Estimation of the frequency and effects of causal mutations on fertility and production traits in Irish dairy cattle

Submitted for the purpose of an MSc award by;

Lyndsey O'Kane, BSc Toxicology

Based on the research carried out under the supervision of

Dr. Michael Mullen (TUS), Prof. Neil Rowan (TUS)

and Dr. Ross Evans (ICBF)

Contents

List of Figures

List of Tables

Declaration

I declare that this thesis is an original report of my research, has been written by me and has not been submitted for any previous degree. The experimental work is almost entirely my own work; the collaborative contributions have been indicated clearly and acknowledged. Due references have been provided on all supporting literatures and resources.

Signed _____ Date _____

List of Abbreviations

AI – Artificial insemination

BLAD – Bovine leukocyte adhesion deficiency

BLUP –Best linear unbiased prediction

CVM – Complex vertebral malformation

DGAT - Diacylglycerol Acyltransferase

DNA – Deoxyribonucleic acid

DUMPs – Deficiency of uridine monophosphate synthase

EBV – Estimated breeding values

FANCI – Fanconi Anemia Complementation Group I

GWAS – Genome wide association study

ICBF - Irish cattle breeding federation

IDB – International dairy and beef genotyping platform

LD – Linkage disequilibrium

LFNG - O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase

MAF – Minor allele frequency

NEB – Negative energy balance

REML – Restricted maximum likelihood

RFLP – Restriction fragment length polymorphism

RNA – Ribonucleic acid

SNP – Single nucleotide polymorphism

STAT - Signal transducer and activator of transcription

PTA - Predicted transmitting ability

QTL - Quantitative trait loci

Abstract

Fertility is a major driver for profitability and sustainability of livestock enterprises. Identifying and estimating the effects of known lethal recessive genetic mutations and genes of major effect on production traits in cattle populations provides additional information to the industry for potential incorporation into breeding programs. Such information may support breeders to make more informed decisions through the identification of carrier animals and the evaluation of potential strategic matings in cases where carrier animals may be of otherwise high genetic merit. This project aims to estimate the frequency and effects of a panel of DNA polymorphisms (n=18) in Irish Holstein Friesian cattle, some of which are validated as causative mutations responsible for lethal recessive disorders (CVM, BLAD, DUMPS and Brachyspina), and some of which have been observed to have major effects on production and functional traits in previous research studies (*STAT1*, *STAT3*, *STAT5*). Genotypes on 21,707 Holstein Friesian dairy cattle were obtained from the ICBF, as were phenotypic data on milk, fertility, carcass and health traits (n=16). Phenotypes, expressed as predicted transmitting abilities (PTAs) were prepared for inclusion in the analysis by removal of parental contributions through a deregression process. Haplotypes were predicted using PHASE for all SNPs with positions on the same chromosome. Subsequently, associations between each SNP/haplotype and PTA were analysed in ASReml using a weighted mixed animal model. Several associations between the genes of major effect and production and functional traits were evident and consistent with previous reports of such associations, for example SNPs within the *DGAT* and Casein genes were associated with milk composition traits as expected, however, they were also associated with fertility and carcass traits. SNPs within the *STAT* genes, of which there has not been extensive previous studies on in cattle populations, were associated with both production and functional traits in the population studied. Additionally, a candidate novel lethal recessive mutation in *LFNG* has been identified. The results from this project will be evaluated by our industry partners, the ICBF, responsible for national genomically assisted breeding programs in Irish cattle.

Chapter One -Introduction

1.0 - The history of animal breeding

Animal breeding can be defined as the selective breeding of domestic animals with the aim of improving desirable traits, which are heritable and so can be improved in this way, in the next generation. Animal breeding strategies where animals are selected for breeding based on their own, or their relatives' performance, were implemented long before population genetics became a common scientific principle. The publication of the 'Origin of species' in 1859 included the following quote by Darwin 'our oldest domesticated animals are still capable of rapid improvement or modification' a viewpoint that has been realised and is currently being expanded upon within the agricultural sciences. This could not be more evident than the example of the chicken, which in 1994 had grew to market size in one third of the time in comparison to 35 years previously, despite consuming less than half of the feed of its ancestors (Havenstein et al., 1994).

Research from archaeological and genetic studies have determined that domestication of cattle, which is a form of mutualism between the human population and a target animal population, occurred 10,500 years ago just after the Neolithic period (Bocquet-Appel and Bar-Yosef, 2008;Zeder et al., 2006). In 1989, Bawden proposed four stages in the evolution of agriculture and animal breeding systems; pioneering, production, productivity and persistence. Pioneering referred to the initial cultivation of lands where the focus was on subsistence for immediate family members, whereas the production phase which started in the 1950s focused on increasing traits of importance using scientific principles for economic benefit. The productivity stage was preceded by the issues associated with overproduction leading to the focus being directed to efficiency of production, with scientific advances in areas such as disease control and pest resistance aiding this advancement. Persistency, or sustainability, is a concept that began in the 1970s and was fuelled by environmental and animal health issues and is even more pertinent in present times due to the demand of feeding a rapidly growing worldwide population.

Traditional animal breeding relied exclusively on the analysis of ancestry data and observable phenotypes to increase genetic gain in livestock populations without any knowledge of the genetic architecture controlling the trait. However, this process has been relatively slow to produce accurate results, particularly in comparison to what can be achieved with more recent approaches such as genomic selection as will be discussed later in this chapter. It was in the 1700s when Sir Robert Bakewell, an agriculturist, revolutionised animal breeding by the introduction of the recording of phenotypic characteristics that could be used in selection for

traits of importance. He also used inbreeding of animals of high merit to propagate desirable phenotypes in the population (Fussell, 1969). Many people followed Sir Bakewell's approach to animal breeding by recording animal phenotypes with the inevitable result of being inundated with data. It became increasingly more difficult to remember the relationships between the animals leading to the establishment of herd books with the inaugural herd book being that of the Shorthorn cattle breed which was founded in England in 1822. Herd books became commonplace in Europe and America in the following years, a progression that greatly advanced cattle breeding approaches.

Regardless of the species considered, animal breeding goals in the current climate, where the emphasis is on ensuring sustenance for an increasing world population while also attempting to decrease adverse environmental impacts, are focused on improving productivity while also improving feed conversion. The improvement of functional traits, such as fertility and health, is fast becoming an important objective within the industry to ensure less replacement animals, and therefore costs, are affecting farm enterprises. The following section describes terminology commonly used in quantitative genetics, with examples of those which are relevant to cattle breeding.

1.2 - What is a phenotype?

A phenotype is any measurable trait or characteristic of an individual. Phenotype expression is a result of the activity of gene products resulting from the expression of those genes. Polymorphisms in genes can result in a change in the function of or a change in the expression levels of a gene product. Polymorphisms in non-coding areas can affect the regulation of gene expression.

In animal breeding programmes, failure to account for all traits of importance can lead to selection for some traits having an antagonistic effect on other vital traits in the population. With regards to dairy cattle, important characteristics include the yield and quality of milk produced, reproductive performance, carcass quality, health status, nutritional requirements, and environmental impact (Berry, 2015).

1.2.1 - Simple traits

A trait that is only influenced by a mutation in a single gene can be referred to as a monogenic, Mendelian or simple trait. A number of diseases observed in cattle populations, including the lethal recessive disorders being investigated in this study, are the result of this underlying genetic mechanism. Other diseases that follow this genetic mechanism may not result in

lethality, such as the case with Bovine leukocyte adhesion deficiency (BLAD), however they can cause decreased fitness to the affected animal and economic loss to the agricultural industry. These disorders are studied using nominal variables that take two possible values (i.e affected/unaffected with regards to disease status or carrier/non-carrier status of a particular mutation). Whereas simple traits are much easier to define and observe, many measurable phenotypes follow different genetic mechanisms which are more difficult to decipher, as discussed below.

1.2.2 - Complex traits

Complex traits are caused by the interaction of multiple genes in combination with environmental influences, also known as multifactorial or polygenic disorders. A common genetic interaction, epistasis, occurs where the effects of one gene are affected by one or several other modifier genes (Phillips, 2008). The identification of loci responsible for complex traits has been less successful than that of simple traits due to incomplete penetrance, environmental effects and the increasing number of loci which may be associated with the trait (Cordell, 2002).

Pleiotropy occurs when a single gene influences multiple phenotypic traits. Consequently, polymorphisms in a pleiotropic gene may influence multiple traits simultaneously. When a gene codes for a protein that is used by numerous cell types or one that has a signalling function, pleiotropy is likely to occur (Mackay, Stone and Ayroles, 2009). These effects may be direct, as in the case with albinism where the genetic mutation causes the suppression of formation of pigment in each organ system in which it occurs (i.e in both the hair and eye) or indirect, as in the case of the *STAT* genes being analysed in this study, which from part of a signal transduction system involved in many biological pathways. The following section discusses how geneticists began to understand the genetic architecture of these traits and applied their findings to improving livestock populations.

1.3– Quantitative genetics in animal breeding

Developments in statistical analysis comprehension and application influenced the agricultural community, greatly enhancing the ability for animal breeders to develop programs that led to more rapid advancements than was previously possible. Notable developments, such as those made by the English geneticist and statistician, Ronald A. Fisher, whose analysis of variance allowed for the separating of genotypic variance into its additive, dominant and epistatic components, and Wright's description of inbreeding and relationship coefficients and their

effect on genetic variation, contributed significantly the field (Fisher, R.A., 1919; Wright, S., 1921).

Lush, often referred to as the father of modern scientific animal breeding, was the first to advocate that animal breeding not only be based on the subjective appearance of the animal, but also utilising quantitative statistics and genetic knowledge. He described heritability in the narrow quantitative sense as the ratio of genetic to phenotypic variance and established the ‘’breeders equation’’ for predicting the response of selection strategies (Lush, 1937). Along with his colleague, Hazel, he developed the selection index theory as a method of artificial selection in which several useful traits are selected for simultaneously and he described the weighting of these traits by their relative importance in a specific breeding program (Lush, 1947). In recent times the weighting on each trait in a multi-trait index is dependent on the amount of additive genetic variation in each trait, the genetic relationships among the traits and their relative economic importance (Simm et al., 2009).

The following section summarises the evolution of the methods and theories that have been developed since Lush started using selection index theory in relation to animal breeding.

1.4 – Estimated Breeding Values

The estimated breeding value is an estimate of the true genetic potential of an animal, which is expressed relative to the population average or a specific cohort within the population. Estimation of breeding values is based on pedigree and performance information and describes the genetic potential of the animal independent to the environment, expressed as values presented in units associated with the trait pertaining to the EBV. Best linear unbiased prediction (BLUP) is a widely used method of calculating these values, with the accuracy of the EBV depending largely on the amount and quality of data collected, which includes the phenotype and pedigree information (Searle, 1997: Schneeberger et al., 1992). This method was developed by Henderson, who was a student of Lush, and was first published in his PhD thesis in 1948. In comparison to the earlier selection methods which used least square estimations, this model also took into account fixed and random effects (Henderson, 1948). Previous estimations of breeding values were performed on animals that were all reared together, and so it was assumed that environmental effects were accounted for, however the advent of artificial insemination (AI) made it more difficult to account for these effects, a difficulty which was circumvented by Henderson’s use of fixed and random effects in the BLUP model (Hill, 2014). BLUP was originally utilised in order to predict the breeding value

of male parents in the sire model, however Henderson further developed this model in the 1970s to allow for the prediction of breeding value for all individuals in a pedigree in what is now termed the animal model (Hill, 2014). BLUP methodology is now incorporated into residual maximum likelihood (REML) estimations and is a standard statistical method in quantitative genetic analysis. The use of the inverse of Wrights numerator relationship matrix (A) is required for best linear unbiased prediction of breeding values, and is determined using pedigree records (Quaas, 1976).

The animal model formula is as follows;

$$Y_i = \mu + a_i + c_i + e_i$$

Where Yi = the phenotypic information, μ = the average population phenotypic mean, a_i = the breeding values (accounts for the influence of the additive effect of the alleles on the phenotype), c_i is the known and measurable environmental effects and e_i = a residual accounting for the rest of the possible variation.

1.4.1 – Heritability

Selective breeding will only be successful in cases where the trait under selection is heritable. Heritability refers to the ratio of the genetic to the total phenotypic variance in a population. Phenotypic variance in a population is a result of both environmental factors and the genes that control traits. Knowledge of the heritability is essential in order to genetically improve any quantitative trait as this parameter highlights the precision in predicting genetic value from phenotypic information.

$$V_P = V_G + V_E$$

Where genetic sources of variance can be segmented into additive (V_A), dominant (V_D) and epistatic (V_I) variance;

$$V_G = V_A + V_D + V_I$$

And environmental variance may be categorised into specific environmental variance (V_{Es}), general environmental variance (V_{Eg}), and genotype by environment interaction (V_{GxE});

$$V_E = V_{Eg} + V_{GxE} + V_{Es}$$

Estimating the heritability of a particular trait allows a comparison to be made on the relative importance of genes and environment to the variation observed for that trait, making it an important factor when investigating the response to selection in agriculture (Visscher, Hill and Wray, 2008). Broad sense heritability (H^2) explains the ratio of genetic variation that is due to dominant and epistatic effects of genes;

$$H^2 = V_G/V_P$$

Whereas narrow sense heritability (h^2) accounts for the ratio of genetic variation attributable to additive genetic variation. This, parameter also significant in animal selection programs as response is dependent on additive genetic variance and the resemblance between relatives is mostly attributed to this influence (Hill, Goddard and Visscher, 2008);

$$h^2 = V_A/V_P$$

Heritability is population specific, due to the fact that environmental and genetic variance will differ depending on the population being studied. Segregation of alleles, their frequencies and effect sizes, all contribute to genetic variance and all these factors are likely to vary among different populations. Environmental conditions are also expected to vary across populations. However, very similar heritability values are often witnessed in populations of the same species, and sometimes across species, and are often higher for production traits than for performance traits (Visscher, Hill and Wray, 2008). In breeding schemes, traits with high heritability can benefit from the animal's own phenotypic information to more accurately predict the animal's breeding value. However for traits with low heritability, information from many relatives is necessary to predict accurate breeding values. The heritability of a trait is not constant and can change over time, which is evidenced by the fact the average heritability for first lactation milk yield in dairy cattle rose from 0.25 in the 1970s to 0.4 in 2003 (Tong, Kennedy and Moxley, 1979; Berry et al., 2003).

1.5 - Marker assisted selection

Traditional selection strategies have led to an immense advancement of productivity in agricultural performance, particularly since the advent of quantitative genetics which has had a profound role in this advancement since the 1930s. Limits to quantitative selection approaches, however, led to interest in utilising molecular genetics approaches to complement the quantitative approached. The advent of DNA based molecular markers in the 1970s also contributed to the eventual molecular technology integration to animal breeding (Guimaraes et

al., 2007). Limitations' to quantitative genetic approaches, including the phenotype having a low heritability, difficult or expensive to measure, expressed later in life as is the case with fertility traits, expressed in only one gender (milk yield in dairy cattle), the phenotypic value only being available after culling of the animal (many carcass traits), or the accurate genetic potential prediction being complicated by epistatic mechanisms, can all be augmented by the application of molecular genetic techniques when estimating breeding values (Dekkers and Hospital, 2002). This led to the initiation of marker assisted selection, which is an indirect selection process based on the recognition of marker genes that indicate the presence of desirable genes through linkage. This selection methodology is implemented by the addition of marker data with the traditional phenotypic evaluations, allowing improved estimations of breeding values and an increase in the rate of genetic gain (Wakchaure and Ganguly, 2015).

This approach led to the identification of markers for traits that were the result of major gene influences, for example *DGAT1* for milk fat percentage and *MSTN* for double muscling in cattle populations. However, most QTL identified only accounted for a small percentage of the genetic variability and are easier to detect for traits with high heritability such as production traits, whereas functional traits such as fertility are more difficult to detect (Hill, 2014). The development of SNP technology allowed researchers to analyse thousands of markers simultaneously and enabled association mapping which led to higher accuracy in QTL discovery.

1.5.1 – Linkage Disequilibrium

In the 1800s the scientist Gregor Mendel proposed a process by which genes are inherited independently of each other, a proposition which was based on his own experimental observations at the time. It is now known that this process is a result of genetic recombination in the gamete and although this observation holds true in many instances, there are circumstances in which genetic loci are likely to be inherited together, particularly if they lie close to each other on a chromosomal region. Linkage disequilibrium (LD) is the non-random association of alleles at different loci and its calculation provides a measure of the deviation from the expectation of non-association of genetic loci. This allows the analysis of haplotypes which are a set of alleles located on a single chromosome, adding statistical power to association studies (de Bakker et al., 2005). The principal motivation for the construction of a linkage disequilibrium (LD) map in the genome is to aid in identifying and characterising genetic variants related to complex traits. LD is important within the agricultural sciences, not only as a way of measuring association through haplotype analysis, but also as a tool for

investigating the evolutionary history of livestock populations and for estimating the response that is likely to occur through artificial selection (Dekkers and Hospital, 2002).

A limitation of genetic association studies is that replication or repeatability is not achieved in subsequent studies investigating the same variants with the disease or trait of interest. This is often due to the fact that the genetic polymorphism being studied is not the actual casual variant, but rather the polymorphism is in LD with the causal mutation. LD is dependent on population history, particularly the genetic make-up of the founder individuals of that population. For example, if all alleles on a section of DNA have been derived from a common recent ancestor then there is little chance of a significant amount of recombination events occurring during meiosis to separate these alleles within that stretch of DNA and they are expected to be inherited together. However, if it has been a longer time since a common ancestor is identified between two populations, recombination events are likely to have disrupted the LD in the genomic region, complicating genetic studies due to the fact that it is possible that a variant may be associated with a disease or trait of interest in one population, but not in another (Hirschhorn et al., 2002).

There are two commonly used equations that measure linkage disequilibrium within a genome. The first one is a measure of D',

$$D = P_{AB} - (P_A P_B)$$

And the second one is a measure of correlation r^2

$$R^2 = D^2 / P_1 P_2 Q_1 Q_2$$

Need to explain what the elements of the equations are i.e. what is P1 and Q1!

1.6- Molecular markers and their use in animal breeding

In recent times, the use of DNA markers such as microsatellites and single nucleotide polymorphisms (SNP) has been successful in identifying quantitative trait loci (QTL) associated with phenotypic traits deemed important in livestock to support animal breeding objectives. However the causal mutation within these genomic loci, termed the quantitative trait nucleotide (QTN) has not been identified in all cases at this stage (Ron and Weller, 2007). Current advancements are focusing on systems biology which aims to account for complex

regulatory relationships between phenotypes and genotypes allowing the full understanding of the genetic architecture underlying phenotypic traits of interest (Berry et al., 2010). This approach has been supported by the sequencing of the bovine genome by Liu et al. in 2009, which now allows the analysis of genes and regulatory genomic sequences for polymorphisms associated with traits of interest.

The advent of DNA markers has revolutionised the biological sciences, allowing for the study of evolutionary pathways and genetic association studies in human and livestock populations. The following section of this chapter explains the genotyping strategies that have been used in the past to detect genetic polymorphisms, their limitations, and the current popular strategy to distinguish between variants within populations which have led to the advent of genome wide association studies (GWAS).

The ability to identify genotypes that are relevant to a phenotype of interest is of vast benefit to animal breeding. A genetic marker provides information about allelic variation at a given locus. In many cases the genotype is not indicative of the underlying biological mechanism involved in the expression of the trait, but instead is in direct linkage to the causative allele (Schlötterer, 2004). Polymorphic DNA markers, including restriction fragment length polymorphisms (RFLPs), microsatellites, and more recently single nucleotide polymorphisms (SNPs) have been extensively used in linkage, association, and gene function studies, allowing the identification of markers which are useful in the selection for desirable traits in cattle populations (Beuzen, Stear and Chang, 2000).

1.6.1 - What is a genotype?

Each living cell contains the genetic information, which is stored as DNA, that codes for all the necessary instructions for an organism to function. Diploid organisms have two copies of every gene, with each copy inherited from its parents, with the possibility of each gene being present in different forms called alleles. Mutations, which occur through many different mechanisms, cause variability in the genomes of species, which may lead to either neutral consequences or observable differences between individuals, depending on the location that the mutation has occurred. The type of genetic variation studied by scientists has depended largely on the information available on the genome of the organism being studied. Genome sequencing has revolutionised this field, allowing for quick and less expensive options to study genetic variation than was previously possible. The following section describes previous and current methods utilised to genotype organisms.

1.6.2 - Restriction fragment length polymorphisms

The identification and isolation of restriction enzymes from bacterial cells in the 1960s, and the realisation of their potential use in cutting DNA at specific sites within the genome has revolutionised the field of genetic markers (Smith and Wilcox, 1970; Danna and Nathans, 1971). As an immune function of bacterial cells to evade viral infection, they cut at precise DNA sites, making them particularly useful in the analysis of DNA variation (Roberts, 2005). A restriction fragment length polymorphism (RFLP) is a genetic marker resulting in the change in the pattern of fragments produced when DNA is cut with Type II restriction enzymes, therefore allowing the analysis of genetic variation through visualisation of band size after gel electrophoresis. Initial genetic dissection of variants in cattle that affect phenotypic traits used this technology in selective breeding programmes. For example, Damiani et al, 1990, discovered polymorphic sites in the Kappa casein gene and described the capability of increasing the allele associated with the cheese making properties of milk. The laborious nature of analysing RFLPs and the advent of new technology allowing simpler and quicker methods to detect genetic variants has largely made their use obsolete.

1.6.3 - Microsatellites

Microsatellite genotyping approaches became popular during the 1990s when polymerase chain reaction (PCR) technology became accessible within laboratories, allowing the primer design and amplification of DNA steps that are pertinent to this method of genotyping. Microsatellites are repetitive DNA motifs, typically 2-6 base pairs in length, which are repeated between 5-50 times allowing the analysis of variations between and within populations and individuals. Their wide distribution throughout the genome, highly polymorphic nature, and codominant transmission makes them ideal genetic markers for many applications in science (See Fig 1.0) (Abdul-Muneer, 2014). The use of dinucleotide, trinucleotide and tetra nucleotide repeats are most practical for molecular genetic investigations. Microsatellites are identified and isolated using primers which are designed to bind to the flanking region on either side of the repeats with subsequent amplification of the locus with PCR. The amplified DNA can then be separated based on its size by gel electrophoresis, allowing the analysis of the difference in the number of tandem repeats (Selkoe and Toonen, 2006). In cattle populations, microsatellite analysis has been utilised to reveal evolutionary history (MacHugh et al., 1997), to assign parentage (Usha, Simpson, and Williams, 2009), and for performing association analysis between specific genes and economically important traits (Sharif et al., 1998; Curi et al., 2005).

Despite their obvious benefits and the widespread use of microsatellites for genetic investigations, there are some drawbacks associated with their use. These include the development of primers that are species specific, with interindividual differences within a species also needing consideration.

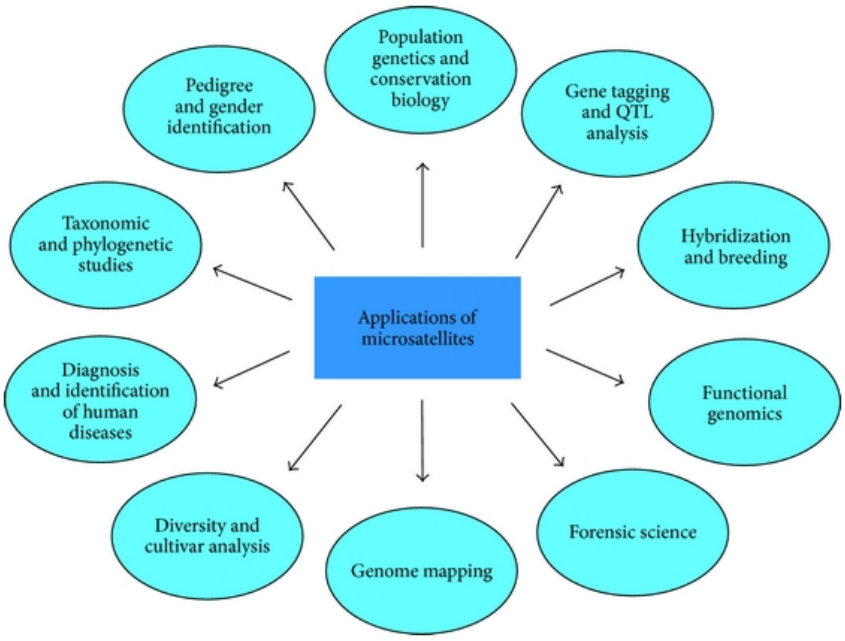


Fig. 1.0 -The application of microsatellites in scientific research (Abdul-Muneer, 2014).

1.6.4 - Single nucleotide polymorphisms (SNPs)

Advancements in DNA sequencing have found genetic variation at the nucleotide level among individuals. SNPs, the most common form of genetic variation within populations, are usually biallelic and as such can distinguish between two variants of a gene, or chromosomal region (Kwok, 2003). When testing for associations between the candidate loci thought to be responsible for quantitative traits, the biallelic nature of SNPs may present a challenge in comparison to the multi-allelic information received from microsatellite analysis, however, this can be overcome by utilising haplotype frequencies from several SNPs in the loci of interest (Vignal et al., 2002), a concept discussed later in this chapter.

SNPs which occur within a gene can affect the protein composition and thus the function of the gene, while SNPs in regulatory regions can affect the regulation of gene expression. Moreover, SNPs in non-coding regions of a chromosome can be tightly linked to causative

mutations affecting phenotypic traits and as such can be utilised in selective breeding programmes (Syvänen, 2001). Genotyping SNPs involves the use of DNA microarray technology, which is based on hybridisation between designed allele specific probes with fluorescently labelled fragmented DNA (See Fig 1.1).

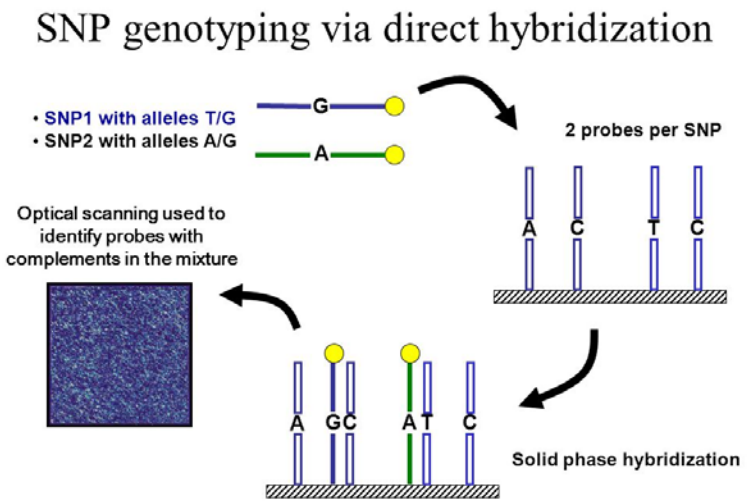


Fig. 1.1– Visual representation of a SNP genotyping panel via direct hybridization (Cd-genomics.com, 2017)

1.7 – The future of animal breeding – Genomic selection

The advent of SNP technology has made genomic selection, which was first described by Meuwissen et al in 2001, an attractive mechanism for attaining faster genetic gain than that achieved based on phenotypic data or marker assisted selection. It is based on the selection for many thousands of SNPs that cover the whole genome simultaneously, taking advantage of the linkage disequilibrium between these SNPs and the causal genetic factors that are responsible for the observed phenotypic variation. The ability to combine close set markers into haplotypes eliminates the need to establish linkage phase in all families, as chromosome segments that contain the same rare haplotype are likely to also carry QTL alleles known to affect important traits. Using genomic information on young animals and genomic and phenotypic data on their older relatives allows the prediction of breeding values on these young animals without phenotypic records, which greatly reduces the generation interval and so increases the rate of

genetic gain (Hill, 2014). An increase in the accuracy of selection can also be achieved for animals that have both genomic and phenotypic data available. It also allows for the replacement of the pedigree relationship matrix (A) with a genomic relationship in BLUP, now termed GBLUP. Van Raden et al (2009) reported an increase in accuracy of prediction with genomic selection (GBLUP) compared to traditional pedigree based selection (A matrix) from 0.28 to 0.47 for cows for milk yield. Many countries around the world have adopted genomic selection strategies, including Ireland, whose approach is discussed below.

1.7.1 – Genomic selection in Ireland

The Irish dairy industry has been using genomic prediction to improve genetic gain in Irish dairy cattle since 2009. The training population in Ireland initially consisted of utilising daughter proven genotyped AI sires, both domestically proven and internationally through access to INTERBULL MACE proofs (Berry, Kearney, and Harris, 2009; Kearney, Cromie and Berry, 2010). Including bulls with progeny in the training population provides a more accurate estimation of the true genetic merit, particularly for low heritable traits such as fertility, since a sire's daughters performance contributes to the traditional estimated breeding value of the sire himself (Spelman, Hayes, and Berry, 2013). More recently informative cows with their own phenotypes have also been added to the training populations. Initially genomic predictions were only implemented in Holstein Friesian dairy cattle but since 2014 the same strategy has been implemented in beef cattle (Cromie et al., 2014). The development of the Irish custom-built genotyping panel, the International Dairy Beef (IDB) 19k chip (discussed below), which added 9,973 variants to the commercial Illumina LD panel, allowed the imputation from a low-density platform to a higher density and aids in the application of genomic selection in Irish cattle (Mullen et al., 2013). From 2009 to 2014 correlation analysis has shown up to 29% improvement in prediction accuracy with the genomically enhanced predicted transmitting ability (PTAs) compared to the parentage average PTA (Cromie et al., 2014). Genomic selection has played a large part in the increase of the genetic merit of dairy herds within the last 8 years with the dairy industry gaining € 750m in net profit overall during the past 15 years. There have been alternative uses for genomic selection strategies proposed which may further accelerate genetic gain in livestock species.

The possibility to further increase genetic gain by genotyping embryos has been described by Fisher et al, 2012. This will allow the implantation of embryos of high genetic merit only and therefore eliminate the expenses associated with rearing animals of insufficient genetic value. Also, of value is the ability of genotyping to decrease the rate of inbreeding in dairy cattle

through the calculation of genomic inbreeding. Inbreeding is known to reduce the rate of future genetic gain. Genotyping also has the on-farm advantage of parentage assignment, avoidance of genetic defects such as CVM, and genomic mate allocation (Spelman, Hayes, and Berry, 2013). The advent of accurate imputation techniques to higher densities is allowing the application of low density genotyping thereby reducing the costs and making it more amenable to industry use.

1.8 - Bioinformatics

The sequencing of genomes has led to the requirement for researchers to develop methods to collect and analyse large quantities of complex biological data efficiently. Luscombe, Greenbaum and Gerstein, (2001) defined bioinformatics as ‘conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying "informatics" techniques (derived from disciplines such as applied maths, computer science, and statistics) to understand and organize the information associated with these molecules, on a large-scale’. The following are some of the tools that were utilised in the bioinformatics analysis of the genetic polymorphisms analysed in this study.

1.8.1 - Databases

1.8.1.0 - Ensemble

Ensembl is an online database, set up and ran by European Molecular Biology Laboratory's European Bioinformatics Institute, whose goal is to annotate and distribute genomic datasets, making the information freely accessible to researchers around the globe (Zerbino et al., 2017). The website provides access to gene sets for various species and sequence alignment between species can be evaluated though its Basic Local Alignment Search Tool (BLAST) function. Comparative, regulatory and variation data are also available through the tools available on the website.

1.8.1.1 -The National Centre for Biotechnology Information (NCBI)

The NCBI aims to advance science and health by providing access to biomedical and genomic information (Ncbi.nlm.nih.gov, 2018). Like ensemble, it provides genome annotation, but also provides a nucleotide database which is a collection of sequences from several sources, including GenBank and RefSeq. This tool provides invaluable information that is required when determining the effect of a SNP within a gene of interest.

1.8.1.2 -Online Mendelian inheritance in animals (OMIA)

The OMIA is a record of disorders of known genetic aetiology. The website provides information on the disorder and the genetic variants associated with them, as well as references to the scientific journals that describe them (Omia.org, 2018).

1.8.1.3 -BLAST (basic local alignment search tool)

BLAST is an application which finds regions of similarity between biological nucleotide and protein sequences. A sequence of interest is scanned against a database of sequences of many

organisms, and those that show similarity above a given threshold are selected (Altschul et al., 1990). It has been shown that sequence conservation indicates functionality, and these sequences tend to show more conservation than other regions of the genome (Frazer et al., 2003). Many of the bioinformatics tools utilised for assessing the impact of a SNP on protein functionality are based on this premise.

1.8.2 - Bioinformatics tools for assessing SNP effects

1.8.2.0-Polyphen 2

Polyphen 2 is a software tool which predicts the consequences of an amino acid substitution on protein structure and function. It is calculated based on predicting the effect of a variant by extracting sequence and structure-based characteristics of the substitution site. Substitutions that occur at pertinent sites, such as the active site of an enzyme, a transmembrane region, within a disulphide bond or crosslink region are identified and the possible functional effect is determined using a Naïve Bayes posterior probability machine learning algorithm. Estimates of false positive rate (FPR, the chance that the mutation is classified as damaging when it is in fact non-damaging) and true positive rate (TPR, the chance that the mutation is classified as damaging when it is indeed damaging) are also reported. A mutation is also appraised qualitatively, as benign, possibly damaging, or probably damaging based on pairs of false positive rate (FPR) thresholds (Adzhubei et al., 2010)

1.8.2.1 -Sift

Sift, similarly to polyphen 2, predicts the consequence of non-synonymous amino acid substitutions based on sequence alignment. The premise is that highly conserved sequences are essential to protein function. The software then calculates the probabilities for all possible amino acid substitutions and positions with normalized probabilities less than 0.05 predicted to be deleterious, those greater than or equal to 0.05 are predicted to be tolerated (Ng and Henikoff, 2001). Sift differs from Polyphen 2 in that it doesn't evaluate structure-based characteristics of the substitution.

1.8.2.2 -SNAP

SNAP (screening for non-acceptable polymorphisms) is a neural network-based method for the prediction of the functional effects of SNPs. SNAP uses both sequence alignment and structural and functional information, where available, to predict the effects of an amino acid substitution. The structural and functional information that is predicted using this tool includes the

biochemical properties of the protein, the change in 1D structure of the protein and information about the family of the protein (Bromberg and Rost, 2007).

1.8.2.3 -PANTHER

PANTHER predicts the impact of non-synonymous mutations based on evolutionary preservation. Like evolutionary conservation, which is known to play a part in protein functionality, homologous proteins are used to recreate the likely sequences of ancestral proteins, where the history of each amino acid is traced back for an estimation on how long that sequence has been preserved (Tang and Thomas, 2016).

1.8.2.4 -MAPP

Multivariate Analysis of Protein Polymorphism (MAPP) predicts the effects of SNPs by quantifying the physicochemical variation in a multiple sequence alignment and calculating the deviation of candidate amino acid replacements from this variation. The larger the deviation the higher the probability is that a SNP will impair protein function (Stone and Sidow, 2005).

1.8.2.5 -PredictSNP

PredictSNP is a consensus classifier that was designed using the six software tools described above, which results in enhanced prediction performance. It has been shown to lead to more accurate results than by the predictions delivered by individual programs (Bendl et al., 2014). A simple to use web interface allows the user to input the protein sequence and SNPs of interest and the output results are normalised to percentage confidence allowing comparisons between the tools and an overall PredictSNP score which considers the result obtained from each tool.

1.8.2.6 -RegRNA

The analysis of the effect that mutations in intronic regions have on gene regulation and expression is made complicated by their very nature and the relatively few tools that have been developed to date. One such tool, RegRNA, evaluates these genetic mutations by the comparison of mRNA sequences to known RNA regulatory motifs. Prediction tools involved in analysing motifs existing in 5’ and 3’UTR regions, those involved in splicing (donor and acceptor site), transcriptional regulation and miRNA target sites are compared to the input mRNA sequence (Chang et al., 2013).

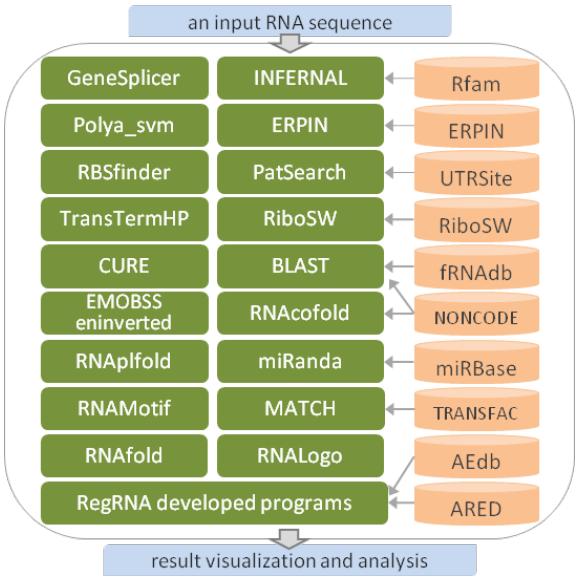


Fig. 1.2– Information flow in the utilisation of the RegRNA software program (Chang et al., 2013)

1.9 - The Irish Cattle Breeding Federation

The Irish Cattle Breeding Federation (ICBF) was established in 1998 for the purpose of providing information on cattle breeding to the dairy and beef industries of Ireland. Their main purpose is the application of genetic gain which has benefits for the farmer, the agri-food industry and ultimately the consumer. This non-profit organisation is committed to utilising the latest technology and scientific principles, which include information technology skills, the agricultural sciences, and the latest developments in genetic and genomic advancements, to achieve this goal. With the main objective of genetic gain in mind, the central aims of the organisation are to ensure that ancestry data, genotypes and phenotypic observations are available for many animals in each generation. Subsequently, the practice of genetic evaluations ensures that superior animals are identified as genetic improvement can only be achieved when the parents of the next generation are genetically greater than their contemporaries (Icbf.com, 2017). The ICBF are also responsible for the running of G€N€ Ireland, Ireland's progeny testing scheme, whose main dairy objectives are to test 70 bulls per year with each bull having 100 heifer replacements recorded in 100 herds by the fourth year of the programme.

To meet the objective of increasing the genetic merit of the Irish dairy and beef herds the ICBF have established a database which holds all information on the ancestry, phenotypic and genotypic data relevant to each animal that is registered in the country. Information flow

between the ICBF and industry partners such as AI companies, veterinary surgeons and laboratories, milk co-operatives, Teagasc, genomic labs, herd book establishments and any other necessary institutions within the agricultural industry, allows the ICBF to have access to relevant data to perform accurate genetic evaluations. This, in turn allows them to provide data to farmers and stakeholders which results in meeting the industries objectives. Herd plus is a service ran by the ICBF by which farmers and relevant personnel can benefit from information contained on the database by allowing them access to reports to complement their breeding strategies.

1.9.1 - The International Dairy and Beef Chip

The development of a custom genotyping panel in 2013, by the ICBF, Teagasc and Weatherbys has supported the development and implementation of genomically assisted breeding programs for cattle in Ireland. Variants of importance, including those used in genetic evaluations, major genes with known large effects, genes responsible for lethal recessive disorders and genes of research interest were added to an Illumina low density bovine genotyping panel. Also included were over 5000 variants to improve imputation to a higher density platform and variants to allow for imputation to microsatellite genotypes, the method that was until recently used in parentage verification, which allows the use of SNPs for this purpose, further decreasing costs by mitigating the need for back pedigrees to be genotyped using this updated SNP technology (Mullen et al., 2013).

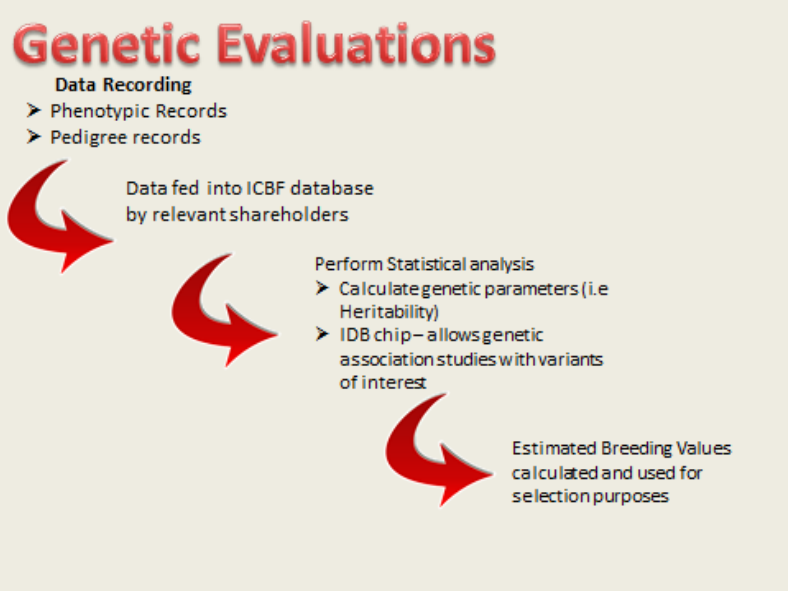


Fig. 1.3– The flow of relevant information through the ICBF that allows for generation of estimated breeding values which are subsequently used for selection purposes.

1.10 - Holstein Friesian Cattle

The Holstein Friesian breed has been the dominant dairy breed in Ireland since the 1950s, with the pure bred being easily identified by its characteristic white and black colour. A 2007 report by the ICBF stated that Holstein Friesians produce on average 5,500 kg of milk per year, containing 203 kg of fat and 185 kg of protein. The average gestation length for this breed is 281.9 days, with difficult births being reported in 6% of cases. Mortality (percentage dead at 28 days) stands at 3.3 %. Beef traits reported include culled cow carcass weight at 319 kg and carcass conformation at 4.7 (1-15 EUROP scoring system). Performance values were reported at 39.1, 21.1, 17.1, 9.8, - 7.8, and - 1.2, for EBI, milk, fertility, calving, beef, and health, respectively (Cattle Breeding in Ireland, 2007). In Ireland herd books for this breed are managed by the Irish Holstein Friesian Association which has been running since 1991 and is licensed by the Department of Agriculture (Ihfa.ie, 2017).

1.11 – Genetic association studies – The candidate gene approach

Genetic association studies aim to detect statistical relationships between genetic polymorphisms and phenotypes or disease traits of interest (Lunetta, 2008). In human populations the emphasis is often based on disease status and the underlying genetic architecture which permits diagnosis and treatment of genetic disorders appropriately. In animal breeding the importance is placed on identifying individuals to be selected for traits that are economically important based on their genotype. However, these studies also allow the identification of animals that carry lethal recessive genetic defects, where historically these animals were likely to be culled, the decreasing cost of genotyping allows for strategic mating for carrier animals that may be of high genetic merit for important traits. Linkage analysis was primarily used to determine associations with genotypes and phenotypes, however, the principle of this technology makes it sufficient for identifying variants associated with the trait in different families, whereas association allows the identification of variants associated with the trait of interest across the whole population (Cordell and Clayton, 2005). Association studies are also more suited to the study of complex traits, where the analysis of more common variants that have a modest effect on a phenotype is warranted (Ott, Wang, and Leal, 2015).

The candidate gene approach assumes that a gene involved in the physiology of the trait may contain a mutation causing a variation in phenotype. The gene, or part of the gene, is sequenced in different animals and variations in the DNA sequence (polymorphisms) are statistically tested for association with variation in the phenotype (Nani et al. 2015).

1.12 – Reproduction in dairy cattle

Reproductive efficiency is an important factor affecting productivity in livestock industries, with reproductive failure being a common reason for culling otherwise healthy animals and consequently causing significant economic losses within the agricultural industry (Fouz et al., 2014). Analysing fertility traits includes measuring the male, female and embryo parameters which considers factors such as embryo mortality, ovulation rates and fertilisation success (Egger-Danner et al., 2014). Neonatal death also influences livestock industries with greater than 50% of perinatal mortality being correlated with dystocia in cattle populations (Walsh, Williams, and Evans, 2011). The past 40 years has seen a decline in the reproductive success of cattle populations with the apparent causes being challenging to decipher but thought to encompass genetic, environmental, and animal management influences (Diskin and Morris, 2008). Historically breeding objectives for cattle focused on an increase in production traits, which includes milk yield, milk components, carcass, and growth traits. Increased selection for milk production has been correlated with an observed decrease in fertility in dairy cattle populations (Berry et al., 2016; Berglund, 2008; Dillion et al., 2006; Pryce et al., 2004). In this section of the chapter some of the non-genetic factors known to be responsible for a decrease in fertility in dairy cattle will be discussed before a brief description of some genetic factors validated as been associated with fertility are acknowledged.

1.12.1 – Non-Genetic factors affecting reproduction in cattle

Fertility in livestock is largely influenced by environmental and physiological factors. The following section concisely reports on the known non-genetic factors that contribute to fertility issues in cattle populations, and includes the influence that uterine capacity, nutritional factors, oocyte development and infectious disease has on fertility parameters in cattle populations.

1.12.2 - Uterine Capacity

The uterine endometrium must provide an appropriate environment to stimulate embryo development. Failure of this environment to provide the necessary conditions may result in failed implantation, early embryonic death or foetal malformations. The transport of nutrients and the secretion of key molecules by uterine epithelial cells support the development of the conceptus during the peri-implantation stage (Forde et al., 2014). From day 4-5 post fertilisation, until implantation occurs on day 19 the embryo is suspended in the uterus and depends on the composition of the secretions for further development. The endometrial glands secrete proteins, sugars, lipids, carbohydrates, with the protein components being responsible for successful implantation and elongation of the trophoblast (Beltman et al., 2014). An

analysis of the protein composition of uterine flushes was performed and correlated to embryo survival in a study performed by Beltman et al, 2014. It was found that specific proteins were more abundant in the group yielding viable embryos with one protein, platelet-activating factor acetylhydrolase IB subunit (PAFAH1B3), only being present in this group. This protein has also been seen to be significant in early pregnancy in other livestock animals. Elevations in the levels of progesterone throughout the luteal stage of the estrous cycle has been shown to be responsible for controlling the expression of endometrial genes and therefore the composition of the uterine secretions (Mullen et al., 2014).

