Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Animal and Aquacultural Sciences

Master Thesis 2015
60 credits

Fine mapping of two QTL regions associated with inguinal hernia in pig

Finkartlegging av to QTL-områder assosiert med pungbrokk hos svin

Inger Helene Heitmann Solheim

Acknowledgements

The practical part of the study presented in the thesis was carried out at CIGENE (Centre for Integrative Genetics) as a part of my Master of Science degree at the Department of Animal and Aquacultural Sciences, Norwegian University of Life Science, Ås.

I would like to express my gratitude to my supervisors Matthew Kent (researcher, CIGENE, NMBU) and Maren van Son (researcher, Norsvin) for expert guidance, valuable advices and constructive suggestions throughout my Master thesis. I am also grateful to Norsvin for providing samples from case and control animals and data on previously genotyped SNPs, and to CIGENE for the financial support needed to perform this study.

In addition, the work in my thesis could not be fulfilled without the help from Tim Knutsen, who patiently answered my thousands of questions regarding my scripting errors; Harald Grove and Tina Graceline Kirubakaran, who provided me with helpful scripts; Terese Andersstuen, for performing and leading me through assay designs; and Mariann Arnyasi, who performed the genotyping and presented me to the laboratory at CIGENE. I am thankful for all the help I have received during the different stages of my thesis, and for the great learning opportunity this has been.

Ås, May 13, 2015

Inger Helene Heitmann Solheim

Abstract

Pigs suffering from hernias are of concern for pig producers as they can lead to poor animal welfare and economic loss for the producer. Inguinal and scrotal hernia describes instances where abdominal contents protrude though a weakness in the abdominal wall and into the inguinal canal or scrotum. The main pig breed in Norway is Norsvin Landrace, and although the incidence of inguinal/scrotal hernia in the population is generally low, selection against the condition is important to decrease the economic loss related to the defect.

The aim of this study was to fine map two previously identified QTL regions (SSC6a and SSC6b) on pig chromosome 6 and to identify SNPs associated with inguinal hernia in the Norsvin Landrace population. Whole-genome sequences generated with Illumina sequencing technology were used to identify putative SNPs. A SNP panel for each of the two QTL regions was designed and polymorphisms were genotyped in 238 case and control animals. An association analysis was performed and the locations of SNPs significantly associated with inguinal hernia were assessed. Linkage disequilibrium between significant SNPs was also evaluated, and Tagger analysis was used to identify tagSNPs for putative use in animal breeding.

In the first QTL region, SSC6a, 13 SNPs were associated with inguinal hernia; the majority of these were clustered close to the uncharacterized gene *LOC102157459*. Moreover, three tagSNPs (ALGA0104695, 6_8476636, and ASGA0027406) in the region captured the alleles of all significant SNPs from this study and can be used in breeding programs.

In the second QTL region, SSC6b, 22 SNPs were associated with inguinal hernia, but the clustering of significant SNPs as well as the LD in the region makes this QTL more difficult to evaluate. Still, several partly characterized genes were affected by the significant SNPs, including *NFIA*, *C6H1orf87*, and two *FGGY* genes. Due to the complexity and size of this region, 12 tagSNPs are necessary to capture the alleles of all significant SNPs, and these can be used in breeding to reduce the incidence of inguinal hernia in Norsvin Landrace pigs.

In total, this study presents fine mapping of two previously identified QTLs for inguinal hernia and identification of SNPs that could be implemented in genetic selection against this defect in Norsvin Landrace. Also, the results indicated several genes that could be involved in susceptibility to inguinal hernia that have not previously had attention as functional candidate genes for hernia formation.

Sammendrag

Brokk i griseindustrien representerer et problem for griseprodusenter da det fører til helse- og velferdsproblemer blant rammede griser, samt økonomiske tap for produsenten. Lyske- og pungbrokk oppstår når tarmer eller annet innhold i buken trenger gjennom et svakt punkt i bukveggen og inn i lyskekanalen, og i noen tilfeller videre til pungen. Norsvins landsvin er den viktigste rasen innen griseproduksjon i Norge, og i den norske landsvinpopulasjonen er forekomsten av lyske- og pungbrokk generelt lav. Det er likevel viktig å selektere mot brokk for å redusere økonomiske tap og øke grisens velferd.

Målet med denne studien var å finkartlegge to områder (SSC6a og SSC6b) på kromosom 6 som i tidligere studier har vært assosiert med lyske- og pungbrokk hos landsvin. Helgenomsekvenser fra 23 dyr ble brukt til å identifisere antatte enkeltnukleotidpolymorfismer (SNPer), og et sett med SNPer ble valgt ut for de to områdene. SNPene ble genotypet i 238 syke og friske dyr, og en assosiasjonsanalyse ble utført for å identifisere SNPer assosiert med lyske- og pungbrokk i landsvinpopulasjonen. Koblingsulikevekt mellom markørene ble evaluert, og Tagger analyse ble brukt for å identifisere SNPer som egner seg til bruk i seleksjon.

Det første området (SSC6a) inneholdt 13 SNPer med signifikant assosiasjon med lyske-/pungbrokk, der størstedelen befant seg inni og i nærheten av et ikke-karakterisert gen (*LOC102157459*). Tre tagSNPer (ALGA0104695, 6_8476636, og ASGA0027406) fanget opp allelene til alle signifikante SNPer i SSC6a og kan brukes videre i seleksjon.

Det andre området (SSC6b) inneholdt 22 SNPer med signifikant assosiasjon med lyske-/pungbrokk. Fordelingen av signifikante SNPer og graden av koblingsulikevekt mellom disse ga ikke et like klart bilde som det gjorde i SSC6a. Flere delvis karakteriserte gener inneholdt signifikante SNPer; *NFIA*, *C6H1orf87* og to *FGGY* gener. Tolv tagSNPer måtte til for å fange opp allelene til alle signifikante SNPer i området, og disse kan brukes i seleksjon.

Totalt sett fremstiller denne studien finkartlegging av to QTL-områder som tidligere har blitt assosiert med lyske-/pungbrokk hos landsvin. Videre har flere markører som kan benyttes i seleksjon mot lyske-/pungbrokk blitt identifisert, samt flere gener som kan være involvert i predisponering for lyske-/pungbrokk som ikke har vært assosiert med brokk i tidligere publikasjoner.

Table of contents

1 Introduction

1.1 The Norsvin Landrace and pig breeding

The pig (*Sus scrofa*) belongs to the family Suidae and is a cloven-hoofed mammal. Pigs are used commercially for the production of meat. In 2013 the total number of pigs in the 28 EU countries was >146 million [1] and close to 22 million tons [2] of pork meat were produced. In Norway, the average number of slaughtered pigs from 2011 to 2014 is about 1.6 million per year [3].

The pig has 18 pairs of autosomal chromosomes in addition to two sex chromosomes. The International Swine Genome Sequencing Consortium (SGSC) sequenced the genome of a Duroc (*Sus scrofa domesticus*) female and published it in 2010 [4]. A high quality draft genome sequence with reviewed annotation was released in 2012, giving the basis for the newest pig reference genome build 10.2 [5].

Norsvin Landrace is the main pig breed in Norway and is known for its white color, hanging ears, small head, narrow shoulders, long and hooked back, and fleshy ham [6]. The Norsvin Landrace originated as a blend of imported Landraces from several other countries, which were subjects to special selection to ensure a unique adaptation to the Norwegian environment [7]. Organized pig breeding has been practiced in Norway since 1958, when the Norwegian pig breeding company (now: Norsvin) was founded. Before the 1990's the focus of the breeding program was mainly on feed conversion rate, meat quality and carcass quality as the breed was bred as a combination line. In the early 1990's the breeding goals changed to focus on maternal traits such as piglet mortality and number of teats, making the Norsvin Landrace a dam line. The background as a combination line has made the Norsvin Landrace a good breed for both maternal traits and production traits [8].

In Norway the Norsvin Landrace is bred in 35 nucleus herds, with a total population of 2100 sows and 50 boars. The nucleus herds produce boars for boar testing, and sows that are sold to multiplier herds [9]. In the multiplier herds, Landrace sows are inseminated with Yorkshire semen to produce a hybrid sow which then is sold to production herds [10]. In the production herds, these hybrid sows are inseminated with Duroc, Landrace, or Hampshire semen to produce slaughter pigs. The production herds are subdivided into three categories; piglet producers produce piglets to sell, finishing herds buy piglets and feed them until slaughter, and combination herds perform both tasks [11].

The pig breeding system in Norway is based on crossbreeding (Figure 1). The pure breeds are bred in small populations, often with different breeding goals for each breed. The production animals are hybrids between two or more breeds. The crossbreeding system takes advantage of heterosis, which is the phenomenon where the offspring of two genetically distant parents shows an increased function of biological traits, for instance growth rate, fertility or size. The function of the heterotic offspring is greater than the average of the parents, and thus the hybrid (LY/LZ hybrid dam) between the dam lines Landrace and Yorkshire is expected to be a better production animal than the pure breeds [12].

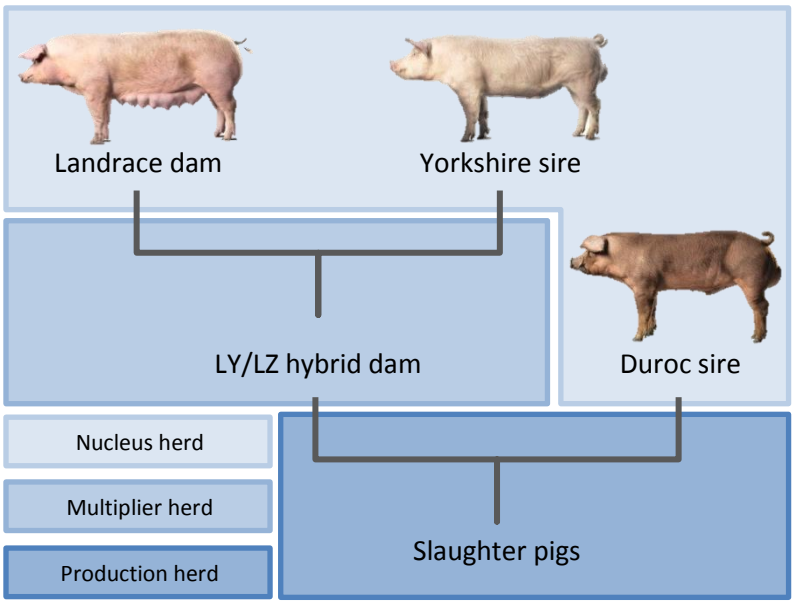


Figure 1: Representation of the breeding system in Norway. In some production herds, Landrace or Hampshire sires are used instead of the Duroc sire. The herd associated with each breed in this representation is the type of herd the pig is born in.

1.2 Anatomical structures and testicular descent

The anatomical structures involved in testicular descent are the same in large common mammals such as pigs and humans [13, 14] and in this section it is assumed that facts about the human testicular descent also holds true for the porcine testicular descent.

The inguinal canal in pigs, as well as in other species, is the potential space between the external and internal abdominal oblique muscles. Each individual has two inguinal canals, one on each side. The opening between the abdominal cavity and the inguinal canal is called the internal (or deep) inguinal ring. At the other end, the external (or superficial) inguinal ring forms the exit of the inguinal canal (Figure 2) [15].

The parietal peritoneum is a membrane that covers the abdominal cavity. The vaginal ring lies over the internal inguinal ring [15] and is the entrance to processus vaginalis; a pouch from the parietal peritoneum that goes through the inguinal canal and into the scrotum during the process of testicular descent [13].

During male development, the testes are formed within the abdomen and descend through the inguinal canals and into the scrotum during a process called testicular descent. The gubernaculum is a fibrous cord connecting the testis to the scrotum, directing the testis during testicular descent and thereafter securing the testis to the scrotum. It is during testicular descent that the processus vaginalis is formed, also mediating in the migration of testes to the scrotum [16, 17].

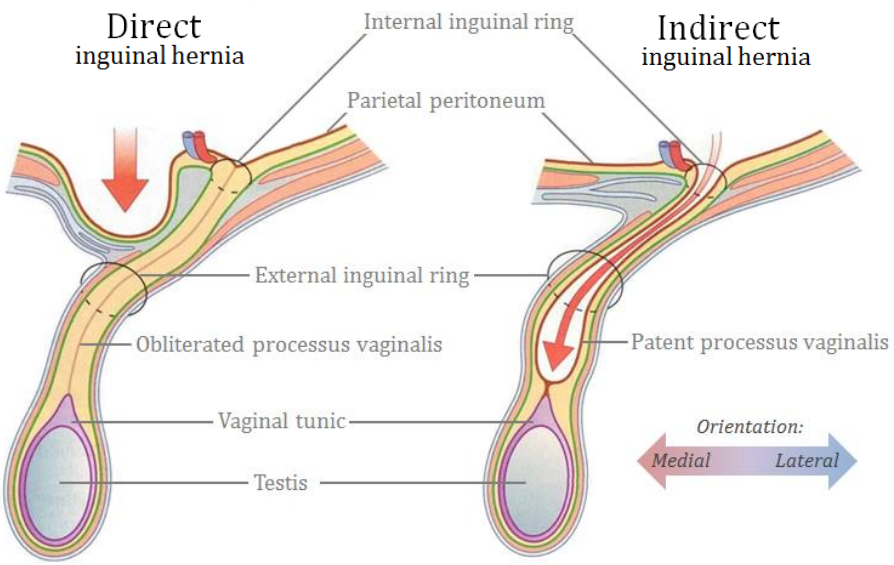


Figure 2: Illustration of human direct and indirect inguinal hernia from an anterior point of view. The illustration is an adaptation of a figure from Elsevier Ltd. Drake et al: Gray's Anatomy for Students.

1.3 Inguinal and scrotal hernia

Inguinal and scrotal hernia is the situation where abdominal content is present in the inguinal canal or the scrotum, respectively. Scrotal hernia can be seen as a more exaggerated version of inguinal hernia, where the abdominal content not only protrudes through the inguinal canal, but all the way to the scrotum.

Inguinal hernia is subdivided into direct and indirect inguinal hernia (Figure 2). Indirect inguinal hernia refers to the protrusion of abdominal content through the vaginal ring, the internal inguinal ring, the inguinal canal, the external inguinal ring and into the vaginal tunic [18]. Direct inguinal hernia describes the situation where abdominal content protrudes through a weakness close to the internal inguinal ring. As the direct inguinal hernia does not go through the vaginal ring, it is not covered by the vaginal tunic or is in other ways associated with processus vaginalis [19].

It is impossible to distinguish between direct and indirect hernia, as well as inguinal and scrotal hernia without clinical examination, which was not performed during sample collection in this study. Pediatric inguinal and scrotal hernias in human are almost exclusively indirect hernias [20]. Likely, this holds true for pigs as well and it is therefore expected that most of the hernia of the study animals is of the indirect type.

Traditionally, the reasons for herniation was said to be a weakness in the abdominal wall, which together with cough, obesity, constipation, pregnancy, or other factors leading to high abdominal pressure would lead to the formation of a hernia. Today, it is said that these factors reveal a hernia, but does not cause it alone, and many efforts have been done to find the underlying defects and conditions that can be seen as the true cause of hernia formation [21].

1.4 Genetic factors affecting inguinal and scrotal hernia

1.4.1 Processus vaginalis and apoptosis

After testicular descent most of the processus vaginalis is obliterated, and only the part closest to the testes remains to form the vaginal tunic. Failed obliteration of the processus vaginalis gives the abdominal content an opportunity to herniate through the vaginal ring and into the processus vaginalis, creating an indirect inguinal hernia. A patent processus vaginalis is a congenital defect and is considered the main reason for development of indirect inguinal hernia [22]. Research confirms that the etiology of indirect inguinal hernia in human is congenital, for adults as well as infants [23].

Apoptosis (programmed cell death) is a key process in the obliteration of the processus vaginalis, and research has been done to find connections between perturbations in the apoptotic pathway and development of inguinal hernia. Calcium (Ca^{2+}) overload is a common event in all types of cell death, and research show that piglets with inguinal and scrotal hernia had significantly less Ca^{2+} in several tissues in the inguinal region than the control group [24]. Thus, it is likely that genes involved in the apoptotic pathway can have an impact on hernia susceptibility.

1.4.2 Development of the testes and testicular descent

Genes for hormones expressed during development of the testes are important candidate genes for congenital abnormalities of testicular descent, such as a patent processus vaginalis or undescended testes. One such hormone is the Leydig insulin-like hormone (*INSL3*) which is expressed in developing testes. Several studies have found associations between the *INSL3* gene and cryptorchidism and other developmental abnormalities in mice [25, 26]. Two studies on pigs found no significant association between polymorphisms in the *INSL3* gene and inguinal hernia [25, 27]. Another study on human found no association between the *INSL3* gene and cryptorchidism (undescended testicle), a defect related to the same anatomical structures as inguinal and scrotal hernia [28].

Another candidate gene is the porcine β-glucuronidase gene (*GUSB*) because of its function within gubernacular tissue during testicular descent. The gubernaculum extends from the undescended testis to the scrotum through rapid growth. This growth is accomplished by forming and depositing hyaluronan within the gubernacular tissue. Then the gubernaculum involutes to allow the testis to descend to the scrotum [26]. This happens due to removal of hyaluronan, a process mediated by the enzymes β-glucuronidase, β-hexosaminidase and

hyaluronidase. A study addressing the *GUSB* gene as a functional candidate for inguinal hernia in pigs found no significant association between polymorphisms in the gene and the presence of inguinal hernia [29].

HOXA10 is an Abdominal B-like homeobox gene that is expressed during the development of the genitourinary system of vertebrates. All male mice homozygous for a targeted disruption of the *HOXA10* gene displayed cryptorchidism [30]. Studies on human aimed to detect polymorphisms in *HOXA10* associated with cryptorchidism remains inconclusive [31, 32]. However, the gene remains as a candidate gene for inguinal and scrotal hernia due to its biological function, and one study found support for association of *HOXA10* with scrotal hernia in pigs [27].

1.4.3 Collagens

Other studies point at altered ratio of different collagens as a reason for development of hernia. In humans, collagen type I is the predominant collagen in skin and is a mechanically stable collagen. Collagen type III is a more unstable and flexible collagen and is important in wound healing. Normal skin is known to have collagen type I and collagen type III in a ratio of 4:1. Skin from patients with inguinal hernia shows an increase in collagen type III, thus decreasing the ratio of collagen type I to collagen type III and making the tissue less rigid. It is proposed that collagen imbalance can serve as an initiating or promoting factor for development of hernias [33, 34]. The ratio of different collagens in different tissues is dependent on multiple genes, and thus collagen imbalance is expected to result from a whole array of genetic alterations rather than one mutated gene, as well as being influenced by environmental factors [35].

One of the genes involved in collagen production is the *COL1A1* gene, which encodes protein chain 1 of collage type I. Research has shown that a specific mutation in the regulatory region recognized by the transcription factor Sp1 increases the transcription of the gene, and that this is associated with occurrence of inguinal hernia in human [36]. Another study on commercial pig lines found a potential association between *COL2A1* and susceptibility of scrotal hernia [27]. Furthermore, alterations of the transcription factors involved in collagen gene regulation may have the same impact on transcription as mutations in the transcription factor binding site. In addition to Sp1, transcription factors such as AP1, Sp3, YB-1, and C/EBP are also important in transcriptional regulation of the collagen genes and may play a role in susceptibility to hernia [35].

Another mechanism that can alter the collagen ratios is the breakdown of collagen. Cleavage of collagens is regulated by the activity of matrix metalloproteinases (MMPs). The MMPs are zinc-dependent endopeptidases that hydrolyze components of the extracellular matrix [33, 34, 37]. Vertebrate MMPs are divided into groups based on their function, and the group collagenases are capable of cleaving the interstitial collagens of type I, II and III. The collagenase group consists of MMP-1, MMP-8, MMP-13, and MMP-18 [38]. In addition, MMP-2 and MMP-14 are also capable of cleaving collagen type I, II and III, but are classified into other groups based on their domain compositions [37].

Studies on human have shown that MMP-2 is important in the pathogenesis of direct inguinal hernia; MMP-2 showed a significant increase in the transversalis fascia (thin membrane between the parietal peritoneum and a muscle in the abdominal wall) of patients with direct inguinal hernia compared to that of patients with indirect inguinal hernia and a control group without hernia [39, 40]. A study on commercial pig lines also showed association between MMP-2 and susceptibility of scrotal hernia [27]. However, it is still unclear whether increased MMP-2 production is the cause of hernia formation, or if it's an effect of it [34]. Association of MMP-1 and MMP-13 with inguinal hernia is not as clear as with MMP-2. Most studies on the subject do not support the involvement of MMP-1 in inguinal hernia, and have failed to detect MMP-13 as this is technically difficult [34].

1.4.4 Quantitative trait loci (QTL) regions on the porcine genome

Many studies have aimed to find QTLs on the porcine genome with association to inguinal and scrotal hernia. Sevillano et al. reports QTLs on *Sus scrofa* chromosome (SSC) 3, 5, 7, 8, and 13 for the breed Large White, and QTLs on SSC1, 2, 4, 10, and 13 for the breed Landrace using the PorcineSNP60 BeadChip [41]. In Pietrain-based lines, QTLs have been found on SSC2 and 12 [42]. Furthermore, for Landrace, Grindflek et al. identified QTLs on SSC1, 5, 6, 7, 15, and 17 using a transmission disequilibrium test (TDT) analysis and QTLs on SSC1, 2, 4, 5, 6, 7, 12, 13, 15, 17, 18, and X using affected sib pair (ASP) test. This study was done using microsatellites, and they identified the QTLs on SSC1, 2, 5, 6, 15, 17, and X to be the most likely to be involved in the etiology of scrotal and inguinal hernia [43]. Another study by Grindflek et al. using the same animal material and the PorcineSNP60 BeadChip identified QTLs on SSC6, 7, 10 and X [44].

The QTLs on SSC2 and SSC12 for both Pietrain [42] and Landrace [43] are in the same genomic region. This is also the case for the QTL on SSC13 for Large White and Landrace [41] as well as the QTL on SSC7 for Large White [41] and Landrace [43].

1.5 Scrotal hernia in the pig industry

Pigs with hernia are of concern for pig producers, as it leads to poor animal welfare as well as economic loss for the producer [43]. In Norway, each case of scrotal hernia costs the producer 170 NOK (~20 euros) [45]. In Europe, the yearly economic loss due to congenital defects in pigs (mainly scrotal hernia, umbilical hernia, cryptorchidism and splay legs) is 200 million euros (Maren van Son, personal communication, 2014).

In the Norsvin nucleus herds, scrotal hernia is the second most common type of hernia, with umbilical hernia as the most common. The observed phenotypic frequency of scrotal hernia is generally low; 0.16 % of all Landrace boars in the nucleus herds born in 2009 had scrotal hernia. However, it is expected that there are underreporting of cases with scrotal hernia and that the true frequency may be higher. For Norsvin Landrace, the heritability of scrotal hernia is estimated to 2.6 % [45].

Since 2008, scrotal hernia has been included in the breeding goals for Landrace under the category "Strength/health" with a relatively low weighting; earlier as 1 % of the total breeding value, but as 2 % of the total breeding value since 2014 [46, 47]. Since both the heritability and the weighting in the breeding goals are low, one cannot expect a large genetic gain [45]. From 1994 to 1999, there has been some genetic gain for scrotal hernia in Norsvin landrace, but from 1999 to 2009 there has been no genetic change [45, 46, 48].

1.6 Whole-genome re-sequencing by Illumina sequencing technology

Next generation sequencing (NGS) is a term used to describe a number of distinct technologies that have unlocked the ability to sequence DNA and RNA in a massively parallel format enabling rapid generation of large data sets (sequence) at a relatively low cost [49].

Today, the US company Illumina is the largest and most successful provider of NGS platforms [50, 51], and their HiSeq2000 model was used for re-sequencing of the pig genome to produce the whole-genome reads used in this thesis [52].

Illumina HiSeq systems utilize an approach called "sequencing-by-synthesis", whereby the identity of single nucleotides is determined as they are sequentially incorporated into a growing DNA strand (Figure 3). "Library preparation" describes a number of steps whereby native DNA is fragmented and specific double-stranded adapters are ligated to each end of the molecule. Individual DNA molecules are then seeded and attach the sequencing flow-cell surface via strand hybridization between adapter sequence and immobilized oligonucleotides. In this environment, it is possible to perform localized bridge-PCR amplification resulting in millions of unique DNA fragment clusters containing approximately one-thousand identical copies of a single progenitor DNA strand [51]. After cluster generation, attached molecules are treated so that they become single stranded and are oriented identically. Decoding the sequence of the strands within a cluster begins with the introduction of a common sequencing primer which anneals to adapter sequence juxtaposed to the native DNA strand. A mixture of four nucleotides is then added to the flow cell, where each type of nucleotide carries a unique fluorescent label, as well as being chemically blocked at the 3'-OH group, preventing elongation. Finally, a high fidelity polymerase is introduced and adds a single nucleotide to the primer dependent upon the DNA template sequence. Fluorescent signal is then detected by imaging before the fluorophore and the chemical block at the 3'-OH group are removed and the cycle is repeated. As all four bases are present at the same time during the reaction, the risk of wrong bases being incorporated is reduced, thereby increasing the sequencing accuracy [50].

