and Imputed382. However, the estimates for genetic variance is lower and the residual variance is higher in Real582 than for the other two sets. The G-matrix for Real582 was built with lower number of SNPs (7 k compared to 777 k). This is likely to be the cause of the lower genetic variance and a higher residual variance seen in Real582. In theory a higher number of markers (SNPs) gives higher accuracies and subsequently lower unexplained variation. Comparing the additive genetic relationship between Real375 built with 777k SNP genotypes and Real582 built with 7k SNP genotypes might therefore not be meaningful.

The additive genetic variance in Real375 was higher in both models than for the Imputed382, resulting in a lower heritability for NC-milk in Imputed382. The difference in additive genetic variance is probably due to the uncertainty intrinsic to imputed data. For the Real582, heritability estimates were much lower than for the two other data sets. The lower heritability reflects the lower genetic variance and the higher residual variance. The fact that we don't have access to the cluster genotypes in the CN region in this study is a pity. We know that the CN cluster have a large effect on NC milk and if we were able to correct for this in this analyse the heritability for these cows are likely to be higher.

In model two (Table 6), including CN cluster, lead to an increase of the genetic variance, and a decrease in the residual variance for Real375. As a result, the estimated heritability increased in model two (0.47) compared to model one (0.39). Interestingly, Imputed382 did not show this trend. This suggest that the results produced by the Imputed382 may contain even more uncertainty than expected, being certainly underestimated and not fully capturing the genetic variation existing in the Real375.

Table 5. *Variances and heritability estimates from DMU, model one.*

Sample set	σ_a^2	σ_e^2	σ_p^2	SE(t)	H^2
Real375	$0.049_{0.22}$	0.078	$0.13_{0.36}$	0.14	0.39
Imputed382	$0.041_{0.20}$	0.079	$0.12_{0.35}$	0.14	0.34
Real582	$0.025_{0.16}$	0.12	$0.15_{0.39}$	0.085	0.17

Lowered value is SD for the additive variance and the phenotypic variance.

Table 6. *Variances and heritability estimates from DMU, model two.*

Sample set	σ_a^2	σ_e^2	σ_p^2	SE	H^2
Real375	$0.056_{0.24}$	0.064	$0.12_{0.35}$	0.15	0.47
Imputed375	$0.038_{0.19}$	0.074	$0.11_{0.33}$	0.15	0.34

Lowered value is SD for the additive variance and the phenotypic variance.

Effect of the genotype

The estimated effects of GG, GC and CC genotypes of the intronic marker in the VPS35 gene affecting NC milk can be seen in Table 7 and 8. In model one (Table 8), the C allele seems to behave in a rather additive manner for Real375 and Imputed382. In addition, the estimated effects for these two data sets are in the same direction, but the underestimation of the Imputed382 is noticeable here too. For Real582, GG and GC estimates are similar suggesting that the effects could not be disentangled from one another. This reflects the lower genetic variance estimated

for Real582 in model one. The inconsistent result between the cows sampled between 2010 through 2011 and 2016 through 2017 may indicate that the results from the earlier study by Duchemin *et al.*, (2016) is a false positive and that the VPS35 SNP genotyped in this thesis is not a causal variant associated with NC milk. Nevertheless, there is also a possibility of sample contamination during the DNA extraction and PCR.

In model two (Table 8), we show the estimated effects of GG, GC and CC genotypes of the intronic marker in the VPS35 gene affecting NC milk when the CN cluster is included. Real375 and Imputed 382 results are in the same direction, however for GG, the estimates seem to be more similar between the two sets (table 8) than the ones estimated without the CN cluster (Table 7). Differences between the estimated effects of GG genotypes between model one and model two (Table 7 and Table 8) might not be caused by differences in the allele frequencies between the two data sets (Table 3). In Real375, the frequency of G i.e., f(G) equals 0.964 and the frequency of C equals 0.036. In Imputed382, f(G) equals 0.953 and f(C) equals 0.029.

Table 7. *Estimated effect of the three genotypes (GG, GC, CC) affecting NC milk in model one.*

	Real375	Imputed382	Real582
CC	0	0	0
GC	$-0.583_{0.263}$	$-0.277_{0.360}$	$0.380_{0.288}$
GG	$-1.010_{0.254}$	$-0.876_{0.352}$	$0.238_{0.283}$

Lowered values are SE for each effect.

