Genome-wide association study for stature in New Zealand dairy cattle

A thesis presented in partial fulfilment of the

requirements for the degree of

Master of Science

in

Animal Science

at Massey University, Palmerston North,

New Zealand

Mun Ee, Tan

2013

Abstract

The objective of this thesis was to perform a genome-wide association study (GWAS) to identify single nucleotide polymorphism (SNP) associated with stature in New Zealand dairy cattle. The phenotype data set used for this study contained the animal key, sire code of the bull, birth date, breed code, proportion of Holstein-Friesian genes, proportion of Jersey genes, percentage of North American Holstein genes, estimated breeding values (EBV) for live weight and stature and their reliabilities of 3140 bulls. The genotype data set contained the genotype of 692,598 SNPs for every bull and another file contained the name and position of the SNPs.

The GWAS was performed on Holstein-Friesian, Jersey and Holstein-Friesian × Jersey crossbred bulls using PLINK software version 1.07. Stature EBV was used as the phenotype. The phenotypes were adjusted for percentage of Holstein-Friesian, Jersey, North American Holstein genes and year of birth using multiple regression. Manhattan plots and multi Manhattan plots of *P*-values adjusted to genomic control against the chromosomes were plotted to identify top SNPs with the most significant *P*-values above the significant threshold line.

Based on the top 50 SNPs according to the *P*-value, this study identified nine chromosomes or BTA in the HF population with SNPs significantly associated with stature, BTA2, 3, 4, 5, 6, 11, 12, 14 and 24. SNPs with significant effect on stature were detected in six chromosomes, BTA9, 10, 12, 18, 19 and 25 in the JE population while the SNPs determined to be significantly associated with stature were located on eleven chromosomes, BTA1, 3, 4, 5, 7, 9, 10, 14, 18, 22 and 24 in the XB population. Several SNPs located above the suggestive threshold in the Manhattan plots were also inspected and kept in view for future studies.

The results from this study suggest that the highlighted SNPs with significant associations to stature can serve as candidate SNPs for further investigation to determine the regions of QTLs and ultimately the exact genes that affect stature with other correlated traits in dairy cattle.

Acknowledgements

This section is to acknowledge and express my gratitude for the help, assistance and encouragements given by many people who made this completion of thesis possible. First of all, I thank God for this given opportunity to study this program in Massey University, and for providing everything, from tangible support to kind people I met here.

Secondly, a special thanks to my supervisor Associate Professor Dr. Nicolas Lopez-Villalobos. Offering me the opportunity to handle this study, giving excellent advices with statistical analysis and spending his time to adjusting the data models were just few endeavours he had done for me throughout this year. His invaluable help is much appreciated. To my other supervisor Dr. Richard Spelman from Livestock Improvement Co., his comments and reviews on the results further enhanced this thesis. His help in providing the data set for this study is also acknowledged.

Not to forget Richard Sherlock, Frances Creagh and Kathryn Sanders from Livestock Improvement Corporation either. Their advices and support with the PLINK program and Manhattan plots package in R have helped me very much in this thesis.

My work partner, Sibyl Chan is an irreplaceable friend especially these two years of study. Her help has also made this completion of thesis possible. Her humour and jokes made this period of time easier to endure too. Acknowledging also my fellow post-graduates and friends at Massey, I cannot make it through this time without them. Especially Judy Lau, a post-graduate fellow and an awesome friend.

To my parents and siblings, thank you for believing in me and support me here in New Zealand to further my studies. Thank you for providing the financial, physical, emotional and spiritual support. I cannot possibly have performed this stint without you all.

Last but not least, I want to give thanks to Zachary, my boyfriend for his constant love, encouragements and prayers especially during the final phase of my thesis.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

CHAPTER 2:

CHAPTER 4:

LIST OF ABBREVIATIONS

ATP : Adenosine triphosphate

BLUP : Best linear unbiased prediction

BTA : *Bos Taurus* autosome

BV : Breeding value

BW : Breeding worth

cm : Centimetres

DNA : Deoxyribonucleic acid

EBV : Estimated breeding value

FCE : Feed conversion efficiency

GEBV : Genomic estimated breeding value

GWAS: Genome-wide association study

HF : Holstein-Friesian

HF_{pct} : Holstein-Friesian genes percentage

ID : Identification

IGF1 : Insulin-like growth factor 1

JE : Jersey

JE_{pct} : Jersey genes percentage

kg : Kilograms

LD : Linkage disequilibrium

Mbp : Mega base pairs

NAH : North American Holstein genes

NAH_{pct}: North American Holstein genes percentage

Others : OT

PLAG1: Pleiomorphic adenoma gene 1

SNP : Single nucleotide polymorphism

TOP : Traits other than production

QTL : Quantitative trait loci

QTN : Quantitative trait nucleotide

XB : Holstein-Friesian and Jersey crossbred

CHAPTER 1

INTRODUCTION

The dairy industry in New Zealand started in the early 1800s, from a few dairy cows kept by early settlers to large herds managed by dairy cooperative groups. The current national dairy herd is made up of about 38.2% of Holstein-Friesian, 40.8% of Holstein-Friesian and Jersey crossbred, 12.1% of Jersey, 0.7% of Ayrshire and other minor breeds such as Brown Swiss, Milking Shorthorn and Guernsey (DairyNZ & LIC 2012).

The climate of New Zealand enables grass to grow all year round in most regions, and the temperature is mild enough for cows to graze on pasture in all seasons. Feeding solely on pasture is a distinctive factor of New Zealand's dairy cattle, and costs of feeding and housing are lower compared to dairy industries in other countries. Milk yield level is the highest during spring (September – November) as it coincides with the high level of pasture growth.

For the production season of 2011-12 there were more than 4.634 million dairy cows in New Zealand distributed in 11,798 herds with a total industry production of 19,129 million litres milk and 1,685 million kg milk solids, consisting of both milk fat plus protein (DairyNZ & LIC 2012). The dairy industry is New Zealand's leading export earner, contributing $11 billion which was 27% of the total $41 billion export value in 2009 (Coriolis 2009), and approximately 95% of milk was exported.

The current national breeding objective for NZ dairy cattle is designed to select for more profitable and efficient cows for feed conversion. Cows and bulls are ranked and selected for breeding based on a Breeding Worth (BW) index defined in unit of net farm income in dollars per 5 tonne of dry matter (NZAEL 2013). BW is the sum of trait breeding values (BV) multiplied by the economic value of each trait. The index was first introduced in 1996 (Harris et al. 1996), and was a selection index for milk fat, protein, yield, live weight, fertility, somatic cell score and residual survival traits (NZAEL 2012). The current BW index for the selection of dairy cattle in New Zealand in 2013 stands as below (NZAEL 2013):

$$BW = \$1.79BV_{\text{milk fat}} + \$8.63BV_{\text{protein}} - \$0.091BV_{\text{milk volume}} - \$1.52BV_{\text{live weight}} + \$7.35BV_{\text{fertility}} - \$38.57BV_{\text{somatic cell score}} + \$0.148BV_{\text{residual survival}}$$

The genetic improvement and breeding schemes being carried out since the 1950s has contributed more than $15 billion to the dairy industry. Milk fat has increased for 0.98kg per year, protein has increased 1.14kg per year and live weight has increased 0.09kg per year (NZAEL 2012). Those are a few examples of the rate of genetic gain from 2001 to 2010 through selection using the BW index. The breeding worth itself has increased to $9.3 per year.

Many countries have evaluation systems where dairy cattle are evaluated for traits other than production (TOP); such as size, dairy character, and mammary system. These traits are considered to have something to offer producers in terms of functional herd life and survival. Several countries have developed indices to combine traits together taking account of the importance of each trait using a selection index approach. TOP are not included in NZ national breeding objective but breeding values are estimated for 16 TOP traits. Stature being one of the TOP evaluated in NZ dairy cattle is genetically correlated with live weight. Selecting against stature will indirectly select against live weight, which leads to genetic gain and profit since live weight is one of the traits included in the BW index and carries a negative economic weight.

Genomic selection promises accelerated genetic gain by selecting elite animals based on their genotypic information for a trait with economic importance. Identifying genetic markers connected to the trait will emphasize the position of underlying genes and breeding values are estimated using the overall effects of the markers in the genome. Therefore identifying the markers for stature will enable genomic selection for this trait and for correlated traits too such as live weight. The objective of this study is to conduct a genome-wide association study for stature in New Zealand dairy cattle to highlight the genetic markers or single-nucleotide polymorphisms (SNP) influencing this trait to improve accuracy of selecting for stature.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Various genome scans and genetic mapping studies have been carried out in an attempt to identify the exact genes that control dairy production traits since the first whole-genome scan of dairy cattle was carried out by Georges et al. (1995). Most studies (Spelman et al. 1996; Coppieters et al. 1998; Velmala et al. 1999; Olsen et al. 2002) have attempted to identify quantitative trait loci (QTL) affecting milk production traits. However in recent times, the focus has been shifted to locate other TOP in dairy cattle such as health, fertility, herd life and conformation traits (Klungland et al. 2001; Casas et al. 2003; Hiendleder et al. 2003; Kühn et al. 2003; Zhang et al. 2004; Druet et al. 2008; Garrett et al. 2008).

Selection for TOP has received more focus nowadays as some have economic importance whereas others are relevant for animal welfare. Although the benefits of selecting for such traits are not evident, they can be used to improve the animal's genetic merit. This is due to the discovery of significant genetic correlation between them and production traits. Moderate to strong genetic correlations indicate that such traits are controlled by similar genes, and hence selecting for one trait is a good indicator of the other trait. The indicator trait should have high heritability as well to ensure an early in life, easy and efficient selection. Identifying TOPs genetically correlated with traits with economic significance has the potential to elevate the selection accuracy for the breeding objective trait.

Traits categorised as linear-type traits are based on measurements instead of opinions and such traits are highly heritable. In addition to that, the cost to measure such traits is very low as these data can be used for marketing and aesthetic purposes as well (Caraviello et al. 2003). Therefore linear-type traits with high heritability can be used as predictors of other more important traits in the selection index of any breeding programme. Indirect selection of linear-type traits and the information can be incorporated to the

breeding value (BV) of the important traits. The more accurate are the estimation of BV the more efficient is the selection. Selection for health and fertility traits has become more cost effective and easier with the inclusion of linear-type traits in national and international breeding programmes (Berry et al. 2004).

2.2 Linear score of stature

The TOP Advisory Committee of New Zealand (a sub-committee of New Zealand Animal Evaluation Limited) categorises 18 non-production traits as TOP. Four of the traits are classified as management traits and are scored by the farmer. The rest are conformation traits (stature included) and are recorded by inspectors. Stature is defined as an estimate of the height at shoulders of the animal according to the evaluation system for TOP for dairy cattle in New Zealand. Again the score of 1 and 9 are for biological extremes with 1 is designated for short and 9 is for tall. Each score represents 5cm in height and ranges from under 105cm to over 140cm. Stature can be easily measured by using this scale, and it follows a normal distribution. The international assessment standard for stature in Holstein-Friesian cattle according to the World Holstein-Friesian Federation is fixed upon the measurement from top of the spine in between hips to ground and categorised in scoring values. The scoring system is from 1 for short and to 9 for tall and each unit score represents 3cm. The reference scale ranges from 130cm (short) to 154cm (tall).

The mean for stature of New Zealand Holstein-Friesian and Jersey is 120cm (score 5.058) and 113cm (score 3.643) with a standard deviation in score of 1.064 and 0.868 respectively (Ahlborn & Demple 1992) while in other studies, the mean score of stature for New Zealand Holstein-Friesian and Jersey is 5.984 and 4.172 with a standard deviation of 1.008 and 0.880 (Cue et al. 1996), and 6.51 for Holstein-Friesian, 5.86 for Jersey and 6.19 for Holstein-Friesian and Jersey crossbred cows (Berry et al. 2005).

Sire statistics from New Zealand Animal Evaluation Limited (NZAEL 2013) stated the average estimated breeding value (EBV) for live weight and stature in Holstein-Friesian is 53.9 kilograms and 1.11 score units. The average EBV for live weight and stature in evaluated Jersey sires is -55.2 kilograms and -1.02 score units. For Holstein-Friesian and Jersey crossbred sires, the average EBV is -5.8 kilograms and -0.18 score units respectively.

Results from the genetic evaluation for stature and live weight (Montgomerie, 2006) showed that mature Holstein-Friesian cows were 25% taller than mature Jersey cows. The records for live weight data also showed that Holstein-Friesian cows weighed 45 kg more than the Holstein-Friesian and Jersey crossbred cows. On average, Holstein-Friesian cows display a well-built and taller body than their Jersey and crossbred counterparts. Cue et al. (1996) showed heterosis affects stature and live weight scores, legs and dairy conformation in their work with New Zealand Holstein-Friesian and Jersey crossbred cows.

2.3 Economic importance of stature

Live weight carries a significant economic value in New Zealand dairy industry (NZAEL 2012). Under the New Zealand pasture-based farming system, maintenance is a function of live weight, and live weight carries a negative economic weight which is minus $1.52 per kilogram. Thus live weight is a trait which is continually recorded for the overall breeding index (Cue et al. 1996), and currently has a moderate weighting of 14% in the BW index (NZAEL 2013). When deriving the selection index and proper economic weights for a multi-trait index, it is more appropriate to appoint negative weighting for live weight and correlated conformation traits. By doing so, cows with higher body mass and increased requirements for feed intake will be curbed, and more production-efficient cows will be indirectly selected. Profit per cow can be increased by reducing the feed costs.

Farm profit gained can be defined as the function below:

Profit = Income – Costs

The main sources of income are milk and beef revenues while costs can be breakdown into feed and operational costs. To maximise profit, either the contributing factors for income must be increased or the components for costs must be decreased.

The level of feed intake is primarily determined by the size of cattle. The larger the cow is, the more feed it consumes. Hence a cow with a big live weight, or tall in stature, contributes more to the cost. In addition to the fact that live weight and stature affects feed consumption, feed conversion efficiency will also undermine the efforts to gain maximum profit from the animals. If the feed consumed by the dairy cows are not converting to milk

efficiently, the weight of feed cost will be heavier than the weight of income from milk production, thus profit will be decreased. Feed conversion efficiency is concluded to be important for profit in dairy industry. Thus ensuring a maximum rate of conversion of consumed feed into milk is very important to increase overall profit and to avoid wastage.

Feed conversion efficiency (FCE) for New Zealand dairy cattle is defined as kilograms of milk solids produced per kilogram of dry matter intake (Holmes et al. 1993). A major factor limiting genetic improvement for FCE in dairy cattle is that this trait is difficult and expensive to measure in individual animals. Data on milk yield, live weight and feed intake are needed and measurements of daily feed intake in grazing cows are difficult to obtain.

Feed conversion efficiency is affected by live weight (Holmes et al. 1993) as larger cows (phenotypically and genetically) consume more feed. The feed required by a cow goes not only to milk production, but to body maintenance, growth and reproduction. The heavier the animal is the higher the energy requirements for maintenance, thus lowering the overall FCE. Larger cows therefore contribute more to the cost variable and decrease overall farm income. Animals of considerable size may not be desirable for some dairy systems as it implies additional cost in feeding and maintenance. Cassell (2009) reported that due to the selection intensity for higher milk producing cows, there is an increase in feed intake for cows with modern genetics in comparison to cows from 40 years ago. Ahlborn & Dempfle (1992) and Kennedy et al. (1999) concluded that cows with high genetic merit for live weight produces less milk and are less efficient compared to their smaller counterparts while consume more energy at the same time. Moore et al. (1992) reported larger cows generate less profit in their study with Canadian Holsteins. A publication by Cole et al. (2009) also stated that production efficiency reduced considerably for taller, longer and heavier cows compared to their smaller counterparts.

Largely built animals are also related with delayed sexual maturity. A cow composing a high body mass is likely to have a calf with a heavy birth weight too, and it may cause calving difficulties or dystocia for the mother (Zaborski et al. 2009). Dystocia in cattle causes several undesirable results such as a decrease in milk yield (Rajala & Grönn 1998), increased levels of somatic cell count in milk (Berry et al. 2007) and culling of a

cow (López de Maturana et al. 2007). Johanson & Berger (2003) discovered that for 1 kg increase in calf birth weight, the probability of dystocia increases by 13% in Holstein cattle. For Holstein-Friesian and Jersey breed, the probability of dystocia during third calving of a male calf at 20, 30, 40 and 50 kg is 1%, 2%, 5% and 15% respectively (Berry et al. 2007). However some bigger cows may not experience dystocia as they have greater frame sizes that enable them to deliver a heavy calf. On the other hand, smaller cows may have higher risk of dystocia too due to their smaller frame.

A correlation therefore exists for stature, live weight with milk production and dystocia in dairy cattle. Selecting against stature and live weight will affect milk production and dystocia, thus selecting for an overall balanced size, wide rump angle and high yield for milk solids is preferential.

2.4 Heritability of stature

In biological terms, heritability is used to explain the degree of resemblance between related individuals for a given characteristic (Jacquard 1983). Heritability, always used with the symbol h^2, is defined as the proportion of variation in performance among animals that is contributed by gene effects transferred from generation to generation (Cassell 2009), and thus is used to explain the range of differences in the phenotypic performance caused by transmissible genes alone. Heritability value is always positive, and lies within the range of 0.0 and 1.0 as heritability reflects a proportion. A trait with h^2 exceeds 0.40 is considered a highly heritable trait whereas h^2 with a value less than 0.15 is considered as low heritability (Cassell 2009).

Stature is classified as one of the highly heritable traits among the TOP. Estimates of heritability with breed, country of study taken from various studies are shown in Table 2.1.

The paper by Ahlborn & Dempfle (1992) states the heritability for stature is 0.29 for Holstein-Friesian cattle and 0.23 for Jersey cattle. Heritability for stature in US Holstein cattle is 0.39 (Fuerst-Waltl et al. 1998). Heritability for stature in Spanish Holstein cows is reported as 0.43 (Pérez-Cabal & Alenda 2002). The heritability for stature is reported to be

0.38 for Holstein-Friesian and 0.27 for Jersey (Cue et al. 1996). The estimation for heritability varies between literatures as it is influenced by breed, population, management and environmental factors (Fortes et al. 2012). Cue et al. (1996) also reported that heterosis effects are significant for stature as there are significant interactions within breed groups, hence explaining the variation of heritability of stature in different breeds. With its high heritability and high accuracy of measurement, this trait is an effective selection criterion to increase the net economic merit for dairy industry.

Table 2.1: Estimates of heritability (h^2) and standard errors (SE) for stature in beef and dairy cattle.

Study	Country of study	Breed	$h^2 \pm SE$
Ahlborn & Dempfle (1992)	New Zealand	Holstein-Friesian	0.29 ±0.04
		Jersey	0.23 ±0.05
Brotherstone (1994)	United Kingdom	Holstein-Friesian	0.48 ±0.03
Cue et al., 1996	New Zealand	Holstein-Friesian	0.38 ±0.02
		Jersey	0.27 ±0.02
		Holstein-Friesian and Jersey crossbred	0.40 ±0.02
Fuerst-Waltl et al. (1998)	United States of America	Holstein	0.39
Kennedy et al. (1999)	Canada	Holstein	0.42
Schrooten et al. (2000)	The Netherlands	Holstein-Friesian	0.6
Pérez-Cabal & Alenda (2002)	Spain	Holstein	0.43 ±0.01
Cassell (2009)	United States of America	Holstein	0.42
		Jersey	0.39
Brand et al. (2010)	Germany	Holstein	0.41
Zink et al. (2011)	Czech Republic	Holstein	0.39 ±0.02
du Toit et al. (2012)	South Africa	Jersey	0.20 ±0.01
Mc Hugh et al. (2012)	Ireland	Aberdeen Angus, Belgian Blue, Charolais, Friesian, Hereford, Holstein, Jersey, Limousin, Montbeliarde, Normande, Norwegian Red, Simmental	0.38 ±0.03

2.5 Correlations between stature and other traits of economic importance

Genetic correlation takes place when genes that control the two selected traits overlap to some extent. Genes located closely together on the same chromosome are unlikely to be randomly segregated during meiosis (Rauw et al. 1998), such association is known as genetic linkage. Traits affected by closely linked genes are normally inherited together as a consequence of linkage. Pribyl et al. (2010) claimed that the effects of linked genes will be counted when examining the effect of a selected gene for a specific performance. Epistatic effects defined as the interaction of genes which are not alleles also induce a locus to influence several traits. The effects can be either positive or negative. If genetic correlation is positive, then selecting one of the traits will increase the selection of the other trait. Vice versa if genetic correlation is negative, selecting one of the traits will decrease the selection of the other trait.

Genetic correlation between two or more traits can also be explained by the combined effect of linkage and pleiotropy. This is due to the effects of pleiotropy where a single gene, QTL or quantitative trait nucleotide (QTN) influences multiple phenotypic traits. The QTNs are closely linked together in various loci and influences multiple traits.

Table 2.2 showed the estimated genetic parameters and correlations for stature and other traits of economic importance in dairy cattle sourced from various publications. The correlation between live weight and milk yield ranges from 0.29 to 0.39 (Ahlborn & Dempfle 1992; Spelman & Garrick 1997; Pryce & Harris 2006). The body size of the cow and udder has a positive genetic correlation of 0.26. Cassell (2009) stated "Such a correlated response is responsible for deeper udders or wider front teat placement on higher producing cows.", therefore a cow with a larger body size has a high possibility of higher milk production that comes with a deeper udder. Short and Lawlor (1992) too have identified a correlation between stature and production, hence any selection concerning stature will indirectly affect selection of milk production as well.

A moderately high genetic correlation of 0.46 for stature and protein yield in Jersey has been reported by Ahlborn & Demple (1992). They too discovered stature has a positive genetic correlation (a range of 0.25 to 0.34 for Holstein-Friesian and 0.43 to 0.46 for Jersey)

Table 2.2: Estimates of genetic parameters for stature and traits∓ of economic importance in dairy cattle. Heritabilities shown on the diagonal and in bold; genetic correlations shown below the diagonal.