Maintenance of the corpus luteum (CL) is critical for the establishment of pregnancy. Prostaglandin $F_{2\alpha}$ ($PGF_{2\alpha}$) causes luteolysis of the CL if implantation does not occur (Silvia, 1991). Interactions between the conceptus and the uterine endometrium play a role in the suppression of $PGF_{2\alpha}$ secretion at the end of the luteal stage of the estrous cycle. Embryo expression of interferon – τ (IFN-τ) regulates the cyclooxygenase-2 (COX-2) pathway and therefore the secretion of $PGF_{2\alpha}$. This may suggest that an inability of the conceptus to inhibit $PGF_{2\alpha}$ secretion may be a cause of early embryo death (Thatcher et al., 2001).

1.12.3 - Nutritional factors

Metabolic changes during the post-partum period may force cows into a state of negative energy balance (NEB) which causes the liver to undergo biochemical and functional changes to adjust to an increased metabolic demand (Wathes et al., 2007). NEB has been shown to be a factor affecting reproductive success in cattle by influencing oocyte quality and resumption of the estrous cycle (Leroy et al., 2006; Llewellyn et al., 2007), through a reduction of luteinising hormone (LH) pulse frequency, low levels of insulin, IGF-I and glucose in the blood stream, which are mediators of estrogen production from dominant follicles (Butler, 2003). NEB has also been associated with uterine function, where it has been found to affect immune response to pathogens leading to inflammatory states that impede reproductive performance (Wathes et al., 2007).

As mentioned previously the selection for greater milk yield has correlated to a decline in reproductive efficiency in dairy cattle. It has been suggested that this reduction is largely due to the greater negative energy balance observed in high-producing cows at the peak of lactation (Kadri et al., 2014).

1.12.4 - Oocyte Development and Quality

Oocyte follicular development and oocyte quality are factors to be considered in the analysis of reproductive performance in cattle populations. Oocyte development relies on the transcription of proteins that aid in its maturation, as well as the involvement of cellular organelles which assist with the ovulatory process (Hyttel et al., 1997). The nutritional status of an animal plays a role in the expression of steroid hormones and mediators of ovulatory processes and can have a profound effect on reproductive success. Follicular growth and development are thought to be negatively affected by NEB and the subsequent decreases in LH pulse and lower circulating concentrations of LH and IGF-I (Lucy, 2001).

1.12.5 - Infectious disease

The presence of infectious disease can affect reproductive success in cattle, either by inhibition of important physiological processes or by causing abortion in already pregnant females. For example, *Campylobacter fetus* is a pathogenic bacterium known to cause temporary infertility in affected animals but also can cause abortion of the fetus at 30-70 days of gestation (Givens, 2006). Bovine viral diarrhoea virus (BVDV) is associated with abortion, or foetal malformations, depending on the time of gestation when infection occurs (Grooms, 2004).

1.13 – Genetic factors affecting reproduction in cattle

Identifying the genetic basis of fertility traits is made complicated by their complex nature with the potential of genetic gain being hampered by their low heritability estimates which lie between 0.02 and 0.04 for female traits, and 0.05 and 0.22 for male traits (Berry, Wall, and Pryce, 2014). However, a significant genetic variation is observed between animals for fertility traits, suggesting the opportunity of improvement through genetic selection strategies (Rodriguez-Martinez et al., 2008). Embryonic lethality and neonatal death are outcomes that have severe economic consequences for the industry and several genes have been found to be associated with these consequences in cattle populations. Mutations that cause embryonic lethality tend to affect gene products that are associated with housekeeping cellular functions, including DNA replication and RNA processing, and so disruptions to these genes inhibits normal development (Charlier et al., 2016). The following table is a list of genes that have been validated in cattle populations as being associated with reproductive success in cattle. Below the table, the genetic aetiology behind these mutational effects is briefly discussed.

Table 1.0 - A list of causal mutations affecting fertility and reproduction in cattle populations as identified from the Online Mendelian Inheritance in Animals (OMIA) database (http://omia.angis.org.au)

Gene	Symbol	BTA	Bp	Effect	Breed	Freq (%)	Ref.
ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	*ATP2A1*	25	C->T Missense mutation	Congenital muscular dystonia 1	Belgian Blue	Not reported	Charlier et al., 2008
Solute Carrier Family 6 (Neurotransmitter Transporter, Glycine), Member 5	*SLC6A5*	29	T->C Missense mutation	Congenital muscular dystonia 2	Belgian Blue	Not reported	Charlier et al., 2008
Annexin A 10	*ANXA10*	8	Deletion of maternal exon 2 to 6	Embryonic lethality	Japanese Black	Not reported	Sasaki et al, 2016
Exosome Component 4	*EXOSC4*	14	p.Arg.64 (Stop-gain)	Embryonic lethality	Belgian Blue	1.33	Charlier et al, 2016
Mediator complex subunit 22	*MED22*	11	frameshift p.Leu38Argfs∗25	Embryonic lethality	Belgian Blue	1.15	Charlier et al, 2016
Transmembrane protein 95	*TMEM95*	19	483 C>A Nonsense mutation	Male subfertility	Fleckvieh	Not reported	Pausch et al, 2014
Armadillo repeat containing 3	*ARMC3*	13	1 bp frameshift mutation - STOP	Sperm, short tail	Swedish Red	Not reported	Pausch et al, 2016
Myosin heavy chain 6	*MYH6*	10	Del p.Lys693	Embryonic lethality	Belgian Blue	4.99	Charlier et al, 2016
MER1 repeat containing imprinted transcript 1	*MIMT1*	18	110 kb deletion	Late abortion/Stillbirth	Ayrshire	Not reported	Flisikowski et al, 2010

Gene	**Symbol**	**BTA**	**Bp**	**Effect**	**Breed**	**Freq (%)**	**Ref.**
Rab geranylgeranyltransferase beta subunit	*RABGGTB*	3	missense p.Tyr195Cys	Embryonic lethality	Holstein-Friesian	2.13	Charlier et al, 2016
Ring finger protein 20	*RNF20*	8	nonsense (stop-gain) p.Lys693	Embryonic lethality	Holstein Friesian	1.82	Charlier et al, 2016
Ribonuclease H2 subunit B	*RNASEH2B*	12	662,463 bp deletion	Embryonic lethality	Nordic Red	Not reported	Kadri et al, 2014
Apoptotic peptidase activating factor 1	*APAF1*	5	p.Q579X Nonsense mutation	Embryonic lethality	Holstein	2	Adams et al, 2016
CWC15 spliceosome-associated protein	*CWC15*	15	C-T Nonsense mutation	Embryonic lethality	Jersey	12	Sonstegard et al, 2013
Oligonucleotide/oligosaccharide-binding fold containing 1	*OBFC1*	26	Frameshift mutation	Embryonic lethality	Jersey	Not reported	Charlier et al, 2016
Glycinamide ribonucleotide transformylase	*GART*	1	A-> C missense mutation	Embryonic lethality	Not reported	Not reported	Fritz et al, 2013
Sex steroid-binding globulin	*SHBG*	19	C->T Nonsense mutation	Embryonic lethality	Not reported	Not reported	Fritz et al, 2013
Solute carrier family 37 member 2	*SLC37A2*	29	C->T Nonsense mutation	Embryonic lethality	Not reported	Not reported	Fritz et al, 2013

1.13.1 – Genetic factors causing neonatal death/embryonic lethality

Congenital muscular dystonia 1 (CMD1) is a disease-causing neonatal death which is caused by a C->T missense mutation in the ATPase, Ca++ transporting, cardiac muscle, fast twitch 1 (*ATP2A1*) gene located on chromosome 25 of the bovine genome. The protein coded by this gene is responsible for pumping calcium ions from the cytoplasm into the sarcoplasmic reticulum, thereby inducing muscle relaxation. The substitution of arginine for cysteine at position 559 in exon 14 affects the function of the protein by inhibiting the binding of ATP (Charlier et al., 2008). Congenital muscular dystonia 2 (CMD2), a condition causing many of the same symptoms in new-born calves is caused by a T->C missense mutation in the Solute Carrier Family 6, Member 5 (*SLC6A5*) gene, a transporter responsible for maintaining levels of the neurotransmitter glycine at the presynaptic neuron (Charlier et al., 2008).

A recent sequencing scan of the bovine genome has identified recessive mutations based on the depletion of homozygotes in the population. Nine putative causal mutations were identified and associated with embryonic lethality in different breeds of cattle. A nonsense mutation in the exosome Component 4 gene (*EXOSC4*) was observed at a frequency of 1.3% in the Belgian blue breed, a frameshift mutation in the Mediator complex subunit 22 (*MED22*) gene was observed at 1.15% of the Belgian blue population, a deletion at position 1730 of the Myosin heavy chain 6 (*MYH6*) gene seen in 4.99% of the Belgian blue population, a missense mutation in the Rab geranylgeranyltransferase beta subunit (*RABGGTB*) was observed in 2.13% of the Holstein-Friesian population and a nonsense mutation in the Ring finger protein 20 (*RNF20*) gene was observed in 1.82% of Holstein Friesian cattle. These were confirmed as lethal mutations through carrier x carrier mating (Charlier et al, 2016).

Apaf1 is a critical molecule involved in programmed cell death through the cytochrome c mediated apoptotic pathway which has been found to be directly involved in the developmental and neurodegenerative disorders seen in knockout mouse models. A nonsense mutation in exon 11 of this gene which truncates over 50% of the protein product has been estimated to cause 525,000 spontaneous abortions in cattle over the past 35 years. The mutation leads to 670 C-terminal amino acids being deleted, including WD40 repeats which are specific regions involved in signal transduction, transcriptional regulation, and apoptosis (Adams et al., 2016). Sonstegard et al, 2013, discovered a nonsense mutation in the *CWC15* spliceosome-

associated protein that was associated with embryo lethality in Jersey cattle. This gene codes for a protein that has been shown to be constitutively expressed emphasising its importance in cellular functions and explaining why this nonsense mutation would contribute to embryo mortality (Harhay et al., 2010).

Frtitz et al, 2013, identified three novel mutations associated with embryonic mortality in dairy cattle. Glycinamide ribonucleotide transformylase (*GART*) is an enzyme involved in the biosynthesis of purines, which are components of key molecules such as DNA and RNA. A C->T nonsense mutation in the *GART* gene which leads to the substitution of an asparagine by a threonine at position 290 has been associated with embryonic mortality in dairy cattle. Asparagine at position 290 within this protein is highly conserved between species and has been suggested to play a key role in *GART* function by acting as a binding site for manganese. Interference in this enzymes function is thought to cause abortion in the early conceptus phase due to its disturbance of essential molecular pathways. The second mutation was identified in the sex steroid-binding globulin (*SHBG*) gene which codes for an androgen transporter involved in regulating the plasma concentration of steroid hormones. This C->T substitution mutation induces a premature stop codon leading to 90% of the protein being truncated. Previous studies in mice have determined that receptors involved in steroid metabolism are critical for the gastrulation stage of embryogenesis (DeYoung et al., 2003). The third novel mutation was identified in the solute carrier family 37-member 2 protein (*SLC37A2*) which is involved with the transport of glucose-6-phosphate, a key molecule in cellular energy metabolism. This C->T nonsense mutation introduces a stop codon at the beginning of the protein.

A 110kb microdeletion which truncates the 3' end of the MER1 repeat that contains the imprinted transcript 1 (*MIMT1*) non-protein coding gene, which is part of the maternally imprinted PEG3 domain, has been shown to be associated with late term abortions and stillbirths in cattle populations (Flisikowski et al., 2010). Flisikowski et al determined that when the mutation is inherited from the sire it has an observed mortality rate of 85%, with the survival of 15% likely due to incomplete silencing of the *MIMT1* alleles inherited from the dam. Further studies by the same author suggested that restriction to the carrier foetuses blood supply may play a part in the pathology observed, which is small stature and collapsed lungs. This were further evidenced by a study by Venhoranta et al in 2013 which showed *MIMT1* heterozygous

foetuses displayed intrauterine growth restriction because of inefficient oxygen and nutrient supply through the placenta. Expression of a protein not usually occurring in placental cells, neuropeptide S receptor 1 (*NPSR1*) was detected in the placental cotyledons of heterozygous foetuses suggesting it may also play a role in the pathological presentation and may also act as a predictive tool in its analysis (Flisikowski et al., 2012).

An example of embryo lethality caused by the genotype of the dam rather than the embryo is the deletion of 34kb encompassing exons 2 -6 of Annexin A 10 (*ANXA10*) on chromosome 8 in the maternal genome which has been associated with embryo loss 30-60 days post artificial insemination (Sasaki et al., 2016). Annexins are involved in homoeostatic cellular functions including membrane scaffolding, calcium dependent processes and inhibition of inflammation (Gerke, Creutz and Moss, 2005), however the exact function of *ANXA10* has not so far been elucidated (Sasaki et al., 2016).

1.13.2 – Genetic factors associated with male fertility

Morphological examination of semen samples for artificial insemination (AI) is always performed to ensure no abnormalities are present, however, the molecular basis for observed aberrations is not usually understood. One important parameter in this evaluation is that of the integrity of the sperm tail, with aberrations, referred to as multiple morphological abnormalities of the flagellin (MMAF), affecting sperm motility which consequently causes impairment of fertilization. A few the underlying genetic factors responsible for MMAF in humans have been characterised (Ben Khelifa et al., 2014, Turner et al., 2001), but it has not been evaluated in cattle until recently. This may be an important parameter given that bull fertility is especially important as one bull can breed thousands of females in artificial insemination breeding programs (Thundathil, Dance and Kastelic, 2016). One such study has identified a nonsense mutation in the armadillo repeat containing 3-encoding gene (*ARMC3*) resulting in 46% of the protein being lost including domains essential for its normal functioning, causing a sperm tail disorder in Swedish Red cattle (Pausch et al.,2016). Identifying the genetic basis of sperm abnormalities would allow the early identification of infertile animals for their exclusion from breeding programs.

Idiopathic subfertility that cannot be detected through semen analysis is a factor thought to affect the deviation in insemination success among AI bulls. One genome

wide association study (GWAS) determined a strong association between male reproductive ability and chromosome 19 on the bovine genome. Further study revealed a causal nonsense mutation in the transmembrane protein 95 encoding gene (*TMEM95).* This protein is located on spermatozoa and is thought to be critical for successful fertilization (Pausch et al., 2014).

1.14 – The lethal recessive and genes with large effects analysed in this study

The following section provides information on the genes, DNA polymorphisms and associated traits that were involved in this research project. First, the milk protein genes, their chromosomal positions and biological importance are discussed. This is followed by an explanation of the lethal recessive disorders that will be considered in this project, to include an analysis of their cause at the molecular level and the resulting phenotype that leads to economic loss for the industry. Finally, the biological relevance of the Signal transducer and activator of transcription genes (1, 3 & 5) will be discussed along with the supporting evidence that these genes play a role in important phenotypic characteristics relevant to cattle breeders.

Table 1.1 - Genes and associated traits investigated in this research project

Gene	BTA	Ensembl ID*	Phenotype
SLC35A3	3	ENSBTAG00000012454	Lethal recessive – Complex vertebral Malformation (CVM)
UMPS	1	ENSBTAG00000013727	Lethal recessive – Deficiency of Uridine Monophosphate Synthase (DUMPS)
FANCI	21	ENSBTAG00000009097	Lethal recessive - Brachyspina
ITGB2	1	ENSBTAG00000017060	Unwanted – Bovine leukocyte adhesion deficiency (BLAD)
STAT1	2	ENSBTAG00000007867	Embryo survival/ Milk traits
STAT3	19	ENSBTAG00000021523	Embryo survival/ Milk traits
STAT5	19	ENSBTAG00000009496	Embryo survival/ Milk traits
DGAT	14	ENSBTAG00000026356	Fat content of milk
Kappa casein	6	ENSBTAG00000039787	Milk protein
Alpha$_1$/Alpha$_2$ Beta Casein	6	ENSBTAG00000002632	Milk protein
LFNG	25	ENSBTAG00000040361	Embryonic lethality

*UMD 3.1 assembly

1.14.1 -Milk protein genes

Since milk from domestic cows has been a stable food source for over 8000 years, the genes that code for these milk proteins have been extensively researched and a vast amount of variation has been identified. It has been found that the genetic makeup of animals has a large effect on milk yield and composition traits, however environmental factors such as nutritional status and animal management strategies also play a role (Kiplagat and Limo and Kosgey, 2012). Milk composition also varies between livestock species and is also affected by factors such as the age, the stage of lactation and the breed of the animal. The major milk proteins consist of four caseins, all of which are located within a 250kb cluster on chromosome six, often referred to as the CN locus: *CSN1S1*, *CSN2*, *CSN1S2*, *CSN3*, which code for α_{S1}-CN, β-CN, α_{S2}-CN and ϰ-CN, respectively. The two whey proteins, α-LA and β-LG are coded by the lactalbumin alpha (*LAA*) and beta-lactoglobulin (*LGB*) genes, mapped to chromosome five and chromosome 11, respectively (Caroli, Chessa and Erhardt, 2009). The analysis of variants in the casein genes, to include the estimation of their frequencies in the Irish cattle population and the evaluation of the effect of these variants on economically important traits is one objective of this study.

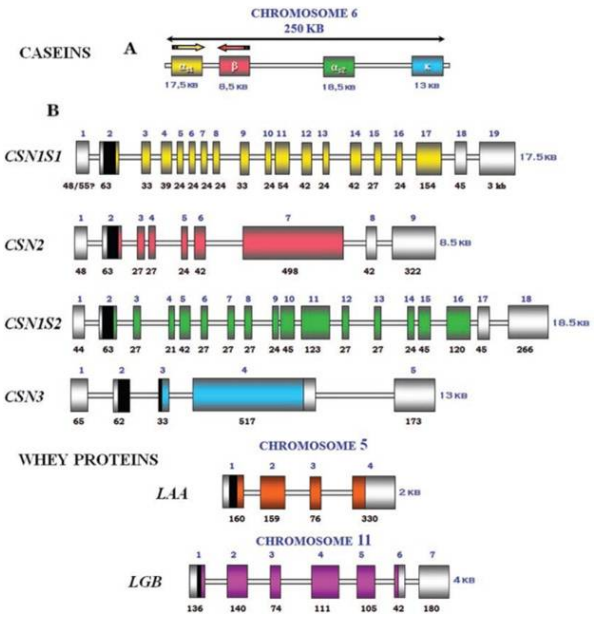


Fig 1.4– Genomic locations of the six main milk protein genes (Caroli, Chessa and Erhardt, 2009)

1.14.3 –The caseins

The caseins are phosphoproteins that constitute ~ 80% of the total protein content in bovine milk which have been classified based on homology of their primary structures (Wong et al., 1996). These proteins constitute the main dairy products as liquid milk, cheese, and yoghurt. Caseins occur in milk as a micelle structure (See Fig 1.5) with its biological function being the transport of calcium phosphate in liquid form to the neonate and to form aggregates which allow for more efficient nutrient release.

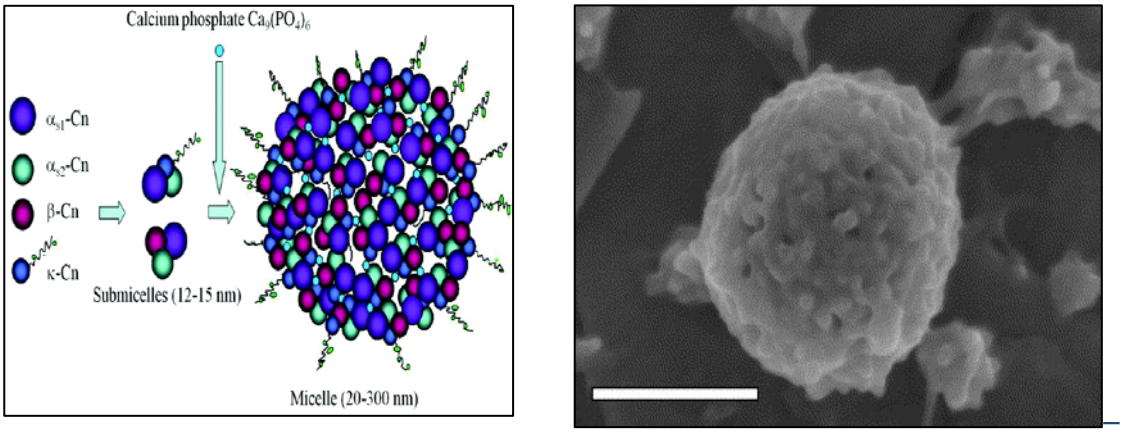


Fig 1.5 - The micelle structure of caseins found in bovine milk (left); A scanning electron microscope image capture of the micelle in bovine milk (right) (Ortega-Requena and Rebouillat, 2015; Dalgleish, Spagnuolo and Douglas Goff, 2004)

Genetic variation within the casein genes has been associated with the cheese making properties of milk due to the effects of casein micelle structure on rennet coagulation properties (Walsh et al., 1998; Aleandri et al., 1990), and has also been associated with milk composition traits in cattle (Ng-Kwai-Hang et al., 1984; Aleandri et al., 1990).

1.14.3 - The fat content of milk

The content of fat in milk is economically important for dairy farmers and nutritionally important for consumers worldwide. Factors that affect the fat composition in the milk of dairy cows include environmental conditions such as the stage of lactation and the health of the dam with mastitis being associated with a decrease in fat percentage. The diet of the animal is also correlated with fat composition because a percentage of fatty acids in milk originate from microbial activity in the rumen (Uoguelph.ca, 2017). Genetics, however, have also been found to play a part in the fat content of milk making it a trait amenable to genetic selection. Phenotypic trends as reported by the

ICBF (2016) have shown an increase of 0.45% in the fat composition of milk in the past 25 years as shown in Fig.

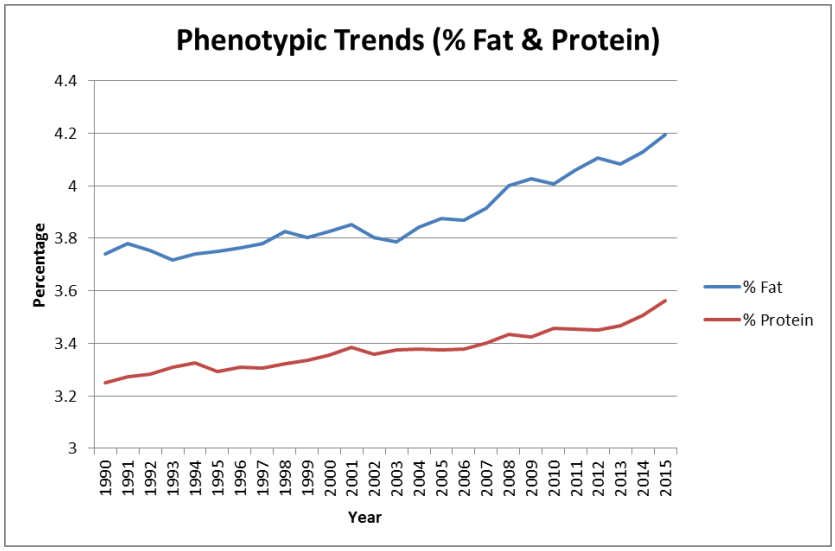


Fig 1.6 – Phenotypic trends showing an increase in the production of milk constituents from 1990-2015 (ICBF, 2016)

Studies aimed at identifying QTL associated with the fat content of milk identified genetic loci associated with this trait (Ashwell, Van Tassell and Sonstegard, 2001; Looft et al., 2001), with some suggesting the *DGAT* gene as the candidate gene associated with this trait (Kim and Georges, 2002). It was Grisart et al, 2002, who were the first to report a causative mutation in the diacylglycerol O-acyltransferase 1 (*DGAT*) gene (ENSBTAG00000026356), to be associated with the fat composition in cattle populations. The gene is located on chromosome 14 of the bovine genome and produces a protein (NP_777118.2) of 485 amino acids. The function of the transmembrane protein is as a metabolic enzyme involved in the conversion of diacylglycerol and fatty acyl CoA to triacylglycerol (Harris et al., 2011). The authors discovered that a lysine to alanine substitution at position 232 was significantly associated with a decrease in the fat composition of milk. Since then, several studies have associated this variant with fat composition in multiple cattle breeds (Spelman et al., 2002; Winter et al., 2002; Thaller et al., 2003) This research project includes the analysis of this variant in the Irish Holstein Friesian cattle population.

1.14.4 - Lethal recessives

Conventionally the only method of discovering carriers of lethal recessive diseases in livestock populations was by the birth of affected offspring after which producers had to make the decision to cull the ancestors of the affected animal rather than risk the birth of another affected animal. Molecular techniques have evolved since then allowing for the genetic testing of animals for lethal recessives with known genetic aetiology, and while these techniques were primarily used to test the carrier status of AI sires or elite bulls, decreasing costs have allowed for the genotyping of whole herds for lethal recessives genetic defects and unwanted traits that affect farm productivity. The IDB chip has included validated probes that screen for these diseases, permitting the evaluation of the frequencies of these mutations within the Irish herd, and aiding in selection decisions for carrier animals that otherwise are of a high genetic merit. As of 2016, approximately 500,000 Irish beef and dairy cattle have been genotyped using the IDB chip which represents 25% of the cattle breeding stock of this country (McClure et al., 2016). This project evaluates a range of lethal recessives and unwanted genetic defects (listed below) from the perspective of their frequency in the Irish population and their potential pleiotropic effects on important productive traits.

1.14.5 – Complex Vertebral Malformation

Complex vertebral malformation (CVM) is an autosomal recessive disorder that manifests during foetal development. It is caused by a missense mutation in the bovine *SLC35A3* gene which encodes for a UDP-N-acetylglucosamine transporter (Thomsen et al., 2006). The SLC35 proteins are enzymes responsible for transporting nucleotide sugars from the cytosol into the endoplasmic reticulum and the Golgi apparatus where they are used by glycosyltransferases to synthesise sugar chains of carbohydrate polymers, glycolipids, and glycoproteins (Song, 2013). Thomsen et al, 2006, determined the causative mutation to be a G/T transversion at position 180 which replaces valine with phenylalanine in the protein. Congenital effects observed in affected calves include malformations of the axial skeleton, with some calves showing minor malformations whereas other showed extensive lesions, malformations of the heart are also observed in up to 50% of cases (Agerholm et al., 2001). Nielson et al, 2003, reports that by gestation day 260 up to 77% of homozygous foetuses are aborted. Agerholm et al 2006 remarked on the familial pattern observed in cases and proposed that a common ancestor was likely to exist in the affected population. Further insight

proved his theory correct when the origin of this disease was traced to a common ancestral bull, Carlin-M Ivanhoe Bell, who has been used extensively in breeding programs over the last two decades due to the high milk yield observed in his daughters, resulting in the disease-causing mutation becoming prevalent among Holstein cattle worldwide (Malher, Beaudeau and Philipot, 2006). On farm, the culling of cows that have aborted calves due to CVM is often warranted as approximately half are aborted between 100-260 days after conception, making it difficult for a cow at this lactation stage to maintain a high yield. This leads to negative economic consequences due to the costs involved with replacement of culled animals (Nielson et al., 2003).

Fig 1.7 - CVM is caused by a transversion mutation (Guanine to Thymine) at position 559 of the SLC35A3 gene resulting in an amino acid substitution (Valine to phenylalanine) at position 180 in the protein

1.14.6 - Bovine leukocyte adhesion deficiency

Bovine leukocyte adhesion deficiency (BLAD) is an autosomal recessive disease which is characterised by severe neutrophil impairment, caused by a reduced expression of the heterodimeric β 2 integrin adhesion molecules on leukocytes, which leads to recurrent bacterial infections particularly of the respiratory and gastrointestinal tracts. The molecular basis, first identified by Shuster et al, 1992, is a missense mutation A->G at position 383 of the *ITGB2* gene, leading to an aspartic acid to glycine substitution at amino acid 128 within this highly conserved extracellular

region of the protein. Affected cattle typically die within the first year of life, although some may survive for longer with impaired fertility and production performance (Meydan, Yildiz and Agerholm, 2010).

1.14.7 - Deficiency of uridine monophosphate synthase

Deficiency of uridine monophosphate synthase (DUMPS) which results in embryonic death of homozygous offspring in early gestation is also inherited in an autosomal recessive manner A genetic basis was first described by Schwenger et al, 1993. The enzyme uridine monophosphate synthase is responsible for catalysing the final two steps in pyrimidine synthesis, converting orotic acid to uridine 5' monophosphate. The molecular basis for this disorder is a point mutation at position 405 of the gene which introduces a stop codon (TGA) resulting in premature termination of translation (Schwenger et al., 1993).

1.14.8 - Brachyspina

Brachyspina (BS) is a rare recessive genetic defect observed in Holstein dairy cattle. The genetic cause of BS development is a 3.3 kb deletion, removing exons 25-27 within the Fanconi anaemia complementation group I (*FANCI*) gene on chromosome 21 (Charlier et al., 2012). Necropsy findings describe the presentation of the disorder as growth retardation and malformation of the vertebral column, with histological findings showing incomplete development of intervertebral discs, or in some cases a complete absence. Abnormalities of the heart, kidneys and testicles were also evident (Agerholm, McEvoy and Arnbjerg, 2006; Testoni et al., 2008). Despite the low occurrence of BS, carrier frequency has been suggested to be as high as 7.4%. The discrepancy between carrier frequency and occurrence was examined by Charlier et al, 2012, and it was determined that approximately 50% of homozygous embryos die during early pregnancy. It was also observed that crosses between carriers and non-carriers resulted in increased pregnancy failure suggesting a polygenic effect associated with the BS mutation.

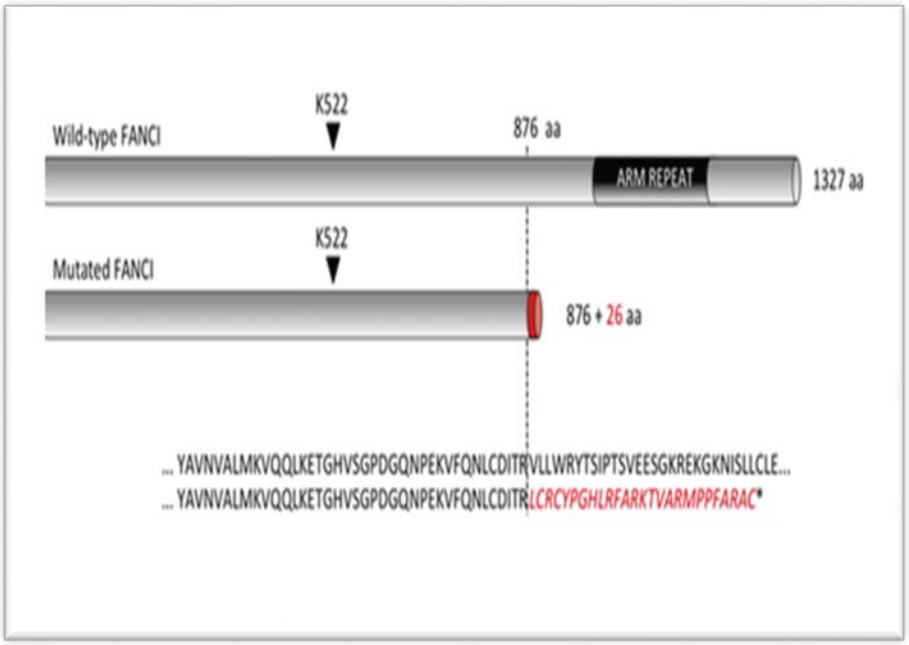


Fig 1.8 – Brachyspina is caused by the deletion of 3.3kb, removing exons 25-27 within the *FANCI* gene

1.15 - Signal Transducer and Activator of Transcription

Signal transducer and activator of transcription (*STAT*) genes code for a family of proteins involved in regulating many gene expression pathways responsible for cell growth, development, and differentiation, acting as signal transducers in the cytoplasm and transcription factors after translocation to the nucleus. They are activated by the binding of a ligand to an extracellular protein, with the different Stat proteins responding to specific extracellular ligands, and subsequent activation of tyrosine kinases which phosphorylate the *STAT*s at a single tyrosine residue located between the 700th and 850th amino acid in the sequence. Homodimer and heterodimer formations of STAT proteins then occur through SH2 interactions (Levy and Darnell, 2002). Facilitated transport to the nucleus is achieved through importing transporter proteins, for which a specific amino acid sequence functions as a nuclear localisation signal (NLS) to bind to the transporter (Iyer and Reich, 2007). Mutations in this sequence may affect nuclear import, while mutations in the DNA binding sequence can affect the stability of the Stat protein binding to promoter regions within genes (Levy and Darnell, 2002). The following paragraphs describe the research that has been done to date with regards to the *STAT* genes, consequently justifying their inclusion as genes of interest for economically important traits in cattle populations.

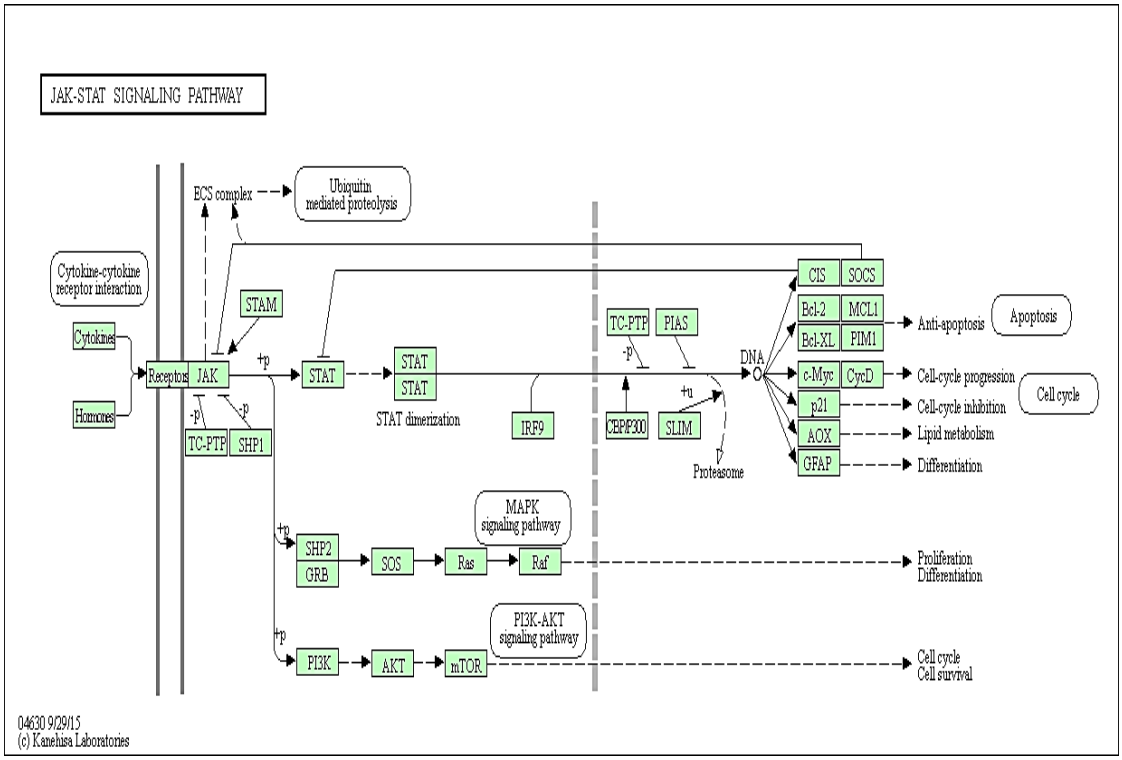


Fig. 1.9- Following the binding of cytokines to their cognate receptor, STATs are activated by members of the JAK family of tyrosine kinases. Once activated, they dimerize and translocate to the nucleus and modulate the expression of target genes. In addition to the activation of STATs, JAKs mediate the recruitment of other molecules such as the MAP kinases, PI3 kinase etc. These molecules process downstream signals via the Ras-Raf-MAP kinase and PI3 kinase pathways which results in the activation of additional transcription factors (KEGG pathway map 04630, www.genome.jp).

1.15.1 – STATs and Fertility

Many of the biological functions of the *STAT* genes have been decided through *in vivo* murine knockout models. Earlier studies showed that *STAT3* was expressed early post implantation which suggested it played a role in early embryogenesis (Duncan et al., 1997). The earliest gene knock out study was performed in 1997 by Takeda et al, where a deficiency of *STAT3* was seen to lead to embryonic lethality. Since then, studies have suggested it plays an essential role in successful implantation. In 2001 Ernst et al demonstrated that the COOH terminal of the gp130 receptor, which acts as a binding site for the Stat proteins, plays an essential role in leukaemia inhibitory factor (LIF) mediated stat signalling during blastocyst implantation. Observations have been made that progesterone (PR) activity is reliant on expression of *STAT3*, and *STAT3*/PR

complexes potentiates transduction pathways in the decidualised mesometrium in early pregnancy (Liu and Ogle, 2002). More recently Lee et al, 2013, also determined that *Stat3* and progesterone crosstalk is essential for successful implantation in the mouse uterus. *Stat3*$^{d/d}$ mice were infertile, with further analysis suggesting that this was due to implantation failure. Conditional knock out studies have shown that the absence of *STAT3* in the CNS leads to infertility with associated hypogonadism suggesting neuroendocrine involvement in reproductive ability through *STAT3* signalling mechanisms (Gao et al., 2004). An *ex vivo* model of human endothelial stromal cells (HESC) proposed a role for Interleukin-11 (IL-11) facilitated *STAT3* activation mechanisms in the initiation and progression of decidualisation (Dimitriadis et al., 2006). A similar study which looked at the effect of *STAT5* expression on decidualisation showed that its expression is important in controlling prolactin expression in differentiating endothelial stromal cells (Mak et al., 2002).

While many studies have shown *STAT3* to be involved in mechanisms ensuring successful post implantation development, cytokine mediated *STAT5* expression has been found to be correlated with preimplantation development in mice (Nakasato et al., 2006), with studies suggesting that this is due to its involvement with the maintenance of the corpus luteum through PRL and placental lactogens (PL) (Teglund et al.,1998: Curlewis et al., 2002). *STAT5* mRNA has also been found to be highly expressed in the preimplantation stage of *in vitro* bovine zygotes (Flisikowski et al., 2013).

STAT1 has been implicated in *in vitro* studies investigating the factors that control porcine ovarian function where it was shown to have an anti-proliferative effect by controlling ovarian secretory activity (Benco et al., 2009).

1.15.2 – STATs and Mammary gland development/ Milk production

Mammary gland development and function has been found to be reliant on several members of the *STAT* family (Watson and Neoh, 2008; Haricharan and Li, 2014). In fact, *STAT5* was originally identified in mammary gland tissue in response to prolactin and was initially referred to as mammary gland factor (MGF) (Wakao, Gouilleux and Groner, 1994). Gene knockout mouse studies have shown that *STAT5* controls the differentiation and proliferation of mammary alveoli during pregnancy, activated through the prolactin receptor (Miyoshi et al.,2001; Teglund et al., 1998). Down

regulation of *STAT5* expression in mammary epithelial cells (MECs) *in vitro* was also shown to impair differentiation and alveolar development (Vafaizadeh et al., 2010). *STAT3* mediates post lactational involution through leukaemia inhibitory factor (LIF) as described by Kritikou et al, 2003.

Not surprisingly, due to the known *STAT5* functions in inducing proliferation of the milk producing mammary alveoli cells, expression of milk protein genes *in vivo* has been found to be correlated with *STAT5* activation. Activated *STAT5* has been found to enhance beta casein expression *in vitro* (Happ and Groner, 1993). The effect on milk production, however, is not just a result of induced proliferation of these cells as shown in a previous *in vivo* study where the analysis of the expression of rat whey acidic protein (WAP) in transgenic mice identified a recognition site for *STAT5* within the promoter region of this gene, with mutation of this site reducing WAP expression (Li and Rosen, 1995). Further *in vitro* analysis suggested that *STAT5* works in cooperation with a specific isoform of the Nuclear factor I (NFI-B) and the glucocorticoid receptor (GR) to regulate WAP gene transcription (Mukhopadhyay et al., 2001). *STAT5* has also been seen to increases the lactational ability of dairy cow mammary gland epithelial cells (DCMECs) *in vitro.* Cells were transfected with the *STAT5* gene which increased gene expression by over three times than that of the control cells. Not only did this higher expression of the *STAT5* gene increase the proliferation of the cells, the synthesis of both beta casein and lactose increased by 20 and 18 percent, respectively (Liu et al., 2012).

1.15.3 - The association of STATs with fertility traits in Cattle

In genetic association studies, polymorphisms within the *STAT* genes and in other genes implicated in their pathways, such as the growth hormone gene and the growth hormone receptor (Mullen et al., 2011; Waters et al., 2011), have been found to be associated with milk production/composition traits in cattle populations. An extensive literature search only yielded results on one published study that investigated a possible association between a polymorphism in *STAT1* and fertility traits, but no association was observed between the SNP at position 3141 in the 3'UTR region of the *STAT1* gene and any of the fertility traits considered (Rychtarova et al, 2014). To the authors knowledge the only published study showing an association with *STAT3* and fertility in cattle is that of Khatib et al, 2009 who investigated the effects of genotypic combinations of the *STAT1* and *STAT3* genes on fertilisation rate and

embryonic survival in Holstein cattle using an *in vitro* fertilisation process. It was found that polymorphisms within these genes had a highly significant effect on embryonic survival (*STAT1/STAT3* 19069). Single SNP analysis also deciphered a significant association with polymorphisms in the STAT3 gene and fertilisation rate. In 2008, the same author performed an *in vitro* fertilisation assessment which discovered that polymorphisms in the *STAT5* gene were significantly associated with embryonic survival (C allele of SNP 12195 associated with higher embryonic survival) in Holstein cattle. Fertilisation rate was affected by genotype at specific SNP sites (for example unfertilised ova (UFO) ratio for genotype AA dams was 41% versus 30% for GG genotype at SNP3117). The embryonic lethality associated with the *STAT5* gene seems to occur much earlier than other known genetic causes of embryo death (such as CVM or DUMPS), which in theory may be regarded as failure to conceive and consequently not lengthen the calving interval significantly.

1.15.4 – The association of STATs with Milk traits in Cattle

Previous studies have detected Quantitative trait loci (QTL) for milk yield and composition traits in the region of the *STAT* genes (Ron et al., 2004, Kemper et al., 2014). Targeted imputation in these regions is permitting the identification of candidate genes associated with economically important traits in dairy cattle (Raven et al., 2015). SNP analysis has uncovered variants within the *STAT* genes that are associated with economically important traits in cattle populations. In Holstein cattle, the C allele of *STAT1* 3141 was associated with an increase in milk protein percentage, milk fat percentage and milk yield in a study performed by Cobanoglu et al, 2006. Khatib et al, 2008, discovered that the *STAT5* 12195 SNP was associated with an increase in milk yield, but a decrease in both fat and protein percentage of 0.01% with the G allele displaying dominance for the two traits. Oikonomou et al, 2011, observed the same results for the *STAT5* 12195 SNP with regards to an increase in milk yield being observed for the G allele, with the GC genotype yielding an intermediate value between the two homozygous genotypes, suggesting no dominance was at play. The same study suggests a suggestive association ($p<0.07$) with this SNP and a decrease in age at first calving by 7.2 days, with the author stating that statistical significance may not have been achieved due to the sample size. A polymorphism in intron 9 of the *STAT5* gene in Jersey cows was also associated with milk yield, fat, and protein content (Brym and Kamiñski and Rusc, 2004). Position 6853 within exon 7 of this

gene was associated with the same traits in Italian brown cattle (Selvaggi et al., 2009), and in Jersey cows (Dario and Selvaggi, 2011) with the C allele yielding higher milk, fat, and protein content.

Chapter Two –Methods

2.0 – Aims and Objectives

The aims and objectives of this research project are as follows.

Aim

To estimate the frequencies of n=18 (causal and candidate causal) DNA polymorphisms in a sample of Irish Holstein Friesian dairy cattle.

Objectives

- Process genetic data into format suitable for calculation of frequencies.
- Calculate the frequencies of the causal and candidate causal DNA polymorphism in the sample population.

Aim

To develop pipelines for processing and preparing the data obtained from the ICBF.

Objectives

- Perform deregression on the phenotypic data,
- Weighting phenotypes based on their reliabilities.
- Perform descriptive statistics for the phenotypes.

Aim

To perform association analysis using a linear mixed models approach to determine if the DNA polymorphisms being studied are associated with economically important production and functional traits in a population of dairy cattle in Ireland.

Objectives

- Use ASReml to perform association analysis between the genotypes and the phenotypic data,
- Use Phase software to construct haplotypes for the genetic data.
- Calculate the haplotype frequencies.
- Use ASReml to perform association analysis between the haplotypes and the phenotypic data,

Aim

To identify and characterise candidate DNA polymorphisms that may also have effects on fertility in Irish cattle for inclusion on subsequent versions of the IDB.

Objectives

- Perform an extensive literature search on candidate genes that may have an effect on fertility in Irish cattle.
- Develop a pipeline to extract the relevant information from the ensembl Biomart database.
- Develop a pipeline to prioritise these mutations based on their predicted consequence.