Table 8. *Estimated effect of the three genotypes (GG, GC, CC) affecting NC milk in model two.*

	Real375	Imputed382
CC	0	0
GC	$-0.529_{0.254}$	$-0.303_{0.351}$
GG	$-0.957_{0.246}$	$-0.916_{0.343}$

Lowered values are SE for each effect.

Effect of the CN cluster

Fifteen different CN clusters were observed of which cluster BC/A^1A^2/AB was represented by one animal and cluster BB/A^1A^2/AA by 68 animals (Table 9). The estimated effect, standard error and the frequency of each CN cluster in data set Real375 can be seen in Table 9. The most frequent CN combined genotype (CN cluster) was represented by 68 animals and it was BB/A1A2/AA. The largest effect on NC milk had CN cluster BB/A1A1/BE with -0.63, only four animals had this CN cluster type. The frequencies of the CN cluster genotypes are identical to Gustavsson *et al.* (2014a) study. However, the estimated effect of the CN cluster for NC milk appears

to be different than their results. Their results suggest that BB/A1A2/AB could have positive effect on the cheese processabilities. The results in this study indicate that CN cluster genotype BB/A1A1/BE is the best for MCP in SR cows. And that BC/A2A2/AB is negative for MCP. But in Gustavsson *et al.*, 2014a, they are only reporting CN cluster with a frequency higher than 5 %. Therefore only nine CN cluster are avaiable in their results. In this analysis, every CN cluster that were available in the SR sample population was included. This was in total 15 CN cluster. Because NC milk in the analyses is treated as a binary trait and there are 15 different combined CN genotypes, it is likely that some combinations are better than others. And that some of the combined CN clusters are exhibiting both NC milk and well coagulated milk. For the BB/A1A1/BE genotype that deviates the most from zero, all animals exhibited well coagulated milk (0) and all animals had genotype GG in the VPS35 SNP. For the CN cluster BC/A2A2/AB that has an estimate of zero, one of the animal exhibit well coagulated milk and the other exhibit NC milk, both animals had genotype GG in the VPS35 SNP.

Table 9. *Estimated genetic effect of the 15 different CN clusters in model two.*

CNcluster	No. obs	Estimated effect	SE
BB/A1A1/AA	19	-0.485738	0.243901
BB/A1A1/AB	5	-0.349680	0.284723
BB/A1A1/AE	42	-0.492395	0.237760
BB/A1A1/BE	4	-0.625927	0.288147
BB/A1A1/EE	11	-0.499366	0.254924
BB/A1A2/AA	68	-0.349093	0.232377
BB/A1A2/AB	29	-0.400412	0.238935
BB/A1A2/AE	67	-0.274211	0.235827
BB/A1A2/BB	7	-0.584496	0.269219
BB/A1A2/BE	25	-0.358466	0.242898
BB/A2A2/AA	55	-0.144899	0.235233
BB/A2A2/AB	38	-0.401326	0.236746
BB/A2A2/BB	2	-0.392296	0.337124
BC/A1A2/AB	1	-0.389184	0.421254
BC/A2A2/AB	2	0.000	0.000

Correlation between the Real375 and Imputed382

The animals genomic breeding value (GEBV) was estimated for the three data sets in DMU. The correlation between the estimates for Real375 and Imputed382 was 0.975 in model one and 0.965 in model two. The high correlation between the Real375 and the Imputed382 supports that the GWAS study conducted by Duchemin *et al.*, (2016) was successful.

4.1.6 Breeding for better MCP

The frequency of NC milk has increased from 18 to 19 % between animals sampled 2010 through 2011 and 2016 through 2017. This is an indication that a measurement of MCP should be included in the breeding programs for SR cows to prevent further increase of NC milk. Unfortunately, the results from this study can't assure that the inclusion of the VPS35 SNP in breeding programmes for SR cows would reduce the frequency on NC milk.

However, there are other things that can be included in breeding programmes for dairy cattle that successfully enhances MCP. For example, in Italy, the bulls are selected upon their genotype in the κ-casein gene. Bulls with the genotype AB or BB are favourable for cheese processing. Genotype AA is not favourable. The genotypes have different properties due to different charge of the amino acids. Variant A in the κ-casein gene have an aspartic acid residue at position 148 as opposed to B that has an asparagine residue at the same position. Aspartic acid provides an extra negative charge of the A variant compare to the B variant. The difference in charge between the two variants results in some different milk processing characteristics (Griffiths, 2010). ANAFI (Italian Friesian Breeders Association) has manage to increase the overall protein and casein content in dairy milk after including an index for MCP in their breeding program (Summer *et al.*, 2014). It would be a good idea to include some measurement of milk quality other than total fat and protein content in the breeding programme for Swedish dairy cows. Evaluation of MCP can be done by measuring RCT and CF which reflects the protein content, fat content, variant of κ-casein and somatic cell count. (Summer *et al.*, 2014; *http://www.anafi.it/*).