Trait	Milk	Protein	Fat	Live weight	Stature	DMI	FCE
Milk							
Ahlborn & Dempfle (1992)	**0.28** (HF)	–	–	–	–	–	–
	0.26 (JE)	–	–	–	–	–	–
Spelman & Garrick (1997)	**0.28**	–	–	–	–	–	–
Veerkamp & Brotherstone (1997)	**0.27**	–	–	–	–	–	–
Kennedy et al. (1999)	–	–	–	–	–	–	–
Veerkamp et al. (2001)	**0.48**	–	–	–	–	–	–
Muller et al. (2006)	–	–	–	–	–	–	–
Pryce & Harris (2006)	**0.36**	–	–	–	–	–	–
Toshniwal et al. (2008)	**0.14**	–	–	–	–	–	–
Protein							
Ahlborn & Dempfle (1992)	0.82 (HF)	**0.26** (HF)	–	–	–	–	–
	0.87 (JE)	**0.24** (JE)	–	–	–	–	–
Spelman & Garrick (1997)	0.80	**0.25**	–	–	–	–	–
Veerkamp & Brotherstone (1997)	0.80	**0.35**	–	–	–	–	–
Kennedy et al. (1999)	–	–	–	–	–	–	–
Veerkamp et al. (2001)	0.84	**0.42**	–	–	–	–	–
Muller et al. (2006)	–	–	–	–	–	–	–
Pryce & Harris (2006)	0.86	**0.29**	–	–	–	–	–
Toshniwal et al. (2008)	0.85	**0.14**	–	–	–	–	–
Fat							
Ahlborn & Dempfle (1992)	0.68 (HF)	0.78 (HF)	**0.26** (HF)	–	–	–	–
	0.77 (JE)	0.85 (JE)	**0.26** (JE)	–	–	–	–
Spelman & Garrick (1997)	0.60	0.70	**0.22**	–	–	–	–
Veerkamp & Brotherstone (1997)	0.55	0.75	**0.40**	–	–	–	–

Trait	Milk	Protein	Fat	Live weight	Stature	DMI	FCE
Fat							
Kennedy et al. (1999)	–	–	–	–	–	–	–
Veerkamp et al. (2001)	0.41	0.58	**0.39**	–	–	–	–
Muller et al. (2006)	–	–	–	–	–	–	–
Pryce & Harris (2006)	0.55	0.69	**0.29**	–	–	–	–
Toshniwal et al. (2008)	0.65	–	**0.04**	–	–	–	–
Live weight							
Ahlborn & Dempfle (1992)	0.39 (HF)	0.37 (HF)	0.34 (HF)	**0.24** (HF)	–	–	–
	0.29 (JE)	0.39 (JE)	0.34 (JE)	**0.16** (JE)	–	–	–
Spelman & Garrick (1997)	0.39	0.37	0.34	**0.24**	–	–	–
Veerkamp & Brotherstone (1997)	0.29	0.28	0.32	**0.44**	–	–	–
Kennedy et al. (1999)	–	–	–	**0.35**	–	–	–
Veerkamp et al. (2001)	–	–	–	–	–	–	–
Muller et al. (2006)	0.19	0.29	0.35	**0.65**	–	–	–
Pryce & Harris (2006)	0.28	0.36	0.33	**0.39**	–	–	–
Toshniwal et al. (2008)	-0.14	-0.07	0.11	**0.46**	–	–	–
Stature							
Ahlborn & Dempfle (1992)	0.34 (HF)	0.32 (HF)	0.25 (HF)	0.92 (HF)	**0.29** (HF)	–	–
	0.43 (JE)	0.46 (JE)	0.42 (JE)	0.85 (JE)	**0.23** (JE)	–	–
Spelman & Garrick (1997)	–	–	–	–	–	–	–
Veerkamp & Brotherstone (1997)	–	–	–	0.52	**0.50**	–	–
Kennedy et al. (1999)	–	–	–	0.61	**0.42**	–	–
Veerkamp et al. (2001)	–	–	–	–	–	–	–
Muller et al. (2006)	–	–	–	–	–	–	–
Pryce & Harris (2006)	–	–	–	–	–	–	–
Toshniwal et al. (2008)	–	–	–	–	–	–	–

Trait	Milk	Protein	Fat	Live weight	Stature	DMI	FCE
DMI							
Ahlborn & Dempfle (1992)	–	–	–	–	–	–	–
	–	–	–	–	–	–	–
Spelman & Garrick (1997)	–	–	–	–	–	–	–
Veerkamp & Brotherstone (1997)	0.59	0.64	0.76	0.37	0.13	**0.44**	–
Kennedy et al. (1999)	–	–	–	–	–	–	–
Veerkamp et al. (2001)	–	–	–	–	–	–	–
Muller et al. (2006)	–	–	–	–	–	–	–
Pryce & Harris (2006)	–	–	–	–	–	–	–
Toshniwal et al. (2008)	1.01	–	–	0.51	–	**0.07**	–
FCE							
Ahlborn & Dempfle (1992)	–	–	–	–	–	–	–
	–	–	–	–	–	–	–
Spelman & Garrick (1997)	–	–	–	–	–	–	–
Veerkamp & Brotherstone (1997)	–	–	–	0.27	–	–	–
Kennedy et al. (1999)	–	–	–	-0.53	-0.36	–	–
Veerkamp et al. (2001)	–	–	–	–	–	–	–
Muller et al. (2006)	–	–	–	–	–	–	–
Pryce & Harris (2006)	–	–	–	–	–	–	–
Toshniwal et al. (2008)	–	–	–	–	–	–	–

− Figures not published in literature

ǂ Milk, protein and fat are milk, protein and fat yield; DMI is dry matter intake; FCE is feed conversion efficiency.

with milk solids and milk volume. Veerkamp (1998) has also pointed out from past researches the correlations between stature and production of milk, fat and protein are 0.22, 0.16 and 0.25 respectively, thus making stature a suitable candidate as a predictor trait when selecting for milk solids and milk yield.

Live weight and feed intake are found to be correlated (Veerkamp & Brotherstone 1997; Kennedy et al. 1999; Toshniwal et al. 2008) and FCE has been reported to be genetically correlated with live weight with figures from -0.53 (Kennedy et al. 1999), to 0.27 (Veerkamp & Brotherstone 1997). The fluctuations in the genetic correlations occur as both of these traits are closely associated with mobilization of body reserves (Veerkamp & Brotherstone 1997), hence the dynamic relationship between live weight and FCE.

With the high genetic correlations between stature, chest width, body depth and rump height with live weight, and the low to moderate genetic correlation between chest width and body depth with dry matter intake estimated by Veerkamp & Brotherstone (1997), they suggested that body conformation traits can be used as indirect predictors for live weight and feed intake. These traits can be used alongside with production traits to construct a selection index since each breed societies score dairy cows for linear type traits due to registration purposes. Given the high accuracy in measurements for stature and it's correlation with live weight, condition score and feed intake; selecting for live weight with FCE determines the selection direction for body conformation traits (stature included), and will subsequently select for the optimum size of the dairy cow (Veerkamp & Brotherstone 1997).

Previous studies have identified significant correlations between stature with production and non-production traits in dairy cattle. In a research by (Schrooten et al. 2000) it was discovered that chromosome 5 contain QTLs for stature, other conformation traits (chest width, body capacity, rump width, udder depth, rear udder height, size), dairy character, birth weight and calving ease. Mc Hugh et al. (2012) discovered a strong genetic correlation with stature and muscularity traits. They also discovered moderate genetic correlations between skeletal type traits and weanling price for cattle in Ireland, indicating taller, deeper and wider cows are paid for at a greater price.

Mc Hugh et al. (2012) observed skeletal type traits (measurement/indicator for body size) have stronger association to the genetic merit of live-weight compared to muscularity type traits. They also identified moderate to strong genetic correlation for skeletal type traits and maternal weanling weight, pointing that larger and wider cows may have higher milk yield. Possible reasoning for this observation could be the animals used for this study has bigger feeding capacity on high forage diets, thus accounting for the high milk yield.

Coffey et al. (2003) found stature, chest depth, body depth and angularity as significant phenotypic predictors for live-weight in dairy cattle. Zink et al. (2011) observed strong genetic correlation for some reproduction traits and chest width, body depth and angularity which is also defined as dairyness in Czech Holstein cattle. They also found that cows portraying genetic extremes in angularity, stature and body depth tend to have poor fertility traits.

Stature is also used to calculate chest width which is related to body size, and body size influences calving ease. Schrooten et al. (2000) deduced genes of body conformation traits may have a pleiotropic effect on calving ease and birth weight and Cole et al. (2009) supported that theory. Meijering (1984) estimated the phenotypic correlations for external body measures with size of internal pelvic opening, and the figures are 0; signalling that large internal pelvic openings which promises easy calving do not come with having a large body size. Taller cows are associated with a lower number of pregnancy rates and require more services (Berry et al. 2004).

In a study for Spanish Holstein cows, Pérez-Cabal & Alenda (2002) found the genetic correlation between profit BV (defined as a function between production and herd lifetime) and stature BV was close to null (0.03). The conclusion was larger animals have higher feed demands which is an increase in cost. If the feeding is lacking for these animals, then fertility and calving will be a problem as well, affecting days in lactation as a consequence. Even though the correlation between profit BV and stature BV was almost zero, but taller animals will not return a negative profit. Therefore larger animals are profitable as well due to taller cows having a higher level of production that have a high correlation between body size and production (Short & Lawlor 1992). However, animals with high body depth values do not prevail in a herd as long as smaller animals, thus their

herd life is shorter. It is most likely the farmers will not keep the foresaid animals for long due to the high feeding cost.

Stature however was observed to have a positive genetic correlation with functional herd life of 0.15 in a study by du Toit et al. (2012). Boichard et al. (2003) suggested incorporating stature into the selection index to estimate BV for longevity. Cue et al. (1996) found indications in their data set that traits related to body size (weight, stature, body capacity) were positively correlated to survival in Jersey cattle. The study by Berry et al. (2005) also showed that cows that have a lower score of stature have a higher possibility of being culled.

Selection of stature affects the production of milk and milk solids, and influences both live weight and feed intake. Taller, deeper and wider cows have higher milk yield, however they tend to have poor fertility traits and low FCE. Smaller cows are more efficient in feed conversion, but possess lower functional herd life in some dairy farms. Identifying genes affecting stature will determine the selection for live weight, thus avoiding the selection for extreme heights (short and tall) in dairy cattle and breed for an overall balanced size of dairy cows that is considered optimum for the industry.

2.6 Finding quantitative trait loci for stature

2.6.1 Definition of quantitative trait loci

Quantitative trait loci are chromosomal segments that carry significant effect on a quantitative trait, and contribute to the variability of the trait (Dekkers & Hospital 2002). QTLs can be detected and mapped using markers for a better understanding of cattle physiology and gene regulation (Georges et al. 1995).

A single QTL however can only capture a small proportion of the genetic variance as complex traits are usually controlled by several QTLs. The accuracy for predicted BV via genetic selection will be less if only a part of the variance is accounted for (Kemper & Goddard 2012). Hence using several markers to construct a compact map of QTLs is important if all or most of the genetic variation needed to be accounted for.

2.6.2 Genetic markers used to detect quantitative trait loci

Genetic markers are deoxyribonucleic acid (DNA) fragments identified within the whole genome transmitted by Mendel's laws of inheritance, and the polymorphism at DNA sequence level labelled as SNP is used as DNA-based marker. SNP is defined as a difference that exists in the form of a single nucleotide/base pair within the DNA sequence (Hayes & Goddard 2010). SNPs are currently considered as the latest and more advanced molecular markers for genotyping studies. The usage of SNPs as genetic markers are now industry standard due to the lower mutation rate, easy plus large-scale genotyping at a reduced cost (Hinds et al. 2005), and these DNA variations are abundantly found in the genome (Snelling et al. 2005). Numerous studies have opted for SNPs to identify QTL for complex traits in many species such as in *Drosophila* fruit fly (Mackay 2001), poultry (van Kaam et al. 1999), red deer (Slate et al. 2002), sheep (Raadsma et al. 2009) and dairy cattle (Khatkar et al. 2004).

Unlike microsatellite mapping which only involves few hundreds of markers, up to tens of thousands of SNPs can be used in genome mapping studies (Cole et al. 2009). Significant associations between SNPs in the genome with a certain phenotype can be determined by conducting tests using the SNP chips and association tests. The assumption made by Hayes & Goddard (2010) is that significant associations arise because the SNP is in linkage disequilibrium (LD) with, and therefore close to, a causative mutation affecting the trait.

Genetic markers however should capture a large proportion of the genetic variance to ensure that the effects of QTL are determined with high accuracy (Kemper & Goddard 2012). However the LD relationship between markers and QTL is not sufficient enough if the genetic variance is affected by additive genetic effects. If the additive effect of causal polymorphism is not accounted for, the accuracy of genome estimated breeding value (GEBV) will be less than predicted. This error can be avoided if appropriate methods of variable selection based on the genetic effects are used (Meuwissen & Goddard 2010).

2.6.3 Linkage disequilibrium

Linkage occurs when the position of various loci that affects different traits are located together on a chromosome, thus it is unlikely that the genes will be distributed randomly during meiosis (Rauw et al. 1998). On the other hand, a single gene influencing two or more traits is called pleiotropy. The phenotype in any organisms is the effect of expression of genes through biochemical reactions controlled by gene regulations and control mechanisms. Thus the expression from different combination and effects of genes causes variation in the phenotypic expression.

Linkage disequilibrium occurs when there are non-random association of alleles at two or more loci (Barendse et al. 2007). The close genetic distance between markers on a chromosome indicate the markers have identical or similar genealogies, thus LD still can be detected even if the alleles are located at different markers (Pritchard & Przeworski 2001). Hayes & Goddard (2010) referred LD in genomic selection as associations between SNP alleles and alleles with mutations that affect the traits of interest. Such mutations that co-occur on segments of in individuals' chromosome can be traced back to a common ancestor with limited intervening recombination. High levels of LD exist in livestock species mainly due to domestication and breed formation (Farnir et al. 2000; Khatkar et al. 2006).

Linkage disequilibrium plays an important part in genome-wide association studies (GWAS) as measures of LD are widely used to quantify the degrees of association between markers (Pritchard & Przeworski 2001). To identify the causal mutation that affects the trait, such information on the structure and pattern of LD is important. Genetic markers that exploits LD with the QTL are detected by constructing either a fine map on genomic regions (Meuwissen et al. 2001) or a fine map on whole-genome QTL (Barendse et al. 2007) using dense markers. A good understanding of LD structure across the genome can enhance our knowledge of the biological processes of recombination and selection in the genome, thus the design and analysis of genome-wide association studies can be gradually improved.

2.7 Genome-wide association studies

The definition of GWAS set by Hayes & Goddard (2010) is finding associations between variants of SNPs with a specific trait by assuming the SNP is in linkage disequilibrium with the causative mutation causing variation in the trait. Using SNP genotypes and associating the differences in phenotypes to find the gene that controls a specific trait, GWAS enable us to study the whole genome instead of just focusing on a single candidate gene.

Gene effects are accumulative with a lot of genes, instead of just a single gene contributing to the effect (a single gene just contribute proportional variation). Individual genes are likely to have small effects and therefore a large amount of genomic data is needed to accurately estimate their effects (Goddard & Hayes 2007). Tracking a small number of these genes through these DNA markers will only explain a small proportion of the genetic variance.

Genome-wide association studies need a large sample of data to find the associations between genes, and also a lot of SNPs to examine all the variations. Statistical associations between a specific trait and the SNPs are being detected by examining a sample of animals recorded for the trait using a genome-wide panel of SNPs.

2.7.1 Genetic markers associated with stature

Table 2.3 that shows a list of markers that have been reported to be associated with stature in cattle chromosomes. This table was constructed by referencing various studies on QTL detection (Spelman et al. 1999; Schrooten et al. 2000; Boichard et al. 2003; Ashwell et al. 2005), QTL mapping (Hiendleder et al. 2003; Kolbehdari et al. 2008; Cole et al. 2009) and GWAS (Bolormaa et al. 2011; Pryce et al. 2011; Littlejohn et al. 2012) for nonproduction traits in both beef and dairy cattle.

Table 2.3: Genetic markers linked with stature in beef and dairy cattle.

Chromosome	Marker name	Position (Mbp)	References
1	rs41581655	95.7	Kolbehdari et al. (2008)
2	TGLA377	36	Boichard et al. (2003)
	TGLA377-URB042	24	Ashwell et al. (2005)
	*	20	Bolormaa et al. (2011)
3	rs41572038	57.6	Kolbehdari et al. (2008)
	*	102.159-109,411	Bolormaa et al. (2011)
4	BMS1634-MAF70	28	Ashwell et al. (2005)
	rs29011323	56.1	Kolbehdari et al. (2008)
5	IGF1-BM315	123	Schrooten et al. (2000)
	BM315	124	Boichard et al. (2003)
	rs41592968	10.2	Kolbehdari et al. (2008)
	rs29014633	74	Kolbehdari et al. (2008)
	*	112.3	Cole et al. (2009)
	*	120.9-121.5	Bolormaa et al. (2011)
6	MCM53	11	Schrooten et al. (2000)
	*	38.4, 38.5, 38.5	Pryce et al. (2011)
	BM1329	54	Boichard et al. (2003)
	*	66	Hiendleder et al. (2003), Bolormaa et al. (2011)
7	rs41591943	39.9	Kolbehdari et al. (2008)

8	*	84.9	Pryce et al. (2011)
	*	111.4	Pryce et al. (2011)
	TGLA13	127	Schrooten et al. (2000)
11	BMS1822	82	Boichard et al. (2003)
	rs29026038	63.3	Kolbehdari et al. (2008)
13	ABS10	74	Boichard et al. (2003)
14	*	22.8	Pryce et al. (2011)
	*	23.5	Pryce et al. (2011)
	BMS1941	36	Spelman et el. (1999)
	ss319607402	8.1	Littlejohn et al. (2012)
15	BMS2684-HBB	37	Ashwell et al. (2005)
18	*	0	Hiendleder et al. (2003)
	ss86324977	57.1	Cole et al. (2009)
19	rs41640016	16.3	Kolbehdari et al. (2008)
	Centro-BM6000	0	Ashwell et al. (2005)
22	BMS875-BM4102	72	Ashwell et al. (2005)
23	Centro-CSSM5	0	Ashwell et al. (2005)
24	*	34.43, 34.45, 34.6	Pryce et al. (2011)
	rs41567447	39.8	Kolbehdari et al. (2008)
27	TGLA179-BM871	6	Ashwell et al. (2005)
29	*	51.5	Cole et al. (2009)

* Marker name is not specified in the literature.

2.7.2 Genetic markers associated with live weight

Table 2.4 displays a list of markers that were reported to be associated with live weight in cattle. The table was compiled using references from studies on QTL detecting (MacNeil & Grosz 2002), QTL mapping (Elo et al. 1999) and GWAS (Peters et al. 2012) for live weight in beef and dairy cattle.

Table 2.4: Genetic markers linked with live weight in beef and dairy cattle.

Chromosome	Marker name	Position (Mbp)	References
1	ss61522637 - ss117966342	109.6 - 109.9	Peters et al. 2012
	ss86318478 - rs41897673	114.0 - 114.1	Peters et al. 2012
	ss117975147 - rs43275053	138.9 - 139.4	Peters et al. 2012
3	ss86303944 - ss64552970	90 - 91.1	Peters et al. 2012
6	ss86338115 - rs42005088	55.0 - 55.2	Peters et al. 2012
9	*	44	MacNeil and Grosz. 2002
	rs42516892 - rs42937117	57.4 - 57.7	Peters et al. 2012
10	ss117969863 - ss86273438	33.1 - 33.4	Peters et al. 2012
12	*	74	MacNeil and Grosz. 2002
16	*	22	MacNeil and Grosz. 2002
	ss86312250 - rs41787407	0.4 - 0.5	Peters et al. 2012
	ss61536681 - ss105263670	3.1 - 3.3	Peters et al. 2012
17	*	52	MacNeil and Grosz. 2002
18	*	78	MacNeil and Grosz. 2002
20	ss117972683 - ss105265377	8.4 - 8.5	Peters et al. 2012
22	rs42011564 - ss86335704	47.2 - 47.4	Peters et al. 2012
23	*	25	Elo et al. 1999
29	*	38	MacNeil and Grosz. 2002
	rs42161771 - ss86320129	2.0 - 2.3	Peters et al. 2012
	rs42171465 - ss86286136	25.9 - 26.0	Peters et al. 2012

* Marker name is not specified in the literature.

2.8 Genomic selection

Genomic selection is the selection of individuals using genomic information in the form of GEBV. Compared to selection based on phenotypic records alone, genomic selection using data on DNA level promises accelerated genetic gain. Genomic selection does not attempt to identify functional mutations but uses a random set of genome-wide markers to predict BV. According to Hayes & Goddard (2010), this technique utilises a panel of dense markers widely distributed across the genome to ensure all QTLs to be in LD with at least one marker (SNP). The genetic variance from QTL that has been found significant with associated markers and contributes to selection is taken into account, and BVs are estimated based on the total effects of molecular markers across the genome (Meuwissen et al. 2001) rather than using progeny testing which is based on family information and phenotypic information or performance of the animal. Linkage disequilibrium between QTL and markers ensures both the QTL and marker alleles persists across generations (de Roos et al. 2008), which are then exploited in genomic selection. With molecular markers highlighting the location of useful traits, it is therefore easier for selection of important traits in any breeding program and to predict BV.

