University of São Paulo
"Luiz de Queiroz" College of Agriculture

Unraveling important genetic associations and differential methylation profiles using reduced genome sequencing in chickens

Fábio Pértille

Thesis presented to obtain the degree of Doctor in Science.
Area: Animal Science and Pastures

Piracicaba
2016

Fábio Pértille
Veterinarian

Unraveling important genetic associations and differential methylation profiles using reduced genome sequencing in chickens

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **LUIZ LEHMANN COUTINHO**

Thesis presented to obtain the degree of Doctor in Science.
Area: Animal Science and Pastures

Piracicaba
2016

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Pértille, Fábio
Unraveling important genetic associations and differential methylation profiles using reduced genome sequencing in chickens / Fábio Pértille. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016.
121 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura “Luiz de Queiroz”.

1. Galinha 2. GWAS 3. MeDIPS 4. Melhoramento animal 5. Saúde e bem-estar animal I. Título

CDD 636.5082
P469u

AKNOWLEDGEMENTS

First of all I will like to acknowledge the College of Agriculture “Luiz de Queiroz” from São Paulo University (ESALQ/USP), thanks for the knowledge and support. For the Animal Biotechnology Laboratory group under the coordination of Dr. Luiz Lehmann Coutinho. Dr. Coutinho I’m grateful for the opportunity to be part of this great group, in addition to that, thank you for all the advises and time that you spent with me, mostly guiding me to achieve this degree. I would like to say thanks for the members of the thematic project “Identificação de locos de interesse zootécnico na galinha doméstica” (FAPESP Grant number: 2014/08704-0) for the valuable discussions and advises, mostly Dr. Mônica Corrêa Ledur, from Embrapa Suínos e Aves. I also want to say thanks to Prof. Dr. Ricardo Zanella which trusted me and helped me to know the Swedish group, where I have conducted part of my Doctorate, and followed my steps during this long journey. I would like to say thanks to Millor Fernandes Rosario, who helped me a lot on this journey, but unfortunately died on June 11th, 2016, go in peace, my colleague and friend.

Thanks for the Coordination for the Improvement of Higher Education Personnel (CAPES) for the scholarship granted both in Brazil and abroad by the Doctoral Sandwich Program (PDSE), which allowed me to conduct part of my program in Sweden. Thanks for all members of the AVIAN group of The Department of Physics, Chemistry and Biology (IFM) of Linköping University (LIU) the helped during the development of my project. Particularly to Dr. Carlos Guerrero Bosagna and Prof. Dr. Per Jensen which have accepted me as a visiting researcher in their group in Sweden. Thanks for the valuable help and time that you have spent in the development of my work.

Special thanks to my parents, Nelita Augusta Pértille and Francisco Natal Pértille and my brothers, Cleysson Pértille and Cristiane Antunes Pértille, who supported, heard and advised me personally. A special thanks to the child who delights my days, my niece and goddaughter Helena Pértille Antunes.

Thanks to my friends all over the world. Those I longstanding met and whom I met this last year. Those that even with the distance were always there. From “PELEGOS” group, from Animal Biotechnology Lab (mainly Gregori Rovadoscki, Ribamar Nunes, Gabriel Costa, Thaís Godoy, Karina Yu, Aline Cesar, Andrezza Felício, Mirele Poleti, Jorge Luís F. de Andrade, Ricardo Brassaloti, Priscilla Villela, Sônia Andrade, Horácio Montenegro, Rosy Simas, Marcela Paduan, Nirlei A. Silva, Pilar Mariani), from Embrapa Suínos e Aves group, from Carlitos´s team and the “chicken’s noise performance group” (mainly Amir Fallahshahroudi, Ann-Sofie

Sundman, Ivana Magic, Nina Rogefors, Petros Batakis, Sofia Eleftheriadis, Tom P. Kunjaparambil, Tâmara Noronha Pasqualotto, Alexandre Tessmann), from Linköping people (especially Daniel Häger and “ai dentooo” group), from Rocked Science Curling Team (Carlos-Julia-Amanda, Francesco and Alice). Thanks people from MANIERO´s family (Denise, Francisco, Marcio S2 Ana Paula, Marcelo S2 family), from “tira zica” group, from “Moxstruosidade group” (mainly my co-apartment partners Renato Alves Prioli, the Biro and Adriano Anselmi, the Nôno; brothers of heart), from Bear’s family friends (Vinícius, Cacarol and Mitra), from my whole family, from my Lavras’ family; thank you very much.

Finally, the most special thanks to the person that have been by my side, my beloved wife Alice de Sousa Cassetari, who shared with me the joys, achievements, successes, but also the sorrows and failures. She makes my life happier. Thank you for sharing your life with me.

To finish, I sheltered myself, my evolution, my decisions, my efforts, my commitment, my physical health and my sanity or insanity. Because it´s all changed so much to kept me alive.

EPIGRAPH

“If you are going through hell,
keep going”

Winston Churchill

CONTENT

RESUMO

Desvendando associações genéticas importantes e perfis de metilação diferenciais utilizando sequenciamento reduzido do genoma da galinha

A galinha é um organismo modelo ideal para melhorar o entendimento de diversas áreas da pesquisa como: filogenética, embriologia, biomedicina, pecuária, e tem sido recentemente sugerida como um modelo promissor para estudos em epigenética. Na pecuária, as galinhas são fonte de proteína para os seres humanos e tem sido alvo de seleção para alcançar um alto padrão de produção com base no melhoramento genético tradicional. Mas agora, estamos na era genômica e epigenômica e as atenções devem ser voltadas para o uso de novas ferramentas para melhorar a seleção não só pensando em produção, mas também na saúde e bem-estar dos animais. O uso de abordagens moleculares, tem sido uma ferramenta fundamental para compreender modelos biológicos e melhorar as estratégias de seleção baseadas na informação genômica em programas de melhoramento. Abordagens moleculares, também tem contribuído para a compreensão da história evolutiva desses modelos e os mecanismos genéticos e epigenéticos envolvidos no processo de evolução e diversificação genética das galinhas. Neste contexto, tecnologias evoluíram para produção de dados de sequenciamento de alto rendimento por sequenciamento de próxima geração (NGS). NGS forneceu uma grande quantidade de informação a ser utilizado para diversos fins, como para detectar polimorfismos de nucleotídeo único (SNPs) e perfis de metilação diferencial do DNA em galinhas. NGS tem permitido também o desenvolvimento de painéis de SNP para testes de associações genômica ampla (GWAS) com fenótipos específicos de interesse. Embora NGS tem poder suficiente para detectar polimorfismos informativos, o seu elevado custo o torna impraticável para ser utilizado em GWAS ou estudos de metilação diferencial por sequenciamento de DNA metilado por imunoprecipitação (MeDIPseq). A procura de um método de genotipagem eficiente, simples, econômico e confiável para descoberta, caracterização e validação de SNPs, foi a razão para o desenvolvimento deste estudo. Utilizamos sequenciamento do genoma reduzido por enzima de restrição (RE) que cliva o genoma alvo para identificação de SNPs nestas bibliotecas reduzidas e aplicação deste método em GWAS. Em seguida, para combinar a representação reduzida do genoma com o método MeDIPS, desenvolvemos uma nova abordagem para a realização de estudos de metilação diferencial utilizando as bibliotecas reduzidas. Estes trabalhos permitiram a identificação de SNPs associados com características de desempenho e janelas de metilação diferencial relacionados a diferentes condições de manejo em galinhas.

Palavras-chave: Galinha; GWAS; MeDIPS; Melhoramento animal; Saúde e bem-estar animal

ABSTRACT

Unraveling important genetic associations and differential methylation profiles in chickens using reduced genome sequencing approach

Chickens are ideal model organism to improve understanding of several research areas as phylogenetic, embryology, biomedicine, livestock, and have recently been suggested as a promising model for epigenetic studies. In the livestock area, chickens are source of protein to humans and had been selected to achieve a high production standards based on genetic breeding by the traditional selection. We are now in the genomics and epigenomics era and it is time be concern about the use of new tools to improve selection not only thinking about production, but also in the health and welfare of animals. The use of molecular approaches, have been a fundamental tool to understand biological models and improve selection strategies based on genomic information in breeding programs. Molecular approaches have also contributed to understanding of the evolutionary history of these models and the genetics and epigenetics mechanisms involved in evolution process and genetic diversification of chickens. In this context, many technologies have emerged to produce high-throughput data using Next-generation sequencing (NGS) approaches. NGS provided a large amount of information for diverse purposes such as to detect single nucleotide polymorphisms (SNPs), and methylated DNA profiles in chickens. In addition, NGS has allowed the development of pre-designed SNP arrays for genome-wide association studies (GWAS) with specific phenotypes of interest. Moreover, although NGS has enough power to detect informative polymorphisms, its high cost makes it impractical to be used in GWAS and Methylated DNA immunoprecipitation sequencing (MeDIPseq) studies. The demand for an economical, efficient, simple-step and reliable genome-wide method of SNPs discovery, validation and characterization, was the reason for the development of this study. We applied reduced representation sequence by restriction enzyme (RE) cleavage of target chicken genome to be applied in GWAS. Thereafter, to combine the reduced representation of the genome with MeDIPseq method, we developed a novel approach to perform differential methylation studies using reduced libraries. These works allowed us to identify SNPs associated with performance traits and differential methylation windows related to different stress conditions in chickens.

Keywords: Animal breeding; Animal health and welfare; Chicken; GWAS; MeDIPS

1 INTRODUCTION

Chicken is an ideal model organism for phylogenetic, embryology (BURT, 2007) studies and biomedical research (WU; KAISER, 2011). The chicken protein is known to has low-fat, high unsaturation degree of fatty acids and low sodium and cholesterol levels that responds to the current consumer demand (PETRACCI; CAVANI, 2011). Therefore, the domestic chicken (*Gallus gallus domesticus*) has been one of the main sources of high-quality protein to humans (MIAO et al., 2013). To achieve this high-quality of meat production standards, great advances in nutrition, management and genetic selection in animal breeding programs has been applied (JORGE et al., 2008). In the genetic background, the use of molecular approaches, has been a fundamental tool to understand genes that control traits of commercial interest to improve selection strategies based on genomic information in breeding programs (JORGE et al., 2008). Molecular approaches also have contributed to understanding the evolutionary history of chickens and the genetics and epigenetics mechanisms involved in evolution process and genetic diversification of this species (RUBIN et al., 2010). This understanding is important in a humanitarian context to improve the animal´s needs and their rearing environments (ROSTAGNO, 2009).

To understand the molecular mechanisms governing these interested traits, especially from the last decade, high-throughput data by Next-generation sequencing (NGS) approaches have emerged providing a large amount of information to be used for diverse purposes such as to detect single nucleotide polymorphisms (SNPs) (GROENEN et al., 2009) and methylated DNA profiles (GUERRERO-BOSAGNA, 2013) in chickens. These markers can be responsible for functional alterations in the chicken genome (GHEYAS et al., 2015), or they can be located at neutral genomic regions being fundamental in many gene processes and activities (BIÉMONT; VIEIRA, 2006). Also, NGS has allowed the development of pre-designed SNP arrays, to widespread testing of associations of SNPs with specific phenotypes of interest (KRANIS et al., 2013). However, pre-designed SNP arrays have limited coverage on functionally important genomic regions in experimental populations (LI et al., 2008). Moreover, although NGS has enough power to detect informative polymorphisms, its high cost makes it impractical to be used in animal breeding, genome-wide association studies (GWAS) (PETERSON et al., 2012; DE DONATO et al., 2013) and Methylated DNA immunoprecipitation (MeDIPS) (GUERRERO-BOSAGNA, 2013) studies.

The demand for a high cost-effective genotyping method was the reason for the development of the second chapter presented here that is titled “High-throughput and Cost-

effective chicken genotyping using Next-Generation Sequencing". The idea was to combine an economical, efficient, simple-step and unbiased genome-wide method of SNP discovery, validation and characterization using reduced representation sequence by restriction enzyme (RE) cleavage of target chicken genome (GLAUBITZ et al., 2014; ZHAI et al., 2015). Thus, from a previously described CornellGBS approach in maize (ELSHIRE et al., 2011), we performed a detailed step-by-step description of the complete reproducible protocol optimization based on the reduced chicken genome sequencing using *PstI* RE. From 462 animals genotyped using this protocol, we carried out the preparation of the third chapter that deals with the manipulation of SNP database generated by CornellGBS approach, SNP imputation, validation and GWAS with performance traits. This third chapter was titled "Genome-wide association study for performance traits in chickens using genotyping by sequencing approach".

For our surprise, the results from the second chapter presented here indicates that the CornellGBS approach showed a pattern of SNP profiling that makes it unique in comparison with other approaches. Such profiling includes not only an enrichment of different functional regions, but also a high interrogation of microchromosomes that are CpG-rich regions (MCQUEEN et al., 1996; SMITH; BURT, 1998; SMITH et al., 2000; HABERMANN et al., 2001) and higher gene density than macrochromosomes (SMITH; BURT, 1998; SMITH et al., 2000; HABERMANN et al., 2001). This leads us to strongly consider using this methodology for the development of another studies involving access of methylation profiles of individuals. This was the part of the PhD process from where it was originated the forth chapter of this thesis and that was held at Linköping University (LIU), Sweden.

The forth chapter of this thesis was titled "Epigenetic marks of rearing conditions detected in red blood cells of adult hens". In this paper, we presented one application in animal welfare area, considering that chickens have recently been suggested as a promising model for epigenetic studies (FRÉSARD et al., 2013). Due to the need for sequencing cost reduction and the advantages in reducing the genome with *PstI* RE in the epigenetic scope, we created a new approach that combines the CornellGBS approach, described here and MEDIP approach previously described elsewhere (GUERRERO-BOSAGNA; JENSEN, 2015). This new approach was named GeDI (Genome Digestion) MeDIP (Methylated DNA Immunoprecipitation) sequencing and, it was performed to handle with differential methylation profile among animal reared in different environmental conditions as an ethical issue in the research. The ethical issue of inducing unnecessary stress in animals and detrimental practices in animal industry has consequences from a human health perspective (ROSTAGNO, 2009).

Therefore, we collected samples from chickens submitted to different stress levels generated by these raring conditions (cage-reared chickens *vs* aviary reared-chickens) that had long term effects in the blood epigenome. For that, we investigated epigenetic marks of stress in red blood cells of chickens reared in cages, in which social isolation stress occurs, versus a complex condition of open aviary.

This thesis refers to the optimization of an approach to generate specific genic and CpG enriched profiles of sequenced fragments. This special profile led us the creation of a new method using the reduced representation libraries to epigenetics studies. Therefore, CornellGBS and GeDI MeDIP methods that enriches both, genes and CpGs, allowed us their application in two situations. First, to detect SNPs to be applied in GWAS with performance traits in chicken and second, to detect genomic windows enriched with methylated DNA to perform differential methylation profile in different rearing conditions of chickens.

References

BIÉMONT, C.; VIEIRA, C. Genetics: Junk DNA as an evolutionary force. **Nature**, London, v. 443, n. 7111, p. 521–524, Oct. 2006.

BURT, D.W. Emergence of the chicken as a model organism: implications for agriculture and biology. **Poultry Science,** Oxford, v. 86, n. 7, p. 1460–1471, 2007.

DE DONATO, M.; PETERS, S.O.; MITCHELL, S.E.; HUSSAIN, T.; IMUMORIN, I.G. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. **PloS One**, San Francisco, v. 8, n. 5, p. e62137, Jan. 2013.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKLER, E.S.; MITCHELL, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS One**, San Francisco, v. 6, n. 5, p. 1–10, 2011.

FRÉSARD, L.; MORISSON, M.; BRUN, J.-M.; COLLIN, A.; PAIN, B.; MINVIELLE, F.; PITEL, F. Epigenetics and phenotypic variability: some interesting insights from birds. **Genetics Selection Evolution**, Amsterdam, v. 45, n. 1, p. 16, 2013.

GHEYAS, A.A.; BOSCHIERO, C.; EORY, L.; RALPH, H.; KUO, R.; WOOLLIAMS, J.A.; BURT, D.W. Functional classification of 15 million SNPs detected from diverse chicken populations. **DNA Research: an International Journal for Rapid Publication of Reports on Genes and Genomes**, Oxford, v. 22, n. 3, p. 205–217, 2015.

GLAUBITZ, J.C.; CASSTEVENS, T.M.; LU, F.; HARRIMAN, J.; ELSHIRE, R.J.; SUN, Q.; BUCKLER, E.S. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **PloS One**, San Francisco, v. 9, n. 2, p. e90346, Jan. 2014.

GROENEN, M.A.M.; WAHLBERG, P.; FOGLIO, M.; CHENG, H.H.; MEGENS, H.-J.; CROOIJMANS, R.P.M.A.; BESNIER, F.; LATHROP, M.; MUIR, W.M.; WONG, G.K.-S.; GUT, I.; ANDERSSON, L. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. **Genome research**, New York,v. 19, n. 3, p. 510–519, Mar. 2009.

GUERRERO-BOSAGNA, C. DNA methylation research methods. **Materials and Methods**, v. 3, Jan. 2013. Disponível em: <http://www.labome.com/method/DNA-Methylation-Research-Methods.html>. Acesso em: 08 set. 2016.

GUERRERO-BOSAGNA, C.; JENSEN, P. Optimized method for methylated DNA immuno-precipitation. **MethodsX**, Amsterdan, v. 2, p. 432–439, 2015.

HABERMANN, F.A.; CREMER, M.; WALTER, J.; KRETH, G.; VON HASE, J.; BAUER, K.; WIENBERG, J.; CREMER, C.; CREMER, T.; SOLOVEI, I. Arrangements of macro- and microchromosomes in chicken cells. **Chromosome Research : an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology**, Oxford, v. 9, n. 7, p. 569–584, 2001.

JORGE, E.C.; FIGUEIRA, A.; LEDUR, M.C.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Contributions and perspectives of chicken genomics in Brazil: from biological model to export commodity. **World's Poultry Science Journal**, Edinburg, v. 63, n. 4, p. 597–610, Feb. 2008.

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. Development of a high density 600K SNP genotyping array for chicken. **BMC Genomics**, London, v. 14, n. 1, p. 59, Jan. 2013.

LI, C.; LI, M.; LONG, J.R.; CAI, Q.; ZHENG, W. Evaluating cost efficiency of SNP chips in genome-wide association studies. **Genetic Epidemiology**, New York, v. 32, n. 5, p. 387–395, 2008.

MCQUEEN, H.A.; FANTES, J.; CROSS, S.H.; CLARK, V.H.; ARCHIBALD, A.L.; BIRD, A.P. CpG islands of chicken are concentrated on microchromosomes. **Nature Genetics**, New York , v. 12, n. 3, p. 321–324, 1996.

MIAO, Y.; PENG, M.; WU, G.; OUYANG, Y.; YANG, Z.; YU, N.; LIANG, J.; PIANCHOU, G.; MITRA, B.; PALANICHAMY, M.G.; BAIG, M.; CHAUDHURI, T.K.; SHEN, Y.; KONG, Q.; MURPHY, R.W.; YAO, Y.; ZHANG, Y. Chicken domestication : an updated perspective based on mitochondrial genomes. **Heredity,** London, v. 110, n. 3, p. 277–282, Mar. 2013.

PETERSON, B.K.; WEBER, J.N.; KAY, E.H.; FISHER, H.S.; HOEKSTRA, H.E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. **PLoS One**, San Francisco, v. 7, n. 5, p. e37135, 2012.

PETRACCI, M.; CAVANI, C. Muscle growth and poultry meat quality issues. **Nutrients**, Basel, v. 4, n. 12, p. 1–12, Dec. 2011.

ROSTAGNO, M.H. Can stress in farm animals increase food safety risk? **Foodborne Pathogens and Disease**, Larchmont, v. 6, n. 7, p. 767–776, Sept. 2009.

RUBIN, C.-J.; ZODY, M.C.; ERIKSSON, J.; MEADOWS, J.R.S.; SHERWOOD, E.; WEBSTER, M.T.; JIANG, L.; INGMAN, M.; SHARPE, T.; KA, S.; HALLBÖÖK, F.; BESNIER, F.; CARLBORG, O.; BED'HOM, B.; TIXIER-BOICHARD, M.; JENSEN, P.; SIEGEL, P.; LINDBLAD-TOH, K.; ANDERSSON, L. Whole-genome resequencing reveals loci under selection during chicken domestication. **Nature**, London, v. 464, n. 7288, p. 587–591, Mar. 2010.

SMITH, J.; BRULEY, C.K.; PATON, I.R.; DUNN, I.; JONES, C.T.; WINDSOR, D.; MORRICE, D.R.; LAW, A.S.; MASABANDA, J.; SAZANOV, A.; WADDINGTON, D.; FRIES, R.; BURT, D.W. Differences in gene density on chicken macrochromosomes and microchromosomes. **Animal Genetics**, New York, v. 31, n. 2, p. 96–103, 2000.

SMITH, J.; BURT, D.W. Parameters of the chicken genome (*Gallus gallus*). **Animal Genetics**, New York, v. 29, n. 4, p. 290–294, 1998.

WU, Z.; KAISER, P. Antigen presenting cells in a non-mammalian model system, the chicken. **Immunobiology**, Stuttgart, v. 216, n. 11, p. 1177–1183, nov. 2011.

ZHAI, Z.; ZHAO, W.; HE, C.; YANG, K.; TANG, L.; LIU, S.; ZHANG, Y.; HUANG, Q.; MENG, H. SNP discovery and genotyping using restriction-site-associated DNA sequencing in chickens. **Animal Genetics**, New York, v. 46, n. 2, p. 216–219, Apr. 2015.

2 HIGH-THROUGHPUT AND COST-EFFECTIVE CHICKEN GENOTYPING USING NEXT-GENERATION SEQUENCING

Abstract

Chicken genotyping is becoming common practice in conventional animal breeding improvement. Despite the power of high-throughput methods for genotyping, their high cost limits large scale use in animal breeding and selection. In the present paper we optimized the CornellGBS, an efficient and cost-effective genotyping by sequence approach developed in plants, for its application in chickens. Here we describe the successful genotyping of a large number of chickens (462) using CornellGBS approach. Genomic DNA was cleaved with the *PstI* enzyme, ligated to adapters with barcodes identifying individual animals, and then sequenced on Illumina platform. After filtering parameters were applied, 134,528 SNPs were identified in our experimental population of chickens. Of these SNPs, 67,096 had a minimum taxon call rate of 90%. Interestingly, 20.7% of these SNPs have not been previously reported in the dbSNP. Moreover, 92.6% of these SNPs were concordant with a previous *Whole Chicken-genome re-sequencing* dataset used for validation purposes. The application of CornellGBS in chickens showed high performance to infer SNPs, particularly in exonic regions and microchromosomes. This approach represents a cost-effective (~US$50/sample) and powerful alternative to current genotyping methods, which has the potential to improve whole-genome selection (WGS), and genome-wide association studies (GWAS) in chicken production.

Keywords: Chicken; GBS; GWAS; Next-generation sequencing; *PstI*; Restriction enzyme

2.1 Introduction

Next-generation sequencing (NGS) analyses have been increasingly employed in production animals, particularly in chickens. NGS generates large amounts of genomic information that can be used to detect genetic variants related to functional alterations (GHEYAS et al., 2015). Single Nucleotide polymorphisms (SNPs) are the most abundant type of molecular markers and their high genomic density facilitates their interrogation by different genetic approaches. These include large-scale genome association analyses, genetic analysis of simple and complex disease states, and population genetic studies (BROOKES, 1999).

The use of NGS has enabled to identify SNPs across genomes and allowed the development of pre-designed SNP chips for widespread testing of SNP associations with specific phenotypes of interest (KRANIS et al., 2013). However, pre-designed SNP chips have limited coverage on functionally important genomic regions in experimental populations. SNP chips generally contain a limited number of SNPs in coding or regulatory regions, rarely contain SNPs with significant effects (LI et al., 2008), and include non-polymorphic SNPs, which difficults tracking their inheritance in specific pedigrees (BURT, 2004). On the other hand, although NGS has enough power to detect informative polymorphisms, its high cost makes its

use impractical in animal breeding and genome-wide selection (ELSHIRE et al., 2011a; GLAUBITZ et al., 2014).

The use of an economical, efficient, and simple-step method of SNPs discovery, validation and characterization that uses reduced representation sequencing generated by restriction enzyme cleavage of target genomes can provide an unbiased genome-wide set of SNP markers in different genomes (GLAUBITZ et al., 2014), including chickens (ZHAI et al., 2015). Reduced representation methods can be grouped in three classes: (1) reduced-representation sequencing, which includes methods such as reduced-representation libraries (RRLs) and complexity reduction of polymorphic sequences (CRoPS); (2) restriction-site-associated DNA sequencing (RAD-Seq); and (3) low coverage genotyping, which includes methods such as multiplexed shotgun genotyping (MSG), genotyping by sequencing from Cornell (CornellGBS) (DAVEY et al., 2011), and genome reducing and sequencing (GGRS) (LIAO et al., 2015). Of these reduced representation methods, RAD-Seq (ZHAI et al., 2015) and GGRS (LIAO et al., 2015) have been employed in chickens (DAVEY et al., 2011). The possibility to reduce the genome complexity using restriction enzymes that generate DNA fragments of specific ranges (KUMAR; YOU; CLOUTIER, 2012) expedite re-sampling and produces coverage levels that are acceptable for solid SNP calling (ALTSHULER et al., 2000).

CornellGBS is a simple reproducible method based on the Illumina sequencing platform (DE DONATO et al., 2013) that requires low input of DNA (100 ng). This method allows for a highly multiplexed approach, which is achieved through the incorporation of unique barcodes that identify individual samples in a DNA pool to be sequenced. This approach avoids the low sequence diversity in which the restriction enzyme overhangs appear at the same position in every read, by employing barcodes of variable lengths (DAVEY et al., 2011). In addition to the methodological simplicity of simultaneously discovering and characterizing polymorphisms, the availability of an open-source analysis tool is a major advantage of the CornellGBS approach (DE DONATO et al., 2013). This methodology is currently being successfully applied in numerous species by a large number of researchers (GLAUBITZ et al., 2014). However, to the best of our knowledge this method has not been applied in chicken.

The present study aims at constructing reduced genome representation sequencing libraries using the CornellGBS approach in chickens. In order to optimize the use of CornellGBS in chickens, cleavage of the chicken genome was tested with two different restriction enzymes, *PstI* and *SbfI*. Two different experimental animal populations were used in the present study: 444 chickens from five families of the EMBRAPA F_2 Chicken Resource Population (Concórdia, SC, Brazil), 8 chickens from the F_1 generation and 18 chickens from

the parental line (F0). In the present article, we have optimized the use of CornellGBS in chickens, which was achieved in part by using the *PstI* restriction enzyme for genomic cleavage. We also provide a new set of chicken SNPs that were detected by using this approach. The application of this methodology will open many possibilities for downstream applications in chickens and facilitate SNP discovery in specific populations of chickens. The relevance of applying a cost-effective genotyping method in chickens is enormous, given their world-wide economic relevance as production animal (UNITED STATES DEPARTMENT OF AGRICULTURE - USDA, 2015).

2.2 Methods

All experimental protocols employed in the present study that relate to animal experimentation were performed in accordance with the resolution number 010/2012 approved by the Embrapa Swine and Poultry Ethics Committee on Animal Utilization, in order to ensure compliance with international guidelines for animal welfare.

2.2.1 Sample selection and preparation

This study was conducted using 464 chickens from an experimental population originated and maintained at the dependencies of the Brazilian Agricultural Research Agency, from EMBRAPA; Concórdia, SC, Brazil. The population includes 446 chickens from five F_2 families of the EMBRAPA F_2 Chicken Resource Population, 10 chickens from their parental lines (5 from each line), and 8 chickens from the F_1 generation.

The F_1 generation individuals were originated from a cross between a parental broiler line (TT) and a layer line (CC), both developed at EMBRAPA. To generate the F_2 population (TCTC), one F_1 male (TC) and three F_1 females (TC) were selected from different F_1 families and were randomly mated with non-related animals. A more detailed description of the population has been previously provided (NONES et al., 2006; ROSÁRIO et al., 2009).

Genomic DNA was extracted from blood samples following proteinase K digestion (Promega), DNA precipitation in absolute ethanol, DNA washing in 70% ethanol and resuspension in ultrapure water. DNA samples were quantified in a fluorometer (Qubit® Fluorometric Quantitation). Sample quality was assessed using the Nanodrop®2000c spectrophotometer and DNA integrity was checked in 1% agarose gel.

2.2.2 Restriction enzymes selection and adapters design

In silico cleavage of DNA with *PstI* and *SbfI* was performed in R using the following Bioconductor (HUBER et al., 2015) packages: *Biostrings, BSgenome.Ggallus.UCSC.galGa14, plyr*, *ggplot2*, *reshape2* and *scales* (https://github.com/) (see Supplementary Fig. S1 online). The *in silico* cleavage was used to generate a dataset of fragments mapped against the galGal4 genome. The dataset of fragments that are predicted to be generated after *in silico* genomic cleavage with *PstI* was named 'Predicted *PstI-Tags'*. The dataset of fragments that are predicted to be generated after *in silico* genomic cleavage with *SbfI* was named '*Predicted SbfI-Tags'*. The dataset of fragments that were obtained from the *in vitro* cleavage of the DNA from all the 462 individuals analyzed was named 'Sequenced *PstI-Tag'* and was generated using sam2bed from BEDOPS v2.4.15 tool. All the fragments either from *in silico* or *in vitro* analyses were aligned against the chicken reference genome (*Gallus gallus 4.0*, NCBI).

We also performed *in vitro* genomic cleavage of chicken DNA samples with the abovementioned restriction enzymes (see Supplementary Fig. S2 online), according to the New England BioLabs® manufacturer´s protocol.

The adapters were designed using the GBS Barcode Generator tool (Deena Bioinformatics) taking into consideration the barcode sequence, in order to maximize the balance of the bases at each position in the defined set (ELSHIRE et al., 2011a).

2.2.3 Preparation of sequencing libraries

After *PstI* digestion, adapters were linked to the cohesive ends of the digested DNA with T4 DNA ligase (New England BioLabs®). Approximately 24 samples were polled and purified using *QIAquick PCR Purification Kit*® (Quiagen). The fragments of each library were amplified by PCR using specific primers for sequencing in the Illumina platform. The purification of PCR reactions was performed using the *Agencourt AMPure XP PCR purification kit*® (Beckman Coulter) (see Supplementary Fig. S3 online). Each library was quantified by quantitative PCR using the *KAPA Library Quantification Kit* (KAPA Biosystems). Two pools of ~24 samples containing equal concentration of DNA were sequenced per flowcell lane totaling ~48 samples sequenced with different barcodes in each flowcell lane. Sequencing libraries were diluted to 16 pM and clustered using the cBOT (Illumina) equipment. Paired-end sequencing with a read length of 100 bp was performed using the HiSeq2500 instrument from Illumina. For the analysis we used the HiSeq Illumina real-time analysis (RTA) software v1.18.61 update. This software

generates a color matrix for the correction of the reads. This is important because HiSeq sequencer uses different lasers to detect G/T and A/C nucleotides. In each cycle, at least one of two nucleotides for each color channel must be read in order to maintain the color balance for each base in the index read sequenced. With this upgrade the color matrix still uses the first four cycles to generate data, like the last version of RTA. However, in the current version the initial matrix is discarded after the template generation is complete. Then, the first 11 cycles of intensity data are used for final estimation of the correction matrix. In order to minimize the issues related to the construction of this matrix, we optimized our protocol using barcodes larger than 4 bp to avoid imbalance between the first bases. The complete laboratory procedures are provided in Supplementary Data S1 online.

2.2.4 Sequence processing

Quality trimming was performed in short sequences with SeqyClean tool v. 1.9.10 (ZHBANNIKOV, 2013) using a Phred quality score ≥24 and a fragment size ≥50. The quality of the *reads* was checked before and after the cleaning by FastQC v.0.11.3 (ANDREWS, 2010).

The Tassel v.3.0 program was used to process the data (GLAUBITZ et al., 2014). For each sample stored in a FASTQ file there is one identification map key file. This key file has the matching information for the sample, flowcell and lane. The reads that begin with one of the expected barcodes (found in the key map) are immediately followed by the expected cut site remnant (CTGCA for *PstI*). Fragments are then trimmed to 64 bases and grouped into a single list called "master" by the TASSEL-GBS Discovery Pipeline.

2.2.5 Alignment and Genetic variants identification

The alignment of quality-trimmed reads was performed using Bowtie2 tool v.2.2.5 (LANGMEAD; SALZBERG, 2012) against the current chicken reference sequence (*Gallus_gallus* 4.0, NCBI). The aligned reads were then imputed in the Tassel v.3.0 default pipeline (GLAUBITZ et al., 2014) for SNP identification. We filtered the polymorphisms initially identified based on the sequencing quality criteria and on the bases identified. The following filters were applied: i) minimum taxon call rate (mnTCov) of 20%, which is a minimum SNP call rate for a taxon to be included in the output, with the call rate being the proportion of the SNP genotypes for a non-N taxon (where N=missing); ii) minimum site coverage (mnScov) of 90%, which is a minimum taxon call rate for a SNP to be included in the

output, with the taxon call rate being the proportion of the taxa with non-N genotypes for that SNP; iii) mismatch rate (misMat) of 5% to minimize the appearance of duplicated SNPs; iv) minimum minor allele frequency (mnMAF) of 0.01. A more detailed description of the default filters has been provided by Glaubitz et al. (2014).

The coverage depth of the "unique sequence tags" file was determined using Samtools v.0.1.19 (LI et al., 2009) with the "depth" option.

2.2.6 Genotyping methods comparison and CornellGBS data validation

We compared the chromosomal positions of the SNPs obtained using the CornellGBS approach with the positions obtained using the following SNP platforms for chickens: Illumina Chicken 60K Beadchip (GROENEN et al., 2011) and 600K HD Affymetrix®Axion® genotyping array for chicken (KRANIS et al., 2013). Bioconductor (HUBER et al., 2015) (GEOquery) and CRAN (data.table, rdrop2 and reshape) repository packages for R were used for the bioinformatics analysis. We validated our method comparing the SNPs obtained (59,205) against a SNP dataset of *Whole Chicken-genome re-sequencing* (WCGR) dataset (BOSCHIERO et al., unpublished results) previously generated with Illumina sequencing with ~11X of sequencing coverage. This dataset contained 12,357,602 filtered SNPs and was generated from the same 10 chickens used in this study (TT and CC parental lines). The comparison between these two datasets was performed using CRAN (data.table and reshape2) repository packages for R. More details of the sequencing process of WCGR SNP data can be found in recent publications (GODOY et al., 2015; MOREIRA et al., 2015).

2.2.7 Functional annotation

The set of unique SNPs obtained from 462 chickens using the Tassel v.3.0 tool was annotated using the Variant Effect Predictor (VEP) tool v.71 (MCLAREN et al., 2010). The SIFT (sorting intolerant from tolerant) scores for the SNPs (NG; HENIKOFF, 2003) were used to predict whether a substitution of an amino acid affects protein function, which is based on sequence homology and the physical properties of amino acids. If the SIFT score lies at or below the 0.05 threshold, the substitution causing the amino acid was considered non tolerated.

2.2.8 Mendelian inheritance of the SNPs

The Mendelian error testing was performed using SNP & Variation Suite v8.4 (BOZEMAN, 2016).

2.2.9 Genetic map construction

SNPs present in all families were filtering using Tassel program (GLAUBITZ et al., 2014). A pseudo-testcross population was used to construct the F_1 linkage map. For the linkage analysis, the SNPs were first tested against the expected segregation ratio. The informative genotypes combination were selected for the map construction. Markers with significant segregation distortion ($P < 0.001$, $\chi 2$ test) were removed.

The genetic map was constructed using R/OneMap package (MARGARIDO; SOUZA; GARCIA, 2007) and JoinMap v.4.1 (VAN OOIJEN, 2011). The R/OneMap was used to join the markers in the linkage groups (LGs). The minimum LOD values of 8 and a maximum recombination fraction of 0.35 were used to organize the markers in each LG with the regression mapping algorithm and the Kosambi mapping function (KOSAMBI, 1943).

The R/OmicCircus package (HU et al., 2014) was used to plot the relationship between the chromosomal and linkage marker groups formed by the abovementioned genetic map

2.3 Results

2.3.1 Enzyme selection and library fragment size distribution

The selection of the appropriate restriction enzyme was based on relevant literature information and took into consideration the number of expected fragments, the fraction of the diploid genome sampled, and the expected number of reads required to obtain a sequencing coverage of ~7X of sequencing coverage (PETERSON et al., 2012b). The library complexity depends on the relation between the enzyme selected and the species' genome under investigation. Therefore, the level of DNA methylation sensitivity and recognition site size in relation to the genome under study had to be tested (POLAND; RIFE, 2012).

We initially selected two enzymes that are insensitive to *dam*, *dcm* and *CpG methylation* according to the manufacturer (NEB BioLabs). These enzymes were *PstI* and *SbfI*. *In silico* cleavage of the chicken genome was performed with both *PstI* and *SbfI* enzymes. Genome

cleavage with *PstI* generated 811,951 fragments, while *SbfI* generated 45,116 fragments. Fragment size distribution obtained with *PstI* showed a low amount of discreet size, which is indicative of low repetitive fragments (DE DONATO et al., 2013) (Figure 2.1). Each enzyme generated a different distribution of fragment lengths across the entire genome.

Fragments ranging between 200-500 bp were generated and mapped against chromosome locations in the chicken genome (galGal 4; GGA). *PstI in silico* digestion generated 159,673 fragments, which were evenly distributed across all the chromosomes, while *SbfI* cleavage generated 1,186 fragments (Supplementary Fig. S1 online). There is a tendency with the *PstI* cleavage to generate clusters with similar range of fragment lengths, particularly between 200 and 500 bp, which is an appropriate length for sequencing by the HiSeq Illumina platform (QUAIL et al., 2012). Cleavage with *SbfI*, however, generated fragments of a variety of sizes and in lower quantity compared to *PstI* (134.6 times less then *PstI* cleavage) (Figure 2.1). Importantly, 40% of the fragments generated by *Sbfl* are outside the range showed on Figure 2.1, representing fragments larger than 15 Mbps.

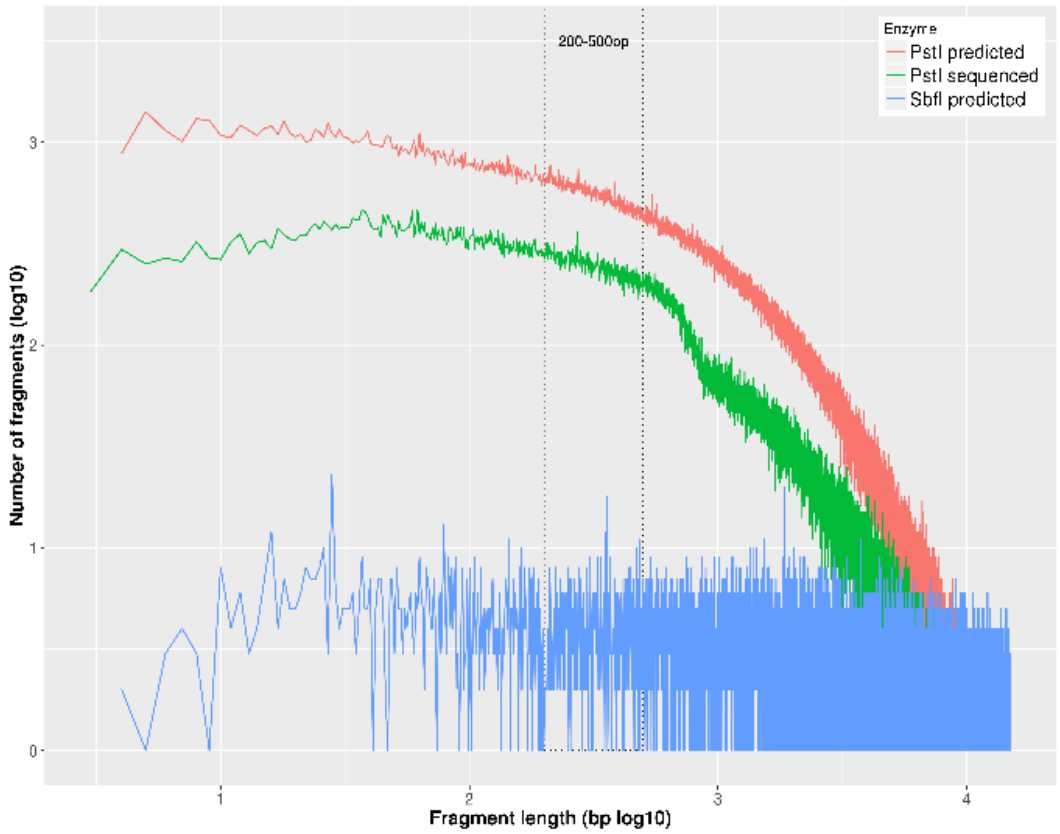


Figure 2.1 - Comparison of patterns of genomic cleavage using *PstI* or *SbfI* restriction enzymes. For cleavage with *PstI* both the predicted (*in silico*) and the obtained pattern after sequencing are shown. Only the predicted (*is silico*) pattern of cleavage is shown for *SbfI* since the pattern generated did not satisfy the requirements for being used in the CornellGBS. The region framed with dashed lines contain fragments in the 200-500 bp length range, which is the range of interest for further Illumina sequencing

We also performed a comparison between the mapping of fragments (*tags*) generated by the *in silico* cleavage (*Predicted PstI-Tags*) and the *tags* generated after the *in vitro* cleavage (*Sequenced PstI-Tags*) of 462 individuals (Figure 2.1). In both cases the tags were aligned against the chicken reference genome (*Gallus gallus 4.0,* NCBI). The number of Predicted *PstI*-Tags obtained was 811,951, while the Sequenced *PstI*-Tags obtained were 287,819. Detailed information on the size categories of the Predicted *PstI*-Tags that were actually sequenced is provided in Supplementary Table S1 online.

Agarose gel electrophoresis of the chicken genomic DNA digested with the *PstI* and *SbfI* restriction enzymes revealed a more efficient cleavage with *PstI* (see Supplementary Fig. S2 online).

2.3.2 Sequencing and alignment

The 48-plex *Pst*I-digested libraries were run in 10 lanes of Illumina flow cells. Approximately 3.6 x 10^9 short reads (100 bp) were generated. After quality trimming by the SeqyClean tool (ZHBANNIKOV, 2013) approximately 1.8 x 10^9 reads (52%) were retained. A high number of short fragments (< 50bp) sequenced were eliminated in the fragment size filtering (37%), as well as contaminants (11%). Approximately 1.4 billion reads were retained after application of the *Tassel* filter (*reads* > 64 bp and properly identified with barcodes).These reads were distributed at an average of 145.6 (±26.5) million reads per lane (Figure 2.2). These results represent 3.1 (±1.7) million reads per individual chicken, of which 3.0 (±1.7) million reads were successfully mapped (plots including read count per animal are provided in Supplementary Figure S4 online).

