Norwegian University
of Life Sciences

Master's Thesis 2019 60 ECTS
Faculty of Biosciences

Linkage disequilibrium in the chromosomal region surrounding the prion protein gene (Prnp) in Norwegian dairy goat

Mie Marie Hammervold
Biotechnology

Acknowledgements

I would like to express my special thanks to my main supervisor Dag Inge Våge for always being ready to answer whatever questions I might have, and my second supervisor Michael A. Tranulis for always sharing his knowledge and enthusiasm.

Secondly, I would like to thank Susan Skogtvedt Røed for information about how the genotyping for the *PRNP*-mutation was done.

Eventually I would like to express my gratitude to my family and friends for always cheering me on - especially to my dad for helping me proof read the final thesis, my roommate of 5 years, Johanna, for our numerous cups of tea and discussions throughout the night, and to my girlfriend putting up with me while providing enough during the final stages of writing.

Table of contents

1. Abstract

1. 1. English abstract

The Norwegian dairy goat is the only known mammal that naturally lacks cellular prion protein (PrP^C); this due to a nonsense mutation in the prion protein gene (*PRNP*). This makes the goats interesting model animals in studies of the physiological functions of the cellular prion protein. Because the *PRNP* gene is both a highly expressed and highly conserved gene, and therefore believed to be an important gene, it is hypothesized that the mutation in the Norwegian dairy goats is a fairly new mutation, and thus have a relatively high amount of homozygosity in the chromosomal region around the mutation.

In this master thesis, I have particularly been investigating the extent of linkage disequilibrium in the region surrounding the prion protein gene (*PRNP*) in the Norwegian dairy goat by. To estimate the length of the haplotype surrounding $PRNP^{ter}$ variant, the SNP-allele frequencies were tested in animals homozygous for this variant, and compared to the corresponding SNP-allele frequencies in heterozygous goats and a control group of unrelated goats without the mutation. Runs of homozygosity (ROH) were also calculated for the complete chromosome 13 to compare the *PRNP*-region to the rest of the chromosome.

The results showed a distinct decrease in the wild type alleles in the area surrounding the *PRNP* gene in the goats with the $PRNP^{ter/ter}$ genotype, which none of the other groups of goats showed. The goats homozygous for the mutation also showed some runs of homozygosity (ROH) surrounding the gene that the goats with *PRNP* does not have; although the length of the runs were short compared to other parts of chromosome 13. Overall, the results showed that the goats might be suitable as model animals, and that the mutated *PRNP* variant reside in a haplotype with a minimum length of around 2.6 MB. However, this is a gene dense region, which is important to consider if the goats with the $PRNP^{ter}$ variant are to be used as model animals on possible phenotypic effects of the *PRNP* gene.

1.2. Norwegian abstract

Norsk melkegeit er det eneste kjente pattedyret som naturlig mangler cellulært prion protein (PrP^C); dette på grunn av en nonsense mutasjon i prionproteingenet (*PRNP*). Dette kan gjøre geitene til interessante modelldyr i studier av de fysiologiske egenskapene til det cellulære prionproteinet; men fordi *PRNP*-genet både er sterkt uttrykt og er et sterkt konservert gen, er en hypotese at mutasjonen i den norske melkegeita er en ganske ny mutasjon, og at det derfor er relativt mye homozygoti i kromosomområdet rundt mutasjonen.

I denne masteroppgaven har jeg spesielt undersøkt utbredelsen av koblingsulikevekt i området rundt *PRNP* på kromosom 13 i norsk melkegeit. For å beregne lengden på haplotypen som $PRNP^{ter}$ varianten er en del av, ble SNP-allelfrekvenser i dette området testet hos geiter som er homozygote for varianten, og deretter sammenlignet med tilsvarende frekvenser hos heterozygote geiter, og en kontrollgruppe ubeslektede geiter uten mutasjonen. «Runs of homozygosity» (ROH) ble også analysert for hele kromosom 13 for å sammenligne *PRNP*-området med resten av kromosomet.

Resultatene viste en tydelig nedgang i villtypealleler i området rundt *PRNP*-genet i geitene med $PRNP^{ter/ter}$ genotypen, noe ingen av de andre gruppene av geiter viste. Geitene som er homozygote for den muterte genvarianten viste også noe runs of homozygosity (ROH - lange rekker med homozygote områder) i området rundt genet som de resterende geitene mangler; på tross av at de homozygote områdene var korte sammenlignet med andre områder på kromosom 13.

Alt tatt i betraktning viser resultatene at geitene kan være gode modelldyr, og også at mutasjonen ligger i en haplotype med en minimumslengde på ca 2.6 MB. Dette er dog et område med stor gentetthet, noe som må tas hensyn til dersom man skal benytte geitene med $PRNP^{ter}$ varianten som modelldyr for å undersøke *PRNP*-genets mulige fenotypiske effekter.

2. Introduction

2. Introduction

The first discovery of the prion protein (PrP) was made in 1982 by Stanley B. Prusiner [1] and his team at the University of California, San Francisco. For the first time, they were able to purify the hypothetical infectious protein from brains of terminally ill hamsters that had been inoculated with sheep scrapie prions [2].

The protein was named a prion, derived from the words **protein** and **infection**, and Prusiner later received the Nobel Prize for his discovery of prions [3] The prion protein and the radical "protein only hypothesis" [4] attracted a lot of attention after an outbreak of the neurodegenerative disease Bovine Spongiform Encephalopathy (BSE, also known as mad cow disease) in the UK in 1986. It was later found that the BSE could cross species barrier and also affect humans, through a new variant of Creutzfeldt-Jakob disease [5]. This led to considerable discussion amongst scientists, not least because of the uncertainty related to how many humans who had potentially been exposed to the BSE prions, and whether or not humans are receptive to them.

BSE is one of the serious, neurodegenerative diseases that can affect both humans and animals, and it was for a long time an enigma how these diseases developed, as no one could identify any disease causing agent. According to the "protein only hypothesis", prions are sub-viral protein particles devoid of nucleic acids. The prion itself consists of aggregates of misfolded "scrapie conformers" (PrP^{Sc}), of the normal cellular prion protein (PrP^{C}). The misfolded prion protein was identified as the pathogen that was causing the diseases, and the quest to figure out how the pathogenic protein (mis)folds itself, and contaminates other cellular prion proteins began, as well as an investigation of the physiological functions of the cellular prion protein.

2.1. The cellular prion protein - PrP^{C}

PrP^{C} is a glycoprotein attached to the outer leaflet of the plasma membrane with a glycosyl-phosphatidyl-inositol (GPI) anchor [6]. The protein is encoded by single-copy gene called *PRNP* and transcription of this gene starts at early embryonal stages [7] and is expressed in most adult tissues of mammals; particularly in neuronal cells of the central and peripheral nervous system [8]. Nuclear magnetic resonance (NMR) studies [9, 10] has revealed that the protein consists of two major domains; one N-terminal unstructured flexible tail with

copper binding octapeptide repeats [11], and a well-structured globular domain that contains two asparagine linked glycans and is stabilized by a disulfide bridge.

The C-terminus of PrP also contains a GPI attachment signal sequence, anchoring PrP to cellular membranes [6]. The structural features of PrP^C are well-conserved among vertebrate lineages indicating that the protein serves important cellular functions [12].

The cellular biology of PrP^C is complex and a detailed description is outside the topic of this thesis. One feature that contributes to the complexity is that PrP^C can undergo several proteolytic processing events [13, 14]. The physiological significance of the various PrP^C derived protein fragments thus generated is unclear and an area of intense studies. For a recent review of this see [15].

Despite that PrPC is a conserved protein expressed at high levels in brain and many other tissues, the first lines of mice with ablation of *Prnp* (PrP-knockout, PrP-KO), showed no major phenotypic aberrations [16, 17]. Quite the contrary, the mice developed normally and had normal life-expectancies. They even seemed to have an advantage to the normal mice; the PrP-knockout mice were completely resistant towards prion infection, as predicted from the “prion only hypothesis”.

The normal, physiological functions of PrP^C were first studied in transgenic mice created from embryonic stem cells from 129 strain mice, which has then been backcrossed to non-129 mice [18], as model animals.

However, backcrossing the 129 strain mice with non-129 strain has led to polymorphic *Prnp*-linked loci. In turn, this may cause false phenotypes in the offsprings because of *genetic hitchhiking*; a change in allelic frequency that happens because the gene is in close proximity to a favorable phenotype that is becoming fixated [19]. This as opposed to random crossing over and random selection. Because of this, finding different model animals has been of interest; however, on a general basis, mice are often used as model organisms, as they have a relatively short re-generation time, and they get many offspring at once. This makes mice very suitable for use as model animals.

The Norwegian dairy goat is a goat breed that during a routine *PRNP* genetic analysis was discovered to be naturally deprived of the cellular prion protein. This is due to what we refer to as a *nonsense mutation* in codon 32 in the *PRNP*-gene, blocking completely the synthesis of PrP^C. A nonsense mutation is a point mutation resulting in the prematurely introduction of a stop codon, leading to an incomplete protein, which usually, as in this case, gives a

nonfunctional protein product. The goats with the nonsense mutation in the *PRNP*-gene (the goats deprived of PrP^C) does not appear to deviate significantly from those without the mutation, which might indicate that there are proteins that compensate for the missing PrP^C. That begs the question:

Is the natural deprivation of the cellular prion protein because PrP^C is so important that the flanking genes can "take over" if something happens to the PRNP-gene, or is the prion protein not as important as its high conservation status proposes?

