Effect of CNV of Production Related Genes on Expression of Production Traits in Dairy Cattle

A Senior Project

presented to

the Faculty of the Dairy Science Department

California Polytechnic State University, San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Science

by

Tyler Kamper

March, 2012

© 2012 Tyler Kamper

ACKNOWLEDGEMENT

This project would not have been possible without the help and support of many people. I would first like to thank my parents, who raised me to have a love for the dairy industry and have supported me throughout my education here at Cal Poly and previously in high school and elementary school. Their continued desire for my success has been a driving force behind my accomplishments. Second, I would like to thank my project supervisor, Dr. Bruce Golden, for his help during this learning experience. His enthusiasm for the topic of interest encouraged me to dig deeper and deeper in order to see this project to completion. I would also like to extend my gratitude to Steve Maddox of Maddox Dairy for providing me with the necessary production information for particular animals. Without this information I would not have been able to make the conclusions that I did.

ABSTRACT

The objective of this project was to determine if copy number variation (CNV) at specific genes related to BST had an effect on milk production. Hair samples from 1000 fresh heifers from Maddox Dairy in Riverdale, CA were sent to Geneseek, a Neogen Company for DNA extraction and processing. The DNA was processed using the BovineSNP50 BeadChip and the data were loaded into the GenomeStudio software. The data were manipulated in GenomeStudio to estimate the CNV at each single nucleotide polymorphism (SNP) of which a vast majority was two. The SNPs with a copy number of two were excluded and all remaining copy number values were converted into a comma separated value file and loaded into an Excel spreadsheet for final processing. This resulted in 283,558 SNPs from 578 individuals. The genes studied were growth hormone (GH1), growth hormone receptor (GHR), growth hormone releasing hormone (GHRH), growth hormone releasing hormone receptor (GHRHR), insulin-like growth factor (IGF-1), and insulin (INS). SNPs located near the start and end of the gene sequences for those genes were found that had a copy number of three. PTAs of animals with SNPs located near these gene sequences were analyzed to determine if copy number had an effect on production. There were 138,411 total SNPs with a copy number of three. Of these 138,411 SNPs, 28 were intronic with the exception of IGF-1 which had no SNPs with CNV in the intron. For IGF-1, the closest exonic SNPs were used. Copy number variation did occur for SNPs that were in introns for all the genes previously mentioned except for IGF-1. No CNV was identified in the SNPs closest to IGF-1.

Key words: copy number variation, growth hormone, growth hormone releasing hormone, insulin-like growth factor, insulin, single nucleotide polymorphism

TABLE OF CONTENTS

LIST OF TABLES vi
LIST OF FIGURES vii
INTRODUCTION 1
LITERATURE REVIEW 1
History of Genetic Testing 1
Single Nucleotide Polymorphisms 3
Quantitative Trait Loci 4
Illumina BovineSNP50 4
Copy Number Variation 5
Bovine Somatotropin 6
Growth Hormone 7
Growth Hormone Receptor 7
Growth Hormone Releasing Hormone 8
Growth Hormone Releasing Hormone Receptor 8
Insulin-like Growth Factor 8
Insulin 9
MATERIALS AND METHODS 9
Data Collection 9
Data Processing 10
CNV Data 12
Genes of Interest 17
DHI Data 24
Critique of Study 25
CONCLUSION 26
REFERENCES 27

LIST OF TABLES

Table 1. The number of SNPs, the average score for those SNPs, the maximum score for those SNPs, the minimum score for those SNPs and the standard deviation for those SNPs for each CN value 13

Table 2. The number of SNPs (Number of SNP) and how many times that number of SNP occurred (Occurrence(s)) when CN=4 14

Table 3. The number of individuals that had CN=0, 1, 3, and 4 for SNPs located in the middle of the gene sequence for the specified genes..... 23

Table 4. Sample ID used during data extraction and interpretation and the corresponding ID of the animal from the farm..... 24

LIST OF FIGURES

Figure 1. Snapshot of the Genome Studio software with data from this experiment. 10

Figure 2. The number of SNPs and the number of times those SNPs had a copy number of 3. 14

Figure 3. The number of SNPs and the number of times those SNPs had a copy number of 1. 15

Figure 4. The number of SNPs and the number of times those SNPs had a copy number of 0. 16

Figure 5. Average score of SNPs located in the middle of the gene sequence for GH1 when CN=3 for two separate animals..... 18

Figure 6. Average score of SNPs located in the middle of the gene sequence for GHR when CN=3 for four animals. 19

Figure 7. Average score of SNPs located in the middle of the gene sequence for GHRH when CN=3 for two separate animals..... 20

Figure 8. Average score of SNPs located in the middle of the gene sequence for GHRHR when CN=3 for three animals..... 21

Figure 9. Average score of SNPs located in the middle of the gene sequence for INS when CN=3 for four animals. 22

INTRODUCTION

The effects of bovine somatotropin (bST) on milk production are evident. This heterogenic peptide hormone can also be used by the dairy farmer to increase initial peak production therefore increasing overall milk for lactation using a recombinant form (rBST). But could an animal peak higher and give more milk throughout lactation based on their genetics? This is what made copy number variation (CNV) an area of interest. With more copies of the gene, maybe an animal would be able to have a more persistent lactation curve or possibly peak higher and give more milk throughout their lactation. Other related genes such as growth hormone releasing hormone (GHRH), insulin-like growth factor (IGF-1), and insulin (INS) are also production related and have different effects on production traits. For this reason, they too were looked at during this study.

The objective of this study was to determine if copy number variation (CNV) has any effect on the intensity of how a trait is expressed, specifically the BST and associated genes that may be related to milk production in dairy cows.

LITERATURE REVIEW

History of Genetic Testing

Breeding cows for superior genetics and production traits has long been a goal for dairy farmers. Much research has been done to allow for easier and simpler ways of detecting good, solid producing cows. Since the mid-1960's the ability to breed for exceptional animals and determine genetic differences has expanded greatly due to the

evolution of using molecular markers to look at genetic variation (Melka et al., 2011). There are numerous genes in the cattle genome. One gene is made up of three structures known as nucleotides. These nucleotides come in five forms. Adenine (A), Thymine (T), Guanine (G), and Cytosine (C) are the nucleotides that are present in DNA. The fifth nucleotide, Uracil (U), is a component of RNA which is paired with Adenine when DNA goes through the process of transcription into messenger RNA (**mRNA**; Berning, 2011).

Some genes have a large effect on a physiological feature, while most genes have a minimal effect on the same type of features. Because of this, a model known as the infinitesimal model has been widely used when dealing with statistical analysis of genes. The infinitesimal model looks at the physiological features and accredits them to a large number of loci with small effects on the perceived physiological features (Liu and Dekkers, 1997).

An older method of determining which genes effected physiological features was by using restriction fragment length polymorphisms (RFLP). These RFLPs can give information about restriction size variation (Powell et al., 1996). RFLPs have been used with success since the 1980's for many plant species (Powell et al., 1996) and have been used for cattle as well. Powell et al. (1996) also determined that because of the fact that some polymorphisms could not be successfully identified, the progress for using RFLPs has slowed. Another reason for the decreased use of RFLPs is because it is a very time consuming process and is also very labor intensive (Powell et al., 1996). Microsatellites have also been used to do the same. These microsatellites, like RFLPs, detect polymorphisms in DNA sequences (Powell et al., 1996). With the advancement of technology and the help of the infinitesimal model, now a sample of hair, blood, or even

nose swabs can be taken and through DNA extraction and analyzing, potential traits of several cattle breeds can be determined.

Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) have now become the standard method for looking at genetic markers in order to analyze large amounts of genetic information in DNA. A SNP is when in a DNA sequence one base pair differs between chromosomes. For example, if one sequence of DNA has the nucleotides AGT on one chromosome of a pair, the other chromosome may have the nucleotides AGA at the same exact point as the previous chromosome. The reason for choosing SNPs as genetic markers is because they are understood to have applicable genetic information (Melka et al., 2011). SNPs are contributors to differences in phenotypes, disease risk, and drug as well as environment responses (Li et al., 2009).

