Genetics of flystrike, dagginess and associated traits in New Zealand dual-purpose sheep

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Animal Science

at Massey University, Palmerston North, New Zealand

Natalie Kathleen Pickering

2013

"The greatest benefit of dag control or elimination may come from reduced fly strike" V. J. Mackereth (1983), an astute breeder.

Abstract

Pickering, N. K. (2013). Genetics of flystrike, dagginess and associated traits in New Zealand dual-purpose sheep. PhD thesis, Massey University, Palmerston North, New Zealand.

A literature review identified breech bareness, dagginess and fibre traits as potential indirect indicator traits for flystrike. Dagginess (faecal accumulation) had the greatest potential as an indirect indicator, and has been identified as an important trait itself. Therefore flystrike and dagginess were investigated for their associations with fibre and production traits. A genome-wide association study (GWAS) was performed to identify regions under selection and associated with these traits. Finally, a genomic selection (GS) analysis was performed for dagginess and dual-purpose production traits to estimate molecular breeding values (MBVs) and to determine their impact on the New Zealand dual-purpose selection index.

Heritability, genetic and phenotypic parameter estimations were performed on a flystrike case-control dataset collected over 2 years. Flystrike had a heritability of 0.37, and high genetic and phenotypic correlations with dag score and a high genetic correlation with the coefficient of variation of fibre diameter. A similar analysis was performed on an existing New Zealand sheep industry dataset of about 2 million pedigree-recorded animals born between 1990 and 2008. The heritability for dag score at 3 and 8 months (DAG3, DAG8) was 0.34 and 0.31 respectively. There were low or nil genetic and phenotypic correlations of DAG3 and DAG8 with the other standard live weight, fleece weight, reproduction and faecal egg count production traits or breech bareness, fibre and wool traits.

A GWAS performed on an industry dataset of 8,705 genotyped animals, using phenotype information on about 3 million pedigree-recorded animals, identified regions on chromosome 6 and 15 associated with DAG3 and DAG8. The lambs from the flystrike case-control dataset with SNPs imputed from 5K to 50K identified a number of immune, diarrhoea and wool/hair growth genes associated with flystrike, dag score and fibre traits in a GWAS. There were no similarities in the genes identified in the industry or case-control GWAS; however, the SNP on chromosome 15 was re-identified in the GS analysis for DAG8. The GS analysis showed that genomic predictions can be

performed for DAG3 and DAG8 and that using MBVs and modifying generation interval can increase the rate of the genetic gain of the dual-purpose index by 84% per year.

Acknowledgements

At least 4 years ago and maybe back the full 28 years, I have had the guidance of a forward thinking father who did not want to crutch another sheep. Also along the last 6 years I have had the opportunity to work with another forward thinking and supportive, first boss then supervisor, John McEwan. Together I have them to thank for one suggesting the topic and the other in moulding my scientific skills so as to answer the question "Can I breed for no dags and flystrike?"

There are a number of people who I've had the opportunity to work with and learn off during this experience. Firstly my full contingent of supervisors: Hugh Blair, Rebecca Hickson from Massey University and Tricia Johnson and John McEwan from AgResearch, Invermay. Thank you for your contributions to shaping me, my experiments, and my thesis. Also for teaching me 'past and present' tense. I would also like to thank Ken Dodds and Benoit Auvray for their guidance and invaluable help with the statistical component of the project, without them the languages, programs and methods such as R, SAS, ASReml, genomic selection and GWAS would still be foreign outer space 'things'! I've also had the pleasure of sharing my workspace with the Invermay Animal Genomics team, some have come and gone since I first started at AgResearch, but all have contributed to the great atmosphere which has been a pleasure to work in.

Secondly, the financial support of Ovita and AgResearch for funding the research and for the provision of my doctoral stipend. Also the IVABS Travel Award, Early Career Scientist Travel Bursary, and the Charles Elgar Scholarship awards which have allowed me to attend and/or present my work at 9th World Congress on Genetics Applied to Livestock Production (2010), The Applied Genomics for Sustainable Livestock Breeding Conference (2011), New Zealand Society of Animal Production (2012), and the 32nd and 33rd International Society of Animal Genetics (2010, 2012) conferences.

Thirdly, this project would not be possible without the support of the farmers that have allowed me access to their flocks via SIL and have allowed me to look at ~ 30,000 dirty and clean lamb bottoms. Also the efforts of those that collected samples and measurements of flystruck lambs for me, a not very pleasant job, were very much

appreciated. A big thank you for those that provide me a bed and great home cooking during my road trips around the country collecting measurements; these provided some of my most memorable moments during my project.

A big thank you to my family; Mum, Dad and brothers for their amazing support throughout my life, and helping me through the PhD by being a willing participant in the case-control study. Finally to Jason, who has been a rock for me for the past 3 years, putting up with my stressed out moments and encouraging me to try new things like scuba diving.

I thank you all.

Table of contents

List of tables

List of figures

List of abbreviations

AFEC	adult faecal egg count
AFW	adult fleece weight
BBREECH	breech bareness
BULK	wool bulk
BV	breeding value
chr	chromosome
CURV	curvature
CW	carcass weight
DAG3	dag score at 3 months of age
DAG8	dag score at 8 months of age
dNTP	deoxynucleoside triphosphate
EWT	adult ewe live weight
FDCV	coefficient of variation of mean fibre diameter
FDSD	standard deviation of mean fibre diameter
FE	facial eczema
FEC1	faecal egg count in summer
FEC2	faecal egg count in autumn
FW12	fleece weight at 12 months
GBLUP	genomic best linear unbiased prediction
GBV	genomic breeding value
GC	genotype call score
GWAS	genome-wide association study
GS	genomic selection
HapMap	haplotype map
ISGC	International Sheep Genomic Consortium
IWTO	International Wool Textile Organisation
LENGTH	wool length

LFW	lamb fleece weight
LW6	live weight at 6 months
LW8	live weight at 8 months
MAF	minor allele frequency
MBV	molecular breeding value
MED%	proportion of medullated fibres
MFD	mean fibre diameter
MT-EBV	multi trait-estimated breeding value
NEM1	nematodirus egg counts in summer
NEM2	nematodirus egg counts in autumn
NLB	number of lambs born
NZWTA	New Zealand Wool Testing Authority
OFDA100	Optical Fibre Diameter Analyser 100
OMIM	Online Mendelian Inheritance in Man
PC	principal components
PETA	People for the Ethical Treatment of Animals
PEV	prediction error variance
QQ	quantile-quantile
RRS	reduced representational sequencing
SAP	shrimp alkaline phosphate
SIL	Sheep Improvement Limited
ST-EBV	single trait-estimated breeding value
SURV	lamb survival
SURVm	survival maternal
TBE	Tris base boric acid and ethylenediaminetetra-acetic acid buffer
TBV	true breeding value
WWT	weaning weight at 3 months
WWTm	maternal weaning weight

Introduction

Myiasis or flystrike is an animal health problem with considerable ethical and welfare implications affecting sheep industries worldwide. Its prevalence in New Zealand has been estimated at 3-5% of the national flock (overt strikes) (Heath and Bishop, 1995), but perhaps double that figure if covert strikes are accounted for. In New Zealand it has been estimated that flystrike costs the sheep industry $40 to $50 million annually (Beef and Lamb New Zealand, 1999). Costs include preventative treatment, weight and wool production losses, animal deaths and staff time. Current preventative strategies including insecticide use, suitable timing of animal husbandry techniques such as shearing, crutching, dagging and tail docking have been effective in reducing flystrike to a low incidence. There is now consumer concern regarding some preventative procedures, such as the use of insecticides in the production of food animals and mulesing in Merinos. Also, flies are now demonstrating increased resistance to insecticides. Consequently, there would be considerable benefit to the New Zealand sheep industry if new tools that reduced labour and insecticide use to prevent flystrike are developed.

The majority of flystrike in New Zealand occurs in the perineum region and dagginess, or the accumulation of faeces around the perineum, is thought to be strongly associated with flystrike. Dagginess is also an undesirable condition in its own right, due to labour costs associated with crutching and dagging before shearing, lamb slaughter and mating and lambing in ewes. These 2 traits would thus benefit from the selection and breeding of less flystrike susceptible and less daggy sheep. Heritability of flystrike in New Zealand dual-purpose sheep has been estimated at 0.18 ± 0.04 (Brandsma and Blair, 1997), and in Australian and South African Merinos at 0.26 ± 0.12 to 0.57 ± 0.28 (Greeff and Karlsson, 2009; Raadsma, 1991a; Scholtz et al., 2010; Smith et al., 2009). Dagginess is reported to have a high positive genetic correlation (0.86 ± 0.17) with flystrike in Merinos (Greeff and Karlsson, 2009).

Identification of an indicator trait for use in selection against flystrike could be beneficial, as it would be more cost effective than trying to measure flystrike directly. A direct measurement of flystrike is reliant on the season's exposure levels and identification of all flystruck cases, and is thus time consuming and costly for the

farmer. Also, for a true measurement, animals should be left untreated; however, purposefully exposing animals is undesirable and raises animal welfare concerns.

A consequence of the sequencing of the sheep genome has been the identification of thousands to hundreds of thousands of single nucleotide polymorphisms (SNP). Sheep can now be genotyped for a proportion of these identified SNPs using either the 50K SNP chip or the 5K SNP chip (Oddy et al., 2007). These can be used to identify genetic regions associated with a trait in genome wide association studies (GWAS), or for predicting an animal's genetic worth using genomic selection (GS).

The objectives of this study were to:

1. Calculate the genetic parameters for flystrike in New Zealand dual-purpose sheep and genetic and phenotypic correlations with dagginess and other potential indirect traits such as wool and fibre traits (Chapter 2)
2. Investigate dagginess and its genetic relationship with dual-purpose production traits (Chapter 3) and other potential flystrike indicator traits; wool and fibre traits (Chapter 4),
3. Determine the best parameters for imputation to provide accurate genotypes for subsequent GWAS, including a summary of deoxyribonucleic acid (DNA) sample type and paternity parentage (Chapter 5),
4. Use Ovine SNP chips to detect genomic chromosomal regions associated with dagginess, production, flystrike and wool and fibre traits (Chapter 6 and 7), and
5. Investigate GS for dagginess, and test the usefulness of SNP Chip technology in a dual-purpose selection index (Chapter 8).

Two animal experiments were conducted as part of this work to gather information on flystrike, dagginess and wool traits to achieve the above objectives:

1. A case-control experiment was performed on 11 farms with Romney and Romney-cross sheep. In 09/10 and 10/11, farmers collected ear tissue and wool samples from flystruck lambs (n = 836) and measured lambs for dag score, breech bareness score and flystrike location. Control contemporaries (n = 745) were collected at the end of the season. Tissue and blood samples were taken from all potential sires, so that paternity testing could be performed. All case and control lambs were run on the 5K SNP panel and the paternity assigned to sires on the Ovine 50K SNP panel, the lambs were imputed up to 50K (Chapter 2 and

5). Imputation is a method for estimating missing SNPs using the information of known SNP variants, linkage disequilibrium and haplotypes present in a reference population.

2. A progeny test was performed on 35 Sheep Improvement Limited (SIL) pedigree recorded flocks, all linked by the use of common sires. Progeny born in 2009 and 2010 (n = 29,419) were measured for dag score, breech bareness score, wool length, wool bulk; and a proportion of animals (~ 30 per flock/year) were measured for mean fibre diameter, standard deviation of mean fibre diameter, coefficient of variation of mean fibre diameter, curvature and proportion of medullated fibres. Tissue and blood samples were collected from the sires (n approximately 700) for use on the Ovine 50K SNP Chips (Chapter 4).

An additional dataset obtained from the SIL database of pedigree recorded animals born between 1990 and 2008 (n ~ 2.4 million) was used to investigate the heritability of dagginess in the industry and the correlations with standard production traits. This was to determine if there maybe any adverse effects of selection against dagginess on these production traits (Chapter 3). This dataset was also combined with the additional information obtained from the progeny test (animals born 2009 and 2010) for the GWAS in Chapter 7 and the genomic selection analysis in Chapter 8.

Chapter 1: Review of literature

1.1 Introduction

Myiasis or flystrike in sheep is an animal health problem that causes significant production losses and death. Prevention also consumes significant labour, and material costs. It also has an ethical dimension, both in relation to the management and treatment of animals, and also the ethics associated with some prevention methods such as mulesing. Finally, there is a consumer issue related to reluctance to purchase wool and consume meat products containing insecticides, anthelmintics and their residues.

There has been some previous research into the genetics of flystrike; results show that it is highly environmentally dependent. With low heritability it is extremely difficult to make genetic gains in resistance by selecting directly for this trait. Some indirect indicators have also been examined. In Australian Merinos, fleece rot and breech wrinkles have been successfully used as indirect indicators in breeding programs (Greeff and Karlsson, 2009; Smith et al., 2009). These traits are quite rare in New Zealand dual-purpose animals and thus other indirect indicators are required.

With the development of new sequencing technologies, and the subsequent decrease in costs, a number of domestic animals have been sequenced, including sheep. As a result, thousands of single nucleotide polymorphisms (SNPs) have been generated, all potentially in linkage disequilibrium with quantitative trait loci. With this new resource, a number of single gene traits have been identified. There is also evidence from other species already sequenced including human, cattle, and chicken that genes associated with polygenic traits can be more rapidly identified compared to traditional linkage studies.

Another new use of SNP information is the calculation of genomic breeding values. These can be estimated from the SNP chip genotypes and can predict the animal's life time potential at birth. The potential genetic gains resulting from this technology have already been demonstrated in dairy cattle (Hayes et al., 2009b). Similar genetic gains will be possible in sheep.

This review initially covers the insect species involved in initiating flystrike and their lifecycle, the economic cost of flystrike in major sheep producing countries, non-genetic control mechanisms, the genetic basis of host resistance and finally, the benefits of indirect and genomic selection (GS) to improve the intrinsic host resistance against infection.

1.2 Biology of flystrike

1.2.1 Species

In New Zealand the 4 main species of fly which initiate or aggravate flystrike are: *Lucilia cuprina*, *Lucilia sericata*, *Calliphora stygia* and C*hrysomya rufifacies*. *L. cuprina*, or as it is commonly known the "Australian sheep blowfly", was introduced into Australia in the late nineteenth, or early twentieth century, from South Africa (Wardhaugh, 2001; Watts et al., 1979). The Australian blowfly is thought to have been in New Zealand since the late 1970's, although it was first detected in 1988 (Heath and Bishop, 1995). *L. cuprina* is a primary strike species, that is, it initiates flystrike rather than depositing eggs after the primary strike has been established. It is the most aggressive fly in Australia and New Zealand and is the main initiator of flystrike in these countries.

L. sericata is the main fly species initiating flystrike in the United Kingdom. It is also present in Australia and New Zealand. It is not as aggressive as *L. cuprina*, however it will initiate strikes. *L. cuprina* can be distinguished from its cousin *L. sericata* (Figure 1.1 A, B) by the increased sclerotisation of the clypeus and the metallic green colouration of the fore femora of *L. cuprina* and the increased number of hairs of the humeral calli and notopleuron area in *L. sericata* (Lang et al., 2001).

C. stygia, the common brown fly, is native to Australia but is also present in New Zealand. Its grey thorax and yellow-brown mottled abdomen, separates this species from the 3 other fly species mentioned here (Figure 1.1, C). It is also a primary strike species, initiating flystrike.

C. rufifacies, the hairy maggot fly, is a secondary species. It does not initiate flystrike but rather targets carrion or live sheep previously struck by one of the primary species. The larvae are competitive, often devouring larvae of other species (Barton, 2007). It is distinguished from the *Lucilia* species by being smaller, but has very similar colouring (Figure 1.1,D), except for a white coloured anterior thoracic spiracle (Lang et al., 2001).

There are numerous other fly species involved in flystrike (Azeredo-Espin and Lessinger, 2006; Farkas et al., 1997; Hall, 1997), however they are not present in New Zealand and will not be discussed further here.

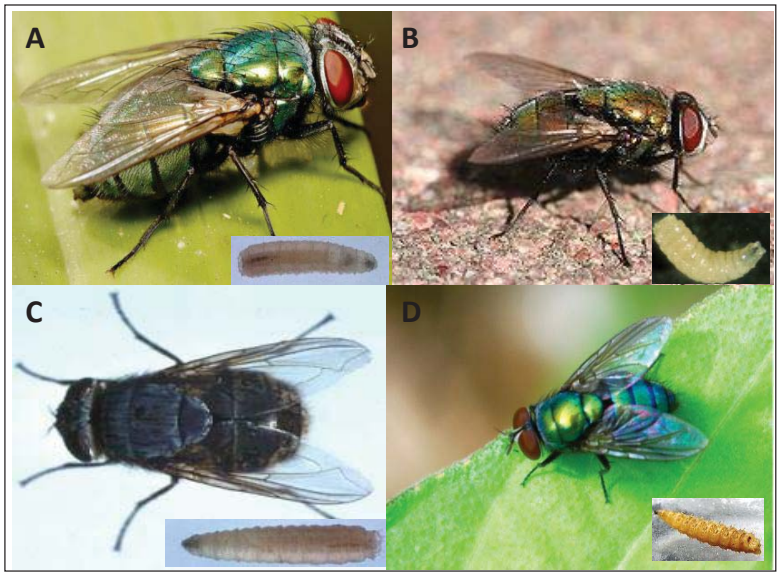


Figure 1.1: Images of the 4 main fly species with larvae inset (not to scale): *L. cuprina* (A), *L. sericata* (B), *C. stygia* (C) and *C. rufifacies* (D). The two *Lucilia* species are separated by the green colouration of the fore femora of *L. cuprina* and the increased number of hairs of the humeral calli and notopleuron area of *L. sericata* (Lang et al., 2001).

1.2.2 Initiation and progression of the flystrike lesion

The fly uses olfactory cues to determine where they are to lay their eggs. Olfactory cues include predisposing bacterial disease, faecal or urine contamination, or pre-existing strike. These cues indicate high humidity and food for the development of the eggs and larvae. The breech region, between the legs to up over the tail, is where the majority of strikes occur (Table 1.1, Figure 1.2). The abdomen, flanks, shoulders, pizzle, belly, head and feet can also be struck to a lesser extent.

Table 1.1: Distribution of strikes, in 3 separate studies from the United Kingdom (UK), New Zealand (NZ), and Australia (AUS).

Study	Breech %	Body %	Head %	Belly %	Foot %
UK (French et al., 1995)	67	14.7	1.8		10.3
NZ (Heath and Bishop, 1995)	81.2	9.3	7.1	2.5	
Aus (Greeff and Karlsson, 2009)	65.4	17.5	17.1		

Flystrike lesions are initiated by the first instar larvae. *L. cuprina* larvae have a group of spines (5µm long) in the anterior end of the oral cavity (Sandeman et al., 1987). These spines protrude beyond the cavity and are thought to inflamme and

damage the skin surface, within 8 hrs of the strike. The first instar larvae release highly active proteases, other enzymes, and possibly toxic substrates which affect coagulation and fibrinolytic pathways (Sandeman et al., 1987). During the first 8 hrs, parts of the basal layer of the skin are exposed due to the removal of the epidermal layer. The second and third instar larvae have full mouthparts that include 2 mouthhooks which project beyond the oral cavity. At 24 hrs after the strike, there is a notable increase in damage to the skin, the epidermis layer is all but gone and the areas of basal layer are damaged, to expose the dermis. By 48 hrs after initiation of the strike, the dermis is completely exposed. By 72 hrs, the whole skin structure including wool follicles are damaged and larvae can be seen to burrow into the dermal tissue (Sandeman et al., 1987).

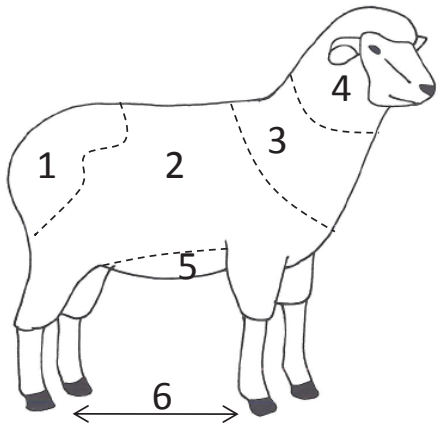


Figure 1.2: Classification of sheep body regions in strike identification. 1: Breech, 2: Body, 3: Shoulders, 4: Head, 5: Belly (and Pizzle in males), 6: Foot. Modified from French et al. (1995).

1.2.3 Lifecycle

There are a number of species in New Zealand which are involved in flystrike, however those of the family Calliphoridae especially *L. cuprina* and *L. sericata*, are the main initiators. The lifecycle of *L. cuprina* and *L. sericata* takes approximately 15 days (Figure 1.3). On day 1, eggs are laid on the site of preference. Female flies use olfactory cues and taste sensors on their feet, mouthparts and ovipositor to seek out a place to lay their eggs, such as a susceptible sheep, fresh carcass, or other source of quality food for

the growing larvae (Anon, 2005). Eggs are laid in batches of up to 200, with adult female flies laying 2 to 3 egg batches during their life, which is about a month.

Eggs take 8 hrs to 3 days to hatch (Anon, 2005; Tellam and Bowles, 1997), when conditions are conducive. Larvae then feed for 2 to 5 days on the skin layer, then drop off the sheep and pupate in the soil. Larvae go through 3 instars before they pupate.

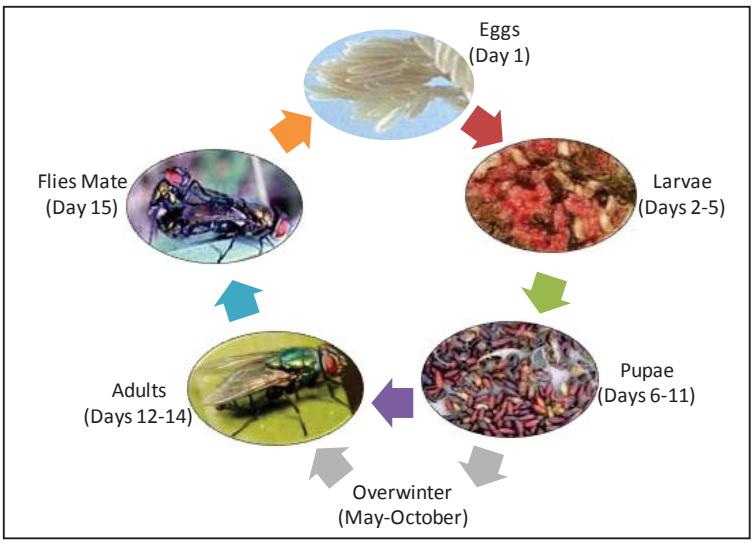


Figure 1.3: Lifecycle of the fly.

Within the soil, the larva skin forms a hard puparium (Barritt and Birt, 1971), within this it takes approximately 6 days for the larvae to metamorphose into a fly. After emergence, the adult fly takes several days to mature. During this time the female flies require a protein meal to develop mature eggs.

Pupa development will arrest and enter an overwintering phase if soil temperatures drop below 13°C (Tellam and Bowles, 1997). In late spring, as temperatures warm up, there is a synchronized emergence of adult flies, generating the first fly prevalence peak. For a comprehensive account of the biochemical and morphological events occurring during the lifecycle of *L. cuprina*, see Barritt & Birt (1971).

All stages of the lifecycle are dependent on at least one of the following; temperature, rainfall, relative humidity (RH), fleece type and faecal soiling (Wall et al., 2000). Humidity is a major factor involved in the survival of eggs and emerging larvae. Optimal RH for egg hatching is around 80% (Davies, 1948). Below 50% RH, *L. sericata* egg survival is limited, and no hatching of eggs occur (Wall et al., 2001).

Newly-hatched larvae will only survive for 1 hr at RH levels of 50% (Wall et al., 2001), at least 7 hrs of relatively high humidity of 60-70% is required. Once larvae reach their second instar, humidity no longer plays a major role in larvae survival.

Temperature is a factor during the pupating or overwintering stage of the lifecycle. Larvae that are to pupate during summer burrow further (~6 cm) to prevent desiccation, than those that overwinter (1-2 cm) (Wardhaugh, 2001). Larvae will not pupate or emerge as adults at temperatures at or below 10°C (Wall et al., 2001). Larvae will instead enter an overwintering stage where they hibernate within the soil as diapausing larvae (Wall et al., 2000). Optimal conditions for pupation are above 15°C, with the upper limit at around 35°C (Wall et al., 2001). It can take 6 days at 30°C or 25 days at 15°C from pupation for adult flies to emerge from pupae (Tellam and Bowles, 1997).

1.2.4 Prevalence and costs

1.2.4.1 New Zealand:

Flystrike was reported as a problem to New Zealand production systems in the 1980s (Heath, 1994), however, it has been present since sheep farming was introduced to New Zealand. There are 4 main fly species which regularly strike sheep in New Zealand; *L. sericata*, *C. stygia, C rufifacies*, and *L. cuprina*. The fly *L. cuprina* was first detected in New Zealand in 1988, though the introduction of this species probably occurred 10 years previously (Heath and Bishop, 1995). This fly has since spread from Northland as far south as Gore (Heath and Bishop, 2006). The introduction of this Australian insect has been linked with a significant increase in flystrike incidence and an extended flystrike season.

Flystrike can occur anytime during the year; however, the main 'season' occurs from October to April, peaking during February and March (Heath and Bishop, 2006). In New Zealand, both *L. cuprina* and *C. stygia* can strike sheep all year round. While *L. sericata* strikes from December to March, *C. rufifacies* strikes between January and March (Heath and Bishop, 1995).

Flystrike is estimated to cost the New Zealand industry $40 to $50 million annually (Beef and Lamb New Zealand, 1999). The majority of this cost is due to production losses through reduced wool growth and yield, reduced live weight, and death as well as control strategies including insecticides and crutching. Other costs involved are labour, pelt damage, and opportunity costs. It has been noted that 1.3% of the pickled lamb

pelts have flystrike and that flystrike faults peaked during July-October (Cooper, 2011). This indicates that a proportion of flystruck lambs survive and that there is a lag between being flystruck and when lambs have recovered to slaughter weight. It is also estimated that the prevalence rate of flystrike is 3-5% of the national flock (Heath and Bishop, 1995). However, it is suggested that this only takes into account overt strikes, and that if covert strikes were accounted for, the prevalence rate could double (Heath and Bishop, 1995).

Recently, an analysis has been conducted evaluating traits which have potential economic benefit to New Zealand farmers if genetic improvement was feasible (Ludemann et al., 2010). Among other traits; flystrike, dag score, and breech bareness were investigated. The financial worth of a 10% improvement in the trait was calculated. Flystrike costs could decrease by $4.5 million/year if a 10% improvement could be made through genetics (Ludemann et al., 2010). Flystrike was also assumed to be reduced indirectly through improvement of dag score and breech and belly bareness. A 10% reduction in faecal soiling (dagginess) was estimated to be worth $21million/year, a portion of which would be due to a reduction in flystrike. While a 10% improvement in breech and belly bareness, was worth $20 million/year, and would also result in a reduction of faecal soiling and flystrike.

1.2.4.2 Australia:

Australia has a different climate to New Zealand with dryer, hotter environments that suit Merinos well. Merinos, however, are susceptible to flystrike and the mulesing operation introduced by Mr JHW Mules in 1931 (Phillips, 2009) has become one of the major strategies used to reduce flystrike. Flystrike was reported as a problem to the Australian sheep industry in the late 1880s, and this was correlated with the increased use of Vermont Merinos with their characteristic body wrinkles (skin folds) (Wardhaugh et al., 2001). The Australian blow fly (*L. cuprina*) is the primary cause of flystrike (Watts et al., 1979) and it was introduced into Australia in the early twentieth century (Sneddon and Rollin, 2010; Watts et al., 1979) from South Africa (Wardhaugh, 2001).

Flystrike costs the Australian industry $161 million, made up of AUS$130 million in control measures and AUS$31 million due to production losses. Control cost is largely due to mulesing, insecticides and mustering (McLeod, 1995). A more recent figure has put the cost up to AUS$280 million/year (Hogan, 2010).

Flystrike incidence differs by region depending on climate and pasture conditions. Overall, it is estimated that 3 million sheep die from flystrike annually in Australia (Wardhaugh, 2001), and a single flock may have up to 10% flystrike mortality (Tellam and Bowles, 1997).

1.2.4.3 United Kingdom:

The United Kingdom is the closest comparison country to New Zealand with regard to flystrike season and climate. The United Kingdom, however, has a different farming system; farming indoors over the coldest months, while New Zealand has an all-year round outdoor farming system. *L. sericata* is the main fly species involved in flystrike in the United Kingdom.

A 1989 survey in the United Kingdom estimated a 1.6% yearly prevalence for flystrike in Wales and England, with approximately 12,000 resulting deaths (French et al., 1992). A latter survey reported an average 1.4% of ewes and 2.8% of lambs struck in 2003 (Bisdorff et al., 2006). They also reported a regional difference with regard to the proportion of farms reporting at least one incidence of flystrike, being highest in south-west England and lowest in Scotland.

There is limited information on the economic cost of flystrike to the industry in the United Kingdom. One model estimates the direct costs of blow-fly strike as between £1 to 4 million/year (Bennett, 2003). This is likely to be an underestimate when compared to the figures quoted previously for New Zealand and Australia above.

1.2.4.4 Non financial penalties: trade barriers

Meat processing companies have to conform to international consumer standards. Policies concerning how animals are treated prior to processing are becoming more stringent. In New Zealand, 'farm assurance' programs are promoted by the various processing companies, to ensure high animal welfare, food safety and traceability standards are followed by farmers (AFFCO, 2010; Asure Quality, 2010).

Consumer view points are also shifting; they are becoming more aware of what is involved in the procurement of their food. Global warming has also played a part; consumers eat local food as concern rises over 'carbon footprints' and are turning away from meat due to methane emissions, and chemical usage. Consumer perceptions on

the ethical treatment of animals are now a major factor in export markets, for example the practice of mulesing in Merinos.

Flystrike in New Zealand is a cost to the farming production system, due to prevention, treatment, death and ethics. There is potential to reduce the incidence and thus costs associated with flystrike using genetic improvement. The current and potential ways for reducing flystrike will be discussed in the following sections.

1.3 Current non-genetic preventative and post-strike treatments

The following section reviews the current non-genetic preventative and post-strike treatments, outlining the effectiveness and current pitfalls of each practice. Practices under discussion include; insecticides, drenching, shearing, dagging, crutching, tail docking, fly traps, mulesing, vaccines, and other farm practices.

1.3.1 Insecticides

Insecticides are the major control method used against flystrike. They are usually applied as a preventative, or in response to the appearance of strike on animals within a flock. Organochlorine insecticides such as dichlorodiphenyltrichloroethane (DDT) and dieldrin were used extensively between 1948 and 1954. Elevated residues in meat and wool forced their withdrawal (Levot, 1995). There are now 3 main chemical families available; organophosphates (OPs), synthetic pyrethroids (SPs), and insect growth regulators (IGRs).

These insecticides can be split into 2 groups; those that act on the nervous system (OPs, SPs) and those that affect the hormone/endocrine system (IGRs). The SPs mimic the naturally occurring chemical pyrethrin that is found in the plant family Compositae. Their mode of action is against the nerve fibre via the voltage-gated sodium channel, causing constant nerve stimulation (Valles and Koehler, 1997). The OPs produce continuous nerve stimulation by binding acetylcholinesterase, an enzyme present in the junction between nerve endings (the synapse) (Valles and Koehler, 1997).

The IGRs are often species specific, and only work during a small window of time in the insects life cycle (Casida and Quistad, 1998). The IGRs, by mimicking juvenile hormone, affect the moulting process of the juvenile insect by keeping the larvae in the immature state (Valles and Koehler, 1997).

1.3.1.1 Insecticide resistance

There are now numerous insecticides available which are becoming less harmful to humans and require shorter withholding periods. However, the improved insecticides have done little to curb the incidence rates of flystrike. This is mainly due to the incorrect application of insecticide or the increase in insecticide resistance (Levot, 1995).

Organochlorides were withdrawn in 1958, but during their use, increased resistance to dieldrin was detected, reaching 70% in 1958 (Levot, 1993). Dieldrin resistance in *L. cuprina* is conferred by an amino acid substitution (Ser to Ala) in the *Rdl* gene on chromosome 4 (McKenzie and Batterham, 1998).

Resistance of *L. cuprina* to OPs is widespread in Australia and New Zealand (Levot, 1993; Wilson and Heath, 1994). Resistance to diazinon was detected in 1965 (Levot, 1993), and by 1970, 95% of flies caught were resistant. Resistance to diazinon and malathion (another OP) is through 2 separate amino acid substitutions in the *Rop-1* gene on chromosome 4 of *L. cuprina* (Campbell et al., 1998). Resistance to OPs has reduced the protection period from 12 weeks to 4-6 weeks, if application is thorough (Levot, 1993, 1995).

In 1979, the IGR cyromazine was registered, and provides 8-10 weeks protection. There is little resistance reported to date for this chemical (Batterham et al., 2006; Tellam and Bowles, 1997). Laboratory mutagenesis studies and work on IGR resistant flies have predicted likely loci on chromosome 4 and 5 for conferred low-level resistance (Batterham et al., 2006).

1.3.1.2 Application

There are numerous ways to apply insecticides; hand jetting, pour on, automatic jetting races, shower dips and plunge dips. The aim is to saturate the skin especially the most at risk areas; the back and breech. Mistakes in application occur when the the wrong dose is applied or not equally treating the whole animal. Timing is also important, application should occur so as to minimise the high risk period and reduce the build up of fly numbers (Anon, 2005). A timely reminder is when the OPs replaced the organochlorides, there was slow acceptance of this class as they were thought not to work effectively. However, this was due to not changing application technique to suit the new chemical composition. OPs do not move through the fleece with the wool

grease like the organochlorides (Levot, 1995). This resulted in the rapid increase in insect resistance towards this group of compounds.

1.3.1.3 Chemical residues

There are consumer concerns about the impact of insecticides on the environment, wool and meat products. The impact on the environment is through the discarding of effluent either from treating sheep (dipwash), or from scouring wool. Increased residues remain if application is on long wool, close to next shearing, or there are repeated applications (Horton et al., 1997; Plant et al., 1999). The withholding period for wool harvesting is recommended for sheep treated with long wool at: 60 days for crossbreds, 100 days for mid-micron, and 180 days for fine wool breeds (Anon, 2005). In New Zealand, meat withholding periods are reviewed by the New Zealand Food Safety Authority, and product labels state their withholding period (Anon, 2005).

1.3.1.4 Summary

Insecticides are the major control method used against flystrike. However, due to incorrect usage, resistance to chemical families is increasing, and is causing ineffective protection. Consumer concerns are also growing about the use of chemicals on the meat and wool products. A move away from reliance on chemicals is needed.

1.3.2 Drenching

Dagginess is often assumed to indicate a high worm burden by farmers and drenching is a common practice performed in response to increased dagginess. Anecdotal evidence suggests that drenching dries up dags and reduces the build up of new ‘wet’ dags. A trial of an ivermectin controlled-released capsule in ewes and weaner lambs showed that dag scores were significantly lower than in untreated sheep (Gogolewski et al., 1997b; Gogolewski et al., 1997a). Drenching can be an effective way of reducing dag scores for the duration of the drenches effectiveness against internal parasites. However, worm burden is not the only factor and in some cases is not a factor contributing towards the presence of dags. Other environmental, physiological, and genetic factors confound the problem; this will be outlined in a later section (1.4.2.1).

1.3.3 Shearing, crutching and dagging

Shearing is a procedure to harvest wool, while crutching and dagging prevent or remove faecal accumulation from around the breech area. These procedures, after insecticides, have historically been the most commonly used for flystrike prevention. They are also used to improve wool income.

Shearing, dagging, and crutching can provide short term protection against flystrike for 4 or more weeks, with no need to chemically treat during this period (Anon, 2005). The primary reason is that the removal of wool allows the drying out of any eggs or larvae deposited on the skin. These procedures also reduce the adherence of faecal material, a prime fly attractant. In addition, if chemical treatment is required then shorter wool allows efficient saturation of the skin with the insecticide of choice.

French et al. (1995) have investigated flystrike incidence rates in ewes and lambs, following different shearing protocols in England and Wales. There was a pattern where flystrike was highest in the ewes prior to shearing and again during late season when wool growth has made them susceptible again. Once ewes are shorn, there was a consequential increase of flystrike in lambs. Shearing was predicted to give a month's protection against flystrike (Wardhaugh and Morton, 1990). In New Zealand, ewes are usually shorn pre-Christmas and lambs within a month of weaning. This reduces the peak of the flystrike season, and dagging and crutching help minimise flystrike risk through the rest of the season. Flystrike risk in New Zealand can be present year round, especially during a mild winter (Heath, 1994). A second shear in winter may be prudent. A second shear has become a common practice on some New Zealand farms; however, this is more to do with generating a biannual wool income than a flystrike control method.

Shearing, crutching and dagging are general farm tasks that need to be carried out for the welfare of the animal. However, the costs of removing wool by any of the 3 techniques are rising. For the 2008/09 year it was estimated that New Zealand farmers spent $4.08 per animal to shear, crutch and dag (Beef and Lamb New Zealand, 2009). This equated to 10% of the gross margin of the sheep enterprise. However, the income generated from wool is decreasing. Between 2007/08 and 2008/09 the fine, medium and strong wool prices decreased by 18.6, 13.7 and 2.3% respectively (Beef and Lamb New Zealand, 2009).

In conclusion, shearing, crutching and dagging are very useful flystrike prevention methods. They primarily work by altering the local environment where eggs are deposited so it is unfavourable for their development, but these also reduce the likelihood of their deposition. However, this comes at a significant cost to the sheep farming enterprise and also in management flexibility.

1.3.4 Tail length

Tail docking is a standard husbandry practice carried out on young lambs. Its purpose is to prevent urine staining and faecal accumulation around the breech, thus reducing flystrike risk. This practice also allows for easier crutching and shearing, two other major sheep management practices.

Tails should be docked below the third palpable joint so that the tail covers the vulva in females and the anus in males (Munroe and Evans, 2009). Docking longer makes the sheep unable to lift the tail as high, leading to urine staining, faecal accumulation, and increase flystrike risk (Fisher et al., 2004; French et al., 1994). Longer tails also prove difficult to shear and crutch.

Ethical concerns arise around the pain docking may cause animals. There is little evidence to suggest that docking has an impact on long term live weight gain (French et al., 1994). However, studies have looked at breeding sheep with shorter tails. Two New Zealand breeding trials estimated heritabilities for tail length at 0.82 and 0.73 (Scobie and O'Connell, 2002; Scobie et al., 2007). Combined with the use of thin tail sheep and increasing the area of bare skin under the tail (heritability estimate of 0.59), this may become a worthwhile alternative.

In conclusion, tail docking is a necessary management practice, to reduce flystrike susceptibility. It does however come with an ethical cost, which may be resolved by genetic selection.

1.3.5 Mulesing

Mulesing is the removal of wrinkly skin and wool from around the breech of Australian Merinos. Though mainly an Australian practice, and a nearly extinct practice in New Zealand, there is a small proportion of mulesing performed on Merinos in New Zealand. It increases the bare patch around the breech, reducing skin wrinkles, faecal accumulation and urine stain, thus reducing the risk of flystrike.

The mulesing operation was developed by Mr JHW Mules in 1931 (Phillips, 2009) in an attempt to control flystrike. The procedure has evolved over time and it is now recommended not to cut the skin in a strip from under the tail to the udder. Strips are removed from either side of the vulva from tail to part way down the hind leg, and across the top of the tail (Lee and Fisher, 2007). The practice is performed on lambs of 2 to 12 weeks of age (Lee and Fisher, 2007).

Mulesing generates some benefits to husbandry farm practices. The presence of the enlarged bare area devoid of wool, reduces the extent of urine staining and faecal accumulation in the breech region. A study on Merino lambs in Tasmania showed that unmulesed sheep had a dag score 0.74 units higher than mulesed sheep at 15 weeks of age (Horton and Iles, 2007). There is a reduction in flystrike incidence rates in mulesed sheep compared to unmulesed sheep (Greeff and Karlsson, 2009).

In 2004, the People for the Ethical Treatment of Animals (PETA) advocated for a termination of the mulesing operation in Australia. In 2005, an agreement was formed to terminate the practice of mulesing by December 2010 in Australia (Sneddon and Rollin, 2010). Concerns were raised over the stress and cruelty of the treatment to sheep. Mulesed animals tend to exhibit abnormal posture (arched backs) and movement immediately following the operation. They also tend to spend a lot more time standing and less time lying and grazing compared to un-mulsed lambs (Fell and Shutt, 1989). Stress responses to mulesing have been studied, however blood cortisol responses are not dissimilar to other farm practices; drafting, drenching, dipping, crutching and shearing (60-70ng/ml) (Fell and Shutt, 1989; Hargreaves and Hutson, 1990). While the stress response to breech strike in Australian Merinos showed peak levels of blood cortisol of 90ng/ml compared to a base line of 20ng/ml (Colditz et al., 2005).

Mulesing has been shown to be an effective way of reducing flystrike in Australian Merinos. Due to the decreased risk of flystrike, there is less need for flock monitoring and preventative and treatment measures i.e. insecticides, and dagging. A large incentive will be needed to reduce the use of this ethically controversial procedure. The resulting agreement to terminate mulesing by December 2010 due to the PETA campaign was a great motivator for alternative methods to be identified. However, this agreement has been abandoned, almost certainly to the detriment of the Australian wool industry. Research is still needed as in the future the ethical treatment debate will arise again.

1.3.6 Blowfly traps

Blowfly traps are used to attract flies away from stock, reduce the number of flies and thus the incidence of flystrike in a particular area. They are also useful in monitoring the fly population over a season to better time preventative treatments for flystrike.

There are numerous blowfly trap models available, either with artificial fly attractants or those that require baiting, often with meat. In trials, synthetic and baited traps have been shown to decrease the incidence of flystrike in ewes and lambs (Broughan and Wall, 2006; Wall et al., 2001).

Based on these studies, using flytraps alone will not reduce the incidence of flystrike as effectively as using insecticides. However, used together with other preventative treatments, traps can act as an efficient monitoring and control method in paddocks or locations which are flystrike prone due to the lie of the land. Unfortunately the downfall of these traps is maintenance. While traps with synthetic lures may not require much maintenance, traps that require baiting, need to be emptied and re-baited regularly.

1.3.7 Other farm practices

Other farm practices that can help in the proactive deterrence of flystrike include the correct disposal of carcasses and considering topography of paddocks when moving stock**.** Carcasses act as a reservoir for flies, and topography can affect the number of flies present in a particular area.

Offal pits are a requirement on New Zealand farms as part of meat processing companies 'farm assurance' schemes (Asure Quality, 2010). However, often carcasses are not noticed, left in paddocks, or thrown down the nearest convenient hole, gully, or scrub bank. This disregard for the correct disposal of carcasses creates a perfect breeding ground for flies. In Tasmania, an average 3,096 flies emerged from 5 sheep carcasses (Lang et al., 2001). Of these, 95% were *C. rufifacies*, and only 4% were *L.* cuprina. A New Zealand study used thawed mice, magpies, possums, and blackbirds. The proportion of *L. cuprina*, *L. sericata*, *C. stygia*, and *C rufifacies* emerging from carcasses were 2.50, 10.10, 33.05, and 7.30% respectively (Heath and Appleton, 2000). *L. cuprina* and *L. sericata*, the main primary egg depositors on live animals, only represented a small proportion of flies emerging from carcasses, due to these species

having a low interspecies competition threshold with secondary species such as *C. rufifacies* (Lang et al., 2001).

Grazing a concentrated mob of sheep in one paddock for a long time is not recommended. Grazing sheep in this manner intensifies the attractants flies seek when laying their eggs (Heath, 1994). It is recommended that where topography allows, sheep are grazed on high ground exposed to strong winds (Heath, 1994). High winds are a deterrent to flies, and fly activity drops as wind speed increases. A review of studies examining distribution and movement of flies in Australia showed flies often clumped together in open pasture, and they tended to follow sheep movements (Wardhaugh, 2001). It has been observed that where farms back on to, or paddocks contain; native scrub, bush gullies, shelterbelts, or forestry, there is an increase in fly populations and flystrike incidence. These areas provide a warm sheltered environment, conducive to fly breeding (R. Barton, Pers. Comm.). Another practice to reduce pasture worm burdens is to rotationally graze pastures with cattle between sheep mobs. The reduction in worm burden helps reduce the risk of sheep developing dags in association with high worm burdens, thus getting rid of one of the fly attractants.

Disposing of carcasses to reduce the reservoirs for flies and the strategic use of paddocks via considering topography and cattle movement are non-chemical ways of reducing flystrike risk. However, this is hard to implement when feed is short, carcasses are not detected, and topography does not allow sufficient paddocks of low fly risk. In these cases, other prevention methods are required.

1.3.8 Vaccines

There have been several studies on developing vaccines against blowfly strike. The level of success has been variable, and even in the field, natural immunity is sparsely reported (Elkington and Mahony, 2007). There is currently no vaccine commercially available.

The lack of success in vaccine development is due to 3 reasons; firstly, during the progression of the strike infection, larvae spend relatively little time on the sheep. This provides a very short space of time for the sheep immune system to provide any protection against the invader (Tellam and Bowles, 1997). Secondly, the larvae's own defence system against the ovine immune system involves a tough cuticle lining the external surface, foregut and hindgut (Tellam and Bowles, 1997). Finally, the excretory-

secretory products of larvae have an immunosuppressive effect on the hosts T cell response and antibody production (Elkington and Mahony, 2007).

Trials involving IgG antibodies showed promise *in vitro*, but failed *in vivo* (Elkington and Mahony, 2007). Cellular immune response to larval antigens showed more promise, inducing hypersensitivity reactions to first stage larvae. Older sheep tended to show some build up of resistance to blowfly strike (Elkington and Mahony, 2007). Experimental repetitive larval infection of sheep showed a similar trend. Immunity has also been shown to be heritable, selective breeding has reduced natural blowfly strike to 1% in resistant lines and increased it to 19% in the susceptible line (Raadsma, 1987).

In conclusion, vaccination could be another worthwhile tool in the prevention of flystrike. However, it will have to produce similar or better results than products already available. Protection time would have to be as long as that awarded by effective insecticides. It would also have to be economically viable.

1.3.9 Summary

There are a number of preventative and treatment options available to reduce flystrike incidence. Even with the use of these options there is still an unsatisfactory number of sheep that suffer from flystrike each year. Some of these treatments, especially insecticides and mulesing, are under increasing public scrutiny for their impact on the animal, environment and consumer end products. Other practices such as shearing, crutching, dagging and tail docking are required husbandry practices unlikely to become obsolete; however, the increasing costs of these procedures warrants research into ways these practices can be minimised. The sequencing of the sheep genome and the vast improvements into genomic technology allows us to exploit these to investigate the genetic control of flystrike and those traits that increase the animal's susceptibility to flystrike.

1.4 Breeding for host resistance

One potential mechanism to reduce flystrike incidence and the costs of associated control and treatment is to breed an animal that is less susceptible to egg deposition and subsequent larvae viability. Historically, this method has been perceived to be slow and potentially conflict with wool production. In addition, the genetics of flystrike resistance

by the host has not been extensively investigated. This is because, in New Zealand, most research funds have been channelled towards reducing internal parasites, the costliest disease problem. In Australia, mulesing has been widely and effectively used and sheep farming has been limited to regions of lower risk. The following section reviews the results from research flocks and examines potential indirect selection approaches. Research flocks have been used because farmers are loath to deliberately expose animals to infection and in some cases, artificial challenges have been used to reduce climatic variability. The current New Zealand industry selection methodologies are then described, followed by an examination of the potential for genome-wide association studies (GWAS) and GS.

1.4.1 Research flock results

Two Australian research flocks have been investigated for the response to selection for resistance and susceptibility to flystrike (Greeff and Karlsson, 2009; McGuirk et al., 1978; Smith et al., 2009). These involve flocks at Trangie (McGuirk et al., 1978), Armidale and Mt Barker (Greeff and Karlsson, 2009; Smith et al., 2009). Other studies have involved short selection trials including crossing more resistant breeds with susceptible breeds (Rathie et al., 1994).

The longest running selection line was the resistant and susceptible selection lines based at Trangie (McGuirk et al., 1978). This experiment was designed to investigate the development of fleece rot and body strike and to determine if resistance can be improved by genetic selection in Merinos. It assumed that fleece rot and flystrike were highly genetically correlated; therefore, selection was based on the indirect predictor fleece rot, both naturally and artificially induced, as the easier trait to record. The animals initially selected for the program came from 15 different Merino strains, with a between flock correlation between body strike and fleece rot of 0.86. After 17 years of selection, mean natural fleece rot incidence was 37% and 10%, and body strike incidence, 8% and 3%, for susceptible and resistant lines respectively. This translates to a divergent rate between lines of 2.8% per annum for natural fleece rot and 0.4% per annum for body strike (Mortimer et al., 1998). The heritability estimate for fleece rot susceptibility was 0.49 ± 0.14 (Raadsma, 1993). Heritability of susceptibility to body strike and of liability to body strike was 0.26 ± 0.12 and 0.53 ± 0.25 respectively (Raadsma, 1991a). Raadsma (1991a) predicted the rate of genetic gain in the reduction of susceptibility using a number of selection criteria. A reduction in body strike from

20% prevalence to less than 1% within 10 years could occur, if the most resistant animals were directly selected from an underlying scale. However, only a 3.9% reduction in body strike prevalence was predicted in the more realistic situation where there is a high variation in flystrike prevalence from year to year.

The second Australian resource, commenced in 2005 and consisted of 3 lines based at Mt Barker and Armidale for the Breeding for Breech Strike Resistance Project (Greeff and Karlsson, 2009; Smith et al., 2009). Mt Barker represents medium wool Merino sheep in a Mediterranean environment, and Armidale represents fine wool Merino sheep in high summer rainfall environment (J Smith, Personal Communication). The 3 lines represent: a plain breech line where both ram and ewe replacements were selected for plain breech (PB); a commercial line where sires were selected on plain breech (CI); and an unselected control flock (UC). By using indicators for breech strike; including breech wrinkles, breech cover (breech bareness), dags, urine stain, and wool colour, the aim was to demonstrate their importance in changes to breech strike incidence. This resource also differs from Trangie, as unmulesed animals were also selected based on breech wrinkle score.

Heritabilities and genetic parameters were calculated for each location separately. For the Mt Barker lines, between selection line incidence of breech strike was not significant before weaning. However, breech strike post weaning to hogget shearing was significantly different ($P<0.01$). Breech strike incidence was 7.7%, 20.7% and 25.6% post weaning for PB, CI and UC lines respectively (Greeff and Karlsson, 2009). Dag score and breech cover were the only 2 indicator traits significantly related to breech strike, and then only in 2008 and 2007, respectively. All indicator traits and breech strike had moderate heritability, but high standard errors. Breech strike on the direct scale had a heritability of 0.57 ± 0.28. Phenotypic correlations between traits were all high, except for a negative correlation between dag score and urine stain. Dag score had the highest genetic correlation with breech strike (0.86 ± 0.17). The incidence of breech strike in different sire groups varied between 2 and 55%.

The experimental selection lines at Armidale again revelled moderate heritabilities for all indicator traits, however, the dag score estimate was only 0.09 ± 0.06, compared to 0.55 ± 0.30 calculated for the Mt Barker experiment (Greeff and Karlsson, 2009; Smith et al., 2009). The heritability of breech strike was 0.32 ± 0.11. Phenotypic correlations with breech strike were low, but dag score and breech wrinkle had the

highest correlations with breech strike, both estimated at 0.22. Breech cover was not related to breech strike as in the Mt Barker lines. Smith et al. (2009) suggested that dag score was not a good indicator trait in high summer rainfall regions as it was 'transient'.

Another way to combat flystrike is to introduce another breed naturally more resistant, such as Wiltshire, which shed their fleece. Rathie et al. (1994) crossed Wiltshire Horn and Merino to examine the effect of shedding on wool production and flystrike incidence. Progeny from crosses were left un-mulsed and un-crutched. Flystrike incidence varied, with lowest in 1/2 Merinos (av 8.45%) compared to 5/8 (av 17.18%) and 3/4 Merinos (av 50.43%). Wool production was least in 1/2 Merinos, but as Merino content increased; wool production increased, and fibre diameter decreased.

The above literature review describes the genetic studies to reduce the incidence of flystrike via selection in Merinos. In addition, heritability and genetic correlations with other traits have been calculated. Fleece rot has been shown as a good indirect indicator for flystrike in the Trangie Merino resource flocks. Fleece rot is not as prevalent in New Zealand dual-purpose breeds as in Australian Merinos, so it would not be a useful indicator. Similar comments apply for breech wrinkles. However, dagginess and breech bareness are both correlated with flystrike and most likely useful for selection in New Zealand dual-purpose breeds. The advantages and potential traits for indirect selection will be discussed below.

1.4.2 Potential for indirect selection

A correlated response occurs when trait X changes due to selection on trait Y. When a change in X is required, it is sometimes more profitable to select for Y than to directly select for X. This is called indirect selection. Indirect selection has an advantage when: a) the 2 traits are highly genetically correlated and the heritability of trait Y is higher than trait X, this leads to a correlated response higher than that achieved by direct selection; b) trait Y is cheaper to measure than trait X, and; c) trait Y is more accurate and easier to measure than trait X, including at a younger age and in both sexes (Falconer, 1989; Turner and Young, 1969). The relative efficacy can be calculated by $Q = r_G.h_Y/h_X$, where r_G is the genetic correlation between the 2 traits and h_Y and h_X are the square root of the respective heritabilities of the indirect and direct traits (Turner and Young, 1969). When $Q > 1$, then indirect selection is more efficient and relatively greater genetic gains can be achieved.

For obvious reasons, commercial farmers and breeders are loath to expose valuable livestock to serious, or indeed any, flystrike challenge. Nobody wants their sheep to suffer from flystrike. This limits the ability to modify the trait, unless indirect selection predictor traits are available. Flystrike is a difficult trait to select for as it is not a major problem in all years, and the level of flystrike risk also varies with season. Selecting a resistant animal is therefore difficult due to only a small portion of animals fully expressing their true flystrike resistance/susceptibility status. This is exacerbated by preventative treatments such as dipping/jetting, shearing, crutching and mulesing. A variety of indirect predictors has been proposed and several have been supported with experimental evidence in at least some breeds. These include dags, breech bareness, breech wrinkles and fleece type.

1.4.2.1 Dag score

Dag score has potential as an indicator trait for flystrike, and the majority of flystrike in New Zealand occurs in the breech region (Heath and Bishop, 1995). Dagginess is especially relevant for New Zealand breeds which are exposed to lush high moisture content ryegrass pasture containing endophytes. Australian research from the Mount Barker experiment had a positive genetic (0.86 ± 0.17) and phenotypic (0.23) correlation of dagginess with flystrike (Greeff and Karlsson, 2009). Dags are generated from the accumulation and build up of faecal material on the wool surrounding the perianal region. The area of dags can cover down the hind legs and up over the tail in the worst cases (Figure 1.4). There are a number of environmental and biological mechanisms that impact on the expression of dags. These need to be understood before genetic tools can be used to reduce dagginess. Factors that will be discussed include immune responses to internal parasites; condensed tannins, moisture and endophytes in pasture; and intestinal function.

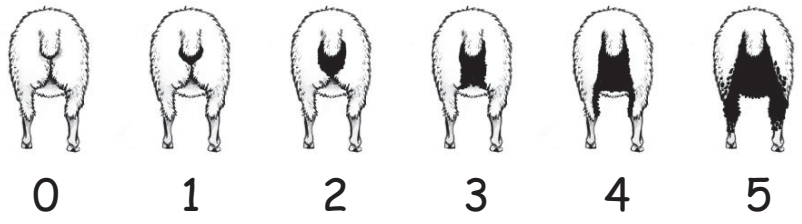


Figure 1.4: Dag scoring system: 0 no dags to 5 most daggy. As proposed by Sheep Improvement Limited (www.sil.co.nz/Files/Tech-Notes/DagScoreV2.aspx).

1.4.2.1.1 Immune response to internal parasites

It is often assumed that dagginess is a result of a high worm burden. Consequently drenching is often timed based on dag score of the mob rather than worm count. However, this theory does not always convert to practice. High dag scores have been associated with both low faecal egg counts (FEC) (Bisset et al., 1997; Morris et al., 1997) and high FEC (McEwan et al., 1992b). The McEwan et al. (1992b) study was performed on lines selected on production traits (number of lambs born, 100 day weight and hogget fleece weight). Between lines: FEC was genetically correlated with dagginess (0.61) and faecal consistency scores (0.54). There was also a positive genetic correlation between dag and faecal consistency scores (0.86). However the relationship between dags and FEC was negative for half-sib genetic correlations (-0.55), indicating low FEC equals high dag score as conferred by Bisset et al. (1997) and Morris et al. (1997). The half-sib heritability estimate for dag score (0.06 ± 0.06) was low compared to other estimates published. Heritability for faecal consistency was moderate 0.27 ± 0.09 and for total FEC was 0.14 ± 0.07. Dags were positively correlated phenotypically (0.46) and genetically (1.37) with faecal consistency. Therefore, in this study, there was an unclear association between dags and FEC.

A study on Australian Merinos calculated heritabilities across 8 recording times from weaning to approximately 400 days old (Pollott et al., 2004). Dag score ranged from 0.07 to 0.32, overall was 0.11 with moderate trait repeatability (0.25). Faecal consistency score had a heritability range from 0.03 to 0.22, overall 0.12 with a repeatability of 0.12. FEC heritabilities ranged from 0.19 to 0.60, overall 0.28, trait repeatability was 0.22. Dag score and faecal consistency score had phenotypic and genetic correlations of 0.33 ± 0.01 and 0.63 ± 0.14, respectively. There were no other significant correlations.

Bisset et al. (1996) went further and noted that the low FEC lines had a higher immunological response to internal parasites, with increases in global leucocytes/mucosal mast cells, connective tissue type mast cells, and eosinophils in the intestinal mucosa. This study did not provide a correlation with dag score; however, it did use animals from the same Wallaceville selection lines used by Bisset et al. (1997) and Morris et al. (1997). In a different study, sheep with severe dags due to the Australian ‘winter flush’ tended to have low FEC counts, high faecal moisture content, and a hypersensitive inflammatory response, characterised by changes in lymphocyte

populations, increased eosinophilia in the gut and small intestine, and mucosal damage (Larsen et al., 1999).

The assumption that the correlation between dags and FEC is due to high worm burden has not been proven. The above literature has shown there is potential for low FECs to increase dag formation. The response of the sheep's immune system to worm burden, independent to the level of burden and FEC, maybe the factor involved in dag formation.

1.4.2.1.2 Pasture

It has been noted that pasture quality, composition and quantity can affect the incidence of dag formation in sheep. Notably when sheep have a 'flush' of grass (i.e. allowed to eat *ab libtum* after a period of little to eat) that there is a marked increase in the proportion of the flock with high dag scores. Several factors about pasture including condensed tannins, fungal endophytes, and pasture composition have been studied in relation to dag formation.

Condensed tannins have been shown to improve animal productivity through reducing; dags, internal parasites and flystrike incidence in sheep (Leathwick and Atkinson, 1995). This has been attributed to the ability of condensed tannins' to bind leaf proteins, creating complexes only released in the abomasum. This increases the amount of amino acids absorbed in the small intestine and thus helps improve animal productivity (Ramirez-Restrepo et al., 2004). A 1994 study by Leathwick and Atkinson (1995) observed lambs grazing endophyte ryegrass had greater dag score, higher average dry weight of dags and higher percentage of lambs struck than those grazing tannin containing *Lotus corniculatus*. In a subsequent study there was an average of 36, 29, 26, 11, and 7% of lambs flystruck when grazing *L. corniculatus* for a period of 0, 4, 7, 10, or 14 continuous days each fortnight respectively (Leathwick and Atkinson, 1998). Mean weight of dags decreased as each grazing treatment period increased; there was a significant difference between grazing for 0 to 4 and for 7 to 14 days. There was also a significant difference between mean dag weight of flystruck and non-flystruck lambs (185 vs 105g respectively, $P<0.01$) (Leathwick and Atkinson, 1998). Ramirez-Restrepo, et al. (2004) also showed reduced dag scores for animals grazing on *L. corniculatus* versus perennial ryegrass/white clover pastures. Condensed tannins also have positive effects on live weight (Leathwick and Atkinson, 1995, 1998; Min et al., 1998; Ramirez-Restrepo et al., 2004). Conflicting results have also been found for the

effect of condensed tannins on wool growth (Leathwick and Atkinson, 1995, 1998; Min et al., 1998; Ramirez-Restrepo et al., 2004).

In New Zealand, the majority of pastures contain perennial ryegrass (Fletcher et al., 1999). There is a mutualistic relationship between asymptomatic fungal endophytes and the ryegrass plant (van Heeswijck and McDonald, 1992), with the presence of endophyte providing the ryegrass with resistance to Argentine Stem Weevil (Schaeffer, 2006). Adverse effects, however, are experienced by herbivores grazing endophyte containing grasses. These include reduced growth rates, ryegrass staggers, diarrhoea, fescue toxicosis, and heat stress (Fletcher et al., 1999; van Heeswijck and McDonald, 1992). High faecal moisture in those animals grazing high endophyte grasses appears to contribute towards diarrhoea and consequently dag formation (Pownall et al., 1993). Pownall et al. (1993) observed that faecal moisture content was 8.6% higher in those animals grazing high endophyte pasture compared to those on nil endophyte, with faecal moisture content only accounting for 15% of the variation in dag score. Fletcher et al. (1999) also observed that animals grazed on endophytpe containing ryegrass developed higher dag scores than those on endophyte free ryegrass. During the summer/autumn grazing trial, animals on high endophyte pasture also had a higher proportion of animals flystruck, 15-30% vs 2% (Fletcher et al., 1999). Endophyte toxins can also affect the body temperature of animals. Flystrike incidence in ryegrass staggers susceptible and resistant flocks was similar (8%), however, the majority of strikes (54%) in the susceptible line occurred on the body (Young et al., 2004). The endophyte toxins can increase body temperature; and it has been shown that increased sweating in the wool increases flystrike risk, due to increased humidity (Davies, 1948).

Spring pastures often contain low levels of neutral detergent fibre (NDF), below the 26% required for optimum rumen function (Davidson et al., 2006). This can lead to softer faeces, and scouring seen during the spring flush. Davidson et al. (2006) supplemented sheep grazing spring pastures with additional fibre. Animals supplemented with fibre had higher worm burdens, higher faecal dry matter percentage, and lower dag scores than control sheep. There was no correlation between worm burden and dags. Only a small proportion of the sheep in this study were examined for flystrike, however, within this group the flystruck sheep had greater median dag scores compared to unstruck sheep (median dag score 4 vs 3). Reid and Cottle (1999), measured mineral and dry matter content of faeces from daggy and non-daggy sheep

from 2 flocks, a high and low FEC selection line and a grazing trial on endophyte or no endophyte pasture. In both experiments, daggy sheep had higher potassium and lower magnesium content. There may be a difference in mineral absorption in animals that tend to accumulate dags, but this needs to be investigated further.

In summary, controlling the composition and species of pasture can help to reduce dags. Pasture that is high in condensed tannins, low in endophytes and with levels of NDF above 26% will provide the highest potential to reduce dags. The potential for certain environments to grow such pasture is limited. There is also conflicting information on the effect condensed tannins have on other major production traits such as live weight and wool growth. These need to be further investigated before pasture replacement can be recommended.

1.4.2.1.3 Intestine

The intestinal tract plays a part in the formation of faeces. In sheep it has been shown that differences in the motility profile of the intestine cause the production of soft, loose, sloppy or hard, pebble size faeces, with higher frequencies of giant contractions in the spiral colon associated with softer faeces (Bedrich and Ehrlein, 2001; Ruckebusch and Fioramonti, 1980). Understanding the mechanism of the large intestine may provide clues as to what genes play a role in this process. These genes could be investigated in GWAS for their effect in reducing the formation of dags.

Cows and sheep have very similar large intestines; however sheep produce faeces containing 40% dry matter, while cattle faeces contain only 20% dry matter (Ruckebusch and Fioramonti, 1980). The sheep large intestine consists of; a large tubular caecum, and a proximal, spiral and distal colon. It has been shown that faecal pellet formation first begins in the spiral colon. Removal of the spiral colon increased the amount of daily faeces (278g to 1042g), while the percentage of faecal dry matter decreased (38 to 12%) (Ruckebusch and Fioramonti, 1980). This study also observed electrical pulses representing contractions along the sheep spiral colon. These pulses are important in the movement, mixing and formation of faeces. One form of these electric pulses are the giant contractions. It was observed that sheep having a higher frequency of giant contractions, produced softer faeces than other sheep (Bedrich and Ehrlein, 2001).

In summary, the spiral colon is the most important part of the large intestine in the formation of faecal material. If disruption to normal processes occurs, it may result in the formation of sloppy loose faeces with the ability to form dags. The ability to select sheep with normal intestinal processes that produce firm pebbles would be beneficial.

1.4.2.1.4 Summary

There are a number or pathways which affect the formation of dags in sheep. Farm management practices can reduce dag formation through, pasture type, drenching and stock movement. Selection for faecal consistency, especially firm pellets, may preferentially select for improved intestinal function. Dagginess is moderately heritable and therefore genetic selection will improve dag score and consequently decrease flystrike susceptibility.

1.4.2.2 Breech bareness

Crutching, dagging and mulesing are all ways of artificially producing a clean bare area in the breech, providing short and long term protection respectively. All of these practices are performed either for, or in part, to reduce the incidence of breech strike. Heath and Bishop (1995) found that 81.2% of flystrike in New Zealand occurred in the breech. Genetic selection for naturally increased bare area around the perineum is one possible alternative to the above practices.

Shedding breeds are naturally bare around the breech and belly, and are less susceptible to flystrike compared to full wool breeds. Reductions in breech strike incidence have been shown in Merinos crossed with Wiltshire Horn sheep in Australia as shown above in section 1.4.1 (Rathie et al., 1994). As the percentage of Merino increased so did flystrike incidence; lowest in 1/2 Merinos (mean 8.45%) compared to 5/8 (mean 17.18%) and 3/4 Merinos (mean 50.43%). This correlated to increasing fleece weight due to less shedding of the wool around the breech, belly and points in crosses with increasing Merino percentage (Rathie et al., 1994). In a comparable New Zealand study, the shedding feral Merino had lower incidence of flystruck lambs compared to the full wool Merinos (0 vs 33%) (Litherland et al., 1992). The shedding feral Merino and Wiltshire also accumulated significantly less dags than the Romney and Merino.

Breech bareness is the area of naturally bare skin around the perineum, and is scored on a 1 to 5 scale (Figure 1.5). Breech bareness is heritable, and current estimates range

from 0.33 to 0.35 (Scobie et al., 2007, 2008). An easy-care sheep resource flock has been established with a breeding goal to produce a polled sheep with no wool on the head, legs, belly, and breech, and with a short tail (Scobie et al., 1999). Using this resource, the genetic parameters for breech bareness, dags, and flystrike incidence have been estimated (Scobie et al., 2007; Scobie et al., 2002). Heritabilities for breech bareness and dag score were estimated at 0.33±0.06 and 0.37±0.07 respectively. Breech bareness had a negative phenotypic and genetic correlation with dag score; -0.17 ±0.02 and -0.3 ±0.13 respectively (Scobie et al., 2007).

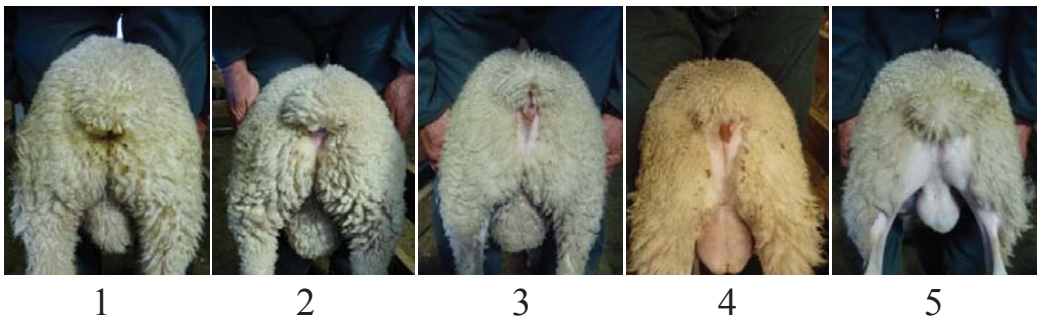


Figure 1.5: Breech bareness scoring system: 1: fleece cover extends to margins of anus, to 5: an extensive bare area either side of the anus. As proposed by Sheep Improvement Limited (www.sil.co.nz/getdoc/a9bb121b-2016-4b34-a163-399ae4b28471/Doc-ID-000010-GW-Bare-Points-Sheep.aspx).

These results have been reported in two New Zealand composite flocks (Scobie et al., 2008). In a Romney based composite, phenotypic and genetic correlations between breech bareness and dag score were -0.18 ± 0.02 and -0.44 ± 0.12, respectively. For a Perendale based composite, the correlations were -0.27 ± 0.03 and -0.72 ± 0.09 respectively. Heritabilities for breech bareness were 0.35 and 0.61, respectively for the Romney and Perendale based flocks. For dag score, heritabilities were 0.31 and 0.34, respectively (Scobie et al., 2008).

Scobie et al. (2002), in their resource flock, showed that as breech bareness increased, the proportion of flystruck lambs decreased. For a breech bareness of 1: 22% lambs were struck post-weaning. This was reduced to 16%, 11% and 0% as bareness increased from score 2 to 4. By breed, no Wiltshires were struck, 6% of Finn x Dorset Down, 21% of Finn x Romney and 12.5% of Feral Merino x Merino lambs were struck.

Mulesing, crutching and dagging are labour intensive practices, to artificially increase the bare patch around the anus. Crutching and dagging are also only temporary and needs to be repeated at least 3 times a year. Mulesing is permanent but ethically questionable. Natural breech bareness has been shown to be moderately heritable, and thus is an alternative tool that can lead to a permanent bare patch in the breech. This will result in reduced need for crutching, dagging, dipping, and drenching to control dags.

1.4.2.3 Wool type

Since domestication, a number of wool breeds have emerged. Wild type fleece was light, and consisted of coarse long fibres forming an outer coat and short fine fibres forming the inner coat (Henderson, 1968). They also tended to shed their coats in summer. From this base, animal selection was made for increased wool growth, around 5 to 10 times more (Henderson, 1968). Secondly, via selection for a uniform fleece, the separation of the inner coat and the outer coat is now no longer identifiable (Henderson, 1968). Thirdly, via selection for the diameter and number of fibres (Henderson, 1968), some breeds have been selected for their fineness (Merinos), others for medium and strong wool for use in carpets and other apparel (Romney, Drysdales). There are still breeds that more closely resemble their ancestors in fleece type, including the Wiltshire and Dorset Horn. However, the selection away from the wild-type fleece has introduced a number of problems including, wool staining, dermatophilosis, fleece rot, and flystrike.

The fleece has its own microclimate, with its own relative humidity and temperature (Henderson, 1968). The architecture and structure of the fleece affect wetting and drying processes of the fleece, and the nature of the microclimate. Wetting of the fleece and presence of water droplets contained within the spaces between staples increases the relative humidity of the microclimate (Henderson, 1968). In a fleece with compact distinct staples, the fleece can dry quickly. However, a fleece that has irregular fibre size and crimp and no defined staples may resemble a confused mass of fibres or become cotted (Henderson, 1968). This type of fleece takes longer to dry out; thus causing wool stains. Another feature of the fleece that is important is termed wool ‘yolk’ (Henderson, 1968). This substance is made up of the wool wax, that when purified is known as lanoline, and suint which consists of mainly dried sweat. The wax helps provide a waterproofing effect on the fleece (Hayman, 1953), while the suint proportion contains nitrogenous material and potassium salts, that influence the growth

of bacteria (Henderson, 1968). It is the combined effect of all these fleece features: microclimate, architecture and structure that influence the growth of microorganisms that lead to conditions such as fleece rot, dermatophilosis and flystrike.

1.4.2.3.1 Fleece rot

Fleece rot is a consequence of prolonged wetting of the fleece, leading to the breakdown of the wax layer on the skin (McGuirk et al., 1978). The skin, sensitive to moisture, exudes serous material which forms horizontal bands within the fleece, yellow in colour and containing hard brittle material (Henderson, 1968; McGuirk et al., 1978). Bacterial growth within the serous material contributes to the colourful staining associated with fleece rot. The bacterium *Pseudomonas aeruginosa* is the most common organism associated with fleece rot (Norris et al., 2008). Other pseudomonads have also been implicated (Australian Wool Innovation and Meat Livestock Australia, 2007; Norris et al., 2008). It is unusual for fibres to become rotten (Henderson, 1968), but if there is a failure of keratinisation in the wool follicle, breaks can occur (McGuirk et al., 1978).

Between family differences in susceptibility to fleece rot have been observed in Merino strains (Dunlop and Hayman, 1958; Hayman, 1953). Heritability estimates for fleece rot range between 0.14 and 0.33 (Atkins, 1979; Li et al., 1999; McGuirk and Atkins, 1984; Mortimer et al., 2009). Correlated responses with production traits include; negative phenotypic and genetic correlation with mean fibre diameter (Li et al., 1999; Mortimer et al., 2009), low phenotypic and a low to moderate genetic correlations with greasy fleece weight and clean fleece weight (Li et al., 1999; Mortimer et al., 2009). Wool colour has high genetic (0.47) and moderate phenotypic (0.18) correlations with fleece rot incidence (Li et al., 1999). All these results imply reduced fleece rot is associated with decreased fleece weight, increased fibre diameter, and brighter whiter wool.

The Trangie research flock was the largest resource focusing on fleece rot and body strike as outlined previously (McGuirk et al., 1978). After a decade, a subset of the animals were used to estimate susceptibility to body strike between 1984 and 1986 (Raadsma, 1991b). Prevalence rates of fleece rot and body strike in each line were 75.8% and 11.3%, respectively, for resistant and 83.3% and 20%, respectively, for susceptible lines. Body strike was only observed in animals with fleece rot (Raadsma,

1991b). After 17 years of selection, the lines had diverged at a rate of 2.8% and 0.4% per annum for fleece rot and body strike respectively (Mortimer et al., 1998).

Fleece rot is correlated to body strike, especially in Australian Merinos, and significant gains can be made in both traits when fleece rot is incorporated into a selection program. In New Zealand, fleece rot is not as prevalent, thus selection for fleece rot will not produce the same effect on body strike. However, the processes that lead to fleece rot, humidity and bacterial growth, are implicated in the attraction of flies to the fleece.

1.4.2.3.2 Dermatophilosis

Mycotic dermatitis or dermatophilosis as it is now known is also caused by prolonged wetting. It involves the zoospores of a fungi *Dermatophilus dermatonomous* (Henderson, 1968; Norris et al., 2008). The disease starts after transmission of the zoospores, free-moving fragments of the mycelium, to the skin (Henderson, 1968). At the skin surface the zoospores develop into the mycelia form. At this stage the skin is tender, red and has a slight swelling. A dome like scab forms as the skin hardens and dries. Beneath the scab the mycelia thrive in the low oxygen environment and exudate is secreted from beneath the scab (Henderson, 1968). The exudate binds the fibres together in pencil like clumps perpendicular to the skin.

The wax content of the fleece with its waterproofing ability is the main barrier to this disease (Roberts, 1963). Removal of the wax layer with petroleum allowed the initiation of dermatophilosis by the zoospores. Outbreaks of the disease also occurred in Merinos soon after birth or shearing, when wax levels are low (Roberts, 1963).

There has been little research done into the genetic component of this disease. Raadsma et al. (1992) investigated induced dermatophilosis in Merinos. Preliminary analysis estimated heritability for scab severity at high dose of 0.25 and heritability at a low dose of 0.42. This study also noted the results of previous work, heritability of liability to induced dermatophilosis of 0.14 (Raadsma et al., 1992). Surveys in Western and Southern Australia observed an association between body strike and dermatophilosis (Gherardi et al., 1983). Laboratory results found *L. cuprina* were attracted to, and deposited eggs on, dermatophilosis lesions. These results were repeated in a field experiment where more strikes occurred on animals affected by dermatophilosis than by fleece rot or on unaffected controls (Gherardi et al., 1983).

1.4.2.3.3 Summary

As outlined above, fleece rot and dermatophilosis, diseases of the fleece, are associated with flystrike. The odours produced by the leaking exudates, and the organisms involved in the two diseases attract the flies to deposit their eggs. However, flystrike can develop in the absence of these two diseases.

The main factor influencing flystrike is humidity. A period of high relative humidity is required for the successful hatching and survival of first stage larvae (Davies, 1948; Wall et al., 2001). Humidity builds up as a combination of the presence of moisture at the skin level and warm temperatures. This can be by rain, water, urine stain, faecal adhesion, or through sweat accumulation. Moisture entering the fleece can become trapped if body conformation prevents drainage down the sides of the animal, for example 'devils grip' the dip present behind the shoulder, and a flat broad back (Henderson, 1968). The resulting odour produced by the high humidity, be it due to bacterial growth, sweat, fleece rot, or dermatophilosis, provides an indication of the moisture and humidity levels of the fleece, optimal for egg viability and hatching. It is the ability of the fleece to reduce moisture accumulation, by body conformation and fleece structure that will be targeted for reduced flystrike susceptibility.

1.4.3 Selection of an indirect indicator for New Zealand

A number of economically important traits have been shown to have a genetic component, and genetic gain can be achieved by direct selection. However, in the case of disease traits, measuring the trait in question is often hard, due to its inconsistent nature, time constraints, money and animal welfare (Blair and Garrick, 2007). In these cases, indirect selection is required, where a trait more easily and accurately measured can be used instead. For example research on facial eczema developed the indirect gamma glutamyltransferase (GGT) test to measure liver damage and an ELISA test has been used for worm burden (Blair and Garrick, 2007).

The uptake of the new phenotype/measure depends on a number of factors (Blair and Garrick, 2007). Firstly, the correlation or the degree to which the indirect trait can predict the disease trait, the higher the correlation, the faster is the genetic gain that can be achieved. Secondly, the difficulty in measuring this new phenotype. The new phenotype needs to be easier to measure than the disease itself. Finally, the cost associated with measuring the new phenotype can be the biggest hurdle. The ram

breeder, who may have to test a large number of potential rams, needs to recoup the total testing cost from the smaller proportion of rams that are sold.

We have documented here the various traits that influence flystrike incidence. The majority of the research has been in Australian Merino resource flocks and translating these results in the New Zealand climate is questionable. The traits selected will also need to be integrated into the New Zealand industry structure.

Table 1.2: Heritability (h^2) of potential indirect selection traits, and their genetic, phenotypic and between flock correlations with flystrike in Australian Merino.

Trait[1]	Genetic	Phenotypic	Between Flock	h^2
Fleece Rot[2]			0.86	
Breech Bareness[3]	0.17(0.19)	0.17		0.42(0.32)
Dag Score[3]	0.86(0.17)	0.23		0.55(0.30)
Urine Stain[3]	0.53(0.22)	0.19		0.49(0.32)
Wool Colour[3]	0.25(0.24)	0.26		0.49(0.32)
Breech Bareness[4]		0.01		0.23(0.09)
Crutch Bareness[4]		0.09		0.47(0.14)
Dag Score[4]		0.22		0.09(0.06)
Urine Stain[4]		0.04		0.30(0.20)
Staple Thickness, mean[5]	0.15	0.13	0.39	0.29
Staple Thickness, sd[5]	-0.63	0.11	0.56	0.16
Staple Thickness, cv[5]	-0.54	0.06	1.02	0.23
Staple Length[5]	-0.69	-0.11	0	0.31
Dust Penetration[5]	-0.34	0.03	-0.76	0.15
Crimp Frequency[5]	0	-0.08	-0.19	0.51
Back fibre dia, mean[5]	0.65	0.09	-0.29	0.56
Back fibre dia, sd[5]	0.35	0.15	0.01	0.53
Back fibre dia, cv[5]	0.04	0.11	0.26	0.46
Birthcoat score[5]	0.08	0.03	0.66	0.62
Neck wrinkle score[5]	0.07	0.1	0.66	0.39
Body wrinkle score[5]	0.22	0.14	0.72	0.38

[1]sd: standard deviation; cv: coefficient of variation; dia: diameter;

[2]Atkins (1979);

[3]Greeff & Karlsson (2009);

[4]Smith et al. (2009);

[5]Raadsma et al. (1993).

Dag score, breech bareness, breech wrinkles and wool type are all contributors to the sheep's risk to flystrike. Apart from fleece rot, dag score has the highest correlation with flystrike (Table 1.2). In New Zealand genetic correlations have not been published, however, observations have shown that flystrike incidence decreases as dag score decreases (Fletcher et al., 1999; Leathwick and Atkinson, 1998; Young et al., 2004) and

breech bareness increases (Scobie et al., 2002). Wool type relationship with flystrike has only been studied in Australia and has mainly investigated the relationship with fleece rot. Raadsma et al. (1993) provided the most comprehensive review of wool production traits and correlations with flystrike, but all these occurred in Australian Merinos. This study is concerned with flystrike in New Zealand dual-purpose and terminal breeds, not the New Zealand Merino. For New Zealand, dag score has the most potential as an indirect indicator of flystrike, however, breech bareness and wool type should also be examined for their effects of flystrike.

1.4.3.1 New Zealand industry structure

The New Zealand sheep industry is well established (>30 million). The industry is divided into 3 breed types: dual-purpose (55.5% of registered ewes), terminal (17.9% registered ewes), and fine wool (13.7% of registered ewes) (Stewart and Garrick, 1996). Each of these types has developed specialised flocks or studs which aim to improve their breed. Dual-purpose breeds are concerned with maternal ability and growth, and mainly include the breeds: Romney, Coopworth and Perendale. While terminal breeds are concerned mainly with growth and carcass quality (mainly Texel, Suffolk and Polled Dorset breeds), and fine wool breeds are concerned with wool (mainly Merino and Corriedale breeds). Genetic gain is passed from these studs to commercial flocks through the sale and use of sires. Selection objectives are used to guide the genetic improvement of the studs. The selection objective outlines the goal traits and the importance of each trait to the animal system (Blair and Garrick, 2007). For each trait, an estimated breeding value (EBV) is calculated, it is weighted by an economic value which may change as the industry changes over time.

Sheep Improvement Limited (SIL) is the current genetic evaluation system available for New Zealand sheep breeders. It superseded the National Flock Recoding Scheme (est. 1967), Sheeplan (est. 1976), and Animalplan (est. 1988) (Young and Wakelin, 2009). It provides genetic evaluations and selection indices for breeders and their clients for dual-purpose, terminal, fine wool and mid-micron flocks (Young and Wakelin, 2009). There is a long list of traits (> 200) available in this database, including those for wool, growth, reproduction, and resistance to internal parasites. It currently provides an across-flock genetic evaluation (SIL-ACE) that includes data from over 300 flocks and 3 million animals (Young and Wakelin, 2009). For dual-purpose flocks, the rate of genetic progress has increased from 23 cents/ewe/year to 84 cents/ewe/year from 1990

to 2006, since SIL was introduced in 1999 (Amer, 2009). For terminal flocks rates of genetic progress have increased from 23 cents to 48 cents/year in the same time period (Amer, 2009).

1.4.3.2 Integration of an indirect indicator

The selection of an indirect indicator, as well as being easy to measure and low cost, needs to able to fit into the current New Zealand sheep industry structure. Currently, dag score and breech bareness are traits that can be recorded in the SIL database. Breeders can request genetic evaluations for dag score, however, this is not yet available for breech bareness (Walker and Young, 2009). Wool quality trait recording and breeding values are also available including; fleece weight, fibre diameter, coefficient of variation (CV) of fibre diameter, curvature, staple length, brightness, and yellowness. Economic weights for these traits are available for mid-micron and fine wool breeds, for dual-purpose and terminal flocks they are only available for fleece weight.

Some potential indirect measurements are therefore already available for use in SIL. Economic weights and genetic models to calculate breeding values for these traits in dual-purpose sheep (with the exception of dagginess breeding values) are not yet available. Before implementation of genomic selection, further development to calculate molecular breeding values for these predictor traits and for flystrike will be required.

1.4.4 Summary

In this section, current research flock results investigating flystrike and indirect indicators have been outlined. As noted these were mainly Australian based, involving Merinos, and have focused primarily on fleece rot as an indirect indicator which is of little relevance to New Zealand. Improvement of flystrike resistance in New Zealand needs to be based on dual-purpose and terminal breeds. Dag score, breech bareness, and wool type have all been examined for potential indirect selection. Dag score is highly correlated with flystrike, and is an economically important trait in its own right. It has also recently been integrated into the New Zealand SIL genetic evaluation system. Breech bareness is also important in reducing dag score and wool stain in the breech region. Currently, values can be recorded in the SIL system but no genetic evaluations are available. Wool trait breeding values are available in SIL, but not for all breed types. No recording or indirect prediction of breeding values for flystrike is available, nor any economically based index. The heritabilities, phenotypic and genetic relationships of

these traits with flystrike in New Zealand first need to be estimated before a flystrike module can be included in SIL.

1.5 Genome-wide association studies

Genome-wide association studies have been used to map and find quantitative trait loci (QTL) associated with traits. Traditional methods involved using linkage studies, candidate gene studies, microarrays and fine mapping. Sequencing of the sheep genome allowed the identification of hundreds of thousands of single nucleotide polymorphisms (SNPs) all along the genome. SNP chips are now available to scan tens of thousands of SNPs selected to have high minor allele frequency and be equally spaced along the genome. All SNPs have the potential to be in linkage disequilibrium with at least one QTL, allowing more rapid identification of QTL associated with the trait of interest. This technology has introduced molecular breeding values, where the information from the SNPs is used to estimate the animal's life time breeding worth soon after birth.

There is huge potential to harness this technology for the investigation of disease traits. These traits often involve the accumulated effect of numerous loci across the genome. The ability to scan the whole genome with a SNP chip will enable these loci to be more easily detected. This section will discuss current and advancing techniques in GWAS, how the information can be harnessed and used for disease traits.

1.5.1 Review of GWAS methodologies

Genome-wide linkage studies were the traditional way to map QTL. QTL were mapped by genotyping families with genetic markers to identify regions associated with the trait of interest. However, when investigating a new trait the initial genome-wide scans used markers widely spaced, typically tens of Mbp apart, requiring further sequencing and genotyping to localise QTL to small genomic regions (Hirschhorn and Daly, 2005). Fine mapping of regions led to marker assisted selection, using single markers or groups of markers (haplotypes). This has been used effectively in animal production, with a number of tests available (McEwan, 2007; Morris et al., 2007). There has been slow progress on disease traits with only 1 or 2 markers available, representing a small proportion of the genetic variation observed (McEwan, 2007). As researchers investigate more traits, we are left with the hard to measure traits or low heritability and

those which involve many loci of small effects. These are harder or impossible to utilise with the traditional mapping techniques.

Regarding flystrike and associated traits there has been little work into identification of QTLs. Dagginess has been investigated in 2 studies (Crawford et al., 2006; MacDonald et al., 1998). MacDonald et al. (1998) found evidence for a QTL, location undisclosed, in 2 half-sib Texel x Coopworth backcross pedigrees, while Crawford et al. (2006) observed no significant loci for dagginess. Genes for pigmentation, medullation, and keratin have already been mapped for wool (Purvis and Franklin, 2005). Many other QTLs affecting wool production and quality have also been observed, however, none concern flystrike risk directly.

Microarray studies have been used in the investigation of QTL for fleece rot. A gene expression study involving 2 ovine-bovine skin cDNA microarrays was used to investigate resistance to fleece rot and body strike (Norris et al., 2008). This was a case-control design with tissue taken from resistant and susceptible animals before, during, and after fleece rot was artificially induced. A list of genes was identified that were differentially expressed, including proteins involved in: apoptosis, adipocytokine signalling, acute inflammatory response, and focal adhesion (Norris et al., 2008). A recent study using microarray and fine mapping of selected genes with SNPs has identified 2 genes, fibulin (FBLN1) and fatty acid binding protein 4 (FABP4), (Smith et al., 2010) associated with fleece rot in Australian Merinos.

The sequencing of many domesticated species: bovine, chicken, and sheep among others, has allowed the implementation of GWAS and GS. Sequencing has allowed the identification of thousands to hundreds of thousands of SNPs. These are then selected for use on SNP chips, with SNPs selected to be equally spaced across the genome and all potential QTL are captured due to linkage disequilibrium with at least 1 of the SNPs (Meuwissen et al., 2001). The ability to genotype thousands of SNPs in single SNP panels has been used to map Mendelain traits, investigate patterns of linkage disequilibrium (LD), map QTL, investigate domestication, and implement GS for genetic gain (Kijas et al., 2009). The sheep and cattle genomes have revealed the genetic structure of these two species (Kijas et al., 2009; The Bovine HapMap Consortium, 2009). A decline in effective population size for all breeds has been observed in cattle (The Bovine HapMap Consortium, 2009). This has occurred recently and is due to the domestication of cattle and the selection of certain breed types,

especially the intense selection for dairy and beef breeds. Distinct selection signatures have been observed (The Bovine HapMap Consortium, 2009). Kijas et al. (2012a; 2009) showed that due to their short evolutionary history, sheep have a weak phylogeographic structure, and high levels of genetic similarity between breeds. There was sufficient variation to cluster breeds based on their geographic origins.

Initial LD mapping in sheep has detected divergence of breeds, mapped breed development, provided evidence of selection sweeps and estimated effective population size (Raadsma, 2010). Selection sweeps due to domestication and breeding have been observed and Mendelian traits such as polledness have already been mapped to a single gene using the Ovine 50K SNP chip (Kijas et al., 2012a; McEwan et al., 2010). One large advantage for the sheep breeder of sequencing the sheep genome and the availability of a dense map of SNPs is the ability to scan an animal's genome and predict its breeding worth over a lifetime; this process is called genomic selection (GS).

Genomic selection is marker assisted selection on a genome-wide scale (Meuwissen, 2007). With the use of SNP chips, estimates are calculated on the effect each SNP has on a trait. The sum of the SNP effects can then be used to predict the animals' molecular breeding value (MBV)[1] (Goddard and Hayes, 2007). Meuwissen et al. (2001) used simulation to show that MBVs could be predicted with an accuracy of 0.85. Combining these MBVs with the EBVs already calculated, generates genomic breeding values (GBV), which could potentially lead to doubling the rate of genetic gain already achieved through using EBVs calculated from phenotypic information alone. These figures depend on the heritability of the trait, if the trait is sex-limited or measured late in life, opportunity for reducing the generation interval, and finally, on the availability of the reference population used in the prediction.

The dairy industry have already been adopting GS to increase genetic gain (Hayes et al., 2009b). Results from preliminary analysis in Australia, New Zealand, United States and the Netherlands have shown that the reliabilities from MBVs are greater than predicting EBVs from parental averages, as was the norm for selecting bulls for progeny testing (Hayes et al., 2009b). Thus it is as accurate as progeny testing, but results can be obtained at birth. The ability to canvass the whole genome using

[1] To keep with the standard nomenclature for the New Zealand sheep industry, in this thesis the breeding value (BV) generated from sum of SNP effects is termed molecular BV (MBV), and the combined MBV + estimated BV (EBV) is termed the genomic BV (GBV). This differs from the literature which uses the term genomic estimated BV (GEBV) for the BV calculated from sum of SNP effects.

thousands of variants enables traits that are; sex-limited, difficult to measure and those measured late in life, to be estimated more accurately (McEwan et al., 2010).

1.5.2 Ovine 50K and 5K SNP chips

After the success of the bovine genome and bovine SNP Chips, the International Sheep Genomics Consortium (ISGC) was brought together to sequence the sheep genome and to design the first Ovine SNP chip (Oddy et al., 2007). Six female sheep from different breeds were each sequenced to 0.5x depth, to give a 3x depth of the whole genome using the Roche 454 FLX sequencer at Otago University, New Zealand and at the Balyor College of Medicine, Human Genome Sequencing Center (BCM HGSC), Houston, USA (McEwan, 2009). Average read lengths were 228 base pairs (bp), which were assembled into contigs of around 800bp in length (Dalrymple et al., 2007), giving the virtual sheep genome (publicly available at www.sheephapmap.org) .

In total, 595,000 SNPs were identified, these were divided into classes A, B, or C, dependent on detection criteria. There were ~240,000 class A Infinium 2 SNPs, approximately 3 every 50,000 base pairs, with an average minor allele frequency (MAF) of 0.31 (McEwan, 2009). Reduced representational sequencing (RRS) of 60 individual sheep from 11 breeds was performed to identify additional SNPs using an Illumina Genome Analyser (Dalrymple, 2009). This was a similar method to cattle RRS (Van Tassell et al., 2008), and resulted in a 20X coverage of ~5% of the genome and 76,000 high quality SNPs were identified (Dalrymple, 2009). Additional SNPs from previous work (n = 1,536) were resequenced in 9 animals from 9 breeds using Sanger sequencing methods.