2.2. PrP^c and its physiological functions

The physiological functions of PrP^c are yet to be completely understood, but regardless of the problems with hitchhiking- and flanking genes, the studies of transgenic mice has led to more knowledge in the area.

Although some lines of transgenic mice with ablation of *Prnp* showed little phenotypic effects, others did. One dramatic phenotype was that of the *Prnd*-gene, a paralog gene in close proximity to the *Prnp*-gene, but which is mostly expressed in testicles. When *Prnp* was removed, the mice showed signs of ataxia, because of loss of cerebral Purkinje cells, some of the largest neurons in the (human) brain, and eventually death.

Upon re-introducing the *Prnp*-gene by crossing the transgenic mice with no-transgenic mice, the mice were back to being healthy, and it was believed that the neurodegeneration in the brain was because *Prnp* has a role in the long term survival of Purkinje cells, and that lack of expression of *Prnp* would result in degeneration of these neurons.

However, later studies [20] researched whether this was actually the case, or if the neurodegeneration was actually happening because of the overexpression of the neighboring gene *Prnd*, which codes for the protein Doppel (Dpl). The result of this study showed that by removing *Prnp*, *Prnd* would be overexpressed, and that not only was it expressed in the testicles; it was also expressed in the brain. When expressed in the brain, Dpl appears to have neurotoxic potential/capacity.

Upon re-introducing the *Prnp*-gene - but still with the over-expression of Dpl - the neurotoxicity "disappeared", arguing that PrP^C can be coexpressed with Dpl, and they are able to interact under certain conditions [20].

By creating a transgenic mouse with a knock-out mutation, the whole chromosome suddenly becomes **one haplotype**, which illustrates the problem with hitchhiking genes. This makes the mice generation 0, and the chromosome will have long areas with flanking genes even after as much as five to six generations. Long haplotypes and areas with flanking genes might fixate some genes close to the targeted gene, leading to hitchhiking genes.

Nuvolone [18] used transcription activator-like effector-nuclease (TALEN)-based genome editing in already fertilized mouse oocytes from C57BL/6J strain to create a new strain of mice with a *Prnp*-ablated allele, called Zurich-3 (ZH3, $Prnp^{ZH3/ZH3}$). The C57BL/6J strain is the most common and highest inbred laboratory strain [21]. These mice failed to show any of the previously observed functions, suggesting that these functions might have been phenotypes inherited by genetic hitchhiking [18].

The study done on $Prnp^{ZH3/ZH3-}$mice [18] is one of the examples of functions found that is believed to be because of PrP^{C}; aged $Prnp^{ZH3/ZH3-}$mice developed a chronic demyelinating neuropathy, suggesting that cellular prion protein is important in myelin maintenance. According to the study by Nuvolone [18], flanking genes are not a problem in the ZH3-mice. Although the mice in this study got chronic demyelinating neuropathy as they aged, they did not show any signs of neurodegeneration, as shown in previous mouse-studies, giving yet another proof that the neurodegeneration did, in fact, show up because of hitchhiking genes.

Because PrP^{C}-cells are highly expressed in immuncells, it has been hypothesized that PrP^{C} plays an important role when it comes to protection against diseases [22].

When treating 8 goats with the $PRNP^{+/+}$ genotype, and 8 goats with the $PRNP^{ter/ter}$ *genotype* with lipopolysaccharide (LPS), all the goats showed signs of sickness, but the $PRNP^{ter/ter}$-goats showed a stronger reaction than the $PRNP^{+/+}$-goats; they were sick longer and showed deeper signs of sickness. The $PRNP^{ter/ter}$-goats were also more sensitive to the LPS than the $PRNP^{+/+}$-goats [23]. The same study conducted a gene ontology enrichment analysis, which showed activation of genes involved in type I interferon signaling; the type I interferon response led to an activation of cytokine-responsive genes.

This might suggest that PrP^{C} plays a role in certain immunity signaling pathways, and contributes to the protection of tissues against inflammatory damage [23, 24]. To learn more about PrP^{C}'s role on the immunsystem, hematological analyses have been done on healthy $PRNP^{ter/ter}$-goat kids [25]. The mean volume of the red blood cells were well within the

normal range; yet overall smaller than the *PRNP*$^{+/+}$-goat kids. This seemed to be compensated for by an increased number of red blood cells in the *PRNP*$^{ter/ter}$-goats, as the mean cell volume corresponded to the number of red blood cells, the number of red blood cells increasing whenever the mean cell volume decreased. The same findings were also previously shown in transgenic *PRNP*$^{-/-}$ Holstein calves [22, 24]

2.3. Prion diseases in mammals

As far as science goes today, cellular prion protein is highly conserved in all mammals, with one exception; **the Norwegian dairy goat**. This even though PrP^{C} can misfold into a pathogenic, isoform protein, PrP^{Sc}, creating such as Creutzfeldt-Jakob disease (CJD) in humans, BSE in cattle, and scrapie in sheep and goats [26] - a group of diseases known as Transmissible Spongiform Encephalopathy (TSE). These are infectious neurodegenerative diseases that are almost always fatal to the host [27].

Prion diseases are known to have long incubation periods and short clinical duration, where PrP^{Sc} will build up over a long period of time (can mean years) without symptoms, but once symptoms begin, the disease becomes rapidly worse [28]. The slow buildup of PrP^{Sc} is a triggering factor of neurodegeneration (and disease), but the mechanism by which PrP^{Sc} is involved in the pathogenesis is otherwise mostly unknown [27, 28].

TSEs can be familial (occurs more frequent in family members than they should by chance; either genetic or environmental), sporadic (strikes randomly - the most common in humans), iatrogenic (caused by previous medical treatment) or transmissible (passed from one individual to another, the one that has gotten the most attention) [27, 29]

2.4. Human prion diseases

There are (at least) five human prion diseases; kuru, Creutzfeldt-Jakob disease (sporadic, iatrogenic and familial versions), variant Creutzfeldt-Jakob disease, fatal familial insomnia and Gerstmann-Sträussler-Scheinker disease. Some of the symptoms vary, but the ones they have in common are neurological progression with dementia (loss of intellect, more forgetfulness, disorientation etc) and ataxia (lack of coordination of muscle movements). Other animals with TSEs also have these symptoms in common [29].

The *PRNP*-gene being highly conserved despite the fatal risks, suggests that the benefits of the protein outweigh the risks of it turning into a pathogen, or else it is likely that the gene would have disappeared during evolution.

Exactly how the PrP^{C} is folded into PrP^{Sc} is not yet known, but the most widely accepted hypothesis is the "**protein only hypothesis**". This hypothesis states that prion diseases are the results of conformational change of the normal cellular prion protein to the isoform, pathogenic form – only needing preexisting PrP^{Sc} to change, no other disease-causing agent [30].

Conformational changes happens when the preexisting PrP^{Sc} comes into contact with PrP^{C}, creating an irreversible change from PrP^{C} to PrP^{Sc}. The PrP^{Sc} aggregate, and keep affecting the PrP^{C}. Normally, the cells have mechanisms making sure damaged proteins, or proteins making plaques, are destroyed. However, this is not the case with PrP^{Sc}, as the proteins instead are allowed their neurodegenerative travel. As of today, a method to cure or break down the pathogenic form is not known, as the pathogen is resistant to heat, radiation, different chemical agents, protease etc.

2.5. Haplotyping and genetic hitchhiking in the Norwegian dairy goat

Typically, the *PRNP*$^{-/-}$ animals does not differ noticeably from other animals even when the gene is altered postnatally [31]. This indicates that either the flanking genes already have a somewhat similar function as the *PRNP*-gene, or the PrP^{C} is not as important as its high conservation status proposes. Neither of these hypotheses make a whole lot of sense though, as either case is suggesting that the gene is not really that important.

This triggers a very interesting discussion; ***Why* is the gene so highly conserved?** Is it not likely that it would be noticed if such an important protein went missing? What other reasons can explain that the gene is still so very well conserved?

In order to investigate this further, the research in this paper will discuss genetic hitchhiking by means of haplotyping. A haplotype is a group of alleles in closely related loci that are inherited together, as they are so close to each other physically that recombination is less likely to happen. The less inbreeding or older population, the shorter the haplotypes are, as

the recombination that happens will more likely have a different set of alleles when they are not closely related, and there has been more time for recombination to happen.

There are positive indications that the Norwegian dairy goat, with its nonsense mutation in codon 32 in the *PRNP*-gene, can be very suitable as basis for research on genetic hitchhiking, and also the physiological functions of PrP^C.

By figuring out the haplotype that the *PRNP*-mutation in the Norwegian dairy goats is part of, we can more easily see what genetic hitchhiking we can expect - although the haplotype still won't necessarily tell us *why* the gene is highly conserved with seemingly not that important status. The haplotype can also tell us whether or not the Norwegian dairy goat is a good model animal for further research on PrP^C and the *PRNP*-gene.

3. Materials and methods

3. Materials and methods

3.1. Animals

In total, SNPs from 96 animals were used in this study. An already existing Illumina 50K SNP data set of 48 AI bucks (born 2009 - 2012) was used as a reference for the Norwegian goat population. 41 of these AI bucks did not have the knockout mutation (hereafter referred to as **the AI bucks**) and 7 were heterozygous (*PRNP*$^{Ter/+}$). The DNA used to make the 50K SNP data set was isolated by Biobank AS, Hamar, Norway [32]. Among the *PRNP*-animals, 21 had the *PRNP*$^{Ter/Ter}$ genotype (**the homozygous *PRNP*-animals**), 27 had the *PRNP*$^{Ter/+}$ genotype (**the heterozygous *PRNP*-animals**), and none were of the wild type.

