GENOME-WIDE ASSOCIATION STUDY OF THE WOODY BREAST MYOPATHY IN BROILER CHICKENS

A Thesis

by

KUBRA CILKIZ

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	David Riley
Co-chair of Committee,	Giridhar Athrey
Committee Member,	Clare Gill
Head of Department,	G. Cliff Lamb

August 2018

Major Subject: Animal Science

ABSTRACT

Increasing demand for poultry meat has encouraged production of the maximum possible yield from chicken, which resulted in broiler chickens being subjected to intense selection for increased Pectoralis major (P. major) breast muscle yield and growth rate. That is presumed to contribute to development of new myopathies, which mainly impairs the nutritional value of the meat and meat quality, and in turn, affects consumer acceptance. A recent and prominent example is woody breast (WB) myopathy characterized by palpably tough breast muscle, which imposes an additional economic burden on the poultry industry. The basis of WB myopathy is not known yet but is suspected to be genetic due to its association with growth rate. Development of the WB myopathy is tightly linked to growth rate, which is under intense selection. On the other hand, growth rate and body mass are complex traits, and WB myopathy is associated with growth rate, which suggests that WB myopathy is more likely a complex trait. Genome-wide association studies (GWAS), which have become more feasible with recent improvements in high-throughput genotyping technologies, were exploited with the aim of the investigating the underlying etiology and genetic mechanism of woody breast myopathy. Using 600 K HD Affymetrix Axiom Array, 10 affected broilers and 10 healthy White Plymouth Rocks (WPR) were genotyped. SNPs and constructed haplotype blocks were examined whether they exhibited statistically significant association with woody breast myopathy. The current study conducted by SNPs and haplotypes suggested that several biological pathways and genes associated with oxidative stress, Ca^{2+} binding, microtubule motor activity and cellular repair could be contributing factors regarding the onset of woody breast myopathy. However, analysis of results clearly showed that the sample

introduced biases related to selection of study participants into the association analysis**.** Due to small sample size, after correcting for population stratification, it was not possible to detect true disease variants or genes genuinely affecting woody breast myopathy.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. David Riley, my committee co-chair, Dr. Giridhar Athrey, and my committee member, Dr. Clare Gill, for their guidance, support and patience throughout the course of this research, and I owe them the greatest degree of appreciation. I also express my gratitude to Animal Science and Poultry Science Departments and all department faculty and staff for making my time at Texas A&M University a great experience. I would like to thank the Ministry of Food, Agricultural, and Livestock and Department of the General Directorate Agricultural Research and Policy in Turkey. Lastly, I am deeply thankful to my mother and father for their encouragement, everlasting support and prayers, and to my husband, Mustafa, for his endless patience and support.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis committee consisting of Dr. Giridhar Athrey of the Department of Poultry Science and Dr. David Riley of the Department of Animal Science and Dr. Clare Gill of the Department of the Animal Science.

All other works conducted for the thesis was completed by the student independently. Ms. Zozik is sponsored by the Ministry of National Education of Turkey.

NOMENCLATURE

P. major	Pectoralis major
kg	Kilogram
%	Percentage
WB	Woody breast
USDA	United States Department of Agriculture
WS	White striping
QTL	Quantitative trait locus
RH	Radiation hybrid
BACs	Bacterial artificial chromosomes
FISH	Fluorescence in situ hybridization
Gb	Gigabyte
Mbp	Million base pair
DNA	Deoxyribonucleic acid
GWAS	Genome-wide association studies
SNP	Single nucleotide polymorphism
K	Thousand
HD	High Density
UTR 3^{I}	Three prime untranslated region
UTR 5^{I}	Five prime untranslated region
ncRNA	Non-coding RNA
WPR	White Plymouth Rock

Ross 708	Broiler chickens
IACUC	Institutional Animal Care and Use Committee
NE	Nebraska
χ^2	Chi-squared
FDR	False discovery rate
FDR-BH	False discovery rate -Benjamini and Hochberg
CMplot	Circle Manhattan Plot
NCBI	National Center for Biotechnology Information
GO	Gene Ontology
IPA	Ingenuity Pathway Analysis
LD	Linkage disequilibrium
EM	Expectation Maximization
IBS	Identity by state
CMH	Cochran Mantel-Haenszel
MDS	Multidimensional scaling
dbSNP	The Single Nucleotide Polymorphism Database
Nov	Novel
e	Euler's number
n	number of SNPs
log	Logarithm
LGE64	Linkage group E64
Q-Q plot	Quantile-quantile plot
W	WPR

R	Ross
IDs	Identifiers
Ca^{2+}	Calsium
bp	Base pair
Chr	Chromosome
G	Guanine
T	Thymine
A	Adenine
C	Cytosine
ATP	Adenosine Tri-Phosphate
GTP	Guanosine-5'-triphosphate

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Page

1. INTRODUCTION

1.1 General Information

Poultry production plays a key role in the economies of developed and developing countries. Chicken is one of the most important agricultural species and a popular source of animal protein around the world. Economical production of poultry products (both eggs and meat), affordability, and nutritional profile have helped chicken consumption increase over the past few decades (Li et al., 2017). In many of the European countries, chicken-based poultry production is a major part of meat marketed, which is more than 60% in Belgium (Xiang et al., 2017), and in the United States consumption of chicken is at an all-time high (over 44 kg/per person/year).

The rising need for animal products is not surprising considering the increase in population and standard of living, especially in developing and underdeveloped countries. Of the total 1,200 million tons of animal production in 2011, 60% was milk and milk products (Szűcs, 2013). The remaining 40% consisted of pork (8.8%), poultry meat (6.7%), sheep and beef meat (6.4), eggs (5.6%), and fish (12.4%). The increases in production have made nutritious sources of animal protein more affordable around the world. Poultry meat and eggs are promoted as healthier alternatives to sheep and beef meat due to their nutritive properties and low-fat content (Szűcs, 2013). Due to the favorable feed conversion ratio, and lower environmental footprint, poultry production is claimed as a sustainable source of animal protein. Additionally, because water availability constitutes a significant concern for the future, the total amount of water that livestock require to produce meat would be a critical factor for this industry. Comparing chicken with other

livestock, like cattle and pigs, they require a significantly lower amount of water for the production of 1 kg meat (AVEC, 2012). Rapid return on investment, no specific need for agricultural land, low feeding cost, and generation of income any time of the year are among the advantages of poultry production compared to other livestock animals (Szűcs, 2013).

Chickens are also a well-studied model organism for vertebrate developmental biology due to their similarity to mammalian embryology and providing a suitable model of human genetic diseases. Chickens have a history of a variety of debilitating diseases, including muscular dystrophy, epileptic seizures, and viral infection (Schmutz & Grimwood, 2004), which are important sources of human morbidity.

1.2 Emergent Muscle Myopathies

Increasing demand for poultry meat has encouraged production of the maximum possible yield from chicken, which resulted in broiler chickens being subjected to intense selection to gain more body weight in the shortest time possible. In 1925, the commercial poultry industry raised broiler chickens to attain 1.1 kg average body live weight for 112 days, and in 2015, broiler chickens can reach 2.8 kg average live body weight in 48 days (National Chicken Council, 2015). Chickens reared for meat production, commonly known as broilers, have undergone intensive genetic selection, which brings particular advantages to the poultry industry like a decrease in generation interval, more body weight gain and increased feed conversion. Some detrimental effects accompany this intense selection, including problems with the quality of meat, like flavor and appearance (Anthony, 1998). Also, greater body size has brought about enhanced metabolic

requirements related to fast growth rate, which constitutes an increased hazard regarding morbidity and commercial profit (Julian, 2005).

One of the major concerns of the poultry meat producer is the increasing incidence of myopathies. It has been shown previously that the amount of muscle fibers is subject to change with growth rate. As such, farm animals with rapid growth rates produce more muscle fibers compared to their counterparts that are growing more slowly (Stickland, 1995). In chicken, a breast muscle, the pectoralis major, is the predominant contributor to the body mass and size; thus, it is of the greatest significance in the broiler industry (USDA, 2016). Broiler chickens have a well-documented record of myopathies, particularly those affecting the pectoralis major muscle (breast fillet).

One recent and prominent example is woody breast (WB) condition in the pectoralis major muscle in commercial broilers. The emerging issue of WB condition is on a global scale, reported in the United States (Bilgili, 2013), Finland (Sihvo et al., 2014) and many other countries. This emergent myopathy is characterized by degenerating muscle fibers, immune cell infiltration, necrosis, and severe fibrosis. In some cases, WB is seen together with another inflammatory myopathy, called white striping (WS), which presents as white stripes across the surface of the breast fillet. Previous studies have characterized other myopathies that resemble WB and WS, such as deep pectoral myopathy, and nutritional myopathy. These studies demonstrated how these myopathies impair the integrity of the pectoralis muscle as well. Deep pectoral myopathy is associated with a condition developing in the deep pectoral muscles – the supracoracoideus muscle or pectoralis minor muscle, in fast-growing chickens and turkeys. The location of these muscles surrounded by the sternum and tough fascia does not provide enough room for the muscles to

expand fully. The muscles induced by exercise get inflated leading to restraint of the blood supply and subsequently to ischemia (Jordan & Pattison, 1998). Deep pectoral myopathy exerts some common microscopic damage on the pectoral muscles (Wight & Siller, 1980) that are similar to those observed in the WB condition.

Another category of myopathies affecting the pectoralis muscle, WS, is characterized by the observation of white lines seen in the same direction as the muscle fibers accompanying the hardness and out-bulging of the breast and thigh muscles (Kuttappan et al., 2013). Histological studies conducted with the WB and the WS condition in chickens have shown that these two myopathies share similar lesions like regenerative alterations, an influx of lymphocytes and macrophages, fibrosis, lipidosis, and necrosis, but fibrosis occurs more severely in the WB condition (Trocino et al., 2015). In many cases, the WB condition is accompanied by the WS condition, and they show similar histopathological lesions, therefore, the WB and the WS conditions together could exhibit a disease spectrum, with the WB condition constituting a more severe mode of the myopathy. However, the etiology and factors initiating the onset of the WS and WB conditions are not clear.

Nutritional myopathy is mainly correlated with a diet that is deficient for selenium and vitamin E (Guetchom et al., 2012). Vitamin E, together with different enzymes including the selenium dependent-glutathione peroxidase, provides protection against the anticipated damages stemmed from free radicals that are leading to degenerative myopathy as well as contributing to many diseases progress (Herrera & Barbas, 2001). Although some typical features of the WS condition and nutritional myopathy are found in in the WB condition, the WB condition is

distinguished by abnormal hardness and rigidity upon palpation, and in contrast to the nutritional myopathy, it does not impair other muscles.

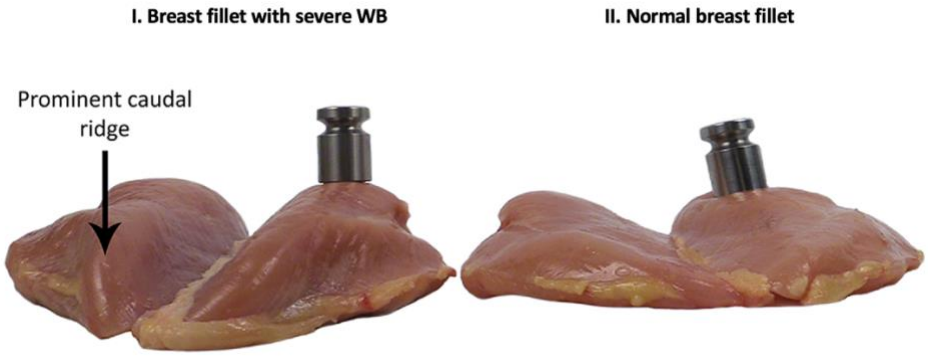


Figure 1. Comparison of fillets with the woody breast (WB) myopathy (I) and normal fillets (II) (Reprinted from Kuttappan et al., 2016). There is no visual effect of compression on the severe WB fillets while the weight compresses the surface of the normal fillet.

According to a Wall Street Journal report, approximately 5-10 % of commercially produced boneless, skinless breast meat shows WB myopathy (Gee, 2016). The occurrence of the WB condition is associated with high growth rate and increased meat muscle yield (Mutryn et al., 2015) in commercial broilers. Prevalence of WB is likely to increase with age and weight of the chicken, with high nutrition diets, and within the male population. The alteration in the muscle causes a lower level of muscle protein, a higher amount of connective tissue and fat content as well as an enhanced water holding capacity (Soglia et al., 2016). Even though the WB myopathy does not constitute a food safety risk, consumers do not prefer to purchase meats with WB because

affected meats do not appeal to taste and have an unsightly visual appearance that is considered unhealthy.

Although the fast growth rate of broiler chickens is associated with WB myopathy, the etiology of WB myopathy remains unclear. The occurrence of the disease is likely affected by many factors (Kuttappan et al., 2016). The basis of WB myopathy is not known yet but is suspected to be genetic due to its association with growth rate. Development of the WB myopathy is tightly linked to growth rate, which is under intense selection. The incidence of WB myopathy in commercial broiler populations is frequently over 50% and as high as 89% (Cruz et al., 2017), which is observed in dominant Mendelian traits. However, dominant inheritance of a disease trait is not a common phenomenon, and furthermore, the four-way crosses that are used in modern broiler production (Bondoc, 2008) make it difficult to answer this question. On the other hand, growth rate and body mass are complex traits, and WB myopathy is associated with growth rate, which suggests that WB myopathy is more likely a complex trait. A recent study by Pampouille et al. (2018, BMC Genomics) reported the identification of QTL for white striping, and suggest that WB and WS are polygenic. Therefore, the genetic basis of WB is one of the leading hypotheses to explain this condition. Differential gene expression analysis has shown that alteration in muscle fiber types, oxidative stress, a higher level of intracellular calcium and localized hypoxia are important indicators of the WB myopathy (Mutryn et al., 2015).

1.3 Background on the Chicken Genome

Recent molecular biology techniques including bacterial artificial chromosomes (BACs) physical mapping, fluorescence in situ hybridization (FISH), fosmid, plasmid paired-end reads,

and radiation hybrid (RH) mapping enhance our comprehension of the avian genome. Analysis of large numbers of partial sequences through the whole genome sequence has dramatically increased our understanding and knowledge of chicken developmental biology and the genetic basis of disease mechanisms over the last decade. The first chicken draft genome sequenced from a female inbred Jungle Fowl (*Gallus gallus*), the ancestor of domesticated chickens, was assembled and released in 2004 utilizing a genome-wide sequencing strategy that provided a high-quality assembly. The assembled chicken genome sequence (1.2 Gb in size) is relatively smaller in size than a typical mammal genome, like humans (Hillier et al., 2004). A further comparison between the chicken and mammalian genome indicates the model bird genome has a relatively low repetitive DNA content that facilitated the more accurate final assembly.

Genome-wide sequencing in chicken has contributed to the understanding of vertebrate evolution because birds shared a common ancestor with mammals ~310 million years ago (Burt, 2005). Sequencing of the chicken genome plays a crucial role in facilitating agricultural studies, especially in poultry science with relation to breeding tools, animal health studies, identification of mammalian and non-mammalian vertebrate diseases, and developing stages of the vertebrate.

In the chicken genome, there are approximately 1200 million base pairs (Mbp) of DNA on their single set of chromosomes. Chicken has a diploid genome with 38 autosomes including 28 micro-chromosomes, five intermediate chromosomes, and five macro-chromosomes, and 1 pair of sex chromosomes denoted by Z and W with females being heterogametic (ZW) and males being homogametic (ZZ). The micro-chromosomes have been shown to be gene rich but have also been more challenging to characterize compared to the macro-chromosomes. Micro-chromosomes contain 26% higher identical substitutions in coding DNA regions and 18% higher DNA sequence

differences in non-coding regions compared to macro-chromosomes (Meunier & Duret, 2004). A mutation mechanism stimulated by recombination and biased gene conversion in favor of one allele over another may explain this discrepancy in mutation rates.