Milk quality traits, such as NC milk, is a quantitative trait regulated by several genes. Lemay *et al.* (2009) found 197 protein coding genes in the mammary gland. One can imagine the large variations that exists between those genes with different genotype combination patterns, which adds to the complexity to find specific genes responsible for milk quality. NC milk in this report and in many other studies is treated as a binary trait even though the trait is measured as a continuous value. This is because it is convenient for the analyses to treat it as a binary trait. Determine milk that doesn't coagulate within 40 minutes as NC milk is reasonable to use as evaluation of the trait. Future studies could include rheological test in the standard procedure for individual milk sample test conducted by the Swedish cow data base (kokontrollen). Measurements of such kind would make it possible to include MCP in breeding programmes. The large number of observations that this would result in could be combined with genotyped cows. It is then possible to make accurate decision in genomic selection for dairy cows with better MCP.

The expression ‘genomic selection’ was coined by Meuwissen *et al.* (2001) that showed in a simulation study that it was possible to make very accurate selection decisions based on genomic breeding values (GEBV). The article is one of the most cited in this area. Based on suggestions from the study Schaeffer (2006) predicted quicker genetic gain and shorter generation interval, which would subsequently, lower costs for breeding companies since less phenotype records will be needed. However, the GEBV must be based on a reference population with known phenotype and genotypes to get accuracies of the predictions. Since the breakthrough, genomic selection have replaced much of the progeny testing within cattle breeding. In plant breeding, genomic selection is currently being used as well as for human disease research. (de Koning, 2016). The cost of genotyping animals has also decreased dramatically in recent years, which contributes to more frequent usage. It’s going to be interesting to follow the development of animal breeding in the near future since new technology makes it possible to sequence whole genomes. And a myriad of information, hidden in the genome, is yet to be explored.

4.2 Conclusion

The frequency of NC milk has increased in the SR population. The validation of the VPS35 SNP, that were associated to NC milk in the study by Duchemin *et al.*, 2016 was successful, showing high correlation between the imputed SNP and the real genotyped SNP. However, the results from this study can’t assure that inclusion of the VPS35 SNP in breeding programmes for SR cows would enhance MCP. This is due to inconsistent results of the effect of the VPS35 SNP between animals sampled between 2010 - 2011 and 2016 - 2017. The inconsistent result between the old and the new samples could be because of a false positive result from the earlier study. The VPS35 SNP may not be associated to NC milk in SR cows as previously suggested. MCP and QTLs affecting MCP should be further investigated to prevent increase of NC milk in dairy cattle.

References

Alberts, B., Bray, D., Hopkin, K. & Johnson, A. (2010). *Essential Cell Biology*. 3rd ed. USA: Garland Science. ISBN 978-0-8153-4130-7.

Anema, S. G. (2009). *The whey proteins in milk: thermal denaturation, physical interactions and effects on the functional properties of milk.*(Thompson, A., Boland, M., & Singh, H., Eds) San Diego: Elsevier Academic Press Inc. ISBN 978-0-08-092068-9.

Bittante, G., Penasa, M. & Cecchinato, A. (2012). Invited review: Genetics and modeling of milk coagulation properties. *Journal of Dairy Science*, 95(12), pp 6843–6870.

Blott, S. C., Williams, J. L. & Haley, C. S. (1998). Genetic relationships among European cattle breeds. *Animal Genetics*, 29(4), pp 273–282.

Brouard, J.-S., Boyle, B., Ibeagha-Awemu, E. M. & Bissonnette, N. (2017). Low-depth genotyping-by-sequencing (GBS) in a bovine population: strategies to maximize the selection of high quality genotypes and the accuracy of imputation. *Bmc Genetics*, 18, p 32.

Caroli, A. M., Chessa, S. & Erhardt, G. J. (2009). Invited review: Milk protein polymorphisms in cattle: Effect on animal breeding and human nutrition. *Journal of Dairy Science*, 92(11), pp 5335–5352.