Genomic selection is a very appealing method for animal selection. Selecting elite animals based on their genotype information rather than phenotype increases the selection efficiency and the rate of genetic gain tremendously (Schaeffer 2006) as the intensity of selection is increased and generation interval is decreased. Dekkers & Hospital (2002) stated that genomic selection has fewer limitations than quantitative genetic selection as extraction of DNA is not restricted by age or sex. Thus genomic selection is most effective to select for traits early in life as it is not too affected by micro-environmental variation (Soller 1994). Estimating BV for traits like milk production that is only restricted to one sex can be implemented as well with genomic selection.

A reference population of animals is first put together for phenotypic measurements of the target trait and each individual is genotyped in each of thousands of SNP widely distributed in the genome. Substitution effects at each of the SNPs are estimated assuming that there is LD between a gene affecting the trait and SNP.

The model proposed by Meuwissen et al. (2001) to estimate the substitution effect at each SNP in the reference population is as below:

$$y_i = \mu + \sum_j X_{ij} b_j + e_i$$

where y_i = mean phenotypic measurements of animal i;

μ = fixed effects

$\sum_j$ = net effects of all SNP

X_{ij} = number (0,1 or 2) of copies of allele '1' (versus '0') that animal i carries at SNP j

b_j = Allele substitution for SNP j

e_i = random residual

The prediction equation will then be applied on a validation population to assess the accuracy of the equation. The validation population can have lesser animals than the reference population (Goddard & Hayes 2007), but still needed to be assayed for phenotypic and genotypic data to test for the accuracy. Then the equation will be implemented on the selection candidates that have genotypes but with no phenotypes to predict their BVs, which is known as GEBV (Meuwissen et al. 2001). The accuracy of the GEBV of the selection candidates is assumed to be the same as the validation population (Goddard & Hayes 2007).

According to Lillehammer et al. (2011), the correlation between GEBV and the true BVs of an animal is calculated as the accuracy of the GEBV. Therefore the correlation between the GEBV and the true BVs of the selection population can be estimated. The square value of the correlation can be used to measure the reliability of the GEBV (Lopez-Villalobos 2012), thus ensuring the accuracy and reliability of the GEBV. The market value of an animal increases with the accuracy of the BV as an accurate figure of BV guarantees the performance of the animal and reduces the risk of a large variation between selected animals. Needless to mention is the more accurate the BV is, the more definite rate of genetic gain can be achieved for the breeding scheme.

A reference population of individuals with both genotypic and phenotypic records is vital to implement genomic selection (Hayes & Goddard 2010). A large number of animals with marker genotypes and phenotypes are required for an accurate estimate of QTL effects. Meuwissen et al. (2001) found that for a trait with a heritability of 0.5, the accuracy of GEBV was higher when 2000 records were used to estimate QTL effects than when 1000 records were use. If the structure of LD is unknown or not favourable (for example larger ancestral effective population size), more records might be needed (Goddard & Hayes 2007).

The accuracy of GEBV depends on three factors provided if the markers are dense enough. First, the number of genotyped and phenotyped animals from the reference population (Meuwissen et al. 2001). The second factor is the heritability of the trait, which can be improved by either using phenotypes of higher accuracy (Pryce & Daetwyler 2012), repeated measurements on individuals (Aguilar et al. 2011), related animals but with no genotypes (Misztal et al. 2009) or including genetically correlated traits for bivariate analysis (Bolormaa et al. 2010). Thirdly, the number of loci that influences the trait affects the accuracy of GEBV too (Kemper & Goddard 2012).

2.9 Genome imputation

Utilising the genome map of fully genotyped animals as a reference to identify and predict the missing alleles in genome sequences of animals that are being studied is known as genome imputation. The reference data set comprises of very dense set of SNPs and the animals in the study data set are genotyped using a subset of the SNPs (Marchini & Howie 2010). The genotypes from the study data set will be compared with the reference set and imputation attempts will take place by predicting and imputing the missing genotypes in the studied population. This process involves running a population genetic model to deduce the allele correlations measured in the reference set (Howie et al. 2009). Even if the individuals are not related, they are still identical by selection therefore the haplotypes in short genome sequences are related to one another.

Howie et al. (2009) stated that the data sets used for imputation of SNPs can be separated into two sets, namely as typed data set (T) and untyped data set (U). Data set T consists of haplotypes identified in both reference and study set while data set U only has haplotypes in the reference set. Most imputation methods align the study set to data set T to identify identical or similar haplotypes in the SNPs. It is then assumed that haplotypes corresponding with SNPs at data set T are identical with the haplotypes in data set U too.

Several important issues raised from this basic principle of genome imputation must be scrutinised before proceeding with the imputation. First, the phasing of the study set to data set T must be accurate to ensure the correct haplotypes are inserted in the missing positions. The accuracy of imputation at SNPs at data set U is likely to be influenced by this issue too. The second issue is it takes huge computational effort to for the phasing stage; however the imputation stage at U will be faster when the unknown haplotypes are imputed when phased with T. Thirdly, all available information should be used in the model algorithm to account for all unknown haplotypes of the study set with T.

Genome imputation completes missing haplotypes in incomplete genotype files. A more complete genotype data improves the association signal of SNPs by 10% (Spencer et al. 2009) and decreases the false discovery association rate, thus ensuring better results for GWAS of complex traits. This imputation method can be used too in focusing on a certain genome region in fine mapping studies and increase the probability to directly identify a causal SNP. It is also able to reconstruct genotypes from low coverage genome sequencing results.

Improvements for genotype sequencing and GWAS are eminent, with larger sample sizes, more incomplete genotype data and more reference sets of SNPs involved (Howie et al. 2009). Having access to more haplotypes available on larger populations in the future will enable such improvements to take place, and offering a more detailed view of the genetic variation in the population.

2.10 Impacts of genomic selection

Using genomic selection for the prediction of BV will remove the limitations that have troubled animal breeders for a long time. Genetic merit of animals for sex-limited traits like milk yield in dairy cattle or hard-to-measure traits like feed efficiency can be evaluated using genomic markers, and then predict the GEBV from the prediction equation.

The rate of genetic gain for selection can be derived from a function of selection intensity with accuracy of selection and genetic standard deviation over the generation interval between parents and progeny (Falconer & Mackay 1996). The selection intensity is determined by the proportion of available animals selected as replacements. Accuracy of selection is a factor derived from the regression of true on estimated genetic merit. The genetic standard deviation measures the extent of genetic variation among animals in the population. This factor provides a real amount of genetic material available in the population. The generation interval is a measure of the time the progeny will be allowed to replace parents.

Using genomic markers to predict BV has been proven to increase the rate of genetic gain by two-fold in the dairy industry (Schaeffer 2006) as genetic evaluations can take place once the progenies are born. This increases the selection intensity for complex yet relevant traits that is not usually recorded on the herd like fertility traits. At the same time the progeny test is also omitted and this greatly shortens the generation interval for dairy cattle. The same study by Schaeffer (2006) showed the leap in genetic gain is mainly due to the reduced generation interval as elite sires were selected at a young age based on their GEBV alone.

Genomic selection provides more accurate BV for young animals compared with conventional best linear unbiased prediction (BLUP) BVs (Daetwyler et al. 2007). Daetwyler et al. (2007) also state that the high rate of genetic gain contributed by genomic selection will not elevate inbreeding as increased accuracy of BV at an early stage makes it possible to increase genetic gain without increasing selection intensity.

BV with low accuracy increases the risk of animals with poor performance, rate of production, and at the same time increases the variation between selected animals. Thus

highly accurate BV is important to improve the genetic gain and also for marketing purposes.

A breeding programme normally selects for various traits that comes with a wide range of heritability, and for the dairy cattle industry in New Zealand, the total merit index for BW has a heritability of about 0.22 (NZAEL 2012). Genomic selection has more advantage especially when selecting for traits with low heritability. The study conducted by Lillehammer et al. (2011) showed that the total genetic gain increases for traits with a range of heritability when genomic selection is implemented. Heritability does not affect the accuracy of GEBV compared to the conventional BVs predicted using BLUP, making genomic selection more suitable when selecting for a trait with low heritability than compared to progeny testing. When the heritability for a desirable trait is low, the information from related animals will be more relevant than the phenotypic information of the selected animal itself. The information from relatives in GEBV which is made possible by the marker data, will be incorporated into selection. Using SNP data, more detailed information at DNA level is used to calculate the BV, hence guaranteeing its accuracy.

As the rate of genetic gain improves the most for traits with low heritability, the implementation of genomic selection will tremendously increase the total genetic gain for a breeding programme (Lillehammer et al. 2011). This is vital for a more sustainable breeding scheme for dairy cattle as most functional traits such as disease resistance and reproduction have low heritabilities and they are important for economical purposes and animal welfare.

Goddard & Hayes (2007) proposed that a more accurate BV can be predicted than is possible with pedigree and phenotypes alone because genomic selection can explain the actual relationship matrix than the average relationship as the Mendelian sampling effect during gamete formation is taken into account. Non-additive genetic effects like epistasis and dominance are included as well if markers are used to forecast the genotypic and phenotypic value (Goddard & Hayes 2007).

Lillehammer et al. (2011) also pointed out that omitting progeny testing and using purely genomic selection to obtain BV is more cost effective. Daetwyler et al. (2007)

showed that genomic selection reduces the rate of inbreeding as well compared to using progeny testing as closely related animals feature different BV.

When evaluating the effect of a certain gene on a specific performance, effects of nearby genes will too be accounted for due to LD (Pribyl et al. 2010). This adds up to the overall effect caused by the identified gene and the relationship between markers and QTLs. However if the markers are too sparse to cover the entire genome, effects of the used markers are normally overestimated. This problem is solved with genomic selection that implements a panel of dense genetic markers (SNP) to analyse the whole genome. More QTLs can be identified and thus taking polygenic effects into account as well to determine the underlying causal gene that influences the target trait.

Genomic selection can disregard the genotype × environment interaction (G×E) that affects breeding programmes. Different countries have their own breeding programmes, and thus the genetic merit of their animals is different. When genetic correlations between productions are more than 0.8 in two separate environments, such as two countries are detected, different animals tend to be selected for the two countries. If the evaluation of animals can only be carried out in their own environment, the two populations are considered to be diverged. However, if the evaluation of animals can be carried out equally in both populations, then animals that are adaptive to either environment can be selected from one population, thus decreasing the size of the total effective population (Goddard 1998). This means that all selection candidates with their known prediction equation can be evaluated no matter which environment they are at. Therefore many divergent populations where each is specialized to a particular environment can be substituted by one general population, thus causing the overall effective population size to be smaller.

A few disadvantages of genomic selection exist. Pribyl et al. (2010) pointed out that at least one generation separates the reference population and the selection population. Therefore the linkage between SNP and QTLs discovered in the reference population may not be able to be applied fully on the younger evaluated population, as the markers may also not be dense enough to fully discover the QTLs that influences the traits. The effects of selection, mutation and intense moving of breeding sires for artificial insemination can alter

the QTL information across generations. Other factors that also will affect this are the development of the commercial population under selection and environmental changes.

It is definitely enticing to run genomic selection completely and widely in livestock breeding to achieve maximum genetic gain. However the total cost of genotyping the whole reference population in high-density SNP panels is still considerably high and that deters most farmers and breeders so far. One of the current solutions is to use low-density panels to predict the GEBV. In a study by Weigel et al. (2009), the results show that low-density panels with selected markers outperforms low-density panels with evenly spaced markers for the lifetime net merit. Genome imputation can be used to predict and insert the missing haplotypes in the genotype results if the genome coverage from the low-density SNP panel is too low.

Even though the genetic gain promised by genomic selection outperforms the genetic gain from progeny testing, progeny testing still carries some importance as the conventional BVs derived from BLUP are more accurate for older animals that are progeny tested. Genomic selection however should be implemented for young animals as their GEBV are more accurate than the BVs estimated from pedigree information (Lillehammer et al. 2011).

CHAPTER 3

MATERIALS AND METHODS

3.1 Data

3.1.1 Phenotypes

The data set used for this study was provided by Livestock Improvement Corporation (Hamilton, New Zealand) and included the animal key, sire code of the bull, birth date, breed code, proportion of Holstein-Friesian (HF) genes, proportion of Jersey (JE) genes, percentage of North American Holstein (NAH) genes, EBV for live weight and stature and their reliabilities of 3140 bulls.

The main breed groups were Holstein-Friesian (HF), Jersey (JE) and HF × JE crossbred (XB), bulls from other breeds and their crosses were classified into a single group called Others (OT). Due to the small number of bulls available for the OT, they were not included for GWAS in this study.

3.1.2 Genotypes

The DNA genotypes were originally from a 50k SNP chip, and were imputed up to a 700k SNP chip by the Livestock Improvement Corporation (Hamilton, New Zealand). The 700k SNP panel has higher density of markers to cover the genome more effectively compared to 50k SNP panel. The 700k SNP panel has 692598 validated markers for this study.

Two files were obtained from the Livestock Improvement Corporation (Hamilton, New Zealand) for the genotypes in this study. They were files in the standard text based format of PED and MAP. The PED file contained SNP genotype information while the MAP file has the name and position of markers in the PED file.

The PED file consists of 2770425 columns and 3140 rows. Each row represents each of the 3410 bulls in this study. The first column is the family identification (ID) which

is not included in this study and has the value of 0. The second column is the individual ID, the third column is the paternal ID, and the fourth column is the maternal ID. The fifth column specifies the sex of the animals, and for this study the value is all 1 which stands for male as all the animals are bulls. The sixth column is for the phenotype. The values for this column is -9, which stand for missing values as the phenotypes for this study are in a separate file. The rest of the columns onwards are the genotype of each of the markers. The markers must be biallelic; therefore all haplotypes detected in the genome have the values of either 1 or 2 specified in the PED file. The markers must have two alleles specified whether they are haploid or not, therefore the numbers "1" and "2" represent the different alleles (allele "1" versus "2") the individual carries.

The MAP file consists of 692598 lines and 4 columns. Each line is the SNP marker found in the 700k SNP panel. The first column represents the number of chromosome, which in this case, has value from 1 to 29 as the cattle genome consists of 29 autosomal chromosomes. The sex chromosomes are not investigated for this study. The second column is the identifier or name of the SNP markers. The third column stands for the genetic distance in the unit of Morgans. The value for this column is set at 0 for the association test in this study. The last column represents the base-pair position of the SNP detected in each chromosome in the unit of base pairs.

3.2 Statistical analysis

Descriptive statistics of bull EBV for live weight and stature were obtained by year of birth and breed. The medians, means and standard deviations were derived using R software version 2.14.1 (R Development Core Team, 2011). The first and third quartiles, plus the minimum and maximum values were calculated as well.

Scatter plots plotting the average EBV for live weight and stature against the years for each breed were plotted using Microsoft Excel. A linear regression model for stature and live weight EBV of all breeds was derived using R software (Ver. 2.14.1; R Development Core Team, 2011), and a correlation graph depicting the correlation between stature and live weight EBV was plotted as well.

3.3 Genome-wide association studies

3.3.1 Adjustment of phenotype

Before running the GWAS, the EBV for stature were adjusted with a multiple regression model using the MIXED procedure of SAS software version 9.3 (SAS Institute, Cary, NC) to adjust for percentage of HF (HF_{pct}), JE (JE_{pct}), NAH (NAH_{pct}) genes and year of birth – (scaled to 1994).

The multiple regression model used is as below:

$$y_i = \beta_0 + \beta_1 HF_{pct} + \beta_2 JE_{pct} + \beta_3 NAH_{pct} + \beta_4 year + e_i$$

where:

y_i	is the EBV of stature of bull i where i = 1 – 3140;
$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$	are the regression parameters; and
e_i	is the random residual effect unique to y_i

Predicted and residuals values from the multiple regression model were plotted in SAS software (Ver. 9.3; SAS Inst., Cary, NC). Residual values were used as phenotypes for GWAS.

3.3.2 Estimation of single nucleotide polymorphism effects

The PED and MAP files were converted to a compact binary PED file format (BED) in PLINK software version 1.07 (Purcell et al., 2007, http://pngu.mgh.harvard.edu/purcell/plink/) for the association analysis. Two other files were also created in the conversion process, the FAM and BIM file. The FAM file stores the pedigree/phenotype information while the BIM file is an extension of the MAP file, and contains information on the allele names.

The BED file is a compressed binary file with codes, while FAM and BIM files are standard text files. The FAM file in this study comprises of 3410 rows and 6 columns. Each row represents the bulls used for this study, and the 6 columns are the same as the first 6

columns found in the MAP file, which comprises of the family ID, individual ID, paternal ID, maternal ID, sex of the animal, and the phenotype. The BIM file contains of 692598 lines and 6 columns. Like the MAP file, each line represents the SNP markers. The first 4 columns stand for the chromosome number, SNP identifier, genetic distance and base pair position, and the extra two columns represents the allele names.

Genome-wide association studies to identify QTL affecting stature were conducted by running a quantitative trait association test with PLINK software (Ver. 1.07; Purcell et al., 2007) for HF, JE and XB populations. Asymptotic significance values using likelihood ratio test and Wald test were implemented to test for association in quantitative traits. A standard linear regression of phenotype (stature EBV) on genotype (SNPs) that ignores family structures within the population was used for this association test. The test ignores family structures by implementing a permutation procedure to avoid the dependence value among individuals within a family. Genotypes were decomposed from a family structure to form a new pseudogenotype value for each individual for the association analysis.

Results in the form of .qassoc file were generated with 692598 lines and 9 columns. Every line is the SNP marker found in the 700k SNP panel. The first column is the chromosome number for the SNP marker, the second column is the SNP identifier while the third column denotes the physical position of the SNP in base pair format. The fourth column shows the number of non-missing genotypes. The fifth column represents the regression coefficient of every SNP. The sixth column shows the standard error and the seventh column is the r-squared value of the regression. The eighth column denotes the *t*-value from the Wald test (based on *t*-distribution) while the last column consists of the asymptotic *P*-values of the same Wald test. The *P*-values obtained from the analysis were adjusted according to the genomic control factor in PLINK software (Ver. 1.07; Purcell et al., 2007).

3.3.3 Manhattan plots

Manhattan plots of *P*-values adjusted to genomic control against the chromosomes were obtained using R software (Ver. 2.14.1; R Development Core Team, 2011) with a code written by Turner (2011). Two threshold lines were plotted over the Manhattan plots to highlight the significant SNP that is associated with stature or live weight, and have already been adjusted for HF_{pct}, JE_{pct}, NAH_{pct} and year of birth. One is the genome-wide significant threshold line and was set at a negative $\log_{10}$ value of 5×10^5. The other is the suggestive threshold line and was set at a negative $\log_{10}$ value of 1×10^4.

Three Manhattan plots for HF, JE and XB were plotted, and multi Manhattan plots for each breed were plotted as well using gwasplot package (Creagh and Sherlock, 2012) in R (Ver. 2.14.1; R Development Core Team, 2011). The top 50 SNP with the most significant *P*-values above the significant threshold line were then highlighted.

CHAPTER 4

RESULTS

4.1 Descriptive statistics

Descriptive statistics for EBVs for stature and live weight of bulls born between 1994 and 2006 was shown in Table 4.1. HF had the highest values for both live weight (103.8kg) and stature (2.48 score units) EBV compared to the other breeds. JE had the lowest EBV for live weight (-103.3kg) and stature (-2.12 score units) in the data set. Overall, HF is the only breed from the data set that had positive values for live weight and stature EBV.

Table 4.1: Descriptive statistics for estimated breeding values (EBV) for live weight and stature of Holstein-Friesian (HF), Jersey (JE) and HF × JE crossbred (XB) bulls born between 1994 and 2006.

EBV	Mean	Std Dev.	Median	Min	Max	1st Qu.	3rd Qu.
HF							
Live weight (kg)	44.7	18.3	43.7	-13.7	103.8	32	57.1
Stature (score)	0.82	0.46	0.8	-0.54	2.48	0.52	1.12
JE							
Live weight (kg)	-58.6	13.5	-58.6	-103.3	-21.9	-67.8	-49.7
Stature (score)	-1.04	0.29	-1.04	-2.12	-0.17	-1.23	-0.84
XB							
Live weight (kg)	-8.9	20.6	-8.7	-71.3	58.6	-23.2	4.2
Stature (score)	-0.2	0.41	-0.2	-1.71	1.17	-0.48	0.05

Scatter plots for the average EBV plotted against year of birth for each breed were shown in Figures 4.1 and 4.2.

The average EBV for live weight in HF population (Figure 4.1) throughout year 1994 until 2006 had shown an overall reduction though there are peaks of increase on year

1997 and 2006, and also a slight curve upwards from 2001 to 2004. The trend for JE population (Figure 4.1) however showed a total opposite as the average EBV increases for the 13 years. There were fluctuations as well with decreasing values in 1995 and 1996, then the numbers increased until 2000. The figures gradually lowered from 2001 to 2004, and picked up in 2005 and 2006. A general positive trend for live weight EBV was observed in XB bulls (Figure 4.1). There was a rapid increase from 1994 to 2000 and thereafter the average EBVs remained constant until 2006.

For average stature EBVs, the HF data showed an overall downward trend from 1994 to 2006 (Figure 4.2), with a significant increase in both 1997 and 2006, and also an upward curve from 2000 and peak in 2004. The average stature EBV for JE population in Figure 4.2 showed an increase from 1994 to 2006 with a very significant drop in 1996 and a continuous decrease in 2001 until 2004 before a sharp increase in 2005. The R^2 value of the regression line was 0.29, indicating the trend line is questionable. The trend of stature EBVs in XB population also showed large fluctuations (Figure 4.2) especially from 1997 to 2000, with sudden rises and falls in the average EBV. After that, the values increased gradually at a stable rate, possibly due to more datas collected between 2000 to 2006. The R^2 value for the linear regression line was 0.14, indicating that this linear trend is questionable too.

Similar trends were observed for the trends in average EBV for live weight and stature. Average stature EBVs in HF population decreased from 1994 to 1996, which increased in 1997, then a slight curve growth through out 1998 until it reached a peak in 2004, an average drop in 2005, and rebound in 2006. The trend in average stature EBV in JE population is similar to the trends in average live weight EBVs. The lowest average was in 1996, and a gradual rise to a peak in 2000, then a decrease until 2004, and finally another high peak in 2005. The average EBV for live weight and stature in XB population also shared a common pattern apart from the data collected in 1994, where the average EBV for live weight was very low compared to the figure in 1995 while the average EBV was higher than the value in 1995 for stature.