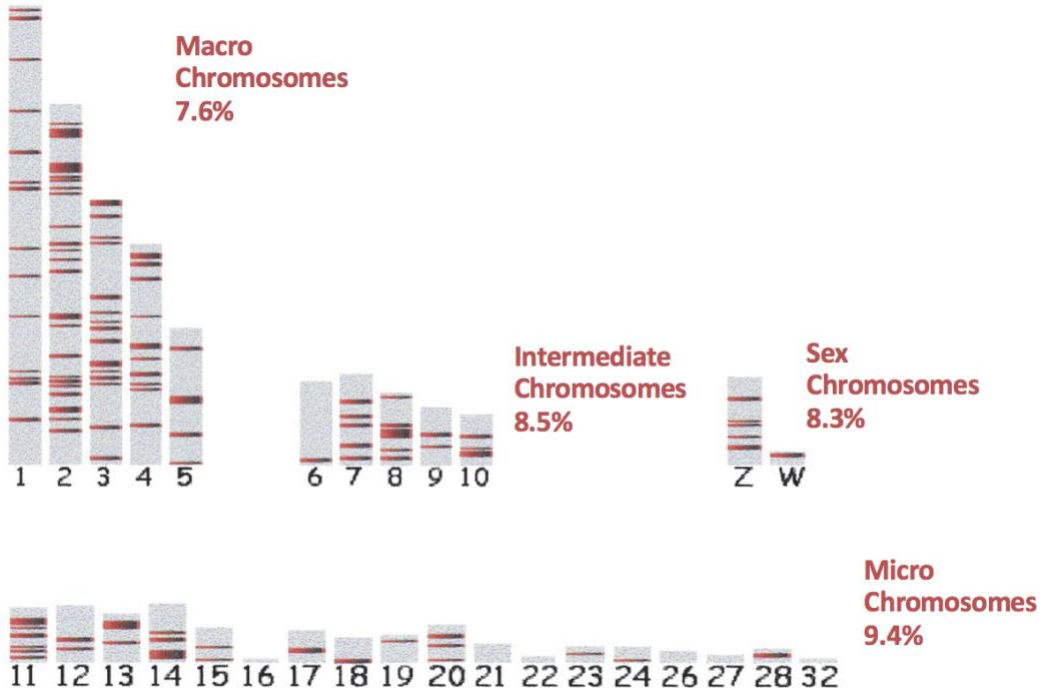


Figure 2**.** Chicken chromosomes. They are categorized into macro, intermediate, micro, and sex chromosomes with the gene desert regions that are lack of protein-coding genes (plotted as red lines) (Reprinted from Ovcharenko et al., 2005).

The relatively small size of the chicken genome - approximately 45% size of the mouse genome and 39 % of the human genome- can be attributed to the paucity of repetitive DNA elements. The amount of repetitive DNA is approximately 15% (Schmid et al., 2000). Along with the reduced number of repetitive regions, there are explanatory factors for the discrepancy in

genome sizes like reduced number of gene duplications, pseudogenes, and segmental duplications (Hillier et al., 2004), which are responsible for around 25% of the differences. The frequency of repetitive regions in chicken is very low in comparison to those in the human genome but these repeats complicate the assembly of the genome since being arranged as tandem repeats (Wicker et al., 2004). A study based on the comparison between orthologous sequences of chicken and turkey indicates that chromosomes varying in size are exposed to different evolutionary pressure (Axelsson et al., 2005).

1.4 Genome-wide Association Study

Genome-wide association studies (GWAS) have become more feasible with recent improvements in high-throughput genotyping technologies. Investigating associations across the whole genome is a powerful approach to examine the genetic architecture of diseases and underlying genetic mechanisms of quantitative traits. The basis of GWAS is the common disease, common variant approach. According to this approach, the effects of genetic differences on many common diseases will be at least partly attributed to a limited number of allelic variants occurring in more than 1-5% of the population (Collins et al., 1997). Many significant disease-related variants would be less frequent than this and, thus, the likelihood of determination of these variants is low with this approach. That is why exploiting sufficient density of markers is essential for detection of the majority of the trait or disease-related variants. On the other hand, the assessment of a sufficient number of individuals' genomes is essential to identify variations with modest effect on the phenotype. These two factors mainly affect the power of GWAS.

In GWAS, hundreds of thousands of genetic variants are independently tested for statistical association with a disease or a complex trait whose occurrence depends on environmental and genetic factors with small contributions to the related phenotype. The power to deal with a complex disease by analyzing the interaction among genetic variants makes it an ideal tool to enlighten mechanisms of polygenic traits or diseases (Hardy and Singleton, 2009). Preferably, single nucleotide polymorphisms (SNP) are used as genetic variants due to their abundance and distribution throughout the genome, and genome-wide high-density SNP arrays are often available for GWAS. A genetic variant or SNP that is detected as associated with a disease or located in the same haplotype as the associated SNP should have a higher frequency in the affected group compared to the control group. Since a haplotype comprises SNPs that are inherited together, other SNPs sharing the same haplotype with a disease-associated SNP can also be related to the disease. Likewise, GWAS are also conducted based on haplotypes as well as independent SNPs that enhances the robustness of association studies (Calus et al., 2009).

There are many genome-wide high-density SNP arrays available for many organisms. For chicken, the 600 K HD Affymetrix Axiom chicken array (Affymetrix, Inc., Santa Clara, CA) was developed by sequencing 24 different chicken breeds consisting of egg laying breeds and broiler breeds exploited for both research and commercial purposes. Identification of single nucleotide polymorphisms (SNP) was achieved by the alignment of the reads to *Gallus gallus* version 4.0 genome assembly. In the final array, SNPs are of high-quality score and evenly distributed regarding genetic distance among different breeds. Approximately two-thirds of SNPs show polymorphism with more than 98% high call rates. Other criteria taken into account while selecting

SNPs for the array were stable Mendelian inheritance and reasonable deviation from Hardy-Weinberg equilibrium (Kranis et al., 2013).

Table 1**.** Summary of annotation of SNPs in the 600 K panel to predict the genomic effect (Reprinted from Kranis et al., 2013).

	Count	Percent
Total Number of SNPs in the Panel	580,954	
Annotation Possible	492,572	84.79
Annotation Result		
Intergenic	266,636	54.13
Intronic	189,128	38.40
Exonic		
Non-Synonymous	9,345	1.90
Synonymous	12,069	2..45
Stopgain/Stoploss	120	0.02
1 Kb Upstream	5,892	1.20
1 Kb Downstream	6,456	1.31
UTR 3	2,497	0.51
UTR 5	302	0.06
Splicing	83	0.02
Non-Coding RNA (ncRNA)	44	0.01

Power calculations of GWAS were considered among genetic models, tag SNP selection, and the required population to select the proper sample size and genotyping platform. The power of a study depends on sample size along with the selected tag SNPs. The tagged and genotyped SNPs can be used to predict the power of GWAS (Klein, 2007). According to these parameters, a sample size of twenty (10 each of cases and controls) genotyped with the 600 K Affymetrix array has approximately 30% power to identify trait associated loci. Due to the limited understanding of the genetic mechanisms underlying WB, this study with small sample sizes will serve as an exploratory analysis to identify genes or genomic regions that can be a risk factor for the commercial broiler chickens to develop woody breast myopathy by using individual SNPs and haplotypes exhibiting statistical association.

2. MATERIALS AND METHODS

2.1 Sample Collection

Ten commercial broiler chickens (Ross 708 line) affected with the woody breast myopathy were selected as a case group, and ten White Plymouth Rocks (a heritage chicken breed) were chosen as the negative control group. Affected broiler chickens were diagnosed by manual palpation of the breast muscle while they were alive. There were 14 males and 6 females (all in the control group) in the sample The White Plymouth Rock (WPR) breed was selected as it is the progenitor breeds of modern broilers (Cornish x Rock crosses), and it is also one of the most closely related breeds to this commercial line that does not exhibit woody breast. Therefore, the White Plymouth Rock was expected to function as a control group in this analysis. All the samples for this study were collected as part of a study focusing on woody breast incidence in commercial broilers (IACUC 2016-0065). This project focuses on the generation and analysis of genomic data from DNA samples that were made available from this study.

2.2 DNA Isolation

Blood was sampled from all individuals after euthanasia at 42 days of age by puncturing the brachial vein and stored in Longmire Buffer until further processing. Genomic DNA was isolated from whole blood using the DNeasy Blood and Tissue kit (Qiagen Inc, Valencia, CA) using standard protocols.

2.3 Genotyping

Equimolar DNA isolates (10 cases and 10 controls) were submitted to the molecular genetics lab at Geneseek (Lincoln, NE) for genotyping on the 600K Affymetrix High-Density Genotyping array (Kranis et al., 2013).

2.4 SNP Analysis

2.4.1 Quality Control

Before statistical analysis, evaluation of the data quality and pre-processing of the data are essential to avoid introducing bias into the analysis (Anderson et al., 2010). Raw SNP data first were subjected to quality control by removing loci without a known chromosomal location. Further filtering of the SNP data was conducted using PLINK v1.9 (http://pngu.mgh.harvard.edu/purcell/plink/). SNP variants having less than 0.05 minor allele frequency and more than 0.1 missing genotype frequency and showing a significant deviation from Hardy-Weinberg equilibrium (p-value < 0.001) were removed.

2.4.2 SNP Association

Following quality control steps, the remaining SNPs were analyzed to detect loci that were associated with the woody breast myopathy as the case status. Basic case-control association analysis was implemented with PLINK v1.9 (Purcell et al., 2007), giving asymptotic p-values according to the basic allelic χ^2 test. Adjusted p-values for each SNP also were obtained as an extension of the basic case-control analysis. The false discovery rate (FDR) was controlled at a level of 0.05 per the step-up method of Benjamini and Hochberg (1995) becouse it adjusts p-values

by compensating the elevation in the likelihood of incorrectly rejecting a null hypothesis which is called type I error. The association test was repeated with permutation, and empirical p-values of each SNP were also obtained. To visualize the significant SNPs identified from association tests, Manhattan plots were generated based on both FDR-BH adjusted p-values and empirical p-values. The results were plotting using CMplot R package (Yin LiLin, 2018; https://CRAN.R-project.org/package=CMplot).

2.4.3 Analysis of Significant SNPs

SNPs that were found to be associated with the disease based on statistical significance were classified functionally according to genic locations, including 5' untranslated region (5' UTR), introns, and exons. The effects of variants on the coding sequence regions, like synonymous, missense, and nonsynonymous changes were queried using the Ensembl Variant Effect Predictor (McLaren et al., 2016). To be able to incorporate all SNPs into the analysis regardless of whether they were assigned a reference SNP (rs) number or not, the chromosomal coordinates of SNPs were used as an identifier instead of rs numbers. This analysis conducted by the Ensembl Variant Effect Predictor has also provided genes that harboring the SNPs found significant at 1.0e-6 level according to the FDR adjusted association test, whose chromosomal coordinates were used as input. Among 268 significant SNPs, 18 SNPs are located on unplaced contigs; therefore, 250 SNPs were included in the analysis.

The DNA sequences of the relevant genes to the significant SNPs were retrieved from Ensemble to exploit them for the annotation analysis using the Blast2GO software (Conesa et al.,

2005). The input file of sequences first was analyzed with blast function against NCBI database to obtain the homologous sequences, and then InterPro and mapping functions were used to retrieve relevant gene ontology (GO) terms. The results were used for annotation step to select reliable functions for query sequences. The annotation result was compiled with GOSlim function. This Blast2GO analysis was performed by the genes obtained from the Ensembl Variant Effect Predictor.

The significant SNPs identified by association testing with permutation were further investigated through the use of Ingenuity Pathway Analysis (IPA, QIAGEN Inc., Redwood, CA). The number of significant SNPs below the genome-wide threshold of 1.0e-6 is 10,742. However, a more stringent threshold of 1.79e-9 was used to narrow down the significant SNPs, which eliminates false positive results to a certain extent. At 1.79e-9 significance level, 762 SNPs were obtained, but of these only 336 SNPs included pathway analysis because the rest of the SNPs have not yet been assigned rs numbers. Analysis of IPA compiles the information regarding correlations among proteins and genes from prior research and articles. It does not solely focus on associations, but instead investigates the consequences of the variants on biological pathways and functions. The logic behind IPA is causal analysis approaches that identify upstream regulators related to the input data and investigate any connection among them that play a role in the same causal mechanism as well as the analysis of downstream effects (Krämer et al., 2014). By the use of IPA, it is aimed to determine the most related pathways to the genes in which the significant SNPs are located.

2.5 Construction of Haplotype Blocks

To identify the combination of SNPs that were located on the same chromosome and that are inherited together from a single parent, construction of haplotype blocks was carried out in PLINK v1.9, and pairwise linkage disequilibrium (LD) was calculated for SNPs within 200 kb. While estimating haplotype blocks, PLINK uses an accelerated Expectation Maximization (EM) algorithm. To perform this analysis, all SNP loci with missing genotypes were discarded to minimize the computational burden and error rate that stemmed from missing genotypes.

2.6 Haplotype Association

A basic case-control χ^2 test was conducted with haplotype blocks to identify haplotypes with a statistically significant association to the WB myopathy. Haplotypes were visualized as an LD heatmap using the r^2 metric for pairs of SNPs with the R package "snp.plotter" (Luna & Nicodemus, 2007). For only coding sequence variants, haplotypes and LD heatmaps were plotted within 50 kb window size with the relevant genes.

2.7 Population Stratification Analysis

Population stratification analysis was based on identity by state (IBS) that is calculated by the proportion of identical alleles that two individuals share. First, for the 20 birds, a 20 x 20 matrix was generated to obtain pairwise IBS distances of the sample based on the autosomal genome-wide SNP data by the use of Plink 1.9. IBS ranges from 1 to 0; 1 indicates that two individuals are completely identical (a duplication of sample or maternal twin); 0 indicates two completely

genetically different individuals with respect to the dataset. The multidimensional scaling (MDS) plot was generated based on the IBS distances (1-IBS) between individuals.

In stratified analyses by Plink 1.9, the Cochran Mantel-Haenszel (CMH) χ^2 test was applied to eliminate false positive associations and increase the power of the association analysis. CMH tests whether a SNP is associated with a disease conditional on population stratification (Cochran, 1954; Mantel & Haenszel, 1959). SNPs were examined though the CMH test by using 2x2xK table in which K represents two clusters (WPR and broiler) and IxJxK table in which the roles of clusters and phenotypes (cases and controls) are switched.

As an alternative way to control population stratification, logistic regression was performed by using clusters and sex status as covariates with Plink 1.9. By this approach, another issue regarding the sample was being addressed. Our sample consists of 6 females in the control group, and differences stemmed from sex are needed to be controlled. Logistic regression analysis allows for different covariates while investigating whether a SNP is associated with a disease trait. Along with the cluster numbers (WPR=0, broiler=1), the MDS components calculated to generate MDS plot were used as covariates in two logistic regression analyses separately. Each analysis was performed with and without the sex covariate.

3. RESULTS

3. 1 Pre-processing and Quality Control Results

Of the 580,961 SNPs genotyped by 600K Affymetrix array, 450 SNPs were discarded due to being assigned to unmapped contigs, and 62 duplicate SNPs were deleted from the data. Four regions assigned to unmapped contigs were retained as they harbored a considerable number of SNPs, which were given chromosome identifiers (ID) 42, 43, 44 and 45. In the data set, approximately half of the SNPs have not yet been issued with a dbSNP number (281,354). Those without dbSNP numbers were given identification values comprising the chromosome number and the location of SNP of interest respectively with the addition of "Nov" at the beginning, which refers to “novel”.

The retained 580,449 SNPs were subject to quality control. There were 6,479 SNPs removed due to having more than 10% genotype data missing, and 101,217 SNPs could not pass the filter that requires more than 0.05 minor allele frequency (Anderson et al., 2010). No SNPs showed a severe violation of Hardy Weinberg equilibrium test at 1.0e-3 level. Finally, 472,753 SNPs for the total data set (20 individual chickens) passed filters and quality control, with 0.99 total genotyping rate, and were used for the further analysis.