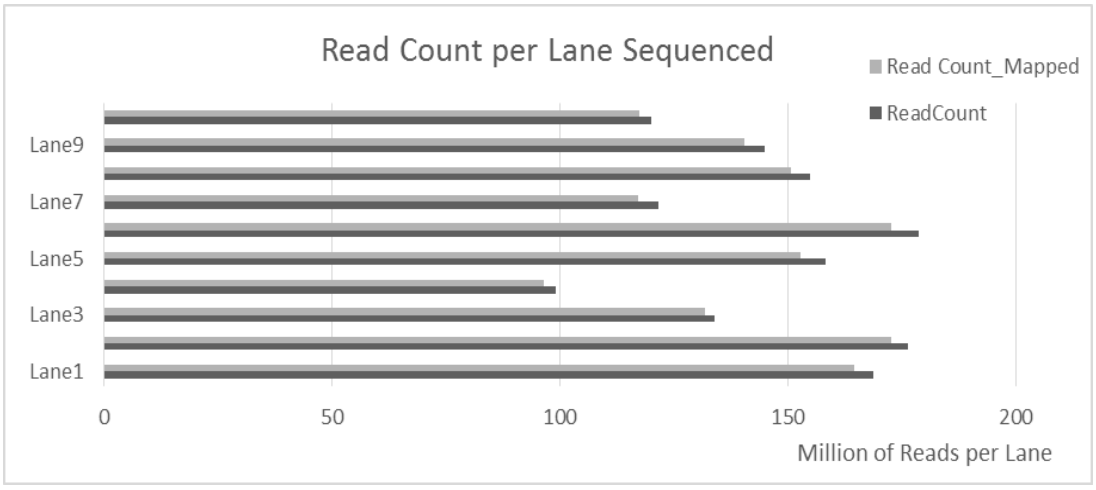


Figure 2.2 - Distribution of the number of sequenced reads counted and mapped per flowcell lane

The number of unique sequence tags (from 464 individuals altogether) that aligned against the chicken reference genome (*Gallus gallus 4.0*, NCBI) was ~5.4 million and 92.8% of them could be mapped. The average sequencing coverage depth was ~264 reads per tag (locus) in these ~5.4 million unique tags.

These ~5.4 million unique sequence tags represent a 4.66% coverage of the whole chicken-genome (~50 million bp). The average coverage for the 464 chickens was 5.6 X for the targeted regions.

2.3.3 SNP discovery

From these ~5.4 million unique sequence tags, 327,240 SNPs were identified considering a minimum minor allele frequency (mnMAF) of 1%. Two of the 464 individuals showed a minimum taxon coverage (mnTCov) of less than 20% and were eliminated from the analysis. The minimum site coverage (mnScov) filter was used to evaluate the taxon call rate. The number of remaining *PstI*-derived SNPs was 134,528 after applying a mnScov filter of 70%, and 67,096 SNPs after applying a mnScov filter of 90%. After filtering with a mnScov of 90% the average taxon call rate per individual was 97% (Figure 2.3).

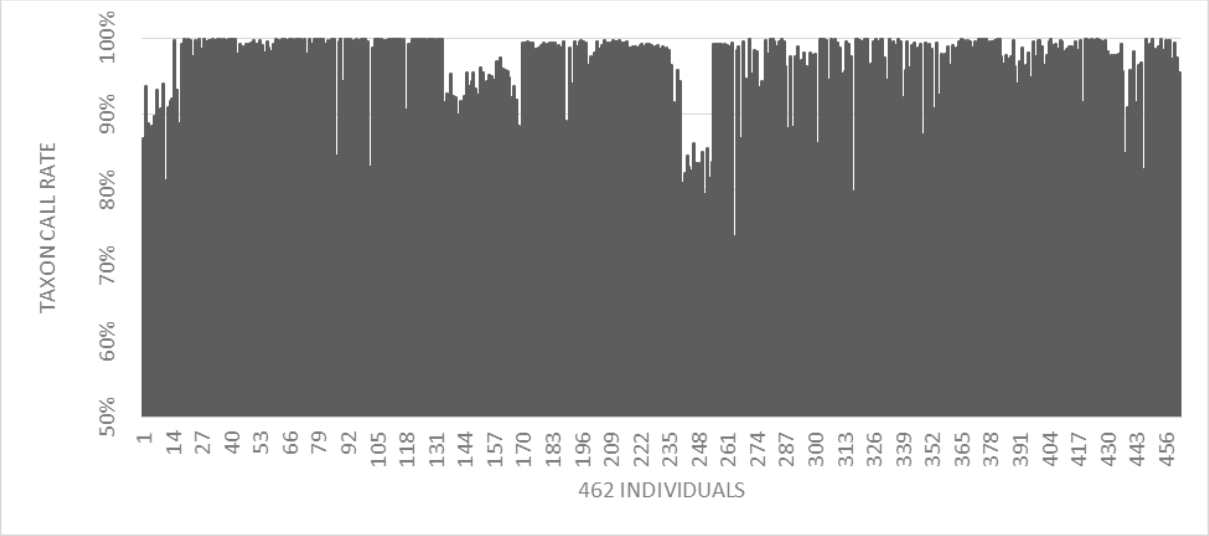


Figure 2.3 - Distribution of the 462 taxon call rates representing the percentage of total SNPs called. The x-axis represents the 462 individuals (taxon) and the y-axis represents the taxon call rate

Additionally, we also tested a mnMAF of 0.05, which generated 300,777 SNPs, as well as a combination of mnTCov of 20% and mnScov of 90%, which generated 61,618 SNPs.

2.3.4 Comparison of genotyping methods and CornellGBS validation

When considering a mnMAF of 1% and mnScov filter of 90% the *PstI*-derived SNPs are shown to be separated by a distance of 15 Kb in average, with a median of 55 bp. This indicates clusters of SNPs in regions represented by the cleaved areas (tags). A comparison among the different genotyping methods is shown on Table 2.1, Figure 2.4 and Supplementary Table S4 online. The distances between SNPs ranged between 1 bp - 1.8 Mb (Table 2.1), and the majority of the SNPs were separated by distances <1 Kb (Figure 2.4).

Table 2.1 - Basic statistical parameters of SNPs distance. SNPs were mapped against chromosome locations of the chicken genome after being detected with 600K Affymetrix, CornellGBS or 60K Illumina

Platform	Mean (Kb)	Median (Kb)	Min (bp)	Max (Mb)
600K Affymetrix	1.7	1.2	1	1.4
CornellGBS	15	0.05	1	1.8
60K Illumina	21	18	1	2

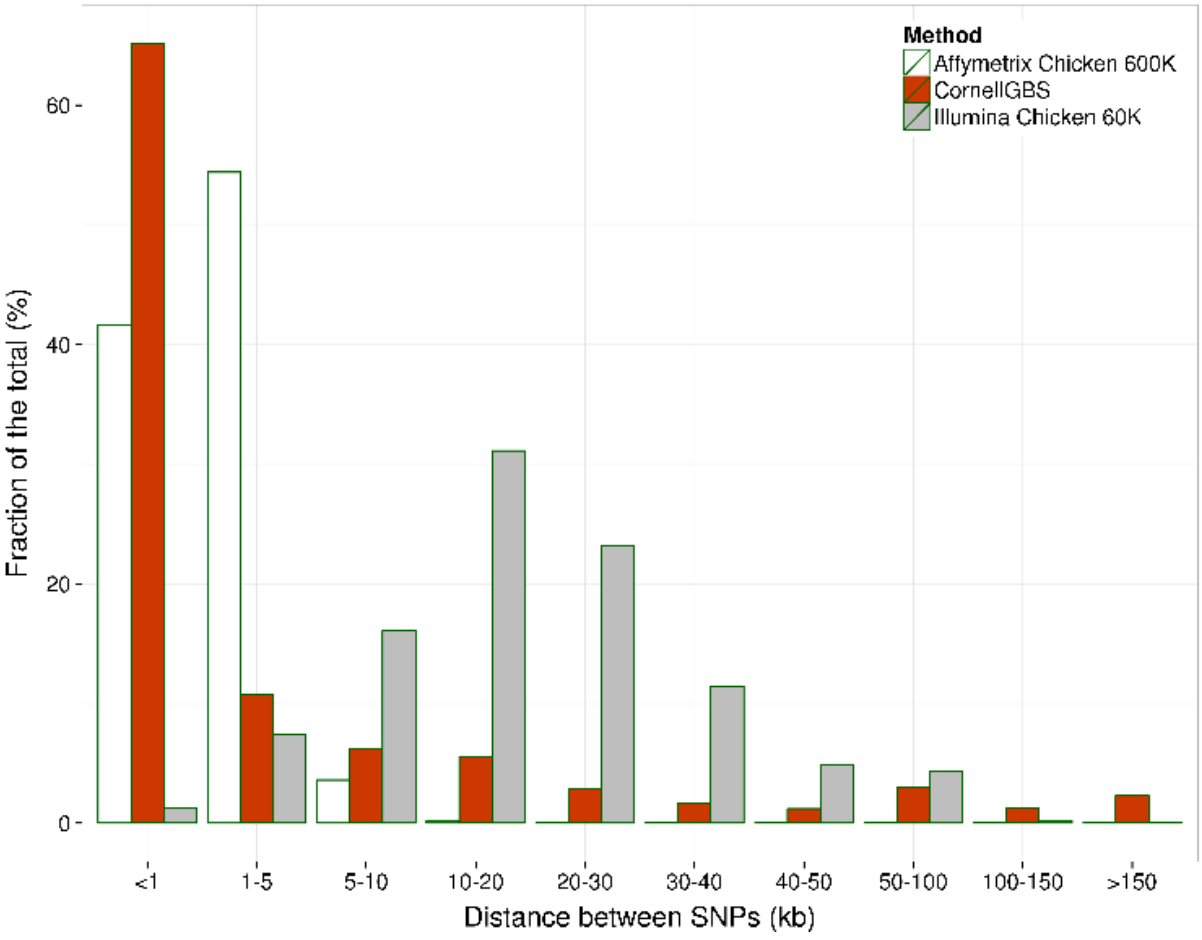


Figure 2.4 - Distribution of distance ranges between SNPs. SNPs were mapped against chromosome locations of the chicken genome after being detected with Affymetrix 600K, CornellGBS or Illumina Chicken 60K Beadchip. The *x*-axis represents the distances between adjacent SNPs (Kb) and the *y*-axis represents the fraction of the total SNPs called

Differences were found between SNP numbers and density (SNP/Mbp) inferred by the three methods used for SNP detection, namely Affymetrix 600K, Cornell GBS and Illumina chicken 60K bead chip (Supplementary Table S4 online). In order to test for differential representation of the SNPs obtained across the chromosomes, the chicken genome was divided into three categories: large chromosomes (GGA1-5, Z), corresponding to ~68% of the chicken genome; medium-size chromosomes (GGA6-10) corresponding to 15% of the chicken genome, and microchromosomes corresponding to 17% of the chicken genome (HABERMANN et al., 2001). The representation of SNPs in each chromosomal category is shown in Table 2.2.

The set of 67,096 SNP chromosomal positions obtained with the CornellGBS (mnMAF 1% and mnScov 90%) was compared to the 12,357,602 filtered SNPs from a WCGR dataset in order to perform a validation of the method since both sets were obtained from the same 10 animals (TT and CC lines). The SNPs with more than one alternative allele (less frequent) were eliminated from this analysis. A concordance of 83.91% (49,680) in the chromosomal positions of the SNPs detected was observed between the two methods. We found that 92.64% of these concordant markers had concordant genotypes between CornellGBS and WCGR datasets. Also, the consistency in the calls of heterozygosity was tested between these two approaches. This test was performed due to the general assumption that reduced representation methods, like CornellGBS, have limitations in the calling of heterozygous SNPs (GLAUBITZ et al., 2014). It was observed that 71.32% of all heterozygous SNPs evaluated here (149,741 genotype comparisons) were validated against the WCGR dataset. However, 86.88% of the non-concordant genotypes occurred because the CornellGBS considered the genotype as homozygotic, and WCGR as heterozygtic. In addition, we found that when both methodologies were able to call heterozygous (106,906 genotype comparisons), 99.90% of the genotypes were in agreement. Interestingly, the number of heterozygous calls in the region assessed was similar between the CornellGBS (112,435) and the WCGR (144,112) approaches, corresponding to 24.15% and 29.18%, respectively, of all the genotype comparisons.

Table 2.2 - Proportion of SNPs detected in each chromosomal size category after using three different genotyping platforms: 600K Affymetrix, CornellGBS and 60K Illumina

Platform	**Large GGA%**	**Medium GGA%**	**Micro GGA %**
600K Affymetrix	54.81	17.00	28.19
CornellGBS	33.50	16.37	50.13
60K Illumina	50.30	17.00	32.70

2.3.5 Homozygous and heterozygous genetic variants

Out of 31 million possible genotypes (462 taxon x 67,096 sites), the proportion of heterozygous SNPs was 31%, with 3.1% being missing data (see Supplementary Table S2 online). The average heterozygosity observed ranged between 9.7 - 48.5%, with 18% of coefficient of variation (CV).

A lower proportion of heterozygous SNPs was found in both parental lines CC (0.20±0.01) and TT (0.26±0.01), followed by the F_2 (0.31± 0.05) and the F_1 generations (0.32±0.10) (Table 2.3). The F_1 generation had the highest CV due to the fact that it represents a heterozygous population. The family F_2-7816 had a higher CV (25%) compared with the other F_2 families due to the low heterozygous call rate for some individuals (25 from 94) in this family.

Table 2.3 - SNP heterozygosity of the genotyped populations (parental, F_1 and F_2 generations)

Population	Number of individuals	Number Heterozygous SNPs	Proportion heterozygous (SD)	CV
Paternal CC	5	11888	0.20 (±0.01)	3%
Paternal TT	5	15244	0.26 (±0.01)	4%
F_1	8	20101	0.32 (±0.10)	30%
F_2-7765	72	21658	0.33 (±0.03)	8%
F_2-7810	82	20323	0.31 (±0.03)	11%
F_2-7816	94	22016	0.34 (±0.09)	25%
F_2-7971	100	18865	0.29 (±0.05)	16%
F_2-7978	96	19982	0.30 (±0.03)	11%

2.3.6 Functional Annotation

The unique set of 67,096 *PstI*-derived SNPs (after filtering) from the 462 chickens were annotated against the known genes from the ENSEMBL database (see the graphical representation of SNPs distributed in genic and intergenic regions of the chicken genome on Figure 2.5). Among the variants found, 20.7% (13,918) were new, while 79.3% (53,178) were already described. Functional annotation of these novel SNPs was performed using the chromosomal positions of the most recent update of chicken genome (*Gallus gallus 4.0,* NCBI) as a reference. The results are available in the supplementary materials (Supplementary Spreadsheet S1).

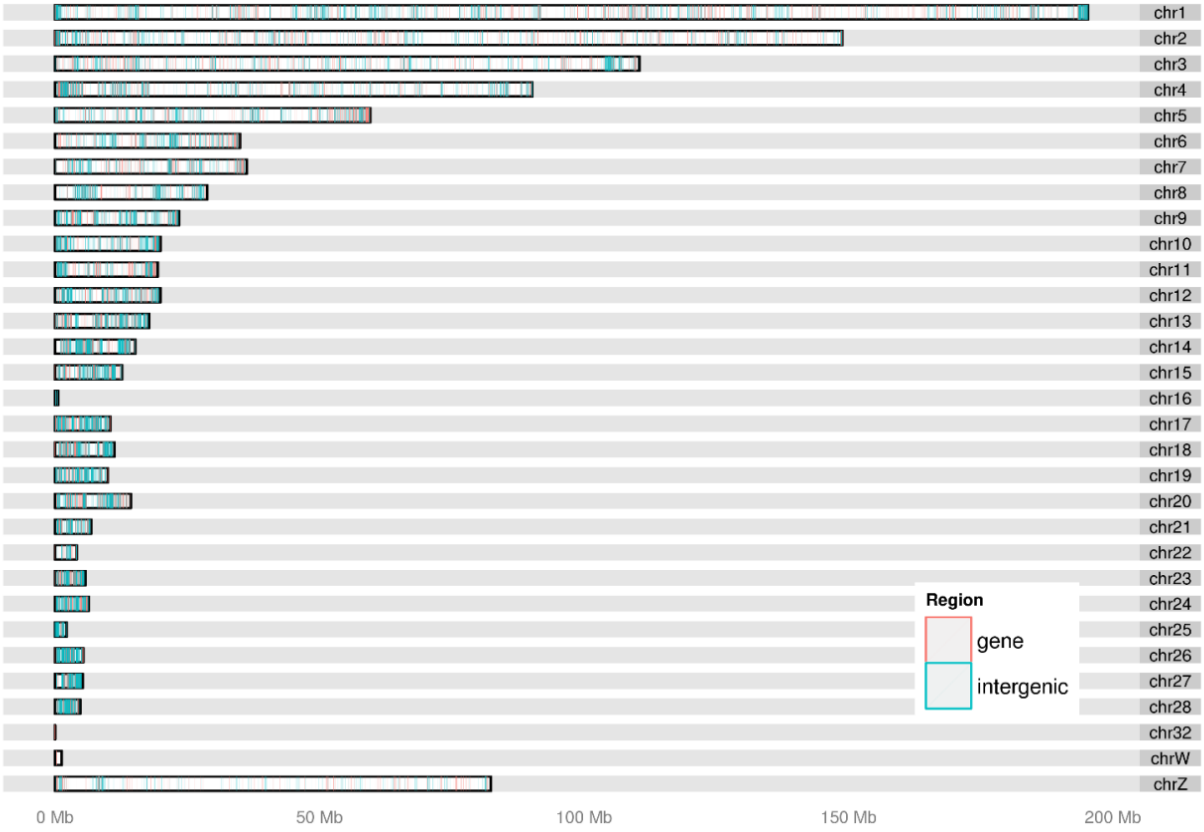


Figure 2.5 - Karyotype of the SNP distribution in genic (red) and intergenic (blue) regions of the *Gallus gallus* genome. The *x*-axis represents the chromosome size (Mbp). The *y*-axis represents the chromosomes

From these 67,096 *PstI*-derived SNPs, 11,372 SNPs had multiple annotations (totalizing 78,399 annotations) as they could be considered into multiple variant classifications (Table 2.4). The non-synonymous SNPs were analyzed by the SIFT algorithm, which predicts whether genetic variants can affect protein function. This is performed by assessing the level of conservation in homologous protein sequences (NG, 2003). The program predicted the SIFT score for 650 SNPs from the 907 non-synonymous SNPs. From these 650 SNPs, 155 SNPs (23.8%) were non-tolerated variants (SIFT score ≤ 0.05) (see Supplementary Table S3 online).

Table 2.4 - Annotation results of the complete set of 67,096 PstI-derived SNPs (after filtering) obtained after genotyping 462 chickens

Variants	Total no.	%
All variants	78,399	100
Intronic	28,181	35.95
Intergenic	22,116	28.21
Exonic	2,590	3.30
Splicing	256	0.11
ncRNA	6	0.01
5'-UTR	268	0.34
3'-UTR	1,328	1.69
Upstream (1kb)	11,516	14.69
Downstream (1kb)	12,306	15.70
Exonic		
Synonymous	1,671	64.52
Non-synonymous	907	35.02
Startlost	5	0.19
Stopgain	3	0.12
Stoplost	4	0.15

2.3.7 Mendelian inheritance of the SNPs

In addition to the SNP validation we also tested for Mendelian errors in the markers obtained in each population used in this study. This test was performed in the complete dataset of 67,096 *PstI*-derived SNPs, as well as in the subset of 13,543 novel SNPs. The results are shown in Table 2.5.

Table 2.5 - Assessment of Mendelian errors in the dataset of 67,096 *PstI*-derived SNPs (after filtering) and in the subset of 13,434 novel SNPs identified. Results are shown separately for each generation of animals studied and for the different families within the F_2 generation population

Family	Nº of Individuals	Mendelian Errors	% of markers	Mendelian Errors (novelSNPs)	% of markers
F1	8	6,488 ± 2,835	9.7	1,313 ± 567	9.7
F2	444	5,947 ± 1,169	8.9	1,216 ± 609	9.0
F2-7765	72	5,060 ± 1,527	7.5	1,039 ± 633	7.7
F2-7810	82	5,464 ± 1,084	8.1	1,124 ± 695	8.3
F2-7816	94	7,872 ± 919	11.7	1,575 ± 757	11.6
F2-7971	100	5,125 ± 471	7.6	1,075 ± 907	7.9
F2-7978	96	6,212 ± 918	9.3	1,268 ± 71	9.4
TotalAnalysed	452	67,096		13,543	

2.3.8 Genetic map construction

We performed a linkage analysis in which the SNPs were tested against the expected segregation ratio. Three genotype combinations in the parental lines were informative for the construction of a genetic map: two combinations when one parent was heterozygous and the other was homozygous (AA × AB or AB × AA) and one combination when both parents were heterozygous (AB × AB). The SNPs following each of these segregation patterns in the parents were retained and markers with significant segregation distortion ($P < 0.001$, $\chi 2$ test) were removed from the map construction. A total of 6,037 SNPs were retained for linkage map construction after filtering, with 387 of these SNPs being classified as female heterozygous, 2,143 SNPs classified as male heterozygous, and 3,507 SNPs classified as heterozygous in both genders.

From the retained 6,037 *PstI*-derived SNPs, 5,982 generated 53 linkage groups (LG) that corresponded to the chromosomes GGA1-28 and Z (see Supplementary Fig. S5). We had no informative markers for chromosomes GGA32 and GGAW LGs. From these 5,982 SNPs that originated LGs, 5,842 markers formed 29 non-fragmented LGs, i.e. markers in agreement with their respective described chromosomes (shown in the physical map, Figure 2.6). Of the remaining markers, 140 formed fragmented LGs, while 55 were considered unlinked. Within these 29 LGs originated, 73 markers were in disagreement with their respective LGs (Figure 2.6).

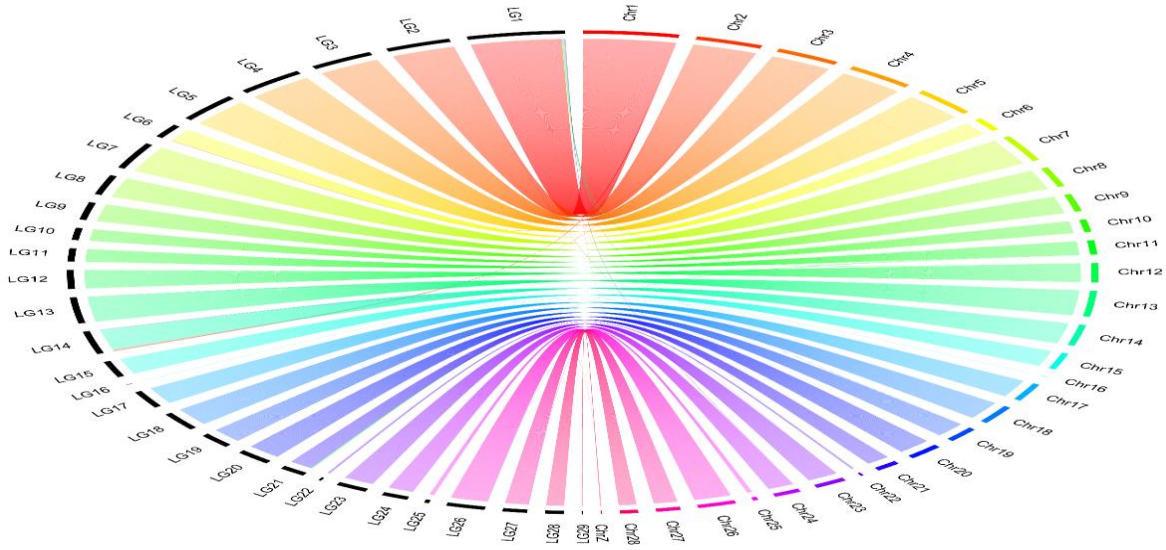


Figure 2.6 - Whole-genome synteny between the physical maps obtained from 5,842 *PstI*-derived SNPs that formed non-fragmented LGs. Each line represents a connection between the chromosomal placement of a particular marker in our linkage map (black; scale in cM) and a homologous sequence in the physical map (non-black colors; scale in Mb)

2.4 Discussion

CornellGBS is a widely employed method for genotyping large genome species because it is simple, fast, specific, reproducible, and interrogates important regions of the genome that are inaccessible to sequence capture approaches (HE et al., 2014). Although this methodology was first reported in maize (ELSHIRE et al., 2011a), its application was recently expanded to bovine (DE DONATO et al., 2013). Moreover, a similar technique called GGRS was recently applied in chickens (LIAO et al., 2015). In the present study we have adapted the CornellGBS successfully to be applied in chickens using a restriction enzyme that generates an appropriate genomic shearing range for this species. This work describes for the first time the application of the CornellGBS method for chicken genotyping. This is a cost-effective genotyping method that was performed here in a large number of individuals (462 chickens).

The GBS approach involves four steps: (1) genomic DNA cleavage, (2) adapter ligation with specific barcodes, (3) sequencing of short reads, and (4) bioinformatics analysis.

The first step in the method adaptation for its use in chickens is the selection of an appropriate restriction enzyme to shear the chicken genome in a suitable range of fragments for sequencing by the Illumina platform. We performed *in silico* (Figure 2.1) and *in vitro* (see Supplementary Fig. S2 online) genomic fragmentation tests to compare the digestion profiles of two restriction enzymes, *PstI* and *SbfI*.

The CornellGBS approach is flexible enough to be applied on different genomes. However, the choice of a restriction enzyme that cleaves the DNA generating a suitable fragment length range is of particular importance. Moreover, genomes of different species will produce distinct cleavage patterns with the same enzyme, reason why optimization is required for the genomic cleavage in each species (GLAUBITZ et al., 2014). It is also important to consider whether the restriction enzyme is sensitive to DNA methylation in its restriction site (POLAND; RIFE, 2012; GUERRERO-BOSAGNA, 2013). *PstI* showed here the best fragmentation profile among the two enzymes tested for cleaving the chicken genome, both *in vitro* and *in silico*. The next step after the selection of the appropriate restriction enzyme was to optimize the binding reactions between the fragments, adapters and barcodes.

After sequencing of the CornellGBS libraries the next step was the bioinformatics analysis. Using the Tassel pipeline ~5.3 million of unique tags were obtained and aligned against the last chicken reference genome (*Gallus_gallus*_4.0, NCBI). Although 48% of the reads were discarded, which can be considered a drawback of CornellGBS approach (LIAO et al., 2015), the number of unique tags obtained (1.4 billion) is sufficient for an accurate

identification of SNPs. As a matter of comparison, a similar study generated ~0.5 million unique tags using 47 individuals (DE DONATO et al., 2013). The multiplexing capability is an advantage of the CornellGBS approach that increases the catalog size of unique tags.

Most of genotyping methods have limitations when it comes to detection of heterozygous SNPs, due to the low coverage of these sites (QUAIL et al., 2012). For a coverage of less than 5X per site per individual the probability that only one of the two chromosomes of a diploid individual is sampled for a particular site is generally high (NIELSEN et al., 2011). The tassel-GBS pipeline compensates low coverage data and under-calling of heterozygotes with the redundant coverage of haplotypes at high marker density, which facilitates imputation of missing genotypes (GLAUBITZ et al., 2014). This is possible because in the Tassel-GBS pipeline the tag catalog is created from individuals pooled altogether, rather than from separate individuals. The latter is the case for the Stacks software, a program commonly used to handle GBS data (CATCHEN et al., 2011).

Different filtering parameters on SNP calling were tested in the present study. Since using a mnMAF of 5% (Tassel default) generated 5,478 less SNPs than using a mnMAF of 1%, we proceeded with a mnMAF of 1%. Moreover, because parental pure lines featuring only 5 individuals per strain were used, the previous mnMAF ≥ 5% would eliminate many important SNPs that might be present in the parental lines.

When considering a mnMAF of 1% and mnScov filter of 90%, our study generated a reliable SNP dataset of 67,096 *PstI*-derived SNPs, out of which 20.7% have not been previously described in the dbSNP (based on the last update of the dbSNP, NCBI, September 2015). A previous study that used RAD-Seq in chickens (ZHAI et al., 2015) found 28,895 *HindIII*-derived SNPs candidates with 53.3% of them newly reported (based on a previous version of dbSNP database, which contains fewer SNPs). Therefore, a reasonable number of novel SNPs were obtained here using the CornellGBS approach (13,434).

In spite of the different SNP calling methodologies used (Stacks vs Tassel), the number of *PstI*-derived SNPs reported here was higher than *HindIII*-derived SNPs previously reported (ZHAI et al., 2015). This is probably explained by the difference in the number of 'tag counts' observed after cleavage (*in silico)* by *HindIII* (~700 K, as previously reported (ZHAI et al., 2015)) and *PstI* (~ 1.2 million, reported here), or by the larger number of genotyped animals in the present study.

The chromosomal position of the SNPs identified in this study (using CornellGBS and considering a mnMAF of 1% and mnScov filter of 90%) was compared with the Illumina chicken 60K Beadchip (GROENEN et al., 2011) and with the 600K HD Affymetrix®Axion®

genotyping array for chicken (KRANIS et al., 2013). We found that the average distances between markers obtained using the CornellGBS or the 60K approaches were similar (15 and 21 Kb, respectively), although lower than with the 600K (1.7 Kb), which showed less distance between markers.

Differences between mean and median were detected only using the CornellGBS approach. This indicates that SNP cluster formation occurs in spite of the markers obtained by the CornellGBS being well spread throughout the genome (Figure 2.5). With the 60K or the 600K panel, however, uniform SNP distribution occurs without cluster formation. Also, differences between mean and median are not observed (Table 2.1). The detection of SNP clusters by the Cornell GBS approach lead us to perform functional annotation of the markers and compare the results between the methodologies tested. When the distribution of SNP distances was evaluated (Figure 2.4) we noticed that the GBS and 600K approaches had a similar proportion of SNPs that corresponded to the fraction of <1 Kb SNP distance between markers. Within GBS clusters the SNP density was higher and approximately 76% of SNPs were <5 Kb apart.

We also investigated SNP density differences related to chromosome size (see Supplementary Table S4 online) using the three methodologies (CornellGBS, 60K Illumina and 600K Affymetrix). CornellGBS detected about one-third more SNPs than the other two methods in regions of the microchromosomes GGA11-32 and W. The microchromosome GGA16 showed a higher representation of SNPs using the CornellGBS approach compared to the 60K panel (0.19% - 0.05%). The GGAW microchromosome in the CornellGBS approach had two-thirds of SNP representation compared to the 600K Illumina panel. Interestingly, SNPs in this microchromosome are not detected by the 60K panel. SNPs in the GGA32 microchromosome were detected only by the CornellGBS approach. Interestingly, microchromosomes have 2-4 times higher gene density than macrochromosomes (SMITH et al., 2000; HABERMANN et al., 2001) and ~48% of genes in microchromosomes have a high CpG island content (MCQUEEN et al., 1996; MCQUEEN; SIRIACO; BIRD, 1998; SMITH; BURT, 1998; SMITH et al., 2000; HABERMANN et al., 2001). This suggests *PstI* RE genomic cleavage would be appropriated for DNA methylation profiling, since it apparently enriches for regions of high CpG content.

A set of SNPs from the CornellGBS dataset obtained in our study was compared with a WCGR SNPs dataset obtained from sequencing the same 10 animals. Substantial chromosomal position (~84%) and genotype (~93%) concordances were observed between the two methods. However, the concordance was reduced to ~71% when considering only the heterozygous

SNPs. In spite of this, 99.90% of the genotypes were concordant in regions where both methodologies were able to call heterozygous. Therefore, although the CornellGBS had fewer calls of heterozygous in comparison with WCGR, those genotypes that are called are quite reliable.

We also tested for Mendelian errors in the markers obtained in each population used in this study. Mendelian inheritance errors are likely to result from erroneous genotype calls (PILIPENKO et al., 2014). The errors found were <10% between the parental (F_0) and the F_1 generation, and the same between the F_1 and the F_2 generations. The exception is, family F_2-7816 that presented slightly higher Mendelian errors (11.9%). These error rates are in agreement with the low heterozygous call rate (0.90>call rate>0.95) and high heterozygous CV (>15%) observed in individuals from this family (25 from 94) when compared to the others four families (0.95>call rate>1.0). Therefore, the Mendelian errors observed were minimal and do not compromise the quality of the genotyping performed in the present study. In addition, the linkage map obtained from markers with Mendelian segregation obtained from the five F_2 families were grouped in LGs. This grouping generated a fairly dense linkage map. These markers (~99% of them) grouped according to their respective described chromosomes (Figure 2.6 and Supplementary Fig. S5).

We also found a small increase in the proportion of SNPs (3.3%) in exonic regions compared to a recent functional classification of 15 million SNPs detected from diverse chicken populations (2.2%) (GHEYAS et al., 2015), or when compared to the WCGR. These newly discovered SNPs in exonic regions include a QTL region on chromosome 3 associated with fatness in chickens (0.98%) (MOREIRA et al., 2015) and another on chromosome 2 associated with muscle deposition (0.59%) (GODOY et al., 2015). These *exonic* variants (2,590) were classified into functional categories due to their potential to alter the tri-dimensional structure and function of the translated protein (NG; HENIKOFF, 2003). These exonic variants detected in the present study were classified as non-synonymous, startlost, startgain or stopgain (Table 2.4).

When comparing the CornellGBS and the 60K Illumina approaches (which have similar SNP density), it was observed that 60K Illumina detects half (51.6%) of the exonic variants detected by CornellGBS. However, that difference is reduced when only non-synonymous SNPs are considered (907 SNPs detected by CornellGBS; 888 SNPs detected by 60K Illumina). When comparing CornellGBS and 600K Illumina, the proportions of SNPs in exonic regions are similar (3.3% and 3.5%, respectively). This shows that Cornell GBS is as powerful as the

600K panel in detecting SNPs in exonic regions, which is remarkable considering that the 600K panel was designed prioritizing coding regions (KRANIS et al., 2013).

On the downside, CornellGBS seems to be less powerful in detecting SNPs in intergenic regions compared to either the 60K Illumina or the 600K Affymetrix approaches (28.21%, 43.68% and 41.77% respectively). On the other hand, CornellGBS presents a high proportion of SNPs in regions 1kb up- or downstream from UTR compared to either the 60K Illumina or the 600K Affymetrix approaches (14.69 and 15.70; 7.94 and 7.54; 7.58 and 7.28; respectively). This is interesting because UTR regions are highly relevant for transcriptional regulation (CHO et al., 1998).

These results indicate that the Cornell GBS approach shows a pattern of SNP profiling that is unique in comparison with other approaches. The unique characteristics of Cornell GBS include better interrogation of specific functional regions, of microchromosomes and of CpG-rich regions compared to other methodologies (60K Illumina or 600K Affymetrix). In particular, we believe that the restriction enzyme used in the present study (*PstI*) is responsible for enriching the cleaved genome for microchromosomic or CpG-rich regions.

The present study shows for the first time the application of CornellGBS in chickens, which will allow for the use of a cost-effective (~US$50/sample) genotyping approach in poultry. The method described is capable of performing a reliable SNP profiling in chickens using a large number of animals. In the present study a number of SNPs were discovered, which were well spread throughout all the chromosomes of the chicken genome (Figure 2.5). This study describes a highly multiplexed sequencing method in chicken, with potential for application in studies involving genome-wide association and genomic selection.

References

ALTSHULER, D.; POLLARA, V.J.; COWLES, C.R.; VAN ETTEN, W.J.; BALDWIN, J.; LINTON, L.; LANDER, E.S. An SNP map of the human genome generated by reduced representation shotgun sequencing. **Nature**, London v. 407, n. 6803, p. 513–516, 2000.

ANDREWS, S. **FASTQC:** a quality control tool for high throughput sequence data; reference source. 2010. Disponível em <http://http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Acesso em: 09 sept. 2016.

BOZEMAN, M.T. **Golden helix**. 2016. Disponível em: <http://goldenhelix.com/>. Acesso em: 10 abr. 2016.

BROOKES, A.J. The essence of SNPs. **Gene**, Amsterdam, v. 234, n. 2, p. 177–86, July 1999.

BURT, D.W. Chicken genomics charts a path to the genome sequence. **Briefings in Functional Genomics and Proteomics**, London,v. 3, n. 1, p. 60–67, Jan. 2004.

CATCHEN, J.M.; AMORES, A.; HOHENLOHE, P.; CRESKO, W.; POSTLETHWAIT, J.H. Stacks: building and genotyping Loci de novo from short-read sequences. **G3,** Bethesda, v. 1, n. 3, p. 171–182, Aug. 2011.

CHO, R.J.; CAMPBELL, M.J.; WINZELER, E.A.; STEINMETZ, L.; CONWAY, A.; WODICKA, L.; WOLFSBERG, T.G.; GABRIELIAN, A.E.; LANDSMAN, D.; LOCKHART, D.J.; DAVIS, R.W. A genome-wide transcriptional analysis of the mitotic cell cycle. **Molecular Cell**, Cambridge, v. 2, n. 1, p. 65–73, July 1998.

DAVEY, J.W.; HOHENLOHE, P.A.; ETTER, P.D.; BOONE, J.Q.; CATCHEN, J.M.; BLAXTER, M.L. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. **Nature Reviews. Genetics**, London, v. 12, n. 7, p. 499–510, July 2011.

DE DONATO, M.; PETERS, S.O.; MITCHELL, S.E.; HUSSAIN, T.; IMUMORIN, I.G. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. **PloS One**, San Francisco v. 8, n. 5, p. e62137, Jan. 2013.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKLER, E.S.; MITCHELL, S.E. A Robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. **PLoS One**, San Francisco v. 6, n. 5, p. e19379, May 2011.

GHEYAS, A.A.; BOSCHIERO, C.; EORY, L.; RALPH, H.; KUO, R.; WOOLLIAMS, J.A.; BURT, D.W. Functional classification of 15 million SNPs detected from diverse chicken populations. **DNA Research: an International Journal for Rapid Publication of Reports on Genes and Genomes**., Oxford, v. 22, n. 3, p. 205–217, 2015.

GLAUBITZ, J.C.; CASSTEVENS, T.M.; LU, F.; HARRIMAN, J.; ELSHIRE, R.J.; SUN, Q.; BUCKLER, E.S. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **PloS One**, San Francisco v. 9, n. 2, p. e90346, Jan. 2014.

GODOY, T.F.; MOREIRA, G.C.M.; BOSCHIERO, C.; GHEYAS, A.A.; GASPARIN, G.; PADUAN, M.; ANDRADE, S.C.S.; MONTENEGRO, H.; BURT, D.W.; LEDUR, M.C.; COUTINHO, L.L. SNP and INDEL detection in a QTL region on chicken chromosome 2 associated with muscle deposition. **Animal Genetics**, New York, v. 46, n. 2, p. 158–163, 2015.

GROENEN, M.A.M.; MEGENS, H.-J.; ZARE, Y.; WARREN, W.C.; HILLIER, L.W.; CROOIJMANS, R.P.M.A.; VEREIJKEN, A.; OKIMOTO, R.; MUIR, W.M.; CHENG, H.H. The development and characterization of a 60K SNP chip for chicken. **BMC Genomics**, London, v. 12, n. 1, p. 274, Jan. 2011.

GUERRERO-BOSAGNA, C. DNA methylation research methods. **Materials and Methods**, v. 3, Jan. 2013. Disponível em: <http://www.labome.com/method/DNA-Methylation-Research-Methods.html>. Acesso em: 08 set. 2016.

HABERMANN, F.A.; CREMER, M.; WALTER, J.; KRETH, G.; VON HASE, J.; BAUER, K.; WIENBERG, J.; CREMER, C.; CREMER, T.; SOLOVEI, I. Arrangements of macro- and microchromosomes in chicken cells. **Chromosome Research: an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology**, Oxford, v. 9, n. 7, p. 569–584, 2001.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z.-X.; LIU, H.; LI, Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Frontiers in Plant Science**, Lausanne, v. 5, p. 484, Sept. 2014.

HU; YAN, C.; HSU, C.-H.; CHEN, Q.-R.; NIU, K.; KOMATSOULIS, G.; MEERZAMAN, D. OmicCircos: a simple-to-use R package for the circular visualization of multidimensional Omics data. **Cancer Informatics**, Auckland, v. 13, suppl. 5, p. 13, Jan. 2014.

HUBER, W.; CAREY, V.J.; GENTLEMAN, R.; ANDERS, S.; CARLSON, M.; CARVALHO, B.S.; BRAVO, H.C.; DAVIS, S.; GATTO, L.; GIRKE, T.; GOTTARDO, R.; HAHNE, F.; HANSEN, K.D.; IRIZARRY, R.A.; LAWRENCE, M.; LOVE, M.I.; MACDONALD, J.; OBENCHAIN, V.; OLEŚ, A.K.; PAGÈS, H.; REYES, A.; SHANNON, P.; SMYTH, G.K.; TENENBAUM, D.; WALDRON, L.; MORGAN, M. Orchestrating high-throughput genomic analysis with bioconductor. **Nature methods,** New York, v. 12, n. 2, p. 115–121, 2015.

KOSAMBI, D.D. The estimation of map distances from recombination values. **Annals of Eugenics**, London, v. 12, n. 1, p. 172–175, Jan. 1943.

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. Development of a high density 600K SNP genotyping array for chicken. **BMC Genomics**, London, v. 14, n. 1, p. 59, Jan. 2013.

KUMAR, S.; YOU, F.M.; CLOUTIER, S. Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. **BMC Genomics**, London, v. 13, n. 1, p. 684, Jan. 2012.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, London, v. 9, n. 4, p. 357–359, Mar. 2012.

LI, C.; LI, M.; LONG, J. R.; CAI, Q.; ZHENG, W. Evaluating cost efficiency of SNP chips in genome-wide association studies. **Genetic Epidemiology**, New York, v. 32, n. 5, p. 387–395, 2008.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G. D.R. The sequence alignment/map format and SAMtools. **Bioinformatics,** Oxford, v. 25, n. 16, p. 2078–2079, 2009.

LIAO, R.; WANG, Z.; CHEN, Q.; TU, Y.; CHEN, Z.; WANG, Q.; YANG, C.; ZHANG, X.; PAN, Y. An efficient genotyping method in chicken based on genome reducing and sequencing. **Plos One**, San Francisco v. 10, n. 8, p. e0137010, Aug. 2015.