Approximately 60,000 evenly spaced SNPs were chosen for the final Illumina Infinium based SNP chip (MAF > 0.2). SNPs were selected on a step-wise basis, the next SNP was chosen as the best SNP within a certain region from the last SNP (Dalrymple, 2009). Spacing ranged from 8 to 100 kilo base pairs (kbp), with an average of 1 SNP per 46,000 base pairs. About 5,300 SNPs failed manufacturing, and another 5,100 were discarded after quality control using the HapMap samples. The ovine SNP chip has in effect 49,034 useful SNPs (Kijas, 2010).

Already over 20,000 animals world-wide have been genotyped with the 50K ovine SNP chip, including the HapMap samples. The HapMap samples consist of 3,400 animals from 74 breeds, and greater than 20 countries, collected by the ISGC to help

understand the genetic structure of the world's sheep breeds (Kijas et al., 2009). Initial analysis shows separation of breeds from different world regions, selection sweeps of single gene traits (Kijas, 2010), and LD profiles for genome and population structure (Raadsma, 2010). In New Zealand, ~15,000 animals have been genotyped using the Ovine 50K bead chip over a number of resources (J. McEwan, Pers. Comm.). Most of these samples are industry sires but also include selection lines, extreme phenotypes, and progeny tested animals. Some of the results will be discussed below.

A low density SNP chip of 5,998 SNPs (5,409 pass Illumina's quality control) has been designed. The SNPs chosen, were a subset of the 50K SNPs and were chosen to be equally spaced across the genome and with a high minor allele frequency (Anderson et al., 2012a). This has the potential to be a low cost option for genotyping, where parents and unrelated animals from the same population are genotyped on the higher density SNP chip and the progeny or selection of animals from the same population are genotyped on the 5K SNP chip. The missing genotypes can be imputed, using known SNP allele frequencies in the population and SNP haplotypes to estimate the missing genotypes. It is suggested that the 5K SNP chip, with imputed genotypes, can provide 60-80% of the molecular BV prediction accuracy of the 50K SNP Chip when the sire has not been genotyped (Anderson et al., 2012a), and 95% when the sire has been genotyped.

1.5.3 Sampling strategies

There are a number of sampling and analysis strategies available to investigate the genetic basis of animal traits. Each of these strategies has potential for investigating disease traits. Outlined below are: progeny tests, selection lines, selection sweeps, feral versus improved breeds, and case-control studies.

1.5.3.1 Progeny tests

The New Zealand sheep industry has been proactive in the use of group breeding schemes, commencing in the early 1970's (Blair and Garrick, 2007). They were promoted as increasing the rate of genetic gain by 10 - 18% over more traditional closed and within-flock selection and breeding schemes. Their advantages were significant before the implementation of BLUP genetic evaluation and the availability of intrauterine artificial insemination. Nowadays there are fewer and smaller of these schemes running, but the industry is fully aware of the importance of linking flocks

across the industry and have implemented a large scale progeny testing scheme, the Central Progeny Test (CPT) (McLean et al., 2006). The CPT was implemented in 2001 to improve carcass traits and to genetically link the existing sire referencing schemes.

Progeny testing is useful for those traits that cannot be measured on the individual itself, because it is sex-limited, or is measured after death. In some circumstances it also allows a greater rate of genetic gain as it increases the correlation between the true and predicted genetic merit for the parent. The accuracy of a trait increases with the number of progeny measured, and is thus an advantage for those traits with low heritability. A trait with high heritability will gain less from progeny testing as the parents' phenotype is sufficiently close to their true genetic merit. For disease traits and other low to moderate traits a more accurate estimate of the parent is gained by progeny testing. These values can be used to predict genomic associations when parents are genotyped.

1.5.3.2 Selection lines

There are various methods to determine that a quantitative trait has a genetic component. One way to do this is to generate selection lines, where the lines diverge for a selected trait. Selection lines allow estimation of the heritability and genetic variance of a trait in the population (Hill and Caballero, 1992). The changes in the rate of response reflect the number, size and frequency of the genes influencing the trait. In selection lines that are highly diverged, their crosses can be used to establish the biological and physiological basis of the trait as well as help map the genes affecting the trait (Hill and Caballero, 1992).

There are a number of sheep selection lines for disease resistance that have been developed and many from New Zealand have been described in Morris et al. (2007). Selection for the resistance or susceptibility of certain disease traits have been successful in estimating the heritability of disease resistance and the realised genetic gain that can be made by selection, for example fleece rot (Raadsma, 1991a). However, to identify and map genes, selection lines need to be kept many generations. For example the Trangie fleece rot selection lines were established in the late 1970's and a summary of the genetic gain after 17 years selection was published in 1998 (McGuirk et al., 1978; Mortimer et al., 1998). The resistant and susceptible lines had diverged for natural fleece rot by only 2.8% per year (Mortimer et al., 1998), and no genes have been mapped or identified from this resource as yet.

1.5.3.3 Selection sweeps

Selection sweeps are identified by fixation of alleles in the population under selection. Selection sweeps occur when a mutation with a large effect occurs on a single haplotype. Selection increases this mutated haplotype frequency within the population. The mutation haplotype 'drags' with it closely linked variants (Sabeti et al., 2006). Eventually, the mutation becomes fixed and resides as a homozygous allele. As time passes the haplotype block length diminishes due to recombination and new mutations. The length of the haplotype can be used to estimate the time since selection sweep.

Selection sweeps in domesticated animals arise from three main sources: domestication, breed formation and trait selection (Pickering et al., 2009a). SNP chips are very powerful in detecting monogenic mutations affecting a trait, arising from such selection sweeps. Often, only a small number of animals are required to find these selection sweeps. In sheep, selection sweeps for polledness, muscling, microphthalmia, and yellow fat have already been identified (Kijas et al., 2012a; Pickering et al., 2009a), and many more are under investigation.

1.5.3.4 Feral versus improved

In some cases it is possible to compare feral and 'improved' sheep. The differences are in part determined by those traits that are important for survival outside of domestication. Feral livestock are domesticated animals that have reverted to the wild state. They may carry novel or rare alleles or relics of ancestral undomesticated breeds (van Vuren and Hedrick, 1989). Resistance to disease is often an important adaptation of feral sheep, however, the researcher needs to account for environmental differences. Feral animals may appear to be disease resistant, simply because they are not subjected to high levels of challenge. To compare feral and improved sheep for selection sweeps, the animals need to experience similar environments.

1.5.3.5 Case-control studies

Case-control studies have been used extensively in human disease research (Cardon and Bell, 2001). They have also been used successfully in cattle and sheep studies to investigate such traits as facial eczema in cattle and sheep (Morris et al., 2007), and extremes have been used for yellow fat and microphthalmia (McEwan et al., 2010).

In the design of a case-control study there are two important considerations, the selection of controls and the number of cases and controls needed. Controls need to be

matched to cases, by gender, breed, geograph, mating practices and other population characteristics (Cardon and Bell, 2001), also they need to be true controls devoid of any disease potential. If these steps are not followed, the allele frequency differences between the case and controls can appear as if they are disease related but may be actually due to breed or other genetic stratification effect mentioned above. Similarly, power will be reduced if controls contain samples that are disease susceptible but have not been challenged. The number of samples affects the power of the study. The power is the ability to correctly identify a genuine association (Cardon and Bell, 2001). With disease traits, there are often a lot of loci of small effect which contribute to the genetic variation observed. Also, traits will have some degree of genotype-environmental interaction. However this should not discourage the effectiveness of case control studies for investigating disease traits.

1.5.3.6 Summary

All of the above strategies show potential for the investigation of disease traits. Progeny tests can be used to generate an accurate estimate of the genetic merit of a parent, usually the sire. These traits may not necessarily be the disease itself, but indirect indicators such as dag score, breech bareness or wool type for flystrike. The information from the progeny test can also be used to calculate accurate heritability estimates for the indirect indicators and genetic and phenotypic correlations between these traits. The increased accuracy of measurements can be used for generating estimated breeding values for use in GWAS where the sires are genotyped.

A case-control study choosing diseased animals and controls from the same population is also a good way of investigating a disease trait when deliberate exposure of industry animals is not feasible due to animal ethical and production concerns. A potential difficulty of this approach is collecting the required number of cases and selecting true, and matched, controls. However, a case-control study can provide estimates of heritability for the disease trait and genetic and phenotypic correlations with indicator traits if appropriate adjustments are made using incidence rates within contemporary groups. The phenotypes can also be used in genome wide association studies.

Selection sweeps and feral versus ‘improved’ breed methods may not be as helpful in disease studies if the comparison groups have been selected for several criteria, not just the disease in question. However, this analysis can provide a low-cost, independent,

validation of GWAS results from independent experiments where flystrike and dagginess traits are measured.

1.5.4 Summary

The introduction of the Ovine 50K SNP chip, has allowed GWAS to be used for detecting QTL for disease traits. The advent of this technology allows a quicker and more rapid discovery of loci affecting these difficult to measure traits. To harness this, the correct sampling strategies are required. Current strategies involve progeny tests, selection lines, selection sweeps, feral versus improved sheep and case-control studies.

Most disease traits other than Mendelian genetic disorders are unlikely to involve single gene traits, therefore detection by selection sweeps is unlikely. Domestication has led to the adaptation of sheep so that they are reliant on human intervention for disease prevention. Feral sheep have evolved back from domestication towards the ancestral type (Young et al., 2011). Therefore some QTLs may be discovered using this approach. Selection lines are costly, as they involve commitment for many years to breed lines divergent enough for association studies. Progeny tests, enable a large number of animals to be measured, and allow calculation of breeding values close to the true genetic merit of the trait for the parents. These parents can then be used in GS and GWAS. Case-control studies involve collection of measurements directly from diseased and non-diseased animals, which can also be used directly in association studies.

A combination of progeny testing and case-control collections is the most cost effective option for the investigation of flystrike. Industry animals can be measured for potential indirect indicator traits to allow for calculation of genetic parameters. Flystruck animals from the same flocks and related flocks can be used in a case-control study investigating flystrike directly.

1.6 Conclusions

Myiasis or flystrike in sheep is a major animal health problem. It costs the New Zealand industry around $40 to $50 million per annum and affects the animal, farmer, and the downstream processing industry. The current preventative and treatment practices, though well developed, do not adequately treat this disease. There is also

consumer concern about some practices, especially mulesing and insecticide residues in meat and wool products.

Flystrike is highly environmentally dependent and thus has low heritability when the incidence is low. Selection for flystrike directly is unlikely in industry and indirect selection is probably the best method. Australia has led research into investigating the potential for indirect selection, and has estimated correlated responses in other production traits. It is unfortunate that much of this research has been in the Australian Merino, hence the results are unlikely to translate to New Zealand dual-purpose and terminal breeds.

Three traits have been identified as potential indirect indicators for flystrike in New Zealand dual-purpose breeds. Dag score, breech bareness and wool type have all been investigated. Dag score is the most highly correlated trait with flystrike, in Australian studies. It has moderate heritabilities both in Australia and New Zealand, and has been shown to be associated with flystrike in New Zealand conditions. Breech bareness is a trait largely investigated in New Zealand; Australia have more recently also studied this trait as an alternative to mulesing. It has a moderate heritability and is associated with both dag score and flystrike. Wool type has largely been investigated via the fleece rot trait in Australian Merinos. In New Zealand fleece rot is much less prevalent. Fleece structure and body conformation traits that lead to humidity build up in the fleece will be the main traits relevant in New Zealand.

With the sequencing of the sheep genome and development of the Ovine50K SNP chip, a new technology is available to investigate disease and production traits. The advantage over traditional technology are speed and low cost. Promising results have been shown with a number of single gene traits already identified. Evidence from the bovine SNP chip has shown an animal's life time genetic worth can be accurately predicted from birth by scanning the whole genome. The potential genetic gains this will have on the New Zealand sheep industry will exceed the current progress already obtained from genetic evaluation schemes (Sise and Amer, 2009).

This review has outlined the current technology and knowledge of the factors surrounding flystrike and its associated traits. There is now the potential to expand on this current knowledge and identify QTL associated with flystrike and its associated traits. A genomic breeding value could also be calculated and integrated into the current SIL genetic evaluation system. This will allow breeders and commercial famers to

genetically improve their flocks for flystrike resistance and thus reduce their costs and reliance on the current treatments.

Chapter 2: Case-control experiment: estimation of genetic parameters and a summary of two flystrike seasons

Foreword: A subset of this chapter was presented at the New Zealand Society of Animal Production conference held at Lincoln University 2-5th July 2012. Pickering, N. K., K. G. Dodds, H. T. Blair, R. E. Hickson, P. L. Johnson and J. C. McEwan. 2012. Brief communication: Estimates of genetic parameters for flystrike in New Zealand Romney and Romney cross sheep. p 189-191 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

2.1 Abstract

Eleven industry farms were used to estimate the heritability and prevalence of flystrike in Romney based flocks and to assess the potential of certain traits including dagginess (faecal accumulation around the perineum), breech bareness (BBREECH, area of naturally bare skin around the perineum), and wool fibre traits such as: wool bulk, wool length, mean fibre diameter (MFD), standard deviation of mean fibre diameter (FDSD), coefficient variation of mean fibre diameter (FDCV), proportion of medullated fibres (MED%) and curvature (CURV) as indirect indicators for flystrike resistance. The average prevalence rate across all flocks and both seasons was 2.15%, with 88% occurring as breech strike. Heritability on the observed scale for flystrike was 0.37 ± 0.10. There were moderate heritabilites for dag score, BBREECH, FDSD, FDCV, MED% and CURV (0.23 to 0.38) and there were high heritabilities for LENGTH, BULK and MFD, (0.45 to 0.88). Flystrike had high positive genetic correlations with dag score and FDCV (0.71 and 0.62), moderate positive genetic correlations with LENGTH, BULK, and FDSD, and low to negative genetic correlations with BBREECH, MFD, MED% and CURV. Flystrike had a high phenotypic correlation with dag score (0.62), and low to negative phenotypic correlations with all other traits. The high genetic and phenotypic correlations between flystrike and dag score make this trait a viable option for indirect selection for flystrike resistance in Romney and Romney cross dual-purpose sheep.

2.2 Introduction

Myiasis or flystrike is a significant animal welfare problem for the New Zealand sheep industry, estimated to cost $40 to $50 million annually (Beef and Lamb New Zealand, 1999). Insecticides have been a relatively effective means for controlling flystrike and mulesing has been an effective technique for controlling breech strike in Australia. However, due to increasing consumer concern for insecticide use and mulesing (Sneddon and Rollin, 2010) there has been more emphasis placed on how genetics can be used to control flystrike.

There have been a number of studies on Australian Merinos and some studies in New Zealand on selecting for flystrike resistance through the use of direct and indirect

indicators (Brandsma and Blair, 1997; Greeff and Karlsson, 2009; McGuirk et al., 1978; Rathie et al., 1994; Scobie and O'Connell, 2010; Scobie et al., 2002; Smith et al., 2009). These studies list a number of traits as potential indirect indicators including breech and wool traits. All studies showed that flystrike resistance is heritable and that breeding for flystrike resistance is a potential permanent control strategy.

The objective of this experiment was to investigate flystrike incidence and factors affecting incidence variability in 11 properties throughout New Zealand. The data collected was then used to calculate the heritability, genetic and phenotypic correlations of flystrike and associated traits, with the aim of identifying an associated trait suitable for indirect selection for flystrike resistance in New Zealand dual-purpose sheep.

2.3 Materials and methods

2.3.1 Experimental Design

Animal ethics approval for this experiment was obtained from the Invermay AgResearch Animal Ethics Committee. A case-control test design (Cardon and Bell, 2001) was used for this study. Individual farmers (n = 11) observed fly struck lambs and collected ear tissue, wool samples and measurements (described in measurement section) from each lamb. On each farm normal farm management practices to prevent flystrike were undertaken and recorded. Control lambs, matched by flock and sex, were collected at the end of the flystrike season, which in New Zealand is from October to April (Spring through Autumn) (Figure 2.1). This trial took place across 2 seasons, 2009/10 and 2010/11.

2.3.2 Experimental Site

Farms were located across New Zealand, and involved ram breeding flocks and commercial sheep farmers flocks (Figure 2.2). In the North Island, 4 farms were located in the Wairarapa, 3 were commercial farmers (A-C) and 1 was a ram breeder (D). Two farms were located in Manawatu-Wanganui, both were ram breeders (E, F). There was 1 farm in each of the Waikato (G), Hawkes Bay (H) and Taranaki (I) regions, with the first 2 being ram breeders and the Taranaki farmer a commercial farmer. The Hawkes-Bay breeder was involved only in the 2009/10 season.

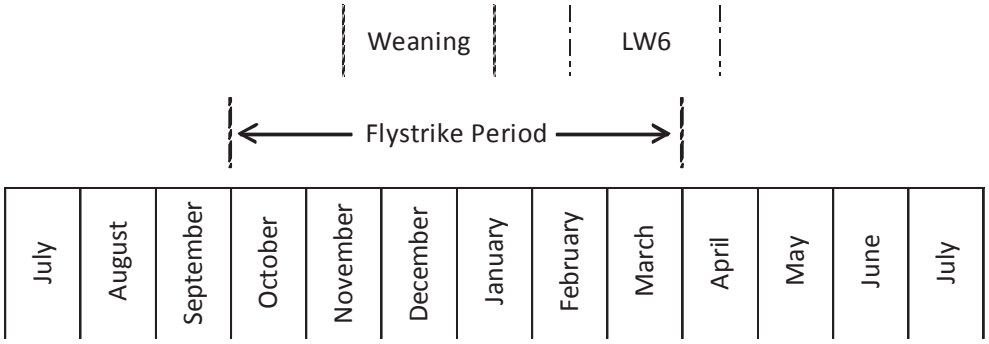


Figure 2.1: The 6-month flystrike season overlapped by the timing of weaning and live weight at 6 months (LW6) recording, when the majority of flystruck lambs were observed and measured by the farmer.

In the South Island, 1 commercial farmer (J) was involved in the case-control test, for the 2009/10 season, from Otago. There was 1 breeder in the Canterbury (K) region during the 2010/11 season. All commercial farmers were linked, by use of sires, to ram breeders involved in this experiment and a concurrent progeny test (Chapter 4). All ram breeders, but not commercial farmers, participated in the progeny test.

2.3.3 Measurements

2.3.3.1 Collections

This experiment involved extensive farmer participation. Farmers were given a pack containing: ear tags for tissue collection (TypiFix™ (Prionics, Switzerland) or FlexoPlus (Caisley, Germany)), ear tag applicators, plastic bags, and a notebook for recording of measurements. For fly struck animals, farmers were asked to measure a range of traits (see below), tag animals for identification and tissue collection, and take a wool sample from as close to the site of flystrike as possible to evaluate the fibre characteristics of the flystrike region. The samples were collected during the flystrike season, but mainly coincided with weaning and other husbandry times (i.e. shearing, insecticide treatment) when lambs were physically handled. The wool and tissue samples were stored together in the freezer until farms were visited at the end of the season, at which time control samples were collected. At this point, missing measurements (described in the following paragraph) on retained flystruck lambs were taken. Samples were collected more regularly from those breeders also involved in the

progeny test. Control lambs were selected by matching birth year and flock. Sex was also used when cases were taken from only one sex.

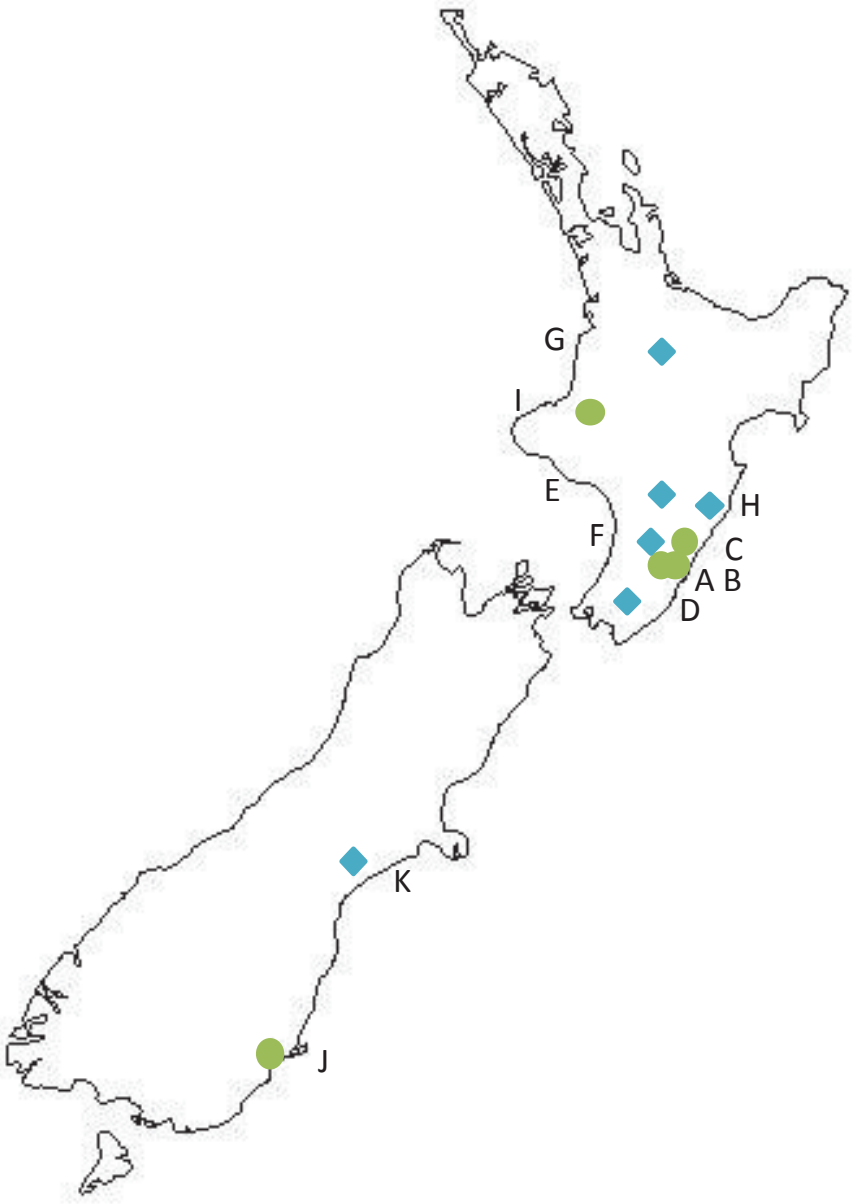


Figure 2.2: Map of experimental sites. Properties involved in case-control test; commercial farmers (●) and Breeders (◆).

The measurements taken were: dag score, breech bareness score (BBREECH (Scobie et al., 2007)), flystrike location, and treatment. Wool length (LENGTH) and hand assessed wool bulk (BULK) information was available only from lambs also measured as part of the progeny test (Chapter 4). Dag score was subjectively assessed

on a 0 (no dags) to 5 (complete dags over the tail, under the crutch, and down the legs) scale (see Figure 1.4). Breech bareness was subjectively assessed on a 1 (wool up to and covering the anus) to 5 (no wool around the breech or down the leg) scale (see Figure 1.5). Wool length was measured in millimetres and measures the relaxed length of the staple under no tension on the spine, above the last rib. Wool bulk was subjectively hand assessed, the wool was compressed by hand and level of resistance scored from 1 (easily compressed, no resistance) to 4 (hard to compress, strong resistance).

Flystrike location was classed into 6 regions (Figure 2.3); breech, body, belly/pizzle, shoulders, head, and feet. Treatment was listed as 1 or more of 4 options; treated with an insecticide, shorn, culled, or not treated. A wool clipping of 5 or more staples was taken as close to the strike position as possible. Wool samples were sent to New Zealand Wool Testing Authority in Napier (NZWTA) for analysis on the Optical Fibre Diameter Analyser 100 (OFDA 100, IWTO-47) (Baxter et al., 1992). At the beginning of the experiment (2009), 120 samples were measured at Massey University on an OFDA 100. Samples were measured for mean fibre diameter (MFD, μm), MFD standard deviation (FDSD, μm), coefficient of variation (FDCV), proportion of medullated fibres (MED, %), and curvature (CURV, ° per mm).

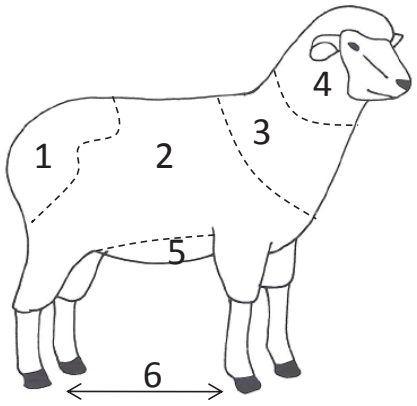


Figure 2.3: Classification of sheep body regions in strike identification. 1: Breech, 2: Body, 3: Shoulders, 4: Head, 5: Belly (and Pizzle in males), 6: Foot. Modified from French et al. (1995).

Additional data were obtained from Sheep Improvement Limited (SIL), for those lambs measured from SIL recorded flocks. These data included parentage, birth and rearing rank, age of dam, and birth date. Sires of those lambs from unrecorded

commercial flocks were determined, and sires of lambs from recorded flocks were confirmed, by paternity parentage testing through a 115 SNP parentage assay (Chapter 5, Clarke et al., 2011).

Each farmer participated in a survey (see Appendix 1) conducted at the end of each flystrike season, to gather a description of the farm and management practices. Rainfall records, where available, were collected from each farmer for the months of October to April.

2.3.3.2 Meteorological observations

In addition to the rainfall records collected from each farmer, weather and rainfall data were also downloaded from the National Institute of Water and Atmospheric Research (NIWA) Virtual Climate Station (VCS) dataset. Virtual stations are located in a 0.05° latitude/longitude grid over New Zealand. The data are generated using spline interpolations of daily climate data observed at the 300 actual climate stations throughout the country (Cichota et al., 2008; Tait et al., 2006).

Data were downloaded for the period July 1st to June 30th for both the 2009/10 and 2010/11 seasons. Daily measurements downloaded included; rainfall (mm), minimum and maximum air temperature (°C) and 9am relative humidity (%). The farmers' rainfall data were used preferentially, while the data downloaded from NIWA's VCS dataset were used to fill in the gaps where no measurements were available from the farmers. Farmers' rainfall data was taken as the most accurate recording, as the VCS dataset at a particular point is estimated from the nearest actual climate station.

Exposure time was calculated as the number of potential flystrike days before collection. Potential flystrike days were calculated from weather data for each contemporary group (farm property by year), using a number of assumptions:

- Pupae require minimum day temperatures above 10°C for adults to emerge (Wall et al., 2001);
- It takes 62 degree-days above a threshold of 11°C for the first batch of eggs to mature and 28 day-degrees for subsequent egg batches to mature (Pitts and Wall, 2004);
- The minimum temperature at which flystrike on lambs will occur is 9.5°C (Broughan and Wall, 2007a);

- Rainfall in the preceding 3 weeks is conducive to flystrike (Wardhaugh and Morton, 1990); and
- The 28 days post-shearing were classed as non-potential days, regardless of day temperature, as sheep with less than 1 month's wool growth are less susceptible to flystrike, (Wardhaugh and Morton, 1990). The fleece is less likely to generate the humidity required for successful egg hatching and 1st stage larvae growth.

For each day from July 1st 2009 for 09/10 season and July 1st 2010 for 10/11 season, the day was scored on its potential to induce flystrike based on the above criteria. Each lamb was then assigned an index of the number of potential flystrike days before it was collected as a case or control.

2.3.4 Statistical analysis

To calculate the significance of the differences for each trait between cases and controls a mixed model procedure (SAS, 2004) was fitted which included flystrike (binary, 0 denoting no flystrike or 1 denoting lamb was flystruck) as a fixed effect and contemporary group (flock by year) and the interaction contemporary group by flystrike as random effects. The least square means and the differences of least square means were calculated.

To estimate the genetic parameters the following procedure was followed. The data were cleaned and traits scaled as appropriate to homogenise the variance (Brown et al., 2005). If there were less than 5 records for any trait for each flock by year contemporary group, data were excluded (n = 10) for that trait. For dag score, BBREECH, BULK, FDCV and CURV, no scaling was applied. For LENGTH, MFD, FDSD, and MED%, data were scaled via means of the contemporary group (flock by birth year). For flystrike, case data were excluded if there were no matching controls collected. This occurred when flock and sex had not been taken into account when selecting controls (n = 32). Flystrike was scaled by incidence rate per contemporary group (flock by birth year), following equation 1 (Mosteller et al., 1970) where *i* is the proportion of flystrike (incidence rate) for the contemporary group. The scaled results were then log transformed using $\log_{10}$(flyscale+1).

$$Flyscale = \frac{Flystrike\ score}{\sqrt{i \times (1-i)}} \qquad \text{Equation 1}$$

Data analysis models developed in Chapter 4 were modified as follows for all traits except flystrike.

- The contemporary group of the pre-existing models were modified to that of flock by birth year;
- For the unrecorded flocks, default values were used for missing values of explanatory variables, as standard practice for SIL. These were: for birth rank and rearing rank, value equal 1, for age of dam, value equal 3, for birthday deviation, value equal 0; and
- For flystrike, fixed effect models were initially determined using the general linear model procedure (SAS, 2004). Fixed effects fitted included: flock, sex (F or M), and contemporary group (flock by birth year).

Insecticide use, scored on a binary scale with 0 denoting no preventative treatment, and 1 denoting preventative treatment used, and exposure time were fitted as covariates for flystrike. Interactions between these effects were tested and discarded from the final model, if not significant at the $P<0.05$ level, via backwards elimination. Traits were then fitted in bi-variate analyses with ASReml (Gilmour et al., 2009) using full animal random effects.

The heritability of liability (h_l^2) to flystrike was derived from the heritability of flystrike on the observed scale (h_o^2) using the following transformation from Lee et al. (2011):

$$h_l^2 = h_o^2 \left[\frac{K^2(1-K)^2}{z^2P(1-P)}\right] \qquad \text{Equation 2}$$

Where K is the proportion of affected animals in the population (prevalence), z is the height of the normal curve at the truncation point and P is the proportion of cases to controls (Lee et al., 2011). The s.e. was similarly adjusted using the equation (Lee et al., 2011);

$$s.e.(h_l^2) = s.e.(h_o^2) \left[\frac{K^2(1-K)^2}{z^2P(1-P)}\right]^2 \qquad \text{Equation 3}$$

2.3.5 Survival data analysis

Data were also analysed using the Survival Kit V6 (Ducrocq et al., 2010). Survival Kit V6 enables analyses of time-dependent variables and censored events. The data were analysed using a Cox proportional hazards model, with flystrike (0 or 1) used as the censor and exposure days as time. Sex, contemporary group (flock by birth year) and contemporary group by sex were used as explanatory variables. Flystrike

heritability on the observed scale was estimated by fitting an animal model (Gábor et al., 2010).

2.4 Results

2.4.1 Climatic records

2009/10 experienced a typical El Niño summer weather pattern, with above normal rainfall in Hawkes Bay, Wairarapa, and Southland (NIWA, 2010a). Temperatures were initially very cold in December followed by heat waves in February (NIWA, 2010a). The north and east of the North Island were very dry during March until May (NIWA, 2010b). Total rainfall for the year 1st July to 30th June 2009/10 ranged from 721 mm on the Otago farm to 1,989 mm on the Taranaki farm (Figure 2.4). Total rainfall over the flystrike 2009/10 season (1st October to 31st March) ranged from 235 to 956 mm (Figure 2.4), with an average of 39 to 159 mm per month per farm. Average daily maximum temperature ranged from 14 to 23°C for the flystrike season, with the hottest month being February (Figure 2.5). Of the properties involved: the farm at Taihape had the highest minimum relative humidity (RH%) between November and March, while the lowest minimum RH% for the same time period was at the Dunedin property (Appendix 2). These 2 properties were the highest and lowest contributors of flystrike cases respectively for the 2009/10 season.

For the 2010/11 season the weather pattern had turned to a La Niña weather system. The summer was a mixture of heat waves and deluge rainfalls (NIWA, 2011b). There were above average to near-record high summer rainfalls in Otago, while temperature means were above average for all of the North Island and the top of the South Island (NIWA, 2011b). Autumn was very warm and wet for most of the North Island, Otago and South Canterbury (NIWA, 2011a). Total rainfall for the year 1st July to 30th June ranged from 759 mm on the Otago farm to 2,575 mm on the Taranaki farm (Figure 2.4). Total rainfall over the flystrike 2010/11 season (1st October to 31st March) ranged from 360 mm in Otago to 862 mm in Taranaki for the 2010/11 season (Figure 2.4), with an average of 60 mm to 143 mm per month. Average daily maximum temperature ranged from 15 to 23°C for the flystrike season, with the hottest months being January and

February which had average daily maximums of 22.3 and 22.9°C respectively (Figure 2.5).

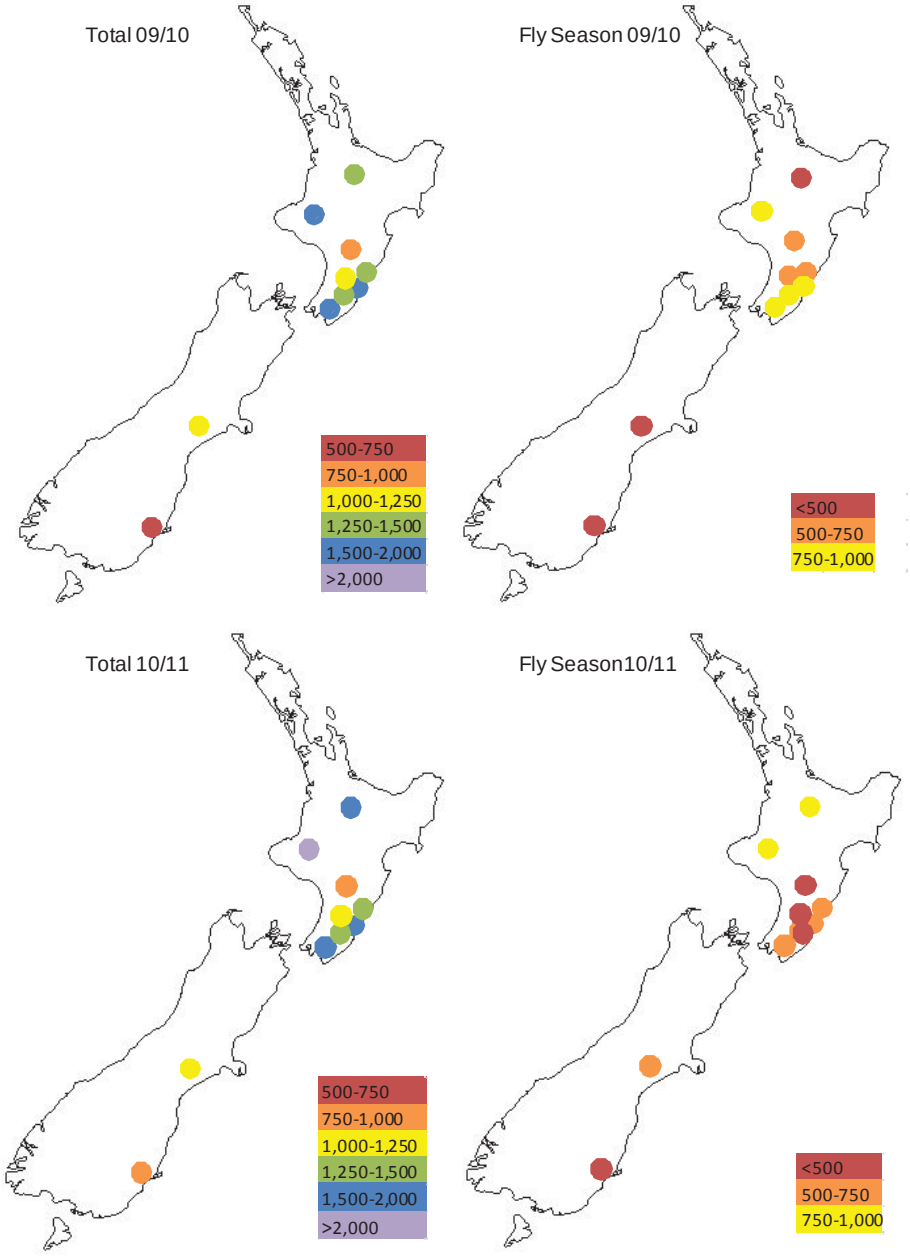


Figure 2.4: Total rainfall (mm) from July 2009 to 30th June 2010 (top left), and over the flystrike season, October 2009 to 31st March 2010 (top right), July 2010 to 30th June 2011 (bottom left), and over the flystrike season, October 2010 to 31st March 2011 (bottom right).

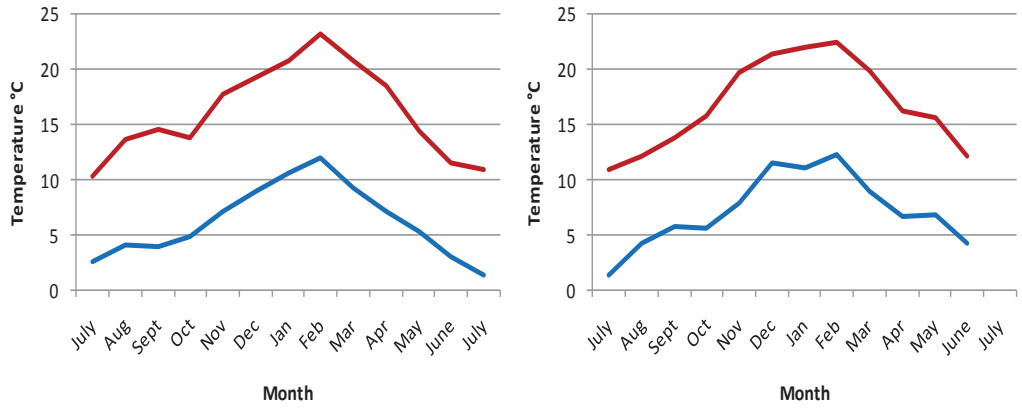


Figure 2.5: Average maximum (red line) and minimum (blue line) temperature (°C) for the 2009/10 season (left) and 2010/11 season (right) across all farms.

2.4.2 Summary of 2009/10 season

Over the 10 farms involved in the 2009/10 season (A to J), 484 cases and 414 controls were collected. The average incidence rate per farm was 1.76%, with a range from 0.47% to 2.95%. The worst hit property was in Taihape, with 137 cases collected in total (Table 2.1, E); 125 in their unrecorded flock and 12 in their recorded flock. In contrast, in Otago (Table 2.1, J) only 10 (2.0 %) of the ram lambs which were left unshorn throughout the flystrike season were struck. The peak of the season was between January and March where 87% of the cases were recorded.

The majority of flystrike occurred in the breech (90.1 %) (Table 2.2). The next most frequent site with 8.2% of cases was the body. The shoulder and belly/pizzle were only recorded in 2 cases each. There were no cases recorded on the head or feet, however, there were 10 cases where the anatomical location of the strike was not recorded.

2.4.3 Summary of 2010/11 season

Over the 9 farms involved in the 2010/11 season (A to G, J-K), 352 cases and 331 controls were collected. The average incidence rate per property was 2.54% with a range from 0.43% to 8.18%. The flock with the highest incidence rate was at Tirau (G), the recorded ram lamb flock had an incidence rate of 8.18% (Table 2.1). Property F, located near Ashhurst, had the lowest incidence rate for the 2010/11 season. The peak of the season was February, where 52% of the flystrike was recorded; both March and April also contributed 38% of the flystrike recorded.

For the 2010/11 season, the majority of flystrike again occurred in the breech (82.7 %), with the next most struck region being the shoulders (8.8%), and then body (7.2%) (Table 2.2). There were 3 records of flystrike on the belly/pizzle and 1 on the head. There were 34 cases where anatomical location of the strike was not recorded.

Table 2.1: Number of flystrike cases (n), incidence rate (I, %), proportion of flystrike located on the breech (B, %) and preventative insecticide use for lambs born 2009 and 2010 per flock.

		2009 born			2010 born			
Farm	Flock[1]	n	I (%)	B (%)	n	I (%)	B (%)	Insecticide use
A	Com	24	1.33	71	na	na	na	Y
B	Com	50	1.92	90	53	1.96	89	N
C	Com[2]	73	2.64	86	6	0.43	100	Y
D	Com[3]	12	1.40	67	31	3.49	97	Y
	Rec	31	2.77	94	31	2.47	97	Y
E	Com	125	3.02	98	35	1.03	94	Y
	Rec	12	2.21	100	9	1.88	89	N
F	Com	16	0.26	86	6	0.09	83	Y[7]
	Rec1	46	4.68	98	15	1.72	83	N
	Rec2	7	2.82	100	14	5.88	92	N
G	Com	10	0.72	90	na	na	na	Y
	Rec[4]	16	3.81	88	18	8.18	56	Y
H	Rec[5]	29	1.75	97	na	na	na	Y
I	Com	21	0.47	81	102	2.34	56	Y
J	Com[6]	10	2.00	80	na	na	na	Y[8]
K	Rec1[4]	na	na	na	28	3.50	100	Y
	Rec2[4]	na	na	na	4	1.40	100	Y

[1]Com: commercial, Rec: recorded flocks;
[2]Ewe replacement lambs only born 2010;
[3] Ewe replacement lambs only;
[4] Recorded ram lambs only born 2010;
[5] Recorded ewe lambs only born 2009;
[6] Unrecorded ram lambs only born 2009;
[7] Preventative used on mob by mob basis;
[8] Preventative used on born 2010.

Table 2.2: Summary of strikes by area for each flystrike season.

Strike Region	2009/10[1]						2010/11[1]					
	E	R	C	U	Total	Prop (%)	E	R	C	U	Total	Prop (%)
Head					0	0.0		1			1	0.3
Body	24	5	10		39	8.2	6	5	12		23	7.2
Breech	262	52	110	7	431	90.1	106	85	71	1	263	82.7
Feet					0	0.0					0	0.0
Shoulders	2				2	0.4	2	4	22		28	8.8
Belly/Pizzle		1	1		2	0.4	1	2			3	0.9
Unknown	1			9	10		29	5			34	
Total	289	58	121	16	484		144	102	105	1	352	

[1]E: ewe lambs, R, entire ram lambs, C: cryptorchids, U: unknown sex, Prop: percentage of total known strikes.

2.4.4 Measurements

There were 836 cases and 745 controls collected over the 2 flystrike seasons. A summary of the least square means and the significant difference between cases and controls for each trait is shown in Table 2.3. There was a significant difference between cases and controls for dag score, FDCV and MED%. On average the control lambs had lower dags, lower FDCV, and lower proportion of MED% than cases.

Table 2.3: Least square means of each trait for cases and controls.

Trait[1]	Control n=745		Case n=836		Significance[2]	
	Mean	s.e.	Mean	s.e.		
Dag score	0.98	0.13	2.64	0.13	<0.0001	***
BBREECH	1.25	0.05	1.26	0.05	0.901	ns
LENGTH (mm)	59.11	2.03	60.66	1.95	0.283	ns
BULK	2.07	0.17	2.04	0.17	0.845	ns
MFD (μm)	35.31	0.47	34.34	0.48	0.052	ns
FDSD (μm)	8.26	0.12	8.32	0.12	0.674	ns
FDCV (%)	23.37	0.30	24.32	0.31	0.008	*
MED (%)	5.59	0.67	7.57	0.70	0.020	*
CURV (°/mm)	44.80	0.76	46.60	0.78	0.060	ns

[1]BBREECH: breech bareness, LENGTH: length of relaxed wool staple on spine, BULK: wool bulk, MFD: mean fibre diameter, FDSD: standard deviation of MFD, FDCV: coefficient of variation of MFD, MED: proportion medullation, CURV: curvature.

[2]*** P<0.001, * P<0.05, ns not significant. Obtained from a mixed model including flystrike (0: control, 1: case) as a fixed effect and contemporary group (flock by year) and contemporary group by flystrike as random effects.

2.4.5 Model fitting

Not all cases had perfectly matched controls (by flock and birth year). After deleting un-matched cases there were 1549 animals remaining, 804 cases and 745 controls. Significant fixed effects for log scaled flystrike were: sex, exposure time, contemporary group (flock by birth year) and contemporary group by sex. The last 2 effects explained 16% and 4% of the variation, respectively. Insecticide use was not significant. Exposure time, however, was discarded from the final ASReml model to determine heritabilities and genetic correlations because it was incompatible in this context, unless it is used in a censored survival model, because most animals were not flystruck before the end of the season. The censored analysis was undertaken separately using Survival Kit V6. The final models used for the ASReml analyses are presented in Table 2.4.

Table 2.4: Final mixed models and fixed effects used for individual trait analysis

Trait[1]	Fixed Effects[2]	Random Effects
Flystrike	sex flk.byr flk.byr.sex	Animal
Dag score	sex brr aod bdev flk.byr	Animal
BBREECH	brr aod bdev flk.byr flk.byr.brr	Animal
LENGTH	aod bdev flk.byr.brr	Animal
BULK	brr aod bdev flk.byr flk.byr.brr	Animal
MFD	brr bdev ofda flk.byr	Animal
FDSD	brr bdev ofda flk.byr flk.byr.brr	Animal
FDCV	brr ofda flk.byr	Animal
MED%	no model	Animal
CURV	brr bdev ofda flk.byr	Animal

[1]Flystrike: log scaled flystrike; BBREECH: breech bareness; LENGTH: length of relaxed wool staple on spine; BULK: wool bulk; MFD: mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; MED%: proportion of medullated fibres; CURV: curvature; WWT: weaning weight at 3 mo; LW8: autumn BW; FW12: fleece weight at 12 mo.
[2] flk: flock; byr: birth year; brr: birth rearing rank; aod: age of dam as linear and quadratic; bdev: birthday deviation; ofda: site where fibre measurement done.

2.4.6 Genetic parameters

2.4.6.1 ASReml heritabilities

Heritabilities, correlations and phenotypic standard deviations are presented in Table 2.5. Flystrike, on the observed scale, had a moderate heritability (0.37 ± 0.10). After transformation to the underlying liability scale using equation 1, heritability for flystrike was 0.26 ± 0.05. There were moderate heritabilites for dag score, BBREECH, FDSD,

FDCV, MED% and CURV (range 0.23 to 0.35), heritabilities for LENGTH, BULK and MFD were high (range 0.45 to 0.88).

2.4.6.2 Survival data analysis heritability

The separately estimated heritability of using the Cox hazard and an animal model in Survival Kit V6 was 0.08 ± 0.05.

2.4.6.3 ASReml genetic and phenotypic correlations

There were high genetic and phenotypic correlations between flystrike and dag score. Flystrike also had a high genetic correlation with FDCV. There were moderate genetic correlations with LENGTH, BULK and FDSD and low to negative correlations with BBREECH, MFD, MED% and CURV. There were low to zero phenotypic correlations with BBREECH, LENGTH, FDSD, FDCV, MED%, and CURV. There was a moderate negative genetic correlation with MFD. However the majority of these correlations had high standard errors.

Dag score had positive moderate genetic correlations with all traits except with BBREECH, BULK and MFD which were moderately negative and with FDSD, and CURV which were close to zero. Phenotypic correlations were low and positive with all traits, except with BBREECH, BULK, MFD, FDSD and CURV. Breech bareness had moderate positive genetic correlations with BULK, MFD, and FDSD all other genetic correlations were moderately negative. Phenotypic correlations with BBREECH were all close to zero, except for a low positive correlation with BULK.

Genetic and phenotypic correlations between LENGTH and BULK could not be estimated, however, they both had moderate positive genetic correlations with the fibre traits except with FDCV which was negative and with CURV which was highly negative with LENGTH but had a positive and low correlation with BULK. Phenotypic correlations with fibre traits were close to zero for BULK except for CURV which was positive and moderate. Wool length had a moderate positive phenotypic correlation with MFD, low positive correlation with FDSD and MED% and negative phenotypic correlations with FDCV and CURV.

Table 2.5: ASReml estimates of heritabilities (diagonal), phenotypic (above diagonal), genotypic (below diagonal) correlations, and phenotypic standard deviation (σ_p) ± s.e.

Trait1	LFLY	DAG	BBREECH	LENGTH	BULK	MFD	FDSD	FDCV	MED%	CURV
LFLY	**0.37 ±**	0.62 ±	-0.03 ±	0.15 ±	-0.05 ±	-0.20 ±	-0.06 ±	0.09 ±	0.09 ±	0.00 ±
	0.10	0.02	0.03	0.08	0.08	0.03	0.03	0.03	0.03	0.03
DAG	0.71 ±	**0.23 ±**	-0.10±	0.05 ±	-0.15 ±	-0.12 ±	-0.02 ±	0.08 ±	0.09 ±	-0.01 ±
	0.13	**0.09**	0.03	0.07	0.07	0.03	0.03	0.03	0.03	0.03
BBREECH	-0.13 ±	-0.33 ±	**0.35 ±**	-0.07 ±	0.12 ±	0.05 ±	-0.01 ±	-0.06 ±	-0.02 ±	-0.02 ±
	0.21	0.25	**0.11**	0.06	0.06	0.03	0.03	0.03	0.03	0.03
LENGTH	0.36 ±	0.33 ±	-0.13 ±	**0.88 ±**	n.e.[3]	0.35 ±	0.13 ±	-0.12 ±	0.12 ±	-0.48 ±
	0.36	0.44	0.37	**0.32**		0.07	0.09	0.09	0.08	0.06
BULK	0.57 ±	-0.38 ±	0.13 ±	n.e.[3]	**0.45 ±**	0.00 ±	0.03 ±	-0.01 ±	0.00 ±	0.25 ±
	0.49	0.58	0.43		**0.31**	0.09	0.09	0.09	0.09	0.08
MFD	-0.12 ±	-0.54 ±	0.34 ±	0.65 ±	0.75 ±	**0.58 ±**	0.62 ±	-0.04 ±	0.22 ±	-0.65 ±
	0.19	0.22	0.20	0.29	0.56	**0.14**	0.02	0.03	0.03	0.02
FDSD	0.38 ±	-0.03 ±	0.19 ±	0.52 ±	0.48 ±	0.72 ±	**0.24 ±**	0.75 ±	0.37 ±	-0.31 ±
	0.28	0.31	0.28	0.51	0.72	0.14	**0.11**	0.01	0.03	0.03
FDCV	0.62 ±	0.53 ±	-0.33 ±	-0.38 ±	-0.38 ±	-0.33 ±	0.39 ±	**0.23 ±**	0.29 ±	0.03 ±
	0.27	0.27	0.28	0.56	0.70	0.25	0.27	**0.12**	0.03	0.03
MED%	0.05 ±	0.30 ±	-0.17 ±	0.29 ±	0.79 ±	0.38 ±	0.51 ±	0.15 ±	**0.25 ±**	-0.15 ±
	0.24	0.28	0.26	0.43	0.58	0.21	0.25	0.32	**0.10**	0.03
CURV	-0.32 ±	-0.08 ±	-0.37 ±	-1.14 ±	0.18 ±	-0.65 ±	-0.79 ±	-0.22 ±	-0.43 ±	**0.38 ±**
	0.22	0.26	0.22	0.21	0.50	0.13	0.22	0.30	0.25	**0.12**
σp	0.51	1.5	0.43	8	0.67	2.93	1.17	2.56	10.13	7.72

[1] LFLY: flystrike, DAG: dag score, BBREECH: breech bareness, LENGTH: length of relaxed wool staple on spine, BULK: wool bulk, MFD: mean fibre diameter, FDSD: standard deviation of MFD, FDCV: coefficient of variation of MFD, MED%: percentage medullation, CURV: curvature.

[2] flystrike heritability on the liability scale is 0.27 ± 0.05.

[3] n.e. non estimable.

Genetic correlations between the fibre traits were positive and moderate except between MFD and FDCV, and between CURV and MFD, FDSD, FDCV and MED% which were moderately negative. The genetic correlation between MFD and FDSD was highly positive. Phenotypic correlations between fibre traits were similar to the genetic correlations.

2.4.7 Survey response

All but 1 farmer answered the survey. Farm size ranged from 80ha to 4291ha, the extremes being a small runoff block and a high country station with 1000ha of rolling downs and flats. The average number of stock units (ignoring the runoff which carried 500 ram lambs) was 8,628 per farm. The majority (91%) of farmers used some insecticide prevention technique for flystrike. Flystrike management began at docking, with three of the 5 farmers who used a hot iron to remove tails also using an insecticide spray on the breech. The remaining farmers (n = 4) who responded used rings and no insecticide. Dipping, or application of insecticide to the whole body, occurred during November to end of March. Six farmers treated every lamb on the property, 3 farmers treated lambs once only while the other 3 treated lambs twice (2 months apart). Two farmers only treated the unrecorded mobs, with 1 treating on an as needed basis. One commercial farmer did not use any insecticide as a preventative treatment, stating that shearing was the main preventative technique used. Another commercial farmer located near Dunedin only used insecticide on a seasonal basis. In the 2009/10 season the ram lambs were left unshorn and untreated until sold in autumn. While in the 2010/11 season all ram lambs were treated. Every farmer used an insecticide to treat flystruck animals.

Products from all three classes of insecticides were used, with the 2 most used insecticides being an organophosphate (Topclip®) and an insect growth regulator (Vetrazin®) (Table 2.6).

Table 2.6: Insecticide use by farmers (A to K) for the prevention and treatment of flystrike. IGR: insect growth regulator, OP: organophosphates, SP: synthetic pyrethoids.

Type	Farmers	Preventative	Treatment
IGR	6	A, D, E, H, I, K	I
OP	7	A, C, D, F	A, B, D, F, G, I
OP/SP	1		E

2.5 Discussion

2.5.1 Summary of 2009/10 and 2010/11 seasons

The 2 flystrike seasons encompassed both an El Niño (2009/10) and La Niña (2010/11) weather pattern. The 2010/11 season with its La Niña weather resulted in a higher average incidence rate (2.54%), than the 2009/10 season (1.75%). This suggests that during a La Niña summer, more care is required, as the increased rain and temperatures affecting the whole country are more conducive to flystrike.

The use of insecticide was not significant when fitted in the flystrike model. As all animals within a flock are treated the same way, contemporary group absorbs any effect. In both seasons within the three most affected flocks, only 1 flock (F, Rec1 (09/10) and Rec2 (10/11)) had not used preventative insecticide treatment. Farmers E, F, and B have been selecting for flystrike resistance, by limiting their use of insecticides, for many years, and have potentially made some progress towards a naturally resistant flock.

There was a lower incidence of flystrike in the commercial flocks, especially those on the same property as a recorded flock. Heath and Bishop (1995), suggest that if covert strikes were accounted for then the incidence rate would be up to double that of the overt strikes. For the commercial flocks in this study it is likely there are a number of covert strikes not discovered, as these flocks are not handled as often as recorded flocks. The recorded flocks require more frequent handling to record measurements for selection purposes. Therefore, the chance of identifying lambs that are flystruck is much greater in the recorded flocks.

The prevalence rate of flystrike in New Zealand was estimated at 3-5% of the national flock (Heath and Bishop, 1995). Studies from the last decade in New Zealand industry and research flocks have estimated incidence rates in the range of 0.5 to 33% (Brandsma and Blair, 1997; Litherland et al., 1992; Scobie and O'Connell, 2010; Scobie et al., 2002). From these studies, estimates for Romney based flocks ranged from 5 to 24% (Litherland et al., 1992; Scobie and O'Connell, 2010; Scobie et al., 2002).

An industry report by Cooper (2011) for New Zealand Leather and Shoe Research Inc. showed that 1.3% of pickled lamb pelts had flystrike faults for the 2009/2010 season. This indicates that a proportion of flystruck lambs survive to slaughter. It was noted in the report that flystrike faults peaked during July – October, suggesting there is

a lag between treatment and healing of flystrike and when lambs have recovered to slaughter weight. The prevalence estimates from this study were at the lower end of the previous estimates. One reason is that the majority of the previous studies were more thorough in detecting flystrike events. Thus the covert strikes in this current study would have been identified if the husbandry activities of the previous studies were applied here. Another reason is the degree of preventative treatment was lower in these previous studies.

2.5.2 Genetic parameters

Heritability estimates for flystrike in New Zealand are limited; Brandsma and Blair (1997) estimated heritability in 3 Perendale flocks as 0.18 ± 0.04. There are several estimates in Australian Merinos. Raadsma (1991a) estimated the heritability of body strike as 0.26 ± 0.12 and on the liability scale as 0.53 ± 0.25. Estimates of heritability of on the observed scale for breech strike were 0.32 ± 0.11 and 0.57 ± 0.28 (Greeff and Karlsson, 2009; Smith et al., 2009). In South African Merinos, the heritability for breech strike was 0.33 ± 0.16 in mulesed, and 0.46 ± 0.23 in unmulsed Merinos (Scholtz et al., 2010). The heritabilities estimated in this study on the observed and liability scale are similar to estimates reported in Australian and South African Merinos.

Flystrike is technically better modelled as a censored trait using time from start of the flystrike season to the day strike occurs as the trait. The estimate from Survival Kit V6 (Ducrocq et al., 2010) was very low compared to estimates generated from ASReml. It is however a much more conservative estimate of the heritability of flystrike resistance, and the true heritability is probably between the heritability estimated on the observed or liability scale and that estimated using Survival Kit V6 estimated heritability. The Survival Kit V6 is not compatible with the current SIL system compared to ASReml which limits its usefulness in the New Zealand national sheep genetic evaluation scheme. Random regression models (Veerkamp et al., 2001) and COXF90 (Misztal, 2009) are other programs/methods that could be investigated. Given the flexibility of using ASReml, especially when including correlated traits in breeding value estimations, for example dagginess as in the current study, it is suggested that the industry continues to use a simple model of presence or absence of flystrike over the first spring, summer and autumn period of the lamb's life and to not implement a survival model.

Previous[2] estimates of the genetic and phenotypic correlations between breech strike and dag score in Australian Merinos, are 0.86 and 0.23 for genetic and 0.22 for phenotypic correlation (Greeff and Karlsson, 2009; Smith et al., 2009). In England it was noticed there was a significant difference between dag score in flystrike cases and controls (French et al., 1996), with flystruck animals tending to have higher dag scores and softer faeces. Leathwick and Atkinson (1998) in New Zealand dual-purpose coarse wool sheep also noted a phenotypic association between dag weight and flystrike (r = 0.58 to 0.82).

The correlations found in this study are high and similar to the previous reports and indicate that dag score would be useful as an indirect predictor of flystrike. Turner and Young (1969) demonstrated how to compare the relative efficacy of using indirect selection (on dag score) compared to direct selection (on flystrike).

Using the observed heritability from ASReml, the relative efficacy is 0.55, indicating that direct selection is more efficient. However, using the conservative heritability calculated using the Cox hazard model, relative efficacy is 1.2, and indirect selection on dag score is more efficient. Even with the range of 0.55 to 1.2 for estimates of relative efficacy, for practical reasons it is still more effective to measure and select on dag score for the following reasons. Flystrike is seasonally dependent, it is hard to measure, and for a true measure animals should be left untreated. However, exposing untreated animals raises animal welfare concerns. Therefore indirect selection for this trait is more appropriate.

Correlations of flystrike with breech cover score in Australian Merinos were 0.17 (genetic) and 0.01 to 0.17 (phenotypic) (Greeff and Karlsson, 2009; Smith et al., 2009). In a Merino cross Wiltshire experiment, it was shown that wool shedding, and thus breech bareness, increased as the proportion of Wiltshire in the cross increased, with a concomitant decrease in the proportion of flystruck animals (Rathie et al., 1994). In a New Zealand study it was shown that the proportion of flystrike decreased as breech bareness score increased (Scobie et al., 2002). This study was in a mixed breed resource and could have been a consequence of using some short wool (Finnish Landrace), naturally bare sheep (Wiltshire), and naturally resistant (feral) breeds in the crosses. Genetic and phenotypic correlations of breech bareness with flystrike were low in the

[2] This chapter is concerned with the correlation of indicator traits with flystrike. Correlations between the indicator traits will be further discussed in Chapter 4.

current study, possibly due to the study involving Romney base flocks, with little variation in breech bareness. Breech bareness would not be a good trait to use for indirect selection of flystrike resistance, in these flocks.

There are only a few published estimates of genetic correlations between wool and fibre traits with flystrike. This is surprising as it is known that wool plays a large part in flystrike susceptibility. Raadsma (1993) estimated genetic and phenotypic correlations of fibre traits with body strike incidence in Australian Merinos. Phenotypic correlations with staple length, MFD, FDSD, and FDCV were -0.11, 0.09, 0.15, and 0.11 respectively. The genetic correlations were -0.69, 0.65, 0.35 and 0.04 respectively. Brandsma and Blair (1997) found that there was a significant association with fleece weight; sheep with lower fleece weight are less at risk of flystrike than those with a heavier fleece. In the current study, FDCV shows the highest potential as an indirect predictor for flystrike, the relative efficacy of indirect selection was estimated at 0.49. The genetic correlations of flystrike with BULK, FDSD, and FDCV suggest that the degree of variation in the fleece and the degree of compactness are important fleece characteristics for flystrike resistance. The high standard errors of the genetic correlations and the low phenotypic correlations mean more work in this area is required before their use as indicators of flystrike.

2.6 Conclusion

Flystrike is a seasonally dependent trait; the observations over the 2 seasons involving an El Niño and a La Niña summer demonstrated the effect of variation in weather on flystrike across the country, with the main effect being the peak of the flystrike season being pushed one month later during the La Niña summer. In the El Niño summer, the flystrike peak was spread across three months (to March), while for the La Niña summer the peak was mainly during February with some in March and April. The average incidence rate of flystrike in the flocks in this study was similar to that previously estimated by Heath and Bishop (1995). The results from this study indicate that it should be possible to reduce flystrike susceptibility by genetic selection. Flystrike has a moderate heritability and it is not surprising that there was a high genetic and phenotypic correlation with dag score, as the majority of strike occurs in the breech. It is, therefore, suggested that dag score could be used as an indirect indicator for

flystrike, while avoiding the ethical issues of exposing untreated animals and recording the incidence of flystrike. Wool bulk and FDSD and FDCV are other potential indirect predictors in dual-purpose sheep. However, these traits need further investigation before industry use.

Chapter 3: Genetic parameters for production traits in New Zealand dual-purpose sheep, with an emphasis on dagginess

Forward: This chapter was published in the Journal of Animal Science and is as published here with format changes to fit in with the required thesis layout Pickering, N. K., K. G. Dodds, H. T. Blair, R. E. Hickson, P. L. Johnson and J. C. McEwan. 2012. Genetic parameters for production traits in New Zealand dual-purpose sheep, with an emphasis on dagginess. J. Anim. Sci. 90: 1411-1420.

3.1 Abstract

Genetic and phenotypic parameters were estimated for production and disease traits (including dagginess) from about 2 million pedigree recorded animals born between 1990 and 2008 in New Zealand dual-purpose ram breeding flocks. This is the most comprehensive study of genetic parameter estimates for the New Zealand sheep industry to date and includes estimates that have not previously been reported. Estimates of heritability were moderate for live weight at 8 months (LW8), fleece weight at 12 months (FW12), dag score at 3 and 8 months (DAG3, DAG8) (0.31 to 0.37), typical for weaning weight (WWT), faecal egg count in summer (FEC1) and autumn (FEC2), and analogously nematodirus counts (NEM1, NEM2) (0.17 to 0.21) and low for number of lambs born in ewes (NLB) (0.09). The genetic correlations among production traits, WWT, LW8, and FW12, were positive and moderate to high. Correlations of DAG3 and DAG8 with production and disease traits were low and mostly negative. NLB had low, but typically positive, correlations with other traits. Disease traits also had low, but positive correlations with production traits (WWT, LW8, and FW12), and were highly correlated among themselves. In general the heritability estimates for live weights and dagginess were higher than currently used in the New Zealand genetic evaluation service (Sheep Improvement Limited) and the availability of accurate estimates for dagginess plus parasite resistance and their genetic correlations with production traits will enable more accurate breeding values (BVs) to be estimated for New Zealand sheep.

3.2 Introduction

Dagginess, or the accumulation of faecal material around the perineum region of sheep, is a trait of interest to New Zealand sheep breeders and farmers for a variety of reasons. Firstly, flystrike is a major animal welfare problem, and it is thought that the presence of dagginess is a strong indicator of flystrike susceptibility. This assumption is supported by dagginess being reported as having a positive genetic (0.86± 0.17) and phenotypic correlation (0.23) with flystrike in Merinos (Greeff and Karlsson, 2009). Secondly, there are financial penalties for presenting daggy sheep for slaughter and the quality thresholds are being driven higher by the shift from frozen to chilled meat,

reducing allowable levels of carcass faecal contamination. Third, the number of stock units managed per labour unit has increased by 21% from 1819 (1980/81) to 2198 (2007/08) (Beef and Lamb New Zealand, 2009) and dagging sheep is a management constraint. Finally, the costs of activities such as dagging and dipping have increased substantially. Between 2000 and 2009, shearing expenditure increased by 74.1%, and animal health expenses by 87.2% (Beef and Lamb New Zealand, 2009).

Currently, selection against dagginess is based on measurements of lambs at 3 or 8 months of age or both, to predict a lifetime dag score. However, the current heritability estimate used by Sheep Improvement Limited (SIL) for dagginess was based on small datasets and omitted key correlations with production and disease traits. Similarly, current SIL estimates for production and disease traits use research results published more than 15 years ago (Bisset et al., 1992; Johnson et al., 1989; McEwan et al., 1995). Presently, the New Zealand industry is dominated by Romney, Coopworth, Perendale, Texel and composite crosses of these breeds which are merging to form a single gene pool. As the industry gene pool evolves and new traits are being measured, the current genetic estimates are becoming outdated. Here we present new genetic estimates, including new fixed effects models for dag traits, calculated from current industry data, for production, disease and dag traits. These estimates will be integrated into SIL to improve the accuracy of prediction of breeding values (BVs) and thus improve the rate of genetic gain in the New Zealand sheep industry.

3.3 Materials and methods

The work reported here was undertaken using records sourced from New Zealand sheep breeders and stored in the SIL database. The animals were managed on commercial farms in accordance with the provisions of the Animal Welfare Act 1999, and the Codes of Welfare developed under sections 68-79 of the Act.

3.3.1 Data

Information on more than 2 million pedigree-recorded animals, born 1990 to 2008, was obtained from SIL. The SIL database contains records on New Zealand sheep breeder and research flocks, and SIL conducts across-flock genetic evaluations (Young and Newman, 2009). Flocks involved in this study were a subset of the New Zealand

sheep industry that were involved in Ovita-funded genomic selection programmes conducted by AgResearch. These flocks cover a range of breeds and crosses, but are predominantly Romney, Coopworth, Perendale, Texel, and some composites (Gosey, 1991).

Data were obtained for weaning weight at 3 months (WWT), live weight at 6 and 8 months (LW6, LW8), fleece weight at 12 months (FW12), and dag score at 3 and 8 months (DAG3, DAG8). The data also included the following traits treated as repeatable measures: number of lambs born in ewes (NLB), faecal egg count of Strongyle (primarily *Ostertagia* spp, *Trichostrongylus* spp, *Cooperia curticei,* and in a proportion of farms *Haemonchus contortus* (Vlassoff et al., 2001)) in summer (FEC1) and autumn (FEC2), and analogously *Nematodirus* spp egg count in summer and autumn (NEM1, NEM2).

The WWT, LW6, LW8, FW12, and NLB are standard production traits recorded by the majority of flocks. The DAG3 and DAG8 are subjective, visually-assessed traits commonly scored on a 6-point scale: 0 (no dagginess) to 5 (complete coverage of the breech and down the legs by faecal material). In a few flocks, where an older 4-point scale had been used; 0, 1, 2, and 3 were transformed to 0, 1, 3, and 5 respectively. The FEC and NEM were scored as a summer (FEC1/NEM1) and/or autumn (FEC2/NEM2) challenge. Two samples (a and b), were often but not always collected at each time point, several days apart. Further details about these traits are presented in SIL technical notes (www.sil.co.nz/getdoc/f7e39437-f465-403d-9565-d1d605da676a/ HealthIntParasite.aspx).

Data were cleaned and traits scaled as appropriate to homogenise the variance (Brown et al., 2005). For WWT, data were scaled via means of the contemporary group (flock.birth year.sex.wwtmob, where "." is used to indicate an interaction: see statistical analysis section) to the overall average of the data set. Either LW6 or LW8, was used as the autumn live weight, taking the trait with the maximum amount of data for the contemporary group (flock.birth year.sex.lw8mob.lw6mob.wwtmob). The new autumn (LW8) values were then scaled via means of its new contemporary group (flock.birth year.sex.lw8mob.wwtmob) to the overall average of the data set. For FW12, data were scaled via means of the contemporary group (flock.birth year.sex.fw12mob.lw8mob. wwtmob) to the overall average of the data set. To offset the differences in age at

measurement for WWT, LW8, FW12, DAG3, and DAG8, birthday deviation from the mean of the contemporary group was used as a covariate in analysis.

For DAG3 and DAG8, data were first screened to ensure average age at measurement, for flock and year, fell within stated bounds (pre-weaning, DAG3, less than 150 days of age and post-wean, DAG8 more than 150 days). Examination of the data identified there was a reduced variance in contemporary groups (flock.birth year.sex) having dag scores below a certain threshold. To mitigate this, log and arcsin transformations were tested, but these did not improve the fit to normality or homoscedasticity. The data were therefore censored, with the trait set to missing for those individuals belonging to contemporary groups whose average fell below a specified threshold (0.5 for DAG3 and 0.25 for DAG8). The assumption was that these contemporary groups were measured before the trait could be fully expressed and thus little variability was seen among the animals tested. For DAG3 22% and for DAG8 15% of the contemporary groups were censored.

The FEC and NEM traits are stringently edited before entering into SIL, following strict guidelines so as to overcome problems of insufficient challenge, and appropriate data transformations have been previously explored (McEwan, 1994). Therefore, we followed the current SIL methodology and used $\log_e(x+50)$ transformation for these traits. For NLB there was no scaling, however, if no NLB data were entered for an animal, the pregnancy ultrasound scan data (NLS) were used as a proxy when available. Typically, industry estimates suggest 98% concordance between the 2 traits (Farmer and Davis, 1999). NLS data contributed 1.4% of the data.

3.3.2 Statistical analyses

Data analysis models were configured for each trait separately. Fixed effect models were initially determined using the general linear model procedure (SAS, 2004). Fixed effects fitted included; flock (flk), sex (Female or Male), birth-rearing rank (brr, born single, twin, or triplet, reared as single, twin, or triplet), and grazing mob (trait mob), or contemporary group (birth year.traitmob.WWTmob). Birthday deviation (bdev) from the mean of contemporary group flock.birth year.sex.WWTmob and age of dam as linear (aod) and quadratic (aod^2) effects, were fitted as covariates. Interactions between these effects were tested and discarded from the final model if not significant via backwards elimination. To assess the significance of maternal (random) and breed (covariate) effects: these were fitted in uni-variate models for each trait using ASReml

(Gilmour et al., 2009). A maternal genetic effect was retained only for WWT and breed was discarded for all traits because genetic variance estimates varied by less than 1 percent. For the FEC and NEM traits, due to the data structure where some of the cohorts had single measurements at each challenge and other cohorts had 2 measurements, repeatability estimates within challenge were estimated. Traits were then fitted in bi-variate analyses with ASReml using full animal random effects as well as maternal and permanent environmental effects where suited (Table 3.1). Genetic and phenotypic parameter estimates and their standard errors were created by averaging across all relevant bi-variate runs.

Table 3.1: Final mixed models and fixed effects used for individual trait analysis.

Trait[1]	Fixed effects[2]	Random effects
WWT	brr, sex, aod, bdev, flk.yr.WWTmob.sex	Animal, Maternal
LW8	brr, sex, aod, bdev, flk.yr.LW8mob.WWTmob.sex	Animal
FW12	brr, sex, aod, bdev, flk.yr.FW12mob.WWTmob.sex	Animal
DAG3	brr, sex, aod, bdev, flk.yr.DAG3mob	Animal
DAG8	brr, sex, aod, bdev, flk.yr.DAG8mob	Animal
NLB	flk.time.sex.lambyr	Animal, eperm[3]
FEC1	flk.time.sex.yr.FEC1mob	Animal, eperm[3]
FEC2	flk.time.sex.yr.FEC2mob	Animal, eperm[3]
NEM1	flk.time.sex.yr.FEC1mob	Animal, eperm[3]
NEM2	flk.time.sex.yr.FEC2mob	Animal, eperm[3]

[1] WWT: weaning weight; LW8: live weight at 8 months; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 months and 8 months; NLB: number of lambs born in ewes; FEC1, FEC2, NEM1, NEM2: faecal egg count in summer and autumn, and analogously *Nematodirus* counts.

[2] brr: birth rearing rank; aod: age of dam as linear and quadratic; bdev: birthday deviation; flk: flock; yr: birth year.

[3] Permanent environmental effects.

Three flocks were used to compare the current SIL dagginess model with the DAG3 and DAG8 models fitted in this analysis. Each flock was analysed separately for each model, with ASReml (Gilmour et al., 2009), using data from animals born 1990 to 2010. The current SIL model scales DAG3 and DAG8 via means of its contemporary group (flock.birth year.sex.DAG3mob.DAG8mob) to the overall mean of the dataset. For analysis, the current SIL model treats dagginess as a repeatable trait generating one estimated breeding value (EBV). Birthday is used as a covariate, flock and sex are used as fixed effects.

Table 3.2: Least square means and ANOVA summary for traits[1]

Item	WWT	LW8	FW12	DAG3	DAG8	NLB	FEC1[2]	FEC2[2]	NEM1[2]	NEM2[2]
No. Obs	1,747,837	1,010,961	768,012	47,957	45,309	929,295	132,131	130,819	123,296	124,135
LS mean	27	40	3.1	1.09	1.26	1.71	6.5	6.63	4.47	4.35
Residual SD	3.93	4.34	0.41	1.26	1.39	0.62	0.76	0.74	0.64	0.61
CV %	14.56	10.84	13.32	115.11	110.22	36.04	11.65	11.14	14.29	13.94
R Square	0.33	0.21	0.04	0.14	0.18	0.21	0.41	0.45	0.27	0.23
Fixed effects[3]	0.02[4]	0.02[4]	0.003[4]	0.12[4]	0.16[4]	0.21[5]	0.41[6]	0.45[6]	0.27[6]	0.23[6]
Contrasts[7]										
brr 11	7.87 ***	5.96 ***	0.19 ***	-0.02 ns	0.07 *					
brr 12	5.24 ***	4.14 ***	0.12 ***	-0.03 ns	0.02 ns					
brr 21	5.61 ***	4.19 ***	0.11 ***	-0.08 *	0.06 ns					
brr 22	2.58 ***	2.07 ***	0.07 ***	-0.02 ns	0.05 ns					
brr 31	3.45 ***	2.36 ***	0.04 ***	-0.07 ns	0.04 ns					
brr 32	1.73 ***	1.27 ***	0.03 ***	0.05 ns	0.07 ns					
brr 33	0	0	0	0	0					
Covariates[8]										
bdev	-0.18 ***	-0.18 ***	-0.01 ***	-0.02 ***	0.004 ***					
aod	2.04 ***	1.87 ***	0.06 ***	0.08 ***	-0.07 ***					
aod_2	-0.22 ***	-0.21 ***	-0.01 ***	-0.01 ***	0.01 **					

[1] WWT: weaning weight; LW8: live weight at 8 months; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 months and 8 months; NLB: number of lambs born in ewes; FEC1, FEC2, NEM1, NEM2: faecal egg count in summer and autumn, and analogously *Nematodirus* counts.

[2] $Log_e(x+50)$

[3] Proportion of variance explained by fixed effects.

[4] flk.yr.mob.sex; flk: flock; yr: birth year; mob: grazing mob.

[5] flk.time.sex.lambyr.

[6] flk.time.sex.yr.mob.

[7] Contrasts birth rearing rank (brr) compared to born and raised as a triplet (brr 33).

[8] bdev: birthday deviation; aod: age of dam as linear; aod_2: age of dam as quadratic.

***P<0.001, **P<0.01, *P<0.05, ns not significant.

3.4 Results

Table 3.2 shows least squares means and an ANOVA summary for the traits used in the analysis. Birthday deviation, age of dam (both linear and quadratic) were significant for all relevant traits ($P<0.01$ to $P<0.001$). Each level of birth-rearing rank was compared against lambs born and reared as triplets. Weaning weight, LW8 and FW12 were all affected by birth-rearing rank ($P<0.001$). However, for DAG3 and DAG8 birth -rearing rank was not significant ($P<0.05$) for most comparisons. Significant fixed effects for DAG3 and DAG8 included flock, grazing mob and sex. For DAG3 (results not shown), sex had very little effect ($P<0.05$), as at time of measurement the animals were yet to be weaned and run in separate mobs. Sex did have a significant effect ($P<0.001$) on DAG8, however, much of this effect was possibly accounted for by mob as post weaning they are run in different mobs and subjected to different environments.

Heritability, correlations, and repeatabilities are presented in Tables 3.3 and 3.4. The production traits of LW8 and FW12 have moderate heritabilities. The high correlation between LW8 and WWT made it difficult to separate the maternal and direct variance components. After removing LW8, maternal heritability of WWT was 0.20 ± 0.001, and WWT direct heritability was 0.14 ± 0.002. Direct and maternal estimates were constrained to have no relationship. The DAG3 and DAG8 had moderate heritabilities, while the heritability estimates for the internal parasite traits of FEC1, FEC2, NEM1, and NEM2 cluster between 0.18 ± 0.01 to 0.21 ± 0.01.

Genetic correlations were positive between WWT and all traits assessed except for DAG3 and DAG8, a similar positive trend was seen with LW8. FW12 genetic correlations were positive with all traits except NLB. The DAG3 trait was negatively genetically correlated with all FEC and NEM traits. The DAG8 trait was negatively correlated with NEM2, all other estimates were close to zero except with DAG3. The NLB trait was positively correlated with all traits except FW12; however, all estimates were close to zero. The FEC and NEM traits were all positively correlated with each other.

Table 3.3: Estimates of heritabilities (diagonal), phenotypic (above diagonal), genotypic (below diagonal) correlations, and phenotypic standard deviations (σp) and repeatability estimates (last row) ± s.e.

Trait[1]	WWT	LW8	FW12	DAG3	DAG8	NLB	FEC1	FEC2	NEM1	NEM2
WWT	**0.14 ± 0.00**[2]	0.73 ± 0.00	0.21 ± 0.00	-0.14 ± 0.00	-0.02 ± 0.01	0.01 ± 0.00	0.03 ± 0.00	0.03 ± 0.00	0.00 ± 0.00	0.02 ± 0.00
LW8	0.91 ± 0.00	**0.35 ± 0.00**	0.36 ± 0.00	-0.09 ± 0.01	-0.05 ± 0.01	0.01 ± 0.00	0.03 ± 0.01	0.04 ± 0.01	0.02 ± 0.01	0.03 ± 0.01
FW12	0.19 ± 0.01	0.21 ± 0.01	**0.37 ± 0.00**	0.00 ± 0.01	-0.03 ± 0.01	-0.01 ± 0.00	0.03 ± 0.01	0.05 ± 0.01	0.02 ± 0.01	0.03 ± 0.01
DAG3	-0.13 ± 0.03	-0.10 ± 0.03	0.07 ± 0.03	**0.34 ± 0.01**	0.40 ± 0.01	0.00 ± 0.01	-0.01 ± 0.01	-0.01 ± 0.01	-0.03 ± 0.01	-0.05 ± 0.02
DAG8	-0.01 ± 0.03	-0.01 ± 0.03	0.09 ± 0.03	0.71 ± 0.02	**0.31 ± 0.01**	0.01 ± 0.01	0.01 ± 0.01	0.03 ± 0.02	0.00 ± 0.01	-0.03 ± 0.02
NLB	0.08 ± 0.01	0.07 ± 0.01	-0.06 ± 0.01	0.01 ± 0.04	0.04 ± 0.04	**0.09 ± 0.00**	-0.01 ± 0.01	0.01 ± 0.01	-0.01 ± 0.01	0.01 ± 0.01
FEC1	0.17 ± 0.02	0.13 ± 0.02	0.11 ± 0.02	-0.04 ± 0.05	0.05 ± 0.05	0.03 ± 0.03	**0.18 ± 0.01**	0.19 ± 0.00	0.22 ± 0.00	0.09 ± 0.01
FEC2	0.16 ± 0.02	0.15 ± 0.02	0.17 ± 0.02	-0.04 ± 0.06	0.10 ± 0.06	0.04 ± 0.03	0.74 ± 0.02	**0.19 ± 0.01**	0.10 ± 0.01	0.21 ± 0.00
NEM1	0.03 ± 0.02	0.08 ± 0.02	0.08 ± 0.02	-0.12 ± 0.05	0.01 ± 0.05	0.01 ± 0.03	0.41 ± 0.02	0.43 ± 0.03	**0.21 ± 0.01**	0.25 ± 0.01
NEM2	0.1 ± 0.03	0.11 ± 0.02	0.11 ± 0.02	-0.17 ± 0.06	-0.13 ± 0.06	0.02 ± 0.03	0.42 ± 0.03	0.44 ± 0.02	0.86 ± 0.02	**0.21 ± 0.01**
σ_p	4.07	4.51	0.42	1.28	1.43	0.62	0.77	0.76	0.65	0.56
Rep						0.12 ± 0.00	0.62 ± 0.00	0.57 ± 0.00	0.59 ± 0.00	0.49 ± 0.00

[1] WWT: weaning weight, LW8: live weight at 8 months, FW12: fleece weight at 12 months, DAG3, DAG8: dag score at 3 months and 8 months, NLB = number of lambs born in ewes, FEC1, FEC2: faecal egg count in summer and autumn, and NEM1, NEM2: analogously *Nematodirus* counts in summer and autumn (NEM1, NEM2).

[2]Direct effect estimate from average of bi-variate estimates (excluding LW8) the maternal effect estimate was 0.20± 0.00.

Table 3.4: Estimates of permanent environmental correlations (above diagonal) and genetic plus permanent environmental correlations (below diagonal) ± s.e.

Trait[1]	WWT	LW8	FW12	DAG3	DAG8	NLB	FEC1	FEC2	NEM1	NEM2
NLB	0.07 ± 0.01	0.06 ± 0.01	-0.05 ± 0.01	0.01 ± 0.03	0.04 ± 0.03		-0.13 ± 0.05	-0.00 ± 0.06	-0.13 ± 0.05	0.06 ± 0.07
FEC1	0.09 ± 0.01	0.07 ± 0.01	0.06 ± 0.01	-0.02 ± 0.03	0.03 ± 0.03	-0.04 ± 0.02		0.12 ± 0.01	0.24 ± 0.01	0.01 ± 0.02
FEC2	0.10 ± 0.01	0.09 ± 0.01	0.10 ± 0.01	-0.02 ± 0.03	0.06 ± 0.04	0.02 ± 0.02	0.31 ± 0.01		0.03 ± 0.02	0.21 ± 0.01
NEM1	0.02 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	-0.07 ± 0.03	0.01 ± 0.03	-0.05 ± 0.02	0.30 ± 0.01	0.17 ± 0.01		0.20 ± 0.02
NEM2	0.06 ± 0.02	0.07 ± 0.01	0.07 ± 0.01	0.00 ± 0.04	-0.09 ± 0.04	0.03 ± 0.03	0.16 ± 0.01	0.30 ± 0.01	0.47 ± 0.01	

[1] Single measure traits: WWT: weaning weight, LW8: live weight at 8 months, FW12: fleece weight at 12 months, DAG3, DAG8: dag score at 3 months and 8 months. Repeated traits: NLB: number of lambs born in ewes, FEC1, FEC2: faecal egg count in summer and autumn, and NEM1, NEM2: analogously *Nematodirus* counts in summer and autumn (NEM1, NEM2).

The majority of the phenotypic correlations were less than 0.2 except for WWT with LW8, WWT and LW8 with FW12, DAG3 with DAG8, and among some of the FEC/NEM traits. The remaining correlations were close to zero for DAG8. The DAG3 trait was negatively correlated with WWT and LW8, and except with DAG8 all other

correlations were close to zero. FEC and NEM traits had correlations close to zero for all traits, except FEC1 and FEC2 with WWT.

Permanent environmental correlations were calculated among NLB, FEC and NEM traits (Table 3.4). Correlations among FEC and NEM traits were low and positive. Correlations were higher for traits measured at the same time (FEC1 with NEM1, FEC2 with NEM2), than with similar traits (FEC1 with FEC2, NEM1 with NEM2), while those between different traits at different times were negligible. Genetic (g) plus permanent environmental (ep) correlations (i.e., correlation of g + ep for pairs of repeatable traits, or g + ep with g for repeatable with non-repeatable traits) were measured for NLB, FEC and NEM traits with all other traits (Table 3.4). These are between trait analogues of the repeatability for a single trait. These correlations were low except among FEC and NEM traits.

3.5 Discussion

This study provides estimates of genetic parameters for a number of production and disease traits in the New Zealand dual-purpose sheep. The large dataset of more than 2 million records enabled some unique analyses to be run, with estimates of correlations not previously presented and to address some common industry perceptions. These include; that there were strong positive genetic and phenotypic associations between the dagginess of an individual and FEC/NEM, and between FEC/NEM and live weight gain.

3.5.1 Data cleaning, model fitting

Breed percentage was fitted as a covariate for all traits, to account for potential effects of breed admixture. However, heritability and covariance estimates varied by less than 1 percent, depending on whether breed was fitted or omitted, and thus breed was discarded from the model and unadjusted estimates are reported. This suggests that the variation within breeds is sufficiently similar and breeds are linked through the industry, possibly by the wide uptake of composite breeds. Romney, Coopworth, Perendale and Texel make up 93% of the total breed composition of this dual-purpose industry dataset. Coopworth and Perendale are fixed interbred crosses containing 50%

Romney, thus the small change in estimates when breed was fitted as a covariate was expected.

The DAG3 and DAG8 traits have a high coefficient of variation. This is useful in that rapid genetic change can be made, but raises the issue of whether appropriate scaling has been used so as to allow valid comparisons across contemporary groups. Of particular concern is a potential relationship between observed variability and the contemporary group mean. To address this, 3 methods of scaling/cleaning were investigated; log transformation, arcsin transformation, and censoring contemporary groups by their means. The last method generated the best fit to normality. This method censored those contemporary groups with very low dag score, where the trait did not have sufficient time to be expressed before measurement and observed variability was lower than normal. The implication for this is that EBV's generated for these animals will depend entirely on measurements in relatives, reducing the accuracy of the breeding values in the animals concerned and reducing rate of genetic gain in the flock. This approach encourages the breeder to allow sufficient time for the trait to be expressed before measurement.

3.5.2 Comparison to current SIL estimates

Current SIL parameters are derived from estimates published more than 15 years ago (Bisset et al., 1992; Johnson et al., 1989; McEwan et al., 1995). Table 3.5 shows a summary of the estimates published from these studies. Actual estimates used in SIL vary slightly from those published due to "bending" (Hayes and Hill, 1981) of the estimates and weighting estimates for the number of sires used in the various analyses (J. C. McEwan, Pers. Comm.).

Compared to current SIL values, the present study has higher estimates for heritability of WWT maternal, DAG3, DAG8, WWT direct, and LW8 whereas heritabilities of FW12, NLB, FEC1, FEC2, NEM1 and NEM2 have stayed the same or fall within the range of past estimates.

Genetic and phenotypic correlations between the production traits WWT, LW8 and FW12 have mainly decreased from previous estimates calculated by Johnson et al. (1989). The exception is the genetic correlation between WWT and LW8 which has increased (from 0.8 to 0.91). The genetic correlation between LW8 and NLB has

increased slightly (from 0.05 to 0.07), while phenotypic correlation has decreased (from 0.1 to 0.01).

Table 3.5: Source estimates of current SIL parameters

Trait[1]	WWT	LW8	FW12	NLB	FEC1	FEC2	NEM1	NEM2	DAG
WWT	**0.12**[2]	0.8[2]	0.3[2]		-0.01 - 0.01[3]	-0.02 - 0.04[3]	-0.01 - 0.00[3]	-0.04 - 0.00[3]	
WWTM	**0.18**[2]								
LW8	0.8[2]	**0.25**[2]	0.35[2]	0.1[2]	-0.05 - -0.03[3]	-0.03 - -0.01[3]	0.00 - 0.03[3]	-0.05 - 0.02[3]	
FW12	0.2[2]	0.25[2]	**0.3**[2]		-0.07 - -0.02[3]	-0.83 - 0.02[3]	0.02 - 0.04[3]	-0.04 - 0.06[3]	
NLB		0.05[2]		**0.1**[2]					
FEC1	0.06 - 0.56[3]	0.15 - 0.33[3]	0.02 - 0.14[3]		**0.15 - 0.21**[3]	0.21 - 0.22[3]	0.19 - 0.24[3]	0.04 - 0.07[3]	
FEC2	-0.03 - 0.28[3]	0.04 - 0.18[3]	0.05 - 0.25[3]		0.55 - 0.84[3]	**0.17 - 0.42**[3]	0.08 - 0.21[3]	0.17 - 0.25[3]	
NEM1	-0.01 - 0.20[3]	0.13 - 0.41[3]	0.02 - 0.49[3]		0.24 - 0.44[3]	0.05 - 0.41[3]	**0.17 - 0.24**[3]	0.24 - 0.32[3]	
NEM2	-0.28 - 0.34[3]	-0.01 - 0.26[3]	-0.00 - 0.26[3]		0.11 - 0.34[3]	0.04 - 0.65[3]	0.79 - 0.84[3]	**0.17 - 0.30**[3]	
DAG									**0.24**[4]

[1] WWT: weaning weight, LW8: live weight at 8 months, FW12: fleece weight at 12 months, DAG3, DAG8: dag score at 3 months and 8 months, NLB: number of lambs born in ewes, FEC1, FEC2: faecal egg count in summer and autumn, and NEM1, NEM2: analogously *Nematodirus* counts in summer and autumn (NEM1, NEM2).

[2] Johnson et al. 1989.

[3] McEwan et al. 1995 (3 study populations, range is shown here).

[4] Bisset et al. 1992.

Heritabilities of FEC and NEM traits fall within the range of past estimates calculated from McEwan et al. (1995). Genetic and phenotypic correlations between production traits WWT, LW8, FW12 and FEC and NEM traits are very similar or slightly larger than previous estimates. Correlations between FEC and NEM at different times fall within range of past estimates or are larger, except the phenotypic correlation for FEC1 versus FEC2 which is slightly lower than past estimates.

3.5.3 Comparison to other studies

The heritabilities for WWT, LW8 and FW12 are similar to previously published estimates in European dual-purpose sheep breeds (Bisset et al., 1992; Brash et al., 1994b, b; Conington et al., 1995; Wuliji et al., 2011). For DAG3 and DAG8, heritability estimates are moderate (0.31 to 0.34). Previously published heritability estimates for dagginess from New Zealand Romney and Perendale based flocks range from 0.24 to 0.34 (Bisset et al., 1992; Scobie et al., 2008) and from Australian Merino flocks range from 0.09 to 0.55 (Greeff and Karlsson, 2009; Smith et al., 2009).

The heritability estimate for NLB is slightly greater than previously published estimates in New Zealand and Australian dual-purpose breeds (Brash et al., 1994b, b; Davis et al., 1998). Heritability estimates of resistance to internal parasites, as measured by faecal egg counts, are slightly lower than previously published results (McEwan et al., 1992b; Morris et al., 1997; Pollott et al., 2004).

Correlations of DAG3 and DAG8 with FEC/NEM traits are of particular interest, due to the industry held view that there is a strong positive association between the 2 sets of traits. The current study does not support this view, with genetic correlations ranging from -0.17 to 0.10, and phenotypic correlations close to zero (-0.05 to 0.03). Previous New Zealand genetic correlation estimates are: 0.45±0.19 (Bisset et al., 1992), -0.70 (Watson et al., 1986), -0.26±0.24 (Baker et al., 1991), and -0.03 (Douch et al., 1995), with phenotypic correlations of 0.11 ± 0.03 (Bisset et al., 1992), -0.18 (Watson et al., 1986), -0.09 (Baker et al., 1991), and -0.13 (Douch et al., 1995). The large variation in the estimates previously reported, often coupled with large standard errors, has led to misinterpretation of the correlation. It is of interest to note the prior studies involved only a small sampling of animals from a few industry and selection line research flocks. The current study used a whole industry dataset consisting of greater than 90,000 DAG measurements, and over 100,000 measurements for each of the FEC/NEM traits.

WWT and LW8 were highly correlated, in line with previously published reports in dual-purpose breeds (Clarke et al., 2000). Genetic and phenotypic correlations of NLB with WWT, LW8 and FW12 are similar to the weighted means calculated from literature estimates by Safari et al. (2005). No known estimates have been previously published for correlations of NLB with DAG, FEC, and NEM traits.

Live weight correlations with FEC/NEM traits were positive, but low genetically and phenotypically. These estimates contrast with those published previously by Bisset et al. (1992) and Douch et al. (1995), which range from -0.29 ± 0.22 to -0.01 ± 0.02. These earlier studies involved a small dataset, however, all correlations from these earlier studies were close to zero or had large standard errors.

3.5.4 Repeatability and permanent environmental effects

Repeatability is the intra-class correlation between repeated records of an individual. The permanent environmental variance is the variance component that is constant across repeated measures but is not part of the additive genetic variance. In the current analytical model, this term is comprised of permanent environmental, non-additive genetic (dominance and epistasis), imprinted, and possibly epigenetic effects.