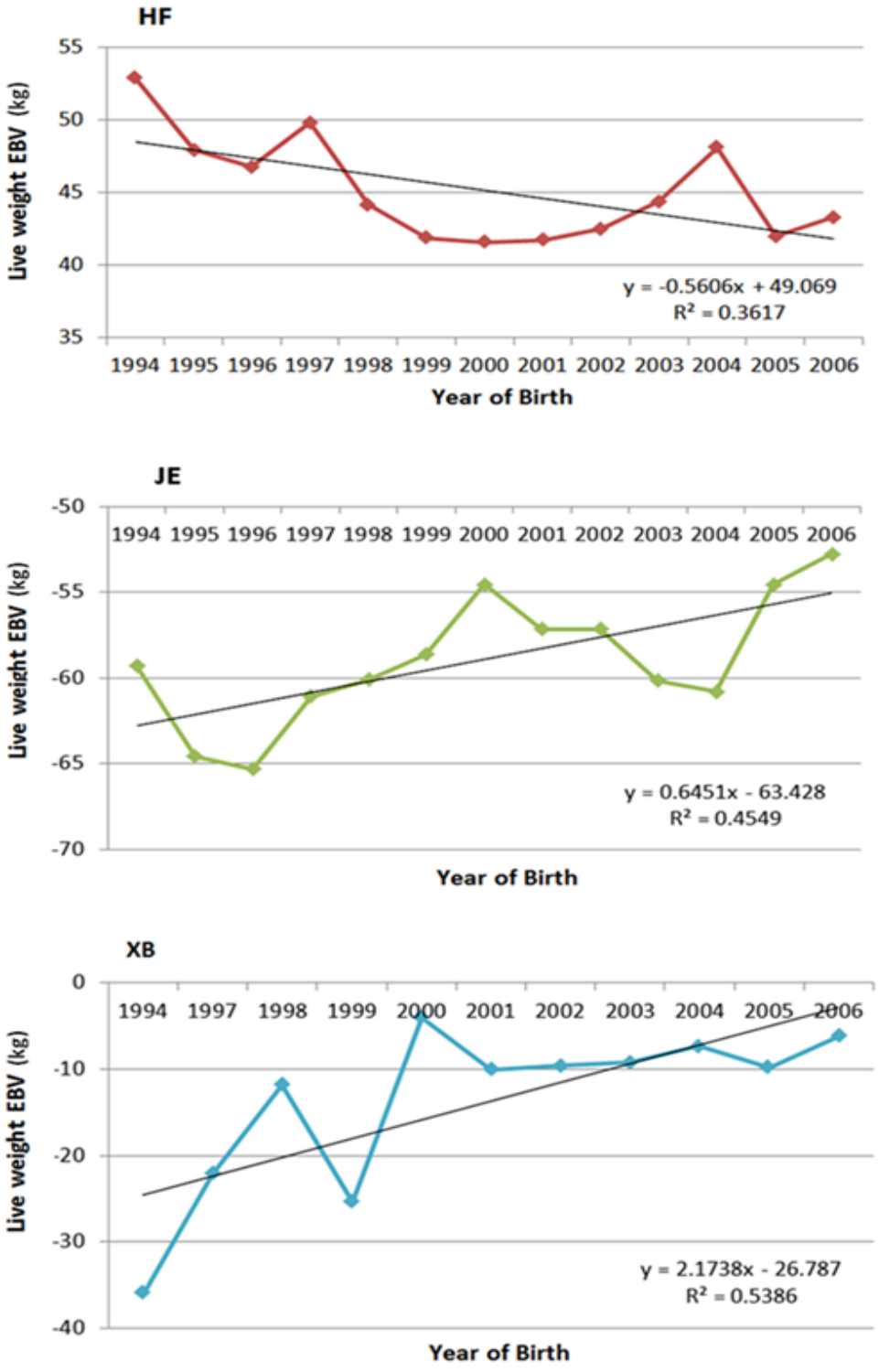

Figure 4.1: The average breeding values for live weight EBV per year of birth (1994-2006) for Holstein-Friesian (HF), Jersey (JE) and Holstein-Friesian × Jersey crossbred (XB) population.

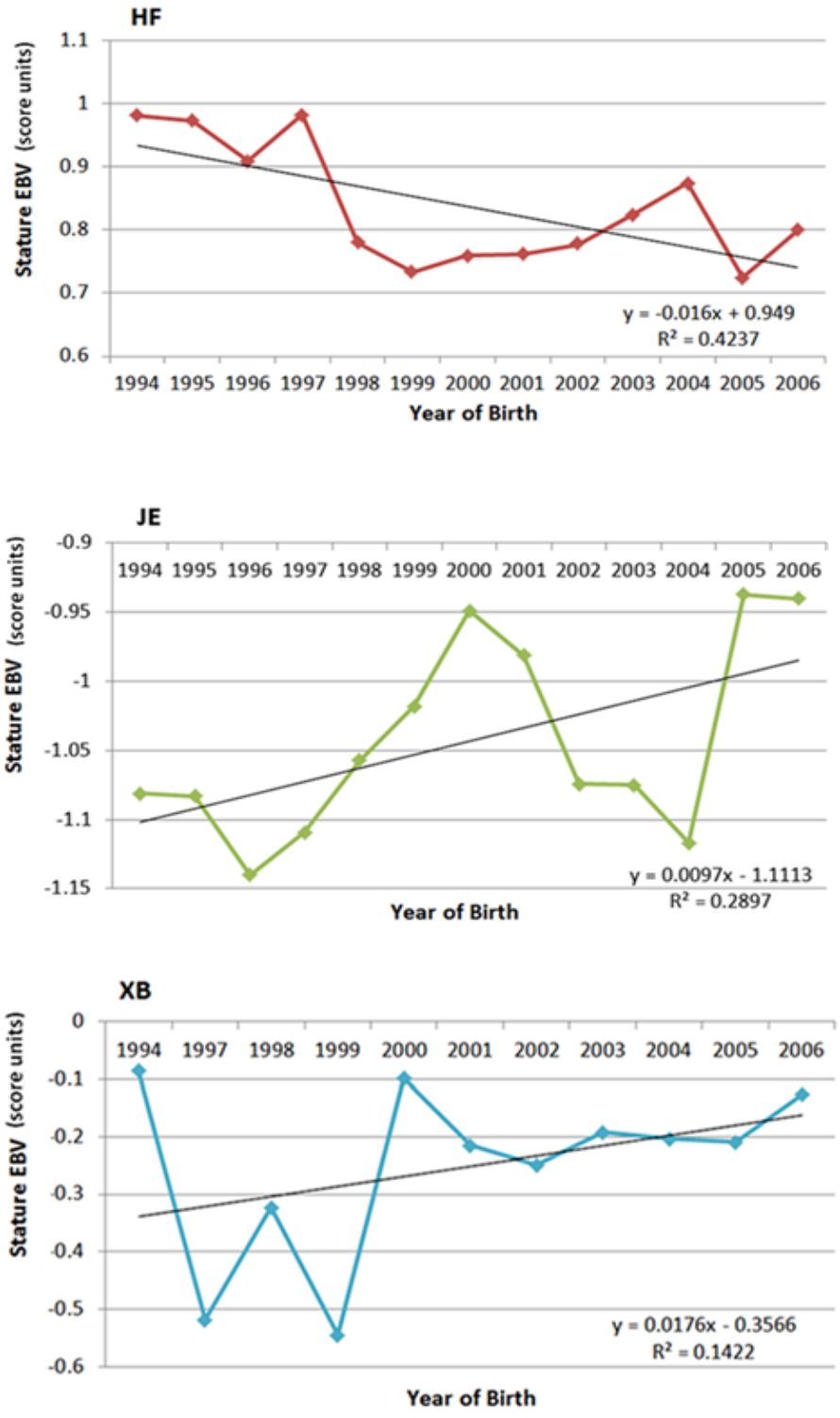

Figure 4.2: The average breeding values for stature EBV per year of birth (1994-2006) for Holstein-Friesian (HF), Jersey (JE) and Holstein-Friesian × Jersey crossbred (XB) population.

The correlation between stature and live weight EBV of bulls from all the breeds in this study was 0.98 with *P*-value of 2×10^{-16} indicating that the correlation is significantly different to zero (Figure 4.3). The regression coefficient of stature EBV on live weight EBV was 0.018 with R^2 value of 0.96, indicating that stature EBV increases by 0.018 per one kilogram of live weight EBV.

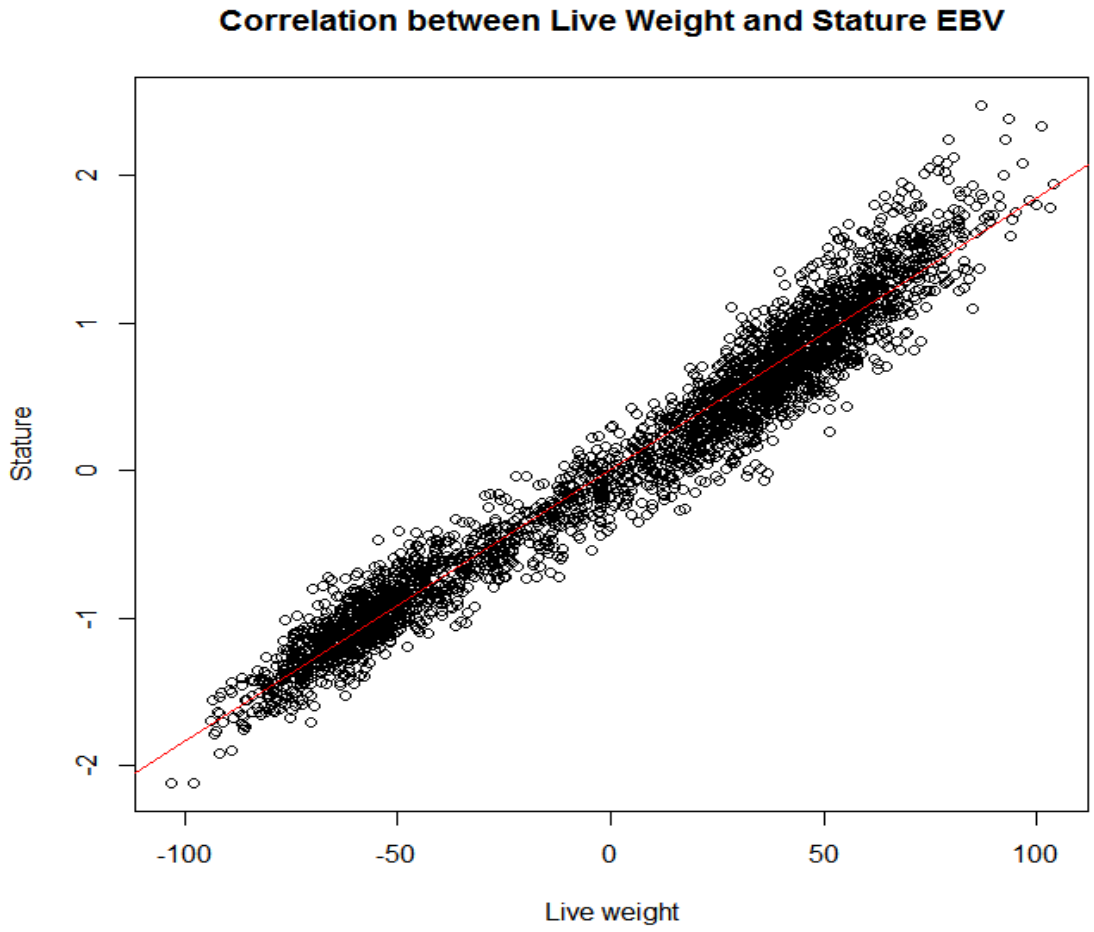

Figure 4.3: Correlation between cow stature and live weight EBV of bulls from all breeds.

Table 4.2: Estimates of regression coefficients of estimated breeding value for live weight on percentages of Holstein-Friesian (HF_{pct}), Jersey (JE_{pct}) and North American Holstein (NAH_{pct}) genes and year of birth (Year).

Parameter	Estimate	Standard Error	*t*-value	*P*-value
Intercept	-0.036	4.732	-0.01	0.994
HF_{pct}	0.228	0.049	4.68	<.0001
JE_{pct}	-0.592	0.047	-12.53	<.0001
NAH_{pct}	0.345	0.017	20.64	<.0001
Year	0.116	0.08	1.45	0.148

Percentage of North American Holstein genes was observed to affect live weight EBV in cattle (P <0.0001). The model indicated that with every 1% increase of NAH genes, live weight EBV increases by 0.345 kilograms. Both HF_{pct} and JE_{pct} also have high significance (P<0.0001) effect on EBV for live weight. Live weight EBV increases by 0.228 kilograms per 1% of HF_{pct}. However, live weight EBV decreases by 0.592 score units with every 1% increase of JE_{pct}. Live weight EBV was not linearly affected by year of birth (P = 0.148).

Table 4.3: Estimates of regression coefficients of estimated breeding value for stature on percentages of Holstein-Friesian (HF_{pct}), Jersey (JE_{pct}) and North American Holstein (NAH_{pct}) genes and year of birth (Year).

Parameter	Estimate	Standard Error	*t*-value	*P*-value
Intercept	-0.2409	0.1047	-2.3	0.022
HF_{pct}	0.0031	0.0011	2.87	0.004
JE_{pct}	-0.0079	0.0011	-7.6	<.0001
NAH_{pct}	0.012	0.0004	32.56	<.0001
Year	-0.0001	0.0018	-0.06	0.952

The regression coefficient of stature EBV on NAH_{pct} was 0.012 (Table 4.3), indicating that stature EBV increases by 0.012 score units per 1% of North American Holstein genes (NAH) in the bull. This shows that NAH_{pct} has a significance in affecting stature EBV (P<0.0001) as shown in Figure 4.4. HF_{pct} was also observed to be highly

significant in determining the stature EBV in dairy cattle with a *P*-value of 0.004, which is less than the significance threshold of 0.05.

An increase of 0.003 score units of stature EBV was determined by the increase of 1% of HF_{pct}, as estimated from the model. Another factor with high significance in affecting stature EBV was JE_{pct}, with a *P*-value of <0.0001. It is observed that JE_{pct} decreases stature EBV overall. With 1% of JE genes in the bull, EBV for stature decreases by 0.008 score units. The model also shows that year of birth does not have a linear effect on stature EBV in dairy cattle in this study.

A scatter plot of stature EBV against NAH_{pct} is shown in Figure 4.4. This shows that the higher percentage of NAH, the higher value for stature EBV.

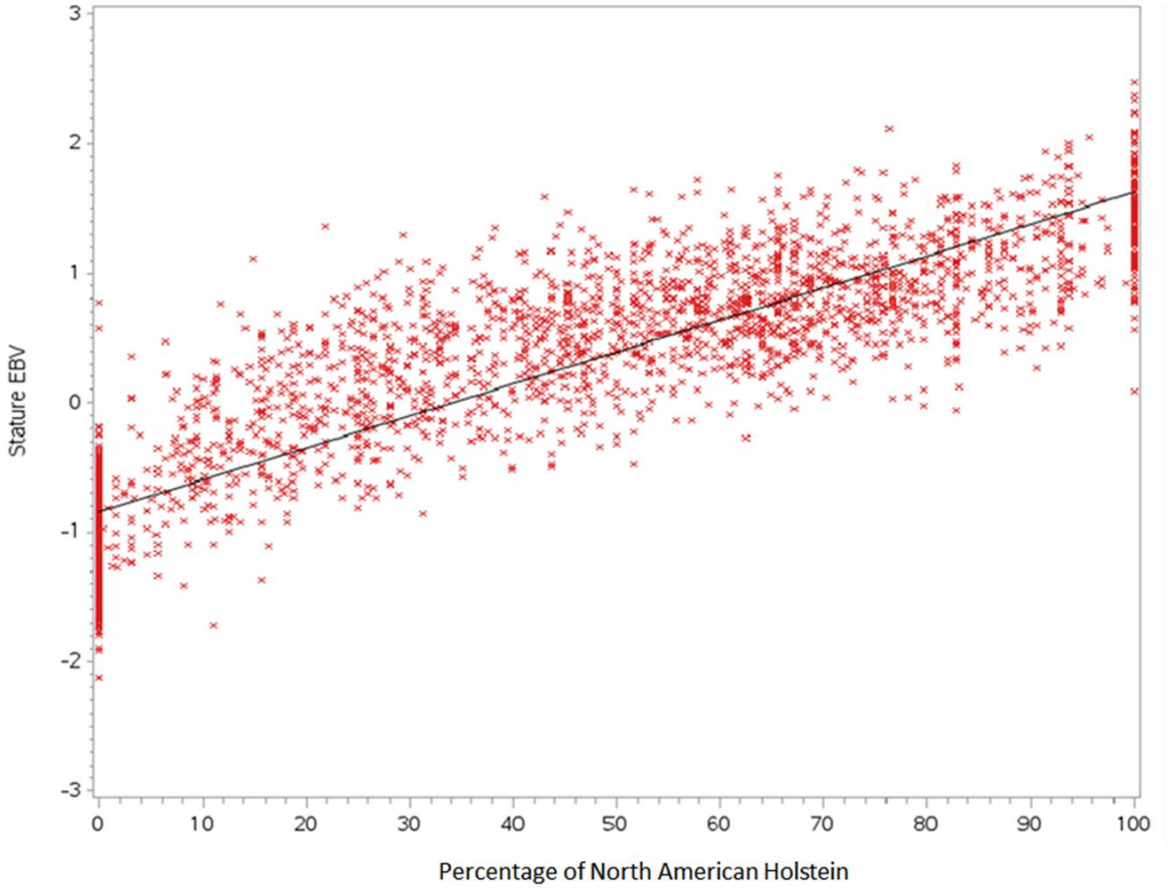

Figure 4.4: Scatter plot of stature EBV against percentage of North American Holstein.

4.2 Genome-wide association analysis

4.2.1 Manhattan plots

Manhattan plots depicting the *P*-values for each of the validated SNP markers for the HF, JE and XB populations were shown in Figures 4.5 to 4.7. The phenotypes were stature EBV adjusted for NAH_{pct}, HF_{pct}, JE_{pct} and year of birth. The *y*-axis of the plots is the negative logarithm *P*-values adjusted to genomic control for the SNPs, and the *x*-axis is the location of each SNP within each chromosome in the data set.

Two threshold lines were set for the Manhattan plots; one is the suggestive threshold which is the negative $\log_{10}$ value of 1×10^4 and the other is the genome-wide significant threshold at the negative $\log_{10}$ value of 5×10^5. As shown in figure 4.5, the chromosomes or *Bos Taurus* autosomes, BTA2, 3, 4, 5, 6, 7, 8, 11, 12, 14 and 24 show peaks of significant SNP effects above the suggestive threshold line in the HF population. Only BTA3, 5, 11 and 12 have significant SNPs that are located above the genome-wide significant threshold.

In Figure 4.6, BTA5, 6, 9, 10, 12, 18, 19, 23 and 25 have significant SNPs located above the suggestive threshold line in the JE population. SNPs on BTA9, 12, 19 and 25 were significant for stature in the JE population. Whereas in XB population, SNPs on BTA1, 3, 5, 7, 9, 10, 14, 22 and 23 displayed some significant SNPs above the suggestive threshold line in Figure 4.7. However, only BTA1, 3, 10, 14, 22 and 24 have significant SNPs that were located above the genome-wide significant threshold.

Manhattan plots for chromosomes with significant and suggested to be significant SNPs for HF, JE and XB populations were shown in Figure 4.8 to 4.10 to indicate the exact location of SNPs that are associated with genes influencing stature. For SNPs that were above the suggestive threshold line in the HF population, the SNPs (Figure 4.8) at BTA2 were located at 0Mbp, 9 – 14Mbp for BTA3, 115Mbp on BTA4, 19Mbp, 85Mbp and 110Mbp for BTA5, 72Mbp and 104Mbp for BTA6, 21Mbp for BTA7, 60Mbp for BTA8, 2Mbp and 17Mbp on BTA12, 51Mbp on BTA14, 31Mbp and 34Mbp on BTA24. For SNPs that have *P*-values above the genome-wide significant threshold line in the HF population,

the SNPs (Figure 4.8) were located at 10Mbp at BTA3, 19Mbp and 110Mbp on BTA5, 3Mbp on BTA11 and 2Mbp for BTA12.

On the other hand, the SNPs having *P*-values above the suggestive threshold line in the JE population (Figure 4.9) were SNPs located at 23Mbp and 109Mbp on BTA5, 93Mbp on BTA6, 43 – 49Mbp and 71Mbp on BTA9, 28Mbp and 30Mbp on BTA10, 9Mbp and 28Mbp for BTA12, 14Mbp on BTA18, 1Mbp, 5Mbp and 12Mbp for BTA19, 53Mbp on BTA23 and finally 33.5Mbp for BTA25. Significant SNPs in the JE population that were above the genome-wide significant threshold line as shown in Figure 4.9 were located at 45 – 49Mbp, 61Mbp and 71Mbp on BTA9, 5Mbp on BTA12, 12Mbp on BTA19 and 33.5Mbp for BTA25.

In the XB population, significant SNPs that were above the suggestive threshold line (Figure 4.10) were located at 52Mbp and 56Mbp on BTA1, 16Mbp on BTA3, 24Mbp for BTA5, 74Mbp for BTA7, 32Mbp on BTA9, and 89Mbp for BTA10, 19Mbp and 26.5Mbp on BTA14, 26Mbp on BTA22 and 27Mbp for BTA24. Significant SNPs that were above the genome-wide significant threshold line in the XB population were located at 78Mbp for BTA1, 16Mbp for BTA3, 89Mbp on BTA10, 25Mbp on BTA14, 26Mbp on BTA22 and 27Mbp for BTA24 as shown in Figure 4.10. The SNPs observed or suggested to be significant in the HF, JE and XB populations were shown in Table 4.4, 4.5 and 4.6 respectively.