MARGARIDO, G.R.A.; SOUZA, A.P.; GARCIA, A.A.F. OneMap: software for genetic mapping in outcrossing species. **Hereditas**, Lund, v. 144, n. 3, p. 78–79, 2007.

MCLAREN, W.; PRITCHARD, B.; RIOS, D.; CHEN, Y.; FLICEK, P.; CUNNINGHAM, F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. **Bioinformatics**, Oxford, v. 26, n. 16, p. 2069–2070, Aug. 2010.

MCQUEEN, H.A.; SIRIACO, G.; BIRD, A.P. Chicken microchromosomes are hyperacetylated, early replicating, and gene rich. **Genome Research,** Cold Spring Harbor, v. 8, n. 6, p. 621–630, 1998.

MCQUEEN, H.A.; FANTES, J.; CROSS, S.H.; CLARK, V.H.; ARCHIBALD, A.L.; BIRD, A.P. CpG islands of chicken are concentrated on microchromosomes. **Nature Genetics**, New York, v. 12, n. 3, p. 321–324, 1996.

MOREIRA, G.C.M.; GODOY, T.F.; BOSCHIERO, C.; GHEYAS, A.; GASPARIN, G.; ANDRADE, S.C.S.; PADUAN, M.; MONTENEGRO, H.; BURT, D.W.; LEDUR, M.C.; COUTINHO, L.L. Variant discovery in a QTL region on chromosome 3 associated with fatness in chickens. **Animal Genetics**, New York, v. 46, n. 2, p. 141–147, 2015.

NG, P.C. SIFT: predicting amino acid changes that affect protein function. **Nucleic Acids Research**, Oxford, v. 31, n. 13, p. 3812–3814, July 2003.

NG, P.C.; HENIKOFF, S. SIFT: predicting amino acid changes that affect protein function. **Nucleic Acids Research**, Oxford, v. 31, n. 13, p. 3812–3814, July 2003.

NIELSEN, R.; PAUL, J.S.; ALBRECHTSEN, A.; SONG, Y.S. Genotype and SNP calling from next-generation sequencing data. **Nature Reviews Genetics**, London, v. 12, n. 6, p. 443–451, 2011.

NONES, K.; LEDUR, M.C.; RUY, D.C.; BARON, E.E.; MELO, C.M.R.; MOURA, A.S.A.M.T.; ZANELLA, E.L.; BURT, D.W.; COUTINHO, L.L. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. **Animal Genetics**, New York, v. 37, n. 2, p. 95–100, Apr. 2006.

PETERSON, B.K.; WEBER, J.N.; KAY, E.H.; FISHER, H.S.; HOEKSTRA, H.E. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. **PloS One**, San Francisco, v. 7, n. 5, p. e37135, Jan. 2012.

PILIPENKO, V.V; HE, H.; KUROWSKI, B.G.; ALEXANDER, E.S.; ZHANG, X.; DING, L.; MERSHA, T.B.; KOTTYAN, L.; FARDO, D.W.; MARTIN, L.J. Using Mendelian inheritance errors as quality control criteria in whole genome sequencing data set. **BMC Proceedings**, London, v. 8, suppl 1, p. S21, 2014.

POLAND, J.A.; RIFE, T.W. Genotyping-by-sequencing for plant breeding and genetics. **The Plant Genome Journal**, Madison, v. 5, n. 3, p. 92, 2012.

QUAIL, M.A.; GU, Y.; SWERDLOW, H.; MAYHO, M. Evaluation and optimisation of preparative semi-automated electrophoresis systems for Illumina library preparation. **Electrophoresis**, Weinheim, v. 33, n. 23, p. 3521–3528, 2012.

ROSÁRIO, M.F. do; LEDUR, M.C.; MOURA, A.S.A.M.T.; COUTINHO, L.L.; GARCIA, A.A.F. Genotypic characterization of microsatellite markers in broiler and layer selected chicken lines and their reciprocal F1s. **Scientia Agricola**, Piracicaba, v. 66, n. 2, p. 150–158, abr. 2009.

SMITH, J.; BRULEY, C.K.; PATON, I.R.; DUNN, I.; JONES, C.T.; WINDSOR, D.; MORRICE, D.R.; LAW, A.S.; MASABANDA, J.; SAZANOV, A.; WADDINGTON, D.; FRIES, R.; BURT, D.W. Differences in gene density on chicken macrochromosomes and microchromosomes. **Animal Genetics**, New York, v. 31, n. 2, p. 96–103, 2000.

SMITH, J.; BURT, D.W. Parameters of the chicken genome (*Gallus gallus*). **Animal genetics**, New York, v. 29, n. 4, p. 290–294, 1998.

VAN OOIJEN, J.W. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. **Genetics Research**, Cambridge, v. 93, p. 343–349, 2011.

ZHAI, Z.; ZHAO, W.; HE, C.; YANG, K.; TANG, L.; LIU, S.; ZHANG, Y.; HUANG, Q.; MENG, H. SNP discovery and genotyping using restriction-site-associated DNA sequencing in chickens. **Animal Genetics**, New York, v. 46, n. 2, p. 216–219, Apr. 2015.

ZHBANNIKOV I.Y., H.S.S.; SETTLES, M.L. **SEQYCLEAN:** user manual. 2013. Disponível em: <https://github.com/ibest/seqyclean>. Acesso em: 09 set. 2016.

Acknowledgements

The authors are grateful to Ph.D. Sonia Andrade, Ph.D. Priscilla Marqui Schmidt Villela, M.Sc. Ricardo Augusto Brassaloti and M.Sc. Gabriel Costa Monteiro Moreira from Luiz de Queiroz College of Agriculture (ESALQ/USP) for helpful suggestions. Financial support was provided by the University of São Paulo and the Embrapa Swine and Poultry National Research Center and the ERC GeneWell project to PerJensen. The development of the F_2 population and genotyping was funded by the PRODETAB project (no. 038-01/01). Fábio Pértille received a fellowship from CAPES. Luiz L. Coutinho is recipient of productivity fellowship from CNPq. Clarissa Boschiero received a fellowship from the program *Science Without Borders* – National Council for Scientific and Technological Development (CNPq). .

Author Contributions Statement

F.P. analyzed the data and wrote the manuscript with the valuable help of. C.G.B. and C.B. C.B., V.H.S. and J.R.N. helped with bioinformatics analysis and graphical development. J.R.N. helped with the implementation of the CornellGBS method. M.C.L. provided the biological material and reviewed the manuscript. P.J provided laboratory and computational facilities

for data analyses, and reviewed the manuscript. L.L.C provided overall supervision of the research. All authors approved the last version of the manuscript.

Supplementary Information

The supplementary information can be obtained on-line through the link: http://www.nature.com/articles/srep26929, where this article is published and where information from this chapter should be cited.

3 GENOME-WIDE ASSOCIATION STUDY FOR PERFORMANCE TRAITS IN CHICKENS USING GENOTYPE BY SEQUENCING APPROACH

Abstract

Performance traits are economically important and are targets for selection in breeding programs, especially in the poultry industry. To identify regions on the chicken genome associated with performance traits, different genomic approaches have been applied in the last years. The aim of this study was the application of CornellGBS approach (134,528 SNPs generated from a *PstI* restriction enzyme) on Genome-Wide Association Studies (GWAS) in an outbred F_2 chicken population. We have validated 91.7% of these 134,528 SNPs after imputation of missed genotypes. Out of those, 20 SNPs were associated with feed conversion, one was associated with body weight at 35 days of age (P<7.86E-07) and 93 were suggestively associated with a variety of performance traits (P <1.57E-05). The majority of these SNPs (86.2%) overlapped with previously mapped QTL for the same performance traits and some of the SNPs also showed novel potential QTL regions. The results obtained in this study have indicated their importance as basis for further refinements of regions controlling performance traits.

Keywords: Chicken; GBS; GWAS; Next-generation sequencing; PstI; Restriction enzyme; Selection

3.1 Introduction

Production efficiency in the poultry industry is constantly improving as a result of selection for growth rate, feed efficiency and carcass traits for broilers, and egg production and quality traits for layers (BLACKBURN, 2006; FULTON, 2012). The understanding of genomic information of loci controlling those traits are important to improvement of selection efficiencies of breeding programs (FULTON, 2012).

Therefore, the population used in this study was developed in 1999 for QTL mapping by Embrapa Suínos e Aves. Several studies have been conducted in this population using markers randomly distributed in the genome (microsatellites), which have allowed the identification of several QTLs for production traits (NONES et al., 2006, 2012; AMBO et al., 2008, 2009; CAMPOS et al., 2009). Following these, some of the studies have focused their attention on the identification of SNPs in functional and positional candidate genes and to test their association on target QTL regions (BOSCHIERO et al., 2013; FELÍCIO et al., 2013a, 2013b; PÉRTILLE et al., 2015) . With the advent of next-generation sequencing (NGS), it was possible to identify a global SNP profile and to perform genome-wide association studies (GWAS) to find novel QTL regions and also to refine the previously published regions (GODOY et al., 2015; MOREIRA et al., 2015). Parallel to our group researches, several

manuscripts have been published in this field related with: QTLs mapping (TATSUDA; FUJINAKA, 2001; KONING et al., 2003; IKEOBI et al., 2004; ZHOU et al., 2006; NASSAR; GORAGA; BROCKMANN, 2012, 2013), SNP identification in candidate genes (NIE; ZHANG; LEI, 2003; SHEN et al., 2012), SNP identification in target QTLs (AHSAN et al., 2013; ROUX et al., 2014; LI et al., 2015) and GWAS (XIE et al., 2012; LUO et al., 2013a, 2013b; MORRIS et al., 2013; PARK et al., 2013; SUN et al., 2013, 2014) using the chicken as a model.

Despite the high-throughput data generation by NGS, which have facilitated the identification of SNPs in several populations, the use of this method for GWAS is still a limitation. This is caused by the high cost involved with the generation of data to be applied in a large number of individuals. To solve this high cost problem, SNPs panels and arrays were designed to be applied in GWAS (GROENEN et al., 2011; KRANIS et al., 2013). However, some important regions in the genome are inaccessible to sequence capture approaches (HE et al., 2014) mainly because they are based on predesigned SNP profiles. To overcome those limitations, and to present a unique SNP profile, we used the *PstI*-derived SNPs dataset from CornellGBS optimized approach. This dataset was originated from a SNP call from the reduced representation of the sequenced genome (~5%) through *PstI* restriction enzyme (PÉRTILLE et al., 2016). This SNP dataset is reliable and reproducible, showing a unique profile of SNPs with microchromosome enrichment (PÉRTILLE et al., 2016) that contains 2-4 times higher gene density than macrochromosomes (SMITH et al., 2000; HABERMANN et al., 2001).

The aim of this study was to identify genetic markers using *PstI*-derived SNPs dataset, and further use that information to conduct a GWAS with performance traits in chickens. In addition, we have performed a linkage disequilibrium (LD) analyses in the parental, F_1 and F_2 generations, to better understand the segregation of haplotype blocks, and the population structure, from the associated and suggestively associated SNPs identified. Finally, we have compared the location of these mentioned SNPs with known QTLs, with the objective to validate and to refine the regions known QTLs.

3.2 Methods

All experimental protocols employed in the present study that relate to animal experimentation were performed in accordance with the resolution number 010/2012 approved by the Embrapa Swine and Poultry Ethics Committee on Animal Utilization to ensure compliance with international guidelines for animal welfare.

3.2.1 Animals and Phenotypes

This study was conducted using 464 chickens from a F_2 populations originated from a reciprocal cross-experimental population from Embrapa Swine and Poultry National Research Center, Concórdia, SC, Brazil. We also included in the analysis, 10 chickens from their parental lines TT and CC (5 from each one), and eigth from the F_1 generation. This F_2 population was developed for QTL mapping studies, and was originated from the crossing of seven males from a broiler line (TT) and seven females from a layer line (CC), resulting in seven full-sib families (F_1 generation). Then, twenty-one F_1 females were artificially inseminated with seven F_1 males (3:1 ratio) to generate the F_2. The F_2 population comprised seven paternal half-sib families composed of three full-sib families of approximately 100 individuals each, produced in 17 hatches, totaling 2,063 F_2 chickens (ROSÁRIO et al., 2009). . For this study, we selected the five most informative families based on the previously QTL studies (AMBO et al., 2009; PÉRTILLE et al., 2015; BARON et al., 2011).

The TT broiler line was selected over six generation to improve body weight, feed conversion, carcass and breast yields, viability, fertility, eclodibility, reduction of abdominal fat and metabolic syndromes (ROSÁRIO et al., 2009). The CC layer line was selected over eight generation to improve egg production, egg weight, feed conversion, viability, sexual maturity, fertility, eclodibility, egg quality and reduction of body weight (ROSÁRIO et al., 2009).

The F_2 chickens were reared as broilers with free access to corn and soybean meal-based diet and water up to 42 days of age. From 35 to 41 days, they were transferred to cages to collect feed intake, and to computate the conversion and efficiency. Body weight was recorded at 1 (birth weight), 35, 41 and 42 (after fasting) days of age. The body weight (BW) at 41 days of age was collected at the end of the conversion measurement and, BW42 days was collected after 6-h fasting period and transportation to the slaughterhouse. More details were previously provided (NONES et al., 2006; ROSÁRIO et al., 2009; PÉRTILLE et al., 2015). We analysed these six performance traits in this study (feed conversion, feed intake and feed efficiency between 35 to 41 days, birth weight, and body weight at 35 and 41 days of age). A total of 23 missed values from the selected traits were imputed by mean using Tassel v.5.2.26 tool (GLAUBITZ et al., 2014) among the 446 F_2 animals selectect to be evaluated in this study.

3.2.2 DNA, genotypic data and imputation

Genomic DNA was cleaved with the *PstI* enzyme, ligated to adapters with barcodes identifying individual animals, and then sequenced on Illumina platform. After filtering parameters were applied, 134,528 SNPs were identified from 462 individuals in our experimental population of chickens using minimum minor allele frequency (mnMAF) of 1%, minimum taxon coverage (mntCov) of 20% of and minimum site coverage (mnScov) of 70% filter parameters. All procedures to obtain the data were previously described (PÉRTILLE et al., 2016). After filtering parameters, the number of missing genotypes increased from 0.9 to 5.8 million. This number represents SNPs identified multiplied for the number of individuals genotyped (67,096 derived *PstI*-SNPs x 462 individuals using 90% of loci call rate and 134,528 derived *PstI*-SNPs x 462 individuals using 70% of loci call rate). The imputation of missing genotypes was performed using Beagle 4.1 software (BROWNING; BROWNING, 2007) using default parameters, which uses empirical LD model. This model adapts to the local structure in the data using iterative approach to haplotype phasing in which an initial prediction of haplotype phase is made, then the model is fit, and improved estimates of haplotype phase are obtained and the model is refit (BROWNING; BROWNING, 2007).

3.2.3 SNPs validation

The validation was performed by the comparison of the filtered 134,528 *PstI*-derived SNPs dataset with the 600K Affymetrix® HD genotyping array SNP dataset using five individuals from F_2-7810 family. The SNPs in both datasets were located on GGA1-28, 32, W and Z chromosomes on the Gallus-gallus-4.0 reference genome. The validation standards used in this study were based on a methodology previously proposed (ECK et al., 2009) considering chromosomal positions and genotype concordances. With these concordances, we estimate the accuracy for homozygous and heterozygous SNPs.

3.2.4 Principal component analysis

The principal component analyses (PCA) of imputed genotype data were performed using Tassel v.5.2.26 (GLAUBITZ et al., 2014) considering five principal components. This tool transforms a set of correlated variables into successive orthogonal PCs accounted for the maximum variance providing a way to highlight groups of individuals differing at the level of

minor allele frequency (GLAUBITZ et al., 2014). The PCA graph was produced using the plot function in R 2.13.2 software (http://www.r-project.org/).

3.2.5 Genome-Wide Association Study

The compressed mixed linear model (MLM) implemented in TASSEL v.5.2.26 software (BRADBURY et al., 2007) was used for GWAS. The statistical model is represented by the following form:

$$y = \boldsymbol{X}\beta + \boldsymbol{Z}u + e \quad (1)$$

where y is the vector of the dependent variables, β is the vector containing fixed effects, including the sex (male/female), hatch (1-17) and SNP effects, and the covariate body weight at 35 days for traits measured from 35 to 41 days of age (feed intake, feed efficiency and feed conversion); u is the vector of random additive genetic effects from background QTL for individuals, X and Z are design matrices, and e is the vector of random residuals. We assumed that u and e vectors are normally distributed with null mean and variance of

$$Var\begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \quad (2)$$

where $G = \sigma_a^2 K$; σ_a^2 is an unknown additive genetic variance and K is the kinship (co-ancestry) matrix calculated from SNPs and provided by the same software using centered identity by state (IBS) method. For the residual effect, homogeneous variance was assumed, with $R = \mathbf{I}\sigma_e^2$, where $\mathbf{I}$ is an identity matrix and σ_e^2 is the unknown residual variance. The Restricted Maximum Likelihood (REML) estimates of σ_a^2 and σ_e^2 were obtained by the Efficient Mixed--Model Association (EMMA) algorithm [70]. Heritability (h^2) was calculated as the ratio of the additive genetic variance (σ_a^2) to the phenotypic variance ($\sigma_a^2 + \sigma_e^2$). Tassel program does not provide the standard errors of the estimates. Thus, standard errors were estimated using the REML method with an average information (AI) algorithm by AIREMLF90 software. Standard errors for additive genetic and residual variance were computed as square roots of diagonal elements of the inverse of the average information matrix. For Heritability, standard deviations obtained from the repeated sampling approach were considered as their standard errors. Each SNP allele was fit as a separate class with heterozygotes fit as additional

SNP classes. And also, the total SNP effect was not decomposed in additive and dominance effects but tested for overall significance (BRADBURY et al., 2007).

Quantile-quantile (Q-Q) plots for each trait and Manhattan plots of genome wide association analyses were performed in R using ggd.qqplot and Manhattan functions. The thresholds were corrected for multiple testing (Bonferroni) established by the estimated number of independent SNPs and LD blocks (pairwise r^2 values> 0.40)(NICODEMUS et al., 2005) that was 63,640 SNPs. We set two thresholds from our data: P < 1.57E-05 (1/ 63,640) for suggestive significance and P < 7.86E-07 (0.05/63,640) for genome-wise significance for 5% genome-wise significance level (GU et al., 2011; LIU et al., 2013).

3.2.6 Linkage disequilibrium analysis

The linkage disequilibrium (LD) analysis was performed with Haploview 4.2 (BARRETT et al., 2005) as well as the LD graphs. To perform the pairwise comparison of ours SNPs considering 1Mb apart, we selected 94 genome-wise associated and suggestively associated SNPs with the traits analyzed in this study (feed conversion, feed intake, feed efficiency, and weight gain between 35-41 days of age, birth weight and body weight at 35 and 41 days). We have defined the haplotype blocks by the solid spine of LD and the family structure of parental lines (CC and TT pure lines), F_1 and F_2 generations (Figure 3.3) using Haploview 4.2.

3.2.7 QTL overlapping with SNPs

The QTL data was obtained from QTLdb (HU et al., 2013) and the overlapping test was performed using the GenomicRanges package from R 3.3.1 software (http://www.r-project.org/). For the data presentation, we designed a figure using ggplot2 package karyogram layout from R 3.3.1 software.

3.3 Results

3.3.1 Animals and Phenotypes

The descriptive statistics for the eight performance traits analyzed can be observed in Table 3.1. Detailed description of these animals and traits were provided elsewhere (CAMPOS et al., 2009; PÉRTILLE et al., 2015). The large variability is expected since the animals are from a broiler x layer F_2 population.

Table 3.1 - Means, standard deviations (SD), maximum (max) and minimum (min) values for performance traits of 444 individuals from the F_2 population

Traits	Average	(SD)	max	min
feed conversion from 35-41 days	2.87	± 0.71	7.55	1.52
feed intake from 35-41 days (g)	603.43	± 139.74	1176.00	258.00
feed efficiency from 35-41 days	0.36	± 0.07	0.66	0.13
birth weight (g)	44.96	± 4.15	55.50	34.60
body weight at 35 days (g)	789.86	± 138.60	1309.00	480.00
body weight at 41 days (g)	1006.82	± 188.34	1686.00	407.00

* g is the weight measured in grams

3.3.2 Genotypes

In our previously work (PÉRTILLE et al., 2016), using a minimum taxon call rate of 90%, we have identified 67,096 SNPs originated from 462 chickens using the GBS approach. However, in this study, different filter parameters were applied. We have reduced the loci call rate filtering criteria to 70%. %. This parameter is the minimum threshold of individuals call rate for each loci to be included in the output. This reduction had minimal impact on sample call rate (proportion of missing genotypes per individual) and large impact on number of SNPs. The sample call rate reduced from 99.96% ± 0.04% to 99.90% ± 0.1% and the number of SNPs increased from 67,096 to 134,528. This allowed us to capture more SNPs, but the number of missing genotypes increased. To overcome this, we have imputed the missing genotypes using Beagle 4.1 software (BROWNING; BROWNING, 2007). This approach resulted in a panel of 134,528 derived *PstI*-SNPs present in all animals.

3.3.3 SNPs validation

The dataset of 134,528 SNP chromosomal positions obtained with the CornellGBS before and after the imputation analysis was compared with the 600K Affymetrix® HD genotyping array dataset in order to perform a method validation, since both sets were obtained from the same animals (5 individuals from F_2-7810 family). The genotype concordance of the SNPs with concordant chromosomal positions detected between the two methods is shown in Table 3.2. On average, 91.80%, and 91.66% of the SNPs had concordant genotypes between the CornellGBS and 600K datasets before and after imputation, respectively. The accuracy of the heterozygous genotypes was slightly lower after the imputation. Reduced representation methods, like CornellGBS, has limitations calling the heterozygous markers (GLAUBITZ et al., 2014). In our study, we have observed that 82.14 and 82.30% of heterozygous SNPs, while 97.97 and 97.65% of all homozygous SNPs were validated before and after imputation, respectively

Table 3.2 - Assessment of genotype concordance between 134,528 *PstI*-derived filtered SNPs before and after imputation and genotyped SNPs dataset from 600K Affymetrix® HD genotyping array from five F_2 individuals (F_2-7810 family); and genotype validation percentages for homozygous and heterozygous SNPs

		Before Imputation		**After Imputation**		**Total validated genotypes***	
F_2-7810 family	**SNP type**	**Concordant physical position†**	**Validated genotypes * (%)**	**Concordant physical position†**	**Validated genotypes *(%)**	**Before Imputation *(%)**	**After Imputation *(%)**
ID-1209	homoz[1]	2,410	97.67	2,493	97.47	93.08	92.93
	heteroz[2]	1,323	85.56	1,370	85.47		
ID-1786	homoz[1]	2,379	97.73	2,427	97.65	94.04	93.96
	heteroz[2]	1,422	88.60	1,440	88.47		
ID-1787	homoz[1]	2,296	97.99	2,356	97.87	93.72	93.42
	heteroz[2]	1,470	87.68	1,512	87.17		
ID-2301	homoz[1]	1,589	98.48	2,409	97.50	87.62	87.67
	heteroz[2]	948	70.25	1,457	72.06		
ID-786	homoz[1]	2,383	97.98	2,438	97.74	90.57	90.33
	heteroz[2]	1,395	78.63	1,422	78.34		

[1]homozygous genotype; [2]heterozygous genotype; † number of concordant physical positions between both datasets used for validation (from CornellGBS vs 600K Affymetrix® HD genotyping array); *is the percentage of † that had concordant genotypes.

3.3.4 Homozygous and heterozygous SNPs

Out of 62 million possible genotypes (462 samples x 134,528 sites), the average frequency of heterozygous SNPs was 25.32% (± 5.6%) before the imputation and after the imputation, it increased to 27.70% (± 5.2%). The average heterozygosity observed per chickens before imputation ranged from 8.30- 44.69% and after the imputation between 11.38-44.67%. The proportion of heterozygous SNPs remained virtually unchanged before and after imputation among the lines/generations (Table 3.3).

Table 3.3 - SNP heterozygosity of genotyped populations (parental, F_1 and F_2 generations) after and before imputation

Population	**Number of individuals**	**Proportion of heterozygous (SD)**	
		Before Imputation	After Imputation
Parental CC	5	0.16 ± 0.01	0.17 ± 0.01
Parental TT	5	0.21 ± 0.01	0.22 ± 0.01
F1	8	0.28 ± 0.09	0.29 ± 0.08
F2-7765	72	0.29 ± 0.03	0.29 ± 0.03
F2-7810	82	0.28 ± 0.04	0.28 ± 0.04
F2-7816	94	0.30 ± 0.08	0.30 ± 0.08
F2-7971	100	0.25 ± 0.05	0.26 ± 0.04
F2-7978	96	0.26 ± 0.04	0.27 ± 0.03

3.3.5 Principal component analyses

From the list of imputed genotypes we have conducted a principal component analysis (PCA), based on covariance, using Tassel v.5.2.26 (GLAUBITZ et al., 2014) to check the F_2 population structure. This plot was useful for visualizing internal structure explained by the variance from *PstI*-derived SNPs dataset of 134,528 SNPs using eigenvector-based multivariate analyses. Each individual lies in its proper group (Figure 3.1).

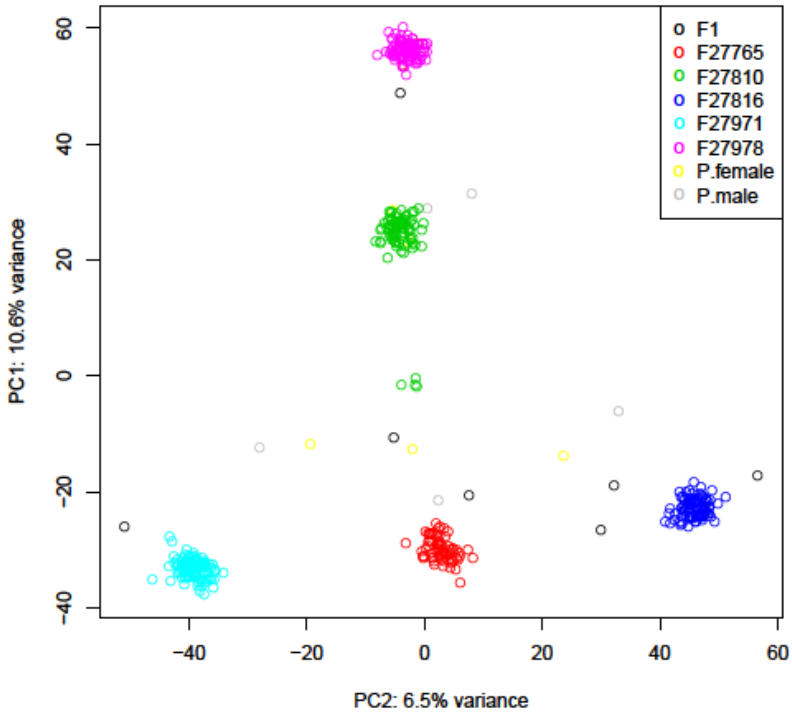


Figure 3.1 - The 444 chickens from five F_2 families and18 chickens from the parental lines (P.female and P.male - N=10) and F_1 generation (N=8) shown in the 2D plane spanned by their first two principal components

3.3.6 Descriptive Statistics of Heritability

The genetic and residual variance for each trait, and their genomic heritability are shown in Table 3.4. Heritabilities ranging from moderate to high, as is expected (GAYA et al, 2006; ROVADOSCKI, et al., 2016), were observed for feed intake and body weights traits, respectively. Low heritabilities were observed for the traits evaluated in short period (between 35 and 41 days) as feed conversion, and feed efficiency, because they are complex traits influenced by several environmental factors (ROVADOSCKI, et al., 2016).

Table 3.4 - Genetic and residual variances, and genomic heritability for each trait analyzed in this study

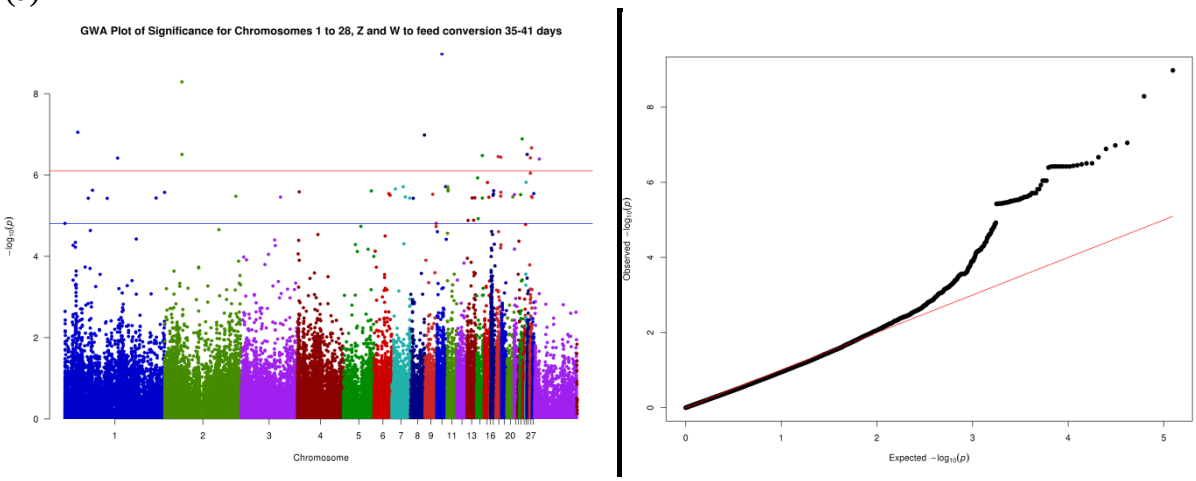
Traits	**Var_genetic ± SE**	**Var_error ± SE**	**Heritability ± SE**
feed conversion 35adj	0.0045 ± 0.00282	0.4363 ± 0.01797	0.01 ± 0.006
feed intake from 35adj	1492.3 ± 855.94	7296.0 ± 830.22	0.17 ± 0.094
feed efficiency 35adj	0.0005 ± 0.00002	0.0036 ± 0.00002	0.11 ± 0.005
birth weight	2.5719 ± 0.50599	3.1009 ± 0.33925	0.45 ± 0.073
body weight at 35 days	10403 ± 1575.9	1879.9 ± 785.7	0.85 ± 0.073
body weight at 41 days	16790 ± 2936.7	5639 ± 1701.1	0.75 ± 0.087

35adj is adjusted to body weight at 35 days

3.3.7 Genome-wide association study

We have identified 20 significant SNPs (P < 7.86E-07) associated with feed conversion adjusted to body weight at 35 days (adj35) and one significant SNP associated with body weight at 35 days of age (Figure 3.2). We also have identified 92 suggestive (P <1.57E-05) SNPs associated with feed conversion adj35, feed intake adj35, feed efficiency adj35, birth weight, and body weight at 35 and 41 days of age (see APPENDIX A for the effects of associated SNPs; Manhattan and QQ plots are available on APPENDIX B).

(a)



(b)

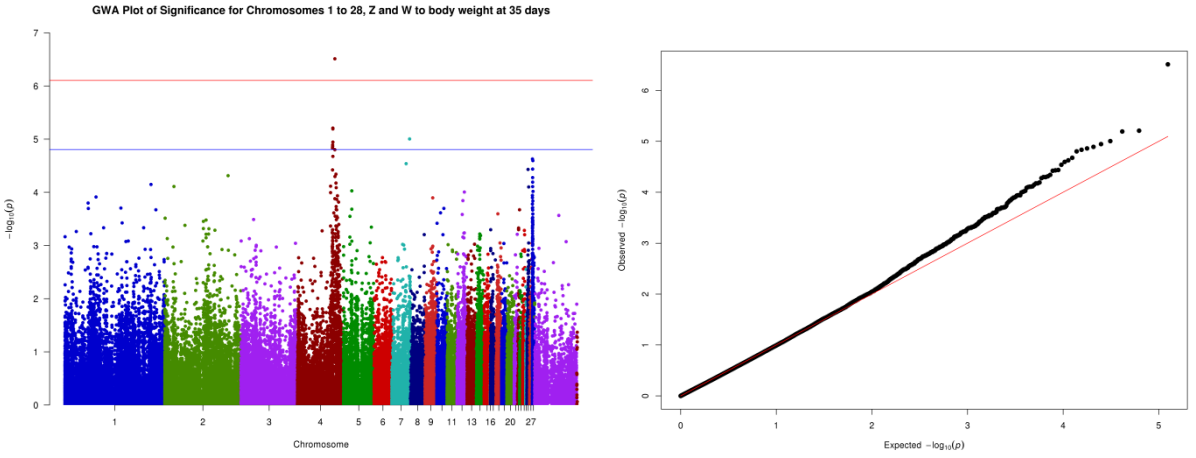


Figure 3.2 - SNPs associated with feed conversion adj35 (a) and body weight at 35 days (b) are presented by Manhattan (left side) and QQ (right side) plots. The y-axis is shown as –log10 (p-value). On the left, the red line indicates genome-wise association (P<7.86E-07) and the blue line, suggestive genome-wise association (P <1.57E-05) with the respective trait. On the right side, the QQ-plots show the relation of normal theoretical quantiles of the probability distributions between expected (x-axis) and observed (y-axis) p-values from each respective associated trait (feed conversion adj35 and body weight at 35 days)

3.3.8 Linkage disequilibrium analysis

Seventeen haplotype blocks were generated from the associated and suggestively associated SNPs from the F_2 population (see Figure 3.3, APPENDIX C for haplotypes details and APPENDIX E to SNPs´ Mendelian descriptions). We noticed a standard block pattern between the SNPs that matched with the F_2 population structure (Figure 3.3). Interestingly, we have checked the genotype frequency of blocks formed by LD analysis to determine if the blocks were fixed or not in the parental lines. From the haplotype blocks, we checked the origin of the haplotype variation (fixed or variable) and haplotype frequency of each block from F_2 generation in the parental lines (APPENDIX C and APPENDIX F for a more detailed description of frequencies). We also determined the advantageous haplotype for each trait in the F_2 generation (Table 3.5). This information enabled us to identify from which parental line (TT or CC) comes the genotypic variation observed in F_2 for each block. All blocks with $r^2 >$ 0.56 had the most frequent haplotype agreeing with the advantageous phenotype in the F_2 individuals (Figure 3.3; APPENDIX C and Supplementary and APPENDIX F), and this advantageous haplotype (lower feed conversion and higher values of other evaluated traits) was fixed in one of the parental lines, except in blocks 2 and 13. This information is also available for each genome-wide suggestive and/or associated SNPs in APPENDIX A, as well as the number of genotype observations obtained per SNP.

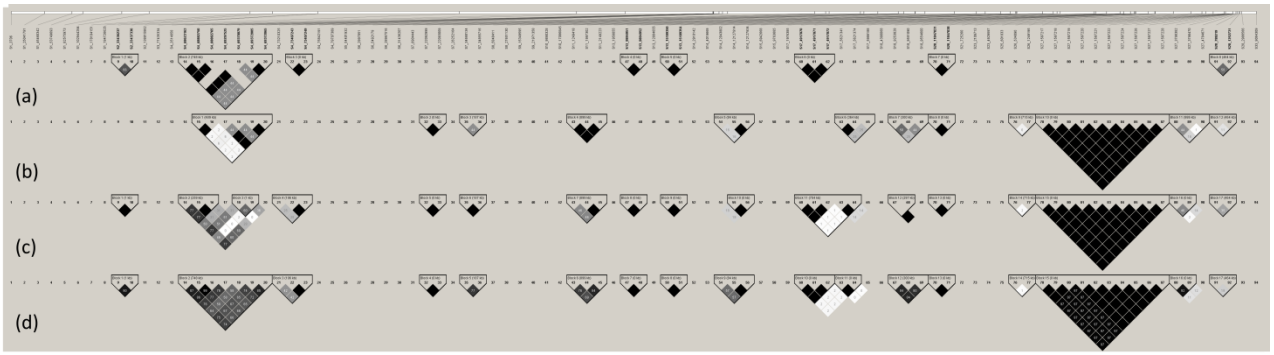


Figure 3.3 - Haplotype blocks obtained by the solid spine of LD and family structure using Haploview 4.2. The header represents the physical position of each 94 SNPs selected on chicken genome (*Gallus gallus 4.0,* NCBI). The family structure is disposed as: maternal CC (a) and paternal TT (b) parental pure lines, F_1 (c) and F_2 (d) generations. The black filled squares indicate $r^2 = 1$ and square numbers represents r^2*100 with black color gradient accordingly

3.3.9 QTL overlapping SNPs

Through Animal QTLdb, we have selected all the 1,458 known QTLs (HU et al., 2013) mapped for body weight, feed efficiency, feed conversion and growth, all evaluated in different chicken lines and ages. Out of those, we have observed that 253 QTLs overlapped with 81 of the 94 associated and suggestively associated SNPs with performance traits obtained from GWAS in this study: 206 QTLs associated with body weight, 39 with postnatal growth, 4 with feed intake, 3 with feed conversion, and 1 with feed efficiency. The complete QTL list that overlapped with these regions can be seen in the APPENDIX G and the graphical representation of the suggestive and significant SNPs distribution in relation to the QTLs can be observed in Figure 3.4.

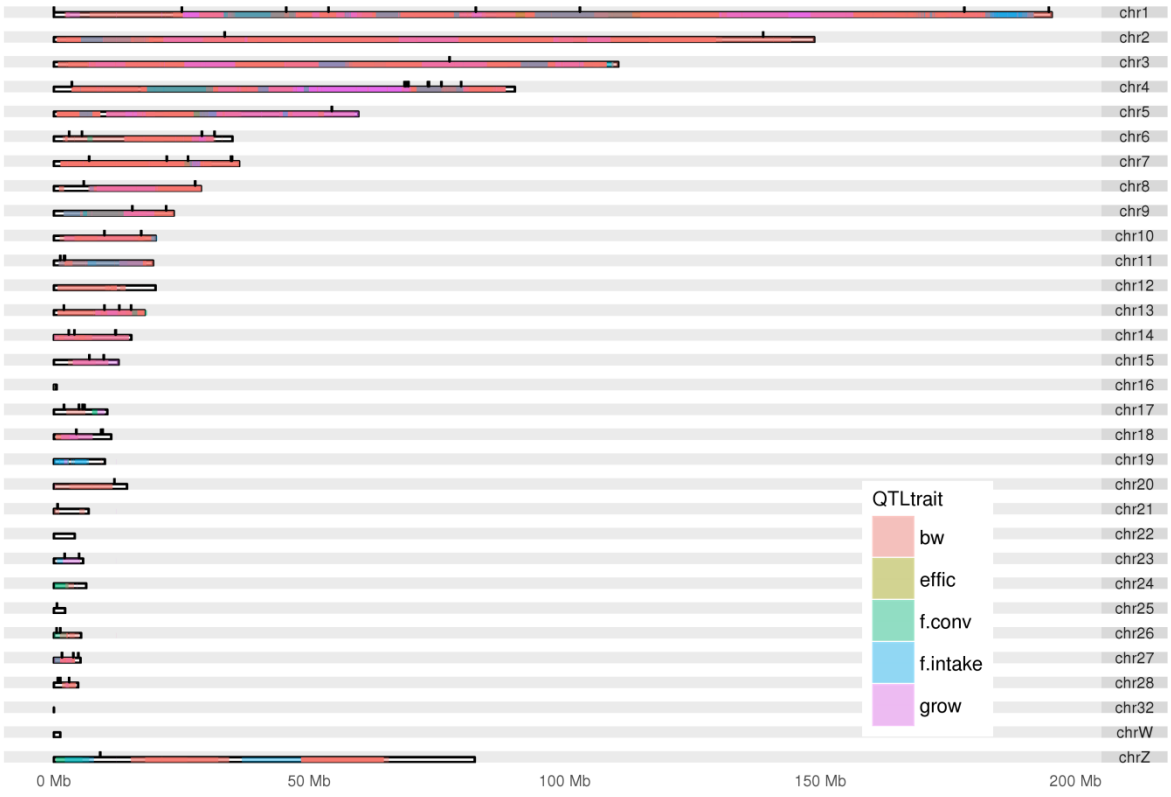


Figure 3.4 - Karyotype of the QTLs (from Animal QTLdb) distribution regions of the chicken genome overlapping suggestive and significant SNPs associated with performance traits (black marks). These QTLs were mapped for body weight (bw), postnatal growth (grow), feed intake (f.intake), feed conversion (f.conv), and feed efficiency (effic). The colors becoming denser according to the number of QTL previously mapped superposing each other.