Fairly new to the world of animal genetics is genome-wide selection (GWS), a variation on genome-wide association (GWA). Genome-wide association can be used in cattle to find the genomic regions that are contributors to natural genetic variation in phenotypic traits. These regions can then be fine-mapped using a higher marker density in order to better identify candidate genes. The density of the markers to be used is based on the average length of chromosome "blocks" that possess a high level of linkage disequilibrium (LD). This LD between two loci is measured by an r^2 statistic. This statistic is the squared correlation between alleles present at two loci which are usually on the same chromosome. Within these LD blocks, some markers could act as indicators to predict most of the haplotypes that would be present in the entire block (Matukumalli et

al., 2009). This helps to improve our knowledge on Mendelian sampling because rather than looking at the parent averages we can look at the genetics of the animal in question and see what is actually the case. Matukumalli et al. (2009) explained that dairy farmers can use this information to predict genetic merit values to select what direction they want to go in their breeding program.

Quantitative Trait Loci

As first discussed by Neiman-Sorensen and Robertson (1961) and Smith (1967), single loci were thought to contribute to metric traits (Abdel-Azim and Freeman, 2002). A Quantitative Trait Loci (QTL) is a stretch of DNA that either contains or is linked to specific phenotypes. This means that QTL can be used to determine where in the genome specific genes are located that deal with phenotypic traits that are expressed. Quantitative trait loci for traits having to do with milk production have been found on the bovine genome on all of the autosomes (Khatkar et al., 2004). Abdel-Azim and Freeman (2002) also cited that many other studies that have been conducted have found many QTL are associated with both production and type traits in dairy cattle. Liu et al. (2004) states in the introduction of the paper *Quantitative Trait Loci Mapping for Dairy Cattle Production Traits Using a Maximum Likelihood Method* that many milk production traits are thought to be influenced by many polygenes and are subject to quantitative inheritance.

Illumina BovineSNP50

One analyzing technique to view SNPs utilizes the Illumina BovineSNP50 BeadChip. Illumina has several different BeadChips which can be used in various animal

breeds including bovine, ovine, goat and porcine (Illumina, 2012). The bovine BeadChips can either be processed using the Illimina Infinium® HD Assay or the Illumina Golden State® Assay. The BovineSNP50 is one of the BeadChips that utilizes the Illumina Infinium® HD Assay (Illumina, 2010). This piece of technology contains over 54,000 evenly spaced SNP probes (Melka et al., 2011) that can give users the genetic information about the sampled animals. The Illumina BovineSNP50 BeadChip has been used for various studies and has proven itself by boasting a call rate of 99.6% according to a study done by Matukumalli et al. (2009) on parentage SNPs. Matukumalli et al. (2009) also determined that in their study the overall call rate for all markers was greater than 99.1% and over 90% of the markers had call rates above 99.98%. Call rate is essentially the percent of a DNA sample that can be read. A higher call rate means that more of the DNA in the sample is pure and not contaminated by impurities that would make it unreadable. With a call rate of this magnitude, it can be seen that the BovineSNP50 BeadChip can be an effective tool. This BeadChip is also cost-effective (Illumina, 2011).

Copy Number Variation

Copy number variation (CNV) of genes occurs in both human and bovine (Hou et al., 2011). According to Hou et al. (2011) CNV, like SNPs, is another key source of genetic variation. Most genes come with two copies; however there are times when that number is changed. Either an addition or deletion mutation occurs within the genetic makeup of the animal resulting in an abnormal number of copies of a particular gene. In humans, CNV is believed to have links to various diseases including mental retardation, neurological disorders, and cancer (Koike et al., 2011). Rincon et al. (2011) found that

CNV analyses were more definite when using the Illumina BovineHD platform compared to the BOS 1 platform because there was a greater marker density which had much less background noise and increased $\log_2$ ratio ranges. It was also seen in the study by Rincon et al. (2011) that there was a difference between CNV detected between Holsteins and Jerseys.

Bovine Somatotropin

One particular gene that is of special interest is the gene that codes for bovine somatotropin (bST). Somatotropin has been shown to affect the growth of the animal as well as maintaining an animal's lactation. There have already been four different types of bST found in cattle (Lee et al., 1996). Bovine somatotropin is a heterogenic peptide hormone, meaning it is made up of a number of different amino acids. It acts by affecting a target organ, in this case the liver. When it binds to bST receptors in the liver, stimulation from Growth Hormone (**GH**; Mullen et al., 2010) causes a release of Insulin-like Growth Factor (IGF-1). This protein is responsible for conducting the movement of nutrients throughout the body. In the non-lactating animal, this process works to physically grow the animal in which the process is being performed. However, in a lactating animal, those nutrients are transported to the mammary gland to aid in milk production (Berning, 2011). In theory, the more bST that is being produced by the animal, the more milk a producer should be seeing in the tank, given there are enough receptors in the liver for all the bST that is being produced. To take this one step further, if more bST and receptors means more milk then there is a possibility that more copies of the genes that code for bST and its receptors also means more milk. According to

Molento et al. (2002) the plasma level of bST decreases as days in milk (DIM) increases. This could be the reason why the lactation curve of a milking cow diminishes over time.

Growth Hormone

Growth Hormone (GH1), which is released from the anterior pituitary, plays a large role in control of lactation because it stimulates the release of IGF-1 (Mullen et al., 2010). Because Schlee et al. (1994) found evidence of an association between GH1 genetic variants and GH plasma levels, it was concluded in that study that some variation of the levels of GH could be due to mutations in the GH1 gene. This makes it a good potential marker for improving or at least looking at milk production traits (Mullen et al., 2010) because different variations of the gene have different effects on milk production as a whole. Growth Hormone can be found on chromosome nineteen starting at 48732931bp and ending at 48825660bp.

Growth Hormone Receptor

Growth Hormone Receptor (GHR) is what controls the effects of GH (Kobayashi et al., 1999). This receptor is part of the cytokine-hematopoietin receptor family and is found mostly in the liver, however it is present throughout many tissues. Different gene promoters regulate how the GHR messenger RNA (mRNA) are expressed. The three that have been found in cattle are promoters 1, 2, and 3 (**P1, P2, P3**; Kobayashi et al., 1999). The promoter active only in the liver is P1 and could be the overall reason for concentration of GHR in the liver. Growth Hormone Receptor can be found on chromosome twenty starting at 31784034bp and ending at 32633244bp.

Growth Hormone Releasing Hormone

Growth Hormone Releasing Hormone (GHRH) was identified in 1982. It is released from the hypothalamus and is responsible for the release of GH from the anterior pituitary. It is in competition with somatostatin, which works to inhibit the release of GH (Vance, 1990). It also has an indirect effect on IGF-1. This is because it releases the hormone (GH) which stimulates the release of IGF-1. Growth Hormone Releasing Hormone can be found on chromosome thirteen starting at 66567169bp and ending at 66927386bp.

Growth Hormone Releasing Hormone Receptor

Growth Hormone Releasing Hormone Receptor (GHRHR) is what GHRH binds to in order to allow it to control the secretion of GH. Mutations of GHRHR have been seen to cause individuals to have a short stature in humans (Wajnrajch, et al., 1996). Growth Hormone Releasing Hormone Receptor can be found on chromosome four starting at 65667089bp and ending at 65984140bp.

Insulin-like Growth Factor

Insulin-like Growth Factor (IGF-1) like the name implies is quite similar to insulin in its molecular structure. This compound is a mitogenic polypeptide present in all mammalian species which is identical in human, bovine, porcine, and ovine species. By connecting to IGF-1 receptors it acts by stimulating growth, differentiation and metabolism in many types of cells. Insulin-like Growth Factor can be found on chromosome five starting at 66293057bp and ending at 67062354bp.

Insulin

Blood plasma concentrations of insulin (INS) have appeared to be related to the availability of glucose in lactating dairy cows (Bines et al., 1980; Hove and Blom, 1983; Vasilatos and Wangsness, 1981). Bines et al. (1980) also hypothesized that the amount of glucose and other nutrients for milk production may be controlled by the ratio of GH to INS. Insulin can be found on chromosome twenty-nine starting at 49955760bp and ending at 50065231bp.

MATERIALS AND METHODS

Data Collection

Data for this project were collected from Maddox Dairy in Riverdale, California during the Summer and Fall of 2011. This operation is home to about 3500 registered cows of which some are grade-ups. The cows are on a three time a day milking rotation with about 3400 going into the bulk tank and 100 cows contributing to hospital milk. The owner uses the DHI-Plus computer software to keep records of all the livestock associated with the operation.

Hair sampling cards were provided by Geneseek, a Neogen Company for sampling of 1000 registered Holstein fresh heifers from Maddox Dairy. Hair was pulled from the tail switch of these 1000 animals and placed on the hair sampling cards. The plastic slip cover was lifted and the follicle end of the hair was placed on the left side of the card with the hair sticking out the right side. The hair sticking out of the card was then cut even with the edge of the card. Twenty to thirty good hair follicles were required to have an adequate amount of sample.