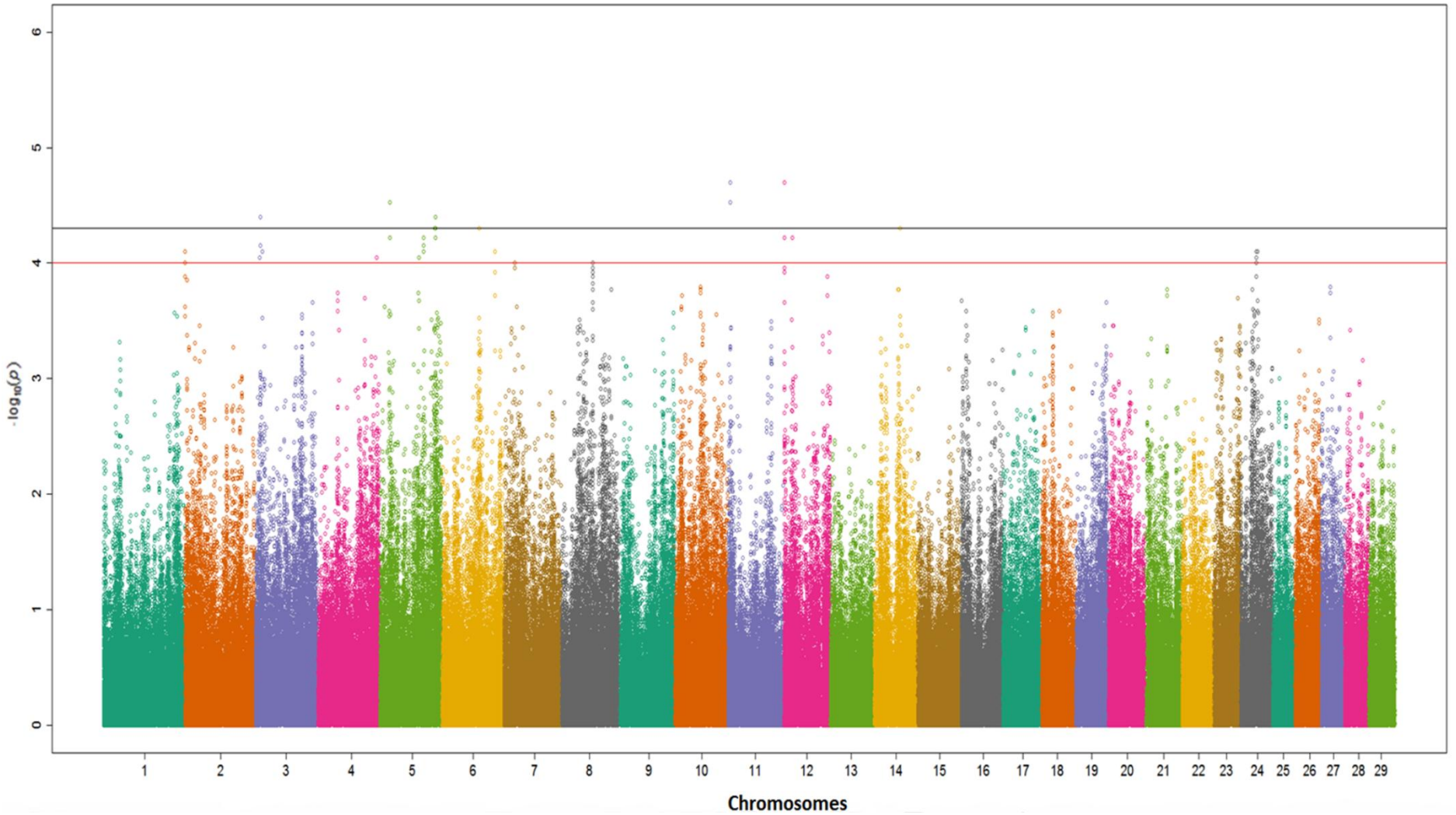

Figure 4.5: Genome-wide Manhattan plot of all associated SNP with adjusted stature EBV for Holstein-Friesian (HF).

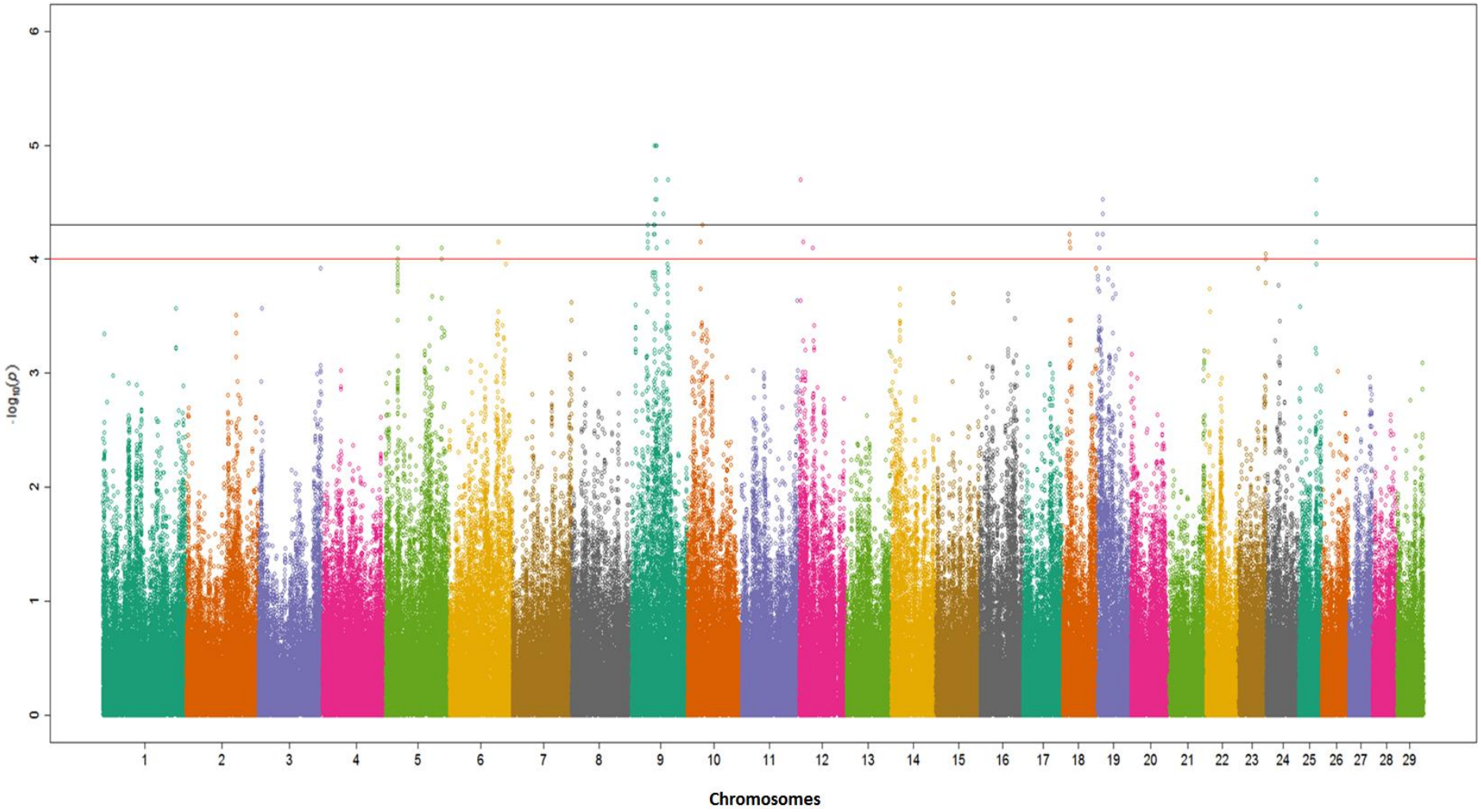

Figure 4.6: Genome-wide Manhattan plot of all associated SNP with adjusted stature EBV for Jersey (JE).

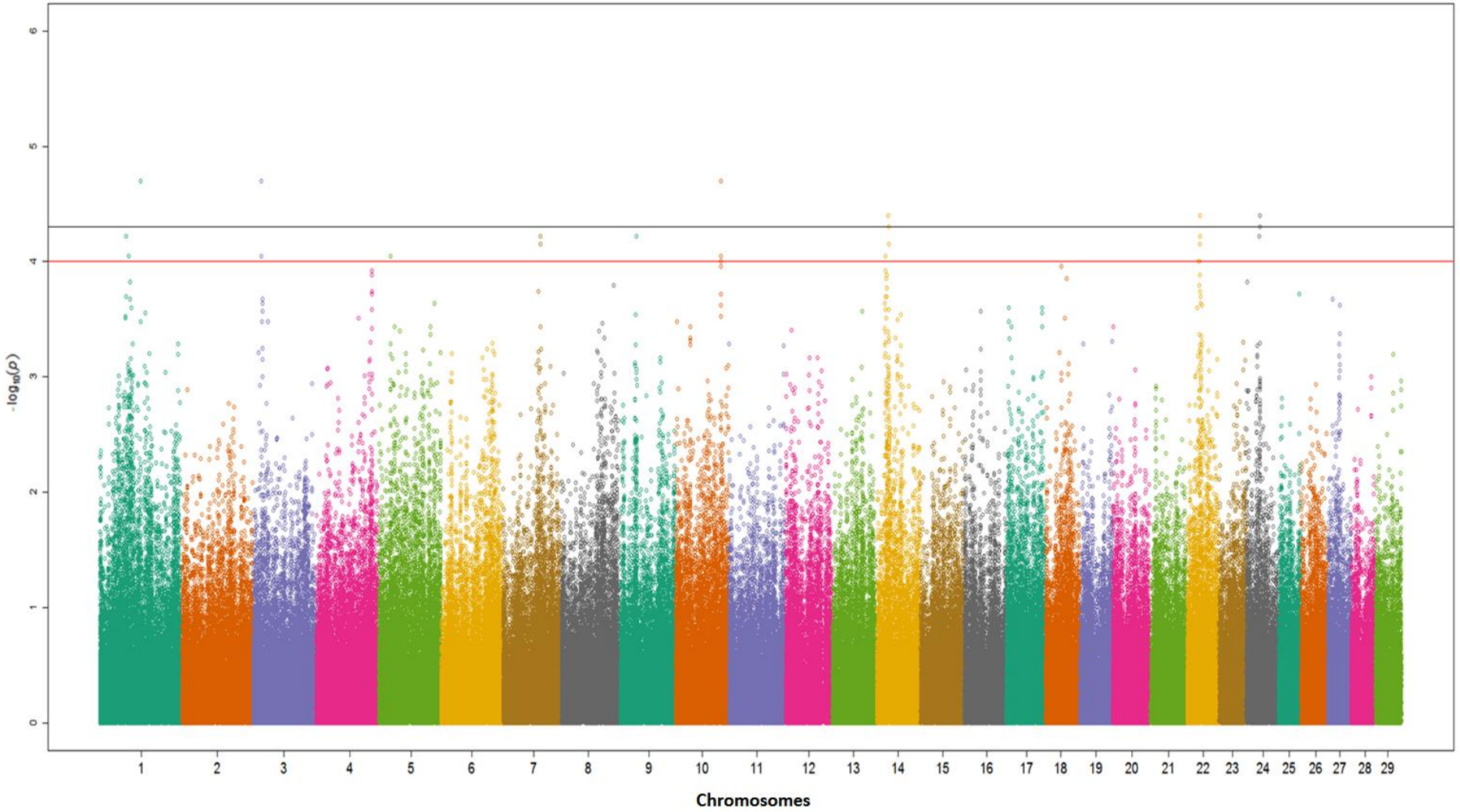

Figure 4.7: Genome-wide Manhattan plot of all associated SNP with adjusted stature EBV for Holstein-Friesian and Jersey crossbreed (XB).

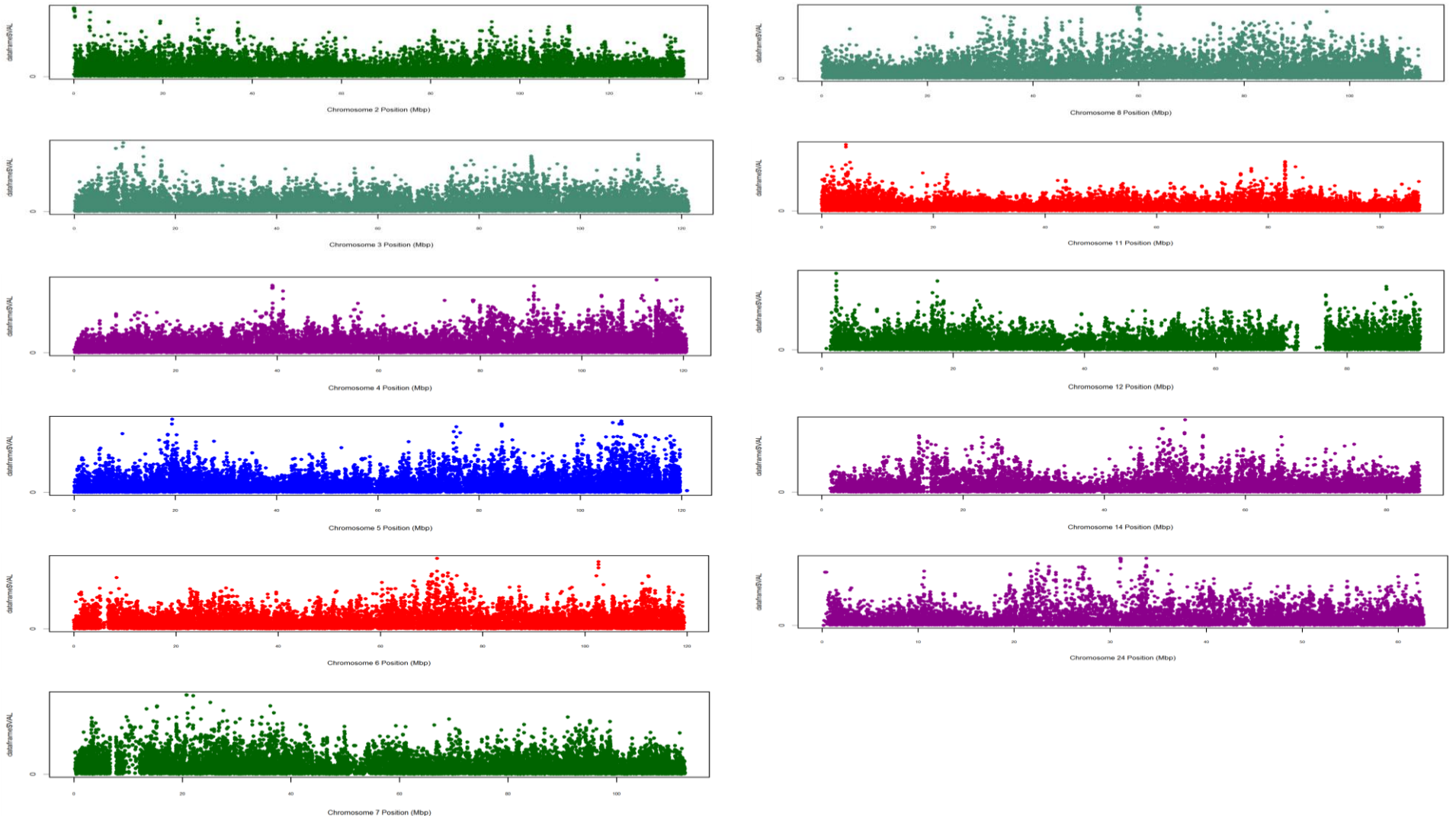

Figure 4.8: Multi Manhattan plots for BTA2, 3, 4, 5, 6, 7, 8, 11, 12, 14 and 24 in Holstein-Friesian (HF) population.

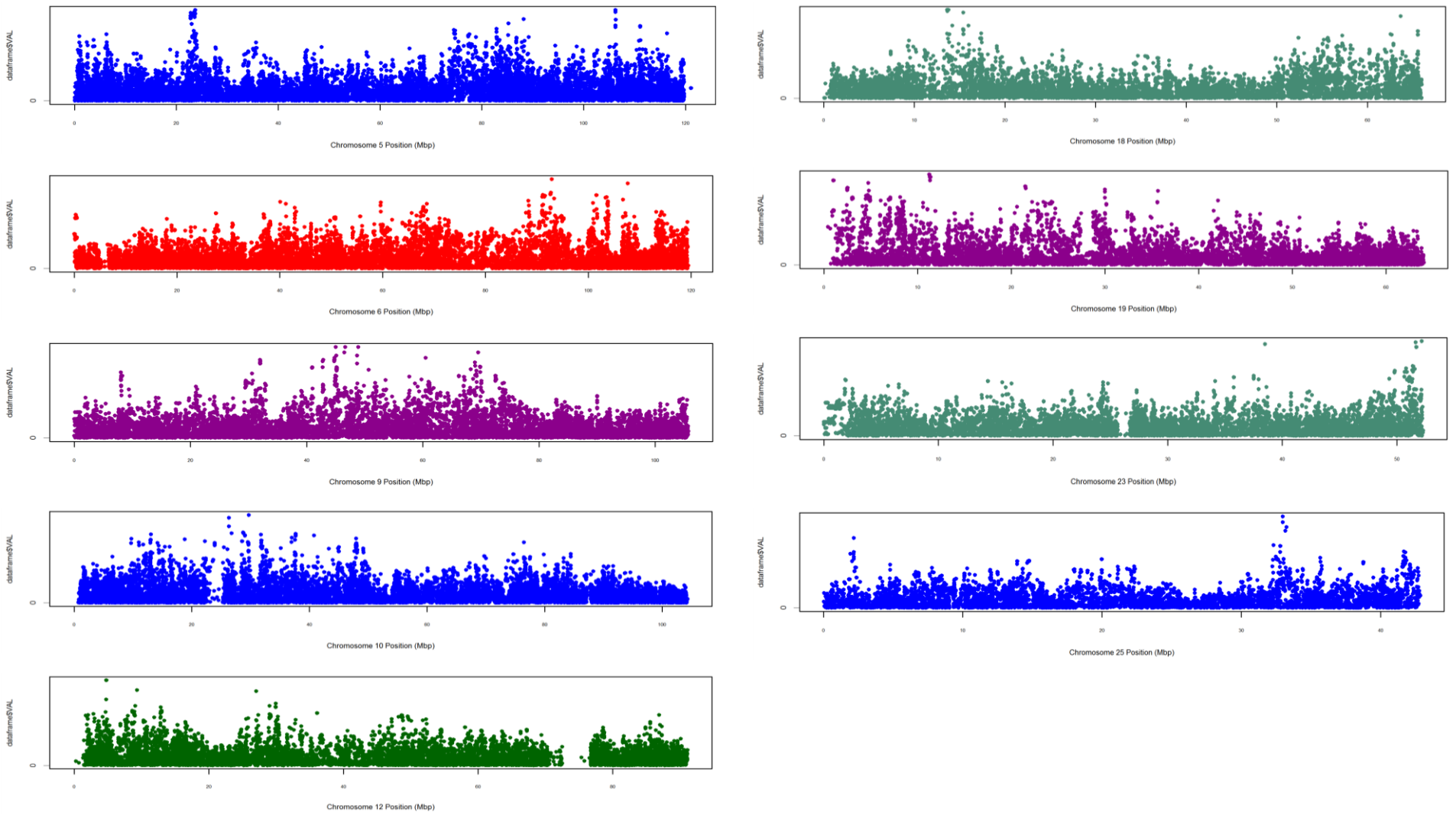

Figure 4.9: Multi Manhattan plots for BTA5, 6, 9, 10, 12, 18, 19, 23 and 25 in Jersey (JE) population.