Table 3.5 - TagSNP significance levels from MLM analyses of each 17 blocks obtained by solid spine of LD in the F_2 population using Haploview 4.2

HaplotypeNº (GGA)	Haplotype	Haplotype Frequency (%)				Trait	p-value		add p-value		dom p-value	
		Parental CC	Parental TT	F_1	F_2							
1 (2)	**AC** †	50	100	75	84	eff.adj35	1.36E-05	*	2.46E-03	*	2.11E-05	*
						f.conv.adj35	5.11E-09	**	2.04E-07	**	1.05E-09	**
2 (4)	**CAGAATG** †	0	60	31	42	bw35	6.19E-06	*	3.26E-05	*	1.13E-02	*
	AGTGTCA	80	0	0	40	bw41	4.70E-06	*	5.31E-06	*	1.25E-02	*
						bw42	4.26E-06	*	5.85E-06	*	9.54E-03	*
3 (4)	**TAG** †	0	80	31	44	bw35	3.08E-07	**	1.95E-07	**	8.31E-02	*
	CGA	90	0	0	38	bw41	1.40E-06	*	1.16E-06	*	6.62E-02	*
						bw42	1.24E-06	*	1.11E-06	*	5.91E-02	*
4 (7)	**GC** †	100	70	94	94	f.conv.adj35	1.94E-06	*	3.38E-07	**	7.30E-07	**
5 (7)	AG	100	80	88	93	bw35	9.91E-06	*	3.63E-06	*	4.53E-03	*
						f.conv.adj35	3.70E-06	*	6.96E-07	**	1.18E-06	*
6 (11)	**GGA** †	100	90	88	93	f.conv.adj35	1.94E-06	*	3.98E-07	**	7.29E-06	*
7 (13)	**CC** †	80	90	75	78	f.conv.adj35	3.67E-06	*	5.76E-07	**	9.02E-07	*
8 (13)	**TC** †	90	100	88	90	f.conv.adj35	3.63E-06	*	7.39E-07	**	7.51E-07	**
9 (14)	**TGC** †	100	60	68	85	f.conv.adj35	3.31E-07	**	8.88E-08	**	6.11E-08	**
10 (17)	**CGT** †	70	100	88	85	f.conv.adj35	3.15E-06	*	5.18E-07	**	1.48E-06	*
11 (17)	**CC** †	100	90	94	95	f.conv.adj35	2.94E-06	*	7.38E-07	**	6.73E-07	**
12 (18)	**ACC** †	90	70	94	93	f.conv.adj35	3.31E-06	*	5.16E-07	**	3.09E-06	*
13 (20)	**TC** †	70	90	75	78	f.conv.adj35	3.45E-06	*	5.45E-07	**	1.07E-06	*
14 (26)	**CA** †	100	60	75	79	f.conv.adj35	3.12E-07	**	4.54E-08	**	1.75E-06	*
						f.int.adj35	1.04E-05	*	1.36E-04	*	1.73E-06	*
15 (27)	**CAGGGAACCA** †	100	90	88	95	f.conv.adj35	9.00E-07	*	3.26E-07	**	1.77E-05	*
16 (27)	**CG** †	100	80	63	94	f.conv.adj35	2.15E-07	**	1.38E-07	**	4.91E-08	**
17 (28)	CA †	70	10	44	35	bw41	2.22E-06	*	4.57E-07	**	1.58E-01	*
	TG	20	50	31	30	bw42	3.16E-06	*	6.73E-07	**	1.77E-01	*
	CG	10	40	25	34							

(GGA) is the *Gallus gallus* chromosome number; †Most frequent haplotype in the population; bolded is the advantage haplotype to individuals compared to the assessed trait; additive (add) and dominant (dom) p-values; * suggestive genome-wise significance (P < 1.57E-05) and ** genome-wise significance (P < 7.86E-07). Abbreviations: eff.adj35 (feed efficiency adjusted for body weight at 35 days), f.conv.adj35 (feed conversion adjusted for body weight at 35 days), f.int.adj35 (feed intake adjusted for body weight at 35 days), all between 35 and 41 days of age and adjusted for 35 days; bw35, bw41, bw42 is body weight at 35, 41 and 42 days of age, respectively

3.4 Discussion

For better understanding of complex traits control in a segregating F_2 population, our research group have focused the attention on genetic association and linkage analyses using different approaches, as: candidate genes (BOSCHIERO et al., 2013; FELÍCIO et al., 2013a, 2013b; PÉRTILLE et al., 2015) and QTL mapping (NONES et al., 2006, 2012; AMBO et al., 2008, 2009; CAMPOS et al., 2009), respectively, and more recently, NGS approaches (GODOY et al., 2015; MOREIRA et al., 2015). We have presented here the first study using a higher density of SNPs in this F_2 population with GWAS purpose. Therefore, we have optimized a method called CornellGBS in chickens (PÉRTILLE et al., 2016) to overcome the concept of pre-designed panels, since we planned a method for genotyping efficiently a specific dataset of SNPs in our specific population.

CornellGBS is a widely employed method to genotype large genomes of model and non-model species exploring important regions in the genome (HE et al., 2014) as microchromosomes previously mentioned. This is due to the high coverage of *tags* (contigs) depending on the number of sequenced individuals of the reduced genome by restriction enzyme cleavage providing a specific SNP profile (PÉRTILLE et al., 2016). The CornellGBS technique was previously developed for inbreeding population and it is known by its general low sequencing coverage, which can cause significant loss of SNPs, mainly heterozygous (GLAUBITZ et al., 2014). For our outbreed population, we used a reasonable multiplex of individuals (~48 animals per lane of Illumina flowcell) to maintain a reasonable sequencing coverage per individual (~5X). We also reduced the loci call rate and use imputation to increase the number of SNPs genotyped. The reduction in the loci call rate was also applied in a recent study that used the same *PstI* restriction enzyme to cleave the cattle genome (DE DONATO et al., 2013). Furthermore, it was already mentioned that the combination of GBS and imputation of missing internal SNPs in haplotype blocks procedures can promote a cost reduction by allowing further reduction of the filtering criteria or sequencing coverage without causing losses in SNP calls (HE et al., 2014). Using this strategy, we doubled the number of SNPs, successfully imputing all lost genotypes (increasing the individual call rate to 100%), the validation ratio remained >90%, and the percentage of heterozygous genotypes in our population had an increase of approximately 2% after the imputation.

The use of the GBS SNP panel for GWAS in our outbred F_2 crosses resulted in 20 SNPs associated ($P < 7.86E-07$) with feed conversion adj.35, one SNP associated with body weight at 35 days (BW35) and other 93 SNPs suggestively associated ($P < 1.57E-05$) with different

performance traits (Table 3.1). Additionally, we noticed that all the evaluated traits, presented an up deviation of the theoretical quantiles (Figure 3.2 and APPENDIX B) of the probability distributions between expected and observed p-values, indicating the existence of QTLs. These results corroborated the Manhattan plot peaks of associated SNPs, indicating that these traits had part of the phenotypic variation significantly explained by the genetic component (BRADBURY et al., 2007). Interestingly, we detected association for several new QTLs located in microchromosomes (GGA11-28). This was only possible because of the distribution of the SNPs used. From our set of SNPs, 38.93% are located in large chromosomes (GGA1-5), 14.15% in intermediate size (GGA6-10), and most, 46.90% are located in microchromosomes (GGA11-28), confirming the microchromosome enrichment mentioned before (PÉRTILLE et al., 2016). Feed conversion, for exemple, had a high number of significant SNPs ($P < 7.86E-07$), mainly located in microchromosomes (GGA8, 10, 14, 18, 23, 26, and 27). However, for this trait, the SNP peaks observed by Manhatan plot, in large and intemediate size chromosomes (Figure 3.2a), were not well defined, as is usually observed for QTLs peaks (FRAGOMENI, et al., 2014). We believe that this is explained by the SNP profile used in this study, which has a lower density of SNPs on large chromosomes compared to microchromosomes (PÉRTILLE et al., 2016). Moreover, feed intake is a complex trait subject to a high residual effect and controlled by several genes with a small effect, which require a large sample size to detect associations (YANG et al., 2011; YUAN, et al., 2015). This small effect also was previously attributed to the short period used to measure this trait (between 35 and 41 days of age) impairing the animal adaptation to the new enviromental condiction (AMBO et al., 2009). On this account, reliable QTLs for this trait were detected in studies that used a larger sampling size (1,534 individuals) for a longer feed intake evaluation period (~ 4 weeks) (TUISKULA-HAAVISTO et al., 2004, 2011) than used in this study.

The GBS strategy can result in clusters of SNPs next to each other (PÉRTILLE et al., 2016) that can indicate genomic regions sheltering possible causal mutations (LEDUR et al., 2010). In order to better define the QTL regions we performed LD analysis. The block pattern between the SNPs matched with our F_2 population structure (Figure 3) (ROSÁRIO et al., 2009). This allowed us to define plenty of possibilities for genetic selection of the lines that did not present the genotype fixed giving attention to the different phenotypic abilities between the CC layer line and the TT broiler line (NONES et al., 2006; ROSÁRIO et al., 2009). As for the blocks 2 and 13, which had variable genotypes for both the CC and the TT lines, and the paternal line presented the favorable genotype most frequent in both cases (Appendix C and Figure 3.3).

Also is important to check if these SNPs are within QTL regions previously published. In the past years, many studies identified QTLs associated with performance traits in different chickens populations (TATSUDA; FUJINAKA, 2001; KONING et al., 2003; IKEOBI et al., 2004; ZHOU et al., 2006; NONES et al., 2006, 2012; AMBO et al., 2008, 2009; BOSCHIERO et al., 2013; NASSAR; GORAGA; BROCKMANN, 2012, 2013) (1,458 QTLs described in the Animal QTLdb) aiming to map loci that control these traits. Recently, to better understand these loci, studies have also applied GWAS with performance traits in chickens (ZHOU et al., 2006; GU et al., 2011; XIE et al., 2012). The validation of single SNP position obtained by GWAS overlapping with QTL regions can confirm interesting genomic regions to explore. From the 94 genome-wise associated and suggestively associated SNPs with the performance traits analyzed in this study, most of them were fairly distributed in mapped QTL regions in the chicken genome (Figure 3.4). Only 13 SNPs did not overlap with QTL regions previously mapped. From these 13 SNPs, one was located on chromosome 1, one in chromosome 8 (GGA1 and 8), one in the Z sex chromosome (GGAZ) and 10 were located in microchromosomes (GGA17, 18, 20, 25, 27 and 28), which confirms the microchromosome enrichment profile obtained by this approach (PÉRTILLE et al., 2016) and suggests novel QTLs to be explored in these regions. It is also important to mention that most of these 13 SNPs (those located on GGA1, 8, 18, 20, 25 and 27) were associated ($P < 7.86E-07$) or suggestively associated ($P < 1.57E-05$) with feed conversion adj35, one with feed efficiency adj35 (GGA17), and two with body weight at 41 days (GGA27) (see APPENDIX A for details). The genes where these SNPs are located are mainly related with cell cycle and metabolic pathways (according to the Reactome pathways - http://www.reactome.org/PathwayBrowser) and were within introns, upstream and downstream of these genes (see APPENDIX H for functional annotation). Despite the importance of the overlap test performed here, previous studies in QTL mapping usually had large confidence intervals (>1Mbps) and often encompassing several genes, making difficult the selection of candidate genes (LEDUR et al., 2010). Therefore, we also checked the overlap of these SNPs only with QTLs mapped using specifically the same traits and the same F_2 population (NONES et al., 2006; AMBO et al., 2009; BOSCHIERO et al., 2013; FELÍCIO et al., 2013b) used in this study. From 23 different QTL intervals, we identified 12 SNPs overlapping with seven of them (see the QTLs bolded in APPENDIX G and I to check the QTL list). It is worth mentioning the SNPs located near the QTL regions, or flanking regions (APPENDIX D). On GGA1, for exemple one SNP (marker 6, APPENDIX A) associated with feed intake ($P=3,83E-07$) overlaped with one QTL mapped for the same trait (TUISKULA-HAAVISTO et al., 2004) in another population, and also with 6 QTLs mapped for body weight

at different ages (LIU et al., 2007) and one with feed efficiency (HANSEN et al., 2005). On the other hand, on GGA4, a well studied chromosome in chickens (KONING et al., 2003; RABIE et al., 2005; AMBO et al., 2009; BARON et al., 2011; GU et al., 2011; LIU et al., 2013, NASSAR; GORAGA; BROCKMANN, 2012, 2013; SUN et al., 2013; LUO et al., 2013a; PÉRTILLE et al., 2015), we identified three SNPs composing the haplotype 3 (markers 21-23), in which one was associated with BW35 (P < 7.86E-07) and suggestively associated with BW41 (P < 1.57E-05), and two SNPs suggestively associated with BW41 (P < 1.57E-05). These three SNPs overlapped with one QTL region previously mapped in this same population for these same traits (AMBO et al., 2009) (QTL_IDs from ChickenQTLdb = 7157; 7162 and 7185). The boundary SNPs from haplotypes 2 and 3 are separated by a short distance (less than 4 Mbps), but these QTLs are not linked, beside they have effect on the same traits (BW35 and BW41) (Figure 3) in our F_2 population. It is worth to mention, the haplotype 2 that overllaped with QTLs mapped for different BW in different ages (SEWALEM et al., 2002; CARLBORG et al., 2004; PODISI et al., 2011; PODISI et al., 2013; NASSAR et al., 2015) and growth (CARLBORG et al., 2004; ZHOU et al., 2006; NASSAR et al., 2015) traits in different populations.

To the best of our knowledge, we showed the application of the CornellGBS *PstI*-derived SNPs to a GWAS for the first time in chickens. We showed a strategy, changing filtering criteria and subsequent genotype imputation, to increase the number of reliable SNPs to be analyzed. We found 13 SNPs indicating new regions associated with performance traits, mainly in microchromosomes, that have not been previously reported. We improved the available information about loci controlling performance traits and we refined these regions to discover novel candidate regions to be explored. Finally, by demonstrating that GBS is a valid strategy for QTL mapping in a species that has genome sequence and SNP panel available, we can argue the validity of GBS in species without genome resources.

References

AHSAN, M.; LI, X.; LUNDBERG, A.E.; KIERCZAK, M.; SIEGEL, P.B.; CARLBORG, Ö.; MARKLUND, S. Identification of candidate genes and mutations in QTL regions for chicken growth using bioinformatic analysis of NGS and SNP-chip data. **Frontiers in Genetics**, Lausanne, v. 4, p. 1–8, Nov. 2013.

AMBO, M.; CAMPOS, R.L.R.; MOURA, A.S.M.T.; BOSCHIERO, C.; ROSÁRIO, M.F. do; LEDUR, M.C.; NONES, K.; COUTINHO, L.L. Genetic linkage maps of chicken chromosomes 6, 7, 8, 11 and 13 from a Brazilian resource population. **Scientia Agricola**, Piracicaba, v. 65, n. 5, p. 447–452, 2008.

AMBO, M.; MOURA, A.S.A.M.T.; LEDUR, M.C.; PINTO, L.F.B.; BARON, E.E.; RUY, D.C.; NONES, K.; CAMPOS, R.L.R.; BOSCHIERO, C.; BURT, D.W.; COUTINHO, L.L. Quantitative trait loci for performance traits in a broiler x layer cross. **Animal Genetics**, New York, v. 40, n. 2, p. 200–208, Apr. 2009.

BARON, E. E.; MOURA, a S. a M. T.; LEDUR, M. C.; PINTO, L. F. B.; BOSCHIERO, C.; RUY, D. C.; NONES, K.; ZANELLA, E. L.; ROSÁRIO, M. F.; BURT, D. W.; COUTINHO, L. L. QTL for percentage of carcass and carcass parts in a broiler x layer cross. **Animal Genetics**, New York, v. 42, n. 2, p. 117–124, 2011.

BARRETT, J.C.; FRY, B.; MALLER, J.; DALY, M.J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, Oxford, v. 21, n. 2, p. 263–265, 2005.

BLACKBURN, H.D. The National Animal Germplasm Program: challenges and opportunities for poultry genetic resources. **Poultry Science**, Oxford, v. 85, n. 2, p. 210–215, 2006.

BOSCHIERO, C.; JORGE, E.C.; NINOV, K.; NONES, K.; DO ROSÁRIO, M.F.; COUTINHO, L.L.; LEDUR, M.C.; BURT, D.W.; MOURA, A.S.A.M.T. Association of IGF1 and KDM5A polymorphisms with performance, fatness and carcass traits in chickens. **Journal of Applied Genetics**, Poznan, v. 54, n. 1, p. 103–112, Feb. 2013.

BRADBURY, P.J.; ZHANG, Z.; KROON, D.E.; CASSTEVENS, T.M.; RAMDOSS, Y.; BUCKLER, E.S. TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**, Oxford, v. 23, n. 19, p. 2633–2635, 2007.

BROWNING, S.R.; BROWNING, B. . Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. **American Journal of Human Genetics**, Baltimore, v. 81, n. 5, p. 1084–1097, 2007.

CARLBORG, O.; HOCKING, P. M.; BURT, D. W; HALEY, C. S. Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. **Genetical research**.Cambridge, v.83**,** n.3, p197–209, 2004.

CAMPOS, R.L.R.; NONES, K.; LEDUR, M.C.; MOURA, A.S.A.M.T.; PINTO, L.F.B.; AMBO, M.; BOSCHIERO, C.; RUY, D.C.; BARON, E.E.; NINOV, K.; ALTENHOFEN, C.A.B.; SILVA, R.A.M.S.; ROSÁRIO, M.F.; BURT, D.W.; COUTINHO, L.L. Quantitative trait loci associated with fatness in a broiler-layer cross. **Animal Genetics**, New York, v. 40, n. 5, p. 729–736, Oct. 2009.

DE DONATO, M.; PETERS, S.O.; MITCHELL, S.E.; HUSSAIN, T.; IMUMORIN, I.G. Genotyping-by-sequencing (GBS): a novel, efficient and cost-effective genotyping method for cattle using next-generation sequencing. **PloS One**, San Francisco v. 8, n. 5, p. e62137, Jan. 2013.

ECK, S.H.; BENET-PAGÈS, A.; FLISIKOWSKI, K.; MEITINGER, T.; FRIES, R.; STROM, T.M. Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. **Genome Biology**, London, v. 10, n. 8, p. R82, 2009.

FRAGOMENI, B. D. O.; MISZTAL, I.; LOURENCO, D. L.; AGUILAR, I.; OKIMOTO, R.;

MUIR, W. M. Changes in variance explained by top SNP windows over generations for three traits in broiler chicken. **Frontiers in Genetics**, Lausanne, v. 5, n. October, p. 1–7, 1 out. 2014.

FELÍCIO, A.M.; BOSCHIERO, C.; BALIEIRO, J.C.C.; LEDUR, M.C.; FERRAZ, J.B.S.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Polymorphisms in FGFBP1 and FGFBP2 genes associated with carcass and meat quality traits in chickens. **Genetics and Molecular Research: GMR**, Ribeirão Preto, v. 12, n. 1, p. 208–22, Jan. 2013a.

FELÍCIO, A.M.; BOSCHIERO, C.; BALIEIRO, J.C.C.; LEDUR, M.C.; FERRAZ, J.B.S.; MICHELAN FILHO, T.; MOURA, A.S.A.M.T.; COUTINHO, L.L. Identification and association of polymorphisms in CAPN1 and CAPN3 candidate genes related to performance and meat quality traits in chickens. **Genetics and Molecular Research: GMR**, Ribeirão Preto, v. 12, n. 1, p. 472–482, jan. 2013b.

FULTON, J.E. Genomic selection for poultry breeding. **Animal Frontiers**, Champaign, v. 2, n. 1, p. 30–36, Jan. 2012.

GAYA, L.G.; FERRAZ, J.B.S.; REZENDE, F.M.; MOURAO, G.B.; MATTOS, E.C.; ELER, J.P.; MICHELAN FILHO, T. Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. **Poultry Science,** Oxford, v. 85, n. 5, p. 837–843, May 2006.

GLAUBITZ, J.C.; CASSTEVENS, T.M.; LU, F.; HARRIMAN, J.; ELSHIRE, R.J.; SUN, Q.; BUCKLER, E.S. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. **PloS One**, San Francisco, v. 9, n. 2, p. e90346, Jan. 2014.

GODOY, T.F.; MOREIRA, G.C.M.; BOSCHIERO, C.; GHEYAS, A.A.; GASPARIN, G.; PADUAN, M.; ANDRADE, S.C.S.; MONTENEGRO, H.; BURT, D.W.; LEDUR, M.C.; COUTINHO, L.L. SNP and INDEL detection in a QTL region on chicken chromosome 2 associated with muscle deposition. **Animal Genetics**, New York, v. 46, n. 2, p. 158–163, 2015.

GU, X.; FENG, C.; MA, L.; SONG, C.; WANG, Y.; DA, Y.; LI, H.; CHEN, K.; YE, S.; GE, C.; HU, X.; LI, N. Genome-wide association study of body weight in chicken F2 Resource population. **PloS One**, San Francisco, v. 6, n. 7, p. e21872, July 2011.

HABERMANN, F.A.; CREMER, M.; WALTER, J.; KRETH, G.; VON HASE, J.; BAUER, K.; WIENBERG, J.; CREMER, C.; CREMER, T.; SOLOVEI, I. Arrangements of macro- and microchromosomes in chicken cells. **Chromosome Research: an International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology**, Oxford, v. 9, n. 7, p. 569–584, 2001.

HANSEN, C.; YI, N.; ZHANG, Y. M.; XU, S.; GAVORA, J.; CHENG, H. H. Identification of QTL for production traits in chickens. **Animal biotechnology**, New Yourk, v. 16, n. 1, p. 67–79, 2005.

HE, J.; ZHAO, X.; LAROCHE, A.; LU, Z.-X.; LIU, H.; LI, Z. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. **Frontiers in Plant Science**, Lausanne, v. 5, p. 484, Sept. 2014.

HU, Z.-L.; PARK, C.A.; WU, X.-L.; REECY, J.M. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, Oxford, v. 41, n. D1, p. D871–D879, Jan. 2013.

IKEOBI, C.O.; WOOLLIAMS, J.; MORRICE, D.; LAW, A.; WINDSOR, D.; BURT, D.; HOCKING, P. Quantitative trait loci for meat yield and muscle distribution in a broiler layer cross. **Livestock Production Science**, Amsterdam, v. 87, n. 2/3, p. 143–151, May 2004.

KONING, D.J. de; WINDSOR, D.; HOCKING, P.M.; BURT, D.W.A.; LAW, C.S.H.A.; MORRIS, J.; GRIFFIN, V.H. Quantitative trait locus detection in commercial broiler lines using candidate regions. **Journal of Animal Science**, Champaign, v. 81, p. 1158–1165, July 2003.

LEDUR, M.C.; NAVARRO, N.; PÉREZ-ENCISO, M. Large-scale SNP genotyping in crosses between outbred lines: how useful is it? **Heredity**, London, v. 105, n. 2, p. 173–182, Aug. 2010.

LI, X.; LIU, X.; NADAF, J.; LE BIHAN-DUVAL, E.; BERRI, C.; DUNN, I.C.; TALBOT, R.; DE KONING, D.J. Using targeted re-sequencing for identification of candidate genes and SNPs for a QTL affecting the pH value of chicken muscle. **bioRxiv**, Cold Spring Harbor, v. 5, p. 17186, Oct. 2015.

LIU, X.; LI, H.; WANG, S.; HU, X.; GAO, Y.; WANG, Q.; LI, N.; WANG, Y.; ZHANG, H. Mapping quantitative trait loci affecting body weight and abdominal fat weight on chicken chromosome one. **Poultry science**, Oxford, v. 86, n. 6, p. 1084–9, jun. 2007.

LIU, R.; SUN, Y.; ZHAO, G.; WANG, F.; WU, D.; ZHENG, M.; CHEN, J.; ZHANG, L.; HU, Y.; WEN, J. Genome-wide association study identifies loci and candidate genes for body composition and meat quality traits in Beijing-you chickens. **PLoS One**, San Francisco, v. 8, n. 4, p. e61172, Apr. 2013.

LUO, C.; QU, H.; MA, J.; WANG, J.; LI, C.; YANG, C.; HU, X.; LI, N.; SHU, D. Genome-wide association study of antibody response to Newcastle disease virus in chicken. **BMC Genetics**, London, v. 14, n. 1, p. 42, Jan. 2013a.

LUO, C.; QU, H.; WANG, J.; WANG, Y.; MA, J.; LI, C.; YANG, C.; HU, X.; LI, N.; SHU, D. Genetic parameters and genome-wide association study of hyperpigmentation of the visceral peritoneum in chickens. **BMC Genomics**, London, v. 14, n. 1, p. 334, Jan. 2013b.

MOREIRA, G.C.M.; GODOY, T.F.; BOSCHIERO, C.; GHEYAS, A.; GASPARIN, G.; ANDRADE, S.C.S.; PADUAN, M.; MONTENEGRO, H.; BURT, D.W.; LEDUR, M.C.; COUTINHO, L.L. Variant discovery in a QTL region on chromosome 3 associated with fatness in chickens. **Animal Genetics**, New York, v. 46, n. 2, p. 141–147, 2015.

MORRIS, G.P.; RAMU, P.; DESHPANDE, S.P.; HASH, C.T.; SHAH, T.; UPADHYAYA, H.D.; RIERA-LIZARAZU, O.; BROWN, P.J.; ACHARYA, C.B.; MITCHELL, S.E.; HARRIMAN, J.; GLAUBITZ, J.C.; BUCKLER, E.S.; KRESOVICH, S. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 110, n. 2, p. 453–458, 2013.

NASSAR, M.K.; GORAGA, Z.S.; BROCKMANN, G.A. Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: II. Muscle weight and carcass composition. **Animal Genetics**, New York, v. 43, n. 6, p. 739–745, Dec. 2012.

_____. Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: III. Fat deposition and intramuscular fat content. **Animal Genetics**, New York, v. 44, n. 1, p. 62–68, Feb. 2013.

_____. Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: IV. Growth performance. **Animal Genetics**, New York, v. 46, n. 4, p. 441–46, Nov. 2015.

NICODEMUS, K.K.; LIU, W.; CHASE, G.A.; TSAI, Y.-Y.; FALLIN, M.D. Comparison of type I error for multiple test corrections in large single-nucleotide polymorphism studies using principal components versus haplotype blocking algorithms. **BMC Genetics**, London, v. 6, suppl. 1, p. S78, 2005.

NIE, Q.-H.; ZHANG, X.-Q.; LEI, M.-M. Single nucleotide polymorphism and its use in chicken QTL mapping. **Yi chuan = Hereditas / Zhongguo yi chuan xue hui bian ji**, Beijing, v. 25, n. 6, p. 729–734, Nov. 2003.

NONES, K.; LEDUR, M.C.; RUY, D.C.; BARON, E.E.; MELO, C.M.R.; MOURA, A.S.A.M.T.; ZANELLA, E.L.; BURT, D.W.; COUTINHO, L.L. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. **Animal Genetics**, New York, v. 37, n. 2, p. 95–100, Apr. 2006.

NONES, K.; LEDUR, M.C.; ZANELLA, E.L.; KLEIN, C.; PINTO, L.F.B.; MOURA, A.S.A.M.T.; RUY, D.C.; BARON, E.E.; AMBO, M.; CAMPOS, R.L.R.; BOSCHIERO, C.; BURT, D.W.; COUTINHO, L.L. Quantitative trait loci associated with chemical composition of the chicken carcass. **Animal Genetics**, New York, v. 43, n. 5, p. 570–576, Oct. 2012.

PARK, M.N.; CHOI, J.A.; LEE, K.-T.; LEE, H.-J.; CHOI, B.-H.; KIM, H.; KIM, T.-H.; CHO, S.; LEE, T. Genome-wide association study of chicken plumage pigmentation. **Asian-Australasian journal of animal sciences**, Seoul, v. 26, n. 11, p. 1523–1528, Nov. 2013.

PÉRTILLE, F.; ZANELLA, R.; FELÍCIO, A.M.; LEDUR, M.C.; PEIXOTO, J.O.; COUTINHO, L.L. Identification of polymorphisms associated with production traits on chicken (*Gallus gallus*) chromosome 4. **Genetics and Molecular Research**, Ribeirão Preto, v. 14, n. 3, p. 10717–10728, 2015.

PÉRTILLE, F.; GUERRERO-BOSAGNA, C.; DA SILVA, V.H.; BOSCHIERO, C.; NUNES, J.R.; LEDUR, M.C.; JENSEN, P.; COUTINHO, L.L. High-throughput and cost-effective chicken genotyping using next-generation sequencing. **Scientific Reports**, London, v. 6, p. 26929, Jan. 2016.

PODISI, B. K.; KNOTT, S. A.; DUNN, I. C.; LAW, A. S.; BURT, D. W.; HOCKING, P. M. Overlap of quantitative trait loci for early growth rate, and for body weight and age at onset of sexual maturity in chickens. **Reproduction**, Cambridge, v. 141, n. 3, p. 381–389, 1 mar. 2011.

PODISI, B. K.; KNOTT, S. A.; BURT, D. W.; HOCKING, P. M. Comparative analysis of

quantitative trait loci for body weight, growth rate and growth curve parameters from 3 to 72 weeks of age in female chickens of a broiler–layer cross. **BMC Genetics**, London, v. 14, n. 1, p. 22, 2013.

ROSÁRIO, M.F. do; LEDUR, M.C.; MOURA, A.S.A.M.T.; COUTINHO, L.L.; GARCIA, A.A.F. Genotypic characterization of microsatellite markers in broiler and layer selected chicken lines and their reciprocal F1s. **Scientia Agricola**, Piracicaba, v. 66, n. 2, p. 150–158, abr. 2009.

ROUX, P.F.; BOUTIN, M.; DÉSERT, C.; DJARI, A.; ESQUERRÉ, D.; KLOPP, C.; LAGARRIGUE, S.; DEMEURE, O. Re-sequencing data for refining candidate genes and polymorphisms in QTL regions affecting adiposity in chicken. **PLoS ONE**, San Francisco, v. 9, n. 10, p. e111299, 2014.

ROVADOSCKI, G. A.; PETRINI, J.; RAMIREZ-DIAZ, J.; PERTILE, S. F. N.; PERTILLE, F.; SALVIAN, M.; IUNG, L. H. S.; RODRIGUEZ, M. A. P.; ZAMPAR, A.; GAYA, L. G.; CARVALHO, R. S. B.; COELHO, A. A. D.; SAVINO, V. J. M.; COUTINHO, L. L.; MOURÃO, G. B. Genetic parameters for growth characteristics of free-range chickens under univariate random regression models. **Poultry Science**, Oxford, v. 95, n. 9, p. 1989–1998, set. 2016.

SEWALEM, A.; MORRICE, D. M.; LAW, A.; WINDSOR, D.; HALEY, C. S.; IKEOBI, C. O. N.; BURT, D. W.; HOCKING, P. M. Mapping of quantitative trait loci for body weight at three, six, and nine weeks of age in a broiler layer cross. **Poultry science**, Oxford, v. 81, n. 12, p. 1775–81, dez. 2002.

SHEN, X.; ZENG, H.; XIE, L.; HE, J.; LI, J.; XIE, X.; LUO, C.; XU, H.; ZHOU, M.; NIE, Q.; ZHANG, X. The GTPase activating Rap/RanGAP domain-like 1 gene is associated with chicken reproductive traits. **PloS one**, San Francisco, v. 7, n. 4, p. e33851, Jan. 2012.

SMITH, J.; BRULEY, C.K.; PATON, I.R.; DUNN, I.; JONES, C.T.; WINDSOR, D.; MORRICE, D.R.; LAW, A.S.; MASABANDA, J.; SAZANOV, A.; WADDINGTON, D.; FRIES, R.; BURT, D.W. Differences in gene density on chicken macrochromosomes and microchromosomes. **Animal Genetics**, New York, v. 31, n. 2, p. 96–103, 2000.

SUN, Y.; LIU, R.; ZHAO, G.; ZHENG, M.; SUN, Y.; YU, X.; LI, P.; WEN, J. Genome-wide linkage analysis and association study identifies loci for polydactyly in chickens. **G3**, Bethesda, v. 4, n. 6, p. 1167–1172, June 2014.

SUN, Y.; ZHAO, G.; LIU, R.; ZHENG, M.; HU, Y.; WU, D.; ZHANG, L.; LI, P.; WEN, J. The identification of 14 new genes for meat quality traits in chicken using a genome-wide association study. **BMC Genomics**, London, v. 14, n. 1, p. 458, Jan. 2013

TATSUDA, K.; FUJINAKA, K. Genetic mapping of the QTL affecting body weight in chickens using a F 2 family. **British Poultry Science**, Edinburgh, v. 42, n. 3, p. 333–337, July 2001.

TUISKULA-HAAVISTO, M.; DE KONING, D.-J.; HONKATUKIA, M.; SCHULMAN, N. F.; MÄKI-TANILA, A.; VILKKI, J. Quantitative trait loci with parent-of-origin effects in chicken. **Genetical research**, .Cambridge, v. 84, n. 1, p. 57–66, ago. 2004.

TUISKULA-HAAVISTO, M.; HONKATUKIA, M.; VILKKI, J.; DE KONING, D. J.; SCHULMAN, N. F.; MÄKI-TANILA, A. Mapping of quantitative trait loci affecting quality and production traits in egg layers. **Poultry science**, Oxford, v. 81, n. 7, p. 919–27, jul. 2002.

YANG, J.; WEEDON, M. N.; PURCELL, S.; LETTRE, G.; ESTRADA, K.; WILLER, C. J.; SMITH, A. V; INGELSSON, E.; O'CONNELL, J. R.; MANGINO, M.; MÄGI, R.; MADDEN, P. A.; HEATH, A. C.; NYHOLT, D. R.; MARTIN, N. G.; MONTGOMERY, G. W.; FRAYLING, T. M.; HIRSCHHORN, J. N.; MCCARTHY, M. I.; GODDARD, M. E.; VISSCHER, P. M. Genomic inflation factors under polygenic inheritance. **European Journal of Human Genetics**, New York, v. 19, n. 7, p. 807–812, 16 jul. 2011.

YUAN, J.; WANG, K.; YI, G.; MA, M.; DOU, T.; SUN, C.; QU, L.-J.; SHEN, M.; QU, L.; YANG, N. Genome-wide association studies for feed intake and efficiency in two laying periods of chickens. **Genetics Selection Evolution**, Amsterdam, v. 47, n. 1, p. 82, 16 dez. 2015.

XIE, L.; LUO, C.; ZHANG, C.; ZHANG, R.; TANG, J.; NIE, Q.; MA, L.; HU, X.; LI, N.; DA, Y.; ZHANG, X. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. **PloS One**, San Francisco, v. 7, n. 2, p. e30910, Jan. 2012.

ZHOU, H.; DEEB, N.; EVOCK-CLOVER, C.M.; ASHWELL, C.M.; LAMONT, S.J. Genome-wide linkage analysis to identify chromosomal regions affecting phenotypic traits in the chicken. I. Growth and average daily gain. **Poultry Science**, Oxford, v. 85, n. 10, p. 1700–1711, Oct. 2006.

4 DNA METHYLATION PROFILES DETECTED IN RED BLOOD CELLS OF ADULT HENS CORRELATE TO THEIR REARING CONDITIONS

Abstract

Stressful conditions are common in the environment where production animals are raised. Stress in animals is usually determined by the levels of stress-related hormones. A big challenge, however, is in determining the history of exposure of an organism to stress, because the release of stress hormones can show an acute (and recent) but not a sustained exposure to stress. Epigenetic tools provide an alternative option to evaluate past exposure to long-term stress. Chickens provide a unique model to study stress effects in the epigenome of red blood cells (RBC), a cell type of easy access and nucleated in birds. The present study investigates in chickens whether two different rearing conditions can be identified by looking at DNA methylation patterns in their RBCs later in life. These conditions are rearing in open aviaries versus in cages, which are likely to differ regarding the amount of stress they generate. Our comparison revealed 115 genomic windows with significant change in RBCs DNA methylation between experimental groups, which were located around 53 genes and within 22 intronic regions. Our results set the ground for future detection of long-term stress in live production animals by measuring DNA methylation in a cell type of easy accessibility.

Keywords: Stress; Red blood cells; Epigenetics; Chicken; DNA methylation; Animal welfare

4.1 Introduction

Stress in production animals generated by unsustainable production methods is a frequent issue of concern. Besides the ethical issue of inducing unnecessary stress in animals, detrimental practices in the animal production industry have consequences from a human health perspective (ROSTAGNO, 2009). The environment where production animals are reared influences not only their later health and wellbeing but also the quality of the food originating from them (BROOM, 2010). Stressful conditions to which production animals can be subjected include extreme illumination patterns (MORGAN; TROMBORG, 2006; OLANREWAJU et al., 2006), social isolation or crowding (GOERLICH et al., 2012), food restriction (SAVORY; LARIVIERE, 2000; MORGAN; TROMBORG, 2006), too high or too low temperatures, restriction of movement, barren environments, and lack of appropriate substrates for foraging, exploration and manipulation (MORGAN; TROMBORG, 2006).

Stress in animals is associated with a cascade of hormonal responses (HENRY, 1992). The primary physiological stress response observed is an increase in the hypothalamic-pituitary-adrenal (HPA) axis activity, which results in elevated levels of the glucocorticoids (FALLAHSHAROUDI et al., 2015). Initially, increases in testosterone levels related to increased anxiety are observed (HENRY, 1992). Subsequently, decreases in the

noradrenaline/adrenaline ratio are observed, concomitant with increases in adrenaline, prolactin and fatty acids (HENRY, 1992). In conditions of further distress, adrenocorticotropic hormone and cortisol levels will increase (HENRY, 1992). Due to this plethora of hormonal changes generated by stressful conditions, stress in animals is usually determined by the levels of stress-related hormones such as cortisol and adrenaline (ISHIBASHI et al., 2013; MULLER et al., 2013). A big challenge, however, is in determining the history of the exposure of an organism to stress, given that the release of stress hormones can show an acute (and recent) but not a sustained exposure to stressful conditions (HENRY, 1992).

An alternative option to the use of hormonal measurements to evaluate past exposure to long-term stress could be to utilize epigenetic tools instead. Epigenetics involves studying how environmental exposures affect gene regulation during the lifetime of organisms. Epigenetic changes are defined as accessory chemical modifications of the DNA that regulate gene expression and are mitotically stable (SKINNER et al., 2010). These modifications include DNA methylation or hydroxymethylation of CG dinucleotides, chemical modifications of histones, interaction of DNA with small RNAs, or states of chromatin condensation (FEIL; FRAGA, 2011;TEPEREK-TKACZ et al., 2011; DENHAM et al., 2014). Altering epigenetic states can lead to distinguishable phenotypic consequences such as changes in the coat color (DOLINOY et al., 2007) or increased propensity to diseases (GUERRERO-BOSAGNA; SKINNER, 2012). A variety of organism models has been used in epigenetic research, including laboratory rodents (DOLINOY et al., 2007; GUERRERO-BOSAGNA et al., 2008; SUSIARJO et al., 2013), flies (SEONG et al., 2011), honey bees (LYKO et al., 2010; DICKMAN et al., 2013), plants (CUBAS et al., 1999; MANNING et al., 2006) and yeast (ZHANG et al., 2013). However, in spite of the importance of epigenetic mechanisms in biology in general, epigenetic studies in production animals are scarce. Among production animals, chickens have been suggested as a promising model for epigenetic studies (FRESARD et al., 2013). Two important reasons for this are that chickens have had their genome extensively sequenced (RUBIN et al., 2010) and have historically been an important model for translational research with implications for human health and physiology (KAIN et al., 2014).

Long term stress is known to generate life-long changes in stress susceptibility that is correlated to epigenetic changes (JENSEN, 2014). Thus, it is expected that if animals are constantly subjected to stress and systemic hormonal changes, this exposure will imprint the epigenome of blood cells. Epigenetic changes in blood cells will then serve as markers of past exposure to stress. Research in humans (MALAN-MULLER et al., 2014) and monkeys (PROVENCAL et al., 2012) have shown that stress affects DNA methylation in blood cells.

The epigenome of blood cells can provide a meaningful assessment of biological processes involved in stress because disruptions of the HPA-axis have systemic consequences (ZANNAS; WEST, 2014). Since different practices in the production environment will generate different levels of stress in animals, it is practical (from the perspective of evaluation of long term stress) to understand how stress correlates with specific profiles in production animals.

The present study aims at investigating in chickens whether two different rearing conditions can be identified by looking at epigenetic patterns in their red blood cells (RBCs) later in life. It is important to point out that unlike mammals, birds contain nucleated RBCs. The conditions tested are rearing in a system of open aviaries versus rearing in cages. These two different rearing conditions are likely to differ with regards to the amount of stress to which birds are exposed, as suggested by the observation that they cause long-term differences in fearfulness (BRANTSAETER et al., 2016) and cognitive function (TAHAMTANI et al., 2015). The objective of using this model is to generate a proof-of-concept for future detection of long-term stress in production animals using epigenetic measurements in cell types of easy accessibility in live animals. The identification of a correlation between RBCs epigenetic profiles and long-term stress will overcome limitations that exist when evaluating stress through hormonal levels or visual health assessments, which do not provide reliable accounts of long-term stress.

In order to identify DNA methylation profiles related to different rearing conditions in chickens, we compared RBCs DNA methylation in a group of birds reared in cages (a common housing system, with low environmental complexity) with that of birds reared in open aviaries (which represents a complex environment). Previous studies have shown that chickens reared in a complex aviary system are less fearful, use elevated areas of the pen more often as adults (BRANTSAETER et al., 2016), and have better spatial working memory (TAHAMTANI et al., 2015) than laying hens reared in a simpler cage environment (BRANTSAETER et al., 2016). The present study tests whether the different rearing conditions applied, which associate with different levels of environmental complexity, stimulation of cognitive capabilities and responses to stress, will have long term effects in the blood methylome of chickens. Our comparison revealed 115 genomic windows with significant change in RBCs DNA methylation between experimental groups, which were located within or in the vicinity of 53 genes and within 22 intronic regions. Our results set the ground for future detection of long-term stress in live production animals by measuring DNA methylation in a cell type of easy accessibility. The present results can be used as a proof-of-concept for the future identification of epigenetic marks related to past stress conditions that occur in the production environment.

4.2 Methods

4.2.1 Subjects and rearing treatments

The study was conducted using non beak-trimmed, female Dekalb white chickens (*Gallus gallus domesticus*), aged 0–23 weeks with normal health status. Birds were hatched at a commercial hatchery and immediately brought to a rearing farm. All birds were housed within the same room. Initially, all birds were kept confined inside the aviary row, with access to food and water. When the birds were four weeks of age, access to the aviary corridors was given to half of them, as this is the normal procedure in aviary rearing systems. This group was named "aviary reared-birds" (AV). The other half of the birds was kept under confinement at the aviary row for the entire rearing period. The group of birds staying inside the cages was named "cage-reared birds" (CG). These two rearing conditions were maintained until the birds were 16 weeks of age. After the rearing period had ended, a random subset of birds from each treatment was moved to the experimental facilities for blood sampling, which occurred at 24 weeks of age. A schematic representation of the experimental design is shown in Figure 4.1.