Aim

To perform bioinformatics analysis which may suggest whether the DNA polymorphisms considered in this study have a direct effect on the trait in question, or whether it is more likely that they are linked to the QTN affecting the trait.

Objectives

- To utilise software tools, such as PolyPhen and Sift to estimate the effect of the mutations location within the gene (i.e regulatory, intronic, exonic regions).

2.1 – Phenotypes Processing

Predicted transmitting abilities (PTAs) on n=16 routinely recorded traits across 21,707 Irish Holstein Friesian dairy cattle were obtained from the ICBF. Deregression was performed and simultaneous weighting of the phenotypes based on their reliabilities was also performed as per Garrick et al, 2009. Two datasets were produced – one based on an adjusted reliability cut off of 0.1 and a second based on an adjusted reliability cut off of 0.2.

2.2 – Summary statistics for phenotypic data

Descriptive statistics, which included the mean, standard deviations and variances were analysed in R studio (See script in appendix II, page 169). The Anderson darling test was used to test normality of the data. Graphs were produced in R allowing the

visualisation of the distribution of the raw phenotypes, before deregression and removal of parental contributions.

Table 2.0 - The 16 traits examined in this research study

Trait	Definition
Milk traits	
Milk yield	The quantity of milk produced in kilograms (kg). The ICBF use parities 1-10 in genetic evaluations. The cow must have a known sire and produce a minimum of 1500 kg in 305 days.
Milk fat yield	The quantity of fat in milk measured in kilograms (kg). The ICBF use parities 1-10 in genetic evaluations. The cow must have a known sire and produce a minimum of 40 kg in 305 days.
Milk protein yield	The quantity of protein, in kilograms, contained in milk. The ICBF use parities 1-10 in genetic evaluations. The cow must have a known sire and produce a minimum of 40 kg in 305 days.
Milk fat percentage	Measured as a percentage of the total quantity of milk produced each year.
Milk protein percentage	Measured as a percentage of the total quantity of milk produced each year.
Fertility traits	
Calving interval	The time in days between consecutive calving's in a cow.
Survival	Whether the calf is dead or alive at 28 days. Combined with the trait for calf perinatal mortality (calf mortality shortly before, during and up to 48 h after parturition), in genetic evaluations.
Calving difficulty	The difficulty in calving due to the genes of the calf. Numerical values for this trait range from 1 (easy) to 5 (veterinary assistance). Sire must be known.
Gestation length	The number of days between conception and calving date. Sire must be known.
Mortality	Calf mortality shortly before, during and up to 48 h after parturition.
Maternal calving difficulty	Difficulty in calving due to the maternal genes of the dam. Numerical score from 1(easy) to 4 (veterinary assistance).
Carcass traits	
Carcass weight	The cold weight of the carcass taken within 2 hours of slaughter after removal of the limbs, skin, external genitalia, head, tail, kidneys, and udder. Measured in kilograms (kg). Cow must have a known sire.
Carcass fat	The quantity of fat on the carcass of a slaughtered animal. Measured through the mechanical grading of cattle carcasses (VBS2000 technology), with phenotypes scaled from 1(leanest) to 15(fattest). Cows must have a known sire.
Carcass conformation	Thickness of muscle on a slaughtered animal. Measured through the mechanical grading of cattle carcasses (VBS2000 technology).
Culled carcass weight	The carcass weight of a cow slaughtered for meat.
Health	
Somatic cell score	Log base 2 of number of somatic cells (leukocytes) within the milk of a cow.

2.3 – Genotypes Processing

The following table lists all SNPs analysed and their corresponding IDB probe number.

Table 2.1 - List of candidate research SNPs analysed in this study

IDB code	Gene	BTA	Chromosomal location*	HGVS_protein
IDBV20100000193	*UMPS*	1	698756880	p.Arg405X
IDBV22100003530	*FANCI*	21	21184869	p.Val876Leufs26X
IDBV20300000706	*SLC35A3*	3	43412427	p.Val180Phe
IDBV20200000628	*STAT1_2697*	2	79888611	-
IDBV21900002527	*STAT3_19069*	19	43070296	-
IDBV21900002523	*STAT3_25042*	19	43063963	-
IDBV21900002492	*STAT5_3117*	19	43036729	-
IDBV21900002499	*STAT5_12195*	19	43045807	-
IDBV21900002505	*STAT5_13244*	19	43046856	-
IDBV21900002510	*STAT5_13319*	19	43046931	-
IDBV21900002515	*STAT5_13516*	19	43047128	-
IDBV20600001270	*Kappa casein*	6	87390198	p.Arg31His
IDBV20600001276	*Kappa casein*	6	87390448	p.Thr114Thr
IDBV20600001459	*Kappa casein*	6	87390670	p.Thr188Thr
IDBV20600001462	*Kappa casein*	6	87390673	p.Ala189Ala
IDBV20600001205	*Kappa casein*	6	87181501	p.His121Gln
IDBV21400002069	*DGAT1*	14	1802264	p.Ala232Lys
IDBV20600001247	*A1/A2 Beta casein*	6	87181619	p.Pro82His

*UMD 3.1 *Bos Taurus* assembly

2.3.1 – Minor allele frequencies

The allele frequencies for all SNPs were calculated in excel.

2.3.2 – Hardy Weinberg Equilibrium (HWE)

HWE was calculated in excel using the chi squared goodness of fit test.

$p^2 + 2pq + q^2 = 1$

where p is the frequency of the dominant allele, and q is the frequency of the recessive allele; therefore, p^2 represents homozygous dominant, q^2 represents homozygous recessive and 2pq represents the heterozygous genotype.

$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

H_0 = There is no significant difference between observed and expected genotypic frequencies

H_A= there is a significant difference between the observed and expected genotypic frequencies

2.3.3 – Linkage disequilibrium analysis

Pairwise LD was calculated using the software programme, Haploview, was utilised to perform LD calculations (r^2 and D'). The software calculates several pairwise measures of LD which are then used to create a graphical representation **(**Barrett et al., 2004).

2.4 – Association analysis

Linear mixed models are statistical models that assume normal distribution and include the incorporation of both fixed effects, which may account for biological effects that may bias the results such as the pedigree of the animal, and random effects, such as the genotype of the animal (Bolker et al., 2009). This statistical approach has the advantage of being able to handle missing data, which is advantageous in the analysis of the data being analysed in this study (Wang and L. A. Goonewardene, 2004).

Univariate SNP and haplotype analysis was performed using a weighted mixed animal model in ASREML (Gilmour et al., 2009). In this animal model, genotyped individuals were included as a random effect and the average expected relationships amongst animals was accounted for through the numerator relationship matrix, which was generated using six generations of back pedigree. Percent Holstein of the animal was included as a fixed effect in the model. The dependant variable was de-regressed PTA weighted by their respective reliabilities. Genotype/haplotype was included in the analysis as a continuous variable coded as the number of copies of a given allele.

Nominal P values are presented unless otherwise noted.

2.5 – Haplotype Analysis

PHASE v2.1.1 was utilised to reconstruct haplotypes from the data available (Stephens, Smith, and Donnelly, 2001). Haplotypes were calculated for the six SNPs located in *STAT*5 and *STAT*3 located on chromosome 19.

2.6 – Bioinformatics analysis

SNPs within exonic regions were analysed using PredictSNP which is a consensus classifier that was designed using the six software tools, Polyphen, Sift, MAPP, PANTHER and SNAP. SNPS known to occur in regulatory regions were analysed using the RegRNA software described in the previous chapter.

2.7 – Comparative Genomics

To facilitate the ICBF with the advancement of the IDB chip, with version 4 due to be released in early 2019, a comprehensive literature review was undertaken to identify genes and mutations that have been implicated in fertility and embryo development in several species. Databases, including PubMed and science direct, were used to search for keywords in the identification of these genes. From this, an excel sheet which included the gene names, ensembl numbers and the phenotype was composed (See appendix II, Page 187, Fig. 5). Genes were prioritised for addition to the IDB chip as follows.

- Candidate genes – deleterious mutations within the candidate genes of focus in this study were top priority
- Based on the amount of journal articles and different species the mutation was found to be significant in for the fertility phenotypes of interest in this study
- Genes that have been associated with fertility in cattle populations in previous studies
- Genes associated with embryonic lethality in mammalian species
- Male and female fertility parameters – these were prioritised based on the amount of journal articles they were referenced in

Subsequently, the programming language R, was utilised to filter through the ensemble Biomart database to extract the relevant information pertaining to the above generated list. The R script utilised can be found in the appendix II, Page 185. This script takes the gene list of ensemble IDs as obtained from the literature review accesses of the Biomart database and pulls all mutations from the database that match the gene ID, while also pulling other relevant information (variant alleles, Sift score and consequence), and outputting all this to a readable file to allow further analysis.

The prioritisation of SNPs was based on the predicted consequence of the mutation referred to 'high' and 'moderate' on the ensemble database, described below.

- High = Transcript ablation, splice acceptor variant, splice donor variant, stop gained, frameshift variant, stop lost, start lost
- Moderate = inframe insertion, inframe deletion, missense variant, protein altering variant

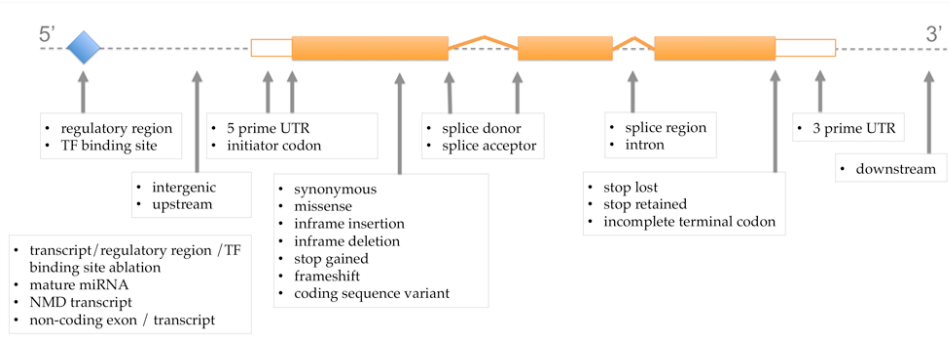


Fig 2.0 – The locations of each of the mutations for consideration for addition to the IDBv4 (Ensembl.org, 2017)

A python script (See appendix II, Page 185 and 186) was then utilised to prioritise the candidate gene and research gene mutations based on their high or moderate consequence as well as the number of mutations to be pulled per gene. The flanking sequences of these SNPs were accessed from the Bovine genome assembly database UMD 3.1. Computational limitations required that this step was performed using a cluster at UCD. A template was prepared for submission to the ICBF to include all relevant information.

2.8 – The identification of a mutation with candidate novel consequences on fertility in Holstein Friesian cattle

A literature review highlighted a possible connection with a SNP in the *LFNG* gene and fertility in mammalian species. As there was genotype data available on 10,707 cattle for this SNP, an association study was carried out in ASReml for this SNP and the traits included in this study.

Chapter Three - Results

3.0 - Phenotype summary statistics

Heritability is an important factor in animal breeding through its use in helping plan breeding programs, estimated breeding values and predicting response to selection. Traits related to fertility tend to have low heritability whereas production traits tend to have higher heritability values. Table 3.0 represents the heritability of each trait, values which were obtained from the ICBF (Ross Evans, personal communication), analysed in this study with values ranging from 0.01 for survival to 0.38 for carcass weight. Summary statistics for all traits analysed in this study are presented in Table 3.1 and Table 3.2. The results presented here are for the data after deregression (removal of parental contributions) and weighting. Fig. 3.0 to Fig. 3.2 graphical representations of the distribution of all traits in the 21k dataset of Irish Holstein Friesian cattle analysed. Table 3.3 represents the variance values after deregression and removal of parental contributions in the 0.1 and 0.2 reliability cut off datasets. The samples sizes remaining in the 0.1 and 0.2 datasets can be viewed in Table 3.4.

Table 3.0 – Heritability values were obtained from the ICBF for phenotypes examined in this study

Trait	Heritability
PPC	0.35
FPC	0.35
MKG	0.35
FKG	0.35
PKG	0.35
CIV	0.02
SU	0.01
CD	0.09
GEST	0.36
MORT	0.04
MCD	0.04
CWT	0.38
CCF	0.33
CFT	0.30
CCWT	0.29
SCS	0.15

Table 3.1– Summary statistics for the milk and fertility phenotypes (Raw data before deregression and removal of parental contributions and weighting).

	PPC	FPC	MKG	FKG	PKG	CIV	SU	CD	GEST	MORT	MCD
Minimum	-0.49	-1.02	-1638.50	-59.46	-61.10	-51.83	-13.74	-8.42	-8.18	-9.26	-6.63
1st quartile	-0.05	-0.07	-252.89	-5.06	-6.15	-6.73	2.47	1.58	-1.83	-2.46	4.42
Median	0.02	0.06	-44.16	2.21	-0.16	2.56	6.97	3.82	0.00	-0.58	6.13
Mean	0.02	0.07	-42.85	2.28	-0.18	7.41	6.61	5.29	0.55	0.65	6.52
3rd Quartile	0.09	0.19	165.86	9.71	6.01	13.26	10.85	7.56	2.38	2.75	8.20
Max	0.51	1.11	1500.54	71.23	57.08	147.37	26.59	46.69	39.11	50.83	51.65
Variance	0.01	0.03	99312.24	131.69	88.79	775.17	39.65	34.38	13.17	30.25	13.51

PPC = Protein Percentage, FPC = Fat percentage, MKG = Milk (kg), PKG= Protein (kg), CIV = Calving Interval, SU= survival, CD= Calving Difficulty, GEST = Gestation Length, MORT =Mortality, MCD = Maternal Calving Difficulty.

Table 3.2 - Summary statistics for the carcass and health phenotypes (Raw data before deregression and removal of parental contributions and weighting)

	CWT	CCF	CFT	CCWT	SCS
Minimum	-152.37	-13.03	-3.15	-266.85	-0.79
1st quartile	-29.72	-1.76	-0.84	-33.69	-0.19
Median	-11.55	-1.16	-0.37	-18.76	-0.06
Mean	-16.79	-1.64	-0.37	-23.71	-0.04
3rd Quartile	0.98	-0.65	0.10	-7.29	0.08
Max	49.46	1.29	1.91	37.79	1.37
Variance	759.10	4.25	0.56	757.42	0.04

CWT =Carcass Weight, CCF=Carcass Conformation, CFT =Carcass Fat, CCWT =Culled Carcass Weight, SCS = Somatic Cell Count

Table 3.3– Variance values after deregression and removal of parental contributions in the 0.1 and 0.2 relibaility cut off datasets

Trait	**Phenotypic variance 0.1**	**Phenotypic variance 0.2**
PPC	0.129	0.011
FPC	0.139	0.037
MKG	0.314	0.138
FKG	0.314	0.138
PKG	0.314	0.138
CIV	16.8	5.11
SU	58.7	164.80
CD	0.89	1.16
GEST	0.290	0.212
MORT	9.38	16.56
MCD	6.68	7.78
CWT	0.063	0.059
CCF	0.082	0.128
CFT	0.098	0.180
CCWT	0.362	0.357
SCS	0.193	0.216

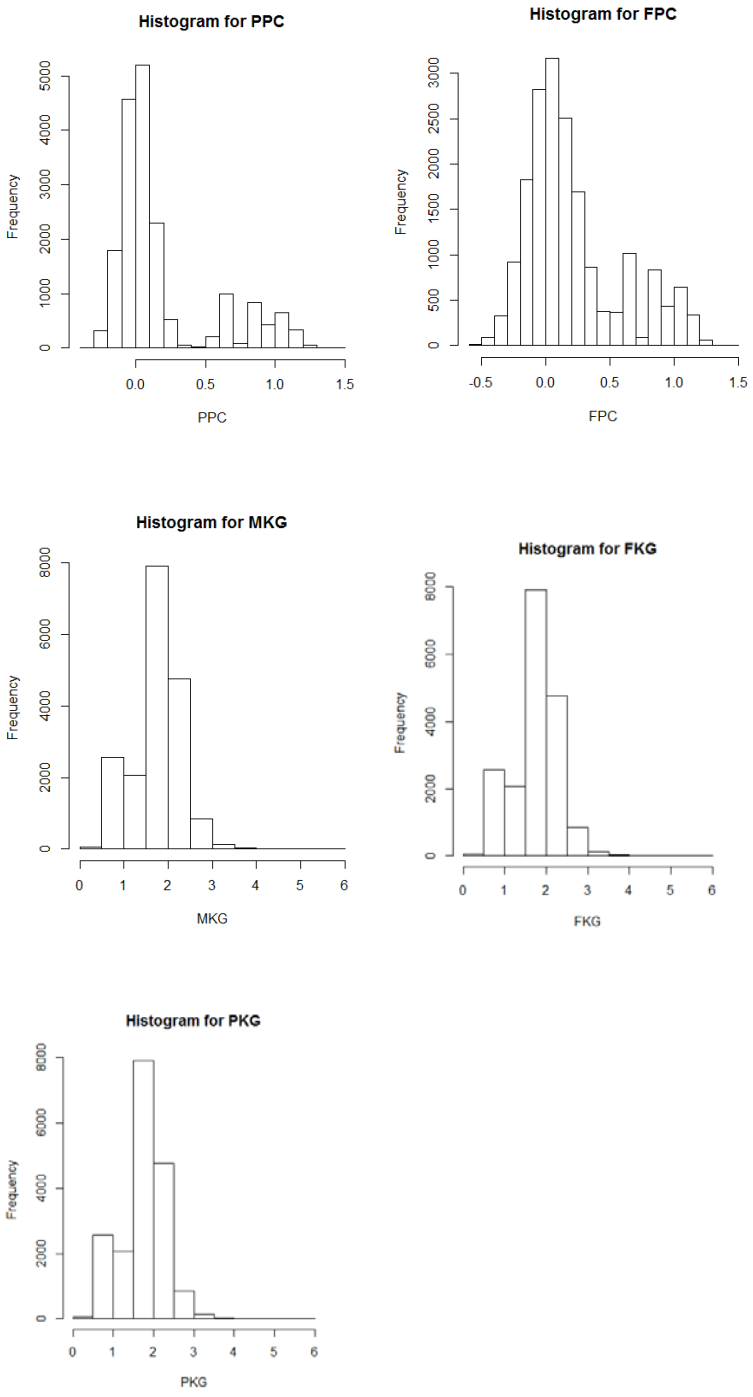


Fig. 3.0 – Graphical representation of the distribution of milk production and composition traits in the 21k dataset of Irish Holstein Friesian cattle. *(PPC = Protein Percentage, FPC = Fat percentage, MKG = Milk (kg), PKG= Protein (kg))*

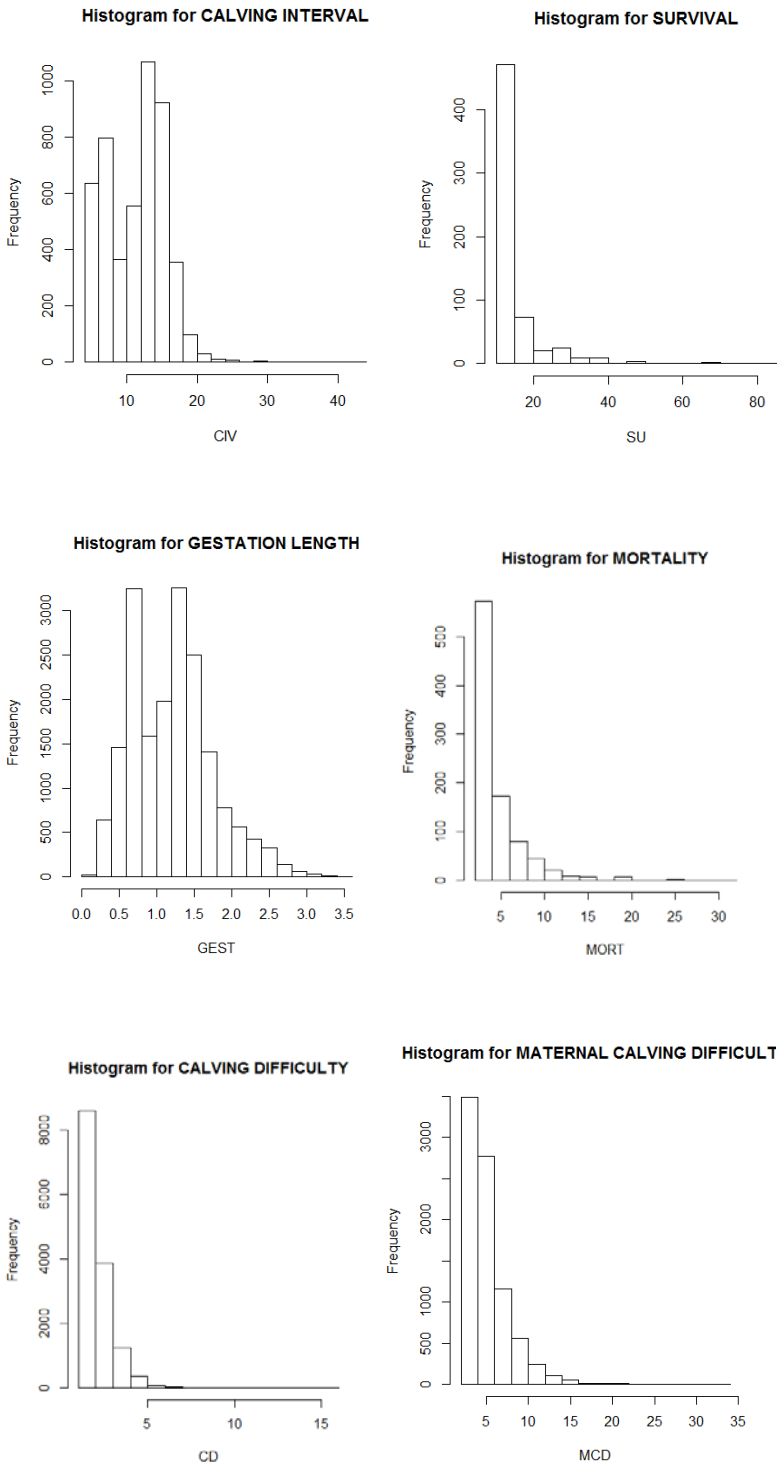


Fig. 3.1 - Graphical representation of the distribution of fertility traits in the 21k dataset of Irish Holstein Friesian cattle

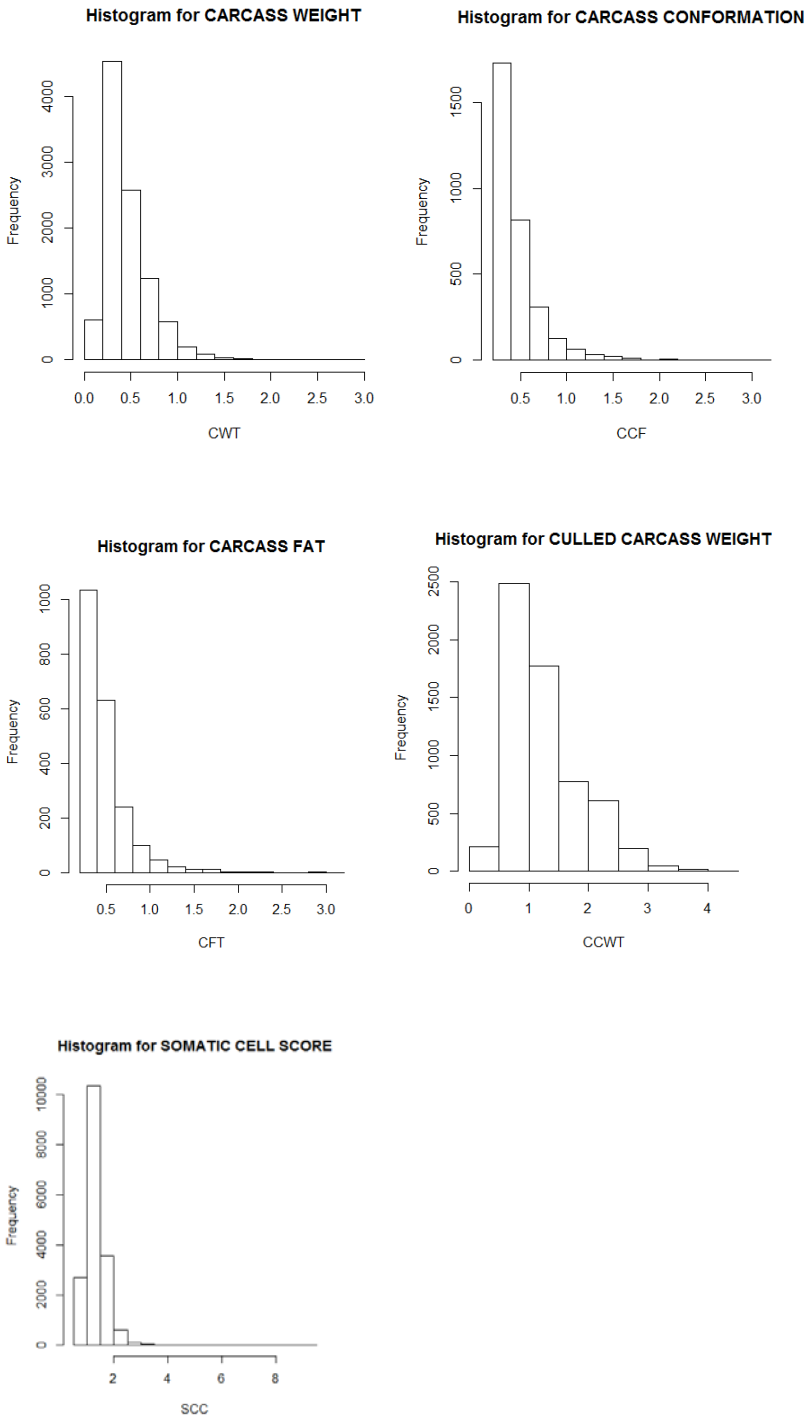


Fig. 3.2 - Graphical representation of the distribution of carcass and health traits in the 21k dataset of Irish Holstein Friesian cattle

Table 3.4 – Sample sizes remaining after deregression and adjustment for reliability at 0.1 and 0.2 cut off

Trait	**Adjusted Reliability cut off**	
	0.1	**0.2**
PPC	18450	18423
FPC	18450	18423
MKG	18450	18423
FKG	18450	18423
PKG	18450	18423
CIV	4917	2712
SU	623	56
CD	14457	3653
GEST	18552	17737
MORT	931	187
MCD	8543	2450
CWT	9983	4981
CCF	3152	978
CFT	2138	563
CCWT	6172	5913
SCC	17609	6751

3.2 -Summary statistics – Lethal recessives

The following table represents the results obtained in the calculation of the frequencies of the mutations responsible for the lethal recessive disorders being analysed in this study. Mutations responsible for DUMPS, Brachyspina and CVM were found to be 0.0004, 0.01 and 0.01, respectively. Are there MAF quoted below the full dataset MAF or one of the 0.1 or 0.2 ??

Table 3.6 - Frequencies of the genetic mutations responsible for the lethal recessive disorders being examined in this study

Gene	Genetic mutation	Protein consequence	Phenotype	OMIA	Minor allele frequency (%)
UMPS	c.1213C>T	p.Arg405X	DUMPS	000262-9913	0.02
FANCI	Deletion of exons 25-27	p.Val876Leufs26X	Brachyspina	000151-9913	1
SLC35A3	c.538G>T	p.Val180Phe	CVM	001340-9913	1.7

3.3 -Association analysis –Lethal recessives disorders

The following section reports on the results obtained in the analysis of the impact of carrier status of the lethal recessive mutations on the production and functional traits included in this study. Results are presented at reliability cut offs of 0.1 and 0.2. Results could not be obtained for carriers of the *UMPS* mutation since carriers of the causative allele did not meet the reliability cut off values.

For carriers of the mutation responsible for Brachyspina, the following table represents the results obtained in the association analysis at 0.1 reliability. Results suggest that carriers of this mutation display a decrease in the protein percentage of milk, yet an increase in the quantity of milk produced. This result was consistent for both adjusted reliability cut offs.

Table 3.3.1– Association between the Brachyspina mutation and production traits in Irish Holstein Friesian cattle (Adjusted reliability cut off-0.1)

Trait	Effect size	P value	Phenotypic variance (%)
PPC	-1.72×10^{-2} (5.0×10^{-3})	0.0006	0.0010
MKG	32.23 (15.78)	0.0400	0.0004

Table 3.3.2 - Association between the Brachyspina mutation and production traits in Irish Holstein Friesian cattle (Adjusted reliability cut off-0.2)

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-1.66×10^{-2} (5.02×10^{-3})	0.0009	0.0010
MKG	32.86 (15.86)	0.0300	0.0004

For carriers of the mutation responsible for CVM, a decrease in both the protein and fat percentage was observed. This was consistent for both reliability cut offs, however

an association between this mutation and survival was observed for the 0.2 cut off value.

Table 3.3.3 - Association between the CVM mutation and production traits in Irish Holstein Friesian cattle (Adjusted reliability cut off-0.1)

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-1.65×10^{-2} (3.78×10^{-3})	1.33×10^{-5}	0.0020
FPC	-1.75×10^{-2} (7.11×10^{-3})	0.01	0.0006

Table 3.3.4 - Association between the CVM mutation and production and functional traits in Irish Holstein Friesian cattle (Adjusted reliability cut off-0.2)

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-1.62×10^{-2} (3.77×10^{-2})	1.65×10^{-5}	0.0020
FPC	-1.70×10^{-2} (7.10×10^{-3})	0.01	0.0006
SU	4.68 (2.02)	0.02	0.0490

3.4 – Summary statistics for the *STAT* genes

The following table represents the results obtained in the calculation of the summary statistics relating to the *STAT* genes

Table 3.4.1 - Summary statistics obtained for the seven *STAT* genes analysed in this study

SNP	BTA	SNP gene Position	Allele substitution	Genotype	Genotype Frequency	MAF	HWE[1]	Rs Identifier[2]	Region
STAT1 2697	2	79888611	A→G	G/G G/A A/A	0.06 0.37 0.57	0.25	1.000	rs43705173	3' UTR
STAT3 19069	19	43070296	T→C	T/T T/C C/C	0.14 0.46 0.40	0.37	0.999	rs110942700	Exon 13 Synonymous variant
STAT3 25402	19	43063963	T→G	T/T T/G G/G	0.10 0.43 0.46	0.32	0.999	rs134279188	Intron 20
STAT5 13516	19	43047128	T→G	T/T T/G G/G	0.20 0.49 0.31	0.42	0.999	rs110495396	Intron 9
STAT5 13319	19	43046931	G→A	G/G A/G A/A	0.90 0.10 0.00	0.04	0.996	rs208753173	Intron 9
STAT5 13244	19	43046856	A→G	A/A A/G G/G	0.20 0.49 0.31	0.42	1.000	rs109788842	Intron 9
STAT5 12195	19	43045807	C→G	C/C C/G G/G	0.16 0.48 0.36	0.4	1.000	rs137182814	Exon 8 Synonymous variant/ splice region variant

[1]Hardy Weinberg equilibrium; [2] DbSNP

Association analysis – *STAT* genes

The following table shows the significant results obtained in the analysis for the *STAT1* (A→G) gene variant. No significant results were observed at the 0.1 cut off, however a tentative association between somatic cell score and this variant was observed at the 0.2 reliability cut off.

Table 3.5.1– Association between *STAT1*` (A→G) and somatic cell score at the 0.2 adjusted reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
SCC	7.19×10^{-3} (4.16×10^{-3})	0.08	0.015

A number of significant associations were observed between the *STAT3* 19069 (T→C) variant and production and functional traits. These results were consistent at both the reliability cut off points.

Table 3.5.2 – Association between *STAT3* 19069 (T→C) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
FPC	-4.99×10^{-3} (2.20×10^{-3})	0.02	0.010
MKG	7.34 (3.69)	0.04	0.010
PKG	0.77 (0.28)	0.01	0.140
GEST	-6.25×10^{-2} (2.99×10^{-2})	0.03	0.006
CCWT	0.89 (0.37)	0.01	0.020

Table 3.5.3 – Association between *STAT3* 19069 (T→C) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
FPC	-5.18×10^{-3} (2.20×10^{-3})	0.010	0.010
MKG	8.14 (3.69)	0.020	0.010
PKG	0.30 (0.11)	0.007	0.020
GEST	-5.10×10^{-2} (2.94×10^{-2})	0.080	0.004
CCWT	0.88 (0.36)	0.010	0.020

A number of associations are presented below for the analysis between *STAT3* 25402 (T→G) and the production and functional traits analysed. Associations between this variant and protein percentage, fat percentage, milk (kg) and gestation length were

observed for both cut offs, however, calving difficulty was also significant at the 0.1 cut off. Protein (kg) and somatic cell score were also observed at the 0.2 cut off.

Table 3.5.4– Association between *STAT3* 25402 (T→G) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-4.49×10^{-3} (1.20×10^{-3})	0.0001	0.030
FPC	-1.05×10^{-2} (2.27×10^{-3})	3.74×10^{-6}	0.050
MKG	10.65 (3.81)	0.0050	0.020
CD	-9.91×10^{-2} (5.86×10^{-2})	0.0900	0.005
GEST	-6.71×10^{-2} (3.07×10^{-2})	0.0200	0.006

Table 3.5.5 – Association between *STAT3* 25402 (T→G) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-4.52×10^{-3} (1.20×10^{-3})	0.0001	0.030
FPC	-1.06×10^{-2} (2.27×10^{-3})	2.74×10^{-6}	0.040
MKG	11.20 (3.80)	0.0030	0.020
PKG	0.103 (0.11)	0.0900	0.007
GEST	-5.14×10^{-2} (3.02×10^{-2})	0.0800	0.003
SCS	-8.12×10^{-3} (4.03×10^{-3})	0.0400	0.020

The *STAT5* 12195 (C→G) variant was found to be associated with calving difficulty at the 0.1 reliability cut off (p=0.03).

Table 3.5.6 – Association between *STAT5* 12195 (C→G) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
CD	$-0.11(5.36 \times 10^{-2})$	0.030	0.008

The *STAT5* 13244 (A→G) was associated with calving difficulty at the 0.1 reliability cut off and carcass fat at the 0.2 cut off.

Table 3.5.7 – Association between *STAT5* 13244 (A→G) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
CD	-0.11 (5.36×10^{-2})	0.030	0.008

Table 3.5.8 – Association between *STAT5* 13244 (A→G) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
CFT	-5.78×10^{-2} (3.29×10^{-2})	0.070	0.130

Several significant associations were observed with the *STAT5* 13319 (G→A) variant and the production and functional traits analysed. These results were consistent between the 0.1 and 0.2 reliabilities.

Table 3.5.9 – Association between *STAT5* 13319 (G→A) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	1.85×10^{-2} (2.45×10^{-3})	5.62×10^{-14}	0.02
FPC	2.21×10^{-2} (4.63×10^{-3})	1.80×10^{-6}	0.009
FKG	0.77 (0.29)	0.0070	0.003
PKG	0.47 (0.24)	0.0400	0.002
MCD	-0.23 (0.12)	0.0700	0.003
CWT	2.16 (0.65)	0.0009	0.004
CCWT	1.98 (0.78)	0.0100	0.004

Table 3.5.10 – Association between *STAT5* 13319 (G→A) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	1.90×10^{-2} (2.45×10^{-3})	1.12×10^{-14}	0.020
FPC	2.18×10^{-2} (4.63×10^{-3})	2.46×10^{-6}	0.009
FKG	0.75 (0.29)	0.008	0.003
PKG	0.50 (0.24)	0.030	0.002
MCD	-0.50 (0.18)	0.005	0.010
CWT	2.44 (0.76)	0.001	0.006
CCWT	2.10 (0.76)	0.005	0.004

An association between *STAT5* 13516 (T→G) and calving difficulty was observed for the 0.1 reliability cut off whereas an association between this variant and carcass fat was observed for the 0.2 cut off.

Table 3.5.11 – Association between *STAT5* 13516 (T→G) and production and functional traits at 0.1 reliability cut off

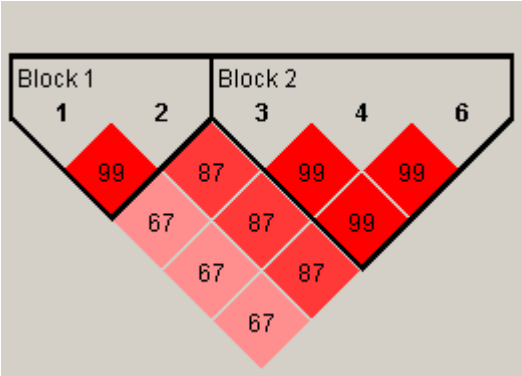
Trait	Effect size	P value	Phenotypic Variance (%)
CD	-0.11 (5.36×10^{-2})	0.030	0.008

Table 3.5.12 – Association between *STAT5* 13516 (T→G) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
CFT	-5.78×10^{-2} (3.29×10^{-2})	0.07	0.13

3.6 -Linkage disequilibrium analysis

The following figure depicts the results obtained in the analysis of linkage disequilibrium using Haploview 4.1 software. As can be observed, many SNPs show high LD.



Gene variant	Genomic location
1 – *STAT3* 19069	43070296
2 – *STAT3* 25402	43063963
3 – *STAT5* 13516	43047128
4 – *STAT5* 13244	43046856
5 – *STAT5* 13319	43046931
6 – *STAT5* 12195	43045807

Fig.3.3– Output from Haploview depicting R^2 values for SNPs located in *STAT3* and *STAT5* on BTA19

#	ObsHET	PredHET	HWpval	%Geno	FamTrio	MendErr	MAF	Alleles
1	0.456	0.465	0.0053	100.0	0	0	0.368	C:T
2	0.428	0.436	0.0094	100.0	0	0	0.321	G:T
3	0.476	0.48	0.2171	100.0	0	0	0.401	G:T
4	0.476	0.48	0.1726	100.0	0	0	0.401	G:A
5	0.092	0.088	2.6296E-21	99.8	0	0	0.046	G:A
6	0.476	0.481	0.2114	100.0	0	0	0.402	G:C

Fig. 3.4 – Haploview output for the six variants located in STAT3 and STAT5 on BTA 19

3.7 – Haplotype Analysis –*STAT5*

Analysis of haplotype frequencies for the SNPs located in the *STAT3* and *STAT5* genes on BTA19 revealed that there were 20 haplotypes, although many of these were at very low frequencies as can be observed in the table below.

Table 3.7.1– Haplotype frequencies in the Irish Holstein Friesian population studied for the *STAT3* and *STAT5* SNPs located on BTA19

Haplotype	**Frequency**
H1	0.290
H2	<0.001
H3	<0.001
H4	0.100
H5	<0.001
H6	<0.001
H7	<0.001
H8	<0.001
H9	<0.001
H10	<0.001
H11	<0.001
H12	<0.001
H13	<0.001
H14	<0.001
H15	<0.001
H16	<0.001
H17	<0.001
H18	0.023
H19	<0.001
H20	<0.001

The following tables represent all results obtained in the association analysis between haplotypes of high or moderate frequencies; H1, H4 AND H18 with production and functional traits in Irish Holstein Friesian cattle. H1 was observed at a frequency of 29% and consists of the SNP *STAT5* 12195. H4 was observed at a frequency of 10% and consists of SNPs *STAT5* 13244, *STAT3* 25402 and *STAT3* 19069. H18 was observed at a frequency of 2.3% and consists of all the six SNPs analysed on BTA19.

H1 was significantly associated with a decrease in protein percentage and milk yield. An increase in gestation length and somatic cell count was also observed in the animals in which this haplotype was segregating. H4 was associated with an increase in protein percentage and milk yield whereas a decrease in fat yield was observed in animals carrying this haplotype. A decrease in carcass fat, an increase in culled cow carcass weight and somatic cell count was also observed. Animals carrying the H18 haplotype exhibited an increased measure of protein and an increased fat yield, whereas a reduction in calving difficulty, gestation length and maternal calving difficulty was observed.

Table 3.7.2 – Association analysis results between H1 and production and functional traits in Irish Holstein Friesian cattle

Trait	**Effect size**	**P value**
PPC	-0.13(0.59)	5.8×10^{-106}
MKG	-6.43(3.84)	0.09
GEST	$0.53 \times 10^{-1}(0.30 \times 10^{-1})$	0.08
SCS	$0.53 \times 10^{-2}(0.28 \times 10^{-2})$	0.06

Table 3.7.3– Association analysis between H4 and production and functional traits in Irish Holstein Friesian cattle

Trait	**Effect size**	**P value**
PPC	$0.38 \times 10^{-1}(0.12 \times 10^{-2})$	2.1 E-207
MKG	20.78 (5.36)	0.0001
FKG	-0.45 (0.19)	0.0200
CFT	-9.60 (3.93)	0.0100
CCWT	$0.13 (0.24 \times 10^{-1})$	2.36×10^{-8}
SCS	$1.00 (0.33 \times 10^{-2})$	<0.0001

Table 3.7.4 – Association analysis between H18 and production and functional traits in Irish Holstein Friesian cattle

Trait	Effect size	P value
PPC	0.93×10^{-1} (0.83×10^{-3})	<0.01
FKG	0.22 (0.13)	0.08
CD	-0.13 (0.53×10^{-1})	0.01
GEST	-0.51×10^{-1} (0.28×10^{-1})	0.06
MCD	-0.10 (0.59×10^{-1})	0.08

3.8 – Summary Statistics – Milk protein genes

The following table provides the summary statistics on the milk protein genes that were significant for associations with the phenotypic traits analysed in this study.

Table 3.8.1 – Summary statistics for the milk protein genes included in this study

SNP	BTA	SNP gene Position	Genotype	Genotype Frequency	MAF	Rs Identifier[2]	Region
CSN3 Kappa casein	6	87390632	AA AG GG	0.940 0.060 0.001	0.03	rs43703017	Exon 4
CSN2 Beta casein a1/a2	6	87181619	GG GT TT	0.160 0.470 0.370	0.39	rs43703011	Exon 7
DGAT1	14	1802264	GG GA AA	0.350 0.480 0.160	0.39	rs473009810	Exon 9

3.9 – Association Analysis – Milk protein genes

The following table represents the results obtained in the association analysis for the variant in the kappa casein gene. The results were consistent for both reliability cut offs for protein percentage, milk production (kg), calving interval and culled carcass weight. At the 0.1 reliability cut off carcass weight and carcass conformation were associated with this gene variant, whereas at the 0.2 reliability cut off calving difficulty, gestation length and carcass fat were also evident.

Table 3.9.1 - Association between *CSN3* (Kappa casein) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-1.25×10^{-2} (3.02×10^{-3})	5.04×10^{-5}	0.0001
MKG	16.77 (9.50)	0.0700	2.83×10^{-5}
CIV	3.70 (1.17)	0.0010	0.0001
CWT	-1.47 (0.80)	0.0600	2.84×10^{-5}
CCF	-0.33 (8.48×10^{-2})	0.0001	0.0002
CCWT	-2.20 (0.95)	0.0100	6.39×10^{-5}

Table 3.9.2 - Association between *CSN3* (Kappa casein) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-1.15×10^{-2} (3.03×10^{-3})	0.0001	0.0002
MKG	17.50 (9.50)	0.0600	3.07×10^{-5}
CIV	3.97 (1.44)	0.0050	0.0002
CD	0.40 (0.20)	0.0400	4.65×10^{-5}
GEST	0.19 (8.02×10^{-2})	0.0100	2.74×10^{-5}
CFT	0.15 (8.77×10^{-2})	0.0700	0.0004
CCWT	-2.25 (0.93)	0.0100	6.68×10^{-5}

A number of associations were observed between the beta casein A1/A2 allele and production and functional traits. Protein percentage, milk production (kg), fat content (kg), protein content (kg), carcass conformation, and somatic cell score were associated with this variant at both reliability cut offs. Additional associations for mortality and carcass fat were observed at the 0.1 cut off.

Table 3.9.3 - Association between *CSN2* (beta casein) and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	3.17×10^{-3} (1.08×10^{-3})	0.0030	0.02
MKG	11.96 (3.42)	0.0004	0.03
FKG	0.25 (0.13)	0.0500	0.01
PKG	0.58 (0.10)	2.40×10^{-8}	0.08
MORT	0.44 (0.22)	0.0400	0.02
CCF	-6.64×10^{-2} (3.44×10^{-2})	0.0500	0.02
CFT	-7.6×10^{-2} (2.25×10^{-2})	0.0006	0.22
SCS	1.23×10^{-2} (2.52×10^{-3})	1.09×10^{-6}	0.07

Table 3.9.4 - Association between *CSN2* (beta casein) and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	3.07×10^{-3} (1.08×10^{-3})	0.0040	0.01
MKG	11.43 (3.43)	0.0008	0.02
FKG	0.28 (0.13)	0.0200	0.01
PKG	0.58 (0.10)	2.82×10^{-8}	0.08
CCF	-6.74×10^{-2} (3.84×10^{-2})	0.0700	0.02
SCS	1.22×10^{-2} (3.67×10^{-3})	0.0008	0.07

The *DGAT1* gene variant associations are presented in the tables below. Protein percentage, fat percentage, milk production (kg), fat content (kg), protein content (kg), calving interval, carcass fat and culled carcass weight were significant at both reliability cut offs. Mortality, and somatic cell score were also significant at the 0.1 reliability cut off whereas carcass weight and carcass conformation were significant at the 0.2 cut off.