The breed referred to in the study is the Norwegian dairy goat, and comes from a Norwegian goat breeding system. The goats from the experimental flock (hereafter referred to as the ***PRNP*-animals**) are also Norwegian dairy goats, but collected from the research herd belonging to the Center for livestock production (SHF) – Ås gård (NMBU)

3.2. Genotyping and data filtering

To genotype for the *PRNP*-mutation, the open reading frame (ORF) of the *PRNP*-gene was first amplified by PCR, before analyzing by sequencing the PCR product with Big Dye Primer chemistry (Applied Biosystems, Foster City, USA) [33].

Then a high-resolution melting curve-analysis (HRM analysis) was done. This is a method based on the melting behavior of double stranded DNA (dsDNA). The melting point (Tm) of the DNA is the temperature where 50% of the DNA is single stranded (ssDNA), and the guanine-cytosine base pair (GC) has a higher Tm than the adenine-thymine base pair (AT). This is because GC are bound by three hydrogen bonds, whereas AT are bound only by two, and thus it takes more energy to break all the bonds between GC than AT.

Analyzing the melting behavior while having used a highly concentrated saturation dye, makes it possible to label the PCR product, and thus make a detailed profile of the product [34]. As little as one base pair difference can be detected if the amplicon is very short. The difference can be spotted from the normalized melting curves, but when a difference plot is calculated the differences are even more visualized.

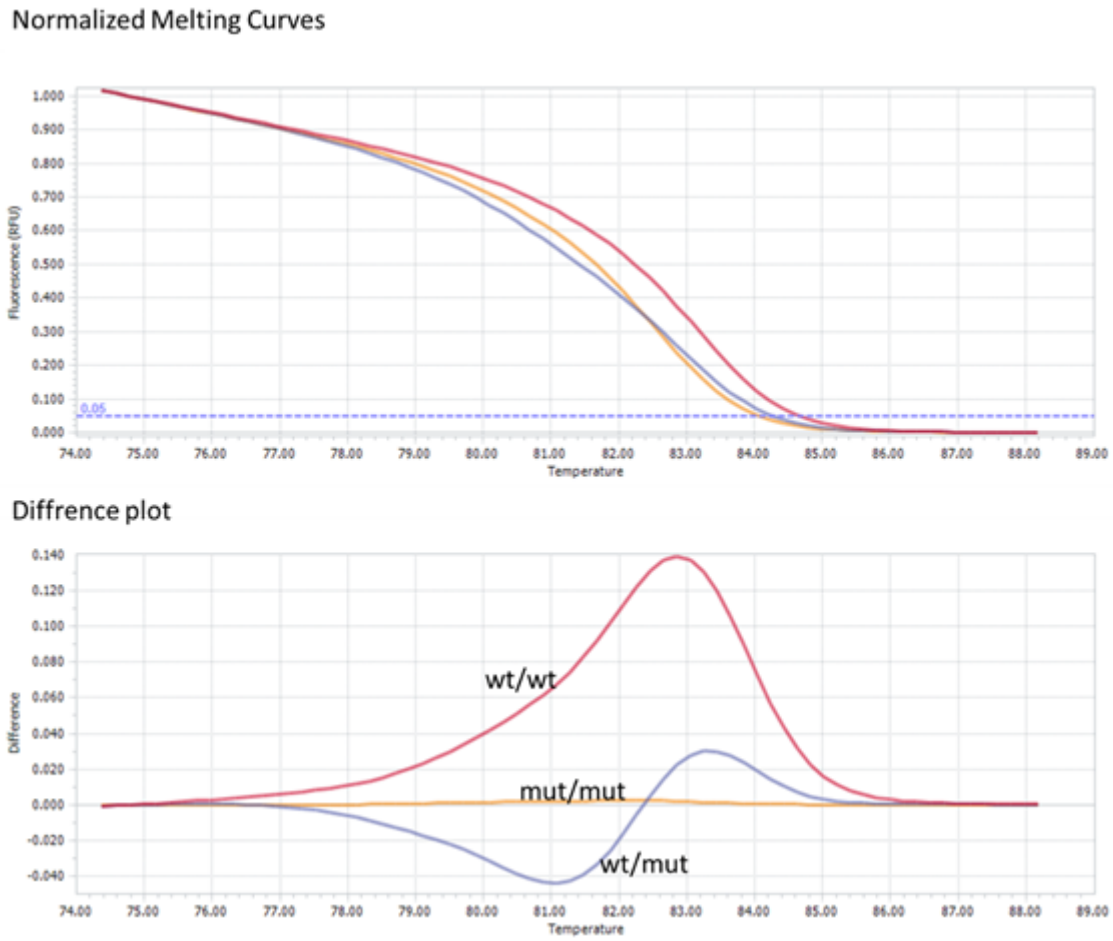


Figure by Susan Skogtvedt Røed

To ensure the quality of the Illumina genotyping data, the following filtering parameters used were: call rate > 90% both for individual markers and per animal, maf > 1% and hwe p-value > 1e-5, using the PLINK 2.0 software (**--geno**, **--mind**, **--maf** and **–hwe**) [35].

The final number of markers in the *PRNP*-animals after the initial filtering was 49 277. One heterozygous and one homozygous animal were removed due to low call rate. The AI bucks had identical quality filtering, and were left with 49 124 (of a total of 5202) markers. No animals were removed in the reference population (AI bucks). After filtering, the datasets were compared, and showed 47 847 identical markers. The datasets were then merged into one dataset containing the 47 847 markers and 94 animals.

3.3. Analysis

The first thing that was done was to update the SNP positions to the ARS1 assembly version. The files were converted into binary files with the PLINK 2.0 ***--make-bed*** function. The binary files were then used to make a kinship matrix in PLINK 2.0, and the resulting data were plotted using R. This was done to show the relationship (kinship) within the homozygous *PRNP*-animals, within the heterozygous *PRNP*-animals, within the AI bucks, and between the three groups.

PLINK 2.0 was also used to calculate minor allele frequencies (MAFs) and runs of homozygosity (ROH). MAFs show the frequency of the least frequent allele at a given SNP, and was calculated to find, and help illustrate, the position and length of the *PRNP*-haplotype. The MAFs were calculated by using PLINK 2.0 ***--freq***, and the function ***--chr 13*** was added to only calculate chromosome 13, instead of the whole genome; this because the *PRNP*-gene and -mutation is located at chromosome 13.

After calculating the MAFs, the ROH was calculated by using PLINK 2.0 ***--homozyg***. Runs of homozygosity are contiguous lengths of homozygosity over the chromosome that are inherited from the parents. As with the MAFs, ROH can be used to find the position and length of the *PRNP*-haplotype in the homozygous *PRNP*-animals. It can also be used to compare the homozygous *PRNP*-animals to the same position at the chromosome in the heterozygous *PRNP*-animals and AI-bucks.

Even though a number of different parameters were tested, only SNP threshold 50 (the minimum amount of SNPs a run of homozygosity can contain) was used in the final analysis. This was due to the fact that any threshold above threshold 50 returned too low number of runs, and any threshold below did not add more ROH.

The rest of the filtering parameters were found to work best at their default thresholds: the length was set to 1000 kb, at least 1 SNP per 50 kb, a scanning window of 50 kb, a scanning window can *at most* contain 1 heterozygous call and 5 homozygous missing calls, and the hit rate for all scanning windows containing the SNP must be at least 0,05 [36]. All the calculations were executed to the datasets containing the homozygous *PRNP*-animals, the heterozygous animals, and the AI bucks.

As PLINK is usually used for analysis of the human genome, it gets a bit confused by the chromosome number in goats. This was fixed by adding the function ***--chr-set 30*** to any command.

4. Results

4. Results

4.1. Allele frequencies on chromosome 13 in homozygous and heterozygous individuals

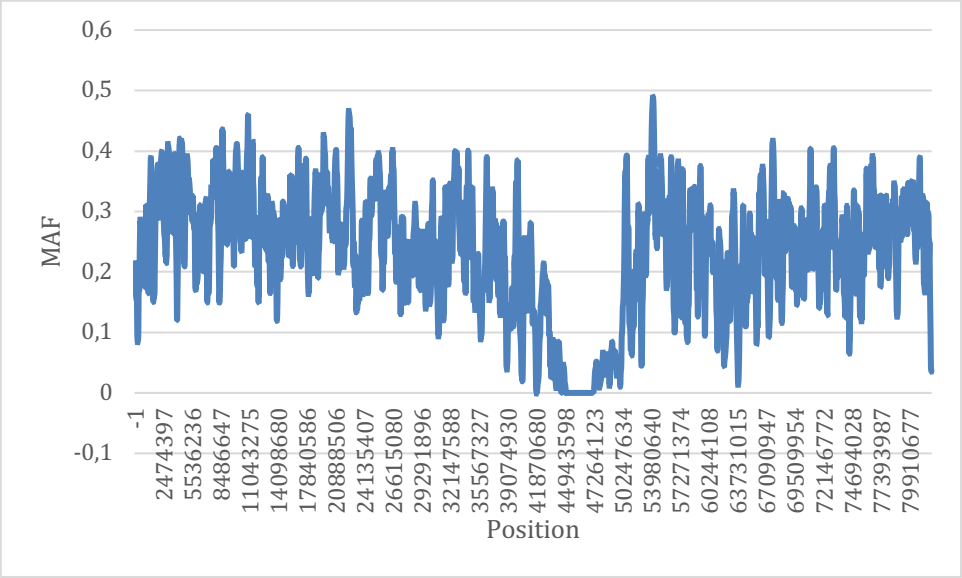


Figure 1a: Minor allele frequency (MAF) at chromosome 13 in the homozygous PRNP-animals.