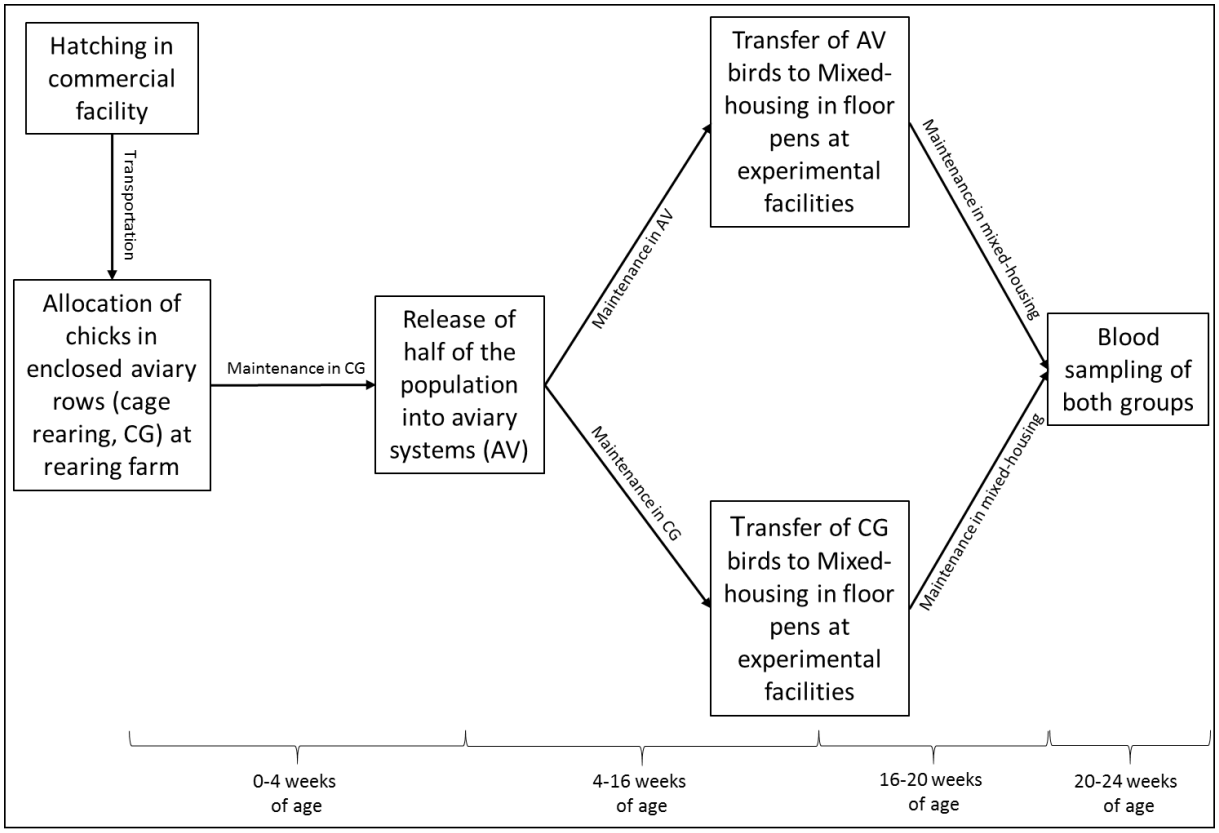


Figure 4.1 - Schematic representation of the housing conditions in the two experimental treatments

4.2.2 Rearing system conditions

The housing system in the single room in which all birds were housed was Natura Primus 1600 (Big Dutchman; http://www.bigdutchmanusa.com) designed for aviary-rearing of laying hen pullets. This system consisted of cages stacked in three tiers placed on either side of a corridor for allowing inspection by the caretaker. Cage dimensions were 120 cm × 80 cm × 60 cm (length × width × height). Each aviary cage contained a 120 cm feed trough, one 120 cm perch, and five drinking nipples. All the cages could be opened at the front, allowing the birds to move freely between each tier and the floor of the corridor. Ramps run from the floor to the second tier to increase ease of access for pullets. When cage doors are in the open position, perches extend from the front of the first and second tiers. The density was 25 birds/m^2 for both treatments during the first four weeks of life. Chick paper covered 30% of the wire mesh floor of the cages in sufficient amounts to last until the birds were released out in the corridors.

During rearing, all birds were exposed to the same light intensity, light schedule, and temperatures, as recommended by the General Management Guide for Dekalb White Commercial Layer (HENDRIX, 2015). They were provided with ad libitum access to feed using a chain dispersal system and ad libitum access to water. The feed type was conventional pullet feed produced and sold by Felleskjøpet, Norway ("Kromat oppdrett 1" for 0- to 6-week-old birds, "Kromat avl egg 1" for 6- to 8-week-old birds, and "Kromat oppdrett 2" for 8- to 15-week-old birds).

4.2.3 Blood collection and DNA extraction

Blood samples were collected from 21 individuals (9 AV and 12 CG) of 24 weeks of age. Before blood sampling chickens were sedated using 0.5 ml/Kg Zoletil mix, which contains 10 ml Rompun (Xylazine 20 mg/ml) and 0.75 ml Butomidor (Butorphanol 10 mg/ml) mixed with Zoletil powder (Tiletamine HCL 125 mg and Zolazepam HCL 125 mg). Blood samples were collected as soon as the birds were considered unconscious, which occurred within a maximum timeframe of 10 minutes. After the birds were considered unconscious they were humanely euthanized by cervical dislocation. Each blood sample was collected using a 1 ml syringe and a BD Microlance cannula (21G x ½", 0.80 x 40 mm). A total of 160 µg of blood was then transferred from each sample into two heparinized glass capillaries, which were then centrifuged at 3000 RPM for 5 minutes. After centrifugation, the tubes were manually broken

into two pieces, one of them containing the hematocrit fraction, which was placed inside 1.5 mL micro-centrifuge tubes and stored in a -80°C freezer until further analyses.

DNA extraction was performed through proteinase K digestion. Initially 10 µL of the hematocrit fractions were incubated with 200 µl of extraction buffer (1M Tris-HCL, 0.5 M EDTA, 10% SDS) and 20 µL of 0.1 M DTT at 65°C for 15 min. Then, incubation with 20 µL of proteinase K (20mg/mL) was performed overnight at 55°C under rotation. After proteinase K digestion samples were incubated with Protein Precipitation Solution (Promega) for 15 min on ice and centrifuged for 20 min at 13000 rpm and 4°C in a benchtop microcentrifuge. The supernatants (1 mL) were transferred to new tubes and DNA was precipitated with equal amounts of 100% isopropanol. In addition, 3 µL of glycogen (5 mg/mL) was added to improve further visualization of DNA pellets. After 30 min of incubation at 4°C the samples were centrifuged at 13000 rpm and 4°C for 30 min. The supernatants were discarded and the DNA pellets were washed with ice cold 70% ethanol, followed by centrifugation at 13000 rpm and 4°C for 10 min. The supernatants were discarded again and the pellets in the tubes were dried out in a heating block at 55°C for 5 min. DNA pellets were re-suspended in 200 uL of ultrapure water.

4.2.4 DNA methylation analyses

In order to perform DNA methylation analyses in a cost effective manner, we have combined a Genotyping by Sequencing method (PÉRTILLE et al., 2016) with the technique of Methylated DNA immunoprecipitation (GUERRERO-BOSAGNA; JENSEN, 2015). We have recently described the optimization of each of these two methodologies separately for its use with chicken DNA (GUERRERO-BOSAGNA; JENSEN, 2015; PÉRTILLE et al., 2016). This combination of methods was needed because current methods that assess DNA methylation in reduced genomes perform such a reduction through enzymatic digestion targeting restriction sites that contain CpG sites (GU et al., 2011). Moreover, such an approach is highly biased towards CpG islands (GU et al., 2011). Our approach, instead, reduces the genome by digesting on restriction sites unrelated to CpGs and is unbiased towards CpG islands.

We first digested the genome with *PstI* as previously described (PÉRTILLE et al., 2016). After this fragmentation had generated a significantly reduced genome (approximately 2% of its original size) and enrichment of small fragments in a suitable range for Illumina sequencing (200-500 bp) (PÉRTILLE et al., 2016), the methylated fraction was captured by an anti-methyl-cytosine antibody (MeDIP) as previously described (GUERRERO-BOSAGNA; JENSEN,

2015). The output of the MeDIP was used as the input of GBS. The GBS method uses ligation steps in which a barcode adapter (identifying individual samples) and a common adapter for Illumina sequencing barcoding system are ligated at each end of the digested DNA fragments (POLAND; RIFE, 2012). Due to the barcoding system, the GBS technique enables the creation of a sequencing library with DNA pooled from several individuals (ELSHIRE et al., 2011; POLAND; RIFE, 2012). Once the barcodes and adaptors are ligated, PCR is performed followed by clean-up of primer dimers and unbound adapters (ELSHIRE et al., 2011; POLAND; RIFE, 2012). A detailed description of the method for its use in chickens has been previously reported (PÉRTILLE et al., 2016). The use of the present approach, in which these two methodologies are combined, allowed us to scan the RBCs methylome of 21 chickens using only half of an Illumina sequencing lane. Sequencing was performed paired-end with read length of 125 bp on the Illumina HiSeq2500 platform.

4.2.5 Bioinformatic analyses

For the methylated DNA sequencing, data quality trimming was performed in paired-end short reads with the SeqyClean tool v. 1.9.10 (ZHBANNIKOV et al., 2013) using a Phred quality score ≥24 and a fragment size ≥50. The quality of the *reads* was checked before and after the cleaning by FastQC v.0.11.3 (ANDREW, 2010). The Stacks v.1.39 program was used for data de-multiplexing (CATCHEN et al., 2011). For each sample stored in a FASTQ file there is one identification map key file. This key file contains the matching information for the respective sample, flow-cell and lane. The expected reads begin with one of the individual barcodes and are followed by the cut site remnant for *PstI*, which contains the sequence CTGCA. Fragments are then grouped into lists, which correspond to individuals identified by their respective barcodes. The alignment of quality-trimmed reads was performed using the Bowtie2 tool v.2.2.5 (LANGMEAD; SALZBERG, 2012) against the chicken reference sequence *Gallus_gallus* 4.0 (NCBI). The coverage depth of each sample was checked using Samtools v.0.1.19 (LI et al., 2009) with the "depth" option.

Because low methylated DNA material is obtained after MeDIP of the *PstI*-reduced genome, some samples will contribute with very low DNA amounts to be sequenced. These individuals will show low total number of reads distributed in a few genomic regions, generating a skewed distribution of methylated sites along the genome. This will result in an overestimation of the coverage values in those CpG sites that happened to be covered by reads. To prevent this, we defined a minimum cut-off in order to select high quality sequenced samples

for further testing of differences between experimental groups. We established a cut-off index, which was defined by dividing the 'percentage of the Chicken Genome covered' by the 'sequencing coverage average for each sample'. Individuals showing index below 1.1 were discarded from further analyses.

Following read alignment, all analyzes were performed using bioinformatics packages from the "R" Bioconductor repository. The Medips package was used for basic data processing, quality controls, normalization, and identification of differential coverage. The BSgenome.Ggallus.UCSC.galGal4 package was uploaded as the reference genome. The edgeR and heatmap.2 packages (and extensions) were used for the confection of plots.

The internet-based tool Consensus PathDB (KAMBUROV et al., 2013) (http://cpdb.molgen.mpg.de) was used to perform an analysis of biological pathways enriched by the genes with differentially methylated regions found in our study, as well as gene ontology analyses of these genes. Consensus PathDB (KAMBUROV et al., 2013) integrates interaction networks based on published information in humans. These interaction networks include complex protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target interactions, as well as biochemical pathways (KAMBUROV et al., 2013). Another internet-based tool used in this study to identify over -represented pathways related to our gene list was Reactome (CROFT et al., 2011), which is an open source curated bioinformatics database of human pathways and reactions (http://www.reactome.org). The advantage of Consensus PathDB over Reactome is that it is capable of accessing a variety of databases that contain previously described biological pathways (e.g., Kegg, Biocarta, Reactome, Wikipathways). However, for using Consensus PathDB the genes in the chicken genome had to be extrapolated to humans, since it does not accept the ENSEMBL chicken genome annotation. Therefore Reactome, which did accept the input of chicken genes with the ENSEMBL identifier, was also used. These two tools therefore provided complementary information about our gene list.

4.3 Results

The present experiment compared the RBCs methylome of chickens reared in open aviaries versus in cages, to detect whether epigenetic profiles in RBCs could be identified as correlating to each of these rearing conditions. The experimental procedures are summarized in Figure 4.2. RBCs of 21 chickens were extracted in total, being 9 reared in open aviaries and 12 reared in cages. A combination of the Genotype by Sequencing (GBS) and MeDIP methods was used to identify genome-wide changes in DNA methylation.

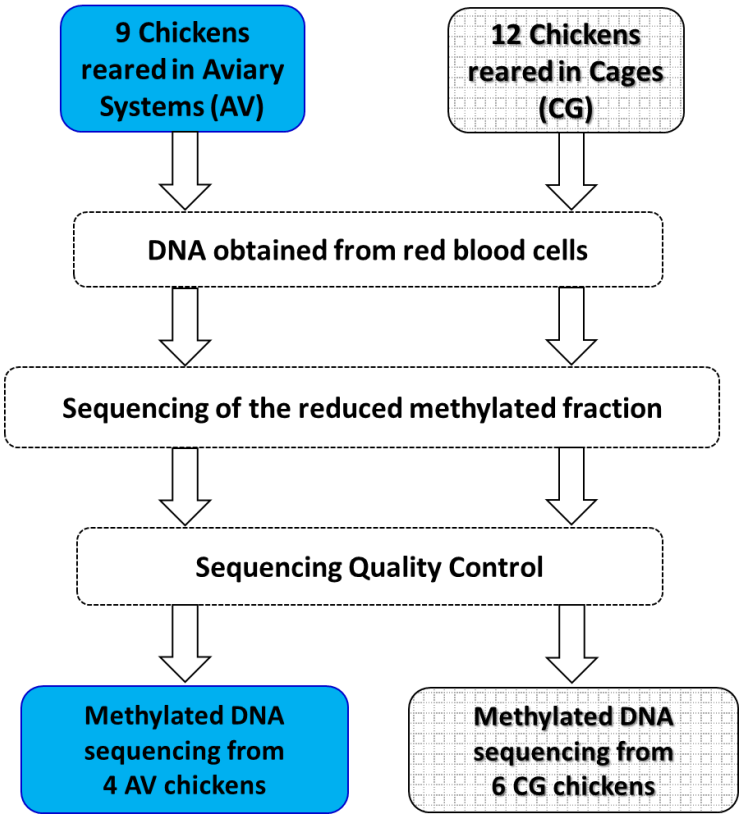


Figure 4.2 - Diagram summarizing the processing of samples from individuals in each treatment group

After sequencing of the reduced-methylated DNA fraction from RBCs of these animals, bioinformatic analyzes were performed and filter parameters were applied. Our quality control procedure selected sequencing data from 4 AV and 6 CG animals for further statistical analyses. Our method interrogated changes in DNA methylation of 810,186 CpG sites per individual, which corresponds to ~7.6% of all CpGs in the chicken genome. An MA plot showing the log-fold change of AV/CG counts per 300 bp genomic windows, which represents changes in DNA methylation, against the normalized window counts is shown in Figure 4.3. Genomic windows with significant changes in DNA methylation between groups ($P < 0.0005$) are depicted in red dots. A principal components analysis (PCA) performed using the windows with significant differences in counts ($P < 0.0005$) between the AV and CG groups confirmed that all individuals in the analysis match the initial experimental group separation (Figure 4.4). Our comparison revealed that 115 windows showed significant change in DNA methylation between experimental groups (APPENDIX J). A heat map showing the windows with significant changes is shown in Figure 4.5. Nearby windows were merged into differentially methylated regions (DMR) between the experimental groups, which were located within or in the vicinity of 53 genes and within 22 intronic regions. APPENDIX K describes the chromosomal location

of all DMR, the number of CpGs within them, their annotation, as well as their location within or in the vicinity of genes. Figure 4.6 6 summarizes the location of these regions regarding genes (Figure 4.6a), as well as their chromosomal location (Figure 4.6b). The fold changes in DNA methylation of the DMR and the direction of these, e.g. hyper- or hypo-methylation of CG reared versus AV, are shown in Figure 4.7.

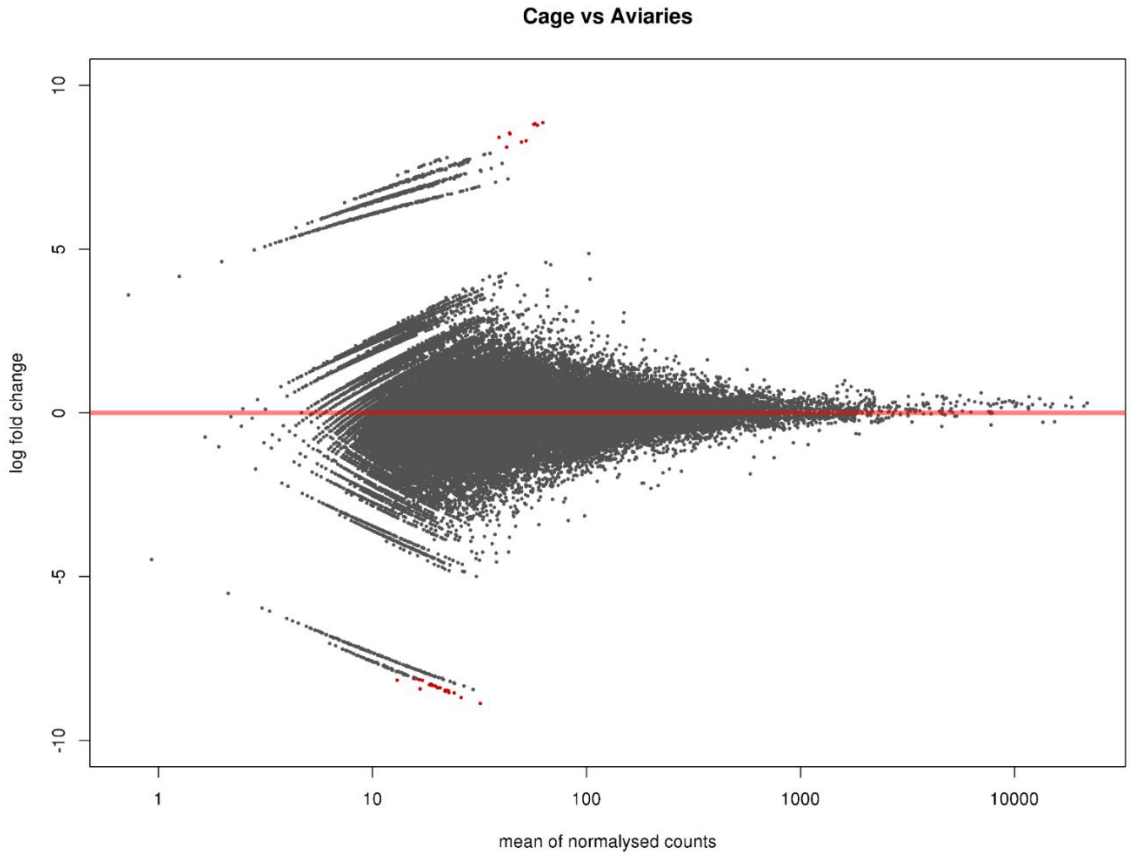


Figure 4.3 - MA plot showing the log-fold change of AV/CG counts (changes in DNA methylation) per 300 bp genomic windows against the normalized window counts. Windows with significant changes ($P < 0.0005$) between experimental groups are depicted as red dots

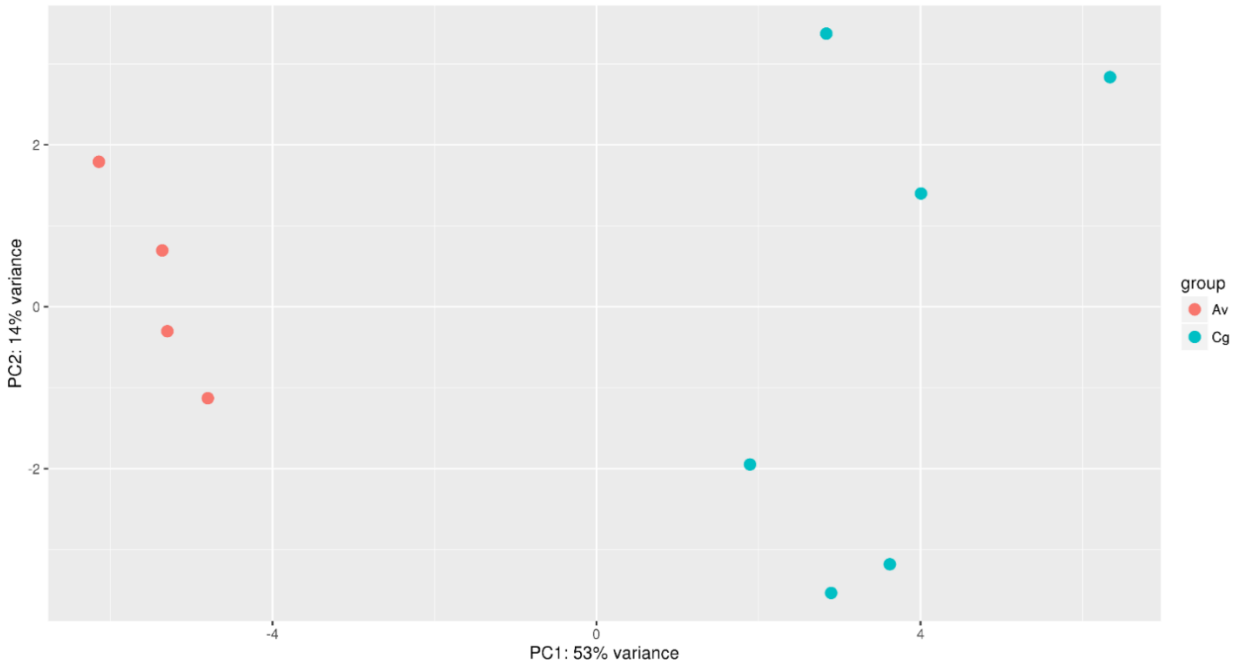


Figure 4.4 - Principal components analysis (PCA) performed using the genomic windows with significant differences in counts (P < 0.0005) between the AV and CG groups

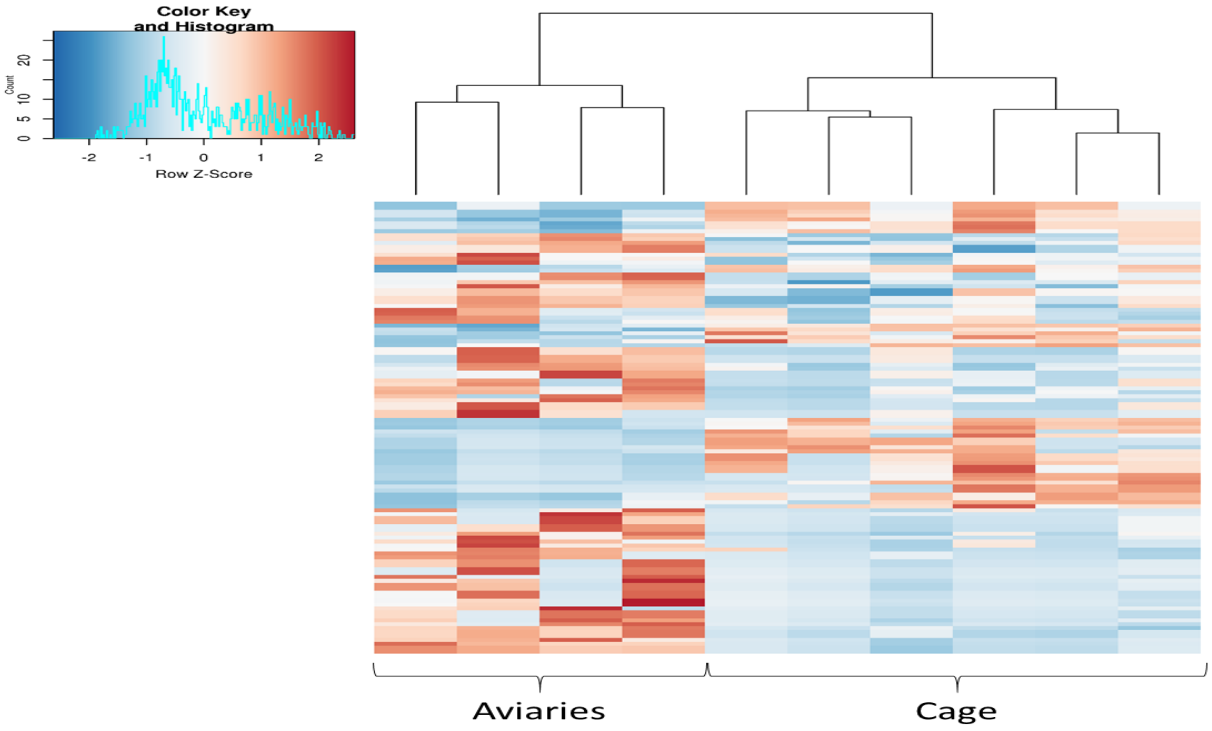


Figure 4.5 - Heat map showing the genomic windows with significant changes in DNA methylation between experimental groups

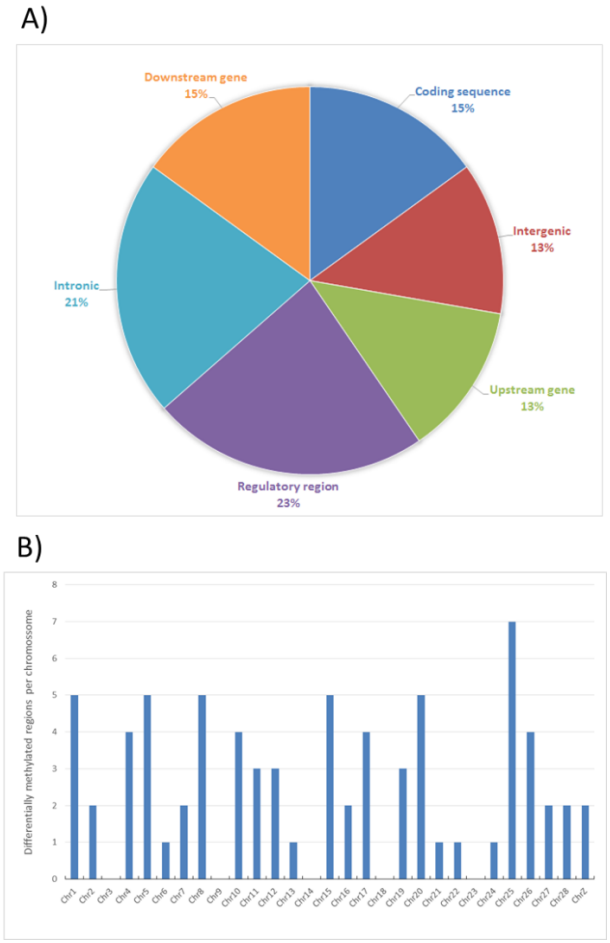


Figure 4.6 - Location of DMRs regarding (a) nearby or associated genes, and (b) chromosomes.

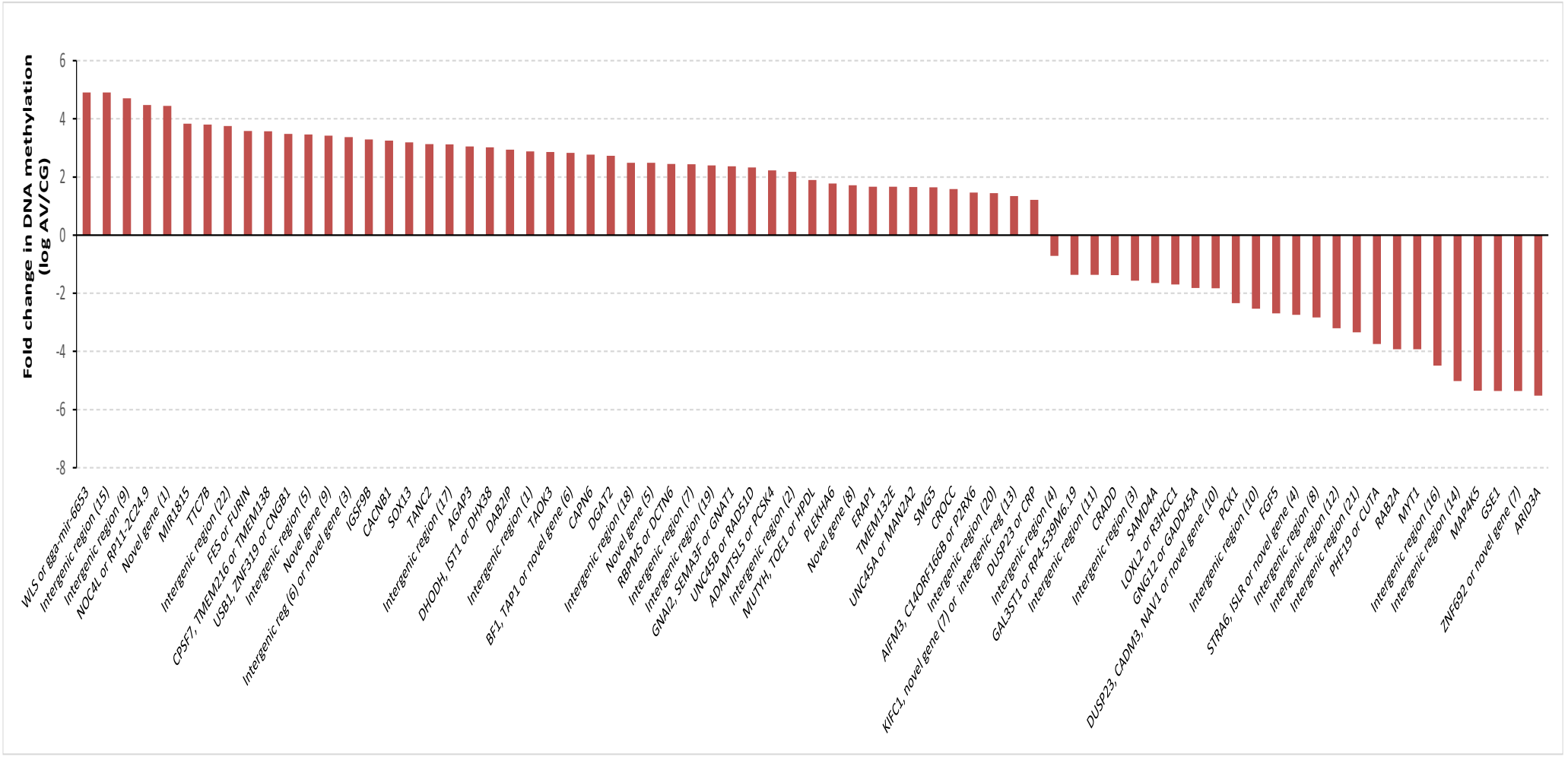


Figure 4.7 - Fold change representation of DMR-associated genes or intergenic regions between the experimental groups

A network analysis performed with the DMR associated genes in Consensus PathDB, which connects biological pathways and gene ontology information, is shown in Figure 4.8 and APPENDIX L. A simplified pathway is shown in Figure 4.8, where redundant information was discarded (e.g., same biological processes showing as being affected by different databases). This analysis shows that DMR associated genes are mainly enriched in biological processes such as G-protein activation (comprising ~10% of the genes in that pathway), mitogen-activated protein kinase (MAPK) signaling (where five genes in our list participate) and purine ribonucleotide binding (where 14 genes in our list participate). P-and q- values of all significantly affected pathways are shown in APPENDIX L. In addition to these main affected pathways, less enriched pathways are shown in APPENDIX L. Of interest is also the appearance in the network of processes such as 'visual photo-transduction', 'opioid signaling', mRNA processing and cytoskeleton organization.

The network analysis performed with Reactome, in turn, shows that genes with altered DNA methylation in our list primarily target pathways in the immune system (Figure 4.9a), followed by signal transduction pathways involved in opioid signaling, regulation of the photo-transduction cascade and G-protein activation (Figure 4.9b). A less affected pathway was the 'metabolic', which showed some effects in the sub-pathways 'inhibition of insulin secretion by adrenaline and noradrenaline' and 'metabolism of Abacavir' (Figure 4.9c).

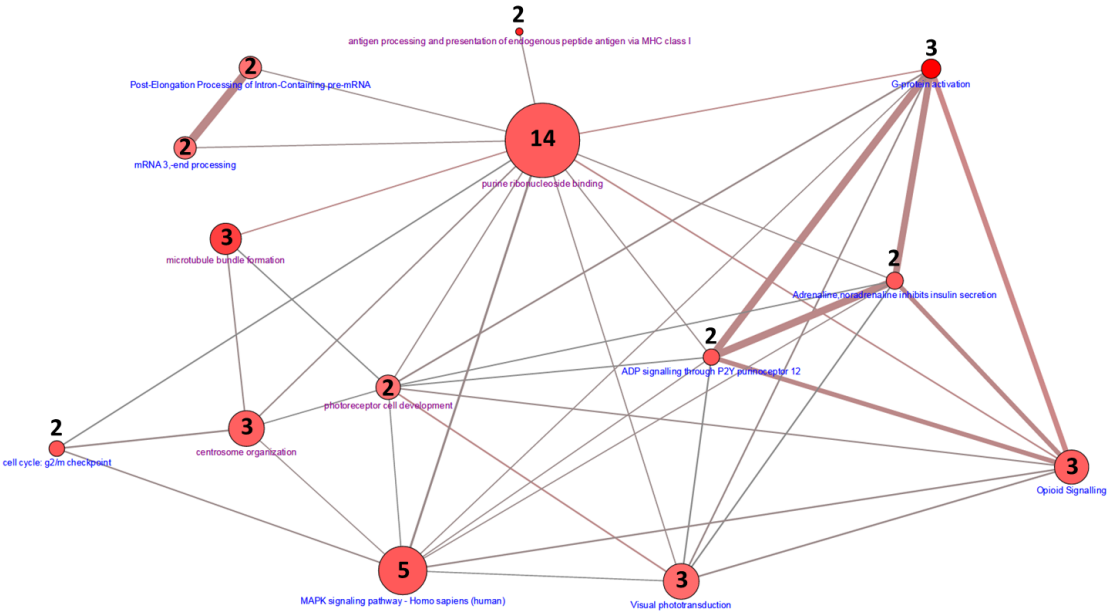
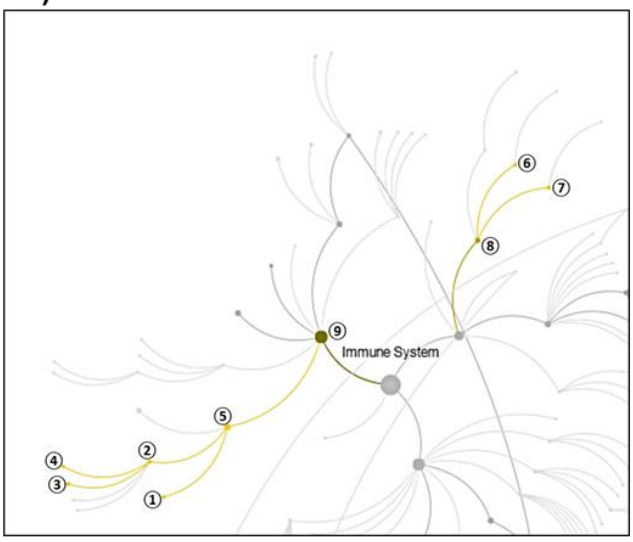


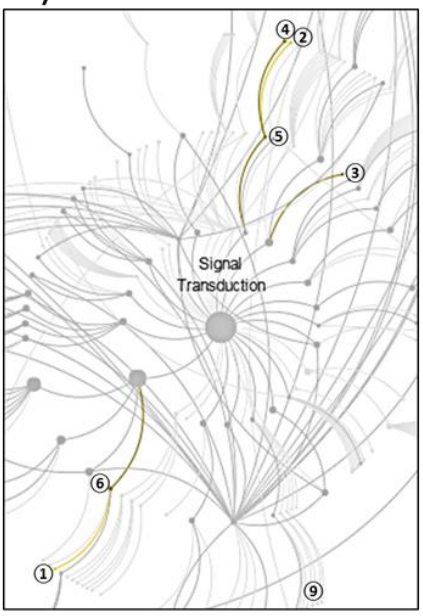
Figure 4.8 - Network analysis performed with Consensus PathDB showing how the DMR associated genes relate to biological pathways and gene ontology information. Significantly enriched pathways and GO terms are shown. The numbers within circles correspond to DMR-associated genes within a specific affected biological pathway (circles with blue annotations) or GO terms (circles with pink annotations). The size of the circles correspond to the total number of genes in database for that specific pathway

A)



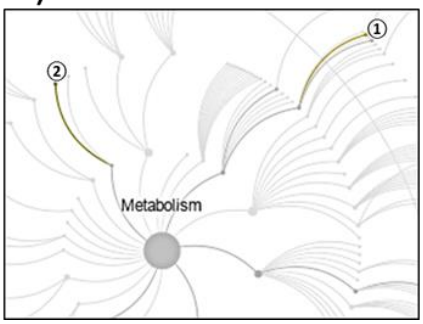
1- Antigen Presentation: Folding, assembly and peptide loading of class I MHC
2- Antigen processing-Cross presentation
3- ER-Phagosome pathway
4- Endosomal/Vacuolar pathway
5- Class I MHC mediated antigen processing & presentation
6- Interferon alpha/beta signaling
7- Interferon gamma signaling
8- Interferon Signaling
9- Adaptive Immune System

B)



1- G-protein activation
2- Activation of the phototransduction cascade
3- Nerve growth factor (NGF) processing
4- Inactivation, recovery and regulation of the phototransduction cascade
5- The phototransduction cascade
6- Opioid Signaling

C)



1- Adrenaline, noradrenaline inhibits insulin secretion
2- Abacavir metabolism

Figure 4.9 - Network analysis performed with Reactome showing how the DMR associated genes relate to biological pathways. Significantly enriched pathways are shown for (a) the immune system, (b) signal transduction, and (c) metabolism

4.4 Discussion

Stress has been reported to associate with DNA methylation specific alterations in brain. For example, infant rats exposed to parental maltreatment present long term DNA methylation and gene expression changes in the BDNF (brain-derived neurotrophic factor) gene in the frontal cortex (ROTH et al., 2009). However, from the perspective of using epigenetic tools to determine the history of stress in live animals, it is of interest to determine whether epigenetic changes can also be observed in cell types of easy access such as blood cells.

A few studies have reported epigenetic changes in blood related to stress. For example, adult rats previously exposed to traumatic conditions during early life exhibit altered microRNA profile in the blood, brain and spermatozoids compared to non-traumatized individuals (GAPP et al., 2014). In humans (MALAN-MULLER et al., 2014) and monkeys (PROVENCAL et al., 2012) DNA methylation in peripheral blood cells has been shown to be altered in correlation with previous stress. Since birds have nucleated RBCs, they represent an organism model in which DNA methylation can be measured in live individuals, and in an easily accessible and simple to purify cell-type.

The present study evaluates the effects of early life conditions on adult DNA methylation patterns in a farm animal. This was performed in RBCs of adult hens after they had been reared in groups exposed to different levels of environmental complexity. Avian RBCs contain nucleated RBCs, which allows for accurate epigenetic profiling because it is simple to purify this cell type that can be obtained from live animals. The aim was to identify in adult hens epigenetic profiles in RBCs associated with different rearing conditions. The rearing conditions to which hens were subjected in the current study cause long-term differences in fearfulness as indicated by differences in inhibition of behavior and avoidance of a human and a novel object in a novel test arena (BRANTSAETER et al., 2016). Although we have not documented stress-related physiological differences between the treatment groups during the rearing phase (first 16 weeks of age), the fact that fear responses are per definition associated with physiological stress suggests that the rearing treatments induce distinct long-term alterations in the stress response. On the one hand, birds in the complex aviary environment are likely to be exposed to a higher degree of mild intermittent stress. On the other hand, confinement in the more barren cage environment may generate a sustained and long-term stress due to deprivation. Interestingly, evidence indicates that the aviary environment may be harsher and more challenging than the cage environment, as indicated by the fact that mortality of aviary-housed birds is normally twice as high as that of cage-housed birds (JANCZAK; RIBER, 2015). In

addition to fear responses, these rearing conditions also associate with different levels of cognitive capabilities observed later in life in birds from the same groups as in the present experiment (TAHAMTANI et al., 2015).

A number of genomic regions presented changes in RBCs DNA methylation between the different rearing conditions tested. These DMR are more present in regulatory regions and less present in intergenic regions (figure 4.6a). DMR were only absent in chromosomes 3, 9, 14, 18, 23, 32 and W. All other chromosomes presented a fairly even distribution of DMRs, although chromosome 25 is the one that contained the higher number (figure 4.6b). The genes associated to these DMR were tested in pathway network analyses to determine whether they would significantly affect biological processes.

We used Consensus PathDB and Reactome to inquire for biological pathways enriched by the genes found associated to the DMRs reported here. For this, we tested whether at least two of the genes in our list would belong to a single biological pathway previously described in the associated databases. Within Consensus PathDB we also performed a Gene Ontology analysis to determine possible common functional roles of these genes. Consensus PathDB analyses demonstrate that differentially methylated genes are involved in pathways related to G-protein activation (in particular, involved in opioid response and the photo-transduction cascade), MAPK signaling and purine ribonucleoside binding (related to post-transcriptional processes). MAPK are known to regulate a wide array of cell functions relating to regulation of gene expression in cellular processes such as proliferation, differentiation, mitosis, apoptosis and survival (PEARSON et al., 2001). Interestingly, MAPKs such as p38, MK2 and MK3 are known to mediate stress response, regulating the transcriptional activation of so-called ‘immediate early genes’ in mammalian cells (RONKINA et al., 2011). The involvement of purine ribonucleoside (i.e., AMP and GMP) binding has been given minor attention in research investigating stress responses. However, of interest is recent data showing the mediation of purine ribonucleoside binding in the antidepressant side-effects of phosphodiesterase inhibitors (i.e., etazolate, an anxiolytic drug; sildenafil, a drug used in the treatment of erectile dysfunction) in mice (WANG et al., 2014).

Pathway analysis with Reactome gave similar results, since equivalent signal transduction pathways were shown to be affected. In addition, many sub-pathways were affected within the immune system. Altered signal transduction pathways include opioid signaling, regulation of the photo-transduction cascade, and G-protein activation. Interestingly, opioid signaling has for a long time been related to housing conditions in farm animals. For example, in pigs opioid receptor density is affected by the housing conditions and is inversely correlated to stereotypic

behavior duration (ZANELLA et al., 1996). Also in pigs, the expression of opioid receptors in the amygdala is substantially different between individuals maintained in enriched versus conventional housing environment (KALBE; PUPPE, 2010). Although not much research has been done on the role of opioids in chickens, it has been reported that opioid systems modulate social attachment and isolation stress (SUFKA et al., 1994; WARNICK et al., 2005). This is concordant with finding in rats showing that social isolation increases the responsiveness of the kappa opioid receptor (KARKHANIS et al., 2016). What emerges as an interesting finding in the present paper is that the opioid system could be affected not only in the central nervous system but also in peripheral cells. Further research needs to be done to understand the role of peripheral opioids systems in the modulation of stress response. Although not many studies have focused on the correlation between photo-transduction and stress, research in chickens has shown that immune response varies with light cycles in a circadian fashion, controlled in part by the pineal gland, which among other cell types contain B-lymphocytes (BAILEY et al., 2003). Vasotocin receptors, which belong to the G-protein receptor family, have been reported to mediate stress response in chickens. The recently characterized neuropeptides in this family (VT2R and VT4R) are known to be involved in stress response, particularly within the cephalic lobe of the anterior pituitary (KUENZEL et al., 2013). Again, how these neuronal effects translate to peripheral signaling is an interesting matter of future investigation.