Cecchinato, A., Ribeca, C., Chessa, S., Cipolat-Gotet, C., Maretto, F., Casellas, J. & Bittante, G. (2014). Candidate gene association analysis for milk yield, composition, urea nitrogen and somatic cell scores in Brown Swiss cows. *animal*, 8(7), pp 1062–1070.

Duchemin, S. I., Glantz, M., de Koning, D.-J., Paulsson, M. & Fikse, W. F. (2016). Identification of QTL on Chromosome 18 Associated with Non-Coagulating Milk in Swedish Red Cows. *Frontiers in Genetics*, 7, p 57.

Elsik, C. G., Tellam, R. L. & Worley, K. C. (2009). The Genome Sequence of Taurine Cattle: A window to ruminant biology and evolution. *Science (New York, N.Y.)*, 324(5926), pp 522–528.

Fox, P. F. & McSweeney, P. L. H. (2003). *Advanced Dairy Chemistry* [online]. Third edition. SPRINGER US. Available from: https://books.google.se/books?id=Pfe_7Me5_w4C&pg=PR4&lpg=PR4&dq=0-306-47271-6&source=bl&ots=a2Va9aumHs&sig=MCQGru1TZa5vyNEVCIYv2Y2MwFs&hl=en&sa=X&ved=0ahUKEwijpsuDq_DYAhUChiwKHZcaAZkQ6AEIKzAB#v=onepage&q=0-306-47271-6&f=false.

Gibbs, R. A., Taylor, J. F., Van Tassell, C. P., Barendse, W., Eversoie, K. A., Gill, C. A., Green, R. D., Hamernik, D. L., Kappes, S. M., Lien, S., Matukumalli, L. K., McEwan, J. C., Nazareth, L. V., Schnabel, R. D., Taylor, J. F., Weinstock, G. M., Wheeler, D. A., Ajmone-Marsan, P., Barendse, W., Boettcher, P. J., Caetano, A. R., Garcia, J. F., Hanotte, O., Mariani, P., Skow, L. C., Williams, J. L., Caetano, A. R., Diallo, B., Green, R. D., Hailemariam, L., Hanotte, O., Martinez, M. L., Morris, C. A., Silva, L. O. C., Spelman, R. J., Taylor, J. F., Mulatu, W., Zhao, K., Abbey, C. A., Agaba, M., Araujo, F. R., Bunch, R. J., Burton, J., Gill, C. A., Gorni, C., Olivier, H., Harrison, B. E., Luff, B., Machado, M. A., Mariani, P., Morris, C. A., Mwakaya, J., Plastow, G., Sim, W., Skow, L. C., Smith, T., Sonstegard, T. S., Spelman, R. J., Taylor, J. F., Thomas, M. B., Valentini, A., Williams, P., Womack, J., Wooliams, J. A., Liu, Y., Qin, X., Worley, K. C., Gao, C., Gill, C. A., Jiang, H., Liu, Y., Moore, S. S., Nazareth, L. V., Ren, Y., Song, X.-Z., Bustamante, C. D., Hernandez, R. D., Muzny, D. M., Nazareth, L. V., Patil, S., Lucas, A. S., Fu, Q., Kent, M. P., Moore, S. S.,

Vega, R., Abbey, C. A., Gao, C., Gill, C. A., Green, R. D., Matukumalli, L. K., Matukumalli, A., McWilliam, S., Schnabel, R. D., Sclep, G., Ajmone-Marsan, P., Bryc, K., Bustamante, C. D., Choi, J., Gao, H., Grefenstette, J. J., Murdoch, B., Stella, A., Villa-Angulo, R., Wright, M., Aerts, J., Jann, O., Negrini, R., Sonstegard, T. S., Williams, J. L., Taylor, J. F., Villa-Angulo, R., Goddard, M. E., Hayes, B. J., Barendse, W., Bradley, D. G., Boettcher, P. J., Bustamante, C. D., da Silva, M. B., Lau, L. P. L., Liu, G. E., Lynn, D. J., Panzitta, F., Sclep, G., Wright, M. & Dodds, K. G. (2009). Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science*, 324(5926), pp 528–532.

Glantz, M., Gustavsson, F., Bertelsen, H. P., Stalhammar, H., Lindmark-Mansson, H., Paulsson, M., Bendixen, C. & Gregersen, V. R. (2015). Bovine chromosomal regions affecting rheological traits in acid-induced skim milk gels. *Journal of Dairy Science*, 98(2), pp 1273–1285.