Table 3.9.5 - Association between *DGAT1* and production and functional traits at 0.1 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-4.55×10^{-2} (1.09×10^{-3})	<0.001	4.41
FPC	-0.15 (1.83×10^{-3})	<0.001	13.08
MKG	126.6 (3.50)	6.50×10^{-276}	3.71
FKG	-3.85 (0.13)	9.05×10^{-183}	2.59
PKG	2.18 (0.11)	4.04×10^{-87}	1.23
CIV	0.77 (0.78)	0.090	0.01
MORT	0.47 (0.23)	0.040	0.16
CFT	5.71×10^{-2} (2.41×10^{-2})	0.010	0.13
CCWT	1.08 (0.37)	0.002	0.03
SCS	6.81×10^{-3} (2.68×10^{-3})	0.010	0.02

Table 3.9.6 - Association between *DGAT1* and production and functional traits at 0.2 reliability cut off

Trait	Effect size	P value	Phenotypic Variance (%)
PPC	-4.59×10^{-2} (1.09×10^{-3})	<0.001	4.48
FPC	-0.15 (1.84×10^{-3})	<0.001	13.08
MKG	126.7 (3.50)	6.31×10^{-277}	3.72
FKG	-3.77 (0.13)	9.82×10^{-175}	2.48
PKG	2.22 (0.11)	1.11×10^{-89}	1.29
CIV	0.99 (0.56)	0.070	0.02
CWT	0.79 (0.35)	0.020	0.01
CCF	0.12 (4.07×10^{-2})	0.003	0.07
CFT	6.74×10^{-2} (3.58×10^{-2})	0.060	0.18
CCWT	0.98 (0.36)	0.005	0.02

3.10 – Association analysis –*LFNG*

The *LFNG* SNP was segregating in the heterozygous state, at a very low frequency (MAF <0.01), with no cows homozygous for this variant identified. The table below represents the results obtained for the association analysis between this SNP and all traits included in this study. A tentative association ($p<0.1$) was observed between this trait and an increase in calving interval.

Table 3.10.1 – Association between *LFNG* and production and functional traits at the 0.1 reliability cut off

Trait	Effect size	P value
CIV	7.71 (4.47)	0.08

3.11 -Bioinformatics

The following section represents the results obtained in the bioinformatics analysis of the mutations analysed in this study. Mutations that occur in exonic regions were investigated for their effect on protein structure, whereas those that occur in non-coding regions were analysed with regards to their role in gene expression by their proximity to regulatory regions of the gene.

3.11.1 -Bioinformatics analysis of the STAT variants

STAT1 2697

The *STAT1* variant (rs43705173) is located at position 3132 in the 3'UTR of the gene. The results obtained using the software tool RegRNA suggest that this variant does not affect the regulation of this gene as no miRNAs, splicing enhancers or silencers, bind to the region in which this nucleotide occurs.

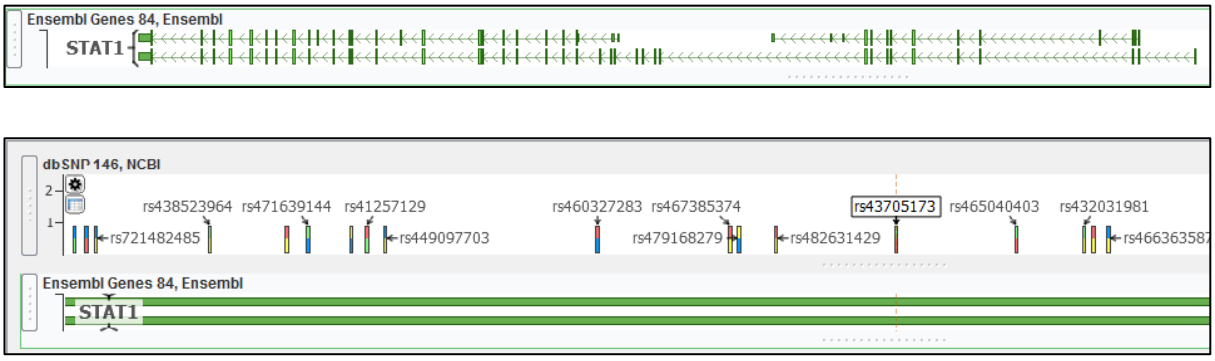


Fig. 3.5 – The *STAT1* variant, rs43705173, is located in the 3'UTR region of the gene (GoldenHelix GenomeBrowse 2.1.2)

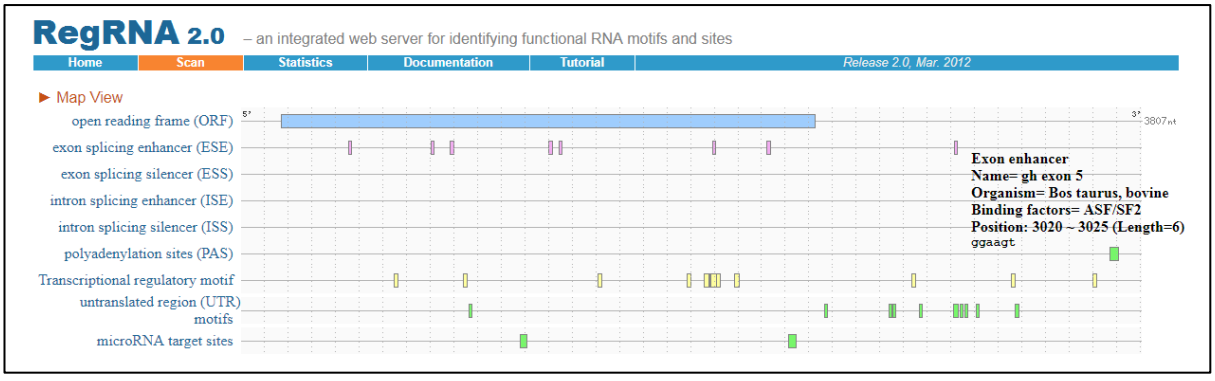


Fig. 3.6 – RegRNA results for *STAT1* determined that no miRNA binds to the region in which this variant is located

STAT3 19069

This T->G substitution occurs in exon 13 at position 1402 of the mRNA transcript of the *STAT3* gene. It results in a synonymous mutation as no change in amino acid sequence results from this substitution.

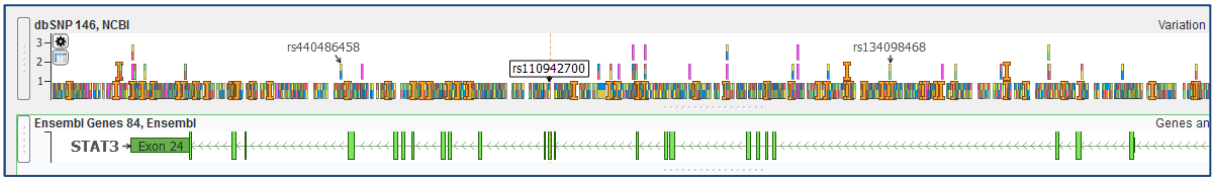


Fig. 3.7 - STAT3 19069 is located in exon 13 of the *STAT3* gene

STAT3 25402, STAT5 13516, STAT5 13319, STAT5 13244 are all located in introns, none of which are close to splice site regions.

STAT5 12195

This variant is located in exon 8, position 329 in the stat5 protein, resulting in a synonymous substitution.

3.11.2 -Bioinformatics analysis of the milk protein genes

*CSN3 (*Kappa casein)

The kappa casein variant (rs43703017) is in exon 4 of the gene, with the A->G substitution leading to a serine to glycine amino acid change at position 176 of the resultant protein. Table represents the results obtained in the bioinformatics analysis of this gene variant.

Fig. 3.8 – The *CSN3* variant is located in exon 4 of the gene (GoldenHelix GenomeBrowse 2.1.2)

Table 3.11.1– Bioinformatics results for the *CSN3* gene variant analysed in this study

Gene	SNP	Amino acid	PredictSNP	MAPP	PhD-SNP	Polyphen 2	SIFT	SNAP	PANTHER
CSN3	c.706G>A	p.ser176Gly	65%	57%	72%	60%	53%	62%	48%

■ Deleterious ■ Neutral % Confidence

Beta casein *CSN2* a1/a2 allele

The *CSN2* A1/A2 G->A substitution leads to a missense mutation, changing the amino acid from proline to histidine. The results for the effect of this mutation are presented in the table below. The SNP was suggested to be deleterious when analysed using PredictSNP, MAPP, Polyphen 2, SIFT and SNAP. It was predicted as neutral using PhD-SNP and PANTHER.

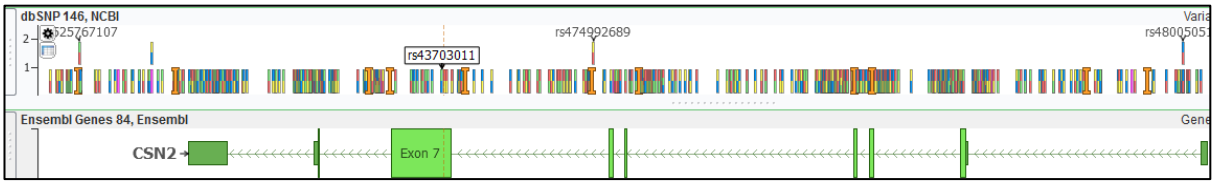


Fig. 3.9 – The *CSN2* variant is located in exon 7 (GoldenHelix GenomeBrowse 2.1.2)

Table 3.11.2 - Bioinformatics results for the *CSN2* gene variant analysed in this study

Gene	SNP	Amino acid	PredictSNP	MAPP	PhD-SNP	Polyphen 2	SIFT	SNAP	PANTHER
CSN2	c.245C>A	p.Pro82His	55%	78%	58%	55%	43%	56%	67%

■ Deleterious ■ Neutral % Confidence

DGAT1

The *DGAT1* SNP leads to an alanine to lysine substitution at position 232 of the protein. The results in the analysis of the effects of this SNP on protein function are presented in the table below. This SNP was predicted to be neutral when analysed using all tools, except for MAPP which predicted it to be deleterious, although with low confidence.

Fig. 3.10– The *DGAT1* gene variant analysed in this is located in exon 8 (GoldenHelix GenomeBrowse 2.1.2)

Table 3.11.3 - Bioinformatics results for the *DGAT1* gene variant analysed in this study

Gene	SNP	Amino acid	PredictSNP	MAPP	PhD-SNP	Polyphen 2	SIFT	SNAP	PANTHER
DGAT1	c.694GC>AA	p.Ala232Lys	74%	46%	66%	87%	90%	71%	64%

■ Deleterious ■ Neutral % Confidence

3.12 – Comparative genomics

A comprehensive literature review was conducted in order to identify candidate genes related to fertility traits in mammalian species. A file was generated based on this research which included the ensembl ID number for the genes of interest, before a R script (See Appendix II, Page 185 and 186) was used to generate results based on the consequence type of each SNP located in these genes. This script ran an input file containing the ensembl IDs and the output file contained all SNPs based on consequence type from the Biomart database (See Appendix II Fig. 5, Page 187). A

total of 9364 DNA polymorphisms were selected and submitted for consideration of inclusion into the IDBv4 design.

Chapter 4 - Discussion

The objectives of this study were to estimate the frequencies of a panel of mutations in Irish Holstein Friesian dairy cattle, to report on any observed associations between these mutations and traits of importance in dairy cattle and to identify novel candidate genes that may be involved in fertility and reproductive success in cattle populations. The following section discusses the results obtained in relation to the objectives of this research study.

4.1 –Lethal recessive disorders

The segregation of mutations with lethal effects are of significant economic importance in cattle production. The elimination from or at least the management of such mutations in the national breeding herd is a desirable objective, however, estimation of the potential pleiotropic or associated effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous. This would enable management and maintenance of animals carrying deleterious mutations which are of otherwise high genetic merit mitigating replacement costs and improving sustainability for farming enterprises. The following section explains the findings associated with the lethal recessive disorders analysed in this study, which includes their frequencies in the Irish Holstein Friesian population and any observed pleiotropic or associated effects that carriers of these mutations may exhibit on the milk, fertility and heath traits examined.

4.1.1 -Complex vertebral malformation

CVM is an autosomal lethal recessive disease that is found in the Holstein breed. The *SLC35A3* mutation associated with the development of complex vertebral malformation was observed at a frequency of 1.7% in the Irish Holstein Friesian cattle population studied, being previously reported at a frequency of 4% (Mullen et al., 2013)**.** The results of the association study suggest that carriers of this mutation in the population studied have a decreased percentage of both protein and fat in their milk at the 0.1 and 0.2 reliability cut offs. At the 0.2 cut off level, survival was also associated with this mutation, with an increase in this trait observed. All significant associations were less than 0.1% phenotypic variance observed. No associations between the other milk and fertility traits were evident. With regards to health and carcass traits no

associations between CVM carriers and either somatic cell score, carcass weight, cull cow carcass weight, carcass conformation and carcass fat were observed in the sample set tested, suggesting that carriers do not significantly influence performance with regards to these traits.

Confounding results have been detected and reported by other research teams. Chu et al, 2010, determined that the EBVs for milk production traits of CVM carriers were significantly higher than those of non-carriers (n=555). However, Cole et al, 2016 reported similar findings to what has been reported here, with their study suggesting heterozygous carriers for CVM exhibit lower fat yield in comparison to non-carriers (n = 1,868) (Cole, Null and VanRaden, 2016). Berglund, Persson and Stålhammar, 2004, presented a study in which they investigated the effect of carrier status on fertility traits in both bulls and daughters, which concluded a significantly lower breeding value for non-return rates (unfavourable?) for carrier bulls compared to non-carriers, however, no effect was observed for daughter fertility traits in this study.

4.1.2 -Brachyspina

Brachyspina, caused by a 3.3kb deletion in the bovine *FANCI* gene, was found at a frequency of 1% in the Irish Holstein Friesian population, being previously reported at being identified in 2% of the population in 2013 (Mullen et al., 2013). In the association analysis, carriers of Brachyspina did not exhibit any effects on the carcass and heath traits examined in this study, however, carriers of this genetic defect did exhibit effects on milk yield and composition traits (effect sizes less than 0.01% phenotypic variance). A decrease in the protein component of milk was observed in addition to an increase in milk yield. These results were consistent across both reliability cut offs. A previous study reported an association between carriers of Brachyspina and reduced fat content of milk (Cole, Null and VanRaden, 2016).

No associations with any of the fertility traits examined were evident in this study. In contrast, a study by Cole, Null and VanRaden, 2016, found that direct genomic values were significantly lower for heifer conception rate in Brachyspina carriers, a finding consistent with Charlier et al in 2012, who observed increased pregnancy failure when carriers were mated with non-carriers.

4.1.3 -Deficiency of uridine monophosphatase

DUMPs, a lethal recessive disorder, causes early embryonic death in Holstein cattle. Earlier studies observed that heterozygote carriers of this mutation had higher genetic merit for milk production traits (Shanks and Greiner, 1992). Detection of carriers of this mutation, which introduces a stop codon which halts translation of the protein, has greatly reduced the frequency of carriers in the Holstein population and a frequency of 0.02% was found to be present in the Irish Holstein Friesian population in this study. Not sure how the detection of carriers is related to greatly reduced frequency? Maybe reword. As of 2016, the frequency of DUMPs in the US was reported as 0.01% (Cole et al., 2016). The association analysis could not be performed to determine the effect of carriers of this mutation on the phenotypes presented in this study as no carriers of the mutation met the minimum adjusted reliability cut off value of 0.1 (10%).

4.1.4 – *LFNG*

The *LFNG* gene codes for an evolutionarily conserved glycosyltransferases that acts in the Notch signalling pathway which is vital during embryonic development for controlling the regulation of the formation and patterning of somites in vertebrates (Zhang and Gridley, 1998). Research studies in both human and mice have shown that this gene plays a role in haematopoiesis and also regulated the formation of many cells and organ systems (Harper et al., 2003). Knockout mice display distinct segmentation defects during embryogenesis (Zhang, Norton and Gridley, 2002). In zebrafish, *lfng* has been shown to be involved in the formation of segment boundaries in the hindbrain and somites (Prince et al., 2001). In humans, mutations in this gene have been associated with the development of Spondylocostal Dysostosis, which is a vertebral malformation disorder arising during embryonic development (Sparrow et al., 2006).

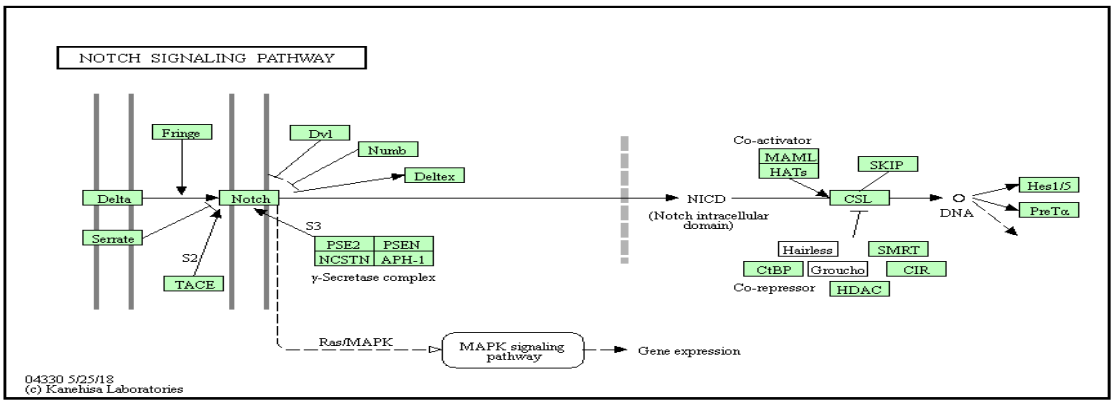


Fig 4.0 – The Bos Taurus Notch signalling pathway, of which the *LFNG* gene is a modulator (Kegg.jp, 2018)

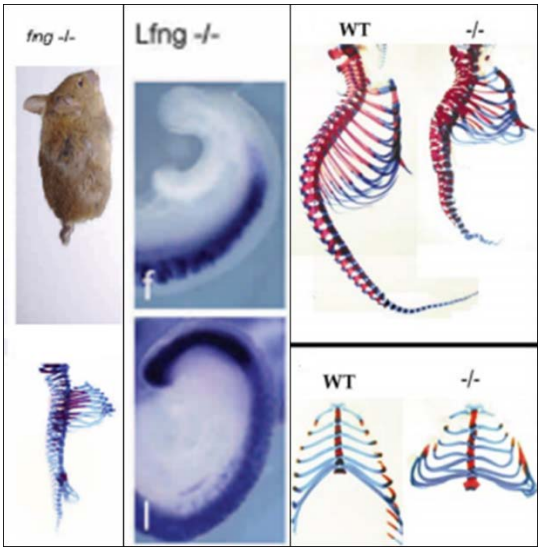


Fig. 4.1 - Knockout *lfng* mice display skeletal and vertebral malformations with the axial skeleton being severely affected, numerous vertebral and rib fusions and incompletely formed vertebrae are also evident (Serth et al., 2003; Sparrow et al., 2006).

The results of the association analysis revealed a tentative association with a nonsense SNP in *LFNG* and an increase in calving interval in Holstein Friesian dairy cattle albeit a very small effect representing <0.001 % of phenotypic variance. This result suggests this SNP may play a role in early embryonic loss in dairy cattle. However, this effect would need to be validated in a larger dataset of animals with more carrier animals coupled with *in vitro* and *in vivo* functional studies to establish *LFNG*s potential role in fertility.

4.2 -Signal transducer and activator of transcription

The *STAT* genes are members of the signal transduction pathways involved in pre and post implantation, mammary gland development and expression of milk protein genes. In genetic association studies, polymorphisms within the *STAT* genes and in other genes implicated in their pathways, such as the growth hormone gene (GH) and the growth hormone receptor (Mullen et al., 2011; Waters et al., 2011), have been found

to be associated with economically important traits in cattle populations. The present study focuses on several previously reported polymorphisms within the *STAT1, STAT3* and *STAT5* genes and their association with production and performance traits in Irish Holstein Friesian dairy cattle and the following section explains the results that have been obtained in the association analysis between these gene variants and the traits of interest in this study.

Minor allele frequencies within the *STAT* genes ranged from 4% to 43%, with comparable MAF results for SNP *STAT5* 12195 being observed in Holstein populations by Khatib, 2008, and Oikonomou, 2011. Direct comparison of MAF for all SNPs is not feasible as studies have not been previously published on the same SNPs analysed in this study. SNPs located in the *STAT5* and *STAT3* genes, both of which are located on chromosome 19, showed high LD between most pairwise comparisons. *STAT*5 13319 was in low LD with all other SNPs analysed, suggesting it may be segregating independently, with its low MAF yet significant association with important traits warranting further investigation.

4.2.1 -*STAT1*

STAT1 has been proven to regulate immune related genes (Durbin et al., 1996; Takeda et al., 2003). It has been suggested that *STAT1* plays a role in the clearance of fungal infections and in particular *Candida* species, which is supported by the association with a gain of function *STAT1* mutation with chronic mucocutaneous candidiasis disease in humans (Dotta et al, 2016; Takezaki et al, 2012). *Candida* species are frequently observed in milk from cattle with cases of mastitis (AalbæK et al.,1994; dos Santos et al., 2005).

The G allele of the *STAT1* variant analysed in this study was found to be at a frequency of 0.25 in the population studied. Substitution of the A allele for the G allele was tentatively associated with an increased somatic cell score (effect size, 0.015% phenotypic variance, p=0.08) in the current study. In clinical cases of mastitis, a measure of the white blood cells (somatic cells) increases in response to the presence of pathogenic micro-organisms. The association observed here with this genetic variant in the *STAT1* gene and an increase in somatic cell score is interesting due to previous research linking this gene and *Candida* immunity. Reanalysis in a larger dataset would need to be undertaken to further analyse this observation.

No association was observed between the *STAT1* variant and any of the other traits investigated in this study, which is in conflict to the studies by Cobanoglu et al, 2006 and Rychtarova et al., 2014, who all reported that this *STAT1* variant was associated with an increase in milk fat and protein content of milk in Holstein (n=762) and Fleckvieh (n=419), however, Fontanesi et al., 2015 reported a decrease in protein percentage in Reggiana (n=128) breeds.

Bioinformatic analysis of this variant revealed the position of this variant to be within the 3'UTR region of the gene, suggesting its ability to affect gene regulation by disturbing miRNA binding, however *in silico* analysis through use of the software tool RegRNA did not detect a miRNA binding site at this position.

4.2.2 -*STAT3*

As previously discusssed the *STAT3* gene has been linked to fertility, and in partiular early embryogenises and successful implantation. Additionally, it has been associated with fertility traits in studies performed by Khatib et al., 2009.

The C allele of *STAT3* 19069 was significantly associated with a decrease in fat percentage, an increase in milk and protein production, a decrease in gestation length and an increase in culled carcass weight. These results were consistent for both the reliability cut offs. The findings of an association with this gene and milk production and composition traits is one which has not been reported in the literature previously, however, it should be noted the effect sizes are small and all less than 0.15% of phenotypic variance.

Ex vivo studies have been performed which have associated SNPs within the *STAT* genes with fertility traits. Khatib et al, 2009, investigated the effects of the SNP–SNP interactions of *STAT* genes on embryonic survival and fertilization rate. A genotype combination of SNP C/T at position 3141 in the 3′ UTR of *STAT1* and *STAT3* (19069) within the heterodimer complex formed in the signalling pathway was found to be significantly associated with higher embryonic survival. Leptin has been shown to exert its effects through the Jak/STAT pathway (Procaccini et al., 2009) and considering leptins effect on energy balance (Liefers et al., 2003), this may suggest an indirect role of *STAT1* on milk traits and fertility traits.

Bioinformatics analysis of the *STAT3* 19069 variant revealed its location in exon 13 of the *STAT3* gene, with the T->G substitution results in a synonymous mutation, therefore it is not expected to have any effect on protein structure or function.

The G allele of *STAT3* 25402 was also significantly associated with several traits in this study. A decrease in protein percentage, fat percentage, calving difficulty and gestation length, whereas an increase in milk yield, was observed, individually representing less than 0.06% of phenotypic variance. At the 0.2 reliability cut off calving difficulty was not significant, likely due to the difference in the sample size tested (n = 14457 for 0.1, n= 3653 for 0.2), however, protein production (kg) and somatic cell score were significant. Khatib at el., 2009 performed single-SNP analysis on this variant which revealed a statistically significant association between *STAT3* 25402 and fertilization rate.

Bioinformatics analysis of the *STAT3* 25402 variant revealed its position within intron 20 of the gene, not near any splice site regions, suggesting the associations observed with this variant is due to linkage with a QTN in the same genomic region.

4.2.3 - *STAT5*

There are a number of published research studies demonstrating a relationship between the *STAT5* gene and milk traits in various cattle breeds, one example being a polymorphism in intron 9 of the *STAT5* gene in Jersey cows being associated with milk yield, fat and protein content (Brym and Kamiñski and Rusc, 2004). Position 6853 within exon 7 of this gene was associated with the same traits in Italian Swiss brown cattle (Selvaggi et al., 2009), and in Jersey cows (Dario and Selvaggi, 2011) with the C allele yielding higher milk, fat and protein content. He et al, 2012, determined an association with polymorphisms in the *STAT5* gene and protein composition in Holstein cattle. With regards to fertility traits, Oikonomou et al, 2011, reported an association (p=0.07) between the G allele of *STAT5* 12195 and a decrease in age at first calving by 7.2 days, with the author stating that statistical significance may not have been achieved due to the sample size. Khatib et al., 2008, describes an association between the G allele of *STAT5* 12195 with a decrease in embryonic survival rates.

In the current study *STAT5* 12195 and *STAT5* 13244 were associated with a decrease in calving difficulty (0.008% phenotypic variance). *STAT5* 13244 and *STAT5* 13516 were also both associated with a decrease in carcass fat at the 0.2 reliability cut off.

Bioinformatics analysis determined that *STAT5* 12195 is located in exon 8, position 329 in the stat5 protein, resulting in a synonymous substitution, whereas *STAT5* 13244 is located in an intronic region, not close to any splice sites or regulatory regions currently known.

Associations between the A allele of the *STAT5* 13319 SNP and an increase in protein percentage, fat percentage, fat yield, protein yield, carcass weight and culled carcass weight were observed, at a maximum of 0.2% phenotypic variance. A tentative decrease in maternal calving difficulty (0.003 % phenotypic variance, p=0.07) was observed at the 0.1 reliability cut off, however, this effect size increased and achieved greater significance (effect size -0.5, 0.01% phenotypic variance, p=0.005) at the 0.2 reliability cut off. Although this SNP is located in an intron, making thorough analysis of its effects unreliable using the currently available tools to analyse these variants, its low MAF and high significance with the traits analysed make it a candidate for further investigation.

The observations presented here are further evidence of the part that the *STAT* genes have been found to play *in vivo* in relation to mammary gland development (Watson and Neoh, 2008; Haricharan and Li, 2014), and milk protein gene expression (Li and Rosen, 1995; Mukhopadhyay et al., 2001). *STAT5* is a key mediator of PRL and GH **(**Gallego et al., 2001), both of which have also been associated with milk traits in cattle populations (He et al., 2006; Dybus, 2002). There has been less emphasis placed on the *STAT3* gene and its relation to fertility traits in cattle despite early *in vivo* studies determining that disruption to the *STAT3* gene led to embryonic lethality in mice (Takeda et al., 1997) and it has since been found to be critical in successful implantation (Lee et al., 2013), suggesting its inclusion in the analysis of fertility traits is warranted.

Recent research into the regulatory functions of non-coding RNAs (ncRNAs) has emphasised their importance in regulating genes involved in both mammary gland development and lactation, and micro RNAs (miRNAs) have been found to regulate genes implicated in pathways that involve the *STAT* genes (Do and Ibeagha-Awemu,

2017), with some of these miRNAs being located in overlapping regions with QTL for milk traits (Ogorevc et al., 2009). The *STAT1* variant analysed in this study was located within the 3'UTR region implicating it as a possible target for miRNA binding, however the SNP was not found to be located in a known miRNA site, nor was it located in a polyadenylation site and therefore is not thought to affect mRNA degradation (Chang et al., 2013). None of the investigated intronic SNPs (*STAT3* 25402, *STAT5* 13516, *STAT5* 13516, *STAT5* 13319 and *STAT5* 13244) were located in close proximity to splice sites and as such are not thought to affect splicing. Researchers are however investigating the role of introns as potential ncRNA molecules that can modulate gene expression (Kornienko et al., 2013; Quinn and Chang, 2015). Two SNPs *STAT3* 19069 and *STAT5* 12195 were located in exons but were not located in positions targeted by splicing enhancers or silencers (Chang et al., 2013).

4.2.4 – Haplotype analysis

The use of haplotypes in association analysis, rather than SNP effects, has been proposed to have several advantages including increasing the power of the study, and it has also been suggested that haplotypes may capture causal variants themselves (Akey, Jin and Xiong, 2001). For this study, haplotypes within the *STAT3* and *STAT5* genes, located on BTA19 were constructed and subsequently tested for their association with the traits of interest.

Using Phase to construct the haplotypes predicted 20 haplotypes in this region of the bovine genome, however, most of these were at very low MAF (<0.001 percent). H1, H4 and H18 were observed at moderate to high frequencies of 29, 0.1 and 2.4 percent and so were subsequently analysed for their associations with the traits of interest. H1 consists of the SNP *STAT5* 12195. H4 consists of SNPs *STAT5* 13244, *STAT3* 25402 and *STAT3* 19069 and H18 consists of all the six SNPs analysed on BTA19. Should really be in results.

Animals with the H1 haplotype were associated with reduced milk protein composition (0.2% phenotypic variation) with tentative associations for decreased milk yield (0.1% phenotypic variation), with an increase in gestation length (phenotypic variation <0.001%) and somatic cell score (0.2% phenotypic variation)

also evident. No associations between this haplotype and any of the fertility or carcass traits were observed.

Animals with the H4 haplotype displayed an increase in protein composition (0.6 % phenotypic variation) and milk yield, whereas a decrease in fat yield was observed. No associations with this haplotype and any of the fertility traits examined was evident. With regards to the carcass traits, this haplotype was associated with a decrease in carcass fat (2.6% phenotypic variation) and an increase in culled carcass weight. A significant association was also observed between this haplotype and an increase in somatic cell score (1 % phenotypic variation).

The H18 haplotype was significantly associated with an increase in protein percentage (0.1% phenotypic variation) and a decrease in calving difficulty. Tentative associations between this haplotype and an increase in milk fat, a decrease in gestation length and maternal calving difficulty were also observed.

4.3 -Milk protein genes

The selection for increased production traits in dairy cattle has had an antagonistic effect on fertility traits in dairy cattle worldwide and this fact has been well documented in the literature (Oltenacu and Broom, 2010; Lucy, 2001; Walsh, Williams and Evans, 2011). Negative genetic correlations between milk production and fertility and health traits and post-partum negative energy balance due to an increased metabolic demand in high producing cows are both factors that are thought to play a part in this observation. The follwing section explains the results obtained in the analysis of the milk protein genes included in this study, namely the *CSN3*, *CSN2* and *DGAT1* genes. Particular interest lies in the associated effects observed between these gene variants and fertility traits in the population of Holstein Friesian cattle.

4.3.1 -KAPPA CASEIN

The G allele of the *CSN3* variant was found at a frequency of 3% in the Holstein Friesian population studied. As expected, this gene was found to be associated with both milk production and composition traits albeit representing very little of the phenotypic variation (<0.001%). Results were consistent for these traits at both reliability cut offs with a decrease in protein percentage and an increase in milk yield observed. These results are consistent with those found in the literature where variants

in this gene are associated with effects on milk production and composition traits (Lechniak et al., 2002; Boettcher et al., 2004; Rachagani and Gupta, 2008).

In addition, the G allele of this variant was associated with decreased carcass weight, carcass conformation and culled carcass weight. With regards to fertility traits, at both reliability cut offs a small increase in calving interval was observed whereas at the 0.2 reliability cut off calving difficulty and gestation length also achieved significance exhibiting increases in these traits. However, these results are in conflict with previously reported studies that suggested variants within the Kappa casein gene had no effect on fertility traits in Holstein Friesian cattle (Demeter et al., 2010; Tsiaras et al., 2005).

Bioinformatics analysis of the *CSN3* variant revealed its location in exon 4 of the gene, position 595 in the mRNA transcript, leading to a non-synonymous mutation at position 176 of the resultant protein. The change from the polar amino acid serine, which is often a highly reactive residue due to its hydroxyl group, to the simple amino acid glycine, which often plays a vital role in phosphorylation reactions, makes it likely that this mutation has an effect on protein structure and function (Barnes, 2003). Analysing this variant through the software PredictSNP, a tool which provides results based on all the bioinformatics tools discussed in the introduction chapter, determined this mutation to be deleterious with a confidence value of 65%.

4.3.2 -BETA CASEIN A1/A2

The beta casein constitutes up to 45% of bovine milk total casein, and presents as 12 genetic variants, however in dairy cattle the A1 and A2 types are most common (Jianqin et al., 2015). The metabolic breakdown of the A1 variant yields the peptide β-casomorphin-7 which has been proposed to cause adverse effects associated with milk consumption, many of which are similar to the symptoms associated with lactose intolerance (Massella et al., 2017). Therefore, the selection for the A2 allele is desirable due to the observed adverse effect of the A1 allele on human health. However, it is important to assess the effect of this allele on other important traits in cattle populations. The frequency of the A2 allele was found at a frequency of 39% in the population of Irish Holstein Friesian analysed.

The association analysis revealed that the A2 allele of this variant was associated with milk protein percentage, milk, fat and protein yield, individually all less than 0.1% of

phenotypic variance. For fertility traits, one copy of the A2 allele was associated with increased mortality of the order of 0.02 % of phenotypic variance in the 0.1 reliability cut off dataset. This allele was also associated with a decrease in carcass conformation, carcass fat and an increase in somatic cell score, all less than 0.3% of phenotypic variance, using the 0.1 reliability cut off, whereas carcass fat failed to achieve significance in the 0.2 reliability cut off dataset. These results highlight the relevance of estimating the effects of increasing the frequency of the A2 beta casein should consumer demand increase for A2A2 milk and milk products. PredictSNP determined this allele substitution to be deleterious, with a confidence level of 55%.

4.3.3 -*DGAT1*

DGAT1 functions as a metabolic enzyme catalysing the last step in triglyceride synthesis. In 1998, Coppieters et al, reported a QTL with major effects on milk traits on bovine chromosome 14. Since then this QTL has been associated with milk traits in a number of studies (HEYEN et al., 1999; Ashwell et al., 2004; Rodriguez-Zas et al., 2002; Boichard et al., 2003). Researchers proposed the *DGAT1* gene located within the QTL as the causative gene for the effects being observed. Candidate gene studies were then performed based on these findings, as well as *in vivo* studies which confirmed the genes role in biological processes linked to milk traits,an exmaple being $Dgat^{-/-}$ female mice displaying defective lactation (Smith et al., 2000). In 2003, Grisart et al, performed functional studies which confirmed the *DGAT1* K232A mutation as being the QTN responsible for the effect on milk traits observed in this genomic region. The amino acid lysine at position 232 of the protein leads to higher milk composition traits in cattle populations.

In the present study the frequency of the *DGAT1* G allele, which produces amino acid alanine at position 232 of the protein, was found to be 0.39. As expected, this allele was associated with significant proportions of phenotypic variance in milk related traits. The DGAT G allele was associated with a decrease in fat percentage (13% phenotypic variance), a finding which agrees with the findings of previous studies where the A allele was associated with increased fat percentage. A decrease in protein percentage (4.41% phenotypic variance) and fat yield (2.59% phenotypic variance) were also observed, whereas an increase in milk and protein was detected (3.71 and 1.23% phenotypic variance respectively). These results were consistent at both reliability cut offs. For fertility traits, an increase in calving interval was associated

with this allele at the 0.1 reliability cut off, where mortality was also increased in the 0.2 reliability dataset. Previous studies have reported a negative correlation between the mutation in *DGAT1* responsible for increased milk composition traits and fertility traits in Holstein Friesian cattle (Kaupe et al., 2007; Zabolewicz et al., 2011). However, Komisarek et al., 2011, reported no association with this mutation and fertility traits.

While associations with this mutation and carcass traits and fertility and udder health traits were evident they represented no more than 0.2% of phenotypic variance. At both reliability cut offs an increase in carcass fat and culled carcass weight was observed. At the 0.2 cut off an increase in carcass weight and carcass conformation was also observed, whereas at the 0.1 cut off an increase in somatic cell score was evident. Variants in this gene were also associated with carcass traits in a number of breeds of chinese beef cattle (Yuan et al., 2013) and in Nellore cattle (Borges et al., 2013), however, no studies have been published associating this gene with carcass traits in Holstein Friesian cattle.

These results suggest that increasing the frequency of the mutation responsible for the K323A QTN may have small anatagonistic effects on fertility and other production traits analysed. Interestingly, PredictSNP predicted this mutation to be neutral, with a confidence level of 74%.

4.4 – Comparative genomics

In order to support the ICBF with the development of the latest version of the IDB chip a comprehensive literature review was undertaken to identify DNA mutations with potentially novel roles in fertility in cattle and other mammalian species, to allow inclusion of these variants on the chip for future analysis. Other variants in the genes currently being investigated (lethal recessives, *STAT*s, milk protein genes) were also suggested to be added to the chip with attention being given to variants that cause functional change in the proteins being studied, allowing the possible idenfication of caustive variants in these genes. The addition of these SNPs will also permit further haplotype analysis to be undertaken. A further 200+ genes were also identified as being possible candidate genes involved in fertility pathways in cattle populations. The variants put forward to the ICBF for inclusion on the chip were prioritised

according to the amount of evidence available from the literature to suggest their link with fertility and by the consequence type of the actual variant itself.

4.5 – Conclusion

The main objectives of this research study were to estimate the frequencies and effects of a panel of mutations, some of which are validated as causative for the disease or trait of interest, and some which are known to be involved in biological pathways that are associated with the traits being studied. As described above, a number of expected and novel associations between these mutations and the traits of interest have been observed in the population of Irish Holstein Frieisan cattle studied.

Aiming for genetic improvement in farm animals involves the selection of parents that when mated are expected to produce progeny that perform better than that of the current generation. Results presented here suggest the possibility of incorporation of these genetic variants into genetic selection programmes in order to increase the rate of genetic gain for production and functional traits in Holstein Friesian cattle, however, care should be taken in view of possible antagonistic effects. Whether all the SNPs analysed in this study are causative for the phenotypic trait or linked to the causative allele is unknown in some cases (i.e., the *STAT* gene variants). Validation of the effects of these mutations is warranted and could include reanalysis of the effects in independent populations coupled with *in vitro* and *in vivo* analysis.

.

References

AalbæK, B.E.N.T., Stenderup, J.Ø.R.G.E.N., Jensen, H.E., Valbak, J.E.S.P.E.R., Nylin, B.R.I.T.T.A. and Huda, A.N.N.A., (1994). Mycotic and algal bovine mastitis in Denmark. *Apmis*, *102*(1-6), pp.451-456.

Abdul-Muneer, P. (2014). Application of Microsatellite Markers in Conservation Genetics and Fisheries Management: Recent Advances in Population Structure Analysis and Conservation Strategies. *Genetics Research International*, 2014, pp.1-11.

Adams, H., Sonstegard, T., VanRaden, P., Null, D., Van Tassell, C., Larkin, D. and Lewin, H. (2016). Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *Journal of Dairy Science*, 99(8), pp.6693-6701.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. and Sunyaev, S.R., (2010). A method and server for predicting damaging missense mutations. *Nature methods*, *7*(4), p.248.

Agerholm, J., Bendixen, C., Andersen, O. and Arnbjerg, J. (2001). Complex Vertebral Malformation in Holstein Calves. *Journal of Veterinary Diagnostic Investigation*, 13(4), pp.283-289.

Agerholm, J., McEvoy, F. and Arnbjerg, J. (2006). Brachyspina Syndrome in a Holstein Calf. *Journal of Veterinary Diagnostic Investigation*, 18(4), pp.418-422.

Akey, J., Jin, L. and Xiong, M. (2001). Haplotypes vs single marker linkage disequilibrium tests: what do we gain?. *European Journal of Human Genetics*, 9(4), pp.291-300.

Aleandri, R., Buttazzoni, L., Schneider, J., Caroli, A. and Davoli, R. (1990). The Effects of Milk Protein Polymorphisms on Milk Components and Cheese-Producing Ability. *Journal of Dairy Science*, 73(2), pp.241-255.

Altschul, S., Gish, W., Miller, W., Myers, E. and Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), pp.403-410.

Ans.iastate.edu. (2017). *Robert Bakewell | Department of Animal Science*. [online] Available at: http://www.ans.iastate.edu/about/history/people/robert-bakewell [Accessed 5 Sep. 2017].

Ashwell, M., Van Tassell, C. and Sonstegard, T. (2001). A Genome Scan To Identify Quantitative Trait Loci Affecting Economically Important Traits in a US Holstein Population. *Journal of Dairy Science*, 84(11), pp.2535-2542.

Ashwell, M., Heyen, D., Sonstegard, T., Van Tassell, C., Da, Y., VanRaden, P., Ron, M., Weller, J. and Lewin, H. (2004). Detection of Quantitative Trait Loci Affecting Milk Production, Health, and Reproductive Traits in Holstein Cattle. *Journal of Dairy Science*, 87(2), pp.468-475.

Barnes, M. (2003). *Bioinformatics for geneticists*. John Wiley & Sons, pp.291 - 316.

Barrett, J., Fry, B., Maller, J. and Daly, M. (2004). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), pp.263-265.

Bawden, R.J. (1991). Towards action research systems. *Action research for change and development*, pp.10-35.

Beltman, M., Mullen, M., Elia, G., Hilliard, M., Diskin, M., Evans, A. and Crowe, M. (2014). Global proteomic characterization of uterine histotroph recovered from beef heifers yielding good quality and degenerate day 7 embryos. *Domestic Animal Endocrinology*, 46, pp.49-57.

Ben Khelifa, M., Coutton, C., Zouari, R., Karaouzène, T., Rendu, J., Bidart, M., Yassine, S., Pierre, V., Delaroche, J., Hennebicq, S., Grunwald, D., Escalier, D., Pernet-Gallay, K., Jouk, P., Thierry-Mieg, N., Touré, A., Arnoult, C. and Ray, P. (2014). Mutations in DNAH1, which Encodes an Inner Arm Heavy Chain Dynein,

Lead to Male Infertility from Multiple Morphological Abnormalities of the Sperm Flagella. *The American Journal of Human Genetics*, 94(1), pp.95-104.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E., Zendulka, J., Brezovsky, J. and Damborsky, J. (2014). PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Computational Biology*, 10(1), p.e1003440.

Berglund, B., Persson, A. and Stålhammar, H. (2004). *Acta Veterinaria Scandinavica*, 45(3), p.161.

Berglund, B. (2008). Genetic Improvement of Dairy Cow Reproductive Performance. *Reproduction in Domestic Animals,* 43, pp.89-95.

Berry, D. (2015). Breeding the dairy cow of the future: what do we need? *Animal Production Science*, 55(7), p.823.

Berry DP, Kearney JF, Harris BL (2009) Genomic selection in Ireland. Interbull Bulletin 26–29(January), 29–34.

Berry, D., Meade, K., Mullen, M., Butler, S., Diskin, M., Morris, D. and Creevey, C. (2010). The integration of 'omic' disciplines and systems biology in cattle breeding. *animal*, 5(04), pp.493-505.

Berry, D., Wall, E. and Pryce, J. (2014). Genetics and genomics of reproductive performance in dairy and beef cattle. *Animal*, 8(s1), pp.105-121.

Berry, D.P., Friggens, N.C., Lucy, M. and Roche, J.R., (2016). Milk production and fertility in cattle. *Annual review of animal biosciences*, *4*, pp.269-290.

BEUZEN, N., STEAR, M. and CHANG, K. (2000). Molecular markers and their use in animal breeding. *The Veterinary Journal*, 160(1), pp.42-52.

Bocquet-Appel, J. and Bar-Yosef, O. (2008). *The Neolithic demographic transition and its consequences*. [Dordrecht]: Springer, pp.pp 179-205.

Boettcher, P., Caroli, A., Stella, A., Chessa, S., Budelli, E., Canavesi, F., Ghiroldi, S. and Pagnacco, G. (2004). Effects of Casein Haplotypes on Milk Production Traits in

Italian Holstein and Brown Swiss Cattle. *Journal of Dairy Science*, 87(12), pp.4311-4317.

Boichard, D., Grohs, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M. and Levéziel, H. (2003). Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution*, 35(1), p.77.

Bolker, B., Brooks, M., Clark, C., Geange, S., Poulsen, J., Stevens, M. and White, J. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), pp.127-135.

Borges, B.O., Curi, R.A., Baldi, F., Feitosa, F.L.B., Andrade, W.B.F.D., Albuquerque, L.G.D., Oliveira, H.N.D. and Chardulo, L.A.L., (2014). Polymorphisms in candidate genes and their association with carcass traits and meat quality in Nellore cattle. *Pesquisa Agropecuaria Brasileira*, *49*(5), pp.364-371.

Bromberg, Y. and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research*, 35(11), pp.3823-3835.

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

Butler, W. (2003). Energy balance relationships with follicular development, ovulation and fertility in postpartum dairy cows. *Livestock Production Science*, 83(2-3), pp.211-218.

Calus, M. and Veerkamp, R. (2011). Accuracy of multi-trait genomic selection using different methods. *Genetics Selection Evolution*, 43(1), p.26.