The estimated average heterozygosity for the total dataset was 29.45. When separated by breed, the WPR had average heterozygosity of 26.48 with standard deviation equal to 1.28, and the Ross 708 line had average heterozygosity of 32.43 with standard deviation of 0.28 (Figure 3). Analysis of SNP density (using 1 Mb window sizes) showed a higher density of SNPs with decreasing chromosome size (Figure 4).

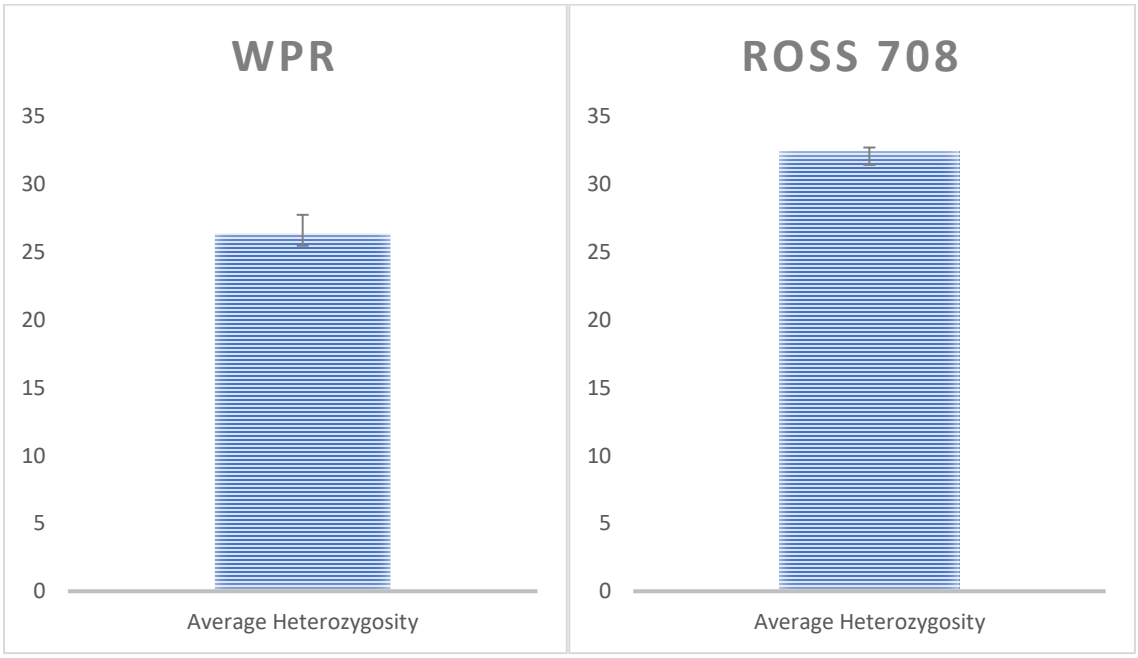


Figure 3. Average heterozygosity rates observed in WPR and Ross 708

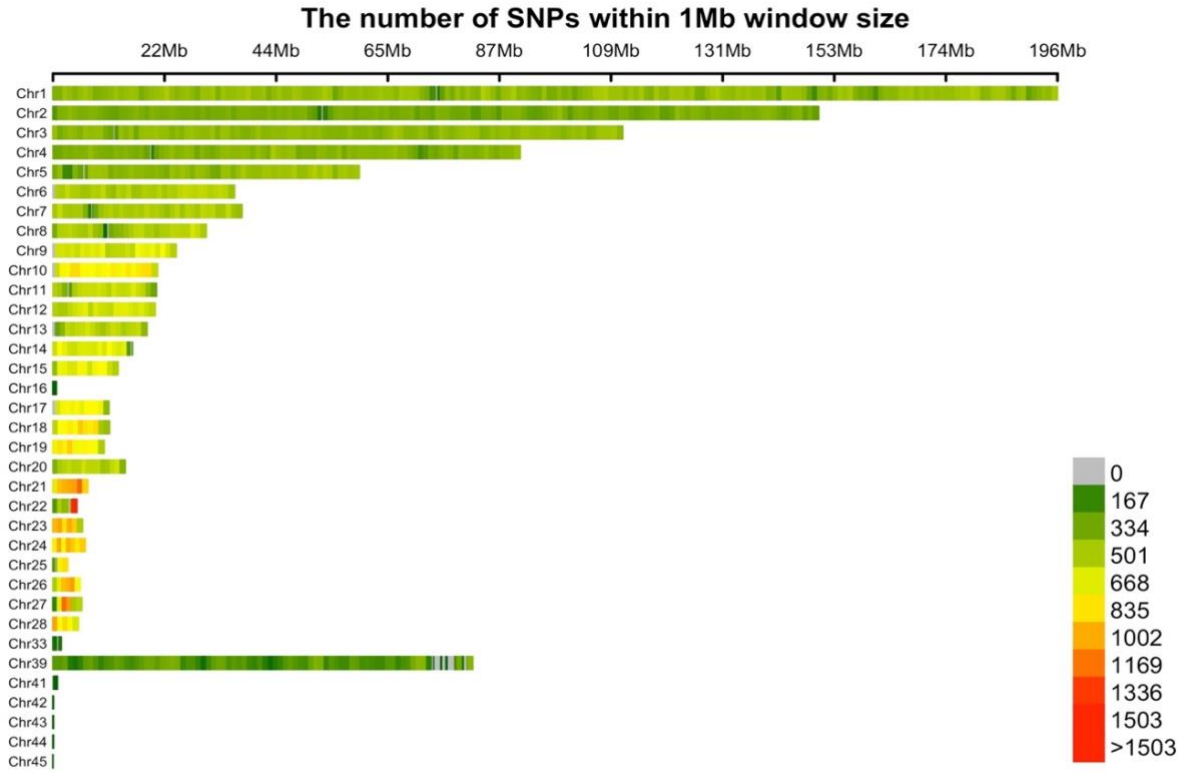


Figure 4. SNP density (n = 472,753 that passed quality editing) across the chicken genome. The legend represents the number of SNPs within one megabase (Mb) windows. The dark red regions on the chromosomes indicate the highest SNP density that is more than 1,503 SNPs located within 1 Mb window. The gray regions that are mainly located on chromosome 39 (Z) indicated the absence of SNP.

3.2 Genome-Wide Association Study Results

Basic allelic association tests were conducted with and without permutation. The association test with adaptive permutation resulted in 10,742 significant SNPs with empirical P-values under the genome-wide threshold of 1.0e-6. The adaptive permutation test was used to improve the precision of the estimate of the significant SNPs. Of those significant SNP, 143 were located on the unplaced contigs, 2,053 SNPs on the sex chromosome Z (SNPs on the chromosome W were removed during quality control process), and 8,546 SNPs on the autosomes. On the other hand, the association test without permutation was used to generate adjusted P-values to provide better control for false positive associations. In this case, we used the FDR-BH adjusted P-values, which revealed 268 significant SNPs at 1.0e-6 genome-wide significance level. Among 268 significant SNPs, none were detected on the sex chromosome Z. There were 250 significant SNPs are located on the autosomes, but 18 SNP with detected association were on the unplaced contigs. These 268 SNPs obtained from the FDR-BH adjusted association test were a full subset of the 10,742 significant SNPs identified through the permutation-based empirical P-value filtering.

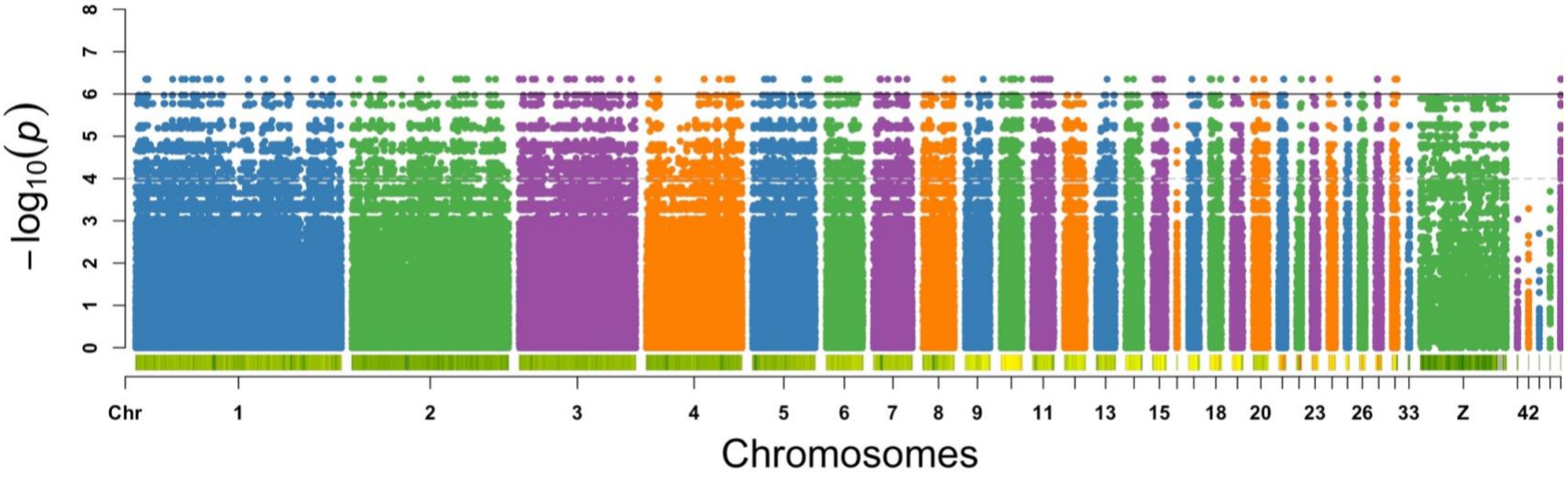


Figure 5. Manhattan plot from the GWAS for the WB myopathy with FDR-BH adjusted P-values. 1.0e-6 is considered as a significant P-value threshold that is indicated by the black horizontal line. A SNP density plot is incorporated into the plot underneath the SNPs with the color scale legend.

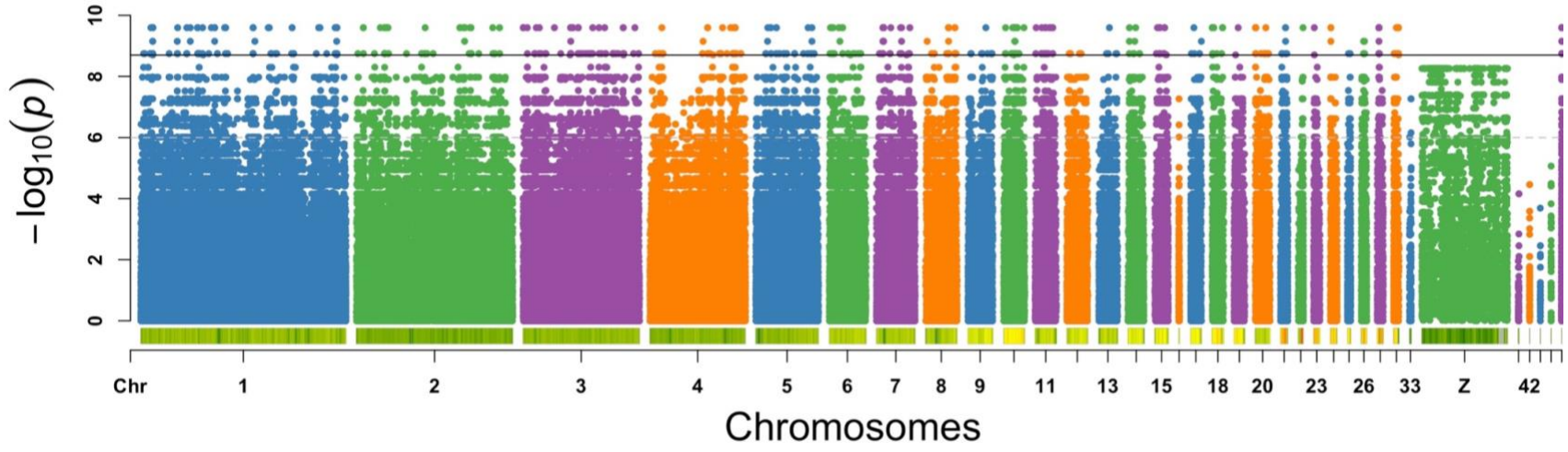


Figure 6. Manhattan plot from the GWAS for the WB myopathy with empirical P-values. The gray dotted line is the genome-wide P-value threshold, and the black line is more stringent P-value threshold, which is set to 1.0e-6 and 1.79e-9 respectively.

A Manhattan plot is a scatter plot, which facilitates the representation of the SNPs across the whole genome chromosome by chromosome. Lower P-values of tested loci are associated with stronger statistical significance of association tests because the P-value is the likelihood of the surveyed SNP distribution by chance. The use of -log10 P-values means that significant association tests are observed as high values in the Manhattan plot.

The first Manhattan plot (Figure 5) shows the FDR-BH adjusted P-values of SNPs obtained from the association test. A total of 268 SNPs had P-values smaller than 1.0e-6. In the second Manhattan plot (Figure 6) with the empirical P-values, a genome-wide threshold of 1.79e-9 (solid black line) was chosen to increase the stringency of the significant SNPs identified based on empirical P-values. In total, 762 significant SNPs were detected below the threshold. These 762 also comprise all SNPs (268) found to be significant from the FDR-BH adjusted association test, which is depicted in Figure 6 above the P-value threshold.

Overall, many significant SNPs above the threshold were detected along the whole length of all chicken autosomes except LGE64 and three unmapped regions indicated as chromosome number 42, 43, 44 in the Manhattan plots. The significant SNP show an equal distribution pattern along all chromosomes that could be an indicator of confounding factors stemming from differences in DNA sampling and genotyping between case and control groups, that is, from population structure.

A quantile-quantile (Q-Q) plot of $-\log_{10}$ P values was generated to compare association test results with the theoretical distribution of χ^2 tests of no association.

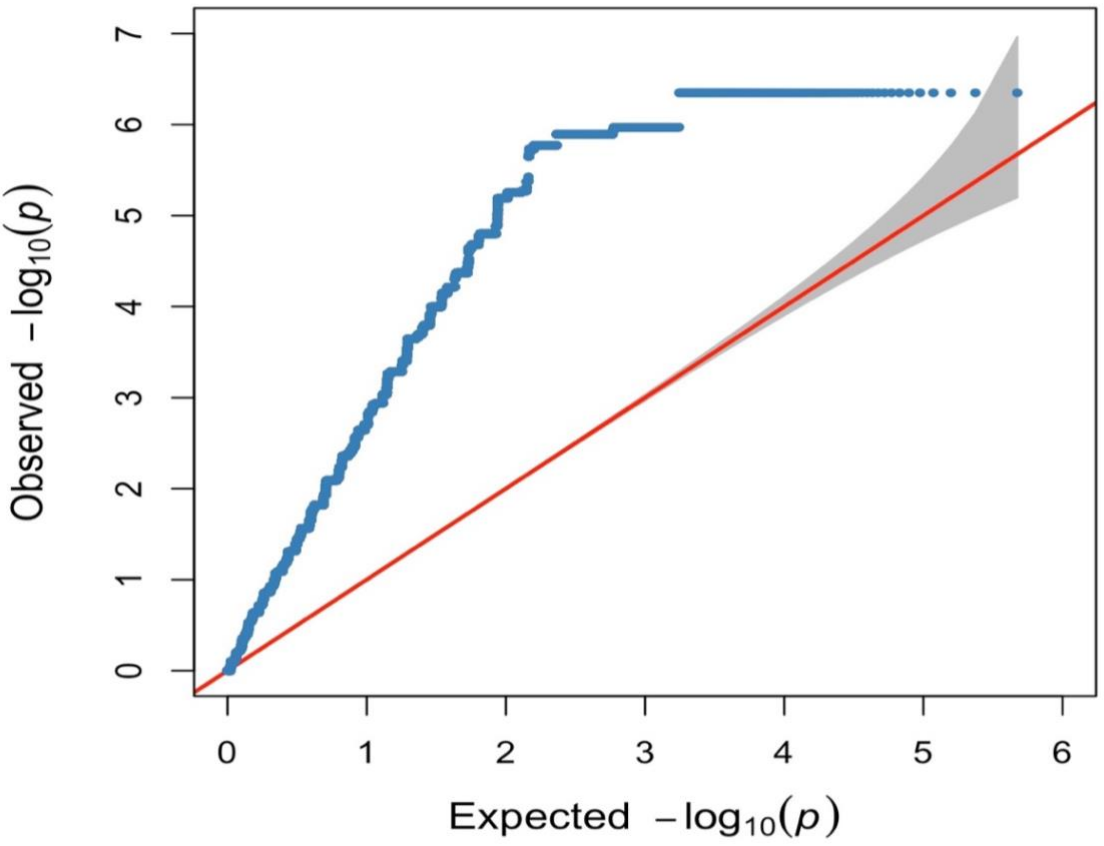


Figure 7. The Q-Q plot of WB association test P-values

The generation of the Q-Q plot in Figure 7 was based on two steps, observed P-values of the thousands association test are sorted into ascending order and then the distribution of the observed values is plotted against the known χ^2 distribution of the expected values. Deviations from the diagonal line can be an indicator of a true association of SNPs with the WB myopathy, however, reaching out a final decision regarding the detection of true associations is not straightforward. Genomic inflation factor (λgc based on median χ^2) = 7.11238, which allows for quantification of the excess of false positive association in our analysis. The λgc value is much bigger than the expected value of 1 and indicates an enormous systematic bias.