The estimated repeatability of NLB measurements was 0.12. This value was only slightly higher than the heritability estimate of 0.09 and suggests that any similarity between repeated measurements was mainly due to additive genetic effects. The repeatability estimates of the FEC and NEM traits, however, range between 0.49 and 0.62 while the heritability estimates were 0.18 to 0.21. The repeatability measure was for the same parasitic challenge, therefore the greater repeatability estimate, compared to the heritability estimate, is an indication of any permanent (within the time period measured) environmental and non-additive genetic effects.

In some cases, permanent environmental sources of variation affect multiple traits resulting in a significant correlation. In the current work, these values were low to moderate and positive between the FEC and NEM traits, suggesting that a common underlying non-additive-genetic cause affects all of these traits in the same direction. In contrast, the permanent environmental correlations were low or negative between NLB and FEC1 and NEM1. This suggests that a common underlying non-additive-genetic cause has economically beneficial effects on these traits. Correlations between the

genetic plus permanent environmental effects for the FEC and NEM traits were intermediate to the genetic correlations and permanent environmental correlations.

3.5.5 Comparison between DAG models

A comparison between the current SIL dagginess model and the DAG3 and DAG8 models fitted in this study was undertaken. Table 3.6 summarises the results of the comparison of the current and new dagginess models across 3 exemplar flocks. Plotting the mean for each contemporary group (flock-year-sex) after normalisation, versus standard deviation, shows that the new models drastically reduce the variation in standard deviation (Figure 3.1), resulting in EBVs that also have markedly reduced variability (Figure 3.2). The standard deviation for the majority of contemporary groups range between 1 and 1.4 for the new models, while for the current SIL model the majority of groups ranged between 2 and 4, with a maximum standard deviation of 7. The new models more accurately reflect the untransformed values for DAG3 and DAG8 and suggest that the previous scaling transformation was inappropriate. There was also a large change in EBV and rank for each flock. Correlations (r) between the current model and new DAG3 and DAG8 models range between 0.62 and 0.77 (Table 3.6).

Table 3.6: Summary of estimated breeding values (EBVs) results for 3 flocks run with the current Sheep Improvement Limited (SIL) and new dag (DAG3, DAG8) models.

Flock	No.	Run	EBV average	[1]EBV max	[1]EBV min	[1]EBV SD	[2]EBV r^2	[2]Rank r^2
A	33478	SIL	-0.16	8.74	-3.78	0.90		
		DAG3	-0.10	1.52	-0.99	0.27	0.77	0.75
		DAG8	-0.10	1.33	-1.05	0.26	0.77	0.74
B	29120	SIL	0.01	2.79	-1.24	0.28		
		DAG3	-0.01	1.15	-0.67	0.12	0.60	0.60
		DAG8	-0.02	1.15	-0.77	0.12	0.62	0.65
C	23376	SIL	-0.45	41.07	-4.56	1.04		
		DAG3	-0.02	1.40	-1.12	0.21	0.74	0.69
		DAG8	-0.08	1.46	-1.28	0.29	0.66	0.62

[1]EBV max: maximum EBV; EBVmin: minimum EBV; SD: standard deviation.

[2]Correlation with SIL EBV estimates or rank.

It is not surprising that there is a large difference in the EBVs between the current SIL and new models presented here for dagginess. This is due to 3 factors; firstly the current SIL model uses a repeated trait model, treating DAG3 and DAG8 as the same trait, measured at different times. The models used in this study separate DAG3 and DAG8 by treating them as different traits, DAG3 being a pre-weaning measure and DAG8 a post-weaning measure, both differing in environmental and animal conditions. Secondly, the current SIL model uses data that have been scaled by the mean of its contemporary group, while the method in this analysis used a threshold approach, discarding measurements from contemporary groups with very low mean dag scores. This is emphasised by flock C, where using the current SIL method a maximum EBV of 41.07 was observed. This animals' own record was subsequently censored in this study due to low contemporary group mean and thus contributed to the dramatic difference between current SIL and new DAG3 and DAG8 EBVs for this flock.

Lastly, the models and genetic parameter estimates calculated using the industry dataset are considerably different to those currently used in SIL. The combination of all of these differences explains why the current and new BVs are not very well correlated in any of the 3 flocks, with considerable re-ranking, and large differences in standard deviation between the current and new models.

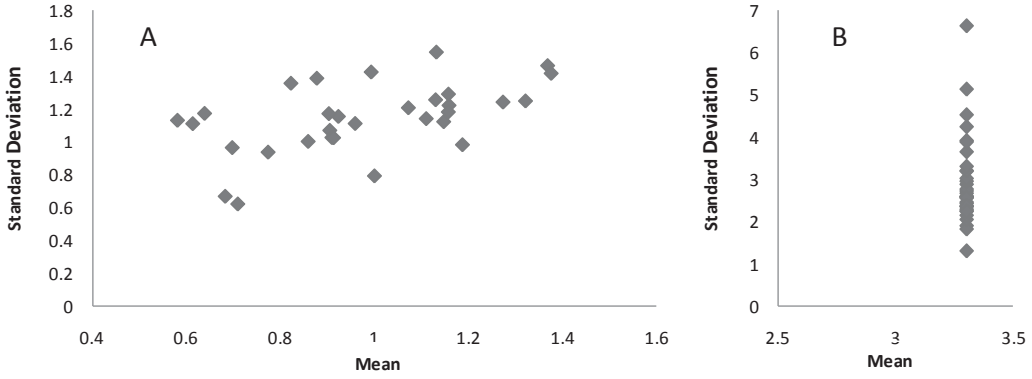


Figure 3.1: Plot of standard deviation versus mean for contemporary groups (flock-year-sex) for new dag score at 3 months model (A) and current Sheep Improvement Limited dagginess model (B), after scaling for all 3 flocks tested.

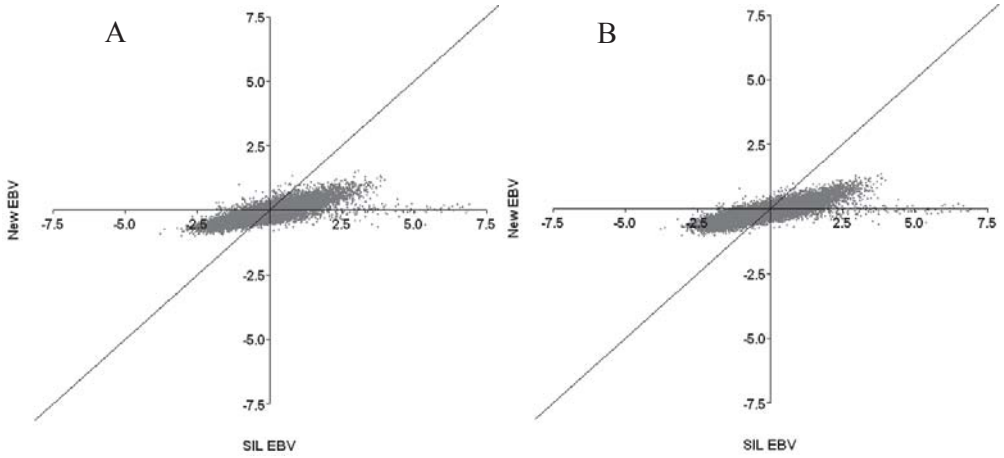


Figure 3.2: Plot of current Sheep Improvement Limited dagginess EBV value (x axis) versus new dag score at 3 months (DAG3; A), and at 8 months (DAG8; B) EBV values (y axis) for flock A.

3.5.6 Implications of current estimates

Selection against dagginess in the economically desirable direction would mostly cause low and favourable responses in other traits. The negative genetic and phenotypic correlations of DAG3 with WWT and LW8 mean that less dagginess is associated with a greater live weight. Selection on DAG8 will have little or no effect on WWT or LW8. Both DAG3 and DAG8 have positive genetic correlations with FW12; this is suggestive of longer fleeces attracting more dagginess. This could be classed as unfavourable; however, the genetic and phenotypic correlations with FW12 are close to zero and selection against dagginess should have little effect on FW12.

The DAG3 trait has low negative genetic correlation with NEM1 and NEM2, whereas DAG8 has low negative genetic correlation with NEM2. Phenotypic correlations among DAG3 and DAG8 with FEC and NEM traits are negative and close to zero. Selection on DAG traits will have little effect on internal parasite load. Thus, in conclusion, there is little if any effect of internal parasite load on dag score. Also in general, relationships with productive traits and parasite measures are close to zero and best addressed via index selection.

The majority of genetic and phenotypic correlations presented here with NLB have not been published before in a New Zealand industry dataset. Most of these correlations are close to zero and positive, negative only for FW12. The genetic and phenotypic

estimates calculated indicate there would be negligible impact on NLB if selection pressure was placed on these traits.

3.6 Conclusions

The genetic parameters estimated in this study are based on the largest available New Zealand industry dataset. It is anticipated that these estimates, and in some cases new analytical models, will be used to update New Zealand genetic evaluations. The impact that the new estimates and models will have on current index values was shown in the extreme by the comparison of the dag BVs. Specifically, this work in retrospect identified an inappropriate contemporary group variance adjustment. The new DAG3 and DAG8 models have reduced the variation in the dagginess index by half on average and for one industry flock the resulting BVs were only 0.2 standard deviations of the original BVs. The differences for the other traits will be much less marked, but important nonetheless. This study also removes any industry doubts that the genetic parameters used are inappropriate and out of date for the current New Zealand dual-purpose flocks. These concerns have been raised, because the industry is increasingly a complex mixture of breeds. This study also demonstrates that the genetic relationships between dagginess, parasite traits and production traits are generally low and of little material impact when index selection is used. Finally, we have reported for the first time, genetic correlations for dags and parasite traits with NLB.

Chapter 4: Genetic relationships between dagginess, breech bareness and wool traits in New Zealand dual-purpose sheep

Foreword: This chapter has been re-submitted to Journal of Animal Science after second review (July 2012). It is as submitted with formatting changes to fit in with the thesis layout.

4.1 Abstract

Genetic and phenotypic parameters were estimated for dagginess, breech, wool and fibre traits from approximately 29,500 progeny born in 2009 and 2010 in New Zealand dual-purpose ram breeding flocks. Dagginess is adherence of faecal matter to the wool and this study investigates the genetic and phenotypic correlations between dagginess, and breech and wool traits. Estimates for heritability were moderate (0.21 to 0.44) for the following traits: dag score at 3 and 8 months (DAG3, DAG8), breech bareness (BBREECH), wool length (LENGTH), wool bulk (BULK), mean fibre diameter (MFD), mean fibre diameter standard deviation (FDSD), mean fibre diameter coefficient of variation (FDCV), curvature (CURV), weaning weight at 3 months (WWT), and autumn live weight (LW8). Heritability estimates for fleece weight at 12 months (FW12) and proportion of medullated fibres (MED%) were high (0.49 and 0.53 respectively). Dag score at 3 months and DAG8 had low genetic and phenotypic correlations with all traits. Breech bareness had positive genetic and phenotypic correlations with CURV and BULK, and mostly negative genetic correlations with all other wool traits. In summary, the quantity and attributes of wool were not primary causative factors in faecal accumulation, leaving faecal consistency and composition as the major factors.

4.2 Introduction

Dagginess or faecal soiling is known to be associated with flystrike (Greeff and Karlsson, 2009; Smith et al., 2009). With the majority of strike occurring in the breech in New Zealand sheep (Heath and Bishop, 1995), a reduction in dagginess could have an effect on reducing flystrike. For the prevention of flystrike a detailed understanding of the causes of dagginess is required.

Dagginess can arise from factors affecting consistency of the faeces and factors affecting adherence to the breech (Broughan and Wall, 2007b; Wall et al., 2000). Factors affecting consistency can be due to worm burden, immune response, feed composition or intestine function (Bedrich and Ehrlein, 2001; Bisset et al., 1996; Davidson et al., 2006; Larsen et al., 1999; Leathwick and Atkinson, 1995; Pownall et al., 1993; Ramirez-Restrepo et al., 2004; Ruckebusch and Fioramonti, 1980). Factors

affecting adherence have been shown to be related to breech bareness, wool length and type in some studies (French et al., 1998; Scobie et al., 2008).

It has already been shown that there was no genetic or phenotypic correlation between dagginess and faecal egg counts in New Zealand dual-purpose sheep (Chapter 3). A study on digestibility and intestinal activity would involve in depth and potentially invasive measurement techniques, which is not feasible on the large numbers of animals needed for the accurate estimation of genetic parameters collected in this study. This leaves breech bareness, wool type and related wool traits to be investigated for their genetic and phenotypic association with dagginess. The concentration on wool traits also has the added benefit in that several wool traits have also in turn been independently implicated in flystrike susceptibility (Greeff and Karlsson, 2009; McGuirk et al., 1978; Rathie et al., 1994; Smith et al., 2009).

This study aimed to collect measurements on dagginess, breech, fibre, fleece and live weight traits using a progeny test design in the New Zealand dual-purpose sheep industry. The data were used to calculate heritabilities and genetic and phenotypic correlations. The New Zealand dual-purpose sheep industry is dominated by Romney, Coopworth, Perendale, Texel, and composite crosses of these breeds.

4.3 Material and methods

4.3.1 Experimental design

Animal ethics approval for this experiment was obtained from the AgResearch Animal Ethics Committee. A progeny test design was used for this study. Twenty-one sheep breeders, with a total of 35 recorded flocks, were recruited into the study with all flocks linked by common sires. Progeny (n = 29,419) born in 2009 and 2010 were assessed representing approximately 700 sires. Progeny were measured for dag score, breech bareness, wool length, wool bulk; a proportion (~10%) of animals also had wool samples collected (as described in the measurements section). The progeny were measured once for the above traits, between weaning and live weight at 8 months, during the 2009/10 and 2010/11 seasons. All progeny measured as part of the progeny test, across all farms, were measured by the same person. Additional pedigree and performance information was obtained from Sheep Improvement Limited (SIL).

4.3.2 Experimental sites

Farms (n = 21) were located across New Zealand, and involved sheep breeders registered on SIL (Figure 4.1). In the North Island 5 farms were located in the Wairarapa and 3 farms located in Manawatu-Wanganui. There was 1 farm in the Waikato, and Hawkes Bay regions respectively. In the South Island the majority of farms (n = 9) were from the Southland region. One farm was also involved from each of the Otago and Canterbury regions. The majority of flocks were of Romney base; however, there were 4 Perendale, 2 Texel, 1 Coopworth, and 1 Dorset breed-based flocks.

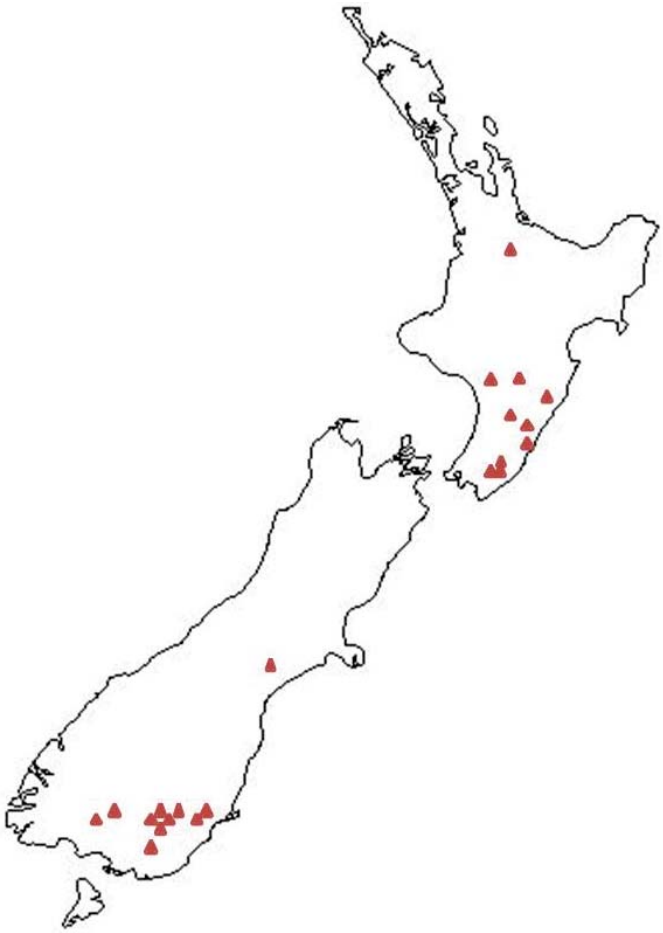


Figure 4.1: Map of experimental sites; properties involved in the progeny test (▲).

4.3.3 Measurements

At weaning (3 months of age) or post-weaning (live weight at 6 or 8 months of age), measurements were taken on the progeny. The measurements taken in the field were:

dag score (DAG3 at weaning, DAG8 at post weaning (Chapter2)), breech bareness score (BBREECH (Scobie et al., 2007)), wool length (LENGTH), and hand assessed wool bulk (BULK). Dag score was subjectively assessed on a 0 (no dags) to 5 (complete dags over the tail, under the crutch, and down the legs) scale (see Figure 1.4). Breech bareness was subjectively assessed on a 1 (wool up to and covering the anus) to 5 (no wool around the breech and down the leg) scale (see Figure 1.5). Wool length was measured in millimetres and reflects the relaxed length of the staple under no tension on the spine, above the last rib. Wool bulk was subjectively hand assessed, the wool was compressed, and the level of resistance scored from 1 (no structure, no resistance) to 4 (wool is hard to compress, strong resistance).

To obtain an indication of the quality of fleeces, 30 to 50 wool samples were collected from each flock. This involved snipping a sample of 5 to 7 staples from the mid-side on the anterior-posterior axis (Craven et al., 2009). The samples were sent to New Zealand Wool Testing Authority (NZWTA) in Napier, New Zealand, for analysis on the Optical Fibre Diameter Analyser 100 (OFDA 100, IWTO-47) (Baxter et al., 1992). At the beginning of the trial (2009) the first 550 samples taken were measured at Massey University on the OFDA 100. Samples were assessed for mean fibre diameter (MFD, μm), MFD standard deviation (FDSD, μm), MFD coefficient of variation (FDCV, %), proportion of medullated fibres (MED%), and curvature (CURV, °/mm). All on-farm measurements (dag score, bare breech, wool length, wool bulk, and staple sample) were taken by the same person throughout the experiment. The same person also processed and measured the wool samples at Massey University; they did not process the samples at NZWTA.

Additional data were obtained from SIL, including data from contemporaries of the animals measured from the flocks and birth years involved in this study. These data included parentage, birth and rearing rank, age of dam, birth day and grazing mob history, as well as weaning weight at 3 months (WWT), live weight at 6 and 8 months (LW6, LW8), and fleece weight at 12 months (FW12). Dag score at 3 and 8 months (DAG3, DAG8) were also included so as to incorporate any dag measurements recorded by the breeder on other mobs or at a different time to when data collection took place. It is to be noted that all lambs are crutched; removing dags and wool from the breech, within a month of weaning, thus resetting the dag score back to zero before

measurement of DAG8. A summary of raw data available on the 2009 and 2010-born progeny is shown in Table 4.1.

From model fitting the maternal effect was significant for WWT and autumn live weight (generated from LW6 and LW8 data; see below). Dam information was needed to correctly estimate maternal effects. This was demonstrated by Maniatis and Pollott (2003), who stated that as the proportion of dams without records increases there was an increased bias in estimates of maternal genetic effects. To complete the pedigree of our data set, additional pedigree and performance data on the dams and their contemporaries were downloaded from SIL so that maternal and direct genetic effects could be separated for these traits.

Table 4.1: Summary of 2009/2010 progeny dataset before data cleaning, transformation and addition of dam information.

Trait[1]	n	Mean	SD	Minimum	Maximum
DAG3	17,215	1.1	1.2	0	5
DAG8	25,498	1.1	1.4	0	5
BBREECH	28,814	1.9	0.9	1	5
LENGTH (mm)	25,420	57.7	16.3	25	145
BULK	25,406	2.4	1.0	1	4
MFD (µm)	1623	33.4	3.8	24.2	48.5
FDSD (µm)	1452	8.6	1.3	5.2	15.4
FDCV (%)	1452	26.8	6.7	15.8	78.2
MED%	1378	8.1	11	0.1	63.9
CURV (°/mm)	1452	48.2	9.9	21.3	90.1
WWT (kg)	70,763	27.7	5.8	6.6	54
LW6 (kg)	36,823	41.0	7.8	11.5	76
LW8 (kg)	16,096	40.1	6.4	11.2	74.8
FW12 (kg)	8,105	3.1	1.1	0.74	6.9

[1] DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; LENGTH: length of relaxed wool staple on spine; BULK: wool bulk; MFD: mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; MED%: proportion of medullated fibres; CURV: curvature; WWT: weaning weight at 3 months; LW8: autumn live weight; FW12: fleece weight at 12 months.

4.3.4 Data cleaning

Data were subjected to quality checks, for example, to ensure that subjectively assessed trait values were within the ranges expected, and traits were scaled as appropriate to homogenise the variance (Brown et al., 2005). To homogenise the variance, 3 transforming methods were tested and the most suitable method chosen for

each trait: $\log_{10}$, scaled to the mean of contemporary group, or arcsin on the score divided by the maximum score. A reduced variation was identified for contemporary groups (DAG3: flock by birth year by sex by DAG3mob, DAG8: flock by birth year by sex by DAG8mob) having dag scores below a certain threshold. To mitigate this, log and arcsin transformations were tested; however the best method to homogenise the variance was to delete dag observations from those contemporary groups with mean dag score below 0.5 for DAG3 and 0.25 for DAG8. This discarded mobs that had little variation in dag score at the time of measurement; it was assumed that these contemporary groups were measured before the trait could be fully expressed. For DAG3, 23% and for DAG8, 25% of the data were removed from the final data set.

For BBREECH, contemporary group **(**flock by birth year by sex by BBREECHmob), with a low mean, had little variation. Thus, as with DAG3 and DAG8, contemporary groups with mean BBREECH score less than 1.2 had their BBREECH data deleted. This excluded 1 purebred Romney flock which had no variation in their flock for this trait. For LENGTH, data were scaled via means of the contemporary group (flock by birth year by sex by LENGTHmob) to the overall mean of the dataset. No scaling was applied for BULK.

For the wool fibre traits; MFD, FDSD, and MED% were all scaled via means of the contemporary groups (flock by birth year by sex by MFDmob) to the overall average of the dataset. For MED%, observations from contemporary groups with mean MED score greater than 20 were deleted as they were outliers. The wool fibre traits of FDCV, and CURV were not scaled.

For WWT, data were scaled via means of the contemporary group (flock by birth year by sex by wwtmob) to the overall average of the data set. Either LW6 or LW8 was used as the autumn live weight (LW8), taking the weight with the maximum amount of data for the contemporary group (flock by birth year by sex by LW8mob by LW6mob by WWTmob). The new LW8 were then scaled via means of the new contemporary group (flock by birth year by sex by LW8mob by WWTmob) to the overall average of the dataset. For FW12, data were scaled via means of the contemporary group (flock by birth year by sex by FW12mob by LW8mob by WWTmob) to the overall average of the data set. To offset the differences in age at measurement, birth day deviation from the mean of the contemporary group was used as a covariate for all traits analysed.

4.3.5 Statistical analysis

For each trait, the following processes were followed to derive the final analytical model. Fixed-effect models were initially determined using the general linear model procedure (SAS, 2004). Fixed effects fitted included; flock (flk), sex (F or M), birth-rearing rank (brr, born single, twin, or triplet, reared as single, twin, or triplet), site of OFDA100 measurement (NZWTA or Massey University) and grazing mob (trait mob), or contemporary group (mainly flk by birth year by sex by traitmob). Birth day deviation (bdev) from the mean of contemporary group (flock by birth year by sex by WWTmob) and age of dam as linear (aod) and quadratic (aod^2) effects were fitted as covariates. Interactions between these effects were tested and discarded from the final model via backwards elimination if not significant. Regression coefficients and fixed effect contrasts were obtained from the final model retained.

To assess the significance of maternal (random) and breed (covariate) effects, they were fitted in uni-variate models for each trait using ASReml (Gilmour et al., 2009). Breed effect consisted of 4 covariates (romperc, coopperc, perenperc, texelperc), each calculating the proportion of a breed (Romney, Coopworth, Perendale, or Texel, respectively) in the animal. A maternal genetic effect was retained only for WWT and LW8. Breed covariates were discarded for all traits except FW12 as they influenced genetic variance estimates by less than 2 percent. Traits were then fitted in bi-variate analyses with ASReml using full animal random effects as well as maternal effects where appropriate. Genetic and phenotypic parameter estimates and their standard errors were obtained by averaging across all relevant bi-variate runs.

4.3.6 Selection response

To estimate the direct and correlated response in dagginess explained by DAG3, DAG8, BBREECH, fibre and wool traits, the multi-trait selection tool provided by van der Werf (2006a) was used, with the objective of reducing dagginess. The heritabilities, genetic and phenotypic correlations of the traits estimated in this chapter were used to estimate the selection response for different scenarios. In scenario 1, 1 measurement on each of DAG3 and DAG8 were taken. For scenario 2, 1 measurement was taken on each of BBREECH, LENGTH, BULK, MFD, FDSD, FDCV, CURV, and MED%. While in scenario 3 there was 1 measurement on each of the 10 traits. The selection response for DAG3 and DAG8 was predicted for each scenario.

4.4 Results

4.4.1 Measurements

Table 4.1 shows a summary of the traits measured, before data cleaning, for both the 2009- and 2010-born progeny. The final number of LW8 measurements (which were generated from LW6 and LW8 data; see above) was 46,212 (av: 41 kg, SD: 7.3). For WWT and LW8, which have maternal components, measurements and pedigree information of the dams and their contemporaries were included. This brought the total number of WWT and LW8 measurements to 287,247 and 191,109 respectively.

4.4.2 Model fitting

Table 4.2 describes the final linear models applied for each characteristic. Table 4.3 shows means and an ANOVA summary for the traits examined. No fixed effects were significant for MED%. Age of dam (both linear and quadratic) were significant for production and visual traits, WWT, LW8, FW12, DAG3, DAG8, BBREECH and BULK, ($P < 0.05$ to $P < 0.001$). Birthday deviation was significant for all traits except DAG8 and FDCV. Birth-rearing rank was significant for each trait except LENGTH. Each level of birth-rearing rank was compared against lambs born and reared as triplets. The traits WWT and LW8 were affected by birth-rearing rank ($P < 0.001$). However, for all other traits birth-rearing rank was not significant ($P > 0.05$) for most comparisons. Contemporary group fixed effects (mainly of flk by birth year by traitmob) were significant for all traits except for MFD, FDCV and MED%.

Breed percentage was fitted as a covariate for all traits; however, it was only significant for FW12. Heritability for fleece weight changed from 0.52 to 0.46 when breed was omitted from the uni-variate run. Heritability estimates varied by less than 2 percent, depending on whether breed was fitted or not, for traits other than FW12 and thus breed was discarded from the model for these traits.

Table 4.2: Final mixed models and fixed effects including contemporary groups used for individual trait analysis.

Trait[1]	Fixed Effects[2]	Random Effects	Contemporary group (cg)[3]
DAG3	sex brr aod bdev cg	Animal	flk.byr.DAG3mob
DAG8	sex brr aod bdev cg	Animal	flk.byr.DAG3mob.dag8mob
BBREECH	brr aod bdev cg cg.brr	Animal	flk.byr.sex.BBREECHmob
LENGTH	aod bdev cg.brr	Animal	flk.byr.sex.LENGTHmob
BULK	brr aod bdev cg cg.brr	Animal	flk.byr.sex.BULKmob
MFD	brr bdev ofda cg	Animal	flk.byr.sex.MFDmob
FDSD	brr bdev ofda cg cg.brr	Animal	flk.byr.sex.MFDmob
FDCV	brr ofda cg	Animal	flk.byr.sex.MFDmob
MED%	no model	Animal	
CURV	brr bdev ofda cg	Animal	flk.byr.sex.MFDmob
WWT	brr aod bdev cg cg.brr	Animal, Maternal	flk.byr.sex.WWTmob
LW8	brr aod bdev cg cg.brr	Animal, Maternal	flk.byr.sex.LW8mob.WWTmob
FW12	brr aod bdev breed[4] cg cg.brr	Animal	flk.byr.sex.FW12mob.LW8mob.WWTmob

[1] DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; LENGTH: length of relaxed wool staple on spine; BULK: wool bulk; MFD: mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; MED%: proportion of medullated fibres; CURV: curvature; WWT: weaning weight at 3 month; LW8: autumn live weight; FW12: fleece weight at 12 months.

[2] brr: birth rearing rank; aod: age of dam as linear and quadratic; bdev: birthday deviation; ofda: site where fibre measurement done.

[3] flk: flock; byr: birth year; DAG3mob: DAG3 grazing mob; DAG8mob: DAG8 grazing mob; BBREECHmob: BBREECH grazing mob; LENGTHmob: LENGTH grazing mob; BULKmob: BULK grazing mob; MFDmob: MFD grazing mob; WWTmob: WWT grazing mob; LW8mob: LW8 grazing mob; FW12mob: FW12 grazing mob.

[4] Breed: romperc, coopperc, perenperc, texelperc.

Table 4.3: Means and ANOVA summary for traits[1] after cleaning, transformation and addition of dam information for WWT and LW8.

Item	DAG3	DAG8	BBREECH	LENGTH	BULK	MFD	FDSD	FDCV	MED%	CURV	WWT	LW8	FW12
No. Obs.	13225	6173	26337	24210	24209	1271	1271	1271	1134	1271	284916	190174	7931
Mean	1.22	1.19	1.95	57.65	2.36	33.3	8.6	25.8	8.1	48.69	28.4	40.02	3.02
Residual StD	1.23	1.33	0.76	9.15	0.78	2.54	1.1	2.59	9.67	6.58	4.07	4.17	0.39
CV%	100.99	111.51	39.11	15.87	32.94	7.62	12.81	10.05	119.4	13.52	14.35	10.41	13.04
R Square	0.15	0.11	0.31	0.08	0.37	0.17	0.23	0.54	0	0.54	0.36	0.67	0.13
Fixed Effects	0.13^2	0.10^2	0.09^2		0.08^2	0.10^2	0.04^2	0.33^2	0.00^5	0.27^2	0.01^2	0.12^2	0.01^2
			0.02^3	0.03^3	0.01^3		0.18^3				0.02^3	0.13^3	0.05^3
Contrasts[4]													
brr 11	-0.01 ns	0.13 ns	0.13 ns		0.23 ns	1.62 ***	-3.74 **	-0.50 ns		-3.00 *	12.88 ***	7.72 ***	-0.26 ***
brr 12	1.07 **	1.31 **	-0.44 ns		-0.46 ns						12.41 **	6.43 ns	-0.72 ns
brr 21	0.02 ns	0.03 ns	-0.00 ns		0.13 ns	2.09 ***	-0.01 ns	0.41 ns		-3.44 **	8.71 ***	4.18 ***	-0.18 ns
brr 22	0.01 ns	0.06 ns	0.00 ns		0.3 ns	0.67 *	-0.02 ns	-0.47 ns		-0.73 ns	4.41 ***	4.07 ***	0.08 ns
brr 31	-0.14 ns	-0.07 ns	-0.48 ns		-0.13 ns	1.61 *	1.13 ns	-0.31 ns		-1.90 ns	12.18 **	10.36 *	0.54 ns
brr 32	-0.09 ns	-0.02 ns	0.20 ns		-0.1 ns	0.95 *	-0.59 ns	0.16 ns		-1.04 ns	5.36 ***	3.90 **	-0.48 ns
brr 33	0	0	0		0	0	0	0		0	0	0	0
Covariates													
bdev	-0.01 ***	0.00 ns	-0.01 ***	-0.28 ***	0.00 ***	-0.06 ***	-0.02 ***			0.07 **	-0.19 ***	-0.17 ***	-0.01 ***
aod	0.13 ***	0.13 *	0.09 ***	0.52 ns	0.09 ***						1.66 ***	1.60 ***	0.05 ***
aod[6]	-0.01 **	-0.01 *	-0.01***	-0.06 ns	-0.01 ***						-0.18 ***	-0.17 ***	-0.01 **

[1] DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; LENGTH: length of relaxed wool staple on spine; BULK: wool bulk; MFD: mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; MED%: proportion of medullated fibres; CURV: curvature; WWT: weaning weight at 3 months; LW8: autumn live weight; FW12: fleece weight at 12 months.
[2] Proportion of variance explained by fixed effect contemporary group (see Table 4.1);
[3] Proportion of variance explained by fixed effect contemporary group and brr interaction (see Table 4.1);
[4] Contrasts birth rearing rank (brr) as born single (1), twin (2) or triplet (3) and reared as single, twin or triplet, compared to born and raised as a triplet (brr33)
[5] Contemporary group (flock by birth year by MFD grazing mob) shown to demonstrate no fixed effects significant for MED%.
[6] Age of dam as quadratic.
*** P <0.001, ** P<0.01, *P<0.05, ns not significant

4.4.3 Genetic parameters

Heritabilities, correlations and phenotypic standard deviations are presented in Table 4.4. The traits WWT and LW8 have moderate (0.21 ±0.01 to 0.23 ± 0.01, respectively) heritabilities. Maternal heritability estimates for these traits were 0.22 ± 0.005 and 0.13 ± 0.005, respectively. The genetic correlations between the direct and maternal effects for WWT were -0.45 ± 0.01 and between the direct and maternal effects for LW8 were -0.28 ± 0.02. The genetic correlation between WWT direct and LW8 maternal effects were -0.31 ± 0.02 and between WWT maternal and LW8 direct effects were -0.11 ± 0.02. These estimates are normally set to zero for SIL genetic evaluations, as non-zero values result in computationally exhaustive genetic evaluation runs. However, the correlation (r) between the resulting estimated breeding values when co-variances between maternal and direct were fitted, or not fitted, in this dataset, ranged between 0.93 and 0.96 for WWT, WWT maternal, LW8 and LW8 maternal. Therefore, there would be no significant re-ranking of animals when co-variances are not fitted. High heritabilities were estimated for FW12 and MED% (0.49 ± 0.05 and 0.53 ± 0.11, respectively), while the remaining traits had moderate heritabilities between 0.23 ± 0.09 to 0.44 ± 0.02.

Of the visually assessed traits, there was a high positive genetic correlation between DAG3 and DAG8 (0.74 ± 0.04) and between BBREECH and BULK (0.66 ± 0.03). There was a moderate and high negative correlation between BBREECH and LENGTH and between LENGTH and BULK, respectively. The traits DAG3 and DAG8 have a low negative genetic correlation with BBREECH and BULK and a low positive correlation with LENGTH.

There were moderate to high genetic correlations between wool fibre traits, MFD, FDSD, FDCV, and MED%. Curvature had negative correlations with the wool fibre

traits except with MED%, which was close to zero. There were moderately negative correlations between wool fibre traits and BBREECH except for CURV where there was a high positive correlation. Correlations of wool fibre traits were positive with LENGTH, except for CURV which had a high negative correlation.

Genetic correlations of wool fibre trairts were low or negative with BULK, DAG3 and DAG8, except for CURV, which had a high positive correlation with bulk. There was also a moderate positive correlation between FDSD and DAG3.

Weaning weight and LW8 have low and negative genetic correlations with all other traits except with BULK, BBREECH and MFD, which were moderately positive. There was a high correlation between WWT and LW8 (0.76 ± 0.01). Fleece weight at 12 months had moderate correlations with most traits except with DAG8, BBREECH, BULK, and WWT, and a moderate negative correlation with CURV.

Phenotypic correlations among visually assessed traits, DAG3, DAG8, BBREECH, LENGTH and BULK were close to zero except for a moderate positive correlation between DAG3 and DAG8 (0.41 ± 0.01) and between BBREECH and BULK (0.52 ± 0.01), a low negative correlation between BBREECH and LENGTH and moderate negative correlation between LENGTH and BULK. Curvature had moderate positive correlation with visually assessed traits of BBREECH and BULK, and a moderate negative correlation with LENGTH. All other wool fibre traits have little to no correlation with all visually assessed traits, except moderate positive correlations between MFD and DAG8, between MFD and LENGTH and between FDSD and LENGTH. Between the wool fibre traits, there were moderate to high positive phenotypic correlations. Exceptions were between MFD and FDCV which was low, and between CURV and all other wool fibre traits. The phenotypic correlations between CURV and other wool fibre traits were moderately negative, except with FDCV which was close to zero.

The traits, WWT and LW8, have low to negative phenotypic correlations with all other traits, except LW8 with FW12 (0.23 ± 0.01). There was a high phenotypic correlation between WWT and LW8 (0.74 ± 0.00). Fleece weight at 12 months had low to negative correlations with all other traits, except a moderate positive correlation with MFD (0.25 ± 0.04).

Table 4.4: Estimates of heritabilities (diagonal), phenotypic (above diagonal), genotypic

Trait[1]	DAG3	DAG8	BBREECH	LENGTH	BULK	MFD
DAG3	**0.44 ± 0.03**	0.41 ± 0.01	-0.08 ± 0.01	0.04 ± 0.01	-0.08 ± 0.01	0.01 ± 0.03
DAG8	0.74 ± 0.04	**0.33 ± 0.02**	-0.06 ± 0.01	0.03 ± 0.01	-0.03 ± 0.01	0.11 ± 0.04
BBREECH	-0.14 ± 0.06	-0.09 ± 0.05	**0.32 ± 0.02**	-0.19 ± 0.01	0.52 ± 0.01	0.01 ± 0.03
LENGTH	0.18 ± 0.05	0.12 ± 0.05	-0.48 ± 0.04	**0.44 ± 0.02**	-0.38 ± 0.01	0.21 ± 0.03
BULK	-0.13 ± 0.06	-0.08 ± 0.05	0.66 ± 0.03	-0.76 ± 0.02	**0.43 ± 0.02**	0.01 ± 0.03
MFD	-0.01 ± 0.13	0.09 ± 0.13	-0.10 ± 0.12	0.23 ± 0.11	-0.05 ± 0.11	**0.40 ± 0.10**
FDSD	0.17 ± 0.18	0.07 ± 0.17	-0.40 ± 0.16	0.24 ± 0.15	0.05 ± 0.16	0.78 ± 0.12
FDCV	0.00 ± 0.18	0.04 ± 0.17	-0.42 ± 0.16	0.11 ± 0.15	0.10 ± 0.16	0.23 ± 0.24
MED%	0.08 ± 0.12	0.07 ± 0.11	-0.23 ± 0.10	0.03 ± 0.10	-0.13 ± 0.10	0.28 ± 0.17
CURV	0.05 ± 0.16	-0.09 ± 0.15	0.77 ± 0.13	-0.71 ± 0.10	0.76 ± 0.10	-0.57 ± 0.16
WWT	-0.12 ± 0.05	-0.04 ± 0.04	0.25 ± 0.04	-0.03 ± 0.04	0.19 ± 0.04	0.24 ± 0.12
LW8	-0.03 ± 0.05	-0.01 ± 0.04	0.26 ± 0.04	-0.03 ± 0.04	0.20 ± 0.04	0.16 ± 0.11
FW12	0.24 ± 0.08	0.04 ± 0.07	-0.07 ± 0.07	0.32 ± 0.06	-0.05 ± 0.06	0.20 ± 0.15
σp	1.28	1.26	0.79	9.33	0.79	2.56

[1] DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; LENGTH: length of relaxed FDCV: coefficient of variation of MFD; MED%: proportion of medullated fibres; CURV: curvature;
[2] Direct genetic effect estimate from bi-variate estimates; the maternal effect estimate was 0.21 ± 0.00

4.4.4 Selection response

The selection response for DAG3 and DAG8 in scenario 1 (selection on DAG3 and DAG8) was 0.59 and 0.47, respectively. For scenario 2 (selection on BBREECH wool and fibre traits), the response for DAG3 and DAG8 was 0.32 and 0.08 standard deviations per unit of selection intensity, respectively. For scenario 3 (selection on 10 traits) the response was 0.65 and 0.47 for DAG3 and DAG8, respectively.

4.5 Discussion

The potential for accumulation of faeces on the breech region can arise from a number of factors affecting consistency of faeces and the adherence of faeces to the

(below diagonal) correlations, and phenotypic standard deviations (σ_p) ± s.e.

FDSD	FDCV	MED%	CURV	WWT	LW8	FW12
0.05 ± 0.04	0.02 ± 0.04	0.00 ± 0.04	-0.01 ± 0.04	-0.10 ± 0.01	-0.06 ± 0.01	0.06 ± 0.02
0.05 ± 0.04	0.02 ± 0.04	0.00 ± 0.04	-0.06 ± 0.04	-0.01 ± 0.01	-0.03 ± 0.01	-0.02 ± 0.01
0.00 ± 0.03	0.01 ± 0.03	-0.04 ± 0.03	0.22 ± 0.03	0.15 ± 0.01	0.14 ± 0.01	-0.01 ± 0.02
0.21 ± 0.03	0.08 ± 0.03	0.05 ± 0.03	-0.37 ± 0.03	0.06 ± 0.01	0.05 ± 0.01	0.18 ± 0.01
0.04 ± 0.03	0.04 ± 0.03	-0.06 ± 0.03	0.33 ± 0.03	0.14 ± 0.01	0.12 ± 0.01	-0.02 ± 0.02
0.64 ± 0.02	0.10 ± 0.03	0.27 ± 0.03	-0.55 ± 0.02	0.20 ± 0.03	0.12 ± 0.03	0.25 ± 0.04
0.27 ± 0.11	0.82 ± 0.01	0.40 ± 0.03	-0.31 ± 0.03	0.15 ± 0.03	0.06 ± 0.03	0.14 ± 0.06
0.83 ± 0.10	**0.23 ± 0.1**	0.33 ± 0.03	-0.03 ± 0.03	0.03 ± 0.03	-0.02 ± 0.03	0.00 ± 0.05
0.29 ± 0.21	0.18 ± 0.22	**0.53 ± 0.11**	-0.16 ± 0.03	0.03 ± 0.03	0.02 ± 0.03	0.08 ± 0.05
-0.31 ± 0.25	0.03 ± 0.28	-0.20 ± 0.20	**0.31 ± 0.10**	-0.08 ± 0.03	-0.03 ± 0.03	-0.13 ± 0.05
0.01 ± 0.16	-0.10 ± 0.16	0.05 ± 0.09	-0.06 ± 0.14	**0.21 ± 0.01**[2]	0.74 ± 0.00	0.13 ± 0.01
-0.10 ± 0.15	-0.18 ± 0.16	-0.05 ± 0.09	0.06 ± 0.14	0.76 ± 0.01	**0.23 ± 0.01**[3]	0.23 ± 0.01
0.32 ± 0.20	0.12 ± 0.20	0.13 ± 0.13	-0.25 ± 0.17	0.08 ± 0.05	0.12 ± 0.05	**0.49 ± 0.05**
1.11	2.6	9.49	6.62	4.2	4.27	0.41

wool staple on spine; BULK: wool bulk; MFD: mean fibre diameter; FDSD: standard deviation of MFD
WWT: weaning weight at 3 months; LW8: autumn live weight; FW12: fleece weight at 12 months.
[3] Direct genetic effect estimate from bi-variate estimates; the maternal effect estimate was 0.11 ± 0.00

breech. This study investigated breech cover, wool and fibre traits for their association with dagginess. Nearly 29,500 progeny representing approximately 700 sires were measured and genetic and phenotypic parameter estimates calculated.

4.5.1 Model fitting

The traits DAG3, DAG8 and MED% had high coefficients of variation (Table 4.3) This was also seen in Chapter 3, for DAG3 and DAG8, and it was shown that deleting contemporary groups with low means was the best method to stabilise the variance across contemporary groups. In this dataset, the means were higher than that previously shown in Chapter 3. One reason was that traits were allowed more time to be expressed before measurement; however 23% and 25% of DAG3 and DAG8 measurements, respectively, were deleted due to low contemporary group means. This stabilised the

variance across contemporary groups as seen in Chapter 3, and therefore the data were on an appropriate scale for analysis.

Breed percentage was fitted as a covariate for all traits to account for potential effects of breed admixture in addition to the fixed effects fitted; however, it was only significant for FW12. For all other traits, genetic estimates differed by less than 2%, depending on breed percentage being fitted or omitted. This suggests that the breeds used were sufficiently linked through the industry, possibly by the wide uptake of composite breeds. Romney, Coopworth, Perendale and Texel made up 93% of the total breed composition of this dual-purpose industry dataset. Coopworth and Perendale are fixed interbred crosses containing 50% Romney, thus the small change in estimates when breed was fitted as a covariate was not unexpected.

4.5.2 Comparison to other studies

Heritability estimates for dagginess in New Zealand dual-purpose sheep breeds range from 0.24 to 0.34 (Bisset et al., 1992; Scobie et al., 2008; Chapter 3) and from Australian Merino flocks range from 0.09 to 0.55 (Greeff and Karlsson, 2009; Pollott et al., 2004; Smith et al., 2009; Chapter 3). The heritability estimate for DAG3 (0.44) in this study was higher than the other New Zealand estimates, however, the DAG8 heritability estimate was similar to those previously reported from New Zealand.

Heritabilities for WWT and LW8 were similar to previously published estimates (Bisset et al., 1992; Brash et al., 1994b; Conington et al., 1995; Douch et al., 1995; Pollott et al., 2004; Wuliji et al., 2011). Heritability for FW12 was a little higher than previous New Zealand estimates (Bisset et al., 1992; Wuliji et al., 2011), but was similar to Australian Merino estimates (Huisman and Brown, 2008).

Breech bareness heritability was similar to other New Zealand dual-purpose (Scobie et al., 2007, 2008) and Australian Merino estimates (Brown et al., 2010; Greeff and Karlsson, 2009; Smith et al., 2009). In this study, LENGTH and BULK were subjectively assessed traits, while in other studies staple length and bulk were machine tested. However, heritability estimates were similar to those calculated by Wuliji et al. (2011) in New Zealand Romneys.

The heritability estimates for the fibre traits; MFD, FDSD, FDCV, and CURV were low compared to dual-purpose and Merino estimates (Brash et al., 1994a, c; Brown et al., 2010; Huisman et al., 2008; Mortimer et al., 2009; Safari et al., 2005; Sumner et al.,

2007; Wuliji et al., 2011). Heritability estimates for MED% have not been previously reported, however, a summary from the New Zealand Central Progeny Test, on 2 year old ewes, reported an average sire mean MED% of 5.8 % (Craven et al., 2010), range 3.5 to 7.6 %.

The genetic correlation between WWT and LW8 was less than that reported in Chapter 3), however, the phenotypic correlation was similar and consistent with previously published estimates for dual-purpose breeds (Clarke et al., 2000). Genetic correlations between DAG3 and DAG8 were similar to previous estimates in Chapter 3.

Breech bareness genetic and phenotypic correlations with DAG3 and DAG8 were lower than previous estimates by Scobie et al. (2007, 2008). Genetic correlations were also lower than estimates in Australian Merinos, however, phenotypic correlations were similar (Greeff and Karlsson, 2009; Smith et al., 2009). This could be a consequence of the breed composition of this study, with the majority of flocks (21 of 34) being Romney-based. The Romney breed has only moderate breech bareness and there is little variation in pure-bred Romney for this trait. Romneys have traditionally been selected for wool growth and thus the longer wool of a Romney can counteract the benefit of a bare breech. This was consistent with the genetic correlation between LENGTH and BBREECH (-0.48 ± 0.04), which indicates longer wool is correlated with less breech bareness. For a Romney purebred flock, Scobie et al. (2008) calculated genetic and phenotypic correlations between breech bareness and dag score at weaning of 0.89 ± 0.46 and -0.12 ± 0.05 respectively. While combining all flocks in this study, which also included Perendale and TEFRom (Texel, East Friesian and Romney composite) and Suffolk sheep, had genetic and phenotypic correlations of -0.59 ± 0.07 and -0.24 ± 0.02, respectively.

There are no known published estimates on genetic and phenotypic correlations between dagginess and the wool traits LENGTH and BULK, and the fibre traits FDSD, FDCV, MED%, and CURV. The only known published genetic correlation between late dag score (first recording at either yearling, hogget or adult age) and mean fibre diameter was 0.07 ± 0.10 and between staple length and late dag score was -0.25 ± 0.14 (Brown et al., 2010). In this dataset, we found dagginess at weaning or 8 months had little or no genetic or phenotypic correlation with most wool or fibre traits. The two exceptions were the moderate genetic correlations of DAG3 with FDSD and LENGTH, but their phenotypic correlations were close to zero. In neither case were the estimates

significant. Dag score at 3 months had a moderate genetic correlation with FW12, higher than the previous estimate in Chapter 3. The phenotypic correlation was again close to zero. In a case-control study in young lambs, it was noted that longer fleeces with low crimp frequency accumulated more faeces than short wool with tighter crimp frequency (French et al., 1998). In a more recent study looking at factors affecting faecal accumulation, wool length was only significant in 2 out of 4 inspections (Broughan and Wall, 2007b).

There are no known published genetic correlation estimates of BBREECH with LENGTH, BULK or the fibre traits. The genetic and phenotypic correlations shown here between BBREECH, LENGTH and BULK indicate that sheep with shorter bulkier wool will have larger bare breeches than sheep with longer, less bulky, wool. As there were only low genetic and phenotypic correlations between BBREECH, wool and fibre traits with DAG3 and DAG8 this suggests that faecal accumulation was not associated with wool length or type.

The majority of estimated parameters for genetic and phenotypic correlations between fibre traits were based on Australian Merinos (Huisman and Brown, 2008; Safari et al., 2005; Taylor et al., 1999), with a few studies in New Zealand Romneys (Wuliji et al., 2011), Perendales (Sumner et al., 2007), and Australian Coopworths (Brash et al., 1994c). There are no known parameter estimates for MED%. The genetic and phenotypic correlations reported here between fibre traits and subjectively-assessed traits of LENGTH and BULK were similar or less than those reported previously. Of interest was the high genetic correlation between BULK and CURV (0.86 ± 0.11), between BBREECH and CURV (0.76 ± 0.10) and the negative genetic correlation between LENGTH and CURV (-0.71 ± 0.10) and the corresponding moderate phenotypic correlations of 0.33 ± 0.03, 0.22 ± 0.03, and -0.37 ± 0.03 respectively. Compression of the fleece is dependent on the geometry of the fibre such that CURV and MFD account for 85 % of the variation seen in core bulk (Sumner et al., 2007). In this study the high correlation between hand-assessed BULK and CURV suggests that hand-assessed BULK was an effective estimate of core bulk. Additionally, the genetic and phenotypic correlations of BULK and CURV with LENGTH and BBREECH were similar. The low negative genetic correlation between BULK and MFD indicates selecting for bulky fleeces may give slightly lower MFD.

There are few known published genetic parameter estimates between fibre traits (CURV, MED%, FDSD, FDCV) and WWT and LW8 in dual-purpose sheep. Genetic and phenotypic correlations between WWT and wool traits (MFD, staple length, and CURV) reported by Huisman and Brown (2008) in Australian Merinos were similar to those reported here. For post-weaning live weight, their phenotypic correlations were similar, but their genetic estimates were greater than those reported in this study. Genetic and phenotypic correlations between WWT and wool traits (MFD, staple length and BULK) in New Zealand Romneys were reported by Wuliji et al. (2011; 1998). These estimates were similar to those reported here except for a greater estimate for genetic correlation between WWT and MFD and a smaller phenotypic correlation between WWT and BULK. Combining the results from this study and Wuliji et al. (2011; 1998), selecting for fibre traits should have no impact on WWT or LW8, except that selecting for decreased MFD may decrease WWT and LW8.

Fleece weight at 12 months (FW12) correlations with fibre traits were either smaller than or similar to other estimates published in New Zealand and Australia (Safari et al., 2005; Sumner et al., 2007; Taylor et al., 1999; Wuliji et al., 2011; Wuliji et al., 1998). Heavier fleeces were correlated with longer wool staples, higher MFD, higher FDSD and less CURV; however genetic and phenotypic correlations with BULK were close to zero. If any of the fibre traits were used as indirect indicators for flystrike, there should be no detrimental effect on FW12.

4.5.3 How much response in dagginess do the wool and fibre traits explain?

In order to examine the effectiveness of direct versus indirect selection for dagginess, the genetic parameters estimated in this study for DAG3, DAG8, BBREECH, fibre and wool traits (LENGTH, BULK, MFD, FDSD, FDCV, MED%, CURV) were used in a selection index tool (van der Werf, 2006a), with the sole objective being to reduce dagginess, through DAG3 and/or DAG8. Approximately 47% of the genetic response achieved by direct selection for DAG3 was achieved by indirect selection on BBREECH, fibre and wool traits, but only 17% of the direct genetic response in DAG8. Thus, for both dagginess traits, BBREECH, fibre and wool traits provided little or no additional predictive power to measure dagginess directly. The difference between DAG3 and DAG8 results was most likely due to animals measured at DAG8 either being crutched or shorn soon after weaning, thus reducing the predictive

ability of the wool traits. It would appear from these results that a considerable proportion of the genetic variation in dagginess was independent of wool weight, fibre traits and BBREECH.

Given the above observations, the likely additional factors affecting the genetic variation in dagginess are; pasture quality, composition, and quantity offered. These factors can affect incidence of dag formation in groups of animals, however, to affect individuals they would have to respond differently to the ingested feed or alternatively, act via diet selection. Including the result from Chapter 3, which shows dagginess was not associated with faecal egg count or nematodirus egg count, this leaves factors affecting faecal consistency as the likely cause for faecal accumulation. Previous studies have shown moderate to high genetic and phenotypic correlations between faecal consistency score and dag score in New Zealand Romney (Douch et al., 1995; McEwan et al., 1992b) and Australian Merinos (Pollott et al., 2004). Differences in faecal consistency have been attributed to pasture composition and neutral detergent fibre levels (Davidson et al., 2006; Leathwick and Atkinson, 1995), immune response to larvae (Larsen et al., 1999) and fungal endophytes (Fletcher et al., 1999; Pownall et al., 1993), and large intestine function (Bedrich and Ehrlein, 2001; Ruckebusch and Fioramonti, 1980). The traits DAG3 and DAG8 were moderately heritable. Thus, there would likely be a genetic component underlying each of the factors listed above.

4.6 Conclusions

Dagginess at 3 and 8 months of age in New Zealand dual-purpose sheep show little to no genetic or phenotypic correlations with breech bareness or wool or fibre traits. When all wool traits and BBREECH were combined together in an index, they can only achieve a minor fraction of the genetic gain observed when selecting against dags alone. As the previous chapter shows there was no correlation with faecal egg counts this then leaves internal processes; such as intestinal function, immune response and feed composition, as potential causes underlying dagginess. It is well known that dagginess is a major factor in flystrike susceptibility; therefore any studies into the above factors which may elucidate the major causes for dagginess and develop strategies to reduce dagginess and thus flystrike would be beneficial.

Chapter 5: Evaluation of DNA sampling, paternity parentage and imputation

5.1 Abstract

Paternity parentage and imputation was performed on 1,533 2009- and 2010- born lambs and 527 potential sires from the flystrike case-control dataset (Chapter 2). Ear tissue samples collected using the ear tagging systems were validated as a DNA sampling method for single nucleotide polymorphism (SNP) genotyping on 2 platforms; a 115 SNP assay and the Low Density (5K) Illumina Ovine SNP Chip. The Caisley and TypiFix^{TM} ear tagging systems produced, on average, 33 and 22 µg of DNA respectively. Genotyping success rate for the 115 SNP assay was 99.7%, with 97% of animals with a call rate greater than 80%. There were 14 SNPs discarded for parentage errors in a validation set consisting of 4 animals from a three generation family. The 2 paternity parentage programs Cervus and Partial Pedigree assigned the same sire to 75% of the progeny. There were 313 sires and 1,593 progeny genotyped on the 5K and Illumina OvineSNP50 BeadChip (50K). The SNP Chips confirmed 85% of the paternity parentage assigned by Cervus and Partial Pedigree, in total, 84% of progeny were assigned a sire. Imputation was most successful when paternity parentage information and a subset of unrelated industry animals were included with the progeny and sires using BEAGLEv3.0.4 (error rate: 0.045). Total computation time for imputation of the flystrike case-control dataset took 18 days 11 hrs and 41 mins. Finally, it was concluded that the 'best-guess' genotypes would be used for further analysis in subsequent chapters.

5.2 Introduction

The advance in high throughput genotyping technologies has now allowed genome-wide association studies (GWAS) and genome selection (GS) to dissect the underlying genes controlling traits of interest. This advance has also meant a wider variety of experimental designs can be used. In the current work, a case-control study for flystrike using primarily unrecorded commercial animals was investigated.

However, genotyping costs, though reducing rapidly, are still expensive and often do not allow the screening of the entire population on the high density (~>50K) single nucleotide polymorphism (SNP) chip. To reduce the loss of genotypes and information from using a lower density SNP chip (e.g. 5K), imputation can be used to estimate the

missing genotypes. For example, a reference population is genotyped on the 50K SNP chip and this is used to identify the allele frequencies and variates of all SNPs in the population and all potential SNP haplotypes. This information is then combined with the genotypes of the 5K SNP chip to estimate the 45K missing genotypes. Thus, the animals genotyped on the 5K SNP chip will have the same amount of information as those genotyped on the 50K SNP chip. There are numerous programs available to impute missing genotypes including BEAGLE (Browning and Browning, 2009), and PHASEBOOK (Druet and Georges, 2010), which are evaluated here. The resulting posterior genotype probabilities, or the 'best-guess' genotype, can be integrated into both GWAS and GS. It has been shown that imputed genotypes used in GWAS can be more powerful than analysing the un-imputed dataset (Anderson et al., 2008).

An option of most imputation programs is the use of pedigree and/or an unrelated reference population data to help distinguish haplotypes for phasing and imputing missing genotypes. This is extremely useful where the parents and distant relatives have already been genotyped with the high density SNP chip

The shift to high density SNP genotyping has also required a change in DNA sampling. These methods work best with high quality DNA. Currently, there is a move away from venipuncture using 10ml heparinised blood tubes, which require a veterinarian or suitably qualified person to collect, in order to comply with Animal Welfare legislation. An ear tagging system can be used by anyone, including the farmer, who is competent to insert ear tags in an animal's ear. Ear tagging (and hence this sampling method) is considered to be a routine on-farm procedure and is not classed as a manipulation that must be carried out by or under the supervision of a veterinarian. However, the quantity and quality of DNA available from a single ear punch needs to be validated for use in high-throughput genotyping technology.

The aim of this work was to create a dataset of 50K genotypes to examine flystrike and associated traits, measured in Chapter 2, in a GWAS. To be cost-effective, four steps needed to be performed and evaluated. Firstly, evaluation of the sample types collected for DNA extraction. Secondly, paternity assignment of the case-control lambs, using a low-cost, low-density (~100 SNPs) parentage panel. Thirdly, genotyping of the identified sires using the Illumina OvineSNP50 BeadBhip and the case-control animals on the Low Density (5K) Illumina Ovine SNP chip. Finally, evaluation of imputation parameters for estimation of missing genotypes. .

5.3 Material and methods

5.3.1 Animals

Samples were collected from lambs and sires as part of the flystrike case-control experiment (Chapter 2). Sample types collected included heparinised whole blood, various ear tag tissue collection systems, fast technology for analysis of nucleic acids card (FTA™, www.nfstc.org/pdi/Subject03/pdi_s03_m04_02_d.htm), spleen, and ear pieces. Some samples had been collected as part of other Ovita - AgResearch research and were available for use, including the one FTA™ sample.

To evaluate the accuracy of imputation, and select a method for use on the case control experiment, a subset of progeny, sires, and unrelated animals were selected from ~8,000 sheep previously genotyped with the Illumina OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA). These consisted of animals from other Ovita projects. The animals were selected from the same Sheep Improvement Limited (SIL) recorded flocks present in the flystrike case - control dataset and in the case of commercial flocks genetically-related recorded flocks of Romney or Romney cross breeds. Hereafter, these animals are called the "imputation evaluation dataset".

5.3.2 DNA isolation from various sample types

DNA was isolated from heparinised whole blood, FTA™ Blots (easiTrace™), and from various tissue samples including the ear tag systems TypiFix™ (Prionics, Switzerland) and FlexoPlus Geno (Caisley, Germany), ear pieces, and spleen (Table 5.1).

Table 5.1: Sample types collected as part of the case-control experiment.

Animal	Blood	FTA™	TypiFix™	Caisley	Ear	Spleen	Total
Case			698	97	45		840
Control			563	175	25		763
Sire	473	1	3		32	18	527
IMF controls	4						4
Total	477	1	1264	272	102	18	2134

Heparinised whole blood samples (~10 ml from each animal) were obtained from 473 industry sires and potential sires of the lambs measured in this study. In addition there were 4 control samples extracted as part of the International Mapping Flock

(IMF). DNA was extracted using proteinase K digestion, NaCl, and ethanol precipitation (Montgomery and Sise, 1990).

An FTA™ blot (easiTrace™ Sampler) containing blood was the only available sample from 1 potential sire. The FTA™ was extracted using GenVault - GenoSolve kit (GVR-100, Millennium Science) and purified using Qiagen® QIAamp DNA blood mini kit (51104/51106, Thermo Fisher) using the manufacturer's instructions.

Ear tissues were collected, from 1,533 lambs and 3 sires (Table 5.1), via tissue punches, TypiFix™ (Prionics, Switzerland) and FlexoPlus Geno (Caisley, Germany). Both contain desiccant for short term storage. The TypiFix™ punch yielded a 3 mm diameter punch, while the Caisley punch yielded a 4 mm diameter punch. Ear pieces were collected from 102 animals, including 70 lambs and 32 sires, when the ear tagging system was unavailable. Spleen was also collected from sires that were culled (n = 18). A 3-4 mm^3 portion of the ear and spleen pieces were used for extraction. All tissue types were held at -20°C until extraction, and were extracted using proteinase K digest, NaCl and ethanol precipitation (see Appendix 3).

All DNA samples were quantified using Quant-iT™ PicoGreen® dsDNA Assay Kit (P11496, Invitrogen), and measured on the Victor3™ (PerkinElmer) plate reader. Samples were diluted with TE buffer (10 mM Tris-EDTA) pH 8.0 to the required concentration for Sequenom® PCR (10 to 20 ng/µl), and for Illumina BeadChip genotyping (30 to 100 ng/µl).

5.3.3 Genotyping

5.3.3.1 Paternity assignment ~100 SNP assay

An iPLEX™ genotyping assay for the Sequenom® MassARRAY® platform was used. Primer assays were designed by Sequenom®. The process has been described by Clarke et al. (2011). SNPs were selected on their performance in other assays including; genotype call rate, minor allele frequency (MAF), and genome spacing. Sequences were sent to Sequenom® for design. The final design included 2 separate assays, a 58 plex and a 57 plex, each SNP has a forward and reverse amplification primer and an extension primer.

Each 5 µl amplification PCR reaction uses the Sequenom® PCR kit (SQM11324, Thermo Fisher) consisting of, 1 µl of template DNA, 0.5 µl of 10x PCR buffer, 0.4 µl of $MgCl_2$ (25 nmol), 0.1 µl of dNTPmix (25 mM each dNTP), 1 µl of primer assay (0.5

µM), and 0.2 µl of PCR Enzyme. PCR was performed on an Eppendorf® 384 well PCR cycler with an initial hold at 95°C for 2 min, followed by 45 cycles (95°C for 30 s, 56°C for 30 s, and 72°C for 1 min), and a final extension of 72°C for 5 min.

The amplification reaction was followed by a shrimp alkaline phosphate (SAP) treatment to dephosphorylate any residual dNTPs. Each 2µl reaction uses the Sequenom® iPLEX Gold Kit (SQM10136, Thermo Fisher) and contains 0.17 µl of 10x SAP buffer, and 0.3 µl of SAP enzyme (1 U/µl). Each 2µl SAP reaction was added to the wells of the PCR amplification plate and SAP treatment was performed in an Eppendorf® 384 well PCR cycler consisting of 37°C for 40 min, and 85°C for 5 min.

The SAP treatment was followed by an iPLEX Gold extension PCR. Each 2 µl reaction uses the Sequenom® iPLEX Gold Kit (SQM10136, Thermo Fisher) and contains 0.2 µl of 10x iPLEX buffer, 0.2 µl of iPLEX Termination mix, 0.94 µl of extend primer pool, and 0.041 µl of iPLEX enzyme. Each 2 µl reaction was added to the SAP treated plate and PCR was performed in an Eppendorf® 384 well PCR cycler with an initial hold at 95°C for 30 s, 60 cycles (94°C for 5 s, 4 cycles of 52°C for 5 s, and 80°C for 5 s), and a final extension of 72°C for 3 min.

Before genotyping, the reaction, products were cleaned using 6 mg of ion exchange resin. The iPLEX reaction products were then spotted onto a Sequenom® SpectroCHIP II array (SQM10117-2, Thermo Fisher), using the MassARRAY Nanodispenser. Spectra were acquired using the MassARRAY Compact Analyzer mass spectrometer and analysed using MassARRAY Typer version 4.0 (Sequenom®).

5.3.3.2 5K and 50K genotyping

All case-control sires identified by parentage using the assays above were genotyped using the Illumina OvineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA), which has usable probes for 54,977 SNPs. All case-control progeny were genotyped using the Low Density (5K) Illumina Ovine SNP Chip, which has usable probes for 5,409 SNPs. Genotyping was performed using the Illumina Infinium HD Ultra BeadChip Protocol (Illumina, Inc., San Diego, CA) at AgResearch (Invermay, New Zealand) using the manufacturers protocols. Hereafter, the Illumina OvineSNP50 BeadChip and Low Density (5K) Illumina Ovine SNP Chip are known as the 50K and 5K SNP Chips respectively.

5.3.4 Analysis

5.3.4.1 Paternity assignment

For quality control, a three generation family consisting of 4 animals (a maternal grandsire, sire, dam and a progeny) were genotyped randomly throughout the plates to test for genotyping errors. These animals were also genotyped during the validation stage of the Sequenom® assay design (Clarke et al., 2011).

Genotypes were analysed using Cervus v3.0 (Kalinowski et al., 2007). Parentage was also analysed using Partial Pedigree (Dodds et al., 2005), which is based on the original Cervus program (Marshall et al., 1998). Results from the two tests were compared and the highest probability sires selected for running on the 50K SNP panel.

Partial Pedigree, based on the original Cervus program (Marshall et al., 1998), was modified so male-female pairs could be assessed as parents of an offspring (Dodds et al., 2005). The error model assumes that the probability of observing the genotyping error is equal to proportion of the observed frequency of that genotype in the population. When performing a paternity test, dummy values are used for the dam. A dummy sire is also added to each subpopulation analysed. If the dummy sire is more likely to be the true sire than any of the potential sires, then no paternity is assigned.

Cervus v3.0 is a new version with 2 main modifications. It can perform maternal, paternal or parent pair parentage. The other modification is a new error model. The old model was shown to over inflate the rate at which genotyping errors are expected to occur (Kalinowski et al., 2007). The new model assumes that when a genotyping error has occurred, the error is equal to the proportion of the Hardy-Weinberg frequency of that genotype.

5.3.4.2 Independent paternity assignment using 5K SNP Chip

Parentage assignments from the ~100 SNP test for the case and control experiment were separately checked using the ~5K SNPs common to both the 50K and 5K SNP Chips. Progeny – sire pairs resulting from the above parentage analysis were checked, any remaining progeny with no sire assigned were assessed by a ‘mix and match’ process with all available sires using SAS 9.1 (SAS, 2004).

5.3.4.3 Imputation evaluation

The imputation evaluation animals were separated into a pedigree file consisting of 227 progeny and 129 sires, a file of 60 progeny with no known sire, and an 'unrelated' file consisting of 1,130 animals treated as unrelated (i.e. no pedigree information used) from the same population. The genotypes of the 287 progeny (227 progeny with known sire and 60 progeny with unknown sire) were masked, such that they would resemble results from a 5K SNP Chip. After imputation using a variety of methods and parameters, the imputed genotypes (best guess) were then compared to their original genotypes from the 50K SNP Chip in order to assess error rate.

The genotyped SNPs were ordered on version 1.0 of the sheep genome assembly (OARv1, http://www.livestockgenomics.csiro.au/sheep/oar1.0.php), chromosome X and unknown chromosome SNPs were discarded, leaving 47,011 SNPs. Two programs were assessed; PHASEBOOK (LinkPHASE and DAGPHASE) (Druet and Georges, 2010) which uses all available pedigree information, and BEAGLE v2.1.3 and v3.0.4 (Browning and Browning, 2009).

A number of combinations were compared, using genotypes from chromosome 26 (n SNPs = 823), to identify the best program and parameter settings.

1. LinkPHASE, DAGPHASE and BEAGLEv2.1.3. The 287 progeny and 129 sires were used including available pedigree information.
2. BEAGLEv3.0.4. The 287 progeny and 129 sires were treated as unrelated.
3. BEAGLEv3.0.4. The 227 progeny and 129 sires in a related 'pairs' file.
4. BEAGLEv3.0.4. As option 2 above with an extra 1130 unrelated animals.
5. BEAGLEv3.0.4. As option 3 above; the 60 unrelated progeny were added to the extra 1130 unrelated animals to generate the unrelated file.

The imputed best guess genotypes were compared to the original genotypes, and an error rate per SNP and per animal calculated.

5.3.4.4 Imputation in case-control 5K animals

Option 5 above was identified as having the lowest error rate. This was then applied to the flystrike case-control dataset. In the case-control dataset, information from chromosome 1 – 3 was split in half where the adjacent SNPS were more than 50 kbp apart. This was because they are at least twice the volume of data of the next biggest chromosome (chr 4) and took excessively long to compute. Chromosome 1 was split

between SNP 2,277 and 2,278, ~94 kbp apart, generating 2 parts containing 2,277 and 3,008 SNPs respectively. Chromosome 2 was split between SNP 2,231 and 2,232, ~51 kbp apart, generating 2 parts containing 2,231 and 2,744 SNPs respectively. Chromosome three was split between SNP 2,143 and 2,144, ~61 kbp apart, generating 2 parts containing 2,143 and 2,360 SNPs respectively. For the animals, these were split into a pairs file containing 1,291 progeny sire pairs (number of Sires, 268), and an unrelated file containing 253 progeny and 1062 industry sires.

Run parameters for each chromosome (e.g. for chr 26) used were:

java_1.6.0 –Xmx100000m -jar beagle.jar unphased=chr26.unr.bgl pairs= chr26.pair.bgl markers=chr26.markers out=beagle missing=? log=beagle

The genotype probabilities or the best guess genotypes can be used for further analysis, however it is recommended that the genotype probabilities are used (Pei et al., 2010). To check if combining the results from the paternity pairs file and the genotype probabilities from the unrelated file could be used for further analysis the genomic relationship matrix (VanRaden, 2008; G, Equation 1) was estimated and the diagonal estimate was plotted for each animal. The G matrix can be used to structure the (co)variances between the random additive genetic effects of the different animals, using their genotypes to derive a measure of genetic similarity i.e. identical by descent (IBS). This was compared against combining the pairs file and the best guess from the unrelated file.

$$G = \frac{ZZ'}{2\sum p_i(1-p_i)} \qquad \text{Equation 1}$$

Where Z is the SNP matrix $-2p_i$, $1-2p_i$, and $2-2p_i$ for BB, AB, and AA respectively, and p_i is the frequency of A allele of the i^{th} SNP in the population.

5.4 Results

5.4.1 DNA isolation from various sample types

All blood samples and the FTA™ sample were successfully extracted to DNA. The 10 ml blood samples generated highly concentrated DNA > 1000 ng/µl in 1 ml of TE, while the 12 2 mm^2 FTA™ sample punches from the single FTA™ sample produced

25.1 ng/µl in 25 µl of TE. Of the ear tag systems 9 samples (1 Caisley, and 8 TypiFix™) failed to collect a tissue sample. Table 5.2 and Figure 5.1 show the average and range of DNA amounts (µg) obtained from the tissue samples. A selection of samples were run on a 1.5 % TBE agarose gel to check for degradation, an example is seen in Figure 5.2. No degradation was observed in the samples and all samples contained strands greater than 12,000 bp, the highest rung of the molecular ladder.

Table 5.2: Average, standard deviation (SD), minimum (Min) and maximum (Max) amount (µg) of DNA from tissue samples successfully extracted.

Sample Type	Average (µg)	SD	Min	Max
TypiFix™	20.20	12.00	0.55	176.50
Caisley	32.53	14.38	7.28	163.50
Ear	25.21	18.66	3.27	102.09
Spleen	138.63	96.81	16.20	388.90

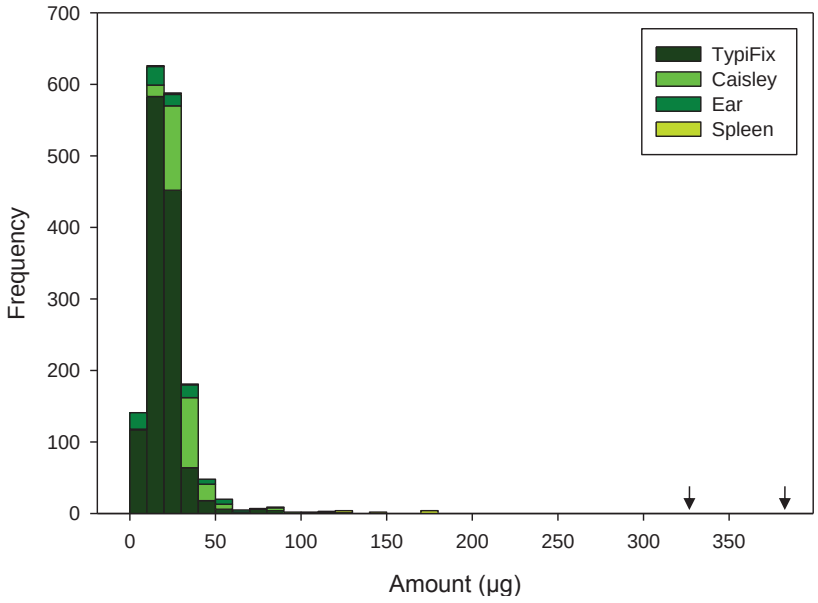


Figure 5.1: Frequency distribution of DNA concentrations (µg) from tissue extractions. Arrows indicate 1 spleen sample at 322 µg and 389 µg.

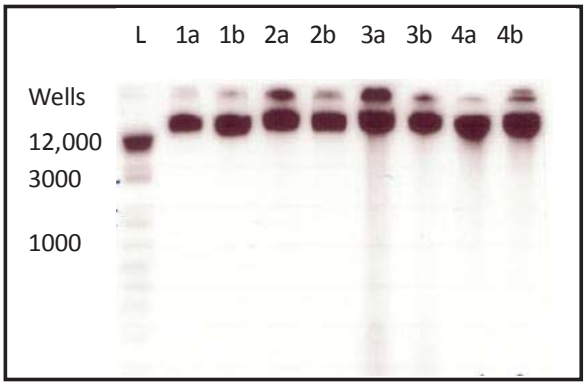


Figure 5.2: Gel picture of DNA from TypiFix™ ear tissue extractions. Samples were run in duplicate, on a 1.5% agarose gel with 1kb plus ladder.

5.4.2 Parentage

5.4.2.1 Genotyping

Table 5.3 shows the number of animals and sample types that were genotyped through the Sequenom® assay. Six samples (all progeny) failed genotyping and 55 samples had a call rate less than 80% (Figure 5.3). For paternity assignment, successful assignment can occur with 50% of the SNPs, if both the progeny and potential sire have results for the same SNPs.

Table 5.3: Sample type and number of animals that were genotyped by the Sequenom® assay.

Animal	Blood	FTA™	TypiFix™	Caisley	Ear	Spleen	Total
Case			681	96	45		822
Control			561	144	25		730
Sire	473	1	3		32	18	527
IMF control	4						4
Total	477	1	1245	240	102	18	2083

The average call rate for the 115 SNPs was 94.5%; Figure 5.4 shows the distribution of call rates among the 115 SNPs. Four IMF controls; consisting of a maternal grandsire, sire, dam and a progeny were used to test for genotyping errors. Ten SNPs were identified as having pedigree errors (greater than 20%) and were thus discarded from further analysis. Four more SNPs were discarded due to call rates under 80%.

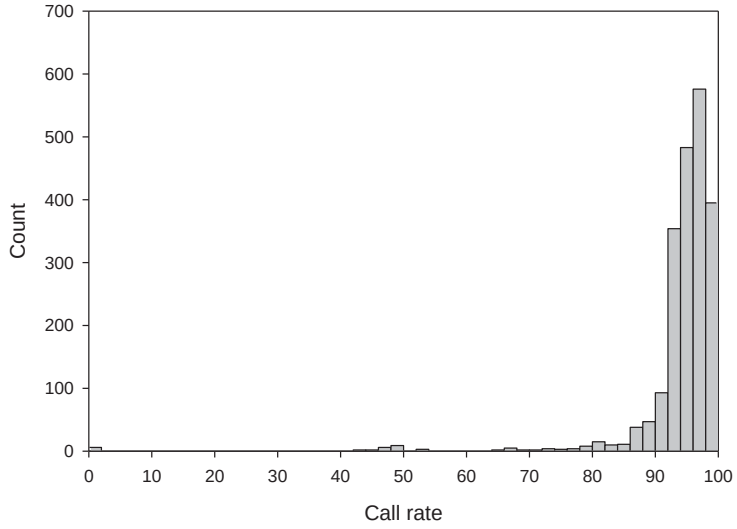


Figure 5.3: Frequency distribution of call rate (%) of all samples genotyped through the Sequenom® assay derived from the 115 SNPs genotyped.

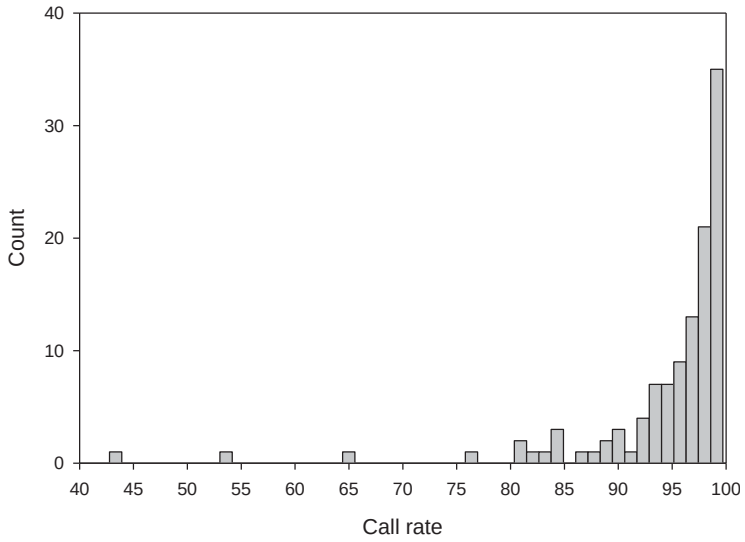


Figure 5.4: Frequency distribution of call rate (%) of all 115 Sequenom® SNPs across all animals.

5.4.2.2 Analysis

5.4.2.2.1 Cervus v3.0

Genotypes (n = 101 SNPs) for progeny and their potential sires (based on mating groups) were analysed for paternity in Cervus v3.0. Every progeny – potential sire pair was given a logarithm of the odds (LOD) score on its potential as being a likely match.

A positive LOD score suggests the potential sire is more likely to be the true sire than not. Of the 1,546 flystrike cases and controls, successfully genotyped, 1,250 were assigned a sire match with a positive LOD score (average LOD = 14, range 0.05 to 51.74, confidence score of 0.95). There were 295 unique sires identified. The validation IMF animals were successfully assigned.

5.4.2.2.2 Partial pedigree

Progeny and their potential sires (based on mating groups) were also analysed through Partial Pedigree, an AgResearch in-house parentage program. All potential progeny-sire matches were given a probability value. Of the 1,546 flystrike cases and controls run through the Partial Pedigree analysis, 1,379 received a sire match and 1,256 were assigned a sire match with greater than 90% probability. There were 314 unique sires identified. The validation IMF animals were successfully assigned.

5.4.2.2.3 Combined

To select the best sires to run on the 50K SNP Chip, the results from Cervus v3.0, Partial Pedigree, and if available, the sire (as identified in Sheep Improvement Limited (SIL) for those lambs belonging to a SIL recorded flock) were compared. Cervus v3.0 and Partial Pedigree identified the same sire for 1,161 of the progeny. Of the remainder the Cervus v3.0 result was used for 22 progeny, and Partial Pedigree for 194 progeny based on their LOD and probability score. In total 1,377 progeny were assigned a sire with confidence (Cervus: v3.0 LOD average 14; Partial Pedigree: > 90% probability). There were 313 unique sires identified for further genotyping on the 50K SNP Chip. See section 5.4.3.3 for the accuracy of these results when they were later compared to 5K assignments.

5.4.3 Imputation

5.4.3.1 Genotyping of case-control and sires

There were 1,593 case-control progeny run on the 5K SNP Chip and 335 sires run on the 50K SNP Chip. This included all the progeny and sires assigned in the previous section the unassigned progeny and some potential sires that had been run as part of separate Ovita research. Table 5.4 shows the genotyping success of the animals. Of the 1,928 animals, run 94.5% had a call rate greater than 98% (95.5% of the 5K SNP panel animals, 94.7% of 50K SNP panel animals). Of the failed genotyping animals, only the

5K SNP Panel animals were due to contamination. The main reason for 50K SNP panel animals that failed to genotype was incorrect dilution of samples for genotyping. Figures 5.5 and 5.6 show the distribution of minor allele frequencies in the 5K SNP and 50K SNP genotyped chips. Average MAF for the 5K SNPs was 0.35 and for 50K SNPs was 0.28. Note the 5K SNPs were selected for their high minor allele frequency in the HapMap 50K dataset of 74 breeds (Anderson et al., 2012a).

Table 5.4: Genotyping success rate of the flystrike case-controls run on the 5K SNP Chip and the sires run on the 50K SNP Chip.

	50K	5K
>98% CR	320 (96%)	1,509 (95%)
<98% CR	11 (3%)	34 (2%)
Fail	4 (1%)	50 (3%)
Total	335	1,593

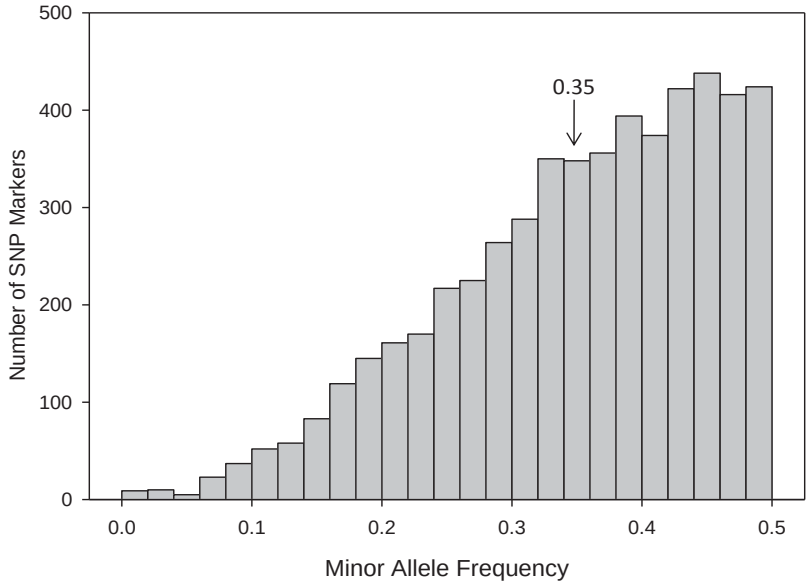


Figure 5.5: Minor allele frequency (MAF) of SNPs genotyped over the progeny on the 5K SNP Chip. The average MAF noted by arrow.

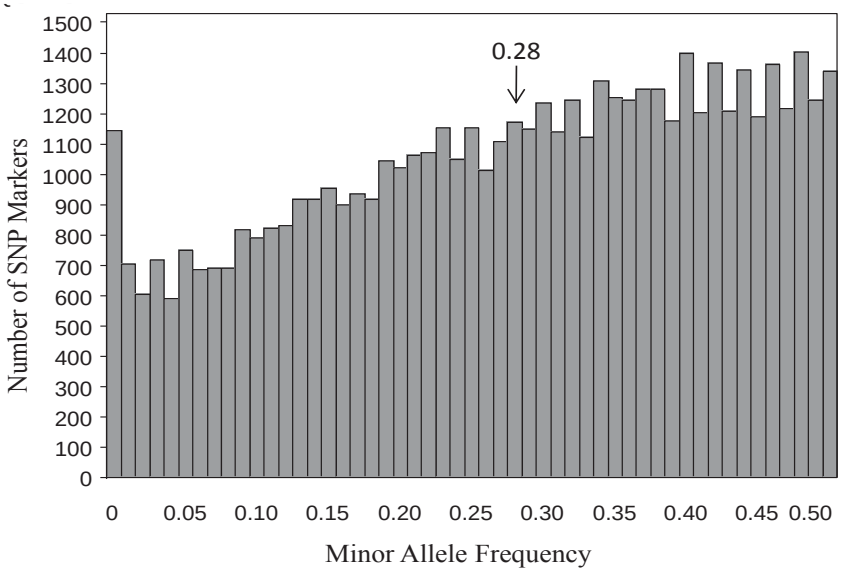


Figure 5.6: Minor allele frequency (MAF) of SNPs genotyped over the sires on the 50K SNP Chip. The average MAF noted by arrow.

5.4.3.2 Imputation evaluation

These results were derived from the separate 50K dataset described in section 5.3.4.3. The error rates calculated from the combined pairs and unrelated best guess results of the 5 options are shown in Table 5.5. The best combination for imputation was Option 5 which used BEAGLE v3.0.4, with a pairs file containing 277 progeny-sire pairs, and an unrelated file with 60 progeny plus 1,130 unrelated industry animals (Table 5.5). For option 5, the error rate across all SNPs for the unrelated progeny was 0.063 and for the 'pairs' progeny 0.042, overall it was 0.046. The distribution of error rates for animals and SNPs are shown in Figure 5.7 for Option 1 (poor) and Option 5 (best). Note the error rate variation for Option 1 was also greater than that for Option 5. The error rate for each imputed SNP is plotted against the minor allele frequency (MAF) of the imputed SNP for the imputed progeny from Option 1 and Option 5 in Figure 5.8. For option 1 the error rate is close to the MAF, as the MAF becomes greater than 0.15 the error rate is less than the MAF. For Option 5 the error rate was consistently less than the MAF.

Table 5.5: Average proportion of errors for each imputed SNP and imputed progeny for each program option.

Option[1]	Option 1	Option 2	Option 3	Option 4	Option 5
Error proportion					
Progeny	0.177	0.178	0.099	0.065	0.046
SNPs	0.139	0.179	0.099	0.065	0.046
Number of animals and SNPs					
n SNPs	724	724	724	724	724
n unrelated progeny	287	287		287	60
n pairs progeny			227		227
n unrelated animals	129	129		1259	1130
n pairs sires			129		129
Computation time	1 min	1min 17s	1min 35s	46 mins	27mins

[1] Option1: LinkPHASE, DAGPHASE and BEAGLEv2.1.3. The 287 progeny and 129 sires were used including available pedigree information. Option2: BEAGLEv3.0.4. The 287 progeny and 129 sires were treated as unrelated. Option3: BEAGLEv3.0.4. The 227 progeny and 129 sires in a related 'pairs' file. Option4: BEAGLEv3.0.4. As option 2 above with an extra 1130 unrelated animals. Option5: BEAGLEv3.0.4. As option 3 above; the 60 unrelated progeny were added to the extra 1130 unrelated animals to generate the unrelated file.

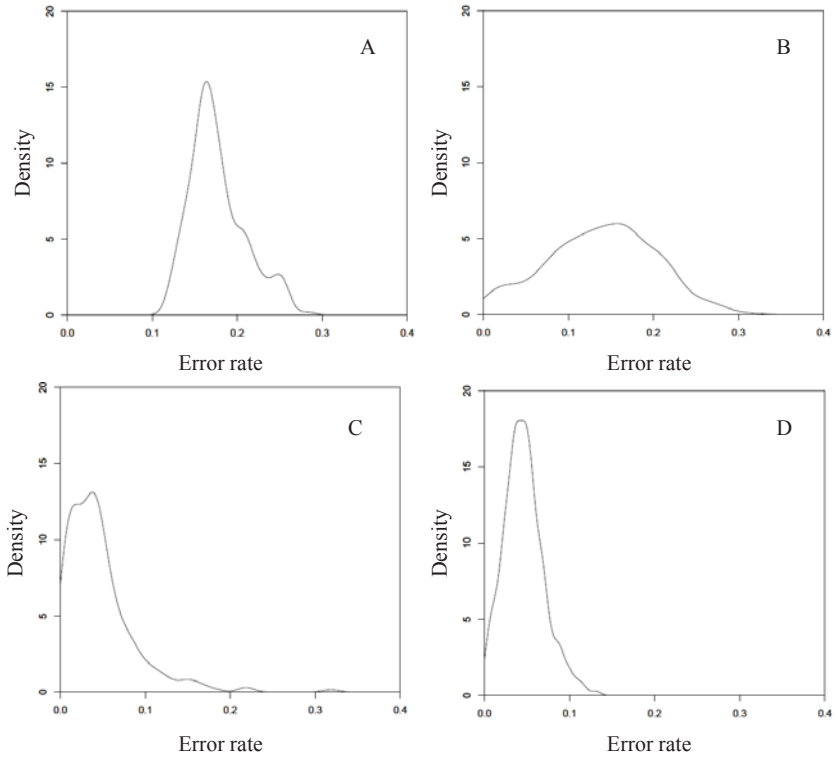


Figure 5.7: Density graph of error rates for Option1: animals (A), SNPs (B), and Option 5: animals (C), SNPs (D). Option 1: LinkPHASE, DAGPHASE and BEAGLEv2.1.3: 287 progeny and 129 sires, including pedigree information. Option 5: BEAGLEv3.0.4: 227 progeny and 129 sires ('pairs' file) plus the 60 unrelated progeny and 1130 unrelated animals (unrelated file).

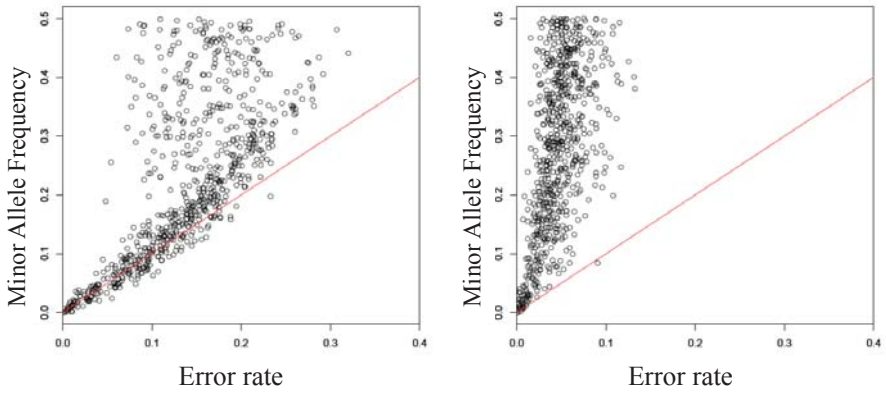


Figure 5.8: Comparison of error rate for the imputed progeny per SNP with the minor allele frequency in the imputed data set for each SNP (0-1 line, red). Option 1: LinkPHASE, DAGPHASE and BEAGLEv2.1.3: 287 progeny and 129 sires, including pedigree information (left). Option 5: BEAGLEv3.0.4: 227 progeny and 129 sires ('pairs' file) plus the 60 unrelated progeny and 1130 unrelated animals (unrelated file) (right).

Table 5.6: Comparison of paternity testing programs. Number of paternity assigned progeny per program on the diagonal (percentage). Below the diagonal, actual values of paternity tested progeny with the same result in 2 tests (percentage), and number of total assigned progeny per SNP platform, number of unassigned progeny per test (percentage) and number of progeny that passed quality control on the 2 platforms; Sequenom (Cervus and Partial Pedigree) and 5K SNP Chip.

Test	Cervus	Partial Pedigree	5K SNP Panel
Cervus	**1250 (81)**		
Partial Pedigree	1161 (84)	**1256 (81)**	
5K SNP Chip	1086 (79)	1162 (84)	**1291 (84)**
Total assigned	1377 (89)		1291
Not assigned	169 (11)		252 (16)
N passed genotyping	1546		1543

5.4.3.3 Flystrike case-control paternity assignment

The paternity assignments derived from the 100 SNP panel were compared with the ~5K SNPs common to both the 50K and 5K SNP Chips (Table 5.6). Of the 1,377 progeny assigned a sire from the paternity testing using Cervus v3.0 and Partial Pedigree, 1,165 (84.6%) were confirmed with the 5K SNPs, and 46 (3.3%) were

assigned to another sire. For the remaining paternity assigned progeny (n = 122), 116 did not match their assigned sire and for 6 their paternity assigned sire failed 50K genotyping. Using the mix and match process with the initially unmatched progeny (n = 233), a further 80 progeny were matched to a sire, with less than 5 mismatches out of 5409 SNPs genotyped. There were 268 unique sires identified. After 5K genotyping there were 253 progeny with no identified sire.