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

Cattle Breeding in Ireland. (2007). [pdf] Available at: https://www.icbf.com/publications/files/eaap_cattle_breeding_booklet.pdf [Accessed 12 Sep. 2017].

Cd-genomics.com. (2017). *SNP Microarray - CD Genomics*. [online] Available at: http://www.cd-genomics.com/snp-microarray.html [Accessed 14 Sep. 2017].

Chang, T.H., Huang, H.Y., Hsu, J.B.K., Weng, S.L., Horng, J.T. and Huang, H.D., (2013), January. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. In *BMC bioinformatics* (Vol. 14, No. 2, p. S4). BioMed Central.

Charlier, C., Agerholm, J., Coppieters, W., Karlskov-Mortensen, P., Li, W., de Jong, G., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., Ahariz, N., Mullaart, E., Georges, M. and Fredholm, M. (2012). A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. *PLoS ONE*, 7(8), p.e43085.

Charlier, C., Li, W., Harland, C., Littlejohn, M., Coppieters, W., Creagh, F., Davis, S., Druet, T., Faux, P., Guillaume, F., Karim, L., Keehan, M., Kadri, N., Tamma, N., Spelman, R. and Georges, M. (2016). NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Research*, 26(10), pp.1333-1341

CHU, Q., ZHANG, Y., SUN, D., YU, Y., WANG, Y. and ZHANG, Y. (2010). Identification of the complex vertebral malformation gene in Chinese Holstein and its association with dairy performance traits. *Hereditas (Beijing)*, 32(7), pp.732-736.

Cobanoglu, O., Zaitoun, I., Chang, Y., Shook, G. and Khatib, H. (2006). Effects of the Signal Transducer and Activator of Transcription 1 (STAT1) Gene on Milk Production Traits in Holstein Dairy Cattle. *Journal of Dairy Science*, 89(11), pp.4433-4437.

Cole, J., VanRaden, P., Null, D., Hutchison, J., Cooper, T. and Hubbard, S. (2016). *Haplotype tests*. [online] Aipl.arsusda.gov. Available at: https://aipl.arsusda.gov/reference/recessive_haplotypes_ARR-G3.html [Accessed 1 Nov. 2017].

Cole, J., Null, D. and VanRaden, P. (2016). Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. *Journal of Dairy Science*, 99(9), pp.7274-7288.

Coppieters, W., Riquet, J., Arranz, J.J., Berzi, P., Cambisano, N., Grisart, B., Karim, L., Marcq, F., Moreau, L., Nezer, C. and Simon, P., (1998). A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mammalian Genome*, *9*(7), pp.540-544.

Cordell, H. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*, 11(20), pp.2463-2468.

Cromie, A.R., Kearney, F., Evans, R. and Berry, D.P., (2014), August. Genomics for pedigree and cross-bred beef cattle populations; some experiences from Ireland. In *10th World Congress of Genetics Applied to Livestock Production, Vancouver'. Available at https://asas.org/docs/default-source/wcgalp-proceedings-oral/260_paper_10495_manuscript_1361_0. pdf.*

Curi, R., Oliveira, H., Silveira, A. and Lopes, C. (2005). Effects of polymorphic microsatellites in the regulatory region ofIGF1andGHRon growth and carcass traits in beef cattle. *Animal Genetics*, 36(1), pp.58-62.

Dalgleish, D., Spagnuolo, P. and Douglas Goff, H. (2004). A possible structure of the casein micelle based on high-resolution field-emission scanning electron microscopy. *International Dairy Journal*, 14(12), pp.1025-1031.

DAMIANI, G., FERRETTI, L., ROGNONI, G. and SGARAMELLA, V. (1990). Restriction fragment length polymorphism analysis of the κ-casein locus in cattle. *Animal Genetics*, 21(2), pp.107-114.

Danna, K. and Nathans, D. (1971). Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. *Proceedings of the National Academy of Sciences*, 68(12), pp.2913-2917.

Dario, C. and Selvaggi, M. (2011). Study on the STAT5A/AvaI polymorphism in Jersey cows and association with milk production traits. *Molecular Biology Reports*, 38(8), pp.5387-5392.

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

Dekkers, J. and Hospital, F. (2002). MULTIFACTORIAL GENETICSTHE USE OF MOLECULAR GENETICS IN THE IMPROVEMENT OF AGRICULTURAL POPULATIONS. *Nature Reviews Genetics*, 3(1), pp.22-32.

Demeter, R., Markiewicz, K., van Arendonk, J. and Bovenhuis, H. (2010). Relationships between milk protein composition, milk protein variants, and cow fertility traits in Dutch Holstein-Friesian cattle. *Journal of Dairy Science*, 93(11), pp.5495-5502.

DeYoung, R., Baker, J., Cado, D. and Winoto, A. (2003). The Orphan Steroid Receptor Nur77 Family Member Nor-1 Is Essential for Early Mouse Embryogenesis. *Journal of Biological Chemistry*, 278(47), pp.47104-47109

Dillon, P., Berry, D.P., Evans, R.D., Buckley, F. and Horan, B., (2006). Consequences of genetic selection for increased milk production in European seasonal pasture based systems of milk production. *Livestock Science*, *99*(2), pp.141-158.

Diskin, M. and Morris, D. (2008). Embryonic and Early Foetal Losses in Cattle and Other Ruminants. *Reproduction in Domestic Animals*, 43, pp.260-267.

Do, D.N. and Ibeagha-Awemu, E.M., (2017). Non-Coding RNA Roles in Ruminant Mammary Gland Development and Lactation. In *Current Topics in Lactation*. InTech.

dos Santos, R.D.C. and Marin, J.M., (2005). Isolation of Candida spp. from mastitic bovine milk in Brazil. *Mycopathologia*, *159*(2), pp.251-253.

Dotta, L., Scomodon, O., Padoan, R., Timpano, S., Plebani, A., Soresina, A., Lougaris, V., Concolino, D., Nicoletti, A., Giardino, G. and Licari, A., (2016). Clinical heterogeneity of dominant chronic mucocutaneous candidiasis disease: presenting as treatment-resistant candidiasis and chronic lung disease. *Clinical Immunology*, *164*, pp.1-9.

Duncan, S., Zhong, Z., Wen, Z. and Darnell, J. (1997). STAT signaling is active during early mammalian development. Developmental Dynamics, 208(2), pp.190-198.

Durbin, J.E., Hackenmiller, R., Simon, M.C. and Levy, D.E., (1996). Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell*, *84*(3), pp.443-450.

Dybus, A., (2002). Associations of growth hormone (GH) and prolactin (PRL) genes polymorphisms with milk production traits in Polish Black-and-White cattle. *Anim Sci Pap Rep*, *20*(20), pp.203-212.

Ensembl.org. (2017). *Predicted Data*. [online] Available at: http://www.ensembl.org/info/genome/variation//predicted_data.html [Accessed 22 Nov. 2017].

Ernst, M., Inglese, M., Waring, P., Campbell, I., Bao, S., Clay, F., Alexander, W., Wicks, I., Tarlinton, D., Novak, U., Heath, J. and Dunn, A. (2001). Defective Gp130-Mediated Signal Transducer and Activator of Transcription (Stat) Signaling Results in Degenerative Joint Disease, Gastrointestinal Ulceration, and Failure of Uterine Implantation. *The Journal of Experimental Medicine*, 194(2), pp.189-204.

Egger-Danner, C., Cole, J., Pryce, J., Gengler, N., Heringstad, B., Bradley, A. and Stock, K. (2014). Invited review: overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal*, 9(02), pp.191-207

Fernando, R.L. and Garrick, D., 2013. Bayesian methods applied to GWAS. In *Genome-wide association studies and genomic prediction* (pp. 237-274). Humana Press, Totowa, NJ.

Fisher, R.A., (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *52*(2), pp.399-433.

Fisher PJ, Hyndman DL, Bixley MJ, Oback FC, Popovic L, McGowan LT, Berg BC, Wells DN (2012) Potential for genomic selection of bovine embryos. *Proceedings of the New Zealand Society of Animal Production* 72, 156–158.

Flisikowski, K., Venhoranta, H., Nowacka-Woszuk, J., McKay, S., Flyckt, A., Taponen, J., Schnabel, R., Schwarzenbacher, H., Szczerbal, I., Lohi, H., Fries, R., Taylor, J., Switonski, M. and Andersson, M. (2010). A Novel Mutation in the Maternally Imprinted PEG3 Domain Results in a Loss of MIMT1 Expression and Causes Abortions and Stillbirths in Cattle (Bos taurus). *PLoS ONE*, 5(11), p.e15116.

Flisikowski, K., Venhoranta, H., Bauersachs, S., Hanninen, R., Furst, R., Saalfrank, A., Ulbrich, S., Taponen, J., Lohi, H., Wolf, E., Kind, A., Andersson, M. and Schnieke,

A. (2012). Truncation of MIMT1 Gene in the PEG3 Domain Leads to Major Changes in Placental Gene Expression and Stillbirth in Cattle. *Biology of Reproduction*, 87(6), pp.140-140.

Fontanesi, L., Scotti, E., Samorè, A., Bagnato, A. and Russo, V. (2015). Association of 20 candidate gene markers with milk production and composition traits in sires of Reggiana breed, a local dairy cattle population. *Livestock Science*, 176, pp.14-21.

Forde, N., Simintiras, C., Sturmey, R., Mamo, S., Kelly, A., Spencer, T., Bazer, F. and Lonergan, P. (2014). Amino Acids in the Uterine Luminal Fluid Reflects the Temporal Changes in Transporter Expression in the Endometrium and Conceptus during Early Pregnancy in Cattle. *PLoS ONE*, 9(6), p.e100010.

Fouz, R., Yus, E., Sanjuán, M.L. and Diéguez, F.J., (2014) Reasons for culling among Holstein dairy cattle in herds in the Dairy Herd Improvement Program. *ITEA*, *110*(2), pp.171-186.

Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I. and Hardison, R.C., (2003). Cross-species sequence comparisons: a review of methods and available resources. *Genome Research*, *13*(1), pp.1-12.

Fritz, S., Capitan, A., Djari, A., Rodriguez, S., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D. and Boichard, D. (2013). Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PLoS ONE*, 8(6), p.e65550.

Gallego, M., Binart, N., Robinson, G., Okagaki, R., Coschigano, K., Perry, J., Kopchick, J., Oka, T., Kelly, P. and Hennighausen, L. (2001). Prolactin, Growth Hormone, and Epidermal Growth Factor Activate Stat5 in Different Compartments of Mammary Tissue and Exert Different and Overlapping Developmental Effects. *Developmental Biology*, 229(1), pp.163-175.

Garrick, D.J., Taylor, J.F. and Fernando, R.L., (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, *41*(1), p.55.

Genome.jp. (2018). GenomeNet Database Resources. [online] Available at: https://www.genome.jp/entry/map04630 [Accessed 2 Jul. 2018].

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R. and Butler, D., (2009). ASReml user guide release 3.0. *VSN International Ltd, Hemel Hempstead, UK*.

Givens, M. (2006). A clinical, evidence-based approach to infectious causes of infertility in beef cattle. *Theriogenology*, 66(3), pp.648-654.

Goddard, M. and Hayes, B. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6), pp.381-391.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P. and Spelman, R., 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome research*, *12*(2), pp.222-231.

Grisart, B., Farnir, F., Karim, L., Cambisano, N., Kim, J., Kvasz, A., Mni, M., Simon, P., Frere, J., Coppieters, W. and Georges, M. (2004). Genetic and functional confirmation of the causality of the DGAT1 K232A quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences*, 101(8), pp.2398-2403.

Grooms, D. (2004). Reproductive consequences of infection with bovine viral diarrhea virus. *Veterinary Clinics of North America: Food Animal Practice*, 20(1), pp.5-19.

Guimaraes, E., Ruane, J., Scherf, B., Sonnino, A. and Dargie, J., (2007). Marker assisted selection. *Food and Agricultural Organization of the United Nations, Rome, Italy*, pp.427-440.

Harhay, G., Smith, T., Alexander, L., Haudenschild, C., Keele, J., Matukumalli, L., Schroeder, S., Van Tassell, C., Gresham, C., Bridges, S., Burgess, S. and Sonstegard, T. (2010). An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. *Genome Biology*, 11(10), p.R102.

Haricharan, S. and Li, Y. (2014). STAT signaling in mammary gland differentiation, cell survival and tumorigenesis. *Molecular and Cellular Endocrinology*, 382(1), pp.560-569.

Harper, J., Yuan, J., Tan, J., Visan, I. and Guidos, C. (2003). Notch signaling in development and disease. *Clinical Genetics*, 64(6), pp.461-472.

Harris, C., Haas, J., Streeper, R., Stone, S., Kumari, M., Yang, K., Han, X., Brownell, N., Gross, R., Zechner, R. and Farese, R. (2011). DGAT enzymes are required for triacylglycerol synthesis and lipid droplets in adipocytes. *Journal of Lipid Research*, 52(4), pp.657-667.

HAVENSTEIN, G., FERKET, P., SCHEIDELER, S. and RIVES, D. (1994). Carcass Composition and Yield of 1991 vs 1957 Broilers When Fed "Typical" 1957 and 1991 Broiler Diets. *Poultry Science*, 73(12), pp.1795-1804.

He, X., Chu, M., Qiao, L., He, J., Wang, P., Feng, T., Di, R., Cao, G., Fang, L. and An, Y. (2011). Polymorphisms of STAT5A gene and their association with milk production traits in Holstein cows. *Molecular Biology Reports*, 39(3), pp.2901-2907.

HEYEN, D., WELLER, J., RON, M., BAND, M., BEEVER, J., FELDMESSER, E., DA, Y., WIGGANS, G., VanRADEN, P. and LEWIN, H. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. *Physiological Genomics*, 1(3), pp.165-175.

Hirschhorn, J., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. *Genetics in Medicine*, 4(2), pp.45-61.

Hyttel, P., Fair, T., Callesen, H. and Greve, T. (1997). Oocyte growth, capacitation and final maturation in cattle. *Theriogenology*, 47(1), pp.23-32.

Icbf.com. (2017). *About ICBF | ICBF*. [online] Available at: https://www.icbf.com/wp/?page_id=27 [Accessed 11 Sep. 2017].

ICBF.com (2016). *Phenotypic trends in Dairy | ICBF*. [online] Icbf.com. Available at: https://www.icbf.com/wp/?p=6939 [Accessed 14 Sep. 2017].

Ihfa.ie. (2017). *IHFA - IHFA Website*. [online] Available at: http://ihfa.ie/ [Accessed 12 Sep. 2017].

Jianqin, S., Leiming, X., Lu, X., Yelland, G., Ni, J. and Clarke, A. (2015). Erratum to: 'Effects of milk containing only A2 beta casein versus milk containing both A1 and A2 beta casein proteins on gastrointestinal physiology, symptoms of discomfort, and cognitive behavior of people with self-reported intolerance to traditional cows' milk'. *Nutrition Journal*, 15(1).

Kadri, N., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., Nielsen, U., Panitz, F., Aamand, G., Schulman, N., Georges, M., Vilkki, J., Lund, M. and Druet, T. (2014). A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLoS Genetics*, 10(1), p.e1004049.

Kaupe, B., Brandt, H., Prinzenberg, E. and Erhardt, G. (2007). Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle1. *Journal of Animal Science*, 85(1), pp.11-21.

Khatib, H., Monson, R., Schutzkus, V., Kohl, D., Rosa, G. and Rutledge, J. (2008). Mutations in the STAT5A Gene Are Associated with Embryonic Survival and Milk Composition in Cattle. *Journal of Dairy Science*, 91(2), pp.784-793.

Khatib, H., Huang, W., Mikheil, D., Schutzkus, V. and Monson, R. (2009). Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. *Journal of Dairy Science*, 92(12), pp.6186-6191.

Kearney JF, Cromie AR, Berry DP (2010) Implementation and update of genomic selection in Irish Holstein-Friesian dairy cattle. *In '9th world congress on genetics applied to livestock production, Leipzig, Germany'*. Available at http://www.kongressband.de/wcgalp2010/assets/pdf/0889. pdf [verified 4 April 2013]

Kegg.jp. (2018). KEGG PATHWAY: Notch signaling pathway - Bos taurus (cow). [online] Available at: https://www.kegg.jp/kegg-

bin/show_pathway?org_name=bta&mapno=04330&mapscale=1.0&show_description=hide [Accessed 3 Sep. 2018].

Kim, J. and Georges, M. (2002). Evaluation of a New Fine-mapping Method Exploiting Linkage Disequilibrium: a Case Study Analysing a QTL with Major Effect on Milk Composition on Bovine Chromosome 14. *Asian-Australasian Journal of Animal Sciences*, 15(9), pp.1250

Kiplagat, S., Limo, M., and Kosgey, I. (2012). Genetic Improvement of Livestock for Milk Production. *Milk Production - Advanced Genetic Traits, Cellular Mechanism, Animal Management and Health*

Komisarek, J., Michalak, A. and Walendowska, A., (2011). The effects of polymorphisms in DGAT 1, GH and GHR genes on reproduction and production traits in Jersey cows. *Animal Science Papers and Reports*, *29*(1), pp.29-36.

Kornienko, A., Guenzl, P., Barlow, D. and Pauler, F. (2013). Gene regulation by the act of long non-coding RNA transcription. *BMC Biology*, 11(1), p.59.

Kwok, P. (2003). *Single nucleotide polymorphisms*. Totowa, N.J.: Humana Press.

Lechniak, D., Strabel, T., Przybyła, D., Machnik, G. and Świtoński, M. (2002). GH and CSN3 gene polymorphisms and their impact on milk traits in cattle. *Journal of Animal and Feed Sciences*, 11(1), pp.39-45.

Lee, J., Kim, T., Oh, S., Yoo, J., Akira, S., Ku, B., Lydon, J. and Jeong, J. (2013). Signal transducer and activator of transcription-3 (Stat3) plays a critical role in implantation via progesterone receptor in uterus. *The FASEB Journal*, 27(7), pp.2553-2563.

Leroy, J., Vanholder, T., Opsomer, G., Van Soom, A. and Kruif, A. (2006). The In Vitro Development of Bovine Oocytes after Maturation in Glucose and beta-Hydroxybutyrate Concentrations Associated with Negative Energy Balance in Dairy Cows. *Reproduction in Domestic Animals*, 41(2), pp.119-123.

Llewellyn, S., Fitzpatrick, R., Kenny, D., Murphy, J., Scaramuzzi, R. and Wathes, D. (2007). Effect of negative energy balance on the insulin-like growth factor system in

pre-recruitment ovarian follicles of post-partum dairy cows. *Reproduction*, 133(3), pp.627-639.

Li, S. and Rosen, J. (1995). Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. *Molecular and Cellular Biology*, 15(4), pp.2063-2070.

Liefers, S.C., Veerkamp, R.F., Te Pas, M.F.W., Delavaud, C., Chilliard, Y. and Van der Lende, T., (2003). Leptin concentrations in relation to energy balance, milk yield, intake, live weight, and estrus in dairy cows. *Journal of dairy science*, *86*(3), pp.799-807.

Liu, Y., Qin, X., Song, X.Z.H., Jiang, H., Shen, Y., Durbin, K.J., Lien, S., Kent, M.P., Sodeland, M., Ren, Y. and Zhang, L., 2009. Bos taurus genome assembly. *BMC genomics*, *10*(1), p.180.

Looft, C., Reinsch, N., Karall-Albrecht, C., Paul, S., Brink, M., Thomsen, H., Brockmann, G., Kühn, C., Schwerin, M. and Kalm, E. (2001). A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine Chromosome 14. *Mammalian Genome*, 12(8), pp.646-650.

Lucy, M. (2001). Reproductive Loss in High-Producing Dairy Cattle: Where Will It End?. *Journal of Dairy Science*, 84(6), pp.1277-1293.

Luscombe, N., Greenbaum, D. and Gerstein, M. (2001). What is Bioinformatics? A Proposed Definition and Overview of the Field. *Methods of Information in Medicine*, 40(04), pp.346-358.

Lush, J.L., (1937). Animal breeding plans. Collegiate press Inc, Ames, IA.

Lush, J.L., (1947). Family merit and individual merit as bases for selection. Part I. *The American Naturalist*, *81*(799), pp.241-261.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P. and Bradley, D.G., (1997). Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (Bos taurus and Bos indicus). *Genetics*, *146*(3), pp.1071-1086.

Mackay, T., Stone, E. and Ayroles, J. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, 10(8), pp.565-577.

MacLeod, I.M., Bowman, P.J., Vander Jagt, C.J., Haile-Mariam, M., Kemper, K.E., Chamberlain, A.J., Schrooten, C., Hayes, B.J. and Goddard, M.E., (2016). Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC genomics*, *17*(1), p.144.

Malher, X., Beaudeau, F. and Philipot, J. (2006). Effects of sire and dam genotype for complex vertebral malformation (CVM) on risk of return-to-service in Holstein dairy cows and heifers. *Theriogenology*, 65(6), pp.1215-1225.

Massella, E., Piva, S., Giacometti, F., Liuzzo, G., Zambrini, A. and Serraino, A. (2017). Evaluation of bovine beta casein polymorphism in two dairy farms located in northern Italy. *Italian Journal of Food Safety*, 6(3).

McClure, M.C., Mullen, M., Waters, S.M., Kearney, F., McClure, J., Flynn, P. and Weld, R., 2016. P6001 Effectively managing bovine genetic disease risk via genotyping the Irish national herd. *Journal of Animal Science*, *94*(7supplement4), pp.148-148.

Meuwissen, T.H.E and Hayes, B.J. and Goddard, M.E., (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), pp.1819-1829.

Meydan, H., Yildiz, M. and Agerholm, J. (2010). Screening for bovine leukocyte adhesion deficiency, deficiency of uridine monophosphate synthase, complex vertebral malformation, bovine citrullinaemia, and factor XI deficiency in Holstein cows reared in Turkey. *Acta Veterinaria Scandinavica*, 52(1), p.56.

Mukhopadhyay, S., Wyszomierski, S., Gronostajski, R. and Rosen, J. (2001). Differential Interactions of Specific Nuclear Factor I Isoforms with the Glucocorticoid Receptor and STAT5 in the Cooperative Regulation of WAP Gene Transcription. *Molecular and Cellular Biology*, 21(20), pp.6859-6869.

Mullen, M., Lynch, C., Waters, S., Howard, D., O'Boyle, P., Kenny, D., Buckley, F., Horan, B. and Diskin, M. (2011). Single nucleotide polymorphisms in the growth hormone and insulin-like growth factor-1 genes are associated with milk production,

body condition score and fertility traits in dairy cows. *Genetics and Molecular Research*, 10(3), pp.1819-1830.

Mullen, M.P., McClure, M.C., Kearney, J.F., Waters, S.M., Weld, R., Flynn, P., Creevey, C.J., Cromie, A.R. and Berry, D.P., (2013). Development of a custom SNP chip for dairy and beef cattle breeding, parentage and research. *Interbull Bulletin*, (47).

Mullen, M., Bazer, F., Wu, G., Parr, M., Evans, A., Crowe, M. and Diskin, M. (2014). Effects of systemic progesterone during the early luteal phase on the availabilities of amino acids and glucose in the bovine uterine lumen. *Reproduction, Fertility and Development*, 26(2), p.282.

Nani, J.P., Raschia, M.A., Carignano, H. *et al.* Single nucleotide polymorphisms in candidate genes and their relation with somatic cell scores in Argentinean dairy cattle. *J Appl Genetics* **56,** 505–513 (2015). https://doi.org/10.1007/s13353-015-0278-5

Ncbi.nlm.nih.gov. (2018). *National Center for Biotechnology Information*. [online] Available at: https://www.ncbi.nlm.nih.gov/ [Accessed 2 Jul. 2018].

Phillips, P. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, 9(11), pp.855-867.

Ng-Kwai-Hang, K., Hayes, J., Moxley, J. and Monardes, H. (1984). Association of Genetic Variants of Casein and Milk Serum Proteins with Milk, Fat, and Protein Production by Dairy Cattle. *Journal of Dairy Science*, 67(4), pp.835-840.

Ng, P. and Henikoff, S. (2001). Predicting Deleterious Amino Acid Substitutions. *Genome Research*, 11(5), pp.863-874.

Ogorevc, J., Kunej, T., Razpet, A. and Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*, 40(6), pp.832-851.

Oikonomou, G., Michailidis, G., Kougioumtzis, A., Avdi, M. and Banos, G. (2011). Effect of polymorphisms at the STAT5A and FGF2 gene loci on reproduction, milk yield and lameness of Holstein cows. *Research in Veterinary Science*, 91(2), pp.235-239.

Oltenacu, P.A. and Broom, D.M., (2010). The impact of genetic selection for increased milk yield on the welfare of dairy cows. *Animal welfare*, *19*(1), pp.39-49.

Omia.org. (2018). *OMIA - Online Mendelian Inheritance in Animals*. [online] Available at: http://omia.org/home/ [Accessed 2 Jul. 2018].

Ortega-Requena, S. and Rebouillat, S. (2015). Bigger data open innovation: potential applications of value-added products from milk and sustainable valorization of by-products from the dairy industry. *Green Chem.*, 17(12), pp.5100-5113.

Pausch, H., Kölle, S., Wurmser, C., Schwarzenbacher, H., Emmerling, R., Jansen, S., Trottmann, M., Fuerst, C., Götz, K. and Fries, R. (2014). A Nonsense Mutation in TMEM95 Encoding a Nondescript Transmembrane Protein Causes Idiopathic Male Subfertility in Cattle. *PLoS Genetics*, 10(1), p.e1004044.

Pausch, H., Venhoranta, H., Wurmser, C., Hakala, K., Iso-Touru, T., Sironen, A., Vingborg, R., Lohi, H., Söderquist, L., Fries, R. and Andersson, M. (2016). A frameshift mutation in ARMC3 is associated with a tail stump sperm defect in Swedish Red (Bos taurus) cattle. *BMC Genetics*, 17(1).

Prince, V., Holley, S., Bally-Cuif, L., Prabhakaran, B., Oates, A., Ho, R. and Vogt, T. (2001). Zebrafish lunatic fringe demarcates segmental boundaries. Mechanisms of Development, 105(1-2), pp.175-180

Procaccini, C., Lourenco, E.V., Matarese, G. and Cava, A.L., (2009). Leptin signaling: A key pathway in immune responses. *Current signal transduction therapy*, *4*(1), pp.22-30.

Pryce, J., Royal, M., Garnsworthy, P. and Mao, I. (2004). Fertility in the high-producing dairy cow. *Livestock Production Science*, 86(1-3), pp.125-135

Quaas, R. (1976). Computing the Diagonal Elements and Inverse of a Large Numerator Relationship Matrix. *Biometrics*, 32(4), p.949.

Quinn, J. and Chang, H. (2016). Unique features of long non-coding RNA biogenesis and function. *Nature Reviews Genetics*, 17(1), pp.47-62.

Rachagani, S. and Gupta, I. (2008). Bovine kappa-casein gene polymorphism and its association with milk production traits. *Genetics and Molecular Biology*, 31(4), pp.893-897.

Regrna2.mbc.nctu.edu.tw. (2018). *RegRNA 2.0*. [online] Available at: http://regrna2.mbc.nctu.edu.tw/documentation.html [Accessed 2 Jul. 2018].

Roberts, R. (2005). How restriction enzymes became the workhorses of molecular biology. *Proceedings of the National Academy of Sciences*, 102(17), pp.5905-5908.

Rodriguez-Martinez, H., Hultgren, J., Båge, R., Bergqvist, A.S., Svensson, C., Bergsten, C., Lidfors, L., Gunnarsson, S., Algers, B., Emanuelson, U. and Berglund, B., (2008). Reproductive performance in high-producing dairy cows: can we sustain it under current practice. *IVIS Reviews in veterinary Medicine*.

Rodriguez-Zas, S., Southey, B., Heyen, D. and Lewin, H. (2002). Interval and Composite Interval Mapping of Somatic Cell Score, Yield, and Components of Milk in Dairy Cattle. *Journal of Dairy Science*, 85(11), pp.3081-3091.

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

Rychtarova, J., Sztankoova, Z., Kyselova, J., Zink, V., Stipkova, M., Vacek, M. and Stolc, L., (2014). Effect of DGAT1, BTN1A1, OLR1, and STAT1 genes on milk production and reproduction traits in the Czech Fleckvieh breed. *Czech Journal of Animal Science*, *59*(2), pp.45-53.

Sasaki, S., Ibi, T., Akiyama, T., Fukushima, M. and Sugimoto, Y. (2016). Loss of maternal ANNEXIN A10 via a 34-kb deleted-type copy number variation is associated with embryonic mortality in Japanese Black cattle. *BMC Genomics*, 17(1).

Schlötterer, C. (2004). Opinion: The evolution of molecular markers — just a matter of fashion?. *Nature Reviews Genetics*, 5(1), pp.63-69.

Schwenger, B., Schöber, S. and Simon, D. (1993). DUMPS Cattle Carry a Point Mutation in the Uridine Monophosphate Synthase Gene. *Genomics*, 16(1), pp.241-244.

Selvaggi, M., Dario, C., Normanno, G., Celano, G. and Dario, M. (2009). Genetic polymorphism of STAT5A protein: relationships with production traits and milk composition in Italian Brown cattle. Journal of Dairy Research, 76(04), p.441.

Selvaggi, M., Albarella, S., Dario, C., Peretti, V. and Ciotola, F. (2016). Association of STAT5A Gene Variants with Milk Production Traits in Agerolese Cattle. Biochemical Genetics, 55(2), pp.158-167.

Serth, K., Schuster-Gossler, K., Cordes, R. and Gossler, A. (2003). Transcriptional oscillation of Lunatic fringe is essential for somitogenesis. Genes & Development, 17(7), pp.912-925.

Shanks, R. and Greiner, M. (1992). Relationship Between Genetic Merit of Holstein Bulls and Deficiency of Uridine-5'-Monophosphate Synthase. *Journal of Dairy Science*, 75(7), pp.2023-2029.

Sharif, S., Mallard, B., Wilkie, B., Sargeant, J., Scott, H., Dekkers, J. and Leslie, K. (1998). Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) alleles with occurrence of disease and milk somatic cell score in Canadian dairy cattle. *Animal Genetics*, 29(3), pp.185-193.

Shorthorn.org. (2017). *History – American Shorthorn Association*. [online] Available at: https://shorthorn.org/history/ [Accessed 13 Sep. 2017].

Shuster, D., Kehrli, M., Ackermann, M. and Gilbert, R. (1992). Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proceedings of the National Academy of Sciences*, 89(19), pp.9225-9229.

Smith, H. and Wilcox, K. (1970). A Restriction enzyme from Hemophilus influenzae. *Journal of Molecular Biology*, 51(2), pp.379-391.

Smith, S., Cases, S., Jensen, D., Chen, H., Sande, E., Tow, B., Sanan, D., Raber, J., Eckel, R. and Farese, R. (2000). Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. *Nature Genetics*, 25(1), pp.87-90.

Song, Z. (2013). Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Molecular Aspects of Medicine*, 34(2-3), pp.590-600.

Sonstegard, T., Cole, J., VanRaden, P., Van Tassell, C., Null, D., Schroeder, S., Bickhart, D. and McClure, M. (2013). Identification of a Nonsense Mutation in CWC15 Associated with Decreased Reproductive Efficiency in Jersey Cattle. *PLoS ONE*, 8(1), p.e54872.

Sparrow, D., Chapman, G., Wouters, M., Whittock, N., Ellard, S., Fatkin, D., Turnpenny, P., Kusumi, K., Sillence, D. and Dunwoodie, S. (2006). Mutation of the LUNATIC FRINGE Gene in Humans Causes Spondylocostal Dysostosis with a Severe Vertebral Phenotype. *The American Journal of Human Genetics*, 78(1), pp.28-37.

Spelman, R.J., Hayes, B.J. and Berry, D.P., (2013). Use of molecular technologies for the advancement of animal breeding: genomic selection in dairy cattle populations in Australia, Ireland and New Zealand. *Animal Production Science*, *53*(9), pp.869-875.

Spelman, R., Ford, C., McElhinney, P., Gregory, G. and Snell, R. (2002). Characterization of the DGAT1 Gene in the New Zealand Dairy Population. *Journal of Dairy Science*, 85(12), pp.3514-3517.

Stephens, M., Smith, N. and Donnelly, P. (2001). A New Statistical Method for Haplotype Reconstruction from Population Data. *The American Journal of Human Genetics*, 68(4), pp.978-989.

Stephens, M. and Balding, D. (2009). Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics*, 10(10), pp.681-690.

Stone, E. and Sidow, A. (2005). Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Research*, 15(7), pp.978-986.

Syvänen, A. (2001). Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews Genetics*, 2(12), pp.930-942.

Takeda, K., Noguchi, K., Shi, W., Tanaka, T., Matsumoto, M., Yoshida, N., Kishimoto, T. and Akira, S. (1997). Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proceedings of the National Academy of Sciences*, 94(8), pp.3801-3804.

Takeda, A., Hamano, S., Yamanaka, A., Hanada, T., Ishibashi, T., Mak, T.W., Yoshimura, A. and Yoshida, H., (2003). Cutting edge: role of IL-27/WSX-1 signaling for induction of T-bet through activation of STAT1 during initial Th1 commitment. *The Journal of Immunology*, *170*(10), pp.4886-4890.

Takezaki, S., Yamada, M., Kato, M., Park, M.J., Maruyama, K., Yamazaki, Y., Chida, N., Ohara, O., Kobayashi, I. and Ariga, T., (2012). Chronic mucocutaneous candidiasis caused by a gain-of-function mutation in the STAT1 DNA-binding domain. *The Journal of Immunology*, *189*(3), pp.1521-1526.

Tang, H. and Thomas, P. (2016). PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, 32(14), pp.2230-2232.

Testoni, S., Diana, A., Olzi, E. and Gentile, A. (2008). Brachyspina syndrome in two Holstein calves. *The Veterinary Journal*, 177(1), pp.144-146.

Thaller, G., Krämer, W., Winter, A., Kaupe, B., Erhardt, G. and Fries, R. (2003). Effects of variants on milk production traits in German cattle breeds. *Journal of Animal Science*, 81(8), p.1911.

Thatcher, W., Guzeloglu, A., Mattos, R., Binelli, M., Hansen, T. and Pru, J. (2001). Uterine-conceptus interactions and reproductive failure in cattle. *Theriogenology*, 56(9), pp.1435-1450.

Thomsen, B., Horn, P., Panitz,, F., Bendixen, E., Petersen, A., Holm, L., Nielsen, V., Agerholm, J., Arnbjerg, J. and Bendixen, C. (2006). A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Research*, 16(1), pp.97-105

Thundathil, J., Dance, A. and Kastelic, J. (2016). Fertility management of bulls to improve beef cattle productivity. *Theriogenology*, 86(1), pp.397-405.

Tsiaras, A., Bargouli, G., Banos, G. and Boscos, C. (2005). Effect of Kappa-Casein and Beta-Lactoglobulin Loci on Milk Production Traits and Reproductive Performance of Holstein Cows. *Journal of Dairy Science*, 88(1), pp.327-334.

Turner, R.M., Musse, M.P., Mandal, A.R.A.B.I.N.D.A., Klotz, K.E.N., Jayes, F.C., Herr, J.C., Gerton, G.L., Moss, S.B. and Chemes, H.E., (2001). Molecular genetic analysis of two human sperm fibrous sheath proteins, AKAP4 and AKAP3, in men with dysplasia of the fibrous sheath. *Journal of andrology*, *22*(2), pp.302-315.

Usha, A., Simpson, S. and Williams, J. (2009). Probability of random sire exclusion using microsatellite markers for parentage verification. *Animal Genetics*, 26(3), pp.155-161.

VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. and Schenkel, F.S., (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of dairy science*, *92*(1), pp.16-24.

Venhoranta, H., Bauersachs, S., Taponen, J., Lohi, H., Taira, T., Andersson, M., Kind, A., Schnieke, A. and Flisikowski, K. (2013). Fetal growth restriction caused by MIMT1 deletion alters brain transcriptome in cattle. *International Journal of Developmental Neuroscience*, 31(7), pp.463-467.

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34(3), p.275.

Wakchaure, R. and Ganguly, S. (2015). Marker Assisted Selection (MAS) in Animal Breeding: A Review. *Journal of Drug Metabolism & Toxicology*, 06(05).

Walsh, C., Guinee, T., Reville, W., Harrington, D., Murphy, J., O'Kennedy, B. and FitzGerald, R. (1998). Influence of κ-Casein Genetic Variant on Rennet Gel Microstructure, Cheddar Cheesemaking Properties and Casein Micelle Size. *International Dairy Journal*, 8(8), pp.707-714.

Walsh, S., Williams, E. and Evans, A. (2011). A review of the causes of poor fertility in high milk producing dairy cows. *Animal Reproduction Science*, 123(3-4), pp.127-138.

Wang and L. A. Goonewardene, Z. (2004). The use of MIXED models in the analysis of animal experiments with repeated measures data. *Canadian Journal of Animal Science*, 84(1), pp.1-11.

Waters, S., McCabe, M., Howard, D., Giblin, L., Magee, D., MacHugh, D. and Berry, D. (2011). Associations between newly discovered polymorphisms in the Bos Taurus growth hormone receptor gene and performance traits in Holstein-Friesian dairy cattle. *Animal Genetics*, 42(1), pp.39-49.

Wathes, D., Fenwick, M., Cheng, Z., Bourne, N., Llewellyn, S., Morris, D., Kenny, D., Murphy, J. and Fitzpatrick, R. (2007). Influence of negative energy balance on cyclicity and fertility in the high producing dairy cow. *Theriogenology*, 68, pp.S232-S241.

Watson, C. and Neoh, K. (2008). The Stat family of transcription factors have diverse roles in mammary gland development. *Seminars in Cell & Developmental Biology*, 19(4), pp.401-406.

Winter, A., Kramer, W., Werner, F., Kollers, S., Kata, S., Durstewitz, G., Buitkamp, J., Womack, J., Thaller, G. and Fries, R. (2002). Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. *Proceedings of the National Academy of Sciences*, 99(14), pp.9300-9305.

Wong, D., Camirand, W., Pavlath, A., Parris, N. and Friedman, M. (1996). Structures and functionalities of milk proteins∗. *Critical Reviews in Food Science and Nutrition*, 36(8), pp.807-844.

Wright, s., (1921). Systems of mating, 1-5. *Genetics* 16:97-`159

Yuan, Z., Li, J., Li, J., Gao, X., Gao, H. and Xu, S., (2013). Effects of DGAT1 gene on meat and carcass fatness quality in Chinese commercial cattle. *Molecular biology reports*, *40*(2), pp.1947-1954.

Zabolewicz, T., Czarnik, U., Strychalski, J., Pareek, C. and Pierzchała, M. (2011). The association between microsatellite Bm6438 and milk performance traits in Polish Holstein-Friesian cattle. *Czech Journal of Animal Science*, 56(No. 3), pp.107-113**.**

Zeder, M., Emshwiller, E., Smith, B. and Bradley, D. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends in Genetics*, 22(3), pp.139-155.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G. and Gil, L., (2017). Ensembl 2018. *Nucleic acids research*, *46*(D1), pp.D754-D761.

Zhang, N. and Gridley, T. (1998). Defects in somite formation in lunatic fringe-deficient mice. *Nature*, 394(6691), pp.374-377.

Zhang, N., Norton, C. and Gridley, T. (2002). Segmentation defects of Notch pathway mutants and absence of a synergistic phenotype in lunatic fringe/radical fringe double mutant mice. *genesis*, 33(1), pp.21-28.

Zhao, H., Pfeiffer, R. and Gail, M. (2003). Haplotype analysis in population genetics and association studies. *Pharmacogenomics*, 4(2), pp.171-178.

Zhu, M., Yu, M. and Zhao, S. (2009). Understanding Quantitative Genetics in the Systems Biology Era. *International Journal of Biological Sciences*, pp.161-170.

Appendix I

The following section of this appendix displays all the abstracts that have been submitted to various conferences (ENVIRON, ASAS, EAAP, ESDAR), for publication based on the results obtained to date in this research study.

No evidence of an association between lethal recessives CVM and Brachyspina and carcass and health traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Mathew McCLure[2], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland

Corresponding author email: l.ratclife@research.ait.ie

The frequency of genes with lethal effects on embryo survival is economically important in livestock production. The maintenance of such deleterious mutations in cattle populations is partly due to the intense selection for milk yield in dairy cattle in recent decades. The elimination from or management of such mutations in the national breeding herd is a desirable objective, however, estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous. Therefore, the objective of the current study was to estimate the effects of two such lethal recessives, CVM and Brachyspina, on carcass and health related traits in Irish dairy cattle. CVM and Brachyspina SNP genotypes on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). Phenotypes for carcass and health traits, also obtained from the ICBF, were expressed as predicted transmitting abilities (PTAs). Associations between each SNP and PTA was analysed in ASREML using a weighted mixed animal model. No associations (P>0.05) between CVM and Brachyspina and either somatic cell score (n=5747), carcass weight (n=3194), cull cow weight (n=1374), carcass conformation (n=518) and carcass fat (n=360) were observed in the sample set tested. No evidence was obtained to support maintenance or strategic matings for carrier animals suggesting that elimination of carriers of either CVM or Brachyspina would not reduce the genetic merit of the national herd in relation to carcass and health related traits.

Keywords: lethal mutations, pleiotropy, dairy cattle, genetics

No evidence of an association between lethal recessives CVM and Brachyspina and carcass and health traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Mathew McClure[2], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland
Corresponding author email: l.ratcliffe@research.ait.ie

Introduction

The frequency of genes with lethal effects on embryo survival is economically important in livestock production. The maintenance of such deleterious mutations in cattle populations is partly due to the intense selection for milk yield in dairy cattle in recent decades (Fritz et al., 2013) . The elimination from or management of such mutations in the national breeding herd is a desirable objective, however, estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous.

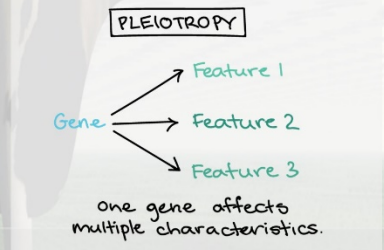
The objective of the current study was to estimate the effects of two such lethal recessives, complex vertebral malformation (CVM) and brachyspina, on carcass and health related traits in Irish dairy cattle.

CVM

- Caused by a missense mutation in the bovine SLC35A3 gene which encodes for a UDP-N-acetylglucosamine transporter (Thomsen et al., 2006).

Brachyspina

- The genetic cause of BS development is a 3.3 kb deletion, removing exons 25-27 within the Fanconi anaemia complementation group I (FANCI) gene on chromosome 21 (Charlier et al., 2012).



Materials and Methods

CVM and Brachyspina SNP genotypes on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). The phenotypes for carcass and health traits, also obtained from the ICBF, were expressed as predicted transmitting abilities (PTAs The association between each SNP and PTA was analysed in ASREML using a weighted mixed animal model.

Results

Table 1.0 – Minor allele frequencies observed for CVM and brachyspina in the Holstein Friesian population

Mutation	Minor allele frequency
CVM	0.015
Brachyspina	0.008

No associations (P>0.05) between CVM and Brachyspina and either somatic cell score (n=5747), carcass weight (n=3194), cull cow weight (n=1374), carcass conformation (n=518) and carcass fat (n=360) were observed in the sample set tested.

Discussion

No evidence was obtained therefore to support maintenance or strategic matings for carrier animals suggesting that elimination of carriers of either CVM or Brachyspina would not reduce the genetic merit of the national herd in relation to carcass and health related traits.

References
- Charlier, C., Agerholm, J., Coppieters, W., Karlskov-Mortensen, P., Li, W., de Jong, G., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., Ahariz, N., Mullaart, E., Georges, M. and Fredholm, M. (2012). A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. PLoS ONE, 7(8), p.e43085.
- Fritz, S., Capitan, A., Djari, A., Rodriguez, S., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D. and Boichard, D. (2013). Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. PLoS ONE, 8(6), p.e65550.
- Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A., Holm, L., Nielsen, V., Agerholm, J., Arnbjerg, J. and Bendixen, C. (2006). A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. Genome Research, 16(1), pp.97-105

A novel association between a STAT1 mutation and carcass conformation in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McCLure[2], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co.Cork, Ireland

Corresponding author email: l.ratclife@research.ait.ie

Signal transducer and activator of transcription (*STAT*) genes encode for a family of proteins involved in pre and post-natal growth and development. In cattle, variants in these genes have been associated with economically important traits including milk production and embryonic survival. The objective of this study was to estimate the effects of polymorphisms in the *STAT1*, *STAT3* and *STAT5* genes on carcass and health traits in dairy cattle. *STAT* genotypes (n=8) on 10,707 dairy cattle were obtained through the Irish cattle breeding federation (ICBF). Phenotypes for carcass and health traits, also obtained from the ICBF, were expressed as predicted transmitting abilities (PTAs). The association analysis included n=5747, 3194, 1374, 518 and 360 cows for somatic cell score, carcass weight, cull cow weight, carcass conformation and carcass fat, respectively. Associations between each SNP and PTA were analysed in ASREML using a weighted mixed animal model. A significant association (P<0.0001) between *STAT1* (2697) and carcass conformation was observed with the A allele associated with an increase of 0.93. No association was observed between the remaining seven mutations in *STAT3* and *STAT5* with carcass conformation. No association was observed between any of the *STAT* mutations examined with carcass weight, carcass fat, cull cow weight or somatic cell count. Results suggest a multifaceted role of the *STAT* family in growth and development and the potential for increasing the frequency of this allele in the national herd without negative effects in relation to other carcass traits tested and somatic cell count.