Data Processing

After being collected, the hair samples were sent to GeneSeek, a Neogen Company. GeneSeek took the hair sampling cards and extracted the DNA from them. The DNA was then processed by GeneSeek using the BovineSNP50 BeadChip. Intensity files were also supplied by GeneSeek and loaded into the GenomeStudio software (Fig. 1).

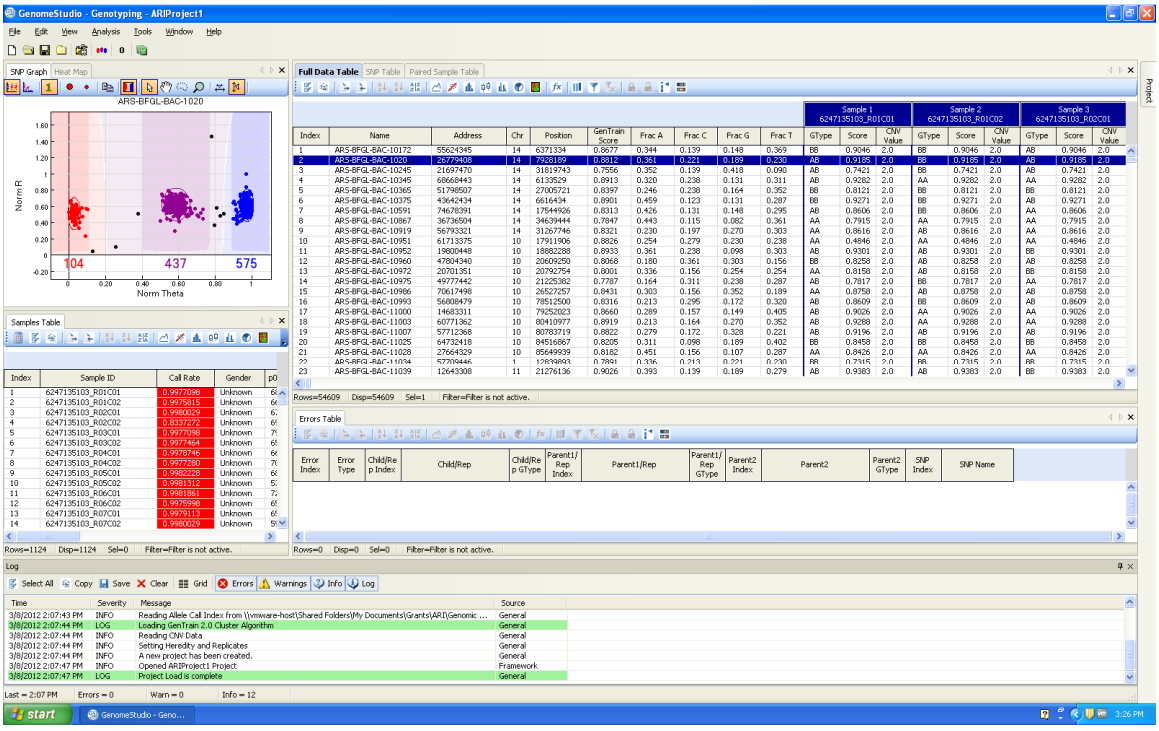

Figure 1. Snapshot of the Genome Studio software with data from this experiment.

The GenomeStudio software converted the intensity files information into CNV information at each marker on the BovineSNP50 BeadChip. Information from close to 50,000 loci for over 1000 animals, a total of 61,380,522 pieces of data, was converted from intensity files information into CNV information. The data were manipulated in the GenomeStudio software to estimate the copy number at each SNP of which a vast

majority were two copies. Default parameters were used. A report was exported from the GenomeStudio software to a comma separated value (CSV) file. This file was too large to export into an Excel spreadsheet so we filtered out all the observations where copy number was equal to two using AWK. This final file was loaded into Excel for final processing. A standard deviation (std), median, and average of the 283,558 scores of SNPs were calculated. A pivot table was produced in Excel that calculated Number of SNP, Average Score, Maximum Score, Minimum Score, and Standard Deviation for each copy number (CN) value. The CN values were 0, 1, 3, and 4. Where CNV was found, correlation was determined between copy number of production related genes and production traits for the sampled animals.

The chromosome and location of six genes considered important to milk production were found and used to find target locations to look at for SNP data that may have been related to milk production. These genes were growth hormone (GH1), growth hormone receptor (GHR), growth hormone releasing hormone (GHRH), growth hormone releasing hormone receptor (GHRHR), insulin-like growth factor (IGF-1), and insulin (INS). The SNPs that were discovered in the middle of the gene sequence for the aforementioned genes were matched with the animals from which the sample was taken. Only SNPs that had a copy number value of three were looked at since these were the most common of the atypical copy number value. Once these data were found, the Predicted Transmitting Abilities (PTAs) of the animals with a copy number of three for SNPs located in the middle of the gene sequence for production related genes were studied to determine if there was a relationship between the number of copies and the production of each individual.

RESULTS AND DISCUSSION

CNV Data

Different values for the counts and scores of each different CN were calculated for copy numbers of zero, one, three, and four. These values were calculated for all SNPs, not just the ones in around the genes that were studied. After narrowing the copious amount of data collected for all CN values into only data for SNPs with zero, one, three or four copy numbers, 283,558 SNPs remained. A CN of three was found most with 138,411 SNPs having three copies. Next were SNPs that had no copies. A total of 108,034 SNPs did not have any copies. A copy number of one and four were far behind those two with 37,001 SNPs and 112 SNPs, respectively. The average score for CN increased as CN increased from zero to two, which is usually close to 1.0. As CN increased from two to four however, the average score for CN decreased. The maximum scores of each CN were all above 0.90, the lowest being 0.9319 at CN=4 and the highest being 0.9728 for CN=3. The minimum scores were zero for all CN values. The std were all quite small with a range from 0.17856782 to 0.347109302 with an average std of 0.2487 (Table 1).

Table 1. The number of SNPs, the average score for those SNPs, the maximum score for those SNPs, the minimum score for those SNPs and the standard deviation for those SNPs for each CN value

CN Value	Number of SNP	Average of Score	Max of Score	Min of Score	StdDev of Score
0	108034	0.688035273	0.9678	0	0.250547403
1	37001	0.753977852	0.9701	0	0.17856782
3	138411	0.720000798	0.9728	0	0.261043224
4	112	0.595604464	0.9319	0	0.347109302
Total	283558	0.712206617	0.9716	0	0.248740444

An average, median, and standard deviation (std) were calculated using the scores of all of the data that were loaded into the Excel spreadsheet. The scores for each SNP were based on a proprietary algorithm (Illumina, 2008) and ranged between zero and one with one indicating a very high degree of quality of the call for the CNV of the SNP. Thus, a score near 1.0 also meant that the score for that particular SNP was relatively reliable. The average score was 0.7122. This average was not very close to one however 20,012 values equaling zero were calculated within this average. Having this many zeros in this average had a negative effect on the outcome. The median was 0.802. This median was low as well but the zero scores were also used in this calculation. A std of 0.2487 was also calculated showing that the variation between scores was high.

Data from each CN value was also placed into separate pivot tables according to CN value (CN = 0, 1, 3, or 4). These data showed how many times a certain number of loci appeared for each CN. For example, when looking at CN=4 (Table 2), 46 SNPs had one occurrence of CN=4. In other words, 46 SNPs had four copies once. The data for CN=3, 1, and 0 were also included in this section (Figures 2, 3, and 4, respectively). "Blank" referred to the fact that there were no data for the number of corresponding loci and Grand Total referred the total number of SNPs (Table 2, Figures 2, 3, and 4).