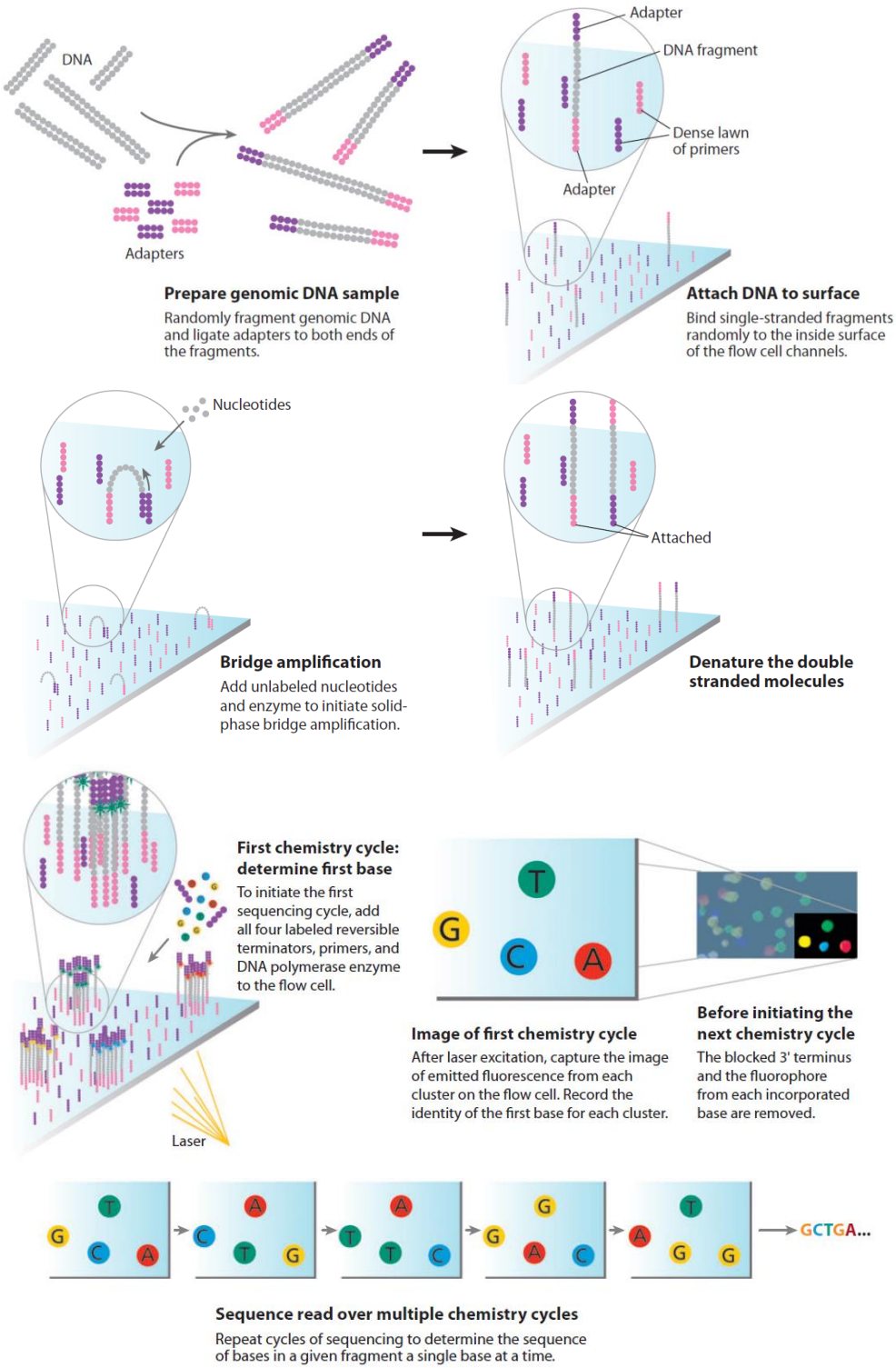


Figure 3: The Illumina sequencing-by-synthesis approach. Cluster strands created by bridge amplification are primed and all four fluorescently labeled 3'-OH blocked nucleotides are added to the flow cell with DNA polymerase. The cluster strands are extended by one nucleotide. Following the incorporation step, the unused nucleotides and DNA polymerase molecules are washed away, a scan buffer is added to the flow cell, and the optics system scans each lane of the flow cell by imaging units called tiles. Once imaging is completed, chemicals that effect cleavage of the fluorescent labels and the 3'-OH blocking groups are added to the flow cell, which prepares the cluster for another round of fluorescent nucleotide incorporation. Figure and text retrieved from Mardis, E.R., *Next-generation DNA sequencing methods*. Annu. Rev. Genomics Hum. Genet., 2008. 9: p. 392-393.

1.7 Genotyping by MALDI-TOF mass spectrometry technology

The genotyping in this study was performed using the MassARRAY® System provided by Agena Bioscience. For each run, as many as 384 samples can be genotyped for 40 SNPs, providing information on more than 15 000 SNPs [53].

To prepare samples for genotyping, a set of up to 40 selected primer pairs are added to each DNA sample and PCR is performed to amplify the fragments containing the SNPs of interest (Figure 4). Next, a cocktail of so-called extension primers (ext-primer), each specific for a single SNP is added. The ext-primers hybridize to a sequence within the PCR amplicons directly adjacent to the SNP position. After hybridization, mass-modified di-deoxynucleotides (ddNTP) are added together with polymerase and a single ddNTP is incorporated into the ext-primer dependent upon the identity of the SNP base in the sample template. Unincorporated ddNTP are removed using gel-filtration and the molecular mass of each ext-primer in a reaction is determined by measuring the molecules mass detected by the MassARRAY® 4 System; MA4 [54].

The MA4 detects which terminator has been added to the extension primer with MALDI-TOF (matrix-assisted laser desorption/ionization – time-of-flight) mass spectrometry (Figure 5). An aliquot of each sample is separately deposited onto a silica chip and, and once placed within the MA4 it is exposed to a short laser pulse. This energy ionizes the extension primers which then accelerate through an electrical field into a field-free drift region. A detector placed at the end of this region registers the time of arrival for all the ionized primers. As smaller molecules moves faster and has a shorter time-of-flight than larger molecules the data analysis software can determine which terminator has been added to the primer by the small mass difference of the different terminators. This enables us to assess which genotype the sampled animal has for all the selected SNPs [54].

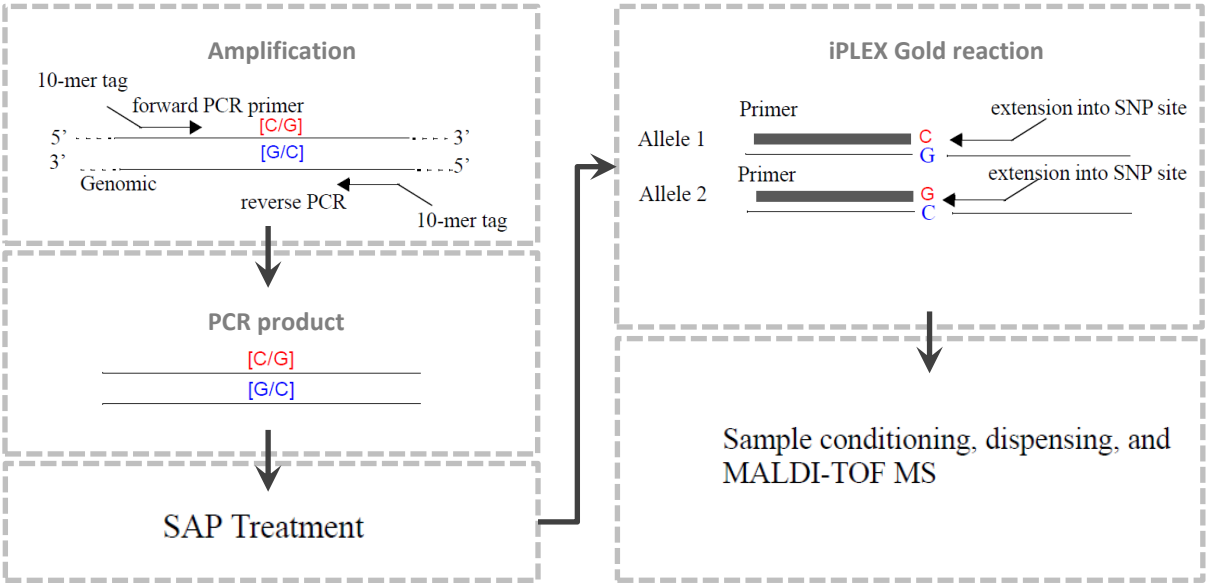


Figure 4: Overview of the iPLEX Gold Reaction performed prior to MALDI-TOF mass spectrometry. First, the fragments containing the SNPs are amplified and treated with SAP to neutralize unincorporated dNTPs. iPLEX Gold reaction is performed to extend extension primers with a mass-modified terminator nucleotide. The figure is an adaptation of a figure in the "iPLEX Gold Application Guide" by Sequenom Bioscience.

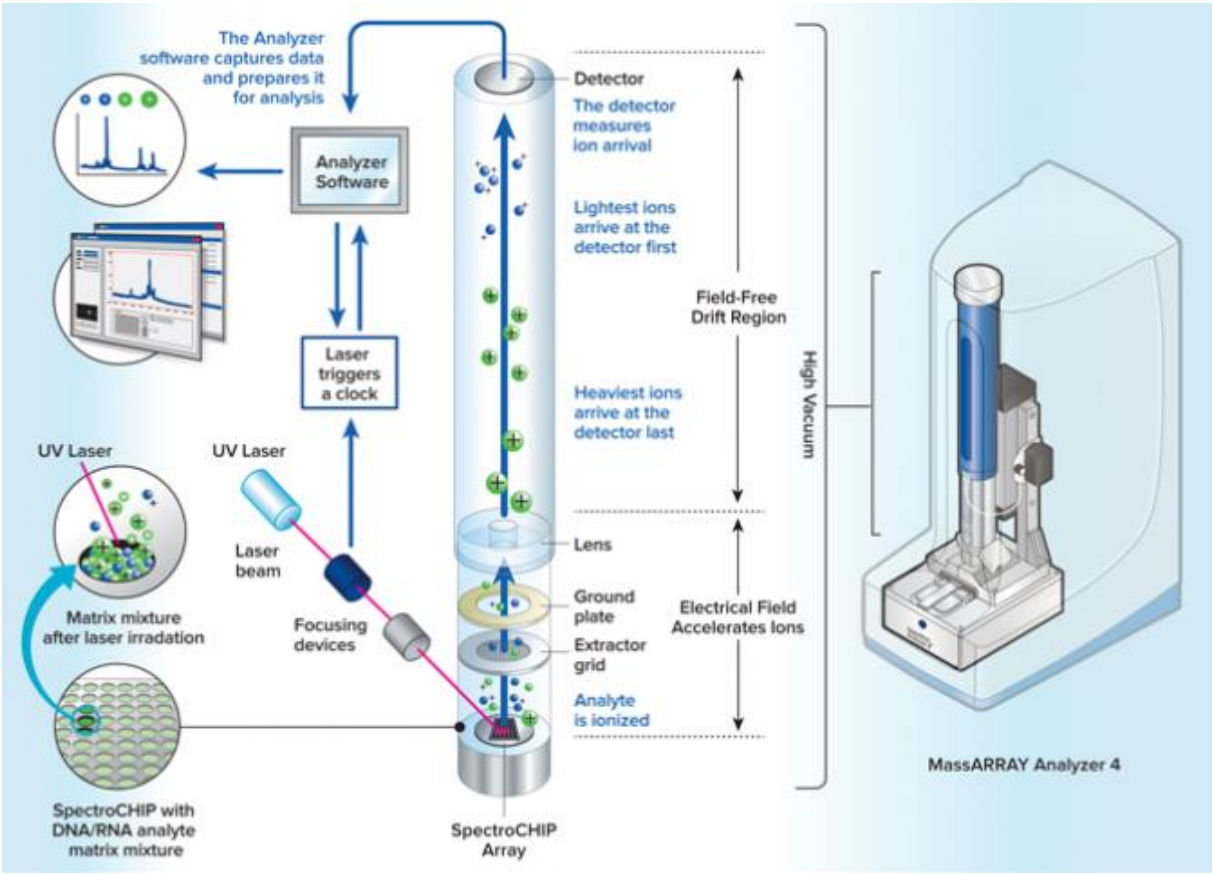


Figure 5: The MALDI-TOF Process in the MassARRAY Analyzer 4. Figure retrieved from http://agenabio.com/genetics.

1.8 A brief introduction to Marker Assisted Selection and Genomic Selection

Marker assisted selection (MAS) is a method of indirect selection, where molecular markers, such as SNPs, linked to the trait of interest is utilized rather than the trait itself. This takes advantage of association of the markers with the trait; the markers have no direct effect on the trait, but are in linkage disequilibrium (LD) with genes or loci with an effect on the trait [55].

Two of the main advantages with utilizing MAS is that animals can be given an estimated breeding value at a very young age compared to traditional methods for obtaining a breeding value, and that animals can be given a breeding value based on traits that are unmeasurable in certain sexes (i.e. milk production cannot be measured in males), obtained late in life or after death (i.e. reproduction traits and carcass quality), and traits that cannot be tested without simultaneously making the animal unfit for production (i.e. disease resistance) [56, 57].

Genomic selection (GS) is a form of MAS, and is also based on indirect selection that utilizes molecular markers. The difference with GS is that instead of focusing on the few markers with high enough association with the trait to be considered significant, it considers all markers simultaneously. The idea is that all QTLs will be in LD with at least one marker, given a sufficiently dense marker map, and that markers in LD with QTLs with little effect on the trait will receive small estimates. Similarly, markers in LD with QTLs with large effects on the trait will receive high estimates [58]. The effects of all markers is used to calculate a genomic estimated breeding value (GEBV) and the selection decisions will be based on these values [59].

1.9 Aims of the study

A previous study has identified several QTLs on the porcine genome that are associated with scrotal hernia. Two QTLs were discovered on SSC6, in addition to one on each of SSC7, SSC10, and SSCX [44]. To identify genes and possible causal SNPs involved in the etiology of scrotal hernia, the two QTLs on SSC6 were chosen for further investigation. The regions that contain the two QTLs were defined and will in this thesis be referred to as SSC6a and SSC6b (Table 1).

Table 1: Overview of the two QTL-containing regions on porcine chromosome 6 chosen for fine mapping.

	SSC6a	**SSC6b**
Chromosome position	6 800 000 – 10 000 000	138 300 000 – 144 800 000
Length of region	3.2 Mb	6.5 Mb

The aim was to do fine mapping of the QTL regions to identify SNPs associated with inguinal hernia, and the procedure in this study was to utilize whole-genome sequences obtained with Illumina sequencing technology to identify SNPs present on the porcine genome in the studied population. Furthermore, case and control animals would be genotyped for a selected SNP panel of SNPs within the QTL regions, and the obtained data supplemented with pre-existing marker data on SNPs within the same chromosomal regions. This would allow for association analysis and identification of SNPs and genes associated with the studied trait. Moreover, it was necessary to assess whether or not the SNPs with statistical significance are situated within a gene or a regulatory region, possible impact on the gene or gene product, and if the biological function of the gene is related to hernia formation. The results of this study provide information about the complex genetic mechanisms underlying inguinal and scrotal hernia in pigs. By implementing the new knowledge about SNPs related to a trait in genomic selection in animal breeding, one might manage to improve the trait in the population [60].

2 Materials and methods

2.1 Animals and phenotypic records

The frequencies of inguinal and scrotal hernia are generally low, thus all 35 Norsvin breeding farms were involved in the sample collection to get a sufficient number of animals. The sampling was performed between 1999 and 2007. Affected animals were reported by the farmers, and diagnostic procedures were performed by breeding consultants from Norsvin. Blood samples from a total of 238 pigs were collected and genomic DNA was extracted with phenol and chloroform. The DNA was quantified using PicoGreen and normalized before genotyping. The quality of the DNA was examined using an agarose gel. The samples were previously genotyped using microsatellites [43] and the PorcineSNP60 BeadChip [44, 61].

The sampled pigs consisted of 111 piglets with inguinal or scrotal hernia and 45 unaffected full- and half-sibs distributed on 56 litters. The samples also included 82 unaffected parents. Each litter contained 2 to 5 piglets where 1 to 5 was defected. In 11 litters all piglets were affected, and 34 litters contained one unaffected sibling. In 12 cases the sire was father to more than one litter.

2.2 SNP detection and filtering

Whole-genome resequencing data from 23 Norsvin Landrace pigs were available from a previous project [52] and were used for SNP detection in the present study. Data (2x100 bp, paired-end reads) were generated using an Illumina HiSeq2000 by a commercial provider (Aros, Denmark). On average, 300 M reads of 100 nucleotides each was available from each pig, which translates roughly to a haploid coverage of 10x.

Reads were trimmed using Trimmomatic v0.32 [62], using ILLUMINACLIP to remove adapter and other Illumina-specific sequences from the reads. The parameters were specified as recommended in Trimmomatic Manual: V0.30 or as the default value, with the exception of “keepBothReads” which was set to “true”. This was necessary to ensure that the files containing forward and reverse reads were balanced, as some of the downstream tools cannot handle unbalanced files (files containing both paired and unpaired reads).

The trimmed reads were aligned to the pig reference genome Build 10.2 [5] using BWA-MEM v0.7.10 with default parameters [63]. Samtools fixmate was used to fill in mate coordinates, ISIZE and mate related flags from a name-sorted alignment [64, 65]. This was

done to prevent potential problems in downstream operations as BWA sometimes leaves unusual flag information; the information given for each mate in paired reads should reflect the information given for the other, and Samtools fixmate provides correct information in cases where the information for a mate has not been updated . Next, Samtools sort was used to sort the alignments by chromosomal coordinates. Samtools calmd was used to calculate a Base Alignment Quality score (BAQ) for each position, which is a a *phred*-scaled probability [66] for each position of being misaligned to the reference sequence. Further, Samtools calmd combines BAQ and the base quality, so that the new quality score contains the probability of the base being wrong, both by sequencing error and misalignment. The Samtools calmd step was included by recommendation from the Samtools manual, and aims to reduce false heterozygotes around INDELs [65].

Freebayes v0.9.18 was used for SNP calling [67]. Mapping quality was included in the calculation of data likelihoods, and the minimum mapping quality was increased from 0 (default) to 1, as reads with a mapping quality of 0 are reads that cannot be uniquely mapped to the reference sequence. A total quality score of at least 30 was required of observations supporting an alternate allele within a single individual in order to evaluate the position. The total quality score is listed under the header line "QUAL" in the Variant Call Format (VCF) files that is given as output from Freebayes. This quality score (Q_s) is a *phred*-scaled probability [66] that the alternative base may be wrong (Eq. 1).

$$Q_s = -10log_{10}Pr\{called\ alternative\ base\ is\ wrong\} \qquad \text{Eq. 1 [68]}$$

A quality of 30 means that on average 1 in 1,000 bases is wrong. As the quality considers both base quality and mapping quality, as well as other factors, a quality of 30 implies that the alignment has few mismatches as well as the base quality of the read being generally good. Lastly, SNPs were only considered for further analysis if they were detected in at least two of the 23 individuals.

Since our goal was to genotype a modest number of known and novel SNPs (≈100 per region) in specific QTL regions, further filtering of all putative SNPs was necessary. Standard VCF files from Freebayes were reformatted in Microsoft Excel 2010 to be able to sort and filter SNPs based on information such as read depth, allele frequency, etc.

"High Quality" putative SNPs were identified based on minor allele frequency (minimum MAF = 0.125), total read depth at the locus (values between the first and fourth quartile, 295-

375 for SSC6a and 330-420 for SSC6b), being strictly bi-allelic, and sequenced in 20 or more of the individuals. A total number of 6434 and 18,099 SNPs met the quality requirements for SSC6a and SSC6b, respectively.

SnpEff v2.1a was used to annotate the SNP variants and predict their impacts on genes [69]. The annotation gives information about amino acid changes, premature stop codons, etc., while the impact gives information on the severity of the change on the protein's function or expression, given as one of the following degrees (in order of severity); "modifier", "low", "moderate" or "high". A list of priority SNPs was created, consisting of SNPs with high impact, missense SNPs, nonsense SNPs, synonymous SNPs, as well as the SNPs present on the PorcineSNP60 BeadChip that were significantly associated with inguinal hernia in a previous project [44].

To ensure an even physical distribution across the regions, and simultaneously reduce the number of SNPs to a more practical target number, a locally developed python script (Appendix I) was used to select evenly distributed SNPs while also giving priorities for selection to SNPs on the priority list. Because proximal SNPs can interfere with Sequenom genotyping, if a selected SNP had proximal SNPs closer than 20 bases on both sides it was manually replaced with a more suitable SNP. Priority SNPs that were not chosen by the script were manually added to the list of selected SNPs.

2.3 Genotyping

For genotyping design, 208 SNPs were chosen from the region corresponding to SSC6a and 152 for the region corresponding to SSC6b. For each selected SNP, a sequence composed of 201 nt (primary SNP at position 101) was extracted from the reference genome, and any secondary SNPs present were manually denoted as `N`.

The online software tool for designing genotyping assays "Assay Design Suite 1.0" from Agena Bioscience was used to design primer pairs for PCR amplification and single extension primers for the iPLEX Gold reaction. The "Mass Range (Da): Upper Limit" was set to 9200 Da, and the "Min Peak Separation" was set to 20 Da; default values were used for all other options. The adjustments were done to accommodate the high sensitivity and detection range of the genotyping equipment used (Sequenom MassArray 4).

For region SSC6a, 208 SNP sequences were provided as input; seven SNPs produced an error report (mainly high dimer potential) and were excluded. The remaining 201 SNPs were

successfully combined into eight multiplexes; four containing the maximum possible of 40 SNPs (40-plexes), one 23-plex, one 12-plex, one 5-plex and one single-plex. To maximize efficiency two of the 40-plexes were chosen for genotyping (Appendix II).

For region SSC6b, 152 SNPs were initially provided as input; six SNPs produced an error report and were excluded. The remaining 146 SNPs were combined into six multiplexes; two 40-plexes, one 35-plex, one 21-plex, one 7-plex and one 3-plex. The three largest multiplexes (two 40-plexes and one 35-plex) were selected for genotyping 115 SNPs (Appendix III).

Genotyping was done using the MassARRAY system (Agena Bioscience (formerly Sequenom), USA). DNA was isolated using Qiagen DNA extraction columns Kit (Qiagen, Germay). After spectrophotometric quantification using NanoDrop 8000 (Thermo Scientific, USA) DNA was normalized to 10 ng/µl. For each of the five multiplexes, DNA was used as template in a PCR reaction containing a master mix of forward and reverse primers prepared by pipetting robots. See Table 2 for a description of the reaction composition.

Table 2: Composition of the PCR mix used for amplification of fragments containing SNPs for genotyping.

	Working concentration (in 5 µl)	Per sample (µl)
dH_2O	NA	1.250
HotStar Taq buffer 10X	1.25X	0.625
dNTPs (25 mM)	500 µM	0.100
PCR primer mix*	0.1 µM	0.500
$MgCl_2$ (25 mM)	1.625 mM	0.325
HotStar Taq Plus 5 U/µl	0.2 U/µl	0.200
Total		**3.000**
gDNA (5-10 ng/µl)	**2-4 ng/µl**	**2.000**
Final reaction volume:		**5.000**

*PCR primer mix contains 1.0 µM forward primer and 1.0 µM reverse primer

Fragments were amplified with the following thermocycler program:

95.0 °C for 5:00 minutes

94.0 °C for 0:30 minute
56.0 °C for 0:30 minute
72.0 °C for 1:00 minute
} 45 cycles

72.0 °C for 1:00 minute

4.0 °C for ever

The PCR product from each reaction was treated with 2 µl of shrimp alkaline enzyme (SAP) master-mix (Table 3) to neutralize unincorporated dNTPs. The reaction volume was 7 µl.

Table 3: Composition of the shrimp alkaline enzyme (SAP) master-mix used for neutralization of unincorporated dNTPs.

	Working concentration (in 7 µl)	Per sample (µl)
dH_2O	NA	1.53
SAP buffer (10X)	0.24X	0.17
Shrimp alkaline phosphatase (SAP) (1.7 U/µl)	0.072 U/µl	0.30
Total volume added to 5 µl PCR products		**2.00**

The combined reaction was incubated as shown below:

37.0 °C for 40:00 minutes

85.0 °C for 5:00 minutes

4.0 °C for ever

Subsequently, 2 µl of the corresponding master mix containing extension primers was added and an iPLEX Gold reaction performed in a total volume of 9 µl. Table 4 details the composition of the iPLEX GOLD cocktail.

Table 4: Composition of the iPLEX GOLD cocktail used for extension of extension primers with mass-modified terminator nucleotides.

	Working concentration (in 9 µl)	Per sample (µl)
dH_2O	NA	0.619
iPLEX Buffer Plus (10X)	0.222X	0.200
iPLEX Termination mix	1X*	0.200
Primer mix (7 µM/9 µM/11 µM/14 µM)**	0.731X/0.94X/1.148X/1.462X	0.940
iPLEX enzyme	1X*	0.041
Total		**2.000**

*Concentration was according to the "iPLEX Gold Application Guide"

**The extension primers for each multiplex were divided into four groups based on their masses. The concentration of each primer in the mix is dependent on the group; low mass primer-groups were added in lower concentrations than high mass primer-groups.

Mass-modified ddNTPs contained in the Termination mix were incorporated into the extension primers while executing the following thermocycler program:

94.0 °C for 0:30 minute

94.0 °C for 0:05 minute
52.0 °C for 0:05 minute
80.0 °C for 0:05 minute
} 5 cycles } 45 cycles

72.0 °C for 3:00 minutes

4.0 °C for ever

Once completed, excess salts and unincorporated ddNTPs were removed from the iPLEX Gold reaction products using a gel-filtration type strategy. Briefly, 25 µl dH_2O was added to each sample together with 6 mg dried gel-filtration resin. The plate was re-sealed and rotated 360° about the short axis for five minutes before being centrifuged at 3200 g for five minutes.

A small volume of each reaction was dispensed onto a SpectroCHIP array using a MassARRAY Nanodispenser. Subsequently, a MALDI-TOF mass spectrometer (MassArray 4, Agena Biosciences, USA) was used to measure the mass of individual extension primers and produce a mass spectrum for each sample and multiplex. Analysis of the mass spectrum is done using TyperAnalyzer software (v4.0.20, Agena Bioscience, Germany). Automatic genotype assignment was manually assessed to ensure accuracy and to identify SNP assays generating anomalous cluster profiles. For a more detailed protocol, please refer to the "iPLEX Gold Application Guide" by Sequenom Bioscience (now: Agena Bioscience) [53].

2.4 Preparation of genotype data

To remove markers with inferior quality, filtering based on the following thresholds was performed; genotype call rate 0.95, MAF 0.025, proportion genotyped limit 0.25 and proportion genotype errors limit 0.025. A total number of 64 and 83 SNPs for SSC6a and SSC6b respectively were successfully genotyped and met the above quality requirements.

The genotype data assessed from the genotyping performed in this study were merged with existing genotype data on the same sample of animals, as to increase the density of markers in the studied QTL regions. The supplemented SNPs comprised all SNPs located within the QTL regions that were previously genotyped with Illumina's PorcineSNP60 BeadChip in the sample population [44]. Beagle was used for imputation of missing genotypes [70].

Chromosome positions for the supplemented SNPs were based on the same reference genome build that the positions of novel SNPs were based on. For SSC6a and SSC6b, 6 and 8 SNPs respectively were genotyped both in this study and in the previous study using the PorcineSNP60 BeadChip [44], and all SNPs with identical chromosome position yielded the same p-value in the analysis. The duplicate SNPs were removed, and for easy comparison with other studies the remaining SNP name was that of the PorcineSNP60 BeadChip.

2.5 Association analysis

ASReml was used to perform the statistical analyses [71]. It uses Residual Maximum Likelihood (REML) when fitting linear mixed models, and was utilized to associate SNP genotypes with the phenotypes. The following model was used when estimating SNP effects:

$$y = \mu + \alpha_i + \beta_j + \gamma_j + x_j + \varepsilon$$

Where y is the binomial trait "case" or "control"

μ is the overall mean

α_i is the SNP effect of SNP i

β_j is the sex of animal j

γ_j is the date of sampling of animal j

x_j is the ID of animal j

ε is the residual

The response variable y is 0 for "controls" and 1 for "cases"; all pigs displaying an inguinal or scrotal hernia are given the response value 1, and pigs not displaying signs of inguinal or scrotal hernia are given the response value 0. SNP effect (α_i), sex (β_j) and date (γ_j) were fitted as fixed effects in the model, while ID (x_j) was fitted as a random effect. A p-value of 0.05 was considered significant.

Haploview 4.2 [72] was used to calculate and visualize linkage disequilibrium (LD) between SNPs. An LD plot for each region was constructed and R^2 was used as a measurement of LD. The Tagger algorithm [73] implemented in Haploview was used for identification of a set of tagSNPs in each region that captured the variation of all significant SNPs in that region.

3 Results

3.1 SNP detection and filtering

SNP detection, performed using Freebayes [67] on whole-genome sequences from 23 pigs, revealed 21,348 SNPs in the SSC6a QTL (3.2 Mb) and 61,392 SNPs in the SSC6b QTL (6.5 Mb). This gives an average of 6.67 SNPs/Kb for SSC6a and 9.44 SNPs/Kb for SSC6b (Table 5).