3.3 Results of Significant SNPs

Out of 472,753 SNPs, 268 were detected as a result of the FDR adjusted association test, but 18 were discarded due to absence of chromosomal coordinate information. The results of the analysis conducted with the 250 significant SNPs by the ENSEMBL Variant Effect Predictor are summarized in Figure 8 and Table 2. This tool permits investigation genomic location and kind of variation by using dbSNP number or chromosomal coordinates as identifiers. In this study, chromosomal coordinates of the SNPs were used. As indicated in the Figure 8, only 2% of the significant SNPs are located on the exons of the genes, which is corresponding to five SNPs; Nov1_11278152, Nov10_10081846, Nov10_19760216, Nov20_497735, rs315521303. Table 2 displays genes having those 5 coding sequence variants, gene descriptions, and details about the SNPs. The allele column refers to the alternate alleles corresponding coding sequence variants. Only one SNP is a nonsynonymous variant that leads to a change in amino acid from isoleucine to valine.

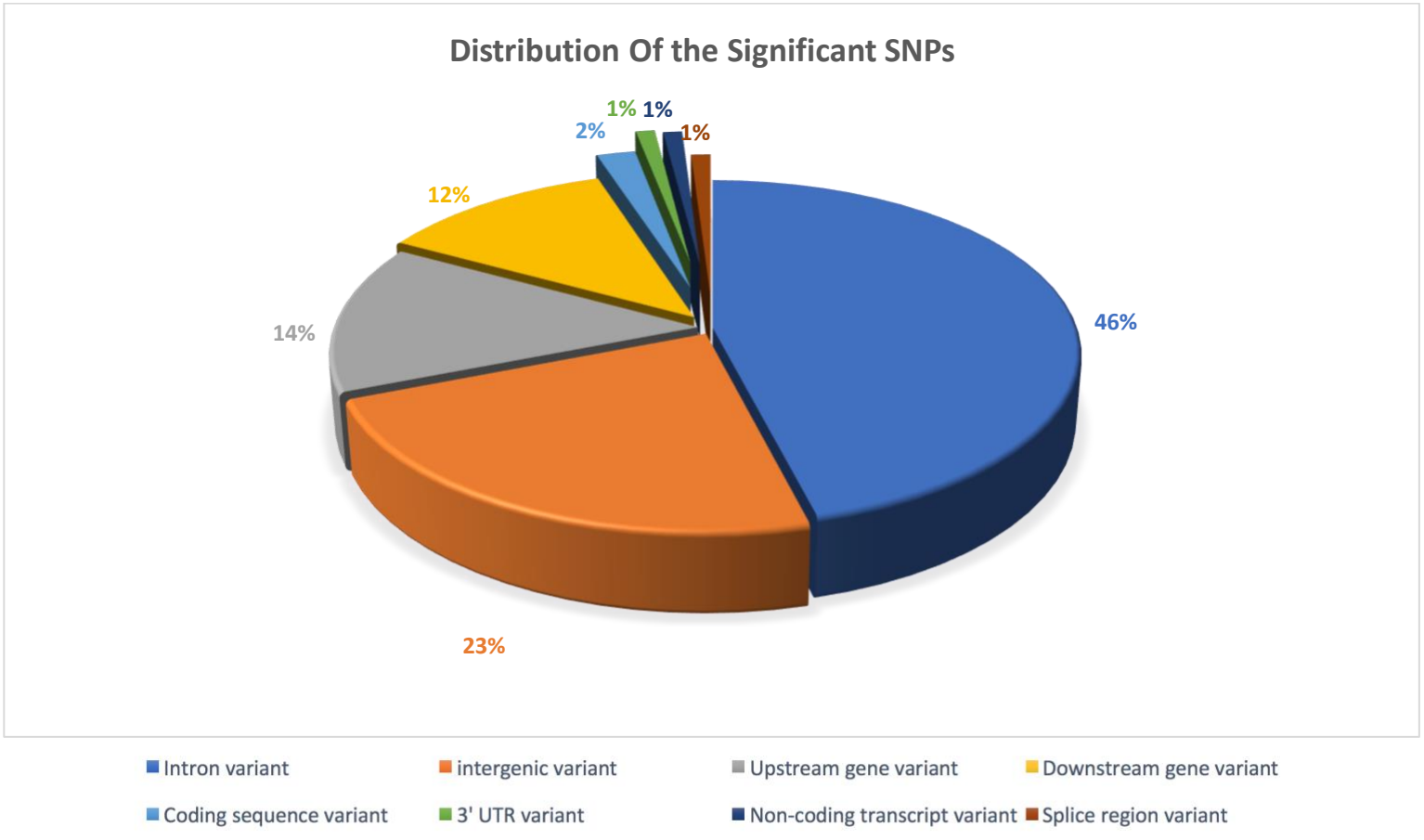


Figure 8. Distribution of the significant SNPs according to the consequence of the variants

Table 2. SNPs categorized as coding sequence variant with their corresponding genes.

Gene Symbol	Description	Location SNP ID	Allele	Ensemble ID	Consequence	Effect
GNAT3	Guanine nucleotide-binding protein G(t) subunit alpha-4	1:11278152 Nov1_11278152	A	ENSGALG00000008427	Coding sequence variant	Synonymous variant
SLC12A1	Solute carrier family 12 member 1	10:10081846 Nov10_10081846	C	ENSGALG00000004945	Coding sequence variant	Synonymous variant
SPG11	Spatacsin	10:19760216 Nov10_19760216	T	ENSGALG00000008180	Coding sequence variant	Missense variant
C20H20ORF24	Uncharacterized protein	20:497735 Nov20_497735	C	ENSGALG00000001054	Coding sequence variant	Synonymous variant
MYO1D	Unconventional myosin-Id	27:1606640 rs315521303	G	ENSGALG00000000674	Coding sequence variant	Synonymous variant

Table 3. Genotypes of the most significant SNPs located on the GNAT3, SLC12A, SPG11, C20H20ORF24 and MYO1D genes

SNP ID	GENE	W	W	W	W	W	W	W	W	W	W	R	R	R	R	R	R	R	R	R	R
Nov1_11273473	GNAT3	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
Nov1_11278152	GNAT3	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
Nov1_11283000	GNAT3	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
Nov1_11293049	GNAT3	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
Nov10_10081846	SLC12A1	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
Nov10_19760216	SPG11	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
Nov20_497735	C20H20ORF24	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs315880270	MYO1D	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
rs316269099	MYO1D	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
rs317270020	MYO1D	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
rs316887638	MYO1D	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs315521303	MYO1D	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
rs315181886	MYO1D	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG
rs314212462	MYO1D	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC
rs16206894	MYO1D	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT
rs313313466	MYO1D	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG

Table 3 shows the opposite homozygosity of Ross 708 (R) and WPR (W) individuals for the most significant SNP of this work. The genotypes of the SNPs indicated above were retrieved from the raw SNP data. The other most significant SNPs located on the same genes with the coding sequence variants are included in this table. A number of significant SNPs are positioned on MYO1D gene and GNAT3 gene with completely different genotypes between cases and controls.

3.3.1 GO Annotation

Annotation of the genes associated with significant SNP were indicated in Table 4 with corresponding Ensembl IDs of the genes, the significant SNPs and their positions within the genes. Many SNPs have not yet been assigned dbSNP identifiers, and therefore could not be used for Ingenuity Pathway Analysis (IPA) analysis. The objective of the GO analysis was to expand the number of candidate genes with potential responsibility for WB myopathy, or given systematic bias identified, the differences between the two lines of chickens. Among the annotated genes, the prevalence of the genes exerting oxidoreductase activity, related to Ca^{2+} binding, cellular repair, and fiber type switching is particularly noteworthy. These chicken genes investigated by the GO analysis could contribute the onset of the WB myopathy.

Table 4. Annotation of the genes harboring significant SNPs

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
AACS	Catalytic activity Acetoacetate-CoA ligase activity	Lipid metabolic process Metabolic process		Acetoacetyl-CoA synthetase	Nov15_4474884	ENSGALG00000002899	Downstream
ACOX3	Acyl-CoA oxidase activity Oxidoreductase activity acting on the CH-CH group of donors FAD binding	Fatty acid metabolic process Fatty acid beta-oxidation Oxidation-reduction process	Peroxisome	Peroxisomal acyl-coenzyme A oxidase 3 isoform X1	rs317697112	ENSGALG00000015591	Intron
ACTR6		Chromatin remodeling	Nucleus	Actin-related protein 6	Nov1_47207134	ENSGALG00000040976	Intron
AIG1				Androgen-induced gene 1 protein isoform X1	Nov3_52780143	ENSGALG00000013780	Upstream
ALYREF	Nucleic acid binding			THO complex subunit 4	Nov18_10028916	ENSGALG00000007237	Downstream
AMN				Protein amnionless	rs316881096	ENSGALG00000011391	Upstream
ANGPT4	Receptor tyrosine kinase binding	Angiogenesis		Angiopoietin-4	Nov20_10103110	ENSGALG00000040755	Downstream
AP1S2		Protein transport		AP-1 complex subunit sigma-2 isoform X1	Nov1_121878877	ENSGALG00000016551	Intron
ARHGAP31		Signal transduction		Rho GTPase-activating protein 31	Nov1_80732748	ENSGALG00000015077	Intron
ARHGDIA	Rho GDP-dissociation inhibitor activity		Cytoplasm	Rho GDP-dissociation inhibitor 1	Nov18_10028916	ENSGALG00000008443	Intron
ATPAF1		Protein-containing complex assembly	Mitochondrion	ATP synthase mitochondrial F1 complex assembly factor 1	Nov8_21919591	ENSGALG00000010444	Intron
BRI3BP			Mitochondrion	BRI3-binding protein	Nov15_4514099	ENSGALG00000002913	Upstream
BRSK2	Protein kinase activity ATP binding	Protein phosphorylation		Serine/threonine-protein kinase BRSK2 isoform X4	rs316707833	ENSGALG00000006681	Intron
C12ORF43				Predicted: Uncharacterized protein C12orf43 homolog	Nov15_9145634	ENSGALG00000006965	Intron
C1H11ORF54			Nucleus	Ester hydrolase C11orf54 homolog isoform X1	Nov1_185874346	ENSGALG00000017220	Downstream
C20H20ORF24				Uncharacterized protein C20orf24 homolog	Nov20_497735	ENSGALG00000001054	Coding sequence

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
CAPZB		Barbed-end actin filament capping	F-actin capping protein complex	F-actin-capping protein subunit beta isoform X1	rs313715793	ENSGALG00000004034	Intron
CCDC88A	Actin-binding Protein binding	Cell migration Regulation of actin cytoskeleton organization		Girdin isoform X14	Nov3_97197 Nov3_107605 rs315666869 Nov3_125412 Nov3_128311	ENSGALG00000043587	Intron Intron Intron Intron Downstream
CD276				CD276 antigen	Nov10_2182652	ENSGALG00000001729	Intron
CDK19	Protein kinase activity ATP binding	Protein phosphorylation		Cyclin-dependent kinase 19 isoform X2	Nov3_66637424	ENSGALG00000015051	Intron
CDK6	Protein kinase activity Cyclin-dependent protein serine/threonine kinase activity ATP binding	Protein phosphorylation Cell cycle Regulation of cell proliferation Cell dedifferentiation		Cyclin-dependent kinase 6	Nov2_22815766	ENSGALG00000009476	Intron
CETN2	Calcium ion binding			Centrin-2	rs15494233	ENSGALG00000007510	Intron
CHERP	Nucleic acid binding RNA binding	RNA processing		Calcium homeostasis endoplasmic reticulum protein isoform X1	rs313429355 rs318170209 rs14307755	ENSGALG00000003824	Upstream Intron Intron
COL18A1	Structural molecule activity	Cell adhesion	Extracellular matrix	Collagen alpha-1(XVIII) chain isoform X2	Nov7_6666747	ENSGALG00000038311	Downstream
COL6A1				collagen alpha-1(VI) chain	Nov7_6743219	ENSGALG00000005974	3’ UTR
COQ5	Methyltransferase activity			2-methoxy-6-polyprenyl-1,4-benzoquinol methylase	Nov15_9334468	ENSGALG00000007185	Downstream
CRMP1	Hydrolase activity Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds	Microtubule cytoskeleton organization Nervous system development Axon guidance	Cytoplasm	Dihydropyrimidinase-related protein 1 isoform X1	rs315239587	ENSGALG00000041203	Upstream
CYP4A22	Iron ion binding Oxidoreductase activity, acting on paired donors Heme binding	Oxidation-reduction process		Cytochrome P450 4B1-like	Nov8_21937348	ENSGALG00000020688	Upstream
CYP4B7	Iron ion binding Oxidoreductase activity, acting on paired donors Heme binding	Oxidation-reduction process		Cytochrome P450 4B1-like	Nov8_21950413	ENSGALG00000010469	Upstream

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
DCPS	Hydrolase activity	Deadenylation-dependent decapping of nuclear-transcribed mRNA		M7GpppX diphosphatase	rs314257517 rs314803280	ENSGALG00000031143	Intron Intron
DHRS7C				Dehydrogenase/reductase SDR family member 7C	rs312875825	ENSGALG00000039712	Upstream
DHX37	Nucleic acid binding Helicase activity ATP binding			Probable ATP-dependent RNA helicase DHX37	Nov15_4520607	ENSGALG00000002964	Upstream
DLGAP4		Signaling		Disks large-associated protein 4	Nov4_344630 Nov20_351099	ENSGALG00000001046	Intron Intron
DOCK9	Guanyl-nucleotide exchange factor activity Binding	Small GTPase mediated signal transduction		Dedicator of cytokinesis protein 9 isoform X12	Nov1_144728718 Nov1_144743920	ENSGALG00000016883	Splice region and Intron Intron
DYNLL1		Microtubule-based process	Dynein complex	Dynein light chain 1, cytoplasmic	Nov15_9334468	ENSGALG00000020999	Upstream
EFCAB14				EF-hand calcium-binding domain-containing protein 14 isoform X1	rs316613723	ENSGALG00000010446	Downstream
ERLIN1	Ubiquitin-protein ligase binding	Ubiquitin-dependent ERAD pathway	Endoplasmic reticulum	Erlin-1 isoform X1	rs317432965	ENSGALG00000003308	Downstream
F2RL3	G-protein coupled receptor activity Thrombin-activated receptor activity	G-protein coupled receptor signaling pathway Thrombin-activated receptor signaling pathway Blood coagulation	Integral component of membrane	Proteinase-activated receptor 4	rs14307705	ENSGALG00000044485	Upstream
FAM110A				Protein FAM110A	Nov20_10103110	ENSGALG00000046425	Downstream
FAM96A				MIP18 family protein FAM96A isoform X2	Nov10_19760216	ENSGALG00000026448	Upstream
FBLN1	Calcium ion binding Peptidase activator activity	Extracellular matrix organization	Extracellular region	Fibulin-1 isoform X1	Nov1_70565519 Nov1_70569407 Nov1_70571806 Nov1_70574070	ENSGALG00000014233	Intron Intron Intron Intron
FREM2		Cell communication	Integral component of membrane	FRAS1-related extracellular matrix protein 2	rs314056895 Nov1_185874346	ENSGALG00000033671ENSGALG00000017219	Intron Upstream