Keywords: genetics, STATs, DNA polymorphism, dairy cattle

A novel association between a STAT1 mutation and carcass conformation in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McCLure[2], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon, Co.Cork, Ireland
Corresponding Author email: l.ratcliffe@research.ait.ie

Introduction

Signal transducer and activator of transcription (STAT) genes encode for a family of proteins that are involved in pre and post natal growth and development. In cattle, variants in these genes have already been associated with economically important traits including milk production and embryonic survival (Cobanoglu et al., 2006; Khatib et al., 2009).

The objective of this study was to estimate the effects of polymorphisms in the *STAT1*, *STAT3* and *STAT5* genes on carcass and health traits in dairy cattle.

Materials and Methods

STAT genotypes (n=8) on 10,707 dairy cattle were obtained through the Irish cattle breeding federation (ICBF). Phenotypes for carcass and health traits, also obtained from the ICBF, were expressed as predicted transmitting abilities (PTAs). The association analysis included n=5747, 3194, 1374, 518 and 360 cows for somatic cell score, carcass weight, cull cow weight, carcass conformation and carcass fat, respectively. Associations between each SNP and PTA were analysed in ASREML using a weighted mixed animal model. Bioinformatic analysis was carried out to estimate the effect of allelic substitution G->A in the *Bos Taurus STAT1* gene using the variant effect predictor located at www.ensembl.org.

Results

Summary statistics, including genotype frequencies, minor allele frequencies and Hardy Weinberg equilibrium (HWE) for all SNPs

SNP	Genotype	Genotype Frequency	MAF	HWE
STAT1_2697	G/G G/A A/A	0.05 0.35 0.60	0.21	0.998
STAT3_19069	T/T T/C C/C	0.19 0.49 0.31	0.43	0.999
STAT3_25402	T/T T/G G/G	0.16 0.48 0.36	0.38	0.999
STAT5_12195	C/C C/G G/G	0.16 0.49 0.35	0.34	0.975
STAT5_13244	A/A A/G G/G	0.20 0.49 0.31	0.42	0.998
STAT5_13319	A/A A/G G/G	0.00 0.09 0.90	0.04	0.997
STAT5_13516	T/T T/G G/G	0.20 0.49 0.31	0.42	0.998
STAT5_3117	A/A A/G G/G	0.00 0.00 1.00	0	1.000

Genotype-phenotype association study

A significant association ($p<0.0001$) between STAT1 (2697) and carcass conformation was observed with the A allele associated with a decrease in carcass conformation with an effect size of 0.93 as shown in Table 1.1.

No association was observed between the remaining seven mutations analysed in STAT3 and STAT5 with carcass conformation. No association was observed between any of the STAT mutations examined in this study and carcass weight, carcass fat, cull cow weight or somatic cell count.

Table 1.1 – Effect of allelic substitution in the STAT1 gene on carcass traits in Holstein Friesian dairy cattle

SNP	Allele Substitution	Carcass weight	Carcass conformation	Carcass fat	Culled carcass weight
STAT1_2697	G→A	-	-0.93(0.19)	-	-

Bioinformatic analysis

Preliminary bioinformatic analysis revealed that the *STAT1* variant is located in the 3'UTR of the gene.

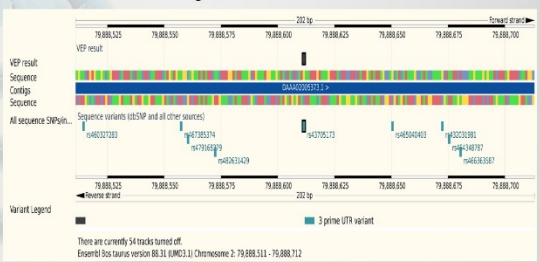


Fig 1.0 – *STAT1* variant (rs43705173) is located in the 3'UTR region of the gene

Discussion

The results of this study suggest a multifaceted role of the STAT family in growth and development and the potential for increasing the frequency of the *STAT1* G allele in the national herd without any negative effects in relation to other carcass traits tested and somatic cell count.

Further bioinformatic studies will be conducted to determine the effect of the variant on miRNA binding sites in the 3'UTR of the *STAT1* gene.

References
- Cobanoglu, O., Zaitoun, I., Chang, Y., Shook, G. and Khatib, H. (2006). Effects of the Signal Transducer and Activator of Transcription 1 (STAT1) Gene on Milk Production Traits in Holstein Dairy Cattle. *Journal of Dairy Science*, 89(11), pp.4433-4437.
- Khatib, H., Huang, W., Mikheil, D., Schutzkus, V. and Monson, R. (2009). Effects of signal transducer and activator of transcription (STAT) genes STAT1 and STAT3 genotypic combinations on fertilization and embryonic survival rates in Holstein cattle. *Journal of Dairy*

Estimation of the effects of mutations causing Complex Vertebral Malformation and Brachyspina on milk production, milk composition and fertility traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Mathew McClure[2], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland

The frequencies of mutations with lethal effects are of significant economic importance in cattle production. The elimination from or at least management of such mutations in the national breeding herd is a desirable objective, however, estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous. Therefore, the objective of the current study was to estimate the effects of two such lethal recessives, Complex Vertebral Malformation (CVM) and Brachyspina (BY) on milk and fertility traits in Holstein-Friesian dairy cattle. CVM and BY SNP genotypes and phenotypes (expressed as predicted transmitting abilities (PTA)) on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. Only cows with an adjusted reliability of >0.1 were included in the analysis and included n=6876 for milk yield and composition traits and n = 1193, 264, 4566, 8564, 152 and 2380 cows for calving interval, survival, calving difficulty, gestation length, calf mortality and maternal calving difficulty, respectively. CVM (MAF 1.7 %) was associated with both increased milk protein (0.019, s.e. 0.006, $p<0.01$) and increased milk fat concentration (0.039, s.e. 0.0132, $p<0.01$) whereas no associations ($p>0.05$) were observed between CVM and any of the other milk traits (milk yield, milk fat yield, milk protein yield) or fertility traits (calving interval, survival, calving difficulty, gestation length, calf mortality and maternal calving difficulty). Significant associations were observed between BY (<1 %) and decreased milk protein concentration (0.024, s.e. 0.008, $p<0.01$) and increased milk yield (73.21 kg, s.e. 30.12, $p<0.05$). No associations ($p>0.05$) were observed between BY and any of the fertility traits considered. These results provide additional evidence that carriers of these recessive mutations exhibit effects on milk production and/or composition in Holstein Friesian cattle, however, with no evidence of effects on the fertility traits examined. Cognisance and monitoring of the potential pleiotropic effects of lethal recessives such as examined in this study will aid livestock breeders when considering elimination of carriers to minimise reduction of the genetic merit of farm enterprises and inform the benefits of strategic matings in controlling these mutations in the population while also sustaining productivity.

Estimation of the effects of mutations causing Complex Vertebral Malformation and Brachyspina on milk production, milk composition and fertility traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland

Introduction

- The frequencies of mutations with lethal effects are of significant economic importance in cattle production.
- The elimination from or at least management of such mutations in the national breeding herd is a desirable objective.
- Estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous.

Methods

CVM and BY SNP genotypes and phenotypes (expressed as predicted transmitting abilities (PTA)) on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. Only cows with an adjusted reliability of >10% were included in the analysis.

Results

- CVM was associated with both increased milk protein (0.019, s.e. 0.006, p<0.01) and increased milk fat concentration (0.039, s.e. 0.0132, p<0.01).
- Significant associations were observed between BY and decreased milk protein concentration (0.024, s.e. 0.008, p<0.01) and increased milk yield (73.21 kg, s.e. 30.12, p<0.05).

Introduction

Complex vertebral malformation (CVM) is a lethal recessive disorder that manifests during fetal development. It is caused by a missense mutation in the bovine SLC35A3 gene which encodes for a UDP-N-acetylglucosamine transporter (Thomsen et al., 2006). Nielson et al, 2003, reports that by gestation day 260 up to 77% of homozygous foetuses are aborted.

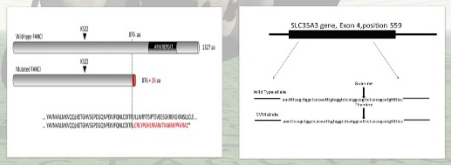


Fig. 1.0 – BY is caused by the deletion of exons 25-27 within the *FANCI* gene *(left)*, and CVM is caused by a transversion mutation (Guanine to Thymine) at position 559 of the SLC35A3 gene *(right)*

Brachyspina (BY) is a rare recessive genetic defect observed in Holstein dairy cattle. The genetic cause of BS development is a 3.3 kb deletion, removing exons 25-27 within the Fanconi anaemia complementation group I (FANCI) gene on chromosome 21 (Charlier et al., 2012).

Conclusion

These results provide additional evidence that carriers of these recessive mutations exhibit effects on milk production and/or composition in Holstein Friesian cattle, however, with no evidence of effects on the fertility traits examined.

Monitoring of the potential pleiotropic effects of lethal recessives such as examined in this study will aid livestock breeders when considering elimination of carriers to minimise reduction of the genetic merit of farm enterprises and inform the benefits of strategic matings in controlling these mutations in the population while also sustaining productivity.

References

- Charlier, C., Agerholm, J., Coppieters, W., Karlskov-Mortensen, P., Li, W., de Jong, G., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., Ahariz, N., Mullaert, E., Georges, M. and Fredholm, M. (2012). A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. PLoS ONE, 7(8), p.e43085.
- CHU, Q., ZHANG, Y., SUN, D., YU, Y., WANG, Y. and ZHANG, Y. (2010). Identification of the complex vertebral malformation gene in Chinese Holstein and its association with dairy performance traits. Hereditas (Beijing), 32(7), pp.732-736.
- Garrick, D., Taylor, J. and Fernando, R. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41(1), p.55.
- Nielsen, U., Aamand, G., Andersen, O., Bendixen, C., Nielsen, V. and Agerholm, J. (2003). Effects of complex vertebral malformation on fertility traits in Holstein cattle. Livestock Production Science, 79(2-3), pp.233-238.
- Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A., Holm, L., Nielsen, V., Agerholm, J., Arnbjerg, J. and Bendixen, C. (2006). A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. Genome Research, 16(1), pp.97-105.

Estimation of the effects of mutations causing Complex Vertebral Malformation and Brachyspina on milk production, milk composition and fertility traits in Holstein Friesian dairy cattle
Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland
Introduction
- The frequencies of mutations with lethal effects are of significant economic importance in cattle production.
- The elimination from or at least management of such mutations in the national breeding herd is a desirable objective.
- Estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous.
Methods
CVM and BY SNP genotypes and phenotypes (expressed as predicted transmitting abilities (PTA)) on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. Only cows with an adjusted reliability of >10% were included in the analysis.
Results
- CVM was associated with both increased milk protein (0.019, s.e. 0.006, p<0.01) and increased milk fat concentration (0.039, s.e. 0.0132, p<0.01).
- Significant associations were observed between BY and decreased milk protein concentration (0.024, s.e. 0.008, p<0.01) and increased milk yield (73.21 kg, s.e. 30.12, p<0.05).

Methods

CVM and BY SNP genotypes on 10,707 dairy cows were obtained from the ICBF.

Phenotypes for milk production, composition and fertility traits, expressed as predicted transmitting abilities(PTAs) were also obtained from the ICBF.

PTAs were deregressed as per Garrick et al, 2009, with only animals with an adjusted reliability of >10% being included in the subsequent association analysis.

This included n=6876 for milk yield and composition traits and n = 1193, 264, 4566, 8564, 152 and 2380 cows for calving interval, survival, calving difficulty, gestation length, calf mortality and maternal calving difficulty, respectively.

Associations between both mutations and the phenotypic traits were analysed using a weighted mixed animal model in ASREML.

In this model, the average expected relationships amongst animals was accounted for through the numerator relationship matrix, generated using six generations of back pedigree. Percent Holstein of the animal was included as a fixed effect in the model. The dependant variable was de-regressed PTA weighted by their respective reliabilities.

Estimation of the effects of mutations causing Complex Vertebral Malformation and Brachyspina on milk production, milk composition and fertility traits in Holstein Friesian dairy cattle
Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland
Introduction
- The frequencies of mutations with lethal effects are of significant economic importance in cattle production.
- The elimination from or at least management of such mutations in the national breeding herd is a desirable objective.
- Estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous.
Methods
CVM and BY SNP genotypes and phenotypes (expressed as predicted transmitting abilities (PTA)) on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. Only cows with an adjusted reliability of >10% were included in the analysis.
Results
- CVM was associated with both increased milk protein (0.019, s.e. 0.006, p<0.01) and increased milk fat concentration (0.039, s.e. 0.0132, p<0.01).
- Significant associations were observed between BY decreased milk protein concentration (0.024, s.e. 0.008, p<0.01) and increased milk yield (73.21 kg, s.e. 30.12, p<0.05).

Results

Table 1.0 – Minor allele frequencies of CVM and BY observed in the Irish Holstein Friesian population examined

Mutation	Minor allele Frequency (%)
Complex vertebral malformation	1.5
Brachyspina	0.8

Table 1.1 – Milk composition and production traits associated (P<0.05) with the brachyspina mutation, with results presented as effect sizes with standard error denoted in brackets

Mutation	Protein composition (%)	Milk Yield (kg)
Brachyspina	-2.4 (0.8)	73.21(30.12)

No associations (p>0.05) were observed between brachyspina and any of the fertility traits considered (calving interval, survival, calving difficulty, gestation length, calf mortality and maternal calving difficulty).

Table 1.2 – Milk composition and production traits associated (P<0.01) with the complex vertebral malformation mutation, with results presented as effect sizes with standard error denoted in brackets

Mutation	Protein composition (%)	Fat composition (%)
Complex vertebral malformation	1.9 (0.6)	4(1.3)

No associations (p>0.05) were observed between CVM and any of the other milk traits (milk yield, milk fat yield, milk protein yield) or fertility traits (calving interval, survival, calving difficulty, gestation length, calf mortality and maternal calving difficulty).

Estimation of the effects of mutations causing Complex Vertebral Malformation and Brachyspina on milk production, milk composition and fertility traits in Holstein Friesian dairy cattle
Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon Co. Cork, Ireland
Introduction
- The frequencies of mutations with lethal effects are of significant economic importance in cattle production.
- The elimination from or at least management of such mutations in the national breeding herd is a desirable objective.
- Estimation of the potential pleiotropic effects on other traits of economic importance would ascertain if strategic matings of carrier animals would be advantageous.
Methods
CVM and BY SNP genotypes and phenotypes (expressed as predicted transmitting abilities (PTA)) on 10,707 dairy cows were obtained through the Irish Cattle Breeding Federation (ICBF). The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. Only cows with an adjusted reliability of >10% were included in the analysis.
Results
- CVM was associated with both increased milk protein (0.019, s.e. 0.006, p<0.01) and increased milk fat concentration (0.039, s.e. 0.0132, p<0.01).
- Significant associations were observed between BY decreased milk protein concentration (0.024, s.e. 0.008, p<0.01) and increased milk yield (73.21 kg, s.e. 30.12, p<0.05).

Conclusion

These results provide additional evidence that carriers of these recessive mutations exhibit effects on milk production and/or composition in Holstein Friesian cattle, however, with no evidence of effects on the fertility traits examined.

Monitoring of the potential pleiotropic effects of lethal recessives such as examined in this study will aid livestock breeders when considering elimination of carriers to minimise reduction of the genetic merit of farm enterprises and inform the benefits of strategic matings in controlling these mutations in the population while also sustaining productivity.

References

- Charlier, C., Agerholm, J., Coppieters, W., Karlskov-Mortensen, P., Li, W., de Jong, G., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., Ahariz, N., Mullaert, E., Georges, M. and Fredholm, M. (2012). A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. PLoS ONE, 7(8), p.e43085.
- CHU, Q., ZHANG, Y., SUN, D., YU, Y., WANG, Y. and ZHANG, Y. (2010). Identification of the complex vertebral malformation gene in Chinese Holstein and its association with dairy performance traits. Hereditas (Beijing), 32(7), pp.732-736.
- Garrick, D., Taylor, J. and Fernando, R. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41(1), p.55.
- Nielsen, U., Aamand, G., Andersen, O., Bendixen, C., Nielsen, V. and Agerholm, J. (2003). Effects of complex vertebral malformation on fertility traits in Holstein cattle. Livestock Production Science, 79(2-3), pp.233-238.
- Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A., Holm, L., Nielsen, V., Agerholm, J., Arnbjerg, J. and Bendixen, C. (2006). A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. Genome Research, 16(1), pp.97-105.

Single nucleotide polymorphisms in the Signal transducer and regulator of transcription (STAT) genes are associated with milk production, milk composition and fertility traits in Holstein Friesian cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co.Cork, Ireland

Signal transducer and activator of transcription (STAT) genes encode for a family of proteins that are involved in pre- and post-natal growth and development. In cattle, variants in these genes have been associated with economically important traits including milk production and embryonic survival. The objective of this study was to estimate the effects of polymorphisms in the *STAT1*, *STAT3*, and *STAT5* genes on milk production, composition, and fertility traits in Holstein Friesian dairy cattle. *STAT* genotypes (n=8) on 10,707 dairy cattle were obtained through the Irish Cattle Breeding Federation (ICBF). The phenotypes (n=16) for milk production, milk composition and fertility traits also obtained from the ICBF and expressed as predicted transmitting abilities (PTAs). The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. The association analysis included n=6876 for milk yield and composition traits. The analysis for fertility traits included n = 1193, 264, 4566, 8564, 152, and 2380 cows for calving interval, survival, calving difficulty, gestation length, calf mortality, and maternal calving difficulty, respectively. In the analysis of the *STAT* variants with milk traits (milk protein concentration, milk fat concentration, milk yield, milk fat yield, milk protein yield) a significant association ($p<0.05$) was observed between *STAT3* (25042) and *STAT5* variants (12195, 13244, 13319, 13516) and milk protein percentage. *STAT3* variants (19069, 25042) were associated ($p<0.001$) with milk fat percentage, additionally *STAT5* variants (13244, 13516) were also found to be associated ($p<0.05$) with this trait. The G allele of *STAT3* (25042) was also associated with increased milk yield (17.01 kg, s.e. 6.708, $p<0.05$). No associations were observed between *STAT1* and the remaining polymorphisms analysed in either *STAT3* or *STAT5* with the milk production and milk composition traits examined. Associations were observed between *STAT3* (19069) and gestation length (0.11 days, s.e. 0.056, $p<0.05$) and *STAT5* (12195) with calf mortality (2.04, s.e. 1.017, $p<0.05$). None of the six remaining polymorphisms considered in this study within the *STAT* genes were associated with any of the aforementioned fertility traits. These results support a multifaceted role of the *STAT* family in milk production, composition, and fertility which warrants further functional analysis and consideration for incorporation into genetic evaluation programs for maximising the rate of genetic gain.

Abstract No:190

Single Nucleotide Polymorphisms in the Signal Transducer and Regulator of Transcription (*STAT*) Genes Are Associated with Milk Production, Milk Composition and Fertility Traits in Holstein Friesian Cattle

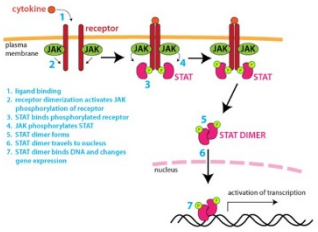
Lyndsey Ratcliffe[1], Francis Kearney[2], Jennifer McClure[2], Matthew McCLure[2], Michael P. Mullen[1]

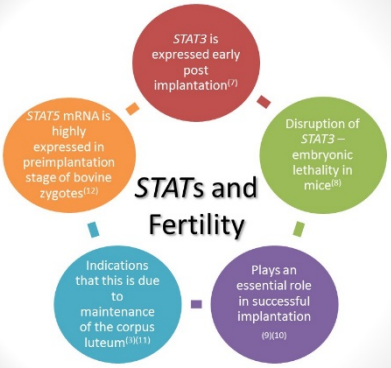
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co.Cork, Ireland

Introduction

- Signal transducer and activator of transcription (*STAT*) genes code for a family of proteins involved in regulating many gene expression pathways





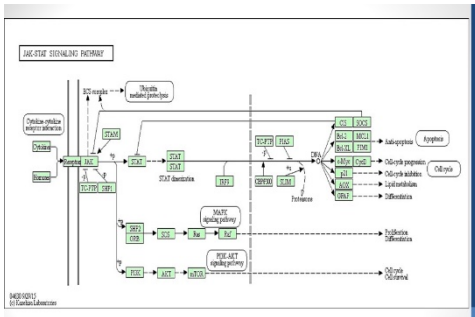


Fig 1.0 – *Bos taurus JAK-STAT* signalling pathway(21)

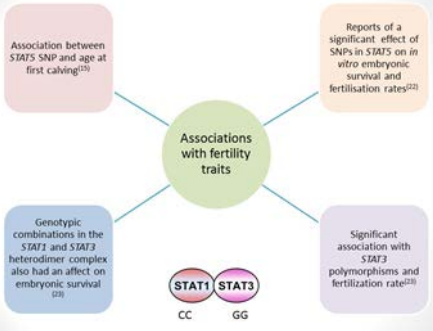
Research Question

Are polymorphisms within the *STAT1, STAT3* and *STAT5* genes associated with economically important traits in Irish dairy cattle?

What do previous research studies suggest?

- QTL for milk traits have been observed in the region of *STAT* genes(13) (14)
- Polymorphisms within the *STAT5* gene have been associated with milk production and composition traits in a number of cattle breeds;

Holstein (15)(16) Jersey (17)(18) Brown Swiss(19) Agerolese(20)



Methods

Genotypes for *STAT* polymorphisms (n=7) located on the *STAT1*, *STAT3* and *STAT5* genes for 10,707 dairy cattle were obtained from the Irish Cattle Breeding Federation (ICBF)

Phenotypic data on milk and fertility traits (n=11) were also obtained from the ICBF and were expressed as predicted transmitting abilities (PTA)

The removal of parental contribution and deregression of the PTA values was performed as per Garrick et al, 2009. Cows with a reliability of >10% were retained for inclusion in the association analysis

Univariate SNP analysis was performed using a weighted mixed animal model in ASREML (Gilmour et al., 2009)

Genotyped individuals were included as a random effect and the average expected relationships amongst animals was accounted for through the numerator relationship matrix, which was generated using six generations of back pedigree

% Holstein was included as a fixed effect. Dependant variable was de-regressed PTA weighted by their respective reliabilities. Genotype was included in the analysis as a continuous variable

Phenotypes

- Milk traits (n=6876)
  - Milk production – m(kg), f(kg), p(kg)
  - Milk composition – protein percentage, fat percentage
- Fertility traits
  - Calving interval (n=1193)
  - Survival (n=264)
  - Calving difficulty (n=4566)
  - Gestation length (n=8564)
  - Mortality (n=152)
  - Maternal calving difficulty (n=2380)

Results

1. Summary statistics
2. Linkage disequilibrium analysis
3. Association study – Milk production and composition
4. Association study – Fertility

1. Summary statistics

Table 1.0 – Summary statistics for the 7 *STAT* SNPs examined in this study

SNP	Rs Identifier	BTA	Position	MAF	HWE[1]
***STAT1* 2697**	rs43705173	2	79888611	0.21	1.000
***STAT3* 19069**	rs110942700	19	43070296	0.43	0.999
***STAT3* 25402**	rs134279188	19	43063963	0.38	0.999
***STAT5* 13516**	rs110495396	19	43047128	0.42	0.999
***STAT5* 13319**	rs208753173	19	43046931	0.04	0.996
***STAT5* 13244**	rs109788842	19	43046856	0.42	1.000
***STAT5* 12195**	rs137182814	19	43045807	0.34	1.000

[1] Deviation from HWE

2. Linkage disequilibrium

Table 1.1- Linkage disequilibrium as measured by r^2 between the 6 SNPs located on BTA19

SNP	*STAT3* 19069	*STAT3* 25402	*STAT5* 13516	*STAT5* 13319	*STAT5* 13244
***STAT3* 25402**	0.992				
***STAT5* 13516**	0.998	0.992			
***STAT5* 13319**	0.176	0.244	0.180		
***STAT5* 13244**	0.998	0.992	0.998	0.180	
***STAT5* 12195**	0.994	0.998	0.994	0.228	0.994

3. Milk Composition and production

Table 1.2 - Effect of allelic substitution of SNPs in *STAT* genes that are significantly associated with **milk composition** in Irish dairy cattle

SNP	Allelic substitution	Protein composition (%)[1]	Fat composition (%)[1]
STAT3 19069	T→C	-0.03(0.002)†	**-0.15(0.003)****
STAT3 25402	T→G	**-0.07(0.002)****	**-0.19(0.003)****
STAT5 13516	T→G	**-0.05(0.001)****	**-0.08(0.004)***
STAT5 13319	G→A	**-0.17(0.004)****	-0.14(0.008)
STAT5 13244	A→G	**-0.05(0.002)****	**-0.09(0.003)***
STAT5 12195	C→G	**-0.05(0.002)****	-0.07(0.004)†

() denotes standard error [1]multiplied by 100
†P<0.1; *P<0.05; ** P<0.01

Table 1.3 - Effect of allelic substitution of SNPs in *STAT* genes that are significantly associated with **milk production** in Irish dairy cattle

SNP	Allelic substitution	Milk yield (kg)	Fat yield (kg)
STAT3 19069	T→C	12.58(6.65)†	
STAT3 25402	T→G	**16.9(6.7)***	
STAT5 13516	T→G		-0.42(0.24)†
STAT5 13244	A→G		-0.42(0.24)†

() denotes standard error; †P<0.1; *P<0.05

4. Fertility traits

Table 1.4 - Effect of allelic substitution of SNPs in *STAT* genes that are significantly associated with fertility traits in Irish dairy cattle

SNP	Allelic substitution	Calving interval	Gestation length	Mortality	Maternal calving difficulty
STAT3 19069	T→C	-2.62(1.52)†	-0.11(0.005)*		
STAT3 25402	T→G			-1.84(1.01)†	
STAT5 13516	T→G			-1.74(0.96)†	-0.24(0.13)†
STAT5 13244	A→G			-1.74(0.96)†	-0.24(0.13)†
STAT5 12195	C→G	-2.83(1.60)†		-2.04(1.02)*	-0.25(0.14)†

() denotes standard error; †P<0.1; *P<0.05

Summary of results

- Milk traits
  - Protein composition - *STAT3* 25402, *STAT5* 13516, *STAT5* 13319, *STAT5* 13244, *STAT5* 12195 (p<0.01)
  - Fat composition - *STAT3* 19069, *STAT3* 25402 (p<0.01) *STAT5* 13516, *STAT5* 13244 (p<0.05)
  - Milk yield - *STAT3* 25402 (p<0.05)
- Fertility traits
  - Gestation length - *STAT3* 19069 (p<0.05)
  - Mortality - *STAT5* 12195 (p<0.05)

Conclusion

- Results are consistent with previous research proposing associations between the *STAT* genes and milk and fertility traits.
- Incorporation of the presented variants into genomic selection programmes requires careful consideration due some antagonistic relationships observed.
- Whether the variants investigated are causative or linked to causative loci is currently unknown and future work could include functional analysis of the SNPs and validation in independent populations.

References

1. Wakao H, Gouilleux F, Groner B. Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. The EMBO journal. 1994 May 1;13(9):2182.
2. Miyoshi K, Shillingford JM, Smith GH, Grimm SL, Wagner KU, Oka T, Rosen JM, Robinson GW, Hennighausen L. Signal transducer and activator of transcription (Stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium. The Journal of cell biology. 2001 Nov 12;155(4):531-42.
3. Teglund S, McKay C, Schuetz E, Van Deursen JM, Stravopodis D, Wang D, Brown M, Bodner S, Grosveld G, Ihle JN. Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. Cell. 1998 May 29;93(5):841-50.
4. Vafaizadeh V, Klemmt P, Brendel C, Weber K, Doebele C, Britt K, Grez M, Fehse B, Desrivières S, Groner B. Mammary epithelial reconstitution with gene-modified stem cells assigns roles to Stat5 in luminal alveolar cell fate decisions, differentiation, involution, and mammary tumor formation. Stem cells. 2010 May 1;28(5):928-38.
5. Li S, Rosen JM. Nuclear factor I and mammary gland factor (STAT5) play a critical role in regulating rat whey acidic protein gene expression in transgenic mice. Molecular and Cellular Biology. 1995 Apr 1;15(4):2063-70.
6. Mukhopadhyay SS, Wyszomierski SL, Gronostajski RM, Rosen JM. Differential interactions of specific nuclear factor I isoforms with the glucocorticoid receptor and STAT5 in the cooperative regulation of WAP gene transcription. Molecular and cellular biology. 2001 Oct 15;21(20):6859-69.
7. Duncan SA, Zhong Z, Wen Z, Darnell JE. STAT signaling is active during early mammalian development. Developmental dynamics. 1997 Feb 1;208(2):190-8.
8. Takeda K, Noguchi K, Shi W, Tanaka T, Matsumoto M, Yoshida N, Kishimoto T, Akira S. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proceedings of the National Academy of Sciences. 1997 Apr 15;94(8):3801-4.
9. Liu T, Ogle TF. Signal transducer and activator of transcription 3 is expressed in the decidualized mesometrium of pregnancy and associates with the progesterone receptor through protein-protein interactions. Biology of reproduction. 2002 Jul 1;67(1):114-8.
10. Lee JH, Kim TH, Oh SJ, Yoo JY, Akira S, Ku BJ, Lydon JP, Jeong JW. Signal transducer and activator of transcription-3 (Stat3) plays a critical role in implantation via progesterone receptor in uterus. The FASEB Journal. 2013 Jul 1;27(7):2553-63.

Relationships between mutations responsible for Holstein Haplotype 1, 3 and 4 and Bovine Leukocyte Adhesion Deficiency (BLAD) and production traits in Holstein Friesian cattle

Matthew McClure[1], Jennifer McClure[1], Lyndsey Ratcliffe[2], Francis Kearney[1], Michael P. Mullen[2]

[1]Irish Cattle Breeding Federation, Highfield road, Bandon Co. Cork, Ireland

[2]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

Identification of carriers of mutations with lethal effects in cattle populations enables more informed decision making by the farmer be it elimination from breeding stock or management through strategic mating schemes for high genetic merit carriers. In order to best advise farmers on the use of this information, estimation of the effects of these mutations on routinely recorded production traits in carrier animals is needed. Therefore, the objective of this study was to estimate if the mutations associated with HH1, 3, 4 or BLAD showed any evidence of effects across any production traits (milk, fertility, carcass and health traits (n=16)) in dairy cows. Genotypes and phenotypes (expressed as predicted transmitting abilities (PTAs)) on 10,707 dairy cows were obtained from the Irish Cattle Breeding Federation (ICBF) database. Only animals with an adjusted reliability of >0.1 were included in the analysis which included n=6876, 1198, 264, 4566, 8564, 152, 2280, 3194, 518, 360, 1374, 5747 cows for milk traits(n=5), calving interval, survival, gestation length, calf mortality, maternal calving difficulty, carcass weight, carcass conformation, carcass fat, cull cow weight, and somatic cell score, respectively. The association between each SNP and PTA (deregressed) was analysed in ASREML using weighted mixed animal models. BLAD carriers were associated with increased somatic cell score ($p<0.05$) and calf mortality ($p<0.05$), however, there was no association ($p>0.05$) with any of the other milk, fertility or carcass traits analysed in this study. No association ($p>0.05$) was observed between HH1 and any of the traits examined. Cows with a HH2 allele were associated ($p<0.05$) with decreased gestation length with no other effects identified. Cows with a HH3 allele were associated ($p<0.05$) with increased calving interval with no other effects observed. Unless carriers of either BLAD, HH1, HH3 or HH4 are of otherwise high genetic merit these results provide no evidence to support the maintenance of carriers on farm or in the national herd.

Associations between mutations in genes affecting milk composition and quality and production traits in Holstein Friesian cattle

Michael P. Mullen[1], Lyndsey Ratcliffe[1], Jennifer McClure[2], Francis Kearney[2], Matthew McClure[2]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Highfield road, Bandon Co. Cork, Ireland

DNA mutations that affect milk composition are of particular interest not only to producers i.e. the livestock breeding industry but also consumers. In dairy cattle, two such genes, κ-casein and β-casein, harbour mutations which have been associated with positive effects on cheese production and human health, respectively. The objective of this study was to estimate the effects of polymorphisms in *κ-casein* and *β-casein* on milk, fertility, carcass and health traits (n=16) in dairy cows. Genotypes and phenotypes on 10,707 dairy cows were obtained through the Irish cattle breeding federation (ICBF). Phenotypes were expressed as predicted transmitting abilities (PTAs). Only animals with an adjusted reliability of >0.1 were included in the analysis which included n=6876, 1198, 264, 4566, 8564, 152, 2280, 3194, 518, 360, 1374, 5747 cows for milk traits (n=5), calving interval, survival, gestation length, calf mortality, maternal calving difficulty, carcass weight, carcass conformation, carcass fat, cull cow weight, and somatic cell score, respectively. The association between each SNP and deregressed PTA was analysed in ASREML using a weighted mixed animal model. No evidence (P>0.05) was found for an association between a validated κ-casein variant (342T>C) and any of the milk, fertility, carcass or health traits analysed in this population of dairy cows. The β-casein A2 variant (A allele) was associated with: increased milk protein percentage (0.007, s.e. 0.0018, $p<0.0001$); increased milk fat percentage (0.0074, s.e. 0.0038, $p\leq0.05$); increased milk fat (0.51 kg, s.e. 0.24, $p<0.05$); protein yield (0.43 kg, s.e. 0.20, $p<0.05$); decreased carcass fat (-0.17 kg, s.e. 0.06, $p<0.05$) and increased somatic cell score (0.18, s.e. 0.005, $p<0.001$). No association was identified between the β-casein A2 variant and any of the fertility or other carcass traits analysed. These results suggest the potential for increasing the frequency of desirable alleles in the national herd without significant negative effects in relation to the milk, fertility and carcass traits tested.

No evidence of an association between *DGAT1* and fertility traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Katie Quigley[1], David F.G. Flores[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2,3], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co.Cork, Ireland

[3]ABS Global, Inc. 1525 River Rd. DeForest, Wisconsin, USA

Corresponding author email: l.ratclife@research.ait.ie

A polymorphism within the diacylglycerol O-acyltransferase 1 (*DGAT1*) gene, which leads to an alanine to lysine substitution at position 232 in the protein, has previously been associated with an increase in milk composition and production traits in cattle populations. Increasing this desirable allele in the Holstein Friesian cattle population may improve the overall genetic merit of the national herd, however, it is necessary to ensure that there are no antagonistic effects on other traits of importance. Genotypes for the *DGAT* dinucleotide substitution polymorphism (c.694GC>AA) on 10,707 cows were obtained from the Irish Cattle Breeding Federation (ICBF). Phenotypes for fertility traits, expressed as predicted transmitting abilities (PTAs), were also obtained from the ICBF. Only animals with an adjusted reliability of >0.1 were included for analysis and this included n =1193, 264, 4566, 8564, 152 and 2380 for calving interval, survival, calving difficulty, gestation length, mortality and maternal calving difficulty, respectively. Associations between each polymorphism and PTA were analysed in ASREML using a weighted mixed animal model. No associations ($p>0.05$) between the *DGAT1* polymorphism and any of the aforementioned fertility traits were observed in the sample set tested. These results suggest that increasing the frequency of the desirable allele contributing to increased milk composition and production would not have a significant negative effect with regards to the fertility traits analysed in this study.

Keywords: genetics, dairy cattle, *DGAT*, fertility

No evidence of an association between *DGAT1* and fertility traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Katie Quigley[1], David F.G. Flores[1], Francis Kearney[2], Jennifer McClure[2], Matthew McClure[2,3], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland

[3]ABS Global, Inc. 1525 River Rd. DeForest, Wisconsin, USA

Corresponding author email: l.ratcliffe@research.ait.ie

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Eirinn

#LoveIrish Research

Introduction

The diacylglycerol O-acyltransferase 1 (*DGAT1*) gene has previously been associated with milk composition traits in dairy cattle[1].

The objective of this study was to determine if a dinucleotide substitution polymorphism (c.694GC>AA) was associated with fertility traits in Irish dairy cattle.

Significance

Sustainable agricultural practices include selection for desirable alleles that are known to have a positive impact on production traits in farm animals. This is the case with the mutation being analysed in this study, which has been shown to increase milk composition traits in dairy cattle.

It is necessary, however, to ensure that there are no antagonistic effects on other traits of importance.

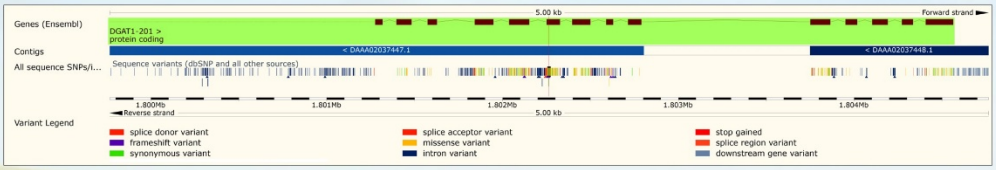


Fig 1 - The *DGAT1* gene is located on BTA 14, with the mutation of focus for this study causing a nonconserative substitution of a positively charged lysine residue with a neutral alanine residue in position 232 of the resultant protein

Methods

- Phenotypic (expressed as predicted transmitting ability (PTAs)) and genotypic data were available from the Irish Cattle Breeding Federation (ICBF) database on 10,707 Irish Holstein Friesian dairy cattle.
- Deregression and removal of parental contributions were performed on the phenotypes.
- Only animals with an adjusted reliability of >0.1 were retained for analysis.
- This included n =1193, 264, 4566, 8564, 152 and 2380 for calving interval, survival, calving difficulty, gestation length, mortality and maternal calving difficulty, respectively.
- Associations between each polymorphism and PTA were analysed in ASREML using a weighted mixed animal model.

Results

The minor allele frequency was 0.43 with significant deviation from Hardy Weinberg equilibrium observed.

Gene	Minor allele frequency	Deviation from HWE
DGAT1	0.43	1.00

No significant associations (p>0.1) were observed between this mutation and any of the fertility phenotypes analysed in this study.

Discussion and Conclusion

The results of this study suggest that increasing the allele in the *DGAT1* gene responsible for increased milk composition traits will not have any negative impact on the fertility traits that have been analysed in this study,

Acknowledgements

Lyndsey Ratcliffe's contribution to this study was supported by the government of Irelands postgraduate scholarship programme.

AIT Connect & Discover

ICBF.com

References

[1] Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P. and Spelman, R., (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome research, 12(2), pp.222-231.

Estimation of the effects of a polymorphism in the *DGAT1* gene with carcass traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Katie Quigley[1], David F.G Flores[1], Francis Kearney[2], Jennifer McClure[2,3], Matthew McClure[2,3], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co.Cork, Ireland

[3]ABS Global, Inc. 1525 River Rd. DeForest, Wisconsin, USA

Corresponding author email: l.ratclife@research.ait.ie

A desirable objective for Irish farmers is to utilise genetic information to increase productivity on farm. Increasing the frequency of alleles associated with increased milk production and composition traits can improve the overall genetic merit of the herd, while also aiding in applying sustainable agricultural practices. It is important, however, that no negative effects on other traits of importance result from the application of this objective. Genotypes for the *DGAT1* dinucleotide substitution polymorphism (c.694GC>AA), a polymorphism previously associated with increased milk production traits, were obtained from the Irish Cattle Breeding Federation (ICBF) for 10, 707 cows. Phenotypes for carcass traits, expressed as predicted transmitting abilities (PTAs), were also obtained from the ICBF. Only animals with an adjusted reliability of >0.1 were included for analysis and this included n= 3194, 1374, 518 and 360 cows for carcass weight, culled cow weight, carcass conformation and carcass fat, respectively. Associations between each polymorphism and PTA were analysed in ASREML using a weighted mixed animal model. Tentative associations ($p<0.1$) were observed between this polymorphism and carcass weight (0.97kg, s.e 0.55) and carcass fat (0.10kg, s.e 0.06). A significant association ($p = 0.01$) was observed for carcass conformation (0.38, s.e 0.15). These results suggest that increasing this desirable allele on farm and in the national herd would have an impact via carcass traits on carrier animals.

Keywords: Genetics, *DGAT1*, Dairy cattle, Carcass traits

Estimation of the effects of a polymorphism in the *DGAT1* gene with carcass traits in Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Katie Quigley[1], David F.G Flores[1], Francis Kearney[2], Jennifer McClure[2,3], Matthew McClure[2,3], Michael P. Mullen[1]
[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland
[2]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland
[3]ABS Global, Inc. 1525 River Rd. DeForest, Wisconsin, USA
Corresponding author email: l.ratcliffe@research.ait.ie

IRISH RESEARCH COUNCIL An Chomhairle um Thaighde in Eirinn

#LoveIrish Research

Introduction

The *DGAT1* gene is located on BTA 14, and encodes a key metabolic transmembrane protein which converts diacylglycerol and fatty acyl CoA to triacylglycerol. The focus of this study is a missense mutation which leads to the positively charged lysine residue being replaced with a neutral alanine residue in position 232 of the protein. This mutation has previously been associated with an increase in milk composition traits in dairy cattle[1].

The objective of this study was to determine if this specific mutation had any effects on carcass traits in Holstein Friesian Irish dairy cattle.

Materials and Methods

- Phenotypic (expressed as predicted transmitting ability (PTAs)) and genotypic data were available from the Irish Cattle Breeding Federation database on 10,707 Irish Holstein Friesian dairy cattle. Summary statistics, including genotype frequencies and Hardy Weinberg equilibrium were calculated.
- Deregression and removal of parental contributions were performed on the phenotypes. Only animals with an adjusted reliability of >0.1 were retained for analysis. This included n= 3194, 1374, 518 and 360 cows for carcass weight, culled cow weight, carcass conformation and carcass fat, respectively.
- Associations between each polymorphism and PTA were analysed in ASREML using a weighted mixed animal model.

Results

The minor allele frequency was determined to be 0.43 with deviation from Hardy Weinberg equilibrium observed. Table 2.0 represents the results obtained for the association analysis.

Table 1.– Summary statistics for *DGAT1 K232A* polymorphism

Gene	Genotype Frequencies		Minor allele frequency	Deviation from HWE
DGAT1	AA	0.19	0.43	1.00
	AG	0.50		
	GG	0.31		

Table 2 – Results obtained for the association analysis: *DGAT1* and carcass traits

Phenotype	Effect size	Standard error	*P* value
Carcass weight	0.97 kg	0.55	0.08
Carcass fat	0.11kg	0.06	0.08
Carcass conformation	0.38	0.15	0.01

The results of the association analysis determined an association (p=0.08) between the dinucleotide substitution polymorphism (c.694GC>AA) and carcass weight with an effect size of 0.97 kg. An association (p=0.08) was also observed between this polymorphism and carcass fat with an effect size of 0.11 kg.
A significant association (p=0.01) was observed with this polymorphism in regards to carcass conformation with a effect size of 0.38 observed.

Discussion and Conclusion

One strategy in applying sustainable agricultural practices is to increase the frequency of genetic polymorphisms that are associated with increases in production and functional traits in farm animal populations. It is imperative to ensure that in doing this, no negative impacts on other important traits are evident,

The results obtained in this research study suggest that increasing the allele that is associated with the *DGAT1 K232A* polymorphism will have an impact on carcass traits in Holstein Friesian dairy cattle, however none of these effects are antagonistic in nature.

Acknowledgements

Many thanks to the Irish Cattle Breeding Federation who supply the data required in order to carry out this important research.

Lyndsey Ratcliffe's contribution to this research was previously supported by the Athlone Institute of Technology's President seed fund and is currently supported by the Irish Research Councils Government of Irelands Postgraduate Scholarship programme.

AIT Connect & Discover

ICBF.com

References
[1]Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P. and Spelman, R., (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome research, 12(2), pp.222-231.

Segregation of a candidate novel lethal recessive in Irish Holstein Friesian dairy cattle

Lyndsey Ratcliffe[1], Francis Kearney[2], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland

The *LFNG* (O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase) gene has previously been observed to be associated with fertility and reproduction *in vivo* in zebrafish, avian and mice. As a member of the NOTCH signalling pathway, this gene has been shown to play a critical role in embryonic development through the regulation of the formation and patterning of somites in vertebrates. These studies suggest *LFNG* may be required for successful early bovine embryo development and any polymorphisms affecting the function of *LFNG* may be contributing to embryonic lethality in cattle. A nonsense mutation in *LFNG* was added to the content of the International Dairy and Beef custom genotyping platform. A total of 10,707 Irish Holstein Friesian dairy cow genotypes at this loci and fertility phenotypes (calving interval and calf mortality expressed as PTAs) were obtained from the Irish Cattle Breeding Federation. The nonsense snp in *LFNG* was segregating in the heterozygous state, albeit at a very low frequency (MAF <0.01), with no cows homozygous for this variant identified. Association analysis, carried out in ASReml using a weighted mixed animal model, revealed a tentative association ($p<0.1$) between this snp and an increase in calving interval which may represent a role in early embryonic loss in carrier animals. Future work includes validation of this association using a larger dataset of cattle including trios where possible with *in vivo* functional studies also warranted.