Glantz, M., Mansson, H. L., Stalhammar, H. & Paulsson, M. (2011). Effect of polymorphisms in the leptin, leptin receptor, and acyl-coenzyme A:diacylglycerol acyltransferase 1 (DGAT1) genes and genetic polymorphism of milk proteins on cheese characteristics. *Journal of Dairy Science*, 94(7), pp 3295–3304.

Gregersen, V. R., Gustavsson, F., Glantz, M., Christensen, O. F., Stålhammar, H., Andrén, A., Lindmark-Månsson, H., Poulsen, N. A., Larsen, L. B., Paulsson, M. & Bendixen, C. (2015). Bovine chromosomal regions affecting rheological traits in rennet-induced skim milk gels. *Journal of Dairy Science*, 98(2), pp 1261–1272.

Griffiths, M. W. (2010). *Improving the Safety and Quality of Milk, Volume 1 - Milk Production and Processing* [online]. Woodhead Publishing. Available from: https://app.knovel.com/web/toc.v/cid:kpISQMVMP3/viewerType:toc/root_slug:viewerType%3Atoc/url_slug:root_slug%3Aimproving-safety-quality?kpromoter=federation. [Accessed 2018-03-12].

Gustavsson, F., Buitenhuis, A. J., Glantz, M., Stalhammar, H., Lindmark-Mansson, H., Poulsen, N. A., Larsen, L. B., Andren, A. & Paulsson, M. (2014a). Impact of genetic variants of milk proteins on chymosin-induced gelation properties of milk from individual cows of Swedish Red dairy cattle. *International Dairy Journal*, 39(1), pp 102–107.

Gustavsson, F., Glantz, M., Poulsen, N. A., Wadsö, L., Stålhammar, H., Andrén, A., Lindmark Månsson, H., Larsen, L. B., Paulsson, M. & Fikse, W. F. (2014c). Genetic parameters for rennet- and acid-induced coagulation properties in milk from Swedish Red dairy cows. *Journal of Dairy Science*, 97(8), pp 5219–5229.

Gustavsson, F., Buitenhuis, A. J., Johansson, M., Bertelsen, H. P., Glantz, M., Poulsen, N. A., Mansson, H. L., Stalhammar, H., Larsen, L. B., Bendixen, C., Paulsson, M. & Andren, A. (2014b). Effects of breed and casein genetic variants on protein profile in milk from Swedish Red, Danish Holstein, and Danish Jersey cows. *Journal of Dairy Science*, 97(6), pp 3866–3877.

Hallen, E., Allmere, T., Naslund, J., Andren, A. & Lunden, A. (2007). Effect of genetic polymorphism of milk proteins on rheology of chymosin-induced milk gels. *International Dairy Journal*, 17(7), pp 791–799.

Hallen, E., Lunden, A., Allmere, T. & Andren, A. (2010). Casein retention in curd and loss of casein into whey at chymosin-induced coagulation of milk. *Journal of Dairy Research*, 77(1), pp 71–76.

Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), pp 433–443.

http://www.anafi.it/. [online]. Available from: http://www.anafi.it/english/. [Accessed 2018-02-26].

de Koning, D.-J. (2016). Meuwissen et al. on Genomic Selection. *Genetics*, 203(1), pp 5–7.

Lemay, D. G., Lynn, D. J., Martin, W. F., Neville, M. C., Casey, T. M., Rincon, G., Kriventseva, E. V., Barris, W. C., Hinrichs, A. S., Molenaar, A. J., Pollard, K. S., Maqbool, N. J., Singh, K., Murney, R., Zdobnov, E. M., Tellam, R. L., Medrano, J. F., German, J. B. & Rijnkels, M. (2009). The bovine lactation genome: insights into the evolution of mammalian milk. *Genome Biology*, 10(4), p R43.

Lindmark-Mansson, H., Fonden, R. & Pettersson, H. E. (2003). Composition of Swedish dairy milk. *International Dairy Journal*, 13(6), pp 409–425.