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
FTCD	Catalytic activity Folic acid binding Transferase activity	Metabolic process Cellular metabolic process		Formimidoyltransferase-cyclodeaminase isoform X1	Nov7_6811648 Nov7_6814900	ENSGALG00000006131	Intron Upstream
GALNT9				Polypeptide N-acetylgalactosaminyltransferase 9	rs316182209	ENSGALG00000002242	Intron
GCGR	Transmembrane signaling receptor activity G-protein coupled receptor activity Glucagon receptor activity	Cell surface receptor signaling pathway G-protein coupled receptor signaling pathway G-protein coupled receptor signaling pathway	Membrane Integral component of membrane	Glucagon receptor	Nov18_10005222	ENSGALG00000011219	Downstream
GFAP	Structural molecule activity		Intermediate filament	Glial fibrillary acidic protein	rs316143928	ENSGALG00000000909	Intron
GNAT3	GTPase activity GTP binding Guanyl nucleotide binding G-protein beta/gamma-subunit complex binding	Signal transduction G-protein coupled receptor signaling pathway Adenylate cyclase-modulating G-protein coupled receptor signaling pathway		Guanine nucleotide-binding protein G(t) subunit alpha-3	Nov1_11273473 Nov1_11278152 Nov1_11283000 Nov1_11293049	ENSGALG00000008427	Intron Coding sequence Intron Downstream
GPM6B			Integral component of membrane	Neuronal membrane glycoprotein M6-b isoform X1	rs313845658	ENSGALG00000016575	Intron
HDAC4	Histone deacetylase activity	Chromatin organization Histone deacetylation	Histone deacetylase complex	histone deacetylase 4 isoform X3	rs313845658	ENSGALG00000016575	Intron
HDGFL3				hepatoma-derived growth factor	rs312523275	ENSGALG00000006036	Intron
HNF1A	DNA binding	Positive regulation of transcription, DNA-templated	Nucleus	Hepatocyte nuclear factor 1-alpha	Nov15_9145634	ENSGALG00000006968	Downstream
ITPKB	Kinase activity	Inositol phosphate biosynthetic process		Inositol-trisphosphate 3-kinase B	rs313697635	ENSGALG00000009068	Intron
KIF18B	Microtubule motor activity ATP binding Microtubule binding	Microtubule-based movement		Kinesin-like protein KIF18B	rs316143928	ENSGALG00000031553	Downstream
LDB3	Protein binding			LIM domain-binding protein 3 isoform X6	rs314384737	ENSGALG00000001977	Intron

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
LEO1	Transcription elongation from RNA polymerase II promoter	Histone modification	Cdc73/Paf1 complex	RNA polymerase-associated protein LEO1 isoform X1	Nov10_8997707	ENSGALG00000004696	Intron
MAP3K4	Protein kinase activity ATP binding	Protein phosphorylation		Mitogen-activated protein kinase kinase kinase 4 isoform X3	Nov3_45390506	ENSGALG00000020003	Intron
MCRIP1				MAPK-regulated corepressor-interacting protein 1	Nov18_10005222	ENSGALG00000042814	Intron
MDM1	Microtubule binding	Negative regulation of centriole replication		Nuclear protein MDM1 isoform X1	rs315517349	ENSGALG00000009905	Intron
MEOX2	DNA binding			Homeobox protein MOX-2	Nov2_28143942	ENSGALG00000010794	Intron
MRPL13	Structural constituent of ribosome	Translation	Ribosome	39S ribosomal protein L13, mitochondrial	Nov2_137044165	ENSGALG00000041863	Intron
MSRA	Peptide-methionine (S)-S-oxide reductase activity	Oxidation-reduction process		Mitochondrial peptide methionine sulfoxide reductase isoform X2	rs14410460	ENSGALG00000028032	Intron
MTBP				mdm2-binding protein	Nov2_137044165	ENSGALG00000031524	Upstream
MTCL1		Regulation of autophagy	Extracellular space	Microtubule cross-linking factor 1	Nov2_99052416	ENSGALG00000038316	Intron
MYEF2	RNA polymerase II transcription factor activity, sequence-specific DNA binding Nucleic acid binding Single-stranded DNA binding			Myelin expression factor 2 isoform X11	Nov10_10045281	ENSGALG00000032546	Downstream
MYH1A	Motor activity Microtubule motor activity ATP binding; Microtubule binding Actin filament binding	Microtubule-based movement	Myosin complex	Myosin heavy chain, skeletal muscle, adult-like	rs315592041 rs315562653 rs313992131	ENSGALG00000037864	Upstream Upstream Upstream
MYL9	Calcium ion binding			Myosin regulatory light polypeptide 9	Nov20_497735	ENSGALG00000028567	Upstream
MYO1D	Motor activity Microtubule motor activity Protein binding ATP binding Microtubule binding	Microtubule-based movement	Myosin complex	Unconventional myosin-Id isoform X2	rs315880270 rs316269099 rs317270020 rs316887638 rs315521303 rs315181886 rs314212462 rs16206894 rs313313466	ENSGALG00000000674	Intron Intron Intron Splice region Coding sequence Intron Intron Intron Intron

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
MYO9B	Motor activity Microtubule motor activity GTPase activator activity Protein binding ATP binding Microtubule binding	Microtubule-based movement Intracellular signal transduction Positive regulation of GTPase activity Signal transduction	Myosin complex	unconventional myosin-IXb isoform X4	rs315203294 rs314198057	ENSGALG00000003717	Intron Intron
NBL1				Neuroblastoma suppressor of tumorigenicity 1	rs317872657	ENSGALG00000040297	Upstream
NSDHL	3-beta-hydroxy-delta5-steroid dehydrogenase activity Oxidoreductase activity	Steroid biosynthetic process Oxidation-reduction process		Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	rs15494233	ENSGALG00000007493	Upstream
NTRK3	Protein kinase activity Protein tyrosine kinase activity Transmembrane receptor protein tyrosine kinase activity Neurotrophin receptor activity Protein binding ATP binding	Protein phosphorylation Transmembrane receptor protein tyrosine kinase signaling pathway	Integral component of plasma membrane	NT-3 growth factor receptor isoform X1	rs317446438 rs317668397 rs315506468 rs13789459 rs14949764	ENSGALG00000040241	Intron Intron Intron Intron Intron
NUAK1	Protein kinase activity ATP binding	Protein phosphorylation		NUAK family SNF1-like kinase 1 isoform X1	Nov1_54014429	ENSGALG00000012662	Intron
OSBPL1A	Protein binding			Oxysterol-binding protein-related protein 1 isoform X1	rs13707291	ENSGALG00000015086	Intron
P4HB	Isomerase activity	Cell redox homeostasis	Endoplasmic reticulum	Protein disulfide-isomerase	Nov18_10015927	ENSGALG00000038290	Intron
PAPPA	Metalloendopeptidase activity	Proteolysis		pappalysin-1 isoform X1	rs317226714	ENSGALG00000007079	Intron
PCBP3	Nucleic acid binding RNA binding			Poly(rC)-binding protein 3 isoform X1	Nov7_6692577 Nov7_6692745 Nov7_6694367	ENSGALG00000004394	Upstream Upstream Intron
PDS5A	Binding			Sister chromatid cohesion protein PDS5 homolog A isoform X1	Nov4_69516300	ENSGALG00000014286	Intron
PHACTR2				phosphatase and actin regulator 2	Nov3_52491675	ENSGALG00000013768	Intron
POLR3A	DNA binding DNA-directed 5'-3' RNA polymerase activity	Transcription, DNA-templated		DNA-directed RNA polymerase III subunit RPC1 isoform X2	rs314789197	ENSGALG00000004947	Upstream
PPP1R27	Protein binding			Protein phosphatase 1 regulatory subunit 27	Nov18_10005222	ENSGALG00000035913	Downstream

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
PUM1	RNA binding Binding			Pumilio homolog 1 isoform X1	rs15192209	ENSGALG00000042598	Intron
RANBP10	Protein binding			Ran-binding protein 10 isoform X1	rs15599960	ENSGALG00000001439	Upstream
RIMS2	Protein binding	Regulation of exocytosis Calcium ion-regulated exocytosis of neurotransmitter	Presynaptic active zone	Regulating synaptic membrane exocytosis protein 2 isoform X2	Nov2_130514583	ENSGALG00000040256	Intron
RLBP1	Visual perception Retinal binding			Retinaldehyde-binding protein 1	rs317366770	ENSGALG00000006676	Downstream
RNF185				E3 ubiquitin-protein ligase RNF185 isoform X1	Nov15_9145634	ENSGALG00000006960	Upstream
ROCK1	Protein kinase activity Protein serine/threonine kinase activity ATP binding Rho GTPase binding GTP-Rho binding;	Protein phosphorylation Rho protein signal transduction Actin cytoskeleton organization Intracellular signal transduction Regulation of stress fiber assembly Regulation of cytoskeleton Regulation of establishment of cell polarity organization		Rho-associated protein kinase 1	rs314226715	ENSGALG00000014922	Intron
ROR1	Protein kinase activity Protein binding ATP binding	Protein phosphorylation Transmembrane receptor protein tyrosine kinase signaling pathway		Tyrosine-protein kinase transmembrane receptor ROR1	rs317593555 Nov8_28008722 Nov9_17536551	ENSGALG0000001101 ENSGALG00000008983	Intron Intron Intron
RPS24	Structural constituent of ribosome	Translation	Ribosome Intracellular	40S ribosomal protein S24 isoform X1	rs314789197	ENSGALG00000004871	Intron
SCO1				protein SCO1 homolog, mitochondrial	rs315279672	ENSGALG00000030475	Intron
SH3GL1	Protein binding	Endocytosis	Cytoplasm	Endophilin-A2 isoform X2	Nov28_2502832	ENSGALG00000001121	Downstream
SIN3B		Regulation of transcription, DNA-templated		Paired amphipathic helix protein Sin3b	rs14307705	ENSGALG00000003771	Downstream
SLC12A1	Transporter activity Cation: chloride symporter activity	Ion transport Transmembrane transport	Membrane Integral component of membrane	Solute carrier family 12 member 1 isoform X1	Nov10_10081846	ENSGALG00000004945	Coding sequence

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
SLC19A1	Folate: anion antiporter activity Methotrexate transmembrane transporter activity Vitamin transmembrane transporter activity	Folic acid transport Vitamin transport Methotrexate transport	Integral component of membrane Integral component of plasma membrane	Folate transporter 1	Nov7_6666747 Nov7_6667828 Nov7_6672285	ENSGALG00000030511	Intron Intron Upstream
SLC24A5	Calcium, potassium: sodium antiporter activity	Ion transport Regulation of melanin biosynthetic process Transmembrane transport	Integral component of membrane Trans-Golgi network	Sodium/potassium/calcium exchanger 5	Nov10_10045281	ENSGALG00000004885	Intron
SLC35E1				Solute carrier family 35 member E1	rs313429355	ENSGALG00000003794	Downstream
SNTG1	Structural molecule activity Protein binding			Gamma-1-syntrophin isoform X1	rs14232568	ENSGALG00000015248	Intron
SP8	Nucleic acid binding			Transcription factor Sp8	Nov2_30245399	ENSGALG00000035725	Downstream
SPG11				Spatacsin isoform X2	Nov10_19760216	ENSGALG00000008180	Coding sequence
SPPL2A	Aspartic-type endopeptidase activity Aspartic endopeptidase activity, intramembrane cleaving	Membrane protein intracellular domain proteolysis Regulation of immune response	Integral component of membrane	Signal peptide peptidase-like 2A isoform X4	rs314300039	ENSGALG00000031348	Downstream
SYNPO2	Protein binding			Synaptopodin-2 isoform X1	Nov4_55204789	ENSGALG00000011994	Intron
TIRAP	Protein binding	Signal transduction		Toll/interleukin-1 receptor domain-containing adapter protein isoform X1	rs314257517	ENSGALG00000040224	Downstream
TMEM132B				Transmembrane protein 132B isoform X1	Nov15_4376122 Nov15_4422532 Nov15_4474884	ENSGALG00000002742	Intron Intron Upstream
TMEM220				Transmembrane protein 220	rs315279672	ENSGALG00000033996	Downstream
TRIAP1				TP53-regulated inhibitor of apoptosis 1	Nov15_9334468	ENSGALG00000007199	Upstream

Table 4. Continued

Gene	Molecular Function	Biological Process	Cellular	Description	SNP ID	Ensemble ID	Position of SNP
TRPM5	Ion channel activity Calcium activated cation channel activity	Ion transport Cation transport Ion transmembrane transport Sensory perception of taste Transmembrane transport	Membrane Integral component of membrane	Transient receptor potential cation channel subfamily M member 5	rs14517355	ENSGALG00000006521	Intron
TSNAXIP1				Translin-associated factor X-interacting protein 1 isoform X1	rs15599960	ENSGALG00000001533	Intron
TSSC4				protein TSSC4	rs14517355	ENSGALG00000006530	Downstream
TTC39C	Protein binding			Tetratricopeptide repeat protein 39C	rs312882159	ENSGALG00000015064	Downstream
TXNRD1	Thioredoxin-disulfide reductase activity Electron transfer activity Oxidoreductase activity Flavin adenine dinucleotide binding	Cell redox homeostasis Oxidation-reduction process		Thioredoxin reductase 1, cytoplasmic	Nov1_54634855	ENSGALG00000035345	Intron
UHRF1BP1L				UHRF1-binding protein 1-like isoform X2	Nov1_47207134	ENSGALG00000011557	Intron
WDFY2	Protein binding Metal ion binding			WD repeat and FYVE domain-containing protein 2	Nov1_170160732	ENSGALG00000017019	Intron
ZMIZ1	Zinc ion binding			Zinc finger MIZ domain-containing protein 1 isoform X7	rs13566721	ENSGALG00000004850	Intron

3.3.2 Pathway Analysis

For ingenuity pathway analysis (IPA), the significant SNPs according to the empirical P-values of the association test were exploited to expand the number of SNPs for the efficiency of the analysis. SNPs issued by a dbSNP identifier can only be used for the recognition of the analysis. That is why among 762 SNPs, 336 SNPs were used for IPA.

Ingenuity pathway analysis indicated that several genes with associated SNP might play important roles in the control the molecular mechanisms of muscle development. These significant genes involved in 5 major canonical pathways identified in human: 1) signaling by Rho Family GTPases, 2) semaphorin signaling in neurons, 3) axonal guidance signaling, 4) 1D-myo-inositol hexakisphosphate biosynthesis, and 5) D-myo-inositol (1,3,4)-triphosphate biosynthesis. The molecular and cellular functions of these pathways influence the assembly, organization, maintenance, movement, morphology, and development of the cells in higher organisms. These pathways are considered to have important physiological roles of the hematological system, nervous system, and in organismal, tissue, and immune cell trafficking (Krämer et al., 2014).

Signaling by Rho Family GTPases mediates multiple biological processes. Earlier work suggested that Rho plays a role in cytoskeletal remodeling, but recent studies also contributed to decipher the biological functions of the Rho GTPases. It is a key molecule to regulate various other cellular processes such as membrane trafficking, transcriptional activation, cell growth control, and development (Van Aelst et al., 1997).