Segregation of a candidate novel lethal recessive gene in Irish Holstein Friesian dairy cattle

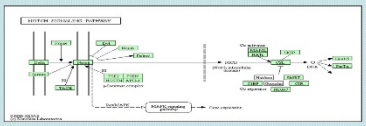
Lyndsey Ratcliffe[1], Francis Kearney[2], Michael P. Mullen[1]

[1]Bioscience Research Institute, Athlone Institute of Technology, Athlone, Co. Westmeath, Ireland

[2]Irish Cattle Breeding Federation, Bandon, Co. Cork, Ireland

Introduction

The *LFNG* (O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase) gene, also known as the lunatic fringe gene, is involved in the NOTCH signalling pathway. The Notch pathway plays multiple roles during vertebrate somitogenesis, functioning in the segmentation clock and during somite patterning[1].



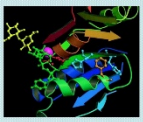


Fig. 1 - The NOTCH signalling pathway [2] , structural model of the *lfng* protein[5]

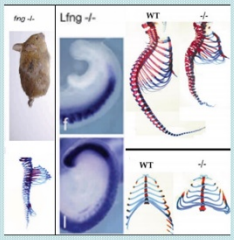


Fig. 2 –Knockout *lfng* mice display skeletal and vertebral malformations with the axial skeleton being severely affected, numerous vertebral and rib fusions and incompletely formed vertebrae are also evident.

Lfng mutant mice display defects in somatic and vertebral patterning during embryogenesis [3][4]. In zebrafish, *lfng* has been shown to be involved in the formation of segment boundaries in the hindbrain and somites [6] . In humans, mutations in this gene have been associated with the development of Spondylocostal Dysostosis, which is a vertebral malformation disorder arising during embryonic development [7] .

Besides these effects observed during embryogenesis, studies suggest it may also affect reproduction. Inactivation of lunatic fringe in mice leads to infertility associated with pleiotropic defects in follicle development and meiotic maturation of oocytes in female mice, whereas lunatic fringe null males mice were found to be subfertile [8] .

Objective

These studies suggest LFNG may be required for successful early bovine embryo development and any polymorphisms affecting the function of LFNG may be contributing to embryonic lethality in cattle.

The objective of this research study was to determine if a polymorphism in the *Bos taurus LFNG* gene was associated with economically important fertility traits in the Irish Holstein Friesian cattle population.

Materials and Methods

- Phenotypic (expressed as predicted transmitting ability (PTAs)) and genotypic data were available from the Irish Cattle Breeding Federation database on 10,707 Irish Holstein Friesian dairy cattle.
- Deregression and removal of parental contributions were performed on the phenotypes. Only animals with an adjusted reliability of >0.1 were retained for analysis. This included n =1193, 264, 4566, 8564, 152 and 2380 for calving interval, survival, calving difficulty, gestation length, mortality and maternal calving difficulty, respectively.
- Associations between each polymorphism and PTA were analysed in ASREML using a weighted mixed animal model.

Results and discussion

The nonsense snp in *LFNG* was segregating in the heterozygous state, albeit at a very low frequency (MAF = 0.002), with no cows homozygous for this variant identified.

Association analysis, carried out in ASReml using a weighted mixed animal model, revealed a tentative association (p<0.1) between this snp and an increase in calving interval which may represent a role in early embryonic loss in carrier animals.

Table 1 – Results obtained in the association analysis of the *LFNG* polymorphism and fertility traits in Irish Holstein Friesian cattle

Minor Allele frequency	Trait	Effect size*	P value
0.002	Calving Interval	7.72(4.47)	0.08

* Calving interval in days, standard error is depicted in brackets

Conclusion

The observations presented in previous studies with regards to the *LFNG* genes role in embryogenesis and reproduction in vertebrates suggests that this gene may be a factor in early embryonic lethality and reproductive success in cattle populations.

Future work includes validation of the association observed in this study using a larger dataset of cattle including trios where possible with in vivo functional studies also warranted.

Acknowledgements

Lyndsey Ratcliffe's contribution to this study was previously supported by the Athlone Institute of Technologys President seed fund and is currently supported by the government of Irelands postgraduate scholarship programme (GOIPG/2017/1801).

1. Shifley, E., VanHorn, K., Perez-Salaguer, A., Franklin, J., Weinstein, M. and Cole, S. (2008). Oscillatory lunatic fringe activity is crucial for segmentation of the anterior but not posterior skeleton. Development, 135(5), pp.899-908.
2. Kegg.jp. (2018). KEGG PATHWAY: Notch signaling pathway - Bos taurus (cow). [online] Available at: https://www.kegg.jp/kegg-bin/show_pathway?org_name=btaa&mapno=04330&mapscale=1.0&show_description=hide [Accessed 3 Sep. 2018].
3. Serth, K., Schuster-Gossler, K., Cordes, R. and Gossler, A. (2003). Transcriptional oscillation of Lunatic fringe is essential for somitogenesis. Genes & Development, 17(7), pp.912-925.
4. Evrard, Y., Lun, Y., Aulehla, A., Gan, L. and Johnson, R. (1998). lunatic fringe is an essential mediator of somite segmentation and patterning. Nature, 394(6691), pp.377-381.
5. Sparrow, D., Chapman, G., Wouters, M., Whittock, N., Ellard, S., Fatkin, D., Turnpenny, P., Kusumi, K., Sillence, D. and Dunwoodie, S. (2006). Mutation of the LUNATIC FRINGE Gene in Humans Causes Spondylocostal Dysostosis with a Severe Vertebral Phenotype. The American Journal of Human Genetics, 78(1), pp.28-37.
6. Prince, V., Holley, S., Sagy-Cuill, L., Prabhakaran, B., Gates, A., Ho, R. and Vogt, T. (2001). Zebrafish lunatic fringe demarcates segmental boundaries. Mechanisms of Development, 105(1-2), pp.175-180.
7. Sparrow, D., Chapman, G., Wouters, M., Whittock, N., Ellard, S., Fatkin, D., Turnpenny, P., Kusumi, K., Sillence, D. and Dunwoodie, S. (2006). Mutation of the LUNATIC FRINGE Gene in Humans Causes Spondylocostal Dysostosis with a Severe Vertebral Phenotype. The American Journal of Human Genetics, 78(1), pp.28-37.
8. Hahn, K., Beres, B., Rowton, M., Skinner, M., Chang, Y., Rawls, A. and Wilson Rawls, J. (2009). A deficiency of lunatic fringe is associated with cystic dilation of the rete testis. REPRODUCTION, 137(1), pp.79-93.

Association analysis results

The following section of this appendix contains graphical representations of the results obtained in the association analysis. Effect sizes versus each trait analysed are presented, with marker sizes depicting the p values obtained.

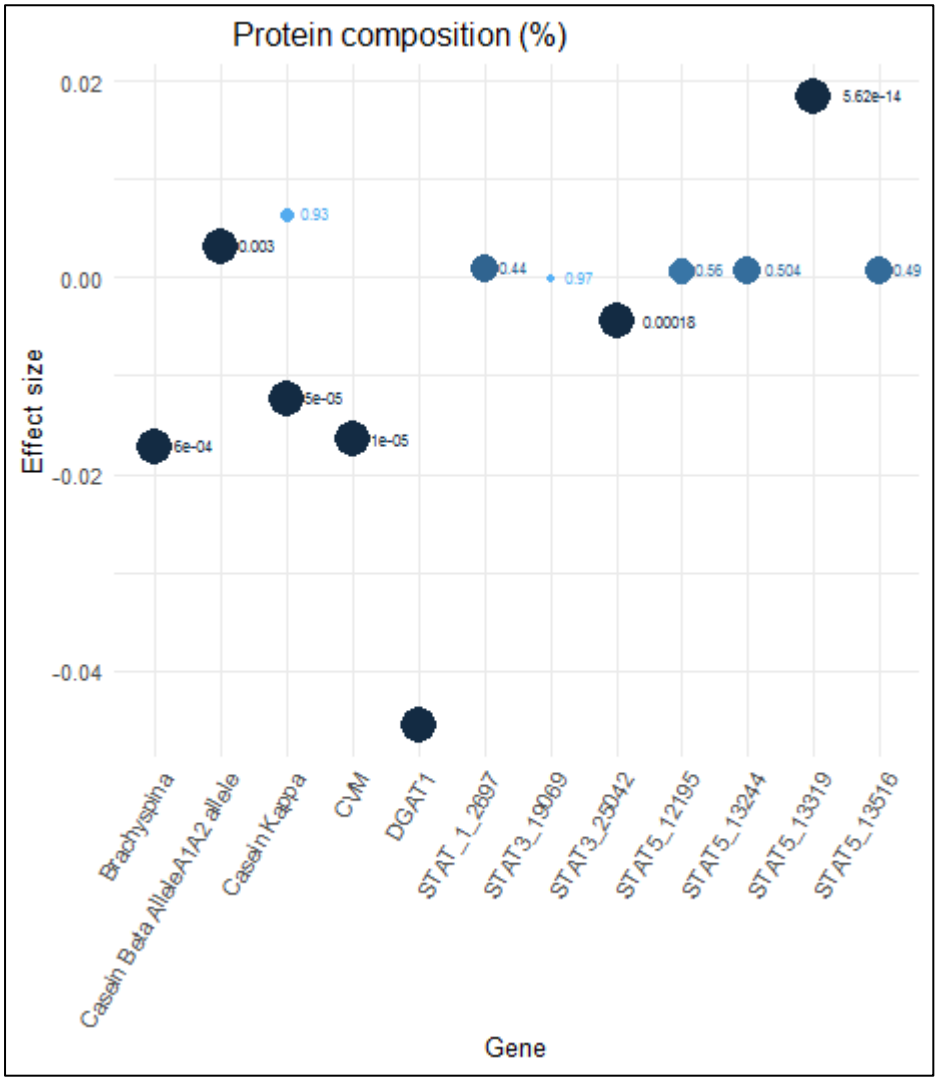


Fig. 1.0 – The effect sizes obtained with regards to protein composition (%) after substitution for the alternative allele in the association study.

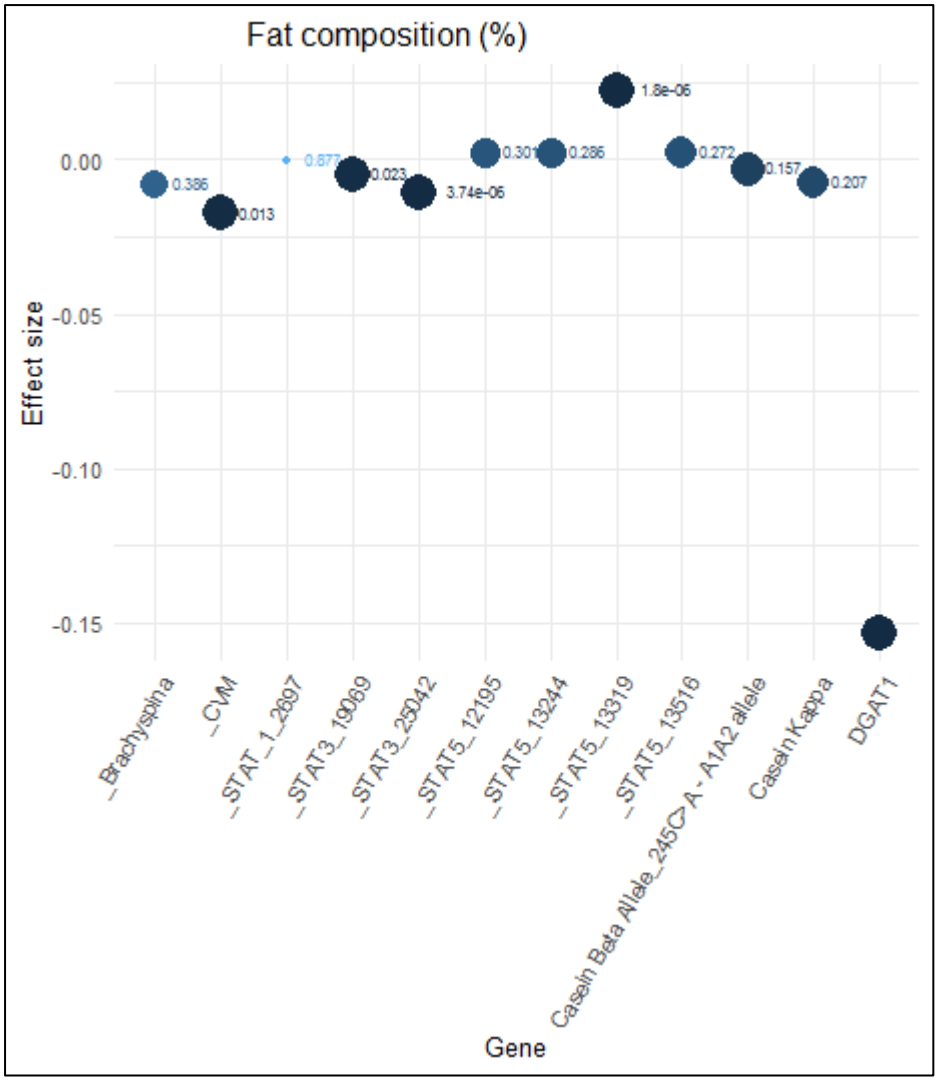


Fig. 1.1 – The effect sizes obtained with regards to fat composition (%) after substitution for the alternative allele in the association study.

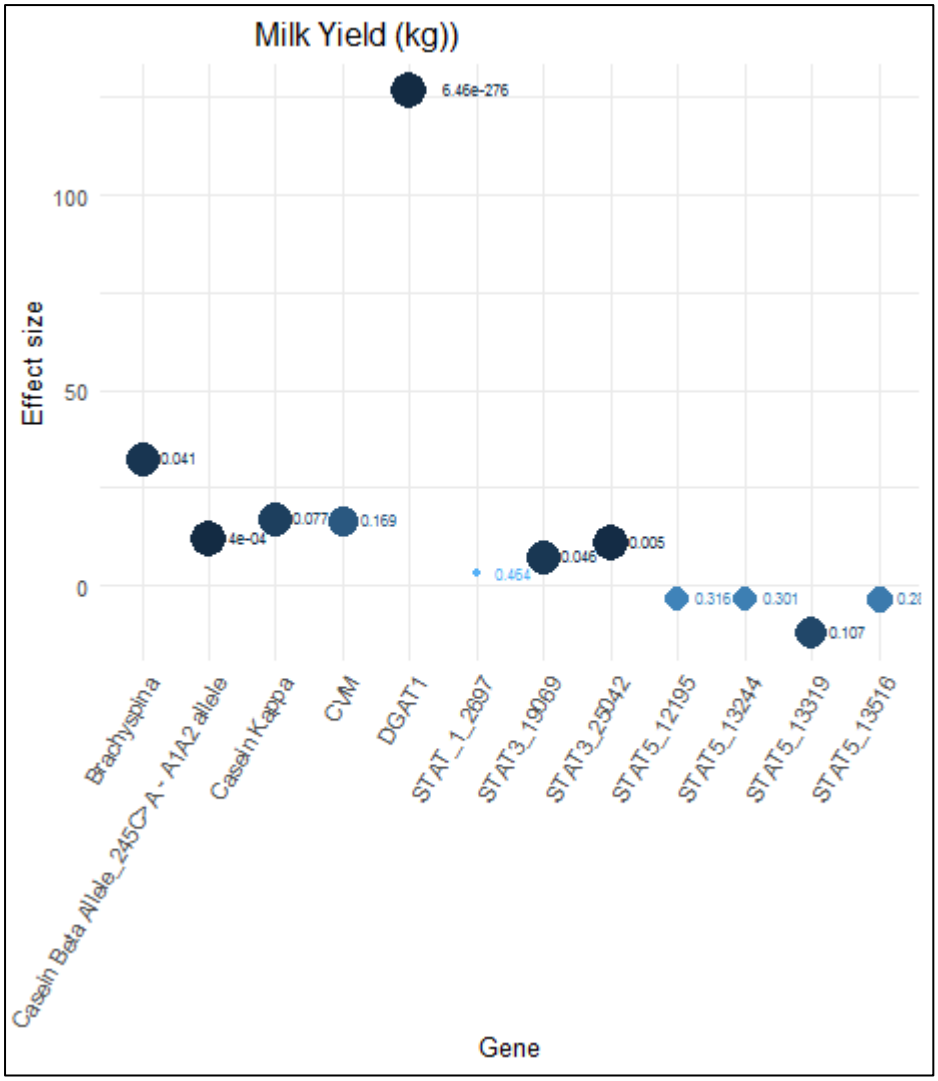


Fig. 1.2 – The effect sizes obtained with regards to milk yield (kg) after substitution for the alternative allele in the association study.

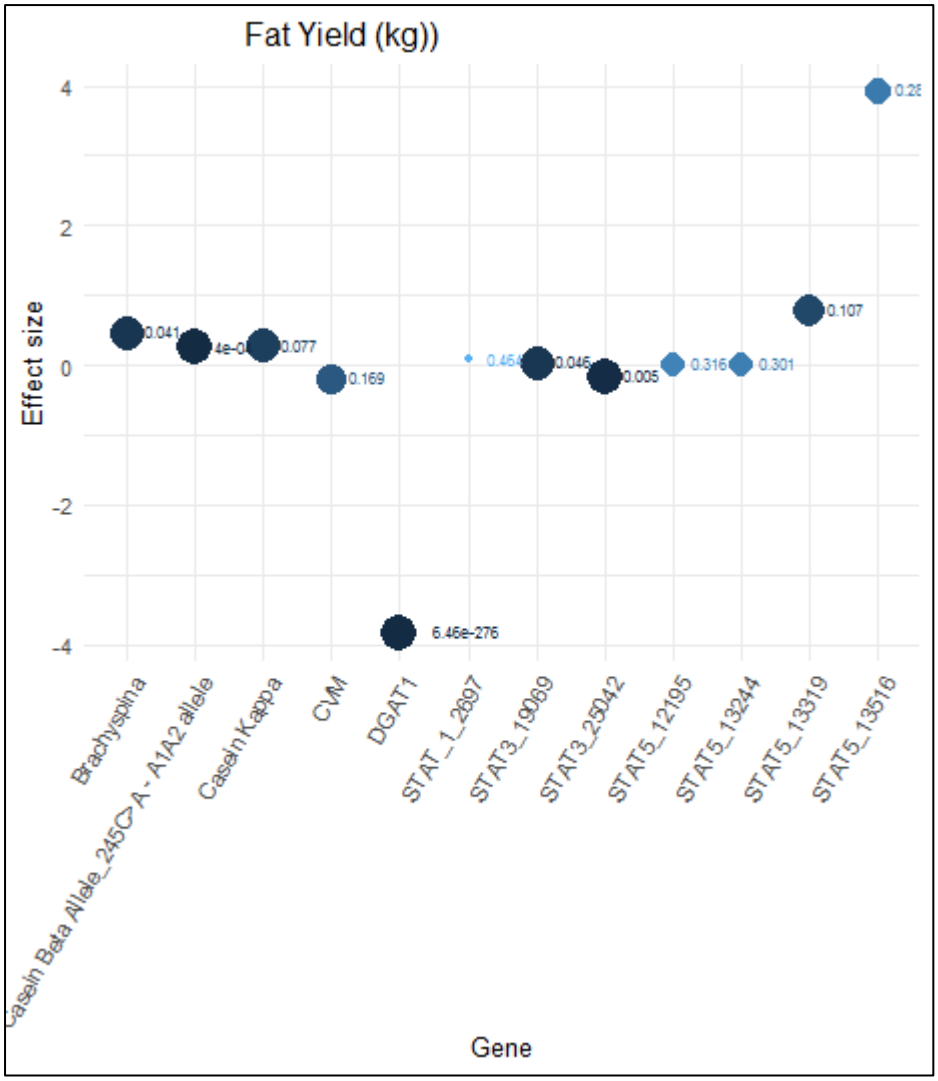


Fig. 1.3 – The effect sizes obtained with regards to fat yield (kg) after substitution for the alternative allele in the association study.

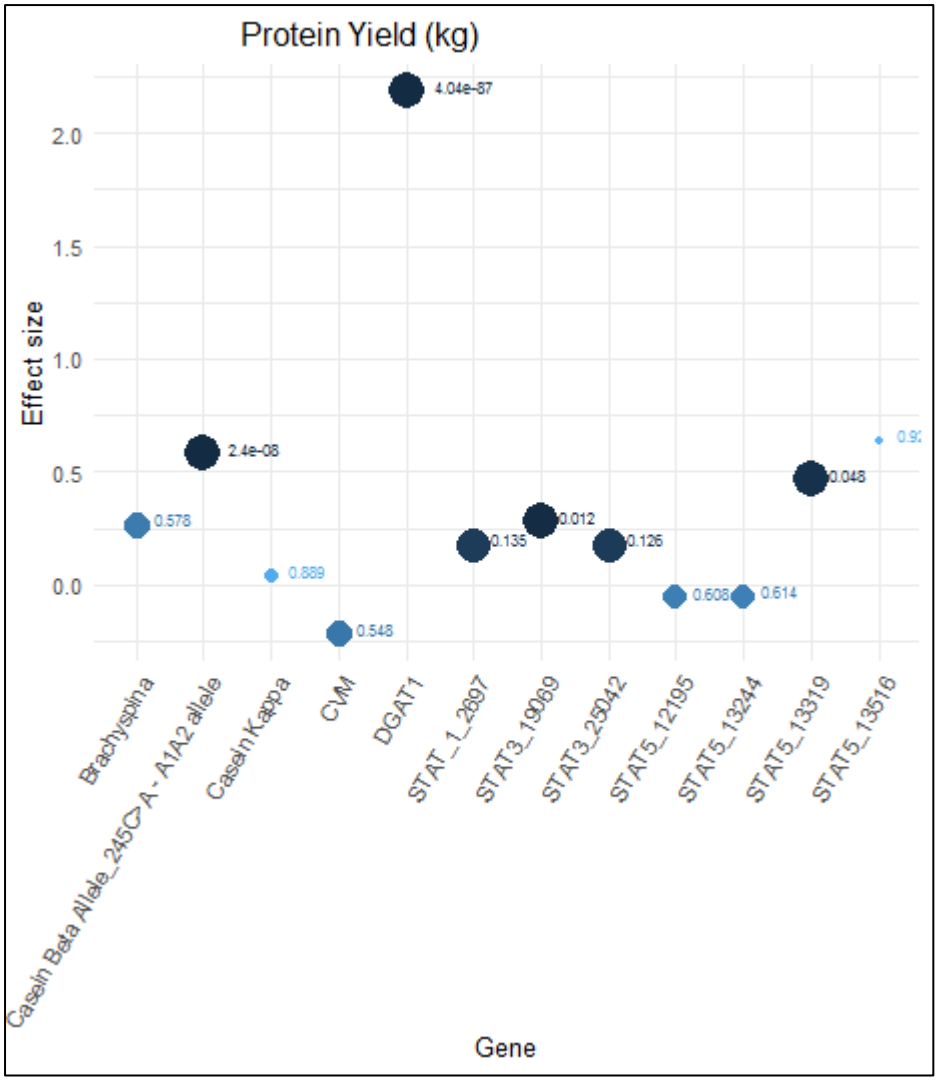


Fig. 1.4 – The effect sizes obtained with regards to protein yield (kg) after substitution for the alternative allele in the association study.

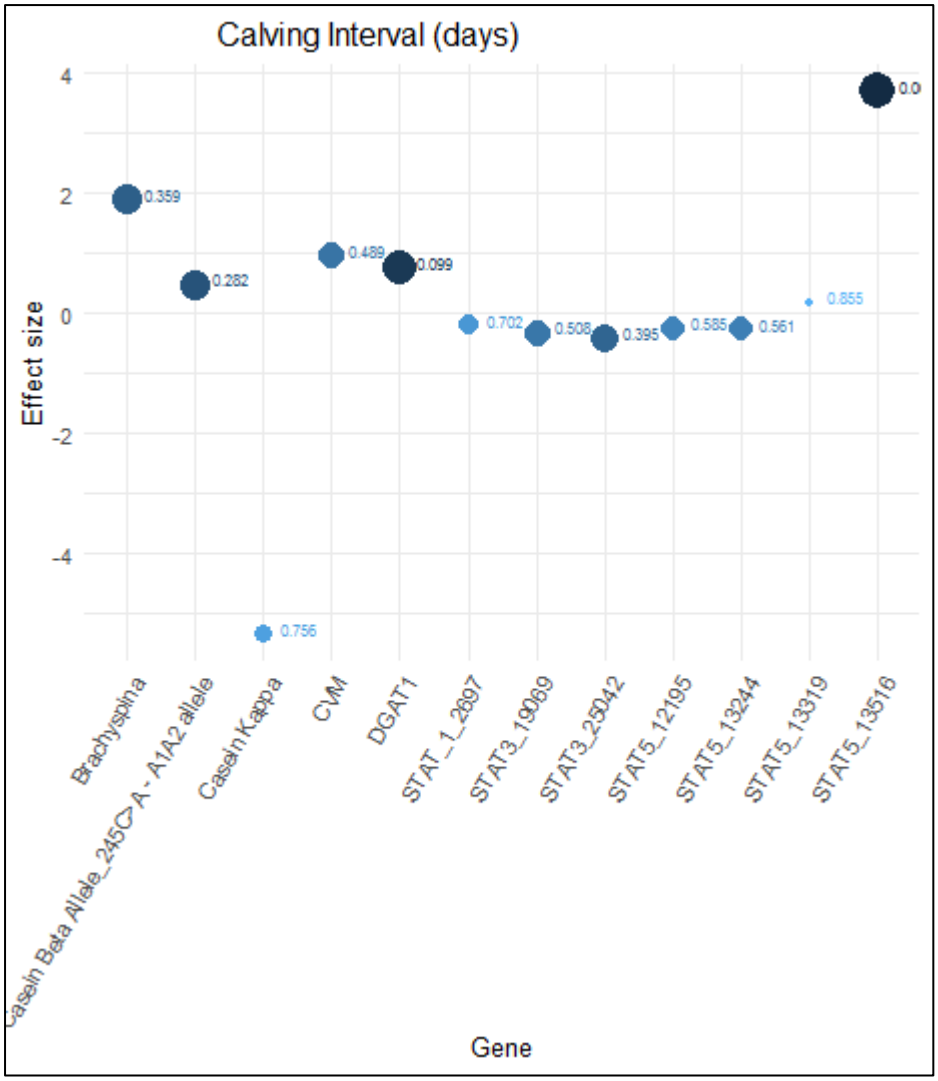


Fig. 1.5 – The effect sizes obtained with regards to calving interval (days) after substitution for the alternative allele in the association study.

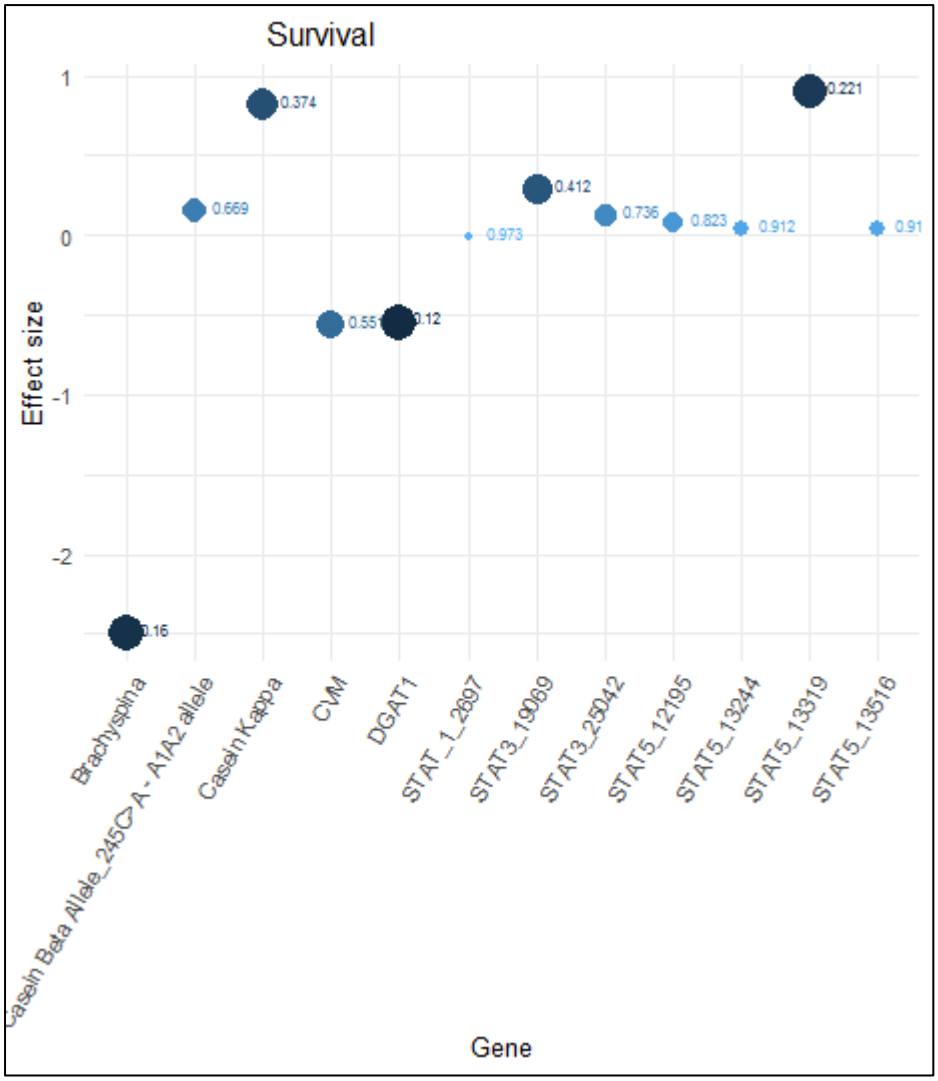


Fig. 1.6 – The effect sizes obtained with regards to survival after substitution for the alternative allele in the association study.

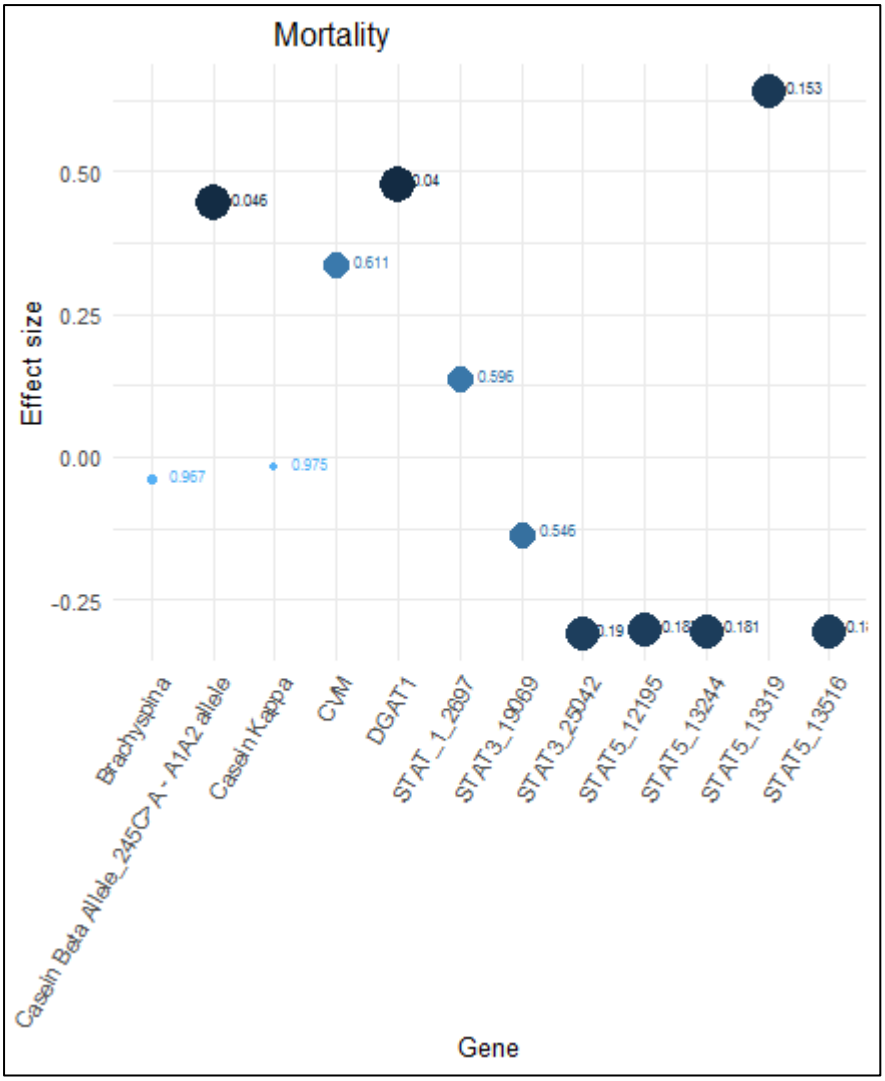


Fig. 1.7 – The effect sizes obtained with regards to mortality after substitution for the alternative allele in the association study.

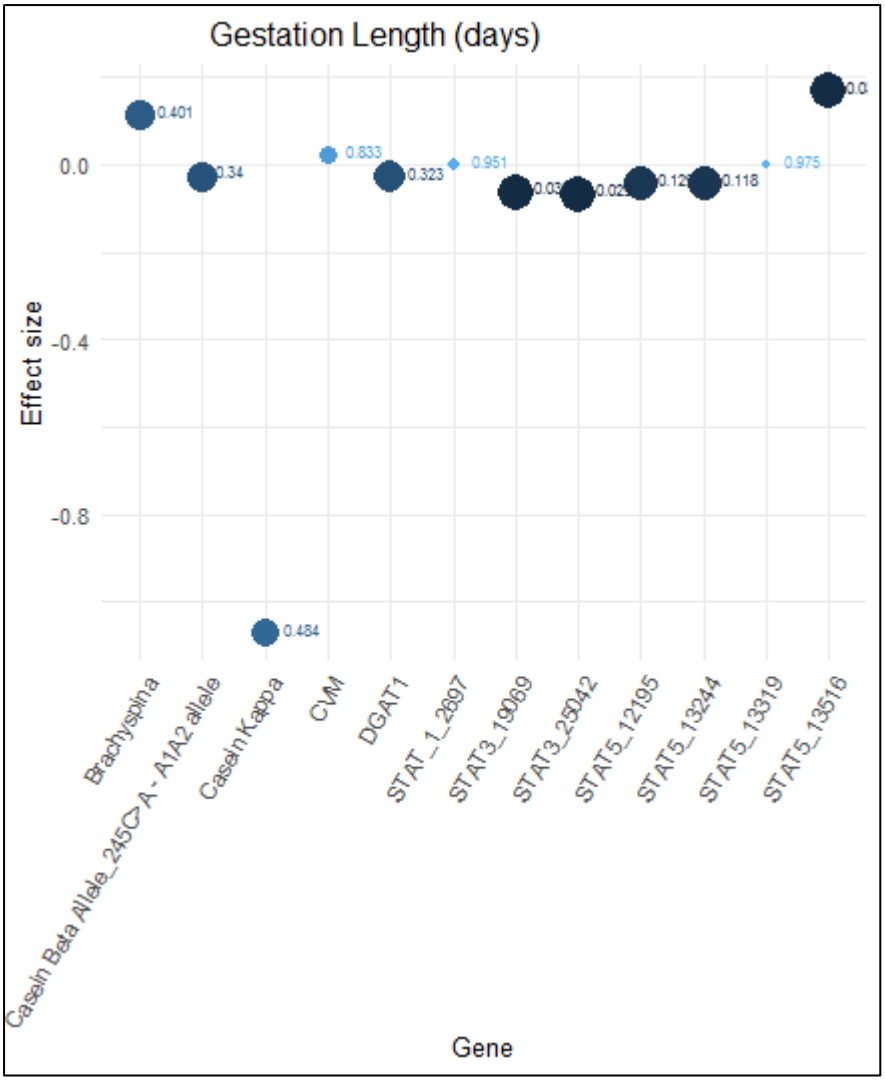


Fig. 1.8 – The effect sizes obtained with regards to gestation length (days) after substitution for the alternative allele in the association study.

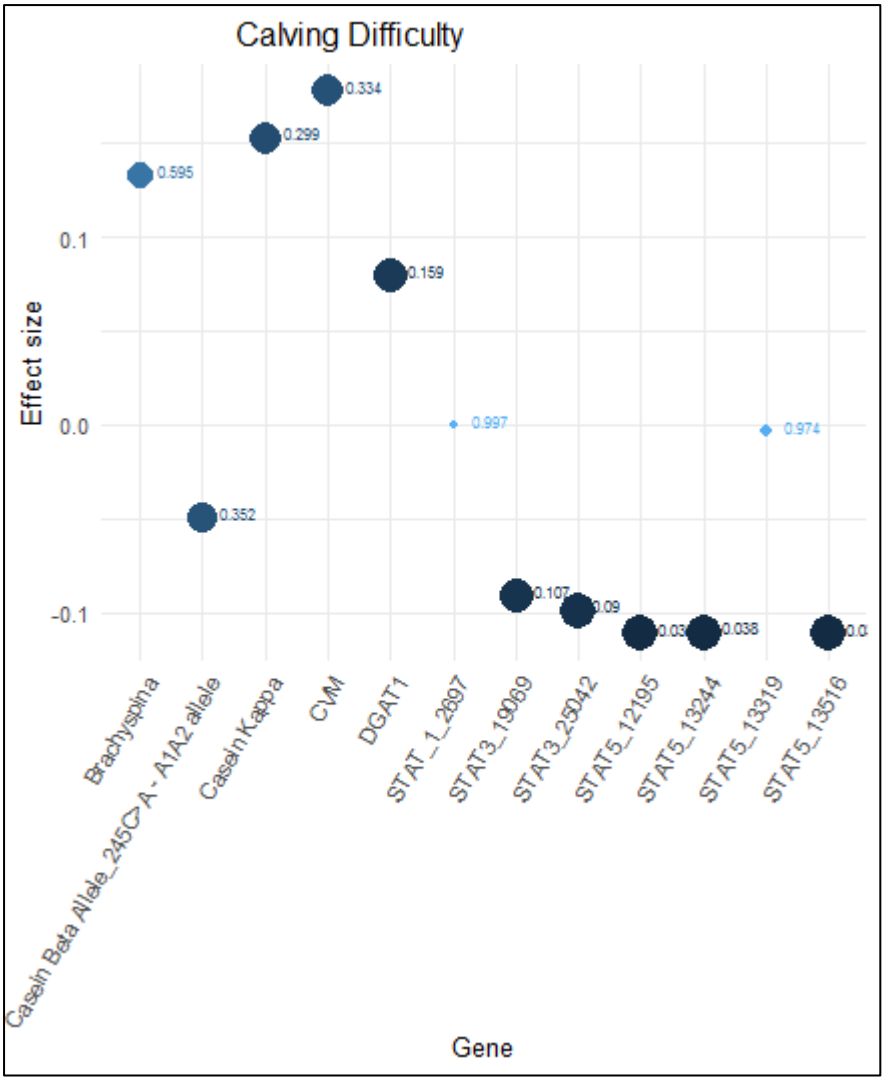


Fig. 1.9 – The effect sizes obtained with regards to calving difficulty after substitution for the alternative allele in the association study.

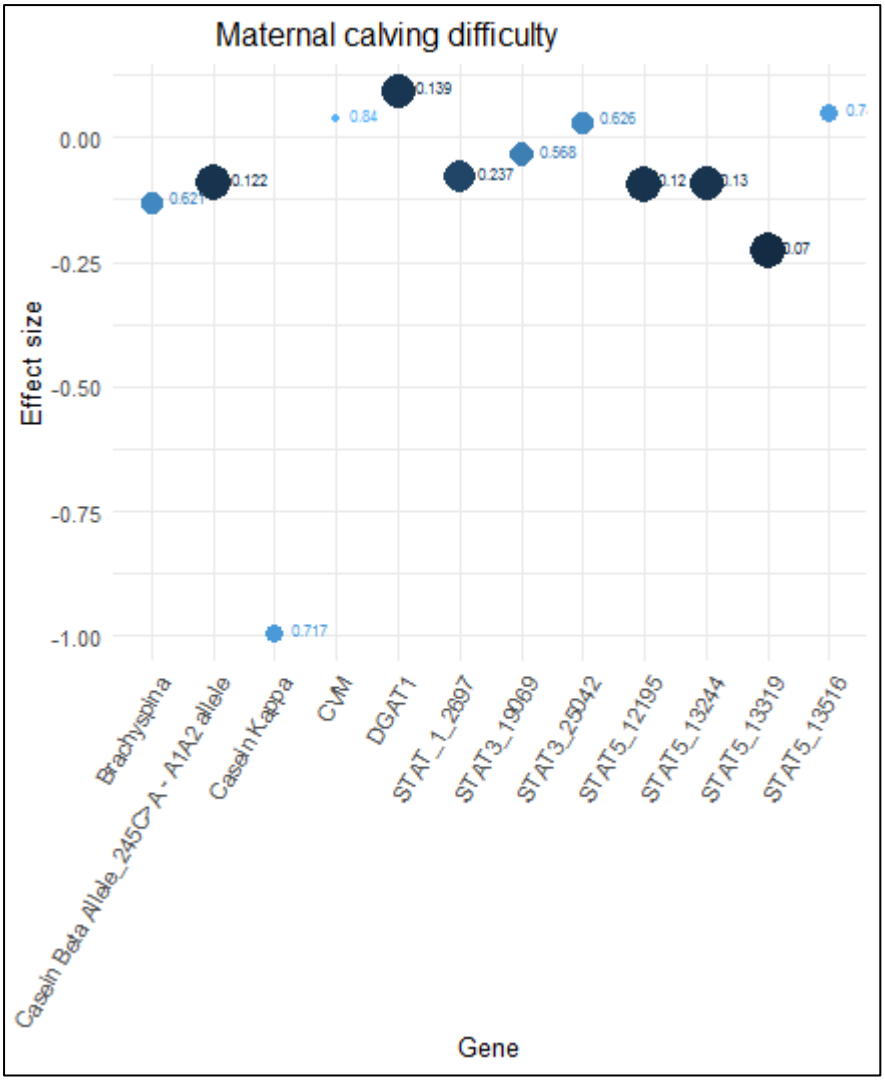


Fig. 1.10 – The effect sizes obtained with regards to maternal calving difficulty after substitution for the alternative allele in the association study.

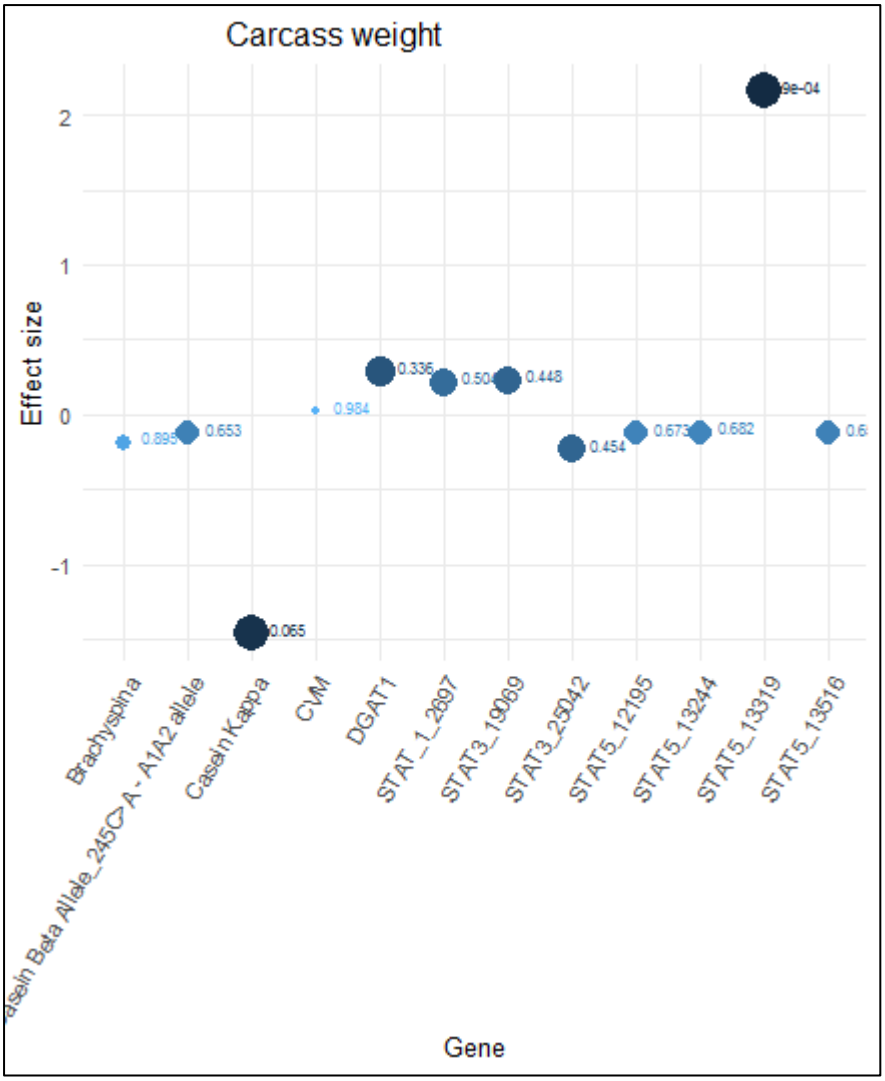


Fig. 1.11 – The effect sizes obtained with regards to carcass weight after substitution for the alternative allele in the association study.