In addition, the Reactome metabolic pathway showed some effects in the inhibition of insulin secretion by adrenaline and noradrenaline, and in the metabolism of Abacavir. Experiments with perfused (canine) pancreas show that insulin secretion is strongly inhibited by adrenaline or noradrenaline (IVERSEN, 1973). In turn, adrenaline and norepinephrine levels are known to vary not only due to stress (HENRY, 1992; ISHIBASHI et al., 2013; MULLER et al., 2013) but also in connection with the conditions under which animals are kept in captivity (MULLER et al., 2013). For example in porpoises, free-ranging animals present higher blood levels of both adrenaline and noradrenaline than animals in rehabilitation or under human care (MULLER et al., 2013). It is not surprising that high adrenaline or noradrenaline levels in free-ranging animals will lead to the inhibition of insulin and a concomitant rapid increase in circulatory glucose levels, concordant with the high energy demands of animals living in free-ranging conditions. However, it is intriguing that such a mechanisms could be epigenetically regulated.

The effects in the pathway ‘metabolism of Abacavir’ (a drug used for HIV treatment in humans) point towards immune responses being affected. For example, Abacavir hypersensitivity syndrome involves changes in immunological tolerance and activation of

specific T cells (ILLING et al., 2012). This is concordant with the main pathway affected in Reactome, which was the immune system. A reason for this effect in the immune system could be that animals living in a confined space would exhibit higher levels of stress, which are known correlate with altered immune response. In humans, for example, individuals with history of post-traumatic stress have compromised immune systems, with reduced number of lymphocytes and T cells, reduced natural killer cells activity, and reduced production of interferon gamma and interleukin-4 (KAWAMURA et al., 2001). Also, housing conditions have been correlated to decreased immune response in farm animals. For example, dairy calves housed in smaller stalls present reduced lymphocyte proliferation in comparison with calves in larger stalls (FERRANTE et al., 1998). In mice, the bedding type is shown to influence the intestinal immune system (SANFORD et al., 2002).

An interesting suggestion from our data is that a compromised immune system response could be imprinted in the epigenome of RBCs after animals are reared under specific conditions of stress. Since DNA methylation patterns are altered, it is suggested that the different rearing conditions leave an epigenetic mark in the red blood cells that will in turn affect the functioning of biological processes such as immune response, maybe in a permanent manner. Further experiments are needed to elucidate whether altered physiological measures of immune responses can correlate to developmentally-altered epigenetic patterns in farm animals.

The aim of the current study was to identify epigenetic profiles of early developmental stress-related environmental effects in RBCs. We identified distinguishable DNA methylation profiles relating to each treatment. A future goal is that the present results can be used as a proof-of-concept for the identification of epigenetic marks related to past stress conditions that occur in the production environment. Future experiments should evaluate whether sets of DMRs could constitute reliable ‘epigenetic signatures’ of specific and controlled stress conditions in extended populations of animals. The present study reports for the first time DNA methylation changes in RBCs of adult hens when reared in conditions of differing environmental complexity. We describe that these changes in DNA methylation associate with genes involved in biological functions such as immune response, and cell signaling related to MAPK, G-protein and opioid pathways. These results open interesting questions regarding the role of early life stimuli in altering epigenetic patterns that could be involved in these mechanisms. Moreover, questions also arise regarding the role RBCs play in G-protein and opioid pathways in stress response.

References

ANDREW, S. **FASTQC:** a quality control tool for high throughput sequence data. 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Acesso em: 09 set. 2016.

BAILEY, M.J; BEREMAND, P.D.; HAMMER, R.; BELL-PEDERSEN, D.; THOMAS, T.L.; CASSONE, V.M. Transcriptional profiling of the chick pineal gland, a photoreceptive circadian oscillator and pacemaker. **Molecular Endocrinology**, Baltimore, v. 17, n. 10, p. 2084-2095, Oct. 2003.

BRANTSAETER, M.; NORDGREEN, J.; RODENBURG, T.B.; TAHAMTANI, F.M.; POPOVA, A.; JANCZAK, A.M. Exposure to increased environmental complexity during rearing reduces fearfulness and increases use of three-dimensional space in laying hens (*Gallus gallus domesticus*). **Frontiers in Veterinary Science,** Lausanne, v. 3, p. 14, 2016.

BROOM, D.M. Animal welfare: an aspect of care, sustainability, and food quality required by the public. **Journal of Veterinary Medical Education,** Knoxville, v. 37, n. 1, p. 83-88, Spring 2010.

CATCHEN, J.M. CATCHEN, J.M.; AMORES, A.; HOHENLOHE, P.; CRESKO, W.; POSTLETHWAIT, J.H. Stacks: building and genotyping Loci de novo from short-read sequences. **G3 (Bethesda),** Bethesda, v. 1, n. 3, p. 171-82, Aug 2011.

CROFT, D.; O'KELLY, G.; WU, G.; HAW, R.; GILLESPIE, M.; MATTHEWS, L.; CAUDY, M.; GARAPATI, P.; GOPINATH, G.; JASSAL, B.; JUPE, S.; KALATSKAYA, I.; MAHAJAN, S.; MAY, B.; NDEGWA, N.; SCHMIDT, E.; SHAMOVSKY, V.; YUNG, C.; BIRNEY, E.; HERMJAKOB, H.; D'EUSTACHIO, P.; STEIN, L. Reactome: a database of reactions, pathways and biological processes. **Nucleic Acids Research**, London, v. 39, p. D691-D697, Jan. 2011.

CUBAS, P.; VINCENT, C.; COEN, E. An epigenetic mutation responsible for natural variation in floral symmetry. **Nature,** London, v. 401, n. 6749, p. 157-11, Sept. 1999.

DENHAM, J.; MARQUES, F.Z.; O'BRIEN, B.J.; CHARCHAR, F.J. Exercise: putting action into our epigenome. **Sports medicine**, Auckland, v. 44, n. 2, p. 189-209, Feb.2014.

DICKMAN, M.J.; KUCHARSKI, R.; MALESZKA, R.; HURD, P.J. Extensive histone post-translational modification in honey bees. **Insect Biochemistry and Molecular Biology,** Oxford, v. 43, n. 2, p. 125-137, Feb. 2013.

DOLINOY, D.C.; HUANG, D.; JIRTLE, R.L. Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. **Proceedings of the National Academy of Sciences of the United States of America,** Washington, v. 104, n. 32, p. 13056-13061, Aug. 2007.

ELSHIRE, R.J.; GLAUBITZ, J.C.; SUN, Q.; POLAND, J.A.; KAWAMOTO, K.; BUCKLER, E.S.; MITCHELL, S.E. A robust, Simple Genotyping-by-Sequencing (GBS) approach for high diversity species. **PLoS One**, San Francisco, v. 6, n. 5, p. e19379, 2011.

FALLAHSHAROUDI, A.; DE KOCK, N.; JOHNSSON, M.; UBHAYASEKERA, S.J.; BERGQUIST, J.; WRIGHT, D.; JENSEN, P. Domestication effects on stress induced steroid secretion and adrenal gene expression in chickens. **Scientific Reports,** London, v. 5, p. 15345, 2015.

FEIL, R.; FRAGA, M.F. Epigenetics and the environment: emerging patterns and implications. **Nature Reviews. Genetics**, London, v. 13, n. 2, p. 97-109, Feb. 2011.

FERRANTE, V.; CANALI, E.; MATTIELO, S.; VERGA, M.; SACERDOTE, P.; MANFREDI, B.; PANERAI, A.E. Preliminary study on the effect of size of individual stall on the behavioural and immune reactions of dairy calves. **Journal of Animal and Feed Sciences,** London, v. 7, n. 1, p. 29-36, 1998.

FRESARD, L.;MORISSON, M.; BRUN, J.M.; COLLIN, A.; PAIN, B.; MINVIELLE, F.; PITEL, F. Epigenetics and phenotypic variability: some interesting insights from birds. **Genetics, Selection, Evolution: GSE**, Paris, v. 45, p. 16, 2013.

GAPP, K.; JAWAID, A.; SARKIES, P.; BOHACEK, J.; PELCZAR, P.; PRADOS, J.; FARINELLI, L.; MISKA, E.; MANSUY, I.M. Implication of sperm RNAs in transgenerational inheritance of the effects of early trauma in mice. **Nature Neuroscience**, New York, v. 17, n. 5, p. 667-669, May 2014.

GOERLICH, V.C.; NÄTT, D.; ELFWING, M.; MACDONALD, B.; JENSEN, P. Transgenerational effects of early experience on behavioral, hormonal and gene expression responses to acute stress in the precocial chicken. **Hormones and behavior.** New York, v. 61, n. 5, p. 711-718, May 2012.

GU, H.; SMITH, Z.D; BOCK, C.; BOYLE, P.; GNIRKE, A.; MEISSNER, A. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. **Nature Protocols**, London, v. 6, n. 4, p. 468-481, Apr. 2011.

GUERRERO-BOSAGNA, C.; JENSEN, P. Optimized method for methylated DNA immuno-precipitation. **Methods X,** Amsterdam, v. 2, p. 432-439, 2015.

GUERRERO-BOSAGNA, C.; SKINNER, M.K. Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. **Molecular and cellular endocrinology**, Amsterdam, v. 354, n. 1/2, p. 3-8, May 2012.

GUERRERO-BOSAGNA, C.M.; SABAT, P.; VALDOVINOS, F.S.; VALLADARES, L.E.; CLARK, S.J. Epigenetic and phenotypic changes result from a continuous pre and post natal dietary exposure to phytoestrogens in an experimental population of mice. **BMC Physiology**, London, v. 8, p. 17, 2008.

HENDRIX, G. **General management guide for Dekalb White Commercial Layer.** 2015. Disponível em: <http://www.isapoultry.com/~/media/Files/ISA/ISA%20product%20information/Dekalb/Parent%20Stock/General%20%20Management%20Guide%20Commercials%20Dekalb%20white.pdf2014>. Acesso em: 09 set. 2016.

HENRY, J.P. Biological basis of the stress response. **Integrative Physiological and Behavioral Science,** New Brunswick, v. 27, n. 1, p. 66-83, Jan./Mar. 1992.

ILLING, P.T.; VIVIAN, J.P.; DUDEK, N.L.; KOSTENKO, L.; CHEN, Z.; BHARADWAJ, M.; MILES, J.J.; KJER-NIELSEN, L.; GRAS, S.; WILLIAMSON, N.A.; BURROWS, S.R; PURCELL, A.W.; ROSSJOHN, J.; MCCLUSKEY J. Immune self-reactivity triggered by drug-modified HLA-peptide repertoire. **Nature,** London, v. 486, n. 7404, p. 554-558, Jun. 2012.

ISHIBASHI, M.; AKIYOSHI, H.; ISERI, T.; OHASHI, F. Skin conductance reflects drug-induced changes in blood levels of cortisol, adrenaline and noradrenaline in dogs. **The Journal of Veterinary Medical Science,** Tokyo, v. 75, n. 6, p. 809-813, 2013.

IVERSEN, J. Adrenergic receptors and the secretion of glucagon and insulin from the isolated, perfused canine pancreas**. The Journal of Clinical Investigation**, New Haven, v. 52, n. 9, p. 2102-2116, Sept. 1973.

JENSEN, P. Behaviour epigenetics: the connection between environment, stress and welfare. **Applied Animal Behaviour Science**, Amsterdam, v. 157, p. 1-7, 2014.

KAIN, K.H.; MILLER, J.W.; JONES-PARIS, C.R.; THOMASON, R.T.; LEWIS, J.D.; BADER, D.M. BARNETT, J.V.; ZIJLSTRA, A. The chick embryo as an expanding experimental model for cancer and cardiovascular research. **Developmental Dynamics: an Official Publication of the American Association of Anatomists**, New York, v. 243, n. 2, p. 216-228, Feb. 2014.

KALBE, C.; PUPPE, B. Long-term cognitive enrichment affects opioid receptor expression in the amygdala of domestic pigs. **Genes, Brain, and Behavior**, Oxford, v. 9, n. 1, p. 75-83, Feb. 2010.

KAMBUROV, A.; STELZL, U.; LEHRACH, H.; HERWIG, R. The ConsensusPathDB interaction database: 2013 update. **Nucleic Acids Research,** London, v. 41, p. D793-800, Jan. 2013.

KARKHANIS, A.N.; ROSE, J.H.; WEINER, J.L.; JONES, S.R. Early-life social isolation stress increases kappa opioid receptor responsiveness and downregulates the dopamine system. **Neuropsychopharmacology,** New York, v. 41, n. 9, p. 2263-2274, Aug. 2016.

KAWAMURA, N.; KIM, Y.; ASUKAI, N. Suppression of cellular immunity in men with a past history of posttraumatic stress disorder. **The American Journal of Psychiatry**, Arlington, v. 158, n. 3, p. 484-486, Mar. 2001.

KUENZEL, W.J.; KANG, S.W.; JURKEVICH, A. Neuroendocrine regulation of stress in birds with an emphasis on vasotocin receptors (VTRs). **General and Comparative Endocrinology**, New York, v. 190, p. 18-23, Sept. 2013.

LANGMEAD, B.; SALZBERG, S.L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, New York, v. 9, n. 4, p. 357-9, Apr. 2012.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R.; 1000 Genome Project data processing subgroup: the sequence alignment/map format and SAMtools**. Bioinformatics**, Oxford, v. 25, n. 16, p. 2078-2079, Aug. 2009.

LYKO, F.; FORET, S.; KUCHARSKI, R.; WOLF, S.; FALCKENHAYN, C.; MALESZKA, R. The honey bee epigenomes: differential methylation of brain DNA in queens and workers. **PLoS Biology,** San Francisco, v. 8, n. 11, p. e1000506, 2010.

MALAN-MULLER, S.; SEEDAT, S.; HEMMINGS, S. M. Understanding posttraumatic stress disorder: insights from the methylome. **Genes, Brain, and Behavior,** Oxford, v. 13, n. 1, p. 52-68, Jan. 2014.

MANNING, K.; TÖR, M.; POOLE, M.; HONG, Y.; THOMPSON, A.J.; KING, G.J.; GIOVANNONI, J.J.; SEYMOUR, G.B. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. **Nature Genetics**, New York, v. 38, n. 8, p. 948-952, Aug. 2006.

MORGAN, K.N.; TROMBORG, C. T. Sources of stressa in captivity. **Applied Animal Behaviour Science,** Amsterdam, v. 102, p. 262-302, 2006.

MULLER, S.; LEHNERT, K.; SEIBEL, H.; DRIVER, J.; RONNENBERG, K.; TEILMANN, J.; VAN ELK, C.; KRISTENSEN, J.; EVERAARTS, E.; SIEBERT, U. Evaluation of immune and stress status in harbour porpoises (*Phocoena phocoena*): can hormones and mRNA expression levels serve as indicators to assess stress? **BMC Veterinary Research,** London, v. 9, p. 145, 2013.

OLANREWAJU, H.A.; THAXTON, J.P.; DOZIER, W.A.; PURSWELL, J.; ROUSH, W.B.; BRANTON, S.L. A review of lighting programs for broiler production. **International Journal of Poultry Science,** Faisalabad, v. 5, n. 4, p. 301-308, 2006.

PEARSON, G.; ROBINSON, F.; BEERS GIBSON, T.; XU, B.E.; KARANDIKAR, M.; BERMAN, K.; COBB, M.H. Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions**. Endocrine Reviews**, Baltimore, v. 22, n. 2, p. 153-183, Apr. 2001.

PÉRTILLE, F.; GUERRERO-BOSAGNA, C.; DA SILVA, V.H.; BOSCHIERO, C.; NUNES, J.R.; LEDUR, M.C.; JENSEN, P.; COUTINHO, L.L.; High-throughput and cost-effective chicken genotyping using next-generation sequencing. **Scientific Reports**, London, v. 7, n. 10, p. 26929, May 2016.

POLAND, J.A.; RIFE, T.W. Genotyping-by-sequencing for plant breeding and genetics. **The Plant Genome Journal,** Madison, v. 5, n. 3, p. 92, 2012.

PROVENCAL, N.; SUDERMAN, M.J.; GUILLEMIN, C.; MASSART, R.; RUGGIERO, A.; WANG, D.; BENNETT, A.J.; PIERRE, P.J.; FRIEDMAN, D.P.; CÔTÉ, S.M.; HALLETT, M.; TREMBLAY, R.E.; SUOMI, S.J.; SZYF, M. The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells. **The Journal of Neuroscience,** Baltimore, v. 32, n. 44, p. 15626-15642, Oct. 2012.

RONKINA, N.; MENON, M.B.; SCHWERMANN, J.; ARTHUR, J.S.; LEGAULT, H.; TELLIEZ, J.B.; KAYYALI, U.S.; NEBREDA, A.R.; KOTLYAROV, A.; GAESTEL, M. Stress induced gene expression: a direct role for MAPKAP kinases in transcriptional activation of immediate early genes. **Nucleic Acids Research**, London, v. 39, n. 7, p. 2503-2518, Apr. 2011.

ROSTAGNO, M.H. Can stress in farm animals increase food safety risk? **Foodborne Pathogens and Disease,** Larchmont, v. 6, n. 7, p. 767-776, Sept. 2009.

ROTH, T.L.; LUBIN, F.D.; FUNK, A.J.; SWEATT, J.D. Lasting epigenetic influence of early-life adversity on the BDNF gene. **Biological Psychiatry**, New York, v. 65, n. 9, p. 760-769, 2009.

RUBIN, C.J.; ZODY, M.C; ERIKSSON, J.; MEADOWS, J.R.; SHERWOOD, E.; WEBSTER, M.T.; JIANG, L.; INGMAN, M.; SHARPE, T.; KA, S.; HALLBÖÖK, F.; BESNIER, F.; CARLBORG, O.; BED'HOM, B.; TIXIER-BOICHARD, M.; JENSEN, P.; SIEGEL, P.; LINDBLAD-TOH, K.; ANDERSSON, L. Whole-genome resequencing reveals loci under selection during chicken domestication. **Nature,** London, v. 464, n. 7288, p. 587-591, Mar. 2010.

SANFORD, A.N.; CLARK, S.E.; TALHAM, G.; SIDELSKY, M.G.; COFFIN, S.E. Influence of bedding type on mucosal immune responses. **Comparative Medicine.** Memphis, v. 52, n. 5, p. 429-432, Oct. 2002.

SAVORY, C.J.; LARIVIERE, J. Effects of qualitative and quantitative food restriction treatments on feeding motivational state and general activity level of growing broiler breeders. **Applied Animal Behaviour Science**, Amsterdam, v. 69, n. 2, p. 135-147, 2000.

SEONG, K.H.; LI, D.; SHIMIZU, H.; NAKAMURA, R.; ISHII, S. Inheritance of stress-induced, ATF-2-dependent epigenetic change. **Cell**, Cambridge, v. 145, n. 7, p. 1049-1061, 2011.

SKINNER, M.K.; MANIKKAM, M.; GUERRERO-BOSAGNA, C. Epigenetic transgenerational actions of environmental factors in disease etiology. **Trends in Endocrinology and Metabolism,** New York, v. 21, n. 4, p. 214-422, Apr. 2010.

SUFKA, K.J.; HUGHES, R.A.; MCCORMICK, T.M.; BORLAND, J.L. Opiate effects on isolation stress in domestic fowl**. Pharmacology, Biochemistry, and Behavior,** Phoenix, v. 49, n. 4, p. 1011-1015, 1994.

SUSIARJO, M.; SASSON, I.; MESAROS, C.; BARTOLOMEI, M.S. Bisphenol a exposure disrupts genomic imprinting in the mouse. **PLoS Genetics,** San Francisco, v. 9, n. 4, p. e1003401, 2013.

TAHAMTANI, F.M.; NORDGREEN, J.; NORDQUIST, R.E.; JANCZAK, A.M. Early Life in a barren environment adversely affects spatial cognition in laying hens (*Gallus gallus domesticus*). **Frontiers in Veterinary Science**, Lausanne, v. 2, p. 3, 2015.

TEPEREK-TKACZ, M.; PASQUE, V.; GENTSCH, G.; FERGUSON-SMITH, A.C. Epigenetic reprogramming: is deamination key to active DNA demethylation? **Reproduction,** Cambridge, v. 142, n. 5, p. 621-632, Nov. 2011.

WANG, C.; ZHANG, J.; LU, Y.; LIN, P.; PAN, T.; ZHAO, X.; LIU, A.; WANG, Q.; ZHOU, W.; ZHANG, H.T. Antidepressant-like effects of the phosphodiesterase-4 inhibitor etazolate and phosphodiesterase-5 inhibitor sildenafil via cyclic AMP or cyclic GMP signaling in mice. **Metabolic Brain Disease**, New York, v. 29, n. 3, p. 673-682, Sept. 2014.

WARNICK, J.E.; MCCURDY, C.R.; SUFKA, K.J. Opioid receptor function in social attachment in young domestic fowl. **Behavioural Brain Research**, Amsterdam, v. 160, n. 2, p. 277-285, 2005.

ZANELLA, A.J.; BROOM, D.M.; HUNTER, J.C.; MENDL, M.T. Brain opioid receptors in relation to stereotypies, inactivity, and housing in sows. **Physiology and Behavior**, Oxford, v. 59, n. 4/5, p. 769-775, 1996.

ZANNAS, A.S.; WEST, A.E. Epigenetics and the regulation of stress vulnerability and resilience. **Neuroscience,** Oxford, v. 264, p. 157-170, 2014.

APPENDIX

Appendix A - Association SNPs stats for each of the 94 markers with genome-wise significance (P < 7.86E-07) in pink highlight color and suggestive genome-wise association (P < 1.57E-05) to the others with the respective trait described in column M. Where: eff.adj35 (feed efficiency), f.conv.adj35 (feed conversion), f.int.adj35 (feed intake), all between 35 and 41 days of age and adjusted for body weight at 35 days; bw35, bw41 is body weight at 35 and 42 days of age, respectively, and birthW is birth weight. Markers overlapping with QTL (TRUE) or not overlapping with QTL (FALSE) regions are indicated in the "overlapping QTLs" column

#	Marker Name	Chr	Pos	F	p	add effect	add F	add p	dom effect	dom F	dom p	Trait	Allele	Obs1	Obs2	Obs3	BlockNº	Most frequent Haplotype	overlapping QTLs
1	S1_3706	1	3706	11,4	1,55E-05	-1,8	22,7	2,57E-06	-1,7	21,4	5,06E-06	f.conv.adj35	C/G/S	349	1	94			FALSE
2	S1_25041791	1	25041791	16,9	8,89E-08	-1,4	31,6	3,42E-08	-1,2	22,6	2,69E-06	f.conv.adj35	C/T/T	346	2	96			TRUE
3	S1_45468342	1	45468342	12,9	3,72E-06	1,7	25,7	6,05E-07	-1,7	23,4	1,85E-06	f.conv.adj35	T/C/Y	392	1	51			TRUE
4	S1_53749693	1	53749693	13,4	2,37E-06	1,7	26,3	4,38E-07	-1,6	20,9	6,46E-06	f.conv.adj35	G/A/R	397	1	46			TRUE
5	S1_82575973	1	82575973	12,9	3,74E-06	1,7	25,6	6,22E-07	-1,7	25,4	6,98E-07	f.conv.adj35	T/C/Y	207	1	236			TRUE
6	S1_102944294	1	102944294	15,3	3,83E-07	-1,3	29,9	7,71E-08	-1,4	29,2	1,11E-07	f.conv.adj35	A/C/M	335	2	107			TRUE
7	S1_178184151	1	178184151	12,9	3,68E-06	1,7	25,5	6,70E-07	-1,7	24,1	1,28E-06	f.conv.adj35	G/A/R	389	1	54			TRUE
8	S1_194726625	1	194726625	13,2	2,68E-06	-1,7	26,2	4,65E-07	-1,6	22,8	2,51E-06	f.conv.adj35	A/C/M	351	1	92			TRUE
9	S2_33436237	2	33436237	11,5	1,36E-05	0,1	9,3	2,46E-03	0,1	18,5	2,11E-05	eff.adj35	A/T/W	4	308	132	1	AC 83.6%	TRUE
9	S2_33436237	2	33436237	20,0	5,11E-09	-1,0	27,9	2,04E-07	-1,2	39,0	1,05E-09	f.conv.adj35	A/T/W	308	4	132			TRUE
10	S2_33437336	2	33437336	15,5	3,12E-07	-1,7	24,3	1,21E-06	-1,8	29,1	1,17E-07	f.conv.adj35	C/G/S	304	1	139			TRUE
11	S2_138815862	2	138815862	13,0	3,33E-06	0,9	25,9	5,35E-07	-0,8	13,9	2,14E-04	f.conv.adj35	G/A/R/	401	39	4			TRUE
12	S3_77435334	3	77435334	12,9	3,49E-06	-1,7	25,8	5,65E-07	-1,7	22,9	2,37E-06	f.conv.adj35	C/T/Y	386	57	1			TRUE
13	S4_3514658	4	3514658	13,3	2,60E-06	-1,7	26,2	4,68E-07	-1,6	22,4	2,95E-06	f.conv.adj35	G/T/K	369	1	74			TRUE
14	S4_68623163	4	68623163	11,4	1,47E-05	-37,7	18,3	2,37E-05	19,4	4,9	2,80E-02	bw35	C/A/M	103	112	229			TRUE
15	S4_68882750	4	68882750	11,6	1,29E-05	35,9	17,4	3,59E-05	21,0	5,5	1,91E-02	bw35	A/G/R	102	130	212			TRUE
16	S4_68882765	4	68882765	11,5	1,38E-05	35,5	17,2	4,14E-05	21,4	5,7	1,72E-02	bw35	G/T/K	103	130	211			TRUE
17	S4_69297525	4	69297525	12,3	6,19E-06	33,4	17,6	3,26E-05	22,8	6,5	1,13E-02	bw35	A/G/R	112	140	192			TRUE
18	S4_69370079	4	69370079	12,3	6,42E-06	27,4	12,6	4,29E-04	34,7	15,1	1,20E-04	bw35	A/T/W	144	110	190			TRUE
19	S4_69372005	4	69372005	11,7	1,14E-05	-36,1	17,2	4,07E-05	24,5	7,8	5,45E-03	bw35	T/C/Y	124	103	217	2	f.conv.adj35G AATG 41,5%	TRUE
15	S4_68882750	4	68882750	11,7	1,08E-05	52,7	20,3	8,66E-06	21,8	3,0	8,27E-02	bw41	A/G/R	102	130	212			TRUE
16	S4_68882765	4	68882765	11,6	1,29E-05	51,7	19,7	1,16E-05	22,8	3,3	7,06E-02	bw41	G/T/K	103	130	211			TRUE
17	S4_69297525	4	69297525	12,6	4,78E-06	48,0	19,7	1,18E-05	28,1	5,0	2,62E-02	bw41	A/G/R	112	140	192			TRUE
18	S4_69370079	4	69370079	13,4	2,19E-06	43,5	16,9	4,73E-05	46,0	13,3	3,05E-04	bw41	A/T/W	144	110	190			TRUE
19	S4_69372005	4	69372005	12,6	4,98E-06	-53,6	20,6	7,39E-06	30,6	6,1	1,39E-02	bw41	T/C/Y	124	103	217			TRUE
20	S4_69372065	4	69372065	12,6	4,70E-06	-55,2	21,3	5,31E-06	31,4	6,3	1,25E-02	bw41	G/A/R	127	98	219			TRUE
21	S4_73210325	4	73210325	15,5	3,08E-07	-38,9	28,0	1,95E-07	15,4	3,0	8,31E-02	bw35	T/C/Y	121	137	186	3	TAG 44.4%	TRUE

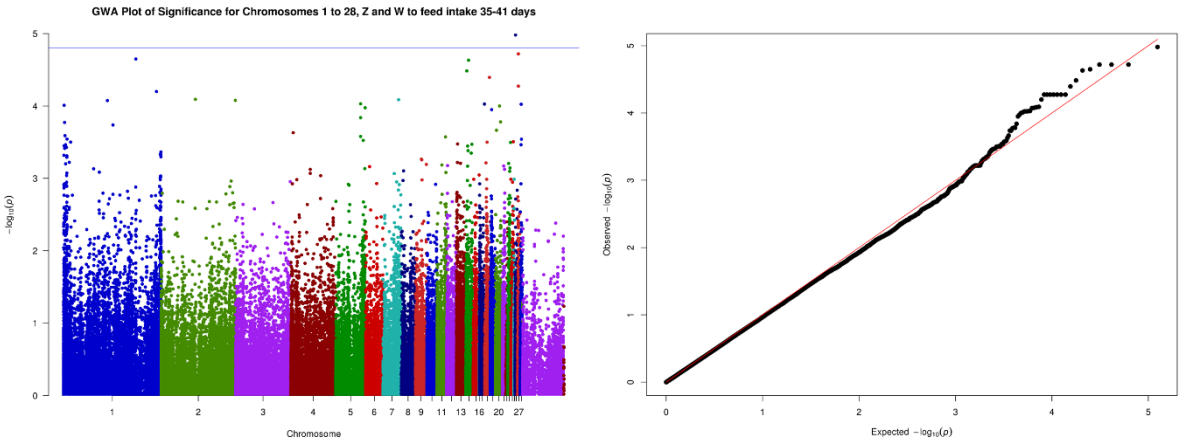
#	Marker Name	Chr	Pos	F	p	add effect	add F	add p	dom effect	dom F	dom p	Trait	Allele	Obs1	Obs2	Obs3	BlockN°	Most frequent Haplotype	overlapping QTLs
21	S4_73210325	4	73210325	13,9	1,40E-06	-49,9	24,4	1,16E-06	23,0	3,4	6,62E-02	bw41	T/C/Y	121	137	186			TRUE
22	S4_73407243	4	73407243	12,2	7,28E-06	60,0	23,2	2,07E-06	28,2	4,4	3,71E-02	bw41	A/G/R	159	82	203			TRUE
23	S4_73407249	4	73407249	12,2	7,28E-06	-60,0	23,2	2,07E-06	28,2	4,4	3,71E-02	bw41	G/A/R	159	82	203			TRUE
24	S4_75842191	4	75842191	12,4	5,97E-06	-35,4	10,1	1,55E-03	43,5	11,2	8,70E-04	bw41	C/A/M	90	167	187			TRUE
25	S4_79707389	4	79707389	11,5	1,36E-05	-22,6	4,5	3,49E-02	48,6	13,2	3,18E-04	bw41	G/C/S	210	162	72			TRUE
26	S5_54401932	5	54401932	13,3	2,45E-06	-1,7	24,9	8,63E-07	-1,8	26,6	3,82E-07	f.conv.adj35	C/T/Y	358	1	85			TRUE
27	S6_2997851	6	2997851	11,4	1,53E-05	-0,7	2,9	8,71E-02	-1,1	5,5	1,95E-02	birthW	G/A/R	370	11	63			TRUE
28	S6_5492179	6	5492179	11,4	1,56E-05	0,8	21,2	5,59E-06	0,4	2,7	1,01E-01	birthW	A/C/M	127	99	218			TRUE
29	S6_28967516	6	28967516	13,2	2,87E-06	-1,7	25,2	7,71E-07	-1,8	26,3	4,51E-07	f.conv.adj35	G/T/K	315	1	128			TRUE
30	S6_31436207	6	31436207	13,1	3,14E-06	1,7	24,6	1,00E-06	-1,7	26,1	4,92E-07	f.conv.adj35	T/C/Y	260	1	183			TRUE
31	S7_6904443	7	6904443	13,4	2,20E-06	1,8	26,9	3,38E-07	-1,6	22,2	3,28E-06	f.conv.adj35	G/A/R	24	1	419			TRUE
32	S7_22092999	7	22092999	13,6	1,94E-06	1,7	26,9	3,38E-07	-1,8	25,3	7,30E-07	f.conv.adj35	G/C/S	393	1	50	4	GC 94.1%	TRUE
33	S7_22093009	7	22093009	13,6	1,94E-06	-1,7	26,9	3,38E-07	-1,8	25,3	7,30E-07	f.conv.adj35	C/T/Y	393	1	50			TRUE
34	S7_26252164	7	26252164	13,0	3,45E-06	-1,7	25,1	7,91E-07	-1,7	25,0	8,23E-07	f.conv.adj35	C/T/Y	384	1	59			TRUE
35	S7_34668134	7	34668134	11,8	9,91E-06	131,5	22,0	3,63E-06	88,7	8,1	4,53E-03	bw35	A/C/M	393	4	47	5	AG 93.2%	TRUE
36	S7_34855716	7	34855716	12,9	3,70E-06	1,7	25,4	6,96E-07	-1,7	24,3	1,18E-06	f.conv.adj35	G/A/R	385	1	58			TRUE
37	S8_5848411	8	5848411	12,9	3,74E-06	1,7	25,7	6,09E-07	-1,7	25,0	8,53E-07	f.conv.adj35	T/C/Y	312	1	131			FALSE
38	S8_27651130	8	27651130	16,7	1,04E-07	1,4	31,4	3,80E-08	-1,2	20,8	6,80E-06	f.conv.adj35	T/C/Y	375	2	67			TRUE
39	S9_15349595	9	15349595	13,1	2,99E-06	1,7	24,8	9,22E-07	-1,8	25,8	5,67E-07	f.conv.adj35	G/A/R	389	1	54			TRUE
40	S9_21971358	9	21971358	11,4	1,54E-05	-1,8	22,7	2,56E-06	-1,7	21,5	4,78E-06	f.conv.adj35	C/T/Y	350	1	93			TRUE
41	S10_9890328	10	9890328	21,7	1,04E-09	-2,2	43,4	1,33E-10	-2,2	41,8	2,77E-10	f.conv.adj35	A/G/R	297	1	146			TRUE
42	S10_17086845	10	17086845	13,6	1,93E-06	1,8	27,0	3,11E-07	-1,7	24,9	8,89E-07	f.conv.adj35	G/A/R	389	1	54			TRUE
43	S11_1249418	11	1249418	13,6	1,94E-06	1,7	26,5	3,98E-07	-1,6	20,6	7,29E-06	f.conv.adj35	G/A/R	396	1	47	6	GGA 93.2%	TRUE
44	S11_1997302	11	1997302	13,3	2,46E-06	1,7	26,4	4,24E-07	-1,6	21,5	4,78E-06	f.conv.adj35	G/A/R	392	1	51			TRUE
45	S11_2149223	11	2149223	13,4	2,23E-06	-1,7	26,5	4,06E-07	-1,6	21,1	5,68E-06	f.conv.adj35	A/G/R	392	1	51			TRUE
46	S13_1958055	13	1958055	11,5	1,31E-05	-1,8	23,0	2,28E-06	-1,7	21,3	5,34E-06	f.conv.adj35	C/T/Y	317	1	126			TRUE
47	S13_9894951	13	9894951	12,9	3,67E-06	-1,7	25,8	5,76E-07	-1,7	24,9	9,02E-07	f.conv.adj35	C/T/Y	252	1	191	7	CC 78.3%	TRUE
48	S13_9894952	13	9894952	12,9	3,67E-06	-1,7	25,8	5,76E-07	-1,7	24,9	9,02E-07	f.conv.adj35	C/T/Y	252	1	191			TRUE
49	S13_12804655	13	12804655	11,6	1,30E-05	0,6	23,1	2,17E-06	-0,6	17,5	3,47E-05	f.conv.adj35	T/C/Y	285	8	151			TRUE
50	S13_15109350	13	15109350	12,9	3,63E-06	1,7	25,3	7,39E-07	-1,7	25,2	7,51E-07	f.conv.adj35	T/C/Y	356	1	87	8	TC 90%	TRUE

#	Marker Name	Chr	Pos	F	p	add effect	add F	add p	dom effect	dom F	dom p	Trait	Allele	Obs1	Obs2	Obs3	BlockN°	Most frequent Haplotype	overlapping QTLs
51	S13_15109354	13	15109354	12,9	3,63E-06	-1,7	25,3	7,39E-07	-1,7	25,2	7,51E-07	f.conv.adj35	C/T/Y	356	1	87			TRUE
52	S14_2918142	14	2918142	14,1	1,18E-06	1,0	23,1	2,09E-06	-0,8	12,5	4,59E-04	f.conv.adj35	C/A/M	345	3	96			TRUE
53	S14_4019669	14	4019669	11,6	1,19E-05	-1,8	22,7	2,57E-06	-1,7	20,1	9,45E-06	f.conv.adj35	G/T/K	326	1	117			TRUE
54	S14_12043802	14	12043802	15,5	3,31E-07	1,3	29,6	8,88E-08	-1,4	30,4	6,11E-08	f.conv.adj35	T/C/Y	313	2	129	9	TGC 84.7%	TRUE
55	S14_12127914	14	12127914	12,9	3,69E-06	1,7	25,6	6,34E-07	-1,7	24,8	9,07E-07	f.conv.adj35	G/A/R	356	1	87			TRUE
56	S14_12127934	14	12127934	12,9	3,69E-06	-1,7	25,6	6,34E-07	-1,7	24,8	9,07E-07	f.conv.adj35	C/T/Y	356	1	87			TRUE
57	S15_6942668	15	6942668	13,8	1,52E-06	1,7	24,9	8,71E-07	-1,8	27,4	2,62E-07	f.conv.adj35	R/A/G	245	1	198			TRUE
58	S15_9763685	15	9763685	12,9	3,52E-06	1,7	25,3	7,32E-07	-1,7	25,3	7,40E-07	f.conv.adj35	C/A/M	318	1	125			TRUE
59	S17_1976385	17	1976385	12,6	4,75E-06	0,0	2,7	1,04E-01	0,0	16,2	6,69E-05	eff.adj35	C/G/S	246	66	132			FALSE
60	S17_4917070	17	4917070	13,1	3,15E-06	-1,7	26,0	5,18E-07	-1,7	23,9	1,48E-06	f.conv.adj35	C/G/S	312	1	131			TRUE
61	S17_4917071	17	4917071	13,1	3,15E-06	-1,7	26,0	5,18E-07	-1,7	23,9	1,48E-06	f.conv.adj35	G/T/K	312	1	131	10	CGT 85%	TRUE
62	S17_4917072	17	4917072	13,1	3,15E-06	1,7	26,0	5,18E-07	-1,7	23,9	1,48E-06	f.conv.adj35	T/G/K	312	1	131			TRUE
63	S17_5621341	17	5621341	13,1	2,94E-06	-1,7	25,3	7,38E-07	-1,8	25,5	6,73E-07	f.conv.adj35	C/T/Y	398	1	45	11	CC 94.7%	TRUE
64	S17_5621374	17	5621374	13,1	2,94E-06	-1,7	25,3	7,38E-07	-1,8	25,5	6,73E-07	f.conv.adj35	C/T/Y	398	1	45			TRUE
65	S17_5986188	17	5986188	13,3	2,43E-06	-1,7	25,0	8,30E-07	-1,8	26,6	3,91E-07	f.conv.adj35	C/T/Y	194	1	249			TRUE
66	S18_4390068	18	4390068	15,4	3,52E-07	-1,4	30,8	5,09E-08	-1,3	23,7	1,56E-06	f.conv.adj35	C/T/Y	397	2	45			TRUE
67	S18_9253828	18	9253828	13,0	3,31E-06	-1,7	26,0	5,16E-07	-1,7	22,4	3,09E-06	f.conv.adj35	A/C/M	391	1	52	12	ACC 93.4%	FALSE
68	S18_9551090	18	9551090	13,2	2,66E-06	-1,7	26,3	4,46E-07	-1,6	21,5	4,75E-06	f.conv.adj35	C/T/Y	393	1	50			FALSE
69	S18_9554600	18	9554600	15,4	3,64E-07	-1,4	30,7	5,33E-08	-1,3	24,4	1,16E-06	f.conv.adj35	C/T/Y	392	2	50			FALSE
70	S20_11847031	20	11847031	13,0	3,45E-06	1,7	25,9	5,45E-07	-1,7	24,5	1,07E-06	f.conv.adj35	T/C/Y	249	1	194	13	TC 77.9%	FALSE
71	S20_11847038	20	11847038	13,0	3,45E-06	-1,7	25,9	5,45E-07	-1,7	24,5	1,07E-06	f.conv.adj35	C/T/Y	249	1	194			FALSE
72	S21_725260	21	725260	13,1	3,02E-06	1,7	26,1	4,87E-07	-1,7	23,4	1,87E-06	f.conv.adj35	G/A/R	345	1	98			TRUE
73	S23_2109713	23	2109713	13,1	3,03E-06	1,7	25,1	7,90E-07	-1,8	26,1	4,92E-07	f.conv.adj35	G/A/R	339	1	104	16.5K b		TRUE
74	S23_4929597	23	4929597	16,5	1,29E-07	-1,6	30,1	6,91E-08	-1,7	32,9	1,84E-08	f.conv.adj35	A/G/R	356	2	86			TRUE
75	S25_601833	25	601833	13,8	1,51E-06	1,7	25,0	8,28E-07	-1,8	27,5	2,52E-07	f.conv.adj35	T/C/Y	305	1	138			FALSE
76	S26_534090	26	534090	15,5	3,12E-07	-1,0	31,0	4,54E-08	-0,9	23,5	1,75E-06	**f.conv.adj35**	C/T/Y	359	4	81	14	f.conv.adj35 79.4%	TRUE
77	S26_1249195	26	1249195	11,8	1,04E-05	-82,6	14,8	1,36E-04	-118,3	23,5	1,73E-06	**cr**	A/G/R	355	5	84			TRUE
78	S27_1587217	27	1587217	14,4	9,00E-07	1,8	26,9	3,26E-07	-1,5	18,9	1,77E-05	**f.conv.adj35**	C/A/M	400	1	43	15	f.conv.adj35G GGAACf.conv. adj35 94.8	TRUE
79	S27_1587218	27	1587218	14,4	9,00E-07	-1,8	26,9	3,26E-07	-1,5	18,9	1,77E-05	**f.conv.adj35**	A/C/M	400	1	43			TRUE
80	S27_1587219	27	1587219	14,4	9,00E-07	1,8	26,9	3,26E-07	-1,5	18,9	1,77E-05	**f.conv.adj35**	G/A/R	400	1	43			TRUE