5.4.3.4 Flystrike imputation results and checking the G matrix

5.4.3.4.1 Running time

Computation time (note, not system time) for each chromosome starting with the smallest chr 26, is reported in Table 5.7. Computation was performed on a HP DL-585 with 4 Opteron 6176 12 core processors (48 cores at 2.3GHz). Note a number of jobs of different computing needs can be run on this machine at the same time, thus computation time as time from start to finish of job is subject to other jobs that were being run. Overall it took 11 days 14 hours and 22 minutes for chromosome 26 to 4 to finish. The computation for the full chr 3 caused a problem, and had to be forced quit, after 2 months. It was then decided to split the last three chromosomes in half and it then took a further 6 days 21 hours and 19 minutes to finish.

5.4.3.4.2 Output files

The resulting file from BEAGLEv3.0.4, using the paternity assignment (pairs) data, was a phased file which consisted of the transmitted haplotype from the sire, the untransmitted haplotype from the sire and the other ungenotyped parent haplotype, where the progeny's genotype equalled the sires transmitted and the ungenotyped parents transmitted haplotype. For the unrelated animals, BEAGLEv3.0.4 produced a phased file containing the 'best guess' genotype for the progeny, and a genotyped probability file containing the probabilities for the AA, AB, and BB genotypes for each animal and each SNP. Minor allele frequencies were calculated for the imputed progeny in the pairs file, and the unrelated file as 'best guess' result or genotype probabilities (Figure 5.9). The average minor allele frequencies were 0.28, for each of the three data sets. The correlation (R^2) between the best guess and the genotype probabilities was 0.99.

Table 5.7: Computational time taken to impute each chromosome (Chr) in days hours (hrs) and minutes (mins).

Chr[1]	Time
26	5 hrs 29 mins
25	8 hrs 25 mins
24	6 hrs 59 mins
23	10 hrs 45 mins
22	7 hrs 37 mins
21	5 hrs 6 mins
20	9 hrs 8 mins
19	10 hrs 6mins
18	11 hrs
17	9 hrs 30 mins
16	13 hrs 2 mins
15	16 hrs 15 mins
14	9 hrs 8 mins
13	13 hrs 38 mins
12	12 hrs 15 mins
11	8 hrs 57 mins
10	11 hrs 24 mins
9	15 hrs 3 mins
8	16 hrs 41 mins
7	21 hrs 11 mins
6	18 hrs 37 mins
5	16 hrs 53 mins
4	21 hrs 13 mins
3a	1 day 2 hrs 49 mins
3b	1 day 1 hr 31 mins
2a	1 day 1 hr 32 mins
2b	1 day 5 hrs 9 mins
1a	1 day 2 hrs 23 mins
1b	1 day 7 hrs 55 mins

[1] Chr 1 to 3 split in half (a, b) for computation, as outlined in section 5.3.4.4.

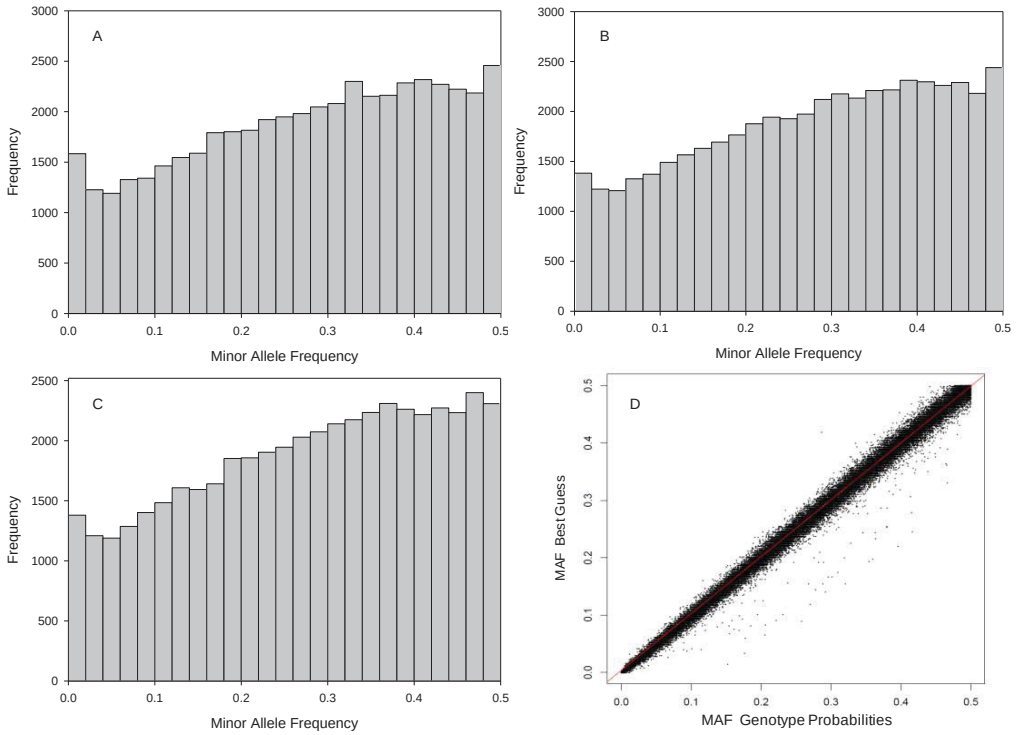


Figure 5.9: Minor allele frequencies (MAF) for unrelated best guess (A), unrelated genotype probabilities (B), pairs (C), and the correlation between best guess and genotype probabilities MAF.

5.4.3.4.3 Genotype probabilities or best guess for further analysis

To check if the best guess or genotype probabilities from the unrelated progeny should be used for further analysis, a G matrix was estimated for each situation. Figure 5.10 shows the diagonal values for the G matrix estimated from the genotype probabilities of the unrelated progeny and the phased pairs genotypes from the paternity assigned progeny. The first 253 animals (unrelated progeny) have diagonal values ranging between 0.5 and 1.1, while the pairs progeny have values around 1. Figure 5.11 shows the diagonal values for the G matrix estimated from the best guess genotypes of the unrelated progeny and the phased pairs genotypes from the paternity assigned progeny. The unrelated animals (first 253) and the pairs animals had values ranging from 0.9 to 1.14.

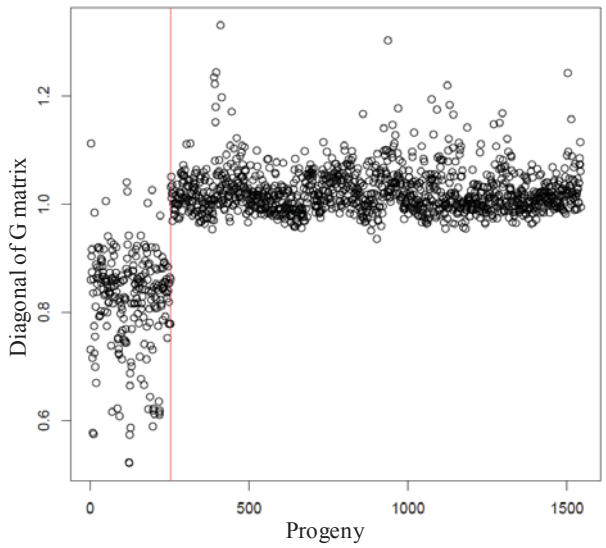


Figure 5.10: Diagonal values of the G matrix using the genotype probabilities for the 253 unrelated progeny (left of red line), and the pairs phased genotypes for the paternity assigned progeny (right of red line).

To minimise the difference in the diagonal values between those calculated from the genotype probabilities and the best guess genotypes for the unrelated progeny, two methods were tested. First, SNPs were removed from the genotype probability calculated G matrix if they had low accuracies. This did not change the diagonal elements for the unrelated progeny (the plotted diagonal values were the same as Figure 5.10). Second, the best guess diagonal values were compared to the genotype probability diagonal values (Figure 5.12). The animals lying on the diagonal were the paternity pairs phased progeny, the unrelated progeny cluster to the left.

For subsequent GWAS and GS analysis, the best guess genotypes for the unrelated progeny and the pairs phased file for the paternity assigned progeny were used.

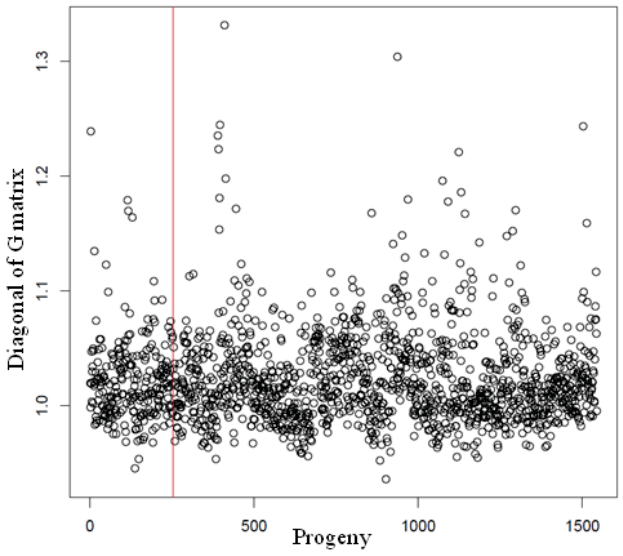


Figure 5.11: Diagonal values of the G matrix using the best guess genotypes for the 253 unrelated progeny (left of red line), and the pairs phased genotypes for the paternity assigned progeny (right of red line).

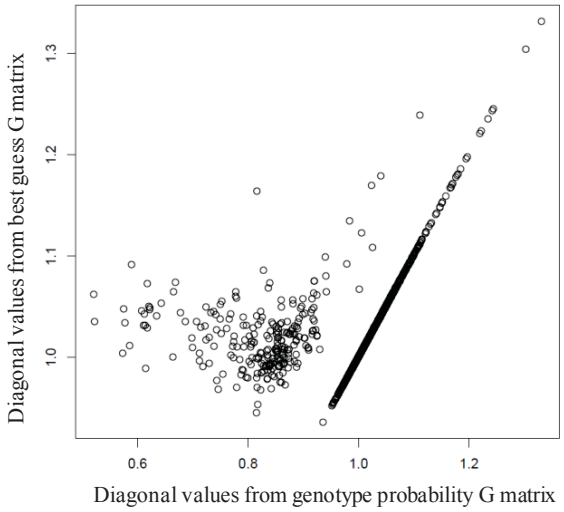


Figure 5.12: The comparison of the diagonal values of the G matrix from the genotype probabilities and the best guess genotypes. Including the paternity pairs phased genotypes.

5.5 Discussion

The objective of this work was to create a set of genotypes to subsequently undertake GWAS and GS for dagginess and flystrike. However, resources were limiting

and not all animals could be genotyped on 50K Chips. Similarly, insufficient ram breeders could be identified who were willing to record natural flystrike challenge in their valuable breeding stock. Commercial farmers linked to willing breeders were therefore recruited. Similarly, in order to do this work cost-effectively, paternity then had to be initially assigned using a cheap test with the samples subsequently genotyped using an optimal mixture of 50K and 5K SNP Chips. Implicit in this work is that the DNA samples collected and subsequently extracted had to provide DNA of sufficient quality and quantity.

5.5.1 Sample type

A number of tissue and blood collection types were extracted and genotyped through the Sequenom iPLEX Platinum Chemistry and 5K SNP Chip. These technologies prefer high quality DNA. Ear tag tissue collection systems are fast becoming the most used sample type for high-throughput genotyping because they typically provide large quantities (>15µg) of high quality DNA. The collection of the 10ml heparinised blood sample by venipuncture, which requires a veterinarian or qualified person to take the sample, used to be the main sample type for research where large quantities of DNA was required. Unfortunately, subsequent DNA extraction is difficult to automate and collection costs were high. In contrast, the ear tagging tissue collection system can be used by the farmer themselves, reducing the costs of sample collection, the desiccated samples can be stored at room temperature during shipping, the bar coded collection tubes aid lab automation and they produce high quality and a large quantity of DNA.

Of the ear tagging systems; the Caisley system yielded the most DNA, 33 µg versus 20 µg (TypiFix^{TM}), due to the larger diameter of the tissue plug collected. During collection, both tagging systems had similar failure rates. Less than 0.7% of TypiFix^{TM} tags failed to collect a sample, while less than 0.4% of Caisley tags failed. The Caisley ear tagging systems also generated similar DNA yield to the alternative tissue collection of spleen or larger ear pieces. In fact both systems produce DNA of adequate quality and quantity for whole genome sequencing (100 ng-4 µg) (BGI Americas, 2012; Wellcome Trust Sanger Institute, 2012).

5.5.2 Paternity

There was a 99.7% genotyping success rate on the Sequenom® platform for the 115 plex assays, and 97.4% of the animals had a call rate above 80%. The average call rate for the SNPs was 94.5% and only 14 SNPs had to be discarded for pedigree errors and low call rates, leaving 101 SNPs for subsequent evaluation. This is similar to results obtained by (Clarke et al., 2011), who evaluated the same assay. They also discarded the same SNPs for parentage error and low call rates. In this study, one additional SNP was discarded due to low call rate. Cervus v3.0 and Partial Pedigree had the same sire match for only 1,161 progeny (75%). This is likely to be due to differences between the 2 programs when dealing with genotyping error. Another problem is collection of all the potential sires. In a breeding flock, care is taken to record which sires are in which mating groups. However, on unrecorded/commercial farms there is little to no recording of which sires were joined, and mating groups are not recorded. This makes collection of a DNA sample from all potential sires problematic and every sire on the property was collected providing a large set of potential sires including a number of half-sibs. With no dam information, the power of paternity only parentage assignment is much less when comparing between half sibs.

From this analysis 313 identified sires were genotyped on the 50K SNP Chip, and 1,593 progeny where genotyped with the 5K SNP Chip. The 5K SNPs shared by the 50K and 5K SNP Chips confirmed 1,165/1,377 or 85% of the parentage results from the 101 Sequenom® SNPs. It was noted that 87% of the unassigned progeny (n= 253) were from unrecorded/commercial flocks. This re-iterates the problem of using unrecorded/commercial flocks in a study where sires and mating groups are unrecorded. A proportion of the unassigned progeny will also be a result of failed genotyping, this occurred when 4 sires failed genotyping on the 50K SNP Chip, relating to 6 unconfirmed progeny-sire matches. It was also found that samples which failed genotyping on the 5K or 50K SNP Chips was mainly due to the sample quality or concentration and to re-run the animal a new sample from the animal would be required. For a research project, the cost of attaining and re-running a new sample was not feasible. A final problem in using 2 parentage platforms, 115 SNPs and 5K SNPs, is that sires that failed to genotype or were not assigned to a progeny using the 115 SNPs were not run on the 50K SNP Chips, thus false-negatives from the 115 SNP platform could not be correctly identified by the parentage checks performed using the 5K SNPs.

5.5.3 Imputation

The imputation evaluation was most successful when both paternity parentage and unrelated industry animals were used in BEAGLEv3.0.4. Including paternity pair information halved the error rate of treating progeny and sires as unrelated. This showed the value of genotyping sires with the 50K SNP Chip. Adding in the unrelated industry animals with the parentage again halved the error rate from just using available parentage information. Adding more animals from the same population allowed for greater sampling of all the potential haplotypes available.

5.5.3.1 Other methods and considerations

BEAGLEv3.0.4 uses a haplotype-frequency model that can make use of large unphased reference panels (Browning and Browning, 2009). The program then has to phase the data which can be computationally intensive. Hickey et al. (2011) report that knowing the phase of marker genotypes is useful and can account for identity by descent or parent of origin alleles. They suggest using long-range phasing and long haplotype library imputation. Phasing was not affected by SNP density, effective population size or family structure. The amount of pedigree information did not affect phasing either, however, it was noted that ignoring pedigree increased computation time (Hickey et al., 2011).

Hayes et al. (2012) simulated low density SNP Chips for imputation based on population-wide linkage disequilibrium (used fastPHASE or BEAGLE with no pedigree information) in Australian sheep breeds. Imputation within breed was more accurate (5K to 50K, 63 - 80% accuracy) than if done across all breeds (63 to 68%). It was more accurate if the genetic diversity within a breed was low, i.e. genetic relationships were closer. They suggest that the 50K SNP Chip is not sufficiently dense and thus the linkage disequilibrium phasing and haplotypes in the reference population do not persist across breeds.

Phasing and consideration of the genetic makeup of the animals to be imputed are 2 areas that need to be further addressed. Pre-phasing the data before imputation may help with the computation intensity of the larger chromosomes. In this flystrike case-control data set, the animals were from Romney or Romney cross flocks, thus the problem faced by Hayes et al. (2012), may not have been as much as an issue for this analysis. For future analyses in other datasets, breed needs to be considered.

5.5.3.2 Computation time

Computation time was calculated as the time from start of the first imputation run to the finish of the last results file. This is not the same as actual system time; however, it is the real running time. The total running time to complete all 26 chromosomes was 18 days 11 hrs and 41 min. This does seem excessive and the chopping of the 3 biggest chromosomes in half to speed up computation is not ideal. This was due to including the paternity information and the extra unrelated industry sires in the imputation. During testing of the different programs with chr 26 (823 SNPs, 724 imputed) the first three options finished within 2 minutes. Using BEAGLEv3.0.4 and treating progeny, sires and extra industry animals as unrelated (Option 4) took longer than splitting the data into a pairs file of progeny and assigned sire and a unrelated file with remaining progeny and industry animals (Option 5). This was potentially due to the pairs information allowing for faster phase haplotyping. Browning and Browning (2009) compared BEAGLEv3.0 with IMPUTEv0.5.0 for imputation in a 5Mb region (1356 SNPs, 746 genotyped) of human chromosome 1 in 188 individuals with reference panels of 300, 600, and 1200 individuals. BEAGLEv3.0 took 2.7 min, 5.5 min, and 12.0 min and IMPUTEv0.5.0 took 60.3 min, 220.2 min and 829.6 min respectively.

If imputation from the 5K SNP Chip to the 50K SNP Chip is to be used for genomic selection in the New Zealand sheep industry, more work is required on what datasets are to be used in BEAGLEv3.0, i.e. pairs or other type of pedigree file, unrelated, and/or reference panels. Especially for the three largest chromosomes, so as to decrease computational time while retaining complete chromosome datasets, and retaining the accuracy of imputation. The current computational capacity allows either the running of each chr one after the other, or simultaneously with the ‘computer’ swapping between runs as memory allows. There is also the option of running the chromosomes in parallel on different nodes.

5.6 Conclusions

The ear tagging system is adequate for use in high-throughput technologies. Parentage testing using the Sequenom® assay of only 101 working SNPs is problematic when used on commercial farms where there is inadequate recording of mating groups. The 5K SNP Chip performed well, however, it was noted there were a number of

contaminated samples from the ear tissue samples. These samples need evaluating to test if the cause is at the extraction stage or the dilution and plating stage before SNP chipping, and may require additional laboratory procedures to be put in place to minimise these risks. Imputation is most accurate when parentage information and a reference set of animals, from the same population as the animals to be imputed, are included.

For future use, sire-progeny pairs should be used as well as a reference panel from the same genetic background. If commercial unrecorded animals are to be used again, emphasis needs to be placed on the recording of what sires are mated each year, or the collection of DNA from sires at the same time as they are joined with the ewes. This would decrease the number of potential sires and help with identifying more progeny-sire pairs. Another option is to increase the initial number of SNPs used for paternity assignment before genotyping on the 5K and 50K SNP Chips. More analysis is needed to find the optimal number of SNPs needed to assign paternity in unrecorded flocks.

Finally, the use of the best guess genotypes resulting from imputation of progeny treated as unrelated are to be used in future GWAS. If the genotype probabilities, were to be used, investigation into appropriate scaling methods to correct for the distortion in the diagonal values from calculating the genomic relationship matrix would be required.

Chapter 6: Genome-wide association study: flystrike case - controls

6.1 Abstract

To identify loci associated with flystrike resistance, a genome-wide association study (GWAS) was performed with 47,011 autosomal SNPs on 738 affected lambs (cases) and 731 controls using an imputed flystrike case-control dataset (Chapter 2 and 5). Associated traits, also measured on the flystrike cases and controls; dag score, breech bareness, wool bulk, wool length, mean fibre diameter (MFD), standard deviation of MFD (FDSD), coefficient of variation of MFD (FDCV), curvature (CURV), and proportion of medullated fibres (MED%), were also analysed by GWAS in an attempt to locate regions associated with these traits as well as identify commonality with flystrike resistance associated regions. No SNPs reached the Bonferroni significance threshold of $P<1.03x10^{-6}$. There were 116 SNPs identified that reached a nominal $P<0.001$ threshold, 196 genes were identified after examination of the regions within 100kbp of each of these SNP. Genes involved in immune response, diarrhoea, and wool/hair growth are reported. There were no genes within the nominal $P<0.001$ threshold that were associated with both flystrike and dag score or both flystrike and FDCV.

6.2 Introduction

Identifying single nucleotide polymorphism (SNPs) or quantitative trait loci (QTL) associated with disease traits is the central purpose of many human genetic studies. Genome-wide association studies (GWAS) are a popular tool for analysing disease traits, and identifying these SNPs and loci (Hirschhorn and Daly, 2005). Traditional methods involved linkage studies, candidate gene studies, and microarrays to identify QTL. However this approach was problematic as the markers were widely spaced necessitating follow up studies via sequencing and fine mapping (Hirschhorn and Daly, 2005), and QTL with small effects were easily missed. To mitigate this, and due to the advent of sequencing and the identification of SNPs across all of the genome, the current standard approach is to genotype large numbers of SNPs, many thousand to millions, using genome arrays across individuals with the disease (cases) and healthy individuals (controls). This has been very successful for disease traits in humans e.g. cancer and Type II diabetes (Sladek et al., 2007; Yeager et al., 2007).

Flystrike is a significant disease for sheep farmers, but is seasonally dependent. Using the case-control data set described in Chapter 2 it was found to be at a prevalence of only 2.15% of lambs on average across 11 farms. It was also estimated to be moderately heritable; 0.38 ± 0.10 on the observed scale and 0.27 ± 0.05 on the underlying liability scale, and had high genetic (0.71) and phenotypic (0.62) correlations with dag score, and a high genetic correlation with FDCV (0.74) (Chapter 4). This dataset was thus a good candidate for further analysis techniques such as GWAS. Only loci with major effects which were at moderate frequency could be detected, given the size of the dataset.

The objectives of this study were to identify genomic regions associated with flystrike and associated traits dag score, breech bareness, wool bulk, wool length, mean fibre diameter (MFD), standard deviation of MFD (FDSD), coefficient of variation of MFD (FDCV), curvature (CURV), and proportion medullation (MED%), and to see if there were any regions represented by more than 1 trait.

6.3 Materials and methods

6.3.1 Phenotypes

The flystrike case and control resource described in chapter 2, was used for this analysis. In summary 1,581 samples were collected from lambs (836 cases and 745 controls) from 11 properties over 2 flystrike seasons. Phenotypic records were available on flystrike, flystrike location, dag score, breech bareness score (BBREECH), and fibre traits, mean fibre diameter (MFD, μm), standard deviation of MFD (FDSD, μm), coefficient of variation of MFD (FDCV, %), proportion of medullated fibres (MED, %), and curvature (CURV, °/mm). Wool length (LENGTH) and wool bulk (BULK) measures were available for lambs similar to both resources described in Chapter 2 and 4 (n = 216). Additional information as described in chapter 2, including; birth rearing rank (brr), age of dam (aod), sex, flock, birth year, flystrike contemporary group (cg, flock by year) and birthday deviation was available for a subset of lambs from fully recorded flocks.

6.3.2 Genotypes

The 5K SNP panel genotypes were available for the cases and controls. The 50K SNP panel genotypes were obtained by imputation, using 50K SNP panel genotypes from sires of the cases and controls, and unrelated Romney-based animals from industry. Chapter 5 describes the methods and results of paternity assignment and imputation of the cases and controls. In summary 47,011 imputed SNP genotypes from BEAGLEv3.0 were available for 1,542 cases and controls; 792 affected lambs (cases) and 750 controls.

6.3.3 Analysis

6.3.3.1 Power analysis

A power analysis was conducted before collection of cases and controls (Chapter 2) to estimate the sample size needed to obtain a desirable level of power in a GWAS. Power calculations were done using Quanto 1.1 (Gauderman and Morrison, 2006), for a dominant and recessive model. The model used was for a two-sided test, with gene only (i.e. no gene by environmental effects), equal numbers of unmatched case-controls and was run over a range of disease allele frequencies (0.05 to 0.30), baseline risk levels (0.03 to 0.05), genetic effect relative risk (odds ratio; 2 to 10) and sample size (333, 667 and 1000 cases) to generate flystrike prevalence in the range of 0.05 to 0.10 and an h^2 up to ~0.12. The type I error rate was set at 5% Bonferroni-corrected (Rice, 1989) i.e. a nominal value of $1x10^{-6}$.

6.3.3.2 Genomic relationship matrix

A genomic relationship matrix was estimated from the SNP markers to structure the variance of the direct genetic random effect.

The genomic relationship matrix was calculated using the equation (VanRaden, 2008):

$$G = \frac{ZZ'}{2\sum p_i(1-p_i)} \qquad \text{Equation 1}$$

Where Z is the SNP matrix $-2(p_i-0.5)$, $1-2(p_i-0.5)$, and $2-2(p_i-0.5)$ for BB, AB, and AA respectively, and p_i is the frequency of A allele of the i^{th} SNP in the population.

The G matrix was used for the pedigree structure with the polygenic model, as described below. The first 4 principal components of the G matrix were plotted for quality assurance.

6.3.3.3 Polygenic model and estimation of P values

The GWAS followed a 2 step procedure, performed using R 2.14.1 (R Development Core Team, 2011). Firstly, a polygenic model was fitted with genomic relationship and fixed heritability,

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e} \qquad \text{Equation 2}$$

Where **X** is an incidence matrix or contains covariates, **β** is a vector of fixed effects, **Z** is an incidence matrix, **u** is a vector of additive genetic effects, with var(**u**)=**G**σ^2_u, and **e** is the vector of residuals, σ^2_u is the additive genetic variance.

The models for each trait were those derived in Chapter 2. Principal components (PC) of the genomic relationship matrix were not fitted. Secondly, the resulting residuals were analysed using an ordinary least square linear regression, fitting each SNP consecutively, to obtain the *P* values.

6.3.3.4 QQ plots

The quantile - quantile (QQ) plots were calculated to check whether the distribution of the observed -$\log_{10}$(*P*) values deviated from the expected distribution (exponential) under the null hypothesis of no genetic association and no LD between SNPs. To do so, the *n* -$\log_{10}$(*P*) values were sorted and plotted against the -$\log_{10}$(1-u) where u = [1 / (n + 1), 2 / (n + 1), …, n / (n + 1)]. If the distributions are similar then the slope, lambda, should be equal to 1. If the observed distribution deviates from the expected distribution with a slope greater than 1, it generally indicates that a proportion of the substructure has not been accounted for. A slope smaller than 1 is more difficult to explain, but could mean that the LD between SNPs and between SNPs and QTL is reducing the number of effective tests performed, thus producing larger *P* values than expected. When the right-hand tail diverges highly upwards from the slope, these SNPs may be considered associated with the trait.

6.3.3.5 Identifying relevant SNP markers

Marker associations were identified by graphing the resulting *P* values as the equivalent -$\log_{10}$(*P*) values in a Manhattan plot. A conservative threshold can be

calculated using the 5% Bonferroni correction (Rice, 1989) 0.05/ n_{SNP}, where n_{SNP} is the number of genotyped SNPs, equal to *P* value of $1.03x10^{-6}$ or -log10(*P*) of ~6. None of the traits reached this nominal conservative threshold, therefore the threshold was trimmed to $P < 0.001$ ($-\log_{10}(P) = 3$). The SNPs that reached this threshold were aligned to the Ovine genome v1.0 map (www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oar1.0/) to identify and report the underlying regions/genes.

6.4 Results

6.4.1 Power analysis

The results of the power analysis identified that the different genetic models resulted in little difference in power for detecting QTL of a certain h^2 when the sample size consists of 1000 cases (Tables 6.1 and 6.2, Figure 6.1). If the parameters used are true, then it is expected that significant results would be obtainable at least 67% of the time if the heritability explained by the QTL was greater than 0.03. The aim therefore was to collect at least 1000 cases and 1000 matched controls.

Table 6.1: Power of analysis for given sample size (n), and parameters[1] for a dominant model.

							Power for n		
p_0	q	r	p_1	Q	k	h^2	333	667	1000
0.05	0.05	2	0.095	0.098	0.054	0.004	0.000	0.001	0.002
0.05	0.05	4	0.174	0.098	0.062	0.023	0.008	0.086	0.281
0.05	0.05	10	0.345	0.098	0.079	0.105	0.282	0.887	0.995
0.04	0.10	2	0.077	0.190	0.047	0.005	0.000	0.003	0.010
0.04	0.10	4	0.143	0.190	0.060	0.029	0.036	0.299	0.675
0.04	0.10	10	0.294	0.190	0.088	0.123	0.592	0.991	0.999
0.03	0.30	2	0.058	0.510	0.044	0.005	0.001	0.006	0.025
0.03	0.30	4	0.110	0.510	0.071	0.024	0.046	0.360	0.746
0.03	0.30	10	0.236	0.510	0.135	0.091	0.483	0.975	0.999

[1] preset values: p_0: probability of flystrike for non-carrier; q: allele frequency of susceptibility allele; r: genetic effect relative risk; values generated by Quanto: p_1: probability of flystrike for carrier; Q: frequency of carriers; k: prevalence of flystrike.

[2] heritability of QTL.

Table 6.2: Power of analysis for given sample size (n), and parameters[1] for a recessive model.

							Power for n		
p_0	q	r	p_1	Q	k	h^2	333	667	1000
0.05	0.3	2	0.095	0.090	0.054	0.003	0.000	0.001	0.002
0.05	0.3	4	0.174	0.090	0.061	0.022	0.007	0.071	0.241
0.05	0.3	10	0.345	0.090	0.077	0.101	0.247	0.855	0.992
0.045	0.5	2	0.086	0.250	0.055	0.006	0.001	0.004	0.017
0.045	0.5	4	0.159	0.250	0.073	0.036	0.051	0.387	0.773
0.045	0.5	10	0.320	0.250	0.114	0.141	0.651	0.995	0.999
0.035	0.7	2	0.068	0.490	0.051	0.005	0.001	0.007	0.026
0.035	0.7	4	0.127	0.490	0.080	0.029	0.050	0.383	0.768
0.035	0.7	10	0.266	0.490	0.148	0.106	0.524	0.983	0.999

[1] preset values: p_0: probability of flystrike for non-carrier; q: allele frequency of susceptibility allele; r: genetic effect relative risk; values generated by Quanto: p_1: probability of flystrike for carrier; Q: frequency of carriers; k: prevalence of flystrike.

[2] heritability of QTL.

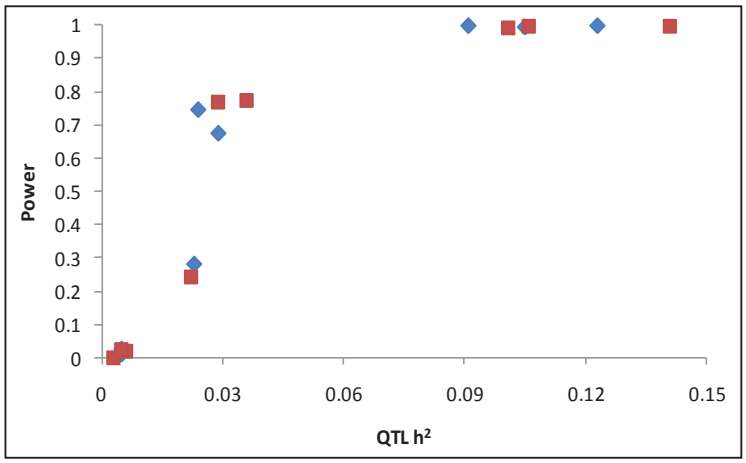


Figure 6.1 Plot of the power for a dominant (blue) and recessive (red) model against the heritability (h^2) of the QTL.

6.4.2 Principal component analysis

The PC plot shows a distinction between farmers, even though the animals are of Romney base (Figure 6.2). Farmers G and I (red and grey) are highly genetically linked as Farmer G is the sole ram provider for farmer I. Farmer E and F (blue and pink) are

also genetically linked with Farmer E using Farmer F's rams. Farmer G is distinct from the rest being from the Auckland Romney Development Group (ARDG) of breeders. Farmers D and H are Wairarapa Romney Development Group (WRIG) breeders (lower North Island). Farmer D has used ARDG rams in his flock. Farmers J, A, C and B also use sires from WRIG Breeders. Farmer K is from Geraldine (South Canterbury) and plots indicate greater links to lower North Island genetics than ARDG genetics. With these distinctions between farmers, they are however a tight group as the axes indicate.

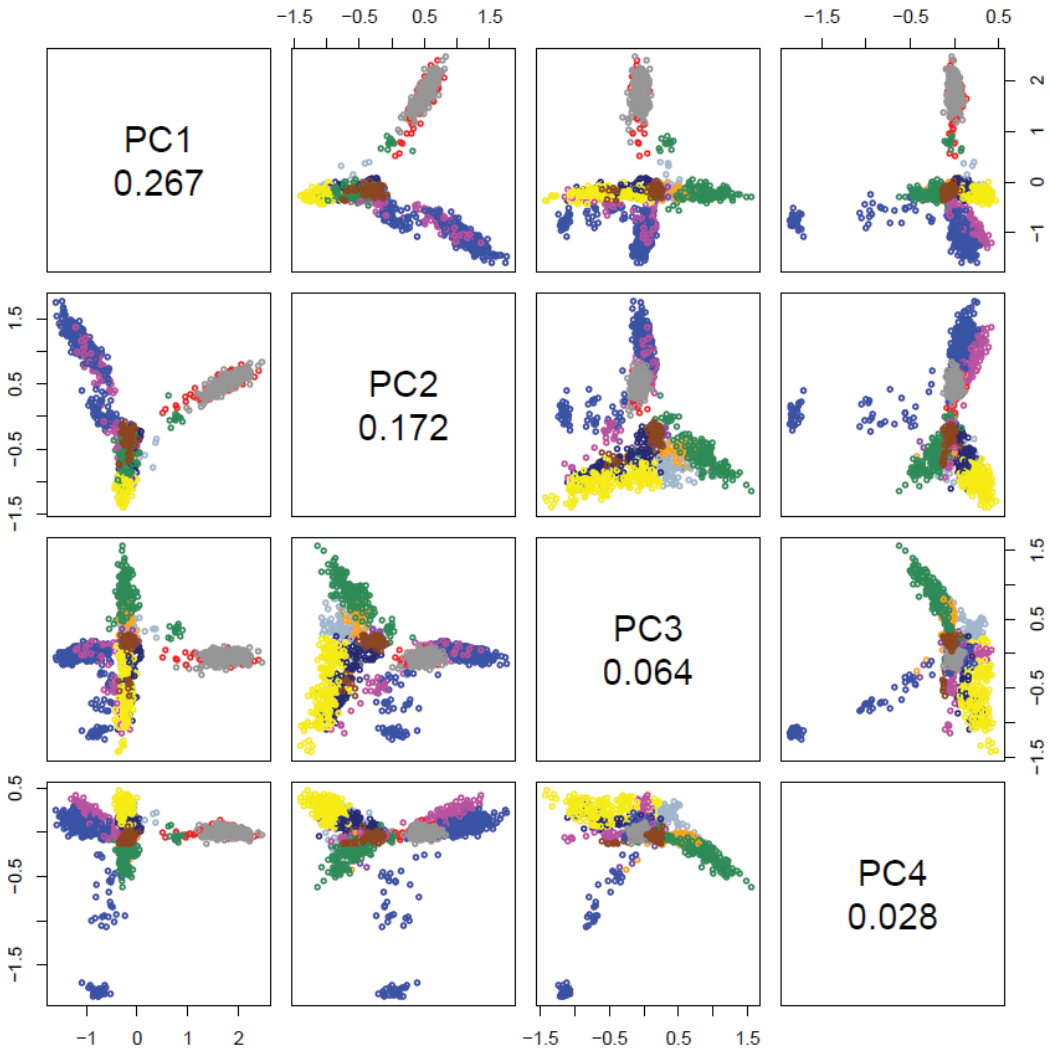


Figure 6.2: Plot of the first 4 principal components (PC) for flystrike case and control derived from genotype information. Coloured by farmer, see Chapter 2. Farmer A (orange), B (yellow), C (Dark blue), D (green), E (blue), F (pink), G (red), H (light grey), I (grey) and K (brown).

6.4.3 GWAS

6.4.3.1 QQ plots of resulting P values

The -$\log_{10}(P)$ values resulting from the GWAS were plotted in a QQ plot against the expected values using an exponential distribution under the null hypothesis of no genetic association and no LD between SNPs. The slopes (lambda) of the observed vs expected -$\log_{10}(P)$ were estimated (Table 6.3). The QQ plots for flystrike and dag score are shown in Figure 6.3, with other traits in Appendix 4. All estimates of lambda are below 1 and range between 0.75 and 0.90. The most likely reason that lambda is below 1 is that there is LD between the SNPs, and between SNPs and QTLs. This reduces the effective number of tests less than the total number of SNPs analysed, thus reducing the observed -$\log_{10}(P)$ values. The traits CURV and MED% have right-hand tail SNPs above the 0-1 line, indicating the best SNP for CURV and the best 3 SNPs for MED% are significantly associated with these traits.

Table 6.3: Estimate of lambda (slope), and their standard error (s.e.) of the linear regression of the observed -$\log_{10}(P)$ on the expected -$\log_{10}(P)$ (QQ plot) for each trait.

Trait	Lambda	s.e.
Flystrike	0.798	0.00004
Dag score	0.856	0.00005
Breech bareness	0.796	0.00004
Wool bulk	0.838	0.00005
Wool length	0.799	0.00008
Mean fibre diameter (MFD)	0.752	0.00009
Standard deviation of MFD	0.873	0.00007
Coefficient of variation of MFD	0.846	0.00010
Curvature	0.855	0.00007
Proportion of medullated fibres	0.917	0.00019

.

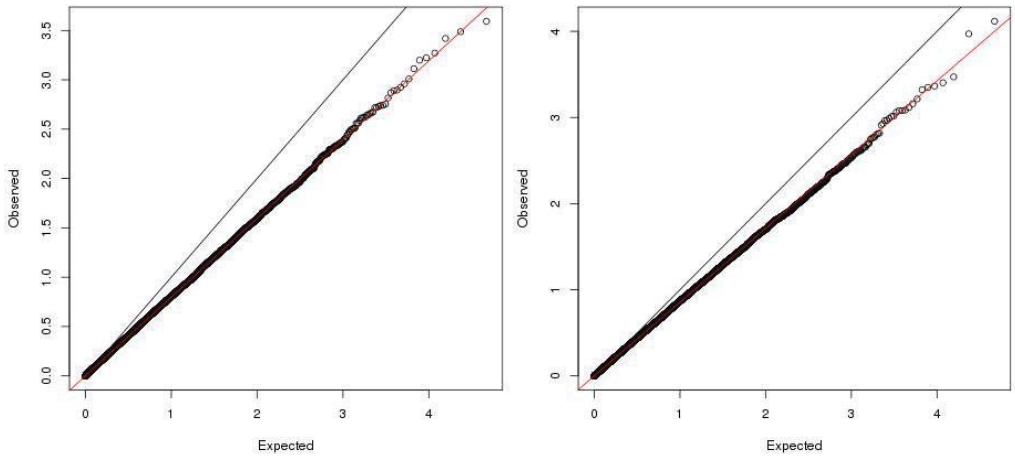


Figure 6.3: QQ plots for flystrike (left) and dag score (right) -log$_{10}$(*P*) values. The 0-1 line is in black and the linear regression in red.

6.4.3.2 Significant SNPs for traits flystrike, dag score, BBREECH, BULK and LENGTH

There were no SNPs below the Bonferroni significance threshold of $P < 1.03 \times 10^{-6}$ for the traits, flystrike, dag score, BBREECH, BULK or LENGTH. At the nominal significant threshold of $P < 0.001$ there were 7 SNPs that reached significance for flystrike, with the best SNP on chromosome 3 (Figure 6.4). There are 16 SNPs that were below the nominal threshold P<0.001 for dag score, with the best 2 SNPs on chromosome 6 and 15 (Figure 6.5). There are 5 SNPs for BBREECH, 9 for BULK and 2 for LENGTH that were below the nominal threshold of P<0.001 (Figure 6.6 to 6.8). The best 5 SNPs for each of these traits are in Table 6.4.

6.4.3.3 Significant SNPs for fibre traits MFD, FDSD, FDCV, CURV and MED%.

There were no SNPs below the Bonferroni significance threshold of $P < 1.03 \times 10^{-6}$ for the fibre traits; MFD, FDSD, FDCV, CURV and MED%. At the nominal significant threshold of P<0.001 there were 4 SNPs for MFD, 15 for FDSD, 9 for FDCV, 17 for CURV and 42 for MED% (Figures 6.9 to 6.13). Table 6.5 shows the best 5 SNPs for each of these traits. Comparing the best 20 SNPs for each trait there are a few shared SNPs. The second best SNP for FDSD is the 6th best SNP for MED%. The 3rd and 4th SNPs for FDSD were also the 10th and 19th SNPs for FDCV and MED% and the 19th SNP for FDCV is the 4th SNP for MED% (Appendix 5).

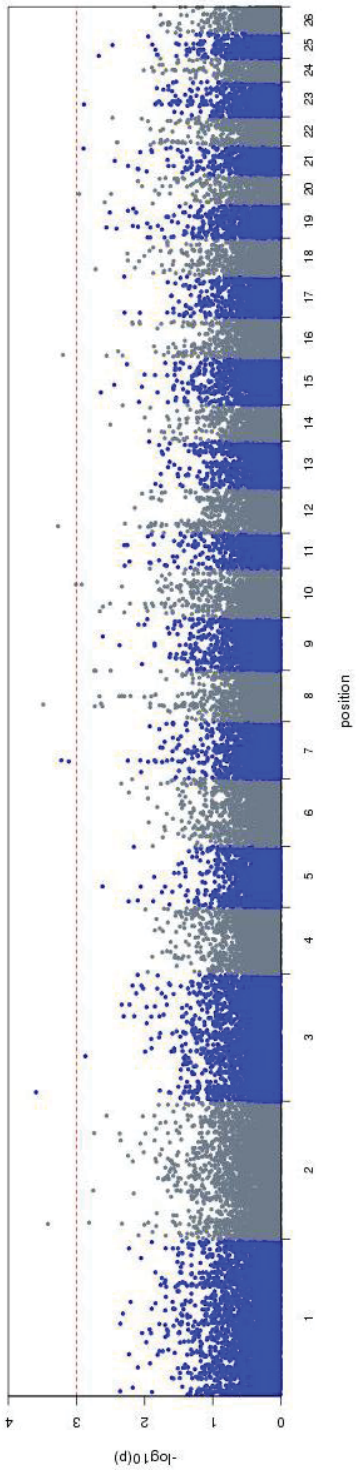


Figure 6.4: Graph of -log$_{10}$(*P*) values of SNPs for flystrike, ordered on ovine genome v1 map, P<0.001 (red dash line).

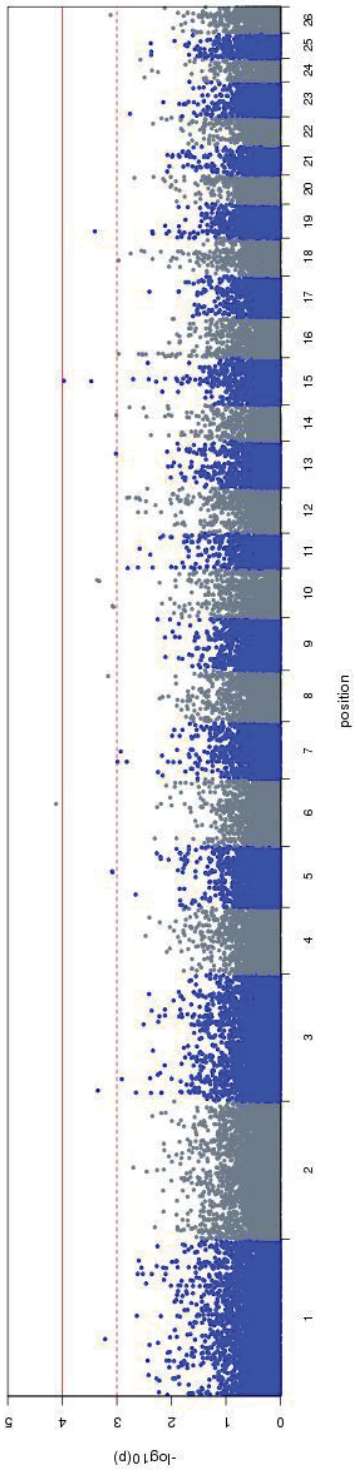


Figure 6.5: Graph of -log$_{10}$(*P*) values of SNPs for dag score, ordered on ovine genome v1 map, P<0.0001 (solid red line), P<0.001 (red dash line).

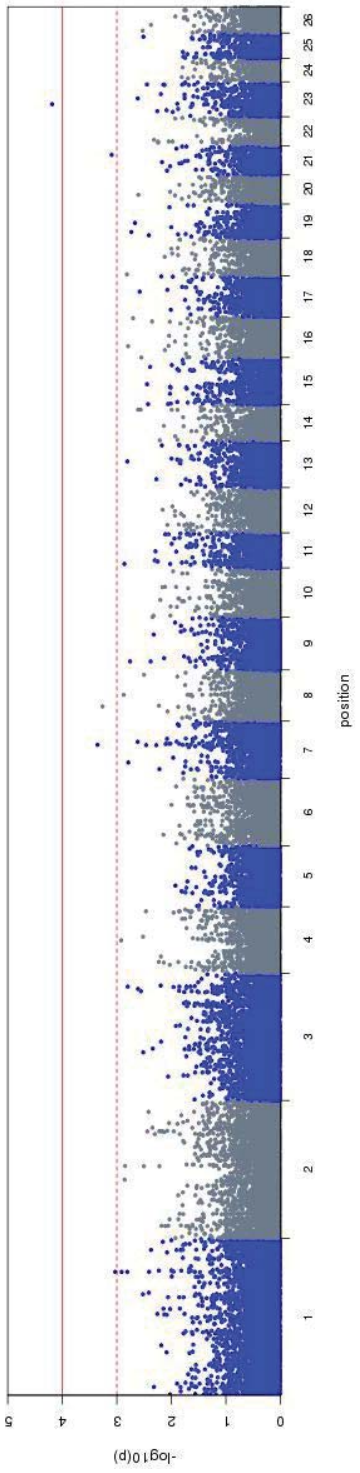


Figure 6.6: Graph of -log$_{10}$(*P*) values of SNPs for breech bareness score, ordered on ovine genome v1 map, P<0.0001 (solid red line), P<0.001 (red dash line).

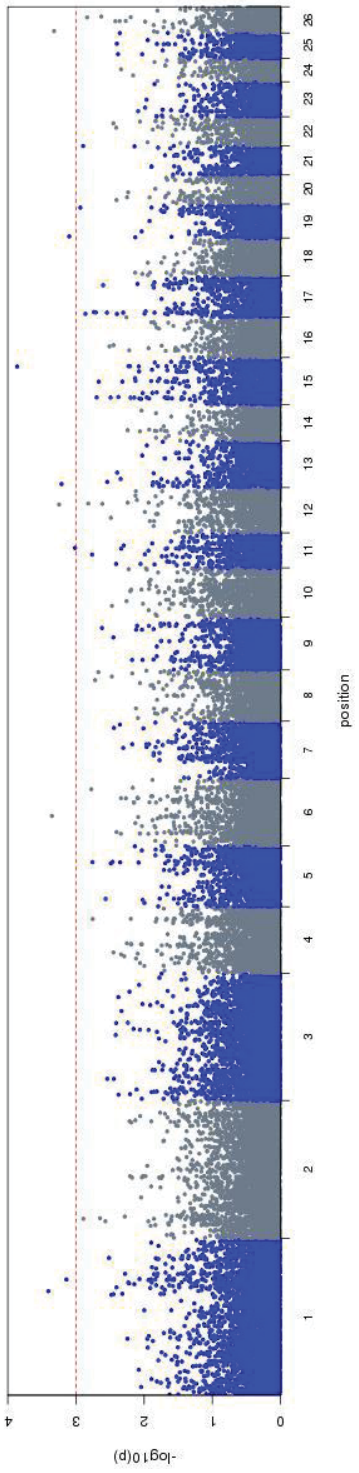


Figure 6.7: Graph of -log$_{10}$(*P*) values of SNPs for wool bulk, ordered on ovine genome v1 map, P<0.001 (red dash line).

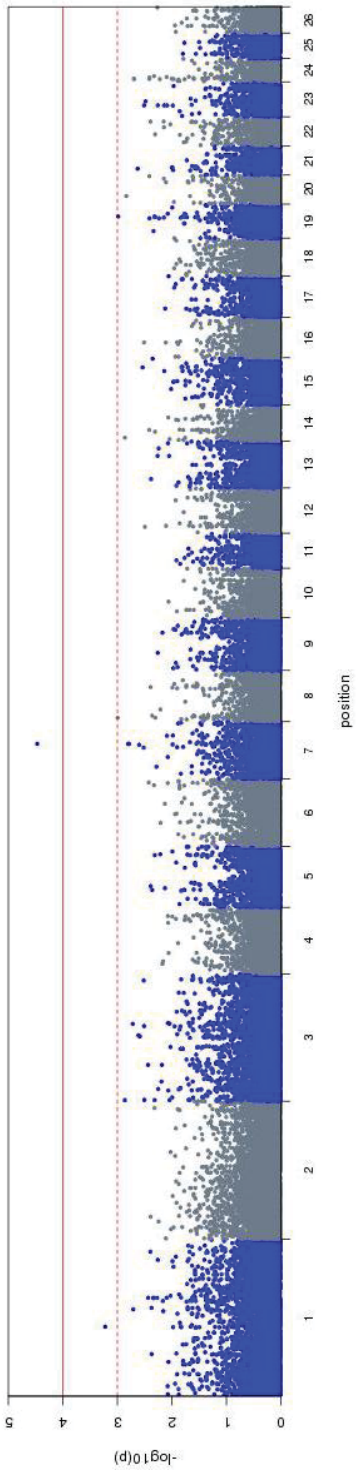


Figure 6.8: Graph of -log$_{10}$(*P*) values of SNPs for wool length, ordered on ovine genome v1 map, P<0.0001 (solid red line), P<0.001 (red dash line).

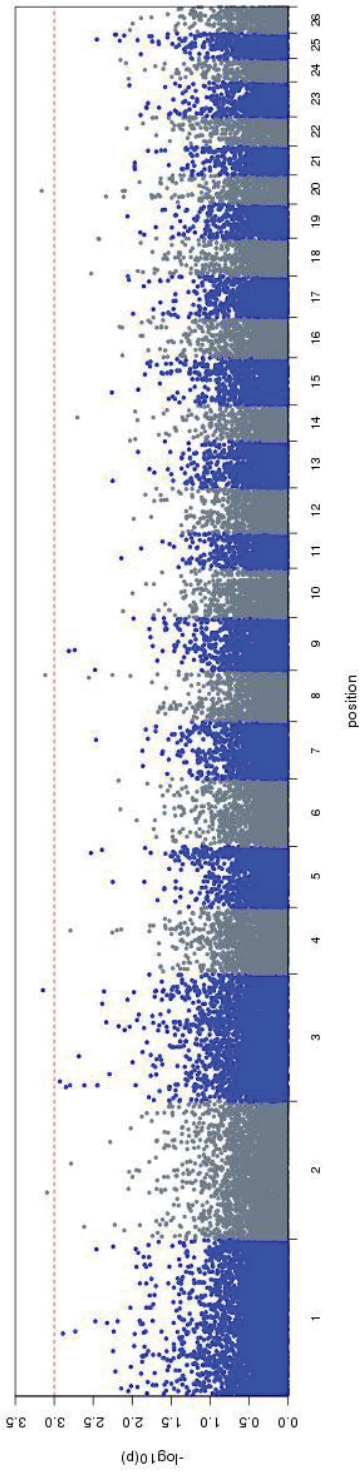


Figure 6.9: Graph of -log$_{10}$(*P*) values of SNPs for mean fibre deviation, ordered on ovine genome v1 map, P<0.001 (red dash line).

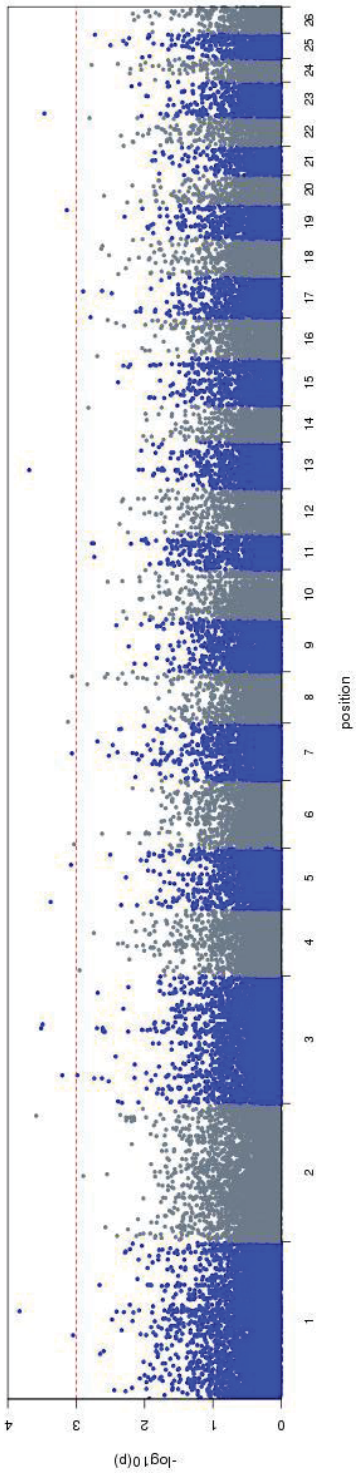


Figure 6.10: Graph of -log$_{10}$(*P*) values of SNPs for standard deviation of mean fibre deviation, ordered on ovine genome v1 map, P<0.001 (red dash line).

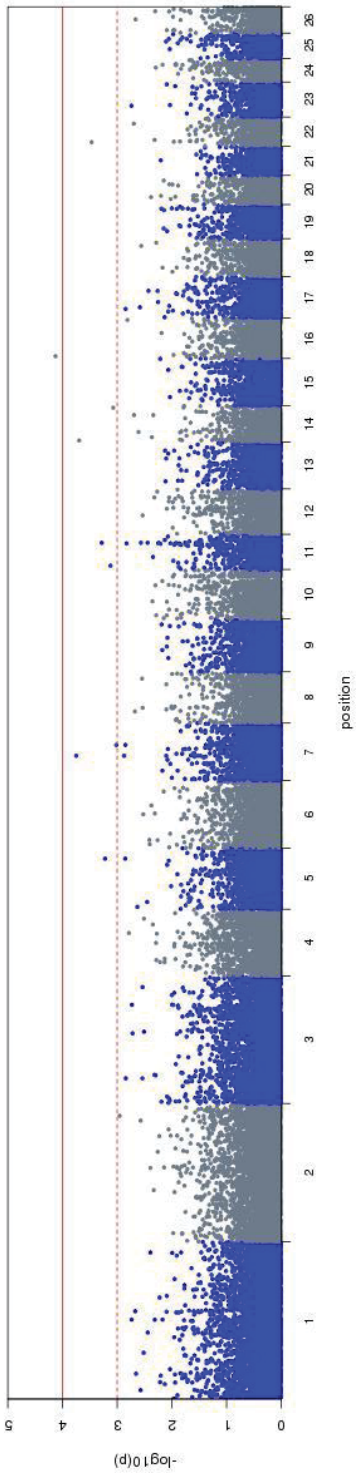


Figure 6.11: Graph of -log$_{10}$(*P*) values of SNPs for coefficient of variation of mean fibre diameter score, ordered on ovine genome v1 map, P<0.0001 (solid red line), P<0.001 (red dash line).

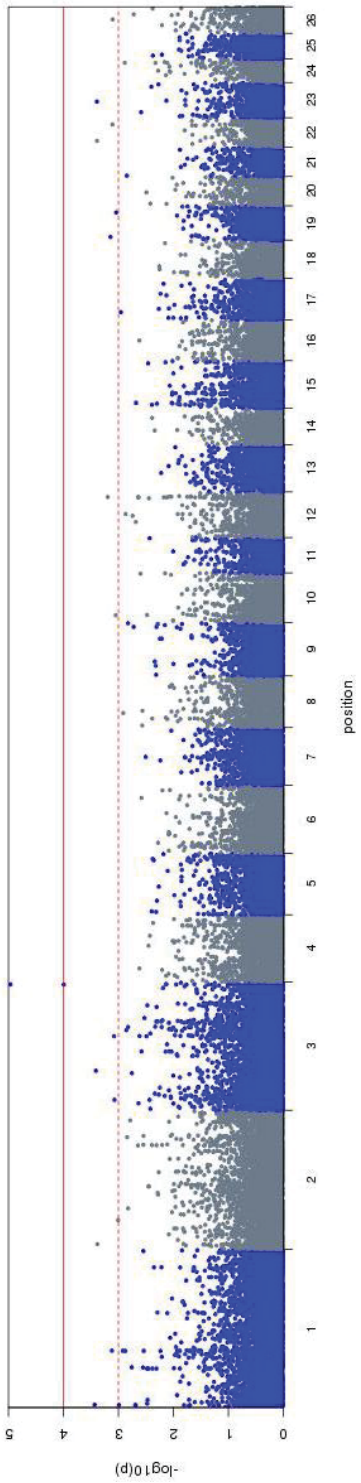


Figure 6.12: Graph of -log$_{10}$(*P*) values of SNPs for curvature, ordered on ovine genome v1 map, P<0.0001 (solid red line), P<0.001 (red dash line).

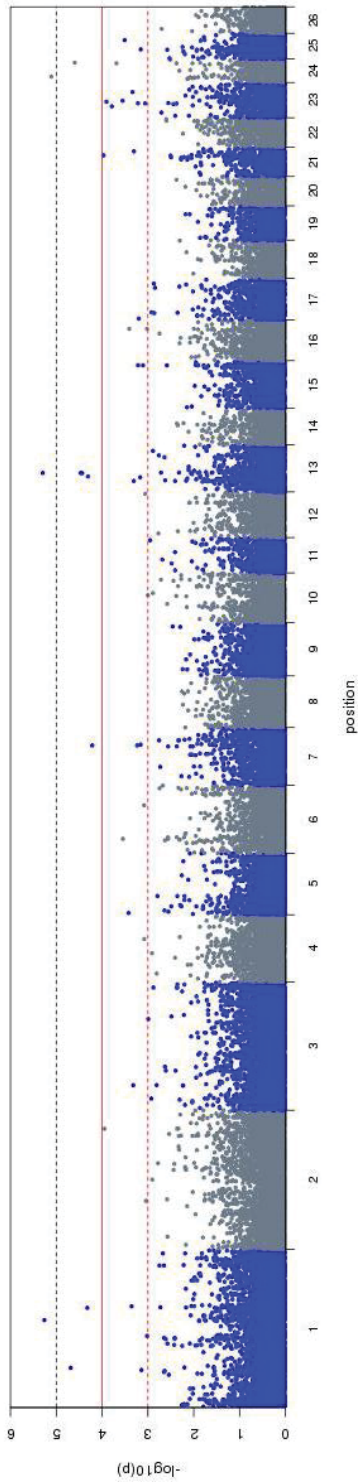


Figure 6.13: Graph of -log$_{10}$(*P*) values of SNPs for proportion medullation score, ordered on ovine genome v1 map, P<0.00001 (black dash line), P<0.0001 (solid red line), P<0.001 (red dash line).

Table 6.4: The trait, number of animals (n), marker name, chromosome (Chr) and chromosome position (base pairs, bp) of the best 5 SNPs for the traits flystrike (as log_{10}(scaled flystrike +1)), dag score (DAG), breech bareness (BBREECH), wool bulk (BULK) and wool length (LENGTH). The *P* value and -log_{10}(*P*) are shown for each marker.

Trait[1]	N	Marker name	Chr	Position (bp)	*P* value	-log_{10}(*P*)[2]
Flystrike	1469	s43785	3	17026008	2.55E-04	3.593
		s50983	8	32402380	3.25E-04	3.488
		OAR2_28934548	2	28934548	3.81E-04	3.419
		OAR12_13334889	12	13334889	5.36E-04	3.271
		OAR7_35146905	7	35146905	5.98E-04	3.223
DAG	1377	OAR6_80988051	6	80988051	7.64E-05	4.117
		s20575	15	45970082	1.07E-04	3.971
		s71120	15	45195724	3.38E-04	3.471
		s08710	19	13029558	3.95E-04	3.403
		OAR10_71200463	10	71200463	4.32E-04	3.365
BBREECH	1308	OAR23_23803976	23	23803976	6.50E-05	4.187
		OAR7_63848145	7	63848145	4.43E-04	3.354
		OAR8_27871953	8	27871953	5.46E-04	3.263
		s60427	21	38184195	8.02E-04	3.096
		s51351	1	235438491	9.33E-04	3.030
BULK	216	OAR15_72588676	15	72588676	1.37E-04	3.863
		OAR1_198827329	1	198827329	3.93E-04	3.406
		OAR6_56978564	6	56978564	4.44E-04	3.353
		s38270	26	3703695	4.77E-04	3.322
		s35163	12	54083096	5.65E-04	3.248
LENGTH	216	OAR7_66581251	7	66581251	3.40E-05	4.469
		s31436	1	132497619	5.99E-04	3.222
		OAR8_7413004	8	7413004	0.0010	2.995
		OAR19_4127468	19	41274682	0.0010	2.985
		s19378	3	1931918	0.0014	2.864

[1]BBREECH: breech bareness; BULK: wool bulk, LENGTH: wool length.

[2] Bonferroni significance $P < 10^{-6}$ or -log_{10}(*P*) = 6, nominal significance $P < 0.001$ or -log_{10}(*P*) = 3.

Table 6.5: The trait, number of animals (n), marker name, chromosome (Chr) and chromosome position (base pairs, bp) of the best 5 SNPs for the fibre traits; mean fibre diameter (MFD), standard deviation of MFD (FDSD), coefficient of variation of MFD (FDCV), curvature (CURV) and proportion medullation (MED%). The *P* value and -log$_{10}$(*P*) are shown for each marker.

Trait[1]	n	Marker name	Chr	Position (bp)	*P* value	-log$_{10}$(*P*)[2]
MFD	1237	OAR3_211858595	3	211858595	0.0002	3.64
		OAR20_25523915	20	25523915	0.0003	3.49
		s33242	1	119335876	0.0009	3.04
		OAR2_88811262	2	88811262	0.0010	3.00
		s54719	14	45266387	0.0012	2.91
FDSD	1237	OAR1_167004318	1	167004318	0.0001	3.83
		OAR13_35686621	13	35686621	0.0002	3.69
		s42564	2	239459013	0.0003	3.58
		s41365	3	143149207	0.0003	3.51
		OAR3_150573902	3	150573902	0.0003	3.49
FDCV	1237	s46992	16	4216284	7.4E-05	4.13
		s53860	7	46907656	0.0002	3.75
		OAR14_3089361	14	3089361	0.0002	3.70
		OAR22_7858171	22	7858171	0.0003	3.47
		s22354	11	50563464	0.0005	3.28
CURV	1237	s57836	3	238116985	1.1E-05	4.96
		OAR3_237829296	3	237829296	0.0001	3.99
		s62307	1	5259866	0.0004	3.43
		OAR3_74732440	3	74732440	0.0004	3.41
		OAR23_30585446	23	30585446	0.0004	3.39
MED%	1230	OAR13_35759190	13	35759190	5.1E-06	5.30
		s58710	1	165442356	5.5E-06	5.26
		s45778	24	11146202	7.8E-06	5.11
		s36857	1	75085887	2.1E-05	4.69
		OAR24_37414016	24	37414016	2.5E-05	4.60

[1]MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

[2] Bonferroni significance $P < 10^{-6}$ or -log$_{10}$(*P*) = 6, nominal significance $P < 0.001$ or -log$_{10}$(*P*) = 3.

6.4.4 Identification of candidate genes

Ovine genome v1.0 (www.livestockgenomics.csiro.au/cgi-bin/gbrowse/oar1.0/) was used to identify genes within 100kbp of the best SNPs that were below the nominal threshold of $P<0.001$. The window of 100kbp was selected based on the known level of LD within the sheep genome, with it being unlikely that the actual loci would be more distant. There were 115 individual SNPs below the nominal significance threshold. Examination of the regions within 100kbp of each individual SNP identified 196 genes

(Appendix 5). Due to the high genetic correlations of flystrike with dag score and FDCV, the genes within the significant regions identified for each trait were compared with each other. There were no genes that were represented by more than 1 trait for flystrike, dag score, or FDCV. Genes involved in immune response, diarrhoea, and wool/hair growth were investigated further (Tables 6.6 to 6.8).

6.4.4.1 Immune, inflammatory and diarrhoea genes

For flystrike, 1 gene with a role in immune status was identified (nominal P = 0.0006 for nearby SNP). This was RAS-guanyl releasing protein1 (RASGRP1), located on chromosome 7 (Table 6.6). For dag score, a number of genes with inflammatory, antibacterial and/or diarrhoea phenotypes were identified. These were signal peptide CUB domain and EGF-like 2 (SCUBE2), A disintegrin and metallopeptidase domain 17 (ADAM17), defensin β 119 and 121 (DEFB119, DEFB121) and serine peptidase inhibitor Kunitz type 2 (SPINT2) (Table 6.6). See section 6.5.3 for a description of these genes.

6.4.4.2 Wool/hair genes

A number of genes associated with keratinocytes, epidermis layer or wool/hair growth were identified for breech bareness and the fibre traits. These included: for bare breech; fibroblast growth factor 7 (FGF7), and WNT1 inducible signalling pathway protein 3 (WISP3), for FDSD; keratin genes (KRT84, 85, 86, 7, 83, 8, 80 and 81), for MED %; dishevelled associated activator of morphogenesis 1 (DAAM1), desmoglein 4 and desmoglein 1 (DSG4, DSG1) (Table 6.7 and 6.8). There were also 2 genes that were associated for both MED% and FDSD, by three different SNPs: enhancer of polycomb homolog 1 (EPC1) and kinesin family member 5B (KIF5B). See section 6.5.3 for a description of these genes.

Table 6.6: List of SNPs identified with immune, inflammatory response and diarrhoea roles (greyed) for the traits flystrike and dag score, which reached nominal significant threshold *P*<0.001.

Chr	SNP name	100kbp region	*P* value	Gene	OMIM	Description	Trait
				CPSF3	606029		
				ADAM17	603639	Protease	
				YWHAQ	609009	Adapter protein	
3	s62291	20195292-20395291	0.00045	RPL36A	180469	Ribosomal protein	Dag score
7	OAR7_35146905	35046905-35246904	0.00060	RASGRP1	603962		Flystrike
				DEFB119		Defensin	
13	OAR13_65478113	65378113-65578112	0.00096	DEFB121		Defensin	Dag score
				SIPA1L3			
				DPF1	601670	Part of neuron-specific chromatin remodelling complex	
				PPP1R14A	608153	Protein phosphatase	
				SPINT2	605124	Serine peptidase inhibitor	
14	OAR14_49536713	49436713-49636712	0.00098	YIF1B		Transmembrane protein	Dag score
				DENND5A		GTPase signalling	
				SCUBE2	611747	Signal peptide	
15	s71120	45195724-45195724	0.00034	NRIP3	613125	Nuclear receptor interactor	Dag score

Table 6.7: List of SNPs identified with wool/hair roles (greyed) for breech bareness (BBREECH), mean fibre diameter standard deviation (FDSD) and proportion medullation (MED%) for chromosome 3, 7 and 8 which reached the nominal significant threshold *P*<0.001.

Chr	SNP name	100kbp region	*P* value	Gene	OMIM		Trait
3	s41365	143049207-143249206	0.00031	KRT84	602766	Keratin	FDSD
				KRT85	602767	Keratin	
				KRT86	601928	Keratin	
				KRT7	148059	Keratin	
				KRT83	602765	Keratin	
				KRT8	148060	Keratin	
				KRT80	611161	Keratin	
				KRT81	602153	Keratin	
7	OAR7_63848145	63748145-63948144	0.00044	FGF7	148180	Fibroblast growth factor	BBREECH
				GALK2	137028	Galactokinase	
	OAR7_74984216	74884216-75084215	6.08E-05	DAAM1	606626	Dishvelled activator	MED%
	OAR7_75022469	74922469-75122468	0.00058	DAAM1	606626	Dishvelled activator	MED%
8	OAR8_27871953	27771953-27971952	0.00055	LAMA4	600133	Glycoprotein	BBREECH
				C6orf225			
				TUBE1	607345	Tubulin	
				WISP3	603400	Wnt signalling	

Table 6.8: List of SNPs identified with wool/hair roles (greyed) for mean fibre diameter standard deviation (FDSD) and proportion medullation (MED%) on chromosome 13 and 23 which reached the nominal significant threshold *P*<0.001.

Chr	SNP name	100kbp region	*P* value	Gene	OMIM		Trait
13	OAR13_35499997	35399997-35599996	3.67E-05	CACNB2	600003	Calcium channel	MED%
				NSUN6		Methyltransferase	
				EPC1	610999	Part of histone acetyltransferase (HAT) complex	
	OAR13_35759190	35659190-35859189	5.05E-06	EPC1	610999	Part of histone acetyltransferase (HAT) complex	MED%
				KIF5B	602809	Kinesin	
	OAR13_35686621	35586621-35786620	3.42E-05	EPC1	610999	Part of histone acetyltransferase (HAT) complex	MED%
				KIF5B	602809	Kinesin	
			0.00021	EPC1	610999	Part of histone acetyltransferase (HAT) complex	FDSD
				KIF5B	602809	Kinesin	
23	OAR23_27233801_X	27133802-27333801	0.00061	DSG4	607892	Desmosomal cadherin	MED%
				DSG1	125670	Desmosomal cadherin	

6.5 Discussion

6.5.1 GWAS

The PCA plots (Figure 6.2) clearly separated out the ARDG breeder from the WRIG and South Island breeder. In this analysis, the PC was not fitted, as it was thought the genomic relationship matrix would account for the population structure as the lambs were all from Romney or Romney based flocks. The first PC explained 27% of the variation. The distribution of the observed $-\log_{10}(P)$ values were plotted against the expected distribution. When the slope, lambda, is greater than 1 this indicates that population substructure has not been completely accounted for. In this analysis, lambda estimates, from the QQ plots, were less than one. The lambda values below one indicate there is some LD between SNPs and between SNPs and QTLs, thus reducing the observed $-\log_{10}(P)$ values. There may also be some over-fitting of the data, and potentially underestimation of the SNP effects. However, it was felt that a conservative test with listing of SNPs below a nominal threshold would allow future researchers with larger datasets to compare and combine their results with those reported here.

Another issue when dealing with a disease such as flystrike is that the incidence rate is season dependent. Thus, the controls selected in this study, are confirmed as non-fly-struck only for the period up until they were collected. Bishop et al. (2012) identify that SNP effects are biased downwards linearly with exposure probability or inaccuracy of diagnosis of disease. Also, that mode of inheritance of the disease changes as the infection level of disease rises in the population. At low infection pressure, the inheritance mode is dominant, at high infection levels the inheritance mode is partially recessive, and for the remainder it is assumed as an additive model. In this analysis, an additive model was assumed, a count of the fly population present on the farms affected would be required to correctly estimate the infection pressure, however, it was assumed to be low to moderate in this study and thus an additive model assumption was acceptable.

Care is needed with experimental design, and at worst, when controls are a representation of the population average (occurs when prevalence is low) the analysis will be less powerful. However, this can be offset by sampling a large proportion of the population average (Bishop et al., 2012). Additionally if fewer cases are collected than required, as estimated from the power analysis, the chance of detecting QTL is reduced.

In this analysis it was estimated that to obtain the power needed to detect QTL 70% of the time 1,000 cases and 1,000 controls should be collected. However, only 836 cases and 745 controls were collected. As can be seen in Figure 6.14, there is a drop (0.70 to ~0.35) in power when the heritability explained by the QTL is ~0.03 if only 667 cases are collected.

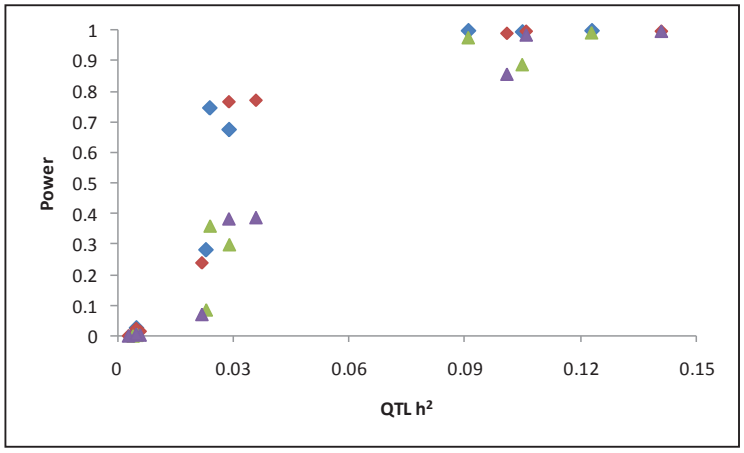


Figure 6.14 Plot of the power for a dominant (blue/green) and recessive (red/purple) model when 1,000 (diamond) or 667 (triangle) cases are collected against the heritability (h^2) of the QTL.

Thus, independent validation of the results of this study is needed, from independent datasets and using different methods of analysis. Cases should be collected during early exposure, as these are the most susceptible, and controls should be taken after peak prevalence, and across-flock information could be incorporated (Bishop et al., 2012). This was in general performed here, when the controls were taken at the end of the flystrike season after they had had full opportunity of being exposed and becoming infected. To improve the power of the analysis another season of collections needed to be undertaken to obtain at least 1,000 cases. The current design could also be improved by increasing the numbers of the controls, the population average, as flystrike prevalence is generally low in New Zealand dual-purpose sheep.

The results could be validated by using other analytical methods. It has been suggested that the single SNP regression approach, as used in this analysis, may not capture all of the genetic variance present in the dataset (Hoggart et al., 2008; Lee et al., 2008). An alternative was to fit SNPs simultaneously using a Bayesian approach

(Hoggart et al., 2008), which has increased power, reduced false positive rates and improved localisation of peaks when compared to single SNP analyses. A single step approach was fitted here as the polygenic models were too complicated to be analysed using other available software. The Bayesian approach would most likely have found the same peaks as seen here, but perhaps with more significance. With the 50K SNP Chip it is unlikely that localisation of the peaks would have been improved. Both approaches would benefit from use of a higher density SNP chip.The methods used in this analysis are very similar to the package GenABEL (Aulchenko et al., 2007), but other packages such as PLINK (Purcell et al., 2007) and Golden Helix (www.goldenhelix.com/SNP_Variation) are alternatives.

6.5.2 Significant peaks

There were a large number of SNPs that reached the nominal threshold of $P<0.001$ (n = 116). There were no SNPs that reached the conservative Bonferroni threshold of $P < 1.03x10^{-6}$. There are a number of genes identified that have roles in immune response, inflammatory stimuli, diarrhoea phenotypes and wool/hair. There were several genes identified involved with myelin, microtubule, fibroblast and actin, associated with the fibre traits. These were not investigated here, however, further investigation is required as there could be cross over between cell structure, muscle fibre and wool/hair fibre structure and growth. For example, the EPC1 gene was associated with three SNPs and 2 traits, MED% and FDSD, muscle fibre shape and variation within the muscle need to be controlled just like wool/hair fibres. Thus, genes that control MED%, a medullated fibre has a core of large fragile cells (may be filled with air) within a normal sheath (Henderson, 1968), and FDSD, variation in fibre diameter, could also control similar traits of muscle fibres. These genes should not be ruled out for a role in hair fibre structure.

Despite the high genetic correlation of dag score and FDCV with flystrike, there were no SNPs or genes, which reached nominal significance, identified for more than one of these traits. The genes affecting these traits possibly have small effect and thus were not identified here.

6.5.3 Description of candidate genes

A literature search and OMIM were used to identify and describe candidate genes identified under the 100kbp window surrounding the best SNPs for each trait. Below is a description of the genes identified in sections 6.4.4.1 and 6.4.4.2 above.

RAS-guanyl releasing protein1 (RASGRP1), located on chromosome 7 is involved in the T-Cell Receptor (TCR) signalling pathway, that enables the generation of a diverse collection of TCRs that allows response towards a vast array of foreign peptides bound to the major histocompatibility complex (MHC) (Priatel et al., 2007). Deficiency in RASGRP1 in mice studies leads to diminished T cell response and delayed pathogen clearance, similar to the functionality of exhausted memory T cells found during chronic infections (Priatel et al., 2007). Microarray studies for response to *Staphylococcus aureus* and *S. epidermidis* or *Escherichia coli* in mammary tissue of sheep (Bonnefont et al., 2011) and cattle (Brand et al., 2011) have identified RASGRP1 for differential expression.

SCUBE2 is a cell-surface protein expressed on the endothelial cells and in the heart, placenta, lung, pancreas, prostate and the small intestine (Yang et al., 2002). Along with its family member SCUBE1, it is down regulated by treatment with interleukin -1β and tumour necrosis factor-α (TNFα) *in vitro* and by lipopolysaccharide *in vivo* indicating a role in inflammatory response (Yang et al., 2002).

ADAM17 is involved in the release of membrane-anchored moieties from the cell-surface by proteolysis and has many substrates including TNFα, p75 TNF-receptor, and L-selectin among others (Peschon et al., 1998). ADAM17 is a target for treatment of TNF-dependent pathologies e.g. rheumatoid arthritis, Crohn's disease and psoriasis (Horiuchi et al., 2007). Loss of ADAM17 expression was found in a family with neonatal inflammatory skin and bowel disease (Blaydon et al., 2011).

The human β-defensins (DEFB) have antibacterial activity, are likely to be involved in innate immunity and are expressed by epithelial cells of tissues exposed to external environment (Yang et al., 2004), including the Paneth cells of intestine. DEFB119 and DEFB121 are expressed in the proximal epididymis (corpus and caput) and testis of human and macaque tissues, suggesting these DEFBs have evolved for reproduction-specific immunity and fertility roles (Radhakrishnan et al., 2005).

SPINT2 is an inhibitor of hepatocyte growth factor activator (HGFAC), a blood coagulation factor (Kawaguchi et al., 1997). The HGFAC is converted to an active

serine protease with strong heparin binding affinity in response to tissue injury. The inhibitor SPINT2 is highly expressed in the kidney, pancreas and other tissues and also expressed in small intestine, colon and lung. Autosomal-recessive congenital sodium diarrhoea has been associated with loss of function of SPINT2 (Heinz-Erian et al., 2009). The disease is characterised by metabolic acidosis, high faecal losses of sodium and watery diarrhoea.

Hair development is controlled by a number of signalling molecules including those of fibroblast growth factors (Kulessa et al., 2000). Rosenquist & Martin (1996) showed FGF7 was localised to the dermal papilla during anagen, the active growth phase of hair follicles.

Also involved in hair development are the signalling molecules bone morphogenetic protein (BMP) and Wnt (Kulessa et al., 2000). WISP3 inhibits BMP and Wnt signalling in mammalian cells by binding to lipoprotein receptor-related protein 6 and Frizzled (Nakamura et al., 2007), while frizzled genes are known to affect hair pattern in mice (Guo et al., 2004). However, WISP3 is associated with an autosomal-recessive skeletal disorder progressive pseudorheumatoid dysplasia (Nakamura et al., 2007), a bone cartilage disease.

Keratin and keratin-associated protein (KAP) genes encode most of the wool and hair proteins (Yu et al., 2011). They are highly sequence-conserved across species, but there are differences in where keratin genes are expressed. In sheep KRT40, 82 and 84 are present in the fibre cuticle, KRT32, 35 and 85 are expressed in cuticle and fibre cortex, while KRT31, 33A, 33B, 34, 36, 38-39, 81, 83 and 86-87 are expressed only in cortex (Yu et al., 2011).

DAAM1 is another gene involved in the Wnt signalling via the Frizzled receptor. Along with Dishevelled it activates Rho in cell polarity signal transduction during gastrulation (Habas et al., 2001).

Desmosomal cadherins are calcium dependent cell adhesion molecules in most epithelia (Jahoda et al., 2004; Rickman et al., 1999). They provide stability to the epithelial cells due to the interactions of the N-terminal extracellular domains with lateral strand dimers and adhesive dimers. A missense mutation in DSG4 results in the *lanceolate hair* phenotype of rats (Jahoda et al., 2004). The phenotype is characterised by the presence of sparse, fragile hair shafts which form a lance head at the tip. DSG4 is a mediator of keratinocyte cell adhesion, coordinating the transition from proliferation

to differentiation. DSG1 is an autoantigen in the autoimmune pemphigus foliaceous, a skin blistering disease (Rickman et al., 1999). The mutation in DSG1 associated with the disease is thought to compromise strand dimer formation.

The EPC1 gene regulates transcription with homeodomain only protein (HOP) to induce skeletal muscle differentiation (Kee et al., 2007). It is highly expressed in embryonic heart and adult skeletal muscles. The kinesin family are molecular motors converting ATP into a motile force along microtubules, indicating a role in motility of organelles towards the cell membrane (Tanaka et al., 1998). KIF5B is associated with the mitochondria indicating a role in mitochondrial transport in nonneuronal cells.

6.6 Conclusion

The number of cases and controls collected did not reach the number required as estimated from the power analysis, thus the chance of identifying QTL of small effect were reduced. However, a number of candidate genetic regions were identified for flystrike, dag score, breech bareness, and the wool and fibre traits in the GWAS. All traits reached the nominal significance threshold of $P<0.001$, but very few regions reached more stringent thresholds. An interesting selection of genes involved in immune response inflammatory stimuli, diarrhoea phenotypes and wool/hair development were identified. RASGRP1 associated with the best SNP for flystrike and ADAM17 associated with a best SNP for dag score are two genes that require validation in an independent analysis.

Further investigation should involve genotyping an independent set of flystrike cases and controls on the 50K SNP panel or at a higher density to validate the results from this study. Also, the independent dataset should involve a larger sampling of the population average as controls collected at the end of the flystrike season after all animals have been subjected to the full exposure present in the season.

Chapter 7: Genome-wide association study: SIL industry data including dagginess, bare breech and fibre traits

Foreword: A subset of this chapter was presented as a poster at the 33rd conference of the International Society of Animal Genetics held in Cairns, Australia 15-20th July 2012. Pickering, N. K., B. Auvray, H. T. Blair, R. E. Hickson, P. L. Johnson and J. C. McEwan. 2012. GWAS for dagginess in New Zealand dual-purpose sheep. P4041.

7.1 Abstract

To identify genetic regions associated with dagginess in the New Zealand population of dual-purpose sheep, a genome-wide association study (GWAS) was performed, using estimated breeding values as phenotypes, on pedigree-recorded animals born 1990-2010 (n ~ 3.5 million, 8,605 genotyped). A selection sweep analysis between 43 daggy Romneys and 40 feral Arapawas was also performed. The GWAS was also performed on a range of production and breech and fibre traits. There were no regions identified in the GWAS that reached the Bonferroni significance threshold ($P<1.03x10^{-6}$) for any of the traits. There were, however, a number of regions that reached the nominal significance threshold of $P<0.001$, specifically on chromosome 6 and 15 for dag score at 3 and 8 months respectively, and chromosome 22 and 6 for weaning weight. An OMIM and literature search identified genes with lipase activity, potential keratin signalling and protease activity within these regions. In the selection sweep analysis comparing daggy Romneys and Arapawas, there were 9 regions identified using a moving window of 5 SNPs and a fixation index threshold greater than 0.3. These regions included the known positive controls consisting of the poll and the agouti loci. There were no similarities between the GWAS and selection sweep analysis, nor with the case-control GWAS (Chapter 6).

7.2 Introduction

The aim of genome-wide association studies (GWAS) is to find quantitative trait loci (QTL) associated with a trait of interest in the population. This approach has been made possible by the development of whole genome arrays which genotype a large number of single nucleotide polymorphisms (SNPs). These SNPs are typically evenly spread across the genome in order to identify the majority of variation in the genome affecting the selected trait via linkage disequilibrium.

When many progeny, are measured it allows for calculation of sire estimated breeding values (EBVs) that are closer to the true merit of the sire in question than its own phenotype, as the accuracy of prediction is higher. Such a progeny test is also useful for hard to measure traits, traits that can only be measured once, are sex-limited, or can only be measured post slaughter. The sires with high accuracy EBVs can then be

used for association studies when genotyped. This approach is extremely powerful when genotyping costs are the major expense. However, unlike genotyping individuals, it assumes that the loci in question have additive effects on the trait.

A related methodology to detect QTL regions of interest is to use diverse strains or breeds of sheep to detect selection sweeps that are contrasting in the traits of interest (Moradi et al., 2012). Specifically, the ability to compare feral and 'improved' sheep potentially allows identification of genetic regions associated with not just domestication but also with survival without anthelmintics and animal husbandry intervention such as shearing, dagging and flystrike intervention. An example of a feral sheep breed is the Arapawa. These are animals that were once domesticated, but allowed to revert back to their unmanaged state. Arapawa sheep were located on Arapawa Island, Marlborough Sounds, New Zealand. They are now thought to have originated from mixed breed sheep from the southern United States brought to New Zealand by United States whalers (Young et al., 2011). The reversion of Arapawa sheep to a feral existence has made them an important reservoir of unique gene variants. Comparing Arapawa sheep with a domesticated breed allows the investigation of genes involved in natural disease resistance to challenges such as, flystrike, footrot and parasites, as well as the presence or absence of dags.

This chapter undertakes a GWAS on production, breech, wool and fibre traits using Sheep Improvement Limited (SIL) recorded sires genotyped with the Illumina OvineSNP50 BeadChip (50K SNP Chip) in order to identify genetic regions associated with these traits and finally, to see if there is any correlation with the results from the smaller case-control GWAS dataset (Chapter 6) for the traits measured in both studies.

Separately, a comparison of feral Arapawa sheep and Romneys with high dag breeding values was undertaken using the fixation index (F_{ST}) metric to identify regions associated with natural resistance and survival. The overall intention of the two studies was to identify common genomic regions associated with flystrike and dagginess resistance.

7.3 Materials and methods

7.3.1 Populations

The animals genotyped with the 50K SNP Chip were sourced from industry and research flocks, involved in Ovita-related research projects, including those involved in the industry genetic parameters study and progeny test described in Chapter 3 and 4 respectively. The animals were predominantly Romney, Coopworth, Perendale or Texel, plus other breeds and various breed crosses and composites. Only those animals that were genotyped and SIL recorded with appropriate phenotype data on themselves or progeny, were used for further analysis (n = 8,705). These were mainly sires and the majority of the resulting EBVs described below were calculated from progeny records. Sires only have own measures for weaning weight (WWT), live weight at 8 months (LW8) and fleece weight at 12 months (FW12). A proportion of the sires may have measures for dag score at 3 and 8 months (DAG3, DAG8) and breech bareness (BBREECH).

The comparison animals for the F_{ST} analysis were from a genetic resource of 60 Arapawa animals as described by Young et al. (2011). Forty of the Arapawa sheep that were the most unrelated that had been genotyped using the 50K SNP Panel were used in this study.

7.3.2 Estimated breeding values GWAS

Estimated breeding values were estimated by SIL, for the following traits: WWT, WWTm, LW8, FW12, DAG3, DAG8 and BBREECH. The analysis, performed on 3,535,557 animal records born between 1990 and 2010 for 233 SIL flocks, used SILs normal data cleaning and scaling procedures for these traits, and the traits were fitted in multi-trait BLUP analyses (S-A. N. Newman pers. comm.) using ASReml (Gilmour et al., 2009). The EBVs from SIL were converted to the original scale using summary information provided by SIL, as the EBVs calculated by SIL, are centred to the year 1995.

For the traits wool length (LENGTH), wool bulk (BULK), mean fibre diameter (MFD), standard deviation of MFD (FDSD), coefficient of variation of MFD (FDCV), proportion of medullated fibres (MED%), and curvature (CURV), the measurements were those detailed in Chapter 4. The traits were cleaned and scaled as detailed in

Chapter 4 and were fitted in uni-variate runs using ASReml (Gilmour et al., 2009) and full animal models.

7.3.3 Genotypes

The 50K genotypes of the 8,705 SIL recorded animals described above were from a data set of 13,559 genotyped industry and research animals. Genotyping was performed using the 50K SNP Chip in 2 laboratories. One set was genotyped by Illumina, the second genotyped at AgResearch, Invermay. The Illumina set included 2,805 AgResearch-Ovita research animals. The AgResearch data set includes all those genotyped since 2009. The 40 Arapawa sheep were genotyped as part of the Ovine HapMap study performed by Illumina (http://www.sheephapmap.org/hapmap.php). All genotypes for this analysis were analysed at AgResearch, Invermay using the manifest and allele calling parameters as defined in the ***Ovine_iSelect_2880.egt*** project. These were the original settings as defined by Illumina. Due to minor changes in chemicals supplied from Illumina since February 2011, new allele calling parameters as defined in the ***OvineFebruary2011.egt*** project were used for the animals genotyped from February 2011 onwards.

7.3.4 Analysis - GWAS

7.3.4.1 Quality control

Genotyping results were processed through a quality control pipeline prior to analysis. The pipeline has been described by Dodds et al. (2009). In summary, SNPs are discarded if they have a call rate <97%, appear non-autosomal (including pseudo-autosomal), MAF = 0, and weighted Gencall 10 score (GC score) <0.422. SNPs that were not retained as part of the Ovine HapMap study (Kijas et al., 2012a) were also discarded.

7.3.4.2 Population structure

Population stratification can give rise to false positive associations; to mitigate this, population structure needs to be accounted for (Lander and Schork, 1994). A genomic relationship matrix was estimated from the SNP markers to structure the variance of the direct genetic random effect. The first 6 principal components of this genomic matrix were used to account for population stratification. The genomic relationship matrix was calculated using the equation (VanRaden, 2008):

$$G = \frac{ZZ'}{2\sum p_i(1-p_i)}$$ Equation 1

Where Z is the SNP matrix $-2p_i$, $1-2p_i$, and $2-2p$ for BB, AB, and AA respectively, and p_i is the frequency of A allele of the i^{th} SNP in the population.

7.3.4.3 Calculating the dependent variables

Dependent variables (y) were calculated taking into account the individuals own and descendants' information. Parent average effects are removed from the EBVs using the method described by Garrick et al. (2009), assuming all genetic variation was explained by the markers (c = 0). The resulting values were de-regressed by the reliabilities of their EBVs with parent contributions removed and only those EBVs with reliabilities greater than and equal to $0.8h^2$ were used as phenotypes.

7.3.4.4 Genome-wide association analysis

The GWAS followed a 2-step procedure, performed using R 2.14.1 (R Development Core Team, 2011). Firstly, a weighted BLUP animal model was fitted to the phenotypes (in this case de-regressed EBVs weighted by equation 3 below), using a genomic relationship matrix calculated from the SNPs, as well as the first 6 principal components (PC) of the relationship matrix as fixed effects (B. Auvray Pers. Comm.).

$$y_i = \mathbf{x_i}\boldsymbol{\beta} + g_i + e_i$$ Equation 2

Where for animal i, y_i is the de-regressed EBV with the parent average removed, $\mathbf{x}_i$ is a vector of the first 6 PC of the G matrix relating to animal i, $\boldsymbol{\beta}$ is a vector of fixed effects of the PC, g_i is the genetic effect and e_i is the residual.

Secondly, the residuals resulting from the BLUP animal model were analysed using an ordinary weighted least square linear regression, fitting each SNP consecutively. The weighting factors (w) used to weight each observation (i) for both steps were (B Auvray pers. Comm.), where r^2 is the reliability:

$$w_i = \frac{r_i^2}{(1-r_i^2)}$$ Equation 3

The reason for the 2 step procedure was that it significantly reduces the time taken compared to doing both steps simultaneously.

7.3.4.5 QQ plot

The quantile - quantile (QQ) plots were calculated to check whether the distribution of the observed -$\log_{10}(P)$ values deviated from the expected distribution (exponential) under the null hypothesis of no genetic association and no LD between SNPs. To do so, the *n* -$\log_{10}(P)$ values were sorted and plotted against the -$\log_{10}$(1-u) where u = [1 / (n + 1), 2 / (n + 1), …, n / (n + 1)]. If the distributions are similar, then the slope, lambda, should be equal to 1. If the observed distribution deviates from the expected distribution with a slope greater than 1, it generally indicates that a proportion of the substructure has not been accounted for. A slope smaller than 1 is more difficult to explain, but could mean that the LD between SNPs and between SNPs and QTL is reducing the number of effective tests performed, thus producing larger *P* values than expected. When the right-hand tail diverges above the slope, the diverging SNPs could be considered associated with the trait.