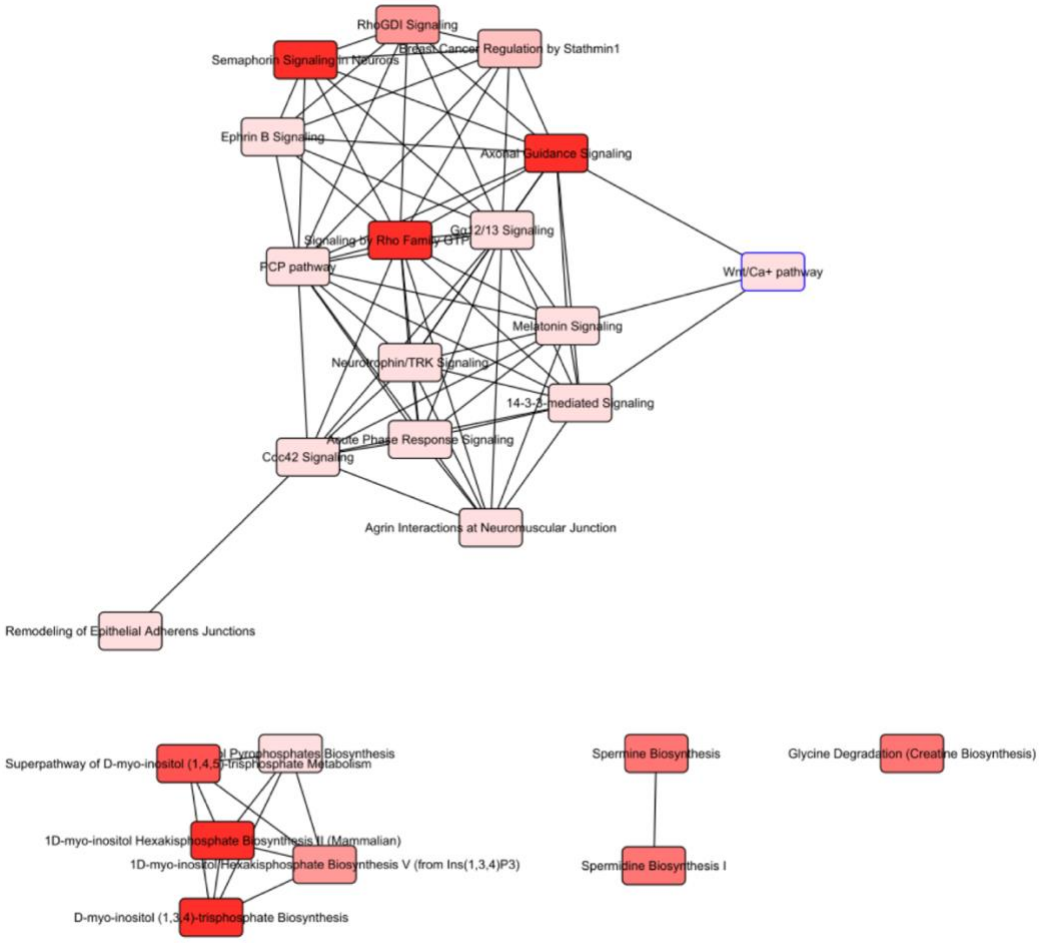


Figure 9**.** Overlapping canonical pathways

Semaphoring signaling in neurons is responsible for regulating many organs and tissue development, and maintenance processes in higher organisms. The alternation to the cytoskeletal and adhesive machinery cause semaphoring signaling to regulate the morphology and motility in diversified cell types by activating Plexin receptors (Alto & Terman, 2017).

Axonal guidance signaling is mainly engaged in the extension of neuronal, embryogenesis, cell proliferation, cell migration, adhesion, apoptosis biological processes (Tang et al., 2014). Activation of the axon guidance signaling pathway triggered by growth cone collapse can trigger repulsion and attraction. The activation also affects the rate of axon extension, which adjusts the dynamics of cytoskeletal rearrangement in the growth cone (Bashaw, & Klein, 2010).

1D-myo-inositol hexakisphosphate (phytate) is one of the most common intermediates of phosphorylated inositols in the cell of higher organisms and have many signaling roles. The cellular distributions of phytate still are not known. The function of phytate in cell might increase the binding affinity of cationic proteins. The molecular roles of phytate might be the regulating nuclear mRNA export, inhibition of clathrin cage assembly, chromatin remodeling, exocytosis of insulin, and changing Ca^{2+} channel activation (Chang et al., 2002).

3.4 Haplotype Analysis Results

In total, 18,539 haplotype blocks were constructed from the whole chicken genome within 200 kb windows. The size of the haplotype blocks differed considerably. The smallest haplotype contains just six bp, the largest spans 199.851 bp. The largest number of SNP comprising a haplotype is 80. As shown in Figure 10, the number of haplotypes located on a chromosome was proportional to the size of the chromosome. The average number of SNP existing in a haplotype and the average haplotype length in kb per chromosome is depicted in Figure 11.

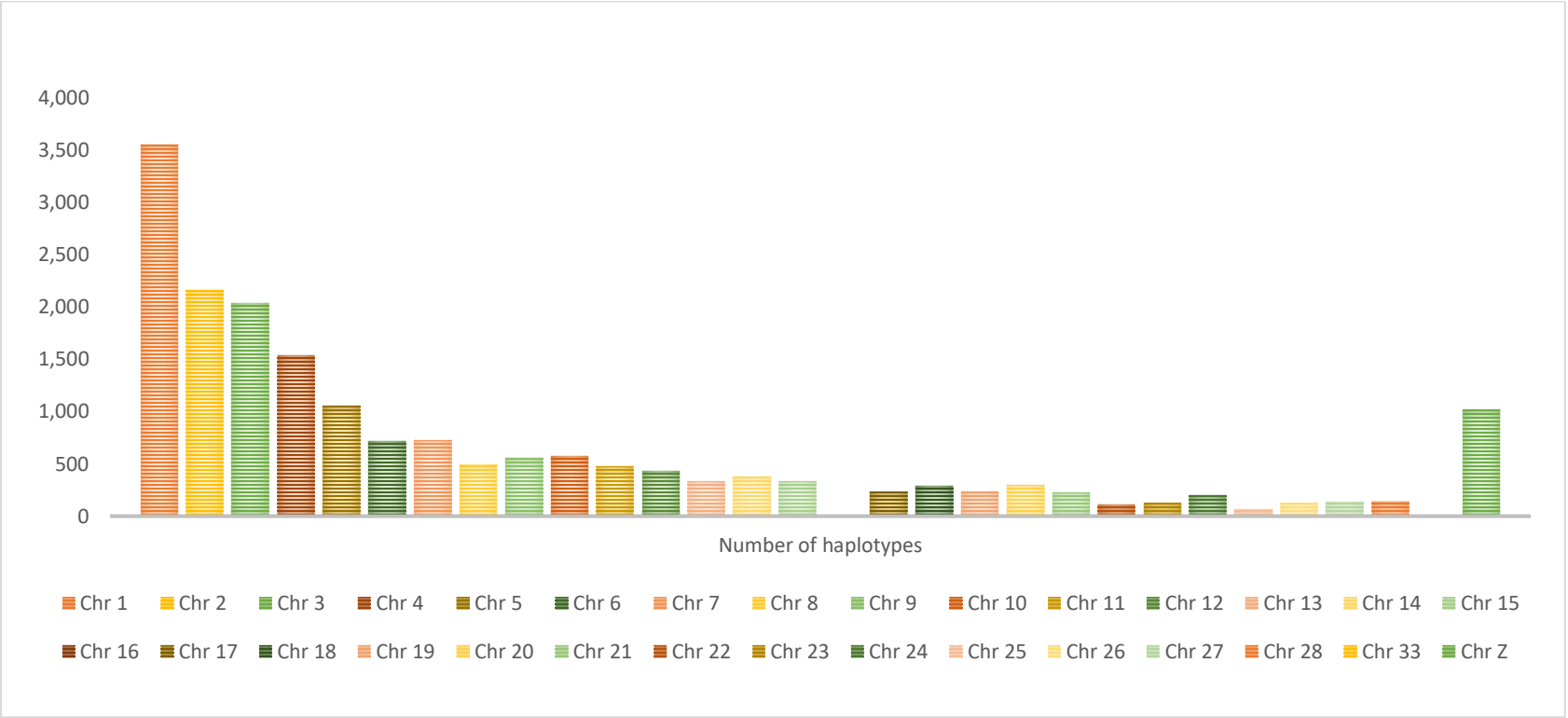
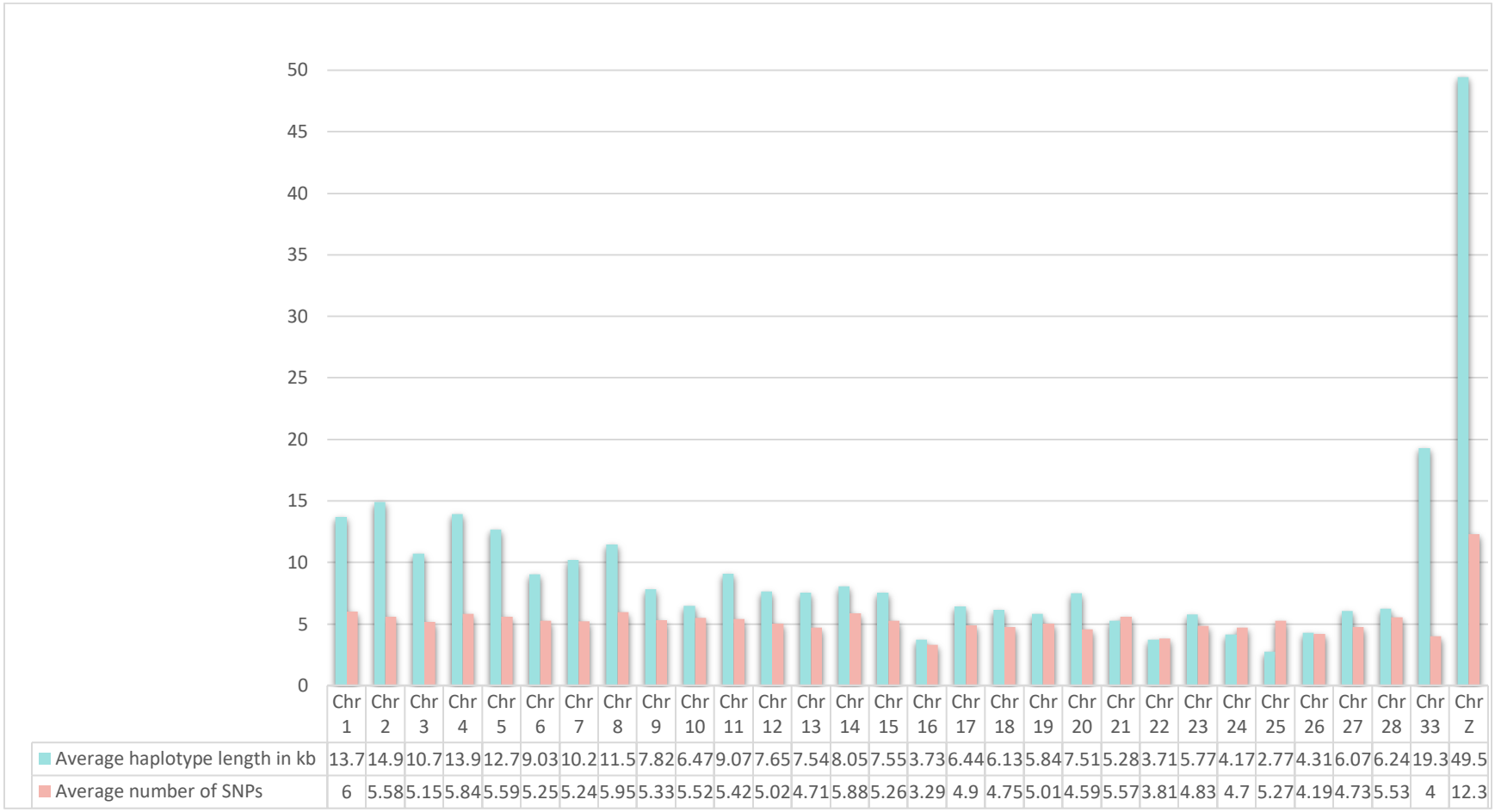


Figure 10. Distribution of the constructed haplotypes per chromosome



	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Chr 6	Chr 7	Chr 8	Chr 9	Chr 10	Chr 11	Chr 12	Chr 13	Chr 14	Chr 15	Chr 16	Chr 17	Chr 18	Chr 19	Chr 20	Chr 21	Chr 22	Chr 23	Chr 24	Chr 25	Chr 26	Chr 27	Chr 28	Chr 33	Chr Z
Average haplotype length in kb	13.7	14.9	10.7	13.9	12.7	9.03	10.2	11.5	7.82	6.47	9.07	7.65	7.54	8.05	7.55	3.73	6.44	6.13	5.84	7.51	5.28	3.71	5.77	4.17	2.77	4.31	6.07	6.24	19.3	49.5
Average number of SNPs	6	5.58	5.15	5.84	5.59	5.25	5.24	5.95	5.33	5.52	5.42	5.02	4.71	5.88	5.26	3.29	4.9	4.75	5.01	4.59	5.57	3.81	4.83	4.7	5.27	4.19	4.73	5.53	4	12.3

Figure 11. Average number of SNPs per haplotype and the average haplotype length

After the haplotype association tests were performed for each haplotype, a global P-value and individual P-values of different combinations of SNP showing variation on the haplotype were recorded. Among the most significant SNPs, only coding sequence variants were visualized with their haplotype blocks. 50 kb regions harboring the most significant SNPs and their related genes were analyzed for linkage disequilibrium which represents nonrandom associations of alleles that are located across a chromosome.

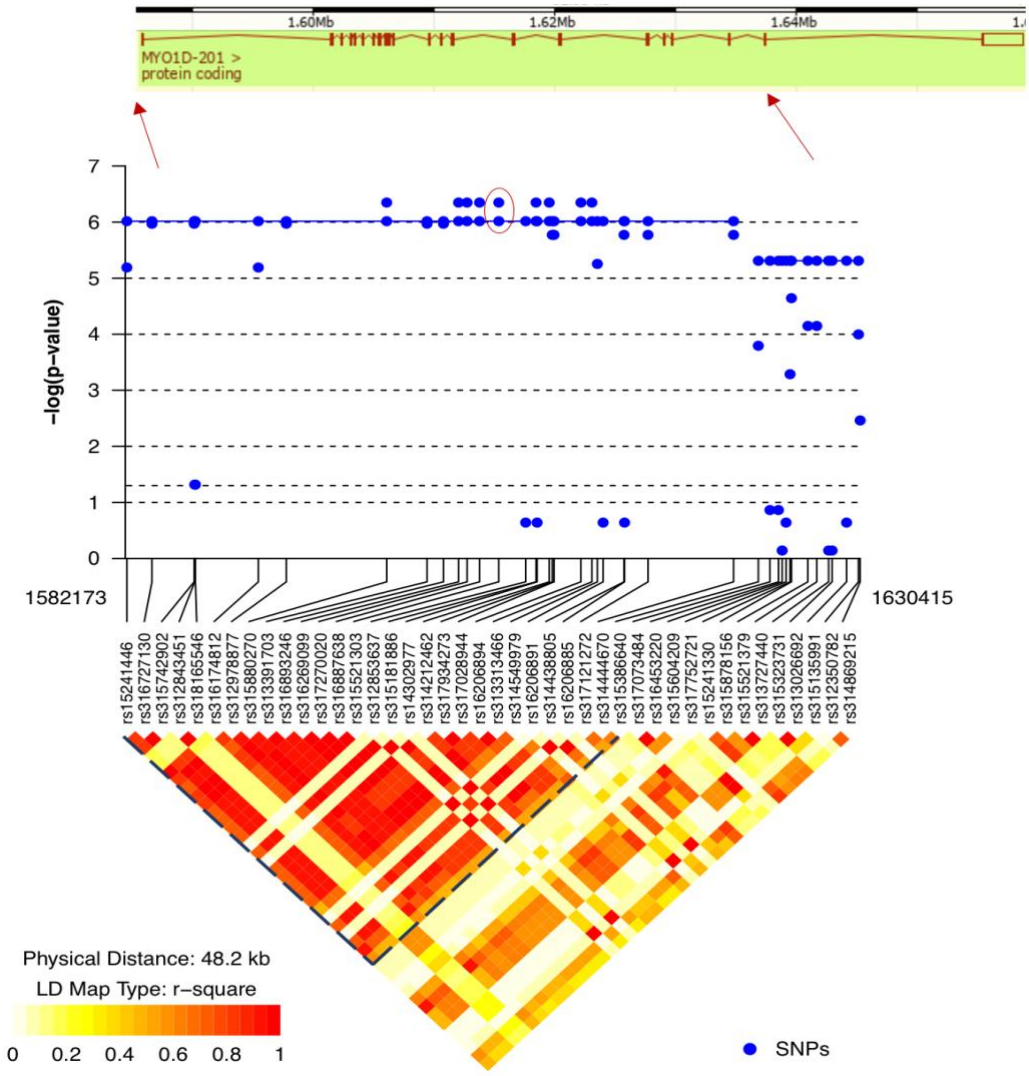


Figure 12. Associations in the MYO1D gene region with the corresponding LD heatmap based on r^2 metric. Marker number rs315521303, located within this gene, is circled; this was one of the most significant coding sequence variants. Haplotypes and SNP portrayed are in the region within 50 kb of this gene. Each point represents one SNP; solid lines connecting SNP represent haplotypes. The individual P-value of the haplotypes and FDR adjusted P-values of the SNP are shown as blue points. In the LD heatmap, the triangle region indicated by the dotted dark blue line shows the pairwise r^2 values of SNP located on the haplotype.