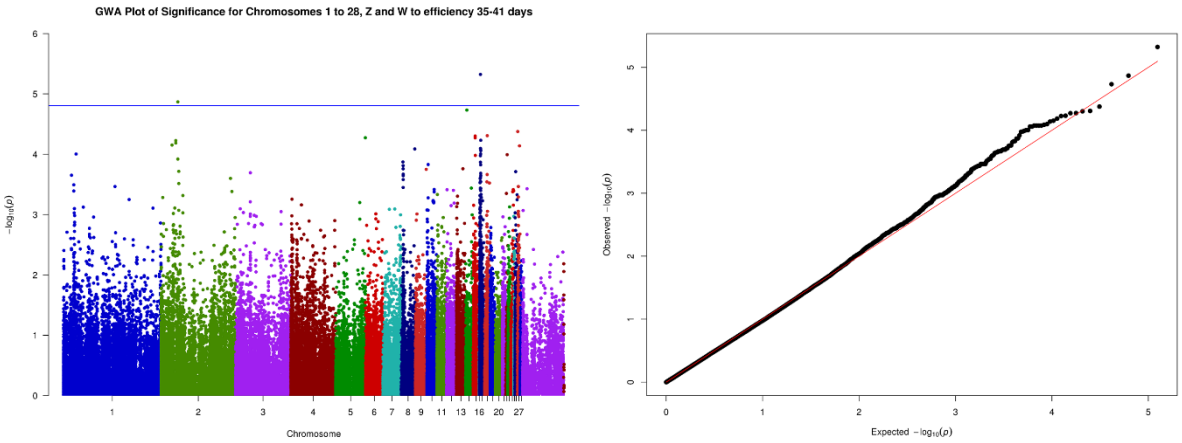
#	Marker Name	Chr	Pos	F	p	add effect	add F	add p	dom effect	dom F	dom p	Trait	Allele	Obs1	Obs2	Obs3	BlockNº	Most frequent Haplotype	overlapping QTLs
81	S27_1587220	27	1587220	15,3	3,78E-07	-1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	G/T/K	399	1	44			TRUE
82	S27_1587221	27	1587221	15,3	3,78E-07	1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	G /C/S	399	1	44			TRUE
83	S27_1587222	27	1587222	15,3	3,78E-07	-1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	A/C/M	399	1	44			TRUE
84	S27_1587224	27	1587224	15,3	3,78E-07	-1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	A/C/M	399	1	44			TRUE
85	S27_1587226	27	1587226	15,3	3,78E-07	1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	C/A/M	399	1	44			TRUE
86	S27_1587227	27	1587227	15,3	3,78E-07	-1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	C/G /S	399	1	44			TRUE
87	S27_1587228	27	1587228	15,3	3,78E-07	-1,8	27,3	2,73E-07	-1,5	17,7	3,12E-05	**f.conv.adj35**	A /G /M	399	1	44			TRUE
88	S27_3798382	27	3798382	15,9	2,15E-07	-1,3	28,7	1,38E-07	-1,5	30,9	4,91E-08	f.conv.adj35	C/G/S	389	2	53	16	CG 93.6%	TRUE
89	S27_3798470	27	3798470	13,0	3,39E-06	1,7	25,2	7,56E-07	-1,8	25,0	8,44E-07	f.conv.adj35	G/A/R	392	1	51			TRUE
90	S27_4784671	27	4784671	12,9	3,55E-06	1,7	25,8	5,57E-07	-1,7	24,5	1,08E-06	f.conv.adj35	G/A/R	275	1	168			FALSE
91	S28_788219	28	788219	11,5	1,41E-05	-50,6	15,7	8,88E-05	24,1	2,6	1,08E-01	bw41	T/C/Y	53	214	177	17	f.conv.adj35 34.5 CG 33.6 TG 30.3	FALSE
92	S28_1252731	28	1252731	13,4	2,22E-06	-66,0	26,2	4,57E-07	20,4	2,0	1,58E-01	bw41	G/A/R	171	48	225			FALSE
93	S28_2986566	28	2986566	13,2	2,86E-06	1,7	26,1	4,81E-07	-1,7	23,8	1,55E-06	f.conv.adj35	G/A/R	298	1	145			TRUE
94	S33_9064309	Z	9064309	15,2	4,04E-07	-1,3	29,8	8,08E-08	-1,4	28,6	1,47E-07	f.conv.adj35	A/G/R	350	2	92			FALSE

Appendix B – Suggestively associated SNPs with feed intake (a), and feed efficiency (b) both adjusted for body weight at 35 days and body weight at 41 (c) days of age and birth weight (d) are presented by Manhattan (left side) and QQ (right side) plots. The y-axis is shown as -log10 (p-value) for both graphs. On the left, the blue line indicates suggestive genome-wise association (P < 1.57E-05) with the respective trait. On the right side, the QQ-plots show the relation of normal theoretical quantiles of the probability distributions between expected (x-axis) and observed (y-axis) p-values from each respective associated trait

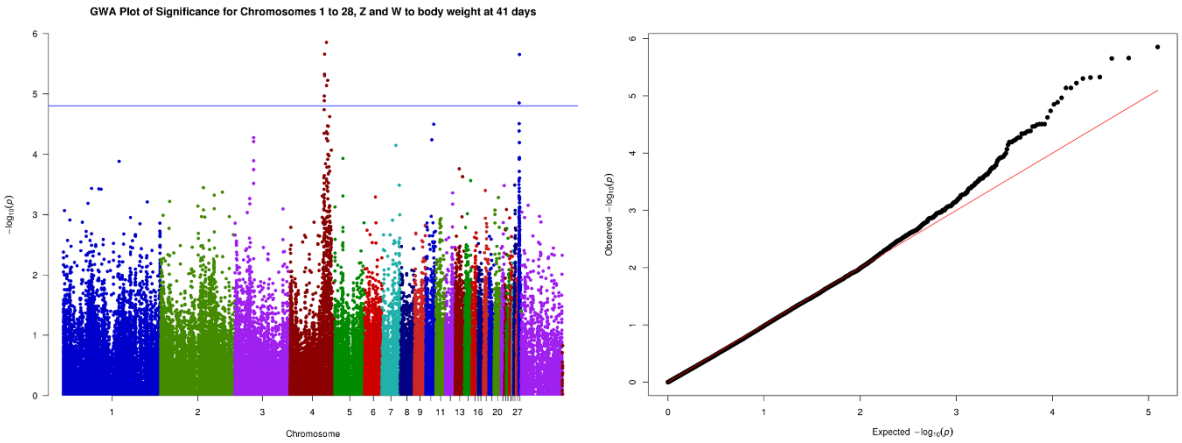
(a)



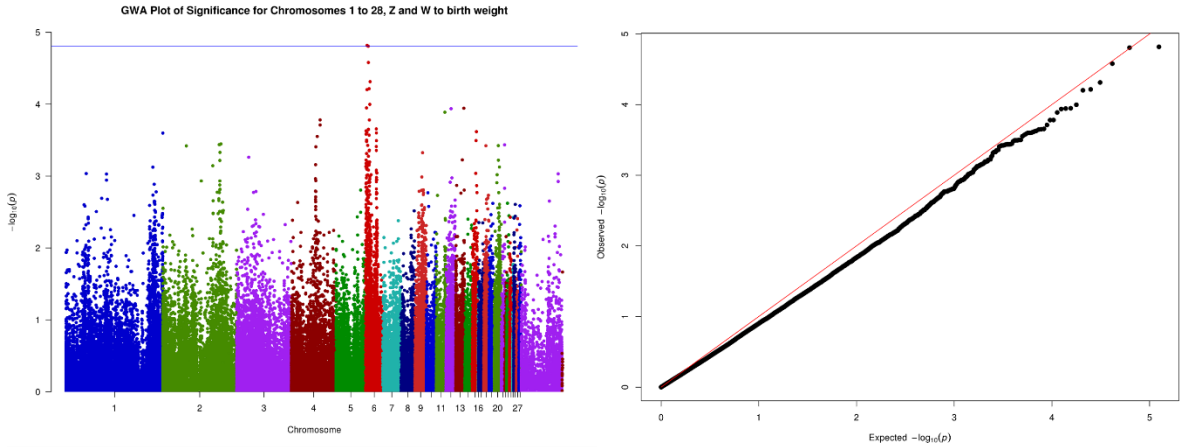
(b)



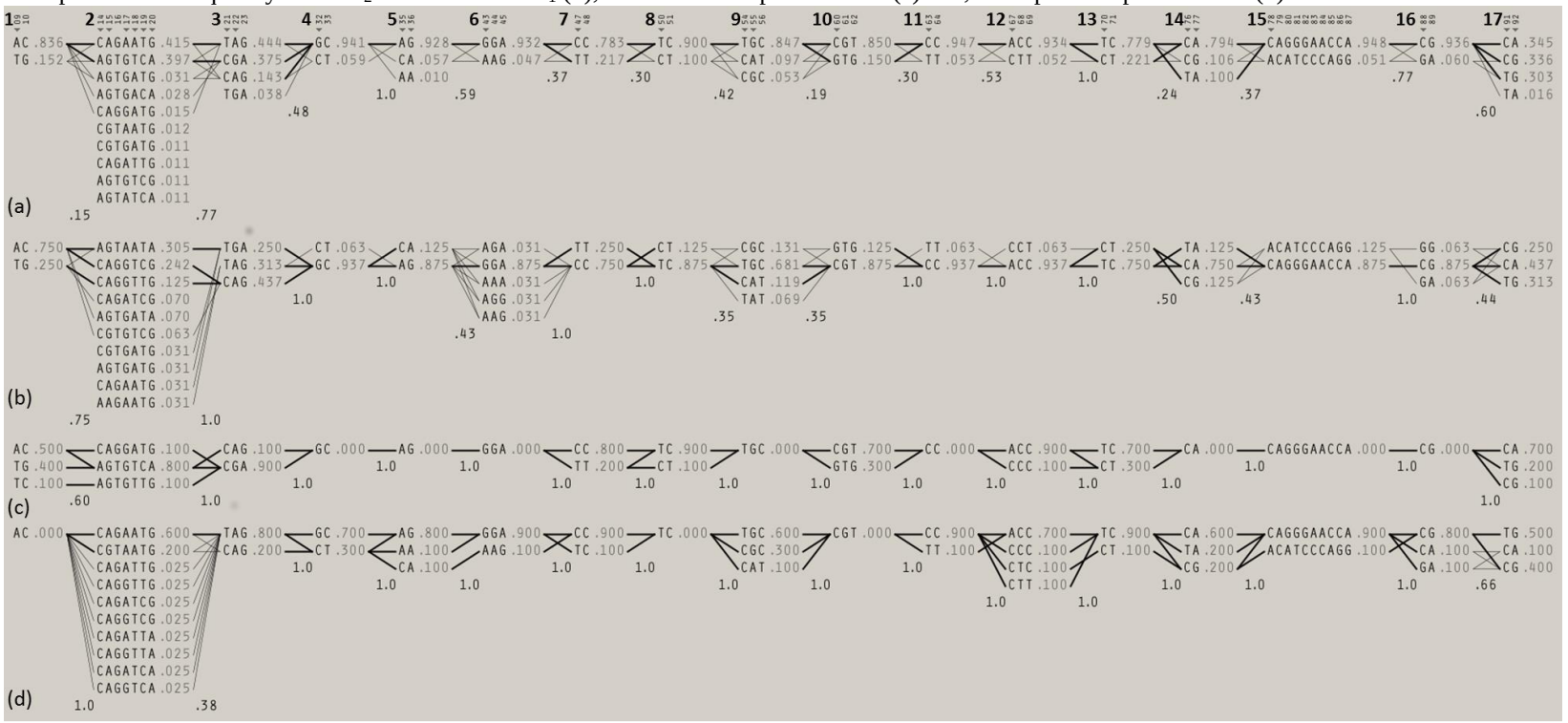
(c)



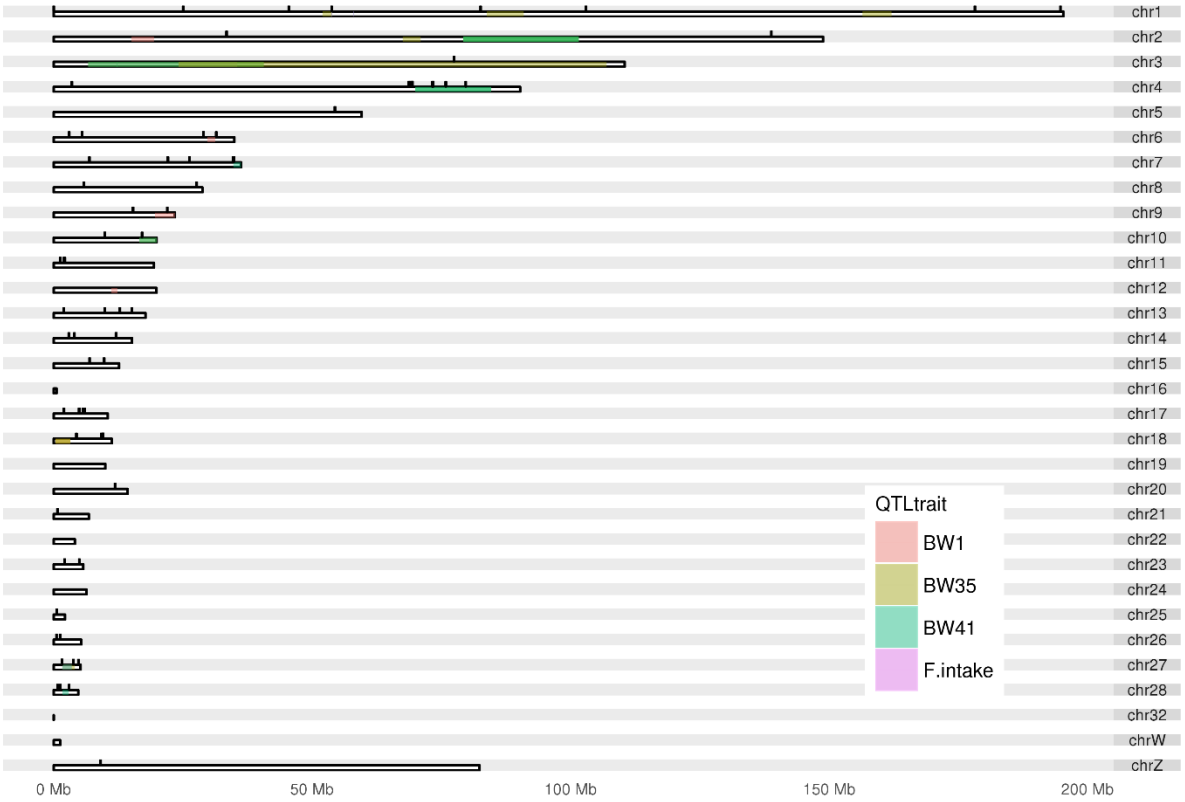
(d)


GWA Plot of Significance for Chromosomes 1 to 28, Z and W to birth weight
−log10(p)
Chromosome
1
2
3
4
5
6
7
8
9
11
13
16
20
27
Observed −log10(p)
Expected −log10(p)

Appendix C – Haplotype blocks obtained by the solid spine of LD and family structure using Haploview 4.2. The header represents the block number and tagSNP for each 17 blocks obtained from 94 markers in LD presented in Figure 3.3 of the manuscript (r^2>0.6 except block 17, r^2>0.2) in 444 F_2 individuals (a). The other letters represent the frequency of the F_2 blocks in the 8 F_1 (b), 5 CC maternal parental line (c) and, 5 TT paternal parental line (d)



Appendix D – Karyotype of the QTLs (from Animal QTLdb) distribution regions of the chicken genome overlapping suggestive and significant SNPs associated with performance traits (black marks). We subset only QTLs mapped for birth weight (BW1), body weight at 35 (BW35), 41 (BW41) and 42 (BW42) days of age, and feed intake (f.intake) traits performed in the same F_2 population used in the presented manuscript and published before



Appendix E – Mendelian descriptions for suggestive (P < 1.57E-05) and/or genome-wise associated (P < 7.86E-07) 94 markers from 444 F_2 individuals

#	Markers Chr_pos	ObsHET	PredHET	HWpval		MAF	Alleles
1	1_3706	0.21	0.19	4.71E-02		0.11	C:G
2	1_25041791	0.22	0.20	1.21E-01		0.11	C:T
3	1_45468342	0.12	0.11	1.00E+00		0.06	T:C
4	1_53749693	0.10	0.10	1.00E+00		0.05	G:A
5	1_82575973	0.53	0.39	7.40E-18	*	0.27	T:C
6	1_102944294	0.24	0.22	3.81E-02		0.13	A:C
7	1_178184151	0.12	0.12	9.36E-01		0.06	G:A
8	1_194726625	0.21	0.19	5.82E-02		0.11	A:C
9	2_33436237	0.30	0.27	1.25E-02		0.16	A:T
10	2_33437336	0.31	0.27	5.87E-05	*	0.16	C:G
11	2_138815862	0.09	0.10	5.43E-02		0.05	G:A
12	3_77435334	0.13	0.12	8.16E-01		0.07	C:T
13	4_3514658	0.17	0.16	2.90E-01		0.09	G:T
14	4_68623163	0.52	0.50	5.79E-01		0.49	A:C
15	4_68882750	0.48	0.50	4.25E-01		0.47	G:A
16	4_68882765	0.48	0.50	3.69E-01		0.47	T:G
17	4_69297525	0.43	0.50	6.80E-03		0.47	G:A
18	4_69370079	0.43	0.50	4.20E-03		0.46	A:T
19	4_69372005	0.49	0.50	7.21E-01		0.48	T:C
20	4_69372065	0.49	0.50	9.01E-01		0.47	G:A
21	4_73210325	0.42	0.50	9.00E-04	*	0.48	C:T
22	4_73407243	0.46	0.49	2.58E-01		0.41	A:G
23	4_73407249	0.46	0.49	2.58E-01		0.41	G:A
24	4_75842191	0.42	0.49	7.00E-03		0.41	A:C
25	4_79707389	0.37	0.45	8.06E-05	*	0.35	C:G
26	5_54401932	0.19	0.18	1.16E-01		0.10	C:T
27	6_2997851	0.14	0.17	1.60E-03		0.10	G:A
28	6_5492179	0.49	0.50	8.21E-01		0.47	A:C
29	6_28967516	0.29	0.25	4.00E-04	*	0.15	G:T
30	6_31436207	0.41	0.33	1.58E-09	*	0.21	T:C
31	7_6904443	0.94	0.50	2.00E-97	*	0.47	G:A
32	7_22092999	0.11	0.11	1.00E+00		0.06	G:C
33	7_22093009	0.11	0.11	1.00E+00		0.06	C:T
34	7_26252164	0.13	0.13	7.40E-01		0.07	C:T
35	7_34668134	0.11	0.12	1.54E-01		0.06	A:C
36	7_34855716	0.13	0.13	7.78E-01		0.07	G:A
37	8_5848411	0.30	0.26	3.00E-04	*	0.15	T:C
38	8_27651130	0.15	0.15	9.07E-01		0.08	T:C
39	9_15349595	0.12	0.12	9.36E-01		0.06	G:A
40	9_21971358	0.21	0.19	5.24E-02		0.11	C:T
41	10_9890328	0.33	0.28	1.46E-05	*	0.17	A:G
42	10_17086845	0.12	0.12	9.36E-01		0.06	G:A
43	11_1249418	0.11	0.10	1.00E+00		0.06	G:A
44	11_1997302	0.12	0.11	1.00E+00		0.06	G:A
45	11_2149223	0.12	0.11	1.00E+00		0.06	A:G
46	13_1958055	0.28	0.25	6.00E-04	*	0.14	C:T
47	13_9894951	0.43	0.34	1.43E-10	*	0.22	C:T
48	13_9894952	0.43	0.34	1.43E-10	*	0.22	C:T
49	13_12804655	0.34	0.31	2.00E-02		0.19	T:C
50	13_15109350	0.20	0.18	9.60E-02		0.10	T:C
51	13_15109354	0.20	0.18	9.60E-02		0.10	C:T
52	14_2918142	0.22	0.20	2.74E-01		0.12	C:A
53	14_4019669	0.26	0.23	2.50E-03		0.13	G:T
54	14_12043802	0.29	0.26	1.90E-03		0.15	T:C
55	14_12127914	0.20	0.18	9.60E-02		0.10	G:A
56	14_12127934	0.20	0.18	9.60E-02		0.10	C:T

#	Markers Chr_pos	ObsHET	PredHET	HWpval		MAF	Alleles
57	15_6942668	0.55	0.40	1.22E-19	*	0.28	G:A
58	15_9763685	0.28	0.25	7.00E-04	*	0.14	C:A
59	17_1976385	0.30	0.42	4.25E-09	*	0.30	C:G
60	17_4917070	0.30	0.26	3.00E-04	*	0.15	C:G
61	17_4917071	0.30	0.26	3.00E-04	*	0.15	G:T
62	17_4917072	0.30	0.26	3.00E-04	*	0.15	T:G
63	17_5621341	0.10	0.10	1.00E+00		0.05	C:T
64	17_5621374	0.10	0.10	1.00E+00		0.05	C:T
65	17_5986188	0.56	0.41	1.79E-20	*	0.28	C:T
66	18_4390068	0.10	0.10	7.77E-01		0.06	C:T
67	18_9253828	0.12	0.11	1.00E+00		0.06	A:C
68	18_9551090	0.11	0.11	1.00E+00		0.06	C:T
69	18_9554600	0.11	0.11	9.82E-01		0.06	C:T
70	20_11847031	0.44	0.34	5.59E-11	*	0.22	T:C
71	20_11847038	0.44	0.34	5.59E-11	*	0.22	C:T
72	21_725260	0.22	0.20	3.02E-02		0.11	G:A
73	23_2109713	0.23	0.21	1.47E-02		0.12	G:A
74	23_4929597	0.19	0.18	2.86E-01		0.10	A:G
75	25_601833	0.31	0.27	7.10E-05	*	0.16	T:C
76	26_534090	0.18	0.18	1.00E+00		0.10	C:T
77	26_1249195	0.19	0.19	1.00E+00		0.11	A:G
78	27_1587217	0.10	0.10	1.00E+00		0.05	C:A
79	27_1587218	0.10	0.10	1.00E+00		0.05	A:C
80	27_1587219	0.10	0.10	1.00E+00		0.05	G:A
81	27_1587220	0.10	0.10	1.00E+00		0.05	G:T
82	27_1587221	0.10	0.10	1.00E+00		0.05	G:C
83	27_1587222	0.10	0.10	1.00E+00		0.05	A:C
84	27_1587224	0.10	0.10	1.00E+00		0.05	A:C
85	27_1587226	0.10	0.10	1.00E+00		0.05	C:A
86	27_1587227	0.10	0.10	1.00E+00		0.05	C:G
87	27_1587228	0.10	0.10	1.00E+00		0.05	A:G
88	27_3798382	0.12	0.12	1.00E+00		0.06	C:G
89	27_3798470	0.12	0.11	1.00E+00		0.06	G:A
90	27_4784671	0.38	0.31	9.31E-08	*	0.19	G:A
91	28_788219	0.40	0.43	1.02E-01		0.32	C:T
92	28_1252731	0.51	0.46	5.19E-02		0.36	G:A
93	28_2986566	0.33	0.28	1.79E-05	*	0.17	G:A
94	Z_9064309	0.21	0.19	1.75E-01		0.11	A:G

*HW p-value cutoff: 0.001, Chr – chromosome, pos – marker position on chromosome

Appendix F – The table is rating who has the advantageous alleles from the parental lines (mother/father) (adv. parental) for the trait evaluated from each of the 17 LD blocks. The first column indicates the block number, the second and third indicate whether this haplotype is fixed in the father/mother and the fourth column indicates the parental which concentrates more frequently the advantageous allele

Block	Father	Mother	adv. parental
1	fixed	variable	father
2	variable	variable	father
3	fixed	variable	father
4	variable	fixed	mother
5	variable	fixed	mother
6	variable	fixed	mother
7	fixed	variable	father
8	fixed	variable	father
9	variable	fixed	mother
10	fixed	variable	father
11	variable	fixed	mother
12	variable	fixed	mother
13	variable	variable	father
14	variable	fixed	mother
15	variable	fixed	mother
16	variable	fixed	mother
17	variable	variable	father/mother

Appendix G – List of 253 QTLs overlapping 81 markers obtained from overlapping test of 94 associated (P < 7.86E-07) and/or suggestive (P < 1.57E-05) associated markers from 444 F_2 individuals in this study against 1,458 QTLs obtained from the Animal QTLdb

chr	start	end	QTL_ID	Trait	Abrev
chr1	2329958	64931354	QTL_ID=9406	Body_weight_(63_days)	Bw
chr1	7010437	27734493	QTL_ID=6752	Body_weight_(70_days)	Bw
chr1	23286118	25486484	QTL_ID=6751	Body_weight_(63_days)	Bw
chr1	23286118	25486484	QTL_ID=6753	Body_weight_(77_days)	Bw
chr1	25275933	48175152	QTL_ID=1775	Body_weight	Bw
chr1	25275933	48175152	QTL_ID=9739	Body_weight_(21_days)	Bw
chr1	25275933	48175152	QTL_ID=9740	Body_weight_(42_days)	Bw
chr1	25275933	48175152	QTL_ID=9741	Body_weight_(63_days)	Bw
chr1	28436329	63038738	QTL_ID=1807	Body_weight	Bw
chr1	28436329	63038738	QTL_ID=1808	Body_weight	bw
chr1	33174045	134863919	QTL_ID=17076	Body_weight_(140_days)	bw
chr1	33279320	47367912	QTL_ID=6675	Feed_conversion_ratio	f.conv
chr1	33279320	47367912	QTL_ID=6674	Residual_feed_intake	f.intake
chr1	42474406	52428237	QTL_ID=1788	Body_weight	bw
chr1	45111848	45996329	QTL_ID=24950	Body_weight	bw
chr1	50511472	82050274	QTL_ID=55913	Body_weight_(105_days)	bw
chr1	50511472	71471379	QTL_ID=55933	Growth_(35-70_days)	grow
chr1	50722023	68558759	QTL_ID=55902	Body_weight_(35_days)	bw
chr1	50722023	69541330	QTL_ID=55908	Body_weight_(70_days)	bw
chr1	50722023	67611281	QTL_ID=55927	Growth_(0-35_days)	grow
chr1	51977644	195276750	QTL_ID=1797	Body_weight	bw
chr1	51977644	55261640	QTL_ID=24872	Body_weight_(168_days)	bw
chr1	51977644	55261640	QTL_ID=24839	Body_weight_(21_days)	bw
chr1	51977644	53798691	***QTL_ID=3324**	Body_weight_(35_days)	bw
chr1	51977644	55261640	QTL_ID=24848	Body_weight_(42_days)	bw
chr1	51977644	53798691	QTL_ID=3325	Body_weight_(42_days)	bw
chr1	51977644	55261640	QTL_ID=14462	Body_weight_(day_of_f.intakerst_egg)	bw
chr1	51977644	55261640	QTL_ID=14467	Body_weight_(day_of_f.intakerst_egg)	bw
chr1	57051322	152887033	QTL_ID=55937	Growth_(70-105_days)	grow
chr1	59332289	134126991	QTL_ID=55919	Body_weight_(140_days)	bw
chr1	81804631	84531095	QTL_ID=9110	Growth_(post-challenge)	grow
chr1	90354029	123032393	QTL_ID=1821	Feed_efficience	effic
chr1	92255503	108492449	QTL_ID=6813	Body_weight_(56_days)	bw
chr1	92255503	108492449	QTL_ID=6814	Body_weight_(56_days)	bw
chr1	92255503	108492449	QTL_ID=6815	Body_weight_(63_days)	bw
chr1	92255503	108492449	QTL_ID=6812	Body_weight_(7_days)	bw
chr1	92255503	108492449	QTL_ID=6816	Body_weight_(77_days)	bw
chr1	92255503	108492449	QTL_ID=6817	Body_weight_(84_days)	bw
chr1	94157976	113198023	QTL_ID=1822	Feed_intake	f.intake
chr1	143677314	191778894	QTL_ID=6583	Body_weight_(112_days)	bw
chr1	143677314	191778894	QTL_ID=6584	Body_weight_(200_days)	bw
chr1	143677314	191778894	QTL_ID=6582	Body_weight_(46_days)	bw
chr1	143677314	191778894	QTL_ID=6579	Body_weight_(8_days)	bw
chr1	143677314	191778894	QTL_ID=6588	Growth_(112-200_days)	grow
chr1	143677314	191778894	QTL_ID=6585	Growth_(1-8_days)	grow
chr1	143677314	191778894	QTL_ID=6587	Growth_(46-112_days)	grow
chr1	143677314	191778894	QTL_ID=6586	Growth_(8-46_days)	grow
chr1	156472083	191778894	QTL_ID=1855	Body_weight	bw
chr1	156472083	191778894	QTL_ID=1858	Body_weight	bw
chr1	156472083	195276750	QTL_ID=24873	Body_weight_(168_days)	bw
chr1	156472083	191778894	QTL_ID=9750	Body_weight_(21_days)	bw
chr1	156472083	191778894	QTL_ID=9751	Body_weight_(42_days)	bw
chr1	178182301	178252485	QTL_ID=16642	Body_weight_(28_days)	bw
chr1	178182301	178252485	QTL_ID=16667	Body_weight_(28_days)	bw
chr1	178182301	178252485	QTL_ID=16643	Body_weight_(35_days)	bw
chr1	178182301	178252485	QTL_ID=16644	Body_weight_(42_days)	bw
chr1	178182301	178252485	QTL_ID=16669	Body_weight_(42_days)	bw
chr1	178182301	178252485	QTL_ID=16645	Body_weight_(49_days)	bw
chr1	178182301	178252485	QTL_ID=16670	Body_weight_(49_days)	bw
chr1	178182301	178252485	QTL_ID=16646	Body_weight_(56_days)	bw
chr1	178182301	178252485	QTL_ID=16671	Body_weight_(56_days)	bw
chr1	178182301	178252485	QTL_ID=16647	Body_weight_(63_days)	bw
chr1	178182301	178252485	QTL_ID=16672	Body_weight_(63_days)	bw
chr1	178182301	178252485	QTL_ID=16648	Body_weight_(70_days)	bw
chr1	178182301	178252485	QTL_ID=16673	Body_weight_(70_days)	bw
chr1	178182301	178252485	QTL_ID=16649	Body_weight_(77_days)	bw
chr1	178182301	178252485	QTL_ID=16674	Body_weight_(77_days)	bw
chr1	178182301	178252485	QTL_ID=16650	Body_weight_(84_days)	bw
chr1	178182301	178252485	QTL_ID=16675	Body_weight_(84_days)	bw
chr2	578373	37938317	QTL_ID=1871	Body_weight	bw
chr2	578373	37938317	QTL_ID=1872	Body_weight	bw
chr2	2452503	37938317	QTL_ID=1873	Body_weight	bw
chr2	15053040	37938317	QTL_ID=1874	Body_weight	bw
chr2	17958970	37831227	QTL_ID=6848	Body_weight_(42_days)	bw
chr2	21441194	71328860	QTL_ID=55934	Growth_(35-70_days)	grow
chr2	29064142	108449411	QTL_ID=55909	Body_weight_(70_days)	bw
chr2	29345262	65036736	QTL_ID=55914	Body_weight_(105_days)	bw
chr2	103056068	144263101	QTL_ID=9415	Body_weight_(63_days)	bw
chr2	129556328	148809762	QTL_ID=1928	Body_weight_(test_end)	bw
chr3	2811981	103804757	QTL_ID=1979	Body_weight	bw
chr3	2811981	103804757	QTL_ID=1980	Body_weight	bw
chr3	6645961	106970113	***QTL_ID=7180**	Body_weight_(35_days)	bw
chr3	6841859	102661494	QTL_ID=55904	Body_weight_(35_days)	bw
chr3	6841859	99827114	QTL_ID=55929	Growth_(0-35_days)	grow
chr3	24160710	79800194	QTL_ID=1957	Body_weight	bw
chr3	37579996	108174898	QTL_ID=1961	Body_weight	bw
chr3	37579996	108174898	QTL_ID=1962	Body_weight	bw
chr3	45203763	84080722	QTL_ID=6611	Body_weight_(112_days)	bw
chr3	45203763	84080722	QTL_ID=6612	Body_weight_(200_days)	bw
chr3	45203763	84080722	QTL_ID=6610	Body_weight_(8_days)	bw
chr3	45203763	84080722	QTL_ID=6613	Growth_(1-8_days)	grow
chr3	45311696	104141066	QTL_ID=9420	Body_weight_(63_days)	bw
chr3	57793506	84765752	QTL_ID=11768	Body_weight_(49_days)	bw
chr3	57793506	84765752	QTL_ID=11772	Body_weight_(63_days)	bw
chr3	58466018	88819995	QTL_ID=1969	Body_weight	bw
chr3	58466018	88819995	QTL_ID=1972	Body_weight	bw
chr3	72032411	84765752	QTL_ID=9127	Growth_(post-challenge)	grow
chr4	3459678	19246773	QTL_ID=1989	Body_weight	bw
chr4	3459678	19246773	QTL_ID=1990	Body_weight	bw
chr4	3459678	19246773	QTL_ID=1991	Body_weight	bw
chr4	3459678	16468784	QTL_ID=24854	Body_weight_(42_days)	bw
chr4	3459678	16468784	QTL_ID=24865	Body_weight_(84_days)	bw
chr4	17039416	80254980	QTL_ID=24875	Body_weight_(168_days)	bw
chr4	17039416	80254980	QTL_ID=24842	Body_weight_(21_days)	bw
chr4	17039416	80254980	QTL_ID=24883	Body_weight_(336_days)	bw
chr4	17039416	80254980	QTL_ID=24855	Body_weight_(42_days)	bw
chr4	17039416	80254980	QTL_ID=24890	Body_weight_(504_days)	bw
chr4	17039416	80254980	QTL_ID=24866	Body_weight_(84_days)	bw
chr4	17039416	80254980	QTL_ID=14457	Body_weight_(day_of_f.intakerst_egg)	bw
chr4	17039416	80254980	QTL_ID=14464	Body_weight_(day_of_f.intakerst_egg)	bw
chr4	17039416	80254980	QTL_ID=14470	Body_weight_(day_of_f.intakerst_egg)	bw
chr4	30753891	88270681	QTL_ID=55905	Body_weight_(35_days)	bw
chr4	32166525	88270681	QTL_ID=55930	Growth_(0-35_days)	grow
chr4	46739707	88408499	QTL_ID=2008	Body_weight	bw
chr4	46739707	88408499	QTL_ID=2015	Body_weight	bw
chr4	46739707	88408499	QTL_ID=2016	Body_weight	bw
chr4	46739707	88408499	QTL_ID=9759	Body_weight_(42_days)	bw
chr4	46739707	88408499	QTL_ID=9760	Body_weight_(63_days)	bw
chr4	46739707	88408499	QTL_ID=9761	Growth_(21-42_days)	grow
chr4	46739707	88408499	QTL_ID=9762	Growth_(42-63_days)	grow
chr4	47497706	81608525	QTL_ID=12498	Growth_(0-14_days)	grow
chr4	47897157	84135947	QTL_ID=12500	Growth_(28-42_days)	grow
chr4	51678002	81608525	QTL_ID=12499	Growth_(14-28_days)	grow
chr4	61658027	87975245	QTL_ID=12501	Growth_(42-56_days)	grow
chr4	67602577	86114364	QTL_ID=55942	Growth_(105-140_days)	grow
chr4	68050486	84928400	QTL_ID=17069	Body_weight_(224_days)	bw
chr4	68567303	80565058	QTL_ID=55938	Growth_(70-105_days)	grow
chr4	69497574	81401798	QTL_ID=55915	Body_weight_(105_days)	bw
chr4	69942858	88408499	QTL_ID=11766	Body_weight_(35_days)	bw
chr4	69942858	84618310	QTL_ID=7181	Body_weight_(35_days)	bw
chr4	69942858	84618310	***QTL_ID=7157**	Body_weight_(35_days)	bw
chr4	69942858	84618310	***QTL_ID=7185**	Body_weight_(41_days)	bw

chr	start	end	QTL_ID	Trait	Abrev
chr4	69942858	84618310	***QTL_ID=7162**	Body_weight_(41_days)	bw
chr4	69942858	88408499	QTL_ID=11769	Body_weight_(49_days)	bw
chr4	69942858	88408499	QTL_ID=11773	Body_weight_(63_days)	bw
chr4	70976493	80254980	QTL_ID=2026	Body_weight	bw
chr4	70976493	80254980	QTL_ID=2023	Feed_intake	f.intake
chr5	13971153	59580361	QTL_ID=55939	Growth_(70-105_days)	grow
chr5	22465118	58876488	QTL_ID=17079	Body_weight_(140_days)	bw
chr5	30162850	59580361	QTL_ID=9763	Growth_(42-63_days)	grow
chr6	1732132	29763072	QTL_ID=55916	Body_weight_(105_days)	bw
chr6	14060241	31613296	QTL_ID=2125	Body_weight	bw
chr6	26971183	31244082	QTL_ID=24843	Body_weight_(21_days)	bw
chr6	26971183	31244082	QTL_ID=9764	Body_weight_(42_days)	bw
chr6	26971183	31244082	QTL_ID=24867	Body_weight_(84_days)	bw
chr6	26971183	31244082	QTL_ID=9765	Growth_(21-42_days)	grow
chr6	28380699	29847153	QTL_ID=12502	Growth_(0-14_days)	grow
chr7	1293359	22946278	QTL_ID=2136	Body_weight	bw
chr7	1293359	27006229	QTL_ID=2149	Body_weight	bw
chr7	1293359	27006229	QTL_ID=2151	Body_weight	bw
chr7	1293359	27006229	QTL_ID=2153	Body_weight	bw
chr7	1293359	28666221	QTL_ID=2147	Body_weight	bw
chr7	1293359	28666221	QTL_ID=2148	Body_weight	bw
chr7	1293359	22946278	QTL_ID=6626	Body_weight_(112_days)	bw
chr7	1293359	22946278	QTL_ID=6627	Body_weight_(200_days)	bw
chr7	4794981	24245453	QTL_ID=2150	Body_weight	bw
chr7	7699967	36245040	QTL_ID=2146	Body_weight	bw
chr7	7699967	36245040	QTL_ID=2158	Body_weight	bw
chr7	7699967	36245040	QTL_ID=2160	Body_weight	bw
chr7	12980232	24245453	QTL_ID=17305	Body_weight_(14_days)	bw
chr7	12980232	24245453	QTL_ID=17308	Body_weight_(35_days)	bw
chr7	12980232	24245453	QTL_ID=17314	Body_weight_(77_days)	bw
chr7	24245453	28666221	QTL_ID=17309	Body_weight_(42_days)	bw
chr7	24245453	28666221	QTL_ID=17310	Body_weight_(49_days)	bw
chr7	24245453	28666221	QTL_ID=17311	Body_weight_(56_days)	bw
chr7	24245453	28666221	QTL_ID=17312	Body_weight_(63_days)	bw
chr7	24245453	28666221	QTL_ID=17304	Body_weight_(7_days)	bw
chr7	24245453	28666221	QTL_ID=17313	Body_weight_(70_days)	bw
chr7	24245453	28666221	QTL_ID=17315	Body_weight_(84_days)	bw
chr7	25665007	26638454	QTL_ID=64529	Feed_conversion_ratio	f.conv
chr7	25675920	26648497	QTL_ID=95407	Body_weight_(21_days)	bw
chr7	28666221	36245040	QTL_ID=6625	Body_weight_(8_days)	bw
chr7	34723350	36245040	***QTL_ID=7163**	Body_weight_(41_days)	bw
chr8	6726971	28767244	QTL_ID=24891	Body_weight_(504_days)	bw
chr8	6967382	28767244	QTL_ID=55922	Body_weight_(140_days)	bw
chr8	6967382	28767244	QTL_ID=55940	Growth_(70-105_days)	grow
chr8	8398681	28767244	QTL_ID=2190	Body_weight	bw
chr8	8398681	28767244	QTL_ID=2199	Body_weight	bw
chr8	8398681	28767244	QTL_ID=2201	Body_weight	bw
chr8	8398681	28767244	QTL_ID=9767	Body_weight_(63_days)	bw
chr8	8398681	28767244	QTL_ID=9768	Growth_(21-42_days)	grow
chr8	19901881	28767244	QTL_ID=24877	Body_weight_(168_days)	bw
chr8	19901881	28767244	QTL_ID=24844	Body_weight_(21_days)	bw
chr8	19901881	28767244	QTL_ID=24857	Body_weight_(42_days)	bw
chr8	19901881	28767244	QTL_ID=24868	Body_weight_(84_days)	bw
chr8	19901881	28767244	QTL_ID=14458	Body_weight_(day_of_f.intakerst_egg)	bw
chr8	19901881	28767244	QTL_ID=14471	Body_weight_(day_of_f.intakerst_egg)	bw
chr9	5124631	18131981	QTL_ID=2217	Body_weight	bw
chr9	6503007	23441680	QTL_ID=9442	Body_weight_(63_days)	bw
chr9	13658592	19669473	QTL_ID=6634	Body_weight_(200_days)	bw
chr9	13658592	23441680	QTL_ID=24886	Body_weight_(336_days)	bw
chr9	13658592	23441680	QTL_ID=24869	Body_weight_(84_days)	bw
chr9	13658592	19669473	QTL_ID=6635	Growth_(46-112_days)	grow
chr9	19669473	23441680	***QTL_ID=7177**	Body_weight_(1_day)	bw
chr10	1114728	15658282	QTL_ID=55923	Body_weight_(140_days)	bw
chr10	1924750	17334850	QTL_ID=55907	Body_weight_(35_days)	bw
chr10	2119755	17529855	QTL_ID=55931	Growth_(0-35_days)	grow
chr10	2142052	19911089	QTL_ID=55911	Body_weight_(70_days)	bw
chr10	3972100	17904865	QTL_ID=55917	Body_weight_(105_days)	bw
chr10	15658282	19911089	QTL_ID=2234	Body_weight_(ascites_conditions)	bw
chr10	16519830	19911089	***QTL_ID=7158**	Body_weight_(35_days)	bw
chr10	16519830	19911089	***QTL_ID=7164**	Body_weight_(41_days)	bw
chr11	888408	19401079	QTL_ID=17080	Body_weight_(140_days)	bw
chr11	975081	8298366	QTL_ID=64559	Feed_intake	f.intake
chr11	1063950	19185191	QTL_ID=55924	Body_weight_(140_days)	bw
chr11	2091236	3659428	QTL_ID=24845	Body_weight_(21_days)	bw
chr11	2091236	3659428	QTL_ID=24858	Body_weight_(42_days)	bw
chr13	721942	17760035	QTL_ID=24859	Body_weight_(42_days)	bw
chr13	721942	17760035	QTL_ID=24870	Body_weight_(84_days)	bw
chr13	721942	17760035	QTL_ID=14472	Body_weight_(day_of_f.intakerst_egg)	bw
chr13	8120084	17760035	QTL_ID=2310	Body_weight	bw
chr13	8120084	17760035	QTL_ID=9769	Body_weight_(21_days)	bw
chr13	8120084	17760035	QTL_ID=9770	Body_weight_(42_days)	bw
chr13	8120084	17760035	QTL_ID=9771	Body_weight_(63_days)	bw
chr13	8120084	17760035	QTL_ID=9772	Growth_(21-42_days)	grow
chr13	9323148	14111077	QTL_ID=9444	Body_weight_(63_days)	bw
chr13	10363506	16327806	QTL_ID=6645	Body_weight_(46_days)	bw
chr13	10363506	16327806	QTL_ID=6646	Growth_(1-8_days)	grow
chr13	12437243	17760035	QTL_ID=17081	Body_weight_(140_days)	bw
chr13	14111077	16327806	QTL_ID=2304	Body_weight	bw
chr14	50307	7445319	QTL_ID=6647	Body_weight_(1_day)	bw
chr14	50307	14696724	QTL_ID=17082	Body_weight_(140_days)	bw
chr14	50307	14696724	QTL_ID=55943	Growth_(105-140_days)	grow
chr14	133818	14696724	QTL_ID=55925	Body_weight_(140_days)	bw
chr14	3987816	7445319	QTL_ID=2328	Body_weight	bw
chr15	2798507	10631416	QTL_ID=24887	Body_weight_(336_days)	bw
chr15	3717446	8184057	QTL_ID=3355	Body_weight_(35_days)	bw
chr15	3717446	10631416	QTL_ID=6648	Body_weight_(46_days)	bw
chr15	3717446	10631416	QTL_ID=6649	Growth_(8-46_days)	grow
chr17	2449728	6121982	QTL_ID=2355	Body_weight	bw
chr18	1430572	4653744	QTL_ID=6650	Body_weight_(8_days)	bw
chr18	1430572	4653744	QTL_ID=6651	Growth_(1-8_days)	grow
chr18	3180498	7587685	QTL_ID=9773	Body_weight_(42_days)	bw
chr18	3180498	7587685	QTL_ID=9774	Growth_(21-42_days)	grow
chr21	13225	979078	QTL_ID=95430	Body_weight	bw
chr23	1779254	5097178	QTL_ID=9124	Growth_(post-challenge)	grow
chr26	58444	2538399	QTL_ID=64571	Feed_conversion_ratio	f.conv
chr26	1193077	4893906	QTL_ID=9453	Body_weight_(63_days)	bw
chr27	81131	4104720	QTL_ID=55906	Body_weight_(35_days)	bw
chr27	81131	4104720	QTL_ID=55932	Growth_(0-35_days)	grow
chr27	81131	4104720	QTL_ID=55944	Growth_(105-140_days)	grow
chr27	1210605	4104720	QTL_ID=2410	Body_weight	bw
chr27	1210605	4104720	QTL_ID=2409	Body_weight	bw
chr27	1969149	4104720	QTL_ID=17084	Body_weight_(140_days)	bw
chr27	2282360	4104720	QTL_ID=55926	Body_weight_(140_days)	bw
chr27	2467966	4104720	QTL_ID=55912	Body_weight_(70_days)	bw
chr27	2467966	4104720	QTL_ID=55936	Growth_(35-70_days)	grow
chr27	2595571	4104720	QTL_ID=55918	Body_weight_(105_days)	bw
chr27	3493381	4104720	QTL_ID=2408	Body_weight	bw
chr27	3493381	4104720	***QTL_ID=7159**	Body_weight_(35_days)	bw
chr28	1657793	4302316	QTL_ID=24893	Body_weight_(504_days)	bw
chr28	2073322	4302316	QTL_ID=55901	Body_weight_(hatch)	bw
chr28	2507636	4302316	QTL_ID=2419	Body_weight	bw