In SSC6a, the average read depth across detected SNPs was 327x and the average MAF of SNPs was 0.213 (± 0.155). In SSC6b, the average SNP read depth considerably higher at 1563x, while the average MAF was similar at 0.237 (± 0.165). A closer inspection of read depth for all SNPs within the SSC6b region revealed a distinct block of 130 kb (pos. 143.33 - 143.46 Mb) with significantly higher values than the rest of SSC6b; a high read depth block (HRD-block). Excluding the HRD-block from SSC6b region revealed an average SNP read depth of 366x which matches well with SSC6a. This feature is more fully discussed in section 4.2; no SNPs from this region were used in our analysis.

After detection, putative SNPs were filtered strongly to approach a number suitable for Sequenom genotyping. Filters included (i) minor allele frequency >0.125, (ii) read depth from 295-375 for SSC6a and 330-420 for SSC6b, (iii) allowing only bi-allelic SNPs, and (iv) requiring that the SNP be detected in ≥20 of the sequenced individuals. The result of this was 6434 SNPs in SSC6a (2.01 SNPs/Kb) and 18 099 in SSC6b (2.78 SNPs/Kb). The portion of SNPs not meeting each of the above criteria was 35 % and 31 % for MAF, 2.1 % and 3.2 % for non-bi-allelic behavior, and 2.1 % and 1.7 % for genotyped pigs for region SSC6a and SSC6b respectively. As the range of acceptable read depths was computed to capture SNPs falling between the first and fourth quartile for each region, it followed that about 50 % of all SNPs did not meet this criteria. In total, as a considerable number of SNPs failed multiple of the above criteria, the number of SNPs was reduced by 70 % for both regions.

Prior to assay design for Sequenom genotyping, a SNP priority list was assembled for both SSC6a and b. For SSC6a this was composed of 7 missense SNPs, 2 high impact SNPs, and 7 SNPs from the PorcineSNP60 BeadChip; 16 SNPs in total. For SSC6b, it was 19 SNPs; 2 synonymous SNPs, 2 missense SNPs, 3 high impact SNPs, and 12 SNPs from the PorcineSNP60 BeadChip. These priority SNPs were supplemented with a larger number of SNPs selected to display even physical distribution. After design, the two best 40-plexes for

SSC6a contained 15 of the 16 priority list SNPs (sacrificing one high impact SNP). The three multiplexes for SSC6b contained 16 of the 19 priority SNPs (sacrificing two high impact and one PorcineSNP60 BeadChip SNP).

3.2 SNP genotyping

For SSC6a, 65 of 80 attempted SNPs (81 %) were successfully genotyped, and for SSC6b it was 96 of 115 attempted SNPs (83 %) (Table 5). The remaining SNPs, 15 for SSC6a and 19 for SSC6b, gave anomalous clusters and/or had low signal intensities and could not be genotyped.

Genotyping efficiency was high with 96 % and 93 % of successful markers, in SSC6a and SSC6b respectively, producing genotypes in >95 % of the samples (Figure 6 and 7). Relatively few samples, 2 and 5 for SSC6a and SSC6b SNP panels respectively, had genotyping efficiency <70 %. Since the sample identities were different between panels, it is unlikely that the poor performance is related to sample quality and perhaps indicates variability in the lab processing.

The multiplexes for SSC6a contained 73 novel markers and 7 known SNPs taken from the PorcineSNP60 BeadChip; of the 15 SNPs with unsuccessful genotyping, 14 were novel markers and one was from the PorcineSNP60 BeadChip. Except from the one SNP from the PorcineSNP60 BeadChip, all SNPs on the priority list were successfully genotyped.

For SSC6b, the multiplexes contained 104 novel markers and 11 known SNPs taken from the PorcineSNP60 BeadChip. Two of the 19 SNPs that were unsuccessfully genotyped were from the PorcineSNP60 BeadChip, and the remaining 17 SNPs were novel markers. Furthermore, one of these was on the priority list as a high impact SNP.

The genotypes for the SNPs present on the PorcineSNP60 BeadChip obtained through Sequenom genotyping matched the genotyped obtained through Illumina genotyping for the same SNPs and the same samples. This indicates that the two different genotyping technologies give the same results, and that the sample identities are the same in both studies.

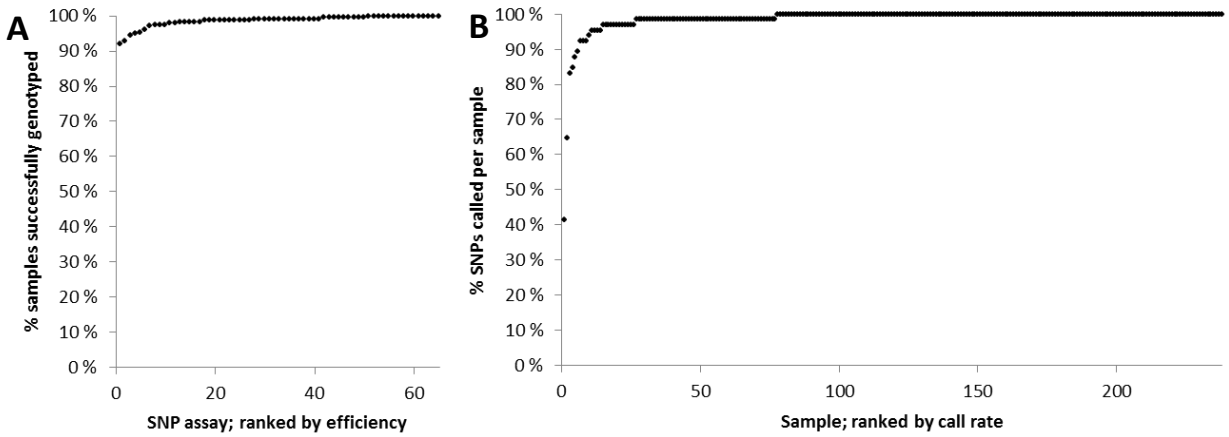


Figure 6: Genotyping efficiency for novel markers within the QTL region SSC6a. **A**. Plot displaying performance of the 65 successfully genotyped SNPs, sorted by fraction of samples genotyped successfully for each SNP. **B**. Plot displaying performance of the 238 samples, sorted by fraction of the 65 SNPs genotyped successfully in each sample.

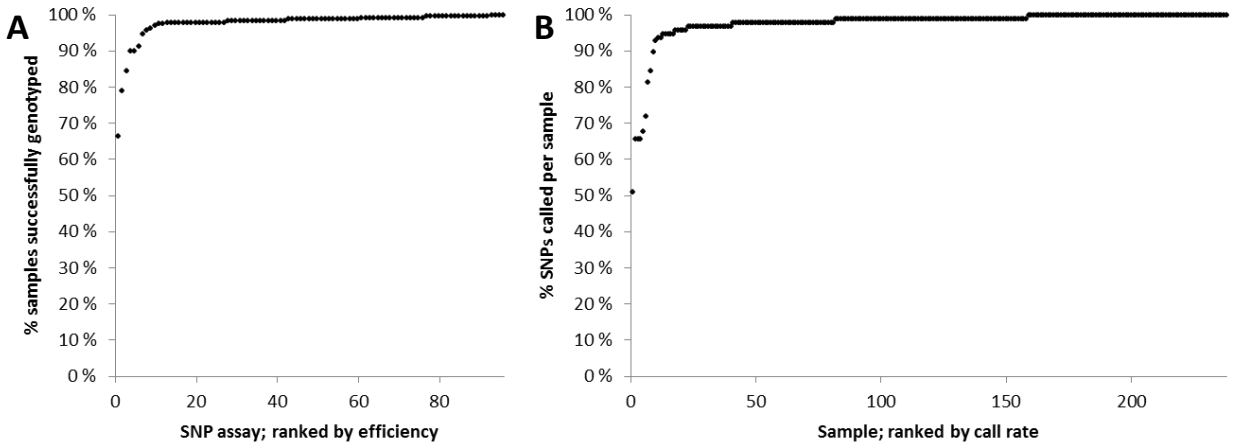


Figure 7: Genotyping efficiency for novel markers within the QTL region SSC6b. **A**. Plot displaying performance of the 96 successfully genotyped SNPs, sorted by fraction of samples genotyped successfully for each SNP. **B**. Plot displaying performance of the 238 samples, sorted by fraction of the 65 SNPs genotyped successfully in each sample.

Table 5: Overview of the number of SNPs present in different stages of the SNP selection process.

	SSC6a		SSC6b	
	SNPs	*SNPs/kb*	*SNPs*	*SNPs/kb*
Detected by Freebayes	21 348	6.67	61 392	9.44
Remaining after quality filtering	6434	2.01	18 099	2.78
Chosen for assay design	208		152	
Included in selected multiplexes	80		115	
Successfully genotyped	65		96	
Suitable for association analysis	64		83	
Novel markers detected by Freebayes*	58		75	
Supplemental markers from PorcineSNP60 BeadChip	82		155	
Total number of SNPs in analyses	**140**	**0.044**	**230**	**0.035**

*Some SNPs were genotyped both in the previous study and in the current study. This number includes SNPs successfully genotyped and suitable for analysis that were solely genotyped in the current study.

3.3 Association and haplotype analyses

3.3.1 Collection of data for association analysis

Some of the successfully genotyped SNPs appeared to be monomorphic or almost monomorphic (MAF < 0.025) and were removed from the data set; one SNP from SSC6a and 13 SNPs from SSC6b. For SSC6b, two SNPs on the priority list were removed for this reason, both from the PorcineSNP60 BeadChip. This was unexpected as previous studies have shown that these SNPs are not monomorphic in the sample population, and it is possible that they actually failed in the assay instead of being successfully genotyped as monomorphic. No SNP from the priority list of SSC6a was removed. A total of 14 SNPs from the priority list of SSC6a were successfully genotyped and suitable for association analysis; 7 missense SNPs, one high impact SNP, and 6 SNPs from the PorcineSNP60 BeadChip. For SSC6b, this number was 11; 2 missense SNPs, 2 synonymous SNPs, and 7 SNPs from the PorcineSNP60 BeadChip.

After combining the data from the novel SNPs with existing data generated from Illumina's PorcineSNP60 BeadChip, the total number of SNPs within the SSC6a region came to 140, and 230 for SSC6b. The average distance between the SNPs within SSC6a was 23.5 (± 28.2) kb, while within SSC6b it was 28.3 (± 29.8) kb. The average MAF was 0.300 (± 0.104) for region SSC6a, and 0.299 (± 0.124) for region SSC6b (Figure 8). There were very few SNPs with low MAF (<0.05) due to the filtering criteria on MAF in both the SNP selection process and filtering of successfully genotyped SNPs.

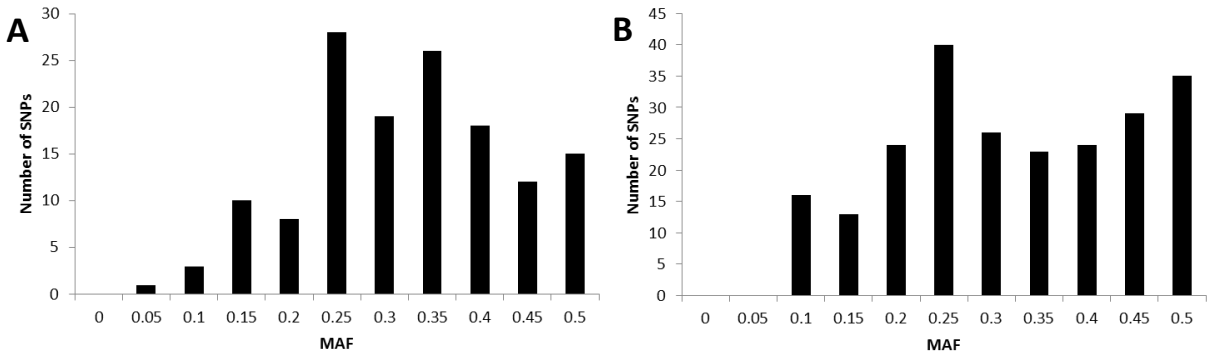


Figure 8: Distribution of minor allele frequencies (MAF) of the SNPs within QTL region SSC6a (**A**) and SSC6b (**B**). Each class has an upper boundary defined by the value (MAF) given for that class, and a lower boundary defined by the upper boundary of the previous class. For instance, in SSC6a, one SNP had a MAF between 0 and 0.05, as shown by **A**.

3.3.2 SSC6a

The association analysis revealed 13 significant SNPs ($P < 0.05$) in the SSC6a region (Table 6, Appendix IV); 6 of these are novel SNPs detected by resequencing, and 7 are pre-existing markers from Illumina's PorcineSNP60 BeadChip. Five of the significant SNPs are located within two uncharacterized genes, *LOC102166616* and *LOC102157459*, found at the beginning and middle of the QTL region respectively. Both genes are coding for mRNA transcripts, but so far there is no evidence for them producing a protein. According to SnpEff (v2.1a) all 13 SNPs have a "modifier" effect, which is the lowest level of impact. None of the novel SNPs on the priority list (7 missense SNPs and one high impact SNP) were significantly associated with inguinal hernia.

Table 6: Association analysis results of the significant SNPs in region SSC6a together with genes affected by the SNPs. SNPs within a gene region are marked with grey.

SNP name	Chr. position	P-value	Region/Genes	Gene description
ALGA0034407	6752216	0.040	LOC102166616 (6586471-6818659)	Uncharacterized gene
ASGA0027406	6813759	0.040		
ALGA0113069	7872494	0.040	Intergenic region	
MARC0056813	7881468	0.040	Intergenic region	
H3GA0017523	7892295	0.040	Intergenic region	
ALGA0124441	8203297	0.016	Intergenic region	
6_8295671	8295671	0.016	Intergenic region	
6_8339436	8339436	0.016	Intergenic region	
6_8372653	8372653	0.016	LOC102157459 (8361115-8559776)	Uncharacterized gene
6_8383536	8383536	0.016		
ALGA0104695	8402621	0.016		
6_8476636	8476636	0.043	Intergenic region	
6_8487631	8487631	0.043	Intergenic region	

The two significant SNPs in *LOC102166616* are in complete LD with each other ($R^2 = 1.0$), however LD between this pair and the other 11 SNPs ranges from $R^2 = 0.187$ to $R^2 = 0.277$ indicating that they are weakly associated with the majority of significant markers. The 11 clustered SNPs are in high LD with each other, with R^2 values ranging from 0.686 to 1.0, and the majority between 0.956 and 1.0. Within this block however, the six SNPs with the lowest p-value are all in complete LD with each other (Figure 9 and 10, red triangle). The two last SNPs in the cluster (6_8476636 and 6_8487631) are in complete LD with each other, but have lower LD with the other SNPs (R^2 between 0.686 and 0.717). The three first SNPs in the cluster are also in complete LD with each other, have a high association with the SNPs with lowest p-value ($R^2 = 0.956$), and a lower association with the two last SNPs ($R^2 = 0.686$) (Figure 10).

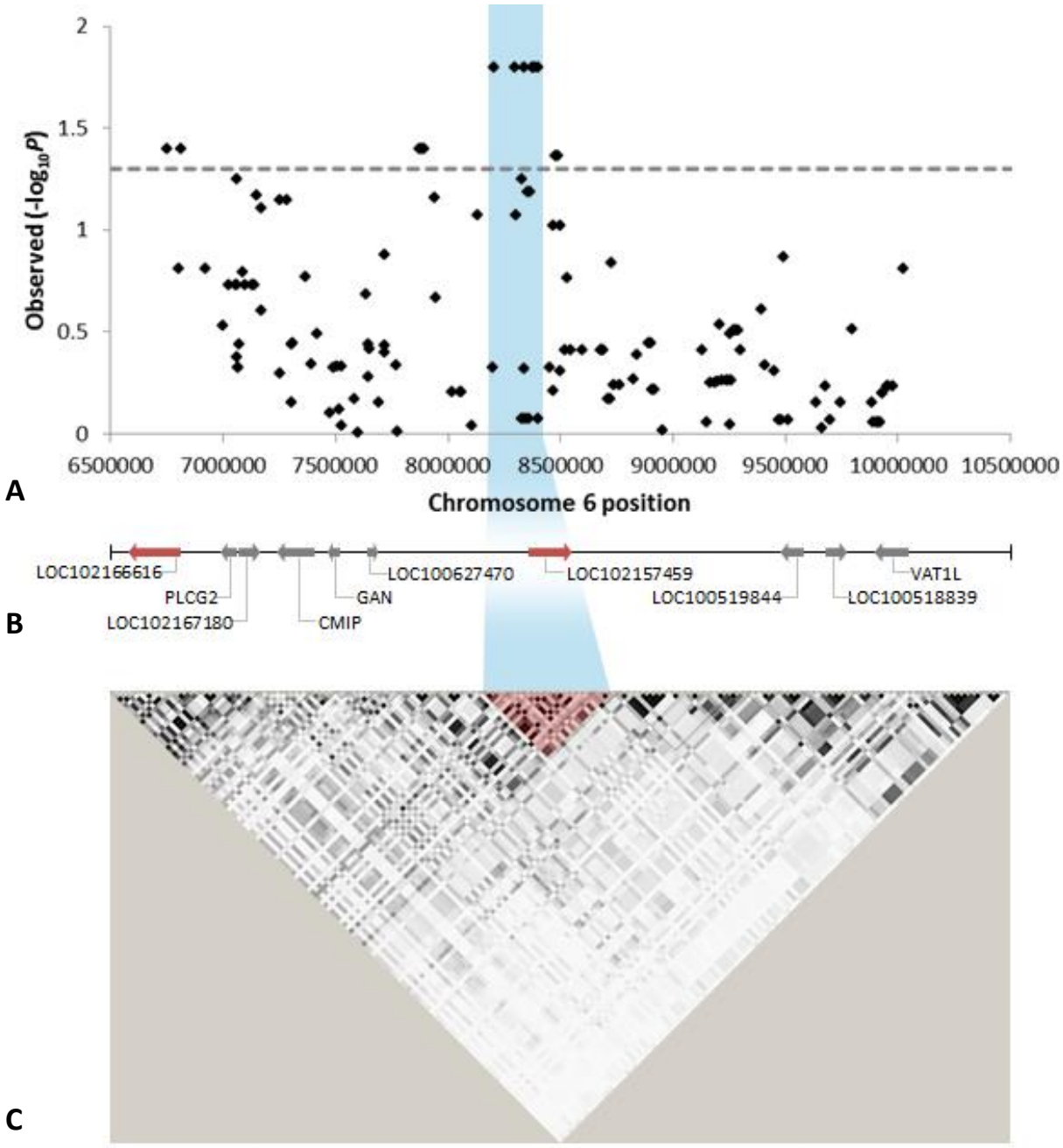


Figure 9: **A**. Association analysis linking 140 SNPs on porcine chromosome 6 to occurrence of hernia. A p-value of 0.05 (grey dashed line) was considered significant. **B**. Map of genes and their orientation in the region, based on information available from NCBI Gene. Genes containing significant SNPs are marked in red. The map is scaled to chromosome positions shown in A. **C**. An LD plot was constructed using Haploview. R^2 was used as a measurement of LD. The color gradient of the diamonds represent the R^2 value, with $R^2 = 0$ shown as white and $R^2 = 1$ shown as black. The red triangle covers the six most significant SNPs, and they are all in complete LD ($R^2 = 1$).

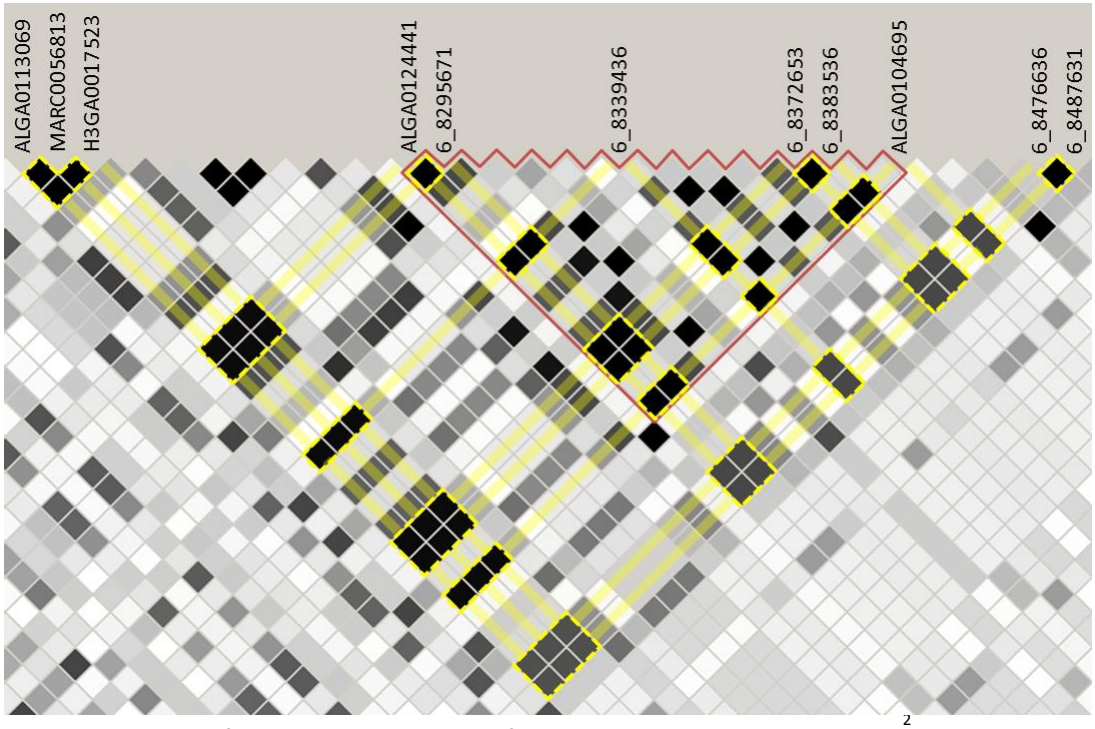


Figure 10: Section of the LD plot constructed for region SSC6a using Haploview. R^2 was used as a measurement of LD. The color gradient of the diamonds represent the R^2 value, with $R^2 = 0$ shown as white and $R^2 = 1$ shown as black. The names of the significant SNPs within this section are shown. The intersections of the yellow lines show the degree of LD between the significant SNPs. The red triangle covers the six most significant SNPs, and they are all in complete LD ($R^2 = 1$).

Table 7 shows the results from the Tagger analysis; ultimately, three SNPs (ALGA0104695, 6_8476636, and ASGA0027406) are capable of capturing 100 % of the alleles of the significant SNPs in SSC6a.

Table 7: Results from Tagger analysis for significant SNPs in region SSC6a. R^2 is the level of LD between the significant SNP (Allele) and the tagSNP (Test).

Allele		Test		
SNP name	**Chr. position**	**SNP name**	**Chr. position**	**R^2**
ALGA0034407	6752216	ASGA0027406	6813759	1.0
ASGA0027406	6813759	ASGA0027406	6813759	1.0
ALGA0113069	7872494	ALGA0104695	8402621	0.956
MARC0056813	7881468	ALGA0104695	8402621	0.956
H3GA0017523	7892295	ALGA0104695	8402621	0.956
ALGA0124441	8203297	ALGA0104695	8402621	1.0
6_8295671	8295671	ALGA0104695	8402621	1.0
6_8339436	8339436	ALGA0104695	8402621	1.0
6_8372653	8372653	ALGA0104695	8402621	1.0
6_8383536	8383536	ALGA0104695	8402621	1.0
ALGA0104695	8402621	ALGA0104695	8402621	1.0
6_8476636	8476636	6_8476636	8476636	1.0
6_8487631	8487631	6_8476636	8476636	1.0

3.3.3 SSC6b

The association analysis revealed 22 significant SNPs ($P < 0.05$) in the SSC6b region (Table 8, Appendix V); 8 of these are novel SNPs detected by resequencing, and 14 are pre-existing markers from Illumina's PorcineSNP60 BeadChip. Twelve SNPs are located in gene regions, affecting a total of 6 genes. Two genes are so far uncharacterized; the remaining genes are nuclear factor I/A (*NFIA*), chromosome 6 open reading frame, human C1orf87 (*C6H1orf87*) and two FGGY carbohydrate kinase domain containing genes (*FGGY*). All SNPs have according to SnpEff (v2.1a) a "modifier" effect, which is the lowest level of impact. None of the novel SNPs on the priority list (2 missense SNPs and 2 synonymous SNPs) were significantly associated with inguinal hernia.

Table 8: Association analysis results of the significant SNPs in region SSC6b together with genes affected by the SNPs. SNPs within a gene region are marked with grey.

SNP name	Chr. position	p-value	Region/Genes	Gene description
MARC0037985	138852864	0.030	NFIA (138764565-139164050)	Nuclear factor I/A
6_138924570	138924570	0.033		
6_139077739	139077739	0.045		
6_139626802	139626802	0.014	Intergenic region	
6_139664663	139664663	0.012	C6H1orf87 (139663734-139782164)	chromosome 6 open reading frame, human C1orf87
ALGA0115609	140397044	0.031	FGGY (140364020-140400250)	FGGY carbohydrate kinase domain containing
ASGA0104579	140475536	0.014	FGGY (140421154-140676162)	FGGY carbohydrate kinase domain containing
ASGA0091723	140480675	0.014		
ALGA0114933	140483422	0.014		
MARC0025692	140511844	0.017		
ASGA0029808	140603308	0.026		
6_140722873	140722873	0.013	LOC102165506 (140676279-140867425)	Uncharacterized gene
6_143795121	143795121	0.050	Intergenic region	
ASGA0095712	143814685	0.050	Intergenic region	
6_143889528	143889528	0.033	Intergenic region	
MARC0016562	143952619	0.029	LOC102158676 (143904566-143995299)	Uncharacterized gene
INRA0022701	144047257	0.029	Intergenic region	
ASGA0029999	144089138	0.004	Intergenic region	
ASGA0030003	144260593	0.033	Intergenic region	
ALGA0037508	144303289	0.044	Intergenic region	
6_144402150	144402150	0.033	Intergenic region	
MARC0093463	144493600	0.025	Intergenic region	

The QTL region SSC6b appears to contain two separate clusters of significant SNPs, one between the chromosome positions 139 Mb and 141 Mb containing 12 significant SNPs; block 1, and one centered on chromosome position 144 Mb containing 10 significant SNPs; block 2 (Figure 11).