Table 2. The number of SNPs (Number of SNP) and how many times that number of SNP occurred (Occurrence(s)) when CN=4

Occurrence(s)	Number of SNP
1	46
2	18
5	6
Blank	46747
Grand Total	46817

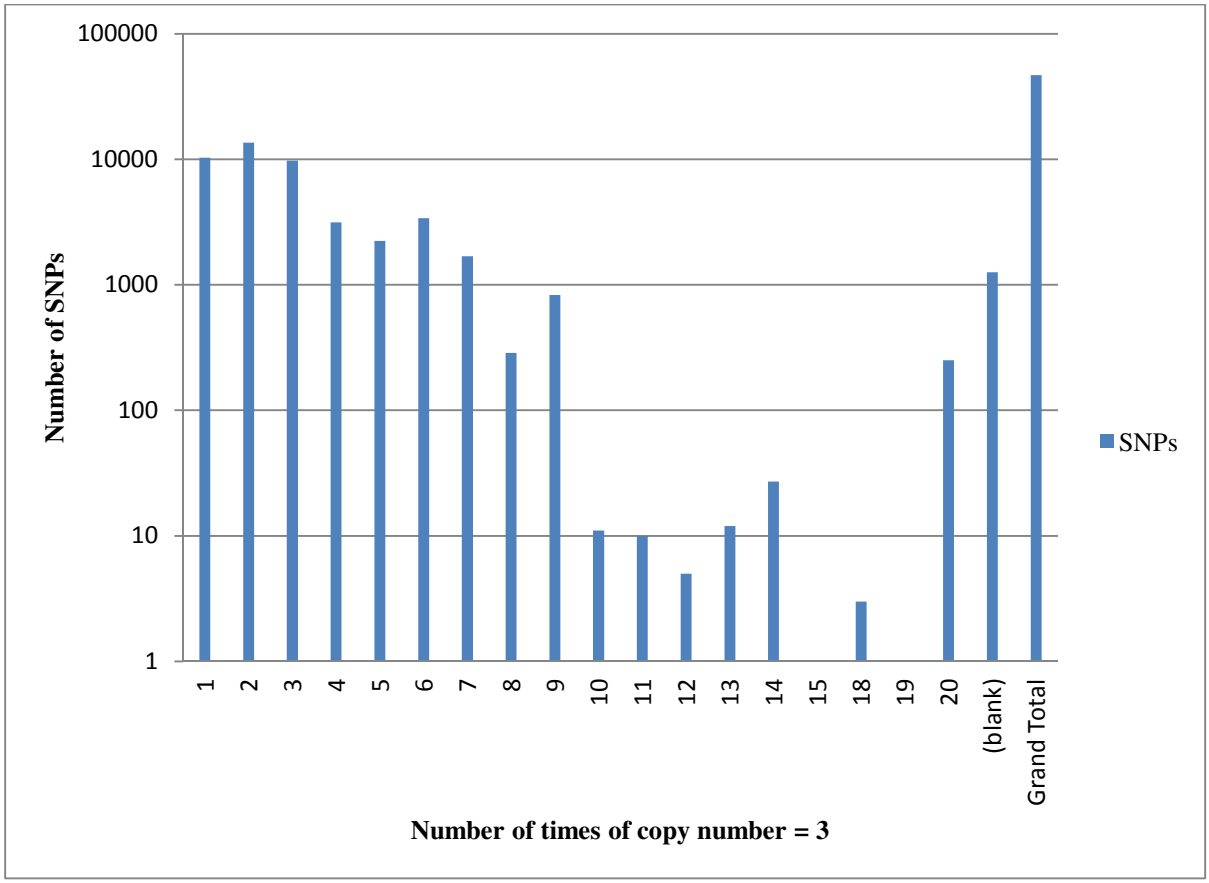

Figure 2. The number of SNPs and the number of times those SNPs had a copy number of 3.

For CN=3 we can see that a high number of SNPs were observed only a few times. Figure 2 also shows that the number of SNPs having multiple occurrences of

CN=3 decreases overall as occurrences increases. This makes sense because one would not expect to have a lot of SNPs to have multiple occurrences of more than two copy numbers, since two copy numbers is standard. However, there is a spike to 251 SNPs that had 20 occurrences of CN=3.

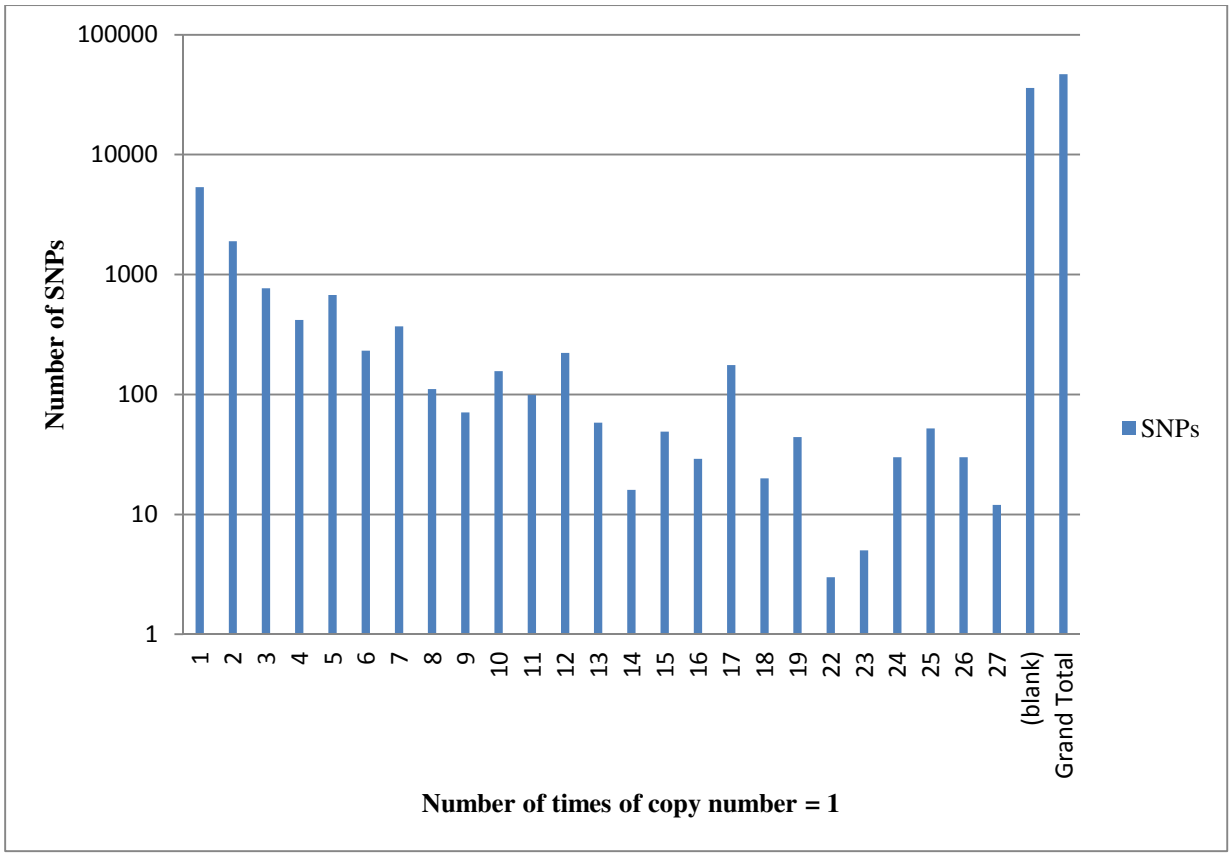

Figure 3. The number of SNPs and the number of times those SNPs had a copy number of 1.

When looking at Figure 3 we can see that the number of SNPs once again decreased as number of occurrences increased. There was no spike at the end of the data.