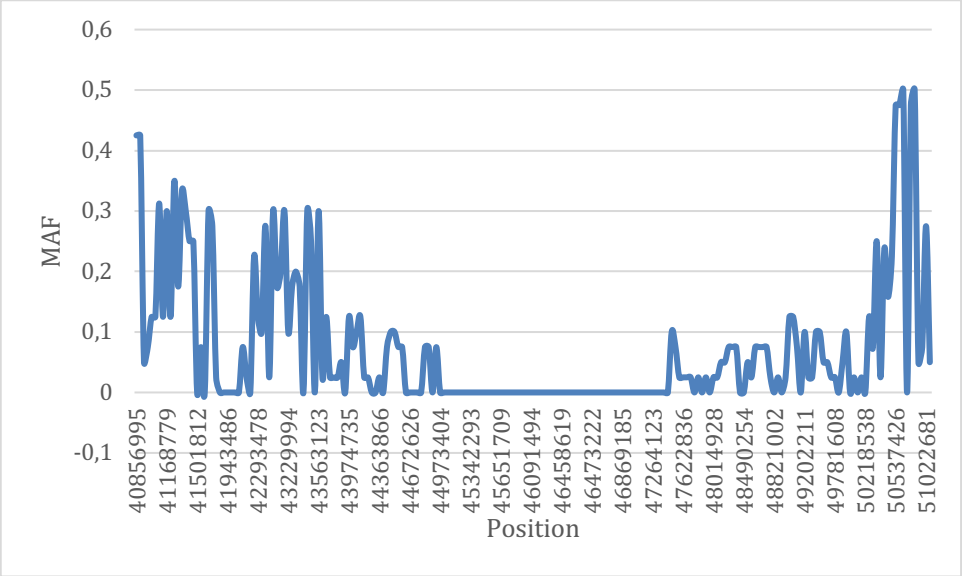


Figure 1b: MAF at the area surrounding the PRNP-gene in the homozygous PRNP-animals.

To illustrate the amount of homozygosity surrounding the *PRNP*-locus in the homozygous *PRNP*-animals, the minor allele frequency (MAF) was estimated across chromosome 13 for all three groups of animals. The minor allele frequency shows the frequency of the wild type allele (wt), and as figure 1a illustrates, the frequency of the wt-alleles is lowered around the

PRNP-locus in the homozygous *PRNP*$^{Ter/Ter}$ animals. The MAF-value is especially low between the positions 44.9 MB to 47.5 MB; but are descending already at around position 40.6, until position 49.3, giving an estimated haplotype of approximately 8.7 MB. Figure 1b shows a zoomed in picture of the figure 1a, from position 40.9 MB to position 51.0 MB.

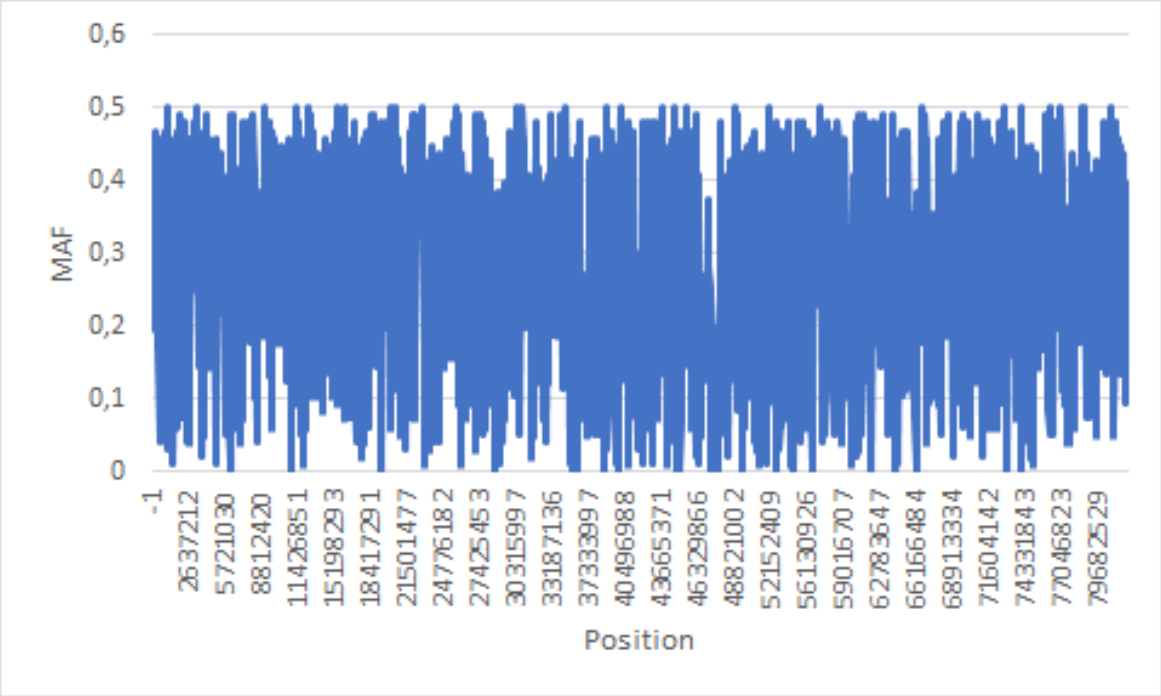


Figure 2: MAF at chromosome 13 in the heterozygous PRNP-animals

To test if the homozygosity shown in figure 1a and 1b is connected to the *PRNP*-allele we also plotted the MAF in the heterozygous individuals and AI-animals at chromosome 13. Figure 2 shows that these animals did not show the distinctive lowered MAF that figure 1a and 1b does at these positions. The AI animals show similar results as the heterozygous *PRNP*-animals; no distinctive lowered MAF-value around the *PRNP* locus (figure 1, appendix).

4.2. Runs of homozygosity

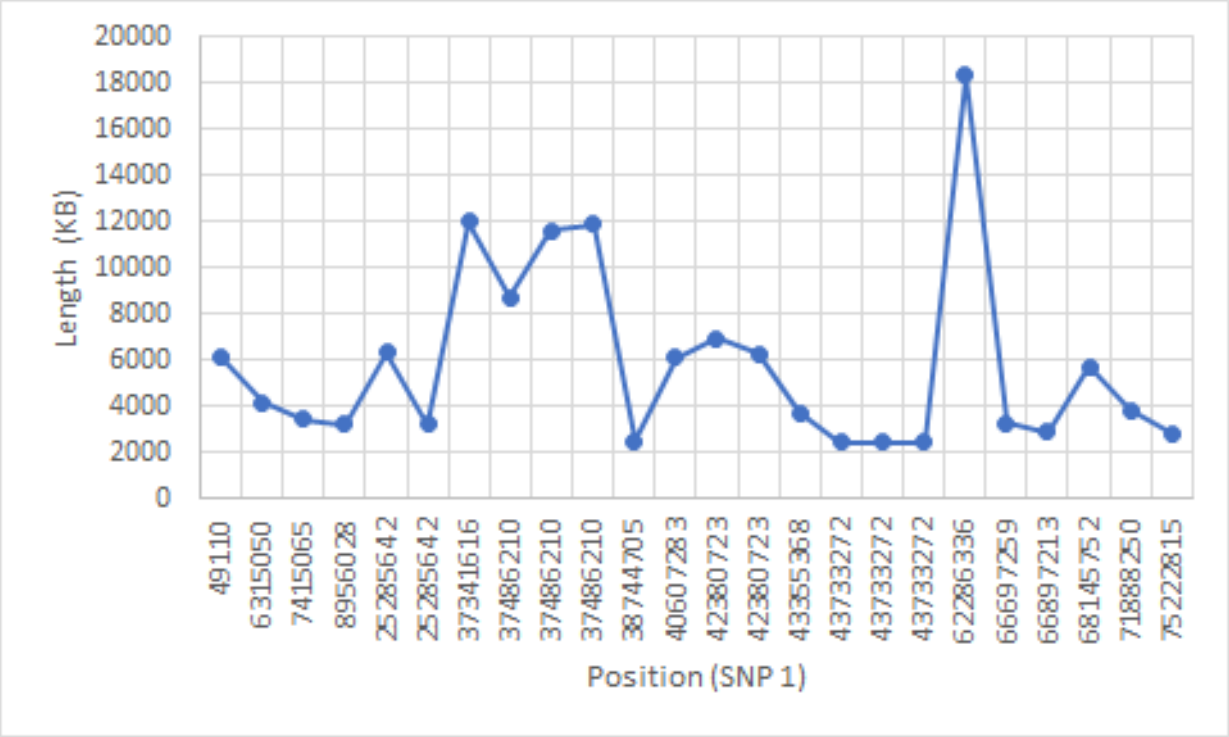


Figure 3: Runs of homozygosity (ROH) at chromosome 13 in the homozygous PRNP-animals. The x-axis shows the position of the first SNP of the run, the y-axis shows the length of the runs in kilo bases (KB).