7.3.4.6 Identifying relevant SNP markers

Marker associations were initially identified by graphing the resulting *P* values as -$\log_{10}(P)$ in a Manhattan plot across the genome. A conservative threshold can be calculated using the 5% Bonferroni correction (Rice, 1989) $0.05/n_{SNPs}$ equal to *P* value of 1.03×10^{-6} or -$\log_{10}(P)$ of ~6. However, for the traits analysed here, the most extreme SNPs had a -$\log_{10}(P)$ values between 1.5 and 5, or a *P* value of 0.03 to 3.4×10^{-6}. Thus, no results that exceeded the -$\log_{10}(P)$ threshold were identified. However, in order for these results to be compared and combined with other future studies we have reported the best SNPs. The nominal threshold level for each trait was first trimmed to a nominal level $P < 0.001$. The SNPs exceeding the nominal thresholds were aligned to the Ovine genome v2.0 map (The International Sheep Genomics Consortium et al., 2010) to identify and report the underlying regions/genes.

7.3.4.7 Comparison to case-control dataset GWAS (Chapter 6)

The -$\log_{10}(P)$ values for each trait were compared to their respective trait values from Chapter 6. The traits DAG3 and DAG8 were both compared to the case-control dataset's dag score. The correlation (r^2) for each comparison was calculated. Note, for LENGTH and BULK the phenotypes used in the case-control dataset are a subset of those used in this chapter.

7.3.5 Analysis – F_{ST}

7.3.5.1 Selection of daggy Romneys

The DAG3 and DAG8 breeding values were calculated by SIL as described in section 7.3.4.3. The average of the DAG3 and DAG8 breeding values were used to identify the top 60 most daggy Romneys. There were 43 purebred Romneys in the top 60 and these were selected for comparing with the 40 genotyped Arapawas. Twenty-seven of the selected Romneys had been genotyped at AgResearch, Invermay, and the remainder at Illumina.

7.3.5.2 Quality control

Genotypes were quality checked before analysis. The SNPs were discarded if the minor allele frequency across both groups (Romney and Arapawa) were less than 2%. Genotypes were also discarded if the call rate in either group was less than 95%. The SNPs were aligned to Ovine genome v2.0. SNPs that were on an unknown chromosome or chromosome Y were discarded. The first 2 principal components were plotted for quality assurance.

7.3.5.3 Wright's F_{ST}

Wright's fixation index F_{ST} (Wright, 1931; 1943; 1965) was calculated for each SNP using Microsoft Office Excel 2007 and the smoothed results plotted by chromosome. For each SNP, firstly, the observed heterozygosity in each population (H_{obs1}: Romney and H_{obs2}: Arapawa) was estimated as the number of heterozygotes in the population divide by the number of animals with genotypes for that SNP in that population. Secondly, the expected heterozygosity was calculated for each population (H_{exp1}: Romney and H_{exp2}: Arapawa) as 2pq, where p and q were the frequency of the first and second allele in the population. Thirdly, the expected heteroygosity for the subpopulation (H_S) was calculated as $(H_{obs1}*n_1 + H_{obs2}*n_2) / n_{total}$, where n_1 and n_2 are the number of Romney and Arapawa sheep with genotypes, respectively for that SNP, respectively and n_{total} is the number of Romney plus Arapawa animals with genotypes for that SNP. Fourthly, the expected heterozygosity for the overall total population (H_T) was calculated as $2\dot{p}\dot{q}$ where $\dot{p}$ and $\dot{q}$ are the frequencies of allele 1 and 2, respectively for Romney and Arapawa as one population. Finally, F_{ST} was calculated as $(H_T - H_S) / H_T$.

The fixation index is the difference in expected heterozygosity between the subpopulations relative to the total population, i.e. the extent of genetic differentiation between subpopulations. The results range from 0 (no differentiation) to 1 (complete differentiation, where subpopulations are fixed for the alternative allele). Smoothed estimates were calculated as the moving average with a window of 5 (WIN5) SNPs. This enables discovery of selection peaks across the 5 marker regions to be identified rather than one-off extreme values. A significance level of a WIN5 F_{ST} value greater than 0.30 was set and identified regions were aligned to the Ovine genome v2 to identify underlying genes. Detailed methods are described in Moradi et al. (2012). Empirical *P* values were calculated by ranking the WIN5 F_{ST} values and dividing the rank of the top individual SNP in the significant region by the number of SNPs with a WIN5 value (n = 46,907).

The resulting F_{ST} values were plotted against the average F_{ST} values from the ovine HapMap study (Kijas et al., 2012a), and the correlation calculated.

7.4 Results

7.4.1 GWAS

7.4.1.1 Quality control

After quality control checks, there were 47,084 SNPs retained for the GWAS. Figure 7.1 shows the minor allele frequency distribution of the final SNPs. The mean minor allele frequency was 0.29. There were 154 monomorphic SNPs and 6 with loss of heterozygote (K. Dodds Pers. Comm.).

7.4.1.2 Data summary

The average reliability after selecting those animals with a reliability greater or equal to $0.8h^2$ ranged from 0.48 to 0.58 for the production and wool traits (Table 7.1), and 0.24 to 0.42 for the fibre traits. Table 7.1 shows a summary of the number of animals, heritability, reliability cut off and resulting average and maximum reliabilities of the EBVs for each trait used in the analysis.

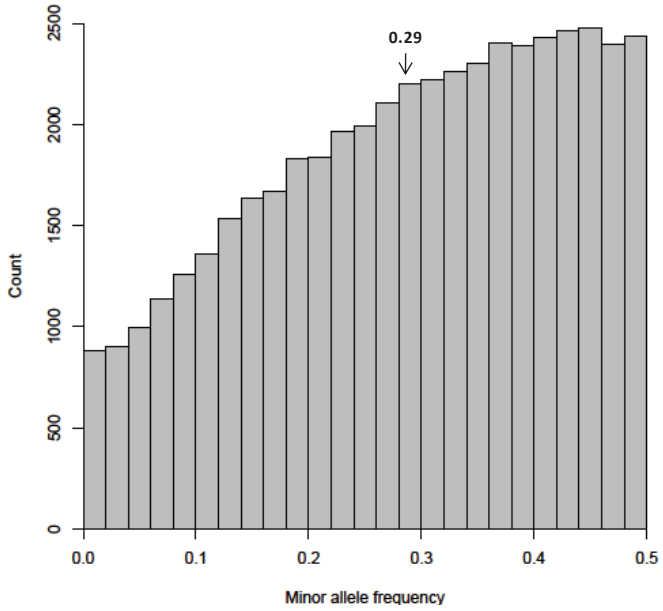


Figure 7.1: Minor allele frequency (MAF) for the Ovine 50K SNP chip, following removal of quality control checks, for the 8579 animals retained for GWAS. Arrow indicates mean MAF.

Table 7.1: Summary table of the number of animals (n), heritability (h^2), reliability cut off ($0.8h^2$), average and maximum reliability (Av Rel, Max Rel) of breeding values used in the analysis.

Trait[1]	n	h^2	$0.8h^2$	Av Rel	Max Rel
WWT	8599	0.177	0.142	0.546	0.998
WWTm	4284	0.115	0.092	0.611	0.980
LW8	8185	0.400	0.320	0.675	0.998
FW12	7088	0.350	0.280	0.582	0.998
DAG3	2800	0.334	0.267	0.474	0.980
DAG8	2114	0.309	0.247	0.514	0.980
BBREECH	1047	0.369	0.295	0.570	0.942
BULK	1245	0.436	0.348	0.545	0.938
LENGTH	1244	0.447	0.358	0.552	0.939
MFD	195	0.418	0.334	0.412	0.644
FDSD	83	0.253	0.203	0.279	0.514
FDCV	105	0.231	0.184	0.263	0.514
CURV	92	0.292	0.233	0.323	0.567
MED%	93	0.202	0.161	0.238	0.485

[1]WWT: weaning weight; WWTm: WWT maternal; LW8: autumn live weight; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; BULK: wool bulk, LENGTH: wool length; MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

Table 7.2: The breed genotypic variance explained by the first 6 principal components (PC) for each trait. Note differences between traits are due to number and selection of animals genotyped that are relevant to the specific analysis.

Trait[1]	PC1	PC2	PC3	PC4	PC5	PC6	Total
WWT	0.521	0.122	0.049	0.036	0.027	0.021	0.775
WWTm	0.467	0.178	0.042	0.021	0.020	0.015	0.744
LW8	0.527	0.124	0.051	0.036	0.024	0.019	0.780
FW12	0.530	0.086	0.053	0.046	0.027	0.021	0.764
DAG3	0.513	0.114	0.036	0.020	0.017	0.015	0.716
DAG8	0.471	0.133	0.032	0.023	0.022	0.021	0.702
BBREECH	0.441	0.060	0.027	0.024	0.018	0.014	0.584
BULK	0.572	0.050	0.021	0.017	0.013	0.012	0.685
LENGTH	0.571	0.051	0.021	0.017	0.013	0.012	0.685
MFD	0.346	0.064	0.037	0.034	0.030	0.022	0.532
FDSD	0.135	0.088	0.076	0.047	0.039	0.033	0.418
FDCV	0.144	0.098	0.070	0.047	0.033	0.026	0.418
CURV	0.147	0.099	0.077	0.046	0.038	0.029	0.436
MED%	0.146	0.105	0.077	0.040	0.035	0.029	0.432

[1]WWT: weaning weight; WWTm: WWT maternal; LW8: autumn live weight; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; BULK: wool bulk, LENGTH: wool length; MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

7.4.1.3 Principal components

Fitting the first 6 PCs accounted for the majority of the population structure. The first 4 PCs for WWT are shown in Figure 7.2. Table 7.2 shows the proportion of genetic variation contained in the genomic relationship matrix explained by the first PC and the additional variance explained by the second to sixth PCs for each trait. The majority of the variance (above 44%) is explained by the first PC for the production and wool traits LENGTH and BULK. For the fibre traits the first PC only explains 14-15 %, except MFD where it explains 35%.

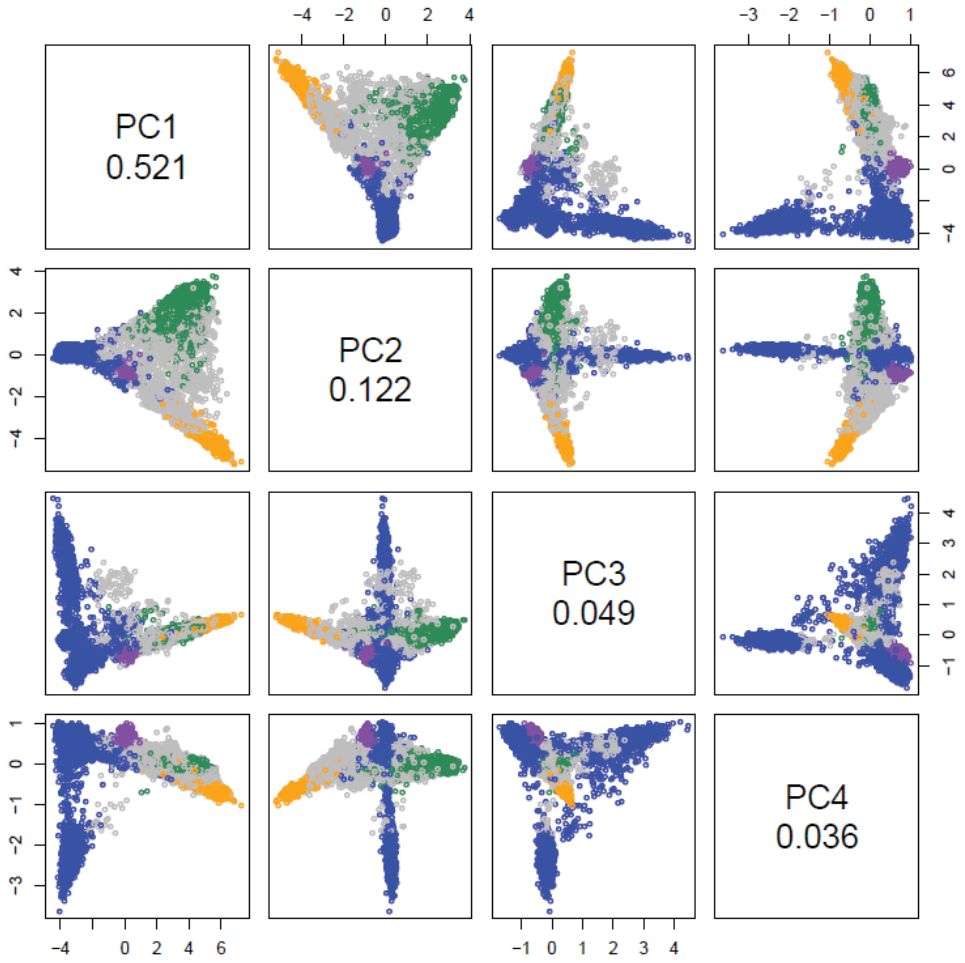


Figure 7.2: The first 4 principal components (PC) calculated from the G matrix for weaning weight. Romney (blue), Coopworth (green), Texel (yellow), Perendale (purple) and others (grey).

7.4.1.4 QQ plots

The observed distribution of $-\log_{10}(P)$ values were plotted against the expected distribution (exponential) under the null hypothesis of no genetic association and no LD between SNPs. The QQ plots for WWT, FW12, DAG3 and DAG8 are shown in Figure 7.3 and for the remaining traits in Appendix 6. The slopes (lambda) for all traits were less than 1 (Table 7.3). The right hand tail deviated above the slope for FW12 (Figure 7.3, B), DAG8 (Figure 7.3, D), and the fibre traits; MFD, FDSD, FDCV, CURV and MED% (Appendix 6). The best SNPs of these traits may be considered significantly above the noise.

Table 7.3: Estimate of lambda (slope), and their standard error (s.e.) of the linear regression of the observed -log$_{10}$(*P*) on the expected -log$_{10}$(*P*)(QQ plot) for each trait.

Trait	lambda	s.e.
Weaning weight	0.66	0.00007
Maternal weaning weight	0.73	0.00006
Live weight at 8 months	0.27	0.00009
Fleece weight at 12 months	0.45	0.00008
Dag score at 3 months	0.58	0.00009
Dag score at 8 months	0.67	0.00011
Breech bareness	0.73	0.00006
Wool bulk	0.70	0.00008
Wool length	0.75	0.00007
Mean fibre diameter (MFD)	0.96	0.00004
Standard deviation of MFD	0.96	0.00012
Coefficient of variation of MFD	0.98	0.00009
Curvature	0.98	0.00014
Proportion of medullated fibres	1.04	0.00006

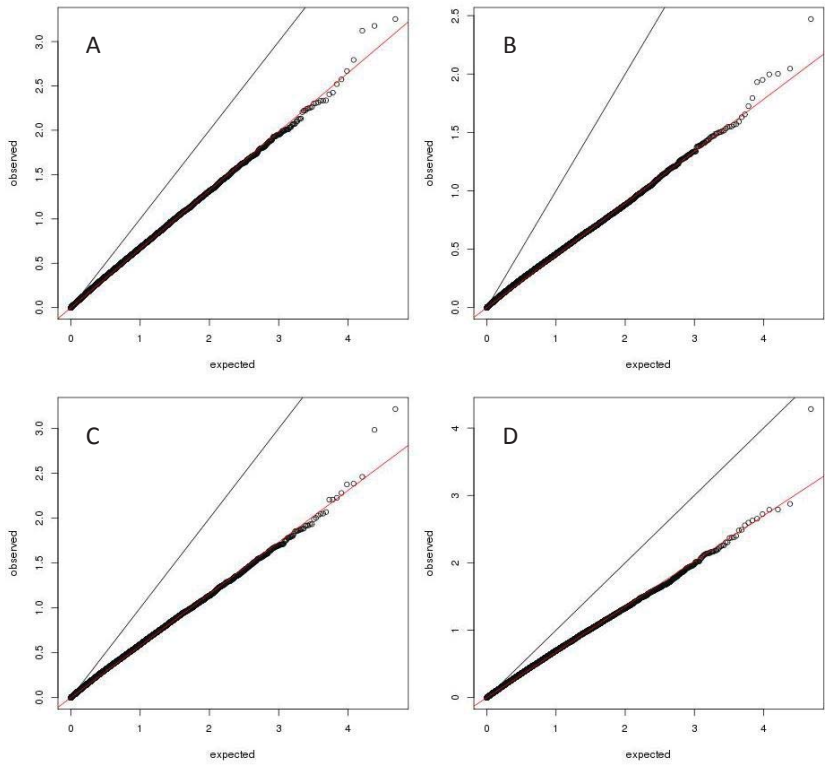


Figure 7.3: QQ plot for weaning weight (A), fleece weight at 12 months (B), dag score at 3 months (C) and 8 months (D) -log$_{10}$(*P*) values. The 0-1 line is in black and the linear regression in red.

7.4.1.5 SNP markers associated with production traits; WWT, WWTm, LW8 and FW12

There were 3 SNPs for WWT and 2 SNPs for WWTm that were identified as reaching the nominal significance threshold of $P < 0.001$ (Table 7.4; Figure 7.4; Appendix 7). There were no SNPS that reached this nominal significance level for LW8 and FW12. There was, however, a common SNP shared between WWT and LW8 that appeared in the top 5 SNPs for these 2 traits: s06773 on chromosome 3 (~21.88Mbp, OARv2). By visualization, there appears to be 1 SNP above the background noise for LW8 and FW12, on chromosomes 2 and 25 respectively (Appendix 8; Figure 7.5**)**.

Table 7.4: The trait, number of animals (n), marker name, chromosome (Chr) and chromosome position (base pairs, bp) of the best 3 SNPs for traits with SIL calculated BVs. The *P* value and -log$_{10}$(*P*) are shown for each marker.

Trait[1]	n	Marker name	Chr	Position (bp)	*P* value	-log10(*P*)
WWT	8599	s28777	22	36112603	5.57E-04	3.254
		s63492	6	70185558	6.66E-04	3.177
		s06773	3	218805503	7.54E-04	3.123
WWTm	4284	OAR1_166829690	1	154879916	4.83E-04	3.316
		OAR1_163628086_X	1	151976634	7.64E-04	3.117
		s54820	2	49622749	0.001	2.977
LW8	8185	s61633	2	205792466	0.025	1.595
		s73440	1	96034095	0.053	1.275
		OAR1_145935121	1	135146835	0.081	1.094
FW12	7088	s71205	25	40615041	0.003	2.471
		OAR3_107450633	3	100836677	0.009	2.046
		OAR19_60381806	19	57107311	0.010	2.002
DAG3	2800	OAR6_91313780	6	83850911	6.07E-04	3.217
		OAR6_57182975	6	51933217	0.001	2.984
		OARUn.2843_16326_X	25	28966531	0.003	2.461
DAG8	2114	s22390	15	40475174	5.23E-05	4.282
		s52289	9	15970168	0.001	2.875
		OAR8_75725850_X	8	70914235	0.002	2.791
BBREECH	1047	s52551	21	2506876	5.13E-04	3.290
		s62387	2	225861497	0.001	2.946
		s42345	7	86430058	0.001	2.927

[1]WWT: weaning weight; WWTm: WWT maternal; LW8: autumn live weight; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness.

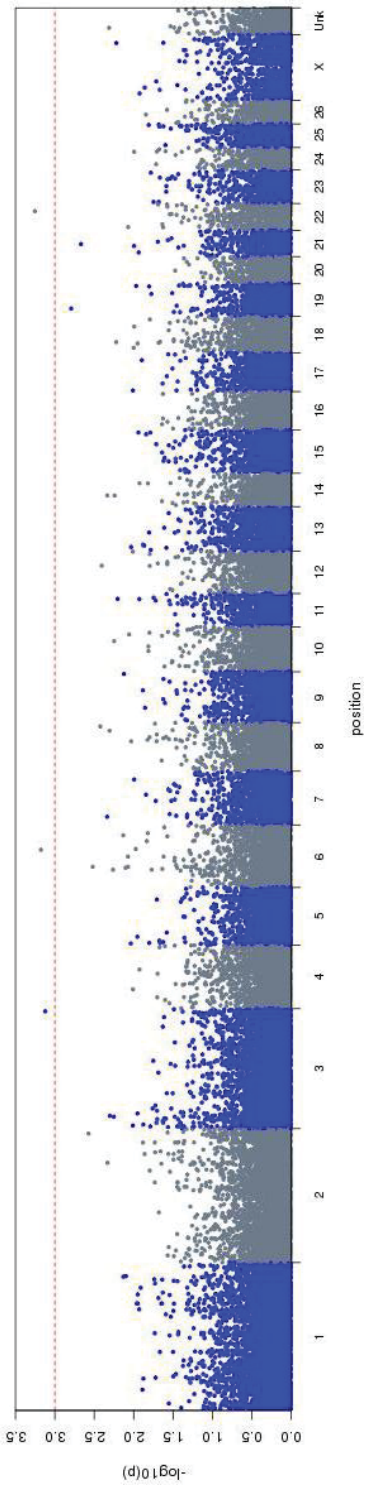


Figure 7.4: Manhattan plot of -log$_{10}$(*P*) values of SNPs for weaning weight. Ordered on the ovine genome v2 map, P<0.001 (red dash line).

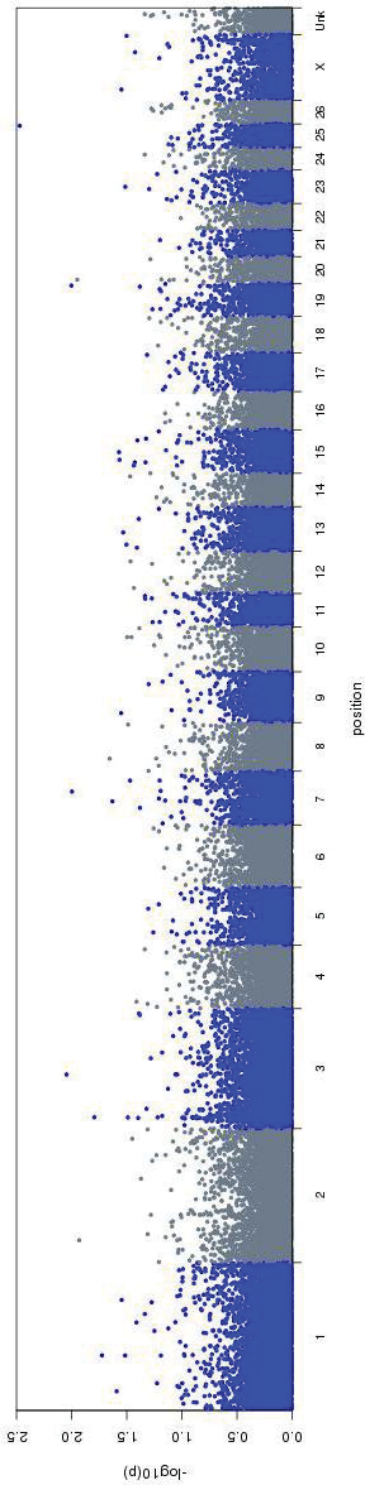


Figure 7.5: Manhattan plot of -log$_{10}$(*P*) values of SNPs for fleece weight at 12 months. Ordered on the ovine genome v2 map.

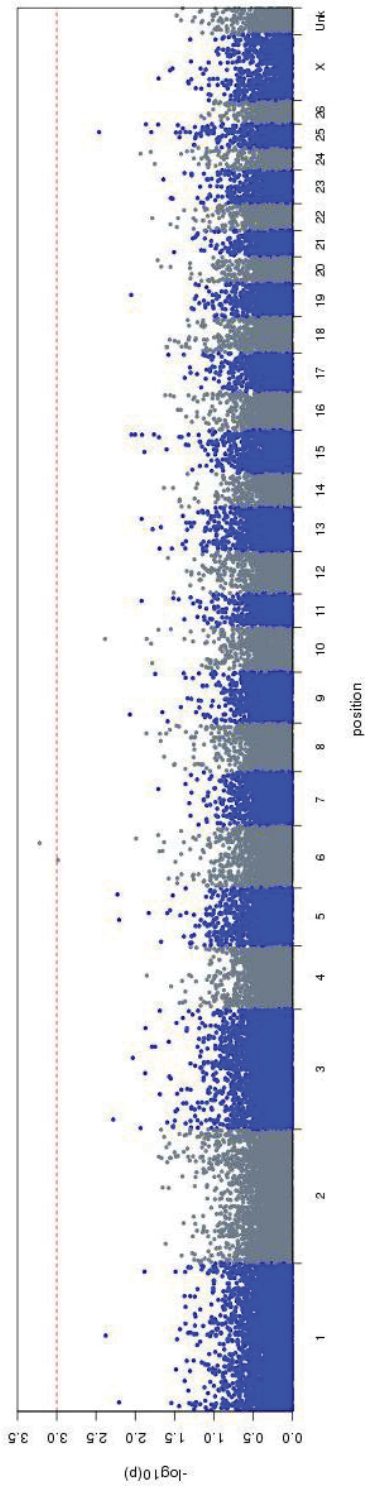


Figure 7.6: Manhattan plot of -log$_{10}$(*P*) values of SNPs for dag score at 3 months. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).

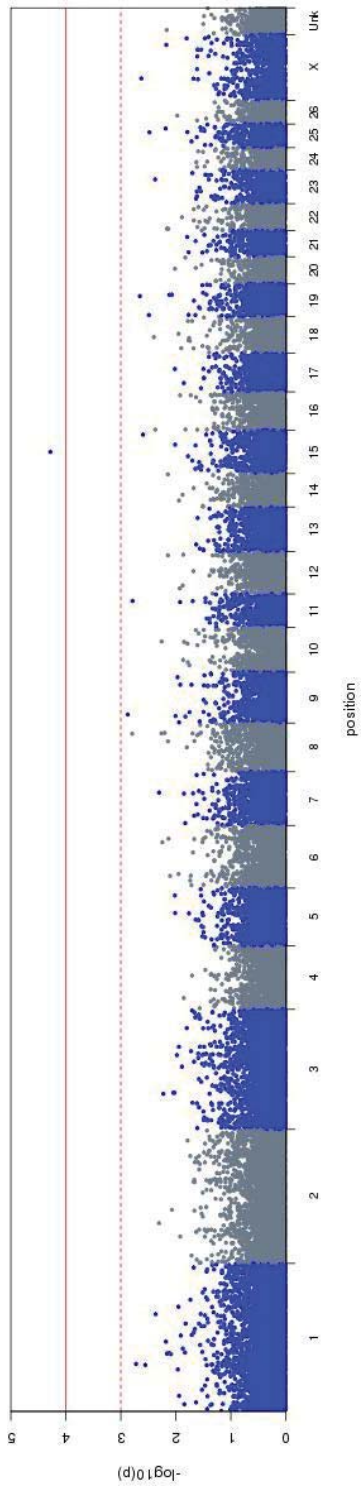


Figure 7.7: Manhattan plot of -log$_{10}$(*P*) values of SNPs for dag score at 8 months. Ordered on the ovine genome v2.0 map, P<0.0001 (solid red line), P<0.001 (red dash line).

7.4.1.6 SNP markers associated with breech traits; DAG3, DAG8 and BBREECH

There was 1 SNP above the significance threshold of P <0.001 for DAG3, DAG8 and BBREECH respectively (Table 7.4, Figures 7.6; 7.7; Appendix 9). Visually, there were 2 SNPs for DAG3 and 1 SNP for DAG8 and BBREECH above the background noise.

7.4.1.7 SNP markers associated with wool and fibre traits: BULK, LENGTH, MFD, FDSD, FDCV, MED%, and CURV

There was 1 SNP that reached the nominal significance threshold (P<0.001) for BULK, and none for LENGTH (Table 7.5; Appendix 10, 11). For the fibre traits; results are indicative as the number of parents genotyped was low (n animals with EBV reliabilities >= $0.8h^2$, < 220). After increasing the nominal significance threshold to P<0.0001, there were 2 SNPs which reached the threshold for MFD, 6 SNPs for FDSD and CURV, 4 for FDCV and 9 SNPs for MED% (Appendix 12-16). For the best 10 SNPs for FDSD and FDCV, 4 were the same for both traits (OAR3_188866593, OAR17_4155475, s36862 and s45116). Table 7.5 lists the best 3 SNPs for each trait.

7.4.1.8 Identification of candidate genes

Ovine genome v2.0 (The International Sheep Genomics Consortium et al., 2010), was used to identify genes within 100kbp of the significant SNP. The window of 100kbp was selected based on the known level of LD within the sheep genome (Raadsma, 2010), with it being unlikely that the actual loci would be more distant. There were 32 regions examined, and a further literature search and Online Mendelian Inheritance in Man (OMIM) were used to identify 39 candidate genes (Appendix 17). Table 7.6 shows the regions of interest for the production and breech traits. There were numerous genes identified in the top SNPs for each trait. Normally, only those genes that fell under SNPs at the significance level of $P < 10^{-6}$ would be further investigated; however, several genes that were associated with the traits WWT, FW12, DAG3 and DAG8 were followed up using a literature search to more deeply examine their attributes, see section 7.5.1.2.

Table 7.5: The trait, number of animals (n), marker name, chromosome (Chr) and chromosome position (base pairs, bp) of the best 3 SNPs for the wool and fibre traits. The *P* value and -log$_{10}$(*P*) are shown for each marker.

Trait[1]	n	Marker name	Chr	Position (bp)	*P* value	-log10(*P*)
BULK	1245	s61646	3	18005487	0.00044	3.359
		OAR7_58737037	7	53007221	0.00114	2.943
		OAR7_58748745	7	53017690	0.00114	2.943
LENGTH	1244	OARX_78982967	X	107668383	0.00107	2.970
		OAR7_97629466	7	89828177	0.00107	2.969
		OAR7_64371135	7	58364778	0.00113	2.948
MFD	195	s33504	2	107187233	1.08E-05	4.965
		OAR8_60310131	8	56543640	2.72E-05	4.566
		s13645	15	34621954	0.00010	3.992
FDSD	83	OAR11_37593909	11	35447352	7.25E-06	5.140
		OAR17_4155475	17	3663786	1.41E-05	4.852
		OAR2_213564294	2	202566737	2.25E-05	4.647
FDCV	105	s03307	3	167353426	2.43E-06	5.614
		s17816	15	28193609	4.52E-06	5.345
		OAR9_8924567	9	9121298	8.81E-06	5.055
CURV	92	OAR15_35337227	15	33814266	4.48E-06	5.349
		OAR5_51178121	5	47404299	1.85E-05	4.733
		s01175	20	40129416	4.69E-05	4.329
MED%	93	s38039	9	71039772	3.46E-06	5.461
		OAR5_88303469	5	80903483	3.73E-06	5.429
		OAR12_14036242	12	11234726	1.02E-05	4.991

[1]BULK: wool bulk; LENGTH: wool length; MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

7.4.1.8.1 Genes associated with production traits: WWT, WWTm, LW8 and FW12

There were 4 genes of interest (PNLIP, PNLIPRP2, PDGFRA and KIT) under the top 2 SNPs for WWT, and 2 genes of interest (GDF2 and GDF10) under the top SNP for FW12 (Table 7.6). There were no genes of interest for WWTm and LW8.

Table 7.6: Candidate genes within 100kbp of the best SNPs for the SIL traits, by chromosome (Chr).

Chr	200kbp region	Top SNP	P value	Candidate genes	OMIM	Function	Trait[1]
1	154779916- 54979915	OAR1_166829690	4.83E-04	No known genes			WWTm
2	205692466-205892465	s61633	0.025	ICOS	604558	T-cell response	LW8[2]
3	218705503-218905502	s06773	7.54E-04	LDOC1L		Placental structure	WWT
6	70085558-70285557	s6349	6.66E-04	PDGFRA	173490	Organ development & platelet growth	WWT[2]
				KIT	164920	Oncogene	
	83750911-83950910	OAR6_91313780	6.07E-04	TMPRSS11D	605369	Mucous membrane	DAG3
				TMPRSS11A	611704	Cellular senescence	
				TMPRSS11G		similar to A & D above	
15	40375174-40575173	s22390	5.23E-05	MICAL2	608881	Secretory pathway	DAG8
				GLEAN predict		unknown	
				DKK3	605416	Embryonic development - WNT signalling	
21	2406876-2606875	s52551	5.13E-04	no genes			BBREECH
22	36012603-36212602	s28777	5.57E-04	PNLIP	246600	Pancreatic lipase	WWT
				PNLIPRP2	604423	Pancreatic lipase	
25	40515041-40715040	s71205	0.003	GDF2	605120	BMP signalling	FW12
				GDF10	601361	BMP signalling	

[1] WWT: weaning weight; WWTm: maternal WWT; LW8: live weight at 8 months; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness.
[2]Genes within 500kbp of the top SNP

7.4.1.8.2 Genes associated with breech traits: DAG3, DAG8 and BBREECH

There were no genes under the top SNP for BBREECH on chromosome 21 (Table 7.6). There were several genes under the DAG3 and DAG8 SNPs. For DAG3 within the 200kbp region there were three genes of the same family of type II transmembrane serine proteases; Transmembrane protease/serine 11D, A and G.

For DAG8 there were 2 known genes and one gene predicted by GLEAN (Elsik et al., 2007) within the 200kbp region. Microtubule associated monoxygenase, calponin and LIM domain containing 2 (MICAL2) and diskkopf 3 homolog (DKK3) are the 2 known genes identified. Comparison of the GLEAN predicted gene with human and bovine genome did not identify any known genes.

7.4.1.8.3 Genes associated with wool and fibre traits: BULK, LENGTH, MFD, FDSD, FDCV, MED% and CURV

There were no known genes under the best SNP for BULK, and the best SNP for LENGTH did not reach nominal significance to require investigation of the region (Appendix 17). Due to the low number of animals (<220) with appropriate EBV reliabilities for the fibre traits there is insufficient power to detect true associated genetic regions. However, even with limited numbers of animals there are a number of interesting genes under the peaks identified including: Guanine nucleotide-binding protein β-4 (GNB4), Rho-assisted, coiled-coil containing protein kinase 2 (ROCK2), and Secreted frizzled-related protein 2 (SFRP2). These genes are not described here, but are worth noting for further investigation.

7.4.1.9 Similarities to case-control dataset GWAS

The correlation of the $-\log_{10}(P)$ values from this test with the results from Chapter 6 GWAS on flystrike case-control dataset ranged from 0.0001 (FDCV) to 0.018 (BBREECH) (Table 7.7). The plotted correlation for dag score (Chapter 6) versus dag score at 3 and 8 months (DAG3, DAG8) is shown in Figure 7.13, see Appendix 18 for remainder of comparisons.

Table 7.7: Correlation (r) of -log$_{10}$(*P*) values between case-control GWAS (C6, Chapter 6) and industry GWAS (C7, this chapter) traits[1].

Trait (C6)	Trait (C7)	R
Dag score	DAG3	0.0046
Dag score	DAG8	0.0006
BBREECH	BBREECH	0.0180
BULK	BULK	0.0123
LENGTH	LENGTH	0.0019
MFD	MFD	0.0013
FDSD	FDSD	0.0012
FDCV	FDCV	0.0001
CURV	CURV	0.0089
MED%	MED%	0.0167

[1]DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; BULK: wool bulk, LENGTH: wool length; MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

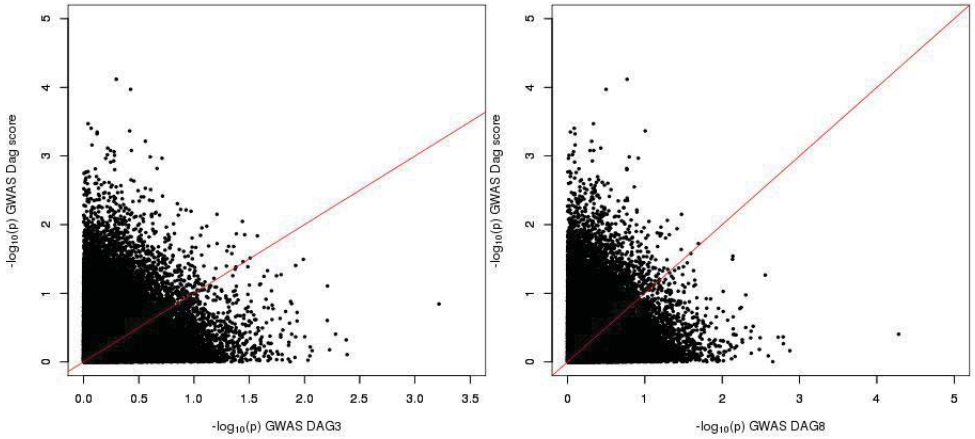


Figure 7.8: Plot of -log$_{10}$(*P*) values for individual SNPs from case-control GWAS (y axis) for dag score versus the industry GWAS (x axis) for dag score at 3 months (DAG3, left), and at 8 months (DAG8, right). The 0-1 line is plotted in red.

7.4.2. Selection sweep in Arapawa and Romney using Fst

There were 53,903 SNPs genotyped with 2,250 discarded for minor allele frequency less than 2% over both Romney and Arapawa groups. A further 3,695 SNPs were discarded for a call rate less than 95% in either Romney or Arapawa groups. Finally, 943 SNPs were discarded as they could not be mapped to a chromosome on Ovine genome v2.0. There were 47,015 SNPs used for F_{ST} analysis.

It is assumed that the sampled Arapawa sheep were representative of a feral breed with a unique reservoir of rare genetic variants absent in modern breeds or useful for adapting to extreme environmental conditions, including high parasite load (van Vuren and Hedrick, 1989), and potential flystrike. The average DAG3 and DAG8 breeding value for the 43 'Daggy Romney's' was 1.08 (range 0.82 to 2.09). The principal component analysis separating the Romneys and Arapawas, for the first 2 principal components (PC1 and PC2) is shown in Figure 7.9. The PC1 and PC2 explained 49% and 11% of the variance respectively. The PC1 component separated the 2 breeds, while adding in PC2 separated one flock of the Arapawa dataset from the rest of its breed. The small spot of Romneys compared to the spread of Arapawas indicate the Arapwas are more genetically diverse and that there is some within flock inbreeding which also causes the spread seen here for Arapawas.

Figure 7.10 shows the moving window of 5 (WIN5) F_{ST} values across the genome. There are 4 significant peaks (>0.3 WIN5F_{ST} value) on chromosome 13, one on each of chromosomes 2, 10, and 17, and 2 peaks on chromosome X. WIN5 F_{ST} values were between 0.3 and 0.6 (Table 7.8).

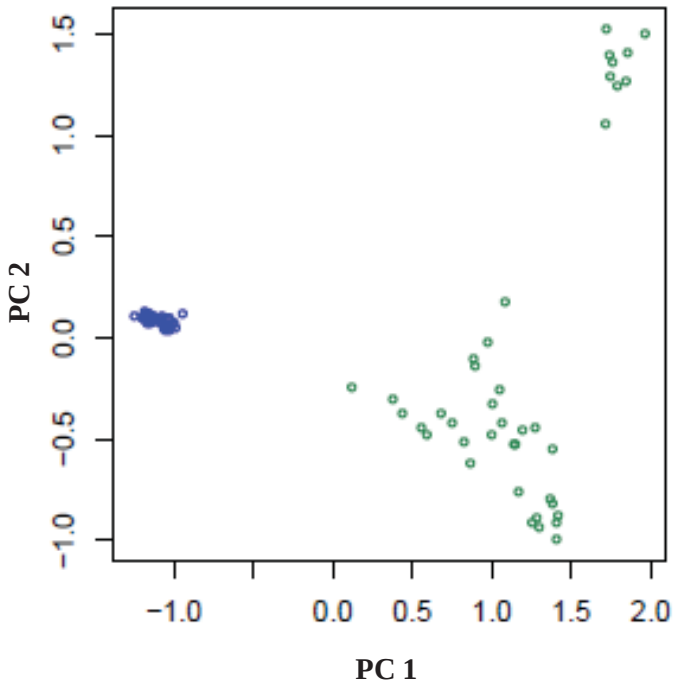


Figure 7.9: Animals clustered on the basis of principal components 1 and 2 (PC 1, PC 2). Romney and Arapawa breeds shown in blue and green respectively.

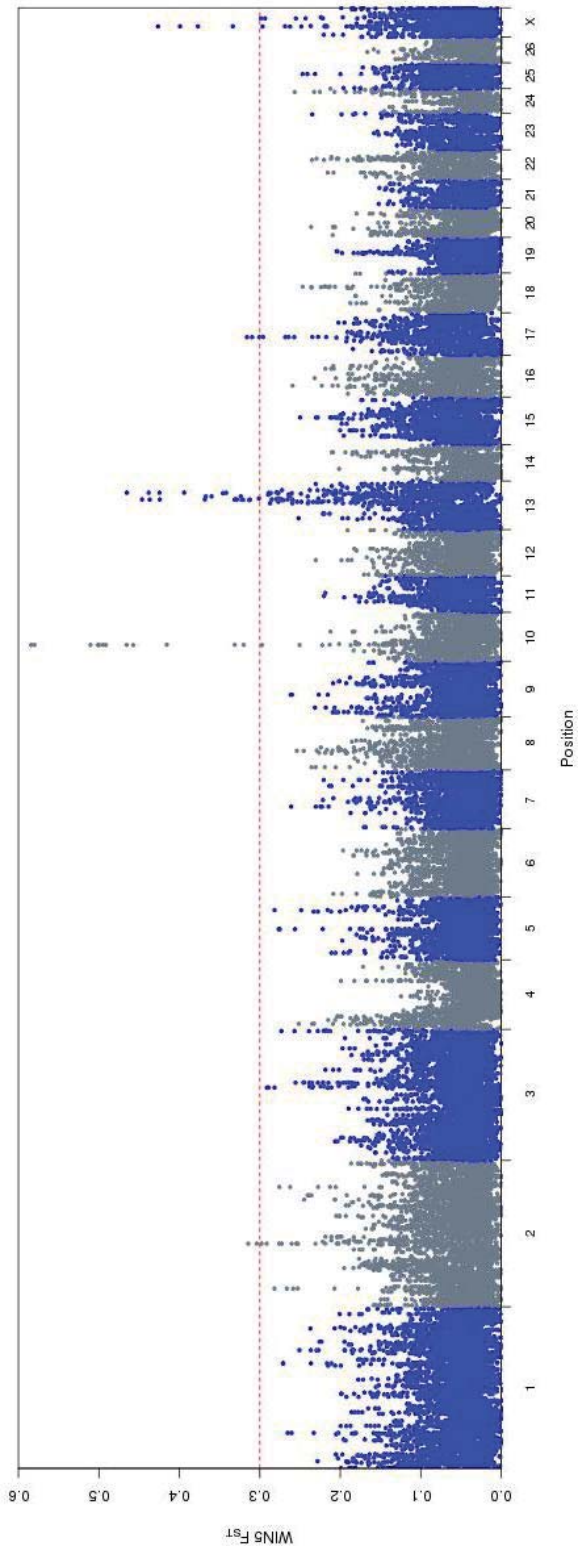


Figure 7.10: A Manhattan plot of the moving window of 5 (WIN5) F_{ST} values between the Arapawa and high dag score breeding value Romneys. Ordered on the ovine genome v2 map, WIN5 F_{ST} = 0.3 (red dash line).

Table 7.8: List of significant peaks and the candidate genes within the regions. The greyed genes are discussed in more detail.

Chr	Region (bp)	Empirical *P* value	Fst Peak - WIN5	Genes in Region	OMIM ID	Gene function
2	107578853 – 107826608	0.00077	0.31	GALNTL6		Glycosyltransferase
10	27496975-28149669	0.00004	0.58	RXFP2	606655	G-protein receptor
				B3GALTL	610308	Glycosyltransferase
13	50863580-51303184	0.00023	0.45	ATRN	603130	Glycoprotein
				C20orf194	614146	
				SLC4A1	109270	Glycoprotein
				ITPA	147520	Insoine triphosphatase
13	53110704 - 53212518	0.00087	0.30	C13H20orf201		
				RGS19	605071	Regulate G protein signalling
				TCEA2	604784	Transcription elongation factor
				SOX18	601618	Transcription factor
				PRPF6	613979	Pre-mRNA processing
				SAMD10		
13	56483792 - 56799235	0.00047	0.37	ZNF831		
				SLMO2		Locomotive behaviour
				ATP5E	606153	Mitochondrial ATP synthase
				TUBB1	612901	Cell structure
				CTSZ	603169	Cysteine protease
13	62188615 - 63220212	0.00019	0.47	PLUNC	607412	Immune response
				LPLUNC1		Immune response
				CDK5RAP1	608200	Cytoarchitecture central nervous system
				SNTA1	601017	Scaffold protein
				CHMP4B	610897	Multivesicular body sorting
				RALY		Coat colour
				EIF2S2	603908	Coat colour
				ASIP	600201	Coat colour
				AHCY	180960	Coat colour
				ITCH	606409	Coat colour
17	30711556 – 31121566	0.00075	0.32	no genes		
X	45318664 - 46859245	0.00030	0.43	ALAS2	301300	Heme biosynthesis
				APEX2	300773	Endonuclease
				PFKFB1	311790	Fructose metabolism
				FGD1	300546	Cytoskeleton and cell growth
				TSR2		pre-rRNA processing
				WNK3	300358	Regulate cell volumne

				GLRX5	609588	Heme biosynthesis
				PHF8	300560	Transcription regulation
				HUWE1	300697	Ubiquitin-protein ligase activity
				HSD17B10	300256	Mitochondrial enzyme
X	82094436 - 82243422	0.00092	0.30	no genes		

7.4.3.1 Chromosome 2

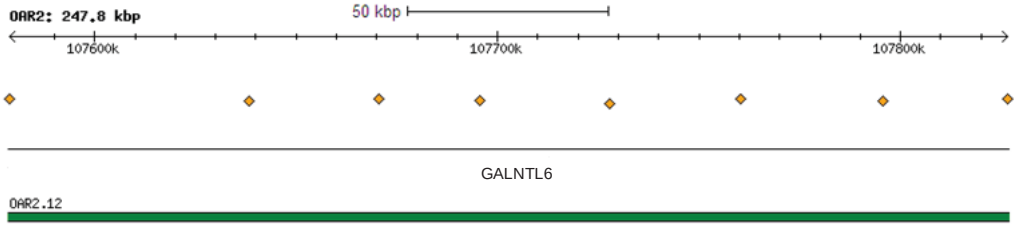


Figure 7.11: Schematic of significant region on chromosome 2, SNPs (diamonds), scaffold (green lines), and bovine ref sequences (black line).

There was one gene within the region (107.5Mbp – 107.8Mbp) on chromosome 2 (Figure 7.11), UDP-N-acetyl-alpha-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-like 6 (GALNTL6).

7.4.3.2 Chromosome 10

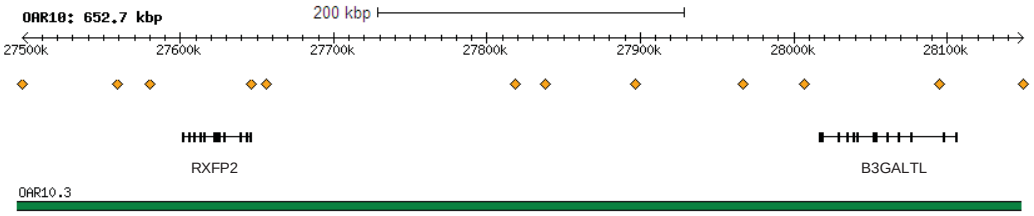


Figure 7.12: Schematic of significant region on chromosome 10, SNPs (diamonds), scaffold (green lines) and bovine ref sequences (black line).

There are 2 genes within the region (27.5Mbp to 28.2Mbp) on chromosome 10 (Figure 7.12): relaxin/insulin-like family peptide receptor 2 (RXFP2) and beta 1,3-galactosyltransferase-like (B3GALTL).

7.4.3.3 Chromosome 13

There were 2 main peaks on chromosome 13 at 50.86 to 51.30Mbp (Figure 7.13**)** and 62.19 to 63.22Mbp (Figure 7.16), with a WIN5 F_{ST} of 0.45 and 0.47 respectively. Another 2 peaks lie in between the 2 main peaks at 53.11 to 53.26Mbp (Figure 7.14) and 56.48 to 56.80Mbp (Figure 7.15), with WIN5 F_{ST} values of 0.30 and 0.37 respectively. There are a number of genes lying within these regions:

Region 1 (50.86 to 51.30 Mbp):

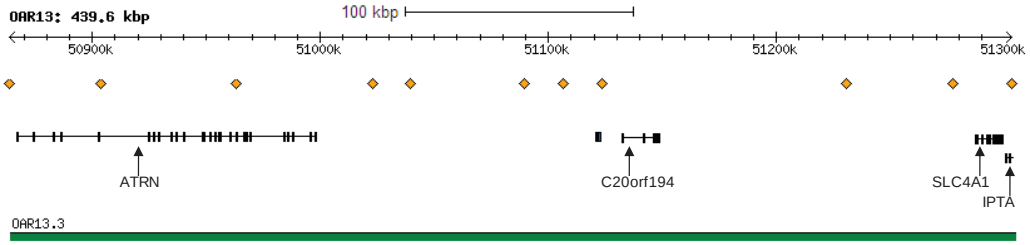


Figure 7.13: Schematic of significant region 1 on chromosome 13, SNPs (diamonds), scaffold (green lines) and bovine ref sequences (black line).

There are 4 genes within this region: attractin (ATRN), Chromosome 20 open reading frame 194 (C20orf194), solute carrier family 4 anion exchanger member 1 (SLC4A1) and inosine triphosphatase (ITPA).

Region 2 (53.11 to 53.21 Mbp):

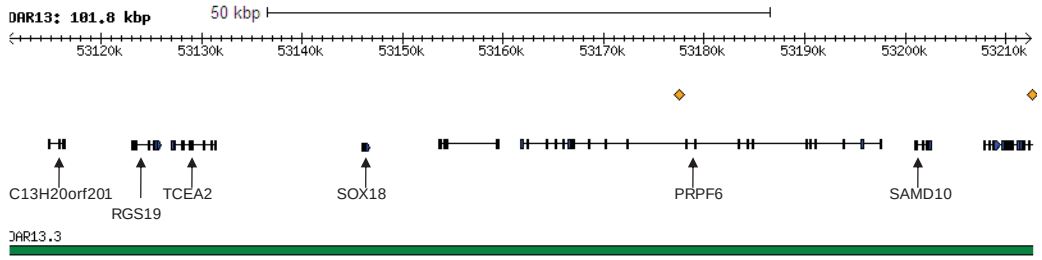


Figure 7.14: Schematic of significant region 2 on chromosome 13, SNPs (diamonds), scaffold (green lines) and bovine ref sequences (black line).

There are 6 genes in this region; Chromosome 13 open reading frame, human (C13H20orf201), regulator of G protein signalling 19 (RGS19), transcription elongation factor A, 2 (TCEA2), sex determining region (SRY) box 18 (SOX18), pre-mRNA

processing factor 6 (PRPF6) and sterile alpha motif domain- containing protein 10 (SAMD10).

Region 3 (56.48 to 56.80 Mbp):

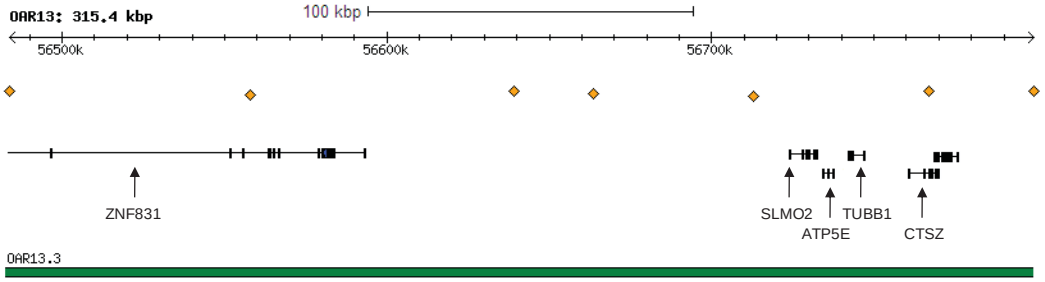


Figure 7.15: Schematic of significant region 3 on chromosome 13, SNPs (diamonds), scaffold (green lines) and bovine ref sequences (black line).

There were 5 genes identified in this region: Zinc finger protein 831 (ZNF831); *Drosophila* slomo homolog 2 (SLMO2); ATP synthase, H+ transporting, mitochondrial F1 complex; epsilon subunit (ATP5E); tubulin, beta 1 (TUBB1) and cathepsin Z (CTSZ).

Region 4 (62.19 to 63.22 Mbp):

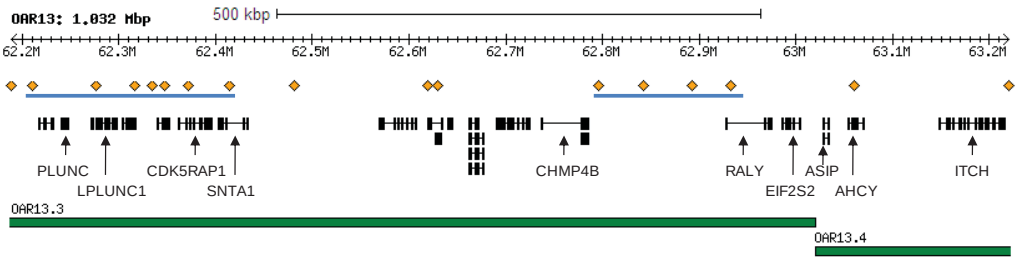


Figure 7.16: Schematic of significant region 4 on chromosome 13, SNPs (diamonds), scaffold (green lines) and bovine ref sequences (black line).

There are 2 sets of genes that have significance in this region (Figure 7.16). The first (WIN5 F_{ST} 0.35) are 2 genes of the bactericidal/permeability-increasing (BPI) protein superfamily: Palate, lung, and nasal epithelium carcinoma-associated protein (PLUNC) otherwise known as BPI fold-containing protein, family A member 1 (BPIFA1) and BPI fold-containing protein, family B member 1 (BPIFB1 or LPLUNC1). Downstream,

other members of the bovine BPI superfamily have been mapped including; BSP30C, BSP30D, SPLUNC2B, SPLUNC2A, BPIL3 (61.73MB to 62.16Mbp) (Wheeler et al., 2007). The second involves 5 genes associated with coat colour; RNA binding protein, autoantigenic (RALY) (hnRNP-associated with lethal yellow homolog (mouse)), eukaryotic translation initiation factor2, subunit 2 beta (EIF2S2), agouti signalling protein (ASIP) and itchy E3 ubiquitin protein ligase homolog (mouse) (ITCH), and Adenosylhomocysteinase (AHCY). These genes lie upstream of the second peak within this region (WIN5 F_{ST} 0.47). Mutations in all 5 genes have an effect on agouti expression.

There are three other genes that lie under or near the 2 peaks; Cyclin-dependent protein kinase 5 regulatory subunit associated protein 1 (CDK5RAP1), Syntrophin, alpha 1 (SNTA1), and Chromatin modifying protein 4B (CHMP4B), however these are not of interest.

7.4.3.4 Chromosome 17

No genes were identified within the region on chromosome 17 (30.71Mbp to 31.12Mbp)

7.4.3.5 Chromosome X

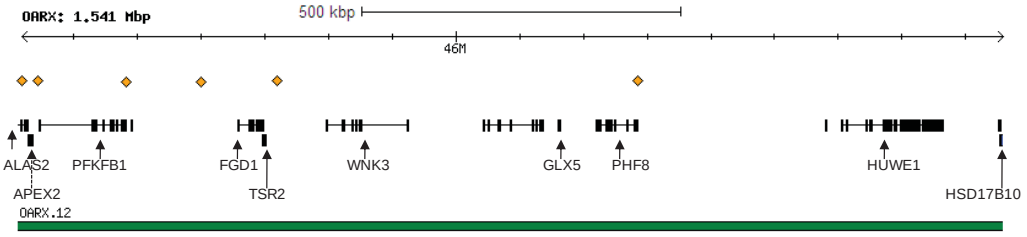


Figure 7.17: Schematic of significant region (45.32 to 46.86 Mbp) on chromosome X, SNPs (diamonds), scaffolds (green lines) and bovine ref sequences (black line).

Two peaks were evident on chromosome X, at 45.32 to 46.86 Mbp (Figure 7.17) and 82.08 to 82.24 Mbp, with WIN5 F_{ST} of 0.43 and 0.30, respectively. The second peak has no genes underlying the region, however the first peak has 10 genes. These are: delta-aminolevulinate synthase 2 (ALAS2), apex nuclease (apurinic/apyrimidinic endonuclease) 2 (APEX2), 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase1 (PFKFB1), FYVE, RhoGEF, and PH domain-containing protein1 (FGD1), 20S rRNA accumulation, homolog (TSR2), Protein kinase, lysine-deficient 3 (WNK3),

Glutaredoxin 5 (GLRX5), PHD finger protein 8 (PHF8), HECT, UBA, and WWE domains-containing protein 1 (HUWE1) and 17-Beta-hydroxysteroid dehydrogenase X (HSD17B10).

7.4.3.6 Comparison with HapMap

The resulting F_{ST} values were also plotted against the F_{ST} values from the ovine HapMap study (Kijas et al., 2012a) (Figure 7.18). The correlation was low at only 0.08, however, the best SNPs for both studies identified the horns locus on chromosome 10 and the agouti locus on chromosome 13.

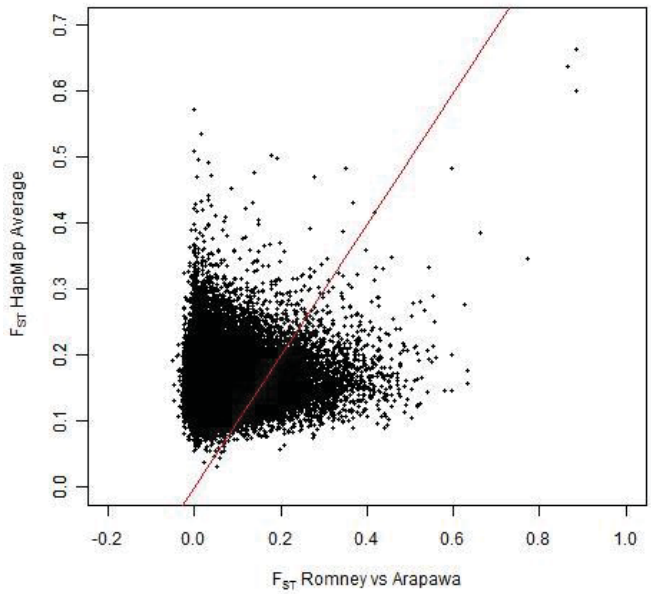


Figure 7.18: Comparison of the F_{ST} values from the Romney versus Arapawa and the average F_{ST} value from the ovine HapMap study (Kijas et al., 2012a). The 0-1 line is plotted in red.

7.5 Discussion

The results of GWAS for 13 production, breech, wool and fibre traits in New Zealand dual-purpose industry sheep are reported in this chapter. Also reported are F_{ST} results from the comparison of a feral breed (Arapawa) with high dag BV Romney sheep. A number of regions were identified from each study. Disappointingly, none of the GWAS results exceeded the conservative Bonferroni threshold. They are tabled and discussed here so that future researchers can combine these results in meta-analyses. In contrast, the selection sweep study identified numerous regions. There were no

similarities between regions identified in the 2 analyses, and there were no similarities between the GWAS done here and the case-control GWAS in the previous chapter.

7.5.1 GWAS

For the GWAS there were 2 SNPs identified in the top 5-7 SNPs for more than 1 trait, namely between FDSD and FDCV, and 1 SNP between WWT and LW8. There were significant SNPs that reached the threshold $P<1.03x10^{-6}$ for the fibre traits (MFD, FDSD, FDCV and MED%), however, visually there was at least 1 SNP significantly above the background noise for most traits.

The GWAS analysis approach used corrects for population and genetic relatedness. However this removes SNPs that are associated with breed differences and it may weaken the power of the analysis by excluding SNPs that have a true effect. This is more apparent for traits where selection has been focused in only one breed e.g. increased muscling in Texel sheep. An alternative approach is to fit pedigree, but, not correct for population stratification, however this can increase the number of false positives as well as identify true associations. Validation of this analysis using a Bayesian approach fitting the SNPs simultaneously, instead of consecutively, could also be implemented.

7.5.1.1 QQ plots

The QQ plots test if the distribution of the observed values deviates from the expected distribution under the null hypothesis of no genetic association and no LD between SNPs. If the distributions are similar then the slope, lambda, should be equal to 1. If the observed deviates from the expected with a slope greater than 1, this indicates that a proportion of the population substructure has not been accounted for. When the slope is less than 1, this is harder to explain but one reason could be that due to LD between SNPs and between QTL and SNPs, which reduces the effective number of tests to less than the number of SNPs analysed, thus the $-\log_{10}(P)$ values are smaller. This may have occurred in this analysis and when there has been intense selection for the trait, more SNPs are associated with the QTLs, increasing the LD between the QTLs and the SNPs and decreasing the effective number of tests.

Figure 7.19 plots the QQ slope versus the selection intensity (estimated from the average EBVs of the animals genotyped, using selection response equation 11.4 in Falconer (1989)) for each trait. This is the reason for why LW8 has a lower slope than

the other traits, especially the fibre traits where no selection has taken place. As the right-hand tail does not deviate significantly above the slope in the QQ plots for each trait, this indicates that there are no significant associations for these traits. Different lambda values were obtained in the genomic selection analysis (Chapter 8) for the traits WWT, WWTm, LW8, FW12, DAG3 and DAG8, and these will be discussed in Chapter 8.

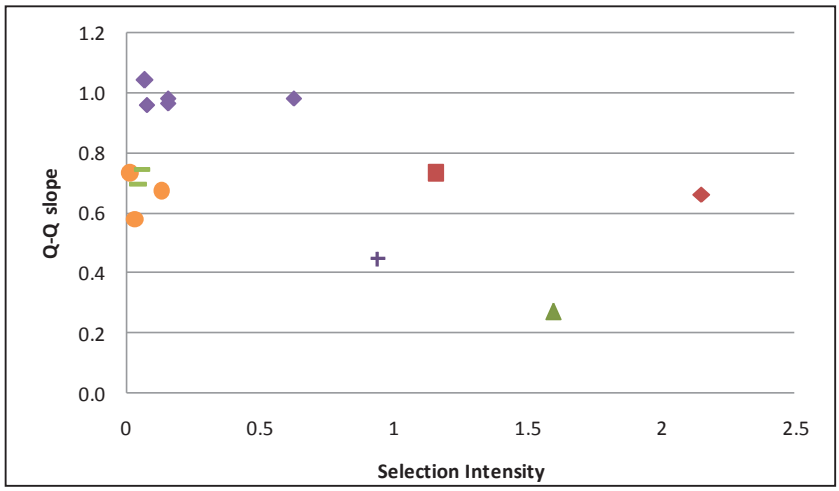


Figure 7.19: Plot of the QQ slope against the selection intensity for each trait; weaning weight direct and maternal (♦ ■), live weight at 8 months (▲), fleece weight at 12 months (+), dag at 3 and 8 months and breech bareness (●), wool bulk and length (—) and fibre traits (♦).

7.5.1.2 Description of identified genes

Pancreatic lipase (PNLIP) and pancreatic lipase-related protein 2 (PNLIPRP2) have a key function in dietary fat absorption. They hydrolyze long chain triglycerides to free fatty acids and monoacylglycerols (Lowe et al., 1989). These products then enter circulation through the lymphatic system. A deficiency of pancreatic lipase leads to mal-absorption of fats and steatorrhea, excessive discharge of fat in the faeces (Gaskin et al., 1982). These enzymes could be important for the digestion of milk fats by lambs.

Platelet-derived growth factor receptor α (PDGFRA) and mast/stem cell growth factor receptor (c-KIT) both belong to the type III family of receptor tyrosine kinases (Wollberg et al., 2003). Platelet-derived growth factor (PDGF) is a mitogen,

chemotactic factor and survival factor for mesenchymally derived cells (Rosenkranz and Kazlauskas, 1999). PDGF induces cell proliferation, migration and cellular survival via 2 subtype receptors, PDGFRA and PDGFRB. PDGFRA expression is low in adult connective tissue compared to PDGFRB, while in embryonic mesenchymal cells expression is high, indicating a critical role in development (Osornio-Vargas et al., 1996).

The receptor c-KIT is important in the stem cell factor (scF)/c-Kit receptor pathway. The c-Kit gene is identical to white spotting locus in mouse (Geissler et al., 1988). Mutations in c-Kit leads to melanogenesis, gametogenesis and haematopoiesis with some of the phenotypes including; anaemia, white coat colour, sterility, mast cell deficiency and reduced gastrointestinal motility (Geissler et al., 1988; Voytyuk et al., 2003; Wollberg et al., 2003). Reduced gastrointestinal motility could lead to inefficient digestion and absorption of food, thus affecting growth.

Growth differentiation factor 2 and 10 (GDF2, GDF10), also known as bone morphogenetic protein 9 and 3B (BMP9, BMP3B) respectively, belong to the TGF-β superfamily. The TGF- β superfamily binds 2 types of transmembrane serine/threonine kinase receptors, and are involved in many aspects of development. BMP signalling is important in the regulation of proliferation and differentiation in the hair follicle (Kulessa et al., 2000). Four BMP proteins have been shown to be expressed in the inner root sheath, outer root sheath or dermal papilla. These are BMP2, BMP4, BMP7 and BMP8a and b (reviewed in Kulessa et al., 2000). Unfortunately, at this stage, no evidence has been found to link GDF2/BMP9 and GDF10/BMP3B to hair follicle development.

GDF2/BMP9 is found to be highly expressed in the liver (Miller et al., 2000), but also plays a big role in the mesenchymal stem cells (MSCs) differentiation process into osteogenic, chondrogenic, adipogenic or myogenic lineages (reviewed in Luther et al., 2011). GDF2/BMP9 also communicates with the WNT/β-catenin signalling pathway, which is known to be involved in hair follicle differentiation (Rutberg et al., 2006), to mediate each other's bone development pathways (reviewed in Luther et al., 2011). GDF10/BMP3B is in the third family of BMPs with BMP3, both proteins have slightly different expression patterns though both are expressed in bone. GDF10/BMP3B is highly expressed in post-neonate and adult cerebellum, neonatal skullcap and femur, adult costa (rib), costicartillage (ribs), femur, trachea and aorta (Takao et al., 1996).

This indicates that GDF10/BMP3B has a role in bone and cartilage development/metabolism and in the central nervous system.

Transmembrane protease/serine 11D (TMPRSS11D) and A belong to the human airway trypsin-like protease/differentially expressed in squamous cell carcinoma (HAT/DESC) subfamily, while G currently has sequence similarity to A and D, but unknown function. TMPRSS11D was found to be almost exclusively expressed in the lower airway including the trachea, localized in submucosal serous glands of trachea and bronchi (Yamaoka et al., 1998; Yasuoka et al., 1997). TMPRSS11A was shown to be highly expressed in the tongue, bladder and eye, and also expressed in trachea, forestomach and glandular stomach in mouse (Sales et al., 2011). Sales et al. (2011) also confirmed expression of TMPRSS11D in mouse trachea, tongue, skin, eye, testis, forestomach and glandular stomach. This suggests there could potentially be a link between immune response in mucosal secretions and faecal consistency leading to faecal accumulation on the breech.

7.5.1.3 Summary of identified genes

Of the genes identified, only KIT had been previously identified as a candidate gene for selection sweep in the sheep HapMap study (Kijas et al., 2012a). Given that these results are based on small number of animals measured for fleece traits, it implies that additional work on these traits may well identify major loci affecting these traits.

Interestingly for WWT, WWTm, and LW8, none of the known genes affecting growth or muscling such as myostatin, callipyge, carwell, GH1 or IGF1 were identified in this study. Instead, 4 other genes, PDGFRA, KIT, PNLIP and PNLIPRP2 were identified. Of these, KIT has been identified in the sheep HapMap study as being under selection (Kijas et al., 2012a), however, it is better known for affecting pigmentation. The implication of KIT in gastrointestinal motility is interesting, and maybe the reason for its association with WWT.

For FW12, again surprisingly, none of the keratin or keratin associated proteins (Purvis and Franklin, 2005) were identified, however, genes belonging to the BMP family on chromosome 25 were found to be associated. BMP signalling genes have been described as regulating hair/wool follicle development (Kulessa et al., 2000).

For the dagginess traits, only 1 family of genes were found to be associated with DAG3, TMPRSS11A, D, and G, which belong to type II transmembrane serine

proteases. TMPRSS11A and D are known to be expressed in the airways and in the stomach. Based on their expression location and association with mucosal secretions, it is not improbable that variants may affect faecal consistency, however, additional research would be required to demonstrate a causal link. The GLEAN predicted gene under DAG8 best SNP, did not map to any known human and bovine genes. Sequencing of this region, and investigation for gene and regulatory element signatures would be beneficial.

7.5.2 Comparison to industry GWAS

The $-\log_{10}(P)$ values from this study were very poorly correlated with the $-\log_{10}(P)$ from the GWAS on case-control dataset (Chapter 6), meaning that the results from the case-control dataset were not validated by this industry and progeny test GWAS.

Several explanations can be proposed. Firstly, the case-control dataset analysed in Chapter 6 is quite independent from this analysis using industry and progeny test measurements. There were only 216 lambs that overlapped between these 2 studies. They were from SIL recorded flocks participating in both studies, and only the wool bulk and wool length measurements were the same. In the case-control GWAS, actual phenotypes were used, with a 2 stage BLUP fitting a polygenic model and genetic relationship matrix fitted. For this analysis, the de-regressed breeding values were used (with reliability greater or equal to $0.8h^2$), and a 2 stage weighted BLUP analysis fitting a genomic relationship matrix and PCs were used. There are no known published results on comparing the use of EBVs and phenotypes in a GWAS. This is an area that requires further work; however, it is beyond the scope of this study.

One additional reason was the use of imputation results in the case-control GWAS. If the imputation was not accurate or an incorrect reference data set that did not contain all the possible haplotypes present in the population was used, potential SNPs in linkage with causative QTL could be incorrectly imputed. However, in Chapter 5, the imputation methods were extensively studied to achieve the highest accuracy. Browning and Browning (2009) suggest using genotype probabilities from the imputation results for association analysis as they can be more powerful than the ‘best-guess’ genotypes that also arise from imputation. In the case-control GWAS, the ‘best-guess’ genotypes were used, as it was found in Chapter 5 that these genotypes generated a G matrix diagonal closer to 1 as is expected.

7.5.3 Selection sweep study

For the feral Arapawa versus domesticated and daggy Romneys, 7 regions were identified with candidate genes involved in domestication and survival. Due to the nature of the study, this will include traits other than dagginess that are under natural or artificial selection. This has benefits and disadvantages. The benefits are that it includes well studied internal controls; the disadvantage is that without matching confirmation from an independent GWAS, a candidate region cannot be definitively assigned as causal for the trait of interest. The study design was originally intended as a low cost confirmation of the GWAS results.

The well studied horns region was again confirmed with a peak on chromosome 10 that included RXFP2 and also implicated B3GALTL gene. This region is well documented for its association with polledness in domesticated sheep (Dominik et al., 2012; Johnston et al., 2011; Kijas et al., 2012a; Pickering et al., 2009b). Defects in RXFP2 cause cryptorchidism (undescended testes), and a reduction in fertility in humans and mice (Gorlov et al., 2002; Overbeek et al., 2001). Castration, the removal of testes from a male, has been shown to be associated with cessation of horn growth in sheep (Marshall and Hammond, 1914). B3GALTL is involved in the transfer of the disaccharide glucosyl-beta-1,3-fucose-O- onto the thrombospondin type-1 repeat glucose (Hess et al., 2008). A splice site mutation in B3GALTL causes Peters-plus syndrome, an autosomal recessive disorder characterised by anterior eye chamber defects, as well as disproportionate short stature, cleft lip and palate, developmental delay and includes in some cases cryptorchidism (Maillette de Buy Wenniger-Prick and Hennekam, 2002). The cryptorchidism phenotype, also seen for RXFP2, indicates this gene could potentially also affect horn presence and size.

It was interesting that in addition to polledness, only the agouti region has previously been reported in either ovine or bovine selection sweeps (Kijas et al., 2012a; The Bovine HapMap Consortium, 2009). In both this study and the Sheep HapMap study, the agouti region was identified. This region has 4 genes that control coat colour; ASIP, RALY, ITCH and AHCY. A different region on the same chromosome contained ATRN which is a negative modulator of ASIP. Agouti signalling protein (ASIP) was first noted in the study of mouse coat colour genes. This gene encodes a paracrine signalling molecule that causes hair follicle melanocytes to synthesize pheomelanin, a yellow pigment, instead of black or brown during the mid-portion of the hair growth

cycle (Morgan et al., 1999). Variation in coat colour and distribution is important for the fitness of animals in the wild. Linnen et al. (2009) showed that the wideband phenotype (light colour) seen in the deer mouse, arose *de novo* after colonization of the Nebraska Sand Hills (estimated at 8000 to 10000 years old), as a selective advantage against predators.

A 120 – 170 kbp deletion in the RALY gene causes the lethal yellow (A^y) mutation (Duhl et al., 1994; Michaud et al., 1994). Heterozygotes of the A^y allele express a number of phenotypes including obesity, increased tumour susceptibility, type II diabetes, and ectopic expression of the agouti gene. Homozygotes of the A^y allele result in embryonic death. The non-agouti-lethal 18H (a^{18H}) mutation is a small inversion that disrupts both agouti and ITCH genes, in which the proximal and distal inversion breakpoints occur respectively (Perry et al., 1998). The mice have the dark agouti phenotype, and develop a variety of immunologic diseases, including inflammation of the lung and stomach, hyperplasia of lymphoid and hematopoietic cells and constant itching in the skin. The lethal non-agouti (a^x) mutation is a ~100kbp deletion that begins 4kb 3' of the agouti last exon, and contains the AHCY gene (Miller et al., 1994). When homozygous, this mutation leads to embryonic death around time of implantation. The AHCY gene encodes an S-adenosylhomocysteine hydrolase (AdoHcy), which inhibits methyltransferase reactions, keeping the ratio of AdoMet/AdoHcy normal. Disruption of AdoHcy can lead to elevated plasma methionine (Barić et al., 2004).

There are 2 products of the ATRN gene as a result of alternative splicing, a soluble protein which helps regulate cell interactions during T-cell activation, and a membrane form which is involved in pigmentation and energy metabolism (Tang et al., 2000). The membrane form is a receptor for the agouti protein (He et al., 2001). Absence of ATRN causes the 'mahogany' phenotype where mice produce mostly eumelanin/black coat colour. The ATRN gene acts as a negative mediator of agouti pigmentation phenotypes. ATRN mutations also give rise to spongy degeneration in the brain, indicating a role in the brain independent of agouti-signalling (He et al., 2001).

SOX18, also on chromosome 13, is implicated in hair development namely growth, and absence and presence of hair. SOX18 is a transcription factor with binding specificity to a *Sox* consensus motif. Mutations in this gene have been associated with 'ragged' mutation in mice (Pennisi et al., 2000), and lymphatic dysfunction in the human hypotrichosis-lymphoedema-telangiectasia syndrome (Irrthum et al., 2003). In

the ‘ragged’ mice model SOX18 expression was found in the developing vascular endothelium and hair follicles. The ’ragged’ mutation has phenotypes ranging from thin, ragged coats to almost baldness with oedema and cyanosis, surviving to approximately weaning (Pennisi et al., 2000). In humans, the phenotype is similar: hypotrichosis is the reduction in hair growth, lymphoedema is the swelling of extremities due to impaired lymphatic drainage and telangiectasia is the presence of small widened blood vessels near the surface of the skin (Irrthum et al., 2003). This gene may be important in the differences seen between the structure of the feral and domesticated fleece types. Arapawa lambs have a hairy coat which moults, while adults have short wool (~ 22 micron) with a tendency to break, and limited shedding (Young et al., 2011).

Also, within the agouti region identified on chromosome 13, in both this study and the sheep HapMap study were PLUNC and LPLUNC1. Downstream were others of the same superfamily, which are part of innate bacterial defence. PLUNC and LPLUNC are expressed in the epithelium of the airways and probably have a role in the immune response to foreign microbes, while others, notable the BSP30 proteins, may have a role within the rumen (Wheeler et al., 2007). These genes could be part of the immune defence differences between feral and domesticated animals. Presence in the airways could help prevent pneumonia, and mucosal secretions in the intestines against foreign organisms could lead to differences in faecal consistency and thus dag formation.

Another gene under a peak implicated in disease is CTSZ, where mutations have led to a resistance against tuberculosis in humans (Cooke et al., 2008). This would be a useful trait for feral/wild sheep which may graze land also in the territory of tuberculosis carriers, however tuberculosis is a very rare disease in New Zealand sheep (Allen, 1988). A GWAS for human tuberculosis susceptibility identified a region containing melanocortin 3 receptor (MC3R) and CTSZ as significant in an African population (Cooke et al., 2008). An T-C SNP in the 3’ UTR of CTSZ and an A-G SNP in MC3R gene were found to be associated. The CTSZ gene belongs to a family of cysteine proteases, a major component of lysosomal proteolytic system, involved in protein degradation and turnover. However, CTSZ is also thought to be involved in macrophages and host defence. CTSZ is active during early phagolysosome biogenesis, but disappears during phagosome maturation (Santamaría et al., 1998).

Several genes, implicated in red blood cell formation, were identified; SLC4A1, ITPA on chromosome 13 and APEX2 on chromosome X. Of these ITPA has evidence

of a role in mitigating the affects of a drug treatment, whose side effects include induced anaemia (Fellay et al., 2010). Anaemia is a common side effect of blood sucking internal parasites such as *Haemonchus contortus* and as such are likely selection targets, albeit their relationship with dagginess is unknown. ITPA hydrolyses inosine triphosphate (ITP), and protects the cell for accumulation of nucleotides such as ITP which could be incorporated into RNA or DNA (Sumi et al., 2002). Deficiency of ITPA leads to accumulation of ITP in red blood cells, and has been shown to protect against ribavirin-induced anaemia, possibly by competing against ribavirin-triphosphate in cellular processes (Fellay et al., 2010). Ribavirin is often used to treat hepatitis C virus affected patients. SLC4A1, also known as band3, is a multifunctional transport protein for glucose, anions and water and is expressed in numerous cell types (reviewed in Kay et al., 1990). The degraded product of this protein also marks senescent and damaged cells for removal by the immune system. Band 3 has been implicated in a number of red blood cell disorders: elliptocytosis/ovalocytosis, spherocytosis, distal renal tubular acidosis (reviewed in Delaunay, 2007), acanthocytosis, choroacanthocytosis and Diego blood group (Bruce et al., 1994; Bruce et al., 1993; Kay et al., 1989). In a bovine anaemia case, Band 3 was shown to contribute to cell membrane stability, CO_2 transport and acid-base homeostasis (Inaba et al., 1996). They found that Band 3 was not essential for survival of the mammal. APEX2 initiates the repair of apurinic/apyrimidinic sites left after base excision repair of damaged bases. Apex-2 null mice exhibit growth retardation, where Apex-2 null male mice have 20% less body weight than wild type (Ide et al., 2004). There was also an indication of reduced proliferation of bone marrow stem cells and abnormal peripheral mature blood cells (Ide et al., 2004).

Of the other genes GALNTL6 may play a role in dag formation, because of its expression in the pancreas and a mutation in this gene in humans has led to its increased expression in the colon epithelia and colon cancer. If there is a disruption in the absorption of water in the colon due to variants in this gene, it could affect faecal consistency and dag formation. GALNTL6; This gene belongs to the family UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase (ppGalNAc-T), which are conserved among species (Bennett et al., 2011), and appears to be one of the first glycosyltransferases in early eukaryote evolution (Kaneko et al., 2001). The ppGalNAc-T genes control the first step of mucin-type O-glycosylation, the transfer of GalNAc from the sugar donor UDP-GalNAc to serine and threonine residues. There are 20

members, each with unique substrate specificity and expression pattern (Peng et al., 2010). The GALNTL6 gene has been found to be expressed strongly in placenta, trachea and mucous cells, and weakly in brain, pancreas and serous cells (Bennett et al., 1999; Bennett et al., 2011). It has also been shown to have increased expression in colon adenocarcinomas, and to play a role in gastric carcinogenesis, and tumour stage ductal breast carcinogenesis (reviewed in Bennett et al., 2011) .

An interesting gene identified was SLMO2, so called for a mutated form showing a phenotype of slow locomotive behaviour of Drosophila larvae (Carhan et al., 2004). It was found to be associated with the mitochondria and was involved in the nervous system during embryogenesis and early larval development. The gene was also associated with the maintenance and survival of the germline in male and female Drosophila (Reeve et al., 2007). This gene may be related to the differences in behaviour seen between feral and domesticated sheep when they are approached. However, a big component in domesticated sheep would be learned behaviour from consistent handling.

PFKFB1 and HSD17B10 were 2 genes implicated in metabolism, with PFKRB1 involved in metabolism under anaerobic conditions. PFKFB1 degrades fructose 2,6-bisphosphate when oxygen becomes limiting and energy generation shifts to a glycolytic mode (Minchenko et al., 2003). PFKB1 is 1 of 4 isozymes that are responsive to hypoxia. PFKB1 is highly expressed in liver, heart, skeletal muscle and testis under hypoxic conditions. HSD17B10, also known as 3-hydroxyacyl-CoA dehydrogenase, short chain is a mitochondrial multifunctional enzyme that helps metabolise estrogen, neuroactive steroids, isoleucine and branched-chain fatty acids (reviewed in Yang et al., 2005). Most mutations in this gene are associated with neurological disorders such as Alzheimer's disease, but also 3-kethothiolase deficiency. The role of these 2 genes in the metabolism of Arapawa and Romney sheep is unknown, but may reflect the type of available feed and geographical location of the sheep breeds before domestication or after reversion back to a feral lifestyle.

For structural fleece traits in both the GWAS and F_{ST} analysis there were no known genes identified in the regions of interest e.g. keratin genes. One gene, fibroblast growth factor 5 (FGF5), that was noted in the sheep HapMap study as having known association with wool (Kijas et al., 2012a) was not detected in this study. This may be because we are comparing sheep with similar fleece types in the GWAS and not the

wide variety of fleece types including hair sheep that was seen across the wide variation of breeds used in the HapMap selection sweep study. However, it does not explain why this gene and others were not identified in the Arapawa and Romney comparison.

There are other analysis techniques for detecting selection sweeps, such as the integrated haplotype score (iHS, Voight et al., 2006) and Cross Population Extended Haplotype Homozogysity (XP-EHH, Sabeti et al., 2007) which are based on linkage disequilibrium. In humans, where these tests were designed and used, millions of SNPs are currently genotyped. Sheep are limited to a 50K SNP Chip, however these methods can be investigated in future.

7.6 Conclusions

Candidate genetic regions were identified for production, breech and fleece traits, in the GWAS. A number of regions for the fibre traits reached the nominal significance threshold of $P<0.001$. The number of animals that had reliabilities greater or equal to $0.8h^2$ were small (< 220), thus there is insufficient power to detect true associated regions and the regions reaching significance cannot be assumed to be true associations even if the QTL was large enough. Independent validation with an increased number of animals (> 1,500) is required. For the other traits, there were only a small number of significant peaks. This could be due to there not being any loci explaining the significant proportion of the variation. Another reason is that the analysis corrects for population and genetic relatedness structure and this has weakened the ability of the study to detect significant regions. Shown in the GS analysis (Chapter 8) over the whole genome, a significant amount of genetic variation can be explained by the SNPs.

The best SNPs found in this analysis did not validate the case-control GWAS (Chapter 6). Reasons for this include; the use of phenotypes versus the use of EBVs as phenotypes in the 2 different studies; and the use of imputed genotypes in this analysis. Another reason for no consistency between studies is that the SNPs detected are false-positives. As seen in Chapter 6, the power analysis identified the need to collected 1,000 cases to obtain a significant result 70% of the time (when QTL h^2 ~0.03), however only ~850 cases were collected and the ability to detect a significant QTL was considerably reduced.

The selection sweep was designed for confirmation of the regions associated in the GWAS, perhaps surprisingly it showed strong and numerous peaks but none overlap the GWAS. Of peaks identified, several were of known selection sweep loci between the breeds used here e.g. horns and coat colour. Several genes were identified with expression in the gastro-intestinal tract that could potentially influence faecal consistency and faecal mucous composition.

Further investigation with additional measurements on progeny, such as the New Zealand Industry Central Progeny Test would be beneficial, for all traits. Validation on a higher density SNP chip which captures more of the linkage disequilibrium between causative QTL and SNPs would be beneficial for these traits, where it is expected they are controlled by a range of SNPs.

Chapter 8: The impact of genomic selection on dual-purpose selection index including dagginess.

8.1 Abstract

Genomic selection (GS) was modelled to identify the impact of using molecular breeding values (MBVs) on dagginess traits, dag score at 3 and 8 months (DAG3, DAG8), and New Zealand sheep dual-purpose index traits. The GS analysis was performed on a range of dual-purpose production traits including live weight, fleece weight, faecal egg count, dagginess, reproduction and survival traits. The training and validation animals were obtained from 8,705 genotyped and pedigree recorded animals from the breeds; Romney, Coopworth, Perendale, Texel and composites of these 4 breeds, categorised in three groups (CompRCP, CompRCPT and CompCRP) according to their breed composition. The breed combined-accuracies ranged from 0.15 to 0.60 for Romney, Coopworth, Perendale and CompRCP which were represented in the training set, and -0.05 to 0.52 for Texel, CompRCPT and CompCRP, which were only in the validation set. The contribution of the MBVs to a dual-purpose index was calculated as effective number of progeny using the combined breed and the individual MBV accuracies. For the majority of production traits, the accuracies of the MBVs contribute 1 to 8 extra measured progeny per sire. For number of lambs born, survival and survival maternal this ranged from 14 to 145 extra progeny per sire. Combined with reducing the generation interval of rams used from 2 years to 1 year, the greatest increase in genetic gain was 84% when using MBVs in a New Zealand dual-purpose index.

8.2 Introduction

DNA testing to genetically improve specific traits is not a new technology. However, these tests took a long time to develop as little information on the species' genome was available and historically, technologies to genotype markers were time consuming and costly. Secondly, it is now known that for many traits, such as disease and production traits, a large number of loci of small effect contribute to the majority of the genetic variance. Thirdly, combining DNA marker information into estimated breeding values (EBVs) is complex, and generated a barrier to the application of marker-assisted selection.

The advent of high-throughput sequencing of numerous domesticated species, including bovine and ovine, has led to the discovery of hundreds and thousands of

single nucleotide polymorphisms (SNPs). This was coupled with decreasing costs for genotyping large numbers of DNA markers. Combining these technologies led to the GS revolution.

Genomic selection is a form of marker-assisted selection on a genome-wide scale (Goddard and Hayes, 2007), by the use of genomic breeding values (GBVs). The animals are genotyped with thousands to hundreds of thousands of SNPs across the genome. These SNPs are assumed to be in linkage disequilibrium (LD) with quantitative trait loci (QTL), which contribute to the variation seen in a particular trait. Therefore, due to the LD, the effect of each SNP for a trait can be estimated, even if the SNP has no true effect itself on the trait (Hayes et al., 2009b). The sum of these effects for each animal generates a molecular breeding value (MBVs) and these MBVs[3] can then be combined with the EBVs calculated from phenotypic information to generate a genomic breeding value (GBV). Meuwissen et al. (2001) used simulation to show that true breeding values (TBVs) could be predicted with an accuracy of 0.85 using modest SNP density in a population similar to dairy cattle. They estimated that using MBVs could potentially lead to the doubling of the rate of genetic gain already achieved through the use of EBVs calculated from phenotypic information alone. This was achieved by a combination of increased accuracy and reduced generation interval (Schaeffer, 2006). However, it is now widely accepted that gains will be smaller than a double of the rate of genetic gain in dairy cattle and that MBV accuracies are generally in the range of 43 to 65% (Pryce and Daetwyler, 2012). For sheep, genetic progress of 5% per annum for a dual purpose index (Sise and Amer, 2009), a 32% increase for a terminal index and a 38% increase for a fine wool index (van der Werf, 2009), have been predicted.

The dairy industry has already adopted GS very successfully. It used to be the norm to choose bulls for progeny testing based solely on the parent-average EBVs. The results from a preliminary review of the Australian, New Zealand, United States and Netherlands dairy indusries has shown the reliabilities from MBVs are greater than predicting EBVs from parent averages, in fact it is nearly as accurate as progeny testing, but results are available from birth (Hayes et al., 2009b).

[3] Note the nomenclature used in this chapter for breeding values calculated from SNP information, differs from that used in literature. Here we use molecular breeding value (MBV) for BVs calculated from SNP information and genomic BV (GBV) for blended BV from estimated BV and MBV. References often use genomic estimated BV (GEBV) for BVs estimated from SNP information.

The implications of GS and the ability to scan the whole genome using thousands of variants; means that breeding values for traits that are difficult to measure, are sex-limited and are observed later in life, can now be estimated (McEwan et al., 2010). This chapter estimates GS parameters for Sheep Improvement Limited (SIL) dual-purpose index selection traits, including dagginess, and investigates the potential benefits they have on selection decisions in the industry.

8.3 Materials and methods

8.3.1 Data

Estimation of breeding values (BVs) was undertaken by SIL (S-A. N. Newman pers. comm.) for the traits: weaning weight at 3 months (WWT), WWT maternal (WWTm), carcass weight (CW), live weight at 8 months (LW8), adult ewe weight (EWT), lamb fleece weight (LFW), fleece weight at 12 months (FW12), adult fleece weight (AFW), dag score at 3 and 8 months (DAG3, DAG8), number of lambs born in ewes (NLB), survival (SURV), SURV maternal (SURVm), and faecal egg count in summer (FEC1), autumn (FEC2) and adult (AFEC). The analysis was performed on 3,535,557 animal records for animals born between 1990 and 2010 for 233 SIL flocks and used SIL's normal data cleaning and scaling procedures for these traits. The traits were fitted in multi or single-trait BLUP analysis (S-A. Newman pers. comm.) using ASReml (Gilmour et al., 2009).

Estimated breeding values were converted back to the original range using summary information provided by SIL. This was done for the EBVs calculated by SIL, as they had been centred to the year 1995.

Apart from CW, EWT, LFW, AFW, SURV, and AFEC, the traits have been previously described in chapter 3. Carcass weight in SIL, is estimated from correlations with other traits, namely LW8, as it is a trait not normally recorded in industry. Adult ewe weight is taken as live weight at 30 months, lamb fleece weight and adult fleece weight are not normally recorded in SIL and their EBVs are estimated from correlations with other traits, primarily FW12. Lamb survival is measured as alive or not at birth, a maternal component is also estimated (SURVm). Adult FEC is another trait not normally recorded in SIL and is estimated from correlations with other traits including FEC1, FEC2, and the analogous *Nematodirus* traits NEM1 and NEM2.

8.3.2 Genotypes

The genotypes consisted of 13,559 animals, mainly sires, genotyped with the Illumina OvineSNP50 BeadChip (50K SNP Chip). Animals were sourced from industry and research flocks, including those involved in the progeny test and flystrike case-control studies described in previous chapters. The animals were predominantly Romney, Coopworth, Perendale or Texel, plus other breeds and various breed crosses and composites. This resource contains animals involved in numerous Ovita-related research projects. Only those animals that were SIL recorded (n = 8,705) and were present in the phenotype data were used for further analysis.

Genotyping of the 13,559 animals was performed using the 50K SNP Chip in 2 sets. One set was genotyped by Illumina, the second set was genotyped at AgResearch – Invermay. The Illumina set included 2,805 Ovita research animals. The AgResearch data set included all those genotyped since 2009. For each set the manifest and allele calling parameters were as defined in the ***Ovine_iSelect_2880.egt*** project. These were the original settings as defined by Illumina. Due to minor changes in chemicals supplied from Illumina since February 2011, new allele calling parameters as defined in the ***OvineFebruary2011.egt*** project was used for those animals genotyped since February 2011.

8.3.3 Quality control

Genotyping results were put through a quality control pipeline before analysis. The pipeline has been described by Dodds et al. (2009). In summary, SNPs are discarded if they have a call rate <97%, appear non-autosomal (including pseudoautosomal), minor allele frequency (MAF) = 0, and weighted Gencall 10 score (GC score) <0.422. SNPs that were not retained as part of the Ovine HapMap study (Kijas et al., 2012a) were also discarded.

8.3.4 Statistical analysis

8.3.4.1 Calculating the dependent variables

Dependent variables (y) were calculated taking into account the individuals own and descendants' information. Parent average effects were removed from EBVs using the method described by Garrick et al. (2009), assuming all genetic variation was explained by the markers (c = 0). In summary, the portion of the EBV derived from the parent

average information was removed. The resulting value was deregressed using the reliability of the EBV, with parent-average removed.

8.3.4.2 Genomic selection analysis and validation

8.3.4.2.1 Breed designation and reliability threshold

Data for each trait was filtered on breed and reliability before analysis. Breed was designated by the following conditions: Romney, Coopworth, Perendale and Texel were reported if their breed composition was greater than or equal to 75%. There were also three composite breeds considered. CompRCPT were those that were greater than 50% of combined Romney, Coopworth, Perendale and at least 25% Texel. CompRCP were those that were greater than 50% of combined Romney, Coopworth, Perendale and less than 25% Texel. Finally CompCRP were those that had greater than 30% and less than or equal to 50% of combined Romney, Coopworth and Perendale. The reliability cut off was 80% of the heritability estimate used for EBV estimation. Animals had to have dependent variable reliabilities equal to or above this cut off to be considered for analysis.