Madsen, P. & Jensen, J. (2007). A user's guide to DMU. A Package for Analysing Multivariate Mixed Models. Version 6, release 5.2. *Aarhus University, Tjele, Denmark.*, p 32.
Mao, J., Ladd, J., Gad, E., Rastetter, L., Johnson, M. M., Marzbani, E., Childs, J. S., Lu, H., Dang, Y., Broussard, E., Stanton, S. E., Hanash, S. M. & Disis, M. L. (2014). Mining the pre-diagnostic antibody repertoire of TgMMTV-neu mice to identify autoantibodies useful for the early detection of human breast cancer. *Journal of Translational Medicine*, 12, p 121.
Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), pp 1819–1829.
Milk Biosynthesis | Food Science. [online]. Available from: https://www.uoguelph.ca/foodscience/book-page/milk-biosynthesis. [Accessed 2018-04-19].
Poulsen, N. A., Bertelsen, H. P., Jensen, H. B., Gustavsson, F., Glantz, M., Mansson, H. L., Andren, A., Paulsson, M., Bendixen, C., Buitenhuis, A. J. & Larsen, L. B. (2013). The occurrence of noncoagulating milk and the association of bovine milk coagulation properties with genetic variants of the caseins in 3 Scandinavian dairy breeds. *Journal of Dairy Science*, 96(8), pp 4830–4842.
Reference SNP (refSNP) Cluster Report: rs379827811. [online]. Available from: https://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?rs=379827811&pt=1TyULNYA1TKLD2VoFd8Iof_Xy6DdRyGSUa8_1LCgwxDJTWyJGSV. [Accessed 2018-03-21].
Sjaastad, O., Sand, O. & Hove, K. (2010). *Physiology of Domestic Animals*. 2nd edition. Olslo: Scandinavian Veterinary Press. ISBN 978-82-91743-07-3.
SRB-föreningen. Available from: http://srb-foreningen.se/. [Accessed 2018-01-22].
Stocco, G., Cipolat-Gotet, C., Bobbo, T., Cecchinato, A. & Bittante, G. (2017). Breed of cow and herd productivity affect milk composition and modeling of coagulation, curd firming, and syneresis. *Journal of dairy science*, 100(1), pp 129–145.
Summer, A., Franceschi, P., Formaggioni, P. & Malacarne, M. (2014). Characteristics of raw milk produced by free-stall or tie-stall cattle herds in the Parmigiano-Reggiano cheese production area. *Dairy Science & Technology*, 94(6), pp 581–590.
Swedish Dairy Association (2016). *Swedish Dairy Association*. Stockholm, Sweden.
Tyriseva, A. M., Vahlsten, T., Ruottinen, O. & Ojala, M. (2004). Noncoagulation of milk in Finnish Ayrshire and Holstein-Friesian cows and effect of herds on milk coagulation ability. *Journal of Dairy Science*, 87(11), pp 3958–3966.
Tyriseva, A.-M., Elo, K., Kuusipuro, A., Vilva, V., Janonen, I., Karjalainen, H., Ikonen, T. & Ojala, M. (2008). Chromosomal Regions Underlying Noncoagulation of Milk in Finnish Ayrshire Cows. *Genetics*, 180(2), pp 1211–1220.
Vilarino-Gueell, C., Wider, C., Ross, O. A., Dachsel, J. C., Kachergus, J. M., Lincoln, S. J., Soto-Ortolaza, A. I., Cobb, S. A., Wilhoite, G. J., Bacon, J. A., Behrouz, B., Melrose, H. L., Hentati, E., Puschmann, A., Evans, D. M., Conibear, E., Wasserman, W. W., Aasly, J. O., Burkhard, P. R., Djaldetti, R., Ghika, J., Hentati, F., Krygowska-Wajs, A., Lynch, T., Melamed, E., Rajput, A., Rajput, A. H., Solida, A., Wu, R.-M., Uitti, R. J., Wszolek, Z. K., Vingerhoets, F. & Farrer, M. J. (2011). VPS35 Mutations in Parkinson Disease. *American Journal of Human Genetics*, 89(1), pp 162–167.
Wattiaux, M. A.MILK SECRETION IN THE UDDER OF A DAIRY COW. Babcock Institute. Available from: http://moodle.najah.edu/pluginfile.php/63715/mod_resource/content/1/Milk%20secretion%20in%20the%20udder.pdf.
Wedholm, A., Larsen, L. B., Lindmark-Mansson, H., Karlsson, A. H. & Andren, A. (2006). Effect of protein composition on the cheese-making properties of milk from individual dairy cows. *Journal of Dairy Science*, 89(9), pp 3296–3305.
Wen, L., Tang, F.-L., Hong, Y., Luo, S.-W., Wang, C.-L., He, W., Shen, C., Jung, J.-U., Xiong, F., Lee, D., Zhang, Q.-G., Brann, D., Kim, T.-W., Yan, R., Mei, L. & Xiong, W.-C. (2011). VPS35 haploinsufficiency increases Alzheimer's disease neuropathology. *Journal of Cell Biology*, 195(5), pp 765–779.
Wilmink, J. (1987). Adjustment of Test-Day Milk, Fat and Protein Yield for Age, Season and Stage of Lactation. *Livestock Production Science*, 16(4), pp 335–348.
Zavodszky, E., Seaman, M. N. J., Moreau, K., Jimenez-Sanchez, M., Breusegem, S. Y., Harbour, M. E. & Rubinsztein, D. C. (2014). Mutation in VPS35 associated with Parkinson's disease

impairs WASH complex association and inhibits autophagy. *Nature Communications*, 5, p 3828.