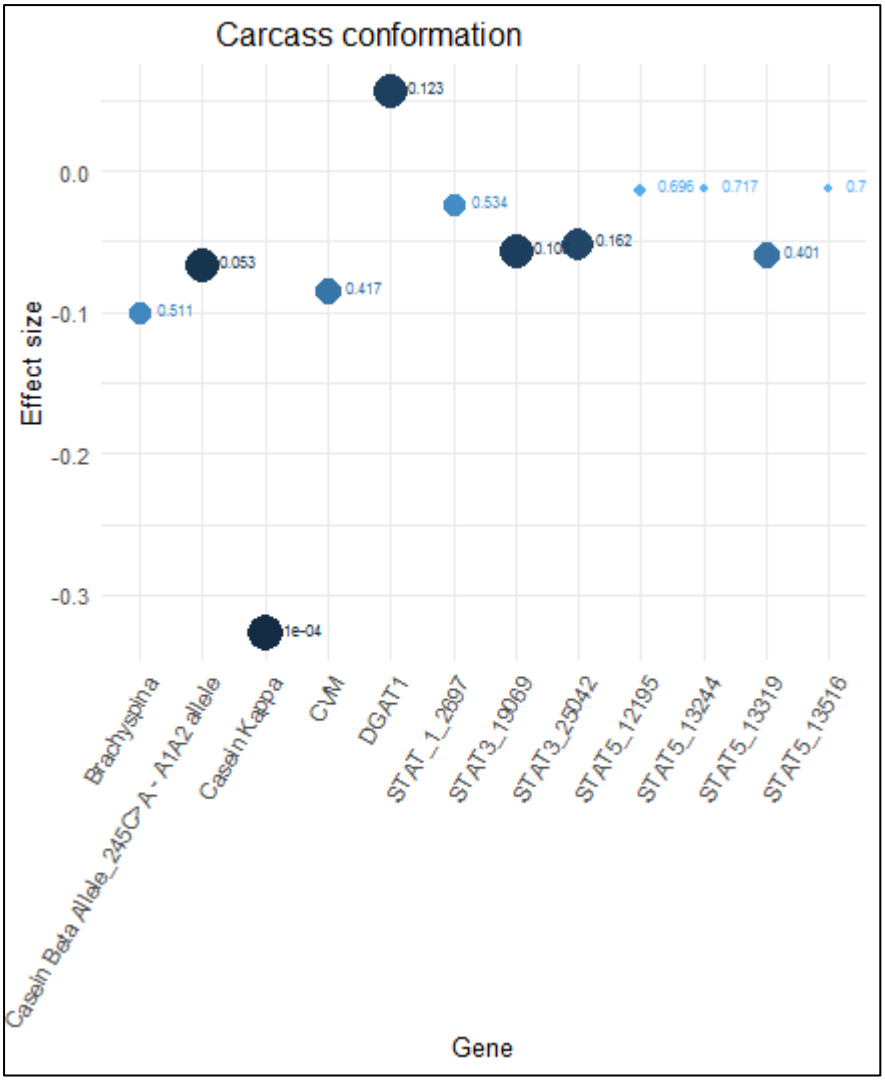


Fig. 1.12 – The effect sizes obtained with regards to carcass conformation after substitution for the alternative allele in the association study.

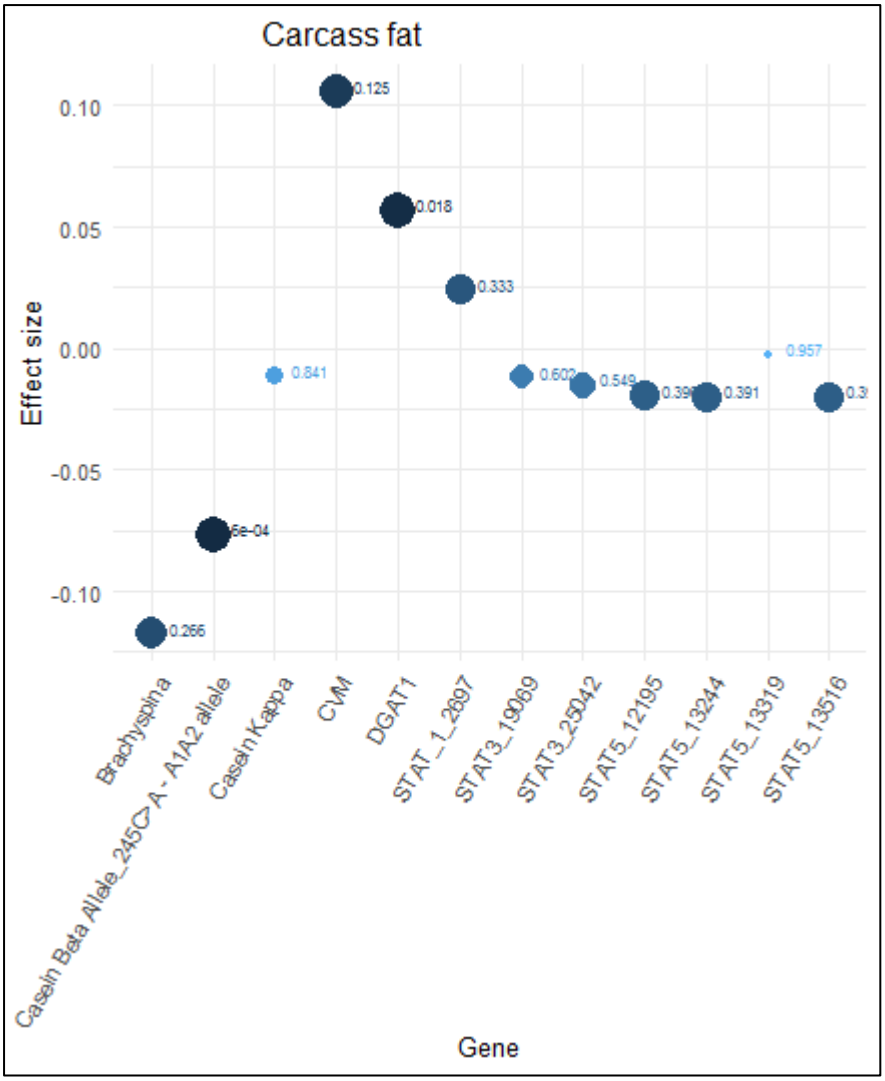


Fig. 1.13 – The effect sizes obtained with regards to carcass fat after substitution for the alternative allele in the association study.

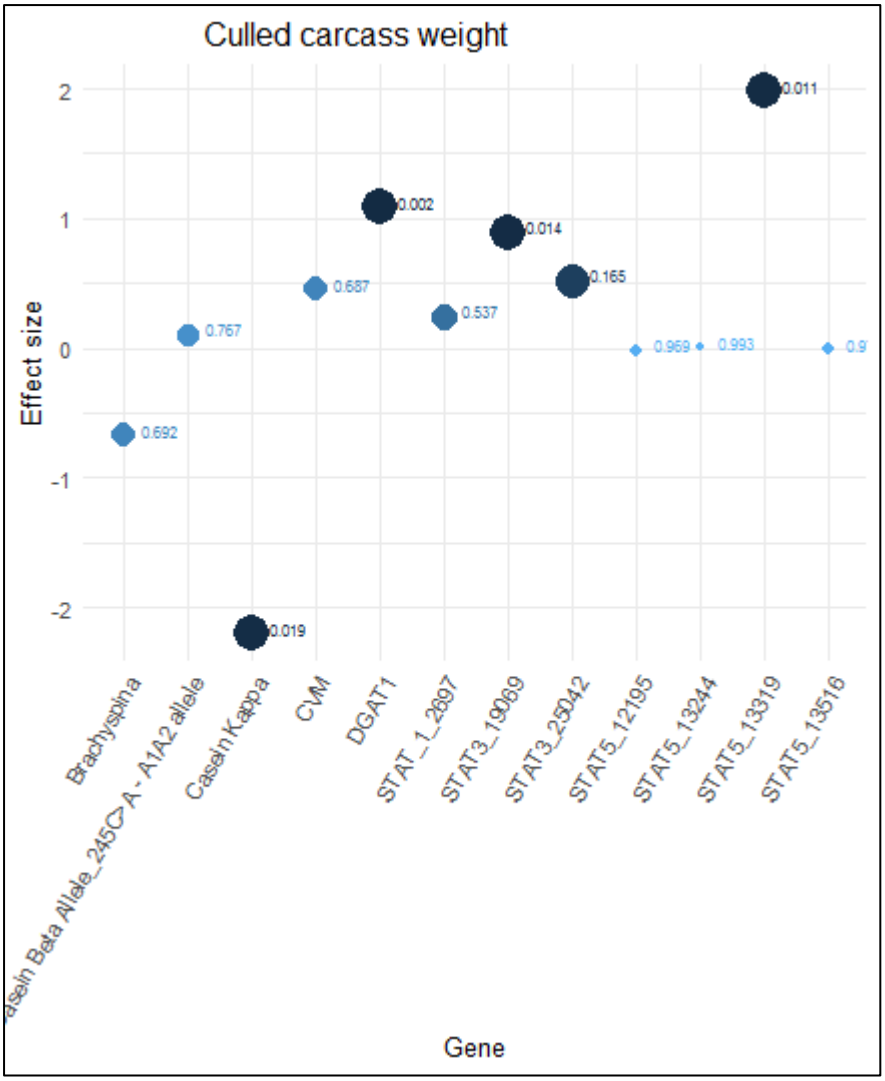


Fig. 1.14 – The effect sizes obtained with regards to culled carcass weight after substitution for the alternative allele in the association study.

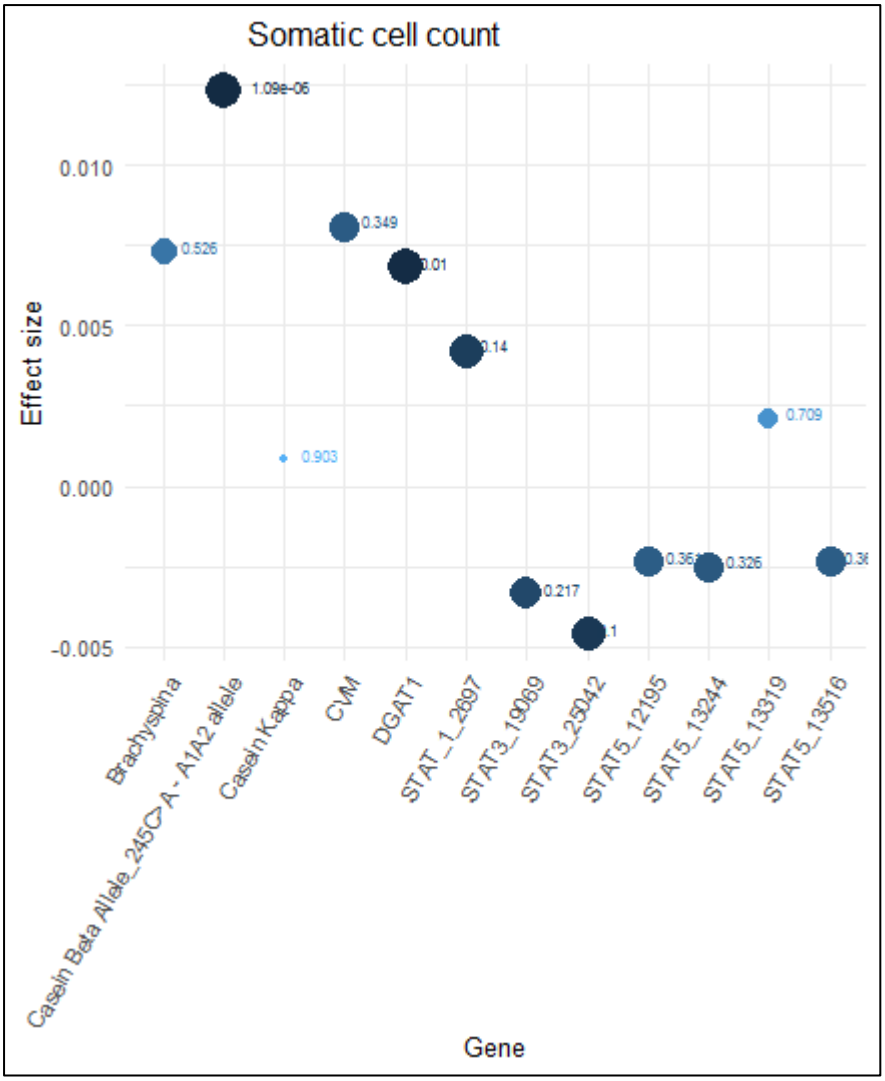


Fig. 1.15 – The effect sizes obtained with regards to somatic cell count after substitution for the alternative allele in the association study.

Appendix II

The following appendix section displays the relevant code that was used to analyse the datasets that were received from the ICBF and the code used for the comparative genomics described in the methods section.

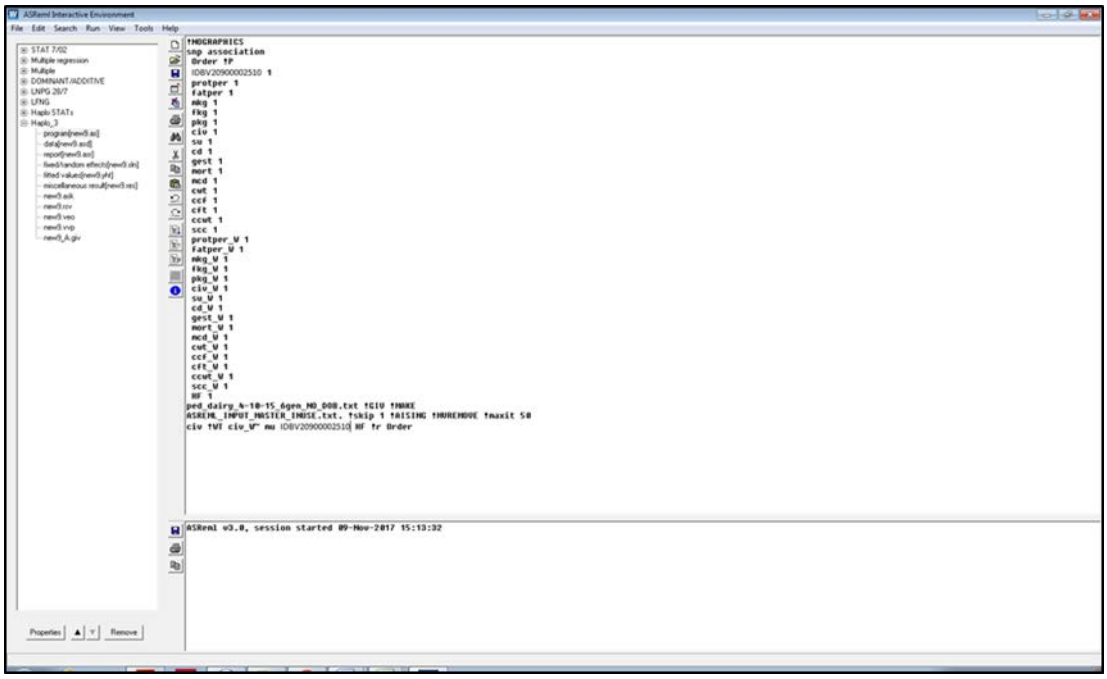


Fig. 1 – ASReml input file used for the association analysis

The following code was used in R to calculate summary statistics for all the phenotypes being analysed.

```
setwd("C:/Users/l.ratcliffe/Desktop")
install.packages("dplyr")
if(!require(devtools)) install.packages("devtools")
devtools::install_github("kassambara/ggpubr")
library("dplyr")
library("ggpubr")
setwd("C:/Users/l.ratcliffe/Desktop/Phenotypes_R_0.2.txt")
my_data <- read.table("Phenotypes_R_0.2.txt", header = TRUE)
dplyr::sample_n(my_data, 10)
sapply(my_data, is.factor)
#HISTOGRAMs
hist(my_data$PPC,main="Histogram for PPC",xlab ="PPC")
summary(my_data)
write.table((summary_table), file = "Summary_stats3", sep = ",", quote = FALSE, row.names = F)
#calculating variance
var(my_data$PPC,na.rm=TRUE)
#Normality test (Anderson darling) for all traits
library(nortest)
ad.test(my_data[,2])
```

The following R script was used to produce the bubble plots respresenting the results obtained in the association analysis.

```
setwd("C:/Users/l.ratcliffe/Documents/R")

install.packages("ggplot2")

data <- read.csv("Bubble_plots.csv")

ggplot(data, aes(x = FKG, y = Effect.size, size = p.value) + geom_point(shape = 16) +
scale_size_area(max_size = 6) + theme_bw() + theme(axis.text.x = element_text(angle=60,
hjust=1)) + ggtitle("Fat (kg)") + labs(x = "SNPs", y = "Effect Size", size="P Value", col="Value")
+ scale_size(trans = 'reverse')
```

The following R script was used to calculate the False Discovery Rate (FDR) q values.

```
setwd("C:/Users/l.ratcliffe/Documents")

list.files()

data=read.table("p_values.txt")

summary(data)

pval=data$V1

pval

install.packages("fdrtool")

library("fdrtool")

fdr=fdrtool(pval,statistic="pvalue")

qval=fdr$qval

fdrval=fdr$lfdr

data2=cbind(pval, qval, fdrval)

summary(data2)

write.table(data2, file="Lyndsey_Qvalues.txt", sep="\t", row.names=FALSE)

summary(data2)
```

The following R script was used to deregress the EBV values that were obtained from the ICBF, before the association analysis was performed.

```
dEBV <- function(trait,Pedigree,genoIDs,dataformat,h2,p.varSNP,outname){
  cat('\n.....  Deregression procedure following Garrick et al. 2009  .....\n')
  # Deregression of EBVs following Garrick et al. 2009
  # Part of the script were sourced from Badke et al. 2014

  #-------------------------------------------------------------------------------------------------------------
---#
  # The deregression scripts requires the following parameters and information to run sucessfully

  #1. trait=""      +++++ The file containing the EBVs, columns = (ID, EBV, reliability)
  #2. Pedigree=""   +++++ The file containing the Pedigree information, columns = (ID, Sire,
Dam)
  #3. genoIDs=""    +++++ The file containing the IDs of the genotyped individuals
  #4. dataformat="" +++++ Specify if the files are in the directory or it an R-object (options are
'DIR','R-object')
  #5. h2=""         +++++ heritabilty of the traits
  #6. p.varSNP=""   +++++ proportion of genetic variance accounted for by marker genotypes
  #7. outname=""    +++++ Output filename
  #-------------------------------------------------------------------------------------------------------------
---#

  if (dataformat=="DIR"){
    Pedigree <- read.table(Pedigree,colClasses=c("character","character","character"))
    cat('\n.....  Pedigree file imported  .....\n')
    trait <- read.table(trait,colClasses=c("character","numeric","numeric"))
    cat('.....  Phenotype file imported  .....\n')
```

```
cat('.....  IDs of genotyped individuals read  .....\n')
  }

  cat('\n.....  Preparing file for deregression  .....\n')

  PedTraits <- merge(x=trait,y=Pedigree,by.x=1,by.y=1)
  PedTraitsgenoIDs <- merge(x=PedTraits,y=genoIDs,by.x=1,by.y=1)
  colnames(x=PedTraitsgenoIDs) <- c("ID","ID_EBV","ID_Rel","SireID","DamID")
  Rel.sire <- merge(x=PedTraits,y=PedTraitsgenoIDs,by.x=1,by.y=4)
  Rel.sire <- Rel.sire[,-4:-5]
  colnames(Rel.sire)[1:3] <- c('SireID','Sire_EBV','Sire_R2')
  Rel.dam <- merge(x=PedTraits,y=Rel.sire,by.x=1,by.y=7)
  Rel.dam <- Rel.dam[,-4:-5]
  colnames(Rel.dam)[1:3] <- c('DamID','Dam_EBV','Dam_R2')
  data <-
Rel.dam[,c('ID','ID_EBV','ID_Rel','SireID','Sire_EBV','Sire_R2','DamID','Dam_EBV','Dam_R2')
]
  dEBV <- data[,-c(1,4,7)]
  dEBV$h2 <- h2
  dEBV$p.varSNP <- 1-p.varSNP

  Debv_Garrick <- function (ebv_mat){
    lambda = (1 - ebv_mat[7])/ebv_mat[7]
    PA = (ebv_mat[3] + ebv_mat[5])/2
    rPA = (ebv_mat[4] + ebv_mat[6])/4
    alpha = 1/(0.5 - rPA)
    delta = (0.5 - rPA)/(1 - ebv_mat[2])
    ZpZPA = lambda*(0.5*alpha - 4) + 0.5*lambda*sqrt(alpha^2 +16/delta)
```

```
ZpZPA = lambda*(0.5*alpha - 4) + 0.5*lambda*sqrt(alpha^2 +16/delta)
   ZpZi = delta*ZpZPA + 2*lambda*(2*delta - 1)
   LHS = rbind(cbind(ZpZPA + 4*lambda, -2*lambda),cbind(-2*lambda, ZpZi + 2*lambda))
   #L1 = solve(LHS)
   RHS = LHS %*% c(PA, ebv_mat[1])
   drgi = RHS[2]/ZpZi
   rdrg = 1 - lambda/(ZpZi + lambda)
   we = (1 - ebv_mat[7])/((ebv_mat[8] + (1 - rdrg)/rdrg)*ebv_mat[7])
   ret = c(ebv_mat[1],ebv_mat[2],round(drgi,3),round(rdrg,3),round(we,3))
   names(ret) <- c("EBV", "r2EBV", "dEBV","r2dEBV","weight")
   return(ret)
 }


 cat('.....  File preparation finished  .....\n\n')
 cat('.....  Deregression process started  .....\n')
 cat('.....  Deregression script is a modified version of Badke et al. 2014  .....\n')


 Deregress <- t(apply(X=dEBV,MARGIN=1,FUN=Debv_Garrick))
 Deregress <- cbind.data.frame(IID=data[,1],Deregress)

write.table(Deregress,paste(outname,".debv",sep=""),col.names=T,row.names=F,quote=F,sep='\t'
)


 cat('.....  Deregression finished    .....\n\n')
 cat(paste('.....  Output files exported as ',outname,'.debv .....',sep=''),'\n')


 return(Deregress)
}
```

The following python script was used to generate the haploview input file.

```
import os
os.chdir("C:\Users\l.ratcliffe\Documents\Haploview")
os.listdir(os.curdir)
LR=open("Haploview pedfile.csv").readlines()
out=open("Haploview.ped.21k.txt","w")
pop=[]
mom=[]
for i in test[1:]:
    line=i.split(",")
    pop.append(line[2])
    mom.append(line[3])
for i in LR[1:]:
    line=i.split(",")
    out.write("L"+line[0]+" "+line[1]+" "+line[2]+" "+line[3]+" ")

    if line[1]in pop:
        out.write("1"+" ")
    else:
        out.write("2"+" ")

    out.write(line[5]+" ")
```

```
for geno in line[6:]:
        if geno == "0":
            out.write("1"+" "+"1"+" ")
        elif geno == "1":
            out.write("1"+" "+"2"+" ")
        elif geno == "2":
            out.write("2"+" "+"2"+" ")
        else:
            out.write("0"+" "+"0"+" ")
    out.write("\n")
out.close()
```

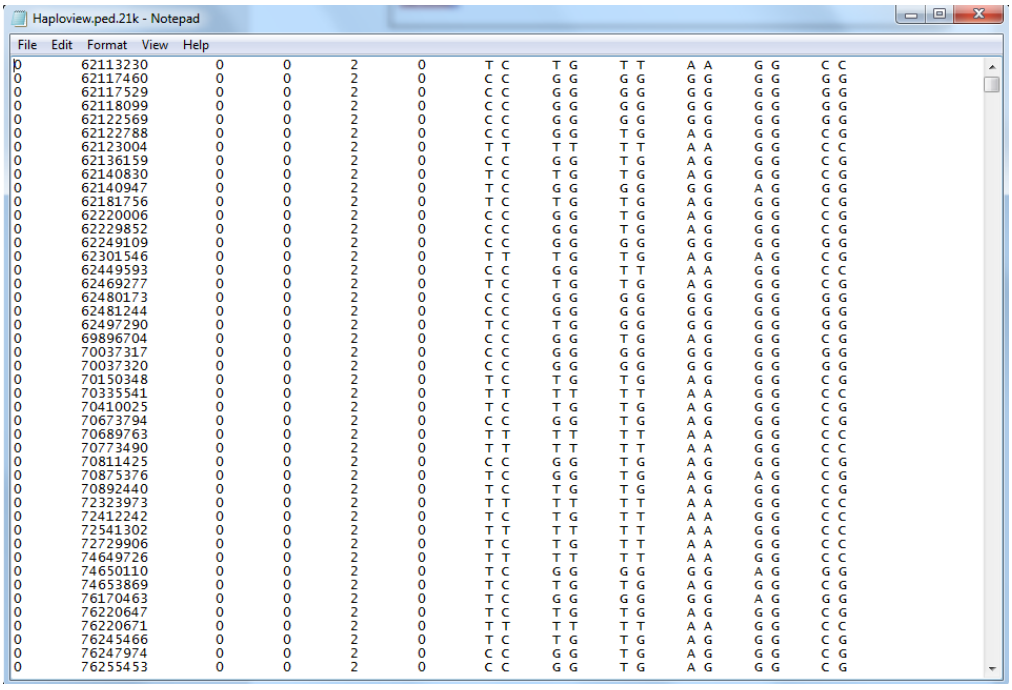


Fig 2. Example of the haplotype input file generated from the above code

The following script was used to generate an input file suitable for Phase software requirements.

```
""" Input file format
The input file is supplied by the user to specify how many individuals there
are to be analysed, how many loci/sites each individual has been typed at,
what sort of loci/sites these are (SNP or microsatellite), and the genotypes
for each individual. Optionally, the file may also specify a group label for
each individual (eg case/control status), and the relative physical positions
of the markers.
There are three possible formats for the input file: this section describes
the default format, as illustrated in the accompanying file test.inp. The
alternative formats are described later (section 7.1).
The default structure for the input file can be represented as follows:
NumberOfIndividuals
NumberOfLoci
P Position(1) Position(2) Position(NumberOfLoci)
LocusType(1) LocusType(2) ... LocusType(NumberOfLoci)
ID(1)
Genotype(1)
ID(2)
Genotype(2)"""


import os
os.chdir("C:\Users\l.ratcliffe\Documents\phase")
os.listdir(os.curdir)
```

```
donor=open("Donor.txt","r").readlines()

out=open("PHASE_INP.inp","w")

header=donor[0].split()

snps=[]

for i in header[1:]:

    snps.append(i)

out.write(str(len(donor)-1)+"\n"+

          str(6)+"\n"+


"P"+"\t"+str(43045807)+"\t"+str(43046856)+"\t"+str(43046931)+"\t"+str(43047128)+"\t"+str(
43063963)+"\t"+str(43070296)+"\n")

for i in range(len(snps)):

    out.write("S"+"\t")


out.write("\n")

for i in donor[1:]:

    line=i.split()

    out.write("#"+line[0]+"\n")

    count=6
for sn in snps:

        if line[count]=="NA":

            out.write("?"+"\t")

        elif line[count]=="0":

            out.write("0"+"\t")

        elif line[count]=="1":

            out.write("0"+"\t")

        else:

            out.write("1"+"\t")

        count-=1

    out.write("\n")
```

```
count=6
    for sn in snps:
        if line[count]=="NA":
            out.write("?"+"\t")
        elif line[count]=="0":
            out.write("0"+"\t")
        elif line[count]=="1":
            out.write("1"+"\t")
        else:
            out.write("1"+"\t")
        count-=1
    out.write("\n")


    count-=1



out.close()
```

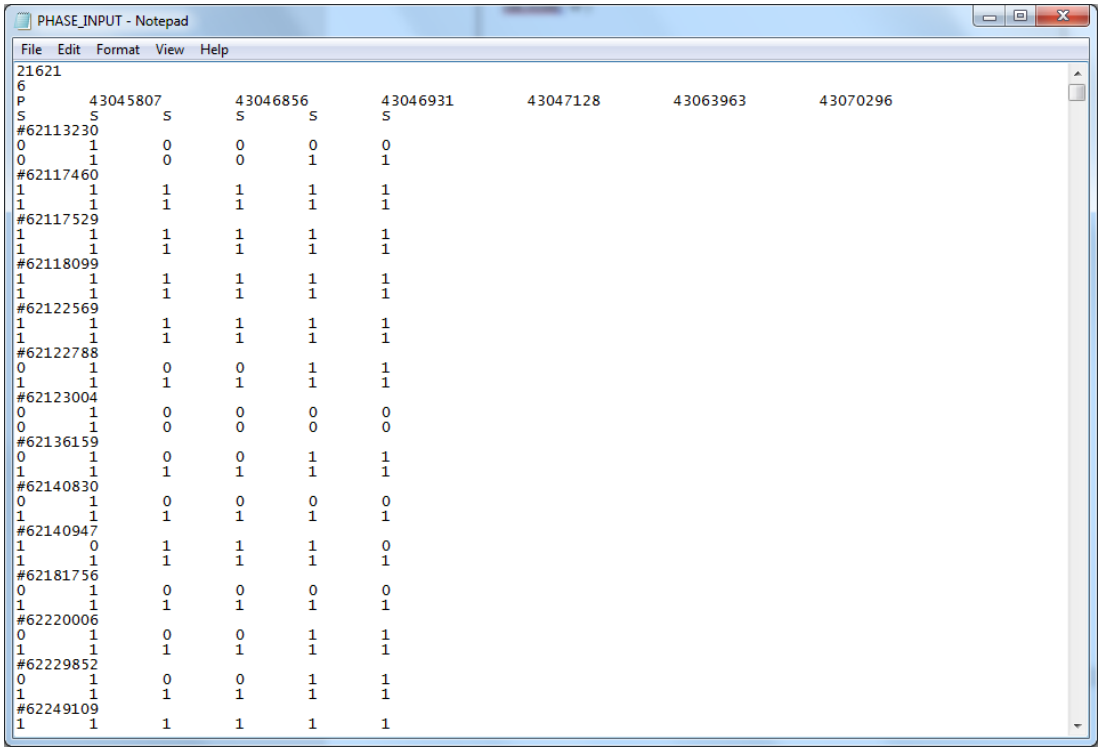


Fig. 3 – Output from the above code with the data reformatted for Phase analysis

The following Python script was used to prepare a file from the haploview output for analysis in Asreml.

```
import os
os.chdir(r"C:\Users\l.ratcliffe\Documents\Phase")
os.listdir(os.curdir)
data=open("Haplotypes_21k.txt","r").readlines()
haps={}
for i in data:
    line=i.split()
    animal=line[0].strip("#")
    animal=animal.strip(":")
    Haps=line[1].strip("(")
    Haps=Haps.strip(")")
    Haps=Haps.split(",")
    haps[animal]={"H1":Haps[0], "H2":Haps[1]}
out=open("Haplotypes_Asreml_format.txt","w")
range(1,21)
out.write("ID"+"\t"+"H1"+"\t"+"H2"+"\t"+"H3"+"\t"+"H4"+"\t"+"H5"+"\t"+"H6"+"\t"+"H7"+"
\t"+"H8"+"\t"+"H9"+"\t"+"H10"+"\t"+"H11"+"\t"+"H12"+"\t"+"H13"+"\t"+"H14"+"\t"+"H15"
+"\t"+"H16"+"\t"+"H17"+"\t"+"H18"+"\t"+"H19"+"\t"+"H20"+"\n")
for i in data:
    line=i.split()
    animal=line[0].strip("#")
    animal=animal.strip(":")
    out.write(animal+"\t")
    H1=haps[animal]["H1"]
    H2=haps[animal]["H2"]
```

```
for i in range(1,21):
        if H1!=H2 and i==int(H1):
            out.write("1"+"\t")
        elif H1!=H2 and i==int(H2):
            out.write("1"+"\t")
        elif H1==H2 and i==int(H1):
            out.write("2"+"\t")
        elif H1==H2 and i==int(H2):
            out.write("2"+"\t")
        else:
            out.write("0"+"\t")
    out.write("\n")


out.close()
```

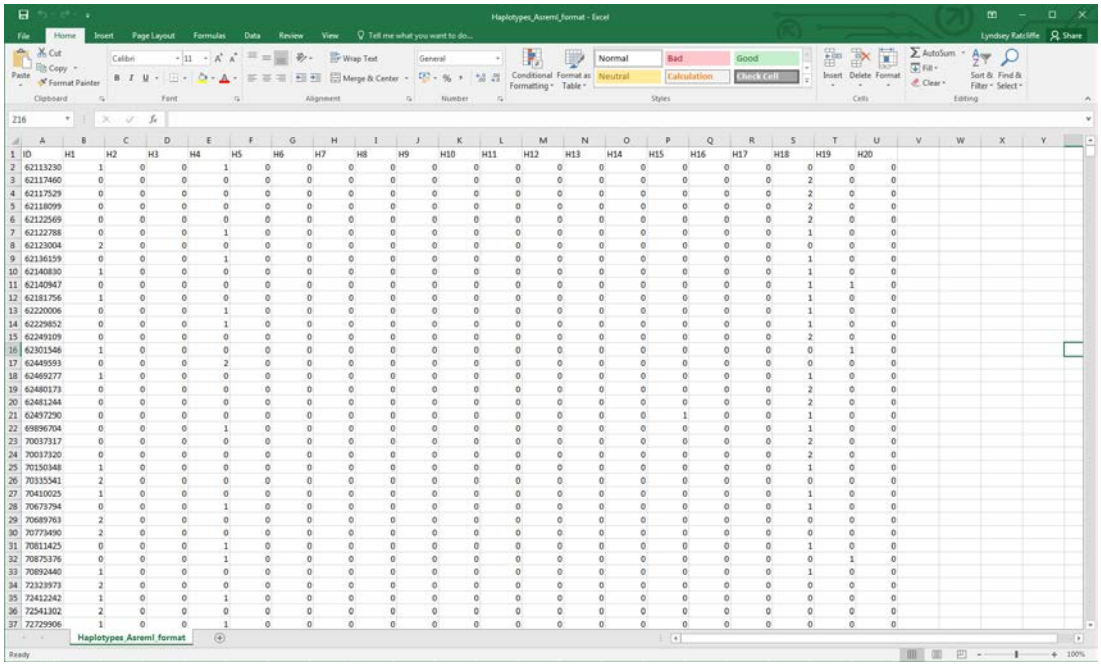


Fig. 4 - Output from the above code, with the haplotypes formated correctly for Asreml

The following Python script was used to reduce the dataset to animals with an adjusted reliability of 0.2 only.

```
# coding: utf-8


# In[1]:


import os
os.chdir("C:\Research\Postgrads\ICBF data Dec17\Dairy_21k")
os.listdir(os.curdir)



# In[46]:


# open current files
phenos=open("21K_Phenos_26-4-18_uniques_only_ADJRel0.1.csv","r").readlines()
rels=open("21K_rels_0.1_26-4-18.csv","r").readlines()
weights=open("Weights_Dereg_PTA_DG2009_c_0.1_21K_rels_0.1_26-4-18.csv").readlines()



# In[47]:


# create new files for adjrel >=0.2 data
weights_new=open("Weights_Dereg_PTA_DG2009_c_0.1_21K_rels_0.2_4-5-18.txt","w")
rels_new=open("21K_rels_0.2_4-5-18.txt","w")
phenos_new=open("21K_Phenos_4-5-18_uniques_only_ADJRel0.2.txt","w")
```

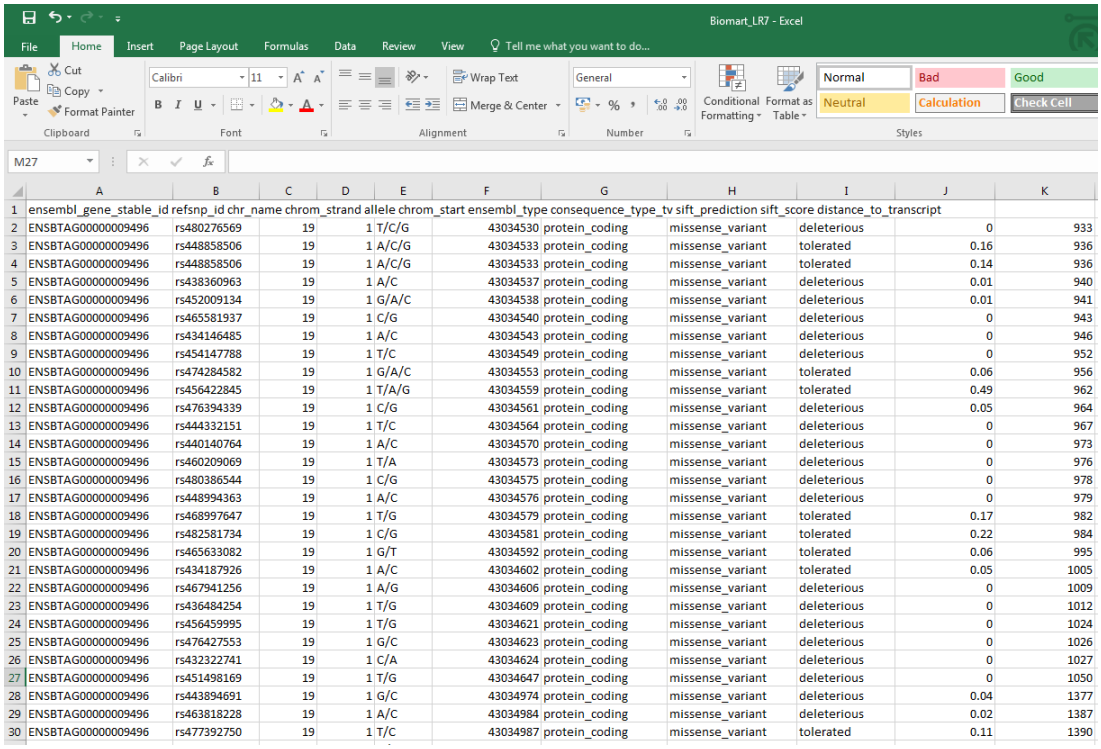
```
# In[48]:


# write header to each new file
weights_head=weights[0]
rel_head=rels[0]
pheno_head=phenos[0]


for i in range(len(pheno_head.split(","))):
    weights_new.write(weights_head.split(",")[i].strip()+"\t")
    rels_new.write(rel_head.split(",")[i].strip()+"\t")
    phenos_new.write(pheno_head.split(",")[i].strip()+"\t")


weights_new.write("\n")
rels_new.write("\n")
phenos_new.write("\n")
# In[49]:


for i in range(1,len(phenos)):
    animal=phenos[i].split(",")[0]
    order=phenos[i].split(",")[1]


    weights_new.write(animal+"\t")
    rels_new.write(animal+"\t")
    phenos_new.write(animal+"\t")


    weights_new.write(order+"\t")
    rels_new.write(order+"\t")
    phenos_new.write(order+"\t")
```

```
for x in range(2,len(pheno_head.split(","))):
        pheno_line=phenos[i].split(",")
        rel_line=rels[i].split(",")
        weights_line=weights[i].split(",")

        try:
            if float(rel_line[x].strip())<0.2:
                weights_new.write("NA"+"\t")
                rels_new.write("NA"+"\t")
                phenos_new.write("NA"+"\t")

            elif float(rel_line[x].strip())>=0.2:
                weights_new.write(weights_line[x].strip()+"\t")
                rels_new.write(rel_line[x].strip()+"\t")
                phenos_new.write(pheno_line[x].strip()+"\t")

        except ValueError:
                weights_new.write("NA"+"\t")
                rels_new.write("NA"+"\t")
                phenos_new.write("NA"+"\t")

    weights_new.write("\n")
    rels_new.write("\n")
    phenos_new.write("\n")

weights_new.close()
rels_new.close()
phenos_new.close()
```

The following R script was used to generate a file containing specific SNPs based on their consequences for the comparative genomics section of this research.

```
##### Example of pulling Ensembl IDS in for loop and writing out snps to file
############################


setwd("C:/Users/l.ratcliffe/Documents/Comparative genomics")

list.files()

####Genes of interest ensemble ID after literature review######

LR_genes=read.table("Biomart_genes_list.txt", sep=" ", header=FALSE)

LR_genes

LR_genes=t(LR_genes)

LR_genes=as.list(LR_genes, sep=" ")

str(LR_genes)


source("http://bioconductor.org/biocLite.R")

biocLite("BiocUpgrade")

biocLite("biomaRt")

library("biomaRt")


btaugen<-useMart("ensembl")

cowinfo<-useDataset("btaurus_gene_ensembl", mart=btaugen)


snps<-useMart("ENSEMBL_MART_SNP")

listDatasets(snps)

btau<-useDataset("btaurus_snp", mart=snps)

listAttributes(btau)
```

```
genes=LR_genes

genes




colnames(test) <- c("ensembl_gene_stable_id refsnp_id chr_name chrom_strand allele
chrom_start ensembl_type consequence_type_tv sift_prediction sift_score
distance_to_transcript")

write(header,file="biomart_LR_new.txt", sep=",",append=TRUE)



for (gene in genes) {

  data=getBM(attributes=c("start_position", "end_position", "chromosome_name"),
filters="ensembl_gene_id",values=gene, mart=cowinfo)

  chr=c(data[,3])

  start=data[,1]

  end=data[,2]

  SNPS<-getBM(c("ensembl_gene_stable_id",
"refsnp_id","chr_name","chrom_strand","allele","chrom_start","ensembl_type","consequence_ty
pe_tv","sift_prediction","sift_score","distance_to_transcript"), filters=c("start",
"end","chr_name"),values=list(start,end,chr), mart=btau)

  TopSNPs=SNPS[SNPS$sift_prediction=="deleterious",]

  #write.table(SNPS, file="test2.csv", sep=",",col.names=FALSE, row.names=FALSE,
append=TRUE)

  write.table(TopSNPs, file="biomart_LR7.txt", sep=",",col.names=FALSE, row.names=FALSE,
append=TRUE)

}
```



Biomart_LR7 - Excel

ensembl_gene_stable_id	refsnp_id	chr_name	chrom_strand	allele	chrom_start	ensembl_type	consequence_type_tv	sift_prediction	sift_score	distance_to_transcript
ENSBTAG00000009496	rs480276569	19	1	T/C/G	43034530	protein_coding	missense_variant	deleterious	0	933
ENSBTAG00000009496	rs448858506	19	1	A/C/G	43034533	protein_coding	missense_variant	tolerated	0.16	936
ENSBTAG00000009496	rs448858506	19	1	A/C/G	43034533	protein_coding	missense_variant	tolerated	0.14	936
ENSBTAG00000009496	rs438360963	19	1	A/C	43034537	protein_coding	missense_variant	deleterious	0.01	940
ENSBTAG00000009496	rs452009134	19	1	G/A/C	43034538	protein_coding	missense_variant	deleterious	0.01	941
ENSBTAG00000009496	rs465581937	19	1	C/G	43034540	protein_coding	missense_variant	deleterious	0	943
ENSBTAG00000009496	rs434146485	19	1	A/C	43034543	protein_coding	missense_variant	deleterious	0	946
ENSBTAG00000009496	rs454147788	19	1	T/C	43034549	protein_coding	missense_variant	deleterious	0	952
ENSBTAG00000009496	rs474284582	19	1	G/A/C	43034553	protein_coding	missense_variant	tolerated	0.06	956
ENSBTAG00000009496	rs456422845	19	1	T/A/G	43034559	protein_coding	missense_variant	tolerated	0.49	962
ENSBTAG00000009496	rs476394339	19	1	C/G	43034561	protein_coding	missense_variant	deleterious	0.05	964
ENSBTAG00000009496	rs444332151	19	1	T/C	43034564	protein_coding	missense_variant	deleterious	0	967
ENSBTAG00000009496	rs440140764	19	1	A/C	43034570	protein_coding	missense_variant	deleterious	0	973
ENSBTAG00000009496	rs460209069	19	1	T/A	43034573	protein_coding	missense_variant	deleterious	0	976
ENSBTAG00000009496	rs480386544	19	1	C/G	43034575	protein_coding	missense_variant	deleterious	0	978
ENSBTAG00000009496	rs448994363	19	1	A/C	43034576	protein_coding	missense_variant	deleterious	0	979
ENSBTAG00000009496	rs468997647	19	1	T/G	43034579	protein_coding	missense_variant	tolerated	0.17	982
ENSBTAG00000009496	rs482581734	19	1	C/G	43034581	protein_coding	missense_variant	tolerated	0.22	984
ENSBTAG00000009496	rs465633082	19	1	G/T	43034592	protein_coding	missense_variant	tolerated	0.06	995
ENSBTAG00000009496	rs434187926	19	1	A/C	43034602	protein_coding	missense_variant	tolerated	0.05	1005
ENSBTAG00000009496	rs467941256	19	1	A/G	43034606	protein_coding	missense_variant	deleterious	0	1009
ENSBTAG00000009496	rs436484254	19	1	T/G	43034609	protein_coding	missense_variant	deleterious	0	1012
ENSBTAG00000009496	rs456459995	19	1	T/G	43034621	protein_coding	missense_variant	deleterious	0	1024
ENSBTAG00000009496	rs476427553	19	1	G/C	43034623	protein_coding	missense_variant	deleterious	0	1026
ENSBTAG00000009496	rs432322741	19	1	C/A	43034624	protein_coding	missense_variant	deleterious	0	1027
ENSBTAG00000009496	rs451498169	19	1	T/G	43034647	protein_coding	missense_variant	deleterious	0	1050
ENSBTAG00000009496	rs443894691	19	1	G/C	43034974	protein_coding	missense_variant	deleterious	0.04	1377
ENSBTAG00000009496	rs463818228	19	1	A/C	43034984	protein_coding	missense_variant	deleterious	0.02	1387
ENSBTAG00000009496	rs477392750	19	1	T/C	43034987	protein_coding	missense_variant	tolerated	0.11	1390

Fig. 5 – Output from the above code where a listing of all SNPs in each gene of interest was produced