8.3.4.2.2 Training and validation assignment

After the above filtering, genotypes were scored on the number of copies of the 'A' allele (based on Illumina AB calling format). Missing genotypes were filled in using the breed mean, estimated using a least squares regression on breed proportions as Romney, Coopworth, Perendale, Texel and other to generate allele frequencies for each SNP within breed. The missing values were then replaced, weighted by the individuals breed proportion of Romney, Coopworth, Perendale, Texel or other. The Romney, Coopworth, Perendale and CompRCP animals were split into validation and training sets (Table 8.1). Texel, CompRCPT and CompCRP animals were only used in validation, to see how well the predictions work for them. Cut off years were chosen for each breed, using a number of criteria. First, at least 200 animals per breed were used for validation. Secondly, if there were less than 400 animals, roughly half were required in each set. Thirdly, if there were between 75 and 100 animals then a small portion (~10) were left in the training set, and the rest in validation. Finally, if there were less than 75 animals, all were added to the training set.

Table 8.1: The year of birth of the first animals placed in the validation set and number (n) of animals in training and validation sets for each breed[1].

	First validation year				n Training				n Validation					
Trait[2]	R	C	P	RCP	R	C	P	RCP	R	C	P	RCP	T	RCPT	CRP
WWT	2009	2009	2006	2009	3600	1345	281	418	637	340	202	357	312	242	263
WWTm	2007	2006	2005	2004	1730	614	184	11	218	320	196	186	241	100	128
CW	2009	2009	2006	2009	3548	1272	267	347	596	333	230	349	312	229	239
LW8	2009	2009	2006	2009	3513	1266	262	347	585	330	221	349	312	229	239
EWT	2008	2007	2005	2007	1998	799	148	125	346	295	191	172	290	148	171
LFW	2008	2005	2006	2004	1898	399	202	11	325	220	161	86	149	73	110
FW12	2009	2009	2006	2009	3397	1110	265	239	495	317	228	275	203	195	198
EFW	2007	2001	2004	2011	1285	182	94	34	351	192	145	0	81	28	56
NLB	2006	2006	2005	2006	1386	607	185	61	479	321	184	133	224	97	113
SURV	2009	2008	2006	2009	2572	1094	262	310	355	270	212	280	299	201	250
SURVm	2008	2007	2005	2007	1899	792	186	144	326	257	191	212	246	138	180
DAG3	2008	2009	2004	2009	624	622	52	188	221	234	56	276	86	158	123
DAG8	2008	2005	2004	2009	715	209	50	72	278	245	53	83	86	85	81
FEC1	2008	2009	2005	2008	1414	1033	164	222	264	239	185	204	124	160	155
FEC2	2008	2009	2005	2007	1168	917	175	101	165	95	193	137	98	97	123
AFEC	2006	2005	2004	2005	771	381	123	10	252	237	170	66	76	54	64

[1]R: Romney; C: Coopworth; P: Perendale; RCP: CompRCP (> 50% combined R, C, P and < 25% T); T: Texel; RCPT: CompRCPT (> 50% combined R, C, P and <25% T); CRP: CompCRP (> 30% and <= 50% combined R, C and P).

[2]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count;

8.3.4.2.3 Genomic relationship matrices

Two genomic relationship matrices were used. The G1 matrix was used to calculate the coefficients, while G6z is used to calculate individual accuracies as described below. The G1 matrix is calculated using the equation (VanRaden, 2008):

$$G1 = \frac{ZZ\prime}{2\sum p_i(1-p_i)} \quad \text{Equation 1}$$

Where Z is the SNP matrix $-2p_i$, $1-2p_i$, and $2-2p_i$ for BB, AB, and AA respectively, and p_i is the frequency of A allele of the i^{th} SNP in the population.

G6z is similar to G1 except the j^{th} row of Z has values $-2p_{ij}$, $1-2p_{ij}$, and $2-2p_{ij}$ for BB, AB, and AA respectively, and p_{ij} is the population allele frequency for the j^{th} animal's breed composition:

$$p_{ij} = \Sigma_b p'_{bi} s_{bj} / \Sigma_b s_{bj} \quad \text{Equation 2}$$

Where p'_{bi} is the allele frequency for SNP i and breed b and s_{bj} is the recorded proportion of breed b for the j^{th} animal. The divisor is; $\Sigma p_{ij}(1-p_{ij}) + \Sigma p_{ik}(1-p_{ik})$ for the $(j,k)^{th}$ element of ZZ', where p_{ij} is the population allele frequency for the j^{th} animal's breed composition. Finally, any negative elements of the relationship matrix were set to zero.

The first (G1) was used to calculate the principal components (PC), as the PCs obtained from G6z do not differentiate the breeds. The first 6 PCs of the genomic matrix (G1, VanRaden, 2008) were used to account for population stratification, in the models below.

8.3.4.2.4 Molecular breeding values

MBVs were calculated using the methods of VanRaden (2008). A mixed model was fitted to *y* as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e} \quad \text{Equation 3}$$

Where $\boldsymbol{y}$ is the de-regressed EBV with parent average removed, $\mathbf{X}$ is a matrix of the first 6 PC of the G1 matrix, $\boldsymbol{\beta}$ is a vector of fixed effects of the PC, $\mathbf{Z}$ is an incidence

matrix and **u** is the animal effects distributed as N(0, G1 σ^2_u), where σ^2_u is the additive genetic variance, and **e** are the residual effects distributed as N(0,R) where R is a diagonal matrix with diagonal elements $(1-r^2)/r^2$ where r^2 is the reliability of the deregressed EBVs.

The analysis used a fixed heritability by fixing σ^2_u at the value used to calculate the EBVs. This analysis is termed a GBLUP and is equivalent to a weighted animal model BLUP, but using a genomic relationship matrix (G1) as the numerator relationship matrix. All animals, Romney, Coopworth, Perendale and CompRCP (training and validation), were used in this analysis.

The MBVs were the predicted animal effects from the above model – see equation 2 of VanRaden (2008). For convenience, these were calculated as a sum of SNP effects. Firstly, the estimated SNP effects were obtained by substituting Z′ for the leftmost G in equation 2 of VanRaden (2008), and then dividing by $2\Sigma p_i(1-p_i)$ (Z and p_i are defined above) (VanRaden, 2008). The MBVs were obtained by multiplying the SNP effects (often termed SNP coefficients in this context) by the SNP genotypes (BB=0, AB=1, AA=2) and summing.

To get the individual accuracies the G1 matrix was substituted for the G6z matrix in the above model.

8.3.4.2.5 Accuracies

The accuracies of the MBVs were derived using 2 different methods. For the first method (using G1);

$$r_A = \frac{cor(y,MBV)}{h_g} \qquad \text{Equation 4}$$

and was weighted, averaged within each breed by $1/(1-r^2)$, and calculated for each breed using the animals in the validation set. The MBV were calculated as above, but only using the training set. The effective heritability from the GS analysis (h^2_g) was used and is equal to the average reliability (r^2).

The second method used the prediction error variance (PEV, Mrode, 2005) from the genomic BLUP analysis (using G6z) giving;

$$r_I = \sqrt{1 - \frac{PEV_i}{\sigma^2_u}} \qquad \text{Equation 5}$$

The PEV is obtained by inverting the LHS of the mixed model equation (Harris and Johnson, 1998). This was calculated for all validation animals and averaged (weighted by $1/(1-r^2)$) within each breed.

The r_A is the realised accuracy and relies on the animals in the validation set, thus, there may be a selection bias. The r_I is a model-based accuracy and can be affected if the model is inaccurately specified. Therefore an average of r_A and r_I was taken; hereafter this is referred to as the breed combined-accuracy (r_C).

8.3.4.2.6 P values

To compare the results from this analysis to the genome-wide association study (GWAS) performed on the same dataset used in Chapter 7, *P* values were calculated for the ith SNP assuming the SNP coefficients (bi), follow a normal distribution with mean zero and variance:

$$var_i = \frac{2pi(1-pi)n_b\sigma_b^2}{\sum(2pi(1-pi))} \qquad \text{Equation 6}$$

where σ^2_b is equal to the empirical variance of b_i, and n_b is the number of SNPs with coefficients.

The $-\log_{10}(P)$ values were graphed in a Manhattan plot on Ovine genome v2, and thresholds set at an initial level calculated using the 5% Bonferroni correction (Rice, 1989) $0.05/n_{SNPs}$ equal to *P* value of $1.03\text{x}10^{-6}$ or –log10(*P*) of ~6, and a lower nominal threshold at $P<0.001$ ($-\log_{10}(P) = 3$).

A comparison of the results from the genome-wide association study on industry and progeny test dataset (Chapter 7) with the estimates derived here was performed as follows. The $-\log_{10}(P)$ values for each trait from the separate analyses were graphed and the linear correlation calculated.

8.3.4.2.7 QQ plots

The quantile - quantile (QQ) plots were calculated to check whether the distribution of the observed $-\log_{10}(P)$ values deviated from the expected distribution (exponential) under the null hypothesis of no genetic association and no LD between SNPs. To do so, the n $-\log_{10}(P)$ values were sorted and plotted against the $-\log_{10}(1-u)$ where u = [1 / (n + 1), 2 / (n + 1), …, n / (n + 1)]. If the distributions are similar then the slope, lambda, should be equal to 1. If the observed deviates from the expected with a slope greater than 1 this indicates that a proportion of the substructure has not been accounted for. A

slope smaller than 1 is more difficult to explain, but could mean the LD between SNPs and between SNPs and QTL is reducing the number of effective tests, thus producing larger *P* values than expected. When the right-hand tail diverges above the slope these diverging SNPs could be considered associated with the trait.

8.3.5 Comparison of genetic gain

To estimate the impact of using MBVs in the New Zealand sheep industry, a simulated breeding scheme was set up to test the impact of adding genotype information in a dual-purpose index, using the Romney breed as an example.

Using the MBV accuracies (r_I and r_C) estimated in the Romneys, the number of effective progeny were estimated using the single trait selection index worksheet (van der Werf, 2006b). Heritability and repeatabilities were from Chapter 3 or were those used for SIL breeding value analysis (S. A. Newman, pers. comm.). Appendix 19 details the genetic and phenotypic correlations. Using the worksheet, the effective progeny numbers were obtained by changing the number of measurements on the progeny (n measures on individuals, parents, or siblings) until the accuracy was similar to the r_C or r_I calculated above for each trait.

The multiple trait selection index worksheet (van der Werf, 2006b) was used to estimate the response per selection round for a given breeding scheme scenario. Selection was on an index including the following objective traits: WWT, WWTm, CW, LW8, EWT, LFW, FW12, AFW, NLB, SURV, SURVm, DAG3, DAG8, FEC1, FEC2, and AFEC.

- Scenario 1 assumed selection was on animal measurements either for a ram hogget (Scenario 1-A) or a 2 year old ram (Scenario 1-B) which is used only once, in a dual-purpose Romney flock.

Depending on the trait, the maximum number of measurements available was 1 measurement on the individual and sire, 2 on the dam and 126 on half siblings (Table 8.2). For CW, LFW, AFW, FEC2 and AFEC no measurements were taken for these traits and these traits were estimated from their correlations with the other traits.

- Scenario 2: ram hoggets have been genotyped with the 50K SNP Chip results and the contribution of the SNP Chip predictions were measured as effective number of progeny. The number of effective progeny were calculated from r_I (2-A) or r_C (2-B). These effective progeny numbers were

added to the animal measurements table as progeny for a ram hogget from Scenario 1-A. Note this method assumes that the errors of the MBVs and traditional EBVs are uncorrelated and this approach is equivalent to simple blending as outlined by Mrode (2005).

Dual-purpose economic weights for the traits were taken from Byrne et al. (2012). To convert the FEC1, FEC2 and AFEC economic weights from Byrne et al. (2012) from % to $\log_e$ the values were multiplied by 100.

Table 8.2: Phenotypic standard deviation (σ_p), heritability (h^2), repeatability (Rep), economic weights (EW), and number of records used for each trait in the dual-purpose selection index for selecting a 2 year old ram.

						Number of records used			
Trait[1]	Units	σ_p	h2	Rep	EW($)	Own	Dam[2]	Sire	Half sibs[3]
WWT	Kg	4.09	0.14	0.14	0.95	1	1	1	126
WWTm	Kg	4.07	0.2	0.2	0.84	0	1	0	0
CW	Kg	2	0.3	0.3	2.60	0	0	0	0
LW8	Kg	4.51	0.35	0.35	0.00	1	1	1	126
EWT	Kg	7	0.45	0.45	-1.04	0	1	0	0
LFW	Kg	0.2	0.15	0.15	1.82	0	0	0	0
FW12	Kg	0.42	0.37	0.37	0.79	1[4]	1	1	63[4]
AFW	Kg	0.5	0.45	0.45	2.28	0	0	0	0
NLB	Lambs	0.62	0.09	0.12	15.55	0	2	0	0
SURV	Lambs	0.4	0.01	0.01	64.45	1	1	1	126
SURVm	Lambs	0.4	0.008	0.008	58.40	0	2	0	0
DAG3	Units	1.28	0.34	0.34	-0.34	1	1	1	126
DAG8	Units	1.43	0.31	0.31	-0.35	1	1	1	126
FEC1	$\log_e$	0.77	0.18	0.62	-3.00	1	0	1	63
FEC2	$\log_e$	0.76	0.19	0.57	-3.00	0	0	0	0
AFEC	$\log_e$	0.74	0.25	0.28	-3.00	0	0	0	0

[1]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

[2]Average dam age at lambing is 3.37, therefore the individual's dam on average has 2 records on NLB and SURVm.

[3]There are ~127 (90*1.41) lambs weaned each year/sire, thus each individual has 126 half siblings (sibs) and as FEC1 is only measured on males ~63 measures are available.

[4]For selection as a ram hogget there will be 0 records on individual and half-sibs.

The results from the multitrait selection are presented as the selection response per 'selection round'. This was converted to selection response per year by multiplying by

i/L, where i is the average female and male selection intensity and L is the average male and female generation interval. Assumptions on the breeding scheme were;

- A Romney flock of 631 ewes (Young and Wakelin, 2009);
- Rams were used once at a ratio of 1:90. For scenario 1-B: 2 year old rams are used. For scenario 1-A and 2-A and B: ram hoggets were used.
- Lambing percentage (lambs weaned/ewes mated) were 141%, based on NLB per ewes present (1.71) from chapter 2, the proportion of lambs weaned per lambs born (0.86) and number of ewes present at lambing to number of ewes mated (0.98) (Jopson et al., 2000; McEwan et al., 1992a);
- Ewes lambed first at 2 years of age and retained to 5 years of age, there was a 10% death and culling rate each year;

Therefore for scenario 1-B: i/L equalled 1.73/2.68 equal to 0.64 and for scenario 1-A and 2-A and B: i/L equalled 1.73/2.18 equal to 0.79.

8.4 Results

8.4.1 Genomic selection

8.4.1.1 Preliminary data analyses and quality assurance steps

Quality control was performed on the animals and SNPs prior to the analysis for this PhD study. This was done as part of the ongoing updating of MBV's for commercial use. Of the 13,599 animals genotyped, only 13,248 had genotypes (call rate >0) and there were 13,157 unique animals genotyped. Of these 8,705 animals were SIL recorded, and used for this analysis. Of the 54,977 loci called, 48,327 autosomal loci passed quality control and were subsequently used.

8.4.1.2 Population structure

Fitting the first 6 PCs, calculated using animals in the training set, accounted for the majority of the population structure (0.51 to 0.73 of the genetic variation contained in the genomic relationship matrix, Table 8.3). The first 4 PCs, calculated using all animals to show the full population in analysis, for WWT are shown in Figure 8.1. Table 8.3 shows the population variation explained by the first PCs and the additional

variance explained by the second to sixth PCs for each trait. The majority of the variance (above 44% for most traits) was explained by the first PC.

Table 8.3: The breed genotypic variance explained by the first 6 principal components (PC) for each trait.

Trait[1]	PC1	PC2	PC3	PC4	PC5	PC6	Total
WWT	0.52	0.07	0.06	0.03	0.02	0.01	0.72
WWTm	0.47	0.05	0.04	0.03	0.03	0.02	0.64
CW	0.52	0.07	0.06	0.03	0.02	0.01	0.72
LW8	0.52	0.07	0.06	0.03	0.02	0.01	0.72
EWT	0.54	0.05	0.03	0.02	0.02	0.02	0.68
LFW	0.37	0.07	0.05	0.04	0.02	0.02	0.57
FW12	0.51	0.08	0.06	0.03	0.02	0.01	0.71
AFW	0.30	0.08	0.05	0.04	0.02	0.02	0.51
NLB	0.47	0.05	0.04	0.03	0.03	0.02	0.65
SURV	0.55	0.06	0.04	0.03	0.03	0.01	0.71
SURVm	0.52	0.05	0.03	0.03	0.02	0.02	0.68
DAG3	0.57	0.06	0.03	0.02	0.02	0.01	0.71
DAG8	0.44	0.06	0.03	0.03	0.02	0.02	0.60
FEC1	0.59	0.04	0.04	0.02	0.02	0.02	0.73
FEC2	0.57	0.05	0.03	0.03	0.02	0.02	0.71
AFEC	0.44	0.06	0.04	0.04	0.02	0.01	0.60

[1]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

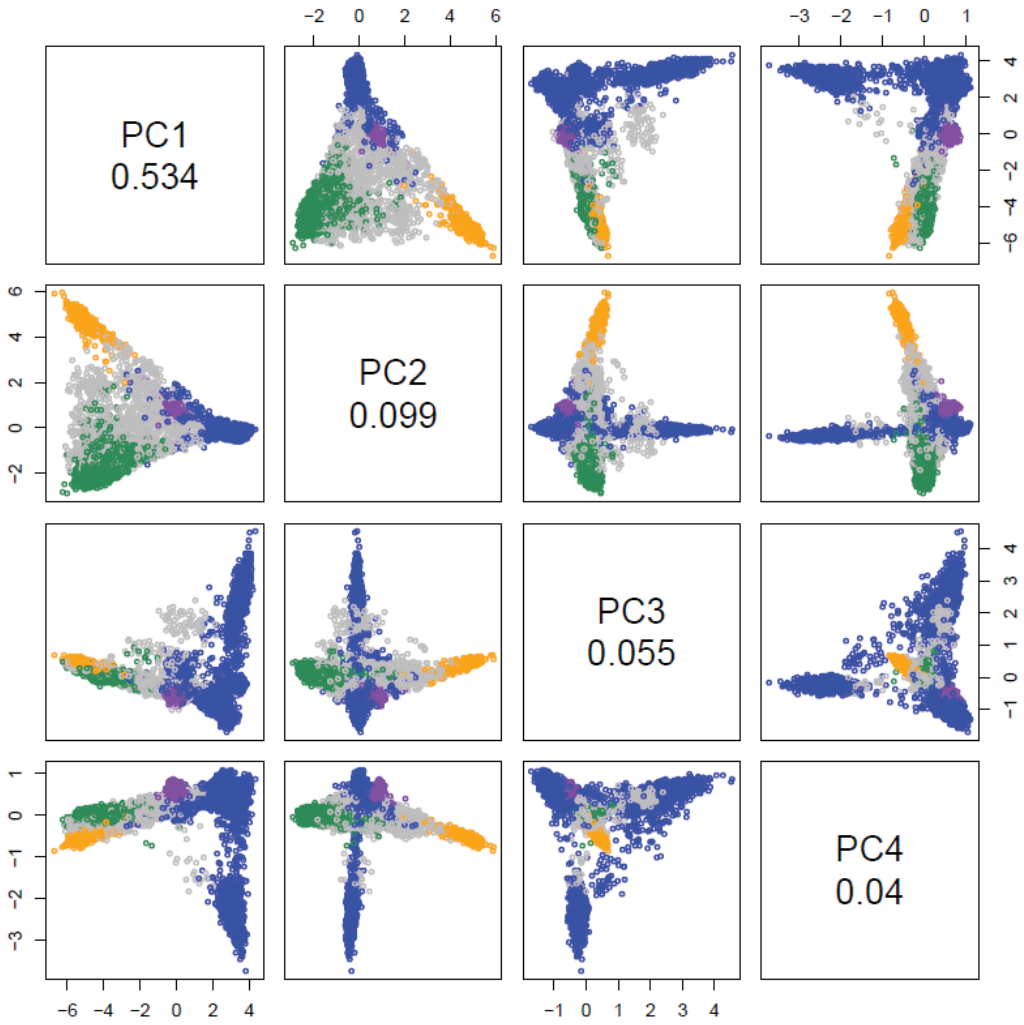


Figure 8.1: The first 4 principal components (PC) calculated from all animals for weaning weight. Romney (blue), Coopworth (green), Texel (yellow), Perendale (purple) and others/composites (grey).

8.4.1.3 Predictive power

The variance components and heritabilities are shown in Table 8.4. Table 8.5 shows the weighted correlations between the MBV and *y* from the G1 and G6z analysis for each trait and breed, for comparison. Table 8.6 shows the accuracy of the weighted accuracy of the correlation between MBV and *y* from the G1 analysis (r_A), the weighted individual accuracy (r_I) and the breed combined-accuracy (r_C). The lower 90% confidence intervals, were calculated from the G1 matrix using Fishers transformation (equation 11.3, Steel and Torrie, 1980), and these are reported in Table 8.7. These

figures provide a threshold of what traits and for what breeds MBVs should be reported. The general threshold is a confidence interval greater than zero. Thus for Texels MBVs will only be reported for live weight traits. Table 8.8 shows the proportion of variance of the trait explained by the MBVs. The proportion is equal to the square of the accuracies; this was calculated using the r_C accuracies as shown in Table 8.6.

Table 8.4: Variance components; additive (σ_u^2) and corrected residual (σ_e^2) variance, and the heritabilities used; (h^2) from Sheep Improvement Limited (SIL) and the effective h^2 for each trait.

Trait[1]	h^2 SIL	σ_u^2	Corrected σ_e^2	Effective h^2_g
WWT	0.18	2.45	2.17	0.53
WWTm	0.12	1.59	1.01	0.61
CW	0.30	1.20	0.90	0.57
LW8	0.40	10.00	5.14	0.66
EWT	0.45	22.05	13.13	0.63
LFW	0.15	0.01	0.02	0.23
FW12	0.35	0.09	0.06	0.58
EFW	0.45	0.11	0.15	0.43
NLB	0.10	0.03	0.04	0.44
SURV	0.01	0.002	0.01	0.17
SURVm	0.008	0.001	0.01	0.12
DAG3	0.33	0.54	0.61	0.47
DAG8	0.31	0.64	0.59	0.52
FEC1	0.16	0.11	0.15	0.43
FEC2	0.20	0.15	0.22	0.40
AFEC	0.25	0.18	0.34	0.35

[1]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

Table 8.5: The weighted correlations from G1 and G6z analysis, between MBVs and dependent variables in the 7 validation breeds[1].

Trait[2]	G matrix	Rom	Coop	Peren	RCP	Texel	RCPT	CRP
WWT	G1	0.26	0.30	0.04	0.34	0.17	0.36	0.29
WWT	G6z	0.21	0.40	0.03	0.33	0.26	0.29	0.27
WWTm	G1	0.10	0.13	0.17	0.36	0.11	0.02	0.28
WWTm	G6z	0.14	0.13	0.14	0.34	0.07	0.03	0.35
CW	G1	0.28	0.30	0.09	0.37	0.18	0.29	0.28
CW	G6z	0.24	0.41	0.06	0.35	0.22	0.26	0.28
LW8	G1	0.30	0.34	0.11	0.41	0.21	0.34	0.33
LW8	G6z	0.28	0.46	0.08	0.37	0.25	0.29	0.28
EWT	G1	0.32	0.29	0.04	0.25	0.12	0.20	0.28
EWT	G6z	0.31	0.32	0.01	0.18	0.16	0.13	0.32
LFW	G1	0.11	0.19	0.11	0.02	0.00	0.08	-0.14
LFW	G6z	0.12	0.19	0.09	-0.02	-0.05	0.19	-0.05
FW12	G1	0.31	0.47	0.17	0.31	0.01	0.26	0.18
FW12	G6z	0.29	0.44	0.22	0.30	0.04	0.30	0.30
EFW	G1	0.16	0.07	0.24	n.v.[3]	-0.05	0.28	0.18
EFW	G6z	0.18	0.12	0.19	n.v.[3]	-0.05	0.31	0.22
NLB	G1	0.36	0.31	0.15	0.20	0.11	0.19	0.19
NLB	G6z	0.38	0.25	0.16	0.21	-0.02	0.18	0.16
SURV	G1	-0.06	-0.03	0.18	-0.03	-0.01	0.16	0.08
SURV	G6z	-0.01	-0.04	0.18	-0.04	-0.08	0.14	0.13
SURVm	G1	0.20	0.12	0.11	0.06	-0.04	0.07	0.19
SURVm	G6z	0.20	0.09	0.06	0.02	-0.08	0.01	0.23
DAG3	G1	0.24	0.38	0.18	0.29	-0.05	0.24	0.27
DAG3	G6z	0.29	0.40	0.20	0.25	0.03	0.20	0.28
DAG8	G1	0.29	0.30	0.08	0.23	0.10	0.29	0.12
DAG8	G6z	0.31	0.27	0.15	0.28	0.03	0.32	0.13
FEC1	G1	0.26	0.46	0.14	0.43	0.02	0.25	0.33
FEC1	G6z	0.29	0.44	0.16	0.43	0.13	0.28	0.33
FEC2	G1	0.31	0.44	0.11	0.43	0.06	0.16	0.42
FEC2	G6z	0.30	0.45	0.12	0.45	0.07	0.15	0.39
AFEC	G1	0.16	0.14	0.14	0.17	0.06	-0.13	0.20
AFEC	G6z	0.18	0.15	0.13	0.19	0.05	-0.11	0.25

[1]Rom: Romney; Coop: Coopworth; Peren: Perendale; RCP: CompRCP; RCPT: CompRCPT; CRP: CompCRP.

[2]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

[3]n.v. no validation set

Table 8.6: Weighted accuracies of the correlation between MBV and dependent variable (r_A), the weighted average individual accuracy (r_I) and the breed combined-accuracy (r_C) calculated for the 16 traits in the 7 validation breeds[1].

	Rom			Coop			Peren			RCP			Texel			RCPT			CRP		
Trait[2]	r_A	r_I	r_C	r_A	r_I	r_C	r_A	r_I	r_C	r_A	r_I	r_C	r_A	r_I	r_C	r_A	r_I	r_C	r_A	r_I	r_C
WWT	0.35	0.60	0.48	0.41	0.60	0.51	0.06	0.49	0.28	0.47	0.56	0.51	0.23	0.31	0.27	0.50	0.54	0.52	0.40	0.48	0.44
WWTm	0.13	0.55	0.34	0.17	0.55	0.36	0.22	0.48	0.35	0.46	0.48	0.47	0.14	0.23	0.18	0.02	0.45	0.24	0.35	0.34	0.35
CW	0.37	0.58	0.48	0.39	0.59	0.49	0.12	0.47	0.29	0.49	0.54	0.52	0.24	0.29	0.26	0.39	0.52	0.45	0.36	0.45	0.41
LW8	0.37	0.62	0.50	0.41	0.63	0.52	0.14	0.50	0.32	0.50	0.58	0.54	0.26	0.32	0.29	0.42	0.57	0.50	0.41	0.50	0.46
EWT	0.41	0.54	0.48	0.36	0.54	0.45	0.06	0.39	0.22	0.32	0.52	0.42	0.15	0.23	0.19	0.25	0.47	0.36	0.35	0.41	0.38
LFW	0.23	0.35	0.29	0.39	0.27	0.33	0.23	0.26	0.25	0.04	0.26	0.15	0.00	0.12	0.06	0.16	0.23	0.19	-0.29	0.18	-0.05
FW12	0.40	0.59	0.49	0.61	0.59	0.60	0.22	0.47	0.35	0.40	0.56	0.48	0.02	0.29	0.15	0.35	0.54	0.44	0.24	0.48	0.36
EFW	0.24	0.39	0.31	0.10	0.26	0.18	0.36	0.26	0.31	n.v.[3]	n.v.[3]	n.v.[3]	-0.07	0.12	0.02	0.42	0.26	0.34	0.28	0.23	0.25
NLB	0.54	0.50	0.52	0.46	0.49	0.48	0.23	0.44	0.33	0.30	0.46	0.38	0.16	0.18	0.17	0.28	0.38	0.33	0.29	0.35	0.32
SURV	-0.14	0.47	0.16	-0.08	0.48	0.20	0.45	0.42	0.43	-0.07	0.44	0.19	-0.03	0.21	0.09	0.41	0.41	0.41	0.20	0.35	0.27
SURVm	0.59	0.36	0.47	0.34	0.42	0.38	0.32	0.37	0.34	0.17	0.36	0.27	-0.13	0.16	0.02	0.20	0.34	0.27	0.56	0.28	0.42
DAG3	0.34	0.46	0.40	0.56	0.52	0.54	0.26	0.35	0.31	0.41	0.44	0.43	-0.07	0.16	0.04	0.35	0.42	0.39	0.39	0.40	0.40
DAG8	0.40	0.47	0.44	0.41	0.43	0.42	0.11	0.31	0.21	0.31	0.35	0.33	0.15	0.13	0.14	0.40	0.33	0.36	0.16	0.30	0.23
FEC1	0.40	0.51	0.46	0.71	0.55	0.63	0.22	0.41	0.31	0.65	0.49	0.57	0.03	0.21	0.12	0.39	0.51	0.45	0.50	0.46	0.48
FEC2	0.49	0.51	0.50	0.69	0.49	0.59	0.18	0.39	0.29	0.68	0.46	0.57	0.09	0.18	0.14	0.26	0.41	0.33	0.66	0.34	0.50
AFEC	0.27	0.35	0.31	0.24	0.35	0.29	0.24	0.28	0.26	0.29	0.33	0.31	0.10	0.11	0.10	-0.22	0.28	0.03	0.33	0.23	0.28

[1]Rom: Romney; Coop: Coopworth; Peren: Perendale; RCP: CompRCP; RCPT: CompRCPT; CRP: CompCRP.

[2]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

[3]n.v.: no validation set

Table 8.7: The lower limit of the 90% confidence interval for the weighted correlation between MBV and dependent variable (r_A), for all traits and all 7 validation breeds[1].

Trait[2]	Rom	Coop	Peren	RCP	Texel	RCPT	CRP
WWT	0.27	0.30	-0.08	0.36	0.11	0.37	0.27
WWTm	-0.01	0.05	0.07	0.32	0.01	-0.19	0.17
CW	0.29	0.28	-0.02	0.39	0.12	0.25	0.23
LW8	0.29	0.31	0.00	0.41	0.15	0.30	0.29
EWT	0.30	0.25	-0.10	0.17	0.03	0.09	0.20
LFW	0.04	0.16	-0.04	-0.34	-0.28	-0.25	-0.60
FW12	0.31	0.52	0.08	0.28	-0.13	0.20	0.08
EFW	0.11	-0.08	0.16	n.v.[3]	-0.35	-0.07	-0.06
NLB	0.44	0.33	0.05	0.09	-0.01	0.03	0.06
SURV	-0.35	-0.33	0.17	-0.31	-0.26	0.12	-0.06
SURVm	0.33	0.04	-0.03	-0.16	-0.43	-0.21	0.21
DAG3	0.19	0.42	-0.06	0.28	-0.33	0.17	0.18
DAG8	0.27	0.27	-0.22	0.06	-0.10	0.16	-0.10
FEC1	0.26	0.58	0.04	0.50	-0.20	0.19	0.31
FEC2	0.30	0.45	-0.01	0.48	-0.17	-0.01	0.45
AFEC	0.10	0.06	0.03	-0.06	-0.22	-0.58	-0.02

[1]Rom: Romney; Coop: Coopworth; Peren: Perendale; RCP: CompRCP; RCPT: CompRCPT; CRP: CompCRP.

[2]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

[3]n.v.: no validation set

Table 8.8: Proportion of variance explained by MBVs in 7 validation breeds[1].

Trait[2]	Rom	Coop	Peren	RCP	Texel	RCPT	CRP
WWT	0.23	0.26	0.08	0.26	0.07	0.27	0.20
WWTm	0.12	0.13	0.12	0.22	0.03	0.06	0.12
CW	0.23	0.24	0.09	0.27	0.07	0.21	0.17
LW8	0.25	0.27	0.10	0.29	0.09	0.25	0.21
EWT	0.23	0.21	0.05	0.18	0.04	0.13	0.14
LFW	0.09	0.11	0.06	0.02	0.00	0.04	0.00
FW12	0.25	0.36	0.12	0.23	0.02	0.19	0.13
EFW	0.10	0.03	0.10	n.v.[3]	0.00	0.12	0.06
NLB	0.27	0.23	0.11	0.14	0.03	0.11	0.10
SURV	0.03	0.04	0.19	0.03	0.01	0.17	0.07
SURVm	0.23	0.14	0.12	0.07	0.00	0.07	0.18
DAG3	0.16	0.29	0.09	0.18	0.00	0.15	0.16
DAG8	0.19	0.18	0.04	0.11	0.02	0.13	0.05
FEC1	0.21	0.40	0.10	0.32	0.02	0.20	0.23
FEC2	0.25	0.35	0.08	0.32	0.02	0.11	0.25
AFEC	0.09	0.09	0.07	0.10	0.01	0.00	0.08

[1]Rom: Romney; Coop: Coopworth; Peren: Perendale; RCP: CompRCP; RCPT: CompRCPT; CRP: CompCRP.

[2]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

[3]n.v.: no validation set

8.4.1.4 Slopes or variance adjustments

The variance of the MBVs were not consistent with what is expected ($r_A^2 \sigma_u^2$) for the BV with the accuracies as calculated above. To correct for this, the MBVs are multiplied by breed by;

$$k = \frac{\sigma_u r_A}{sd(MBV)} \qquad \text{Equation 7}$$

Where *sd(MBV)* is the standard deviation of the MBVs across breed. The k values are an estimate of the consistency between the variance of the MBVs predicted and what is expected given the accuracy estimates obtained. When k=1, the predicted and expected values are consistent. Values greater (or less) than 1 indicate the predictions have too little (or too much) spread i.e. MBVs are conservative (non-conservative) compared to what is expected. Table 8.9 shows the weighted k values for each breed and trait.

Table 8.9: The k values for traits in the 7 validation breeds[1].

Trait[2]	Rom	Coop	Peren	RCP	Texel	RCPT	CRP
WWT	0.48	0.56	0.10	0.64	0.51	0.74	0.65
WWTm	0.18	0.36	0.44	0.80	0.39	0.05	0.71
CW	0.64	0.75	0.31	0.96	0.66	0.73	0.83
LW8	0.65	0.77	0.35	0.95	0.71	0.79	0.89
EWT	0.85	0.87	0.18	0.87	0.52	0.61	0.97
LFW	0.46	0.90	0.54	0.09	0.02	0.46	-0.87
FW12	0.59	0.82	0.38	0.62	0.05	0.54	0.43
EFW	0.56	0.20	0.92	n.v.[3]	-0.35	0.80	0.69
NLB	0.68	0.73	0.58	0.39	0.51	0.47	0.51
SURV	-0.14	-0.11	0.68	-0.08	-0.06	0.67	0.33
SURVm	0.79	0.45	0.47	0.26	-0.41	0.33	0.87
DAG3	0.87	1.06	1.00	0.88	-0.27	0.81	1.03
DAG8	0.87	1.05	0.35	0.69	0.75	1.20	0.46
FEC1	0.52	1.02	0.45	0.99	0.00	0.58	0.69
FEC2	0.74	0.95	0.44	1.31	0.36	0.47	1.15
AFEC	0.62	0.58	0.68	0.92	0.55	-0.53	0.80

[1]Rom: Romney; Coop: Coopworth; Peren: Perendale; RCP: CompRCP; RCPT: CompRCPT; CRP: CompCRP.

[2]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

[3]n.v.: no validation set

8.4.1.5 SNP P values from G1 GS analysis

Figures 8.2 to 8.5 show the Manhattan plots of the resulting $-\log_{10}(P)$ values for the traits: WWT, FW12, DAG3 and DAG8 (other traits in Appendix 20-31). For WWT, there were 89 SNPs below the nominal threshold of $P<0.001$, and 3 SNPs below $P< 1x10^{-5}$. For FW12 there were 103 SNPs below the nominal threshold and 3 below $P< 1x10^{-5}$. For DAG3 and DAG8 there were 54 and 51 SNPs, respectively, below nominal significance and 1 for each trait that reached the Bonferoni $P< 1x10^{-6}$ threshold. The best 5 SNPs for these traits are shown in Table 8.10, (other traits in Appendix 32)

Table 8.10: The best 5 SNPs for the traits: weaning weight (WWT), fleece weight at 12 months (FW12), dag score at 3 and 8 months (DAG3, DAG8), with chromosome (Chr), Chr position (base pairs, bp), and the *P* value and -log$_{10}$(*P*) value.

Trait	Marker name	Chr	Position (bp)	*P* value	-log$_{10}$(*P*)
WWT	s33509	13	59530141	1.85E-06	5.733
	OAR6_42755193	6	38492428	7.33E-06	5.135
	s51888	1	228974087	8.69E-06	5.061
	OAR4_115321021	4	107332365	1.09E-05	4.963
	OAR8_80973749	8	75443214	1.27E-05	4.897
FW12	OAR3_107450633	3	100836677	3.68E-06	5.434
	s14624	20	45761858	3.94E-06	5.404
	s14162	8	23388014	4.50E-06	5.347
	s71134	NA	NA	2.24E-05	4.651
	s74857	NA	NA	3.04E-05	4.518
DAG3	s22390	15	40475174	5.04E-06	5.297
	OAR6_90538374	6	83053269	2.73E-05	4.563
	OAR1_166405159	1	154566172	6.63E-05	4.178
	OAR6_104194705	6	95182020	0.0002	3.699
	s51141	3	104897268	0.0002	3.640
DAG8	s22390	15	40475174	2.72E-10	9.565
	OAR8_75725850_X	8	70914235	1.07E-05	4.970
	s35981	5	1871947	3.61E-05	4.442
	s73301	3	138345356	4.48E-05	4.349
	OAR8_76325701	8	71496359	6.17E-05	4.210

The traits that were in both the GWAS using SIL BV's (Chapter 7) and in this analysis were compared for similarities in resulting -log$_{10}$(*P*) values (Table 8.11, Figures 8.6). The correlations (r) ranged from 0.32 to 0.46, and the number of SNPs that were in the most extreme 20 for both analyses ranged between 2 and 6. For DAG8, the best SNP was the same for both analyses, and the third best SNP from the GWAS analyses was the second best SNP in this GS analysis. The -log$_{10}$(*P*) values for the DAG3 and DAG8 GS analyses were also compared. The correlation (r) between these 2 traits was 0.52; within the most extreme 20 SNPs, 3 were the same including 1 that was the most extreme SNP (s22390) for both traits (Figure 8.7).

Table 8.11: The correlation (r), number of SNPs in top 20 for both analyses (Top 20) and summary of the -$\log_{10}(P)$ values from the 2 analyses (GWAS and GS).

Trait[1]	R	Top 20	Min[2]	1st Qu	Median	Mean	3rd Qu	Max	Analysis
WWT	0.34	3	0.000001	0.09	0.22	0.30	0.42	3.25	GWAS
			0	0.13	0.31	0.45	0.61	5.73	GS
WWTm	0.32	2	0.000008	0.10	0.23	0.33	0.46	3.32	GWAS
			0	0.12	0.30	0.44	0.61	5.52	GS
LW8	0.42	2	0.000001	0.05	0.10	0.13	0.19	1.60	GWAS
			0	0.13	0.31	0.45	0.61	7.52	GS
FW12	0.46	6	0	0.07	0.15	0.21	0.29	2.47	GWAS
			0	0.13	0.30	0.45	0.61	5.43	GS
DAG3	0.46	2	0.000005	0.08	0.19	0.26	0.37	3.22	GWAS
			0	0.13	0.30	0.44	0.60	5.30	GS
DAG8	0.45	4	0.000018	0.09	0.22	0.31	0.43	4.28	GWAS
			0	0.12	0.30	0.43	0.60	9.57	GS

[1]WWT: weaning weight; WWTm: WWT maternal; LW8: live weight at 8 months; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 and 8 months.

[2] Min: minimum; 1st Qu: first quantile; 3rd Qu: third quantile; Max: maximum.

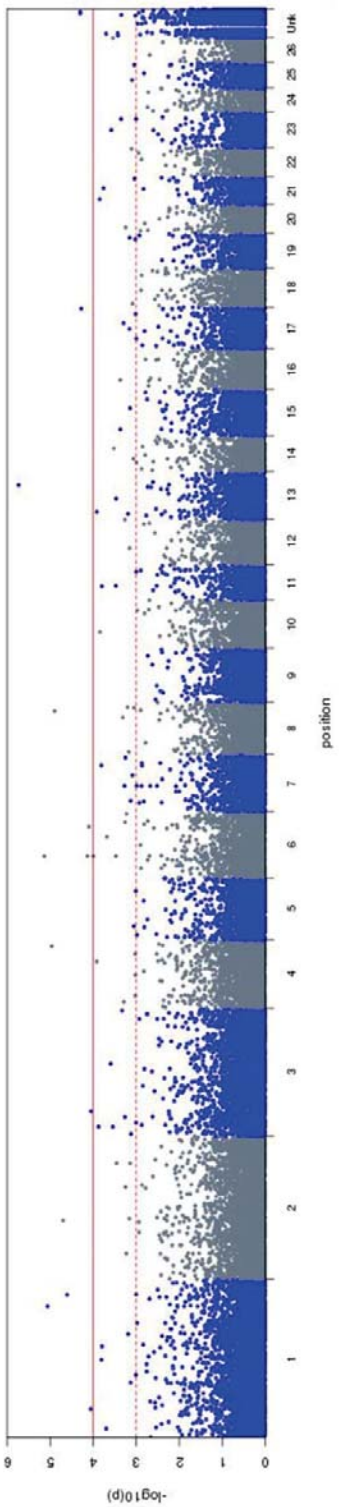


Figure 8.2: Manhattan plot of -log$_{10}$(*P*) values of SNPs for weaning weight. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).

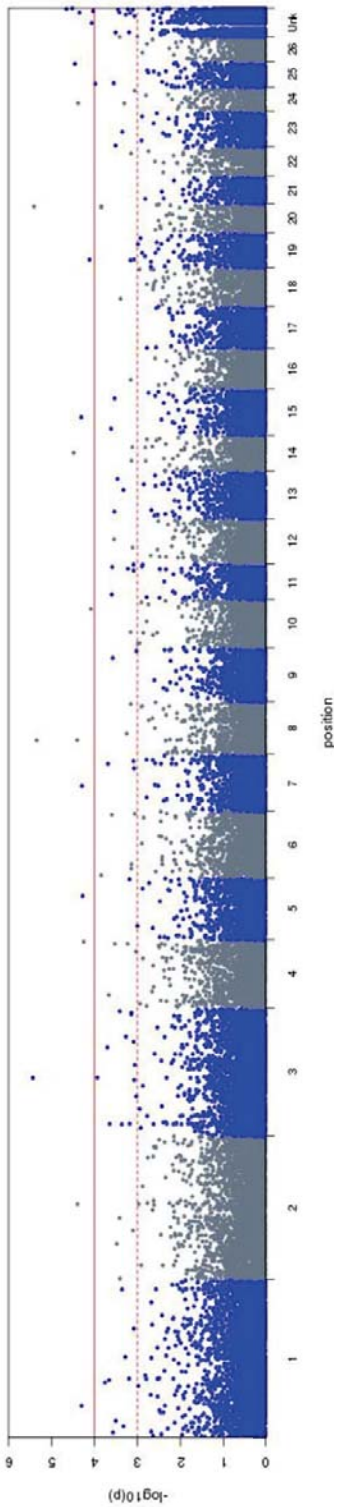


Figure 8.3: Manhattan plot of -log$_{10}$(*P*) values of SNPs for fleece weight at 12 months. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).

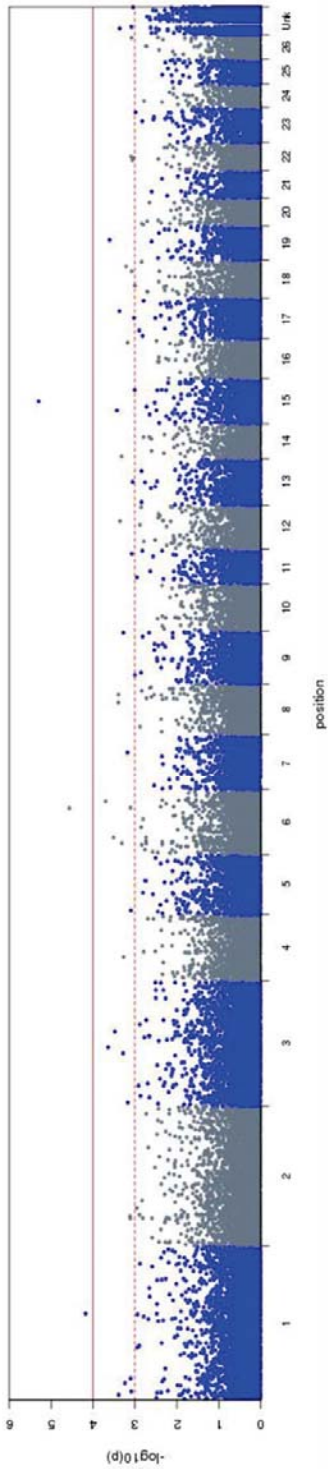


Figure 8.4: Manhattan plot of -log$_{10}$(*P*) values of SNPs for dag score at 3 months. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).

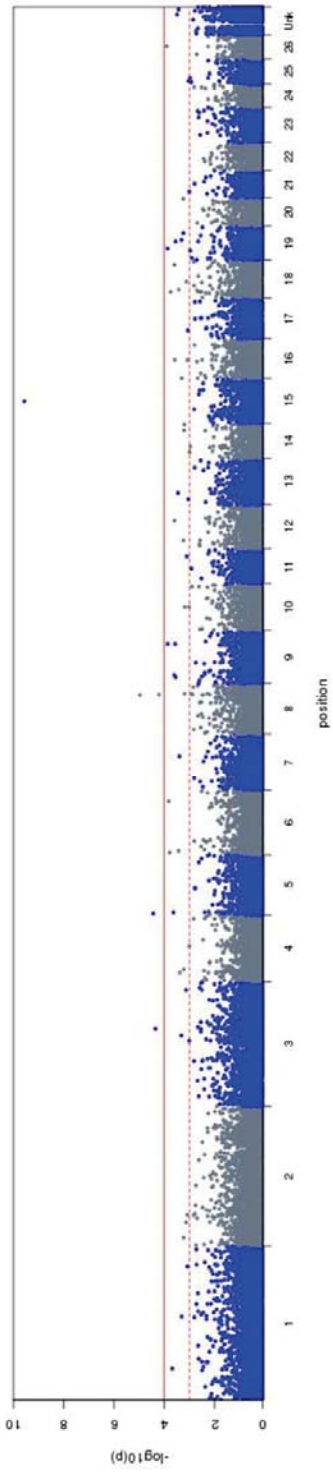


Figure 8.5: Manhattan plot of -log$_{10}$(*P*) values of SNPs for dag score at 8 months. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).

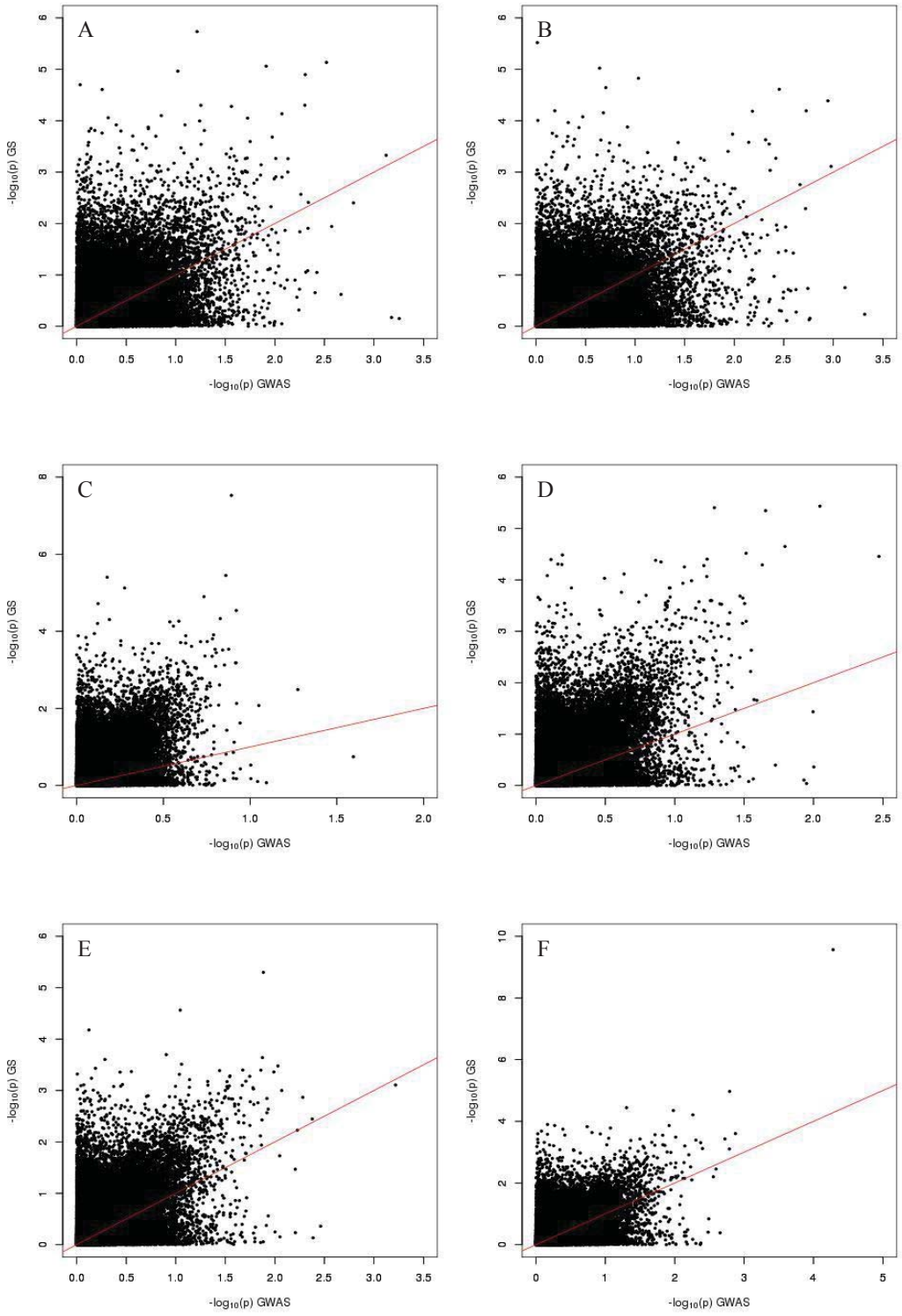


Figure 8.6: Comparison of GWAS and GS -log$_{10}$(*P*) values for weaning weight, (A), maternal weaning weight (B), live weight at 8 months (C), fleece weight at 12 months (D), dag score at 3 (E) and 8 (F) months. The 0-1 line is plotted in red.

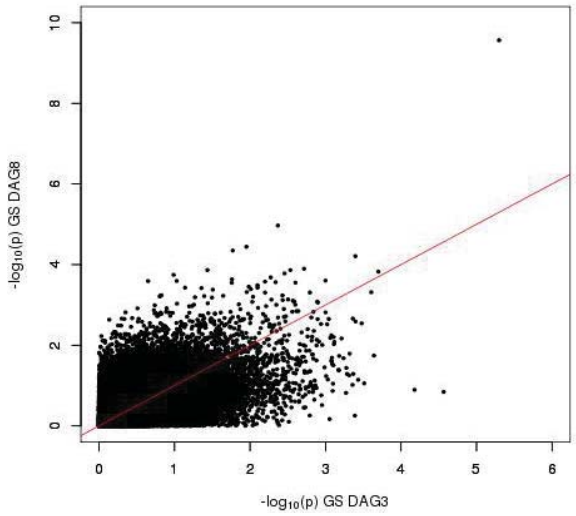


Figure 8.7: Comparison of dag score at 3 and 8 months (DAG3, DAG8) -$\log_{10}(P)$ values from GS analysis. The 0-1 line is plotted in red line.

Table 8.12: Estimate of lambda (slope), and their standard error (s.e.) of the linear regression of the observed -$\log_{10}(P)$ on the expected -$\log_{10}(P)$ (QQ plot) for each trait.

Trait	lambda	s.e.
Weaning weight	1.055	0.00018
Weaning weight maternal	1.027	0.00012
Carcass weight	1.050	0.00017
Live weight at 8 months	1.059	0.00022
Adult ewe weght	1.035	0.00010
Lamb fleece weight	1.019	0.00019
Fleece weight at 12 months	1.059	0.00021
Adult fleece weight	1.020	0.00012
Number of lambs born	1.014	0.00005
Survival	1.026	0.00013
Survival maternal	1.005	0.00003
Dag score at 3 months	1.001	0.00007
Dag score at 8 months	1.006	0.00025
Faecal egg count summer	1.024	0.00034
Faecal egg count autumn	1.017	0.00017
Adult faecal egg count	1.011	0.00014

8.4.1.6 QQ plots

The observed distribution of -$\log_{10}(P)$ values were plotted against the expected distribution (exponential) under the null hypothesis of no genetic association and no LD

between SNPs. The QQ plots for WWT, FW12, DAG3 and DAG8 are shown in Figure 8.8 the remaining traits are in Appendix 33. The slopes (lambda) for all traits are close to 1 (Table 8.12). The right-hand tail deviates above the slope for the traits; DAG3 (Figure 8.12, D), CW, LW8, EFW, FEC1, FEC2 and AFEC (Appendix 33). These can be considered as SNPs associated with these trait.

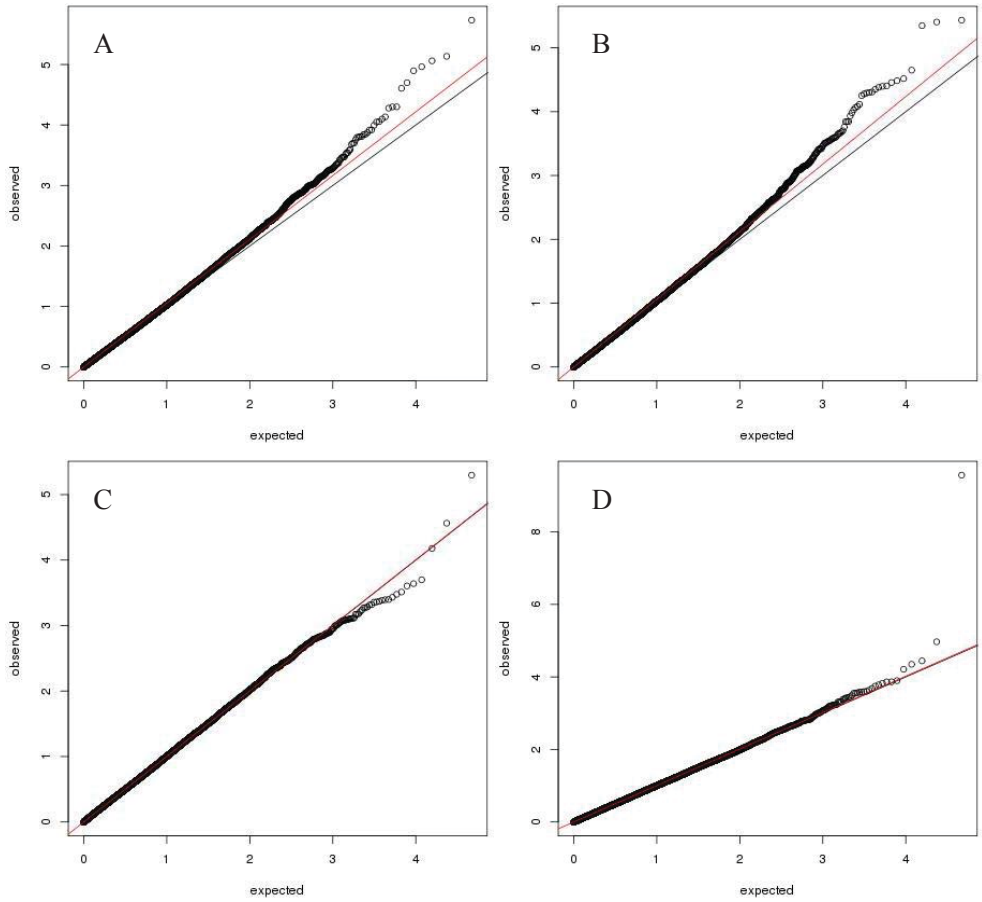


Figure 8.8: QQ plot for weaning weight (A), fleece weight at 12 months (B), dag score at 3 months (C) and 8 months (D) $-\log_{10}(P)$ values. The 0-1 line is in black and the slope in red.

8.4.2 Comparison of genetic gain

8.4.2.1 Effective progeny

The effective numbers of progeny were calculated with the r_I or the r_C accuracies, using the single trait selection worksheet (Table 8.13). For traits that have low

heritability, the effective number of progeny equal to the accuracy of the SNP chip is large, e.g. SURV and SURVm. For traits that are easy to measure and have moderate heritabilities, the SNP chip is only equal to 1 or 2 effective progeny e.g. AFW and EWT.

Table 8.13: The heritability (h^2), repeatability (rep) and genomic selection accuracy (GS acc), effective number of progeny (n prog) and their single trait selection accuracy (Seln acc) estimated from the breed combined-accuracy (r_C) and the individual accuracy (r_I).

			r_C			r_I		
Trait[1]	h2	rep	GS acc	Seln acc	n prog	GS acc	Seln acc	n prog
WWT	0.14	0.14	0.477	0.474	8	0.603	0.606	16
WWTm	0.20	0.20	0.343	0.369	3	0.551	0.544	8
CW	0.30	0.30	0.476	0.495	4	0.583	0.572	6
LW8	0.35	0.35	0.497	0.473	3	0.622	0.634	7
EWT	0.45	0.45	0.476	0.450	2	0.544	0.525	3
LFW	0.15	0.15	0.292	0.269	2	0.351	0.367	4
FW12	0.37	0.37	0.495	0.484	3	0.586	0.581	5
AFW	0.45	0.45	0.315	0.335	1	0.391	0.45	2
NLB	0.09	0.12	0.517	0.519	16	0.496	0.494	14
SURV	0.01	0.01	0.163	0.164	11	0.465	0.465	110
SURVm	0.008	0.008	0.475	0.475	145	0.359	0.359	74
DAG3	0.34	0.34	0.400	0.396	2	0.456	0.467	3
DAG8	0.31	0.31	0.436	0.449	3	0.472	0.502	4
FEC1	0.18	0.62	0.458	0.469	6	0.512	0.523	8
FEC2	0.19	0.57	0.501	0.509	7	0.512	0.509	7
AFEC	0.25	0.28	0.308	0.343	2	0.346	0.343	2

[1]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

8.4.2.2 Selection response

The selection response per 'selection round' for each trait and overall for each selection scenario, for Romneys, is shown in Table 8.14. Scenario 1-B, the most likely scenario representing a farmer's decision without SNP Chip results, where selection was made on a 2 year old ram with records available from itself, its dam, sire and half

siblings, has a standard deviation of index response of $2.22 and an accuracy of 0.34. Scenario 2, where selection was made on a ram hogget with records available from itself, its dam, sire and half siblings and has SNP chip results included as effective progeny numbers calculated from r_I (2-A) or r_C (2-B), has an increase in selection response per round from scenario 1-B of 64% and 50% respectively. Converted to response per year or the rate of genetic gain for each scenario is $1.43, $2.87 and $2.63 respectively, i.e. 100% and 84% more in scenarios 2-A and 2-B, respectively compared to 1-B (Table 8.15). For subsequent discussion the last estimate is used.

The majority of gain was seen in lowly heritable traits, or sex limited traits measured late in life, specifically NLB, SURV, SURVm and EWT and with more modest changes for FEC. The effect of SNP Chips on dagginess was minimal, the response per selection round in units for DAG3 and DAG8 when selection was only on measurements was -0.11 and -0.09 respectively, in dollars this was $0.04 and $0.03 respectively, regardless of 2 year old or hogget. The use of SNP Chip information only slightly improves the response, by 0.03 and 0.02 units, respectively, or $0.01 for both traits, regardless if r_I or r_C are used. This is because dag score is a moderately heritable, easy to measure trait and the SNP Chip does not add as much information compared with a lowly heritable hard to measure trait like NLB or SURVm. However, if there are no measurements and SNP Chip information is unavailable for DAG3 or DAG8, then the response relies on correlations with other traits, the response per selection round for a 2 year old ram is -0.02 and -0.01 units for DAG3 and DAG8, respectively. With addition of SNP Chip information, but no phenotypes, the response per selection round is -0.05 and -0.04 respectively regardless if r_I or r_A are used. There was also an improvement in the accuracy from 0.12 and 0.08 as a 2 year old and no information to 0.51 and 0.52 for SNP chip information only, respectively, for DAG3 and DAG8. Thus, SNP Chip data is better than no information at all.

Table 8.14: Response per selection round per trait (change (Δ) in units and dollars ($)) and accuracies[1], and overall index (standard deviation (SD) index in dollars) and accuracy for a dual-purpose Romney index.

	Scenario 1-A: Ram Hogget[2]				Scenario 1-B: 2 year old Ram[2]				Scenario 2-A: Indiv acc[3]				Scenario 2-B: Average acc[3]			
Trait[4]	Δ units	Δ $	MT-EBV	ST-EBV	Δ units	Δ $	MT-EBV	ST-EBV	Δ units	Δ $	MT-EBV	ST-EBV	Δ units	Δ $	MT-EBV	ST-EBV
WWT	0.49	0.46	0.61	0.58	0.48	0.45	0.61	0.58	0.43	0.41	0.75	0.73	0.42	0.39	0.70	0.67
WWTm	0.25	0.21	0.61	0.23	0.26	0.22	0.61	0.23	0.38	0.32	0.73	0.59	0.32	0.27	0.68	0.44
CW	0.20	0.53	0.57	0.00	0.18	0.46	0.58	0.00	0.13	0.33	0.71	0.55	0.12	0.32	0.67	0.48
LW8	0.56	0.00	0.74	0.71	0.54	0.00	0.74	0.71	0.46	0.00	0.81	0.79	0.45	0.00	0.78	0.75
EWT	0.12	-0.13	0.57	0.33	0.04	-0.05	0.58	0.33	-0.52	0.54	0.71	0.58	-0.30	0.31	0.66	0.52
LFW	0.01	0.02	0.28	0.00	0.01	0.03	0.41	0.00	0.01	0.02	0.47	0.35	0.01	0.02	0.40	0.25
FW12	0.02	0.01	0.43	0.40	0.04	0.03	0.70	0.69	0.02	0.02	0.65	0.62	0.02	0.01	0.59	0.56
AFW	0.06	0.13	0.36	0.00	0.08	0.18	0.52	0.00	0.07	0.16	0.58	0.45	0.06	0.14	0.51	0.34
NLB	0.01	0.20	0.21	0.20	0.01	0.18	0.22	0.20	0.04	0.64	0.52	0.51	0.05	0.75	0.54	0.53
SURV	0.00	0.21	0.27	0.27	0.00	0.20	0.27	0.27	0.01	0.45	0.51	0.51	0.00	0.18	0.31	0.31
SURVm	0.00	0.01	0.06	0.06	0.00	0.01	0.06	0.06	0.00	0.15	0.36	0.36	0.00	0.28	0.47	0.47
DAG3	-0.11	0.04	0.72	0.70	-0.11	0.04	0.72	0.70	-0.08	0.03	0.76	0.74	-0.08	0.03	0.76	0.74
DAG8	-0.09	0.03	0.71	0.68	-0.09	0.03	0.71	0.68	-0.07	0.02	0.76	0.74	-0.07	0.02	0.76	0.74
FEC1	-0.07	0.20	0.56	0.56	-0.06	0.19	0.56	0.56	-0.07	0.21	0.70	0.67	-0.07	0.22	0.68	0.65
FEC2	-0.04	0.13	0.43	0.00	-0.04	0.12	0.44	0.00	-0.06	0.17	0.64	0.50	-0.06	0.19	0.63	0.50
AFEC	-0.07	0.13	0.40	0.00	-0.07	0.13	0.40	0.00	-0.08	0.17	0.57	0.34	-0.09	0.17	0.57	0.34
Response		2.18				2.22				3.64				3.32		
Accuracy		0.33				0.34				0.56				0.51		

[1] accuracy for a multi trait- or single trait- estimated breeding value (MT-EBV, ST-EBV)

[2] Scenario 1: selected on individual, dam, sire and half sibling records as a ram hogget or 2 year old ram;

[3]Scenario 2: a hogget ram selected using individual, dam, sire, half sibling records as well as SNP Chip results calculated as number of effective progeny based on individual (Indiv acc) or average (Average acc) accuracies from genomic selection.

[4]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

Table 8.15: Response per selection round (SD index) and accuracy (acc), conversion factor (i/L, selection intensity / generation interval), and rate of genetic gain (ΔG), for each scenario.

	SD index ($)	acc	i/L	ΔG ($)
Scenario 1				
A: Ram Hogget	2.18	0.33	0.79	1.72
B: 2 yr old Ram	2.22	0.34	0.64	1.43
Scenario 2[1]				
A: r_I	3.64	0.56	0.79	2.87
B: r_C	3.32	0.51	0.79	2.63

[1]weighted average individual accuracy (r_I) and the breed combined-accuracy (r_C)

8.5 Discussion

8.5.1 Genomic selection

There are many ways to estimate parameters for the MBV calculation component of GS. These include GBLUP, BayesA, BayesB, BayesC, SSVS, PLSR, PCR and LASSO (González-Recio et al., 2010; Habier et al., 2011; Meuwissen et al., 2001; Moser et al., 2007; Solberg et al., 2009a; Verbyla et al., 2009). For the current work, GBLUP was used for the following reasons: computational efficiency, ability to use weighted analyses, and there were small differences between methods in actual results (B. Auvray, pers. comm.). However, this may not be the case in all situations nor for all traits. In the current context, the primary limitations to the study were the number of animals genotyped and the density of genotypes (and therefore average LD between genotyped SNPs and the causative QTL variants). In practice, the 50K Ovine SNP Chip (having approximately 60kb spacing), has low LD, depending on breed. For example, New Zealand Texel has LD (r^2) between SNPs of ~0.2, and New Zealand Romney of ~0.17 (Raadsma, 2010). While there is not a major ascertainment breed bias (Kijas et al., 2012a) in the SNPs used, there is a major bias in MAF, because only high MAF SNPs were selected. Combined, these effects lead to SNPs with poor LD with lower

frequency causative QTL. It was found that GBLUP methods can perform better in these circumstances (Lorenz et al., 2011).

8.5.1.1 Quality assurance and population structure

New Zealand sheep breeds are predominantly of European descent. Kjias et al. (2009) showed that the New Zealand Romney and Texel fitted well within the breeds from European origin. The Romney, Coopworth and Texel seem to make the three corners of this New Zealand sheep breed industry dataset. Coopworth and Perendale were generated from crossing Romney with Border Leicester and Cheviot respectively. Therefore, Border Leicester may sit above Coopworth and Cheviot may fit in between Texel and Perendale. There is a large component of composites that fit in the middle, which enables moderately accurate prediction for composite breeds not in the training dataset. The principal component plots were generated from the stated breed proportions in SIL. There were a few animals which defy their stated breed and sit within the boundaries of other breeds. This may be due to parentage errors; there were 9% incorrect parentage detected in this dataset (mainly paternity only tested) (Ken Dodds pers. comm.), and a portion of these may be due to incorrect labelling of samples. This did not affect analysis as relationships and PCs were based on genotype data. Breed boundaries for selecting training and validation sets were based on the recorded breed, which may affect accuracies and k values, if wrong. However, those animals with a large difference between reported breed and breed estimated from genotypes, i.e. were classed as an entirely different breed to its reported breed type, were removed from the analysis.

8.5.1.2 Estimates of predictive power versus theory

Using equation 8 from Goddard (2009), we can calculate the expected accuracy from our population using effective population size (Ne) of 405 (NZ Romney, Table S4, Kijas et al., 2012a), number of records available per trait, and genome length of 30M. In general, the accuracy estimates are higher than those calculated theoretically for Romney and Coopworth (Figure 8.9). The Perendales have a flatter distribution, with points below the line of equality possibly due to the low number of Perendales genotyped with trait information. The reason for the higher accuracies is that the theoretical line is for ‘unrelated’ animals i.e. ~10 or more generations distant. In practice, most of the validation animals have an ancestor 1-3 generations distant in the

training data set and as such will have higher estimated accuracies as expected from theory.

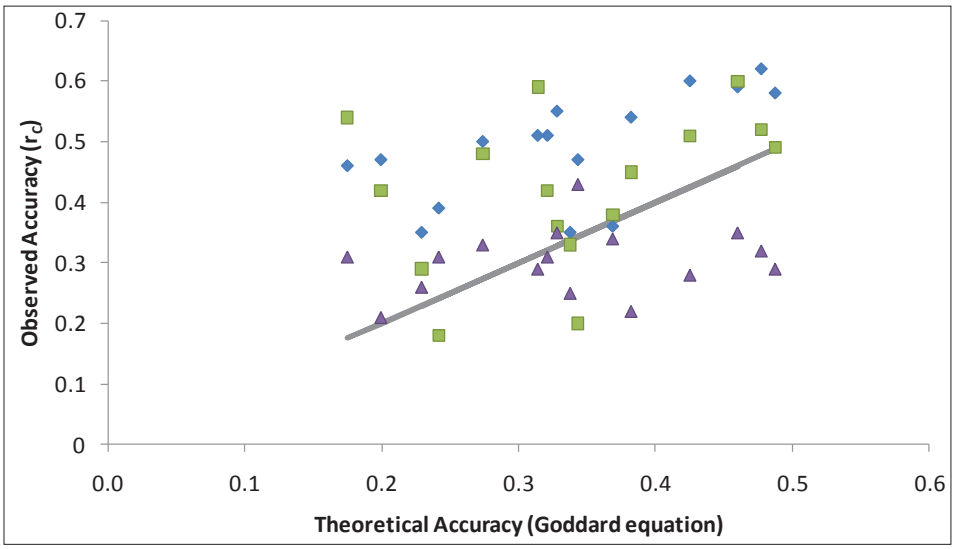


Figure 8.9: Comparison of the observed breed combined-accuracies (r_C) versus the theoretical accuracy using Goddard's equation for each trait, and for the breeds Romney (green), Coopworth (blue) and Perendale (purple). The line of equality is in grey.

Solberg et al. (2009b) estimated the effect of moving away from the training population, with different marker densities. As the validation population moves away from the initial training population, accuracy decreases. However, as marker density increases the effect is reduced. Thus, the training population has to be updated regularly with newer generations so that the validation animals do not become too far removed from the training set.

The number of animals needed to reach an accuracy of 0.70 using the current SNP chip can be calculated using the effective heritability of the trait. Table 8.16 shows how many animals need to be genotyped to reach an accuracy ~0.70 if the animal has only its own measure, or if it has 126 progeny measured for the trait. The 126 progeny was taken from table 8.2 as the average number of progeny per sire for a 629 ewe flock with a mating ratio of 1:90. For the majority of traits, the smallest number of animals that can be genotyped to get an accuracy of ~0.70, is 16,000 with at least 126 progeny records. For the lowly heritable traits such as NLB, SURV and SURVm, firstly, it is unlikely that 126 measures on progeny would be available, especially for NLB and SURVm

which are measured on females only. Secondly, greater than 1 million animals are needed for the survival traits and ~128,000 animals needed for NLB, if only their own measure is available. This can be decreased to 16,000 for NLB and 48-64,000 for the survival traits if 126 progeny are recorded. Another way to increase accuracy would be to use a higher density SNP chip, increasing the LD between markers (Meuwissen et al., 2001). Meuwissen et al. (2001) showed increasing the marker spacing from 1 SNP every 1 cM to 1 every 2 cM or 4 cM in an effective population size of 100 decreased accuracy estimated by BLUP from 0.732 to 0.708 for 1 every 2 cM and to 0.668 for 1 every 4 cM.

Table 8.16: Number of animals genotyped to reach accuracy of ~0.70 if have own measurement or 126 progeny measurements.

	Own measure			126 progeny measure		
Trait[1]	Eff h^2	N	r	Eff h^2	N	r
WWT	0.14	80,000	0.71	0.82	16,000	0.73
WWTm	0.20	64,000	0.73	0.87	16,000	0.74
CW	0.30	48,000	0.75	0.91	16,000	0.75
LW8	0.35	32,000	0.71	0.92	16,000	0.75
EWT	0.45	32,000	0.75	0.94	16,000	0.75
LFW	0.15	80,000	0.72	0.83	16,000	0.73
FW12	0.37	32,000	0.72	0.93	16,000	0.74
AFW	0.45	32,000	0.75	0.94	16,000	0.75
NLB	0.09	128,000	0.71	0.74	16,000	0.72
SURV	0.01	1,104,000	0.71	0.24	48,000	0.71
SURVm	0.008	1,376,000	0.71	0.20	64,000	0.73
DAG3	0.34	32,000	0.70	0.92	16,000	0.75
DAG8	0.31	48,000	0.75	0.91	16,000	0.75
FEC1	0.18	64,000	0.71	0.86	16,000	0.74
FEC2	0.19	64,000	0.72	0.86	16,000	0.74
AFEC	0.25	48,000	0.72	0.89	16,000	0.75

[1]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count;
[2]Eff h^2: effective heritability;
[3]r: accuracy.

8.5.1.3 Ascertainment biases

One issue that has not been fully explored in these analyses is the effect of pre-selection in the animals genotyped. The following identifies some likely situations in the current datasets and their possible effects.

8.5.1.3.1 Sampling of ewes/extremes

The ewes that were genotyped in the current dataset, making up approximately 21% of the animals, consisted of 3 broad groups. The first group were ewes born in recent years and primarily consisting of randomly selected animals measured as part of a green house gas genetic parameter estimation project (Pinares-Patiño et al., 2013). These animals would primarily be part of the validation datasets and as such will provide unbiased data. The second group consisted of older ewes with many recorded lambings, typically 3 or more. These animals would likely have high breeding index values relative to their contemporaries, especially for the live weight traits. The final group was a small number of ewes born in 2007, primarily Coopworths and Coopworth crosses, which were selected based on their extreme breeding values for FEC. In this case, any bias is likely to be restricted to that trait (as it is poorly correlated with other production traits) and breed and is likely to cause an over prediction of the MBV accuracy.

8.5.1.3.2 Culling of rams used

For traits where the animals had been measured prior to selection, it was expected that some ascertainment bias would be present. The extreme would be SURV because all animals measured survived, and perhaps not surprisingly, this was the measurement that showed maximum difference between the accuracies estimated by r_I and r_A (Table 8.6, Figure 8.2). Other traits also measured before selection include early live weights and traits that are associated with early live weights e.g. CW.

8.5.1.3.3 Breeds

The current analysis combined the training animals into one population to estimate the SNP coefficients used in the validation set. The training set used animals of Romney, Coopworth, Perendale and CompRCP breeds. No animals of Texel CompRCPT or CompCRP breed were in the training set. Simulations showed that when there are limited numbers of animals from one population set, then the most accurate genomic predictions are generated when information from all populations are combined

in the training set rather than predicting separately by population (de Roos et al., 2009). Across breed predictions are more preferable when breeds are closely related, the SNP marker density is high, and there are small numbers in the training set (de Roos et al., 2009; Ibánẽz-Escriche et al., 2009). The more genetically diverse the populations are, the less accurate are the genomic predictions for across breed analysis. As the degree of LD between SNP and QTL decreases, the predictions cannot be preserved across populations. Thus higher density SNP chips are required if predictions across diverse populations are required. This has been agreed by Pryce et al. (2011) and Hayes et al. (2009a) in dairy cattle, where combining populations into 1 training set was more accurate than separating training sets by breed. However, if the population in the validation set was not present in the training set, then the accuracy for that population was significantly decreased.

This corroborates the low accuracy for Texels in this analysis, as they are the most divergent breed in the validation set compared to the training set. The Texels originated from Texel Island off Holland, while the Romneys were from England (www.texel.co.uk, www.romneysheepuk.com), their estimated divergence is 160 to 240 generations ago (Figure S10, Kijas et al., 2012a). To increase the accuracies for Texels, more animals are required so that some may be combined in the training set. This may in part be achieved by increasing the number of composites with at least 50% Texel, if pure-breds are hard to collect.

The accuracies for CompRCPT and CompCRP are perhaps higher than would be expected for a breed not present in the training set, however, they are at least 50% and at least 30-50%, respectively, of the breeds that are represented in the training set. This result is similar to Ibánẽz-Escriche et al. (2009), except that they were predicting pure-breds from a cross-bred training population.

8.5.1.4 Genomic selection versus GWAS P values comparison.

The $-\log_{10}(P)$ values were much higher for the GS results than for the GWAS results from Chapter 7. The reason for this is unclear, both methods adjusted for population structure (mainly breed) and both were weighted by the accuracy of the EBVs used. The difference between the two methods is that the *P* values for GS were calculated from the SNP coefficients and SNP effects were estimated simultaneously, while the GWAS *P* values come from fitting the SNPs in a linear model sequentially against the residuals from the GBLUP for each animal. The GS method may have removed the correlation

between tests, because all SNP effects are estimated together, while for the GWAS as the tests are sequential, the correlations remain. Therefore, there is difference in the slopes of the QQ plots, where the slopes are close to 1 for the GS analysis, but are less than 1 for the GWAS (Chapter 7).

The moderate positive correlation estimates from comparing the estimated $-\log_{10}(P)$ values from the GS analysis and the GWAS for DAG8 (Chapter 7) was reassuring. This outcome was most likely due to two reasons; firstly, the analyses used the same phenotypes and genotypes, only differing in analysis methods. Secondly, the same extreme SNP was identified in both analyses, i.e. s22390 (Chr15:40475174bp). As described in Chapter 7, there was no gene directly under this SNP on Ovine genome v2. There was, however, a GLEAN predicted gene, and two genes upstream and downstream of the SNP within 100kbp (MICAL2 and DKK3). In summary, neither of these genes were good candidates for dag formation. The sequence of the predicted gene was blasted against the human and bovine genomes. No known homologous annotated genes were identified by this process. Further investigation of this region is required, as well as validation in other datasets.

For the other traits, WWT, WWTm, LW8 and FW12, there were a few SNPs that were present in the best 20 SNPs from each analysis. The correlations of *P* values between analysis methods for LW8 and FW12 are similar to those obtained for DAG3 and DAG8, however the correlations for WWT and WWTm are slightly lower and there is no clear peak near the 0,1 line in the graphs. This may be an effect of the small differences between the two methods.

8.5.2 Predicted genetic gain

8.5.2.1 Approach

To estimate the impact of using the MBVs in the New Zealand sheep industry, it was most convenient to use selection theory and effective progeny numbers. The current method of estimating effective progeny numbers from the accuracies arising from the GS analysis may not be the most appropriate way of calculating the response. It is, however, a simple way of showing clearly the additional information arising from using the SNP Chip.

It was also assumed that the rams used are all of the same age, where, in fact the rams used will be a mixture of new untested rams, emerging rams used once before and

mature tested rams with complete progeny measurements. In this study, we are therefore testing the high-end extreme breeding program where the 'traditional' is using only 2 year olds and the SNP Chip is testing ram hoggets.

The most likely change arising from using SNP Chips, is the change from using 2 year olds to using ram hoggets in the elite breeding flocks. This will not produce as large a gain as seen in the dairy industry where the generation interval is significantly shortened by changing from use of progeny-tested bulls (5 years of age) to SNP Chip-tested bulls of a young age (~1-2 years of age) (Pryce and Daetwyler, 2012). Rates of genetic gain were estimated to range between 28 and 108% above that achieved by progeny testing (reviewed in Pryce and Daetwyler, 2012). However, the modest decrease in generation interval from 2 year olds to ram hoggets would produce significant genetic gains.

8.5.2.2 Selection response comparison

The dual-purpose selection index shown contains a number of traits; the overall response was an increase in rate of genetic gain of between 84-100% from the traditional approach of selecting a 2 year old on its own and its relative's measurements. Trait by trait the response varied and the majority of gain came from the sex-limited and later-in-life-measured traits such and EWT, NLB and SURVm. This selection objective also omits other disease and adult traits that would greatly benefit from GS. These traits include adult ewe longevity i.e. how long the ewe stays in the flock before she is culled or dies, and disease traits such as flystrike and facial eczema (FE), both of which are seasonally dependent. These traits are poorly recorded in practice and would greatly benefit from GS.

The current comparisons between existing selection and including SNP chip MBVs, do not take into account the cost of genotyping. In practice, to maximise discounted returns, two stage selection would be used and most likely only a proportion (10-20%) of ram lambs would be genotyped (Sise et al., 2011). In addition, costs would be reduced further by the use of lower density chips, such as the 5K SNP Chips used in Chapter 5, coupled with imputation. This would be expected to have a minimal impact on the estimated MBV accuracies. The work in Chapter 5 identified that best guess genotypes were 94 to 96% correct when the sire 50K genotype was available and included with an industry reference set. At the maximum this would result in a change in MBV of ± 4-6%. This was seen by Berry and Kearney (2011), who had an average

97% correlation between imputed SNP allele and real genotypes resulting, in an average 97% correlation between MBVs estimated using imputed or real genotypes (range was 92 to 99%).

A full exploration of 2 stage selection and imputation from low density genotyping along with optimisation of discounted returns and associated estimates of the reduction in the rate of genetic progress under these conditions has not been fully explored. The reason is that the focus was solely on the potential increase in genetic gain from the results of the current work and the costs of genotyping are dramatically reducing, making such estimates only applicable in the short term. For instance, new technologies such as genotyping by sequencing (Elshire et al., 2011) potentially makes it feasible that all animals could be genotyped cost effectively in the near future.

Other breeds and species implementing GS have identified that decreasing generation length, especially on the female side, greatly accelerates genetic gain when combined with GS (Schaeffer, 2006). No such investigation has been conducted here, but, the idea is not novel and was originally called 'velogenetics' (Georges and Massey, 1991). In the case of sheep, reducing the generation length to 1 year of age is feasible now in males, but typically cannot be reduced below an average of 2.5 years in females using natural mating. Normally, the female generation length is closer to 3.5 years (Blair and Garrick, 2007). This makes the overall average generation length, using ram lambs, of the order of 2.25 years. The potential of juvenile *in vitro* embryo transfer (JIVET) mating (Armstrong et al., 1997) to generate enough female embryos to genotype and subsequently implant only those that had the best GBVs, to satisfy an annual female replacement of the entire herd has great potential. This is especially the case when significant selection pressure on females can also be implemented. However, no such system coupled with GS has yet been demonstrated. This is an obvious extension to the current work and potentially has great synergy when combined with GS.

8.5.2.3 Economic weight for dagginess

The SNP chip does not add much benefit to decreasing dag score as it is already a moderately and easily measured trait. However, when examining these results it was identified that the response to selection in dollar terms as estimated by Byrne et al. (2012), did not fully reflect the benefit of decreasing dag score and thus flystrike risk.

Three factors require revising or inclusion into the equation. Firstly, the preventative treatment for flystrike is not included, as this animal husbandry technique is applied regardless of the average dag score of the flock. However, answers from the survey in Chapter 2, suggest there is variation in this practice, with not all farmers involved in the case-control study using preventative insecticide treatments. Secondly, in the current equation, ewes are assumed to be crutched twice a year. However, crutching occurs 2-3 times a year depending on breed and farm locality and is done typically at pre-shearing, before mating and crutched and bellied mid-winter (pre-lambing) (Scobie et al., 1997).

Lastly, there are replacement and reduced carcass weight costs due to flystrike. In a recorded flock where genetic gains are desirable, any animal that gets flystrike should be culled from the flock. In a commercial flock and also for the culled animals from the recorded flock, there is a cost in sending flystruck lambs of lower than average body condition to the abattoir early. The other option is to retain the animal until it reaches optimal body condition and weight to be sent to the abattoir, which also generates additional costs. Using the SIL recorded flocks involved in chapter 2, the flystruck lambs had an average growth rate from WWT to autumn live weight (LW6/LW8) of 101g/day (SD: 42, range -95g to 240g, n:228) and non flystruck lambs had an average growth rate of 118g/day (SD: 50, range -120g to 330g, n:5698); a difference of 17g/day ($P<1.4x10^{-6}$), compared to non flystruck lambs. This indicates that the reduced value of sending a flystruck lamb to the works at autumn live weight is \$8.98 (17g/day * 120 days * \$4.40 (Byrne et al., 2012)).

Modifications suggested are; the assumption of 2 crutchings a year for ewes is increased to 3, the reduced value of sending a flystrike lamb to the works at autumn live weight of \$8.98 and the cost of rearing a replacement ewe lamb from weaning to 2 years old (\$216.30 (Byrne et al., 2012)) be included. Therefore, the equations from Byrne (2012) should be (changes indicated in red):

$$DAG_L = C_T * (T_c - T_D) + CL_F * [(F_c - F_D)F_L] + CL_R * (R_c - R_D)$$

$$DAG_E = -3 * \left[\frac{(C5_T - C1_T) + (C_T * 8/1400)}{4}\right] + CE_F * [(F_C - F_D)E_L]$$

Where C_T, CL_F, CL_R and CE_F are: labour costs, cost of flystrike treatment per lamb, cost of crutching per lamb and cost of flystrike treatment per ewe. T, F, R are time cost for crutching, flystrike likelihood and number of extra crutching for a clean ($_C$) and dirty ($_D$) mob. Finally F_L and E_L are the flystrike costs due to loss in carcass weight for lambs

and replacement cost for ewes respectively. The economic weights for DAG3 (called lamb dag) and DAG8 (called adult dag) from Byrne et al. (2012) were estimated at $-0.34 and $-0.35 respectively. The new economic weights for DAG3 and DAG8 are $-0.62 and $-11.00 respectively.

Using these new estimates for the economic weights (not presented), the response to selection when selecting a 2 year old ram using only measurements increases to $-0.42 and $-0.52 per selection round. If the SNP Chip information was used, the response per selection round would be $-0.41 and -$0.52 for individual accuracies or $-0.42 and $-0.53, if the averaged accuracies are used to estimate the number of effective progeny.

Using the current figures, there is no benefit in using the SNP Chip to select for decreased dag score and flystrike incidence. The SNP Chip is very useful for increasing the rate of genetic gain for traits with low heritability and/or are sex limited and/or are hard to measure. This can particularly be seen for EWT and NLB, which had an average increase in response per selection of $0.51 each after use of the SNP Chip. Thus, if the SNP Chip is to be used and the genomic prediction equations for DAG3 and DAG8 are already estimated, there is no extra cost in getting this information and the increase in accuracy of the resulting GBVs is beneficial.

There is one final point. Dag score at 8 months is referred to as adult dag score in SIL, indicating it is a measure of adult dag score and not of autumn lamb dag score. Using an economic weight based on costs of ewe crutching and flystrike treatment may not be appropriate for a trait measured on a 6 to 8 month old lamb. Conversely, using a trait measured on a 6 to 8 month old lamb might not be an appropriate proxy for adult dag score. Currently, there appears to be no published estimates of heritability and genetic correlations with dagginess in adult sheep. This is a useful future study that should be undertaken.

8.6 Conclusions

Genomic selection was modelled on a number of dual-purpose sheep traits. The resulting breed accuracies were moderate for most breeds and traits. The breeds only in the validation set have lower accuracies and genomic predictions should not be used for these breeds until enough animals have been genotyped to enable a validation set to be included. Genomic selection has a major impact on traits that have low heritability

and/or are sex-limited or measured later in life, such as EWT, NLB, SURV and SURVm. For easily measured traits available on hogget rams, the MBVs do not greatly increase the selection response. Overall, using two-stage selection to select ram hoggets for mating can increase the rate of genetic gain of a dual-purpose index by 84%.

The GS *P* values estimated from the SNP coefficients identified the same best SNP for DAG8 as from the GWAS performed on the same dataset in Chapter 7. There was also a moderate correlation between the 2 dagginess traits for their GS *P* values. There were also moderate correlations between the 2 analyses for the remaining traits, with a few SNPs appearing in the best 20 for both analyses. The best SNP for DAG8 requires validation and further investigation.

A final recommendation is to include reduced live weight and replacement costs due to flystrike into the economic weight for dagginess, due to the high correlation between dag score and flystrike. This increases the response to selection for DAG3 and DAG8 by increasing the importance of this trait to farmers in a monetary value.

Chapter 9: Concluding discussion

9.1 Summary of findings

The objectives of this study were to:

1. Calculate the genetic parameters for flystrike in New Zealand dual-purpose sheep and genetic and phenotypic correlations with dagginess and other potential indirect traits such as wool and fibre traits (Chapter 2). Important outcomes were:
    - Heritability estimates for flystrike ranged between 0.08 ± 0.05 using Survival kit and 0.37 ± 0.10 on the observed scale.
    - Heritability estimates were similar to those reported for Australian and South African Merinos (Greeff and Karlsson, 2009; Scholtz et al., 2010; Smith et al., 2009), and slightly larger than an estimate in New Zealand Perendales (Brandsma and Blair, 1997).
    - Flystrike was highly genetically and phenotypically correlated with dag score, indicating dag score is a trait which could be useful as an indirect indicator against flystrike.
    - Of the wool and fibre traits assessed, FDCV had the most potential as an indicator trait as it was highly genetically correlated with flystrike, however, the phenotypic correlation between flystrike and FDCV was close to zero.

2. Investigate dagginess and its genetic relationship with dual-purpose production traits (Chapter 3) and other potential flystrike indicator traits; wool and fibre traits (Chapter 4). Important outcomes were:
    - Heritability estimates for dag score at 3 and 8 months (DAG3, DAG8) were moderate (0.31 to 0.44).
    - Selection against dagginess would not affect other production traits in an undesirable direction.
    - There were low genetic and phenotypic correlations with faecal egg count traits, indicating little association of dag score with the amount of parasitic eggs shed onto the pasture.
    - Dag score was also not associated with wool length, wool bulk, breech bareness or the fibre traits.

3. Determine the best parameters for imputation to provide accurate genotypes for subsequent GWAS, including a summary of DNA sample type and paternity parentage (Chapter 5). Important outcomes were:
    - Ear tissue collection systems provided adequate DNA quantity and quality for use on the Sequenom® and SNP Chip platforms.
    - Cervusv3.0 and Partial Pedigree had the same sire match for only 75% of the progeny (based on the thresholds used in this study).
    - The 5K SNPs represented on both the 50K and 5K SNP Chips confirmed only 85% of the parentage results from the 101 Sequenom® SNPs and in total 84% of the progeny were assigned a sire by the 5K SNPs.
    - The lowest error rate for imputation occurred when all the assigned sires, the unrelated industry animals and the progeny with known and unknown sires were used.
    - The 'best guess' genotypes were to be used for GWAS.
4. Use Ovine SNP chips to detect genomic chromosomal regions associated with dagginess, production, flystrike and wool and fibre traits (Chapter 6 and 7). Important outcomes were:
    - No SNPs reached Bonferroni significance $P<1x10^{-6}$, and there were no similarities between the case control GWAS (Chapter 6) and the industry GWAS (Chapter 7).
    - For flystrike, the gene RASGRP1 was identified within 100kbp of the best SNP for this trait.
    - For dag score, DAG3 and DAG8, interesting genes identified within 100kbp of the best SNPs included: SCUBE2, ADAM17, DEFB119, DEFB121, SPINT2, TMPRS11D, TMPRS11A and a GLEAN prediction gene.
    - For the wool traits, GDF2, GDF10, FGF7, WISP3, KRT family, DAAM1, DSG4 and DSG1 were identified.
    - Selection sweep analysis identified regions associated with horns and coat colour, as well as the genes PLUNC and LPLUNC (immune response), FGF5 and SOX18 (wool growth) and SLMO2 (locomotive behaviour).

5. Investigate genomic selection (GS) for dagginess, and test the usefulness of SNP Chip technology in a dual-purpose selection index (Chapter 8). Important outcomes were:
    - The individual accuracies for DAG3 and DAG8 prediction equations ranged between 0.35 to 0.52 for the breeds included in the training population (Romney, Coopworth and CompRCP) and 0.30 to 0.42 for CompRCPT and CompCRP. The combined accuracies ranged from 0.4 to 0.54 and 0.23 to 0.40 respectively.
    - For DAG8, 2 of the best 3 SNPs from both the industry GWAS (Chapter 7) and GS were present in both, including the SNP (s22390) which had a GLEAN prediction gene within 100kbp.
    - Using SNP Chip information on Romney ram hoggets increased the rate of genetic gain for a dual-purpose index by 84 to 100% from using 2 year old rams and phenotype information only.
    - Individually, for DAG3 and DAG8, adding in SNP Chip information decreased the response in this trait by only $0.01 when phenotype information was available.