Figure 3 shows runs of homozygosity in the homozygous *PRNP*-animals. It shows the location of the *first* SNP in a run on the x-axis, and the length of the run on the y-axis. Table 1 (appendix) show the position of both the first and last SNP of the runs. Figure 3 and table 1 (appendix) show a 6.10 MB long ROH from SNPs beginning at position 40.60 to position 46.70, which runs over the positions of the *PRNP*-gene, 46.45 to 46.47. Although figure 3 clearly shows a run over the *PRNP*-gene, this run is by no means the longest at chromosome 13; both the 11.99 MB run from position 37.34 to 49.33 and the 18.31 MB run from position 66.69 to 80.61 are longer.

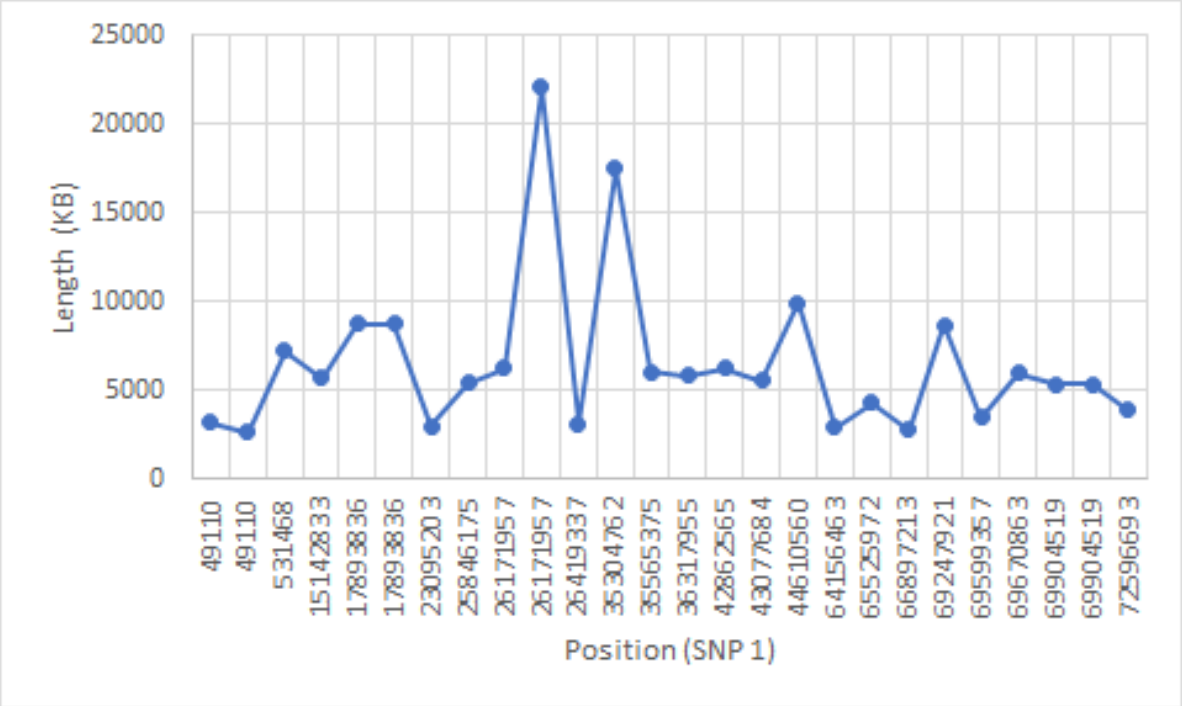


Figure 4: Runs of homozygosity (ROH) at chromosome 13 in the heterozygous PRNP-animals.
The x-axis shows the position of the first SNP of the run, the y-axis shows the length of the runs in kilo bases (KB).

Figure 4 and table 2 (appendix) show the same as figure 3 and table 1 (appendix) - the positions of the SNPs of the ROH; only in the heterozygous *PRNP*-animals. Figure 4 lacks the ROH from SNPs beginning at position 40.60 to position 46.70 that figure 3 shows. Figure 3 (appendix) shows that the heterozygous *PRNP*-animals (figure 4) and the AI-animals (figure 3, appendix) have runs in the same positions.

4.3. Kinship matrix

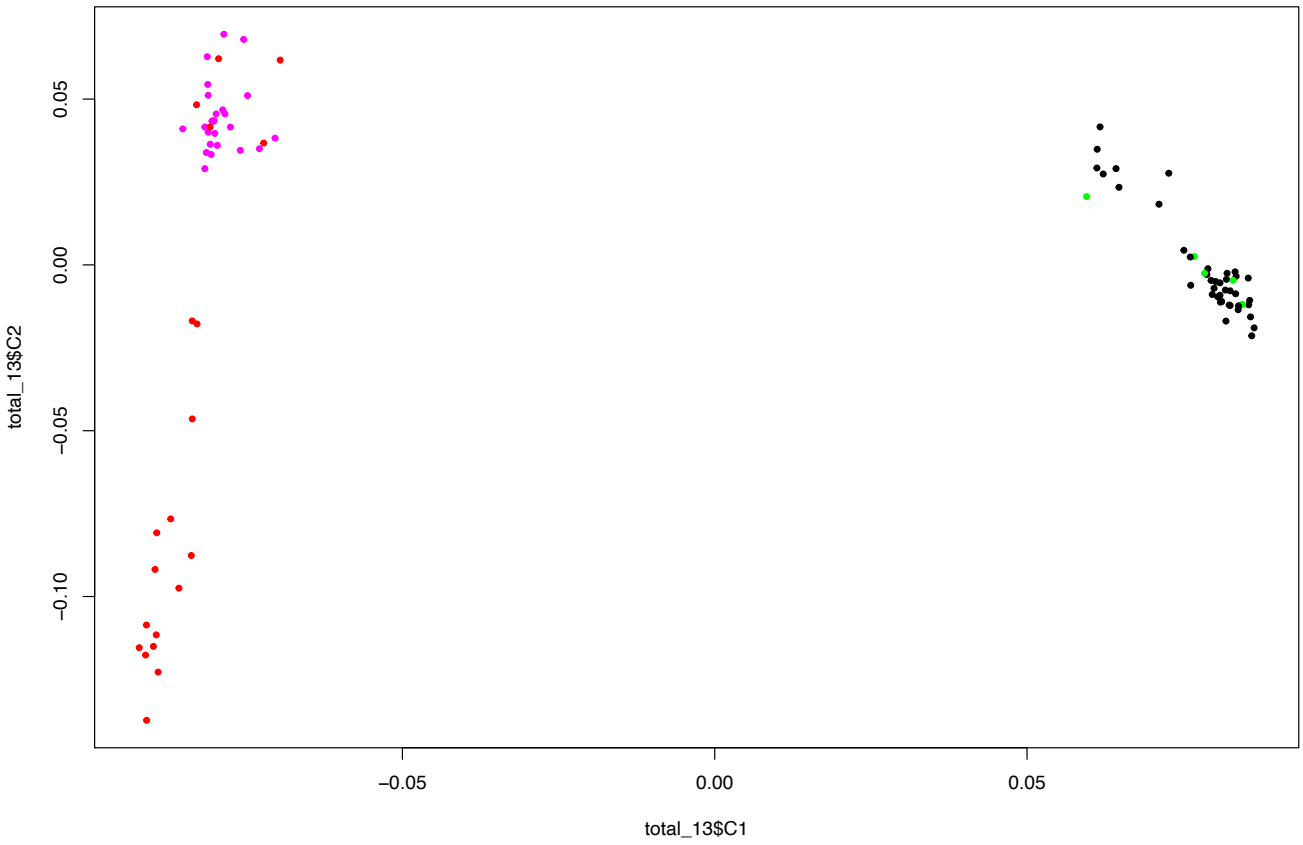


Figure 5: Kinship between the 21 homozygous PRNP-animals (pink), the 27 heterozygous PRNP-animals (red), the 41 homozygous AI bucks (black) and the 7 heterozygous AI bucks (green).

Figure 5 shows a kinship matrix - a plot showing how closely related the homozygous and heterozygous *PRNP*-animals are. The heterozygous animals are more closely related to each other than to the homozygous animals; however, the homozygous animals are more closely related to each other than the heterozygous animals are.

5. Discussion

5. Discussion

In this study, the main aim has been to investigate the amount of linkage disequilibrium surrounding the *PRNP*-gene in chromosome 13 in the Norwegian dairy goat. These goats are naturally deprived of the prion protein because of a nonsense mutation in codon 32 in the *PRNP*-gene.

A major issue when studying the physiological functions of PrP^C in knock-out mice has been the effect of hitchhiking genes. Therefore, investigating the linkage disequilibrium in order to estimate the length of the haplotype surrounding the *PRNP*-gene has been of interest. The length of the haplotype has also been essential when verifying the suitability of the Norwegian dairy goat as a suitable model animal; the same problem with long haplotypes and hitchhiking genes as in the transgenic mice would be a problem in the goats if the animals are too closely related, and the haplotype is too long.

5.1. Haplotypes and runs of homozygosity in inbred animals

When two parents have a common ancestor just a few generations back in time, the offspring is expected to have long stretches of homozygosity due to the inheritance of these common haplotypes through both parents. Since crossing over is limited from one generation to the next, haplotypes will remain fairly long over several generations. However, the haplotypes will be shorter for every generation that passes on from the common ancestor (due to the cross over in every generation) [37].