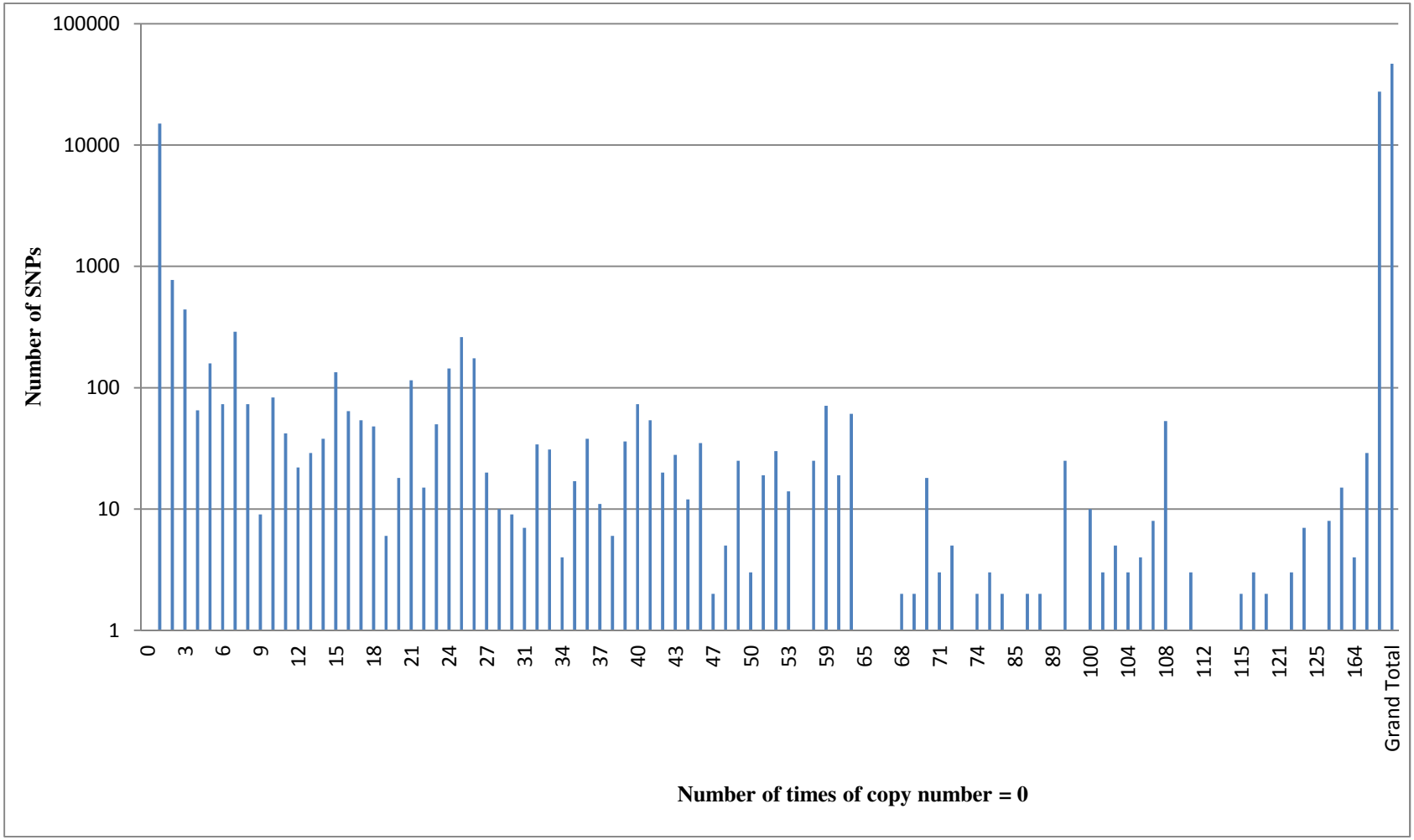

Figure 4. The number of SNPs and the number of times those SNPs had a copy number of 0.

The same trend that occurred when CN=1 also occurred when CN=0. The number of SNPs decreased as the number of occurrences increased. There were fourteen occurrences which had only one SNP at that occurrence. Since only one SNP was at these fourteen separate occurrences, the figure did not show a bar for them. This was because the figure starts out at one and not zero because it is on a logarithmic scale and when making a logarithmic scale in Excel the figures automatically start at one and cannot be changed to start at zero.

Genes of Interest

There were six genes of interest that we chose to study related to BST. These genes were growth hormone (GH), growth hormone receptor (GHR), growth hormone releasing hormone (GHRH), growth hormone releasing hormone receptor (GHRHR), insulin-like growth factor (IGF-1), and insulin (INS). Figures 4, 5, 6, 7, and 8 show the results for the animals with SNPs located in the middle of the sequence of each of the genes mentioned above, excluding IGF-1 because there were no SNPs located in the middle of that gene sequence. Keady et al. (2011) noticed in their study titled *Effect of sire breed and genetic merit for carcass weight on the transcriptional regulation of the somatotropic axis in longissimus dorsi of crossbred steers* that IGF-1could be a potential molecular marker for muscle growth by looking at small RNA regulation, transcription factors, and copy number variation along with SNP variation. Our project showed that there was no copy number variation of IGF-1.

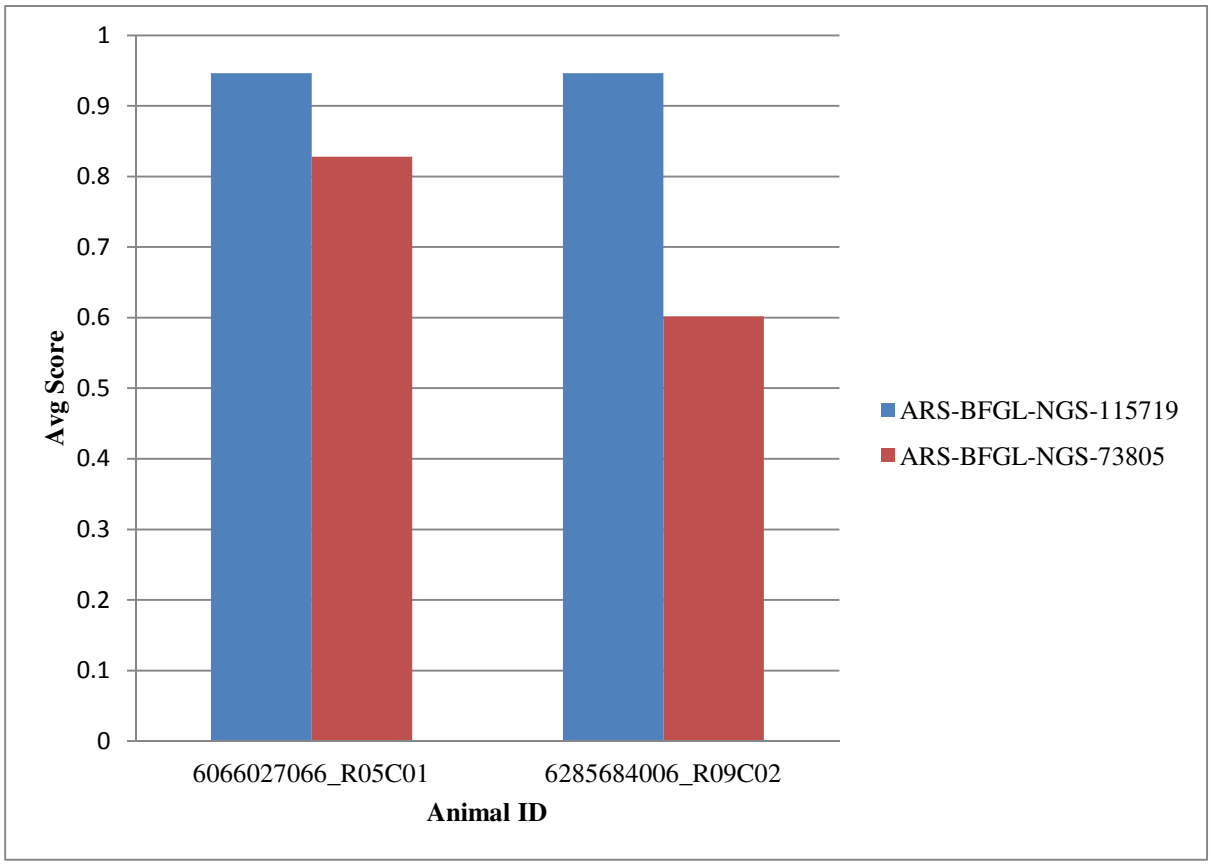

Figure 5. Average score of SNPs located in the middle of the gene sequence for GH1 when CN=3 for two separate animals.

Only two SNPs occurred in the middle of the gene sequence for GH1. These SNPs were ARS-BFGL-NGS-115719 and ARS-BFGL-NGS-73805. There were two animals which these SNPs were found to have the above values. The sample identification of these animals was 6066027066_R05C01 and 6285684006_R09C02. The average CN for each SNP in each animal was three, which was the CN value at which the attention was focused. The average score of ARS-BFGL-NGS-115719 was 0.9462 and the average score for ARS-BFGL-NGS-73805 was 0.7150.