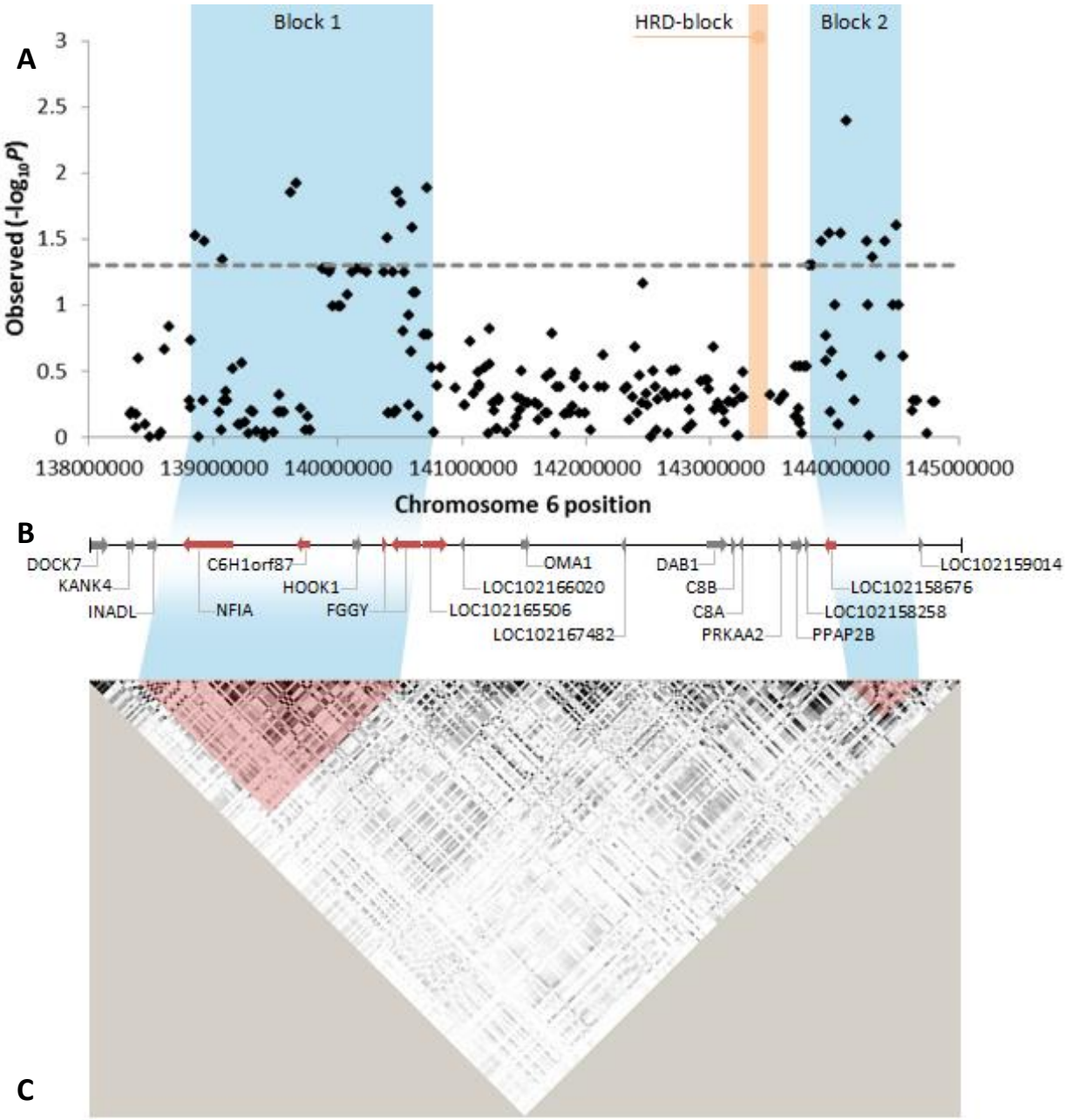


Figure 11: **A**. Association analysis linking 230 SNPs on porcine chromosome 6 to occurrence of hernia. A p-value of 0.05 (grey dashed line) was considered significant. The blue shadowing signifies the two blocks in the region, and the orange shadowing signifies the region with elevated read depth values; HRD-block. **B**. Map of genes and their orientation in the region, based on information available from NCBI Gene. Genes containing significant SNPs are marked in red. The map is scaled to chromosome positions shown in A. **C**. An LD plot was constructed using Haploview. R^2 was used as a measurement of LD. The color gradient of the diamonds represent the R^2 value, with $R^2 = 0$ shown as white and $R^2 = 1$ shown as black.

The pattern of LD between significant SNPs is not as clear for region SSC6b as for region SSC6a. For both clusters of significant SNPs there is high LD between as many as 5 SNPs, but weak LD between those SNPs and other significant SNPs within the same block.

Within the first block of SNPs there are three significant SNPs (Haploview position 14, 40, and 41) that don't show high LD with any other significant SNPs in the same region (Figure 12, A and B), two of which are located within a small region with generally high LD (Figure 12, B). There are five significant SNPs that show generally high LD, three of which are in complete LD with each other. Within the same region, two more significant SNPs are in high LD, but show little or no LD with the other five significant SNPs (Figure 12, C).

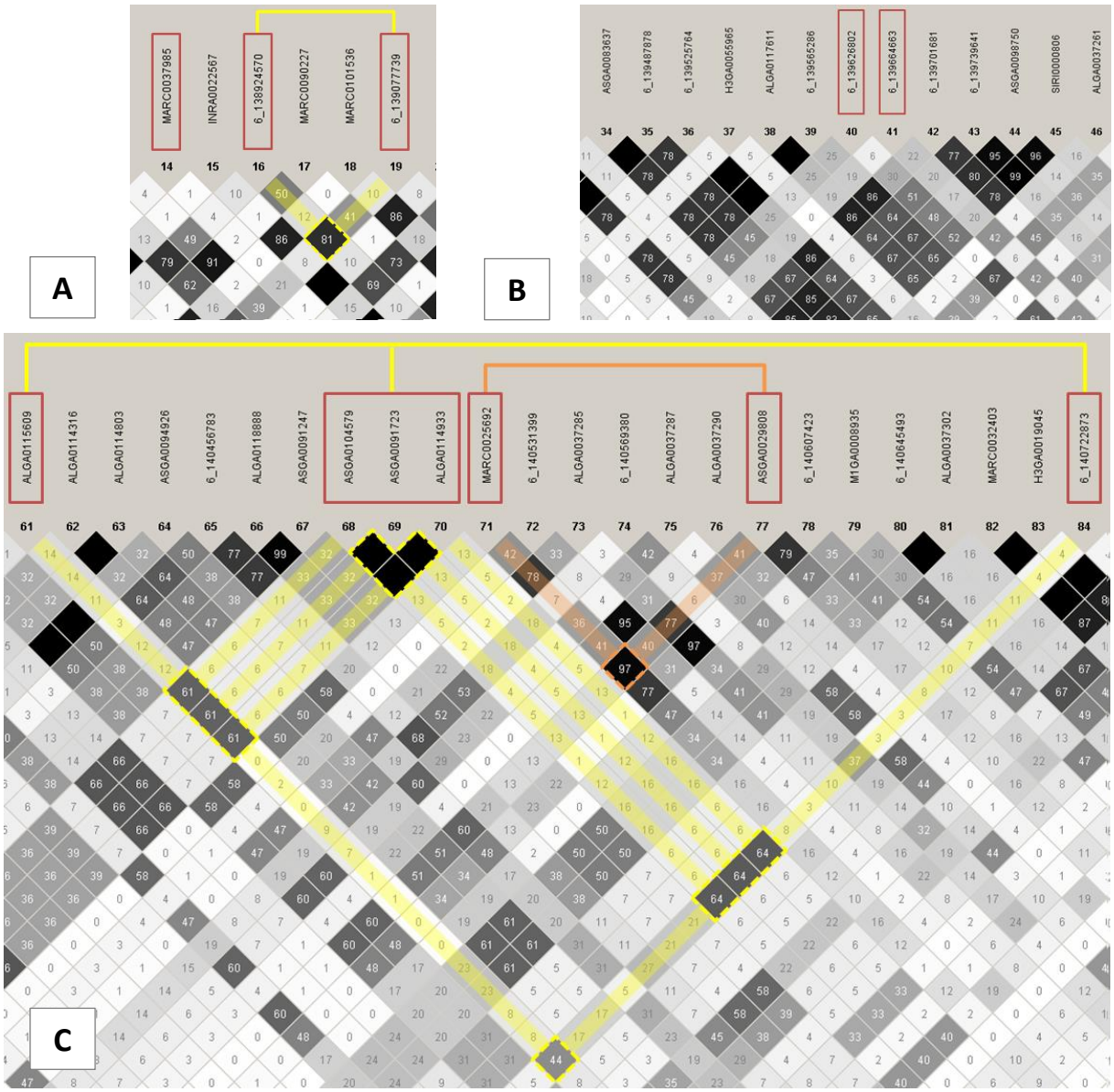


Figure 12: Sections (**A**, **B**, and **C**) of the LD plot constructed for region SSC6b using Haploview, displaying block 1. The sections are from the first cluster of significant SNPs, located between chromosome positions 139 Mb and 141 Mb. R^2 was used as a measurement of LD. The color gradient of the diamonds represent the R^2 value, with $R^2 = 0$ shown as white and $R^2 = 1$ shown as black. The significant SNPs within the sections are indicated with red rectangles. The intersections of the yellow or orange lines show the degree of LD between the significant SNPs, and the topmost yellow and orange lines connect SNPs in high LD.

Within the second cluster of significant SNPs (block 2), five SNPs show very high LD with each other (Figure 13, complete yellow lines), and moderate LD with the SNP with the lowest p-value (Figure 13, dashed yellow line). In the same region, two more significant SNPs are in compete LD, but show little or no LD with the other significant SNPs (Figure 13, orange line). The SNP with the lowest p-value are in moderate LD with one other SNP as well (Figure 13, dashed green line).

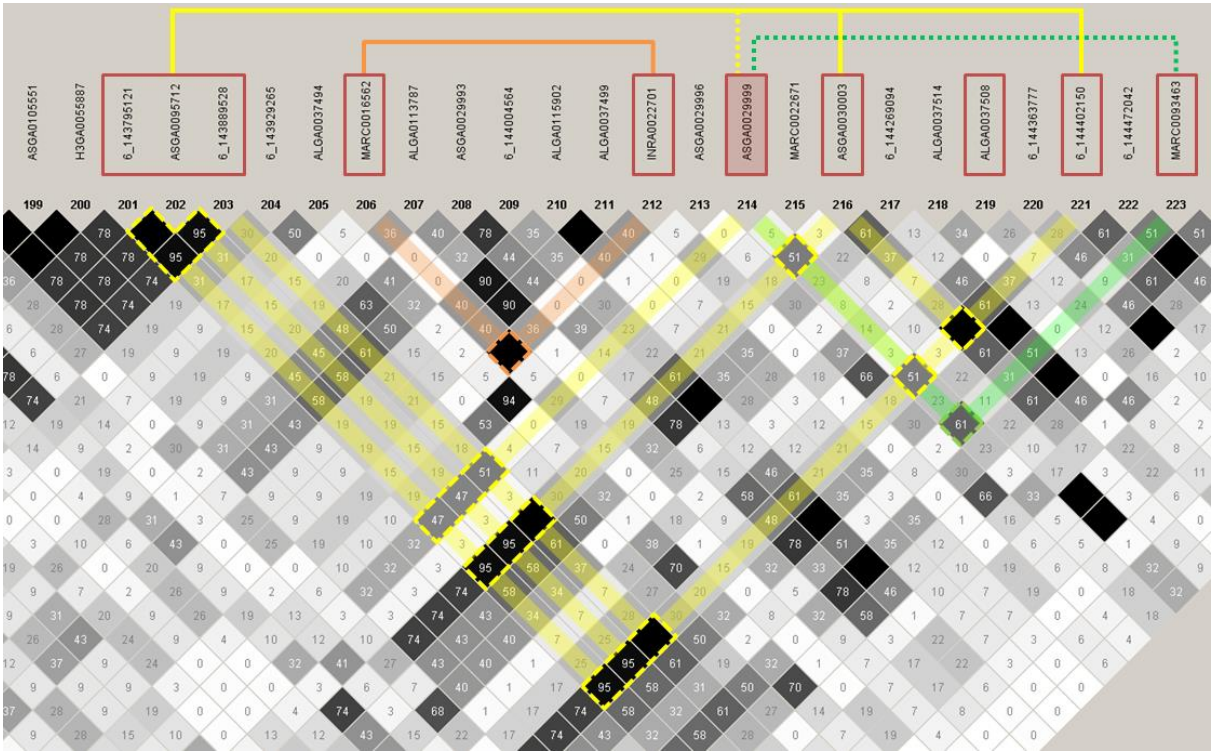


Figure 13: Section of the LD plot constructed for region SSC6b using Haploview, displaying block 2. The section shows the second cluster of significant SNPs; centered on chromosome position 144 Mb. R^2 was used as a measurement of LD. The color gradient of the diamonds represent the R^2 value, with $R^2 = 0$ shown as white and $R^2 = 1$ shown as black. The significant SNPs within the section are indicated with red rectangles. The intersections of the yellow, orange or green lines show the degree of LD between the significant SNPs, and the topmost yellow, orange and green lines connect SNPs in high LD. The SNP with the lowest p-value from the association analysis is indicated with a filled, red rectangle, and is in moderate LD with other SNPs connected by the yellow line, as well as the SNP with Haploview position 223.

Table 9: Results from Tagger analysis for significant SNPs in region SSC6b. R^2 is the level of LD between the significant SNP (Allele) and the tagSNP (Test).

Allele		Test		
SNP name	**Chr. position**	**SNP name**	**Chr. position**	**R^2**
MARC0037985	138852864	MARC0037985	138852864	1.0
6_138924570	138924570	6_139077739	139077739	0.816
6_139077739	139077739	6_139077739	139077739	1.0
6_139626802	139626802	6_139626802	139626802	1.0
6_139664663	139664663	6_140722873	140722873	0.813
ALGA0115609	140397044	ALGA0115609	140397044	1.0
ASGA0104579	140475536	ALGA0114933	140483422	1.0
ASGA0091723	140480675	ALGA0114933	140483422	1.0
ALGA0114933	140483422	ALGA0114933	140483422	1.0
MARC0025692	140511844	MARC0025692	140511844	1.0
ASGA0029808	140603308	MARC0025692	140511844	0.972
6_140722873	140722873	6_140722873	140722873	1.0
6_143795121	143795121	ASGA0030003	144260593	0.952
ASGA0095712	143814685	ASGA0030003	144260593	0.952
6_143889528	143889528	ASGA0030003	144260593	1.0
MARC0016562	143952619	INRA0022701	144047257	1.0
INRA0022701	144047257	INRA0022701	144047257	1.0
ASGA0029999	144089138	ASGA0029999	144089138	1.0
ASGA0030003	144260593	ASGA0030003	144260593	1.0
ALGA0037508	144303289	ALGA0037508	144303289	1.0
6_144402150	144402150	ASGA0030003	144260593	1.0
MARC0093463	144493600	MARC0093463	144493600	1.0

Table 9 shows the results from the Tagger analysis; ultimately, 12 SNPs (Table 10) are capable of capturing 100 % of the alleles of the significant SNPs in SSC6b.

Table 10: Results from Tagger analysis; the 12 significant SNPs that captured 100 % of the alleles of the significant SNPs in SSC6b.

SNP name	**Chr. position**
ASGA0030003	144260593
ALGA0114933	140483422
MARC0025692	140511844
6_140722873	140722873
INRA0022701	144047257
6_139077739	139077739
MARC0037985	138852864
ALGA0037508	144303289
ALGA0115609	140397044
MARC0093463	144493600
6_139626802	139626802
ASGA0029999	144089138

4 Discussion

In this study, whole-genome re-sequencing data was used to detect putative SNPs in two regions on SSC6 (SSC6a and SSC6b) previously associated with inguinal hernia in Norsvin Landrace pigs. After extensive filtering designed to select high quality, informative SNPs, a total of 133 novel markers were successfully genotyped in affected and unaffected animals.

This data was combined with genotypes (n = 237) from the same individuals generated using the Illumina PorcineSNP60 BeadChip, and an association analysis was performed to fine map the QTL regions on SSC6. The results revealed 13 and 22 SNPs significantly associated with inguinal hernia in region SSC6a and SSC6b respectively.

4.1 Variations in inter-SNP distances and SNP frequencies

Even though one of the goals of SNP selection was to select evenly distributed SNPs, there was considerable variation within each region with regards to the physical distance between the novel SNPs used in analyses. One likely reason for the relatively large inter-SNP standard deviations was that SNP selection was not completely random and a number of high impact SNPs (such as those within coding sequence) were specifically included in the assay design even when they fell close to other selected SNPs. The inclusion of markers from Illumina's PorcineSNP60 BeadChip increased the relative size of the standard deviation further as they were included even when they fell close to novel markers. The presence of gaps in the reference assembly also places restrictions on how evenly distributed (per physical distance) SNPs could be. The number of gaps was 12 and 18 for region SSC6a and SSC6b respectively, each gap spanning approximately 50 kb. Taken together, these factors give rise to significant variation in the distances between the SNPs.

The observed frequencies of detected SNPs within the QTL regions were 6.67 and 9.44 SNPs/kb for SSC6a and SSC6b respectively. This difference may be partially attributed to the presence of gaps; the gaps covered 19 % of SSC6a and 14 % of SSC6b. As this was not considered when calculating the SNP frequencies, the frequencies were underestimated, and more for SSC6a than for SSC6b. Furthermore, the HRD-block identified within SSC6b had very high read depths and, as discussed below, SNPs within this region are likely to be false. The SNP frequency within HRD-block is 43.25 SNPs/kb which is 4.6 times higher than that of the whole QTL region. When adjusting the SNP frequencies for gaps and the HRD-block, the calculated frequency is 8.21 and 10.20 SNPs/kb for SSC6a and SSC6b respectively. Compared to the previous frequencies of 6.67 and 9.44 SNPs/kb for SSC6a and SSC6b

respectively, the difference between the regions is smaller, though still evident. However, uneven SNP frequencies across the genome are expected due to uneven distribution of genes as well as the location of the centromeres. Regions closer to a centromere tends to have fewer SNPs as a result of reduced recombination [74], and, as region SSC6a is closer to the centromere than SSC6b, this could also partially explain the observed difference. Region SSC6a did not harbor more gene sequence than region SSC6b, and so the distribution of genes does not appear to offer an explanation of the observed difference.

4.2 The HRD-block in SSC6b

The average total read depth of QTL region SSC6b (1563x) was considerably higher than that of SSC6a (327x), and notably higher that one would expect from 23 animals and a haploid coverage of 10x (expected read depth of ~230x). After closer inspection, a smaller region within SSC6b (the HRD-block) displayed extremely high read depth (as much as 1000 times the average of the rest of SSC6b) (Figure 14). Outside this region, SSC6b had a read depth more similar to that of SSC6a, indicating that the HRD-block was biasing the average total read depth. Indeed, by excluding the HRD-block from calculations, the average total read depth of SSC6b was 367x, which is much more similar to that of SSC6a at 327x.

This region was not identified at the time of filtering and may have created a problem in the SNP filtering process. When filtering on read depth only SNPs with read depth values between the first and fourth quartile were kept, the SNP read depths in the HRD-block were very high and led to an upwards shift in the filtering depths for SSC6b. The applied boundaries were 330-420, however subsequent analysis shows that if all SNPs within the HRD-block were removed the “correct” boundaries would have been 326-406. Fortunately, therefore the impact of few SNPs with extremely high read depth on the filtering boundaries was slight.

Visualizing the HRD-block in UCSC Genome Browser [75] revealed that a series of microsatellite repeats [76] are reported in this region (Figure 15). Curiously, the microsatellite(s) allegedly lie within the first intron of two annotated pig mRNAs (accession AK396675 and AK352268). Such a large repeat structure (116 kb) would have significant consequences on transcription and since these mRNAs are, by definition transcribed, it seems unlikely that the 116 kb repeat element can be truly located between exon 1 and exon 2 as reported and that the reference assembly is incorrect [77].

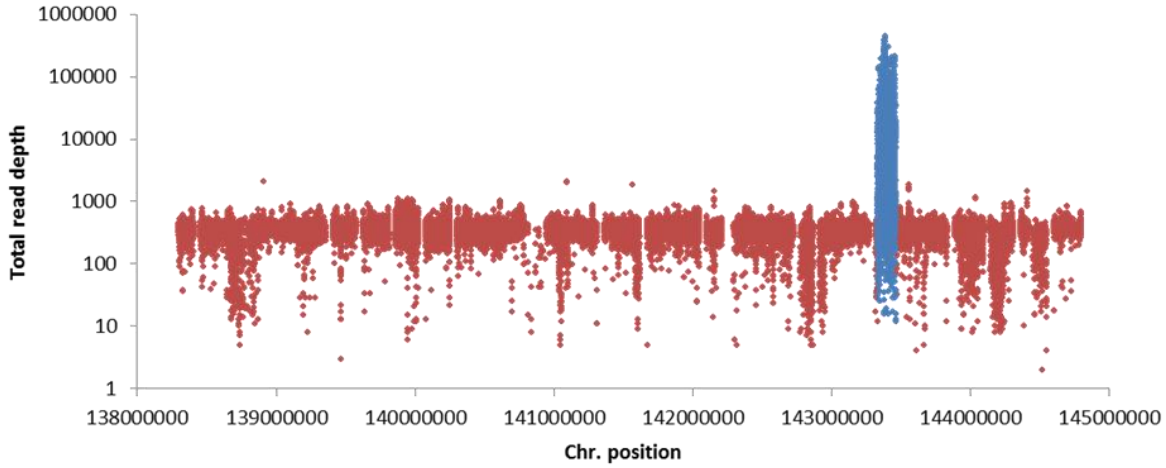


Figure 14: Total read depth at each locus for all SNPs in SSC6b. HRD-block is marked with blue indicators, the rest with red indicators.

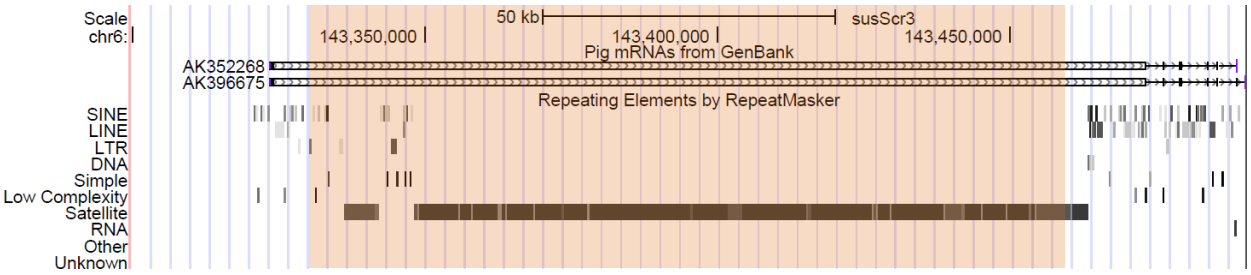


Figure 15: Section from UCSC Genome Browser on "Pig Aug. 2011 (SGSC Sscrofa10.2/susScr3) Assembly" showing chromosome 6, pos. 143,300,000-143,490,000 with the tracks "Base position", "Pig mRNA" and "RepeatMasker". The HRD-block is highlighted in orange and covers 130 kb, while the large satellite covers 116 kb.

Screen shot retrieved from:
https://genome.ucsc.edu/cgi-bin/hgTracks?db=susScr3&position=chr6%3A143300000-143490000&

Correctly assembling and positioning repeats is a major challenge in *de novo* genome assemblies especially when utilizing short NGS sequences. Repeat regions present at multiple locations across the genome can be collapsed by the assembly algorithm, and therefore only present at one location in the reference genome assembly [78]. The pig reference genome was sequenced mainly with use of BACs, supplemented with whole-genome shotgun reads [5]. The use of BACs could go a long way preventing collapse of repeat regions and correctly resolving repeats in the genome, but this is not always so. Thus, it is possible that the repeat region in SSC6b is actually present at multiple locations across the genome and that the assembly algorithm has collapsed all of these regions to the one location on chromosome 6 [78]. In turn, this would lead to increased read depths at the location of the microsatellite complex when whole-genome sequences are aligned to the reference genome. Excluding reads that cannot be uniquely mapped to the reference genome (mapping quality = 0), as was done in this study, would be ineffective at excluding reads wrongly mapped to such a

microsatellite as the microsatellite would only be present at one location in the reference genome. Moreover, when reads from highly similar sequences are wrongly mapped to the same location, SNPs detected in that region are likely to result from minor differences between the regions mapped together and not true polymorphisms present in the studied population.

4.3 Association analysis

None of the novel SNPs detected in the resequencing data displayed a higher significance than that of the previously genotyped SNPs. Our inability to identify SNPs with higher association scores can be the result of a number of limitations. Firstly, we genotyped a relatively small number of new markers, potentially with a denser marker test, a SNP or SNPs could have been revealed with greater association to inguinal hernia. Secondly, the number of animals tested in this study may be a limiting factor, and with a higher number of animals it would be possible to associate the SNPs with inguinal hernia with a higher level of significance. However, it could prove difficult to collect enough animals with inguinal hernia as the incidence of inguinal hernia in the Norsvin Landrace population is low. Thirdly, insecurities in the reference assembly may have meant that the order and/or presence of SNPs may be imperfect.

The relatively large QTL regions in this study contained many genes, and the etiology of the trait may involve several genes; simply put, what may seem to be one QTL might actually be several adjacent QTLs. Indeed, for region SSC6b (6.5 Mb) the significant SNPs appear to roughly divide into two clusters, it is possible that the region actually harbor two QTLs instead of one. Further investigation is required to determine whether the two clusters belong to the same QTL, or to different QTLs within the same region.

None of the significant SNPs located within gene regions appear to have an effect on the amino acid sequence of the protein, and there is no evidence for frameshift variants, nonsynonymous variants, nonsense variants, and splice region variants. While this observation counts towards the exclusion of these specific mutations as being causative, they may still be in linkage disequilibrium with the true causative SNP(s). For SSC6a, this is supported by the degree of LD in the region, which was generally low, but consistently very high between significant SNPs, indicating that a causative SNP might be in high LD with all the significant SNPs. SSC6b did not show such a clear pattern of LD, and together with the two peaks it makes a more complex picture. Furthermore, it is possible that the causative

mutation does not engender a change in protein structure, but has a regulatory impact on gene expression. SNPs located in introns especially may have an effect on the gene expression by altering the way the intron affects mRNA metabolism.

Introns can affect transcription of genes in many ways, and it is known that intron-containing genes have a higher transcriptional efficiency than intronless versions of the same gene. Introns are also known to contain enhancers and repressors that regulate the expression of the gene. In addition, a study showed that mRNA transcribed from cDNA (intronless) was retained in the nucleus, whereas the same mRNA transcribed from constructs containing introns was not. Other ways introns may affect the mRNA is through RNA editing; by changing nucleotides in the pre-mRNA molecule a codon can be reprogrammed to encode a different amino acid, and new splice regions can be introduced [79]. Furthermore, in animals, introns encode almost all small nucleolar RNAs (snoRNA) and a large portion of microRNAs (miRNA), which regulate gene expression [80, 81].

Examples of functional intronic SNPs includes a SNP in the first intron of the human gene *SLC22A4* that was found to be responsible for susceptibility to rheumatoid arthritis (inflammatory disease) [82], a SNP in intron 7 of the human gene *USF1* associated with familial combined hyperlipidemia (disorder that increases blood fats) [83], and a SNP in intron 6 of the human gene *CYP2D6* that causes a nonfunctional mRNA lacking exon 6 to be created [84].

The significant SNPs located in intergenic sequences may also be biologically relevant as they may be located within unmapped genes encoding long non-coding RNA (lncRNA). lncRNAs are RNA transcripts with lengths between 200 b and up to 100 kb that lack significant open reading frames. The human genome is suspected to harbor 7,000 – 23,000 unique lncRNAs, and thus lncRNA represent a large, yet undiscovered, class of transcripts. Some lncRNA have been characterized and seem to be involved in a range of biological functions [81]. Studies in mouse show that the expression of many lncRNAs is restricted to different tissues and different developmental stages, indicating that the expression of lncRNA is highly regulated. Moreover, many lncRNAs seem to regulate expression of nearby protein-coding genes [85]. Further research is needed to assess the possible relation of intronic and intergenic SNPs to inguinal hernia.