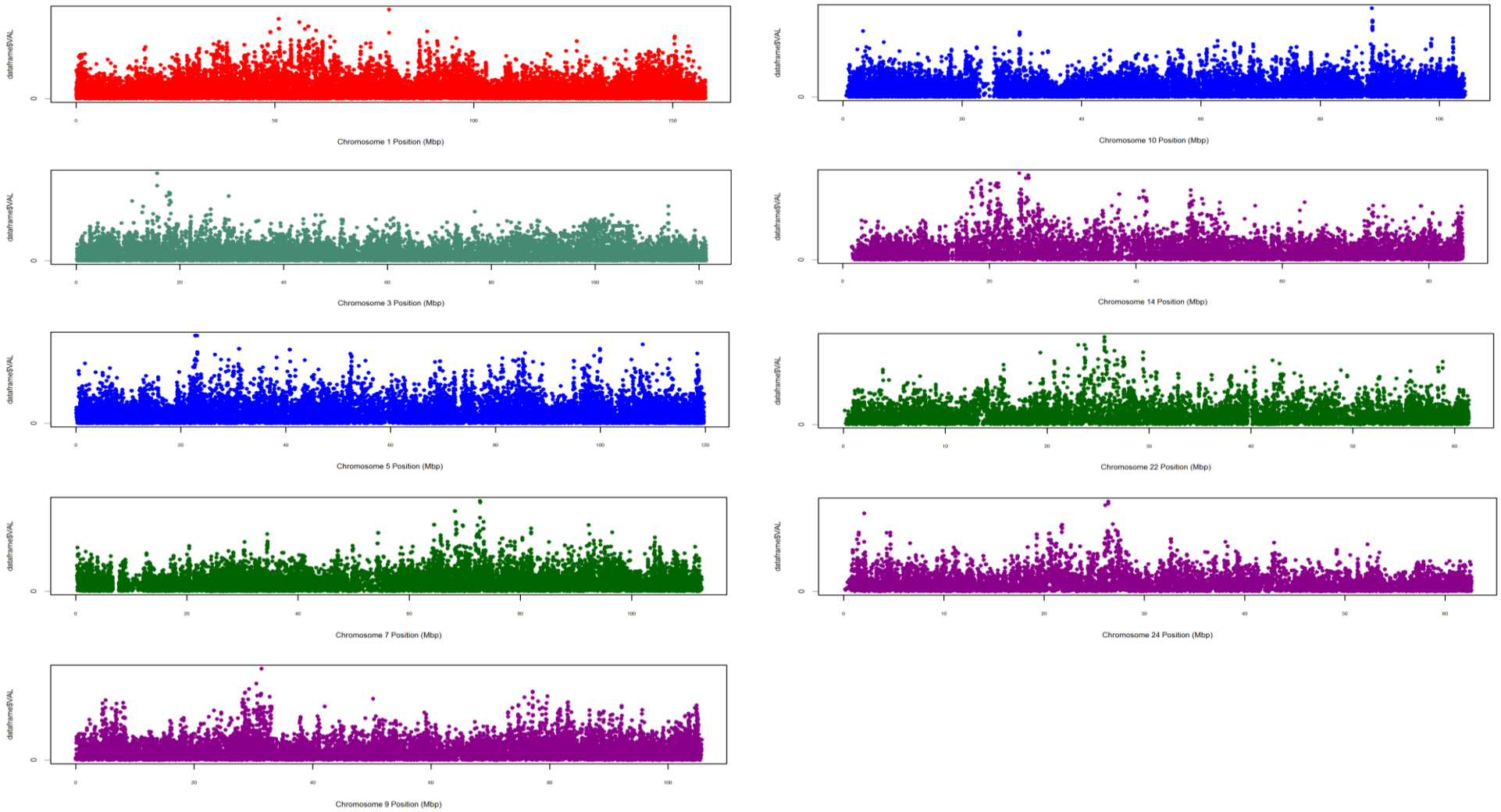

Figure 4.10: Multi Manhattan plots for BTA1, 3, 5, 7, 9, 10, 14, 22 and 24 in Holstein-Friesian × Jersey crossbred (XB) population.

Table 4.4: SNPs located above the genome-wide significant and suggestive threshold for Holstein-Friesian (HF) population.

Chromosome	Base-pair position (Mbp)
2	0
3	10
3	9, 10, 14
4	115
5	19, 110
5	19, 75, 85, 108, 110
6	72, 104
7	21
8	60
11	3
12	2
12	2, 17
14	51
24	31, 34

* SNPs located above the genome-wide significant threshold are highlighted.

Table 4.5: SNPs located above the genome-wide significant and suggestive threshold for Jersey (JE) population.

Chromosome	Base-pair position (Mbp)
5	23, 109
6	93
9	45 - 49, 61, 71
9	43 - 49, 71
10	28, 30
12	5
12	9, 28
18	14
19	12
19	1, 5, 12
23	53
25	33.5
25	33.5

* SNPs located above the genome-wide significant threshold are highlighted.

Table 4.6: SNPs located above the genome-wide significant and suggestive threshold for Holstein-Friesian × Jersey crossbred (XB) population.

Chromosome	Base-pair position (Mbp)
1	52, 56
1	78
3	16
3	16
5	24
7	74
9	32
10	89
10	89
14	25
14	19, 26.5
22	26
22	26
24	27
24	27

* SNPs located above the genome-wide significant threshold are highlighted.

4.3 Top 50 single nucleotide polymorphisms

Tables 4.7 to 4.9 show the top 50 SNPs according to the *P*-values in ascending order for the HF, JE and XB populations.

Out of the top 50 SNPs in HF population (Table 4.7), 4 SNPs were located on BTA2, 4 SNPs were on BTA3, 1 SNP on BTA4, 17 SNPs on BTA5, 2 SNPs on BTA6, 3 SNPs on BTA11, 5 SNPs on BTA12, 1 SNP on BTA14 and lastly, 13 SNPs on BTA24. As shown in Table 4.8, the top 50 SNPs for JE population were 27 SNPs located on BTA9, 1 SNP on BTA10, 4 SNPs on BTA12, 11 SNPs on BTA18, 5 SNPs on BTA19 and 2 SNPs on BTA25. For XB population, the top 50 SNPs were separately located on 11 chromosomes (Table 4.9). 5 SNPs were on BTA1, 2 SNPs on BTA3, 9 SNPs on BTA4, 3 SNPs on BTA5, 2 SNPs on BTA7, 1 SNP on BTA9, 7 SNPs on BTA10, 12 SNPs on BTA14, 1 SNP on BTA 18, 5 SNPs on BTA22 and 3 SNPs on BTA24.

Table 4.7: Top 50 SNPs according to the *P*-value for Holstein-Friesian (HF) population.

Chromosome	SNP identifier	Base-pair position	*P*-value
11	BovineHD1100001551	4348872	0.00002
11	BovineHD1100001552	4350255	0.00002
12	BovineHD1200000615	2199863	0.00002
12	BovineHD1200000616	2201087	0.00002
12	BovineHD1200000617	2202713	0.00002
5	BovineHD0500005588	19378297	0.00003
11	BovineHD1100001553	4351058	0.00003
3	BovineHD0300003148	9717281	0.00004
5	BovineHD0500031142	108124191	0.00004
5	BovineHD0500031143	108130344	0.00004
5	Hapmap26794-BTA-74957	108120416	0.00004
5	BovineHD0500030534	106461363	0.00005
5	BovineHD0500031137	108087792	0.00005
6	BovineHD0600019802	71095600	0.00005
14	BovineHD1400014625	51466187	0.00005
5	BovineHD0500005567	19346717	0.00006
5	BovineHD0500005568	19348038	0.00006
5	BovineHD0500023912	84484874	0.00006
5	BovineHD0500023913	84485450	0.00006
5	BovineHD0500031023	107721367	0.00006
5	BovineHD0500031168	108261151	0.00006
12	BovineHD1200000659	2263070	0.00006
12	BovineHD1200005332	17649405	0.00006
3	ARS-BFGL-NGS-18442	9610758	0.00007
5	BovineHD0500023911	84484139	0.00007
5	BovineHD0500023914	84486136	0.00007
5	BovineHD0500023916	84495554	0.00007
24	BTB-01258307	31086312	0.00008
2	BovineHD0200000008	55364	0.00008
2	BovineHD0200000011	77091	0.00008
2	BovineHD0200000026	185262	0.00008
3	BovineHD0300004436	13663730	0.00008
5	BovineHD0500023918	84498849	0.00008
24	BovineHD2400008456	31057884	0.00008
24	BovineHD2400008467	31102511	0.00008
24	BovineHD2400008470	31108680	0.00008
24	BovineHD2400008471	31109463	0.00008
24	BovineHD2400008472	31113298	0.00008
24	BovineHD2400009358	33776586	0.00008
6	BovineHD4100005491	102674813	0.00008
3	BovineHD0300002687	8224067	0.00009
4	BovineHD0400033209	114781240	0.00009
5	BovineHD0500021517	75545890	0.00009
24	BovineHD2400008462	31093717	0.00009
24	BovineHD2400008463	31094387	0.00009
24	BovineHD2400008464	31096544	0.00009
24	BovineHD2400008465	31097142	0.00009
24	BovineHD2400008473	31114237	0.00009
24	BovineHD2400008474	31115439	0.00009
2	ARS-BFGL-NGS-102158	35126	0.0001

Table 4.8: Top 50 SNPs according to the *P*-value for Jersey (JE) population.

Chromosome	SNP identifier	Base-pair position	*P*-value
9	ARS-BFGL-NGS-119030	46662943	0.00001
9	BTB-00391421	48901956	0.00001
9	BTB-00391456	48926104	0.00001
9	BovineHD0900012512	45014461	0.00001
9	BovineHD0900012888	46655716	0.00001
9	BovineHD0900013491	48921851	0.00001
9	BovineHD0900013492	48926890	0.00001
9	BovineHD4100007411	46661271	0.00001
9	BovineHD0900012862	46511484	0.00002
9	BovineHD0900019210	69542080	0.00002
9	BovineHD0900019211	69543047	0.00002
9	BovineHD0900019212	69543889	0.00002
12	BTB-01435273	4787061	0.00002
12	BovineHD1200001426	4764866	0.00002
12	BovineHD1200001437	4807493	0.00002
12	BovineHD1200001438	4808609	0.00002
25	BovineHD2500009072	32982924	0.00002
9	ARS-BFGL-NGS-61170	45121053	0.00003
9	BovineHD0900013421	48699964	0.00003
19	BovineHD1900003068	11253507	0.00003
19	BovineHD1900003070	11269333	0.00003
9	BovineHD0900016606	60496984	0.00004
9	Hapmap44189-BTA-109443	44815699	0.00004
19	BovineHD1900003075	11303460	0.00004
19	BovineHD1900003082	11359971	0.00004
19	BovineHD1900003087	11414075	0.00004
25	BovineHD2500009069	32979480	0.00004
9	BTB-01737039	32004342	0.00005
9	BovineHD0900011892	42852058	0.00005
9	BovineHD0900012436	44786657	0.00005
9	BovineHD0900012440	44802314	0.00005
9	BovineHD0900012441	44805630	0.00005
9	BovineHD0900012442	44806469	0.00005
9	BovineHD0900012443	44809399	0.00005
10	BovineHD1000009799	29718121	0.00005
9	BovineHD0900008724	32045680	0.00006
9	BovineHD0900011861	42762405	0.00006
9	BovineHD0900011872	42784015	0.00006
9	BovineHD0900031449	44780620	0.00006
18	BovineHD1800004506	13648552	0.00006
18	BovineHD1800004507	13651264	0.00006
18	BovineHD1800004511	13669998	0.00006
18	BovineHD1800004512	13673430	0.00006
18	BovineHD1800004521	13700444	0.00006
18	BovineHD1800004526	13724571	0.00006
18	BovineHD1800004527	13730814	0.00006
18	BovineHD1800004531	13756296	0.00006
18	BovineHD1800004532	13758135	0.00006
18	BovineHD1800004533	13762565	0.00006
18	BovineHD1800004534	13768459	0.00006

Table 4.9: Top 50 SNPs according to the *P*-value for Holstein-Friesian × Jersey crossbred (XB) population.

Chromosome	SNP identifier	Base-pair position	*P*-value
1	BTB-01164879	50775613	<0.00001
1	BovineHD0100022659	78719409	0.00002
1	BovineHD0100022660	78719999	0.00002
3	BovineHD0300005121	15638565	0.00002
10	BovineHD1000025258	88688518	0.00002
14	BovineHD1400006989	24099719	0.00004
14	BovineHD1400006990	24102024	0.00004
14	BovineHD1400024435	24096532	0.00004
22	BovineHD2200007508	25637733	0.00004
24	BovineHD2400007207	26410034	0.00004
14	BovineHD1400007334	25332510	0.00005
24	BovineHD2400007209	26413303	0.00005
1	BovineHD0100014337	50995138	0.00006
7	BovineHD0700021398	72769554	0.00006
9	BovineHD0900008528	31376496	0.00006
22	BovineHD2200007509	25644986	0.00006
24	BovineHD2400007085	26120966	0.00006
7	BovineHD0700021405	72799470	0.00007
14	BovineHD1400007259	25015640	0.00007
14	BovineHD1400007348	25393163	0.00007
22	BovineHD2200007506	25627122	0.00007
14	Hapmap46735-BTA-86653	25401722	0.00007
1	BovineHD0100015799	56142576	0.00009
3	BovineHD0300005119	15629896	0.00009
5	BovineHD0500006626	22796266	0.00009
5	BovineHD0500006628	22806678	0.00009
5	BovineHD0500006722	23119371	0.00009
10	BovineHD1000025288	88789903	0.00009
14	BovineHD1400005415	18870264	0.00009
14	BovineHD1400005416	18874879	0.00009
14	BovineHD1400005418	18884162	0.00009
14	Hapmap48727-BTA-36014	18876940	0.00009
10	BovineHD1000025289	88790676	0.0001
10	BovineHD1000025291	88794427	0.0001
10	BovineHD1000025293	88799353	0.0001
22	BovineHD2200006765	23046933	0.0001
22	BovineHD2200007038	23678284	0.0001
10	BovineHD1000025295	88802343	0.00011
10	BovineHD1000025297	88807592	0.00011
18	BovineHD1800009565	31509520	0.00011
4	BovineHD0400030582	107451941	0.00012
4	BovineHD0400030584	107454754	0.00012
4	BovineHD0400030585	107460385	0.00012
4	BovineHD0400030589	107481811	0.00012
4	BovineHD0400030591	107490155	0.00012
4	BovineHD0400030594	107500383	0.00012
4	BovineHD0400030595	107500976	0.00012
4	BovineHD0400030596	107503166	0.00012
14	BovineHD1400005285	18423308	0.00012
4	BovineHD0400030597	107508974	0.00013

CHAPTER 5

DISCUSSION

Live weight is a trait of economic importance and included in the breeding objective of the breeding program for the genetic improvement of New Zealand dairy cattle. Selection of bulls and cows as parents of the next generation is based on an economic index called BW index that measures the expected capability of a cow or bull to breed replacements which can convert feed into profit efficiently. Cattle with high live weight but low level of production of milk solids are reported to have low FCE as more feed is converted for maintenance than production purposes (Moore et al. 1992; Holmes et al. 1993; Cole et al. 2009). Therefore live weight is a trait that is included the calculation of BW with a negative economic weight of $1.52 per kilogram. The current breeding objective is placing more selection intensity on cows and bulls with optimum live weight and high FCE to generate maximum net profit for the dairy industry.

5.1 Correlation between stature and live weight

Stature is genetically correlated to live weight in cattle and selecting against stature in breeding programs will select against live weight indirectly. Ahlborn & Demple (1992) reported a figure of 0.85 to 0.92 for correlation of stature and live weight, and other literature derived figures are between 0.42 to 0.96 (Cue et al. 1996; Veerkamp & Brotherstone 1997; Coppieters et al. 1998; Kennedy et al. 1999; Toshniwal et al. 2008; Mc Hugh et al. 2012). Proper weights are important in deriving the selection index. Stature is one of the linear-type body conformation traits that provides a good measurement of live weight, thus stature is recommended as an indirect predictor for both live weight and FCE for cattle. Veerkamp (1998) reported a high accuracy of selection from an index that includes linear type traits. He also stated that the overall economic efficiency can be improved by selecting against live weight and other related indicator traits to improve FCE.

Even identifying QTLs associated with stature with the aid of GWAS, preferences for stature in cattle are still up to the consideration and decision of the farmer. In a paper by

Berry et al. (2005) that investigates the phenotypic relationships between TOP and longevity across New Zealand dairy cattle, they discovered that intermediate optimum is preferable for stature, whether in registered or commercial herds, and farmers are not selecting for extreme heights even in pedigree herds. Genomic selection can be implemented to breed cows with the intermediate optimum height which are suitable for the dairy industry and generates maximum profit.

5.1.1 Effects of percentage of North American Holstein genes on stature and live weight estimated breeding value

During 1960s and 1970s, NAH were imported from Canada to be introduced to the genetic base of the HF population in New Zealand. NAH from United States were introduced during the early 1980s. The percentage of HF population in New Zealand with NAH has since then increased from 2% to 38% (Harris & Kolver 2001). Importation of NAH into the HF population was to broaden the genetic base and to incorporate the high productivity traits found in NAH.

The NAH was selected for the high milk yield and later for a total merit index. Dairy cows with 75% NAH have significantly higher milk yield, fat yield and protein yield than cows with 0% NAH (Berry et al. 2003). North American-derived Holstein cows had higher live weight, milk yield and protein yield compared to the New Zealand HF herd (Harris & Kolver 2001).

The descriptive statistics from this study in Table 4.1 shows that the HF population has higher EBV for stature and live weight than the JE and XB population. The evaluated sire statistics (NZAEL 2013) also stated the average EBV for stature and live weight in HF sires are higher than JE and XB sires. As shown in Figure 4.4, stature EBV is affected by NAH_{pct}. The higher the NAH_{pct} found in the bulls, the higher the stature EBV is.

5.2 Effects on feed conversion efficiency when selecting for live weight

Live weight is a moderately high heritable trait as reported with figures from 0.16 to 0.69 (Ahlborn & Demple 1992; Spelman & Garrick 1997; Veerkamp & Brotherstone 1997; Kennedy et al. 1999; Veerkamp et al. 2001; Muller et al. 2006; Pryce & Harris 2006; Toshniwal et al. 2008). Heritability for live weight in New Zealand dairy cattle was reported at 0.35 by NZAEL (2013).

Studies have reported genetic correlations between live weight and feed intake in cattle (Veerkamp & Brotherstone 1997; Kennedy et al. 1999; Toshniwal et al. 2008) with figures from 0.37 (Veerkamp & Brotherstone 1997) to 0.51 (Toshniwal et al. 2008).

Correlation of live weight and FCE have been reported with figures of -0.53 (Kennedy et al. 1999) and 0.27 (Veerkamp & Brotherstone 1997). Veerkamp & Brotherstone (1997) pointed out that fluctuations in the genetic correlations occur as both live weight and FCE are closely related with mobilization of body reserves, hence any strenuous activities by the cattle for example reproduction will cause changes in their efficiency to convert feed to body mass and milk production.

The correlation with live weight and other production traits such as milk yield and milk solids have been overall moderately positive. Figures from -0.14 to 0.39 were reported for correlation between live weight and milk yield while numbers from -0.07 to 0.39 were reported for correlation between live weight and milk solids (Ahlborn & Demple 1992; Spelman & Garrick 1997; Veerkamp & Brotherstone 1997; Kennedy et al. 1999; Veerkamp et al. 2001; Muller et al. 2006; Pryce & Harris 2006; Toshniwal et al. 2008). The fluctuations in the figures are due to live weight being measured during different stages of lactation. Correlations are higher during early lactation period and lower during late lactation period (Veerkamp 1998).

The moderately positive correlations between live weight and the production traits are appealing to select for live weight for the genetic gain in other production traits as well however live weight with a negative economic value of -$1.52 in the BW index (NZAEL 2013) deters such efforts. This is due to cattle with high live weight do not normally convert feed efficiently and are subjected to poor health and reproduction performances

(Veerkamp 1998). Hence live weight has a bearing on profit as feed for maintenance increases with live weight. Therefore placing selection intensity on live weight will affect the genetic gain on FCE.

Live weight varies according to the condition of the cow; it decreases due to negative energy balance during high production periods, and increases when excess feed intake goes to maintenance. Complications occur as live weight varies due to the dynamics of body tissue mobilization and milk production during different stages in dairy cattle. This affects both dry matter intake and FCE. Veerkamp (1998) reported difficulties to obtain appropriate weighting factors for milk yield, feed intake and live weight in breeding schemes. The genetic associations of these traits are strong thus selecting for one trait may lead to the loss of genetic gain for the other trait due to the correlated response. This further complicates the attempts to estimate the BV for FCE.

The derivation of FCE is a ratio, and use of ratio traits in genomic selection is challenging due to the selection pressure exerted disproportionately on the component traits (Arthur et al. 2001). The amount of genetic change on the traits in the future will be hard to estimate. Researchers have to achieve a balance in improving FCE through genomic selection by considering the feed capacity and amount of feed required for milk yield, growth and maintenance.

5.3 Identification of significant single nucleotide polymorphisms affecting stature

PLINK was used to run GWAS to identify SNPs that influences stature and live weight in New Zealand dairy cattle. Before running the GWAS, the phenotype which was the EBV for stature was adjusted for HF_{pct}, JE_{pct}, NAH_{pct} and year of birth using a multiple regression model.

Manhattan plots for the GWAS for stature were plotted with the chromosomes on the *x*-axis and the negative logarithm of the association *P*-value for every SNP on the *y*-axis. SNP with the highest association with the trait has the smallest *P*-value, thus the negative logarithm value will be the greatest and the scatter plot will be at the highest peak in the Manhattan plot. The *P*-values of the SNPs were reduced to a lower level after adjusting for

genomic control factor. The multi Manhattan plots were plotted to highlight the exact locations of the SNPs that is of interest for this study.

Based on the top 50 SNPs according to the *P*-value, this study identified SNPs significantly associated with stature in nine chromosomes or BTA, respectively BTA2, 3, 4, 5, 6, 11, 12, 14 and 24 in the HF population. Significant SNPs for stature were detected in six chromosomes, namely BTA9, 10, 12, 18, 19 and 25 in the JE population whereas the SNPs determined to be significantly associated with stature were located on eleven chromosomes, respectively BTA1, 3, 4, 5, 7, 9, 10, 14, 18, 22 and 24 in the XB population.

5.3.1 Significant single nucleotide polymorphisms

According to the Manhattan plots (Figure 4.7 – 4.9), SNPs (52 and 56Mbp) on BTA1 were located above the genome-wide suggestive threshold in the XB population, while the SNP located at 78Mbp was found to be significant. Kolbehdari et al. (2008) also observed a marker at 95.6Mbp on this chromosome which has significance on dairy cattle stature. HF is the only population found to have a SNP above the genome-wide suggestive threshold on BTA2, and the SNP was located at 0Mbp which is not near with the locations of SNPs identified on BTA2 from various publications (Boichard et al. 2003; Ashwell et al. 2005; Bolormaa et al. 2011).

SNPs were located above the suggestive genome-wide significance threshold on BTA3 in the HF and XB populations. SNPs were located above the suggestive threshold on 9 and 14Mbp of BTA3 in the HF population, and a SNP peak on 10Mbp that lie above the genome-wide suggestive and significance threshold (Figure 4.7). A SNP on 16Mbp that forms a peak on the plot (Figure 4.9) which suggests genome-wide significance on stature EBV was detected in the XB population. However none of the SNPs identified here were close to the SNPs observed on BTA3 from other publications (Kolbehdari et al. 2008; Bolormaa et al. 2011).