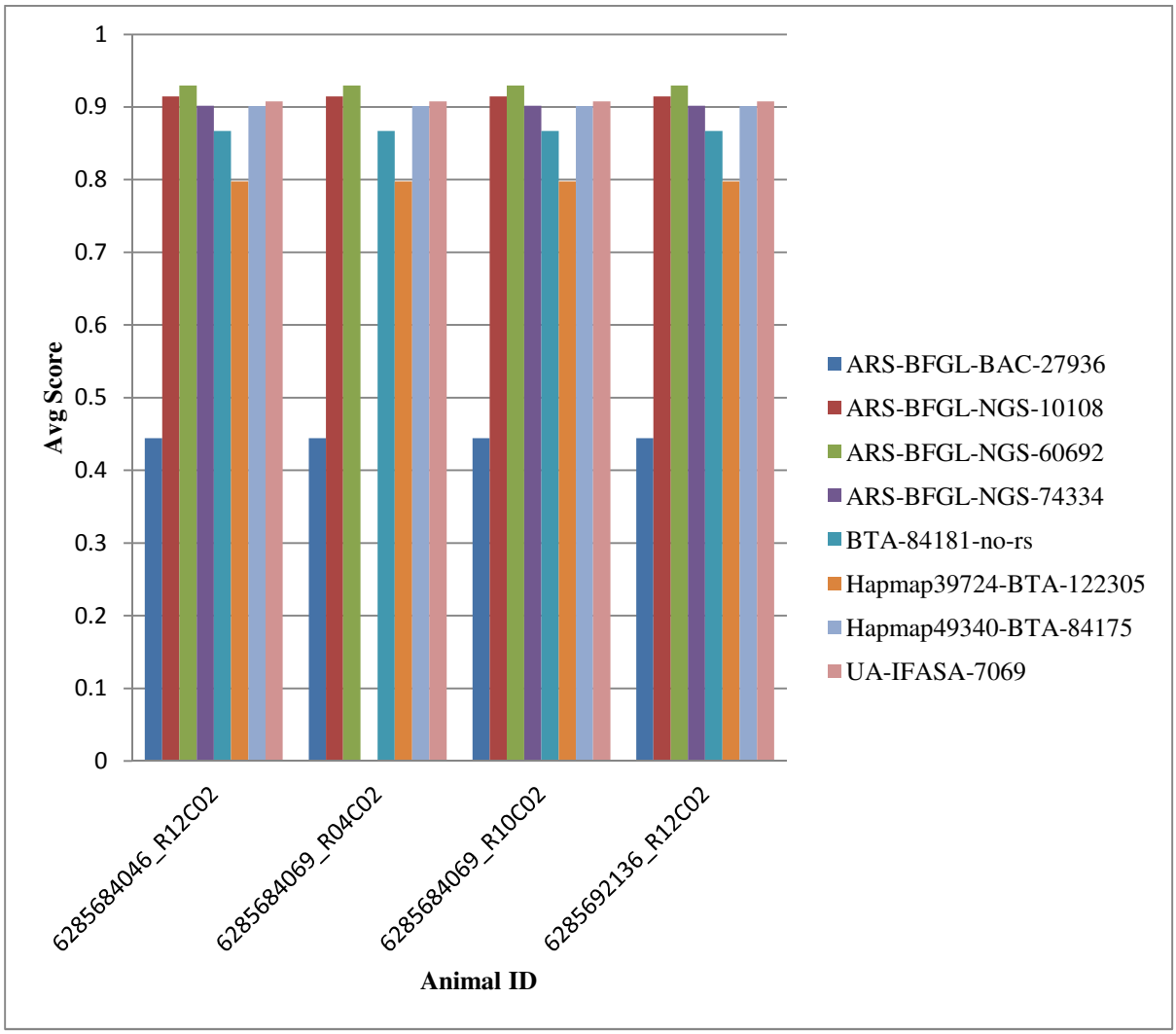

Figure 6. Average score of SNPs located in the middle of the gene sequence for GHR when CN=3 for four animals.

Figure 6 shows the average score (vertical axis) of eight different SNPs that were located in the middle of the gene sequence for GHR. All eight of these SNPs occurred in four animals (horizontal axis). The average CN was again three, because this was the CN value at which attention was focused. Animal number 6285684069_R04C02 had a score of zero for ARS-BFGL-NGS-74334 and SNP ARS-BFGL-BAC-27936 had a low score across the board of 0.4442. However, the rest were all above 0.8000 with the exception

of Hapmap39724-BTA-122305 which had an average score of just below that with 0.7978.

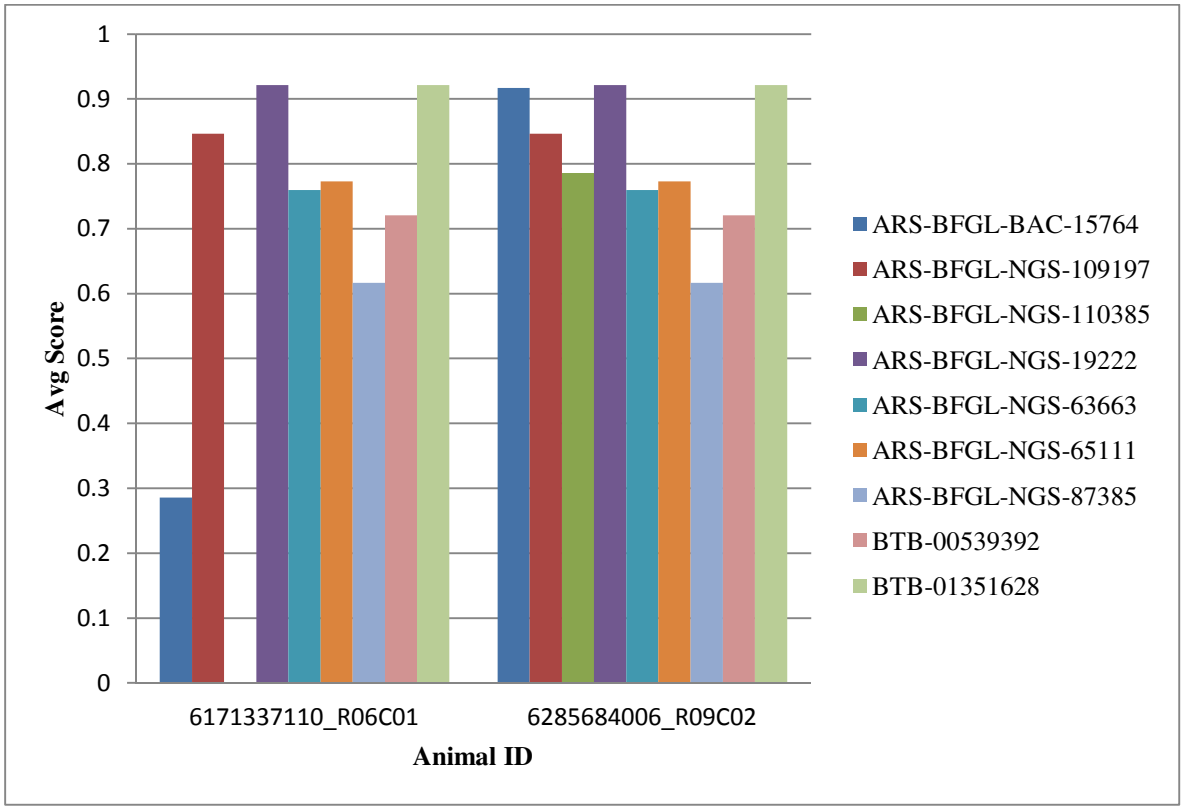

Figure 7. Average score of SNPs located in the middle of the gene sequence for GHRH when CN=3 for two separate animals.

Nine SNPs were located between the start and end locations of the gene sequence for GHRH. Once again, the average CN value was three because that was the CN value at which the attention was focused. Animal number 6171337110_R06C01 had a score of zero for SNP ARS-BFGL-NGS-110385 and had a score of 0.2858 for SNP ARS-BFGL-BAC-15764. All other SNPs scores were fairly good with all but one SNP being below 0.7000 for both animals.