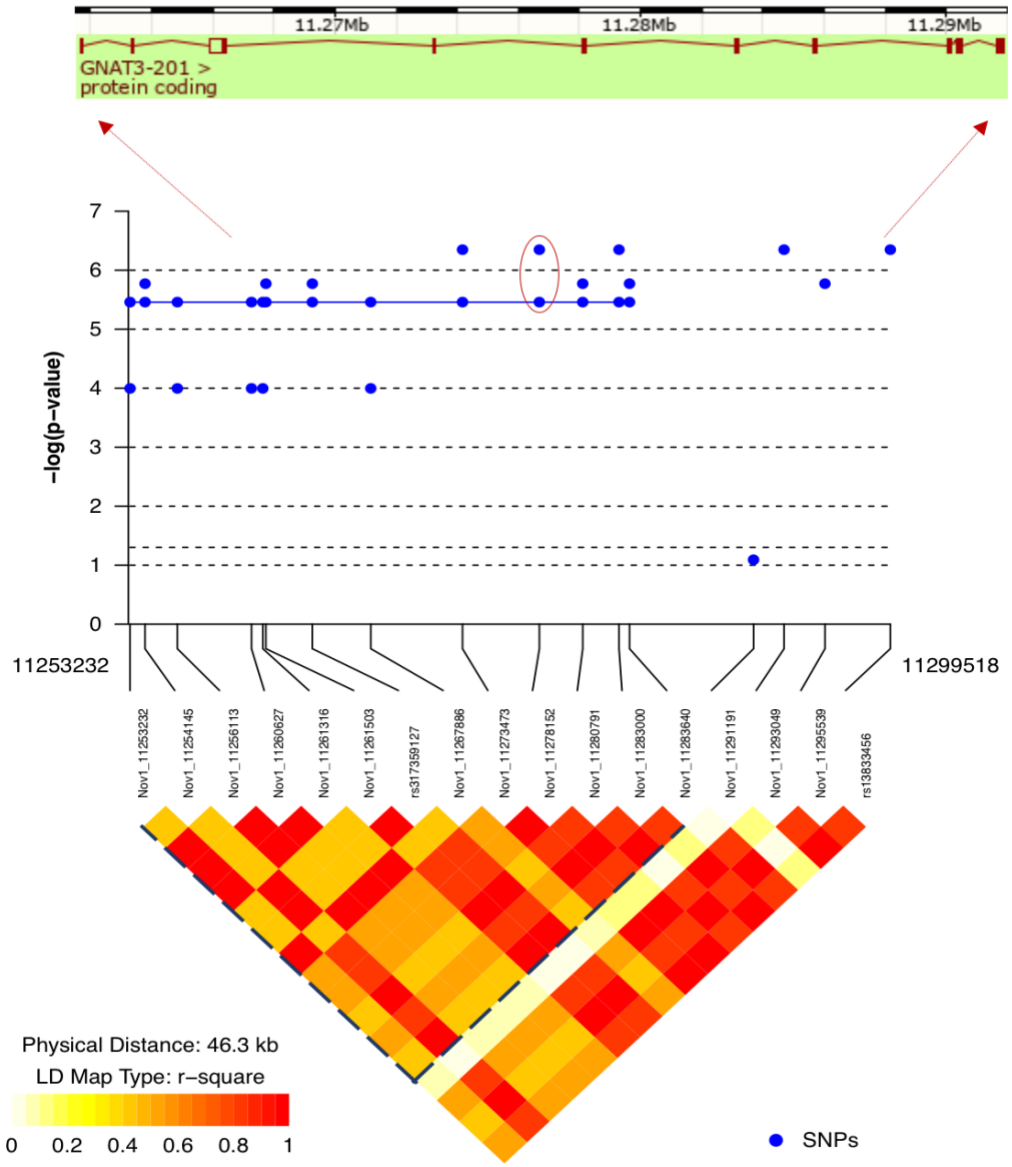


Figure 13. Associations in the GNAT3 gene region with the corresponding LD heatmap. Haplotypes and SNP within 50 kb of the significant coding sequence variant Nov1_11278152 are shown. This leading marker is located within GNAT3. The triangle area indicates the pairwise r^2 values of SNPs existing in the haplotype.

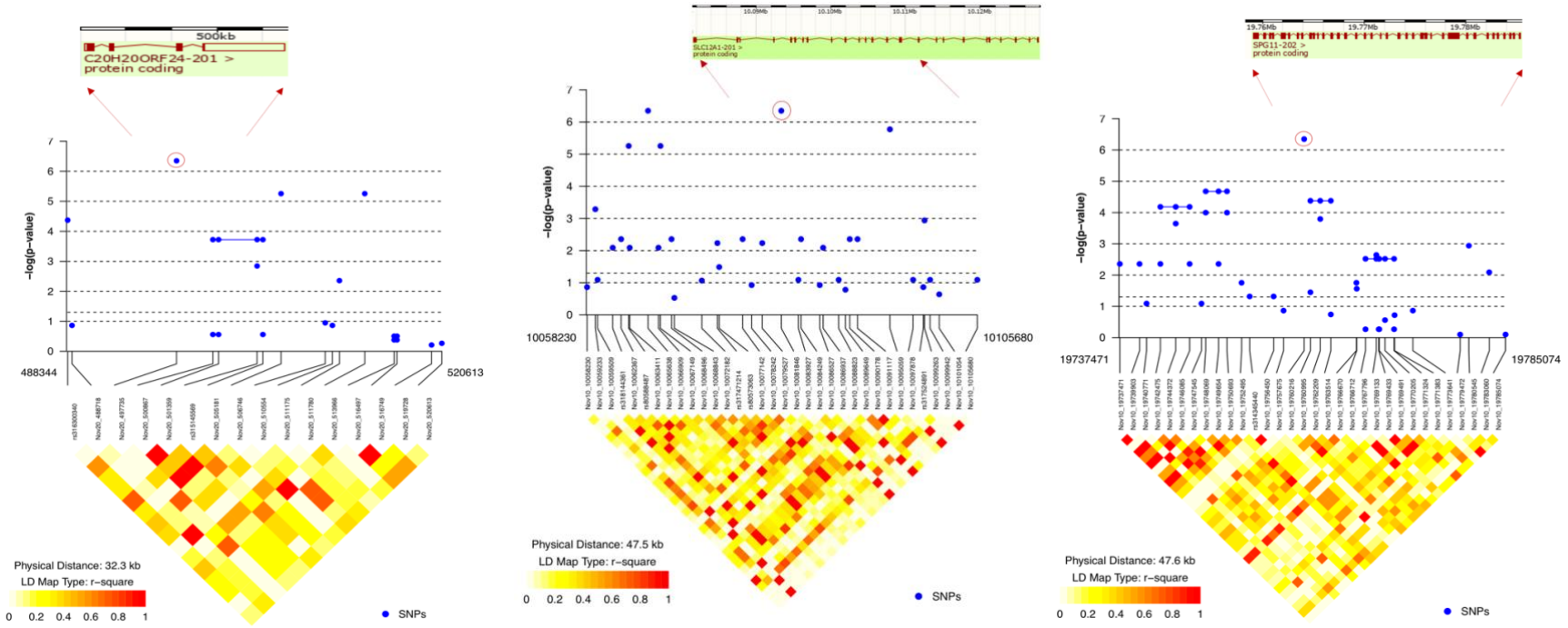


Figure 14. Associations in the C20H20ORF24, SLC12A1, SPG11 gene regions, respectively, with their corresponding LD heatmaps. The coding sequence variants are inside the red circles. As observed in the LD heatmap patterns, SNPs in the flanking regions do not exhibit high correlation. Even though there are a few small haplotype blocks observed, Nov20_497735, Nov10_10081846, Nov10_19760216 do not exist in these haplotypes.

In Figures 12, 13, and 14, the LD heatmap and haplotypes were plotted with SNP that corresponded to the most significant SNP associations with WB that were located within known genes. The metric used in LD heatmap is called r^2 ranging from 0 to 1 to exhibit the relatedness of each SNP pair in a genomic region of interest. As shown in Figures 12, 13, and 14, 1 indicates complete correlation with the intensely shaded red area in the LD heatmap. The decrease in the correlation between SNPs is proportional to the reduction in the intensity of the shaded area to yellow or white, and 0 indicates no correlation between SNPs of interest.

The plots show the association test results as the $-\log_{10}$(P-values). Nov1_11278152 located on GNAT3 gene and rs315521303 located on MYO1D gene is involved in the haplotype blocks seen in the Figure 12 and Figure 13. In the LD heatmaps, the triangle-shaped areas surrounded by the dotted dark blue line represent the pairwise r^2 values of the SNPs that are inherited together as a haplotype block. However, Nov20_497735 located within the gene C20H20ORF24, Nov10_10081846 located within the SLC12A1 gene and Nov10_19760216 located on SPG11 gene do not exist in a haplotype block.

The individual P-values indicated on the plot are haplotype association test statistics instead of global P-values to assure that the haplotypes contain the correct variant found in significant association with the WB condition. The haplotype harboring Marker rs315521303 has six different combinations of the same SNP, and three of them include the correct variant of rs315521303 that is guanine (G) nucleotide. Among three, the most frequent combination is taken into account to use its P-value. Table 5 shows two haplotypes having the coding sequence variants, genes that these haplotypes are positioned, number of SNPs, coordinates of haplotypes, haplotype length, frequency and P-values obtained from haplotype association test.

Table 5. Haplotypes harboring the significant coding sequence variants

Haplotype	SNP ID	Gene	Number of SNPs	Coordinates	Haplotype Length (Kb)	Frequency	P-value
H192	Nov1_11278152	GNAT3	16	1:11242819-11283640	40.822	0.35	0.000003468
H17264	rs315521303	MYO1D	28	27:1582173-1622086	39.914	0.375	9.634E-07

3.5 Population Stratification Results

The pairwise clustering performed with respect to IBS is favorable to identify pairs of individuals that exhibit more differences from one another than expected in a homogeneous sample. Figure 15 shows the substructure of the population, and there is a strong separation observed in the sample based on the values of the first two dimensions (C1 and C2). Each point represents an individual, and the individuals circled are the females existing in the control group. The case and control groups are positioned away from each other as expected. While broiler chickens cluster on the right hand, WPRs form a cluster on the left hand. Broiler chickens gather close to each other, which indicates differences within broilers are lower compared to those in WPRs because they are positioned apart.

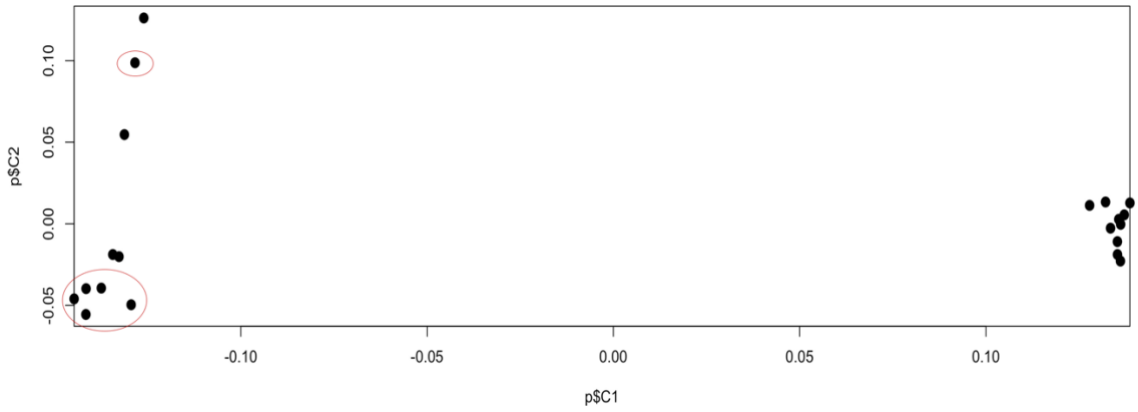


Figure 15. Multidimensional scaling plot.

As the Manhattan plots and the Q-Q plot suggested, the analysis was tremendously affected by population stratification. MDS plot demonstrates that clearly. In order to account for population stratification and sex, CMH test and logistic regression analysis were conducted. CMH test was replicated after removing females from the sample, and logistic regression analysis was performed with sex covariate alone, with cluster covariate alone and with sex and cluster covariates together. However, the result did not change, and there were no significant SNP detected after controlling the population stratification and sex.

4. DISCUSSION

In these analyses, population structure in the form of breed type apparently introduced a bias that is more likely responsible for the many observed significant associations of SNPs along the whole chicken genome. The Q-Q plot (Figure 7) indicates considerable deviation of P values of tests of association from expectation across the whole genome. One of the main confounding factors that affects the accuracy of genome-wide association studies is population structure (Thomas & Witte, 2002). In this case, the two breeds used for association testing are potentially highly diverged due to different rates of selection and show inherent differences in underlying allele frequencies. On the one hand, such differences in the genetic background could lead to a higher probability of false positives compared to associations that are indicative of WB myopathy. On the other hand, up to 89% of commercial broilers (including Ross 708 variety) are observed to express WB (Cruz et al., 2017), which makes it difficult to select a control group from the same genetic background. Although observing deviations from expected P-values is an anticipated result of any GWAS, the predominant part of SNPs included in association test are not supposed to be in association with the disease. Optimally, expected and observed P-values would correspond more or less to the red diagonal line in Figure 7 until the end in which it forms a curve towards the shaded area. Points showing deviation in the shaded area would be most likely to be a true disease variant (Figure 7). Strict adherence to this shows that only a handful of associated loci could be considered to be not false positives. A second consideration is the sample size in this present study, which is not ideal for a thorough association mapping study. As this study was exploratory, the uncertainty surrounding the validity of the association tests is not surprising. Detection of strong

deviations of observed P-values from expected P-values under the null hypothesis that assumes no association between SNPs and the WB condition, cannot be attributed to the strong association of SNP with the disease. Further analysis conducted using only the most significant SNP may eliminate, in part, spurious associations.

Results from haplotype analyses indicate GNAT3 and MYO1D genes have many significant SNP as well as the coding sequence variants that are inherited as haplotype blocks. GNAT3 is a guanyl nucleotide binding protein with GTPase activity and a key molecule for signal transduction involving several pathways. MYO1D is a microtubule and ATP binding protein in myosin complex with the molecular function of catalysis of the force generation leading to movement across a microtubule or a microfilament or torque leading to membrane scission.