4.4 Candidate genes

None of the genes in region SSC6a and SSC6b has previously had attention as functional candidate genes for hernia formation. Out of the total eight genes associated with inguinal hernia in this study, four are uncharacterized and can offer little to our understanding of hernia etiology. The sequences of the uncharacterized genes were used in a sequence BLAST [86] in an attempt to identify the function of similar genes in other organisms, but the results were inconclusive. Of the remaining four genes, two are positioned fairly close to each other and belong to the same gene family, leaving three distinct genes/gene families as potential functional candidates in the SSC6 QTL regions.

4.4.1 NFIA - nuclear factor I/A

The nuclear factor I protein family consists of site-specific DNA-binding proteins. The proteins bind the palindromic sequence 5'-TTGGC(N_5)GCCAA-3' as dimers and play a role in viral DNA replication and regulation of transcription in higher organisms. NFI proteins regulate the transcription of a diverse set of tissue-specific and developmentally regulated genes, thus it appears that NFI proteins is an important factor in development. In vertebrates, the *NFI* gene family consists of four genes (*NFIA*, *NFIB*, *NFIC* and *NFIX*), where each *NFI* gene transcript can be differentially spliced and give rise to several distinct proteins [87].

Studies show that binding sites for NFI proteins are characterized for genes expressed in almost every tissue and organ system, including the testes [87]. *NFIA* is reported to play a role in development of the central nervous system, specifically in development of glial cells [88, 89], brain development, and early postnatal brain maturation [90]. In a study with mice homozygous for a mutant *NFIA* allele, most died shortly after birth and the few that survived developed hydrocephalus (abnormal accumulation of cerebrospinal fluid in the brain) and tremors [87]. *NFIA* also has tumor-promoting functions and may increase cell proliferation and survival. Studies have showed that *NFIA* is abundantly expressed in astroglial tumors in humans [88].

In addition, *NFIA* is reported to play a role in development of the ureter and the urinary tract, and homozygous mutant (*NFIA*$^{-/-}$) mice had abnormalities in the junctions between the bladder and ureter, and kidneys and ureter. Furthermore, humans with microdeletions or translocations involving the *NFIA* gene had urinary tract defects as well as other developmental abnormalities and defects [91, 92]. No studies have shown that this gene is

involved in testicular descent and further studies are needed to clarify the role in inguinal hernia indicated by the results of this study.

4.4.2 C6H1orf87

C6H1orf87 is an open reading frame on pig chromosome 6 that is an ortholog to an open reading frame on the human chromosome 1 (*C1orf87*). The gene is protein coding in both pig and human, but is so far uncharacterized. The human protein is found to have an EF-hand motif, which is known to be calcium ion binding [93].

The number of proteins containing EF-hands is high. Most proteins with EF-hand motifs contain more than one EF-hand, and often they occur in pairs. The functions of the EF-hand containing proteins are diverse, and include buffering of the cytosolic Ca^{2+}, signal transduction, and muscle contractions [94].

It is shown that increased concentration of Ca^{2+} in the cytosol occur both in early and late stages of apoptosis [95], one of the key cellular processes in the formation of inguinal and scrotal hernia. A range of proteins involved in the apoptotic pathway controls or are controlled by the Ca^{2+} concentration. Examples of such proteins that also contain EF-hand motifs are calmodulin (CaM), hippocalcin, and apoptosis-linked gene-2 (ALG-2). CaM contains four EF-hands, and is one of the most studied proteins containing the EF-hand motif. It is a Ca^{2+} sensor protein that participates in a range of cellular processes, including apoptosis [96, 97]. Hippocalcin is expressed in hippocampal neurons and is a known inhibitor of neuronal apoptosis. Its binding of Ca^{2+} by EF-hands is associated with the translocation of the protein to membranes [98]. ALG-2, also known as programmed cell death 6 (PDCD6), is a pro-apoptotic protein that binds Ca^{2+} by five EF-hand structures [99, 100].

Due to the presence of an EF-hand motif in the human ortholog of the *C6H1orf87* it is possible that the protein participates in the apoptotic pathway through interaction with Ca^{2+}. If so, different alleles may contribute to increased risk of inguinal hernia by preventing apoptosis of the cells in processus vaginalis. Further research is required to resolve how *C6H1orf87* may have an impact on increased susceptibility to inguinal hernia.

4.4.3 FGGY - FGGY carbohydrate kinase domain containing

The *FGGY* gene is a protein coding gene, and encodes a protein that phosphorylates carbohydrates. The two *FGGY* genes located on pig chromosome 6 contains a conserved domain that belongs to the Yersinia Pseudotuberculosis carbohydrate kinase-like subgroup of the *FGGY* kinase family, which consists of an uncharacterized carbohydrate kinase named glyerol/xylulose kinase and similar uncharacterized proteins [101].

The *FGGY* kinase family is a large and heterogeneous family, consisting of kinases that phosphorylate a range of sugars, from trioses (C3) to heptoses (C7) [102]. The proteins are composed of two domains, the N-terminal domain (FGGY_N) and the C-terminal domain (FGGY_C), which are structurally related, and the catalytic cleft is formed by the interface between the domains [102, 103].

A *FGGY* gene in human was linked to amyotrophic lateral sclerosis (ALS) in a genome-wide association study (GWAS) and the gene has been proposed as a functional candidate for ALS susceptibility [104]. However, these findings have proved difficult to replicate by other GWASs [105-108]. It is difficult to conclude on the possible association between *FGGY* genes and inguinal hernia due to the diverse cellular functions of the *FGGY* kinase family members and the lack of more specific information on the genes in question.

4.4.4 LOC102157459 – uncharacterized gene

The peak of associated SNPs in SSC6a covers parts of the uncharacterized gene *LOC102157459*, and together with the detected LD in the region this could be a promising functional candidate gene in relation to inguinal hernia in the Norsvin Landrace population. Further research is needed to identify the functions of this gene, and how it can affect susceptibility to inguinal hernia. Also, the Norsvin Landrace population could be more extensively investigated for this gene, for instance by employing a high number of animals to study, identify more SNPs in the relevant region and genotype an even denser marker map to identify the causative SNP.

In this study, a brief BLAST search was performed to identify similar genes in other organisms. Although the results were disappointing, a more comprehensive search for gene homology might shed light on the gene function.

4.5 Implementation of results

This study presents fine mapping of two previously identified QTLs for inguinal hernia and identification of SNPs with association to the trait. None of the SNPs were highly significant, which could be due to the reasons discussed in section 4.3. Moreover, previous studies have identified numerous QTLs underlying the trait [41-44]. Although finding QTLs that explain a large fraction of a phenotype is rare, there are several examples of this in literature and the results implemented in breeding schemes worldwide. One example is a point mutation that explains susceptibility of the porcine stress syndrome (PSS).

PSS covers a group of conditions that includes acute stress, sudden death, and pale, soft, and exudative muscles. Pigs with PSS causes economic losses to the pig industry due to the poor quality of the meat, manifested in the negative effects on color, texture, and loss of fluids [109]. A missense mutation in the ryanodine receptor 1 (*RYR1*) gene was associated with PSS [110] and made it possible to select individuals homozygous for the normal allele for breeding and thereby reducing PSS in commercial populations. The mutation is a recessive mutation and increases the susceptibility of PSS significantly in homozygous carriers. The gene is also known as halothane gene (*HAL*), as homozygous recessive pigs reacts to the anesthetic gas halothane in the same manner as they react to stress, resulting in death [109]. Norsvin used to test for this gene but stopped a few years back as the populations were fixed for the favorable allele (Maren van Son, personal communication, 2015).

Another well-known example is resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon. IPN is a viral disease that causes economical losses to the salmon industry due its high mortality rate. A genome scan using microsatellite markers revealed two genome-wide significant QTLs, one with major effect and one with minor effect. The QTL with major effect was further investigated with linkage-based fine mapping to more precisely define the QTL position. The QTL explained 29 % of the phenotypic variation, and 83 % of the genetic variation [111].

With information about the major QTL, a gene test was developed, and the breeding company Aqua Gen launched IPN-QTL eggs (eggs carrying the IPN resistance allele) in Norway in 2009 [112]. The number of annual outbreaks of IPN prior to the release of IPN-QTL eggs has been stable at 200 for decades. However, after the release, the number of annual outbreaks decreased to 50 in 2013 [113].

Inguinal hernia is not likely to be influenced by one major QTL like the previous examples. Many studies have already been carried out to identify QTLs related to inguinal hernia, both for increased understanding of hernia formation and to identify possible SNPs to use in breeding programs. The significant SNPs of this study could also be used for breeding purposes to reduce the incidence of inguinal hernia. Specifically, the three tagSNPs (ALGA0104695, 6_8476636, and ASGA0027406) in region SSC6a can be used as they capture the alleles of all significant SNPs from this study. Furthermore, if only one SNP is to be genotyped, ALGA0104695 would be the best choice; this SNP alone captured 9 of 13 significant SNPs (69 %), and the SNPs that received the lowest p-value in the association analysis are among the captured SNPs.

The SSC6b region was more complex, with two QTL peaks and a high number of putative SNPs. However, 12 tagSNPs have been suggested and could be used in breeding to reduce the incidence of inguinal hernia in Norsvin Landrace pigs.

4.6 SNP detection and fine mapping of QTLs

Fine mapping of QTL is best undertaken using high-density, population specific polymorphisms, and the abundance of SNPs in the genome makes this an ideal marker for this purpose [114]. Recent advances in sequencing technologies allowing massive parallel sequencing, as well as the availability of reference genome assemblies, makes whole-genome resequencing an efficient way to obtain information on SNPs present in the sequenced population(s) [49, 114]. Technical advances in both sequencing and genotyping are drastically reducing the cost of obtaining sequence and marker data, thus the method of SNP detection through resequencing followed by genotyping and association analysis has become increasingly available as an affordable method of studying complex traits and QTLs [114, 115]. In addition to the use of high-density SNP array, applying association analysis in contrast to linkage analysis improves mapping resolution further as the association analysis utilizes historical recombination events [115, 116].

In this study, aligning re-sequence data on 23 individuals to the reference genome revealed numerous putative SNPs. As genotyping of the selected SNP panel revealed that the majority of SNPs were segregating in the population, it appears that the 23 sequenced individuals were an adequate representation of the population, and that the filtering on MAF gave satisfying results.

In total, the use of resequencing and genotyping allows for fine mapping of multiple targeted QTLs, which in turn facilitates the search for genes and causative mutations underlying complex traits [114, 117]. The methods are relatively new, and with future advances within technology involving sequencing, sequence alignment, SNP detection, multiplexing techniques, and genotyping, the use of SNPs for fine mapping will increase in efficiency [114].

5 Conclusion

This study included the use of whole-genome re-sequencing data for SNP detection, selection of SNPs within two QTL regions for genotyping, and association analysis to identify SNPs with association to inguinal hernia in Norsvin Landrace pigs. In the first QTL region, SSC6a, 13 SNPs were associated with inguinal hernia, and the majority of these SNPs were clustered close to the uncharacterized gene *LOC102157459*. This gene may play a role in susceptibility to inguinal hernia, but further research is required to assess the biological function of the gene. Furthermore, in the second QTL region, SSC6b, 22 SNPs were associated with inguinal hernia, but the clustering of significant SNPs as well as the LD in the region makes this QTL more difficult to evaluate. Still, several partly characterized genes were affected by the significant SNPs, including *NFIA*, *C6H1orf87*, and two *FGGY* genes, and further research would reveal how these may contribute to susceptibility to inguinal hernia.

In conclusion, this study presents fine mapping of two previously identified QTLs for inguinal hernia and identification of SNPs that could be implemented in genetic selection against this defect in Norsvin Landrace. The results indicated several genes that could be involved in susceptibility to inguinal hernia that have not previously had attention as functional candidate genes for hernia formation.

6 References

1. Eurostat. EUROSTAT: Pig population - annual data [Table]. In *Statistical Office of the European Communities*. Available at: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apro_mt_lspig&lang=en (accessed: 02.03.2015).

2. Eurostat. EUROSTAT: Production of meat: pigs [Table]. In *Statistical Office of the European Communities*. Available at: http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&plugin=1&pcode=tag00042&language=en (accessed: 02.03.2015).

3. Animalia. *Gris - Slaktestatistikk*: Animalia. Available at: http://kterm.animalia.no/statistikk/vanligeGrafer/gris (accessed: 23.03.2015).

4. Archibald, A. L., Bolund, L., Churcher, C., Fredholm, M., Groenen, M. A. M., Harlizius, B., Lee, K.-T., Milan, D., Rogers, J., Rothschild, M. F., et al. (2010). Pig genome sequence - analysis and publication strategy. *BMC Genomics*, 11: 438-438.

5. Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, 491 (7424): 393-398.

6. NordGen. *Norwegian landrace*: Nordic Baltic Farm Animal Database. Available at: http://www.nordgen.org/husdyrdb/visrase.php?id=73&langid=1 (accessed: 02.03.2015).

7. *Breeds of Livestock - Norwegian Landrace Swine*. (2002). Oklahoma State University, Department of Animal Science. Available at: http://ansi.okstate.edu/breeds/swine/norwegianlandrace (accessed: 20.04.2015).

8. *Background of Norsvin Landrace*. Swineweb.com. Available at: http://www.swineweb.com/background-norsvin-landrace/.

9. *Foredlingsbesetninger*. (2012). Norsvin. Available at: http://www.norsvin.no/norsvin.no/Avl/Avlsprogram/Avlssystemet/Foredlingsbesetninger (accessed: 11.03.2015).

10. *The Norwegian pig breeding pyramid*. Norsvin. Available at: http://www.norsvin.lt/Genetic-Program/The-Norwegian-pig-breeding-pyramid (accessed: 11.03.2015).

11. *Bruksbesetninger*. (2013). Norsvin. Available at: http://www.norsvin.no/norsvin.no/Avl/Avlsprogram/Avlssystemet/Bruksbesetninger (accessed: 11.03.2015).

12. *Avlsmetode*. (2010). Norsvin. Available at: http://www.norsvin.no/norsvin.no/Avl/Avlslaere/Avlsmetode (accessed: 11.03.2015).

13. Amann, R. P. & Veeramachaneni, D. N. R. (2007). Cryptorchidism in common eutherian mammals. *Reproduction*, 133 (3): 541-561.

14. Heyns, C. F. & Hutson, J. M. (1995). HISTORICAL REVIEW OF THEORIES ON TESTICULAR DESCENT. *Journal of Urology*, 153 (3): 754-767.

15. Dyce K.M., S. W. O., Wensing C.J.G. (1996). Textbook of Veterinary Anatomy. In, pp. 776-785. Philadelphia: Saunders.

16. Williams, M. & Hutson, J. (1991). The history of ideas about testicular descent. *Pediatric Surgery International*, 6 (3): 180-184.

17. Tanyel, F. & Okur, H. (2004). Autonomic nervous system appears to play a role in obliteration of processus vaginalis. *The Journal of Hernias and Abdominal Wall Surgery*, 8 (2): 149-154.

18. *Indirect Inguinal Hernia*. University of Connecticut Health Center. Available at: http://fitsweb.uchc.edu/student/selectives/Luzietti/hernia_inguinal_indirect.htm (accessed: 15.01.2015).

19. *Direct Inguinal Hernia*. University of Connecticut Health Center. Available at: http://fitsweb.uchc.edu/student/selectives/Luzietti/hernia_inguinal_direct.htm (accessed: 15.01.2015).

20. Graf, J. L., Caty, M. G., Martin, D. J. & Glick, P. L. (2002). Pediatric hernias. *Seminars in Ultrasound, CT and MRI*, 23 (2): 197-200.

21. Bendavid, R. (2004). The Unified Theory of hernia formation. *Hernia*, 8: 171-176.

22. Tanyel, F. C., Erdem, S., Buyukpamukcu, N. & Tan, E. (2002). Smooth muscle within incomplete obliterations of processus vaginalis lacks apoptotic nuclei. *Urologia Internationalis*, 69 (1): 42-45.

23. van Wessem, K. J. P., Simons, M. P., Plaisier, P. W. & Lange, J. F. (2003). The etiology of indirect inguinal hernias: congenital and/or acquired? *Hernia*, 7 (2): 76-79.

24. Beuermann, C., Beck, J., Schmelz, U., Dunkelberg, H., Schütz, E., Brenig, B. & Knorr, C. (2009). Tissue Calcium Content in Piglets with Inguinal or Scrotal Hernias or Cryptorchidism. *Journal of Comparative Pathology*, 140 (2–3): 182-186.

25. Knorr, C., Täubert, H., Peters, U. & Brenig, B. (2004). Characterization of Two SNPs (Single Nucleotide Polymorphisms) in the Porcine INSL3 Gene and Their Exclusion as a Common Genetic Basis of Hernia Inguinalis in Pigs. *Biochemical Genetics*, 42 (1-2): 11-19.

26. Klonisch, T., Fowler, P. A. & Hombach-Klonisch, S. (2004). Molecular and genetic regulation of testis descent and external genitalia development. *Developmental Biology*, 270 (1): 1-18.

27. Zhao, X., Du, Z.-Q., Vukasinovic, N., Rodriguez, F., Clutter, A. C. & Rothschild, M. F. (2009). Association of HOXA10, ZFPM2, and MMP2 genes with scrotal hernias evaluated via biological candidate gene analyses in pigs. *American journal of veterinary research*, 70 (8): 1006-1012.

28. Baker, L. A., Nef, S., Nguyen, M. T., Stapleton, R., Nordenskjold, A., Pohl, H. & Parada, L. F. (2002). The insulin-3 gene: lack of a genetic basis for human cryptorchidism. *J Urol*, 167 (6): 2534-7.

29. Beck, J., Bornemann-Kolatzki, K., Knorr, C., Taeubert, H. & Brenig, B. (2006). Molecular characterization and exclusion of porcine GUSB as a candidate gene for congenital hernia inguinalis/scrotalis. *BMC veterinary research*, 2: 14.

30. Satokata, I., Benson, G. & Maas, R. (1995). Sexually dimorphic sterility phenotypes in HoxalO-deficient mice. *Nature*, 374 (6521): 460-463.

31. Kolon, T. F., Wiener, J. S., Lewitton, M., Roth, D. R., Gonzales Jr, E. T. & Lamb, D. J. (1999). ANALYSIS OF HOMEOBOX GENE HOXA10 MUTATIONS IN CRYPTORCHIDISM. *The Journal of Urology*, 161 (1): 275-280.

32. Bertini, V., Bertelloni, S., Valetto, A., Lala, R., Foresta, C. & Simi, P. (2004). Homeobox HOXA10 gene analysis in cryptorchidism. *J Pediatr Endocrinol Metab*, 17 (1): 41-5.

33. Rosch, R., Klinge, U., Si, Z., Junge, K., Klosterhalfen, B. & Schumpelick, V. (2002). A role for the collagen I/III and MMP-1/-13 genes in primary inguinal hernia. *BMC Medical Genetics*, 3 (2).

34. Antoniou, S. A., Antoniou, G. A., Granderath, F. A. & Simopoulos, C. (2009). The role of matrix metalloproteinases in the pathogenesis of abdominal wall hernias. *Eur J Clin Invest*, 39 (11): 953-9.

35. Jansen, P. L., Klinge, U. & Mertens, P. R. (2006). Hernia disease and collagen gene regulation: are there clues for intervention? *Hernia*, 10 (6): 486-491.

36. Sezer, S., Şimşek, N., Çelik, H. T., Erden, G., Ozturk, G., Düzgün, A. P., Çoşkun, F. & Demircan, K. (2014). Association of collagen type I alpha 1 gene polymorphism with inguinal hernia. *Hernia*, 18 (4): 507-512.

37. Nagase, H., Visse, R. & Murphy, G. (2006). *Structure and function of matrix metalloproteinases and TIMPs*, vol. 69. 562-573 pp.

38. Visse, R. & Nagase, H. (2003). Matrix Metalloproteinases and Tissue Inhibitors of Metalloproteinases: Structure, Function, and Biochemistry. *Circulation Research*, 92 (8): 827-839.

39. Bellon, J. M., Bujan, J., Honduvilla, N. G., Jurado, F., Gimeno, M. J., Turnay, J., Olmo, N. & Lizarbe, M. A. (1997). Study of biochemical substrate and role of metalloproteinases in fascia transversalis from hernial processes. *Eur J Clin Invest*, 27 (6): 510-6.

40. Smigielski, J., Kolomecki, K., Ziemniak, P., Drozda, R., Amsolik, M. & Kuzdak, K. (2009). Degradation of collagen by metalloproteinase 2 in patients with abdominal hernias. *Eur Surg Res*, 42 (2): 118-21.

41. Sevillano, C. A., Lopes, M. S., Harlizius, B., Hanenberg, E. H., Knol, E. F. & Bastiaansen, J. W. (2015). Genome-wide association study using deregressed breeding values for cryptorchidism and scrotal/inguinal hernia in two pig lines. *Genetics Selection Evolution*, 47 (1): 18.

42. Du, Z. Q., Zhao, X., Vukasinovic, N., Rodriguez, F., Clutter, A. C. & Rothschild, M. F. (2009). Association and Haplotype Analyses of Positional Candidate Genes in Five Genomic Regions Linked to Scrotal Hernia in Commercial Pig Lines. *Plos One*, 4 (3): 8.

43. Grindflek, E., Moe, M., Taubert, H., Simianer, H., Lien, S. & Moen, T. (2006). Genome-wide linkage analysis of inguinal hernia in pigs using affected sib pairs. *Bmc Genetics*, 7: 12.

44. Grindflek, E., Hansen, M. H. S., Hamland, H., Lien, S. & Son, M. v. *Genome-wide association study for inguinal/scrotal hernias in Norsvin Landrace pigs*. Unpublished manuscript.

45. Andersen-Ranberg, I. M. (2009). Avl for færre defekter og leddsjukdommer hos gris: Norsvin.

46. Grindflek, E., Andersen-Ranberg, I. M. & Hansen, M. (2011). Det genetiske bidraget til pung- og navlebrokk hos gris: Norsvin, CIGENE/UMB.

47. Norsvin. (2014). *Avlsmål Landsvin*.

48. Andersen-Ranberg, I. M. & Heringstad, B. (2007). Avl, helse og gris: Norsvin, UMB, Geno.

49. Hall, N. (2007). Advanced sequencing technologies and their wider impact in microbiology. *Journal of Experimental Biology*, 210 (9): 1518-1525.

50. Bentley, D. R. (2006). Whole-genome re-sequencing. *Current opinion in genetics & development*, 16 (6): 545-552.

51. Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.*, 9: 387-402.

52. Agarwal, R., Van Son, M., Kent, M. P., Lien, S. & Grindflek, E. (2013). Application of whole genome resequencing in the dissection of QTLs affecting boar taint. *EMBnet. journal*, 19 (A): pp. 35-36.

53. Sequenom. (2009). *iPLEX® Gold Application Guide*. Hamburg.

54. *GENETICS* Agena Bioscience. Available at: http://agenabio.com/genetics (accessed: 16.04.2015).

55. Lande, R. & Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124 (3): 743-56.

56. Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142 (1-2): 169-196.

57. Guimarães, E. P. (2007). *Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish*: Food & Agriculture Org.

58. Meuwissen, T. (2003). *Genomic selection: the future of marker assisted selection and animal breeding*. Proc. of International Workshop Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding. 54-59 pp.

59. 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): 433-443.

60. Dekkers, J. C. M. (2012). Application of Genomics Tools to Animal Breeding. *Current Genomics*, 13 (3): 207-212.

61. Ramos, A. M., Crooijmans, R. P., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., Bendixen, C., Churcher, C., Clark, R., Dehais, P., et al. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One*, 4 (8): e6524.

62. Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*.

63. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2 [q-bio.GN]*.

64. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. & Subgroup, G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25 (16): 2078-2079.

65. *Manual page from samtools-1.1*. (2014). Available at: http://www.htslib.org/doc/samtools-1.1.html (accessed: 22.04.2015).

66. Ewing, B. & Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research*, 8 (3): 186-194.

67. Garrison, E. & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv:1207.3907v2 [q-bio.GN]*.

68. *The Variant Call Format (VCF) Version 4.1 Specification*. (2013). The Global Alliance Data Working Group File Formats Task Team.

69. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly*, 6 (2): 80-92.

70. Browning, S. R. & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, 81 (5): 1084-1097.

71. Gilmour, A. R., Gogel, B. J., Cullis, B. R. & Thompson, R. (2006). *ASReml User Guide Release 2.0*. Hemel Hempstead, UK: VSN International Ltd.

72. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21 (2): 263-265.

73. de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J. & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature genetics*, 37 (11): 1217-1223.

74. Stephan, W. & Langley, C. H. (1989). Molecular Genetic Variation in the Centromeric Region of the X Chromosome in Three Drosophila Ananassae Populations. I. Contrasts between the Vermilion and Forked Loci. *Genetics*, 121 (1): 89-99.

75. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler & David. (2002). The Human Genome Browser at UCSC. *Genome Research*, 12 (6): 996-1006.

76. Smit, A., Hubley, R. & Green, P. (2013-2015). RepeatMasker Open-4.0.

77. Hong, X., Scofield, D. G. & Lynch, M. (2006). Intron Size, Abundance, and Distribution within Untranslated Regions of Genes. *Molecular Biology and Evolution*, 23 (12): 2392-2404.

78. Alkan, C., Sajjadian, S. & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nat Meth*, 8 (1): 61-65.

79. Le Hir, H., Nott, A. & Moore, M. J. (2003). How introns influence and enhance eukaryotic gene expression. *Trends in Biochemical Sciences*, 28 (4): 215-220.

80. Mattick, J. S. & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, 15 (suppl 1): R17-R29.

81. Gibb, E. A., Brown, C. J. & Lam, W. L. (2011). The functional role of long non-coding RNA in human carcinomas. *Mol Cancer*, 10 (1): 38-55.

82. Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., et al. (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet*, 35 (4): 341-348.

83. Naukkarinen, J., Gentile, M., Soro-Paavonen, A., Saarela, J., Koistinen, H. A., Pajukanta, P., Taskinen, M.-R. & Peltonen, L. (2005). USF1 and dyslipidemias: converging evidence for a functional intronic variant. *Human Molecular Genetics*, 14 (17): 2595-2605.

84. Toscano, C., Klein, K., Blievernicht, J., Schaeffeler, E., Saussele, T., Raimundo, S., Eichelbaum, M., Schwab, M. & Zanger, U. M. (2006). Impaired expression of CYP2D6 in intermediate metabolizers carrying the* 41 allele caused by the intronic SNP 2988G> A: evidence for modulation of splicing events. *Pharmacogenetics and genomics*, 16 (10): 755-766.

85. Mercer, T. R., Dinger, M. E. & Mattick, J. S. (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet*, 10 (3): 155-159.

86. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol*, 7 (1-2): 203-14.

87. Gronostajski, R. M. (2000). Roles of the NFI/CTF gene family in transcription and development. *Gene*, 249 (1–2): 31-45.

88. Lee, J. S., Xiao, J., Patel, P., Schade, J., Wang, J., Deneen, B., Erdreich-Epstein, A. & Song, H.-R. (2014). A novel tumor-promoting role for nuclear factor IA in glioblastomas is mediated through negative regulation of p53, p21, and PAI1. *Neuro-Oncology*, 16 (2): 191-203.