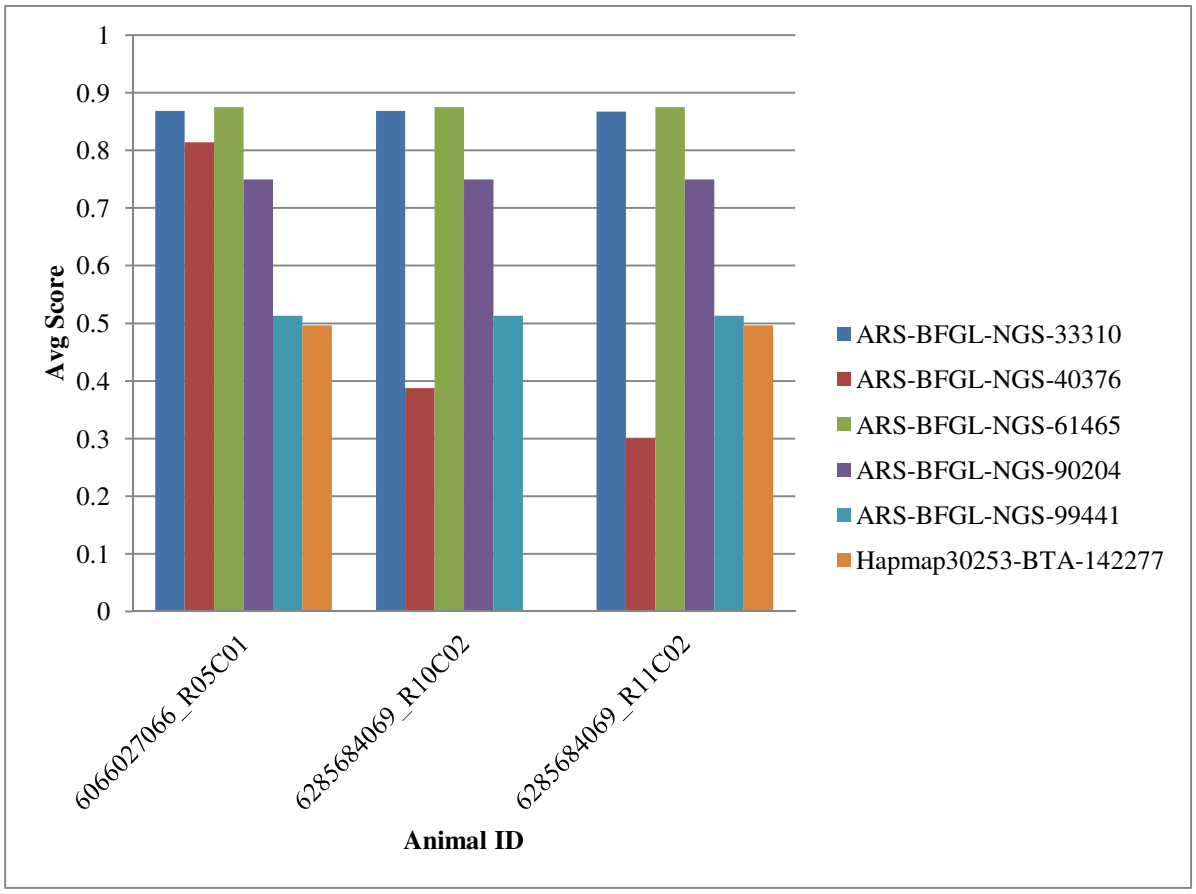

Figure 8. Average score of SNPs located in the middle of the gene sequence for GHRHR when CN=3 for three animals.

Six SNPs were found between the start and end locations of the GHRHR gene sequence that had a CN value of three. These six SNPs were found in three animals and the scores of many of the SNPs were consistent throughout the sampled animals. However, two SNPs did change. First, SNP Hapmap30253-BTA-142277 was the same for both the first and third animal on Figure 8 but the second animal, 6285684069_R10C02, had a score of zero for that SNP. The other SNP with discrepancy between animals was ARS-BFGL-NGS-40376. This SNP decreased by animal from left

to right starting with 0.8141 for the first animal, going to 0.3878 for the second animal, and ending with 0.3005 for the third animal.

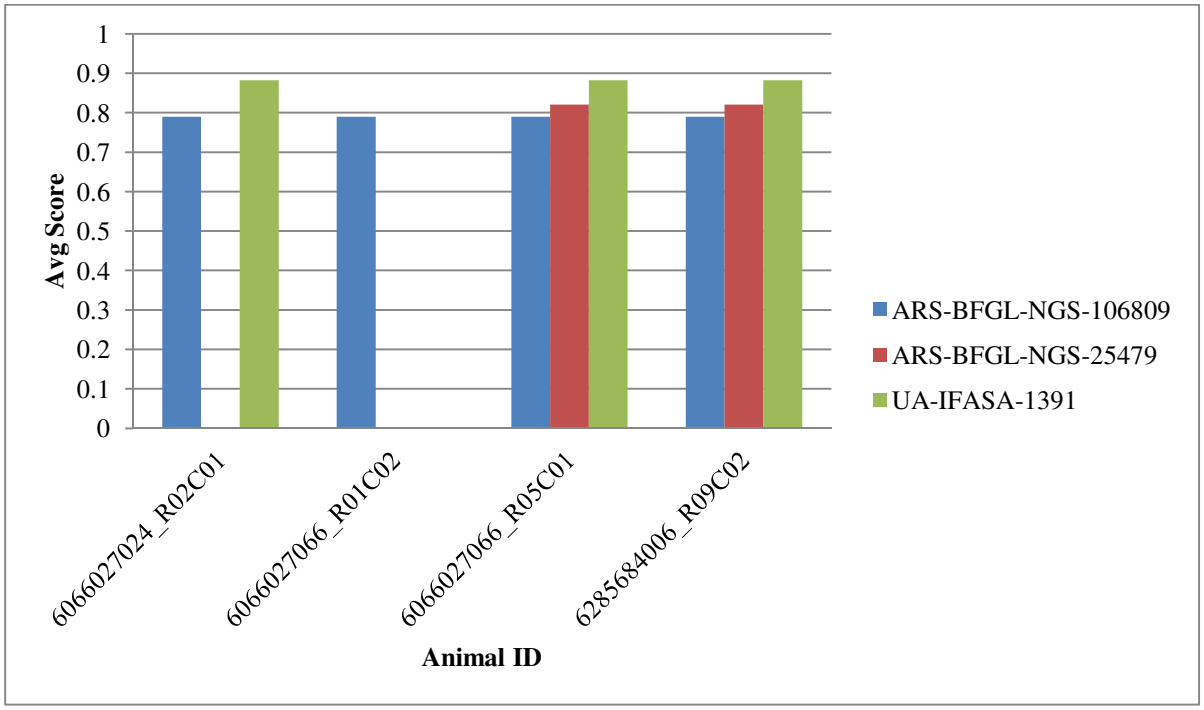

Figure 9. Average score of SNPs located in the middle of the gene sequence for INS when CN=3 for four animals.

There were three SNPs that fell between the start and end locations of the gene sequence for INS. There were multiple animals that had these SNPs however, the animals had either zero or one CN and it was determined that we would look at only those animals with a CN value of three. The SNP ARS-BFGL-NGS-25479 had a score of zero for the first two animals and SNP UA-IFASA-1391 had a score of zero for the second animal. The scores for ARS-BFGL-NGS-106809 were consistent for each animal at 0.7898.

As mentioned previously there were no SNPs located in the middle of the gene sequence for IGF-1. Therefore no information was available to include in Table 3 for IGF-1. The rest of the production related genes did have SNPs in the middle of the gene sequence. Table 3 shows the number of individuals that had a SNP or multiple SNPs in the middle of the gene sequence for the specified genes at the different copy numbers listed. For GH1, there were two SNPs that each had two individuals with three copy numbers. Therefore four individuals had a copy number of three for SNPs located in the middle of the gene sequence for GH1. The GHR gene had two individuals that had zero copies of the SNPs that were intronic and forty-one individuals that had three copies of intronic SNPs. Growth hormone releasing hormone had twenty individuals with a copy number of three SNPs in the middle of the gene sequence. Its receptor, GHRHR, had six animals with copy number zero and eighteen animals with copy number three for SNPs located in the middle of the gene sequence. Insulin had many individuals with a span of copy numbers. There were 116 individuals that had zero copies of SNPs located in the middle of the gene sequence, ten animals with a copy number of one, and twelve animals with a copy number of three for these SNPs.

Table 3. The number of individuals that had CN=0, 1, 3, and 4 for SNPs located in the middle of the gene sequence for the specified genes

Gene	CN=0	CN=1	CN=3	CN=4
GH1	0	0	4	0
GHR	2	0	41	0
GHRH	0	0	20	0
GHRHR	6	0	18	0
IGF-1	N/A	N/A	N/A	N/A
INS	116	10	12	0

Using a cross-reference data sheet, the actual on farm animal identification numbers were associated with the corresponding seventeen character sample identification numbers that were used in the previous figures. Table 4 shows the corresponding identification numbers. Animal 84659 had a copy number of three for SNPs located in the middle of the gene sequence for GH1, GHRH, and INS. Animal 92248 had a copy number of three for SNPs located in the middle of the gene sequence for GHR and GHRHR. Animals 94097 93077 had SNPs with a copy number of three located in the middle of the gene sequence for GHR. For the gene GHRHR animal 93397 had a copy number of three for SNPs located in the middle of the gene sequence.

Table 4. Sample ID used during data extraction and interpretation and the corresponding ID of the animal from the farm.