9.2 Limitations

9.2.1 Case-control study

9.2.1.1 Design

The power analysis performed to estimate the number of cases required to enable the detection of a QTL, with a heritability of at least ~0.03, indicated that 1,000 cases and 1,000 controls would detect the QTL 70 times out of 100. Unfortunately, due to the poor flystrike rates, only ~850 cases and ~780 controls were collected. This reduced the power of detecting significant genomic regions associated with flystrike. To improve the probability of collecting the required number of case samples, future collections could either be run over an increased number of seasons or a greater number of farms could be involved in the trial.

Collecting less than the desired number of cases and controls may have influenced the outcome of the trial. For disease traits, the ability to detect true associations is

dependent on the accuracy of measurement, especially for disease traits where the phenotype is not fully exposed. If the animals had been subjected to the right conditions, potentially every lamb would have been flystruck. That is, there are potentially different levels of susceptibility, and for each flock only one level of challenge (dictated by the incidence rate) has been measured in each season. Biases can arise from either incomplete exposure to the disease, or from imperfect detection of infection. Both can result in a downward bias in estimated single nucleotide polymorphisms (SNP) effects and estimation of heritability (Bishop et al., 2012). However, these were accounted for in the analysis, by adjusting for incidence rates using the methods of Lee et al. (2011).

A case-control experimental design was appropriate for this study, as the cases collected in this study were genetically susceptible. The controls, however, were either truly healthy given they had maximum opportunity to express resistance, or more likely they were only a representation of the population average i.e. a population consisting of resistant animals and susceptible animals that have not had a chance to express their phenotype (a 'Welcome Trust design') (Bishop et al., 2012). This 'Welcome Trust design' is able to obtain information in an association study, however, it is less powerful. To improve the case-control design, Bishop et al. (2012) suggested 4 points that should be included:

- More controls from the population average should be genotyped, i.e. collect at least 1.5 to 2 times as many control samples as case samples from each contemporary group;
- Target controls after peak prevalence, when they have had the maximum opportunity to demonstrate their true resistance;
- Cases should be collected when most likely truly susceptible i.e. when prevalence is low at the start of a season. When prevalence increases, a proportion of the cases will also include those that are least susceptible to infection and the difference between the cases and controls will be smaller; and
- Use multiple flocks which may have differences in prevalence, and adjust for biases by including flock-level prevalence as a covariate in the analysis.

The current case-control design did involve collecting the controls at the end of the season after peak prevalence; cases were truly susceptible as flystrike prevalence was

low (average 2.15% over 2 seasons, Chapter 2) and many of the corrections and adjustments for incidence rates and heritability estimates were undertaken in this study using the rationale of Lee et al. (2011) to adjust for across flock biases. Therefore, the only improvements that could have been made to the current design were to increase the number of controls and to increase the number of farmers or seasons to obtain at least 1,000 cases.

The case-control design may not allow the identification of significant peaks due to the inability to select true controls. An alternative phenotype could be obtained using an artificial challenge, whereby lambs are challenged with either larval implants (Bowles et al., 1992; Bowles et al., 1994; Heath et al., 1987; Sandeman et al., 1985; Seaton and Sandeman, 1991; Steiner et al., 1994) or a high fly population in an enclosed space (Blackwell et al., 1997) to try and affect individuals. The majority of these studies have been done to analyse immune response or insecticide efficacy rather than flystrike susceptibility, and are very expensive per data point collected. Also, to carry out this type of infection on a large scale would be ethically questionable and labour intensive.

The case-control study concentrated on Romney or Romney-cross based flocks. Thus, the correlations for flystrike with other traits can only be applied to Romney-based populations. This was a consequence of the farmers who volunteered to participate in the case-control study. The progeny test which involved flocks of Perendale, Coopworth and Texel base showed that breed did not have an effect on dag score, breech bareness, or wool and fibre traits, which were the common traits measured in both studies. Therefore, the results of the genetic parameter estimates from the case-control dataset may not differ much for other breeds.

9.2.3.2 Analysis

In this study, Survival Kit V6 (Ducrocq et al., 2010) was used as a way to fit flystrike as a longevity or survival trait. The heritability was calculated using a time-dependent variable (exposure time, from start of season until the case or control was collected) and a censored event (flystrike, 0-1 scale) and assumed a Cox proportional hazard model. The resulting heritability was low (0.08 ± 0.05) compared to that obtained using ASReml on the observed or liability scale (0.37 ± 0.10 and 0.26 ± 0.05, respectively). The survival analysis is a more appropriate model for this trait, as it is desirable to select for the most resistant or least susceptible animals that survive the season without infection. The survival model was computationally exhaustive as the

best analysis was to fit all contemporaries in the model, not just the cases and controls collected, and there was not sufficient computer power to undertake this analysis. This model is also limited in its practical use as it could not be incorporated into the current SIL system, which performs the national sheep genetic evaluation. Further investigation using Survival Kit V6 or other programs/methods including random regression models (Veerkamp et al., 2001) and COXF90 (Misztal, 2009) should be performed.

Finally, to improve the power of the experiment, the lamb DNA samples should be analysed on the 50K SNP Chip, instead of imputing from the 5K SNP Chip. This would minimise the genotype error rate due to missing sires in the 'pairs' data file and inaccuracies in the programmes used for imputation. A higher density SNP Chip (~700K SNPs) is planned by the International Sheep Genomics Consortium and if lambs or sires were genotyped on this platform, this would again improve the chance of detecting an association of a SNP in LD with a causative QTL.

9.2.2 Progeny test

Progeny tests are useful for measuring traits that cannot be measured on a live animal or those that are sex-limited. It also increases the accuracy of selection as more progeny are measured, especially for low heritability traits. For high heritability, traits the advantage is less as the sires' (potentially being progeny tested) own phenotype has a sufficiently high correlation with the true genetic merit. The design of the progeny test (Chapter 4), provided reliable genetic parameter estimates for those traits measured on every animal i.e. they had low standard errors for the heritability and phenotypic and genetic correlations between these traits. For the fibre traits, the heritability, genetic and phenotypic correlation estimates were not significant in all cases. This was due to the experimental design in which only a small proportion (30 samples per flock) of the lambs measured for the dag, breech and wool traits also had a wool sample taken to measure the fibre traits. It was expected that 30 random samples would give a good representation of the fleece types present within a flock. In most cases, this equalled 30 random samples per flock. When several flocks were measured together and not individually identifiable, there was a chance that the 30 samples were spread unequally across the flocks and not 30 per flock. Many of the heritabilities for the fibre traits and genetic and phenotypic correlations between fibre traits had low standard errors as the traits were measured on the same sample for each animal. The genetic and phenotypic correlations between the fibre traits and the remainder of the traits had standard errors

similar to or larger than the estimates. To mitigate this problem, the best option would be to collect a wool sample from every animal in the progeny test. This may not be cost effective, and a better solution could be to collect wool samples from at least 5-10 progeny per sire.

9.2.3 Genome-wide association analysis

The GWAS methods for the case-control study (Chapter 6) were very similar to those used by the GenABEL software (Aulchenko et al., 2007). The industry GWAS (Chapter 7) using breeding values as phenotypes required some modification and was similar to parts of the GS methods. The use of breeding values as phenotypes in a GWAS and a comparison with using trait phenotypes has not been previously reported in the literature. This is an area that requires further research, but it is beyond the scope of this study. Other packages such as PLINK (Purcell et al., 2007) and Golden Helix (www.goldenhelix.com/SNP_Variation) are alternatives that could be considered for use.

9.3 Future considerations

9.3.1 Flystrike

To improve the analysis of flystrike, two options could be investigated. Firstly, the current analysis did not fully consider body strike. The shoulders and the middle body were the second most common flystruck regions after the breech. Due to the high proportion of flystrike cases occurring in the breech, the genetic and phenotypic correlations between flystrike and potential indicator traits mainly highlighted those traits correlated with breech strike, particularly dag score. There were some indicative genetic correlations between the fibre traits, FDCV, FDSD and BULK and flystrike, however, the phenotypic correlations were close to zero. It was anticipated that some fibre traits would be associated with body strike. A study investigating body-strike and the structure of the fleece and its impact on humidity and fly attractiveness would potentially highlight the reasons for these strikes. In Australian Merinos, there was a between-flock correlation between body strike and fleece rot of 0.86 (Mortimer et al., 1998). Raasdma (1993) also reported correlations of wool and fibre traits with fleece rot and body-strike, including FDSD, FDCV and staple length. However, these parameter

estimates were in Merinos; there are no known published estimates in New Zealand dual-purpose breeds. Present knowledge would suggest that selection against dagginess would greatly reduce flystrike and associated costs. If there was a consequent increase in body-strike, then investigation into fleece traits would be worthwhile.

Secondly, the majority of sheep in a New Zealand dual-purpose flock are adult ewes. There is potentially a larger cost associated with a flystruck ewe than a flystruck lamb, caused by the need for greater numbers of replacements and the loss of performance e.g. fleece growth, reduced fertility and maternal effects on lamb growth. If a lamb that has recovered from flystrike is kept as a replacement, the impact of the juvenile flystrike event on its future performance and propensity for being flystruck again are important considerations. The first question can be answered by following flystruck ewe lambs through their performance as an adult. If a flystruck lamb can recover and perform competitively with its peers, then should selection be for flystrike resistance or for survival of the flystrike event? The second question; whether the animal would get flystruck again, is influenced by the seasonality of the disease, and would be harder to answer, but a lifetime study of affected and non-affected ewe lambs should address this to a limited extent.

9.3.2 Dagginess

A number of questions on dagginess in lambs remain unanswered. Firstly, is dag score as a lamb (3 or 8 months) related to the individuals adult dag score? Secondly, what do dags consist of? Finally, what are the causative factors for dagginess?

Lamb dag score is important for 2 reasons; firstly, lambs kept as replacements that undergo shearing at weaning and if daggy will require dagging/crutching prior to shearing. Secondly, lambs sent for slaughter must be presented without dags to avoid penalties. Dag score as an adult is also important, as daggy ewes require crutching pre-shearing, pre-mating and pre-lambing; a higher dag score increases the time it takes for the ewes to be crutched, thereby increasing costs (Byrne et al., 2012).

Currently within SIL, the EBVs calculated from DAG8 measures are reported as the adult dag score EBVs. To do this requires the genetic and phenotypic parameters between the juvenile and adult traits to be known. To the authors knowledge, there are no published estimates of the relationships between dag score as a lamb and adult dag score. Adult ewe dag scores can now be recorded in the SIL database. It is

recommended that ewes be measured at the same time each year, and that the average mob dag score should be close to 1 before measurements are taken. When a sufficient number of adult dag scores are available, a preliminary analysis could be performed to estimate the heritability of adult dag score and the likely correlation of adult dag score with DAG8 and DAG3 and the production traits such as WWT, LW8, EWT and NLB. The results of the initial analysis should confirm if the current SIL model is appropriate for reporting adult dag EBVs, or if the model needs revising to include adult dag measures and its correlations with DAG3 and DAG8.

To investigate the composition of dags and the cause of dagginess, an experimental design is required that tests animals selected via phenotype that are from the extremes (very daggy or very clean) under a number of environmental and pasture conditions. A number of studies have considered the effects of pasture quality, composition and quantity on dag score (Davidson et al., 2006; Fletcher et al., 1999; Leathwick and Atkinson, 1995, 1998; Min et al., 1998; Ramirez-Restrepo et al., 2004). A single study examined faecal mineral and dry matter content (Reid and Cottle, 1999) in relation to dagginess. These studies need to be progressed further with a study of faecal consistency and composition over time and on different pasture types. Faecal composition including the proportion of dry matter, minerals, mucus, microbial content and metabolites should be measured. To investigate the digestive system and build on the work Ruckenbusch and Fioramonti (1980) and Bedrich and Ehrlein (2001) performed, a gene expression study of the digestive system and the 4 stomachs, pancreas, spiral colon and small intestine is required. The results of a RNA sequencing or a microarray study can be examined to identify genes in common with the industry GWAS (Chapter 7) performed in this project.

A New Zealand green-house gas project is currently underway, measuring methane emissions from New Zealand dual-purpose sheep selected for either low or high methane emissions (J. C. McEwan pers. comm.). As part of this study, rumen samples, rumen volume measurements by CT scanning and dag scores are being collected. From this study, there is an indication that low methane emitters (gCH_4/kgDMI) had higher dag scores based on genetic correlations albeit phenotypic correlations were close to zero (J. C. McEwan pers. comm.). If such a correlation does exist, it would be supported by higher reported flow rate (Pinares-Patino et al., 2011).

The selection sweep comparing high dag BV Romney sheep with Arapawa sheep identified several genes of interest. To extend this study, additional breeds could be included into the selection sweep analysis. Arapawa sheep were found to cluster with Gulf Coast Native sheep (GCN) in a principal component analysis (Young et al., 2011). There are two strains of the GCN, the Florida and the Louisiana lines with some observed differences between the two (Kijas et al., 2012b). The GCN are thought to have descended from the Spanish Churra and were brought to the Americas as early as the 1500s, they have adapted to high parasitic loads and have natural resistance to internal parasites. A comparison of Romney and GCN animals may be beneficial in identifying new genes and validating genes from the Arapawa comparison involved in dagginess formation.

9.3.3 Genomic Selection

Genomic selection can significantly improve the genetic gains in the New Zealand sheep industry. Chapter 8 demonstrated that by enhancing performance and pedigree records with genomic information, the rate of genetic gain for a dual-purpose index could be improved by up to 84%. These preliminary results require validation and extension to, for example, two-stage selection, while also including the costs of genotyping. Sise et al. (2011) investigated the cost effectiveness of using genomic technologies to assist in selection decisions on ram lambs. This showed that for traits with prediction accuracy of 0.3 and 0.5, the potential increase in the rate of genetic gain was 4% and 10%, respectively, when all animals were tested. They also observed that genetic gain for each trait varied, with those traits that were sex-limited or disease relevant, e.g. reproduction and parasite resistance had a greater increase in the rate of genetic gain than easy-to-measure traits such as growth and wool.

The study by Sise et al. (2011) included assumptions for; MBV accuracies, correlations between selection index values on recorded information and selection index values using recorded information and MBVs and the assumption of costs of 50K SNP chip testing. The analysis can now be repeated using actual values obtained in this study, and using the 5K SNP chip and imputation rather than the 50K SNP chip. The correlation between MBVs estimated from 50K SNP chip and from the 5K SNP chip which has been imputed to 50K information is 95% (Anderson et al., 2012b). Other options to be considered in a more comprehensive study should include genotyping the sires of the ram lambs undergoing two stage selection with the 50K SNP Chip so that

imputation can be performed to increase MBV accuracies. Another aspect worthy of study is to examine the proportion of top ram lambs, as ranked by SIL EBVs, that should be genotyped using the SNP Chip to obtain maximum genetic gain. The costs of ensuring sires of rams are tested and the commercial costs of the 5K SNP Chip and imputation needs to be estimated. Timing of two stage selection should also be considered. The current Ovita 5K Beta test project (L. Archer, pers. comm.) identified a number of stages when MBVs would be a valuable tool in selection decisions: a) genotype ram lambs in June for selection of ram lambs to undergo facial eczema testing, b) genotype ram lambs in January – March for selection of ram lambs for use that year, c) genotype ram lambs in October – November for use as 2 year olds, d) sire-evaluation, where a number of progeny per sire are genotyped, e) screen ewe lambs for entering a nucleus flock and f) for parentage evaluation.

The current accuracies of the traits analysed for genomic selection ranged between 0.0 and 0.6, depending on the trait and breed. When genomic selection was first proposed for livestock and crop species, it was predicted that accuracies of up to 85% would be achievable, however, it is now accepted that accuracies between 40 and 60% are more achievable (Pryce and Daetwyler, 2012). If accuracies of ~70% could be attained, the number of animals genotyped with phenotypic measurements for each trait would need to drastically increase from the current numbers used. As was calculated in Chapter 8, up to a million animals are needed for lamb survival, and ~48,000 if each sire genotyped has 127 progeny records to get an accuracy of 70% for the trait. It is unlikely that this number of animals required could be obtained. An alternative way to increase the accuracy is to genotype animals with a higher density SNP Chip and/or to sequence key sires. Thus, imputation could be used to estimate the missing genotypes of the animals currently genotyped on the 50K SNP Chip to the higher density chip and up to full sequence. Meuwissen et al. (2001) showed that a ~10% increase in accuracy can be obtained by increasing marker density from 1 SNP every 4cM to 1 SNP every 1cM.

With the advent of the SNP Chips and the ability to screen a large number of markers densely spaced across the genome, the way in which traits are examined genomically has been modified. Historically small genome-wide scans were performed and QTL of large effect were identified to a region which required fine mapping to identify the causal mutation or genetic region. Other techniques involved a literature search of the trait and identifying candidate genes which could be tested for association

with the trait. These experiments took a number of years to complete. However, with the SNP chip, genome-wide scans can easily pin-point chromosomal regions, for example the polled region for sheep was first identified to chromosome 10 by Montgomery et al. (1996). More than ten years, later the polled gene, RXFP2, and a mutation was identified by fine mapping the region using recombinant progeny from the Montgomery study (Pickering et al., 2009b). Near the end of the Pickering et al. (2009b) study, the 50K SNP chip was used and pin-pointed the polled gene within 2 weeks of genotyping the animals.

9.4 Implications of these findings

The results of this study have confirmed that flystrike is highly genetically and phenotypically correlated with dag score and that dagginess can be selected against without adversely affecting other production traits in New Zealand Romney cross sheep. Secondly, this study has identified that wool traits and faecal egg count traits are not associated with accumulation of dags, leaving internal processes such as digestion and immune response to be investigated. The results from the genetic parameter analysis performed in Chapters 3 and 4 concerning dag score traits and production traits have been used to update the parameters used in SIL genetic evaluations. For the production traits, some of these parameters had not been updated for at least 15 years and the initial estimates were a result of "bending" (Hayes and Hill, 1981) the then published estimates, and then weighting estimates by the number of sires used in the various analyses. The new DAG3 and DAG8 models were also implemented in SIL and were shown in Chapter 3 to reduce the variation in standard deviation resulting in estimated breeding values (EBVs) that also had reduced variability. Currently, the estimates for breech bareness and the correlations with DAG3 and DAG8 are being used for a bare points module in SIL.

The GWAS, though not reaching Bonferroni significance levels, have identified some genes that require further study to confirm whether they are involved in the genetic basis of dagginess. Especially, the GLEAN predicted gene underlying the best SNP s22390 (OAR15) for DAG8 in the industry GWAS and GS analyses and the immune, inflammatory and diarrhoea genes identified for flystrike and dag score in the case-control analysis, RASGRP1, SCUBE2, ADAM17, DEF119 and 121 and SPINT2.

Genomic selection can be applied to a variety of dual-purpose traits in a number of breed types including DAG3 and DAG8 in the breeds Romney, Coopworth, CompRCP and CompRCPT. The GS analysis can provide significant improvements in genetic gain for New Zealand dual-purpose production trait index. An increase of 84% in the rate of genetic gain from using a 2 year old ram with measurements on itself, parents and half siblings to using a ram hogget with those measurements plus MBVs can be achieved. This is due to the incorporation of information on reproduction, survival, and adult live weights.

References

AFFCO. 2010. Select programmes. Accessed 21 August 2012. http://www.affco.co.nz/indexf595.html.

Allen, G. 1988. Tuberculosis in sheep-a very rare disease. Surveillance 15: 8-9.

Amer, P. R. 2009. Rates of genetic progress being acheived throughout the New Zealand ram breeding industry. p 155-157 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Anderson, C. A., F. H. Pettersson, J. C. Barrett, J. J. Zhuang, J. Ragoussis, L. R. Cardon and A. P. Morris. 2008. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. Am. J. Hum. Genet. 83: 112-119.

Anderson, R., B. Auvray, N. K. Pickering, K. G. Dodds, M. J. Bixley, D. Hyndman and J. C. McEwan. 2012a. Development and characterisation of a low density (5K) Illumina ovine SNP chip. p P0576 in Plant & Animal Genomes XX, San Diego, CA, United States of America.

Anderson, R., B. Auvray, H. Henry, H. Baird, M. Lee, L. Archer and J. McEwan. 2012b. Development and beta test of a low density (5K) Illumina ovine SNP chip. p G22 in Queenstown Molecular Biology Conference, Queenstown, New Zealand.

Anon. 2005. Fly advice, a farmer's guide to blowfly control in sheep, Schering-Plough Coopers Upper Hutt, New Zealand.

Armstrong, D. T., P. Kotaras and C. R. Earl. 1997. Advances in production of embryos *in vitro* from juvenile and prepubertal oocytes from the calf and lamb. Reprod. Fertil. Dev. 9: 333.

Asure Quality. 2010. Farm and Farm Dairy Sevices. Accessed 25 May 2010. http://www.asurequality.com/services-offered-to-farms-and-livestock/services-offered-to-farm-and-farm-dairies.cfm.

Atkins, K. D. 1979. Breeding for resistance to body strike. p 145-156. in National symposium on the sheep blowfly and flystrike in sheep, Sydney, Australia.

Aulchenko, Y. S., S. Ripke, A. Isaacs and C. M. Van Duijn. 2007. GenABEL: an R library for genome-wide association analysis. Bioinformatics 23: 1294-1296.

Australian Wool Innovation and Meat Livestock Australia. 2007. Visual Sheep Scores, Australian Wool Innovation Ltd, Sydney, Australia.

Azeredo-Espin, A. M. L. and A. C. Lessinger. 2006. Genetic approaches for studying myiasis-causing flies: molecular markers and mitochondrial genomics. Genetica 126: 111-131.

Baker, R. L., T. G. Watson, S. A. Bisset, A. Vlassoff and P. G. C. Douch. 1991. Breeding sheep in New Zealand for resistance to internal parasites: research results and commercial application. p 19-32 in Breeding for disease resistance in sheep G. D. Gary and R. R. Woolaston, eds. Australian Wool Corporation, Melbourne.

Barić, I., K. Fumić, B. Glenn, M. Ćuk, A. Schulze, J. D. Finkelstein, S. J. James, V. Mejaški-Bošnjak, L. Pažanin, I. P. Pogribny, M. Radoš, V. Sarnavka, M. Šćukanec-Špoljar, R. H. Allen, S. Stabler, L. Uzelac, O. Vugrek, C. Wagner, S. Zeisel and S. H. Mudd. 2004. S-adenosylhomocysteine hydrolase deficiency in a human: A genetic disorder of methionine metabolism. PNAS 101: 4234-4239.

Barritt, L. C. and L. M. Birt. 1971. Development of *Lucilia cuprina*: correlations of biochemical and morphological events. J. Insect Physiol. 17: 1169-1183.

Barton, N. 2007. Sheep Blowflies of Victoria Agricultural Notes No. AG0081. Department of Primary Industries, Victoria, Australia.

Batterham, P., A. Hill-Williams, G. Levot, N. Sales and J. A. McKenzie. 2006. The genetic basis of high-level resistance to diflubenzuron and low-level resistance to cyromazine in a field strain of the Australian sheep blowfly, *Lucilia cuprina* (Wiedemann) (Diptera: Calliphoridae). Aust. J. Entomol. 45: 87-90.

Baxter, B. P., M. A. Brims and T. B. Taylor. 1992. Description and Performance of the Optical Fibre Diameter Analyser (OFDA). J. Textile Inst. 83: 507-526.

Bedrich, M. and H. Ehrlein. 2001. Motor function of the large intestine and flow of digesta in sheep. Small Ruminant Res. 42: 141-155.

Beef and Lamb New Zealand. 1999. Research and development brief no.54 Combating flystrike and lice through using a model approach. Beef + Lamb New Zealand, Wellington, New Zealand.

Beef and Lamb New Zealand. 2009. Beef + Lamb New Zealand's Economic Service. Beef + Lamb New Zealand, Wellington, New Zealand.

Bennett, E. P., H. Hassan, U. Mandel, M. A. Hollingsworth, N. Akisawa, Y. Ikematsu, G. Merkx, A. G. van Kessel, S. Olofsson and H. Clausen. 1999. Cloning and characterization of a close homologue of human UDP-N-acetyl-α-d-

galactosamine:polypeptide N-acetylgalactosaminyltransferase-T3, designated GalNAc-T6. J. Biol. Chem. 274: 25362-25370.

Bennett, E. P., U. Mandel, H. Clausen, T. A. Gerken, T. A. Fritz and L. A. Tabak. 2011. Control of mucin-type O-Glycosylation – a classification of the polypeptide GalNAc-transferase gene family. Glycobiology 22: 736-756.

Bennett, R. 2003. The 'direct costs' of livestock disease: The development of a system of models for the analysis of 30 endemic livestock diseases in Great Britain. J. Agric. Econ. 54: 55-71.

Berry, D. and J. Kearney. 2011. Imputation of genotypes from low-to high-density genotyping platforms and implications for genomic selection. Animal 5: 1162 -1169.

BGI Americas. 2012. Sample requirements. Accessed 3 October 2012. http://bgiamericas.com/customer-support/sample-requirements/.

Bisdorff, B., A. Milnes and R. Wall. 2006. Prevalence and regional distribution of scab, lice and blowfly strike in Great Britain. Vet. Rec. 158: 749-752.

Bishop, S. C., A. B. Doeschl-Wilson and J. A. Woolliams. 2012. Uses and implications of field disease data for livestock genomic and genetics Studies. Front. Genet. 3: 114.

Bisset, S. A., A. Vlassoff, C. A. Morris, B. R. Southey, R. L. Baker and A. G. H. Parker. 1992. Heritability of and genetic correlations among faecal egg counts and prodcutivity traits in Romney sheep. N. Z. J. Agric. Res. 35: 51-58.

Bisset, S. A., A. Vlassoff, P. G. C. Douch, W. E. Jonas, C. J. West and R. S. Green. 1996. Nematode burdens and immunological responses following natural challenge in Romney lambs selectively bred for low or high faecal worm egg count. Vet. Parasitol. 61: 249-263.

Bisset, S. A., A. Vlassoff, C. J. West and L. Morrison. 1997. Epidemiology of nematodosis in Romney lambs selectively bred for resistance or susceptibility to nematode infection. Vet Parasitol 70: 255-269.

Blackwell, G. L., M. A. Potter, Y. H. Cottam and H. T. Blair. 1997. Susceptibility of Romney and Perendale sheep to flystrike by the Australian green fly, *Lucilia cuprina* (Wied.), and fly attractant trials. p 37-40 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Blair, H. T. and D. J. Garrick. 2007. Application of new technologies in sheep breeding. N. Z. J. Agric. Res. 50: 89-102.

Blaydon, D. C., P. Biancheri, W.-L. Di, V. Plagnol, R. M. Cabral, M. A. Brooke, D. A. van Heel, F. Ruschendorf, M. Toynbee, A. Walne, E. A. O'Toole, J. E. Martin, K. Lindley, T. Vulliamy, D. J. Abrams, T. T. MacDonald, J. I. Harper and D. P. Kelsell. 2011. Inflammatory Skin and Bowel Disease Linked to ADAM17 Deletion. N. Engl. J. Med. 365: 1502-1508.

Bonnefont, C., M. Toufeer, C. Caubet, E. Foulon, C. Tasca, M.-R. Aurel, D. Bergonier, S. Boullier, C. Robert-Granie, G. Foucras and R. Rupp. 2011. Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. BMC Genomics 12: 208.

Bowles, V. M., S. T. Grey and M. R. Brandon. 1992. Cellular immune responses in the skin of sheep infected with larvae of *Lucilia cuprina*, the sheep blowfly. Vet. Parasitol. 44: 151-162.

Bowles, V. M., E. N. T. Meeusen, K. Chandler, A. Verhagen, A. D. Nash and M. R. Brandon. 1994. The immune response of sheep infected with larvae of the sheep blowfly *Lucilia cuprina* monitored via efferent lymph. Vet. Immunol. Immunopathol. 40: 341-352.

Brand, B., A. Hartmann, D. Repsilber, B. Griesbeck-Zilch, O. Wellnitz, C. Kühn, S. Ponsuksili, H. H. D. Meyer and M. Schwerin. 2011. Comparative expression profiling of *E. coli* and *S. aureus* inoculated primary mammary gland cells sampled from cows with different genetic predispositions for somatic cell score. Genet. Sel. Evol. 43: 24.

Brandsma, J. H. and H. T. Blair. 1997. Heritability of resistance to flystrike in New Zealand Perendale sheep. p 41-42 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Brash, L., N. Fogarty and A. Gilmour. 1994a. Genetic parameters for Australian maternal and dual-purpose meatsheep breeds. II. Liveweight, wool and reproduction in Corriedale sheep. Aust. J. Agric. Res. 45: 469-480.

Brash, L., N. Fogarty, S. Barwick and A. Gilmour. 1994b. Genetic parameters for Australian maternal and dual-purpose meatsheep breeds. I. Liveweight, wool production

and reproduction in Border Leicester and related types. Aust. J. Agric. Res. 45: 459-468.

Brash, L., N. Fogarty and A. Gilmour. 1994c. Genetic parameters for Australian maternal and dual-purpose meatsheep breeds. III. Liveweight, fat depth and wool production in Coopworth sheep. Aust. J. Agric. Res. 45: 481-486.

Brash, L., N. Fogarty and A. Gilmour. 1994d. Reproductive performance and genetic parameters for Australian Dorset Sheep. Aust. J. Agric. Res. 45: 427-441.

Broughan, J. M. and R. Wall. 2006. Control of sheep blowfly strike using fly-traps. Vet. Parasitol. 135: 57-63.

Broughan, J. M. and R. Wall. 2007a. Fly abundance and climate as determinants of sheep blowfly strike incidence in southwest England. Med. Vet. Entomol. 21: 231-238.

Broughan, J. M. and R. Wall. 2007b. Faecal soiling and gastrointestinal helminth infection in lambs. Int. J. Parasitol. 37: 1255-1268.

Brown, D. J., K. Atkins and A. E. Huisman. 2005. Expression of body weight, fleece weight and fibre diameter in across flock genetic evaluation. p 84-87 in Proc. Assoc. Advmt. Anim. Breed. Genet., Noosa Lakes, QLD, Australia.

Brown, D. J., A. A. Swan and J. S. Gill. 2010. Within- and across-flock genetic relationships for breech flystrike resistance indicator traits. Anim. Prod. Sci. 50: 1060-1068.

Browning, B. L. and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84: 210-223.

Bruce, L. J., M. M. B. Kay, C. Lawrence and M. J. Tanner. 1993. Band 3 HT, a human red-cell variant associated with acanthocytosis and increased anion transport, carries the mutation Pro - 868 - Leu in the membrane domain of band 3. Biochem. J. 293: 317-320.

Bruce, L. J., D. J. Anstee, F. A. Spring and M. J. Tanner. 1994. Band 3 Memphis variant II. Altered stilbene disulfonate binding and the Diego (Dia) blood group antigen are associated with the human erythrocyte band 3 mutation Pro854-->Leu. J. Biol. Chem. 269: 16155-16158.

Byrne, T. J., C. I. Ludemann, P. R. Amer and M. J. Young. 2012. Broadening breeding objectives for maternal and terminal sheep. Livest. Sci. 144: 20-36.

Campbell, P. M., R. D. Newcomb, R. J. Russell and J. G. Oakeshott. 1998. Two different amino acid substitutions in the ali-esterase, E3, confer alternative types of organophosphorus insecticide resistance in the sheep blowfly, *Lucilia cuprina*. Insect Biochem. Mol. Biol. 28: 139-150.

Cardon, L. and J. Bell. 2001. Association study designs for complex diseases. Nat. Rev. Genet. 2: 91-99.

Carhan, A., S. Reeve, C. T. Dee, R. A. Baines and K. G. Moffat. 2004. Mutation in *slowmo* causes defects in *Drosophila* larval locomotor behaviour. Invertebr. Neurosci. 5: 65-75.

Casida, J. E. and G. B. Quistad. 1998. Golden age of insecticide research: past, present, or future? Annu. Rev. Entomol. 43: 1-16.

Cichota, R., V. Snow and A. Tait. 2008. A functional evaluation of virtual climate station rainfall data. N. Z. J. Agric. Res. 51: 317-329.

Clarke, J. N., J. L. Dobbie, K. R. Jones, A. L. Wrigglesworth and S. M. Hickey. 2000. Genetic parameters for improvement of dual-purpose flock productivity with constraints on mature ewe body weight. p 203-206 in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Clarke, S., H. M. Henry, T. Manley and J. C. McEwan. 2011. Parentage assay development for NZ sheep breeds utilising Sequenom iPLEX platunum chemistry. p Q51 in QMB, Queenstown, New Zealand.

Colditz, I. G., S. W. Walkden-Brown, B. L. Daly and B. J. Crook. 2005. Some physiological responses associated with reduced wool growth during blowfly strike in Merino sheep. Aust. Vet. J. 83: 695-699.

Conington, J., S. Bishop, A. Waterhouse and G. Simm. 1995. A genetic analysis of early growth and ultrasonic measurements in hill sheep. Anim. Sci. 61: 85-93.

Cooke, G. S., S. J. Campbell, S. Bennett, C. Lienhardt, K. P. W. J. McAdam, G. Sirugo, O. Sow, P. Gustafson, F. Mwangulu, P. van Helden, P. Fine, E. G. Hoal and A. V. S. Hill. 2008. Mapping of a novel susceptibility locus suggests a role for MC3R and CTSZ in human tuberculosis. Am. J. Respir. Crit. Care. Med. 178: 203-207.

Cooper, S. M. 2011. Farm faults on pelts and hides for the 2009/2010 season RR192, New Zealand Leather & Shoe Research Association Inc., Palmerston North, New Zealand.

Craven, A. J., M. G. Ashby, J. E. Wildermoth, N. R. Cox, D. R. Scobie and A. J. Nixon. 2009. Variation of wool between and across the fleeces of composite-breed strong-wool sheep in New Zealand. p 90-95 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Craven, A. J., M. G. Ashby and D. R. Scobie. 2010. Wool characteristics in the Meat and Wool New Zealand Central Progeny Test flocks. p 217-221 in Proc. N. Z. Soc. Anim. Prod., Palmerston North, New Zealand.

Crawford, A., K. Paterson, K. Dodds, D. Tascon, P. Williamson, R. Thomson, S. Bisset, A. Beattie, G. Greer and R. Green. 2006. Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. BMC Genomics 7: 178.

Dalrymple, B., E. Kirkness, M. Nefedov, S. McWilliam, A. Ratnakumar, W. Barris, S. Zhao, J. Shetty, J. Maddox, M. O'Grady, F. Nicholas, A. Crawford, T. Smith, P. de Jong, J. McEwan, V. H. Oddy, N. Cockett and the International Sheep Genomics Consortium. 2007. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. Genome Biology 8: R152.

Dalrymple, B. P. 2009. Identification of sheep SNPs using Illumina sequencing and design of the sheep 60K Illumina Iselect^{TM} SNP BeadChip. p W083 in Plant & Animal Genomes XVII, San Diego, CA, United States of America.

Davidson, B. S., S. J. Chaplin and C. Laird. 2006. Effect of fibre supplementation on dag formation and flystrike in sheep grazing spring pastures. Aust. J. Exp. Agric. 46: 783-786.

Davies, L. 1948. Observations on the development of *Lucilia sericata* (Mg.) eggs in sheep fleeces. J. Exp. Biol. 25: 86-102.

Davis, G. H., C. A. Morris and K. G. Dodds. 1998. Genetic studies of prolificacy in New Zealand sheep. Anim. Sci. 67: 289-297.

de Roos, A. P. W., B. J. Hayes and M. E. Goddard. 2009. Reliability of Genomic Predictions Across Multiple Populations. Genetics 183: 1545-1553.

Delaunay, J. 2007. The molecular basis of hereditary red cell membrane disorders. Blood Rev. 21: 1-20.

Dodds, K. G., M. L. Tate and J. A. Sise. 2005. Genetic evaluation using parentage information from genetic markers. J. Anim. Sci. 83: 2271-2279.

Dodds, K. G., B. Auvray, N. Pickering and J. C. McEwan. 2009. Quality control for ovine SNP50 beadchip genotypes. p 296-299 in Proc. Assoc. Advmt. Anim. Breed. Genet., Barrosa Valley, SA, Australia.

Dominik, S., J. M. Henshall and B. J. Hayes. 2012. A single nucleotide polymorphism on chromosome 10 is highly predictive for the polled phenotype in Australian Merino sheep. Anim. Genet. 43: 468-470.

Douch, P. G. C., R. S. Green, C. A. Morris, S. A. Bisset, A. Vlassoff, R. L. Baker, T. G. Watson, A. P. Hurford and M. Wheeler. 1995. Genetic and phenotypic relationships among anti-*Trichostrongylus colubriformis* antibody level, faecal egg count and body weight traits in grazing Romney sheep. Livest. Prod. Sci. 41: 121-132.

Druet, T. and M. Georges. 2010. A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. Genetics 184: 789-798.

Ducrocq, V., J. Sölkner and G. Mészáros. 2010. Survival Kit v6-A software package for survival analysis. p 232 in Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany.

Duhl, D. M., M. E. Stevens, H. Vrieling, P. J. Saxon, M. W. Miller, C. J. Epstein and G. S. Barsh. 1994. Pleiotropic effects of the mouse lethal yellow (Ay) mutation explained by deletion of a maternally expressed gene and the simultaneous production of agouti fusion RNAs. Development 120: 1695-1708.

Dunlop, A. A. and R. H. Hayman. 1958. Differences among Merino strains in resistance to fleece-rot. Aust. J. Agric. Res. 2: 261-266.

Elkington, R. A. and T. J. Mahony. 2007. A blowfly strike vaccine requires an understanding of host-pathogen interactions. Vaccine 25: 5133-5145.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto, E. S. Buckler and S. E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Elsik, C., A. Mackey, J. Reese, N. Milshina, D. Roos and G. Weinstock. 2007. Creating a honey bee consensus gene set. Genome Biology 8: R13.

Falconer, D. S. 1989. Introduction to Quantitative Genetics. Third ed. Longman Scientific & Technical, England. p 192 and 320.

Farkas, R., M. J. R. Hall and F. Kelemen. 1997. Wound myiasis of sheep in Hungary. Vet. Parasitol. 69: 133-144.

Farmer, R. J. and G. H. Davis. 1999. Experiences of a pioneer in the commercial sheep pregnancy scanning business. p 91-92 in Proc. N. Z. Soc. Anim. Prod., Mosgiel, New Zealand.

Fell, L. R. and D. A. Shutt. 1989. Behavioural and hormonal responses to acute surgical stress in sheep. Appl. Anim. Behav. Sci. 22: 283-294.

Fellay, J., A. J. Thompson, D. Ge, C. E. Gumbs, T. J. Urban, K. V. Shianna, L. D. Little, P. Qiu, A. H. Bertelsen, M. Watson, A. Warner, A. J. Muir, C. Brass, J. Albrecht, M. Sulkowski, J. G. McHutchison and D. B. Goldstein. 2010. ITPA gene variants protect against anaemia in patients treated for chronic hepatitis C. Nature 464: 405-408.

Fisher, M. W., N. G. Gregory, J. E. Kent, D. R. Scobie, D. J. Mellor and J. C. Pollard. 2004. Justifying the appropriate length for docking lambs' tails - a review of the literature. p 293-296 in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Fletcher, L. R., B. L. Sutherland and C. G. Fletcher. 1999. The impact of endophyte on the health and productivity of sheep grazing ryegrass-based pastures. Grassland Research and Practice Series 7: 11-17.

French, N. P., R. Wall, P. J. Cripps and K. L. Morgan. 1992. Prevalence, regional distribution and control of blowfly strike in England and Wales. Vet. Rec. 131: 337-342.

French, N. P., R. Wall and K. L. Morgan. 1994. Lamb tail docking: a controlled field study of the effects of tail amputation on health and productivity. Vet. Rec. 134: 463-467.

French, N. P., R. Wall and K. L. Morgan. 1995. The seasonal pattern of sheep blowfly strike in England and Wales. Med. Vet. Entomol. 9: 1-8.

French, N. P., T. D. H. Parkin and K. L. Morgan. 1996. A case control study of blowfly strike in lambs. Vet. Rec. 139: 384-388.

French, N. P., E. Berriatua, G. Kaya and K. L. Morgan. 1998. Case control study of diarrhoea and faecal soiling in two- to six-month-old lambs. Vet. Rec. 143: 408-412.

Gábor, M., P. Judit, D. Vincent and S. Johann. 2010. Heritability of longevity in Large White and Landrace sows using continuous time and grouped data models. Genet. Sel. Evol. 42: 13.

Garrick, D. J., J. F. Taylor and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41: 55.

Gaskin, K. J., P. R. Durie, R. E. Hill, L. M. Lee and G. G. Forstner. 1982. Colipase and maximally activated pancreatic lipase in normal subjects and patients with steatorrhea. J. Clin. Invest. 69: 427-434.

Gauderman, W. J. and J. M. Morrison. 2006. QUANTO1.1: A computer program for power and sample size calculations for genetic-epidemiology studies, http://hydra.usc.edu/gxe.

Geissler, E. N., M. A. Ryan and D. E. Housman. 1988. The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. Cell 55: 185-192.

Georges, M. and J. M. Massey. 1991. Velogenetics, or the synergistic use of marker assisted selection and germ-line manipulation. Theriogenology 35: 151-159.

Gherardi, S. G., S. S. Sutherland, N. Monzu and K. G. Johnson. 1983. Field observations on body strike in sheep affected with dermatophilosis and fleece-rot. Aust. Vet. J. 60: 27-28.

Gilmour, A. R., B. J. Gogel, B. R. Cullis and R. Thompson. 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. p 372.

Goddard, M. and B. Hayes. 2007. Genomic selection. J. Anim. Breed. Genet. 124: 323-330.

Goddard, M. E. 2009. Genomic seletion: prediction of accuracy and maximisation of long term response. Genetica 136: 245-257.

Gogolewski, R. P., D. Rugg, G. R. Allerton, D. Kawhia, R. A. Barrick and J. S. Eagleson. 1997a. Demonstration of the sustained anthelmintic activity of a controlled-release capsule formulation of ivermectin in ewes under field conditions in New Zealand. N. Z. Vet. J. 45: 163-166.

Gogolewski, R. P., G. R. Allerton, D. Rugg, D. Kawhia, R. A. Barrick and J. S. Eagleson. 1997b. Demonstration of the sustained anthelmintic efficacy of a controlled-release capsule formulation of ivermectin in weaner lambs under field conditions in New Zealand. N. Z. Vet. J. 45: 158-161.

González-Recio, O., K. A. Weigel, D. Gianola, H. Naya and G. J. M. Rosa. 2010. L2-Boosting algorithm applied to high-dimensional problems in genomic selection. Genetics Research 92: 227-237.

Gorlov, I. P., A. Kamat, N. V. Bogatcheva, E. Jones, D. J. Lamb, A. Truong, C. E. Bishop, K. McElreavey and A. I. Agoulnik. 2002. Mutations of the GREAT gene cause cryptorchidism. Hum. Mol. Genet. 11: 2309-2318.

Gosey, J. A. 1991. Crossbreeding systems and the theory behind composite breeds. p 33-54 in Proc. Range Beef Cow Symp. XII, Fort Collins, Colorado.

Greeff, J. and L. J. E. Karlsson. 2009. Opportunities to breed for resistance to breech strike in Merino sheep in a mediterranean environment. p 272-278 in Proc. Assoc. Advmt. Anim. Breed. Genet., Barossa Valley, SA, Australia.

Guo, N., C. Hawkins and J. Nathans. 2004. Frizzled6 controls hair patterning in mice. Proc. Natl. Acad. Sci. U. S. A. 101: 9277-9281.

Habas, R., Y. Kato and X. He. 2001. Wnt/Frizzled Activation of Rho Regulates Vertebrate Gastrulation and Requires a Novel Formin Homology Protein Daam1. Cell 107: 843-854.

Habier, D., R. L. Fernando, K. Kizilkaya and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics 12: 186.

Hall, M. J. R. 1997. Traumatic myiasis of sheep in Europe: A review. Parassitologia 39: 409-413.

Hargreaves, A. L. and G. D. Hutson. 1990. The stress response in sheep during routine handling procedures. Appl. Anim. Behav. Sci. 26: 83-90.

Harris, B. and D. Johnson. 1998. Approximate reliability of genetic evaluations under an animal model. J. Dairy Sci. 81: 2723-2728.

Hayes, B., P. Bowman, A. Chamberlain, K. Verbyla and M. Goddard. 2009a. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet. Sel. Evol. 41: 51.

Hayes, B., P. Bowman, A. Chamberlain and M. Goddard. 2009b. Invited review: Genomic selection in dairy cattle: Progress and challenges. J. Dairy Sci. 92: 433.

Hayes, B. J., P. J. Bowman, H. D. Daetwyler, J. W. Kijas and J. H. J. van der Werf. 2012. Accuracy of genotype imputation in sheep breeds. Anim. Genet. 43: 72-80.

Hayes, J. and W. Hill. 1981. Modification of estimates of parameters in the construction of genetic selection indices ('bending'). Biometrics 37: 483-493.

Hayman, R. H. 1953. Studies in fleece-rot of sheep. Aust. J. Agric. Res. 4: 430-468.

He, L., T. M. Gunn, D. M. Bouley, X.-Y. Lu, S. J. Watson, S. F. Schlossman, J. S. Duke-Cohan and G. S. Barsh. 2001. A biochemical function for attractin in agouti-induced pigmentation and obesity. Nat. Genet. 27: 40-47.

Heath, A. C. G., D. M. Bishop and J. D. Tenquist. 1987. The effects of artificially-induced fly-strike on food intake and liveweight gain in sheep. N. Z. Vet. J. 35: 50-52.

Heath, A. C. G. 1994. Ectoparasites of livestock in New Zealand. N. Z. J. Zool. 21: 23-38.

Heath, A. C. G. and D. M. Bishop. 1995. Flystrike in New Zealand. Surveillance 22: 11-13.

Heath, A. C. G. and C. Appleton. 2000. Small vertebrate carrion and its use by blowflies (Calliphoridae) causing ovine myiasis (flystrike) in New Zealand. N. Z. Entomol. 22: 81-87.

Heath, A. C. G. and D. M. Bishop. 2006. Flystrike in New Zealand: an overview based on a 16-year study, following the introduction and dispersal of the Australian sheep blowfly, *Lucilia cuprina* Wiedemann (Diptera: Calliphoridae). Vet. Parasitol. 137: 333-344.

Heinz-Erian, P., T. Müller, B. Krabichler, M. Schranz, C. Becker, F. Rüschendorf, P. Nürnberg, B. Rossier, M. Vujic, I. W. Booth, C. Holmberg, C. Wijmenga, G. Grigelioniene, C. M. F. Kneepkens, S. Rosipal, M. Mistrik, M. Kappler, L. Michaud, L.-C. Dóczy, V. M. Siu, M. Krantz, H. Zoller, G. Utermann and A. R. Janecke. 2009. Mutations in SPINT2 Cause a Syndromic Form of Congenital Sodium Diarrhea. Am. J. Hum. Genet. 84: 188-196.

Henderson, A. E. 1968. Growing Better Wool. A. H. & A. W. Reed, New Zealand. p 108.

Hess, D., J. J. Keusch, S. A. L. Oberstein, R. C. M. Hennekam and J. Hofsteenge. 2008. Peters plus syndrome is a new congenital disorder of glycosylation and involves defective O-glycosylation of thrombospondin type 1 repeats. J. Biol. Chem. 283: 7354-7360.

Hickey, J., B. Kinghorn, B. Tier, J. Wilson, N. Dunstan and J. van der Werf. 2011. A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. Genet. Sel. Evol. 43: 12.

Hill, W. and A. Caballero. 1992. Artificial selection experiments. Annu. Rev. Ecol. Syst. 23: 287-310.

Hirschhorn, J. and M. Daly. 2005. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6: 95-108.

Hogan, L. 2010. Fighting flystrike proactively. Farming Ahead 217: 58-69.

Hoggart, C. J., J. C. Whittaker, M. De Iorio and D. J. Balding. 2008. Simultaneous Analysis of All SNPs in Genome-Wide and Re-Sequencing Association Studies. PLoS Genet 4: e1000130.

Horiuchi, K., T. Kimura, T. Miyamoto, H. Takaishi, Y. Okada, Y. Toyama and C. P. Blobel. 2007. Cutting edge: TNF-α-converting enzyme (TACE/ADAM17) inactivation in mouse myeloid cells prevents lethality from endotoxin shock. J. Immunol. 179: 2686-2689.

Horton, B. J., D. J. Best, L. G. Butler and G. G. Gregory. 1997. Organophosphorus residues in wool grease resulting from specified on-farm lice and fly strike control treatments. Aust. Vet. J. 75: 500-503.

Horton, B. J. and L. Iles. 2007. Dag scores and mean crutching and shearing period for mulesed and unmulesed weaners. Int. J. Sheep Wool Sci. 55: 48-64.

Huisman, A. E. and D. J. Brown. 2008. Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in Merino sheep. 2. Genetic relationships between bodyweight traits and other traits. Aust. J. Exp. Agric. 48: 1186-1193.

Huisman, A. E., D. J. Brown, A. J. Ball and H. Graser. 2008. Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in Merino sheep. 1. Description of traits, model comparison, variance components and their ratios. Aust. J. Exp. Agric. 48: 1177-1185.

Ibánẽz-Escriche, N., R. L. Fernando, A. Toosi and J. C. M. Dekkers. 2009. Genomic selection of purebreds for crossbred performance. Genet. Sel. Evol. 41: 12.

Ide, Y., D. Tsuchimoto, Y. Tominaga, M. Nakashima, T. Watanabe, K. Sakumi, M. Ohno and Y. Nakabeppu. 2004. Growth retardation and dyslymphopoiesis accompanied by G2/M arrest in APEX2-null mice. Blood 104: 4097-4103.

Inaba, M., A. Yawata, I. Koshino, K. Sato, M. Takeuchi, Y. Takakuwa, S. Manno, Y. Yawata, A. Kanzaki, J. Sakai, A. Ban, K. Ono and Y. Maede. 1996. Defective anion transport and marked spherocytosis with membrane instability caused by hereditary total deficiency of red cell band 3 in cattle due to a nonsense mutation. J. Clin. Invest. 97: 1804-1817.

Irrthum, A., K. Devriendt, D. Chitayat, G. Matthijs, C. Glade, P. M. Steijlen, J.-P. Fryns, M. A. M. Van Steensel and M. Vikkula. 2003. Mutations in the transcription factor gene SOX18 underlie recessive and dominant forms of hypotrichosis-lymphedema-telangiectasia. Am. J. Hum. Genet. 72: 1470-1478.

Jahoda, C. A. B., A. Kljuic, R. O'Shaughnessy, N. Crossley, C. J. Whitehouse, M. Robinson, A. J. Reynolds, M. Demarchez, R. M. Porter, L. Shapiro and A. M. Christiano. 2004. The lanceolate hair rat phenotype results from a missense mutation in a calcium coordinating site of the desmoglein 4 gene. Genomics 83: 747-756.

Johnson, D. L., A. L. Rae and J. N. Clarke. 1989. Technical aspects of the Animalplan system. p 197-202 in Proc. N. Z. Soc. Anim. Prod., Palmerston North, New Zealand.

Johnston, S. E., J. McEwan, N. K. Pickering, J. W. Kijas, D. Beraldi, J. G. Pilkington, J. M. Pemberton and J. Slate. 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Mol. Ecol. 20: 2555-2566.

Jopson, N. B., K. G. Dodds, K. J. Knowler, R. Wheeler and J. C. McEwan. 2000. Lamb and ewe performance of East-Fresian x Coopworth relative to pure Coopworths. p 47-50 in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Kalinowski, S., M. Taper and T. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16: 1099-1106.

Kaneko, M., S. Nishihara, H. Narimatsu and N. Saitou. 2001. The evolutionary history of glycosyltransferase genes. Trends Glycosci. Glyc. 13: 147-156.

Kawaguchi, T., L. Qin, T. Shimomura, J. Kondo, K. Matsumoto, K. Denda and N. Kitamura. 1997. Purification and cloning of hepatocyte growth factor activator inhibitor type 2, a kunitz-type serine protease inhibitor. J. Biol. Chem. 272: 27558-27564.

Kay, M. M., N. Flowers, J. Goodman and G. Bosman. 1989. Alteration in membrane protein band 3 associated with accelerated erythrocyte aging. PNAS 86: 5834-5838.

Kay, M. M., J. J. Marchalonis, J. Hughes, K. Watanabe and S. F. Schluter. 1990. Definition of a physiologic aging autoantigen by using synthetic peptides of membrane protein band 3: localization of the active antigenic sites. PNAS 87: 5734-5738.

Kee, H. J., J.-R. Kim, K.-I. Nam, H. Y. Park, S. Shin, J. C. Kim, Y. Shimono, M. Takahashi, M. H. Jeong, N. Kim, K. K. Kim and H. Kook. 2007. Enhancer of polycomb1, a novel homeodomain only protein-binding partner, induces skeletal muscle differentiation. J. Biol. Chem. 282: 7700-7709.

Kijas, J. 2010. The impact of selection on the sheep genome: findings from the ISGC HapMap experiment. p W132 in Plant & Animal Genomes XVIII, San Diego, CA, United States of America.

Kijas, J. W., D. Townley, B. P. Dalrymple, M. P. Heaton, J. F. Maddox, A. McGrath, P. Wilson, R. G. Ingersoll, R. McCulloch, S. McWilliam, D. Tang, J. McEwan, N. Cockett, V. H. Oddy, F. W. Nicholas, H. W. Raadsma and for the International Sheep Genomics Consortium. 2009. A genome wide survey of SNP variation revels the genetic structure of sheep breeds. PLoS One 4: e4668.

Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, S. Paiva, W. Barendse, E. Ciani, H. Raadsma, J. McEwan, B. Dalrymple and other members of the International Sheep Genomics Consortium. 2012a. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10: e1001258.

Kijas, J. W., J. E. Miller, T. Hadfield, R. McCulloch, E. Garcia-Gamez, L. R. Porto Neto and N. Cockett. 2012b. Tracking the Emergence of a New Breed Using 49,034 SNP in Sheep. PLoS One 7: e41508.

Kulessa, H., G. Turk and B. L. M. Hogan. 2000. Inhibition of Bmp signaling affects growth and differentiation in the anagen hair follicle. EMBO J. 19: 6664-6674.

Lander, E. and N. Schork. 1994. Genetic dissection of complex traits. Science 265: 2037-2048.

Lang, M. D., G. R. Allen and B. J. Horton. 2001. The utilisation of carcasses by *Lucilia cuprina* (Wiedemann) as a breeding site in Tasmania. p 71-78 in Proceedings of the FLICS Conference, Launceston, TAS, Australia.

Larsen, J. W. A., N. Anderson and A. L. Vizard. 1999. The pathogenesis and control of diarrhoea and breech soiling in adult Merino sheep. Int. J. Parasitol. 29: 893-902.

Leathwick, D. M. and D. S. Atkinson. 1995. Dagginess and flystrike in lambs grazed on *Lotus corniculatus* or ryegrass. p 196-198 in Proc. N. Z. Soc. Anim. Prod., Dunedin, New Zealand.

Leathwick, D. M. and D. S. Atkinson. 1998. Influence of different proportions of *Lotus corniculatus* in the diet of lambs on dags, flystrike and animal performance. Wool Technol. Sheep Breed. 46: 353-359.

Lee, C. and A. D. Fisher. 2007. Welfare consequences of mulesing of sheep. Aust. Vet. J. 85: 89-93.

Lee, S. H., J. H. J. van der Werf, B. J. Hayes, M. E. Goddard and P. M. Visscher. 2008. Predicting Unobserved Phenotypes for Complex Traits from Whole-Genome SNP Data. PLoS Genet 4: e1000231.

Lee, S. H., N. R. Wray, M. E. Goddard and P. M. Visscher. 2011. Estimating missing heritability for disease from genome-wide association studies. Am. J. Hum. Genet. 88: 294-305.

Levot, G. W. 1993. Insecticide resistance: new developments and future options for fly and lice control on sheep. Wool Technol. Sheep Breed. 41: 108-119.

Levot, G. W. 1995. Resistance and the control of sheep ectoparasites. Int. J. Parasitol. 25: 1355-1362.

Li, Y., A. Swan and I. Purvis. 1999. Genetic variation in resistance to fleece rot in CSIRO's fine wool flock. p 524-527 in Proc. Assoc. Advmt. Anim. Breed. Genet., Mandurah, WA, Australia.

Linnen, C. R., E. P. Kingsley, J. D. Jensen and H. E. Hoekstra. 2009. On the origin and spread of an adaptive allele in Deer mice. Science 325: 1095-1098.

Litherland, A. J., E. Sorenson, J. Niezen and D. Bishop. 1992. A pilot evaluation of shedding sheep breeds compared with non-shedding breeds for susceptibility to nematodes and flystrike. p 233-235 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Lorenz, A. J., S. Chao, F. G. Asoro, E. L. Heffner, T. Hayashi, H. Iwata, K. P. Smith, M. E. Sorrells and J.-L. Jannink. 2011. Chapter Two - Genomic Selection in Plant Breeding: Knowledge and Prospects. p 77-123 in Advances in Agronomy: 110, L. S. Donald, ed. Academic Press.

Lowe, M. E., J. L. Rosenblum and A. W. Strauss. 1989. Cloning and characterization of human pancreatic lipase cDNA. J. Biol. Chem. 264: 20042-20048.

Ludemann, C., P. Amer and J. Sise. 2010. Value propositions and trait prioritisation for the New Zealand sheep industry, AbacusBio Limted, Dunedin, New Zealand.

Luther, G., E. R. Wagner, G. Zhu, Q. Kang, Q. Luo, J. Lamplot, Y. Bi, X. Luo, J. Luo, C. Teven, Q. Shi, S. H. Kim, J. L. Gao, E. Huang, K. Yang, R. Rames, X. Liu, M. Li, N. Hu, H. Liu, Y. Su, L. Chen, B. C. He, G. W. Zuo, Z. L. Deng, R. R. Reid, H. H. Luu, R. C. Haydon and T. C. He. 2011. BMP-9 induced osteogenic differentiation of mesenchymal stem cells: molecular mechanism and therapeutic potential. Curr. Gene Ther. 11: 229-240.

MacDonald, P. A., J. C. McEwan, G. J. Goosen, K. G. Dodds, R. S. Green, R. W. Wheeler, K. J. Knowler, G. J. Greer and A. M. Crawford. 1998. A 'dagginess' QTL in sheep. p E031 in Proceedings of the XXVIth International Conference for Animal Genetics, Auckland, New Zealand.

Mackereth, V. J. 1983. Dags - cost thereof. p 231 in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Maillette de Buy Wenniger-Prick, L. J. J. M. and R. C. M. Hennekam. 2002. The Peters' plus syndrome: a review. Ann. Genet. 45: 97-103.

Maniatis, N. and G. E. Pollott. 2003. The impact of data structure on genetic (co)variance components of early growth in sheep, estimated using an animal model with maternal effects. J. Anim. Sci. 81: 101-108.

Marshall, F. and J. Hammond. 1914. On the effects of complete and incomplete castration upon horn growth in Herdwick sheep. J. Physiol. 48: 171-176.

Marshall, T., J. Slate, L. Kruuk and J. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639-655.

McEwan, J. C., G. H. Davis, K. G. Dodds, P. F. Fennessy, J. N. Clarke, G. D. Bruce and M. G. Hishon. 1992a. Comparison of growth, hogget fleece weight and reproductive traits of three sheep breeds from a flock selected for prolificacy and a

Romney control flock. p 211-215 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

McEwan, J. C., P. Mason, R. L. Baker, J. N. Clarke, S. M. Hickey and K. Turner. 1992b. Effect of selection for productive traits on internal parasite resistance in sheep. p 53-56 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

McEwan, J. C. 1994. Breeders Manual, NZ WormFEC^{TM} Service, NZ Pastoral Agricultural Research Institute Limited, Invermay, Mosgiel, New Zealand.

McEwan, J. C., K. G. Dodds, G. J. Greer, W. E. Bain, S. J. Duncan, R. Wheeler, K. J. Knowler, P. J. Reid, R. S. Green and P. G. C. Douch. 1995. Genetic estimates for parasite resistance traits in sheep and their correlations with production traits. 23rd New Zealand Society for Parasitology. In: N. Z. J. Zool. 22:177

McEwan, J. C. 2007. Developing genomic resources for whole genome selection. p 148-153 in Proc. N. Z. Soc. Anim. Prod., Wanaka, New Zealand.

McEwan, J. C. 2009. Skim sequencing the sheep genome, and SNP discovery, using Roche 454 FLX sequencing. p W081 in Plant & Animal Genomes XVII, San Diego.

McEwan, J. C., N. K. Pickering, K. G. Dodds, B. Auvray, P. L. Johnson, R. Tecofsky and T. Wilson. 2010. Putting the sheep SNP50 Beadchip to work: case studies in gene mapping and genomic selection. p W133 in Plant & Animal Genomics XVIII, San Diego.

McGuirk, B. J., K. D. Atkins, E. Kowal and K. Thornberry. 1978. Breeding for resistance to fleece rot and body strike - the Trangie programme. Wool Technol. Sheep Breed. 26: 17-24.

McGuirk, B. J. and K. D. Atkins. 1984. Fleece rot in Merino sheep. 1. The heritability of fleece rot in unselected flocks of medium-wool Peppin Merinos. Aust. J. Agric. Res. 35: 423-434.

McKenzie, J. A. and P. Batterham. 1998. Predicting insecticide resistance: mutagenesis, selection and response. Philos. Trans. R. Soc. Lond. B. 353: 1729-1734.

McLean, N. J., N. B. Jopson, A. W. Campbell, K. Knowler, M. Behrent, G. Cruickshank, C. M. Logan, P. D. Muir, T. Wilson and J. C. McEwan. 2006. An evaluation of sheep meat genetics in New Zealand: The central progeny test (CPT). p 368-372 in Proc. N. Z. Soc. Anim. Prod., Napier, New Zealand.

McLeod, R. S. 1995. Costs of major parasites to the Australian livestock industries. Int. J. Parasitol. 25: 1363-1367.

Meuwissen, T., B. Hayes and M. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819.

Meuwissen, T. 2007. Genomic selection: marker assisted selection on a genome wide scale. J. Anim. Breed. Genet. 124: 321-322.

Michaud, E. J., S. J. Bultman, M. L. Klebig, M. J. van Vugt, L. J. Stubbs, L. B. Russell and R. P. Woychik. 1994. A molecular model for the genetic and phenotypic characteristics of the mouse lethal yellow (Ay) mutation. PNAS 91: 2562-2566.

Miller, A. F., S. A. K. Harvey, R. S. Thies and M. S. Olson. 2000. Bone morphogenetic protein-9. J. Biol. Chem. 275: 17937-17945.

Miller, M. W., D. Duhl, B. M. Winkes, F. Arredondo-Vega, P. Saxon, G. Wolff, C. Epstein, M. Hershfield and G. Barsh. 1994. The mouse lethal nonagouti (a^x) mutation deletes the S-adenosylhomocysteine hydrolase (Ahcy) gene. EMBO J. 13: 1806.

Min, B. R., T. N. Barry, W. C. McNabb and P. D. Kemp. 1998. Effect of condensed tannins on the production of wool and on its processing characteristics in sheep grazing *Lotus corniculatus*. Aust. J. Agric. Res. 49: 597-605.

Minchenko, O., I. Opentanova and J. Caro. 2003. Hypoxic regulation of the 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase gene family (PFKFB-1–4) expression in vivo. FEBS Lett. 554: 264-270.

Misztal, I. 2009. BLUPF90 family of programs. Accessed 29 Apr 2012. http://nce.ads.uga.edu/~ignacy/newprograms.html.

Montgomery, G. and J. Sise. 1990. Extraction of DNA from sheep white blood cells. N. Z. J. Agric. Res. 33: 437-441.

Montgomery, G. W., H. M. Henry, K. G. Dodds, A. E. Beattie, T. Wuliji and A. M. Crawford. 1996. Mapping the Horns (Ho) Locus in Sheep: A Further Locus Controlling Horn Development in Domestic Animals. J. Hered. 87: 358-363.

Moradi, M. H., A. Nejati-Javaremi, M. Moradi-Shahrbabak, K. G. Dodds and J. C. McEwan. 2012. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC Genet. 13: 10.

Morgan, H. D., H. G. E. Sutherland, D. I. K. Martin and E. Whitelaw. 1999. Epigenetic inheritance at the agouti locus in the mouse. Nat. Genet. 23: 314-318.

Morris, C. A., A. Vlassoff, S. A. Bisset, R. L. Baker, C. J. West and A. P. Hurford. 1997. Responses of Romney sheep to selection for resistance or susceptibility to nematode infection. Anim. Sci. 64: 319-329.

Morris, C. A., A. W. Campbell, N. G. Cullen, G. H. Davies, J. M. Everett-Hincks, R. J. Hall, H. M. Henry, P. L. Johnson, J. C. McEwan, S. H. Phua and T. Wilson. 2007. Current status of QTL and association studies in New Zealand cattle, sheep and deer. p 153-159 in Proc. N. Z. Soc. Anim. Prod., Wanaka, New Zealand.

Mortimer, S. I., K. D. Atkins and H. W. Raadsma. 1998. Responses to selection for resistance and susceptibility to fleece rot and body strike in Merino sheep. p 27: 283-286 in Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, NSW, Australia.

Mortimer, S. I., D. L. Robinson, K. D. Atkins, F. D. Brien, A. A. Swan, P. J. Taylor and N. M. Fogarty. 2009. Genetic parameters for visually assessed traits and their relationships to wool production and liveweight in Australian Merino sheep. Anim. Prod. Sci. 49: 32-42.

Moser, G., R. Crump, B. Tier, J. Sölkner, K. Zenger, M. Khatkar, J. Cavanagh and H. Raadsma. 2007. Genome based genetic evaluation and genome wide selection using supervised dimension reduction based on partial least squares. p 227-230 in Proc. Assoc. Advmt. Anim. Breed. Genet., Armidale, NSW, Australia.

Mosteller, F., R. E. K. Rourke and G. B. Thomas. 1970. Probability with statistical applications. Second ed. Addison-Wesley Publishing Company, Reading, Massachusetts, USA. p 277.

Mrode, R. A. 2005. Linear models for the prediction of animal breeding values. 2nd ed. CAB international, Walingford, Oxfordshire, OX0 8DE, UK. p 51.

Munroe, T. and I. Evans. 2009. Tail length in lambs - the long and short of it. Farming Ahead 211: 88-89.

Nakamura, Y., G. Weidinger, J. O. Liang, A. Aquilina-Beck, K. Tamai, R. T. Moon and M. L. Warman. 2007. The CCN family member Wisp3, mutant in progressive pseudorheumatoid dysplasia, modulates BMP and Wnt signaling. J. Clin. Invest. 117: 3075-3086.

NIWA. 2010a. National Climate Summary - Summer 2009/10: Cold start, hot finish! Accessed 2 Nov 2011. www.niwa.co.nz/our-science/climate/publications/all/cs/seasonal/summer-2009-10.

NIWA. 2010b. National Climate Summary - Autumn 2010: Warm and wet South Island, sunny North Island. Accessed 2 Nov 2011. www.niwa.co.nz/our-science/climate/publications/all/cs/seasonal/autumn-2010.

NIWA. 2011a. National Climate Summary - Autumn 2011: Warm and wet for many regions. Accessed 2 Nov 2011. www.niwa.co.nz/our-science/climate/publications/all/cs/seasonal/autumn-2011.

NIWA. 2011b. National Climate Summary - Summer 2010/2011: Heat waves and deluges. Accessed 2 Nov 2011. www.niwa.co.nz/our-science/climate/publications/all/cs/seasonal/summer-2010-2011.

Norris, B. J., I. G. Colditz and T. J. Dixon. 2008. Fleece rot and dermatophilosis in sheep. Vet. Microbiol. 128: 217-230.

Oddy, H., B. Dalrymple, J. McEwan, J. Kijas, B. Hayes, J. van der Werf, D. Emery, P. Hynd, T. Longhurst, T. Fischer, D. Ferguson, R. Forage, N. Cockett and F. Nicholas. 2007. Sheep genomics and the international sheep genomics consortium. p 411-417 in Proc. Assoc. Advmt. Anim. Breed. Genet., Armidale, NSW, Australia.

Osornio-Vargas, A. R., P. M. Lindroos, P. G. Coin, A. Badgett, N. A. Hernandez-Rodriguez and J. C. Bonner. 1996. Maximal PDGF-induced lung fibroblast chemotaxis requires PDGF receptor-alpha. Am. J. Physiol. Lung Cell. Mol. Physiol. 271: L93-L99.

Overbeek, P. A., I. P. Gorlov, R. W. Sutherland, J. B. Houston, W. R. Harrison, H. L. Boettger-Tong, C. E. Bishop and A. I. Agoulnik. 2001. A transgenic insertion causing cryptorchidism in mice. genesis 30: 26-35.

Pei, Y.-F., L. Zhang, J. Li and H.-W. Deng. 2010. Analyses and comparison of imputation-based association methods. PLoS One 5: e10827.

Peng, C., A. Togayachi, Y. D. Kwon, C. Xie, G. Wu, X. Zou, T. Sato, H. Ito, K. Tachibana and T. Kubota. 2010. Identification of a novel human UDP-GalNAc transferase with unique catalytic activity and expression profile. Biochem. Biophys. Res. Commun. 402: 680-686.

Pennisi, D., J. Gardner, D. Chambers, B. Hosking, J. Peters, G. Muscat, C. Abbott and P. Koopman. 2000. Mutations in Sox18 underlie cardiovascular and hair follicle defects in ragged mice. Nat. Genet. 24: 434-437.

Perry, W. L., C. M. Hustad, D. A. Swing, T. N. O'Sullivan, N. A. Jenkins and N. G. Copeland. 1998. The itchy locus encodes a novel ubiquitin protein ligase that is disrupted in a18H mice. Nat. Genet. 18: 143-146.

Peschon, J. J., J. L. Slack, P. Reddy, K. L. Stocking, S. W. Sunnarborg, D. C. Lee, W. E. Russell, B. J. Castner, R. S. Johnson, J. N. Fitzner, R. W. Boyce, N. Nelson, C. J. Kozlosky, M. F. Wolfson, C. T. Rauch, D. P. Cerretti, R. J. Paxton, C. J. March and R. A. Black. 1998. An Essential Role for Ectodomain Shedding in Mammalian Development. Science 282: 1281-1284.

Phillips, C. J. C. 2009. A review of mulesing and other methods to control flystrike (cutaneous myiasis) in sheep. Anim. Welf. 18: 113-121.

Pickering, N. K., R. Tecofsky and J. C. McEwan. 2009a. Selection sweeps in sheep-preliminary results. p 16 in 5th Molecular Mapping Workshop, Mosgiel, New Zealand.

Pickering, N. K., P. L. Johnson, B. Auvray, K. G. Dodds and J. C. McEwan. 2009b. Mapping the horns locus in sheep. p 88-91 in Proc. Assoc. Advmt. Anim. Breed. Genet., Barrosa Valley, SA, Australia.

Pinares-Patiño, C., S. M. Hickey, E. A. Young, K. G. Dodds, S. MacLean, G. Molano, E. Sandoval, H. Kjestrup, R. Harland, N. K. Pickering and J. C. McEwan. 2013. Heritability estimates of methane emissions in sheep. Plant and Animal Genetics. In: P0622

Pinares-Patino, C. S., S. H. Ebrahimi, J. C. McEwan, K. G. Dodds, H. Clark and D. Luo. 2011. Is rumen retention time implicated in sheep differences in methane emission? p 219-222 in Proc. N. Z. Soc. Anim. Prod., Invercargill, New Zealand.

Pitts, K. M. and R. Wall. 2004. Adult mortality and oviposition rates in field and captive populations of the blowfly *Lucilia sericata*. Ecol. Entomol. 29: 727-734.

Plant, J. W., B. J. Horton, R. T. F. Armstrong and N. J. Campbell. 1999. Modelling pesticide residues on greasy wool: using organophosphate and synthetic pyrethroid survey data. Aust. J. Exp. Agric. 39: 9-19.

Pollott, G. E., L. J. E. Karlsson, S. J. Eady and J. Greeff. 2004. Genetic parameters for indicators of host resistance to parasites from weaning to hogget age in Merino sheep. J. Anim. Sci. 82: 2852-2864.

Pownall, D. B., R. J. Lucas, A. S. Familton, B. G. Love, S. E. Hines and L. R. Fletcher. 1993. The relationship between staggers and diarrhoea in lambs grazing

different components of endophyte-infected ryegrass. p 19-22 in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Priatel, J. J., X. Chen, L. A. Zenewicz, H. Shen, K. W. Harder, M. S. Horwitz and H.-S. Teh. 2007. Chronic immunodeficiency in mice lacking RasGRP1 results in CD4 T cell immune activation and exhaustion. J. Immunol. 179: 2143-2152.

Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. Sölkner, M. E. Goddard and B. J. Hayes. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. J. Dairy Sci. 94: 2625-2630.

Pryce, J. E. and H. D. Daetwyler. 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Anim. Prod. Sci. 52: 107-114.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. De Bakker and M. J. Daly. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559-575.

Purvis, I. and I. Franklin. 2005. Major genes and QTL influencing wool production and quality: a review. Genet. Sel. Evol. 37: 97-107.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statictical Computing, Vienna, Austria.

Raadsma, H. W. 1987. Flystrike control: an overview of management and breeding options. Wool Technol. Sheep Breed. 35: 174-185.

Raadsma, H. W. 1991a. Fleece rot and body strike in Merino sheep. V. Heritability of liability to body strike in weaner sheep under flywave conditions. Aust. J. Agric. Res. 42: 279-293.

Raadsma, H. W. 1991b. The susceptibility to body strike under high rainfall conditions of flocks selected for and against fleece rot. Aust. J. Exp. Agric. 31: 757-759.

Raadsma, H. W., R. M. Sandeman, A. B. Sasiak, C. R. Engwerda and T. J. O'Meara. 1992. Genetic improvement in resistance to body strike in Merino sheep: where are we at with indirect selection. p 143-146 in Proc. Assoc. Advmt. Anim. Breed. Genet., Rockhampton, QLD, Australia.

Raadsma, H. W. 1993. Fleece rot and body strike in Merino sheep. VI. Experimental evaluation of some physical fleece and body characteristics as indirect selection criteria for fleece rot. Aust. J. Agric. Res. 44: 915-931.

Raadsma, H. W. 2010. Linkage disequilibrium in the sheep genome: findings from the ISGC HapMap Initiative. p W134 in Plant & Animal Genomes XVIII, San Diego, CA, United States of America.

Radhakrishnan, Y., K. G. Hamil, S. Yenugu, S. L. Young, F. S. French and S. H. Hall. 2005. Identification, characterization, and evolution of a primate β-defensin gene cluster. Genes Immun. 6: 203-210.

Ramirez-Restrepo, C. A., T. N. Barry, N. Lopez-Villalobos, P. D. Kemp and W. C. McNabb. 2004. Use of *Lotus corniculatus* containing condensed tannins to increase lamb and wool production under commercial dryland farming conditions without the use of anthelmintics. Anim. Feed Sci. Technol. 117: 85-105.

Rathie, K. A., M. L. Tierney and J. C. Mulder. 1994. Assessing Wiltshire Horn-Merino crosses. 1. Wool shedding, blowfly strike and wool production traits. Aust. J. Exp. Agric. 34: 717-728.

Reeve, S., A. Carhan, C. T. Dee and K. G. Moffat. 2007. *slowmo* is required for *Drosophila* germline proliferation. genesis 45: 66-75.

Reid, T. C. and D. J. Cottle. 1999. Dag formation. p 52-54 in Proc. N. Z. Soc. Anim. Prod., Mosgiel, New Zealand.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution 43: 223-225.

Rickman, L., D. Šimrak, H. P. Stevens, D. M. Hunt, I. A. King, S. P. Bryant, R. A. J. Eady, I. M. Leigh, J. Arnemann, A. I. Magee, D. P. Kelsell and R. S. Buxton. 1999. N-terminal deletion in a desmosomal cadherin causes the autosomal dominant skin disease striate palmoplantar keratoderma. Hum. Mol. Genet. 8: 971-976.

Roberts, D. S. 1963. Barriers to *Dermatophilus dermatonomus* infection on the skin of sheep. Aust. J. Agric. Res. 14: 492-508.

Rosenkranz, S. and A. Kazlauskas. 1999. Evidence for distinct signaling properties and biological responses induced by the PDGF receptor α and β subtypes. Growth Factors 16: 201-216.

Rosenquist, T. A. and G. R. Martin. 1996. Fibroblast growth factor signalling in the hair growth cycle: expression of the fibroblast growth factor receptor and ligand genes in the murine hair follicle. Dev. Dyn. 205: 379-386.

Ruckebusch, Y. and J. Fioramonti. 1980. Motor profile of the ruminant colon: hard vs soft faeces production. Experientia 36: 1184-1185.

Rutberg, S. E., M. L. Kolpak, J. A. Gourley, G. Tan, J. P. Henry and D. Shander. 2006. Differences in expression of specific biomarkers distinguish human beard from scalp dermal papilla cells. J Invest Dermatol 126: 2583-2595.

Sabeti, P. C., S. F. Schaffner, B. Fry, J. Lohmueller, P. Varilly, O. Shamovsky, A. Palma, T. S. Mikkelsen, D. Altshuler and E. S. Lander. 2006. Positive natural selection in the human lineage. Science 312: 1614-1620.

Sabeti, P. C., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S. A. McCarroll, R. Gaudet, S. F. Schaffner and E. S. Lander. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913-918.

Safari, E., N. M. Fogarty and A. R. Gilmour. 2005. A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. Livest. Prod. Sci. 92: 271-289.

Sales, K. U., J. P. Hobson, R. Wagenaar-Miller, R. Szabo, A. L. Rasmussen, A. Bey, M. F. Shah, A. A. Molinolo and T. H. Bugge. 2011. Expression and genetic loss of function analysis of the HAT/DESC cluster proteases TMPRSS11A and HAT. PLoS One 6: e23261.

Sandeman, R. M., C. A. Dowse and P. R. Carnegie. 1985. Initial characterisation of the sheep immune response to infections of *Lucilia cuprina*. Int. J. Parasitol. 15: 181-185.

Sandeman, R. M., B. J. Collins and P. R. Carnegie. 1987. A scanning electron microscope study of *L. cuprina* larvae and the development of blowfly strike in sheep. Int. J. Parasitol. 17: 759-765.

Santamaría, I., G. Velasco, A. M. Pendás, A. Fueyo and C. López-Otín. 1998. Cathepsin Z, a novel human cysteine proteinase with a short propeptide domain and a unique chromosomal location. J. Biol. Chem. 273: 16816-16823.

SAS. 2004. SAS/STAT 9.1 User's Guide. SAS Institute Inc., Cary, North Carolina, USA. p 5121.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123: 218-223.

Scholtz, A. J., S. W. P. Cloete, J. B. van Wyk, I. Misztal, E. du Toit and T. C. K. van der Linde. 2010. Genetic (co)variances between wrinkle score and absence of breech strike in mulesed and unmulesed Merino sheep, using a threshold model. Anim. Prod. Sci. 50: 210-218.

Scobie, D. R., A. R. Bray and D. O'Connell. 1997. The ethically improved sheep concept. p 84-87 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Scobie, D. R., A. R. Bray and D. O'Connell. 1999. A breeding goal to improve the welfare of sheep. Anim. Welf. 8: 391-406.

Scobie, D. R. and D. O'Connell. 2002. Genetic reduction of tail length in New Zealand sheep. p 195-198 in Proc. N. Z. Soc. Anim. Prod., Palmerston North, New Zealand.

Scobie, D. R., D. O'Connell, A. R. Bray and P. Cunningham. 2002. Breech strike can be reduced by increased area of naturally bare skin around the perineum of lambs. p 201-204 in Proc. Aus. Soc. Anim. Prod.

Scobie, D. R., D. O'Connell, C. A. Morris and S. M. Hickey. 2007. A preliminary genetic analysis of breech and tail traits with the aim of improving the welfare of sheep. Aust. J. Agric. Res. 58: 161-167.

Scobie, D. R., D. O'Connell, C. A. Morris and S. M. Hickey. 2008. Dag score is negatively correlated with breech bareness score of sheep. Aust. J. Exp. Agric. 48: 999-1003.

Scobie, D. R. and D. O'Connell. 2010. Breech bareness reduces flystrike in New Zealand crossbred sheep. Anim. Prod. Sci. 50: 599-602.

Seaton, D. S. and R. M. Sandeman. 1991. The humoral antibody response of the sheep to repeated infections with *Lucilia cuprina*. p 18-26. 17 ref. in Immunological control of blowfly strike R. M. Sandeman, J. H. Arundel and E. A. Scheurmann, eds. Australian Wool Corporation, Parkville, VIC, Australia.

Sise, J. A. and P. R. Amer. 2009. SNP predictors to accelerate the rate of genetic progress in sheep. p 220-223 in Proc. Assoc. Advmt. Anim. Breed. Genet., Barrosa Valley, SA, Australia.

Sise, J. A., E. M. Linscott and P. R. Amer. 2011. Predicting the benefits of two stage genomic selection in young breeding rams. p 266-269 in Proc. N. Z. Soc. Anim. Prod., Invercargill, New Zealand.

Sladek, R., G. Rocheleau, J. Rung, C. Dina, L. Shen, D. Serre, P. Boutin, D. Vincent, A. Belisle, S. Hadjadj, B. Balkau, B. Heude, G. Charpentier, T. J. Hudson, A. Montpetit, A. V. Pshezhetsky, M. Prentki, B. I. Posner, D. J. Balding, D. Meyre, C. Polychronakos and P. Froguel. 2007. A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 445: 881-885.

Smith, J. L., H. G. Brewer and T. Dyall. 2009. Heritability and phenotypic correlations for breech strike and breech strike resistance indicators in Merinos. p 334-337 in Proc. Assoc. Advmt. Anim. Breed. Genet., Barossa Valley, SA, Australia.

Smith, W., Y. Li, A. Ingham, E. Collis, S. McWilliam, T. Dixon, B. Norris, S. Mortimer, R. Moore and A. Reverter. 2010. A genomics-informed, SNP association study reveals FBLN1 and FABP4 as contributing to resistance to fleece rot in Australian Merino sheep. BMC Vet. Res. 6: 27.

Sneddon, J. and B. Rollin. 2010. Mulesing and animal ethics. J. Agric. Environ. Ethics 23: 371-386.

Solberg, T. R., A. K. Sonesson, J. A. Woolliams and T. H. E. Meuwissen. 2009a. Reducing dimensionality for prediction of genome-wide breeding values. Genet. Sel. Evol. 41: 29.

Solberg, T. R., A. K. Sonesson, J. A. Woolliams, J. Ødegard and T. H. E. Meuwissen. 2009b. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. Genet. Sel. Evol. 41: 53.

Steel, R. G. D. and J. H. Torrie. 1980. Principles and procedures of statistics. A biometrical approach. Second ed. McGraw-Hill, New York. p 279.

Steiner, E. C., S. A. Harrington, K. V. Watson and N. Monzu. 1994. A technique for initiating cutaneous myiasis (flystrike) on sheep. Bull. Entomol. Res. 84: 115-118.

Stewart, J. B. and D. J. Garrick. 1996. Structural characteristics of the sheepbreeding industry and implications for genetic and reproduction technologies. p in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Sumi, S., A. Marinaki, M. Arenas, L. Fairbanks, M. Shobowale-Bakre, D. Rees, S. Thein, A. Ansari, J. Sanderson, R. De Abreu, A. Simmonds and J. Duley. 2002. Genetic basis of inosine triphosphate pyrophosphohydrolase deficiency. Hum. Genet. 111: 360-367.

Sumner, R. M. W., J. N. Clark and N. G. Cullen. 2007. Effect of divergent selection for wool bulk on live weight and wool characteristics in Perendale sheep. p 180-186 in Proc. N. Z. Soc. Anim. Prod., Wanaka, New Zealand.

Tait, A., R. Henderson, R. Turner and X. Zheng. 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. Int. J. Climatol. 26: 2097-2115.

Takao, M., J. Hino, N. Takeshita, Y. Konno, T. Nishizawa, H. Matsuo and K. Kangawa. 1996. Identification of rat bone morphogenetic protein-3b (BMP-3b), a new member of BMP-3. Biochem. Biophys. Res. Commun. 219: 656-662.

Tanaka, Y., Y. Kanai, Y. Okada, S. Nonaka, S. Takeda, A. Harada and N. Hirokawa. 1998. Targeted disruption of mouse conventional kinesin heavy chain *kif5B*, results in abnormal perinuclear clustering of mitochondria. Cell 93: 1147-1158.

Tang, W., T. M. Gunn, D. F. McLaughlin, G. S. Barsh, S. F. Schlossman and J. S. Duke-Cohan. 2000. Secreted and membrane attractin result from alternative splicing of the human ATRN gene. PNAS 97: 6025-6030.

Taylor, P. J., K. D. Atkins and A. R. Gilmour. 1999. Genetic association between fibre curvature, staple crimp and wool production and quality of Merino sheep. p 456-459 in Proc. Assoc. Advmt. Anim. Breed. Genet., Mandurah, WA, Australia.

Tellam, R. L. and V. M. Bowles. 1997. Control of blowfly strike in sheep: Current strategies and future prospects. Int. J. Parasitol. 27: 261-273.

The Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324: 528-532.

The International Sheep Genomics Consortium, A. L. Archibald, N. E. Cockett, B. P. Dalrymple, T. Faraut, J. W. Kijas, J. F. Maddox, J. C. McEwan, V. Hutton Oddy, H.

W. Raadsma, C. Wade, J. Wang, W. Wang and X. Xun. 2010. The sheep genome reference sequence: a work in progress. Anim. Genet. 41: 449-453.

Turner, H. N. and S. S. Y. Young. 1969. Quantitative Genetics in Sheep Breeding. Macmillian, Australia. p 130.

Valles, S. M. and P. G. Koehler. 1997. Insecticides used in the urban environment: mode of action. p ENY282 in SP134: Pests in and around the Florida Home P. G. Koehler, D. E. Short and W. H. Kern, eds. University of Florida, Gainesville, FL, United States of America.

van der Werf, J. H. J. 2006a. Teaching software used for quantitative genetics at UNE: Multiple trait selection index. Accessed 4 Apr 2012. http://www-personal.une.edu.au/~jvanderw/software.htm.

van der Werf, J. H. J. 2006b. Teaching software used for quantitative genetics at UNE: Single trait selection index. Accessed 4 Apr 2012. http://www-personal.une.edu.au/~jvanderw/software.htm.

van der Werf, J. H. J. 2009. Potential benefit of genomic selection in sheep. p 38-41 in Proc. Assoc. Advmt. Anim. Breed. Genet., Barrosa Valley, SA, Australia.

van Heeswijck, R. and G. McDonald. 1992. *Acremonium* endophytes in perennial ryegrass and other pasture grasses in Australia and New Zealand. Aust. J. Agric. Res. 43: 1683-1709.

Van Tassell, C., T. Smith, L. Matukumalli, J. Taylor, R. Schnabel, C. Lawley, C. Haudenschild, S. Moore, W. Warren and T. Sonstegard. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat. Meth. 5: 247-252.

van Vuren, D. and P. W. Hedrick. 1989. Genetic conservation in feral populations of livestock. Conserv. Biol. 3: 312-317.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414-4423.

Veerkamp, R., S. Brotherstone, B. Engel and T. Meuwissen. 2001. Analysis of censored survival data using random regression models. Anim. Sci. 72: 1-10.

Verbyla, K. L., B. Hayes, P. J. Bowman and M. E. Goddard. 2009. Short Note: Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. Genet. Res 91: 307-311.

Vlassoff, A., D. M. Leathwick and A. C. G. Heath. 2001. The epidemiology of nematode infections of sheep. N. Z. Vet. J. 49: 213-221.

Voight, B. F., S. Kudaravalli, X. Wen and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4: e72.

Voytyuk, O., J. Lennartsson, A. Mogi, G. Caruana, S. Courtneidge, L. K. Ashman and L. Rönnstrand. 2003. Src family kinases are involved in the differential signaling from two splice forms of c-Kit. J. Biol. Chem. 278: 9159-9166.

Walker, G. and M. J. Young. 2009. Bare Points Sheep - SIL Technical Note. Accessed 8 Feb 2011. http://www.sheepimprovement.co.nz/getdoc/a9bb121b-2016-4b34-a163-399ae4b28471/Doc-ID-000010-GW-Bare-Points-Sheep.aspx.

Wall, R., N. P. French and A. Fenton. 2000. Sheep blowfly strike: a model approach. Res. Vet. Sci. 69: 1-9.

Wall, R., K. M. Pitts and K. E. Smith. 2001. Pre-adult mortality in the blowfly *Lucilia sericata*. Med. Vet. Entomol. 15: 328-334.

Wardhaugh, K. G. and R. Morton. 1990. The incidence of flystrike in sheep in relation to weather conditions, sheep husbandry, and the abundance of the Australian sheep blowfly, *Lucilia cuprina* (Wiedemann)(Diptera: Calliphoridae). Aust. J. Agric. Res. 41: 1155-1167.

Wardhaugh, K. G. 2001. The biology and ecology of the Australian sheep blowfly, *Lucilia cuprina* (Wiedemann) - an update. p 53-64 in Proceedings of the FLICS Conference, Launceston, TAS, Australia.

Wardhaugh, K. G., R. J. Mahon and D. Bedo. 2001. Factors affecting the incidence of flystrike in sheep - a description and analysis of data from three separate areas in eastern Australia. p 80-87 in Proceedings of the FLICS Conference, Launceston, TAS, Australia.

Watson, T. G., R. L. Baker and T. G. Harvey. 1986. Genetic variation in resistance or tolerance to internal nematode parasites in strains of sheep at Rotomahana. p 23-26 in Proc. N. Z. Soc. Anim. Prod., Dunedin, New Zealand.

Watts, J. E., M. D. Murray and N. P. H. Graham. 1979. The blowfly strike problem of sheep in New South Wales. Aust. Vet. J. 55: 325-334.

Wellcome Trust Sanger Institute. 2012. DNA requirements for sequencing. Accessed 3 October 2012.

http://www.sanger.ac.uk/resources/technologies/dna_requirements_for_sequencing_june2010.pdf.

Wheeler, T., K. Hood, N. Maqbool, J. McEwan, C. Bingle and S. Zhao. 2007. Expansion of the Bactericidal/Permeability Increasing-like (BPI-like) protein locus in cattle. BMC Genomics 8: 75.

Wilson, J. A. and A. C. G. Heath. 1994. Resistance to two organophosphorus insecticides in New Zealand populations of the Australian sheep blowfly, Lucilia cuprina. Med. Vet. Entomol. 8: 231-237.

Wollberg, P., J. Lennartsson, E. Gottfridsson, A. Yoshimura and L. Rönnstrand. 2003. The adapter protein APS associates with the multifunctional docking sites Tyr-568 and Tyr-936 in c-Kit. Biochem. J. 370: 1033-1038.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97-159.

Wright, S. 1943. Isolation by distance. Genetics 28: 114.

Wright, S. 1965. The interpretation of population structure by F-Statistics with special regard to systems of mating. Evolution 19: 395-420.

Wuliji, T., K. G. Dodds, R. N. Andrews, P. R. Turner and R. Wheeler. 1998. Responses to fleece weight selection and heritability estimates of wool characteristics in Romney sheep. p 24: 55-58 in Proc. 6th World Congr. Genet. Appl. Livest. Prod., Armidale, NSW, Australia.

Wuliji, T., K. G. Dodds, R. N. Andrews and P. R. Turner. 2011. Selection response to fleece weight, wool characteristics, and heritability estimates in yearling Romney sheep. Livest. Sci. 135: 26-31.

Yamaoka, K., K.-i. Masuda, H. Ogawa, K.-i. Takagi, N. Umemoto and S. Yasuoka. 1998. Cloning and characterization of the cDNA for human airway trypsin-like protease. J. Biol. Chem. 273: 11895-11901.

Yang, D., A. Biragyn, D. M. Hoover, J. Lubkowski and J. J. Oppenheim. 2004. Multiple Roles of Antimicrobial Defensins, Cathelicidins, and Eosinophil-Derived Neurotoxin in Host Defense. Annu. Rev. Immunol. 22: 181-215.

Yang, R.-B., C. K. D. Ng, S. M. Wasserman, S. D. Colman, S. Shenoy, F. Mehraban, L. G. Kömüves, J. E. Tomlinson and J. N. Topper. 2002. Identification of a Novel Family of Cell-surface Proteins Expressed in Human Vascular Endothelium. J. Biol. Chem. 277: 46364-46373.

Yang, S.-Y., X.-Y. He and H. Schulz. 2005. Multiple functions of type 10 17β-hydroxysteroid dehydrogenase. Trends Endocrinol. Metab. 16: 167-175.

Yasuoka, S., T. Ohnishi, S. Kawano, S. Tsuchihashi, M. Ogawara, K. Masuda, K. Yamaoka, M. Takahashi and T. Sano. 1997. Purification, characterization, and localization of a novel trypsin-like protease found in the human airway. Am. J. Respir. Cell Mol. Biol. 16: 300-308.