89. Fancy, S. P. J., Glasgow, S. M., Finley, M., Rowitch, D. H. & Deneen, B. (2012). Evidence that NFIA inhibits repair after white matter injury. *Annals of neurology*, 72 (2): 224-233.

90. Wong, Y., Schulze, C., Streichert, T., Gronostajski, R., Schachner, M. & Tilling, T. (2007). Gene expression analysis of nuclear factor I-A deficient mice indicates delayed brain maturation. *Genome Biology*, 8 (5): R72.

91. Lu, W., Quintero-Rivera, F., Fan, Y., Alkuraya, F. S., Donovan, D. J., Xi, Q., Turbe-Doan, A., Li, Q.-G., Campbell, C. G., Shanske, A. L., et al. (2007). NFIA Haploinsufficiency Is Associated with a CNS Malformation Syndrome and Urinary Tract Defects. *PLoS Genet*, 3 (5): e80.

92. Rao, A., O'Donnell, S., Bain, N., Meldrum, C., Shorter, D. & Goel, H. (2014). An intragenic deletion of the NFIA gene in a patient with a hypoplastic corpus callosum, craniofacial abnormalities and urinary tract defects. *European Journal of Medical Genetics*, 57 (2–3): 65-70.

93. Paula Delgado, A., Brandao, P., Hamid, S. & Narayanan, R. (2013). Mining the Dark Matter of the Cancer Proteome for Novel Biomarkers. *Current Cancer Therapy Reviews*, 9 (4): 265-277.

94. Lewit-Bentley, A. & Réty, S. (2000). EF-hand calcium-binding proteins. *Current Opinion in Structural Biology*, 10 (6): 637-643.

95. Rizzuto, R., Pinton, P., Ferrari, D., Chami, M., Szabadkai, G., Magalhaes, P. J., Virgilio, F. D. & Pozzan, T. (2003). Calcium and apoptosis: facts and hypotheses.(APOPTOSIS-PART 1)(Review). *Oncogene*, 22 (53): 8619.

96. Chin, D. & Means, A. R. (2000). Calmodulin: a prototypical calcium sensor. *Trends in Cell Biology*, 10 (8): 322-328.

97. Youn, H.-D., Sun, L., Prywes, R. & Liu, J. O. (1999). Apoptosis of T Cells Mediated by Ca2+-Induced Release of the Transcription Factor MEF2. *Science*, 286 (5440): 790-793.

98. O'Callaghan, D. W., Tepikin, A. V. & Burgoyne, R. D. (2003). Dynamics and calcium sensitivity of the Ca(2+)/myristoyl switch protein hippocalcin in living cells. *The Journal of Cell Biology*, 163 (4): 715-721.

99. Vito, P., Lacana, E. & D Adamio, L. (1996). Interfering with apoptosis: Ca2positive-binding protein ALG-2 and Alzheimer's disease gene ALG-3. *Science*, 271 (5248): 521.

100. Jung, Y.-S., Kim, K.-S., Dong Kim, K., Lim, J.-S., Kim, J.-W. & Kim, E. (2001). Apoptosis-Linked Gene 2 Binds to the Death Domain of Fas and Dissociates from Fas during Fas-Mediated Apoptosis in Jurkat Cells. *Biochemical and Biophysical Research Communications*, 288 (2): 420-426.

101. Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C., He, J., Gwadz, M. & Hurwitz, D. I. (2014). CDD: NCBI's conserved domain database. *Nucleic acids research*: gku1221.

102. Zhang, Y., Zagnitko, O., Rodionova, I., Osterman, A. & Godzik, A. (2011). The FGGY Carbohydrate Kinase Family: Insights into the Evolution of Functional Specificities. *PLoS Computational Biology*, 7 (12): e1002318.

103. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. *Nucleic Acids Research*, 42 (D1): D222-D230.

104. Dunckley, T., Huentelman, M. J., Craig, D. W., Pearson, J. V., Szelinger, S., Joshipura, K., Halperin, R. F., Stamper, C., Jensen, K. R., Letizia, D., et al. (2007). Whole-Genome Analysis of Sporadic Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, 357 (8): 775-788.

105. Daoud, H., Valdmanis, P. N., Dion, P. A. & Rouleau, G. A. (2009). Analysis of DPP6 and FGGY as candidate genes for amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, 11 (4): 389-391.

106. Cai, B., Tang, L., Zhang, N. & Fan, D. (2014). The single-nucleotide polymorphism rs6690993 in FGGY is not associated with amyotrophic lateral sclerosisin a large Chinese cohort. *Neurobiology of Aging*, 35 (6): 1512.e3-1512.e4.

107. Li, X.-G., Lin, Y.-C. & Zhang, J.-H. (2013). Polymorphism risk factor study on FGGY gene in Chinese patients with sporadic amyotrophic lateral sclerosis. *Beijing Medical Journal*, 5: 002.

108. Van Es, M. A., Van Vught, P. W. J., Veldink, J. H., Andersen, P. M., Birve, A., Lemmens, R., Cronin, S., Van Der Kooi, A. J., De Visser, M., Schelhaas, H. J., et al. (2009). Analysis of FGGY as a risk factor for sporadic amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, 10 (5-6): 441-447.

109. Davoli, R. & Braglia, S. (2008). Molecular approaches in pig breeding to improve meat quality. *Briefings in Functional Genomics*.

110. Fujii, J., Otsu, K., Zorzato, F., De Leon, S., Khanna, V. K., Weiler, J. E., O'Brien, P. J. & MacLennan, D. H. (1991). Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science*, 253 (5018): 448-451.

111. Moen, T., Baranski, M., Sonesson, A. K. & Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. *BMC genomics*, 10 (1): 368.

112. *QTL Eggs – a Novel DNA Test combats IPN in Atlantic Salmon Farming*.

113. *The gene and mechanism behind IPN resistance are identified*. (2014). Aqua Gen. Available at: http://aquagen.no/en/2014/06/17/identification-of-the-gene-and-mechanism-behind-ipn-resistance/ (accessed: 24.04.2015).

114. Wu, X., Ren, C., Joshi, T., Vuong, T., Xu, D. & Nguyen, H. T. (2010). SNP discovery by high-throughput sequencing in soybean. *BMC genomics*, 11 (1): 469.

115. Zhu, C., Gore, M., Buckler, E. S. & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome*, 1 (1): 5-20.

116. Yu, J. & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Current Opinion in Biotechnology*, 17 (2): 155-160.

117. Goddard, M. E. & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 10 (6): 381-391.

Appendix list

Appendix I: Python script for SNP selection

```
#!/usr/bin/env python

# Copyright Harald Grove, CIGENE, 2009
# This is GNU GPL Software: http://www.gnu.org/

# Description:
# A library with some genotype functions

import sys
import math

class searchSNP(object):
    def __init__(self, datafile,lookupfile):
        """ if markfile and pedfile are not valid files, markobj and pedobj are generated locally and can be exported
        """
        self.data = {}
        self.continfo = {}
        self.look = []
        self.limit = {}
        self.lastSNP = {}
        try:
            self.fdata = open(datafile,'r')
            self.flook = open(lookupfile,'r')
        except IOError:
            sys.stderr.write('Error opening files\n')
            sys.exit(1)
        for line in self.fdata:
            l = line.strip().split('\t')
            try: cont,size,pos,t1,t2,t3,q1,q2,q3,qc,q4 = l
            except:
                print "Wrong input format:",len(l),l
                sys.exit(1)
            if cont not in self.data:
                self.data[cont] = [[pos,qc,q1,q2,q3,q4,t1]] # t1 is going to be the column for SNP type
                self.continfo[cont] = size
                self.lastSNP[cont] = 0 # The next marker to start searching
                continue
            self.data[cont].append([pos,qc,q1,q2,q3,q4])
        for line in self.flook:
            l = line.strip().split('\t')
            start,end,t1,isd,snp,t2 = l
            self.look.append([int(end),int(isd),int(snp)])

    def __getitem__(self, key):
        mode,item = key[0],int(key[1])
        if mode == 'snp':
            for el in self.look:
                if item < el[0]: return el[1],el[2]

    def __setitem__(self,key,value):
        pass
```

```
    def validSNP(self,qc,q1,q2,q3,q4):
        if qc == 'TRUE': return True
        return False

    def findSNP(self,cont,position):
        # Assumes size of contig is position of last SNP
        prevpos = 1000000000
        exonSNP = False # Indicates that a SNP positioned in an exon is chosen, only other exon SNP's will be
preffered
        if position == -1: position = int(self.data[cont][0][0])
        prevel = []
        lastSNP = self.lastSNP[cont]
        start = self.lastSNP[cont]
        for i in xrange(start,len(self.data[cont])):
            el = self.data[cont][i]
            pos,qc,q1,q2,q3,q4 = int(el[0]),el[1],el[2],el[3],el[4],el[5]
            # Skip SNP's that are too far away or not passing QC
            if not self.validSNP(qc,q1,q2,q3,q4) or pos < position-self.limit[cont]:
                if i == len(self.data[cont])-1:
                    self.lastSNP[cont] = lastSNP
                    if pos < position-self.limit[cont]:
                        return [] # Last SNP is too far away from target
                    else: return prevel # Last possible snp, not valid so return previous (if any)
                continue # SNP not good enough
            if (pos > position+(position-prevpos) and exonSNP)\
                or (prevel != [] and pos > position+self.limit[cont]): # An exon-SNP is the best option, or limit is
passed with a SNP found
                break
            if exonSNP and q4 != 'Y': continue
            if math.fabs(position-pos) < math.fabs(position-prevpos)\
                or (exonSNP and pos < position+(position-prevpos))\
                or (not exonSNP and q4 == 'Y'): # Store position currently closest to the wanted position
                prevpos = pos
                prevel = el[:]
                lastSNP = i
                if q4 == 'Y': exonSNP = True
                continue
        self.lastSNP[cont] = lastSNP + 1
        return prevel

    def search(self):
        snpList = {}
        for cont in self.continfo:
            #print "Cont:",cont
            size = int(self.continfo[cont])
            isi,antsnp = self.__getitem__(['snp',size]) # required distance and number of snps
            self.limit[cont] = isi/10
            #print "ISD %s, SNP %s" % (isi,antsnp)
            snpList[cont] = []
            if antsnp == 0: continue
            first = self.findSNP(cont,-1)
            if first == []: continue
```

```
            snpList[cont].append(first)
            if antsnp == 2: # Special case when only 2 SNP's are required, take first and last
                second = self.findSNP(cont,size)
                if second != []: snpList[cont].append(second)
                continue
            while 1:
                #print snpList[cont]
                nextpos = int(snpList[cont][-1][0])+int(isi)
                next = self.findSNP(cont,nextpos)
                if next == []: break
                snpList[cont].append(next)
                if nextpos >= size or int(next[0]) == size: break
        return snpList


    def testSpeed(self):
        import time
        t = time.time()
        for i in xrange(100):
            for animal in self.ped['animals']:
                a1,a2 = self.__getitem__([animal,i])
        print "Option 1: %.3f" % (time.time()-t)
        t = time.time()
        for i in xrange(100):
            for animal in self.ped['animals']:
                a1,a2 = self[animal,i]
        print "Option 2: %.3f" % (time.time()-t)

def main():
    try:
        snp = searchSNP(sys.argv[1],sys.argv[2])
    except IndexError:
        print "Use: %s [snpfile] [lookuptable]" % sys.argv[0]
        sys.exit(1)
    snpList = snp.search()
    #print "Gathered snps"
    for el in snpList:
        for line in snpList[el]:
            print "%s %s %s" % (el,line[0],line[1])

if __name__ == '__main__':
    main()
```

Appendix II: Assay design of genotyped SNPs in SSC6a

Appendix II: Assay design of genotyped SNPs in SSC6a. MP is multiplex 1 or multiplex 2. "SNP_ID" is the SNP name, where the number following the underscore is equal to the chromosome position of the SNP. "Length" is the number of nucleotides in the amplified sequence from the PCR reaction. "Dir." is the direction of the extension primer; R is reverse (extended base is the complementary base of the allele given in this table), and F is forward (extended base is the same base as the allele given in this table). All masses are given in Da.

					Unextended extension primer			Extended extension primer			
								Alternate allele		Reference allele	
MP	SNP_ID	Reverse PCR primer	Forward PCR primer	Length	Sequence	Mass	Dir.	Allele	Mass	Allele	Mass
1	6_8054202	ACGTTGGATGACCCCTAGATCAAAAAGCAG	ACGTTGGATGAGAAGTGGACTTGTGGTTGG	102	AAGCAGAACCACCAC	4524.0	R	G	4771.2	A	4851.1
1	6_7000793	ACGTTGGATGTTGCTGTGGCTGTGATGAAG	ACGTTGGATGAAATCTTTTTAGGGCCACGC	116	GTGGCTTCAGCTCCA	4544.0	R	T	4815.2	C	4831.2
1	6_7068363	ACGTTGGATGACATGAAGCTCCAGGGCAG	ACGTTGGATGCATAGCCTGGCAGCCAAATG	97	GGCAGGCCTCTCGCT	4545.0	F	C	4792.1	T	4872.0
1	6_8351513	ACGTTGGATGCTTAAAGCCACTGATGCCTG	ACGTTGGATGAACTGCTGATGACCAAGCTG	100	CCTGCATCCTACATCC	4737.1	R	T	5008.3	C	5024.3
1	6_7718925	ACGTTGGATGGGAAGGAAGCCTTTTTCGTG	ACGTTGGATGTCAAGGGACAGACTTGACAG	99	GCACAGAAGGAACCAC	4893.2	F	C	5140.4	T	5220.3
1	6_7305244	ACGTTGGATGGGATTATAGTGACGCCCATC	ACGTTGGATGTAACTGTCTTGGCCCAGAAC	119	TCACTGGGAAGCTGGC	4922.2	F	C	5169.4	T	5249.3
1	6_7524638	ACGTTGGATGGACCCACGAGCTGAGGAGT	ACGTTGGATGAAAGAGCTGAGCGCTCGCA	117	cCCGTGTCCGACCCTCA	5067.3	R	G	5314.5	C	5354.5
1	6_7062060	ACGTTGGATGTTTGGCCACACCAAAGTTCC	ACGTTGGATGATGGAACTGGGTGGGTTGAG	118	TCCCTTGGCTGACACTT	5112.3	R	T	5383.5	C	5399.5
1	6_9164561	ACGTTGGATGAGGAATTCACTTTCCTCTGC	ACGTTGGATGTCCCTCATGTTTCAGGATGC	118	gCCTCTGCTTCTGAACAT	5425.5	F	A	5696.7	G	5712.7
1	6_9271639	ACGTTGGATGCCCTAGAATGTAGTCTCCAG	ACGTTGGATGCTGGGTTTAGACTTGAGGAC	100	CCAGAGAAACTGCAGCCC	5462.6	F	A	5733.8	G	5749.8
1	6_9448931	ACGTTGGATGTGCTTCTGTTCCCCAAAGAG	ACGTTGGATGCCCAGAGGTCAGATTCTAAG	99	attGGCAGAAGGCACCTA	5532.6	F	C	5779.8	T	5859.7
1	6_8465599	ACGTTGGATGCGCCGGCAACATAAAATAGC	ACGTTGGATGCTTGTGAGGTGGTATTTCGG	97	gAAAGGGGATTGCTTTCC	5554.6	R	G	5801.8	A	5881.7
1	6_7489504	ACGTTGGATGCACACTTACACCCGCTCTTC	ACGTTGGATGCATGCTCGCCAGCTTCAAG	98	ACCCGCTCTTCGCCACCGA	5669.7	F	C	5916.9	T	5996.8
1	6_7872494	ACGTTGGATGGAAGTGACTCTATGCTTGGG	ACGTTGGATGTCAGGTCACAGAATCCTTTC	115	GGCTTTTGAGAGCTCTAAT	5833.8	F	A	6105.0	G	6121.0
1	6_8383536	ACGTTGGATGTGTTGTTCAACCGTCACCAC	ACGTTGGATGTGGGAAAGCTGGGAAGTAAG	100	CATAACCATCACTCCCATAA	5973.9	F	A	6245.1	G	6261.1
1	6_8203297	ACGTTGGATGAGAGGGTGCTCTTTTAATGG	ACGTTGGATGCCTATGACACAGATTAGCAC	109	CTCTTTTAATGGCAGCTATT	6073.0	F	G	6360.2	T	6400.1
1	6_7068912	ACGTTGGATGTCTGACTCTTCTCGTACTCG	ACGTTGGATGTCCGAGTTAGCGCATCCCTC	115	gggGTCCACGCTGATCATGG	6174.0	F	C	6421.2	T	6501.1
1	6_8842067	ACGTTGGATGGTGTTTGTGTGTTTAGAGAG	ACGTTGGATGCGTTCTGTGCTGTTTCTGTG	99	TGTGTTTAGAGAGAACCAGG	6221.1	R	G	6468.2	A	6548.2
1	6_9511328	ACGTTGGATGTCAGTGCACCTTCTGGTAAG	ACGTTGGATGGGGTAAAGTGACAGAGGGTG	100	ttgcGCATGTCCATCCCATTT	6323.1	R	G	6570.3	A	6650.2
1	6_9280901	ACGTTGGATGAACCGAGGTGTAGGTCACAG	ACGTTGGATGTAGGGGTCGAATTGGAGCTG	110	aggaGTCACAGACATGACTCA	6448.2	F	G	6735.4	T	6775.3
1	6_7114849	ACGTTGGATGACTCATGCTACTCACCAGAG	ACGTTGGATGGCTTGTTACAAACAACCCCC	100	ttccAACAACCCCAAGGTCTAT	6623.3	R	G	6870.5	A	6950.4

1	6_7061387	ACGTTGGATGTCTTTTGCAGCATCAGCCTC	ACGTTGGATGTGTCTTTGAGGAACTTGGCG	107	cttCCGCTGGTCAACTTCAAAG	6670.3	R	G	6917.5	A	6997.4
1	6_9407409	ACGTTGGATGAGAGAGGAAAAAGGTGCTGG	ACGTTGGATGCCGTGATGTTGAAAGAGGAC	118	aaacAGGTGCTGGGGTGTCCAA	6824.4	F	C	7071.6	G	7111.7
1	6_9918959	ACGTTGGATGACAAAGACAGTGCCTCCAAG	ACGTTGGATGGAGAGTAAGTTAAGGGAGGG	87	ccaacGGACACTCATCATCTCAC	6897.5	F	C	7144.7	A	7168.7
1	6_8002940	ACGTTGGATGTGGGTTTACAGTGTTCTGTC	ACGTTGGATGATGATGGAAGACAATGTGAG	114	ctaAACTTCTACTGTACAGCAAA	6975.6	F	C	7222.8	G	7262.8
1	6_7938504	ACGTTGGATGAACAGCCAGGCCTAAATGAC	ACGTTGGATGCTCTTGTGAGTAGGAGAGTC	120	ggaagGTTGCTAAATTGCCCTCC	7039.6	F	A	7310.8	G	7326.8
1	6_9941567	ACGTTGGATGTGTAAGGATAACCTGACCCC	ACGTTGGATGCAAAACATTGGGAAGAACAC	117	ggacTAGCAAGTTCAAGTCCACA	7041.6	F	C	7288.8	T	7368.7
1	6_9252732	ACGTTGGATGAAGTGAGATGTAGAAGGGTG	ACGTTGGATGATACCCAATTCCAAGCTCCC	100	AGATGTAGAAGGGTGAAATAACA	7193.7	R	G	7440.9	A	7520.8
1	6_8772744	ACGTTGGATGGGAGAGATTTGGCTTTGTGG	ACGTTGGATGGCTATCATTTAACTTTCAGC	118	tGAAAATAAGGTTTCTGTCTTAAA	7389.8	F	C	7637.0	T	7716.9
1	6_7037667	ACGTTGGATGTTGCAGCAACATGGATGGAC	ACGTTGGATGTAGATTCCATATACGAGTG	115	cGTAAGTCAGACAAAGAAAAACAA	7404.9	F	A	7676.1	G	7692.1
1	6_7517225	ACGTTGGATGAGGAAAAGCTGGATGGAGGG	ACGTTGGATGTTCCAGGATGGAATTGGAGG	119	gggcgGACCGAAGGGAACATTCAA	7460.9	F	G	7748.1	T	7788.0
1	6_9469229	ACGTTGGATGAGTCGCAGGGCCTAATAATC	ACGTTGGATGTCTGTTTACACAGCTAAGGG	100	ataTCACCACTGTCATATTGACGCT	7575.9	R	G	7823.1	A	7903.0
1	6_9885915	ACGTTGGATGTGTAGCTCCAAATCAACCCC	ACGTTGGATGACCTGGCTCCAGTGTATTAC	115	gtaacCCCCCAGCCTAGGAACGTGC	7596.9	R	T	7868.1	A	7924.0
1	6_6801001	ACGTTGGATGAAACTGACTGCCTACCATTC	ACGTTGGATGGCCACTAAATTTGTGGTTA	107	gcggtCACCATTTCATTTCCCGTACT	7848.1	F	G	8135.3	T	8175.2
1	6_8955625	ACGTTGGATGGCACGTTGATGTCTGTTGTC	ACGTTGGATGCTTTCTTTGGCAAGCACACC	95	aggggTGTCTGTTGTCTCTTCAAACT	7983.2	R	G	8230.4	A	8310.3
1	6_8399066	ACGTTGGATGACCACAGGATTACAGGGAAG	ACGTTGGATGTTCAGATTTCCCCCTTCTCC	101	gtgACACACTGTGAAACCGGATATTA	8003.2	F	C	8250.4	T	8330.3
1	6_9404187	ACGTTGGATGGCTATTGTTACCTGATGGAG	ACGTTGGATGAACTGATAACAGCGTCCGTC	106	tatgGTTACCTGATGGAGTTAGATAC	8040.2	R	G	8287.4	A	8367.3
1	6_9757764	ACGTTGGATGTACCTCACCATCTCCTGGAC	ACGTTGGATGAAATTCTTGGGCCCCACCTC	108	gcgcgGGGCTCCACCTCCAGAAATCCT	8206.3	R	C	8493.5	A	8533.4
1	6_7015995	ACGTTGGATGTTCATCAAAAGGTGCTCTCC	ACGTTGGATGTTCGTTAACCACTGCGCCAC	116	gggttTGGTACAGCAGGTTAAGAATCT	8394.5	F	C	8641.6	G	8681.7
1	6_7881998	ACGTTGGATGCAGGGAATTCCTCTTTGAGC	ACGTTGGATGACAGAGATGAAGTGTGAGTG	99	TTATTTTAAAAATAATCTTCTCTTATCAT	8789.8	R	G	9036.9	A	9116.9
2	6_8339436	ACGTTGGATGAGAGAGTCTCTTTTGCCAGC	ACGTTGGATGAGTCCCTGAATGGAGTTTGG	100	TGCCAGCTCCAAAGA	4546.0	R	G	4793.2	T	4817.2
2	6_7416050	ACGTTGGATGTGAAAAGCCGAATGCCAAGG	ACGTTGGATGTTGCCATGAAGCGGTCCAGT	108	CAGGATGGGCTTCAG	4633.0	F	C	4880.2	A	4904.2
2	6_7147835	ACGTTGGATGCATCCCAGACAAAAAGCAGG	ACGTTGGATGAGCTCTGCTTAGAGACACTG	110	TCCTAGAACCCTCCCG	4762.1	F	C	5009.3	T	5089.2
2	6_7769782	ACGTTGGATGACAAGCAAGATGCTGCTCTG	ACGTTGGATGTGATCTGAGCAACAGACGAG	110	gGGGCGTGCGGCCACC	4924.2	F	C	5171.4	G	5211.4
2	6_9744464	ACGTTGGATGCTCGTCATATTCCTACAGGC	ACGTTGGATGGTCACTGTAGGGTTTGAGAC	100	CATGGAGGGCAAAAGA	4988.3	R	C	5275.5	A	5315.4
2	6_8487631	ACGTTGGATGGAAACCATGTTTGCTTGAAC	ACGTTGGATGTTCATCCAGTATTCGAAAGC	109	cTGCTGTACTACCTTGC	5112.3	F	C	5359.5	T	5439.4
2	6_8761799	ACGTTGGATGGGAAAACACGGATATGCACC	ACGTTGGATGATACCCATCACAGCCAACAC	96	TGGAGTTTCGCCTCTTC	5143.3	F	C	5390.5	T	5470.4
2	6_9636486	ACGTTGGATGATCAGGTGGGCAACTGGAAC	ACGTTGGATGTTCACAAGCTTCACCTTCCC	113	AACTGGAACTGTGAGTC	5234.4	F	G	5521.6	T	5561.5
2	6_8609974	ACGTTGGATGCCTGTAAAGTGGTCTCTTCC	ACGTTGGATGTAAAGATCTAGCCCAGTCCC	110	aTTGGGAGAGTGAGGAC	5339.5	F	A	5610.7	G	5626.7
2	6_8978215	ACGTTGGATGATGGCTGGCTGTTTGTTTCC	ACGTTGGATGAAGTTCATGGAAGGGACCAC	100	cGTTTCCACTGAGCTTTC	5416.5	F	C	5663.7	T	5743.6
2	6_7068369	ACGTTGGATGCATAGCCTGGCAGCCAAATG	ACGTTGGATGACATGAAGCTCCAGGGCAG	97	TGTTGAGTTTTACTGCCA	5495.6	R	T	5766.8	C	5782.8