Gene	Sample ID	Animal ID
GH1	6285684006_R09C02	84659
GHR	6285684069_R04C02	94097
GHR	6285684069_R10C02	92248
GHR	6285692136_R12C02	93077
GHRH	6285684006_R09C02	84659
GHRHR	6285684069_R10C02	92248
GHRHR	6285684069_R11C02	93397
INS	6285684006_R09C02	84659

DHI Data

DHI datasheets were provided by the owner of Maddox Dairy. From these sheets we took the production information to determine if those animals with CNV for

production related genes had increased milk production compared to herd average. Animal 84659 had a lifetime milk total of 29,680 pounds. She had a relative value of 97%. This animal also had the most production related genes, as seen in Table 4, with a copy number of three. Animal 92248 had two production related genes with a copy number of three (Table 4). Her lifetime milk total was 27,730 pounds and she had a relative value of 110%. The final three animals each had one gene with a copy number of three. Animal 93397 had a copy number of three for the GHRHR gene (Table 4). Her lifetime milk total was 23,460 pounds. She had a relative value of 116%. The final two animals, 93077 and 94097 had a copy number of three for the gene GHR (Table 4). These animals also had the lowest lifetime milk totals with 21,970 pounds and a relative value of 130% for animal 93077 and 15,140 pounds with a relative value of 92% for animal 94097.

Critique of Study

In this study six production related genes were studied. Additional studies on this topic should be performed to see what information can be taken from a larger data set. Along with this, additional production related genes should be studied. Additional traits should also be looked at. This study focused on total lifetime production and relative value. Looking into other production traits such as fat, protein, and cheese yield may also be beneficial.

CONCLUSION

After looking at the results, it was apparent that there is copy number variation for SNPs in the middle of the gene sequences for production related genes such as growth hormone (GH), growth hormone receptor (GHR), growth hormone releasing hormone (GHRH), growth hormone releasing hormone receptor (GHRHR), and insulin (INS). This study focused only on a copy number of three for these SNPs based on the fact that from the data collected a copy number value of three was most common among the other atypical copy number values. Having found CNV within these sample animals is a promising step for the dairy cattle industry. If we as an industry were able to find a way to manipulate CNV in our favor, we could see a revolution of how dairying is currently done.

Although the information we got from the limited number of animals we were able to look at DHI records for is not enough to make any definite conclusions, it does show that a study like this should be performed again on a larger population of animals, possibly the Holstein USA database. With a larger number of animals for which there is production information there would be a better conception of whether or not CNV of production related genes has an effect on production.

REFERENCES

Abdel-Azim, G., and A. E. Freeman. 2002. Superiority of QTL-Assisted Selection in Dairy Cattle Breeding Schemes. J. Dairy Sci. 85:1869-1880.

Berning, L. 2011. Lactation Physiology. Lecture material. California Polytechnic State University, San Luis Obispo. 6:31,42-43.

Bines, J. A., I. C. Hart, and S. V. Morant. 1980. Endocrine control of energy metabolism in the cow: the effect on milk yield and levels of some blood constituents of injecting growth hormone and growth hormone fragments. Br. J. Nutr. 43:179.

Hou, Y., G. E. Liu, D. M. Bickhart, M. F. Cardone, K. Wang, E. S. Kim, L. K. Matukumalli, M. Ventura, J. Song, P. M. VanRaden, T. S. Sonstegard, and C. P. Van Tassell. 2011. Genomic characteristics of cattle copy number variations. Bmc Genomics 12:127.

Hove, K., and A. K. Blom. 1973. Plasma insulin and growth hormone in dairy cows: Diurnal variation and relation to food intake and plasma sugar and acetoacetate levels. Acta Endocrinol. 73:289.

Illumina. 2008. GenomeStudio™ Genotyping Module v1.0 User Guide. Illumina, Inc. San Diego, CA.

Illumina. 2010. Revolutionizing genetics in the cattle industry. Extensive toolset. Proven technology. Collaborative approach. Accessed Jan. 26, 2012. http://www.illumina.com/Documents/products/brochures/BovineProductBrochure.pdf .

Illumina. 2011. BovineSNP50 Genotyping BeadChip. Featuring 54,609 evenly spaced SNP probes that span the bovine genome. Accessed Jan. 21, 2012. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp5O.pdf.

Illumina. 2012. Agriculture. Accessed Jan. 26, 2012. http://www.illumina.com/applications/agriculture.ilmn#ag_consortia.

Kang, S. H., J. U. Kim, J. Y. Imm, S. Oh, and S. H. Kim. 2006. The Effects of Dairy Processes and Storage on Insulin-Like Growth Factor-I (IGF-I) Content in Milk and in Model IGF-I-Fortified Dairy Products. J. Dairy Sci. 89:402-409.

Keady, S. M., D. A. Kenny, M. G. Keane, and S. M. Waters. 2011. Effect of sire breed and genetic merit for carcass weight on the transcriptional regulation of the somatotropic axis in longissimus dorsi of crossbred steers. J. Anim. Sci. 89:4007-4016.

Khatkar, M. S., P. C. Thomson, I. Tammen, and H. W. Raadsma. 2004. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. Genet. Sel. Evol. 36(2004) 163-190.

Kobayashi, Y., M. J. Vandehaar, H. A. Tucker, B. K. Sharma, and M. C. Lucy. 1999. Expression of Growth Hormone Receptor 1A Messenger Ribonucleic Acid in Liver of Dairy Cows During Lactation and After Administration of Recombinant Bovine Somatotropin. J. Dairy Sci. 82:1910-1916.

Koike, A., N. Nishida, D. Yamashita, and K. Tokunaga. Comparative analysis of copy number variation detection methods and database construction. BMC Genetics. 12:29

Lee, B. K., G. F. Lin, B. A. Crooker, M. P. Murtaugh, L. B. Hansen, and H. Chester-Jones. 1996. Association of somatotropin (bST) gene polymorphism at the 5th exon with selection for milk yield in Holstein cows. Domest. Anim. Endocrinol. 13(4):373-381.

Li, R., Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, and J. Wang. 2009. SNP detection for massively parallel whole-genome resequencing. Genome Res. 19:1124-1132.

Liu, Y., G. B. Jansen, and C. Y. Lin. 2004. Quantitative Trait Loci Mapping for Dairy Cattle Production Traits Using a Maximum Likelihood Method. J. Dairy Sci. 87:491-500.

Liu, Z., and J. C. M. Dekkers. 1997. Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. Genetics 148:495-505.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and Characterization of a High Density SNP Genotyping Assay for Cattle. PLoS One 4(4):e5350.

Melka, H. D., E. Jeon, S. Kim, J. Han, D. Yoon, and K. Kim. 2011. Identification of Genomic Differences between Hanwoo and Holstein breeds using the Illumina Bovine SNP50 BeadChip. Genomics and Informatics 9(2):69-73.

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

Molento, C. F. M., E. Block, R. I. Cue, and D. Petitclerc. 2002. Effects of Insulin, Recombinant Bovine Somatotropin, and Their Interaction on Insulin-Like Growth Factor-I Secretion and Milk Protein Production in Dairy Cows. J. Dairy Sci. 85:738-747.

Mullen, M. P., D. P. Berry, D. J. Howard, M. G. Diskin, C. O. Lynch, E. W. Berkowicz, D. A. Magee, D. E. MacHugh, and S. M. Waters. 2010. Associations between novel single nucleotide polymorphisms in the *Bos taurus* growth hormone gene and performance traits in Holstein-Friesian dairy cattle. J. Dairy Sci. 93:5959-5969.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2:225-238.

Rincon, G., K. L. Weber, A. L. Van Eenennaam, B. L. Golden, and J. F. Medrano. 2011. Performance of bovine high density genotyping platforms in Holsteins and Jerseys. J. Dairy Sci. 94:6116-6121.

Schlee, P., R. Graml, E. Schallenberger, D. Schams, O. Rottmann, A. Olbrich-Bludau, and F. Pirchner. 1994. Growth hormone and insulin-like growth factor I concentrations in bulls of various growth hormone genotypes. Theor. Appl. Genet. 88:497–500.

Vance, M. L. 1990. Growth-hormone-releasing-hormone. Clin. Chem. 36;3:415-420.

Vasilatos, R., and P. J. Wangsness. 1981. Diurnal variations in plasma *insulin* and growth hormone associated with two stages of lactation in high producing dairy cows. Endocrinology 108:300.

Wajnrajch, M. P., J. M. Gertner, M. D. Harbison, S. C. Chua Jr., and R. L. Leibel. 1996. Nonsense mutation in the human growth hormone-releasing hormone receptor causes growth failure analogous to the little (*lit*) mouse. Nature Genetics 12:88-90.