Zhang, P. Z., Yu, L., Gao, J., Fu, Q., Dai, F. Y., Zhao, Y., Zheng, L. & Zhao, S. Y. (2000). Cloning and characterization of human VPS35 and mouse Vps35 and mapping of VPS35 to human chromosome 16q13-q21. *Genomics*, 70(2), pp 253–257.

Zimprich, A., Benet-Pages, A., Struhal, W., Graf, E., Eck, S. H., Offman, M. N., Haubenberger, D., Spielberger, S., Schulte, E. C., Lichtner, P., Rossle, S. C., Klopp, N., Wolf, E., Seppi, K., Pirker, W., Presslauer, S., Mollenhauer, B., Katzenschlager, R., Foki, T., Hotzy, C., Reinthaler, E., Harutyunyan, A., Kralovics, R., Peters, A., Zimprich, F., Bruecke, T., Poewe, W., Auff, E., Trenkwalder, C., Rost, B., Ransmayr, G., Winkelmann, J., Meitinger, T. & Strom, T. M. (2011). A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease. *American Journal of Human Genetics*, 89(1), pp 168–175.

Acknowledgements

Thanks to Sandrine that have been a great supervisor and support throughout this thesis. Thanks to Kim also for being a great supervisor, especially during the practical part of the thesis. Thanks to the milk group in Lund who I paid a visit during the autumn. It was very interesting, and it added a lot to the understanding of my literature. Thanks to other people who have helped me in the process of my thesis!

Appendix

Scripts

The scripts are descriptions of some edits made in the data sets. Six animals could at first not be matched with the genotype record. But, after some studying of the raw data, they were identified and could be added to the data set. Character values needed to be transcribed to numbers in order to do analyses in DMU.

```
data exjobb.VPS_382_edited1;
set exjobb.VPS_382;
if Lab_sample = "A34" then VPS35_SNP = "GG";
if Lab_sample = "A56" then VPS35_SNP = "GG";
if Lab_sample = "B57" then VPS35_SNP = "GG";
if Lab_sample = "D96" then VPS35_SNP = "GC";
if Lab_sample = "E2" then VPS35_SNP = "GG";
if Lab_sample = "0" then VPS35_SNP = "-99";

*change the clusters to numeric characters;
data exjobb.char_to_numbers;
set exjobb.VPS_375_edited3;
if cncluster= "BB/A1A1/AA" then cnclusternr= 1;
if cncluster= "BB/A1A1/AB" then cnclusternr= 2;
if cncluster= "BB/A1A1/AE" then cnclusternr= 3;
if cncluster= "BB/A1A1/BE" then cnclusternr= 4;
if cncluster= "BB/A1A1/EE" then cnclusternr= 5;
if cncluster= "BB/A1A2/AA" then cnclusternr= 6;
if cncluster= "BB/A1A2/AB" then cnclusternr= 7;
if cncluster= "BB/A1A2/AE" then cnclusternr= 8;
if cncluster= "BB/A1A2/BB" then cnclusternr= 9;
if cncluster= "BB/A1A2/BE" then cnclusternr= 10;
if cncluster= "BB/A2A2/AA" then cnclusternr= 11;
if cncluster= "BB/A2A2/AB" then cnclusternr= 12;
if cncluster= "BB/A2A2/BB" then cnclusternr= 13;
if cncluster= "BC/A1A2/AB" then cnclusternr= 14;
if cncluster= "BC/A2A2/AB" then cnclusternr= 15;

data exjobb.char_to_numbers2;
set exjobb.char_to_numbers;
if VPS35_SNP = "GG" then genotype_no=11;
if VPS35_SNP = "GC" then genotype_no=12;
if VPS35_SNP = "CC" then genotype_no=22;
```

Additional test models in SAS with data set *Real375*

A general linear model (GLM) in SAS with only the CN cluster fitted as a fixed effect also showed significant effect.