2	6_8295671	ACGTTGGATGAGAGAGGAATGACGGAAAGC	ACGTTGGATGGCTGTGGGAGTTAATGAGTC	100	gaCAGCACTGTGTGGAGA	5588.6	R	T	5859.9	C	5875.9
2	6_9187066	ACGTTGGATGCAGCCAGCCTTTTCTTCTTC	ACGTTGGATGCAACTGGTATGTATGTTTG	115	tGTTCCCGCCAACATTTTC	5689.7	R	T	5960.9	C	5976.9
2	6_8734413	ACGTTGGATGGCTTTATGTGGTTTGTCCCC	ACGTTGGATGACTGAGATCAGGAAGTGACC	114	aCTGTGTTGTTAAGGACCC	5818.8	F	C	6066.0	T	6145.9
2	6_8372653	ACGTTGGATGCAATGCCAGCAAGGGAGG	ACGTTGGATGTGTTGTAACGTGTATGTCGC	120	aaccAGGGAGGCCCATTGT	5837.8	R	T	6109.0	C	6125.0
2	6_8889138	ACGTTGGATGAAACACTCTCTACCCCACTC	ACGTTGGATGTGTTGGAGGTGTGTTCCTTG	100	ggGGTGTTCTCTGCGGATG	5906.8	F	A	6178.0	T	6233.9
2	6_8531573	ACGTTGGATGCTACCATTTGTAACCTGGAC	ACGTTGGATGGGATTTGGGTGGAAGGAATG	115	gggcTAGCTCTGTTCCCCCC	6020.9	F	C	6268.1	A	6292.1
2	6_7635427	ACGTTGGATGGGCTCCATCTGCAGAAATAC	ACGTTGGATGTTCTGCTTTGATCCTGGGTG	100	TGCAGAAATACCAATCACAA	6087.0	R	G	6334.2	C	6374.2
2	6_8712157	ACGTTGGATGAACTGAGAGCCTGAAAAGGG	ACGTTGGATGTTAGGGCAGTTTTGGGCCTC	117	gggTGTGCTTGGTGTTCAAT	6210.0	R	G	6457.2	A	6537.1
2	6_8908946	ACGTTGGATGCTGGTTTCCTGAAAGATGGG	ACGTTGGATGCCCCATGAGGAGCTTATATC	104	GTTTATTTCAACCCTAAAGGA	6404.2	F	C	6651.4	T	6731.3
2	6_9976418	ACGTTGGATGCCTGTGGAACAGATTGTAGC	ACGTTGGATGAGGAAAGCCATCGGATTTTG	111	gGTTAATTGTCTAATGGGCTA	6491.2	R	C	6778.4	A	6818.3
2	6_8923553	ACGTTGGATGCAATGCCAGATCCTTAACCC	ACGTTGGATGCGGAAGTGAACCCAGTTATG	100	gggaaACTGAGCGAGGCCAGG	6570.3	F	A	6841.5	G	6857.5
2	6_9301999	ACGTTGGATGTGGAGATCATACAAGCTGTG	ACGTTGGATGCACTTATGCAGTGGTCATGG	99	cTGCTTTACTGCCTTATTGTAG	6682.3	F	C	6929.5	T	7009.4
2	6_8596789	ACGTTGGATGAACAGGAACGTGAGTGATGG	ACGTTGGATGTCCTCCCTCCTTCTTCAGAC	101	cTTTGCGTGTTCTGCTTTTTAG	6704.3	F	C	6951.5	T	7031.4
2	6_8677642	ACGTTGGATGCACAGACATTGAGGGCAAAC	ACGTTGGATGGGGAACTCCTCTGTAGTGTG	119	GGATAAAGTAGGAATTTGGAAT	6886.5	R	G	7133.7	A	7213.6
2	6_7136808	ACGTTGGATGCTCGCCATTGTGTTTTCAGG	ACGTTGGATGCCCTTTTCACTCATATGTGC	100	ggggaGTCAGGCTCGGAGGGAC	6906.5	R	T	7177.7	C	7193.7
2	6_9480658	ACGTTGGATGTGCAGAATCGTGATGTGTCC	ACGTTGGATGTCCCCGCATTTCCTAAGCAG	116	ccgttTCCGATTGTTTCACGTGT	6987.5	R	G	7234.7	A	7314.6
2	6_8130180	ACGTTGGATGCATTGCCACCAACAGTGTAG	ACGTTGGATGTACCACCTCTCACCAGACCA	118	CAGCATTTGTTATTTGTAATGAT	7058.6	F	G	7345.8	T	7385.7
2	6_7306793	ACGTTGGATGAGGAAGGCCTGAGACAGGAC	ACGTTGGATGATCACGATGGGAACTCCAAG	104	ggcgGAGACAGGACCAGACCAGG	7157.7	R	T	7428.9	C	7444.9
2	6_7791147	ACGTTGGATGGAATGCAGCTCCAATTCGAC	ACGTTGGATGTTGTCTTTTTAGGGCCGCAC	90	accaCTAGTCTGAGAACCTCCATA	7265.8	R	G	7512.9	A	7592.8
2	6_9127363	ACGTTGGATGCACCAAGGAGACGACAAAAG	ACGTTGGATGGGTATCACCACTAGGGTTTC	98	cccgAACTTGCGAGCGGCCTTTCA	7289.7	F	C	7536.9	T	7616.8
2	6_8476636	ACGTTGGATGACTCCACAGATGAAGACAGG	ACGTTGGATGTCGGTCCCATTTCCAAAGTC	110	ggacGAAGGTACCTGTGGAGCCAG	7467.9	F	A	7739.1	G	7755.1
2	6_9679825	ACGTTGGATGGGTGGCTGAGGCTGTAAAAG	ACGTTGGATGACCACAATGGTACTGAGTCC	101	tTAAAAGTAAATAAGGACCATTTCA	7681.1	F	C	7928.2	T	8008.1
2	6_8328490	ACGTTGGATGAGCAGAAAGGCTCTCATCTC	ACGTTGGATGGCAGCAGGGTTTGTGTTTAC	120	aaaagCTCCATTTAAATGGAGACAG	7707.1	R	G	7954.2	T	7978.3
2	6_9241597	ACGTTGGATGGAAAAATCACAAATGCTTCTG	ACGTTGGATGTTCATGGCGATCATGCTAAC	110	CATAATTTTATTTAATATCCTCCAAC	7838.1	F	A	8109.3	T	8165.2
2	6_7059895	ACGTTGGATGAGTTCATCTTTGTGTGCGCC	ACGTTGGATGTGTGGTGGACACCAGAAATC	97	cctcGTGTGCGCCAATTTCCTGTAAA	7897.1	F	C	8144.3	T	8224.2
2	6_7169940	ACGTTGGATGTCTCCCAGAGAACTTAACAG	ACGTTGGATGGGTTTATGAGTTCAGGCCAG	102	agagcACTTAACAGAAATCGAGGACA	8030.3	F	C	8277.4	T	8357.4
2	6_9794687	ACGTTGGATGGGTGAGAAGTGTATGCACAG	ACGTTGGATGGTCATGGATGCTTCCACTTC	100	agggCAGAACGATGCCAGGTAATTTA	8068.3	R	G	8315.4	A	8395.4
2	6_9270142	ACGTTGGATGGCCCTAGAAACCTGTAAGTC	ACGTTGGATGTTCAACGTGCTTGTCAGCAG	107	caccaTGTCACCAACTGCGAGATTATC	8188.3	F	C	8435.5	T	8515.4
2	6_9208969	ACGTTGGATGATCCTCGCAGGTAGATACAC	ACGTTGGATGCCCGTGACAGATTTATCCTC	112	aaagCGCAGGTAGATACACAGGCACAT	8335.4	R	G	8582.6	C	8622.7

Appendix III: Assay design of genotyped SNPs in SSC6b

Appendix III: Assay design of genotyped SNPs in SSC6b. MP is multiplex 1 or multiplex 2 or multiplex 3. "SNP_ID" is the SNP name, where the number following the underscore is equal to the chromosome position of the SNP. "Length" is the number of nucleotides in the amplified sequence from the PCR reaction. "Dir." is the direction of the extension primer; R is reverse (extended base is the complementary base of the allele given in this table), and F is forward (extended base is the same base as the allele given in this table). All masses are given in Da.

					Unextended extension primer			Extended extension primer			
								Alternate allele		Reference allele	
MP	SNP_ID	Reverse PCR primer	Forward PCR primer	Length	Sequence	Mass	Dir.	Allele	Mass	Allele	Mass
1	6_139899319	ACGTTGGATGTTAACCACTGAGCCACAACG	ACGTTGGATGATTCCCAAGGAAACTTCGGC	97	CCACAACGGGAACTC	4531.0	R	G	4778.2	A	4858.1
1	6_143221994	ACGTTGGATGAGATACTGCGCCACACAAAC	ACGTTGGATGCATCAGGTACTGATCCAAGG	100	CCCCTGGAGGCCAAG	4563.0	R	G	4810.2	A	4890.1
1	6_141940557	ACGTTGGATGAGAGACAAAGAGGAGACTCG	ACGTTGGATGCTGTTCCTGTCAATGGCAAG	106	TCAGCAGCTGCTGGT	4584.0	F	C	4831.2	T	4911.1
1	6_139001353	ACGTTGGATGTGCTATTTCTTGGGCCGCTC	ACGTTGGATGTTGCTGTGGCTCTGGTGTAG	116	gCCAGGCTAGGGGTTG	4978.2	F	A	5249.4	G	5265.4
1	6_141025042	ACGTTGGATGCCATAATGTGTCCAGAGGTC	ACGTTGGATGTGCCTTCCTTTGTCTCCTTC	95	GGGACAAGCTGGCAAG	4980.3	F	C	5227.4	T	5307.4
1	6_141590341	ACGTTGGATGTCAAAGACTTCCAGTGAAGG	ACGTTGGATGTTTCCTCTATTAGGGAGATG	113	gGCACTCCTACCTTTCC	5057.3	R	T	5328.5	C	5344.5
1	6_143259990	ACGTTGGATGGGATCTTCGAAGATGACTGC	ACGTTGGATGAGGCAGGGACTCACCCCAA	98	ccTGACTGCAGCCAGTA	5155.4	R	G	5402.5	A	5482.5
1	6_141438009	ACGTTGGATGACAGACCAGGTTTTGAATCC	ACGTTGGATGGGTCAAGGCTCAGACAGATA	103	TACTAGCTGCATGGTCT	5176.4	F	C	5423.6	T	5503.5
1	6_143591366	ACGTTGGATGACAGGTTGATCTCTTGCCTC	ACGTTGGATGGGCCAAATTTGACCCATAGG	92	cCTTCAAATCTGGGGGT	5201.4	R	G	5448.6	A	5528.5
1	6_139929703	ACGTTGGATGCCAGCACTCTCCTGCTGGG	ACGTTGGATGGCCCAGGCGGGTAGTTCTT	106	cCTTCTTTGCCGATTTCC	5367.5	R	G	5614.7	DEL (A)	5638.7
1	6_140009596	ACGTTGGATGATAAGTAGGTTAAAAGTAAAA	ACGTTGGATGTTTCTTTGCTTGAACTATAT	120	AAAAGATAGCTGTCGTGG	5587.7	F	C	5834.8	T	5914.8
1	6_140078143	ACGTTGGATGATGGATGTGGCTTGGATCTG	ACGTTGGATGGTGGCATATGAAGTTCCCAG	116	ctGCTTCAGCTCCAATTCA	5698.7	F	A	5969.9	G	5985.9
1	6_143147314	ACGTTGGATGCCTGTCTCTGTCCATCAATG	ACGTTGGATGATTCCCTCTGATAAGAGGGC	101	TCTTTTTCCCCAGAGTGAG	5769.8	R	G	6016.9	T	6041.0
1	6_138376833	ACGTTGGATGAGTCTCACCTTGTCAGTCAG	ACGTTGGATGGTTGCCATGGCAACACTGAC	102	aAGTCAGGCTGCTGTCACA	5812.8	R	C	6100.0	A	6139.9
1	6_138722540	ACGTTGGATGACCTGAGAACCTCTACATGC	ACGTTGGATGAGAATGTGTTTGGGATAAGG	116	agggACATGCCATGGGTGC	5893.8	R	T	6165.0	A	6220.9
1	6_142957033	ACGTTGGATGTTTTGAGGATCGTTCCCCTG	ACGTTGGATGGGTCCTCCAGGCATAAACTA	102	cccctCCCTGCTGTCATCAT	5939.9	R	G	6187.0	A	6267.0
1	6_141791377	ACGTTGGATGACACTGGGCACTAGACATAG	ACGTTGGATGCATCAAAGAGAGGGTTCCAG	115	tTAAGGAGCTGTCCCTTTCT	6074.0	R	G	6321.1	A	6401.1
1	6_138924570	ACGTTGGATGCTTTCCTCCACTCATGATGC	ACGTTGGATGACTAGTCTGGGACCTGGAAG	100	agGTGGGGACTATAGGACAT	6246.1	R	G	6493.3	C	6533.3
1	6_143109193	ACGTTGGATGATCCTAAGTATCTGGTGGGC	ACGTTGGATGCTTGTGATAGGAAGAGGAGC	119	cccCCTTAGGGATCTGGTTCT	6364.1	R	G	6611.3	T	6635.3
1	6_141715343	ACGTTGGATGTGCATCTCTTGTGAGATCAG	ACGTTGGATGGGAATGTTCTTAGCTTATCTC	96	AGAATATACCTTTTCTGGAGA	6444.2	F	C	6691.4	T	6771.3
1	6_139937372	ACGTTGGATGTGTGCAGGAGAATCCTGATG	ACGTTGGATGTGAAAGATGTCTGCATTCC	100	GGCAGTGGGGATGTTAACAAG	6575.3	R	G	6822.5	C	6862.5

1	6_143185874	ACGTTGGATGTCACTTGTGCTACTGATGCC	ACGTTGGATGACTGCAGCAATTAGCAAGTC	103	CCAAATTCAAACATTCCAGATA	6655.4	R	G	6902.6	A	6982.5
1	6_142708528	ACGTTGGATGGCCTTGTAATTGGCTTCCTG	ACGTTGGATGGAAATAACCCTCCCCACATC	102	GGTAGATCAGCTTTCATTCTTG	6731.4	F	A	7002.6	T	7058.5
1	6_141753293	ACGTTGGATGCCATGTTCTCCTCTAAGCTC	ACGTTGGATGGTCAATGGCAAAAATCCCTG	100	cccagACGAATATGTTCACACAC	6961.6	F	C	7208.7	A	7232.8
1	6_141219878	ACGTTGGATGGAGAGAAGAAATGTGGAACC	ACGTTGGATGTTACTTTCACCTCCTCCACC	104	acGGGGATTTTGTTCTGCCCTTT	7027.6	R	G	7274.7	A	7354.7
1	6_141867210	ACGTTGGATGTGTTGTTCTAGCACATCACG	ACGTTGGATGGGCAAACAGAAGCGATTCAG	97	ACGAAATATTCATTTCTGAGAGG	7086.6	F	C	7333.8	T	7413.7
1	6_144150268	ACGTTGGATGCAATTCTGGGTCTTTTGTGG	ACGTTGGATGCCCTATCAAATTACCCATGAC	106	gggcTGTTTGGATTGTTTGTTCT	7113.6	R	T	7384.8	A	7440.7
1	6_138809484	ACGTTGGATGATCCCTGTGTCTTCAAGCTG	ACGTTGGATGCTGCAACGGAATGCATTCTC	101	tttcCAAGCTGCCATTTTATTACA	7252.7	F	A	7523.9	T	7579.8
1	6_138606050	ACGTTGGATGGATTTTAGACTTCTGGCGTT	ACGTTGGATGCTCCACGTGAGCTTCCTATA	120	atttcGACTTCTGGCGTTCAAAAC	7302.8	R	G	7549.9	A	7629.9
1	6_138566115	ACGTTGGATGTCCACTCACCTGATGCAAAC	ACGTTGGATGCAATGTCTGGAGGCAGTTTG	90	cccccATGCAAACAGCATCCCCCAA	7469.9	R	G	7717.1	A	7797.0
1	6_144269094	ACGTTGGATGCCTAAATTGGCAAAAACTATC	ACGTTGGATGGCACATTTCATAGTGTCTCTG	87	ccATTGGCAAAAACTATCTCTATGT	7600.0	F	C	7847.2	G	7887.2
1	6_140867872	ACGTTGGATGTCTGATTAGCACCCATGAGG	ACGTTGGATGCCACAGCTCATGATAACACC	96	gtcaCTGGCCTCGCTCAGTGGGTTA	7665.0	F	C	7912.1	A	7936.2
1	6_141905193	ACGTTGGATGAGGGTTCAAGTCATCCTGTC	ACGTTGGATGCTGGGATGAGGTTAGTTTTC	99	gttaAAGTCATCCTGTCACGTGGTG	7688.0	R	T	7959.2	C	7975.2
1	6_139487878	ACGTTGGATGGACCTTCACAGAAATGCTAAG	ACGTTGGATGACTGATGTCATAGGAATTG	120	atGTTAATGCTAAATGGATGTTAAA	7743.1	F	A	8014.3	G	8030.3
1	6_142099288	ACGTTGGATGGGCACAGGCAACAAAAAGTC	ACGTTGGATGCCATAGCTGAAACAAGTCTG	108	ccctcCAACAAAAAGTCATCAAACTC	7821.1	R	T	8092.3	C	8108.3
1	6_143552748	ACGTTGGATGCTGACAGTACTGATGCTGTG	ACGTTGGATGGTAATCTTAGTCCTGTGGGC	101	ccctCTGCGTAGCTACTACTTTCATT	7823.1	R	G	8070.3	A	8150.2
1	6_141401113	ACGTTGGATGGAGAGTAACCATCTTAAGCC	ACGTTGGATGGCTCTTTCAAGGATCTTGCC	104	cccgAACCATCTTAAGCCTTGAGATTA	8203.4	F	C	8450.5	T	8530.4
1	6_141552175	ACGTTGGATGCACCAGTCAGAATGGCCATC	ACGTTGGATGTGTAGGAGGGTTCCCTCTTC	103	atcagTTAACAAGTCCACAAATAACAA	8237.4	F	C	8484.6	A	8508.6
1	6_139701681	ACGTTGGATGGCATATCTAGTGTTCAAGCA	ACGTTGGATGATGTCAGGGTTGTAGGGTTG	120	atgagTAAACATTACACTAAACGTACA	8268.4	R	C	8555.6	A	8595.5
1	6_141363410	ACGTTGGATGCCTTGAGATTTCTGAAGAC	ACGTTGGATGGTTTTAATGTTCAGAGTCG	104	agaGTACTTTATGGGTATAAAATACTTT	8640.7	R	G	8887.8	A	8967.8
2	6_140342933	ACGTTGGATGGTGGCTACAGCTCCAATTAG	ACGTTGGATGTTTTGCCTTTTCTAGGGCCG	95	CCCCTAGCCTGGGAA	4538.0	F	C	4785.1	G	4825.2
2	6_142332173	ACGTTGGATGAGAAGTCAAGGGAAGAAGGG	ACGTTGGATGCAATGGTCAGTTCTCAGTCC	112	GTGCTGTGATGCAGA	4648.0	F	C	4895.2	T	4975.1
2	6_142816860	ACGTTGGATGTTCCCCGCCCCCAATACCC	ACGTTGGATGTGTGCTGAGAAAGGCAAGTG	103	CCCAATACCCTCTCGG	4762.1	R	G	5009.3	A	5089.2
2	6_143476143	ACGTTGGATGAGGAAACACACTCACAGTGG	ACGTTGGATGGGGACTGACATCATCCAGAC	105	aGGGGTCCACCAGGCA	4916.2	R	T	5187.4	C	5203.4
2	6_140815805	ACGTTGGATGTCCATGAGGATGAGGGTTTG	ACGTTGGATGATCCAAGCTGCATCTGCAAC	117	tGCCTCGATTGGTGGG	4944.2	F	G	5231.4	T	5271.3
2	6_142852976	ACGTTGGATGCATTGAAGGTGCTGAACCTG	ACGTTGGATGGAATTCTGGACCCACTACTC	99	TGCTGAACCTGCACATT	5145.4	R	G	5392.5	A	5472.5
2	6_138885828	ACGTTGGATGTCGGATCTGGTGTTGCTGTG	ACGTTGGATGATGGAGGTTCCCAGACTAGG	99	caTGGCTGTGGTGTAGG	5297.4	F	C	5544.6	T	5624.5
2	6_138339026	ACGTTGGATGCACTCTTATCAACACTGAGG	ACGTTGGATGAGATTAGTTTCTGCTGTGCC	102	TTGTGGGTTTTGGGAGT	5318.5	F	C	5565.6	T	5645.5
2	6_139975133	ACGTTGGATGCCATAACTCATGGCAACACC	ACGTTGGATGGACTAGGAACCATGAGGTTG	100	GCAACACCAGATCCTTAA	5436.6	R	G	5683.8	A	5763.7
2	6_142632617	ACGTTGGATGAGTTGGGCTGTTTGGCTTAG	ACGTTGGATGCCTCTGATCTGTGTTCACTC	104	gggTGGAGGAGGCCCTCT	5596.6	F	C	5843.8	T	5923.7
2	6_139664663	ACGTTGGATGTATGTATGTTGGTGGGTGGG	ACGTTGGATGTGATTTTCCCTTGCCCCATC	105	cttTGGGTGGGTACTGAG	5601.6	F	A	5872.8	G	5888.8

2	6_140762428	ACGTTGGATGGCTTGTTCATGTGTCTGTGC	ACGTTGGATGGGGCTCTCTGAAACCTTTC	91	GTGCATTAACCCCTTTTCG	5729.7	F	C	5976.9	T	6056.8
2	6_142518571	ACGTTGGATGCTGCATTTGCTCTAGTGCTC	ACGTTGGATGCTCTGGTGGATGGAATTTTG	106	TGCTCTTTTGAAGACAATG	5817.8	F	A	6089.0	G	6105.0
2	6_144402150	ACGTTGGATGATCAAAGCTCACATTCCCCC	ACGTTGGATGCTTTTCCAAGAAGCAGCAGG	102	ataCCCCCATTAGTGTACCA	6020.9	F	C	6268.1	T	6348.0
2	6_139779269	ACGTTGGATGAGAATGCTCTGTAGAGAGCG	ACGTTGGATGACTTATTGAGGAATGGCCAC	111	GCGTTAATCACCTCTAAATC	6036.0	F	A	6307.2	G	6323.2
2	6_140722873	ACGTTGGATGCAGCATTACTGTAGGAGGAG	ACGTTGGATGCTCCTTCCTCTGTAGAATG	100	ATGTAAATGACACAGGACAA	6167.1	F	A	6438.3	G	6454.3
2	6_140950224	ACGTTGGATGATACCAAAGCACTCCAAGCC	ACGTTGGATGAGAGCTGGAAAGCACATACC	99	ggggAGACAGTGGGACTACT	6247.1	F	A	6518.3	G	6534.3
2	6_140456783	ACGTTGGATGAAGTCTACCTGTGCTCCACC	ACGTTGGATGTAGAAAGTGCAGAGAGAGCC	106	CCCCAAAACATGTAATTTGAA	6382.2	R	G	6629.4	A	6709.3
2	6_143719226	ACGTTGGATGGGGAGAGAAATTAGGAGAGC	ACGTTGGATGTAAAGGGCCAAGCTTCAGAG	99	GCAAACTATTAATGAGAGCAG	6487.3	F	C	6734.4	A	6758.5
2	6_144547694	ACGTTGGATGGAAATACTGACTCCTGATGC	ACGTTGGATGGAAGCACTTTCAGTGATCTC	95	cccaTGATGCTCATCTCCACTA	6590.3	R	C	6877.5	A	6917.4
2	6_143032834	ACGTTGGATGTGTGTCACGAGTACTTAGTC	ACGTTGGATGCACTGAGTGCATCTGCATAC	105	CGAGTACTTAGTCTTTATCCTT	6666.3	F	A	6937.6	G	6953.6
2	6_140154203	ACGTTGGATGGGATACTGAGATAACTGATGG	ACGTTGGATGGCGAATTTGTCGTTGATACAC	96	AACTGATGGAATCTGAATAATA	6790.5	R	T	7061.7	A	7117.6
2	6_142484363	ACGTTGGATGTTCTGGCCACTGAGACAAAG	ACGTTGGATGGCAGAGGCAGGGAGAAAAAG	99	tggcGGGGTTTCCAGAAAAGAT	6839.5	R	G	7086.6	A	7166.6
2	6_144786659	ACGTTGGATGTAGGGCCAGAGCTTCATTTC	ACGTTGGATGCTCACACTATGAAAGCATGG	115	ggCCACCCTCATGATTATGTTAT	6989.6	R	T	7260.8	C	7276.8
2	6_140492893	ACGTTGGATGCTAGAGAAGACCCATCTCTG	ACGTTGGATGGAGCAAACAAAGGCACAGAC	103	ggccTGACTTACTGTGTGACTTT	7036.6	F	A	7307.8	G	7323.8
2	6_140607423	ACGTTGGATGCCCCACAAAAGCAAAGTCTG	ACGTTGGATGTGGAAAAAGTCCCACTCGTC	101	ccctcGGGAAGTGGGGGAAGAGT	7194.7	F	C	7441.9	T	7521.8
2	6_141978409	ACGTTGGATGTAACTTGCTGCACGTCACAC	ACGTTGGATGCAGTGTGGTAGCATGGAAAG	102	ccctCCAGCAAAGCAAGATCCCCT	7211.7	F	A	7482.9	G	7498.9
2	6_138530222	ACGTTGGATGTAGGAACCTCCAAATGCCAC	ACGTTGGATGCTTCATCACTGTTCAGAGAC	97	gaacCAGGTGCAGCCCTTAAAAAA	7363.8	R	T	7635.0	C	7651.0
2	6_141514143	ACGTTGGATGAATCAGCCTCCTGACTTCAC	ACGTTGGATGGCCTTTTTATGCTGTTCCAC	119	GATACCAACATAAAAACAGAAATA	7363.9	R	G	7611.0	A	7691.0
2	6_140116064	ACGTTGGATGTGCATTTGCTATCCAGTCCC	ACGTTGGATGCTAATGTCGCTAGTGTCCAG	110	tcTCATTATTGCAGAACTTTCTTAC	7556.9	R	G	7804.1	A	7884.0
2	6_143400255	ACGTTGGATGCCATGCAACGTGGTATAATG	ACGTTGGATGGTTCCATTACTTCACGCTG	117	ctcgTCCATGCATCCTGATTTAATG	7582.9	R	G	7830.1	A	7910.0
2	6_144472042	ACGTTGGATGCAGGAGTAAAGGCTGAGTTC	ACGTTGGATGTTTCTGAAGTTGTGGATCTG	86	ggggTCTTAGAAGAGCATCCTTTTA	7712.0	R	G	7959.2	T	7983.2
2	6_138300375	ACGTTGGATGCTAGTCACTTATGATGGAGC	ACGTTGGATGCTACTGTACAGCATGGTGAC	105	GCAAAAATAGAATGTGTACATGTAT	7737.1	F	A	8008.3	G	8024.3
2	6_140192635	ACGTTGGATGCCCCAAATAAAGTAGCTGGA	ACGTTGGATGGGCACTTGGATCCTGAAATC	112	tagaTGGAAAACAAAAGAGAAATCT	7764.1	R	C	8051.3	A	8091.2
2	6_139449445	ACGTTGGATGACGTCCATTGTAACTTCTCC	ACGTTGGATGCCCTTATCCAGACTTATCAAG	102	atggAACTTCTCCTTTTTCATTTCTA	7852.1	R	T	8123.3	A	8179.2
2	6_139565286	ACGTTGGATGCGAGTTCCCACAAATGACTG	ACGTTGGATGAAGTTTCCTGGAGAGGCAAG	99	ggaaTGCTGCCATGTCTATGTACCCT	7937.2	R	C	8224.4	A	8264.3
2	6_139159492	ACGTTGGATGCGCTCATTAATTTCCGCCTC	ACGTTGGATGAGGAAGATTCACCGGATTGG	87	ggagtCTCCCTTTCTCCAGAAAACCCT	8155.3	R	G	8402.5	A	8482.4
2	6_141144207	ACGTTGGATGGTTACACCTAGAATGGGCAC	ACGTTGGATGAGCTCTCGTTATGCACAGTC	108	ggtcCACCTAGAATGGGCACCTAATTG	8284.4	F	A	8555.6	G	8571.6
2	6_138848835	ACGTTGGATGGTGCGGCCCTAAAAAGACAA	ACGTTGGATGTTCGTTAACCACTGAGCCAC	105	gaGGCCCTAAAAAGACAAAATAAATAA	8335.5	F	A	8606.7	G	8622.7
2	6_141475994	ACGTTGGATGATCAGAGAGCGGTATGATCC	ACGTTGGATGCATTCTTCAAGATGTTCTTC	105	TCCAAAATTAGTTTGTTATATTTATATTC	8845.8	R	G	9093.0	A	9172.9
3	6_142994952	ACGTTGGATGTGAAGCCACAGTAAGAGTGC	ACGTTGGATGGATGAACCCACATCAGATGC	87	TGCCTGCTGCTGAAA	4568.0	R	G	4815.2	A	4895.1