*bolded are QTL mapped in the same F_2 population used in this study

Appendix H – Functional annotation of 94 associated (P < 7.86E-07) and/or suggestive associated (P < 1.57E-05) SNPs from 444 F_2 individuals using Variant Effect Predictor (VEP) tool v.71 online [1]. The last column is the classification of the SNPs as overlapping QTLs (TRUE) or not overlapping QTLs (FALSE). These QTLs were the same presented in Table S3 online

Chr	Pos	Allele	#	Consequence	Ensemble gene ID	Gene Symbol	##	Overlapping QTLs
chr1	3706	C/G	1	intron_variant	ENSGALG00000009771			FALSE
chr1	25041791	C/T	2	intergenic_variant				TRUE
chr1	45468342	T/C	3	3_prime_UTR_variant	ENSGALG00000011406	*NTN4*		TRUE
chr1	53749693	G/A	4	upstream_gene_variant	ENSGALG00000012647	*RFX4*		TRUE
chr1	82575973	T/C	5	downstream_gene_variant	ENSGALG00000027337	*CLDND1*		TRUE
chr1	102944294	A/C	6	intergenic_variant				TRUE
chr1	178184151	G/A	7	intergenic_variant				TRUE
chr1	194726625	A/C	8	intron_variant	ENSGALG00000017320	*RAB6A*		TRUE
chr2	33436237	A/T	9	intergenic_variant			1	TRUE
chr2	33437336	C/G	10	intergenic_variant				TRUE
chr2	138815862	G/A	11	intergenic_variant				TRUE
chr3	77435334	C/T	12	intron_variant	ENSGALG00000015860	*UBE3D*		TRUE
chr4	3514658	G/T	13	intron_variant	ENSGALG00000029157	*MBNL3*		TRUE
chr4	68623163	A/C	14	intergenic_variant				TRUE
chr4	68882750	G/A	15	3_prime_UTR_variant	ENSGALG00000014320	*UGDH*		TRUE
chr4	68882765	T/G	16	downstream_gene_variant 3_prime_UTR_variant	ENSGALG00000014312 ENSGALG00000014320	*LIAS* *UGDH*	2	TRUE
chr4	69297525	G/A	17	downstream_gene_variant	ENSGALG00000014312	*LIAS*		TRUE
chr4	69370079	A/T	18	intron_variant	ENSGALG00000013521	*TBC1D1*		TRUE
chr4	69372005	T/C	19	intron_variant	ENSGALG00000013521	*TBC1D1*		TRUE
chr4	69372065	G/A	20	intron_variant	ENSGALG00000013521	*TBC1D1*		TRUE
chr4	73210325	C/T	21	intergenic_variant				TRUE
chr4	73407243	A/G	22	intergenic_variant			3	TRUE
chr4	73407249	G/A	23	intergenic_variant				TRUE
chr4	75842191	A/C	24	intron_variant	ENSGALG00000014485	*LDB2*		TRUE
chr4	79707389	C/G	25	intron_variant	ENSGALG00000028116			TRUE
chr5	54401932	C/T	26	intergenic_variant				TRUE
chr6	2997851	G/A	27	intron_variant	ENSGALG00000002327	*NRG3*		TRUE
chr6	5492179	A/C	28	intergenic_variant				TRUE
chr6	28967516	G/T	29	intergenic_variant				TRUE
chr6	31436207	T/C	30	intergenic_variant				TRUE
chr7	6904443	G/A	31	synonymous_variant				TRUE
chr7	22092999	G/C	32	intron_variant	ENSGALG00000025739	*RUFY4*	4	TRUE
chr7	22093009	C/T	33	intron_variant	ENSGALG00000025739	*RUFY4*		TRUE
chr7	26252164	C/T	34	intergenic_variant				TRUE
chr7	34668134	A/C	35	missense_variant	ENSGALG00000012484	*RIF1*	5	TRUE
chr7	34855716	G/A	36	upstream_gene_variant	ENSGALG00000012511	*CACNB4*		TRUE
chr8	5848411	T/C	37	intron_variant	ENSGALG00000003893	*XPR1*		FALSE
chr8	27651130	T/C	38	downstream_gene_variant upstream_gene_variant	ENSGALG00000027967 ENSGALG00000011238	*MIR6630* *WLS*		TRUE
chr9	15349595	G/A	39	downstream_gene_variant	ENSGALG00000006246			TRUE
chr9	21971358	C/T	40	intron_variant	ENSGALG00000009669	*RSRC1*		TRUE
chr10	9890328	A/G	41	intron_variant	ENSGALG00000005011	*SHC4*		TRUE
chr10	17086845	G/A	42	intron_variant	ENSGALG00000026468	*CHSY1*		TRUE
chr11	1249418	G/A	43	missense_variant	ENSGALG00000117983	*MUC5B*	6	TRUE
chr11	1997302	G/A	44	intron_variant	ENSGALG00000002853	*CFDP1*		TRUE
chr11	2149223	A/G	45	intron_variant upstream_gene_variant	ENSGALG00000003084 ENSGALG00000003071	*NUDT21* *OGFOD1*		TRUE
chr13	1958055	C/T	46	downstream_gene_variant	ENSGALG00000002447	*CTNNA1*		TRUE
chr13	9894951	C/T	47	downstream_gene_variant intron_variant	ENSGALG00000002457 ENSGALG00000003512	*SIL1* *SPINK7*	7	TRUE
chr13	9894952	C/T	48	intron_variant	ENSGALG00000003512	*SPINK7*		TRUE
chr13	12804655	T/C	49	intergenic_variant				TRUE
chr13	15109350	T/C	50	intron_variant	ENSGALG00000006424	*JADE2*	8	TRUE
chr13	15109354	C/T	51	intron_variant	ENSGALG00000006424	*JADE2*		TRUE
chr14	2918142	C/A	52	intron_variant	ENSGALG00000004224	*MAD1L1*		TRUE
chr14	4019669	G/T	53	intron_variant	ENSGALG00000004504	*RADIL*		TRUE
chr14	12043802	T/C	54	downstream_gene_variant 3_prime_UTR_variant	ENSGALG00000007456 ENSGALG00000028691	*CGTHBA* *MPG*	9	TRUE
chr14	12127914	G/A	55	downstream_gene_variant	ENSGALG00000007473	*MRPL28*		TRUE
chr14	12127934	C/T	56	downstream_gene_variant	ENSGALG00000007473	*MRPL28*		TRUE
chr15	6942668	G/A	57	intergenic_variant				TRUE
chr15	9763685	C/A	58	intergenic_variant				TRUE
chr17	1976385	C/G	59	intron_variant	ENSGALG00000008475	*ARRDC1*		FALSE
chr17	4917070	C/G	60	intron_variant	ENSGALG00000005036	*SH2D3C*	10	TRUE
chr17	4917071	G/T	61	intron_variant	ENSGALG00000005036	*SH2D3C*		TRUE

Chr	Pos	Allele	#	Consequence	Ensemble gene ID	Gene Symbol	##	Overlapping QTLs
chr17	4917072	T/G	62	intron_variant	ENSGALG00000005036	*SH2D3C*		TRUE
chr17	5621341	C/T	63	intergenic_variant			11	TRUE
chr17	5621374	C/T	64	intergenic_variant				TRUE
chr17	5986188	C/T	65	intergenic_variant				TRUE
chr18	4390068	C/T	66	intron_variant	ENSGALG00000001971	*UBE2O*		TRUE
				upstream_gene_variant	ENSGALG00000004429	*COG1*		
chr18	9253828	A/C	67	downstream_gene_variant	ENSGALG00000004418	*SS2R*	12	FALSE
				upstream_gene_variant	ENSGALG00000004429	*COG1*		
chr18	9551090	C/T	68	intron_variant	ENSGALG00000006895	*CEP131*		FALSE
chr18	9554600	C/T	69	intron_variant	ENSGALG00000006895	*CEP131*		FALSE
chr20	11847031	T/C	70	intron_variant	ENSGALG00000007636	*PEPCK*	13	FALSE
chr20	11847038	C/T	71	intron_variant	ENSGALG00000007636	*PEPCK*		FALSE
chr21	725260	G/A	72	intergenic_variant				TRUE
				downstream_gene_variant	ENSGALG00000026379	*OPRD1*		
chr23	2109713	G/A	73	downstream_gene_variant	ENSGALG00000021931	*SNORA73*		TRUE
				downstream_gene_variant	ENSGALG00000002871	*PHACTR4*		
				intron_variant	ENSGALG00000027784	*RCC1*		
chr23	4929597	A/G	74	intron_variant	ENSGALG00000026836	*COL16A1*		TRUE
chr25	601833	T/C	75	intron_variant	ENSGALG00000014559	*MEF2D*		FALSE
chr26	534090	C/T	76	intron_variant	ENSGALG00000000427	*KDM5B*	14	TRUE
chr26	1249195	A/G	77	upstream_gene_variant	ENSGALG00000000329	*AHCYL1*		TRUE
chr27	1587217	C/A	78	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587218	A/C	79	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587219	G/A	80	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587220	G/T	81	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587221	G/C	82	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587222	A/C	83	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*	15	TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587224	A/C	84	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587226	C/A	85	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587227	C/G	86	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	1587228	A/G	87	3_prime_UTR_variant	ENSGALG00000000201	*PLEKHM1*		TRUE
				upstream_gene_variant	ENSGALG00000025745	*ARHGAP27*		
chr27	3798382	C/G	88	intron_variant	ENSGALG00000027305	*SKAP1*	16	TRUE
chr27	3798470	G/A	89	intron_variant	ENSGALG00000027305	*SKAP1*		TRUE
chr27	4784671	G/A	90	intron_variant	ENSGALG00000003403	*ZNF385C*		FALSE
chr28	788219	C/T	91	downstream_gene_variant	ENSGALG00000027742	*KLHL33*		FALSE
chr28	1252731	G/A	92	intron_variant	ENSGALG00000026716	*CELF5*		FALSE
chr28	2986566	G/A		intergenic_variant				TRUE
chrZ	9064309	A/G	94	downstream_gene_variant	ENSGALG00000004022	*TLN1*	17	FALSE

marker number; ##block number

Appendix I – List of 40 QTLs mapped using microsatellite markers, from the same F_2 population used in this study. This information was filtered by the Chicken Animal QTLdb

Chr	start	end	QTL_ID	Trait	Abrev
chr1	51977644	53798691	***QTL_ID=3324**	Body_weight_(35_days)	BW35
chr1	53763599	53833783	QTL_ID=19671	Body_weight_(41_days)	BW41
chr1	53763599	53833783	QTL_ID=19672	Feed_intake	F.intake
chr1	58033892	58104076	QTL_ID=19675	Body_weight_(41_days)	BW41
chr1	58033892	58104076	QTL_ID=19676	Feed_intake	F.intake
chr1	83747978	90906992	QTL_ID=12464	Body_weight_(35_days)	BW35
chr1	156472083	162032735	QTL_ID=12469	Body_weight_(35_days)	BW35
chr2	15053040	19379651	QTL_ID=7175	Body_weight_(1_day)	BW1
chr2	67519043	70985270	QTL_ID=7170	Body_weight_(35_days)	BW35
chr2	79199747	101588000	QTL_ID=7155	Body_weight_(35_days)	BW35
chr2	79199747	101588000	QTL_ID=7160	Body_weight_(41_days)	BW41
chr2	79199747	101588000	QTL_ID=7173	Body_weight_(41_days)	BW41
chr2	79199747	101588000	QTL_ID=7179	Body_weight_(35_days)	BW35
chr2	79199747	101588000	QTL_ID=7183	Body_weight_(41_days)	BW41
chr3	6645961	24160710	QTL_ID=7184	Body_weight_(41_days)	BW41
chr3	6645961	106970113	***QTL_ID=7180**	Body_weight_(35_days)	BW35
chr3	12229218	12229258	QTL_ID=24377	Body_weight_(35_days)	BW35
chr3	12229218	12229258	QTL_ID=24378	Body_weight_(41_days)	BW41
chr3	12229218	12229258	QTL_ID=24379	Body_weight_(42_days)	BW42
chr3	24160710	35512024	QTL_ID=7167	Body_weight_(1_day)	BW1
chr3	24160710	35512024	QTL_ID=7171	Body_weight_(35_days)	BW35
chr3	24160710	35512024	QTL_ID=7174	Body_weight_(41_days)	BW41
chr3	35512024	40606300	QTL_ID=7156	Body_weight_(35_days)	BW35
chr3	35512024	40606300	QTL_ID=7161	Body_weight_(41_days)	BW41
chr4	69942858	84618310	***QTL_ID=7157**	Body_weight_(35_days)	BW35
chr4	69942858	84618310	***QTL_ID=7162**	Body_weight_(41_days)	BW41
chr4	69942858	84618310	***QTL_ID=7185**	Body_weight_(41_days)	BW41
chr6	29735045	31244082	QTL_ID=7176	Body_weight_(1_day)	BW1
chr7	34723350	36245040	***QTL_ID=7163**	Body_weight_(41_days)	BW41
chr9	19669473	23441680	***QTL_ID=7177**	Body_weight_(1_day)	BW1
chr10	16519830	19911089	***QTL_ID=7158**	Body_weight_(35_days)	BW35
chr10	16519830	19911089	***QTL_ID=7164**	Body_weight_(41_days)	BW41
chr12	11106924	12275026	QTL_ID=7168	Body_weight_(1_day)	BW1
chr18	378464	3180498	QTL_ID=7169	Body_weight_(1_day)	BW1
chr18	378464	3180498	QTL_ID=7172	Body_weight_(35_days)	BW35
chr18	378464	3180498	QTL_ID=7182	Body_weight_(35_days)	BW35
chr27	1627441	3493381	QTL_ID=7178	Body_weight_(1_day)	BW1
chr27	1627441	3493381	QTL_ID=7186	Body_weight_(41_days)	BW41
chr27	3493381	4104720	***QTL_ID=7159**	Body_weight_(35_days)	BW35
chr28	1674859	2912979	QTL_ID=7187	Body_weight_(41_days)	BW41

*bolded are QTLs overlapped by genome-wise associated and suggestively associated SNPs with the performance traits analyzed in this study

Appendix J

	Location	Gene	SYMBOL	edgeR,logFC	edgeR,logCPM	edgeR,p,value	edgeR,adj,p,value
1	1:1244701-1244701	ENSGALG00000008449	novelgene	4,448	3,344	0,005	4,69E-03
2	1:1245001-1245001			4,448	3,344	0,005	4,69E-03
3	1:44775901-44775901	ENSGALG00000011297	CRADD	-1,373	6,134	0,003	2,79E-03
4	1:81671101-81671101	-	-	2,877	4,270	0,002	1,77E-03
5	1:81671401-81671401			2,877	4,270	0,002	1,77E-03
6	1:143647201-143647201	-	-	2,179	4,363	0,002	2,32E-03
7	1:143647501-143647501			2,179	4,363	0,002	2,32E-03
8	1:193971001-193971001	ENSGALG00000009114	DGAT2	2,541	3,400	0,004	4,08E-03
9	1:193971301-193971301			2,541	3,400	0,004	4,08E-03
10	1:193971601-193971601			3,195	3,370	0,004	3,73E-03
11	2:223501-223501	ENSGALG00000013341	AGAP3	3,053	3,326	0,003	2,74E-03
12	2:112513801-112513801	ENSGALG00000015450	RAB2A	-3,924	4,005	0,003	3,37E-03
13	4:13297201-13297201	ENSGALG00000008006	CAPN6	2,767	3,510	0,003	3,14E-03
14	4:13429201-13429201	-	-	-1,567	4,940	0,003	2,67E-03
15	4:34189801-34189801	ENSGALG00000010291 ENSGALG00000010298	RBPMS DCTN6	2,286	4,055	0,003	3,36E-03
16	4:34190101-34190101			2,557	3,982	0,003	2,71E-03
17	4:44742601-44742601	ENSGALG00000010893	FGF5	-2,695	4,638	0,000	2,65E-04
18	5:402001-402001	ENSGALG00000013298	CPSF7	3,572	3,779	0,002	1,88E-03
		ENSGALG00000022369	TMEM216				
		ENSGALG00000025756	TMEM138				
19	5:5731501-5731501	-	-	-0,713	8,232	0,003	3,36E-03
20	5:5731801-5731801			-0,712	8,514	0,002	1,82E-03
21	5:43370101-43370101	ENSGALG00000010680	TTC7B	3,802	3,585	0,000	4,92E-04
22	5:43370401-43370401			3,802	3,585	0,000	4,92E-04
23	5:56258101-56258101	ENSGALG00000012203	SAMD4A	-1,651	5,633	0,001	1,43E-03
24	5:56258401-56258401			-1,651	5,633	0,001	1,43E-03
25	5:57753301-57753301	ENSGALG00000012321	MAP4K5	-5,357	4,037	0,001	5,13E-04
26	5:57753601-57753601			-5,385	4,055	0,000	4,59E-04
27	6:27634201-27634201	ENSGALG00000025403	MIR1815	3,831	3,561	0,003	2,92E-03
28	6:27634501-27634501			3,831	3,561	0,003	2,92E-03
29	7:10155301-10155301	-	-	3,461	3,442	0,002	2,22E-03
30	7:10155601-10155601			3,461	3,442	0,002	2,22E-03
31	7:34455901-34455901	-	-	3,254	3,374	0,001	1,40E-03
32	7:34456201-34456201			3,254	3,374	0,001	1,40E-03
33	7:34456501-34456501	ENSGALG00000012475	novelgene	3,494	3,455	0,001	1,38E-03
34	8:3895201-3895201	-	-	2,440	3,704	0,002	1,61E-03
35	8:19935901-19935901	ENSGALG00000010226	MUTYH	1,891	4,176	0,005	4,58E-03
		ENSGALG00000010228	TOE1				
		ENSGALG00000023348	HPDL				
36	8:27597601-27597601	ENSGALG00000026591	GNG12	-1,819	4,950	0,002	2,49E-03
		ENSGALG00000025977	GADD45A				

	Location	Gene	SYMBOL	edgeR,logFC	edgeR,logCPM	edgeR,p,value	edgeR,adj,p,value
37	8:27652201-27652201	ENSGALG00000011238	WLS	4,908	3,584	0,004	3,80E-03
		ENSGALG00000027802	gga-mir-6653				
38	8:28043701-28043701	-	-	-2,832	4,513	0,003	3,01E-03
39	10:2100301-2100301	ENSGALG00000001449	STRA6	-2,742	4,917	0,004	4,00E-03
		ENSGALG00000021525	ISLR				
		ENSGALG00000029151	novelgene				
40	10:12681001-12681001	-	-	4,800	3,523	0,004	3,92E-03
41	10:12681301-12681301	-	-	4,616	3,435	0,005	4,78E-03
42	10:19580401-19580401	ENSGALG00000008315 ENSGALG00000008336	UNC45A MAN2A2	1,612	5,653	0,000	2,52E-04
43	10:19580701-19580701			1,708	5,559	0,000	1,90E-04
44	10:19592401-19592401	ENSGALG00000008340 ENSGALG00000008341	FES FURIN	3,584	3,748	0,002	1,99E-03
45	10:19592701-19592701			3,584	3,748	0,002	1,99E-03
46	11:453601-453601	ENSGALG00000000904	USB1	3,478	3,400	0,004	4,02E-03
		ENSGALG00000000999	ZNF319				
		ENSGALG00000001011	CNGB1				
47	11:17139901-17139901	ENSGALG00000014284	GSE1	-5,359	4,016	0,003	3,01E-03
48	11:17140201-17140201			-5,359	4,016	0,003	3,01E-03
49	11:19161901-19161901	ENSGALG00000000802	DHODH	3,019	3,598	0,002	1,81E-03
		ENSGALG00000000811	IST1				
		ENSGALG00000000787	DHX38				
50	12:3190201-3190201	ENSGALG00000004639	GNAI2	2,362	4,114	0,005	4,95E-03
		ENSGALG00000013370	SEMA3F				
		ENSGALG00000028697	GNAT1				
51	12:5752201-5752201	-	-	-2,520	4,337	0,004	4,40E-03
52	12:5752501-5752501			-2,549	4,034	0,001	7,62E-04
53	12:11534101-11534101	-	-	-1,405	5,706	0,003	3,14E-03
54	12:11534401-11534401			-1,338	5,596	0,004	3,93E-03
55	13:15793201-15793201	ENSGALG00000006569	novelgene	2,434	4,351	0,005	4,97E-03
56	13:15793501-15793501			2,551	4,488	0,002	1,99E-03
57	15:2566201-2566201	ENSGALG00000002272	NOC4L	4,472	3,349	0,003	3,14E-03
		ENSGALG00000002336	RP11-2C24.9				
58	15:7104601-7104601	-	-	-3,205	4,148	0,004	3,67E-03
59	15:9882901-9882901	ENSGALG00000007396	TAOK3	2,858	3,697	0,001	9,83E-04
60	15:10068901-10068901	ENSGALG00000007720	AIFM3	1,460	4,850	0,004	4,38E-03
		ENSGALG00000028023	C14ORF166B				
		ENSGALG00000026902	P2RX6				

	Location	Gene	SYMBOL	edgeR,logFC	edgeR,logCPM	edgeR,p,value	edgeR,adj,p,value
61	15:10800301-10800301	ENSGALG00000007781	GAL3ST1	-1,368	5,310	0,004	4,14E-03
		ENSGALG00000007840	RP4-539M6.19				
62	16:70501-70501	ENSGALG00000000178 ENSGALG00000000181 ENSGALG00000026269	BF1 novelgene TAP1	2,832	4,523	0,001	8,84E-04
63	16:70801-70801			2,832	4,523	0,001	8,84E-04
64	16:219901-219901	ENSGALG00000019836 ENSGALG00000028962	ZNF692 novelgene	-5,336	4,053	0,003	2,77E-03
65	16:220201-220201			-5,393	4,088	0,002	2,42E-03
66	17:168901-168901	ENSGALG00000019837	KIFC1	1,344	4,824	0,005	4,74E-03
67	17:8289901-8289901	ENSGALG00000001595 ENSGALG00000001620	PHF19 CUTA	-3,805	5,124	0,001	8,25E-04
68	17:8290201-8290201			-3,677	3,935	0,001	1,25E-03
69	17:8550001-8550001	ENSGALG00000001419	DAB2IP	2,939	4,030	0,004	3,58E-03
70	17:8550301-8550301			2,939	4,030	0,004	3,58E-03
71	17:9845701-9845701	-	-	-5,019	3,840	0,002	1,58E-03
72	19:4471201-4471201	ENSGALG00000002186 ENSGALG00000002212	UNC45B RAD51D	2,321	3,861	0,002	2,13E-03
73	19:4471501-4471501			2,321	3,861	0,002	2,13E-03
74	19:4753501-4753501	ENSGALG00000002312	TMEM132E	1,650	4,849	0,003	2,50E-03
75	19:4753801-4753801			1,669	4,859	0,003	2,59E-03
76	19:8303401-8303401	-	-	4,901	3,525	0,001	9,14E-04
77	19:8303701-8303701			4,901	3,525	0,001	9,14E-04
78	20:2325901-2325901	-	-	-4,852	4,715	0,001	7,05E-04
79	20:2326201-2326201			-4,115	4,730	0,003	2,62E-03
80	20:8868601-8868601	-	-	3,123	3,889	0,000	4,05E-04
81	20:9458401-9458401	ENSGALG00000005932	MYT1	-3,928	4,105	0,005	4,53E-03
82	20:11843101-11843101	ENSGALG00000007636	PCK1	-2,591	4,055	0,001	1,37E-03
83	20:11843401-11843401			-2,453	3,975	0,002	2,27E-03
84	20:11844601-11844601			-2,149	4,166	0,003	2,82E-03
85	20:11844901-11844901			-2,149	4,166	0,003	2,82E-03
86	21:4398301-4398301	ENSGALG00000027085	CROCC	1,579	5,808	0,000	4,36E-04
87	22:1246501-1246501	ENSGALG00000000402	LOXL2	-1,682	5,492	0,002	1,87E-03
		ENSGALG00000000405	R3HCC1				
88	22:1246801-1246801			-1,724	5,320	0,003	2,83E-03

	Location	Gene	SYMBOL	edgeR,logFC	edgeR,logCPM	edgeR,p,value	edgeR,adj,p,value
89	24:2467801-2467801	ENSGALG00000001450	IGSF9B	3,291	3,378	0,004	3,57E-03
90	25:15301-15301	ENSGALG00000009011	SMG5	1,640	4,771	0,004	3,87E-03
91	25:15601-15601			1,640	4,771	0,004	3,87E-03
92	25:209701-209701	ENSGALG00000000443	novelgene	1,467	7,140	0,001	1,06E-03
93	25:210001-210001			1,954	6,799	0,000	3,49E-05
94	25:225301-225301	-	-	2,562	5,493	0,002	2,38E-03
95	25:225601-225601			2,420	5,586	0,002	1,78E-03
96	25:234901-234901	-	-	2,406	5,405	0,004	4,09E-03
97	25:235201-235201			2,394	5,426	0,004	3,86E-03
98	25:243301-243301	-	-	1,441	6,808	0,001	5,96E-04
99	25:660001-660001	ENSGALG00000027046	novelgene	3,421	3,810	0,005	4,91E-03
100	25:1675201-1675201	**ENSGALG00000028854**	**DUSP23**	1,213	5,974	0,004	4,41E-03
		ENSGALG00000022137	CRP				
101	25:1675501-1675501	ENSGALG00000027054 ENSGALG0000002879	novelgene CADM3	1,213	5,974	0,004	4,41E-03
102	26:648601-648601			-1,825	4,777	0,005	4,69E-03
		ENSGALG00000000362	NAV1				
		ENSGALG00000028854	**DUSP23**				
103	26:1633201-1633201	ENSGALG00000000583	SOX13	3,188	3,594	0,002	2,18E-03
104	26:1633501-1633501			3,188	3,594	0,002	2,18E-03
105	26:1707301-1707301	ENSGALG00000000587	PLEKHA6	1,772	4,710	0,002	1,77E-03
106	26:4467901-4467901	-	-	-3,344	3,754	0,004	4,25E-03
107	27:2586001-2586001	ENSGALG00000000478	TANC2	3,125	3,321	0,005	4,69E-03
108	27:4113601-4113601	ENSGALG00000025788	CACNB1	3,252	3,461	0,004	4,33E-03
109	28:2780701-2780701	ENSGALG00000026231	ARID3A	-5,524	4,143	0,001	1,14E-03
110	28:2781001-2781001			-5,524	4,143	0,001	1,14E-03
111	28:3127801-3127801	ENSGALG00000024298	ADAMTSL5	2,225	3,991	0,003	3,27E-03
		ENSGALG00000026384	PCSK4				
112	Z:1446301-1446301	-	-	3,788	3,612	0,003	2,86E-03
123	Z:1446601-1446601			3,707	3,576	0,004	3,74E-03
114	Z:56827201-56827201	ENSGALG00000014684	ERAP1	1,661	4,443	0,003	2,79E-03
115	Z:56827501-56827501			1,661	4,443	0,003	2,79E-03

Appendix K

Merged windows with differential methylation		CpGs within the merged window	Within or nearby gene		Position regarding gene
			ENSEMBL name	Gene symbol	
1	chr1:1244701-1245300	5	ENSGALG00000008449	Novel gene (1)	coding sequence ; intronic ; regulatory region
2	chr1:44775901-44776200	12	ENSGALG00000011297	CRADD	intronic ; regulatory region
3	chr1:81671101-81671700	10	--	Intergenic region (1)	intergenic
4	chr1:143647201-	29	--	Intergenic region (2)	intergenic
5	chr1:193971001-	36	ENSGALG00000009114	DGAT2	upstream gene
6	chr2:223501-223800	19	ENSGALG00000013341	AGAP3	intronic ; regulatory region
7	chr2:112513801-	6	ENSGALG00000015450	RAB2A	upstream gene
8	chr4:13297201-13297500	4	ENSGALG00000008006	CAPN6	intronic ; regulatory region
9	chr4:13429201-13429500	11	--	Intergenic region (3)	intergenic
			ENSGALG00000010298	DCTN6	downstream gene
10	chr4:34189801-34190400	20	ENSGALG00000010291	RBPMS	coding sequence ; intronic ; regulatory region
11	chr4:44742601-44742900	14	ENSGALG00000010893	FGF5	coding sequence ; intronic ; regulatory region
12	chr5:402001-402300	10	ENSGALG00000013298	CPSF7	downstream gene
			ENSGALG00000022369	TMEM216	upstream gene
			ENSGALG00000025756	TMEM138	coding sequence ; intronic ; regulatory region
			ENSGALG00000028882	Novel gene (2)	upstream gene
13	chr5:5731501-5732100	29	--	Intergenic region (4)	intergenic
14	chr5:43370101-43370700	10	ENSGALG00000010680	TTC7B	intronic ; regulatory region
15	chr5:56258101-56258700	34	ENSGALG00000012203	SAMD4A	intronic ; regulatory region
16	chr5:57753301-57753900	6	ENSGALG00000012321	MAP4K5	upstream gene
17	chr6:27634201-27634800	3	ENSGALG00000025403	MIR1815	upstream gene
18	chr7:10155301-10155900	13	--	Intergenic region (5)	intergenic
19	chr7:34455901-34456800	20	--	Intergenic region (6)	intergenic
			ENSGALG00000012475	Novel gene (3)	downstream gene
20	chr8:3895201-3895500	7	--	Intergenic region (7)	intergenic
21	chr8:19935901-19936200	3	ENSGALG00000010226	MUTYH	coding sequence ; intronic ; regulatory region
			ENSGALG00000010228	TOE1	upstream gene
			ENSGALG00000023348	HPDL	downstream gene
22	chr8:27597601-27597900	8	ENSGALG00000026591	GNG12	coding sequence ; intronic ; regulatory region
			ENSGALG00000025977	GADD45A	downstream gene
23	chr8:27652201-27652500	9	ENSGALG00000011238	WLS	upstream gene
			ENSGALG00000027802	gga-mir-6653	downstream gene
24	chr8:28043701-28044000	3	--	Intergenic region (8)	intergenic
25	chr10:2100301-2100600	27	ENSGALG00000001449	STRA6	downstream gene
			ENSGALG00000021525	ISLR	coding sequence ; regulatory region
			ENSGALG00000029151	Novel gene (4)	downstream gene
26	chr10:12681001-12681600	5	--	Intergenic region (9)	intergenic
27	chr10:19580401-19581000	43	ENSGALG00000008315	UNC45A	upstream gene
			ENSGALG00000008336	MAN2A2	coding sequence ; intronic ; regulatory region
28	chr10:19592401-19593000	21	ENSGALG00000008340	FES	intronic ; regulatory region
			ENSGALG00000008341	FURIN	downstream gene
29	chr11:453601-453900	29	ENSGALG00000000904	USB1	upstream gene
			ENSGALG00000000999	ZNF319	coding sequence ; regulatory region
			ENSGALG00000001011	CNGB1	upstream gene
			ENSGALG00000000904	USB1	upstream gene
30	chr11:17139901-17140500	26	ENSGALG00000014284	GSE1	intronic ; regulatory region
31	chr11:19161901-19162200	8	ENSGALG00000000802	DHODH	downstream gene
			ENSGALG00000000811	IST1	downstream gene
			ENSGALG00000000787	DHX38	upstream gene
32	chr12:3190201-3190500	11	ENSGALG00000004639	GNAI2	downstream gene
			ENSGALG00000013370	SEMA3F	downstream gene
			ENSGALG00000028697	GNAT1	coding sequence ; intronic ; regulatory region
33	chr12:5752201-5752800	12	--	Intergenic region (10)	intergenic
34	chr12:11534101-11534700	23	--	Intergenic region (11)	intergenic
35	chr13:15793201-15793800	9	ENSGALG00000006569	Novel gene (5)	downstream gene
36	chr15:2566201-2566500	6	ENSGALG00000002272	NOC4L	coding sequence ; intronic ; regulatory region
			ENSGALG00000002336	RP11-2C24.9	downstream gene
37	chr15:7104601-7104900	5	--	Intergenic region (12)	intergenic

Merged windows with differential methylation		CpGs within the merged window	Within or nearby gene		Position regarding gene
38	chr15:9882901-9883200	6	ENSGALG00000007396	TAOK3	coding sequence ; intronic ; regulatory region
39	chr15:10068901-10069200	24	ENSGALG00000007720	AIFM3	downstream gene
			ENSGALG00000028023	C14ORF166B	start lost; coding sequence ; 5 prime UTR
			ENSGALG00000026902	P2RX6	coding sequence ; intronic ; regulatory region
40	chr15:10800301-10800600	6	ENSGALG00000007781	GAL3ST1	downstream gene
			ENSGALG00000007840	RP4-539M6.19	downstream gene
41	chr16:70501-71100	13	ENSGALG00000000178	BF1	downstream gene
			ENSGALG00000000181	Novel gene (6)	upstream gene
			ENSGALG00000026269	TAP1	coding sequence ; intronic ; regulatory region
42	chr16:219901-220500	11	ENSGALG00000019836	ZNF692	coding sequence ; intronic ; regulatory region
			ENSGALG00000028962	Novel gene (7)	upstream gene
43	chr17:168901-169200	1	--	Intergenic region (13)	intergenic
			ENSGALG00000019837	KIFC1	downstream gene
			ENSGALG00000028962	Novel gene (7)	upstream gene
44	chr17:8289901-8290500	28	ENSGALG00000001595	PHF19	coding sequence ; intronic ; regulatory region
			ENSGALG00000001620	CUTA	downstream gene
45	chr17:8550001-8550600	24	ENSGALG00000001419	DAB2IP	intronic ; regulatory region
46	chr17:9845701-9846000	10	--	Intergenic region (14)	intergenic
47	chr19:4471201-4471800	12	ENSGALG00000002186	UNC45B	coding sequence ; intronic ; regulatory region
			ENSGALG00000002212	RAD51D	upstream gene
48	chr19:4753501-4754100	22	ENSGALG00000002312	TMEM132E	coding sequence ; intronic ; regulatory region
49	chr19:8303401-8304000	11	--	Intergenic region (15)	intergenic
50	chr20:2325901-2326500	13	--	Intergenic region (16)	intergenic
51	chr20:8868601-8868900	37	--	Intergenic region (17)	intergenic
52	chr20:9458401-9458700	2	ENSGALG00000005932	MYT1	coding sequence ; intronic ; regulatory region
53	chr20:11843101-11843700	31	ENSGALG00000007636	PCK1	intronic ; regulatory region
54	chr20:11844601-11845200	17	ENSGALG00000007636	PCK1	intronic ; regulatory region
55	chr21:4398301-4398600	12	ENSGALG00000027085	CROCC	coding sequence ; intronic ; regulatory region
56	chr22:1246501-1247100	19	ENSGALG00000000402	LOXL2	downstream gene
			ENSGALG00000000405	R3HCC1	coding sequence ; intronic ; regulatory region
57	chr24:2467801-2468100	10	ENSGALG00000001450	IGSF9B	upstream gene
58	chr25:15301-15900	37	ENSGALG00000009011	SMG5	upstream gene
59	chr25:209701-210300	34	ENSGALG00000000443	Novel gene (8)	intronic ; regulatory region
60	chr25:225301-225900	21	--	Intergenic region (18)	intergenic
61	chr25:234901-235500	20	--	Intergenic region (19)	intergenic
62	chr25:243301-243600	12	--	Intergenic region (20)	intergenic
63	chr25:660001-660300	17	ENSGALG00000027046	Novel gene (9)	coding sequence ; intronic ; regulatory region
64	chr25:1675201-1675800	54	ENSGALG00000022137	CRP	coding sequence ; regulatory region
			ENSGALG00000028854	DUSP23	upstream gene
65	chr26:648601-648900	6	ENSGALG00000000362	NAV1	coding sequence ; intronic ; regulatory region
			ENSGALG00000028854	DUSP23	upstream gene
			ENSGALG00000027054	Novel gene (10)	downstream gene
			ENSGALG00000028795	CADM3	downstream gene
66	chr26:1633201-1633800	12	ENSGALG00000000583	SOX13	intronic ; regulatory region
67	chr26:1707301-1707600	5	ENSGALG00000000587	PLEKHA6	intronic ; regulatory region
68	chr26:4467901-4468200	8	--	Intergenic region (21)	intergenic
69	chr27:2586001-2586300	5	ENSGALG00000000478	TANC2	intronic ; regulatory region
70	chr27:4113601-4113900	12	ENSGALG00000025788	CACNB1	coding sequence ; intronic ; regulatory region
71	chr28:2780701-2781300	25	ENSGALG00000026231	ARID3A	intronic ; regulatory region
72	chr28:3127801-3128100	8	ENSGALG00000024298	ADAMTSL5	downstream gene
			ENSGALG00000026384	PCSK4	downstream gene
73	chrZ:1446301-1446900	16	--	Intergenic region (22)	intergenic
74	chrZ:56827201-56827800	76	ENSGALG00000014684	ERAP1	upstream gene

Appendix L - Results for Consensus PathDB network analyses

1) Pathway over-representation analysis
47 genes (68.1%) from the input list are present in at least one pathway

pathway name	pathway source	set size	candidates contained	p-value	q-value
G-protein activation	Reactome	28	3 (10.7%)	0,000211	0,0323
Activation of the phototransduction cascade	Reactome	11	2 (18.2%)	0,000926	0,0708
cell cycle: g2/m checkpoint	BioCarta	21	2 (9.5%)	0,00344	0,0899
ADP signalling through P2Y purinoceptor 12	Reactome	22	2 (9.1%)	0,00378	0,0899
Adrenaline,noradrenaline inhibits insulin secretion	Reactome	23	2 (8.7%)	0,00413	0,0899
MAPK signaling pathway - Homo sapiens (human)	KEGG	257	5 (2.0%)	0,00414	0,0899
Visual signal transduction: Rods	PID	24	2 (8.3%)	0,00449	0,0899
Opioid Signalling	Reactome	84	3 (3.6%)	0,00504	0,0899
p38 MAPK signaling pathway	PID	29	2 (6.9%)	0,00652	0,0899
Phototransduction - Homo sapiens (human)	KEGG	29	2 (6.9%)	0,00652	0,0899
Signal amplification	Reactome	31	2 (6.5%)	0,00742	0,0899
Visual phototransduction	Reactome	96	3 (3.1%)	0,00754	0,0899
Inactivation, recovery and regulation of the phototransduction cascade	Reactome	32	2 (6.2%)	0,0079	0,0899
The phototransduction cascade	Reactome	33	2 (6.1%)	0,00838	0,0899
mRNA 3,-end processing	Reactome	35	2 (5.7%)	0,0094	0,0899
Post-Elongation Processing of Intron-Containing pre-mRNA	Reactome	35	2 (5.7%)	0,0094	0,0899

2) Enriched gene ontology (GO) terms
67 genes (97.1%) from the input list are present in at least one GO category

gene ontology term	category, level	set size	candidates contained	p-value	q-value
GO:0019885 antigen processing and presentation of endogenous peptide antigen via MHC class I	BP 5	11	2 (18.2%)	0,000683	0,0936
GO:0001578 microtubule bundle formation	BP 5	67	3 (4.5%)	0,0018	0,123
GO:0019001 guanyl nucleotide binding	MF 5	394	6 (1.5%)	0,00276	0,0412
GO:0032550 purine ribonucleoside binding	MF 5	1828	14 (0.8%)	0,00458	0,0412
GO:0051297 centrosome organization	BP 5	97	3 (3.1%)	0,00512	0,234
GO:0007602 phototransduction	BP 5	115	3 (2.6%)	0,00819	0,26
GO:0042461 photoreceptor cell development	BP 5	41	2 (4.9%)	0,0095	0,26