Yeager, M., N. Orr, R. B. Hayes, K. B. Jacobs, P. Kraft, S. Wacholder, M. J. Minichiello, P. Fearnhead, K. Yu, N. Chatterjee, Z. Wang, R. Welch, B. J. Staats, E. E. Calle, H. S. Feigelson, M. J. Thun, C. Rodriguez, D. Albanes, J. Virtamo, S. Weinstein, F. R. Schumacher, E. Giovannucci, W. C. Willett, G. Cancel-Tassin, O. Cussenot, A. Valeri, G. L. Andriole, E. P. Gelmann, M. Tucker, D. S. Gerhard, J. F. Fraumeni, R. Hoover, D. J. Hunter, S. J. Chanock and G. Thomas. 2007. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat. Genet. 39: 645-649.

Young, E. A., J. W. Kijas, R. McCulloch, D. R. Scobie, K. M. McRae, N. K. Pickering, K. G. Dodds and J. C. McEwan. 2011. Brief communication: Arapawa; a novel New Zealand sheep breed of distinct origin. p 248-250 in Proc. N. Z. Soc. Anim. Prod., Invercargill, New Zealand.

Young, M. J. and S.-A. N. Newman. 2009. SIL-ACE- Increasing access to genetic information for sheep farmers. p 153-154 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Young, M. J. and R. Wakelin. 2009. SIL-progress over a decade. p 148-152 in Proc. N. Z. Soc. Anim. Prod., Lincoln, New Zealand.

Young, S. R., D. R. Scobie, C. A. Morris and N. C. Amyes. 2004. Flystrike in lambs selected for resistance or susceptibility to ryegrass staggers, when challenged with high or low endophyte pastures. p 297-299 in Proc. N. Z. Soc. Anim. Prod., Hamilton, New Zealand.

Yu, Z., J. E. Wildermoth, O. A. M. Wallace, S. W. Gordon, N. J. Maqbool, P. H. Maclean, A. J. Nixon and A. J. Pearson. 2011. Annotation of sheep keratin intermediate filament genes and their patterns of expression. Exp. Dermatol. 20: 528-588.

Appendices

Appendix 1

Flystrike Questionnaire

Owner/Farm:

Farm Facts:

Farm size:_____

Stocking rate:

	November	*January*	*March*
Ewes			
Hoggets			
Rams			
Lambs			
Cattle			

Pasture Type:_____

_____

Docking Technique:_____

_____

Current Flystrike Control Strategy;

Treatment: Dip, Spray, Powder, Crutch, Shear, Cull, Other_____

Products Used:_____

_____

Frequency:_____

Treat: Everything, Individuals, Other_____

Other Strategies:_____

_____

Farm Practices: _____

_____

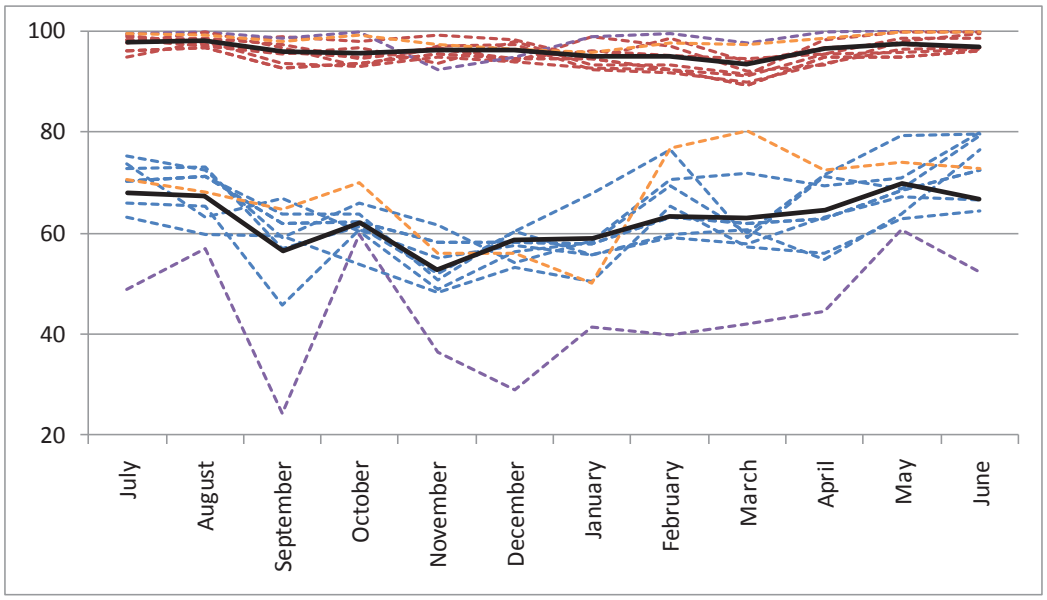
_____

_____

_____

_____

Appendix 2



Minimum (blue) and maximum (red) relative humidity (RH, %) for the 2009/10 season (July2009 to June 2010). The average minimum and maximum (RH) across all farms is in black. Highlight is the farmer with the least RH (purple) and highest RH (orange) for each measurement.

Appendix 3

High-throughput tissue extraction SOP method for ear and spleen samples.

This is the protocol for extracting DNA from a 3mm punch of sheep ear tissue, 3-4 mm^2 pieces of ear and spleen samples. The extraction is carried out in 96 well storage plates and produces 200µl of DNA at a concentration of 50-100/µl (average~100ng/µl).

Solutions (Refer to AGR-GNZ-DNA05 Solutions)

- TNE Buffer
- 1M Tris Buffered Saline (TBS)
- 5 M NaCl
- TE Buffer (Tris-EDTA) pH 8
- Proteinase K
- 95% Ethanol
- 75% Ethanol

Working Solutions

- TNES-PK Buffer
    - TNES 200ml
    - Proteinase K (20mg/ml) 6ml

Mix well in 250ml bottle

Note: TNES may become cloudy in cool temperatures. Place bottle in warm water to dissolve solids, and mix – don't shake - before measuring to use

Transfer Punches to 96-round well U bottom Plate

1. Label a 96 well round well U bottom 1.2ml plate with PCR plate number that has just been created, and mark control wells.
2. Set up work area comfortably with ethanol bottle, piercing tool, lid tool rack of samples and printout of plate. Have a separate rack labelled with punched PCR plate # for samples that have been punched.
3. Starting at A1, pick up sampler and check that the Sample Identification (SampleID) barcode matches the location on plate printout.

4. Open Caisley or Tpifix^{TM} container with side cutters. Use tweezers to extract tissue from sampler and put into the correct well. For ear and spleen pieces, cut a 3-4 mm^2 piece of tissue and place into correct well.
5. Work across the plate, transferring punches into matching well locations checking that the SampleID barcode matches the plate printout. Tick off each sample as you go.
6. If any punches stick to the sides of the wells, push down with a clean pipette tip.
7. If any samples do not have an ear tissue punch circle the well in the plate and highlight on the sheet and continue.
8. Any issues or comment regarding the samples to be written on PCR Plate printed sheet

Digest

1. Take out of the freezer the next 4 plates for extraction.
2. Ensure tissue at or close to the bottom of the well.
3. Turn on incubator (New Brunswick Incubator Shaker Series 25D) and heat to 55^{o}C before use.
4. Add the steel balls supplied to each well of the storage plate using the grinding ball dispenser.
5. Use the Biomek NX^{P} Laboratory Automation Workstation to add 500µl TNES-PK Buffer to each well
   a. Start Robot
   b. Turn on compressed air - remember to turn off after use
   c. Prime wash station, check levels in water and waste carboys
   d. Double-click on the 'Biomek Software' icon on the desktop
   e. <Open Folder>icon
   f. Select BiomekNX MC\Methods\Genomnz\TissueExtn_Plate\TNES-PK Addition for 4 plates
   g. <OK>
   h. <Instrument Setup> shows how the deck should be set up
   i. To identify the appropriate tips, reservoir and titerplates to use, hover the cursor over each item. The item's name will appear and can be cross-checked against the Biomek® NX^{P} labware sheets on the wall.
   j. Check location of tip box, that it is within the spikes

k. Add 6ml Proteinase K to 200ml TNES Buffer, mix gently and pour all the solution into the reservoir
l. <Run>, a dialog box will appear asking for confirmation that the deck has been set up appropriately, <OK> if deck matches layout illustrated
m. Once program is complete, turn off compressed air, retain tips for later use

6. Seal plates with round well sealing mat
7. Shake vigorously on Eppendorf MixMate at 1400rpm for 1-2 minutes
8. Centrifuge plates at 1000rpm for 1 minute. Check plates are well sealed.
9. Incubate plates at 55°C with gentle shaking (120rpm) for a minimum of 12 hours.

Salting

1. Remove extraction plates from the incubator
2. Centrifuge plates at 1000rpm for 1 minute
3. Use the Biomek NX^P to transfer 140µl of 5M NaCl
   a. Start Biomek NX^P and BiomekNX MC\Methods\Genomnz\Tissue Extn Plate\2 Salt Addition for 4 plates
   b. Turn on compressed air
   c. <Instrument Setup> , check plates, tips and reservoir are in correct positions
   d. Add 60ml 5M NaCl to the reservoir
   e. Re-use tips from TNES-PK addition
   f. Remove sealing mats from extraction plates, keep to re-seal plates
   g. <Run> <OK>
   h. Once program is complete, turn off compressed air
4. Re-seal plates and shake plates by hand vigorously for 30 seconds
5. Centrifuge plates at 4000rpm for 30 minutes at room temperature. A pellet of tissue debris and protein should form at the bottom of the wells.
6. Place a covered beaker containing 160ml of 96% ethanol in the freezer.
7. Use the Biomek NX^P to transfer 400µl of each supernatant to a 2.2ml square V bottom storage plate
   a. Label a 2.2ml square V bottom storage plate with PCR plate number
   b. Start Biomek NX^P and BiomekNX MC\Methods\Genomnz\Tissue Extn Plate\3 Supernatant Transfer for 4 plates
   c. Turn on compressed air

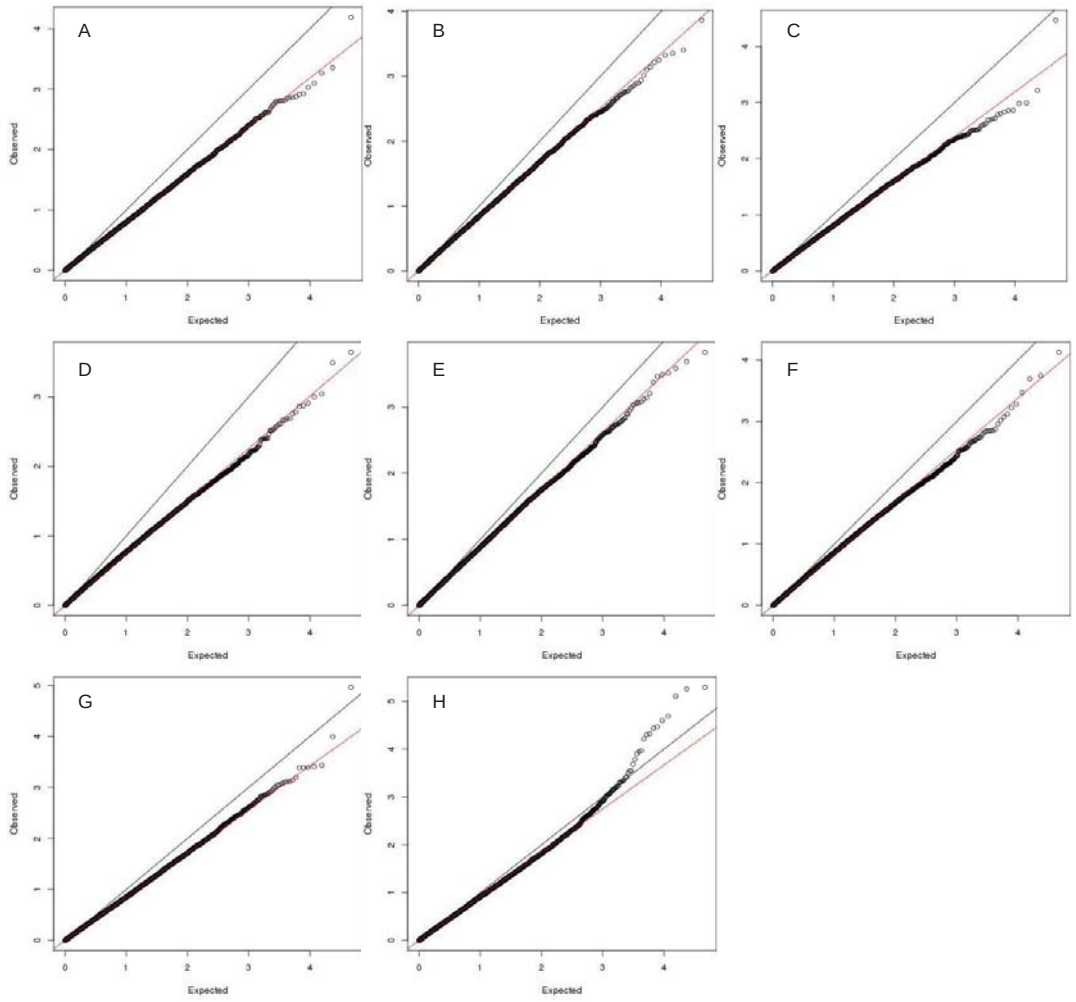
   d. <Instrument Setup> , this program uses the gripper tool so ensure the plates, tips and reservoir are in correct positions
   e. Use sterile, barrier tips for this transfer
   f. Remove sealing mats from extraction plates, keep to re-seal plates after transfer
   g. <Run> <OK>
   h. Check that the pellet is not being sucked up
   i. Supernatant should be clear but may be pale yellow to brown
8. Re-seal the original extraction plate containing residual supernatant and take back to Freezer. Keep until samples have been quantified.
9. Precipitate DNA
10. Use the Biomek NX^P to add 1 x volume (400µl) of cold 96% ethanol to the 2.2ml supernatant plate
11. Start Biomek NX^P and BiomekNX MC\Methods\Genomnz\TissueExtn_ Plate\96% EtOH Addition for 4 plates
   a. <Instrument Setup> , check plates, tips and reservoir are in correct positions
   b. Add 160ml of cold 96% ethanol to the reservoir
   c. <Run> <OK>
   d. Once program is complete, turn off compressed air, retain tips for later use
12. Seal with a square well sealing mat and mix by inverting several times
13. Stand plates in freezer for 30 minutes; set centrifuge to 4°C.
14. Centrifuge plates at 4000rpm for 30 minutes at 4°C. A white to yellow pellet should be visible at the bottom of the wells.
15. Remove seal, drain supernatant into a plastic container and blot over a stack of paper towels.
16. Use the Biomek NX^P to add 900ul of freshly prepared 70% ethanol to the pellet
   a. Start Biomek NX^P and BiomekNX MC\Methods\Genomnz\TissueExtn _Plate\ 70% EtOH Addition for 4 plates
   b. Turn on compressed air
   c. <Instrument Setup> , check plates, tips and reservoirs are in correct positions
   d. Re-use tips from 96% ethanol addition
   e. Add 180ml of 70% ethanol to each reservoir (2 reservoirs are needed)
   f. <Run> <OK>
   g. Once program is complete, turn off compressed air, retain tips for later use

17. Re-seal plates and invert several times
18. Leave plates to stand for a minimum of 1 hour at room temperature
19. Centrifuge plate at 4000rpm for 15 minutes at room temperature
20. Remove seal, drain supernatant as before and blot on paper towels
21. Air-dry the plate and seal on the bench until all the ethanol has evaporated (or incubate at 37°C for 40 minutes).

Dissolve DNA

1. Use the Biomek NX^P to add 100µl TE buffer
2. Start Biomek NX^P and BiomekNX MC\Methods\Genomnz\TissueExtn_ Plate\TE Buffer Addition for 4 plates
   a. Turn on compressed air
   b. <Instrument Setup> , check plates, tips and reservoir are in correct positions
   c. Add 50ml TE buffer to the reservoir
   d. <Run> <OK>
   e. Once program is complete, turn off compressed air
3. Re-seal plates and mix on MixMate at 1500rpm for 1-2 minutes to dislodge pellet from bottom of well
4. Incubate at 37°C for 1 hour
5. Mix plates on MixMate set at 1000rpm for 1 minute
6. Incubate plates overnight on benchtop rocker
7. Mix plate on MixMate set at 1000rpm for 1 minute
8. Put plate of extracted DNA in the freezer.

Appendix 4



QQ plots from case-control GWAS for breech bareness (A), wool bulk (B), wool length (C), mean fibre diameter (D, MFD), standard deviation of MFD (E), coefficient of variation of MFD (F), curvature (G) and proportion medullation (H) $-\log_{10}(P)$ values. The 0-1 line is in black and the slope in red.

Appendix 5

List of regions for each trait that reached nominal significance threshold *P*<0.001, for the case-control GWAS, by chromosome (ch), ordered on ovine genome v1.0.

Ch	SNP name	100kbp region	*P* value	Gene	OMIM	Trait
1	s62307	5159866 - 5359865	0.00037	SH3BP4	605611	CURV
	OAR1_71047743	70947743 - 71147742	0.00072	LRRC8D	612890	MED%
	s36857	74985887 - 75185886	2.1E-05	DNTTIP2	611199	MED%
				GCLM	601176	
				PCBP1	601209	
				ABCA4	601691	
	OAR1_107634967	107534967 - 107734966	0.00076	PSMB4	602177	CURV
				POGZ		
				CGN	609473	
				TUFT1	600087	
	OAR1_108852238	108752238 - 108952237	0.00061	LCE3E	612617	Dag score
				LCE1B	612604	
				LCE1E	612607	
				LCE2C	612611	
				LCE1D	612606	
				LCE2A	612609	
				LCE4A	612618	
				LCE1C	612605	
				LCE1F	612608	
	s33242	119235876 - 119435875	0.00091	NTF4	162662	MFD
	OAR1_121168723	121068723 - 121268722	0.00090	UHMK1	608849	FDSD
				UAP1	602862	
				UAP1L1		
	s31436	132397619 - 132597618	0.00060	LOC150051		LENGTH
	OAR1_135442243	135342243 - 135542242	0.00096	BACH1	602751	MED%
	s58710	165342356 - 165542355	5.5E-06			MED%
	OAR1_167004318	166904318 - 167104317	0.00015			FDSD
	s57353	188506405 - 188706404	4.7E-05	PLCXD2		MED%
				PHLDB2	610298	
	s17535	191435368 - 191635367	0.00044	ZBTB20	606025	MED%
	OAR1_198827329	198727329 - 198927328	0.00039	STXBP5L	609381	BULK
				POLQ	604419	
	OAR1_220822979	220722979 - 220922978	0.00072	CCDC39	613798	BULK

	s51351	235338491 - 235538490	0.00093	ZBBX		BBREECH
2	s11260	9404300 - 9604299	0.00041	RNF183		CURV
				PRPF4	607795	
				CDC26		
				SLC31A1	603085	
				FKBP15		
				SLC31A2	603088	
	OAR2_28934548	28834548 - 29034547	0.00038	FAM20A	611062	Flystrike
	s00331	55038384 - 55238383	0.00099	PAX5	167414	CURV
	OAR2_88811262	88711262 - 88911261	0.00100	C9orf93		MFD
	OAR2_91422909	91322909 - 91522908	0.00090			MED%
	OAR2_228008151	216243787 - 216443786	0.00011	VWC2L		MED%
	s42564	239359013 - 239559012	0.00026	KIAA1486		FDSD
3	s43785	16926009 - 17126008	0.00025			Flystrike
	OAR3_19548373	19448373 - 19648372	0.00085	ID2B		CURV
				KIDINS220		
	s62291	20195292 - 20395291	0.00045	CPSF3	606029	Dag score
				ADAM17	603639	
				YWHAQ	609009	
				RPL36A	180469	
	s05258	47196797 - 47396796	0.00048	VPS54		MED%
				UGP2	191760	
				FLJ36848		
	OAR3_53590166	53490166 - 53690165	0.00063	CTNNA2	114025	FDSD
	OAR3_74732440	74632440 - 74832439	0.00039			CURV
	OAR3_140038094	139938094 - 140138093	0.00083	FGD6	613520	CURV
				VEZT		
	s41365	143049207 - 143249206	0.00031	KRT84	602766	FDSD
				KRT85	602767	
				KRT86	601928	
				KRT7	148059	
				KRT83	602765	
				KRT8	148060	
				KRT80	611161	
				KRT81	602153	
	OAR3_150573902	150473902 - 150673901	0.00032			FDSD
	OAR3_211858595	211758595 - 211958594	0.00023	RERGL		MFD

	OAR3_237829296	237729296 - 237929295	0.00010	ARHGAP8 PHF21B	609405	CURV
	s57836	238016985 - 238216984	1.1E-05	NUP50 C22orf9 UPK3A FAM118A	604646 611559 	CURV
4	s49979	81193293 - 81393292	0.00083	NACAD CCM2 SNORA9 MYO1G PURB H2AFV LOC392352	 607929 613445 608887 	MED%
5	s23633	3216947 - 3416946	0.00038	LOC100130093 ZNF354B		MED%
	OAR5_13675205	13575205 - 13775204	0.00042	OR7E5P SERINC1 OR7E24		FDSD
	OAR5_67014817	66914817 - 67114816	0.00083			Dag score
	s20756	68519566 - 68719565	0.00082	GRIA1	138248	Dag score
	s21747	84212456 - 84412455	0.00084	ODZ2	610119	FDSD
	OAR5_96299808	96299808 - 96299808	0.00060	ARRDC3	612464	FDCV
6	OAR6_7316696	7216696 - 7416695	0.00093			FDSD
	OAR6_27183957	27083957 - 27283956	0.00028	PPP3CA	114105	MED%
	OAR6_56978564	56878564 - 57078563	0.00044			BULK
	OAR6_80988051	80988051 - 80988051	7.6E-05			Dag score
	OAR6_90873545	90773545 - 90973544	0.00083	STAP1	604298	MED%
7	OAR7_32968123	32868123 - 33068122	0.00077	C15orf41		Flystrike
	OAR7_35146905	35046905 - 35246904	0.00060	RASGRP1	603962	Flystrike
	s53860	46807656 - 47007655	0.00018	PLEKHO2 PIF1	 610953	FDCV
	s33221	51144935 - 51344934	0.00087			FDSD
	OAR7_63848145	63748145 - 63948144	0.00044	FGF7 GALK2	148180 137028	BBREECH
	OAR7_66581251	66481251 - 66681250	3.4E-05			LENGTH
	OAR7_67222553	67122553 -	0.00096			FDCV

		67322552				
	OAR7_74984216	74884216 - 75084215	6.1E-05	DAAM1	606626	MED%
	OAR7_75022469	74922469 - 75122468	0.00058	DAAM1	606626	MED%
	OAR7_76625075	76525075 - 76725074	0.00069	MNAT1	602689	MED%
8	OAR8_2281075	2181075 - 2381074	0.00076	TMEM30A FILIP1	611028 607307	FDSD
	OAR8_27871953	27771953 - 27971952	0.00055	LAMA4 C6orf225 TUBE1 WISP3	600133 607345 603400	BBREECH
	s50983	32302380 - 32502379	0.00033	PDSS2 BEND3	610564	Flystrike
	OAR8_86919224	86819224 - 87019223	0.00069	C6orf35 LDHAL6B LOC121498		Dag score
	OAR8_89040570	88940570 - 89140569	0.00087	GATSL2 SOD2 WTAP	 147460 605442	FDSD
10	OAR10_13904809	13804809 - 14004808	0.00089	TSC22D1 LOC641467	607715	CURV
	OAR10_20099465	19999465 - 20199464	0.00087	RNASEH2B GUCY1B2	610326 603695	Dag score
	s61347	22013327 - 22213326	0.00083			Dag score
	s18338	51713524 - 51913523	0.00100	TBC1D4	612465	MED%
	OAR10_62928758	62828758 - 63028757	0.00098			Flystrike
	OAR10_69772232	69672232 - 69872231	0.00048			Dag score
	OAR10_71200463	71100463 - 71300462	0.00043	GPC6	604404	Dag score
11	s65959	6889605 - 7089604	0.00076	SCPEP1		FDCV
	s21591	37822707 - 38022706	0.00097	WFIKKN2 LUC7L3 ANKRD40	610895 609434	BULK
	s22354	50463464 - 50663463	0.00052	DDX42 FTSJ3 SMARCD2 GH1 CD79B SCN4A C17orf72	613369 601736 139250 147245 603967	FDCV
12	OAR12_13334889	13234889 -	0.00054	RGS21	612407	Flystrike

		13434888				
	s35163	53983096 - 54183095	0.00056	PRKCZ GABRD	176982 137163	BULK
	s20362	76454598 - 76654597	0.00064	DTL INTS7 LPGAT1	610617 611350 610473	CURV
	s67459	82580325 - 82780324	0.00087	LHX9	606066	MED%
13	s69760	6564015 - 6764014	0.00061			BULK
	OAR13_20578061	20478061 - 20678060	0.00049			MED%
	OAR13_27572962	27472962 - 27672961	0.00068	ENKUR THNSL1 LOC100128811 GPR158	611025 611260	MED%
	OAR13_29176068	29076068 - 29276067	4.9E-05			MED%
	OAR13_35499997	35399997 - 35599996	3.7E-05	CACNB2 NSUN6 EPC1	600003 610999	MED%
	OAR13_35686621	35586621 - 35786620	3.4E-05	EPC1 KIF5B	610999	MED%
			0.00021	EPC1 KIF5B	610999	FDSD
	OAR13_35759190	35659190 - 35859189	5.1E-06	EPC1 KIF5B	610999 602809	MED%
	OAR13_65478113	65378113 - 65578112	0.00096	DEFB119 DEFB121		Dag score
14	OAR14_3089361	2989361 - 3189360	0.00020			FDCV
	OAR14_49536713	49436713 - 49636712	0.00098	SIPA1L3 DPF1 PPP1R14A SPINT2 YIF1B	 601670 608153 605124	Dag score
	OAR14_65327934	65227934 - 65427933	0.00085	SUV420H2 TMEM150B BRSK1 HSPBP1 PPP6R1 SYT5 C19orf51 TNNI3 TNNT1 TNNT2	613198 609235 612939 610875 600782 191044 191041 191045	FDCV
15	s71120	45195724 -	0.00034	DENND5A		Dag score

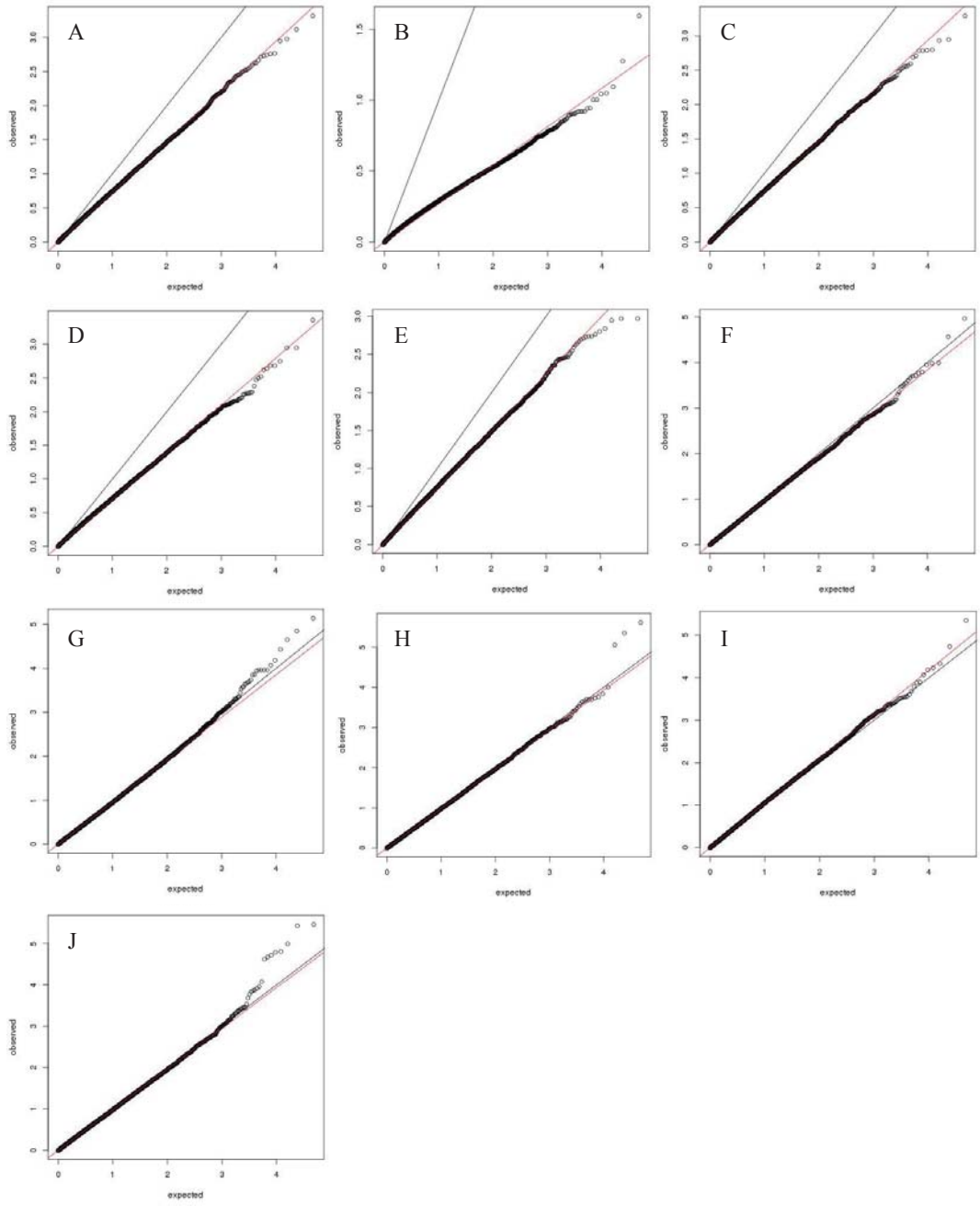
		45195724		SCUBE2	611747	
				NRIP3	613125	
	s20575	45870082 - 46070081	0.00011	LMO1	186921	Dag score
	OAR15_72588676	72488676 - 72688675	0.00014			BULK
	s50822	80917183 - 81117182	0.00078	AMBRA1	611359	MED%
				HARBI1		
	OAR15_81225977	81125977 - 81325976	0.00061	KIAA0652		MED%
				ARHGAP1	602732	
				ZNF408		
				THROMBIN	176930	
				CKAP5	611142	
16	s46992	4116284 - 4316283	7.4E-05	STK10	603919	FDCV
				UBTD2	610174	
				SH3PXD2B	613293	
	s32372	5337547 - 5537546	0.00063			Flystrike
	OAR16_60075688	59975688 - 60175687	0.00096			MED%
	OAR16_60178570	60078570 - 60278569	0.00039			MED%
17	OAR17_2428600	2328600 - 2528599	0.00063			MED%
19	OAR19_2928871	2828871 - 3028870	0.00080	ZCWPW2		BULK
	OAR19_5861149	5761149 - 5961148	0.00072			CURV
	s08710	12929558 - 13129557	0.00040	WDR48	612167	Dag score
				ST6GALNAC6	610135	
				SLC25A38	610819	
				RPSA	150370	
				MOBP	600948	
	s75645	52154944 - 52354943	0.00091	DOCK3	603123	CURV
				MAPKAPK3	602130	
				C3orf18		
	s18836	54093600 - 54293599	0.00073	ATRIP	606605	FDSD
				CCDC72		
				CCDC51		
				PLXNB1	601053	
20	OAR20_25523915	25423915 - 25623914	0.00032	PKHD1	606702	MFD
21	s60427	38084195 - 38284194	0.00080			BBREECH
	OAR21_39837614	39737614 - 39937613	0.00011			MED%
	OAR21_47542072	47442072 - 47642071	0.00050	MALAT1	607924	MED%
				SCYL1	607982	

				LTBP3	602090	
				SSSCA1	606044	
				EHBP1L1		
				KCNK7	603940	
				MAP3K11	600050	
22	OAR22_7858171	7758171 - 7958170	0.00034			FDCV
	OAR22_11944710	11844710 - 12044709	0.00041	RNLS	609360	CURV
				LIPJ	613921	
				LIPF	601980	
	OAR22_43010940	42910940 - 43110939	0.00078	MTUS1	609589	CURV
				NANOS1	608226	
				EIF3A	602039	
				FAM45B		
23	OAR23_6741945	6641945 - 6841944	0.00034			FDSD
	s03573	21320505 - 21520504	0.00016	KIAA1328		MED%
	OAR23_23803976	23703976 - 23903975	6.5E-05	DTNA	601239	BBREECH
	OAR23_27233801_X	27133802 - 27333801	0.00061	DSG4	607892	MED%
				DSG1	125670	
	OAR23_27272887	27172887 - 27372886	0.00087	DSG1	125670	MED%
				LOC100101938		
	OAR23_30351189	30251189 - 30451188	0.00012	CDH2	114020	MED%
				NDUFAF2	609653	
	OAR23_30585446	30485446 - 30685445	0.00041			CURV
	OAR23_32137172	32037172 - 32237171	0.00028	KCTD1	613420	MED%
	OAR23_48748723	48648723 - 48848722	0.00046	PSTPIP2		MED%
				ATP5A1	164360	
				LOC100286956		
				HAUS1	608775	
24	s45778	11046202 - 11246201	7.8E-06	LOC100293550		MED%
				CLEC16A	611303	
	s59571	36088265 - 36288264	0.00021	CLIP2	603432	MED%
				RFC2	600404	
				ELN	130160	
	OAR24_37414016	37314016 - 37514015	2.5E-05	FLJ37078		MED%
				HSPB1	602195	
				HSPBL2		
				YWHAG	605356	
				SRCRB4D	607639	
				ZP3	182889	
25	OAR25_17689768	17589768 - 17789767	0.00070	TMEM26		MED%

	s58718	35256962 - 35456961	0.00031			
26	s38270	3603695 - 3803694	0.00048	CSMD1	609397	BULK
	OAR26_26026748	25926748 - 26126747	0.00079	DLC1	604258	CURV
	s58715	34017225 - 34217224	0.00077	UNC5D		Dag score

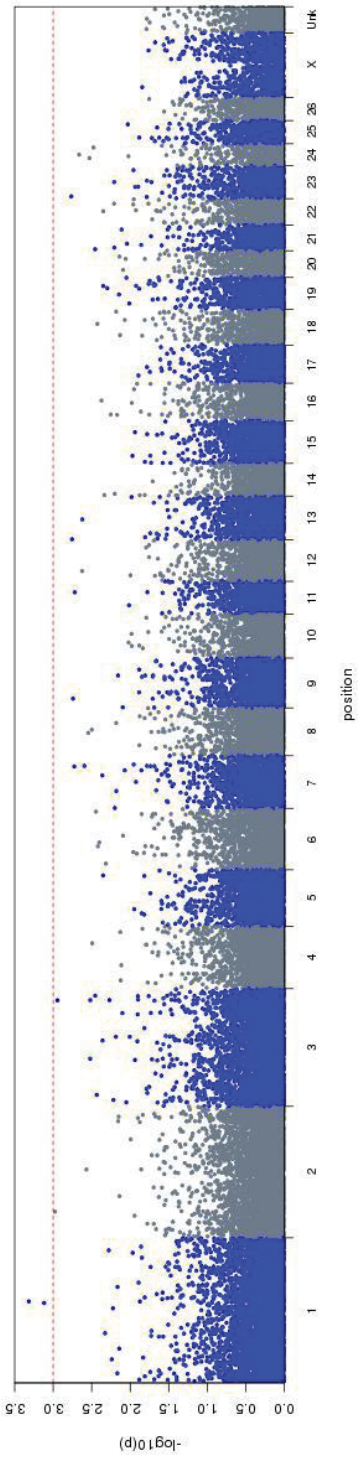
[1]BBREECH: breech bareness; BULK: wool bulk, LENGTH: wool length; MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

Appendix 6

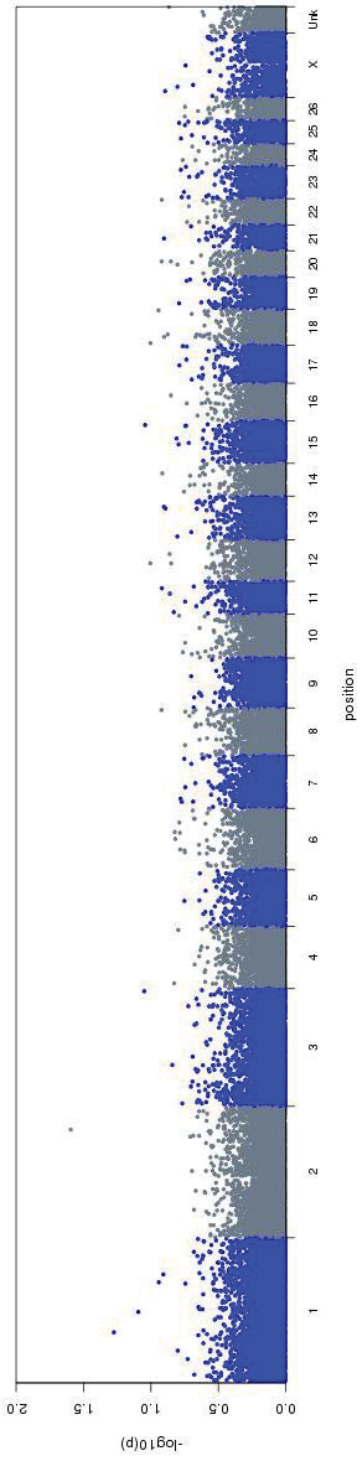


QQ plots from industry GWAS for weaning weight maternal (A), live weight at 8 months (B), breech bareness (C), wool bulk (D), wool length (E), mean fibre diameter (F, MFD), standard deviation of MFD (G), coefficient of variation of MFD (H), curvature (I) and proportion medullation (J) $-\log_{10}(P)$ values. The 0-1 line is in black and the slope in red.

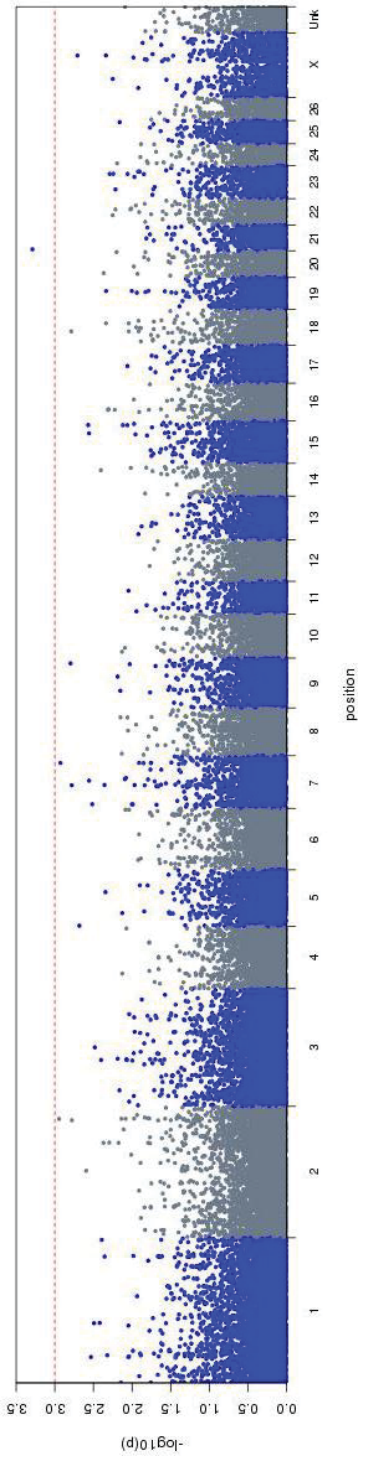
Appendix 7-16



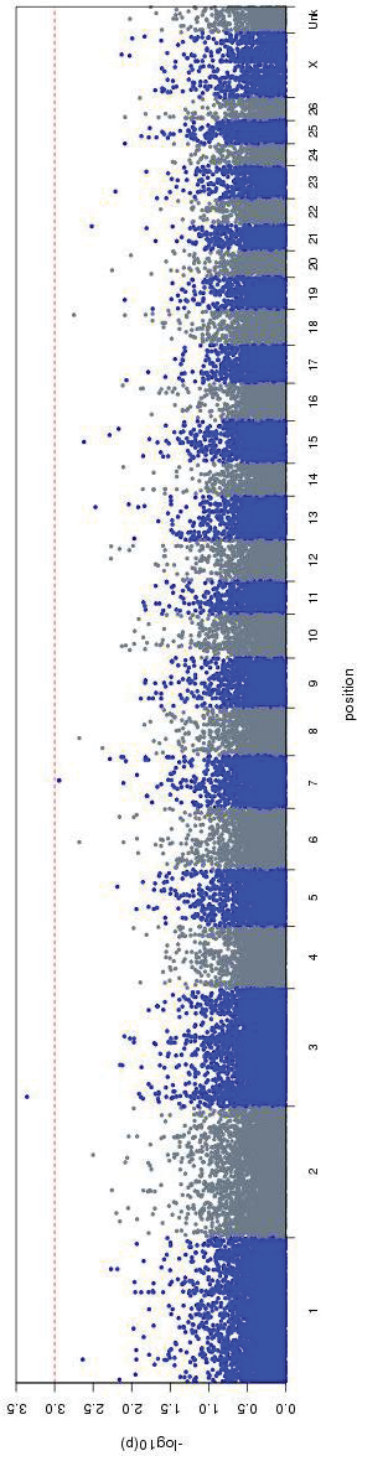
Appendix 7: Manhattan plot of -log$_{10}$(*P*) values of SNPs from industry GWAS for maternal weaning weight. Ordered on the ovine genome v2.0 map, P<0.001 (red dash).



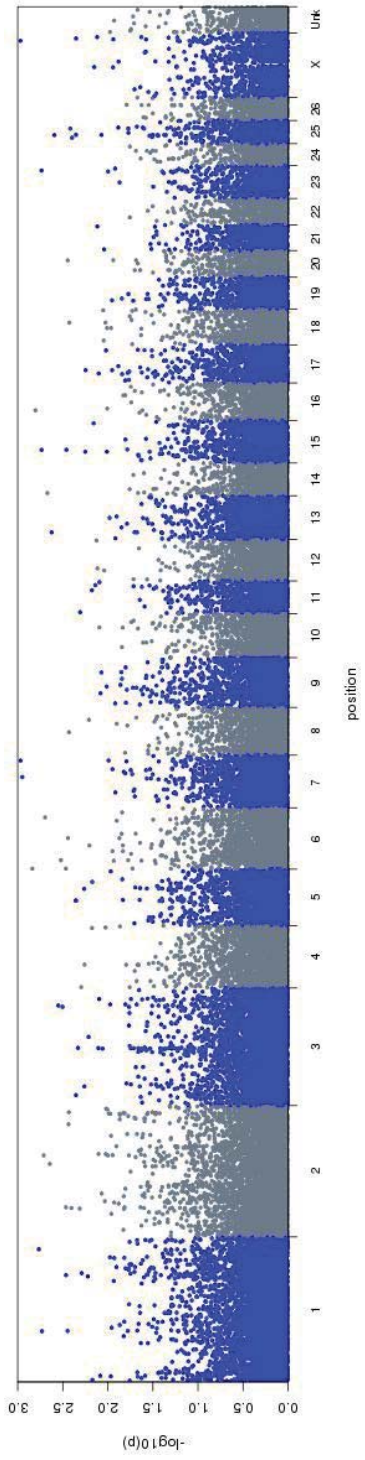
Appendix 8: Manhattan plot of -log$_{10}$(*P*) values of SNPs from industry GWAS for live weight at 8 months. Ordered on the ovine genome v2.0 map.



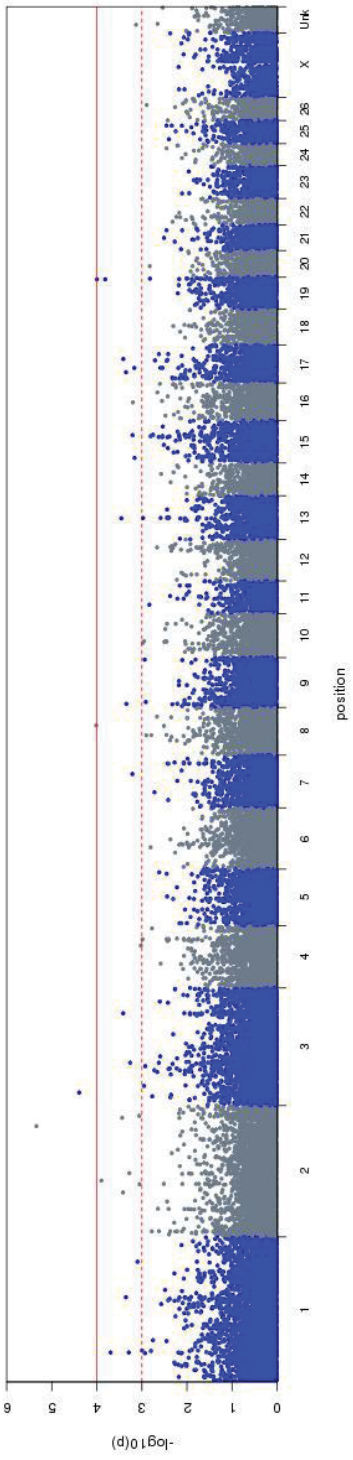
Appendix 9: Manhattan plot of -log$_{10}$(*P*) values from industry GWAS of SNPs for breech bareness. Ordered on the ovine genome v2.0 map, P<0.001 (red dash line).



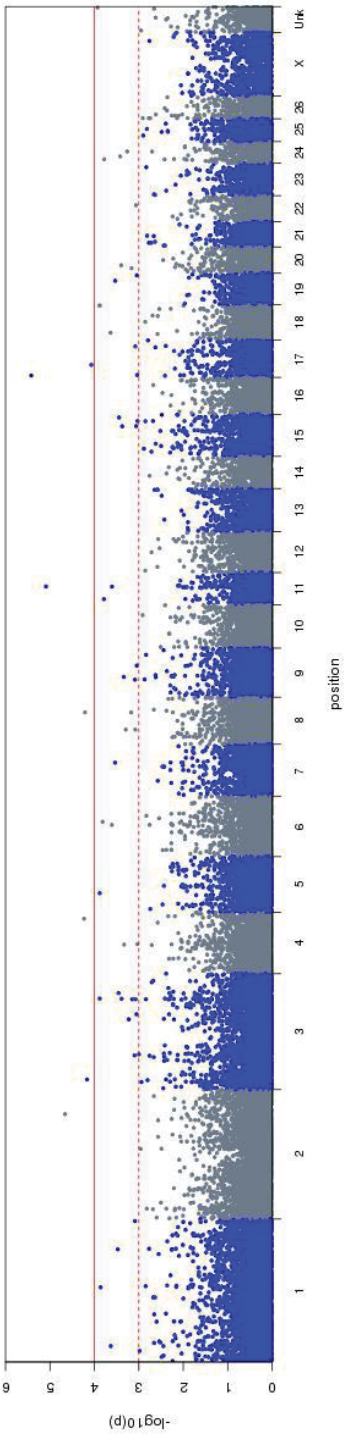
Appendix 10: Manhattan plot of -log$_{10}$(*P*) values from industry GWAS of SNPs for wool bulk. Ordered on the ovine genome v2.0 map, P<0.001 (red dash line).



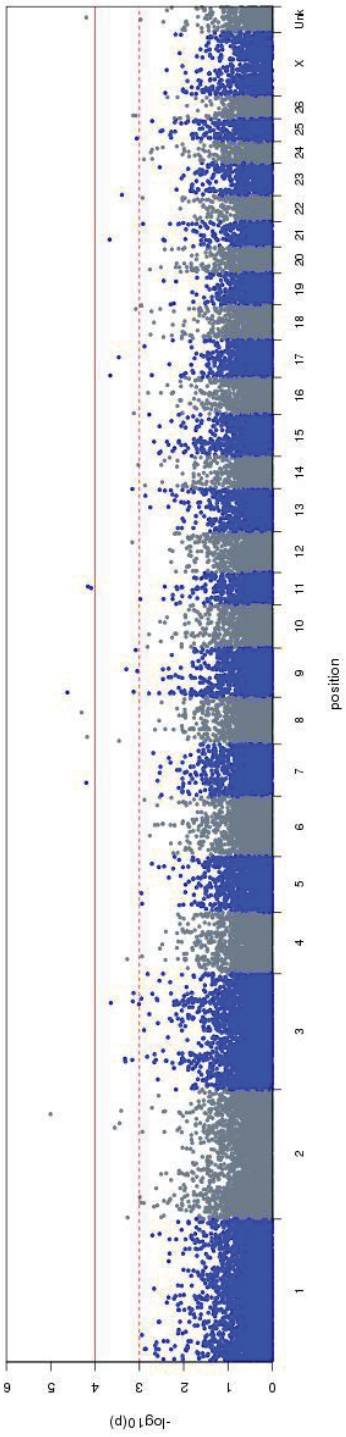
Appendix 11: Manhattan plot of -log$_{10}$(*P*) values from industry GWAS of SNPs for wool length. Ordered on the ovine genome v2.0 map.



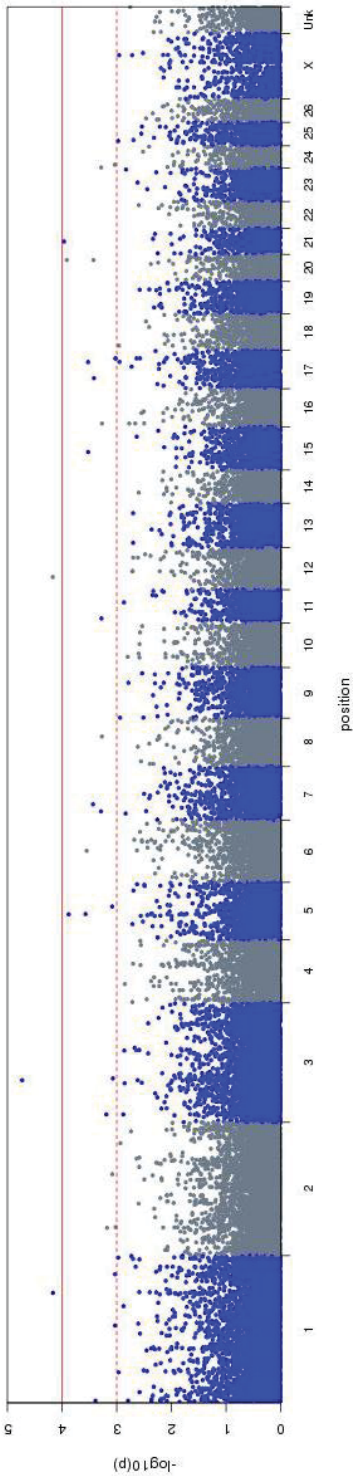
Appendix 12: Manhattan plot of -log$_{10}$(*P*) values from industry GWAS of SNPs for mean fibre diameter. Ordered on the ovine genome v2.0 map, P<0.0001 (solid red line), P<0.001 (red dash line).



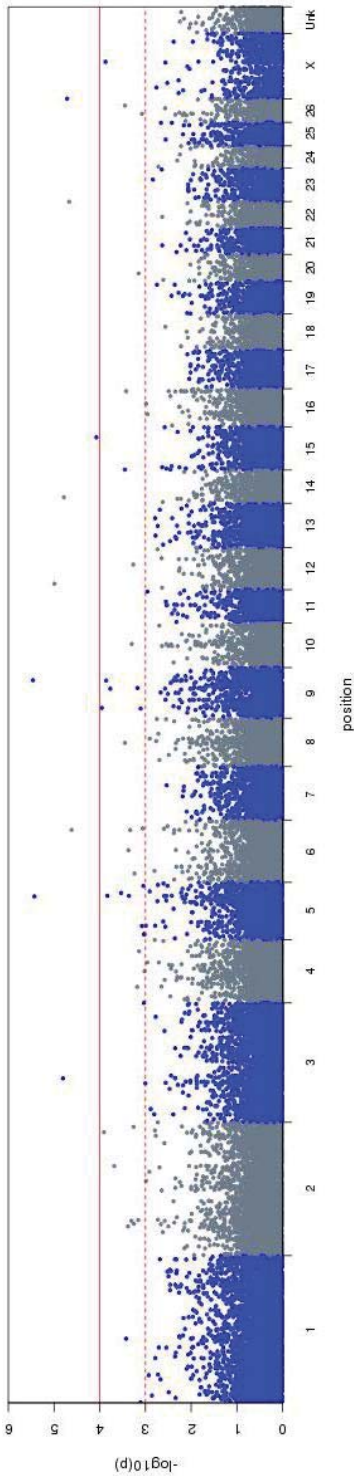
Appendix 13: Manhattan plot of -log$_{10}$(*P*) values of SNPs from industry GWAS for standard deviation of mean fibre diameter. Ordered on the ovine genome v2.0 map, P<0.0001 (solid red line), P<0.001 (red dash line).



Appendix 14: Manhattan plot of -log$_{10}$(*P*) values of SNPs from industry GWAS for coefficient of variation of mean fibre diameter. Ordered on the ovine genome v2.0 map, P<0.0001 (solid red line), P<0.001 (red dash line).



Appendix 15: Manhattan plot of -log$_{10}$(*P*) values of SNPs from industry GWAS for curvature. Ordered on the ovine genome v2.0 map, P<0.0001 (solid red line), P<0.001 (red dash line).



Appendix 16: Manhattan plot of -log$_{10}$(*P*) values of SNPs from industry GWAS for proportion of medullated fibres. Ordered on the ovine genome v2.0 map, P<0.0001 (solid red line), P<0.001 (red dash line).

Appendix 17

Candidate genes within 100kbp of the best 1 - 7 best SNPs for each trait by chromosome (Chr), from industry GWAS (Chapter 7), ordered on ovine genome v2.0.

Chr	200kbp region	Top SNP	*P* value	Candidate genes	OMIM	Function	Trait[1]
1	154779916 - 154979915	OAR1_166829690	4.83E-04	na			WWTm
	206103223 - 206303222	OAR1_222461757	6.83E-05	GNB4	610863	Transmembrane signalling	CURV
				MFN1	608506	Mitochondrial fusion	
				ZNF639		Transcription repressor	
2	202466737 - 202666736	OAR2_213564294	2.19E-05	KCTD18		Potassium channel	FDSD
			9.86E-06	SGOL2	612425	Protect centromeric cohesion	FDCV
				AOX1	602841	Free radical pathway	
	205692466 - 205892465	s61633	0.025	ICOS	604558	T-cell response	LW8[2]
	210744095 - 210944095	s32024	4.57E-06	na			MFD
3	17905487 - 18105486	s61646	4.37E-04	na			BULK
	19531469 - 19731468	OAR3_21630699	6.82E-05	PQLC3		PQ loop repeat	FDSD
				ROCK2	604002	Regulate cytokinesis	
	218705503 - 218905502	s06773	7.54E-04	LDOC1L		Placental structure	WWT
	24674437 - 24874436	OAR3_27227185	4.05E-05	na			MFD

3	78675714 - 78875713	OAR3_83710165	1.87E-05	na			CURV
	82149748 - 82349747	OAR3_87533552	1.57E-05	na			MED
4	106121310 - 106321309	OAR4_114145516	5.86E-05	OR2A2		Olfactory receptor	FDSD
5	80803483 - 81003482	OAR5_88303469	3.73E-06	XRCC4	194363	DNA double strand break repair	MED
6	70085558 - 70285557	s63492	6.66E-04	PDGFRA	173490	Organ development & platelet growth	WWT[2]
				KIT	164920	Oncogene	
	83750911 - 83950910	OAR6_91313780	6.07E-04	TMPRSS11D	605369	Mucous membrane	LDAG
				TMPRSS11A	611704	Cellular senescence	
				TMPRSS11G		similar to A & D above	
7	25511695 - 25711694	DU241785_427	6.33E-05	AVEN	605265	Apopotosis inhibitor	FDCV
				RYR3	180903	Calcium ion release channels	
8	56443640 - 56643639	OAR8_60310131	9.67E-05	EPB41L2	603237	Epithelial tight junctions	MFD
	61475336 - 61475335	OAR8_65932891	6.14E-05	PDE7B	604645	G-protein signalling pathways	FDSD
			4.95E-05				FDCV
9	9021298 - 9221297	OAR9_8924567	2.36E-05	na			FDCV

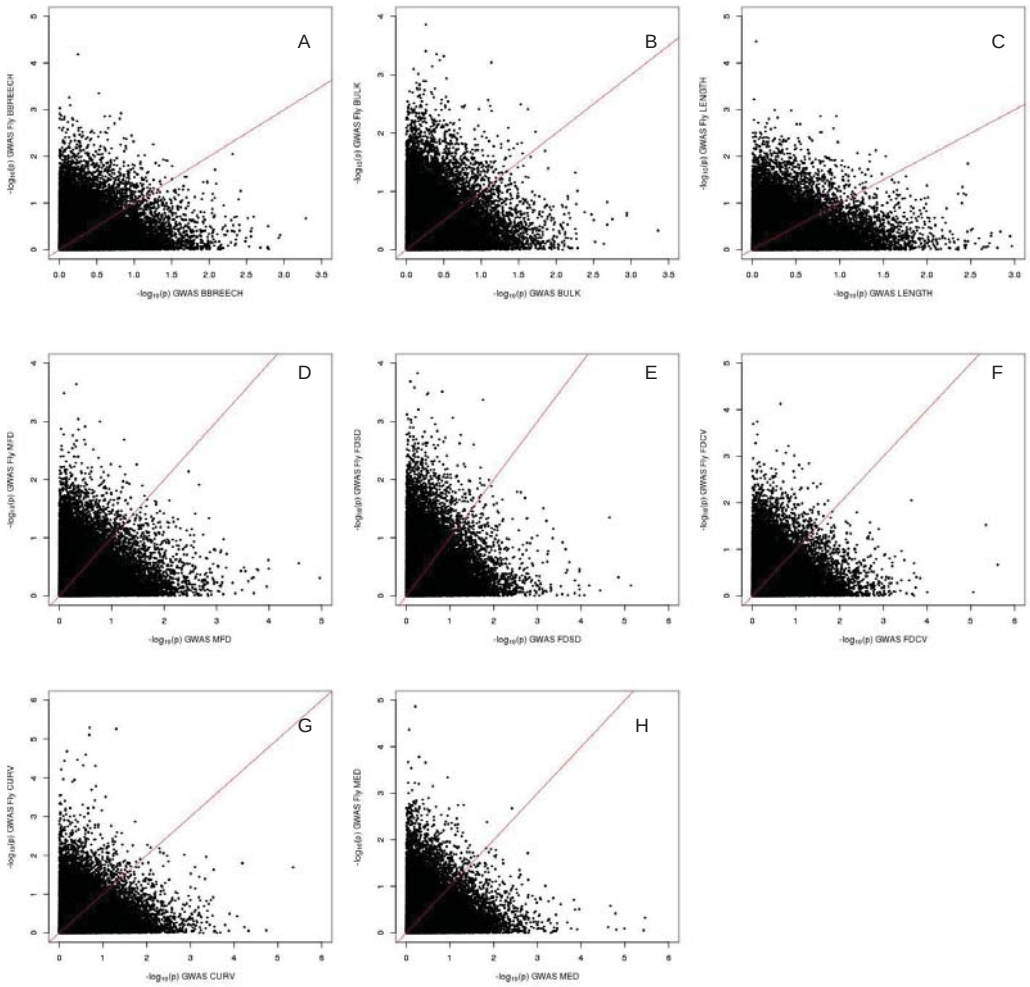
9	70939772 - 71139771	s38039	3.46E-06	OXR1	605609	Oxidative damage protection	MED
11	35347352 - 35547351	OAR11_37593909	8.16E-06	SPAG9	605430	Scaffolding protein	FDSD
			6.92E-05				FDCV
12	11134726 - 11334725	OAR12_14036242	1.02E-05	na			MED
	23520045 - 23720044	s44092	6.77E-05	na			CURV
14	11012765 - 11212765	s57229	1.65E-05	GSE1		Genetic suppressor element	MED
				GINS2	610609	DNA replication	
15	40375174 - 40575173	s22390	5.23E-05	MICAL2	608881	Secretory pathway	ADAG
				GLEAN predict		unknown	
				DKK3	605416	Embryonic development - WNT signalling	
17	3563786 - 3763785	OAR17_4155475	3.80E-06	SFRP2	604157	Cell growth & differentiation - WNT signalling	FDSD
19	56516604 - 56716603	s22636	1.00E-04	RAF1	164760	Cell death regulator	MFD
				MKRN2	608426	Ubiquitin ligase	
21	2406876 - 2606875	s52551	5.13E-04	na			BBREECH

22	36012603 - 36212602	s28777	5.57E-04	PNLIP	246600	Pancreatic lipase	WWT
				PNLIPRP2	604423	Pancreatic lipase	
25	40515041 - 40715040	s71205	0.003	GDF2	605120	BMP signalling	FW12
				GDF10	601361	BMP signalling	
X	119393 - 319392	OARX_8203085	1.92E-05	MXRA5		Matrix remodelling	MED
				ARSE	300180	Sulfatase, bone and cartilage matrix	
				GYG2	300198	Glycogen biosynthesis	
				PRKX	300083	Cell differentiation & epithelial morphogenesis	
UNK		OAR22_42532953	6.32E-05				FDCV

[1] WWT: weaning weight; WWTm: maternal WWT; LW8: live weight at 8 months; FW12: fleece weight at 12 months; DAG3, DAG8: dag score at 3 and 8 months; BBREECH: breech bareness; BULK: wool bulk, LENGTH: wool length; MFD, mean fibre diameter; FDSD: standard deviation of MFD; FDCV: coefficient of variation of MFD; CURV: curvature; MED%: proportion of medullated fibres.

[2]Genes within 500kbp of the top SNP

Appendix 18



Plot of -log$_{10}$(*P*) values from case-control GWAS (y axis, Chapter 6) versus the industry GWAS (x axis, Chapter 7) for breech bareness (A), wool bulk (B), wool length (C), mean fibre diameter (D, MFD), standard deviation of MFD (E), coefficient of variation of MFD (F), curvature (G) and proportion medullation (H). The 0-1 line is plotted in red.

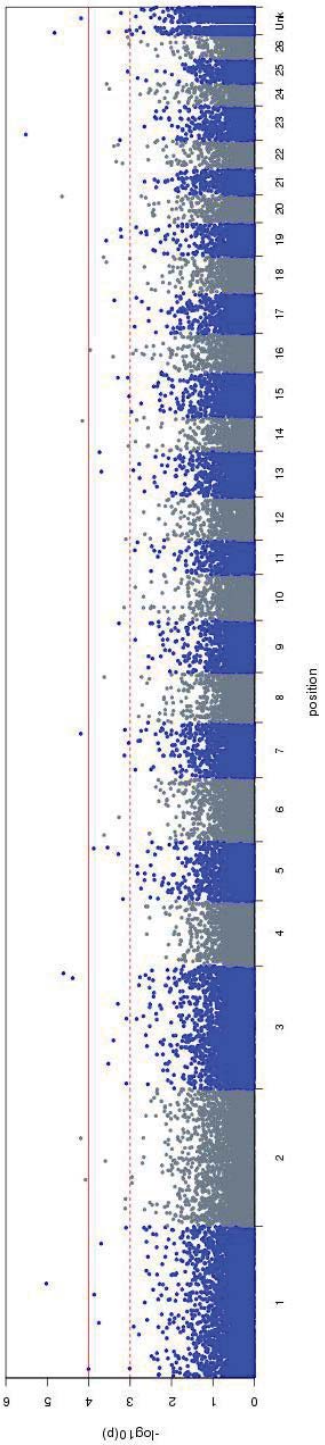
Appendix 19

Estimates of heritabilities (diagonal), genetic (below diagonal), and phenotypic (above diagonal) correlations used for the dual-purpose selection index.

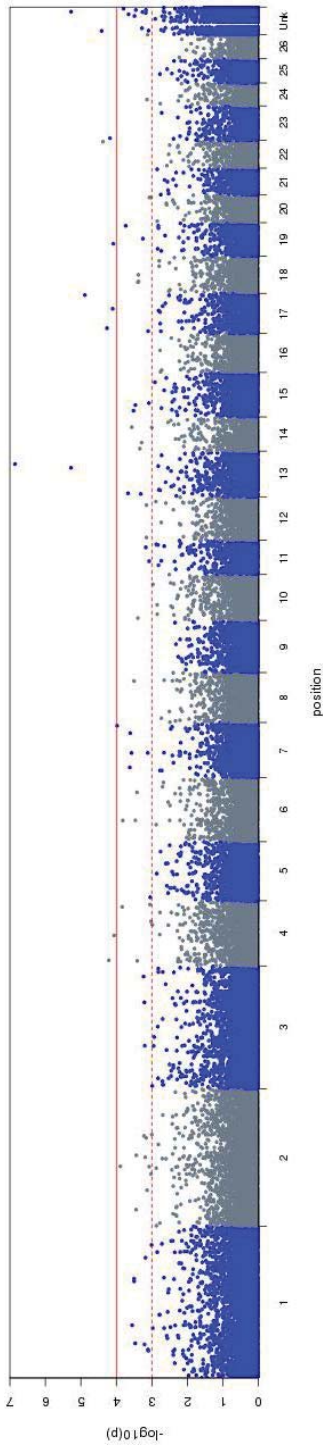
Trait[1]	WWT	WWT_M	CW	LW8	EWT	LFW	FW12	AFW	NLB	SURV	$SURV_M$	DAG3	DAG8	FEC1	FEC2	AFEC
WWT	**0.14**	0	0.7	0.74	0.5	0.2	0.23	0.25	0.01	0	0	-0.158	-0.03	0.02	0.03	0
WWT_M	0	**0.2**	0	0.73	0	0	0.21	0	0.008	0	0	-0.144	-0.02	0.03	0.03	0
CW	0.75	0	**0.3**	0.9	0.7	0	0	0	0	0	0	0	0	0	0	0
LW8	0.87	0.91	0.85	**0.35**	0.7	0.3	0.36	0.25	0.01	0	0	-0.087	-0.05	0.03	0.04	0
EWT	0.5	0	0.75	0.75	**0.45**	0	0	0	0	0	0	0	0	0	0	0
LFW	0.2	0	0	0.3	0	**0.15**	0.4	0.1	0	0	0	0	0	0	0	0
FW12	0.19	0.19	0	0.21	0	0.6	**0.37**	0.5	-0.01	0	0	0	-0.03	0.03	0.05	0
AFW	0.3	0	0	0.35	0	0.4	0.75	**0.45**	0	0	0	0	0	0	0	0
NLB	0.08	0.08	0	0.07	0	0	-0.06	0	**0.09**	0	0	0	0.01	-0.01	0.01	0
SUR	0	0	0	0	0	0	0	0	0	**0.01**	0	0	0	0	0	0
$SURV_M$	0	0	0	0	0	0	0	0	0	0	**0.01**	0	0	0	0	0
DAG3	-0.14	-0.13	0	-0.1	0	0	0.07	0	0.01	0	0	**0.34**	0.4	-0.01	-0.01	0
DAG8	-0.04	-0.01	0	-0.01	0	0	0.09	0	0.04	0	0	1	**0.31**	0.01	0.03	0
FEC1	0.09	0.17	0	0.13	0	0	0.11	0	0.03	0	0	-0.04	0.05	**0.18**	0.19	0.2
FEC2	0.11	0.16	0	0.15	0	0	0.17	0	0.04	0	0	-0.04	0.1	0.74	**0.19**	0.2
AFEC	0	0	0	0	0	0	0	0	0	0	0	0	0	0.71	0.71	**0.25**

[1]WWT: weaning weight; WWT_M: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; $SURV_M$: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

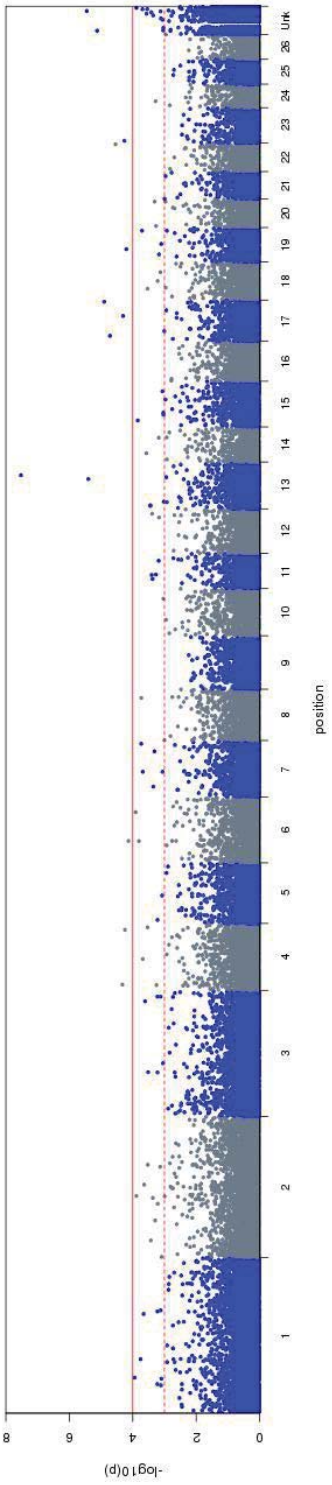
Appendix 20-31:



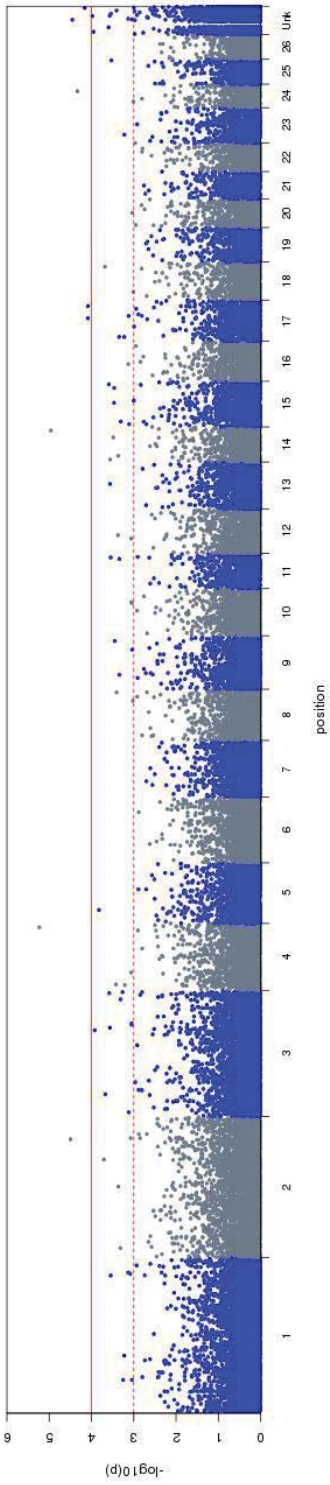
Appendix 20: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for maternal weaning weight. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



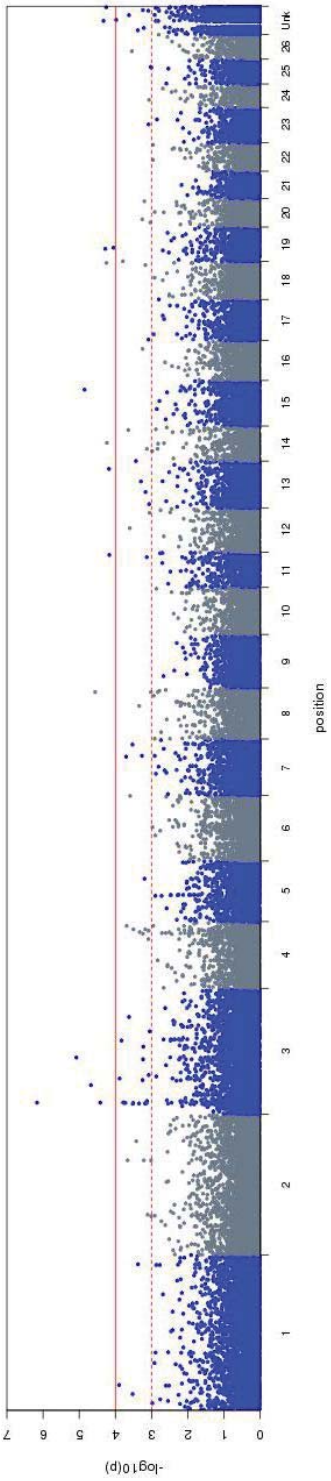
Appendix 21: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for carcass weight. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



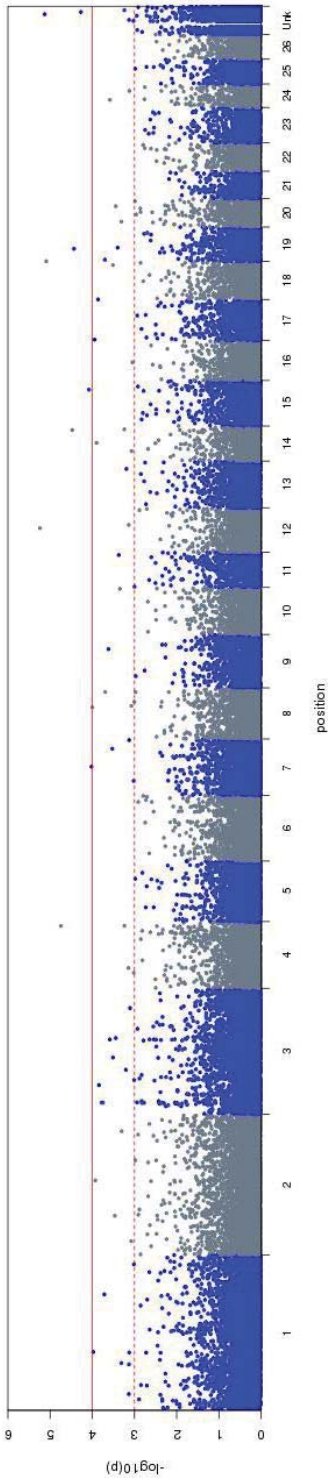
Appendix 22: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for live weight at 8 months. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



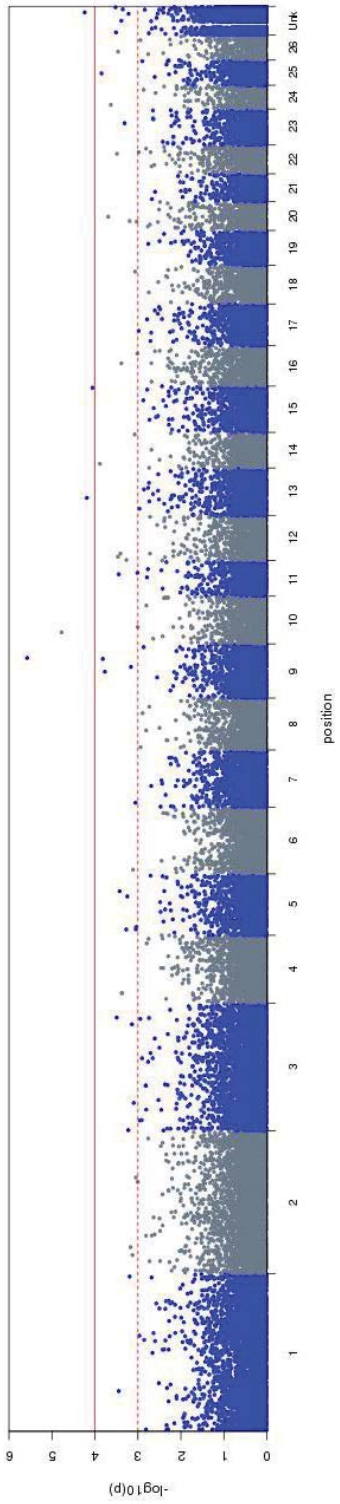
Appendix 23: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for adult ewe weight. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



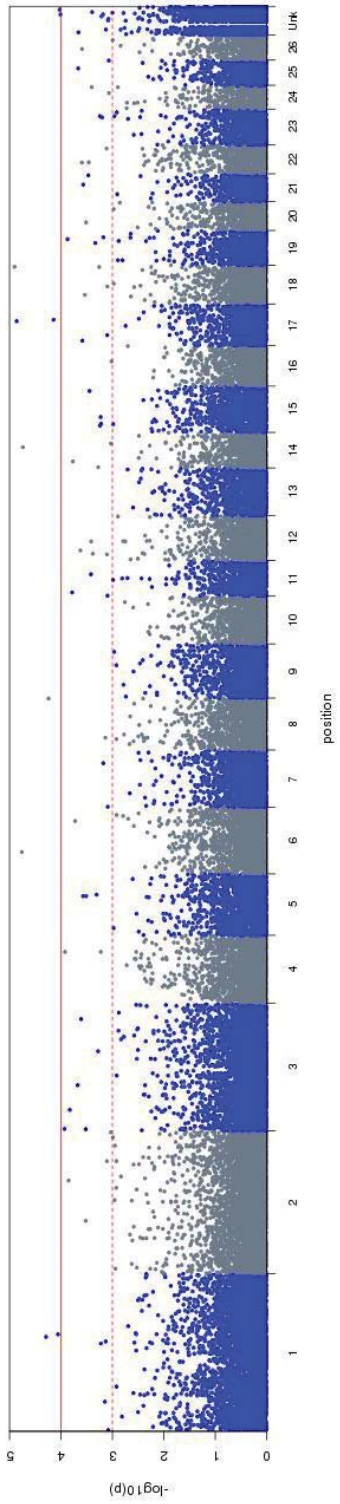
Appendix 24: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for lamb fleece weight. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



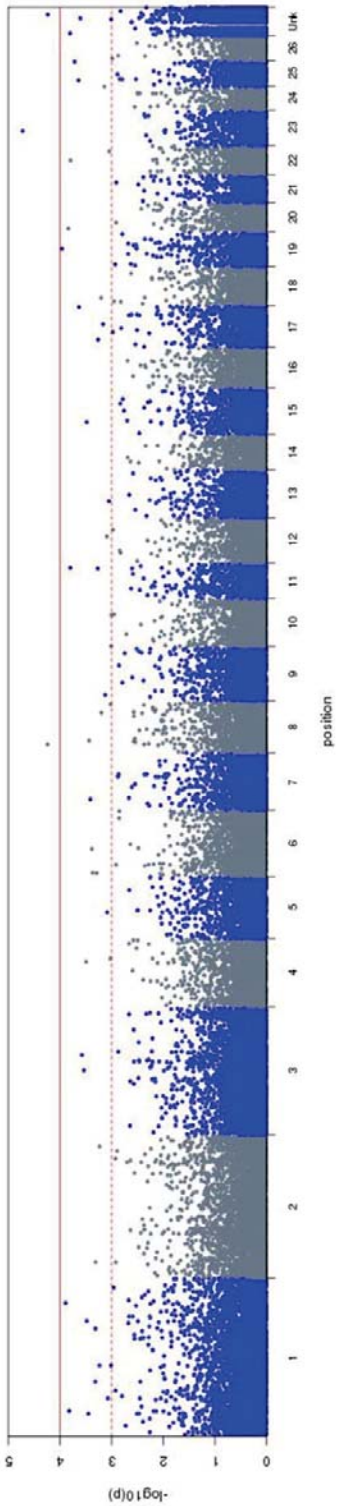
Appendix 25: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for adult ewe fleece weight. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



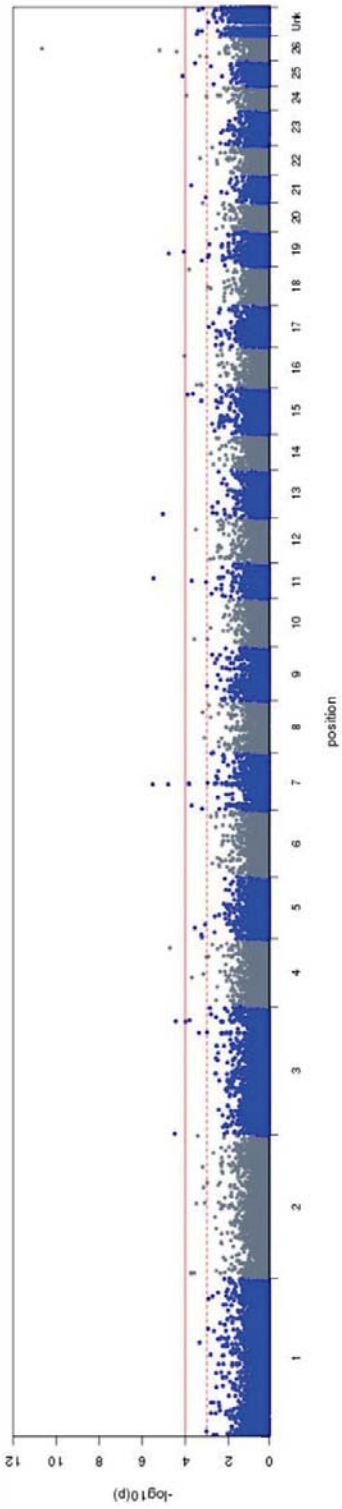
Appendix 26: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for number of lambs born. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



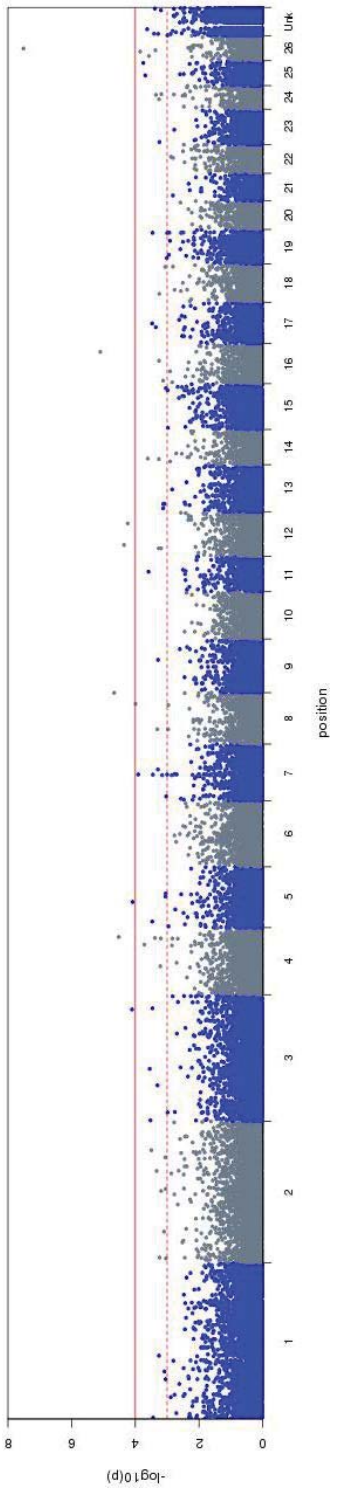
Appendix 27: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for lamb survival. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



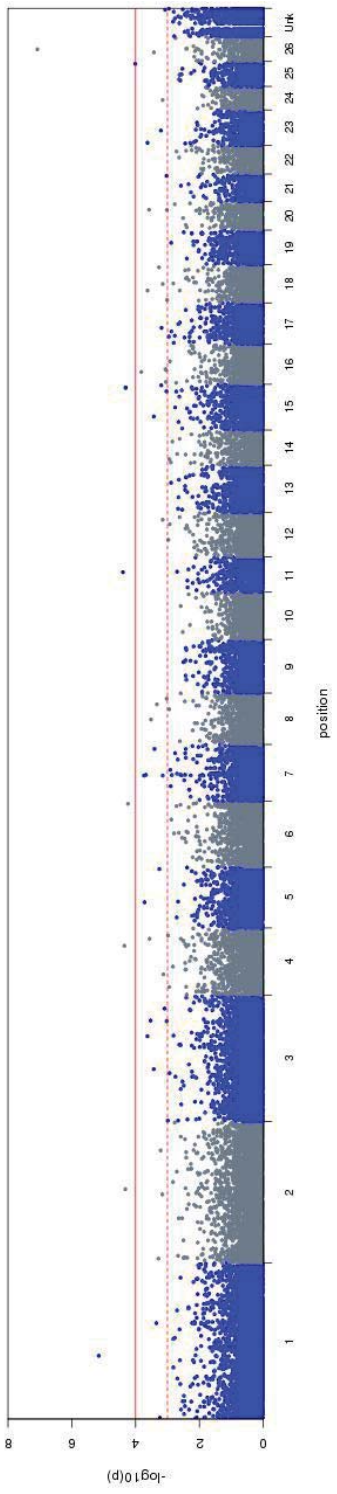
Appendix 28: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for maternal lamb survival. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



Appendix 29: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for faecal egg count in summer. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



Appendix 30: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for faecal egg count in autumn. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).



Appendix 31: Manhattan plot of -log$_{10}$(*P*) values of SNPs from GS analysis for adult faecal egg count. Ordered on the ovine genome v2 map, P<0.0001 (solid red line), P<0.001 (red dash line).

Appendix 32

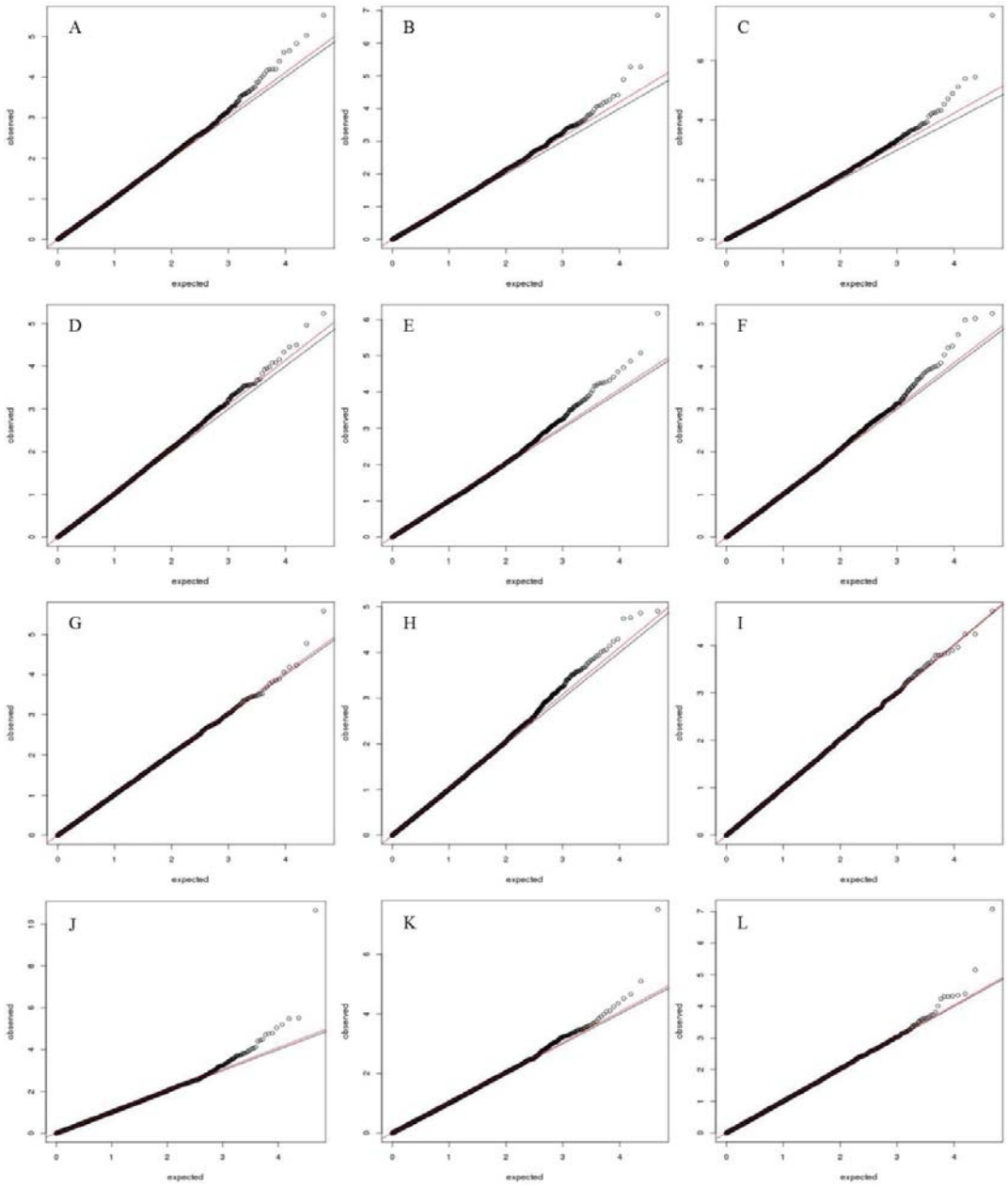
The trait, marker name, chromosome (Chr) and chromosome position (base pairs, bp, Ovine genome v2) of the best 5 SNPs from GS analysis for dual-purpose production traits. The *P* value and $-\log_{10}(P)$ are shown for each marker.

Trait[1]	Marker name	Chr	Position (bp)	*P* value	-log10(*P*)
WWTm	OAR23_11162757	23	10327999	3.03E-06	5.519
	OAR1_185029313_X	1	171685750	9.48E-06	5.023
	OAR11_63183639	UNK		1.49E-05	4.826
	OAR20_52019528	20	47565629	2.27E-05	4.644
	s73027	3	210429191	2.45E-05	4.611
CW	s33509	13	59530141	1.39E-07	6.855
	s53259	UNK		5.28E-06	5.277
	OAR13_57402806	13	52853915	5.29E-06	5.277
	s59808	17	69631709	1.30E-05	4.887
	OAR2_63848688	UNK		3.83E-05	4.417
LW8	s33509	13	59530141	2.99E-08	7.524
	s53259	UNK		3.56E-06	5.449
	OAR13_57402806	13	52853915	3.97E-06	5.402
	OAR2_63848688	UNK		7.51E-06	5.124
	s59808	17	69631709	1.26E-05	4.899
EWT	OAR4_119618487	4	111642446	5.84E-06	5.234
	s76019	14	55734529	1.10E-05	4.959
	s41291	2	209707327	3.20E-05	4.495
	s09930	UNK		3.53E-05	4.453
	s40549	24	28954838	4.67E-05	4.331
LFW	OAR3_22724265	3	20690810	6.84E-07	6.165
	OAR3_107450633	3	100836677	8.35E-06	5.078
	s24485	15	64620511	1.40E-05	4.853
	OAR3_55315847	3	51869303	2.10E-05	4.677
	s72743	8	83389124	2.75E-05	4.560
EFW	s29250	12	43022564	5.77E-06	5.239
	s36765	UNK		7.47E-06	5.127
	s03219	18	67999364	8.14E-06	5.089
	OAR4_118854495	4	110876860	1.81E-05	4.742

	s01263	14	55147830	3.35E-05	4.475
NLB	OAR9_74016375	9	70084773	2.64E-06	5.578
	s46011	10	20136033	1.68E-05	4.776
	s46892	UNK		5.76E-05	4.240
	OAR13_33557008	13	30426470	6.49E-05	4.188
	OAR15_86383219	15	78087390	8.75E-05	4.058
SURV	s67734	18	65005905	1.25E-05	4.904
	s62648	17	42273626	1.37E-05	4.862
	OAR6_42419142	6	38217814	1.74E-05	4.760
	OAR14_37812931	14	36190089	1.81E-05	4.742
	OAR1_178282136	1	165298058	5.10E-05	4.292
SURVm	OAR23_26838339	23	25851387	1.89E-05	4.723
	OAR8_16197437	8	14724849	5.75E-05	4.240
	s38976	UNK		5.79E-05	4.237
	OAR19_32862578	19	31292040	0.0001	3.962
	OAR1_250699047	1	232804957	0.0001	3.893
FEC1	OAR26_25273391	26	21451687	2.18E-11	10.662
	s65809	7	45021712	3.06E-06	5.515
	OAR11_37440114	11	35253844	3.32E-06	5.479
	OAR26_21546761	26	18224368	6.33E-06	5.199
	DU204611_391	13	5921168	9.09-06	5.041
FEC2	OAR26_25273391	26	21451687	3.09E-08	7.510
	OAR16_61003181	16	56107466	7.96E-06	5.099
	s36619	8	90411165	2.17E-05	4.663
	OAR4_109160452	4	102025464	3.04E-05	4.518
	OAR12_23611167	12	20146292	4.43E-05	4.353
AFEC	OAR26_25273391	26	21451687	8.38E-08	7.077
	s22519	1	111781161	7.03E-06	5.153
	OAR11_37440114	11	35253844	4.01E-05	4.397
	OAR4_93103283	4	87035468	4.46E-05	4.351
	OAR2_138309230	2	130400624	4.77E-05	4.321

[1]WWT: weaning weight; WWTm: WWT maternal; CW: carcass weight; LW8: autumn live weight; EWT: adult ewe weight; LFW: lamb fleece weight; FW12: fleece weight at 12 months; AFW: adult fleece weight; NLB: number of lambs born; SURV: lamb survival; SURVm: SURV maternal; DAG3, DAG8: dag score at 3 and 8 months; FEC1, FEC2: faecal egg count in summer and autumn; AFEC, adult faecal egg count.

Appendix 33



QQ plots from GS for weaning weight maternal (A), carcass weight (B), live weight at 8 months (C), adult ewe weight (D), lamb fleece weight (E), adult fleece weight (F), number of lambs born (G), survival (H), survival maternal (I), faecal egg count (FEC) in summer (J), FEC in autumn (K) and adult FEC (L) -$\log_{10}(P)$ values. The 0-1 line is in black and the slope in red.