3	6_140645493	ACGTTGGATGGGACGCTGTGCTTATAAATG	ACGTTGGATGGTCCAGGGCAGGTTATACAA	110	AACACAGGAGGGTGA	4675.1	R	G	4922.3	A	5002.2
3	6_139411637	ACGTTGGATGAAGCCATCTCTGTGACCTAC	ACGTTGGATGTGAGGTTGCAGGTTCAATCC	109	CCTACACAACAGCTCA	4779.1	R	G	5026.3	A	5106.2
3	6_139234216	ACGTTGGATGTCCCTAGACTGAGAACTTCC	ACGTTGGATGTATTTGTGGCACTTATCAG	109	tTTCCGTATGCCTTGG	4854.2	R	C	5141.4	A	5181.3
3	6_139739641	ACGTTGGATGTAAAATCGTTAAGCACAGGG	ACGTTGGATGAATATTTGCTACCCTTTCCC	82	tGGAAGCTGGCTGTGA	4977.2	R	C	5264.5	A	5304.3
3	6_142446166	ACGTTGGATGTTCACCAAAGGGAAGCAAGG	ACGTTGGATGTTAAGATGCCTGTACCCAGC	114	GCCCTTGAATGAAAAGC	5203.4	F	C	5450.6	T	5530.5
3	6_141062948	ACGTTGGATGCTGTTCCTGGGAAATGGAAG	ACGTTGGATGTTCTTGTCCCAGAGCTGTTC	97	gcGCTGCCTGTTGGTTG	5224.4	R	G	5471.6	A	5551.5
3	6_142556587	ACGTTGGATGACAAAGGTGGATGGCCATTC	ACGTTGGATGCTCTACCCACCAAGGACTTC	118	gGAGAGCTGAATGTGGT	5330.5	F	C	5577.7	G	5617.7
3	6_143889528	ACGTTGGATGGCTCCCAAAAGACAGAACAC	ACGTTGGATGGTTCTCTAGGAGAACAGTGG	100	CTCCTTTACCGTGATTTC	5391.5	F	C	5638.7	T	5718.6
3	6_139626802	ACGTTGGATGCCATTCTTAACACTGTGCCG	ACGTTGGATGATCTCAGAGTATCCCAGGTG	112	cGTTCCCCAGCAAAGCAC	5413.5	R	G	5660.7	T	5684.7
3	6_141829392	ACGTTGGATGACACCTATTAAGGCAGGGAC	ACGTTGGATGAGTCTCTGTTATGTGTCTGG	105	cGCAGGGACATTGTCTTT	5505.6	R	G	5752.8	A	5832.7
3	6_138490776	ACGTTGGATGTCCAACTAGGAACCATGAGG	ACGTTGGATGTGTGACTAACACCACAGCTC	114	ccctGGTTAAGGATCCGGC	5804.8	F	A	6076.0	G	6092.0
3	6_139525764	ACGTTGGATGTTAGCAGCACAGTGTCATGG	ACGTTGGATGGTAGTAGTACCAAGGTCTTC	117	CACAGTGTCATGGGTAGGA	5892.8	R	G	6140.0	A	6219.9
3	6_141257795	ACGTTGGATGAGAAAGGTTTGTGGTGCTGG	ACGTTGGATGTGCTGCAGCCAGAGCTAAG	99	ggttCTGGCAGGCAGGAAA	5917.9	F	C	6165.0	T	6245.0
3	6_143437748	ACGTTGGATGCTCGCTGCATAGACTCCACT	ACGTTGGATGGAATATTCCATGCATTGCGG	83	AGACTCCACTGCATATAATC	6045.0	R	C	6332.2	A	6372.1
3	6_142141674	ACGTTGGATGGGTCTTCTGGGAGAGAAGTC	ACGTTGGATGTCTGTTGGCAGGTCGTTCAG	111	tgtAGAGAAGTCCCAGGAAA	6199.1	R	C	6486.3	A	6526.2
3	6_144737962	ACGTTGGATGGACGAGTTCATCTGAGTAGG	ACGTTGGATGGCCTAAGTATCATATTCTC	96	ccccACTCCATGCACAAAGTA	6304.1	F	G	6591.3	T	6631.2
3	6_143795121	ACGTTGGATGAACGGCAGTTTCCTCAACTC	ACGTTGGATGGGAAGAGGCTGTCTAATTGG	102	gactGAAACATGGCTTCCGCT	6406.2	R	G	6653.4	A	6733.3
3	6_140230629	ACGTTGGATGGTTAGAAAATATTGAGTATGG	ACGTTGGATGGATCAATAGATACATACAAC	106	cGTATGGTTTATTCTGCTGTA	6433.2	R	G	6680.4	A	6760.3
3	6_140531399	ACGTTGGATGCCTTGGGTCAAAATAAATTC	ACGTTGGATGGAGTTTTGTTTTAAAACACC	112	ctgCCTAATTACTTCTCCCTCA	6556.3	F	C	6803.5	A	6827.5
3	6_140569380	ACGTTGGATGCTGACACATCTTAGAGGCTG	ACGTTGGATGTGGCCCTTCTTTGAAGCTTG	104	tgtgCCTCCCTTGGGCTGTAAT	6708.3	F	A	6979.6	G	6995.6
3	6_141296104	ACGTTGGATGGGTGCTTAGGGCTAGAAATG	ACGTTGGATGGGCACAGAAAGTTTGTAAACC	99	GAATGGAAAGTGTATTCTGAAG	6862.5	R	G	7109.7	C	7149.7
3	6_143680244	ACGTTGGATGTTGGTGTCAGAGAGATGGTG	ACGTTGGATGCATGTGGTCAGCAACATACG	107	gggaGTGCCCTGGAGAAAGTGA	6889.5	F	G	7176.7	T	7216.6
3	6_143757239	ACGTTGGATGCTTTCGCACCTGTGTTTCTG	ACGTTGGATGCACGATGGTAACTCCTTGAC	113	cccccGCATATGAAAAATGGGCT	7017.6	R	G	7264.8	T	7288.8
3	6_138456283	ACGTTGGATGTCTTGTGTGCTATGCTAACC	ACGTTGGATGCAGAAAAAAAGCAAGTGCC	105	ATGCTAACCTTAAAAAAAATGTA	7047.6	R	T	7318.9	A	7374.7
3	6_144510240	ACGTTGGATGCAGAAGTACATCACAGAGGG	ACGTTGGATGCCATCCAATGGTTTCCTGTC	98	ggggAGAGGGCTGTGTCAGTGAT	7240.7	F	G	7527.9	T	7567.8
3	6_144363777	ACGTTGGATGTAGACCCCTAGACAGACTTG	ACGTTGGATGGGGATGGAGTTAACATTTC	99	aTGCCTCAATCATATGAAAAAAAA	7345.8	R	G	7593.0	T	7617.0
3	6_143514133	ACGTTGGATGTTTAGGAGTTCCCTTGTGGC	ACGTTGGATGTGCCATAGCAGTGAACTGAG	98	cctCAGGTTAAAGGATCTGGTGTT	7398.8	F	G	7686.0	T	7725.9
3	6_139350912	ACGTTGGATGTCAGATCTGGTGTTGCTGTG	ACGTTGGATGCAGCATATGGAGGTTCCCAG	105	atggGGTGTAGGCTGGCAGCTGTA	7504.9	F	G	7792.1	T	7832.0
3	6_141673999	ACGTTGGATGCATGCCATAGAGGCAGAATC	ACGTTGGATGGTCTGATGACATGTAGAACC	111	gaaaaTGTGTATAGTCCTTAGGAAA	7753.1	F	C	8000.3	T	8080.2
3	6_144622895	ACGTTGGATGGTGACACATCCTTAATACC	ACGTTGGATGGTCTTGTGGGAGTTTCTAAC	119	ACACATCCTTAATACCTTTTTAGTGA	7879.2	R	T	8150.4	C	8166.4

3	6_139077739	ACGTTGGATGTAACTGCATTGACCTACGCC	ACGTTGGATGTTGCTCCATCTCAATCAGCC	105	cccaCCAGCTGTGGGGCTGTGTTCGT	7970.2	R	T	8241.4	C	8257.4
3	6_144004564	ACGTTGGATGGCTGCTAGTCTTCAGATCAT	ACGTTGGATGAAGACGCTGAGCTCACTTCC	114	tgccCTTTAAATAGTTGTGGTATTGG	8022.2	R	C	8309.4	A	8349.3
3	6_143929265	ACGTTGGATGCCTAAGGTTGTCACTGTGAG	ACGTTGGATGAACACTGTGGTGAACACTGG	100	ggaaTAGACACTGACATGTCAGATTGG	8372.5	R	C	8659.7	A	8699.6
3	6_142408065	ACGTTGGATGGCTATATAAAACACTGCTAGG	ACGTTGGATGTCACCCCAACAAAGGATTTC	111	TTTTATAATATGCATATGATAAATCTTC	8550.6	R	T	8821.8	A	8877.7

Appendix IV: Results from association analysis on SNPs in SSC6a

Appendix IV: Results from association analysis of SNPs in SSC6a. "SNP_ID" is the SNP name; SNPs previously genotyped are named by the PorcineSNP60 BeadChip SNP name, and the SNPs genotyped in this study are named as ["Chromosome number"_"Chromosome position"]. "Position" is the chromosomal position, "MAF" is the minor allele frequency of the SNP, and P-value is the significance level obtained by association analysis.

SNP_ID	Position	MAF	P-value
ALGA0034407	6752216	0.490	0.040
6_6801001	6801001	0.208	0.154
ASGA0027406	6813759	0.490	0.040
ASGA0027409	6921532	0.208	0.154
ALGA0034436	7000417	0.361	0.294
ALGA0034441	7024721	0.214	0.185
H3GA0017469	7056662	0.214	0.185
6_7059895	7059895	0.294	0.420
6_7061387	7061387	0.214	0.185
6_7062060	7062060	0.492	0.056
6_7068363	7068363	0.282	0.473
6_7068369	7068369	0.282	0.473
6_7068912	7068912	0.342	0.362
ALGA0034455	7087271	0.435	0.160
M1GA0008273	7095483	0.214	0.185
H3GA0017455	7129139	0.214	0.185
6_7136808	7136808	0.214	0.185
6_7147835	7147835	0.490	0.068
ASGA0027461	7167577	0.424	0.248
6_7169940	7169940	0.483	0.078
ALGA0117805	7252505	0.433	0.071
M1GA0008302	7254956	0.176	0.507
ALGA0034463	7283709	0.433	0.071
ASGA0102684	7303586	0.240	0.364
ALGA0034467	7305244	0.166	0.700
6_7306793	7306793	0.231	0.357
M1GA0008329	7363656	0.422	0.170
ALGA0034498	7390710	0.328	0.452
M1GA0008318	7416050	0.246	0.323
ALGA0034480	7476614	0.221	0.792
6_7489504	7489504	0.101	0.474
M1GA0008309	7506712	0.389	0.469
6_7517225	7517225	0.288	0.760
6_7524638	7524638	0.147	0.916
ASGA0027502	7528631	0.389	0.469
ASGA0083742	7584584	0.279	0.677
ASGA0094024	7600738	0.311	0.990
6_7635427	7635427	0.466	0.206
ASGA0095713	7643426	0.183	0.362
ASGA0090879	7647082	0.351	0.528
MARC0007721	7652040	0.176	0.384
M1GA0026539	7691238	0.408	0.702
ASGA0094340	7715574	0.393	0.401
MARC0006941	7717694	0.424	0.132
MARC0114645	7718925	0.183	0.366
6_7769782	7769782	0.357	0.459
ASGA0027522	7772810	0.466	0.977
ALGA0113069	7872494	0.242	0.040
MARC0056813	7881468	0.242	0.040
H3GA0017523	7892295	0.242	0.040
6_7938504	7938504	0.485	0.069
ALGA0034502	7943484	0.307	0.215
H3GA0017528	8019390	0.466	0.627
6_8054202	8054202	0.466	0.627
H3GA0017531	8058591	0.466	0.627
H3GA0017532	8106730	0.242	0.915
6_8130180	8130180	0.319	0.084
ALGA0119278	8198500	0.088	0.470
ALGA0124441	8203297	0.250	0.016
6_8295671	8295671	0.250	0.016
ASGA0092787	8303476	0.319	0.084
6_8328490	8328490	0.288	0.839
ALGA0123364	8328777	0.462	0.056
ALGA0118757	8338626	0.326	0.478
6_8339436	8339436	0.250	0.016
ALGA0116973	8340755	0.288	0.839
6_8351513	8351513	0.338	0.065
MARC0059052	8359798	0.288	0.839
ALGA0102858	8361799	0.338	0.065
6_8372653	8372653	0.250	0.016
6_8383536	8383536	0.250	0.016
6_8399066	8399066	0.288	0.839
ALGA0104695	8402621	0.250	0.016
M1GA0026870	8453407	0.088	0.470
6_8465599	8465599	0.374	0.095
ASGA0090877	8466863	0.311	0.619
6_8476636	8476636	0.317	0.043
6_8487631	8487631	0.317	0.043
ALGA0113531	8497891	0.374	0.095
ASGA0106245	8500002	0.145	0.495
ASGA0100146	8519321	0.143	0.388
6_8531573	8531573	0.462	0.172

SNP_ID	Position	MAF	P-value
MARC0025924	8542652	0.143	0.388
6_8596789	8596789	0.143	0.388
6_8677642	8677642	0.143	0.388
ALGA0110070	8679743	0.143	0.388
ALGA0122972	8687181	0.143	0.388
6_8712157	8712157	0.202	0.674
ALGA0120698	8719089	0.202	0.674
ALGA0107172	8727015	0.462	0.145
6_8734413	8734413	0.359	0.577
6_8761799	8761799	0.359	0.577
ASGA0091175	8822541	0.435	0.540
6_8842067	8842067	0.197	0.410
6_8889138	8889138	0.303	0.360
ALGA0102512	8900829	0.303	0.360
6_8908946	8908946	0.363	0.608
ALGA0114575	8917933	0.363	0.608
6_8955625	8955625	0.334	0.965
6_9127363	9127363	0.370	0.389
ALGA0120964	9149467	0.130	0.876
6_9164561	9164561	0.298	0.563
6_9187066	9187066	0.298	0.563
ALGA0119039	9198558	0.324	0.557
6_9208969	9208969	0.290	0.290
ASGA0106220	9217466	0.330	0.543
H3GA0055185	9237989	0.330	0.543
6_9241597	9241597	0.330	0.543
6_9252732	9252732	0.370	0.323
ALGA0123921	9253224	0.366	0.905
ASGA0098582	9260915	0.330	0.543
6_9270142	9270142	0.265	0.309
ALGA0112212	9271639	0.265	0.309
MARC0019338	9277653	0.265	0.309
MARC0115370	9290031	0.265	0.309
6_9301999	9301999	0.370	0.389
ASGA0084059	9393115	0.408	0.246
6_9407409	9407409	0.460	0.462
ASGA0089652	9448931	0.435	0.494
6_9469229	9469229	0.311	0.858
6_9480658	9480658	0.311	0.858
ALGA0109170	9490990	0.057	0.135
6_9511328	9511328	0.197	0.857
6_9636486	9636486	0.439	0.704
MARC0045581	9663840	0.372	0.933
6_9679825	9679825	0.363	0.584
H3GA0017537	9695895	0.197	0.857
6_9744464	9744464	0.439	0.704
6_9794687	9794687	0.311	0.305
6_9885915	9885915	0.254	0.698
ASGA0102946	9891486	0.227	0.878
MARC0094871	9902739	0.227	0.878
MARC0095910	9908019	0.227	0.878
6_9918959	9918959	0.227	0.878
ASGA0097772	9929286	0.256	0.630
ASGA0101598	9929460	0.256	0.630
ASGA0093244	9952971	0.303	0.586
ASGA0084984	9957369	0.303	0.586
6_9976418	9976418	0.303	0.586
MARC0069210	10024253	0.032	0.154

Appendix V: Results from association analysis on SNPs in SSC6b

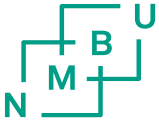
Appendix V: Results from association analysis of SNPs in SSC6b. "SNP_ID" is the SNP name; SNPs previously genotyped are named by the PorcineSNP60 BeadChip SNP name, and the SNPs genotyped in this study are named as ["Chromosome number"_"Chromosome position"]. "Position" is the chromosomal position, "MAF" is the minor allele frequency of the SNP, and P-value is the significance level obtained by association analysis.

SNP_ID	Position	MAF	P-value
ALGA0037191	138334401	0.416	0.676
6_138339026	138339026	0.431	0.651
6_138376833	138376833	0.416	0.676
ASGA0029722	138377709	0.424	0.853
H3GA0018970	138401742	0.445	0.255
6_138456283	138456283	0.420	0.799
6_138566115	138566115	0.363	0.977
ALGA0101321	138579392	0.431	0.920
6_138606050	138606050	0.155	0.218
ALGA0108608	138646379	0.160	0.147
CASI0002165	138810072	0.059	0.531
ALGA0102709	138817138	0.311	0.187
ASGA0091104	138822417	0.218	0.599
MARC0037985	138852864	0.132	0.030
INRA0022567	138921625	0.059	0.531
6_138924570	138924570	0.363	0.033
MARC0090227	139045500	0.225	0.654
MARC0101536	139068649	0.067	0.894
6_139077739	139077739	0.412	0.045
DRGA0006969	139082550	0.059	0.531
ALGA0037236	139099898	0.284	0.456
DRGA0006970	139112031	0.059	0.531
6_139159492	139159492	0.233	0.306
ASGA0029775	139200169	0.120	0.802
ASGA0029778	139221327	0.120	0.802
6_139234216	139234216	0.397	0.275
MARC0062260	139259757	0.057	0.780
ALGA0037242	139288814	0.092	0.937
ALGA0037249	139303054	0.084	0.632
MARC0056678	139322780	0.191	0.652
ALGA0037253	139350533	0.061	0.905
ASGA0106260	139410790	0.477	0.919
ASGA0083637	139413426	0.477	0.919
6_139487878	139487878	0.477	0.919
6_139525764	139525764	0.462	0.653
H3GA0055965	139535744	0.061	0.481
ALGA0117611	139538709	0.462	0.653
6_139565286	139565286	0.462	0.653
6_139626802	139626802	0.332	0.014
6_139664663	139664663	0.170	0.012
6_139701681	139701681	0.426	0.611
6_139739641	139739641	0.490	0.897
ASGA0098750	139760470	0.479	0.702
SIRI0000806	139779269	0.487	0.897
ALGA0037261	139875001	0.282	0.053
ALGA0037265	139899319	0.252	0.053
H3GA0019026	139912090	0.282	0.053
ASGA0029789	139932906	0.233	0.057
6_139937372	139937372	0.282	0.053
DIAS0004575	139944638	0.282	0.053
MARC0005246	139965340	0.240	0.102
ALGA0037272	140009596	0.240	0.102
ALGA0037277	140025895	0.240	0.102
ASGA0090175	140080043	0.229	0.084
6_140116064	140116064	0.233	0.057
6_140154203	140154203	0.282	0.053
6_140192635	140192635	0.282	0.053
ALGA0110546	140238070	0.233	0.057
ASGA0101351	140372952	0.233	0.057
ALGA0115609	140397044	0.273	0.031
ALGA0114316	140405637	0.487	0.657
ALGA0114803	140405801	0.487	0.657
ASGA0094926	140440928	0.233	0.057
6_140456783	140456783	0.378	0.661
ALGA0118888	140462905	0.439	0.637
ASGA0091247	140469852	0.441	0.637
ASGA0104579	140475536	0.206	0.014
ASGA0091723	140480675	0.206	0.014
ALGA0114933	140483422	0.206	0.014
MARC0025692	140511844	0.340	0.017
6_140531399	140531399	0.179	0.158
ALGA0037285	140538736	0.393	0.056
6_140569380	140569380	0.307	0.571
ALGA0037287	140576155	0.160	0.119
ALGA0037290	140591679	0.176	0.225
ASGA0029808	140603308	0.334	0.026
6_140607423	140607423	0.387	0.080
M1GA0008935	140632303	0.319	0.080
6_140645493	140645493	0.397	0.696
ALGA0037302	140649286	0.397	0.696
MARC0032403	140692911	0.204	0.169
H3GA0019045	140704653	0.204	0.169

SNP_ID	Position	MAF	P-value
6_140722873	140722873	0.143	0.013
ASGA0029822	140724273	0.204	0.169
ALGA0037308	140754146	0.183	0.298
ALGA0113744	140776171	0.494	0.923
H3GA0019050	140804651	0.435	0.407
ALGA0037315	140829602	0.183	0.298
6_140950224	140950224	0.139	0.427
ALGA0037327	141025042	0.414	0.576
6_141062948	141062948	0.374	0.189
ALGA0037340	141094271	0.111	0.473
MARC0092172	141126737	0.195	0.325
ALGA0037343	141143654	0.223	0.416
6_141144207	141144207	0.240	0.405
ASGA0029872	141189342	0.202	0.296
H3GA0019067	141213312	0.233	0.948
MARC0057158	141219751	0.158	0.151
6_141219878	141219878	0.355	0.280
ALGA0037350	141246850	0.191	0.555
6_141257795	141257795	0.424	0.631
ALGA0117399	141280500	0.359	0.869
ALGA0037357	141283453	0.359	0.869
6_141296104	141296104	0.481	0.535
ASGA0029874	141297531	0.221	0.512
6_141363410	141363410	0.298	0.927
ASGA0099800	141427571	0.498	0.825
ASGA0082303	141438009	0.284	0.717
ALGA0108422	141442919	0.439	0.496
H3GA0056289	141465780	0.288	0.619
6_141475994	141475994	0.454	0.517
ALGA0111542	141482906	0.414	0.315
6_141514143	141514143	0.397	0.558
ALGA0116958	141530340	0.397	0.558
6_141590341	141590341	0.397	0.558
ASGA0087611	141612059	0.090	0.745
ASGA0087447	141615571	0.300	0.571
6_141673999	141673999	0.162	0.656
ASGA0096677	141680099	0.240	0.348
ASGA0094369	141692873	0.162	0.656
ASGA0090978	141715343	0.338	0.329
ASGA0101341	141723382	0.410	0.166
6_141753293	141753293	0.479	0.948
H3GA0054001	141765398	0.134	0.417
6_141791377	141791377	0.134	0.417
6_141829392	141829392	0.477	0.669
ASGA0105042	141841405	0.162	0.656
6_141867210	141867210	0.162	0.656
ALGA0104495	141892248	0.401	0.581
6_141905193	141905193	0.176	0.357
ASGA0098617	141913167	0.387	0.333
H3GA0056068	141914378	0.387	0.333
MARC0041415	141940557	0.162	0.656
6_141978409	141978409	0.134	0.417
MARC0064117	141988774	0.162	0.656
MARC0066661	142036921	0.460	0.895
6_142099288	142099288	0.134	0.417
ASGA0029876	142137263	0.288	0.239
6_142141674	142141674	0.134	0.417
H3GA0019078	142311498	0.363	0.434
ASGA0029880	142332173	0.134	0.417
ALGA0037369	142342553	0.090	0.745
H3GA0019083	142369946	0.431	0.501
ASGA0029890	142393332	0.277	0.210
6_142408065	142408065	0.221	0.667
ASGA0029893	142431821	0.355	0.344
6_142446166	142446166	0.237	0.548
ALGA0037389	142452858	0.487	0.069
6_142484363	142484363	0.218	0.476
ALGA0037390	142492867	0.477	0.571
ASGA0029901	142518571	0.305	0.994
ALGA0037402	142534365	0.248	0.315
6_142556587	142556587	0.456	0.415
ALGA0037404	142562712	0.326	0.893
CASI0008278	142578125	0.460	0.521
6_142632617	142632617	0.464	0.466
H3GA0056596	142653834	0.378	0.939
MARC0026081	142661463	0.227	0.497
ASGA0094758	142685412	0.466	0.318
ALGA0112024	142719548	0.158	0.308
H3GA0056667	142725268	0.216	0.476
ALGA0037412	142803333	0.447	0.477
ASGA0084628	142814186	0.265	0.864
6_142816860	142816860	0.216	0.476
ALGA0037416	142832114	0.468	0.627
6_142852976	142852976	0.342	0.804
ASGA0029922	142925060	0.265	0.378
6_142957033	142957033	0.460	0.373
ASGA0029925	142957402	0.460	0.373
MARC0114321	142973530	0.460	0.373
ALGA0037432	142989178	0.309	0.437
H3GA0019109	143021550	0.053	0.210
6_143032834	143032834	0.439	0.617
MARC0041406	143056683	0.191	0.549

SNP_ID	Position	MAF	P-value
ALGA0037450	143081104	0.315	0.603
6_143109193	143109193	0.279	0.632
ALGA0037458	143112740	0.321	0.776
H3GA0019137	143137565	0.185	0.538
6_143147314	143147314	0.185	0.538
6_143185874	143185874	0.481	0.554
H3GA0019143	143194905	0.477	0.437
ALGA0037473	143219220	0.412	0.980
6_143221994	143221994	0.412	0.980
ALGA0037484	143241088	0.277	0.497
6_143259990	143259990	0.277	0.497
ALGA0037491	143263519	0.330	0.323
6_143476143	143476143	0.271	0.480
6_143552748	143552748	0.218	0.534
ASGA0102302	143591366	0.271	0.480
6_143680244	143680244	0.273	0.291
ASGA0089837	143684647	0.319	0.695
ASGA0092589	143709272	0.414	0.605
ALGA0112970	143711252	0.311	0.729
ASGA0089288	143719226	0.273	0.291
ASGA0096559	143719797	0.183	0.783
ALGA0116974	143740794	0.490	0.952
6_143757239	143757239	0.273	0.291
ASGA0105551	143767111	0.273	0.291
H3GA0055887	143770523	0.273	0.291
6_143795121	143795121	0.227	0.050
ASGA0095712	143814685	0.227	0.050
6_143889528	143889528	0.223	0.033

SNP_ID	Position	MAF	P-value
6_143929265	143929265	0.481	0.265
ALGA0037494	143932163	0.412	0.173
MARC0016562	143952619	0.347	0.029
ALGA0113787	143962682	0.408	0.645
ASGA0029993	143976085	0.372	0.227
6_144004564	144004564	0.317	0.100
ALGA0115902	144024419	0.431	0.810
ALGA0037499	144026033	0.431	0.810
INRA0022701	144047257	0.347	0.029
ASGA0029996	144060537	0.397	0.344
ASGA0029999	144089138	0.359	0.004
MARC0022671	144155251	0.095	0.531
ASGA0030003	144260593	0.223	0.033
6_144269094	144269094	0.317	0.100
ALGA0037514	144273518	0.433	0.986
ALGA0037508	144303289	0.210	0.044
6_144363777	144363777	0.498	0.245
6_144402150	144402150	0.223	0.033
6_144472042	144472042	0.317	0.100
MARC0093463	144493600	0.477	0.025
6_144510240	144510240	0.317	0.100
ALGA0037516	144547947	0.498	0.245
6_144622895	144622895	0.376	0.635
ALGA0037532	144631809	0.095	0.531
ALGA0037534	144661058	0.095	0.531
6_144737962	144737962	0.128	0.945
6_144786659	144786659	0.246	0.547
ASGA0030017	144807148	0.246	0.547



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no