BTA4 only has a SNP on 115Mbp that was located above the genome-wide suggestive threshold in the HF population. This SNP too was not located near to the SNPs on BTA4 from other literatures (Ashwell et al. 2005; Kolbehdari et al. 2008). SNPs that

affect stature were observed on BTA5 for all three populations in this study. SNPs on 19, 75, 85, 108 and 110Mbp that were above the suggestive threshold and SNPs that peaked above the significant threshold on 19 and 110Mbp were detected in the HF population (Figure 4.7). SNPs located on 23 and 109Mbp were above the suggestive threshold in the JE population while a SNP located on a similar position like the JE population, which is on 24Mbp was observed in the XB population.

Various publications have reported many SNPs on BTA5 that affects stature significantly. Schrooten et al. (2000), Boichard et al. (2003) and Bolormaa et al. (2011) have reported SNPs on BTA5 that were located between 120 and 124Mbp that have an effect on cattle stature. In this study, SNPs on 19 and 75Mbp observed in the HF population (Figure 4.7) coincided with the SNPs reported by Kolbehdari et al. (2008) which were located on 10.2 and 74Mbp respectively. SNP on 110Mbp in the HF population (Figure 4.7) and SNP on 109Mbp in the JE population (Figure 4.8) are near to the SNP reported on 112.3Mbp by Cole et al. (2009).

BTA5 chromosome also has been identified to have an effect on stature by other publications (Cole et al. 2009, Bolormaa et al. 2011, Karim et al. 2011). QTL influencing various body conformation traits including stature were identified by Schrooten et al. (2000) and Boichard et al. (2003) on BTA5 by using microsatellite markers.

SNPs located above the genome-wide suggestive threshold for BTA6 were observed in both HF and JE populations, respectively on 72 and 104Mbp in the HF population and 93Mbp in the JE population. None of the literature reviewed for this study reported SNPs for that region. SNPs above the suggestive threshold on BTA7 were located on 21Mbp in the HF population and on 74Mbp in the XB population. None of these SNPs were near to the SNPs for this chromosome observed on 39.9Mbp by Kolbehdari et al. (2008). Only a SNP above the genome-wide suggestive threshold that affects stature on 60Mbp on BTA8 was observed in the HF population. This reported SNP was not near to any SNPs reported from literature (Schrooten et al. 2000; Pryce et al. 2011) for BTA8.

SNPs affecting stature on BTA9 were observed in both JE and XB populations. A cluster of SNPs around 43 – 49Mbp that were above the genome-wide suggestive threshold

that continues to rise above the significant threshold at 45 – 49Mbp were located in the JE population (Figure 4.8). This is similar to the SNP observed on 71Mbp, with the SNP located both above the suggestive and significant threshold (Figure 4.8). A SNP on 61Mbp that was placed above the significant threshold line was also located in the JE population (Figure 4.8). A significant SNP was located on 32Mbp that lies above the suggestive threshold in the XB population. Similarly, only SNPs on BTA10 that have an influence on cattle stature EBV were found in JE and XB populations. SNPs on 28 and 30Mbp located above the suggestive threshold were detected in JE population whereas a SNP on 89Mbp that were situated above the suggestive threshold and raised up above the significant threshold line was observed in XB population (Figure 4.9).

None of the literature reviewed here reported any SNPs that affect cattle stature on BTA9 and BTA10. It is noted however that most of the literature reported QTL for stature from researches conducted with HF cattle, rarely for JE and XB cattle. The results suggested that these genes on BTA9 and BTA10 which are not found in HF population may have significance on stature in JE and XB population.

The SNP located on BTA11 at 3Mbp in the HF population were located above the genome-wide significant threshold (Figure 4.7). This SNP does not coincide with the QTL positions reported for BTA11 (Boichard et al. 2003; Kolbehdari et al. 2008) either. SNPs affecting stature on BTA12 were observed in both HF and JE populations. SNPs above the suggestive threshold on 2 and 17Mbp were detected in the HF population, with the SNP on 2Mbp continue to peak above the significant threshold line (Figure 4.7). SNPs were located above the suggestive threshold on 9 and 28Mbp, and a SNP located above the significant threshold on 5Mbp in the JE population (Figure 4.8). No literature reviewed so far reported SNPs that affect cattle stature on BTA12.

The present study identified SNPs affecting stature in the HF and XB population on BTA14. The SNP that were above the genome-wide suggestive threshold were located on 51Mbp in the HF population while SNPs that were above the genome-wide suggestive threshold were located on 19 and 26.5Mbp in the XB population. A SNP above the significant threshold on 25Mbp was also located in the XB population (Figure 4.9). This result is similar with the QTL studied by Pryce et al. (2011), who reported QTL significant

for cattle stature between 22.8 and 23.5Mbp. Spelman et al. (1999) and Littlejohn et al. (2012) reported QTL for stature on BTA14 too, on 36Mbp and 8.1Mbp respectively.

Apart from Spelman et al. (1999) who identified QTL for stature on BTA14, Karim et al. (2011) also discovered BTA14 has a genomic interval which is not specified in the literature that affects stature and weight in dairy cattle. The pleiomorphic adenoma gene 1 (*PLAG1*) has been reported to be located at the chromosomal region (25Mbp) of BTA14 (Karim et al. 2011). Littlejohn et al. (2012) and Nishimura et al. (2012) found that *PLAG1* gene has a significant effect on live weight during early life in dairy cattle. Variations of this gene also have been discovered to bring changes in gene expression for growth across mammalian species such as in humans (Pryce et al. 2011).

Hawken and colleagues (2011) point out the locus at BTA14 may carry genes that affect fertility traits. This discovery is confirmed by Fortes et al. (2012) as SNPs with significant association for male and female reproduction traits were discovered on BTA14. Among the SNPs are polymorphisms linked to insulin-like growth factor 1 (*IGF1*), which contributes to both reproduction and growth. Fortes et al. (2012) indicate a possibility of pleiotropic effects for reproductive and growth traits on the region (21 to 28 Mb) on BTA14 across different cattle breeds.

Coppieters et al. (1998) reported QTL affecting milk production for this chromosome too, indicating BTA14 has genomic regions affecting milk production traits in dairy cattle. A study by Berry et al. (2004) showed taller cows are associated with a lower number of pregnancy rates and require more services, hinting possibilities of genetic correlation between body conformation traits with reproduction.

Only a SNP with significant effect on stature on BTA22 was reported in XB population, with the SNP located on 26Mbp placed above both the genome-wide suggestive and significant threshold (Figure 4.9). This however does not comply with the result reported by Ashwell et al. (2005), with a marker for stature located on 72Mbp on BTA22. JE is the only population that reported a SNP for BTA23, with the SNP on 53Mbp located above the suggestive threshold line. This result too does not agree with the position

reported by Ashwell et al. (2005) who reported a marker that affect stature at the centromere (0Mbp) of BTA23.

SNPs on BTA24 found to affect stature were observed in the HF and XB populations, with SNPs above the suggestive threshold line on 31 and 34Mbp in the HF population and on 27Mbp in the XB population. The SNP in the XB population also peaked above the genome-wide significant threshold (Figure 4.9). The SNP observed in the HF population complied with the results by Pryce et al. (2011) who reported QTL significant for stature between 34.4 and 34.6Mbp on BTA24. Kolbehdari et al. (2008) too reported a QTL for stature on 39.8Mbp.

SNPs on BTA25 affecting stature were only reported in the JE population, with SNPs on 33.5Mbp located above the suggestive threshold, and continue above the significant threshold line (Figure 4.8). Similarly for SNPs observed in the JE population on chromosome 9, 10 and 12, no literature reviewed so far reported QTL associated with stature for BTA25.

5.3.2 Reported significant single nucleotide polymorphisms from literature

The paper by Karim et al. (2011) did not state the positions for SNPs detected significant for stature on various chromosomes (BTA2, 5, 7, 8, 10, 14, 25, and 28). Cole et al. (2011) also identified several SNP markers on BTA5 and BTA11 for stature and body depth but the base-pair positions were not stated in the paper. SNPs detected on BTA2, 5, 7, 8, 10, 11, 14 and 25 in this study in the HF, JE and XB populations may coincide with those identified SNPs from those two papers.

Kolbehdari et al. (2008) identified significant chromosome-wise SNPs with stature in BTA1, 3, 4, 5, 7, 11, 19 and 24 in Canadian Holstein bulls especially SNP (rs41572038 and rs41592968) on BTA3 and 5 are significant with values of $P < 0.01$. The SNPs detected on BTA5 coincides with the SNPs detected in the HF population in this study. QTL affecting conformation traits also have been located on BTA18 in HF cattle from studies by Ashwell et al. (2005), Schnabel et al. (2005) and Kolbehdari et al. (2008).

In the paper published by Cole et al. (2009), markers on BTA18 affect several traits such as dystocia (sire and daughter calving ease, sire stillbirth), conformation (stature, rump width, strength, body depth) and efficiency (longevity, net merit). They discovered that the SNP (ss86324977) is located on a gene called sialic acid binding Ig-like lectin (Siglec)-5 that is involved with leptin secretion. Leptin deficiency is associated with delayed parturition (Mounzih et al. 1998) while high levels of leptin indicate high energy reserves of the body and as a result, lowering the feed intake and increase weight loss. Cows that carry their calves for longer period will most likely bear larger calves and have a greater risk for dystocia (Hansen et al. 2004). SNP ss86324977 is then associated with increased calf size due to longer gestation lengths, and affecting both calving ease and conformation traits. The XB population in this study had a SNP on BTA18 defined to be significant according to the *P*-value. However no significant SNPs on this chromosome were detected in all three breeds according to the Manhattan plots in this study.

QTL associated with stature were discovered on chromosome 5, 6 and 8 by Schrooten et al. (2000). QTL suggested being associated with stature; chest width and birth weight were specifically detected on BTA5. They also found indications of QTL on BTA6 that affect stature, body size and dairy character. The definition of dairy character is optimal bone angularity with refined structure in cattle that normally comes with high milk production level. Dairy character depends more on a farmer's judgement, and is more of a subjective criterion. Body size and stature are found to be strongly correlated with dairy character (Berry et al. 2005) and cows that scored high in dairy character are high producers. A paper published by Khatkar et al. (2004) noted that BTA6 showed strong presence of QTLs affecting milk yield.

A study pointed out that a SNP (rs29014633) on BTA5 is significantly associated with dairy strength and angularity while SNP rs41591943 on BTA7 is significantly associated with dairy strength (Kolbehdari et al. 2008), indicating genetic correlation. Casas et al. (2003) identified 5 chromosomes harbouring QTL that are associated with birth weight, in which BTA1, 2, 3 and 5 are also observed to harbour significant SNPs in this study. Kim et al. (2003) also detected the presence of a QTL for birth weight on BTA5. Cole et al. (2009) also reported the SNP for stature on BTA5 (112.3Mbp) is located near to

a microsatellite associated with body size reported by Schrooten et al. (2000) on the same chromosome (123Mbp), hinting a plausible correlation.

The paper by Pryce et al. (2011) identified common SNPs found in beef, dairy cattle and humans that influence stature. Among them are SNPs detected in BTA6, 8, 14 and 24, which also coincide with the significant SNPs detected in this study. Several genes affecting stature in human were found to play a role too in cell division and growth for cattle, indicating both cattle and human share some identical genes for stature.

Bolormaa et al. (2011) reported 9 significant SNPs on BTA3, 5, 6 and 8 for residual feed intake, which is used as a measure for FCE, plus dozens of SNP on BTA3 associated with stature in dairy cattle and different growth traits in beef cattle. Genes located on BTA2 were identified by Barendse et al. (2007) that affect FCE are involved with the secretion of insulin, serotonin and leptin. Though phenotypes are expressed at the hierarchy levels of cells, tissues and organs, but these hormones determines food intake by appetite control and body mass homeostasis. *ATP1A1* (BTA3), *UBE2I* (BTA5), and *RPLP2* (BTA14) are a few of identified genes involved in an animal's basal metabolism. These genes utilize energy such as adenosine triphosphate (ATP) to recycle and form proteins, which in turn affects the behavioural and homeostatic systems that determine feed intake.

A large number of genes on BTA1, 2, 5, 7, 8, 9, 12, 13, 15 and 19 were found to affect extracellular matrix and cell adhesion, which in turn affect residual feed intake (Barendse et al. 2007). One of those genes (*DMD*) on BTA9 is known to have a mutation related to muscle regeneration and muscular hypertrophy (Harper et al. 2002). Pleiotropic effects might be involved, and indicates residual feed intake is related to tissue strength and other musculature aspects. Barendse et al. (2007) stated that this relationship shows the tendency of an animal's food use and metabolic efficiency, as the animal may allocate more energy for reproduction or for other functions and involve genes that affect tissue strength and integrity.

Barendse et al. (2007) also pointed out that DNA sequences involved with genes of hypothetical or unknown function are likely to be involved in metabolic efficiency or feed conversion. Some genes, such as *XP_602409* on BTA3 are similar to the *Cyclin M1* gene

which is related to RFI. However most of the function of these genes are still unknown, and will require more research and analysis in the future. Studying the genetic function of other difficult traits will help to identify the functions of this large number of genes by utilizing the pleiotropy and correlations.

5.4 Differences in genome-wide association studies

Genome-wide association studies can differ in terms of the markers, the dataset, models and methods. Different genome coverage, breed and families selected for every study will affect the detection and identification of QTL too. The number of QTL for body conformation and behaviour identified in previous studies (Ashwell et al. 2001; Schrooten et al. 2000; Spelman et al. 1999) is not directly comparable to the present study because of differences in the number and type of traits analysed, differences in experimental design, and differences in significance thresholds employed. Various studies (Barendse et al. 2007; Spelman et al. 1999) used permutation methodology to set the threshold line. Ashwell et al. (2005) used 95% QTL position confidence intervals. They also used suggestive ($P < 0.05$) and significant ($P < 0.01$) chromosome-wise F-value thresholds for the different traits to identify and summarize the putative QTL. Schrooten et al., (2000) used a genome-wise significant threshold of 10% (that was not adjusted for the number of traits) was used to determine the significant markers for the research.

de Roos et al. (2008) pointed out that SNPs found in LD with a QTL in reference population may not be identified as a significant SNP in the validation population. Such circumstances arises when the LD phase in the two populations are comparable between breeds, countries or generations. Dekkers & Hospital (2002) discussed that gene markers can only be validated if the LD phase between marker and QTL persists between the reference and validation population, therefore a SNP observed to be significant in the HF breed may not be significant in the JE breed. Hence selecting for the SNP in a selection population without determining the LD phase will not bring genetic improvement, and may even bring negative genetic response if the LD phase is reversed. The persistence of LD phase is therefore important for GWAS or genomic selection (Meuwissen et al. 2001). Wiggans et al. (2009) suggested eliminating SNPs that reduces the accuracy of GWAS

evaluation to decrease the analysis effort and improve the stability of estimation of effects from the remaining SNPs.

Hayes & Goddard (2010) claimed SNPs with significant associations with the traits of interest must then be tested in two different dairy cattle breeds in GWAS as the extent of LD phase across cattle breeds is limited, therefore using two breeds to map the causative mutation to approximate genome intervals is better than using only one breed. The SNPs with significant association with the QTL can then be confirmed on specific chromosomes.

5.5 Threshold probability values

There are several methodologies in multiple comparison statistical tests to identify SNPs adequately with the highest association rate and yet not too stringent to cut off other potentially significant SNPs. False discovery rate was introduced by Benjamini and Hochberg (1995) to use the predicted number of false rejections of the null hypothesis over the total proportion of rejections as a cut off value. This method is adaptive to the available data set in the studies, thus it is less stringent. False discovery rate is implemented in most genome scanning and mapping studies to identify the underlying genes for complex diseases and traits of interest. It also allows multiple phenotypes and genotypes to be analysed at the same time (Sabatti et al. 2003).

Setting a significant threshold for GWAS has been a questionable issue for various studies. Most QTL have small effects, thus setting a stringent threshold will cause some of them to go undetected. The polygenic effects may also be overestimated to be significant with a stringent threshold, thus deterring the attempt to predict phenotype just based on the genotype as it will only be restricted to genes with large effects. Predicting phenotype using SNP for most quantitative traits will be challenging too as genes of such traits consist of many QTL with small effects with epistatic interactions (Barendse 2005) pointed out that instead the use of stringent significance thresholds, understanding the replication of the associations is more important to detect QTL. Thus knowing the LD function between the marker and the causative mutation is necessary to avoid failure of SNP validation and overestimation of the significant effects.

As for this study, two threshold lines were set for the Manhattan plots to determine the significant SNPs for stature. One being the suggestive threshold which is the negative $\log_{10}$ value of 1×10^4 and the other is the genome-wide significant threshold at the negative $\log_{10}$ value of 5×10^5. The suggestive threshold line was appointed to highlight the SNPs worth noting for future research.

As shown in the Manhattan plots, the peaks of SNPs that rose above the genome-wide significance threshold were significant and influences stature in the 3 breeds of dairy cattle. SNPs on BTA3, 5, 11 and 12 (Figure 4.7) above the genome-wide significant threshold were located in the HF population in this study. SNPs on BTA9, 12, 19 and 25 above the genome-wide significant threshold were located in the JE population and the SNPs that were significant for stature were located on BTA1, 3, 10, 14, 22 and 24 in the XB population.

5.6 Genomic selection

Compared to selection based on phenotypic records alone, genomic selection using data on DNA level promises accelerated genetic gain. With molecular markers highlighting the location of useful traits, it is therefore easier for selection of important traits in any breeding program and to predict breeding values. Using the explicit information at DNA level, SNP data can be exploited by animal breeders to further enhance livestock. The additional information is based on the genetic variance accounted for by the QTN associated with the markers and contributes to selection. Genomic selection is more advantageous compared to marker-assisted selection or multiple ovulation and embryo transfer methods for genetic improvement.

Estimating BVs using genomic selection is very beneficial as reported by Schaeffer (2006). Pre-selecting young sires by utilizing EBV obtained from their genotypes will greatly reduce the generation interval as the bulls will be selected at the age of one compared to being selected at the age of five or six after their progeny are tested and proven. The generation interval for both sires and dams to breed replacement animals will be reduced as well. Genomic selection will also increase the selection intensity on the dams

to breed replacement bulls' selection pathway (Schaeffer 2006). Compared with the traditional progeny test scheme, genomic selection not only increases the accuracy of EBV, but greatly reduces the cost of genotyping too.

Inbreeding complications can be avoided by utilizing the heterozygosity index derived from the SNP genotypes of the animals. The SNP genotypes between identical animals are an exact match, therefore there will be no genetic improvement for the progenies as the GEBV values will still be the same. Mating between animals can be chosen based on the index rate to maximize the heterozygosity value for the next generation.

Lillehammer et al. (2011) discussed the benefits and methodology to combine progeny testing and genomic selection together by pre-selecting young bulls using their GEBVs for progeny testing. The BVs reported are more accurate and it does not need a complete restructuring of the breeding programme. Genomic selection is more favourable for young animals, but progeny testing increases the accuracy and reliability of the estimated BVs as the performance for the animals will be recorded and evaluated as they grow older.

The benefits of SNPs outweighs using microsatellites for LD mapping seeing SNPs are more abundantly found in the genome and have lower mutation rate (Snelling et al. 2005). SNPs usually come in biallelic form, and a pair of SNP markers will have four possible haplotype formats. Therefore a much larger set of SNPs are required to obtain all represented haplotype formats in the population for the genome analysis (Vignal et al. 2002). This aspect will enable genome-wide selection researches to identify all major QTL with large effects instead of picking out a single QTL that are situated near to a SNP as all panelled SNP across the genome can be detected..

It was pointed out that SNP effects needed to be re-estimated on a yearly basis (Schaeffer 2006) for breeding schemes that uses genomic selection to obtain an accurate GEBV for a selected trait (Sonesson & Meuwissen 2009). The haplotypes effects are re-estimated by adding new genotyped bulls to the reference population annually after obtaining their progeny records.

CHAPTER 6

CONCLUSIONS

Stature is genetically correlated with live weight, an economically important trait that is included in the breeding objective for the genetic improvement of New Zealand dairy cattle. The economic value for live weight is -$1.52 per kg, indicating a negative desired gain for this trait.

Selecting against stature will indirectly select against live weight, and will influence the genetic and economic gain for the dairy industry by elevating FCE. Identifying genetic markers (SNP) that are associated with stature and live weight highlights the QTLs that affect these traits, and thus can be used in breeding programmes to select either for or against stature and live weight using genomic selection. The industry should seek an intermediate optimum height for dairy cattle to avoid the undesirable effects that come with extreme heights such as lower FCE for taller cows or lower milk yield for smaller cows.

In this genome-wide association study for stature, 9 chromosomes (BTA2, 3, 4, 5, 6, 11, 12, 14 and 24) were observed to have SNPs significantly associated with stature in the HF population, 6 chromosomes (BTA9, 10, 12, 18, 19 and 25) in the JE population and 11 chromosomes (BTA1, 3, 4, 5, 7, 9, 10, 14, 18, 22 and 24) in the XB population based on the top 50 SNPs according to the *P*-value. Various SNPs located above the suggestive threshold in the Manhattan plots were also scrutinised and kept in view for future studies.

The SNPs identified from this study with significant associations to stature can serve as candidate SNPs for further investigation to determine the regions of QTLs and ultimately the exact genes that affect stature and other correlated traits in dairy cattle.

REFERENCES

Ahlborn G, Dempfle L 1992. Genetic parameters for milk production and body size in New Zealand Holstein-Friesian and Jersey. Livestock Production Science 31: 205-219.

Aguilar I, Misztal I, Tsuruta S, Wiggans GR, Lawlor TJ 2011. Multiple trait genomic evaluation of conception rate in Holsteins. Journal of Dairy Science 94: 2621-2624.

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. Journal of Animal Science 79: 2805-2811.