Several pathways were identified by using the genes harboring the significant SNPs, which are associated with cellular repair, tissue development, and Ca^{2+} transportation by IPA. These results seem correlated to the symptoms of the WB myopathy identified in the previous studies. According to the result of IPA analysis, one of the top canonical pathways is signaling by Rho Family GTPases contributing to the cellular repair, which is consistent with the differential gene expression analysis of the chicken affected by the WB and WS myopathies (Mutryn et al., 2015). However, at the gene level, differences occurring in cellular repair could be the reason for muscle degradation observed in the affected chicken, rather than results. The result of the GO annotation analysis performed with the genes in which 250 significant SNP identified genes that are part of the Signaling by Rho Family GTPases, including Rho GTPase-activating protein 31 (ARHGAP31), Rho-associated protein kinase (ROCK1), and Rho guanine nucleotide exchange factor 18 isoform. Other significant genes related to cellular repair and fiber type switching that

could contribute to the onset of the WB myopathy are synaptopodin-2 isoform X1 (stress fiber) (SYNPO2), myosin heavy chain skeletal muscle (MYH1A), myosin regulatory light polypeptide 9 (stress fiber) (MYL9), unconventional myosin-Id isoform X2 (MYO1D), unconventional myosin-IXb isoform X4 (MYO9B), and phosphatase and actin regulator 2 (PHACTR2). Another significant pathway, 1D-myo-inositol hexakisphosphate (phytate), is presumed to be involved in the alternation of Ca^{2+} channel activation as well as a number of signaling roles, which may be associated with the WB myopathy. Likewise, various chicken genes were identified in the GO analysis related to Ca^{2+} binding including sodium/potassium/calcium exchanger 5 (SLC24A5), EF-hand calcium-binding domain-containing protein 14 isoform X1 (EFCAB14), calcium ion binding (FBLN1): calcium ion binding (CETN2:). Zambonelli et al. (2016) in their microarray-based differential gene expression analysis, demonstrated increased levels of calcium in affected WB fillets and found the differentially expressed genes associated with the calcium signaling pathway.

Among the annotated significant genes based on GO analyses, the prevalence of the genes exerting oxidoreductase activity is particularly noteworthy; those include peroxisomal acyl-coenzyme A oxidase 3 isoform X1 (ACOX3), dehydrogenase /reductase SDR family member 7C (DHRS7C), cytochrome P450 4B1-like (CYP4A22), cytochrome P450 4B1-like (CYP4B7), sterol-4-alpha-carboxylate 3-dehydrogenase (NSDHL), and thioredoxin reductase 1 (TXNRD1) as detected genes involved in oxidoreductase activity. Oxidative stress is considered as a biomarker of the chicken with the WB myopathy (Abasht et al., 2016), and it is speculated that oxidative stress stimulates changes occurring in the muscles of the affected chickens as a result of intense selection towards high breast yield (Zambonelli et al., 2006).

Several genes and pathways obtained as a result of the SNP and haplotype association analysis seem relevant to WB myopathy but the likelihood of these results being related to differences between two breeds is higher. Therefore, before the population stratification is corrected, the results are not valid.

5. CONCLUSION

The present study is focused on the investigation of the genetic basis that predisposes the commercial broiler chickens to develop the WB myopathy at gene and SNP level by exploiting GWAS. Considering small sample size and structure, this study serves as an exploratory association study. The population structure exhibits a stratification that dramatically influenced the association results. The use of stringent empirical P-value threshold and FDR-BH adjusted association test statistics was employed in effort to mitigate the limitations as much as possible. The haplotype analysis incorporated into the SNP association analysis for MYO1D, GNAT3, SLC12A1, SPG11 and C20H20ORF24 genes may enhance the robustness of the association study. The GO annotation analysis and the pathway analysis conducted with a small set of genes with FDR-BH adjusted P-values, and very small empirical P-values respectively are deemed informative, and the results are consistent with the previous functional studies published by other researchers.

Even though, using P-values corrected for multiple tests and the FDR adjusted P-values may decrease the number of false positive associations, it is hard to draw a conclusion about disease association from the data unless corrected for stratification. After correcting for population stratification no significant SNPs were identified, likely due to a lack of power in the experimental design.

REFERENCES

Abasht, B., Mutryn, M.F., Michalek, R.D., & Lee, W.R. (2016). Oxidative Stress and Metabolic Perturbations in Wooden Breast Disorder in Chickens. *Plos One* 11(4): e0153750.

Alto, L.T., & Terman, J.R. (2017). Semaphorins and their Signaling Mechanisms. *Methods in Molecular Biology* 1493, 1–25.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P., & Zondervan, K.T. (2010). Data quality control in genetic case-control association studies. *Nature protocols* 5(9),1564-1573. doi:10.1038/nprot.2010.116.

Anthony N. B. (1998). A review of genetic practices in poultry: Efforts to improve meat quality. *J Muscle Foods* 9, 2533.

Axelsson, E., Webster, M.T., Smith, N.G.C., Burt, D.W., & Ellegren, H. (2005). Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. *Genome Res.* 15, 120–125.

AVEC (2012): Annual report 2012. Brussels. Retrieved from http://www.avec-poultry.eu/

Bashaw, G.J., & Klein, R. (2010). Signaling from Axon Guidance Receptors. *Cold Spring Harbor Perspectives in Biology 2*(5), a001941.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.* B 57:289–300.

Bilgili, S.F. (2013, April). Broiler Chicken Myopathies: II. Woody Breast? In: Worthw Oper Guidel Suggest.

Bondoc, O.L. (2008). *Animal Breeding: Principles and Practice in the Philippine Context*. Quezon, Philippine: The University of Philippines Press

Burt, D.W. (2005). Chicken genome: current status and future opportunities. *Genome Res.* 15, 1692–1698.

Calus, M., Meuwissen. T., Windig, J., Knol, E., Schrooten, C., Vereijken, A.L.J. & Veerkamp, R.F. (2009). Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genetics Selection Evolution* 41, 11.

Chang, S.C., Miller, A.L., Feng, Y., Wente, S.R, & Majerus, P.W. (2002). The human homolog of the rat inositol phosphate multikinase is an inositol 1,3,4,6-tetrakisphosphate 5-kinase. *J Biol Chem*. 277(46): 43836-43843.

Cochran, W.G. (1954). Some Methods for Strengthening the Common Chi-Squared Tests. *Biometrics* 10 (4): 417–51.

Collins, F.S., Guyer, M.S., & Chakravarti, A. (1997). Variations on a theme: cataloging human DNA sequence variation. Science 278(5343):1580-1581.

Conesa, A., Götz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676.

Cruz, R.F.A., Vieira, S.L., Kindlein L., Kipper M., Cemin H.S. & Rauber S.M. (2017). Occurrence of white striping and wooden breast in broilers fed grower and finisher diets with increasing lysine levels. *Poultry Science* 96, 501–510. doi: 10.3382/ps/pew310.

Gee, K. (2016). Poultry's tough new problem: 'Woody Breast' Wall Street Journal. *Sect. Business and Tech*. CCLXVII: B1.

Guetchom, B., Venne, D., Chenier, S., & Chorfi, Y. (2012). Effect of extra dietary vitamin E on preventing nutritional myopathy in broiler chickens. *Appl Poult Res*. 21(3):548–55.

Hardy, J. & Singleton, A. (2009). Genomewide association studies and human disease. *N Engl J Med*. 360,1759-1768.

Herrera, E., & Barbas, C. (2001). Vitamin E: action, metabolism and perspectives. *J. Physiol. Biochem.* 57 4356.

Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E., Dodgson, J. B., Chinvalla, A. T. Cliften, P. F. Clifton, S.W. & Delehaunty, K.D. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.

Jordan, F.T.W., & Pattison, M. (1998). *Poultry Diseases*. London, UK: Saunders Elsevier.

Julian R.J. (2005). Production and growth-related disorders and other metabolic diseases of poultry-A Review. *Vet. J.* 169 350369.

Krämer, A., Green, J., Pollard, J.Jr., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30(4):523-30.

Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P.M., Fife, M., Salmon, N., Fulton, J., Strom, T.M., Haberer, G., Weigend, S., Preisinger, R., Gholami, M., Qanbari, S., Simianer, H., Watson, K.A., Woolliams, J.A. & Burt, D.W. (2013). Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14(59), 10.1186/1471-2164-14-59.

Kuttappan, V.A., Hargis, B.M., & Owens, C.M. (2016). White striping and woody breast myopathies in the modern poultry industry: a review. *Poult Sci.* 95(11):2724-2733. doi:10.3382/ps/pew216.

Kuttappan, V.A., Shivaprasad, H.L., Shaw D.P., Valentine, B.A., Hargis, B.M., Clark, F.D., McKee, S.R., & Owens, C.M. (2013). Pathological changes associated with WS in broiler breast muscles. *Poult. Sci.* 92(2) 331-8 doi: 10.3382/ps.2012-02646.

Li, S., He, Y., Zhao, H., Wang, Y., Liu, J., Shao, Y., Li, J., Sun, X., Zhang, L., & Xing, M. (2017). Assessment of 28 trace elements and 17 amino acid levels in muscular tissues of broiler chicken

(Gallus gallus) suffering from arsenic trioxide. *Ecotoxicology and Environmental Safety* 144, 430-437.

Luna, A., & Nicodemus, K.K. (2007). snp.plotter: An R-based SNP/haplotype association and linkage disequilibrium plotting package. *Bioinformatics* 23, 6:774-6.

Mantel, N., & W. Haenszel. (1959). Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. *Journal of the National Cancer Institute* 22 (4): 719–48.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology* 17(1), 122. doi:10.1186/s13059-016-0974-4.

Meunier, J. & Duret, L. (2004). Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* 21: 984–990.

Mutryn, M.F., Brannick, E.M., Fu, W., Lee, W.R., & Abasht, B. (2015). Characterization of a novel chicken muscle disorder through differential gene expression and pathway analysis using RNA-sequencing. *BMC Genomics* 16(1), 399. doi:10.1186/s12864-015-1623-0.

Mutryn, M., Fu, W., & Abasht, B. (2015). Incidence of Wooden Breast Disease and its correlation with broiler performance and ultimate pH of breast muscle. Proceeding of XXII European symposium on the quality of poultry meat. Nantes, France.

National Chicken Council (2015). Statistics and research: U.S. Broiler Performance http://www.nationalchickencouncil.org/about-the-industry/statistics/u-s-broiler-performance/.

Ovcharenko, I., Loots, G.G., Nobrega, M.A., Hardison, R.C., Miller, W., & Stubbs, L. (2005). Evolution and functional classification of vertebrate gene deserts. *Genome Res.* 15: 137-145 doi:10.1101/gr.3015505.

Pampouille, E., Berri, C., Boitard, S., Hennequet-Antier, C., Beauclercq, S. A., Godet, E., Praud, C., Jégo, Y., & Le Bihan-Duval, E. (2018). Mapping QTL for white striping in relation to breast muscle yield and meat quality traits in broiler chickens. *BMC Genomics 19*, 202. http://doi.org/10.1186/s12864-018-4598-9

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., De Bakker, P.I.W., Daly, M.J., & Sham, P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* 81: 559-575.

R Development Core Team R. (2006). A language and environment for statistical computing.

Robert, J.K. (2007). Power analysis for genome-wide association studies. *BMC Genetics* 8(58), doi:10.1186/1471-2156-8-58.

Schmid, M., Nanda, I., Guttenbach, M., Steinlein, C., Hoehn, M., Schartl, M., Haaf, T., Weigend, S., Fries, R., Buerstedde, J.M., Wimmers, K, Burt, D.W., Smith, J., A'Hara, S., Law, A., Griffin, D.K., Bumstead, N., Kaufman, J., Thomson, P.A., Burke, T., Groenen, M.A., Crooijmans, R.P., Vignal, A., Fillon, V., Morisson, M., Pitel, F., Tixier-Boichard, M., Ladjali-Mohammedi, K.,

Hillel, J., Mäki-Tanila, A., Cheng, H.H., Delany, M.E., Burnside, J., & Mizuno, S. (2000). First report on chicken genes and chromosomes 2000. *Cytogenet. Cell Genet*. 90**:** 169-218.

Schmutz, J., & Grimwood, J. (2004). Genomes: fowl sequence. *Nature* 432(7018), 679-80.

Sihvo, H.K., Immonen, K., & Puolanne, E. (2014) Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. *Vet Pathol*. 51:619–23.

Smith, J., Bruley, C.K., Paton, I.R., Dunn, I., Jone, C.T., Windsor, D., Morrice, D.R., Law, A.S., Masabanda, J., Sazanov, A., Waddington, D., Fries, R., & Burt, D.W. (2000). Differences in gene density on the Chicken macrochromosomes and microchromosomes: A tool for gene discovery in vertebrate genomes. *Anim. Genet.* 31, 96–103.

Soglia, F., Mudalal, S., Babini, E., Di Nunzio, M., Mazzoni, M., Sirri, F., Cavani, C., & Petracci, M. (2016). Histology, composition, and quality traits of chicken Pectoralis major muscle affected by wooden breast abnormality. *Poult. Sci.* 95, 651-659.

Stickland, N. C. (1995). Microstructural aspects of skeletal muscle growth. 2nd Dummerdorf Muscle Workshop—Muscle Growth and Meat Quality, Rostock, Germany.

Szűcs, I. (2013). *Sectorial Economy II*. Retrieved from https://www.tankonyvtar.hu/en/tartalom/tamop412A/2011_0009_Szucs_et_al-Sectorial_Economy_II

Tang, H., Wei, P., Duell, E. J., Risch, H.A., Olson, S.H., Bueno-de-Mesquita, H. B., Gallinger, S., Holly, E.A., Petersen, G., Bracci, P.M., McWilliams, R.R., Jenab, M., Riboli, E., Tjønneland, A., Boutron-Ruault, M.C., Kaaks, R., Trichopoulos, D., Panico, S., Sund, M., Peeters, P.H., Khaw, K.T., Amos, C.I., & Li, D. (2014). Axonal guidance signaling pathway interacting with smoking in modifying the risk of pancreatic cancer: a gene- and pathway-based interaction analysis of GWAS data. *Carcinogenesis 35*(5), 1039–1045.

Thomas, D.C., & Witte, J.S. (2002). Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev*. 11(6):505-512.

Trocino, A., Piccirillo, A., Birolo, M., Radaelli, G., Bertotto, D., Fil- iou, E., Petracci, M., & Xiccato., G. (2015). Effect of genotype, gender and feed restriction on growth, meat quality and the occurrence of white striping and wooden breast in broiler chickens. *Poult. Sci.* 94:2996–3004.

USDA. (2016). *Broiler Market News Report*. Des Moines, IA: Livestock Poultry Grain Market News Service, 49.

Wicker, T., Robertson, J.S., Schulze, S.R., Feltus, F.A., Magrini, V., Morrison, J.A., Mardis, E.R., Wilson, R.K., Peterson, D.G., Paterson, A.H., & Ivarie, R. (2005). The repetitive landscape of the chicken genome. *Genome Res.* 15, 126–136.

Wight, P.A.L., & Siller, W.G. (1980). Pathology of deep pectoral myopathy of broilers. *Vet. Pathol.* 17 2939.

Xiang, W., Shang, Y., Wang, Q., Xu, Y., Zhu, P., Huang, K., & Xu, W. (2017). Identification of a chicken (Gallus gallus) endogenous reference gene (Actb) and its application in meat

adulteration. *Food Chemistry* 234, 472-478.

Van Aelst L, D'Souza-Schorey C. (1997). Rho GTPases and signaling networks. *Genes and Development* 11(18), 2295-322.

Zambonelli, P., Zappaterra, M., Soglia, F., Petracci, M., Sirri, F., Claudio, C., & Davoli, R. (2016). Detection of differentially expressed genes in broiler pectoralis major muscle affected by White Striping - Wooden Breast myopathies. *Poultry science* 95. 10.3382/ps/pew268.