Table 1. *PROC GLM, model; nc=cncluster. R-square: 0.096, C.V.: 208.11. dataset Real375.*

Source	Df	SS	MS	F-value	P-value
Model	14	5.26	0.36	2.72	0.0008
Error	360	49.77	0.34		
Corrected total	375	55.03			

Association analysis in SAS

An association analyse were performed in SAS to test the effect of the SNP for milk coagulation properties. The cows with genotype GC, heterozygote, exhibited significantly more milk that didn't coagulate within 40 minutes after rennet addition.

Table 2. *Least square means of the effect of the SNP*

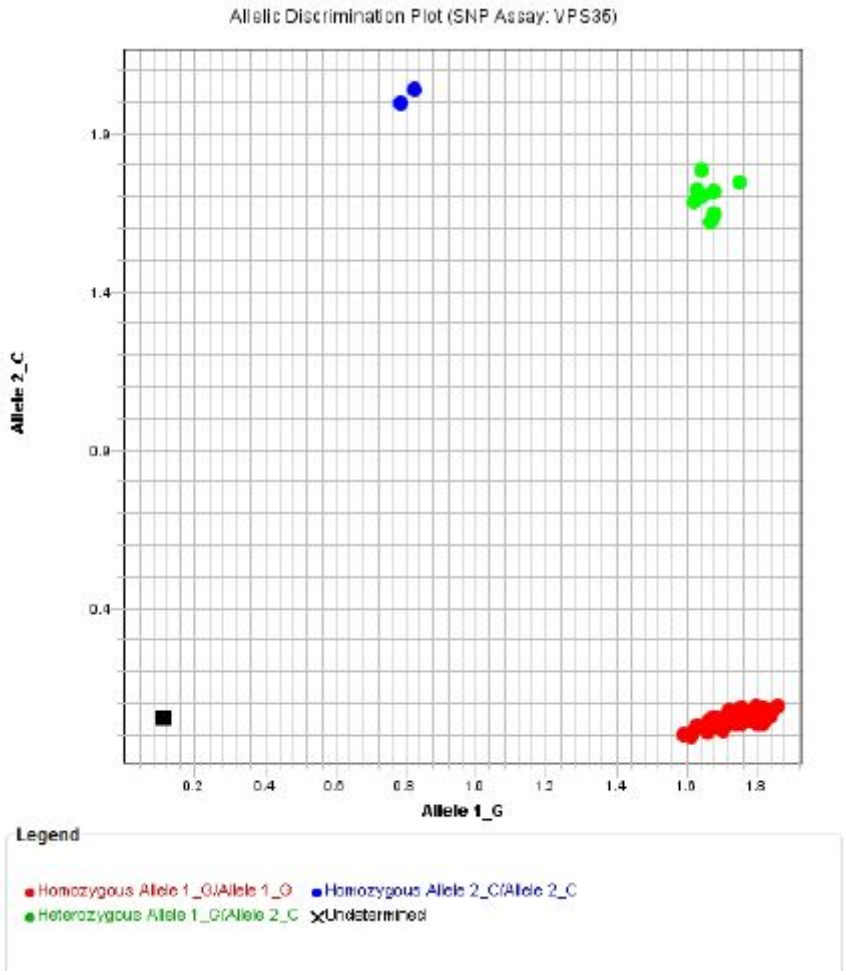
Effect	VPS35_SNP	LSM	SE	DF	T Value	Pr>|t|
VPS35_SNP	CC	0.9565	0.2527	15	3.78	0.0018
VPS35_SNP	GC	0.5878	0.081	15	7.23	<.0001
VPS35_SNP	GG	0.1048	0.041	15	2.55	0.0224

Table 3. *Differences of least square means.*

Effect	VPS35	_Vps35_	ΔLSM	SE	DF	T Value	Pr>|t|
VPS35	CC	GC	0.3687	0.2597	15	1.42	0.1762
VPS35	CC	GG	0.8517	0.2501	15	3.41	0.0039
VPS35	GC	GG	0.4831	0.0765	15	6.32	<.0001

Allelic discrimination plot

One example of an allelic discrimination plot obtained from the genotyping procedure during the data collection part of the project.


Allelic Discrimination Plot (SNP Assay: VPS35)
Allele 2_C
Allele 1_G
1.0
1.4
0.8
0.4
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
Legend
Homozygous Allele 1_G/Allele 1_G
Homozygous Allele 2_C/Allele 2_C
Heterozygous Allele 1_G/Allele 2_C
Undetermined