Ashwell MS, Heyen DW, Weller JI, Ron M, Sonstegard TS, Van Tassell CP, Lewin HA 2005. Detection of quantitative trait loci influencing conformation traits and calving ease in Holstein-Friesian cattle. Journal of Dairy Science 88: 4111-4119.

Barendse W 2005. The transition from quantitative trait loci to diagnostic test in cattle and other livestock. Australian Journal of Experimental Agriculture 45: 831-836.

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB 2007. A validated whole-genome association study of efficient food conversion in cattle. Genetics 176: 1893-1905.

Benjamini Y & Hochberg Y 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57: 289-300.

Berry DP, Buckley F, Dillon P, Evans RD, Rath M, Veerkamp RF 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. Journal of Dairy Science 86: 2193-2204.

Berry DP, Buckley R, Dillon P, Evans RD, Veerkamp RF 2004. Genetic relationships among linear type traits, milk yield, body weight, fertility and somatic cell count in primiparous dairy cows. Irish Journal of Agricultural and Food Research 43: 161-176.

Berry DP, Harris BL, Winkelman AM, Montgomerie W 2005. Phenotypic associations between traits other than production and longevity in New Zealand dairy cattle. Journal of Dairy Science 88: 2962-2974.

Berry DP, Lee JM, Macdonald KA, Roche JR 2007. Body condition score and body weight effects on dystocia and stillbirths and consequent effects on postcalving performance. Journal of Dairy Science 90: 4201-4211.

Boichard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, Rupp R, Amigues Y, Boscher M, Leveziel H 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genetics Selection Evolution 35: 77 - 101.

Bolormaa S, Pryce JE, Hayes BJ, Goddard ME 2010. Multivariate analysis of a genome-wide association study in dairy cattle. Journal of Dairy Science 93: 3818-3833.

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM, Goddard ME 2011. Genome-wide association studies for feedlot and growth traits in cattle. Journal of Animal Science 89: 1684-1697.

Brand B, Baes C, Mayer M, Reinsch N, Seidenspinner T, Thaller G, Kühn C 2010. Quantitative trait loci mapping of calving and conformation traits on Bos taurus autosome 18 in the German Holstein population. Journal of Dairy Science 93: 1205-1215.

Brotherstone S 1994. Genetic and phenotypic correlations between linear type traits and production traits in Holstein-Friesian dairy cattle. Animal Production 59: 183-188.

Caraviello DZ, Weigel KA, Gianola D 2003. Analysis of the relationship between type traits, inbreeding, and functional survival in Jersey cattle using a Weibull proportional hazards model. Journal of Dairy Science 86: 2984-2989.

Casas E, Shackelford SD, Keele JW, Koohmaraie M, Smith TPL, Stone RT 2003. Detection of quantitative trait loci for growth and carcass composition in cattle. Journal of Animal Science 81: 2976-2983.

Cassell B 2009. Using Heritability for Genetic Improvement. Virginia Cooperative Extension. http://pubs.ext.vt.edu/404/404-084/404-084.html [accessed 14 April 2012].

Coffey MP, Simm G, Hill WG, Brotherstone S 2003. Genetic evaluations of dairy bulls for daughter energy balance profiles using linear type scores and body condition score analyzed using random regression. Journal of Dairy Science 86: 2205-2212.

Cole JB, VanRaden PM, O'Connell JR, Van Tassell CP, Sonstegard TS, Schnabel RD, Taylor JF, Wiggans GR 2009. Distribution and location of genetic effects for dairy traits. Journal of Dairy Science 92: 2931-2946.

Cole J, Wiggans G, Ma L, Sonstegard T, Lawlor T, Crooker B, Van Tassell C, Yang J, Wang S, Matukumalli L, Da Y 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. BMC Genomics 12: 408.

Coppieters W, Riquet J, Arranz JJ, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M 1998. A QTL with major effect on milk yield and composition maps to bovine chromosome 14. Mammalian Genome 9: 540-544.

Coriolis 2009. Fonterra and the New Zealand dairy industry: options going forward. Coriolis Research – Reports. http://www.coriolisresearch.com/pdfs/coriolis_firm_dairy_03_0909_fonterra_options_going_forward_101a.pdf [accessed 5 September 2012].

Creagh F, Sherlock R 2012. gwasplot: produces gwas plots in R. R package version 1.0.

Cue RI, Harris BL, Rendel JM 1996. Genetic parameters for traits other than production in purebred and crossbred New Zealand dairy cattle. Livestock Production Science 45: 123-135.

Daetwyler HD, Villanueva B, Bijma P, Woolliams JA 2007. Inbreeding in genome-wide selection. Journal Of Animal Breeding And Genetics 124: 369-376.

DairyNZ, LIC 2012. New Zealand dairy statistics 2011-12. http://www.dairynz.co.nz/file/fileid/45159 [accessed 25 February 2013].

de Roos APW, Hayes BJ, Spelman RJ, Goddard ME 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. Genetics 179: 1503-1512.

Dekkers JCM, Hospital F 2002. Multifactorial genetics: the use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics 3: 22-32.

du Toit J, van Wyk JB, Maiwashe A 2012. Relationships between functional herd life and conformation traits in the South African Jersey breed. South African Journal of Animal Science 42: 47-54.

Druet T, Fritz S, Boussaha M, Ben-Jemaa S, Guillaume F, Derbala D, Zelenika D, Lechner D, Charon C, Boichard D, Gut IG, Eggen A, Gautier M 2008. Fine mapping of quantitative trait loci affecting female fertility in dairy cattle on BTA03 using a dense single-nucleotide polymorphism map. Genetics 178: 2227-2235.

Elo KT, Vilkki J, de Koning D-J, Velmala RJ, Mäki-Tanila AV 1999. A quantitative trait locus for live weight maps to bovine chromosome 23. Mammalian Genome 10: 831-835.

Falconer DS, Mackay TFC 1996. Introduction to quantitative genetics. 4th edition. New York, Longman. 480 p.

Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Research 10: 220-227.

Fortes MRS, Reverter A, Hawken RJ, Bolormaa S, Lehnert SA 2012. Candidate genes associated with hormone levels of inhibin, luteinising hormone, and insulin-like growth factor 1, testicular development, and sperm quality in Brahman bulls. Biology of Reproduction 87: 58-70.

Fuerst-Waltl B, Sölkner J, Essl A, Hoeschele I, Fuerst C 1998. Non-linearity in the genetic relationship between milk yield and type traits in Holstein cattle. Livestock Production Science 57: 41-47.

Garrett AJ, Rincon G, Medrano JF, Elzo MA, Silver GA, Thomas MG 2008. Promoter region of the bovine growth hormone receptor gene: Single nucleotide polymorphism discovery in cattle and association with performance in Brangus bulls. Journal of Animal Science 86: 3315-3323.

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139: 907-920.

Goddard ME 1998. Advances in dairy cattle breeding research. Proceedings intermediate report workshop EU concerted action genetic improvement of functional traits in cattle, Warsaw, Poland 23 August 1998. Interbull Bulletin 19. http://www-interbull.slu.se/ojs/index.php/ib/article/view/642/635 [accessed 20 April 2012].

Goddard ME, Hayes BJ 2007. Genomic selection. Journal of Animal Breeding And Genetics 124: 323-330.

Hansen M, Lund MS, Pedersen J, Christensen LG 2004. Gestation length in Danish Holsteins has weak genetic associations with stillbirth, calving difficulty, and calf size. Livestock Production Science 91: 23-33.

Harper SQ, Hauser MA, DelloRusso C, Duan D, Crawford RW, Phelps SF, Harper HA, Robinson AS, Engelhardt JF, Brooks SV, Chamberlain JS 2002. Modular flexibility of dystrophin: implications for gene therapy of Duchenne muscular dystrophy. Nature Medicine 8: 253-261.

Harris BL, Clark JM, Jackson RG 1996. Across breed evaluation of dairy cattle. Proceedings of the New Zealand Society of Animal Production 56: 12-15.

Harris BL, Kolver ES 2001. Review of Holsteinization on intensive pastoral dairy farming in New Zealand. Journal of Dairy Science 84: E56-E61.

Hawken RJ, Zhang YD, Fortes MRS, Collis E, Barris WC, Corbet NJ, Williams PJ, Fordyce G, Holroyd RG, Walkley JRW, Barendse W, Johnston DJ, Prayaga KC, Tier B, Reverter A, Lehnert SA 2011. Genome-wide association studies of female reproduction in tropically adapted beef cattle. Journal of Animal Science 90: 1398-1410.

Hayes B, Goddard M 2010. Genome-wide association and genomic selection in animal breeding. Genome 53: 876-883.

Hiendleder S, Thomsen H, Reinsch N, Bennewitz J, Leyhe-Horn B, Looft C, Xu N, Medjugorac I, Russ I, Kühn C, Brockmann GA, Blümel J, Brenig B, Reinhardt F, Reents R, Averdunk G, Schwerin M, Förster M, Kalm E, Erhardt G 2003. Mapping of QTL for body conformation and behaviour in cattle. Journal of Heredity 94: 496-506.

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307: 1072-1079.

Holmes CW, Wilson GF, Kuperus W, Buvaneshwa S, Wickham B 1993. Liveweight, feed intake and feed conversion efficiency of lactating dairy cows. Proceedings of the New Zealand Society of Animal Production 53: 95-99.

Howie BN, Donnelly P, Marchini J 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics 5: 1-15.

Jacquard A 1983. Heritability: One word, three concepts. Biometrics 39: 465-477.

Johanson JM, Berger PJ 2003. Birth weight as a predictor of calving ease and perinatal mortality in Holstein cattle. Journal of Dairy Science 86: 3745-3755.

Karim L, Takeda H, Li L, Druet T, Arias JAC, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B, Harris BL, Keehan MD, Littlejohn MD, Spelman RJ, Georges M, Coppieters W 2011. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nature Genetics 43: 405-413.

Kemper KE, Goddard ME 2012. Understanding and predicting complex traits: knowledge from cattle. Human Molecular Genetics 21: R25-51.

Kennedy BW, Dekkers JCM, Moore RK, Jairath L 1999. Genetic and phenotypic parameter estimates between production, feed intake, feed efficiency, body weight and linear type traits in first lactation Holsteins. Canadian Journal of Animal Science 79: 425-431.

Khatkar MS, Thomson PC, Tammen I, Raadsma HW 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genetics Selection Evolution 36: 163-190.

Khatkar MS, Collins A, Cavanagh JAL, Hawken RJ, Hobbs M, Zenger KR, Barris W, McClintock AE, Thomson PC, Nicholas FW, Raadsma HW 2006. A first-generation metric linkage disequilibrium map of bovine chromosome 6. Genetics 174: 79-85.

Kim J-J, Farnir F, Savell J, Taylor JF 2003. Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between Bos taurus (Angus) and Bos indicus (Brahman) cattle. Journal of Animal Science 81: 1933-1942.

Klungland H, Sabry A, Heringstad B, Olsen HG, Gomez-Raya L, Våge DI, Olsaker I, Ødegård J, Klemetsdal G, Schulman N, Vilkki J, Ruane J, Aasland M, Rønningen K, Lien S 2001. Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. Mammalian Genome 12: 837-842.

Kolbehdari D, Wang Z, Grant JR, Murdoch B, Prasad A, Xiu Z, Marques E, Stothard P, Moore SS 2008. A whole-genome scan to map quantitative trait loci for

conformation and functional traits in Canadian Holstein bulls. Journal of Dairy Science 91: 2844-2856.

Kühn C, Bennewitz J, Reinsch N, Xu N, Thomsen H, Looft C, Brockmann GA, Schwerin M, Weimann C, Hiendleder S, Erhardt G, Medjugorac I, Förster M, Brenig B, Reinhardt F, Reents R, Russ I, Averdunk G, Blümel J, Kalm E 2003. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. Journal of Dairy Science 86: 360-368.

Lillehammer M, Meuwissen THE, Sonesson AK 2011. A comparison of dairy cattle breeding designs that use genomic selection. Journal of Dairy Science 94: 493-500.

Littlejohn M, Grala T, Sanders K, Walker C, Waghorn G, Macdonald K, Coppieters W, Georges M, Spelman RJ, Hillerton E, Davis S, Snell R 2012. Genetic variation in PLAG1 associates with early life body weight and peripubertal weight and growth in Bos taurus. Animal Genetics 43: 591-594.

López de Maturana E, Ugarte E, Gonza´lez-Recio O 2007. Impact of calving ease on functional longevity and herd amortization costs in Basque Holsteins using survival analysis. Journal of Dairy Science 90: 4451-4457.

Lopez-Villalobos N 2012. Analysing the genetic basis of milk production traits. CAB Reviews 7: 1-18.

Mackay TFC 2001. Quantitative trait loci in Drosophila. Nature Reviews Genetics 2: 11-20.

MacNeil MD, Grosz MD 2002. Genome-wide scans for QTL affecting carcass traits in Hereford × composite double backcross populations. Journal of Animal Science 80: 2316-2324.

Marchini J, Howie B 2010. Genotype imputation for genome-wide association studies. Nature Reviews Genetics 11: 499-511.

Mc Hugh N, Evans RD, Fahey AG, Berry DP 2012. Animal muscularity and size are genetically correlated with animal live-weight and price. Livestock Science 144: 11-19.

Meijering A 1984. Dystocia and stillbirth in cattle — a review of causes, relations and implications. Livestock Production Science 11: 143-177.

Meuwissen THE, Hayes BJ, Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Meuwissen T, Goddard M 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185: 623-631.

Misztal I, Legarra A, Aguilar I 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. Journal of Dairy Science 92: 4648-4655.

Montgomerie B 2006. Experience with data recording and genetic evaluation of live weight for dairy cows in New Zealand. Interbull Bulletin 36. http://www-interbull.slu.se/bulletins/bulletin36/Montgomerie.pdf [accessed 27 June 2012].

Moore RK, Kennedy BW, Schaeffer LR, Moxley JE 1992. Relationships between age and body weight at calving, feed intake, production, days open, and selection indexes in Ayrshires and Holsteins. Journal of Dairy Science 75: 294-306.

Mounzih K, Qiu J, Ewart-Toland A, Chehab FF 1998. Leptin is not necessary for gestation and parturition but regulates maternal nutrition via a leptin resistance state. Endocrinology 139: 5259-5262.

Muller CJC, Cloete SWP, Olivier JJ, Botha JA, de Waal H 2006. Heritability of live weight and condition score in a Holstein herd and correlations with milk traits – preliminary estimates. South African Journal of Animal Science 36: 79-88.

Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, Sugimoto Y, Takasuga A 2012. Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. BMC Genetics 13: 40-50.

NZAEL 2012. New Zealand Animal Evaluation – All about BW. http://www.nzael.co.nz/all-about-bw [accessed 20 September 2012].

NZAEL 2012. New Zealand Animal Evaluation FAQ. http://www.nzael.co.nz/information/faq/success [accessed 5 September 2012].

NZAEL 2013. New Zealand Animal Evaluation – All about BW – Economic values. http://www.nzael.co.nz/all-about-bw [accessed 19 February 2013].

NZAEL 2013. New Zealand Animal Evaluation – Economic value update February 2013. http://www.nzael.co.nz/files/2013%20AE%20Economic%20Update.pdf [accessed 19 February 2013].

NZAEL 2013. New Zealand Animal Evaluation – Sire statistics. http://www.nzael.co.nz/animal-data/sire-statistics [accessed 31 January 2013].

Olsen HG, Gomez-Raya L, Våge DI, Olsaker I, Klungland H, Svendsen M, Ådnøy T, Sabry A, Klemetsdal G, Schulman N, Krämer W, Thaller G, Rønningen K, Lien S 2002. A genome scan for quantitative trait loci affecting milk production in Norwegian dairy cattle. Journal of Dairy Science 85: 3124-3130.

Pérez-Cabal MA, Alenda R 2002. Genetic relationships between lifetime profit and type traits in Spanish Holstein cows. Journal of Dairy Science 85: 3480-3491.

Peters SO, Kizilkaya K, Garrick DJ, Fernando RL, Reecy JM, Weaber RL, Silver GA, Thomas MG 2012. Bayesian genome-wide association analysis of growth and yearling ultrasound measures of carcass traits in Brangus heifers. Journal of Animal Science 90: 3398-3409.

Pribyl J, Rehout V, Citek J, Pribylova J 2010. Genetic evaluation of dairy cattle using a simple heritable genetic ground. Journal of the Science of Food and Agriculture 90: 1765-1773.

Pritchard JK, Przeworski M 2001. Linkage disequilibrium in humans: models and data. The American Journal of Human Genetics 69: 1-14.

Pryce JE, Harris BL 2006. Genetics of body condition score in New Zealand dairy cows. Journal of Dairy Science 89: 4424-4432.

Pryce JE, Hayes BJ, Bolormaa S, Goddard ME 2011. Polymorphic regions affecting human height also control stature in cattle. Genetics 187: 981-984.

Pryce JE, Daetwyler HD 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. Animal Production Science 52: 107-114.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81: 559-575.

R Development Core Team 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/.

Raadsma H, Jonas E, McGill D, Hobbs M, Lam M, Thomson P 2009. Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. Genetics Selection Evolution 41: 45-59.

Rajala PJ, Grönhn YT 1998. Effects of dystocia, retained placenta, and metritis on milk yield in dairy cows. Journal of Dairy Science 81: 3172-3181.

Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ 1998. Undesirable side effects of selection for high production efficiency in farm animals: a review. Livestock Production Science 56: 15-33.

SAS Institute Inc. 1987. SAS user's guide: Statistics. SAS Institute, Inc., Cary, NC.

Sabatti C, Service S, Freimer N 2003. False discovery rate in linkage and association genome screens for complex disorders. Genetics 164: 829-833.

Schaeffer LR 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123: 218-223.

Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS 2005. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Animal Genetics 36: 408-416.

Schrooten C, Bovenhuis H, Coppieters W, Van Arendonk JAM 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. Journal of Dairy Science 83: 795-806.

Short TH, Lawlor TJ 1992. Genetic parameters of conformation traits, milk yield, and herd life in Holsteins. Journal of Dairy Science 75: 1987-1998.

Slate J, Visscher PM, MacGregor S, Stevens D, Tate ML, Pemberton JM 2002. A genome scan for quantitative trait loci in a wild population of red deer (Cervus elaphus). Genetics 162: 1863-1873.

Snelling W, Casas E, Stone R, Keele J, Harhay G, Bennett G, Smith T 2005. Linkage mapping bovine EST-based SNP. BMC Genomics 6: 74-83.

Soller M 1994. Marker assisted selection - an overview. Animal Biotechnology 5: 193-207.

Sonesson AK, Meuwissen THE 2009. Testing strategies for genomic selection in aquaculture breeding programs. Genetics Selection Evolution 41: 37.

Spelman RJ, Coppieters W, Karim L, van Arendonk JAM, Bovenhuis H 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics 144: 1799-1807.

Spelman RJ, Garrick DJ 1997. Effect of live weight and differing economic values on responses to selection for milk fat, protein, volume, and live weight. Journal of Dairy Science 80: 2557-2562.

Spelman RJ, Huisman AE, Singireddy SR, Coppieters W, Arranz J, Georges M, Garrick DJ 1999. Short communication: quantitative trait loci analysis on 17 nonproduction traits in the New Zealand dairy population. Journal of Dairy Science 82: 2514-2516.

Spencer CCA, Su Z, Donnelly P, Marchini J 2009. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genetics 5: 1-15.

Toshniwal JK, Dechow CD, Cassell BG, Appuhamy JADRN, Varga GA 2008. Heritability of electronically recorded daily body weight and correlations with yield, dry matter intake, and body condition score. Journal of Dairy Science 91: 3201-3210.

Turner S 2011. Annotated Manhattan plots and QQ plots for GWAS using R, revisited. http://gettinggeneticsdone.blogspot.co.nz/2011/04/annotated-manhattan-plots-and-qq-plots.html [accessed 9 July 2012].

van Kaam J, Groenen M, Bovenhuis H, Veenendaal A, Vereijken A, van Arendonk J 1999. Whole genome scan in chickens for quantitative trait loci affecting growth and feed efficiency. Poultry Science 78: 15-23.

Veerkamp RF, Brotherstone S 1997. Genetic correlations between linear type traits, food intake, live weight and body condition score in Holstein Friesian dairy cattle. Animal Science 64: 385-392.

Veerkamp RF 1998. Selection for economic efficiency of dairy cattle using information on live weight and feed intake: a review. Journal of Dairy Science 81: 1109-1119.

Veerkamp RF, Koenen EPC, De Jong G 2001. Genetic correlations among body condition score, yield, and fertility in first-parity cows estimated by random regression models. Journal of Dairy Science 84: 2327-2335.

Velmala RJ, Vilkki HJ, Elo KT, de Koning DJ, Mäki-Tanila AV 1999. A search for quantitative trait loci for milk production traits on chromosome 6 in Finnish Ayrshire cattle. Animal Genetics 30: 136-143.

Vignal A, Milan D, SanCristobal M, Eggen A 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution 34: 275-305.

Weigel KA, de los Campos G, González-Recio O, Naya H, Wu XL, Long N, Rosa GJM, Gianola D 2009. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. Journal of Dairy Science 92: 5248-5257.

Wiggans GR, Sonstegard TS, VanRaden PM, Matukumalli LK, Schnabel RD, Taylor JF, Schenkel FS, Van Tassell CP 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science 92: 3431-3436.

Zaborski D, Grzesiak W, Szatkowska I, Dybus A, Muszynska M, Jedrzejczak M 2009. Factors affecting dystocia in cattle. Reproduction in Domestic Animals 44: 540-551.

Zhang C, de Koning DJ, Hernández-Sánchez J, Haley CS, Williams JL, Wiener P 2004. Mapping of multiple quantitative trait loci affecting Bovine Spongiform Encephalopathy. Genetics 167: 1863-1872.

Zink V, Štípková M, Lassen J 2011. Genetic parameters for female fertility, locomotion, body condition score, and linear type traits in Czech Holstein cattle. Journal of Dairy Science 94: 5176-5182.