It is likely to expect the homozygous *PRNP*-animals to be more related to each other than in the rest of the population, and that the chromosomal region surrounding the *PRNP*-gene is a common haplotype. Depending on how far back in time this mutation occurred, the length of the haplotype carrying this variant can vary. If this mutation occurred for the first time just a few generations back, the haplotype should be long, if the mutation is very old the haplotypes might be shorter.

It is not known how old the *PRNP*-mutation is; however, because *PRNP* is such a highly conserved gene, it points toward an important biological function of the gene product. It seems reasonable to believe that the goats with mutated variant will have a somewhat reduced health, and thus have been selected against under selective breeding, both in the

past, and if the mutation was to happen spontaneously at this point. Our hypothesis is therefore that this is a fairly new mutation, occurring not too many generations ago, and thus being situated in a longer haplotype. This study does not particularly investigate how old the mutation is or how far back one can trace it, although this might be interesting for a future study.

It is important that the haplotype carrying the *PRNP* knock-out variant is not too long in order for these goats to be suitable model animals. This because of the problems with hitchhiking of flanking genes explained in the introduction. In order to investigate the lengths of the haplotypes, this study analyzed the data by looking at minor SNP-allele frequencies in the *PRNP*-region, and runs of homozygosity.

5.2. Genetic comparison of AI-bucks to *PRNP*-animals

One could argue that the investigated animals probably are too related to each other to give a reliable picture of the situation in this chromosomal region. Therefore, we also compared the genetics of the *PRNP*-animals to the panel of AI-bucks, to see how much they differ from the ordinary goat population.

The AI bucks are of course also a selected group that is related to each other, but the alleles found among the AI bucks will also be frequently found in the Norwegian goat population. In that sense, they can be considered representative of the Norwegian goat population.

The kinship matrix (figure 5) is a plot showing how closely related the homozygous and heterozygous *PRNP*-animals are, based on all SNPs except those located on chromosome 13. The heterozygous animals are less related to each other than the homozygous animals are. A closer relationship between the homozygous animals is to be expected since they have the *PRNP*-variant in common.

However, the plot illustrates that the homozygous *PRNP*- animals are about as closely related as the AI bucks are to each other. They are distinctly different from the AI bucks, and also different compared to the heterozygous individuals.; It is interesting to note that the 7 heterozygous AI-bucks are much closer to the other AI bucks than to the other individuals being hetero- or homozygous for this allele.. This indicates that the mutated allele has been around in the population for a while, and might be the first indication that the haplotypes aren’t as long as first expected.

5.3. The behavior of haplotypes in specific chromosomal regions

The haplotype length can be determined using different approaches. Estimations of the runs of homozygosity over chromosome 13 shows examples of ROH that were found around the *PRNP*-gene in the homozygous *PRNP*-animals (figure 3) that is not present in the heterozygous animals (figure 4), nor in the AI bucks (figure 3, appendix). As previously explained, a common reason for long runs of homozygosity is inbreeding. The more closely related the parents are, the more likely they are to give their offspring similar haplotypes, and thus more homozygosity. Based on this information, it would be reasonable to draw the conclusion that the homozygous *PRNP*-animals are not as closely related as first hypothesized; however, we have already determined that the homozygous *PRNP*-animals *are* quite closely related (figure 5). An explanation of this might be that the haplotype carrying the *PRNP*-variant is older than first believed, and have accumulated variation that interrupt the runs of homozygosity.

The minor allele frequencies surrounding the *PRNP*-gene (figures 1a and 1b) clearly illustrates a lowered frequency of the wt-alleles surrounding the *PRNP*-locus; from positions 44.9 MB to 47.5 MB - although the run begins even earlier - at about position 40.7 MB and ends even later - about position 49.3 MB. This shows that at least ~2.5 MB of this region is in complete linkage disequilibrium. The drop in allele frequencies around *PRNP* illustrates the effect of the haplotype becoming homozygote in this chromosomal region. Although one in highly inbred animals, or animals with a very recent mutation, such as mice used as model animals, might expect even longer stretches of lowered allele frequencies, both the 2.6 MB long stretch of MAF=0, and the estimated 8.2 MB long stretch of lowered MAF, most certainly show a distinct haplotype which contains 627 genes. In comparison, one might have to look at stretches down to less than 0.5 MB in animals that are distantly related in order to see any linkage disequilibrium at all.

The of runs of homozygosity in the *PRNP*-area show that there are not many very long runs in the area; in fact, there are several other areas at chromosome 13 with much longer runs. However, the examples of ROH overlap the area where MAF=0, giving an overlapping haplotype of about 2.6 MB running over the *PRNP*-gene. If the goats were to be used as model animals, this is important information. A gene list of chromosome 13 in goats [38] shows 162 genes in this interval, suggesting that this is a gene dense area, which should be accounted for when studying the phenotypes of *PRNP* gene using the Norwegian dairy goat

as model animal, in order to estimate the likelihood of hitchhiking genes, and which genes might cause this problem.

This might argue that the naturally *PRNP*-deprived animals can be good model animals for studying the PrP^C phenotypes, as the shorter ROH means shorter haplotypes, and thus less problems with hitchhiking genes than one might expect in transgenic mice. However, it would be important to take into consideration that the haplotype covers a gene dense area when using the animals studying the phenotypes of the *PRNP* variant.

5.5. Conclusion

A major problem with the traditional transgenic mouse models is that the stretches of homozygosity are so prolonged that there are a lot of flanking genes. This has made it difficult to determine whether the observed phenotypes are due to the of PrP^{C}-deficiency in the mice, or because of effects from neighboring gene variants, being part of the same haplotype.

The Norwegian dairy goats might be well suited as model animals, because of a shorter region that seems to be in linkage disequilibrium with the *PRNP*^{Ter}-variant. However, because the haplotype covers a gene dense area, one might still risk hitchhiking genes. Their haplotypes are shorter compared to mice however, which indicates that the goats have a lower risk of flanking gene problems than mice, and can, if taking into consideration the gene density, be used as model animals. By knowing the haplotype length and position, it is also possible to predict which genes are likely to be inherited together with the mutation by looking into the gene map of this region.

By using the Norwegian dairy goats as model animals and studying the haplotypes surrounding the *PRNP*-gene using different methods, we were able to determine the length of said haplotype with some level of accuracy; the observed minor allele frequencies shows quite a long area with linkage disequilibrium, estimated to be approximately 8.2 MB. The area where there is examples of ROH over the *PRNP*-gene, and where MAF=0 overlap with a haplotype that is about 2.6 MB.

While still clearly a haplotype, this is somewhat shorter than expected, suggesting that **the mutation has been around for longer than previously believed**.

References

References

1. Bolton, D.C., M.P. McKinley, and S.B. Prusiner, *Identification of a protein that purifies with the scrapie prion.* Science, 1982. **218**(4579): p. 1309-11.
2. Prusiner, S.B., *Novel proteinaceous infectious particles cause scrapie.* Science, 1982. **216**(4542): p. 136-44.
3. Prusiner, S.B., *Prions.* Proc Natl Acad Sci U S A, 1998. **95**(23): p. 13363-83.
4. Prusiner, S.B., *Prions: novel infectious pathogens.* Adv Virus Res, 1984. **29**: p. 1-56.
5. Will, R.G., et al., *A new variant of Creutzfeldt-Jakob disease in the UK.* Lancet, 1996. **347**(9006): p. 921-5.
6. National Center for Biotechnology Information. *PRNP prion protein [Homo sapiens (human)]*. 2019, December 10 Available from: https://www.ncbi.nlm.nih.gov/gene/5621.
7. Makzhami, S., et al., *The prion protein family: a view from the placenta.* Front Cell Dev Biol, 2014. **2**: p. 35.
8. Wulf, M.A., A. Senatore, and A. Aguzzi, *The biological function of the cellular prion protein: an update.* BMC Biol, 2017. **15**(1): p. 34.
9. Riek, R., et al., *NMR structure of the mouse prion protein domain PrP(121-231).* Nature, 1996. **382**(6587): p. 180-2.
10. Riek, R., et al., *NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231).* FEBS Lett, 1997. **413**(2): p. 282-8.
11. Viles, J.H., et al., *Copper binding to the prion protein: structural implications of four identical cooperative binding sites.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 2042-7.
12. Bakkebo, M.K., et al., *The Cellular Prion Protein: A Player in Immunological Quiescence.* Front Immunol, 2015. **6**: p. 450.
13. Tveit, H., et al., *Proteolytic processing of the ovine prion protein in cell cultures.* Biochem Biophys Res Commun, 2005. **337**(1): p. 232-40.
14. Lund, C., et al., *Alternative translation initiation generates cytoplasmic sheep prion protein.* J Biol Chem, 2009. **284**(29): p. 19668-78.
15. Salvesen, O., J. Tatzelt, and M.A. Tranulis, *The prion protein in neuroimmune crosstalk.* Neurochem Int, 2019. **130**: p. 104335.
16. Bueler, H., et al., *Normal development and behaviour of mice lacking the neuronal cell-surface PrP protein.* Nature, 1992. **356**(6370): p. 577-82.
17. Manson, J.C., et al., *129/Ola mice carrying a null mutation in PrP that abolishes mRNA production are developmentally normal.* Mol Neurobiol, 1994. **8**(2-3): p. 121-7.
18. Nuvolone, M., et al., *Strictly co-isogenic C57BL/6J-Prnp-/- mice: A rigorous resource for prion science.* J Exp Med, 2016. **213**(3): p. 313-27.
19. Barton, N.H., *Genetic hitchhiking.* Philos Trans R Soc Lond B Biol Sci, 2000. **355**(1403): p. 1553-62.
20. Moore, R.C., et al., *Doppel-induced cerebellar degeneration in transgenic mice.* Proc Natl Acad Sci U S A, 2001. **98**(26): p. 15288-93.
21. Kang, S.K., N.A. Hawkins, and J.A. Kearney, *C57BL/6J and C57BL/6N substrains differentially influence phenotype severity in the Scn1a (+/-) mouse model of Dravet syndrome.* Epilepsia Open, 2019. **4**(1): p. 164-169.
22. Richt, J.A., et al., *Production of cattle lacking prion protein.* Nat Biotechnol, 2007. **25**(1): p. 132-8.
23. Salvesen, O., et al., *LPS-induced systemic inflammation reveals an immunomodulatory role for the prion protein at the blood-brain interface.* J Neuroinflammation, 2017. **14**(1): p. 106.

24. Salvesen, O., et al., *Goats without Prion Protein Display Enhanced Proinflammatory Pulmonary Signaling and Extracellular Matrix Remodeling upon Systemic Lipopolysaccharide Challenge.* Front Immunol, 2017. **8**: p. 1722.
25. Reiten, M.R., et al., *Hematological shift in goat kids naturally devoid of prion protein.* Front Cell Dev Biol, 2015. **3**: p. 44.
26. Wille, H. and J.R. Requena, *The Structure of PrP(Sc) Prions.* Pathogens, 2018. **7**(1).
27. Soto, C. and N. Satani, *The intricate mechanisms of neurodegeneration in prion diseases.* Trends Mol Med, 2011. **17**(1): p. 14-24.
28. Saa, P., D.A. Harris, and L. Cervenakova, *Mechanisms of prion-induced neurodegeneration.* Expert Rev Mol Med, 2016. **18**: p. e5.
29. Burchell, J.T. and P.K. Panegyres, *Prion diseases: immunotargets and therapy.* Immunotargets Ther, 2016. **5**: p. 57-68.
30. Laurent, M., *Prion diseases and the 'protein only' hypothesis: a theoretical dynamic study.* Biochem J, 1996. **318 (Pt 1)**: p. 35-9.
31. Westergard, L., H.M. Christensen, and D.A. Harris, *The cellular prion protein (PrP(C)): its physiological function and role in disease.* Biochim Biophys Acta, 2007. **1772**(6): p. 629-44.
32. Biobank, *Biobank.* 2019.
33. Benestad, S.L., et al., *Healthy goats naturally devoid of prion protein.* Vet Res, 2012. **43**: p. 87.
34. Druml, B. and M. Cichna-Markl, *High resolution melting (HRM) analysis of DNA--its role and potential in food analysis.* Food Chem, 2014. **158**: p. 245-54.
35. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
36. Chang, C., *Identity-by-decent.* 2019 November 30.
37. Ceballos, F.C., S. Hazelhurst, and M. Ramsay, *Assessing runs of Homozygosity: a comparison of SNP Array and whole genome sequence low coverage data.* BMC Genomics, 2018. **19**(1): p. 106.
38. Information, N.C.f.B., *Genome data viewer: Capra Hiercus* 2019 November 14.

Appendix

Appendix

Table 1: Runs of homozygosity (ROH) at chromosome 13 in the homozygous PRNP- animals. POS1 shows the position of the first SNP of the run, POS2 shows the position of the last SNP of the run.

SNP1	SNP2	POS1	POS2
snp7737-scaffold1278-1908535	snp58662-scaffold956-442042	66897213	69780514
snp32537-scaffold371-1929531	snp15101-scaffold1613-34525	38744705	41256063
snp32413-scaffold369-1215213	snp56128-scaffold875-3327766	8956028	12172460
snp7742-scaffold1278-2108489	snp58657-scaffold956-247613	66697259	69974943
snp32493-scaffold371-66953	snp12998-scaffold150-823296	40607283	46738092
snp25738-scaffold265-1319315	snp6114-scaffold1216-1397211	6315050	10472390
snp58772-scaffold959-1474019	snp13006-scaffold150-1182831	43355368	47097627
snp15410-scaffold1634-315669	snp13057-scaffold150-3423510	37341616	49338306
snp7706-scaffold1278-659996	snp8520-scaffold1308-1659018	68145752	73856465
snp23784-scaffold24-2195031	snp48983-scaffold7-757999	25285642	31633995
snp58750-scaffold959-499374	snp13057-scaffold150-3423510	42380723	49338306
snp31406-scaffold348-294256	snp19337-scaffold1952-142453	75222815	78065152
snp23784-scaffold24-2195031	snp49053-scaffold7-3882968	25285642	28509026
snp58750-scaffold959-499374	snp13042-scaffold150-2772062	42380723	48686858
snp12128-scaffold1449-337427	snp45533-scaffold622-19903	62286336	80605116
snp15413-scaffold1634-460263	snp12986-scaffold150-305015	37486210	46219811
snp56268-scaffold881-151577	snp12985-scaffold150-267716	43733272	46182512
snp56268-scaffold881-151577	snp12985-scaffold150-267716	43733272	46182512
snp25762-scaffold265-2419330	snp6123-scaffold1216-1769899	7415065	10845078
snp35289-scaffold423-309098	snp31417-scaffold348-772409	71888250	75700968
snp15413-scaffold1634-460263	snp13051-scaffold150-3186404	37486210	49101200
snp56268-scaffold881-151577	snp12985-scaffold150-267716	43733272	46182512
snp27399-scaffold291-1305664	snp25734-scaffold265-1144215	49110	6139950
snp15413-scaffold1634-460263	snp13058-scaffold150-3456786	37486210	49371582

Table 2: Runs of homozygosity (ROH) at chromosome 13 in the AI-bucks. POS1 shows the position of the first SNP of the run, POS2 shows the position of the last SNP of the run.

SNP1	***SNP2***	***POS1***	***POS2***
snp27399-scaffold291-1305664	*snp44879-scaffold611-1343960*	*49110*	*3222075*
snp27399-scaffold291-1305664	*snp44866-scaffold611-783573*	*49110*	*2661688*
snp27388-scaffold291-823306	*snp32386-scaffold369-29066*	*531468*	*7769881*
snp56060-scaffold875-357393	*snp58707-scaffold958-181795*	*15142833*	*20874441*
snp46281-scaffold638-1608187	*snp49094-scaffold7-5724860*	*17893836*	*26667134*
snp46281-scaffold638-1608187	*snp49094-scaffold7-5724860*	*17893836*	*26667134*
snp23732-scaffold24-4592	*snp49109-scaffold7-6339310*	*23095203*	*26052684*
snp49114-scaffold7-6545819	*snp48991-scaffold7-1097529*	*25846175*	*31294465*
snp49106-scaffold7-6220037	*snp21378-scaffold2086-13656*	*26171957*	*32405749*
snp49106-scaffold7-6220037	*snp13032-scaffold150-2345286*	*26171957*	*48260082*
snp49100-scaffold7-5972657	*snp49028-scaffold7-2826658*	*26419337*	*29565336*
snp5219-scaffold1180-142160	*snp47991-scaffold677-242551*	*35304762*	*52810257*
snp25880-scaffold267-909648	*snp15108-scaffold1613-410647*	*35565375*	*41632185*
snp25863-scaffold267-157068	*snp58745-scaffold959-281479*	*36317955*	*42162828*
snp58760-scaffold959-981216	*snp13051-scaffold150-3186404*	*42862565*	*49101200*
snp58765-scaffold959-1196335	*snp13040-scaffold150-2688016*	*43077684*	*48602812*
snp56288-scaffold881-1028865	*snp51636-scaffold758-1938854*	*44610560*	*54508110*
snp5415-scaffold1188-2221283	*snp7732-scaffold1278-1716051*	*64156463*	*67089697*
snp5383-scaffold1188-851774	*snp58661-scaffold956-400042*	*65525972*	*69822514*
snp7737-scaffold1278-1908535	*snp58665-scaffold956-551693*	*66897213*	*69670863*
snp58676-scaffold956-974635	*snp37975-scaffold471-3461*	*69247921*	*77919139*
snp58667-scaffold956-623199	*snp8502-scaffold1308-877640*	*69599357*	*73075087*
snp58665-scaffold956-551693	*snp31416-scaffold348-721818*	*69670863*	*75650377*
snp58659-scaffold956-318037	*snp31408-scaffold348-360907*	*69904519*	*75289466*
snp58659-scaffold956-318037	*snp31408-scaffold348-360907*	*69904519*	*75289466*
snp8490-scaffold1308-399246	*snp31435-scaffold348-1535832*	*72596693*	*76464391*

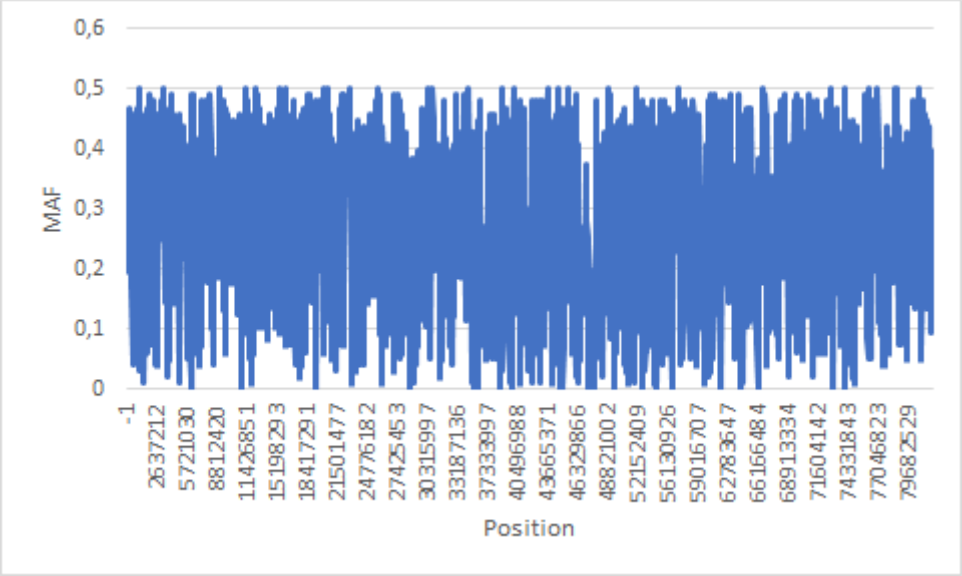


Figure 1, appendix: Minor allele frequency (MAF) at chromosome 13 in the AI-bucks.

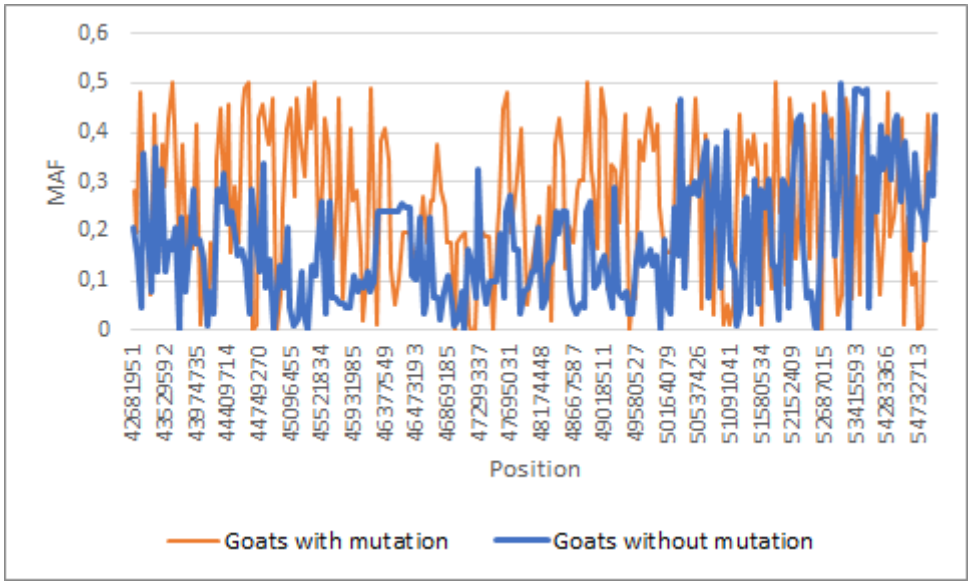


Figure 2, appendix: MAF at the area surrounding the PRNP-gene in both the homozygous PRNP-animals, and in the AI-bucks

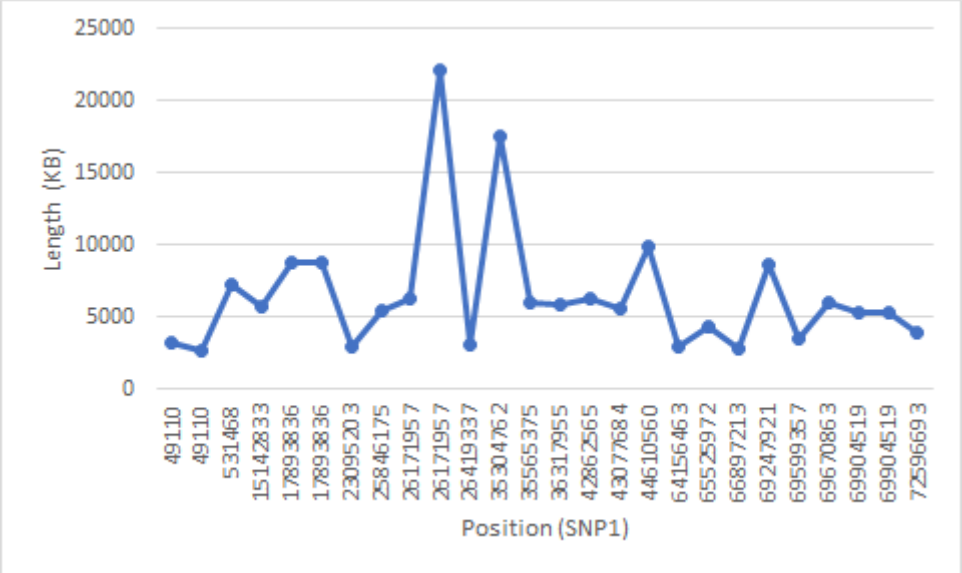


Figure 3, appendix: Runs of homozygosity (ROH) at chromosome 13 in the AI-bucks. The x-axis shows the position of the first SNP of the run, the y-axis shows the length of the runs in kilo bases (KB).

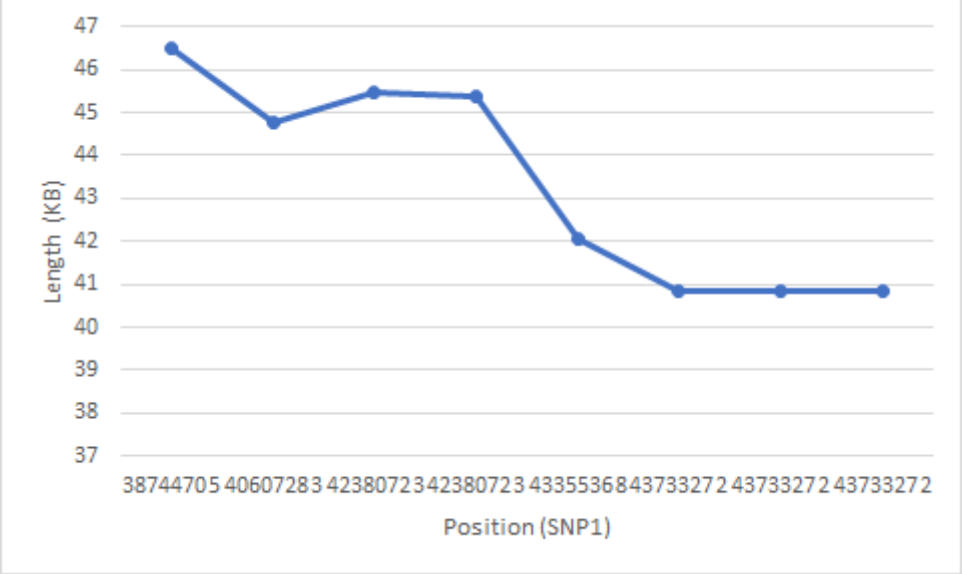


Figure 4, appendix: Runs of homozygosity (ROH) around the PRNP-gene in the homozygous PRNP-animals. The x-axis shows the position of the first SNP of the run, the y-axis shows the length of the runs in kilo bases (KB).

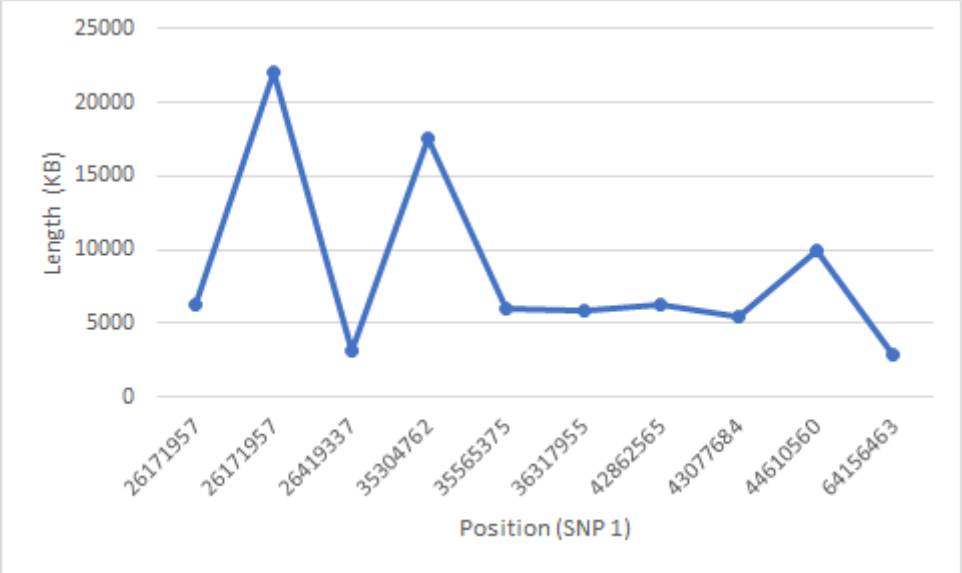
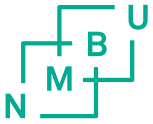


Figure 5, appendix: Runs of homozygosity (ROH) around the PRNP-gene in the AI-bucks. The x-axis shows the position of the first SNP of the run, the y-axis shows the length of the runs in kilo bases (KB).



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway