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

CNV-based genome-wide association studies for performance traits in broilers

Anna Carolina Fernandes

Dissertation presented to obtain the degree of Master in Science. Area: Animal Science and Pastures

Piracicaba
2021

Anna Carolina Fernandes
Animal Scientist

CNV-based genome-wide association studies for performance traits in broilers

versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

Dissertation presented to obtain the degree of Master in Science. Area: Animal Science and Pastures

Piracicaba
2021

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

Fernandes, Anna Carolina

CNV-based genome-wide association studies for performance traits in broilers / Anna Carolina Fernandes. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

49 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. GWAS 2. Desempenho 3. CNVs 4. QTLs 5. qPCR. I. Título

DEDICATION

This dissertation is dedicated to my mom and dad. For your endless love, support and encouragement. You have been my drive, rock and inspiration throughout life.

ACKNOWLEDGMENTS

First and foremost, thank you, God, for always being there for me.

To my parents, Anna and Douglas, for your truly selfless and unconditional love. Thank you both for always giving me strength, for reminding me every day that I can do this and for being a constant source of support in my many many moments of crisis, during the challenges of graduate school and life. You are my world and my everything this life can bring. I could not have done it without you. My little brother André, aunt Kátia and uncle Luciano deserve my wholehearted thanks and appreciation as well.

To my beloved grandpa Pasquale and great-aunts Lola and Beta. You are loved beyond words and missed beyond measures. I will always have this piece of my heart that smiles whenever I think about you. You were a priceless gift in my life and I will be forever grateful for the love, knowledge and values you instilled in me.

I would like to sincerely express my gratitude and appreciation to my academic advisor, Dr. Luiz Coutinho, for his guidance, help and support, and especially for his confidence in me. You have set an example of excellence as a researcher and mentor. I would also like to thank my LBA team, in particular Dr. Gabriel and Dr. Vinicius, for guiding me throughout this study; your discussion, ideas, and feedback have been absolutely invaluable.

To all my friends, in particular Pedro and Thais. You are the people who make me smile brighter, laugh louder and live better. True friendship isn't about being inseparable. It's being separated and nothing changes. Your friendship makes my life a better experience!

Finally, I wish to express my deepest gratitude to University of São Paulo – "Luiz de Queiroz" College of Agriculture (USP/ESALQ) for the privilege granted and for aggregating so much knowledge in my academic history. You challenged me to be fearless in the pursuit of all opportunities and taught me that learning knows no boundaries. A special acknowledgment to CNPq for supporting me with a scholarship (process 134211/2019-7).

Summary

RESUMO

Associação genômica ampla de CNVs com características de desempenho em frangos

A galinha (*Gallus gallus*) é uma importante fonte de proteína animal e é considerada um modelo biológico de pesquisa principalmente na área da genética. É notável que a produção de carne de frango tem extensa relevância na economia brasileira e mundial, consequência sobretudo da constante melhoria das características de desempenho. Neste sentido, a seleção de animais que apresentam tais características fenotípicas de interesse é fundamental. Nesta perspectiva, a identificação de variações genéticas e sua associação com características de importância zootécnica se apresentam como etapas cruciais para melhor compreensão dos mecanismos biológicos que controlam estas características complexas. Uma importante fonte de variação conhecida no DNA são as variações no número de cópias (CNVs), que podem contribuir significativamente com a variação fenotípica em diversas espécies. Neste contexto, foram utilizados os dados genotípicos de aproximadamente 1.500 animais de uma linha paterna de frangos de corte (TT), obtidos utilizando um chip de SNPs de alta densidade (600k, Affymetrix) para identificar regiões genômicas e genes candidatos associados à características de desempenho. Realizamos análises de associação genômica ampla (GWAS) baseadas em CNV usando o programa CNVRanger, ajustando um modelo linear misto, para identificar regiões no genoma associadas à: peso ao nascimento, peso aos 21 dias, peso aos 35 dias, peso aos 41 dias, peso aos 42 dias, consumo de ração, conversão alimentar e ganho de peso. Segmentos CNV significativamente associados ao peso ao nascimento, peso aos 35, 41 e 42 dias e ganho de peso foram identificados. Após as análises de associação, foi feita a validação destes segmentos CNV signifcativos associados por meio da técnica de qPCR (PCR quantitativa). A busca por genes candidatos foi feita dentro de cada região genômica associada, considerando termos de *Gene Ontology (GO)* e também a informação da literatura. Foram identificadas novas regiões no genoma associadas a estas características e importantes genes candidatos para crescimento e desenvolvimento muscular, tais como *KCNJ11*, *MyoD1* e *SOX6*, fornecendo novas informações para melhor compreeensão acerca da regulação do controle genético para desempenho em frangos.

Palavras-chave: GWAS, Desempenho, CNVs, QTLs, qPCR

ABSTRACT

CNV-based genome-wide association studies with performance traits in broilers

Chicken (*Gallus gallus*) is an important source of animal protein, considered a biological research model in the genetics field and a species of considerable economic relevance worldwide, mainly as a consequence of constant improvement of performance traits. In this sense, selection of animals that present phenotypic traits of interest is fundamental. Therefore, the identification of genetic variations and their association with production traits of economic importance are crucial steps for a better understanding of the biological mechanisms that control these complex characteristics. An important source of known variation in the DNA are copy number variations (CNVs), which can contribute significantly to phenotypic variation in several species. In this context, genotypic data from approximately 1,500 animals from a paternal broiler line (TT), obtained using a high-density SNP array (600k, Affymetrix), were used to identify genomic regions and candidate genes associated with performance traits. We performed a CNV-based genome-wide association study (GWAS) using the CNVRanger software, adjusting a linear mixed model, to identify regions in the genome associated with birth weight, body weight at 21 days, body weight at 35 days, body weight at 41 days, body weight at 42 days, feed intake, feed conversion ratio and body weight gain. CNV segments significantly associated with birth weight, body weight at 35, 41 and 42 days and body weight gain were identified. After the association analyses, validation of these significantly associated CNV segments was performed by qPCR. The search for candidate genes was made within each associated genomic region, considering Gene Ontology (GO) terms and also the literature information. We identified novel genomic regions associated with these traits and important candidate genes for muscle growth and development, such as *KCNJ11*, *MyoD1* and *SOX6*, with known role on chicken growth and muscle development, providing new information for a better understanding of the regulation of genetic control for performance in broilers.

Keywords: GWAS, Performance, CNVs, QTLs, qPCR

1. INTRODUCTION

Chicken is an excellent biological model organism (Wolpert, 2004; Ellegren, 2005; Stern, 2005) and a species of considerable economic relevance worldwide. Over the past 50 years, a growing demand for chicken meat in the world has pressured chicken breeders to enhance growth rate, feed efficiency, size of breast muscle and reduce abdominal fat, aiming to produce a carcass with maximal meat-yield and a low-fat content (Petracci & Cavani, 2012). Therefore, poultry breeding programs have been directed to improve growth and muscle deposition (Ismail & Joo, 2017), focusing on commercial broiler breeds with superior traits and higher growth and performance capabilities (Scheuermann et al. 2003; Havenstein et al. 2003).

In comparison to 50 years ago, broilers are ready to market in half the time and twice the body weight (Tallentire et al., 2016), mainly due to genetic improvement and high heritability of body weight and body composition (Le Bihan-Duval et al., 2003). Understanding genetic architecture and regulation of traits of economic interest may contribute to development of new genomic strategies for improving breeding programs.

In Brazil, in 1999 a collaboration was established between EMBRAPA Swine and Poultry National Research Center and ESALQ/USP to conduct research on chicken genomics, in order to investigate quantitative trait loci (QTLs) and identify genes associated with traits of economic interest. An F2 population obtained from reciprocal crosses between a paternal broiler line (TT) and a maternal layer line (CC) was used for research. The TT broiler line has been under multiple trait selection since 1992 to improve body weight, feed conversion, carcass and cuts yields, viability, fertility, eclodibility, reduction of abdominal fat and metabolic syndromes (Marchesi et al., 2018). In the F2 population, several QTLs were mapped and genes associated with performance, carcass and abdominal fat were identified (Nones et al., 2006; Ambo et al., 2009; Campos et al., 2009; Baron et al., 2010; Boschiero et al., 2013, Ragognetti et al., 2015).

In 2008, from an expansion of the TT broiler line, the TT Reference Population was generated to validate results obtained in the F2 population and for advanced genomic studies. In this population, QTLs were mapped for abdominal fat and skin traits on GGA5, 9, 10, 13, 15, 20, 24, 26 and 27 (Moreira et al., 2018). Complementary, other studies have been developed to identify QTLs and single nucleotide polymorphisms (SNPs) associated with performance, carcass and abdominal fat in chickens (de Koning et al. 2004; Navarro et al., 2005; Zhou et al., 2007; Nassar et al., 2012; Felício et al., 2013; Mignon-Grasteau et al., 2015; Wang et al., 2016).

The genomic variation can explain a considerable fraction of the variation observed at phenotypic level, mainly for traits with heritability estimates. In the last two decades, analysis of genomic variation focused on SNPs (Syvänen, 2001) however, it has been increasingly observed that SNPs are not the only source of variation in the genome. A major source of structural variation are the copy number variations (CNVs), which are large regions of the genome, differing in number of copies due to duplication or deletion events (Redon et al., 2006; Mccarroll et al., 2007). CNVs can induce phenotypic variation due to their effect on gene expression, as a consequence of gene disruption, dosage alteration or positional effects (Beckmann et al., 2007, da Silva et al., 2016). Known phenotypes associated with CNVs in chicken include late feathering (Elferink et al., 2008), pea-comb (Wright et al., 2009), dermal hyperpigmentation (Dorshorst et al., 2010), dark brown plumage color (Gunnarsson et al., 2011) and resistance/susceptibility to Marek's disease (Luo et al., 2013; Yan et al., 2015; Xu et al., 2017).

Most CNVs present in the offspring are inherited from their parents, following a Mendelian inheritance (Wang et al., 2008). However, a fraction of them is the result of *de novo* events, representing CNVs that were not

inherited from any of the parents (Wang et al., 2008). *De novo* mutations take place due to mitotic/meiotic recombination or chromosome rearrangements, induced by germinative or somatic cells (Wang et al., 2008).

CNVs can be experimentally detected from comparative genomic hybridization (CGH) assays (Pinkel et al., 2005), new generation sequencing (Zhao et al., 2013) and SNP arrays (Winchester et al., 2009; Xu et al., 2013). A frequently used algorithm to infer CNVs from SNP-chip data is the PennCNV software (Wang et al., 2007).

In this study, the use of CNVs to conduct GWAS for performance-related traits in a broiler population allowed the identification of novel genomic regions associated with these traits and potential candidate genes for growth and development. This dissertation generated results that can contribute to a better understanding of the genetic variants underlying performance regulation in chickens and may be useful for future chicken studies and poultry breeding programs aiming to increase chicken growth .

1.1. GENERAL OBJECTIVES

Identify CNVs and positional candidate genes associated with traits related to performance in a Brazilian broiler population using a CNV-based genome-wide association study (GWAS) approach.

1.2. SPECIFIC OBJECTIVES

(a) Identify CNVs in the whole chicken genome of an experimental broiler population (i.e., TT Reference Population).

(b) Concatenate CNVs into CNV regions (CNVRs).

(c) Identify CNV segments associated with performance traits in broilers using a CNV-based GWAS approach.

(d) Validate significantly associated CNV segments using qPCR.

(e) Annotate QTLs, positional candidate genes, enriched biological processes and metabolic pathways related to performance in significantly associated regions.

1.3. SYSTEMATIZATION TEXT

This project allowed the identification of structural genetic variations associated with performance traits in chickens. CNVs were inferred in the whole genome using high density SNP-chip (Affymetrix®) data from a Brazilian broiler population. We identified 23,214 unique autosomal CNVs, merged into 5,042 distinct CNV regions (CNVRs), covering 12.84% of the chicken autosomal genome. CNV segments were significantly associated with birth weight, body weight at 35, 41 and 42 days of age and body weight gain from 35-41 days of age. These associated CNV segments overlap and/or are in proximity of genes, such as *KCNJ11*, *MyoD1* and SOX6, with known role on chicken growth and muscle development.

Genome-wide association study is a common method to link phenotypic variation to genomic variation. The majority of the GWAS studies have focused only on SNPs. However, the effects of complex structural variants,

such as CNVs, on economically relevant phenotypes could improve our understanding on the genetics underlying their variation.

Moreover, an integration of the results obtained in this project with results obtained in previous studies will help to unravel molecular mechanisms involved in the regulation of performance, contributing to a deeper comprehension of bird physiology. Therefore, this information may help detect metabolic and/or physiological changes in birds that could be associated with the presence of complex variants in the genome.

In summary, the results presented here provide relevant information for future chicken studies and breeding programs, allowing the increase of selection accuracy and, consequently, genetic gain for traits of economic interest. We hope that this study will contribute to poultry genetic improvement programs, to the poultry industry, and also to a better understanding of the genetic variants underlying muscle physiology and body growth.

References

Ambo, M. et al. 2009. Quantitative trait loci for performance traits in a broiler x layer cross. Animal Genetics 40:200–208.

Baron, E.E. et al. 2010. QTL for percentage of carcass and carcass parts in a broiler x layer cross. Animal Genetics 42:117–124.

Beckmann, J.S.; Estivill, X.; Antonarakis, S.E. 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nature Reviews Genetics 8:639–646.

Boschiero, C. et al. 2013. Association of IGF1 and KDM5A polymorphisms with performance, fatness and carcass traits in chickens. Journal of Applied Genetics 54:103–112.

Campos, R.L.R. et al. 2009. Quantitative trait loci associated with fatness in a broiler x layer cross. Animal Genetics 40:729–736.

da Silva, V.H.; Regitano, L.; Geistlinger, L.; Pértille, F.; Giachetto, P.F.; Brassaloti, R.A. et al. 2016. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. PLoS ONE 11(6):e0157711.

de Koning, D.J. et al. 2004. Segregation of QTL for production traits in commercial meat-type chickens. Genetics Research 83:211–220.

Dorshorst, B.; Okimoto, R.; Ashwell, C. 2010. Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. Journal of Heredity 101:339–350.

Elferink, M.G.; Vallee, A.A.; Jungerius, A.P.; Crooijmans, R.P.; Groenen, M.A. 2008. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. BMC Genomics 9:391.

Ellegren, H. 2005. The avian genome uncovered. Trends in Ecology and Evolution 20(4):180–186.

Felício, A.M. et al. 2013. Identification and association of polymorphisms in CAPN1 and CAPN3 candidate genes related to performance and meat quality traits in chickens. Genetics and Molecular Research 12:472–482.

Fornari, M.B.; Zanella, R.; Ibelli, A.M.G.; Fernandes, L.T.; Cantão, M.E.; Thomaz-Soccol, V. et al. 2014. Unraveling the associations of osteoprotegerin gene with production traits in a paternal broiler line. Springplus 3:682.

Grupioni, N.V.; Stafuzza, N.B.; Carvajal, A.B.; Ibelli, A.M.G.; Peixoto, J.O.; Ledur, M.C.; Munari, D.P. 2017. Association of RUNX2 and TNFSF11 genes with production traits in a paternal broiler line. Genetics and Molecular Research 16(1): gmr16019443.

Gunnarsson, U. et al. 2011. The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. Pigment Cell Melanoma Research 24:268–274.

Havenstein, G.B.; Ferket, P.R.; Qureshi, M.A. 2003. Carcass composition and yield of 1957 versus 2001 broilers when fed representative 1957 and 2001 broiler diets. Poultry Science 82:1509–1518.

Ismail, I. & Joo, S.T. 2017. Poultry meat quality in relation to muscle growth and muscle fiber characteristics. Korean Journal for Food Science of Animal Resources 37(6):873-883.

Le Bihan-Duval, E.; Berri, C.; Baeza, E.; Sante, V.; Astruc, T.; Rémignon, H.; Le Pottier, G.; Bentley, J.; Beaumont, C.; Fernandez, X. 2003. Genetic parameters of meat technological quality traits in a grand-parental commercial line of turkey. Genetics Selection Evolution 35:623–635.

Luo, J.; Yu, Y.; Mitra, A.; Chang, S.; Zhang, H.; Liu, G. et al. 2013. Genome-Wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. G3 Genes\Genomes\Genetics 3(2):217–223.

Mccarroll, S.A.; Altshuler, D.M. 2007. Copy-number variation and association studies of human disease. Nature Genetics 39:37-42.

Mignon-Grasteau, S. et al. 2015. Detection of QTL controlling feed efficiency and excretion in chickens fed a wheat-based diet. Genetics Selection Evolution 47:74.

Moreira, G.C.M. et al. 2018. A genome-wide association study reveals novel genomic regions and positional candidate genes for fat deposition in broiler chickens. BMC Genomics 19:374.

Nassar, M.K.; Goraga, Z.S.; Brockmann, G.A. 2012. Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: II. Muscle weight and carcass composition. Animal Genetics 43:739–745.

Navarro, P. et al. 2005. Mapping of quantitative trait loci affecting organ weights and blood variables in a broiler layer cross. Brazilian Journal of Poultry Science 46:430–442.

Nones, K.; Ledur, M.C.; Ruy, D.C.; Baron, E.E.; Melo, C.M.R.; Moura, A.S.A.M.T. et al. 2006. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. Animal Genetics 37:95–100.

Petracci, M. & Cavani, C. 2012. Muscle growth and poultry meat quality issues. Nutrients 4(1):1-12.

Pinkel, D.; Albertson, D. 2005. Comparative genomic hybridization. Annual Review of Genomics and Human Genetics 6: 331–354.

Ragognetti, B.N.N.; Stafuzza, N.B.; Silva, T.B.R.; Chud, T.C.S.; Grupioni, N.V.; Cruz, V.A.R.; Peixoto, J.O.; Nones, K.; Ledur, M.C.; Munari, D.P. 2015. Genetic parameters and mapping quantitative trait loci associated with tibia traits in broilers. Genetics and Molecular Research 14(4):17544-54.

Redon, R.; Ishikawa, S.; Fitch, K.R.; Feuk, L.; Perry, G.H.; Andrews, T.D.; Fiegler, H.; Shapero, M.H.; Carson, A.R. et al. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Scheuermann, G.N.; Bilgili, S.F.; Hess, J.B.; Mulvaney, D.R. 2003. Breast muscle development in commercial broiler chickens. Poultry Science 82:1648–1658.

Stern, C.D. 2005. The chick: A great model system becomes even greater. Developmental Cell 8(1):9–17.

Syvänen, A. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2(12):930–42.

Tallentire, C.W.; Leinonen, I. & Kyriazakis, I. 2016. Breeding for efficiency in the broiler chicken: A review. Agronomy for Sustainable Development 36:66.

Wang, K.; Li, M.; Hadley, D. et al. 2007. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research 17(11):1665-1674.

Wang, K.; Chen, Z.; Tadesse, M.G.; Glessner, J.; Grant, S.F.A.; Hakonarson, H.; Bucan, M.; Li, M. 2008. Modeling genetic inheritance of copy number variations. Nucleic Acids Research 36(21):e138.

Wang, W. et al. 2016. Genome-wide association study of 8 carcass traits in Jinghai Yellow chickens using specific-locus amplified fragment sequencing technology. Poultry Science 95:500–506.

Winchester, L.; Yau, C.; Ragoussis, J. 2009. Comparing CNV detection methods for SNP arrays. Briefings in Functional Genomics and Proteomics 8:353–366.

Wright, D. et al. 2009. Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. PLoS Genetics 5:e1000512.

Wolpert, L. 2004. Much more from the chicken's egg than breakfast - A wonderful model system. Mechanisms of Development 121(9):1015–1017.

Xu, L.; Hou, Y.; Bickhart, D.; Song, J.; Liu, G. 2013. Comparative analysis of CNV calling algorithms: literature survey and a case study using bovine high-density SNP data. Microarrays 2:171–185.

Xu, L. et al. 2017. Characterization of copy number variation's potential role in Marek's disease. International Journal of Molecular Sciences 18(5):1020.

Yan, Y. et al. 2015. Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek's disease. BMC Genomics 16(1):843.

Zhao, M.; Wang, Q.; Wang, Q.; Jia, P.; Zhao, Z. 2013. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 14:S1.

Zhou, H.; Evock-Clover, C.M.; Mcmurtry, J.P.; Ashwell, C.M.; Lamont, S.J. 2007. Genome-wide linkage analysis to identify chromosomal regions affecting phenotypic traits in the chicken. IV. Metabolic traits. Poultry Science 86:267–76.

2. GENOME-WIDE DETECTION OF CNVS AND THEIR ASSOCIATION WITH PERFORMANCE TRAITS IN BROILERS

ABSTRACT

Copy number variations (CNVs) are a major type of structural genomic variants that underlie genetic architecture and phenotypic variation of complex traits, not only in humans, but also in livestock animals. We identified CNVs along the chicken genome and analyzed their association with performance traits. Genome-wide CNVs were inferred from Affymetrix® high density SNP-chip data for a broiler population. We identified 23,214 autosomal CNVs, merged into 5,042 distinct CNV regions (CNVRs), covering 12.84% of the chicken autosomal genome. CNVs were concatenated into segments and association analyses were performed with linear mixed models considering a genomic relationship matrix, for birth weight, body weight at 21, 35, 41 and 42 days, feed intake from 35-41 days, feed conversion ratio from 35-41 days and, body weight gain from 35-41 days of age. One significant CNV segment was associated with BWG on GGA3 (*p*-value=0.00443); one significant CNV segment was associated with BW35 (*p*-value=0.00571), BW41 (*p*-value=0.00180) and BW42 (*p*-value=0.00130) on GGA3, and one significant CNV segment was associated with BW on GGA5 (*p*-value=0.00432). All of the significant CNV segments identified were verified by qPCR. We obtained a validation rate of 92.59%. Such CNV segments are located nearby genes, such as *KCNJ11, MyoD1* and *SOX6*, known to underlie growth and development. Moreover, gene-set analysis revealed terms linked with muscle physiology, cellular processes regulation and potassium channels. Overall, this CNV-based GWAS study unravels potential candidate genes that may regulate performance traits in chickens. Our findings provide a foundation for future functional studies on the role of specific genes in regulating performance in chickens.

Keywords: 1. GWAS, 2. Performance, 3. CNVs, 4. QTLs, 5. qPCR

2.1. Introduction

Gallus gallus is an excellent biological model organism for genetic studies [1-3] and a species of considerable economic relevance worldwide. In 2019, global poultry meat consumption was estimated at 97,000 tons [4], being one of the main sources of protein for humans. Understanding the genetic architecture of performance-related traits may contribute to the development of new genomic strategies to increase production efficiency and sustainability of the chicken industry.

Significant advances have been achieved on chicken genetics [5] since the landmark publication of the first reference genome [6], which has been continuously updated with the most recent genome assembly (GRCg6a) released in 2018. Variations in the genome, especially single nucleotide polymorphisms (SNPs), are known to be associated with phenotypic variation [7]. However, structural variations, such as the copy number variations (CNVs) have been increasingly studied and associated with quantitative traits of economic interest in livestock [8-15].

CNVs associated with phenotypes of economic interest are promising targets for animal breeding programs [16]. They are defined as large DNA fragments (conventionally >1 kb) that, due to deletion or duplication events, display variable copy number between individuals of a population [17, 18]. When compared to SNPs, CNVs encompass more total bases and seem to have a higher mutation rate and potentially greater effects on gene structure, gene regulation and consequently gene expression [19, 20].

Various techniques are available for CNVs detection in humans and other animal species [21]. Most of them depend on the analysis of signal intensity along the genome, such as the comparative genomic hybridization array (aCGH) [22] and high-density SNP chips [23]. Although sequencing-based CNV analysis pipelines have been developed and seem to be a viable alternative [24], SNP chips have been commonly used for CNV detection [25-31]. This technology allows CNVs identification due to the abnormal hybridization that occurs for SNPs located in CNV regions (CNVRs) [32]. Simultaneous measurement of both signal intensity variations, measured for each allele of a given SNP, and changes in allelic composition (i.e., B allele frequency) allow the detection of both copy number changes and copy-neutral loss-of-heterozygosity (LOH) events [33, 34].

Several factors, such as detection algorithm, genotyping platform, SNP density and population genetic background may impact CNV scanning performance [35, 36]. Indeed, different algorithms used for CNV detection [37, 38] may demonstrate variable sensitivity, consistency and reproducibility [39], specially for commercial SNP arrays, such as Illumina and Affymetrix SNP chips. One of the most prominent algorithms for CNV detection is the PennCNV software [38], which has been widely applied in several studies on livestock species, including chickens [40, 41], horses [42], pigs [43-48], cattle [49-53] and sheep [54]. Moreover, PennCNV has better consistency when compared to other CNV calling algorithms [39]. Nevertheless, CNVs identified through SNP-chip platforms can be associated with a considerable rate of false negative and positive results [35]. Therefore, the quantitative polymerase chain reaction (qPCR) is commonly used for CNV validation, being a molecular method to confirm computationally identified *loci* [55-57].

In chickens, several studies have identified quantitative trait loci (QTL) and positional candidate genes flagged by SNPs significantly associated with traits of economic interest, such as performance, carcass and abdominal fat [58-71]. Unsurprisingly, the number of CNV-focused studies is increasing in chicken populations as well [41, 72-81]. CNVs associated with late feathering [82], pea-comb phenotype [83], dermal hyperpigmentation [84], dark brown plumage color [85] and resistance/susceptibility to Marek's disease [76, 81, 86] have been reported. None CNV-association study for performance traits in chickens has been described yet.

Herein, we identified CNVs in the genome of a broiler population, performed a CNV-based GWAS for performance traits and validated associated CNV segments by qPCR. In addition, we identified performance-related genes overlapping significant CNV segments to establish relationships between structural genomic variation and such phenotypes.

2.2. Material and Methods

2.2.1. Population description

A paternal broiler line (TT) belonging to the Chicken Breeding Program of Embrapa Swine and Poultry National Research Center, in Concórdia, Santa Catarina State, South of Brazil, was developed in 1992. This line originated from White Plymouth Rock and White Cornish breeds and has been under multiple trait selection to improve body weight, feed conversion, carcass and breast yields, viability, fertility, reduction of abdominal fat and metabolic syndromes [87]. The experimental broiler population evaluated in this study, called the TT Reference Population, was generated by an expansion of the paternal broiler line TT in 2008 and consisted of approximately

1,500 chickens generated in five hatches from 20 males and 92 females (1:5). More details can be found in Marchesi et al. [87].

2.2.2. Phenotype measurement

Body weight was recorded at 1 (birth weight), 21, 35, 41 and 42 (after fasting) days of age. Over the period between 35 and 41 days of age, chickens were transferred to individual cages for measuring feed intake and body weight gain to evaluate feed conversion. At 42 days of age, the chickens were weighed and euthanized by cervical dislocation followed by exsanguination. By then, a blood sample from each animal was collected for subsequent DNA extraction. In this study, we analyzed eight performance traits: (i) birth weight (BW), (ii) body weight at 21 days of age (BW21), (iii) body weight at 35 days of age (BW35), (iv) body weight at 41 days of age (BW41), (v) body weight at 42 days of age (BW42), (vi) feed intake measured from 35-41 days of age (FI), (vii) feed conversion ratio measured from 35-41 days of age (FCR), and body weight gain measured from 35-41 days of age (BWG). More detailed descriptions on this population, rearing conditions and phenotype measurements are available in Marchesi et al. [87]. The descriptive statistics for the analyzed phenotypes are shown in Table 1.

Table 1. Descriptive statistics from phenotypic values for performance traits analyzed in the TT Reference Population.

Traits[1]	N[2]	Mean	SD[3]	Minimum	Maximum
BW	1448	47.66	3.7	37.4	61.8
BW21	1426	648.43	133.86	256	1034
BW35	1450	1730.96	202.52	776	2444
BW41	1443	2219.2	251.82	1026	2922
BW42	1452	2223.86	260.15	988	2971
FI	1443	1091.45	152.43	508	1590
FCR	1439	2.31	0.47	1.42	5.25
BWG	1439	488.77	106.53	128	802

[1]BW: birth weight in grams; BW21: body weight at 21 days in grams; BW35: body weight at 35 days in grams; BW41: body weight at 41 days in grams; BW42: body weight at 42 days in grams; FI: feed intake from 35-41 days in grams; FCR: feed conversion ratio from 35-41 days; BWG: body weight gain from 35-41 days in grams
[2]Number of animals
[3]Standard deviation of the mean

2.2.3. DNA extraction, genotyping and quality control

Genomic DNA from 1,461 blood samples was extracted using the PureLink® Genomic DNA (Invitrogen, Carlsbad, CA, USA) kit and then quantified using Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). After extraction, DNA integrity was evaluated on agarose gel (1%) and diluted to 10 ng/µL. Diluted genomic DNA was prepared following recommended Affymetrix protocols in order to perform the genotyping analysis using the 600K Affymetrix Axiom Genotyping Array (Affymetrix, Inc. Santa Clara, CA, USA, [88], that contains segregating SNPs for different populations, including commercial broiler lines.

Initially, Axiom™ Analysis Power Tools (Affymetrix®) software v.2.10.2.2 was used to filter genotypes based on DishQC and call rate parameters. A minimum default quality control of 0.82 and a minimum sample call rate of 97% were used. Therefore, only samples with DishQC ≥0.82 and call rate ≥ 97% were kept for following

analyses. SNPs in sex chromosomes, and those not mapped in the chicken genome assembly (GRCg6a) were excluded. Only SNPs annotated to autosomal chromosomes from GGA1 to GGA33 were included in the analysis. From the total of 580,961 SNPs available on the SNP array, 476,254 informative polymorphic SNPs on the autosomal chromosomes (GGA1-33) were kept after filtering.

2.2.4. Input construction and CNV calling

CNV calling was performed using PennCNV v.1.0.5 [38], an integrated hidden Markov model (HMM) that merges various sources of information, including relative signal intensities (log R Ratio, LRR) and relative allele frequencies (B allele frequency, BAF) at each SNP, the distance between adjacent SNPs, and the population frequency of the B allele (PFB).

The files denominated 'summary', 'calls' and 'confidences' that are built during SNP genotyping and initial data filtering, and are required for signal intensities estimation, were used to extract the LRR and BAF values. First, these files were used to generate canonical clusters (Peiffer, 2006) by the PennCNV-Affy 'generate_affy_ geno_cluster.pl' function, which allows the estimation of the LRR and BAF values by the PennCNV-Affy 'normalize_affy_geno_cluster.pl' function. Then, the PFB file was estimated from marker's individual BAF values, using the PennCNV 'compile_pfb.pl' function. Next, the individual-based CNV calling was performed using the -test option with default parameters for the HMM model. Given that the GC ratio content around each SNP marker is known to influence signal strength, creating the so-called genomic waves [89], the LRR of each sample was corrected using the chicken GC content file (i.e., GC content of 1-Mb genomic regions surrounding each SNP) by the -gcmodel option. As long as family structure can be used for generating more accurate CNV calls [38], and pedigree information for a father-mother-offspring trio was available, a family-based CNV detection algorithm was used to jointly update CNV status previously obtained in the individual-based calling step.

For CNV filtering, the default PennCNV standard deviation (SD) criteria for LRR≤0.35, BAF drift<0.01, and waviness factor≤0.05 were used. Note that the waviness factor represents the dispersion in signal intensity over the genome. Moreover, CNVs smaller than 1 kb were excluded and only CNVs consisting of at least three consecutive SNPs were retained in the analysis [41]. Lastly, all duplicated CNVs (i.e., same event in the same parental) were removed. Duplicated CNV entries occurred due to half sib families, as some sires and dams were included more than once in PennCNV analysis. The CNV calling was focused only on autosomal chromosomes GGA1 to GGA33 as PennCNV results for sex chromosomes are unreliable and difficult to interpret [38].

2.2.5. CNVR compilation

Individual CNV calls filtered by PennCNV overlapping at least one base pair were concatenated into CNV regions (CNVRs) using the populationRanges(grl, density=0.1) function from the CNVRanger R/Bioconductor package [90]. Genomic areas with density <10% were deleted to avoid false positive predictions. The CNVRs were classified as gain or loss. The overlapping CNVRs of 'gain' and 'loss' were merged into single regions to account for genomic regions in which both events can occur (i.e., 'both' CNVRs). The frequency of each CNVR was estimated based on the number of samples mapped at the genomic interval covered by the CNVR.

2.2.6. Genome-wide association analyses

Genome-wide association analyses between performance traits and CNV segments were carried out using the CNVRanger R/Bioconductor package [90]. This procedure was originally proposed by da Silva et al. [51]. First, the CNV segments to be used in the association analyses were established. For that, a state was assigned for each of the SNP probes overlapping a CNV call. Then, we estimated the CNV frequency in each probe and selected only those with frequency above 5% [90]. Finally, selected probes were used to construct the CNV segments based on a CNV-genotype similarity, in which subsequent probes with identical genotype in ≥95% of our population were concatenated to CNV segments. A raw *p*-value was independently generated for each probe and, the probe with the lowest *p*-value was selected to represent the CNV segment. Multiple testing correction was carried out using the FDR method [91] generating the *q*-values for each CNV segment. The following statistical models were used:

a) $\boldsymbol{y_{ijkl} = \mu + S_i + H_j + CNV_k + a_l + e_{ijkl}}$

b) $\boldsymbol{y_{ijklm} = \mu + S_i + H_j + CNV_k + b\left(BW35_l - \overline{BW35}\right) + a_m + e_{ijklm}}$

where y_{ijkl} and y_{ijklm} are the phenotypic records on the l^{th} or m^{th} animal, respectively, μ is the overall intercept, S_i is the fixed effect of the i^{th} sex (i = 1, 2), H_j is the fixed effect of the j^{th} hatch (j = 1, 2, 3, 4, 5), CNV_k is the number of copies of a given allele in the genotype of the k^{th} CNV segment (k = 1, …, 191, represented as gain, loss and normal (2n), and coded as 1, − 1 and 0, respectively), b is the linear regression coefficient related to the BW35 effect considered as deviation from the mean ($\overline{BW35}$), a_l and a_m are the random direct additive genetic effect for the l^{th} or m^{th} animal, respectively, and e_{ijkl} and e_{ijklm} are the random residual effect for the l^{th} or m^{th} animal, respectively. Note that sex and hatch were included in the models as class effects for all phenotypes, and BW35 was fit as continuous effect only for FI, FCR and BWG (model b). The random components of the models were distributed as $\mathbf{a} \sim \mathrm{N}(0, \mathbf{G}\sigma_a^2)$ and $\mathbf{e} \sim \mathrm{N}(0, \mathbf{I}\sigma_e^2)$, where σ_a^2 and σ_e^2 are the genetic and residual variances, respectively, $\mathbf{G}$ is the CNV-based genomic relationship matrix, and $\mathbf{I}$ is an identity matrix.

Lastly, we established two different thresholds. The first corresponded to a suggestive association (FDR-corrected *p*-value<0.1) and should be used to identify CNVs for consideration in future studies. The second one corresponded to a significant association (FDR-corrected *p*-value<0.05), consequently, highlighting regions more likely to be truly associated with the investigated phenotypes [92].

2.2.7. Validation by qPCR

Quantitative PCR (qPCR) was carried out to validate significant CNV segments associated with performance traits. Copy number was determined in the 3 significant CNV segments using 3 distinct primer pairs designed to target each CNV segment. We designed 3 primer pairs for each segment as CNV breakpoints may vary. A total of 18 samples, consisting of 3 different reference animals (2n) and 3 different testing animals per CNV segment, were selected for the validation process based on the amount of double-stranded DNA (dsDNA) measured with Qubit® 2.0 Fluorometer. Primers designed using Primer3plus [93] were quality tested through NetPrimer (http://www.premierbiosoft.com/netprimer). Additionally, we used the SNPdb [94] against the Ensembl-Biomart

tool (http://www.ensembl.org/biomart/martview, [95]) to check the presence of SNPs in the genomic region targeted by the primers.

All primers were previously PCR-tested to verify non-specific amplicons and to optimize qPCR conditions. A qPCR solution of a final 10 µl was used consisting of 5.0 µl PowerUp™ SYBR™ Green Master Mix 2x (Applied Biosystems®, catalog number: A25742), 0.5 µl forward primer (10 mM), 0.5 µl reverse primer (10 mM) and 4.0 µl of genomic DNA (2.5 ng/µl). The reference and testing samples were amplified with the designed primers sets in technical triplicates carried out in QuantStudio™ 12k Flex machine coupled to QuantStudio 12K Flex Software v.1.2.2 (Applied Biosystems®). The qPCR thermocycling steps were as follows: 50ºC for 2 min, 95°C for 2 min and 40 cycles of amplification (95°C for 15 seconds, 55–60°C (primer-dependent) for 15 seconds and 72°C for 1 min). The reference samples were randomly chosen from a set predicted by PennCNV to have normal copy number status on each of the tested regions.

Cycle thresholds (Ct) were corrected by primer mean efficiency calculated by LinReg [96] and copy number was estimated from normalized ratio method (NR): $2x2^{-(\Delta\Delta Ct)}$ [97-99]. The primers for the propionylcoenzyme A carboxylase gene (PCCA, GGA1) were used as references ([72, 73, 100, 101]. Moreover, the control value was estimated based on the average value of ΔCt from reference diploid animals, and copy number states were categorized based on the geometric mean between copy number 1, 2 and 3 [98], where lack of amplification was considered as 0n (complete deletion).

2.2.8. CNV segments overlapping known QTLs

Overlaps of significant associated CNV segments with previously mapped QTLs for performance traits were determined using information from the Chicken QTLdb - release 41 [7]. We used the available .gff files with the QTL coordinates based on the last chicken genome assembly (GRCg6a) to check for overlaps using the subsetByOverlaps function from the GenomicRanges R/Bioconductor package [102]. All previously mapped QTLs were reported by QTL ID numbers, available at the Chicken QTLdb [7].

Additionally, we checked the overlapping between the genomic windows covered by the significant CNV segments and the QTLs for growth-related traits reported for the Embrapa F2 Chicken Resource Population, which was originated by crossing sires of the paternal broiler line TT (same line used to obtain the TT Reference Population) and dams of a maternal layer line [103]. Therefore, genomic coordinates were converted from Gallus_gallus-5.0 to GRCg6a using the LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

2.2.9. Identification of candidate genes and gene-set analysis

Gene content of significant CNV segments was assessed using Ensembl Release 101 BioMart tool [95, 104], based on the GRCg6a chicken genome assembly. We investigated genes flanking genomic intervals for significant associated CNV segments, corresponding to 1 Mb windows (500 kb up and downstream).

Enrichment analyses were performed with WebGestalt (http://www.webgestalt.org/), a "WEB-based GEne SeT AnaLysis Toolkit" designed for functional genomics, proteomics and large-scale genetic studies [105, 106]. GO-terms for biological process (BP), cellular component (CC) and molecular function (MF) were

investigated. Multiple testing correction was carried out using the default option (i.e., Benjamini-Hochberg method [91]). In addition, STRING v.11 (http://string-db.org/) [107] was used as a complementary approach to search for enriched pathways and protein domains. Conversely to WebGestalt, the STRING database intends to integrate all publicly available sources of protein–protein direct and indirect interaction information to obtain a comprehensive global network.

2.3. Results

2.3.1. CNV identification

After applying the initial quality control filters, 223 individuals out of 1,461 genotyped chickens from the TT Reference Population presented DishQC<0.82 and call rate<97%, and were excluded from further analyses. Therefore, individual-based CNV calls were performed on the remaining 1,238 samples. Pedigree information on father-mother-offspring trio was used to update the CNV status for the trios, generating more accurate CNV calls [38]. From the total of 1,238 chickens, 709 trios were determined based on complete family information available. Then, the trio-based CNV calling using 779 animals, represented by 709 trios, consisting of 14 sires, 56 dams and 709 offspring, was performed. Several families with incomplete information could not be used as PennCNV is not able to handle trios with missing sire or dam genotypes. After quality control filtering and removal of duplicated CNVs from the dataset, we identified 23,214 unique autosomal CNVs, including 2,905 deletions and 20,309 duplications. Finally, a total of 614 chickens had at least one CNV call after the quality control process.

2.3.2. CNVR compilation

CNVRs represent the concatenation of overlapping CNVs into a consensus genomic region. CNVs showing overlap of at least one base pair among samples in this population were summarized across all individuals into CNVRs. After filtering, 23,214 individual CNVs were merged into 5,042 distinct CNV regions, which cover 12.84% (136.75 million of base pairs - Mb) of the chicken autosomal genome. The number of regions with copy loss and gain were 424 and 4,105, respectively. The presence of both types was observed in 513 regions. The CNVRs had variable sizes ranging from 0.14 kb to 760 kb with an average size of 27.12 kb. The number of chickens with CNVs mapped onto a given CNVR ranged from 1 (0.13%) to 348 (44.67%) from the total of 614 chickens. We identified 656 CNVRs occurring in more than 1% of the population (i.e. 'polymorphic CNVRs', as suggested by Itsara et al. [108]). The relative chromosome coverage by CNVRs ranged from 1.55% for GGA24 to 18.38% for GGA2, while the absolute genomic length overlapped by CNVRs varied from 0.10 Mb for GGA24 to 35.98 Mb for GGA1. Detailed information of all CNVRs detected in our population is provided under request (carolina_fernandes@usp.br).

2.3.3. Association of CNV segments with performance traits

Genome-wide association studies were performed to investigate significant associations of CNV segments, named as CNV-based GWAS, with eight performance-related traits available in our population: BW, BW21, BW35, BW41, BW42, FI, FCR and BWG. There were three distinct CNV segments classified as losses and significantly associated (p-value < 0.05) with BWG, BW35, BW41, BW42 and BW (Table 2).

Table 2. Characterization of significant CNV segments associated with performance traits in the TT Reference Population.

Trait[1]	GGA: first–last position[2]	Number of genes/window[3]
BWG	3: 64169030-64171297	16
BW35	3: 97801202-97809208	3
BW41	3: 97801202-97809208	3
BW42	3: 97801202-97809208	3
BW	5: 12059966-12062666	13

[1]BWG: body weight gain from 35-41 days; BW35: body weight at 35 days; BW41: body weight at 41 days; BW42: body weight at 42 days; BW: birth weight
[2]Map position based on GRCg6a chicken genome assembly
[3]Number of annotated genes within a 1-Mb window of each significant CNV segment associated with performance traits in the TT Reference Population, based on Ensembl Genes 101 Database (https://www.ensembl.org/biomart/martview/)

One CNV segment was significantly associated with BWG (p-value=0.00443); one CNV segment was significantly associated with BW35 (p-value=0.00571), BW41 (p-value=0.00180) and BW42 (p-value=0.00130), and one CNV segment was significantly associated with BW (p-value=0.00432). It is interesting to highlight that the significant CNV segment associated with BW35, BW41 and BW42 was the same (GGA3:97801202-97809208). Manhattan plots for CNV segments across the 33 autosomal chromosomes associated with performance traits are presented in Figure 1. The QQplots for BW, BW35, BW41, BW42 and BWG are in APPENDIX A. Note that none significant CNV segments associated with BW21, FI and FCR were detected.

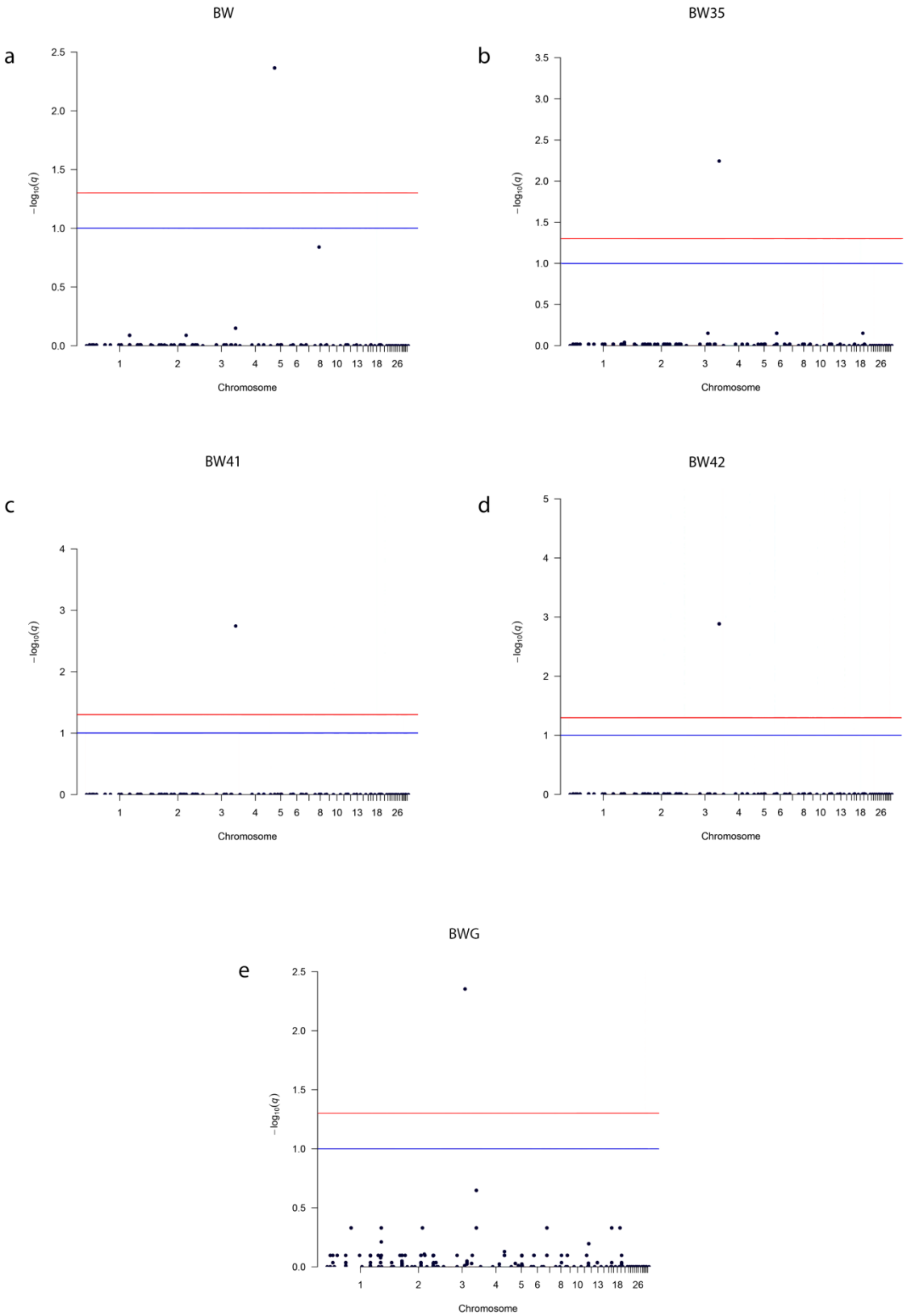


Figure 1. Manhattan plots for CNV segments across the 33 autosomal chromosomes associated with (a) birth weight, (b) body weight at 35 days, (c) body weight at 41 days and (d) body weight at 42 days and (e) body weight gain. The X-axis represents the somatic chromosomes, and Y-axis shows the corresponding -$\log_{10}$ *q*-value. Red and blue lines indicate FDR-corrected *p*-values of 0.05 and 0.1, respectively.

In Figure 2, each dot represents an animal in the corresponding copy number state (0-3n) on the X-axis and the observed phenotypic value on the Y-axis. For the significant CNV segment associated with BW (GGA5: 12059966-12062666), a decrease in copy number is associated with heavier birth weight. The same trend was observed for the significant CNV segment associated with BW35, BW41 and BW42 (GGA3: 97801202-97809208), i.e. higher copy number was observed in animals with lower body weight. Conversely, the significant CNV segment associated with BWG (GGA3: 64169030-64171297) displayed an opposite behavior.

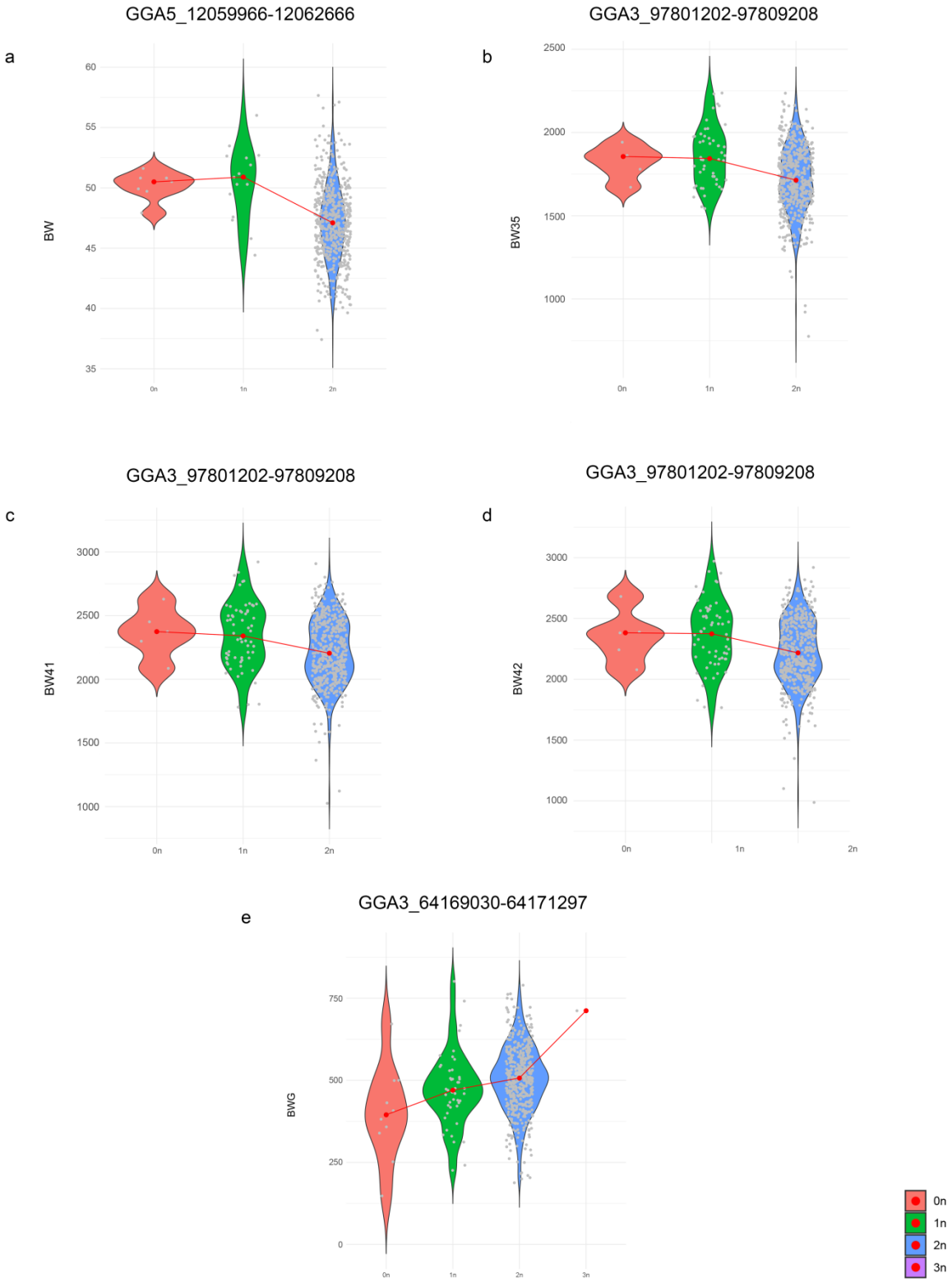


Figure 2. (a) Birth weight, (b) body weight at 35 days, (c) body weight at 41 days and (d) body weight at 42 days and (e) body weight gain distribution in each CN state for the significant CNV segment. Each dot represents an animal in the corresponding copy number state (0-3n) on the X-axis and the observed phenotypic value on the Y-axis. The legend on the right displays the color code for the CN state. See the main text for a detailed description of each segment.

2.3.4. qPCR validation

As CNV breakpoints depended on the segmentation algorithm used in the computational method, results may vary between PennCNV and qPCR. The qPCR results (Figure 3) revealed a validation rate of 92.59%, which confirms the existence of CNV segments that were associated with performance traits. In addition, it showed that for most samples, CNV type was concordant between both methods. For CNVs where at least one breakpoint was within the targeted segment tested, PennCNV results were verified by qPCR (i.e., primers 1-6 and primer 9 for each animal tested). Conversely, in graph c, considering two testing animals (the third animal used had a copy number status estimated by PennCNV of 0n for the segment tested), it is possible that CNV breakpoints for the region tested did not coincide between methods (i.e., primers 7 and 8), therefore, *in silico* results were not verified by qPCR for both primer pairs. Primer information and validation rates are presented in APPENDICES B and C, respectively.

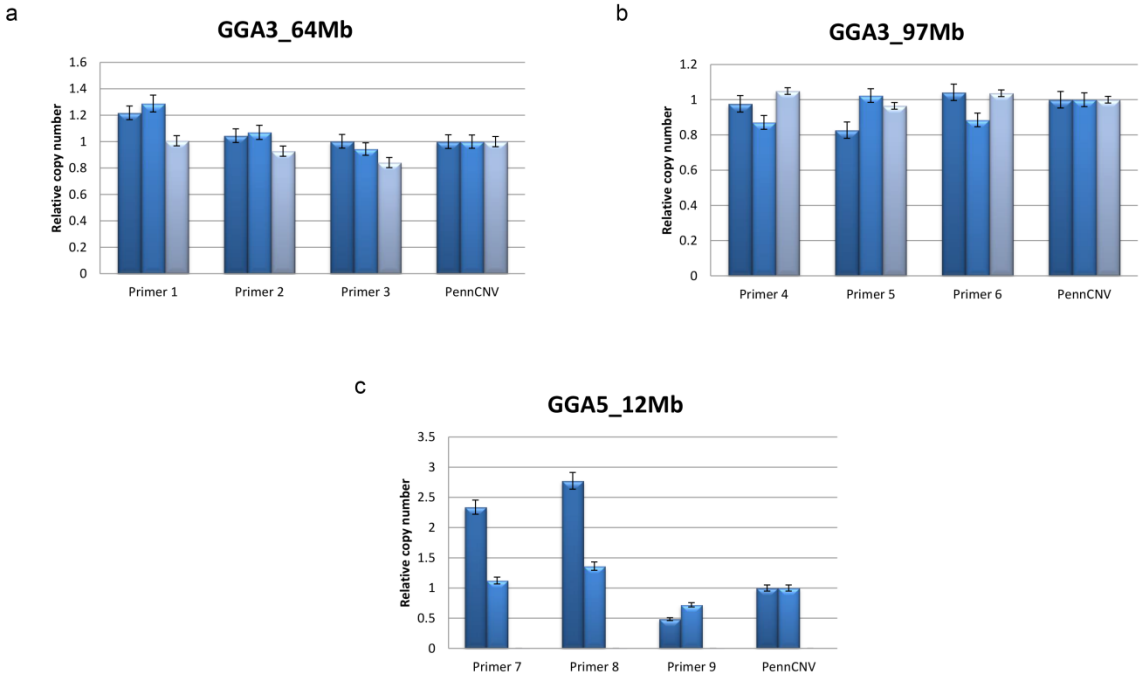


Figure 3. Quantitative PCR was carried out for significantly associated CNV segments on (a) GGA3 at 64Mb, (b) GGA3 at 97Mb and (c) GGA5 at 12Mb using two groups (control (2n) and experimental) with three different animal samples per group and three distinct primer pairs per CNV. In each panel, bars in different colors represent distinct experimental animals for each segment. The right-most bars depict the relative copy number estimated for each animal in PennCNV. Each bar was calculated from three technical replicates. The error bars show the minimum and maximum value encountered among the replicates.

2.3.5. CNV segments overlapping known QTLs

The CNV segment significantly associated with body weight gain (GGA3: 64169030-64171297) overlapped with a QTL associated with residual feed intake on GGA3: 51848188-64944087 (QTL #64556, [68]). Additionally, the significant CNV segments associated with body weight at 35, 41 and 42 days, and body weight gain overlapped with 18 out of 27 previously published QTLs for growth-related traits mapped in the Embrapa F2 Chicken Resource Population ([103], Table 3). None previously reported QTLs overlapped with the CNV segment associated with birth weight (GGA5: 12059966-12062666).

Table 3. CNV segments associated with performance traits overlapping QTL regions previously mapped for growth-related traits.

CNV segments (GGA: first-last position[1])	QTL_IDs	Associated trait
3: 64169030-64171297; 3: 97801202-97809208	**QTL #1979**	**Body_weight**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #1980**	**Body_weight**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #7180**	**Body_weight_(35_days)**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #55904**	**Body_weight_(35_days)**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #55929**	**Growth_(0-35_days)**
3: 64169030-64171297	**QTL #1957**	**Body_weight**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #1961**	**Body_weight**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #1962**	**Body_weight**
3: 64169030-64171297	**QTL #6611**	**Body_weight_(112_days)**
3: 64169030-64171297	**QTL #6612**	**Body_weight_(200_days)**
3: 64169030-64171297	**QTL #6610**	**Body_weight_(8_days)**
3: 64169030-64171297	**QTL #6613**	**Growth_(1-8_days)**
3: 64169030-64171297; 3: 97801202-97809208	**QTL #9420**	**Body_weight_(63_days)**
3: 64169030-64171297	**QTL #11768**	**Body_weight_(49_days)**
3: 64169030-64171297	**QTL #11772**	**Body_weight_(63_days)**
3: 64169030-64171297	**QTL #1969**	**Body_weight**
3: 64169030-64171297	**QTL #1972**	**Body_weight**
3: 64169030-64171297	**QTL #9127**	**Growth_(post-challenge)**

[1]Map position based on GRCg6a chicken genome assembly (NCBI)
CNV segments significantly associated with performance traits located within QTL regions for growth-related traits [103]. QTLs that overlap genomic intervals covered by CNV segments associated with body weight gain (GGA3: 64169030-64171297) and/or body weight at 35, 41 and 42 days (GGA3: 97801202-97809208) are highlighted in bold text.

2.3.6. Candidate genes and gene-set analysis

A total of 32 genes, including *KCNJ11*, *MyoD1* and *SOX6*, were annotated within a 1-Mb window in genomic regions defined by significant CNV segments associated with BWG, BW35, BW41, BW42 and BW. Considering all of the 32 genes, 16 genes were annotated within a 1-Mb window in the genomic region defined by the CNV segment associated with BWG, 3 genes were annotated within a 1-Mb window in the genomic region defined by the CNV segment associated with BW35, BW41 and BW42, and 13 genes were annotated within a 1-Mb window in the genomic region defined by the CNV segment associated with BW (Table 4). No CNV segments were located inside a gene. A list with information about the 32 genes is provided in the APPENDIX D.

Table 4. List of annotated genes within a 1-Mb window of significantly associated CNV segments.

GGA: first-last position[1]	Associated trait (s)[2]	Gene Name (Aliases)	Ensembl Gene ID[3]	Entrezgene ID[4]
3: 64169030-64171297	BWG	RFX6	ENSGALG00000014918	421737
		GPRC6A	ENSGALG00000014925	428620
		FAM162B	ENSGALG00000019941	100857953
		KPNA5	ENSGALG00000014937	421738
		ZUFSP (ZUP1)	ENSGALG00000014940	421739
		SOT3A1L	ENSGALG00000014950	421740
		RWDD1	ENSGALG00000014953	421741
		FAM26D (CALHM4)	ENSGALG00000014955	421742
		TRAPPC3L	ENSGALG00000028539	421743
		FAM26E (CALHM5)	ENSGALG00000038162	769904
		FAM26F (CALHM6)	ENSGALG00000014962	421744
		DSE	ENSGALG00000014963	421745
		NT5DC1	ENSGALG00000014964	421746
		COL10A1	ENSGALG00000014965	100858979
		FRK	ENSGALG00000014979	421747
		HS3ST5	ENSGALG00000026594	428621
3: 97801202-97809208	BW35 BW41 BW42	GREB1	ENSGALG00000016455	421944
		LPIN1	ENSGALG00000016456	421945
		TRIB2	ENSGALG00000016457	378919
5: 12059966-12062666	BW	SOX6	ENSGALG00000006074	423068
		C5H11orf58	ENSGALG00000006077	395520
		PLEKHA7	ENSGALG00000029679	423069
		RPS13	ENSGALG00000006096	414782
		PIK3C2A	ENSGALG00000006121	423070
		NUCB2	ENSGALG00000006147	423071
		KCNJ11	ENSGALG00000020505	428846
		ABCC8	ENSGALG00000006172	423072
		USH1C	ENSGALG00000006192	423073
		FTL	ENSGALG00000028696	378899
		MYOD1	ENSGALG00000006216	374048
		KCNC1	ENSGALG00000006220	423076
		SERGEF	ENSGALG00000006231	423077

[1]Map position based on GRCg6a chicken genome assembly (NCBI)
[2]BWG: body weight gain from 35-41 days, BW35: body weight at 35 days, BW41: body weight at 41 days, BW42: body weight at 42 days, BW: birth weight
[3]Ensembl gene ID based on GRCg6a genome assembly (Ensembl Genes 101 Database)
[4]NCBI gene ID based on GRCg6a genome assembly (http://www.ncbi.nlm.nih.gov/gene)

Gene enrichment analysis was performed using WebGestalt to search for biological processes, cellular components and molecular functions. WebGestalt top 10 most relevant enriched categories for Biological Process, Cellular Component and Molecular Function, based upon genes annotated to each category, can be observed in Table 5. Considering the top 10 most relevant enriched categories for biological processes, it is possible to observe a variety of terms associated with muscle growth and development, such as regulation of striated muscle cell differentiation, regulation of muscle cell differentiation and regulation of muscle tissue development.

Table 5. WebGestalt to 10 most relevant enriched categories for Biological Process, Cellular Component and Molecular Function.

Biological Process GO Terms			
GO ID	**Description**	***p*-value**	**Significant Associated Gene(s)**
GO:0051153	regulation of striated muscle cell differentiation	0.0015	SOX6;MYOD1
GO:0051147	regulation of muscle cell differentiation	0.0030	SOX6;MYOD1
GO:0016202	regulation of striated muscle tissue development	0.0031	SOX6;MYOD1
GO:0048634	regulation of muscle organ development	0.0034	SOX6;MYOD1
GO:1901861	regulation of muscle tissue development	0.0034	SOX6;MYOD1
GO:1901700	response to oxygen-containing compound	0.0048	GPRC6A;RWDD1;KCNJ11;MYOD1
GO:0014070	response to organic cyclic compound	0.0071	RWDD1;KCNJ11;MYOD1
GO:0055026	negative regulation of cardiac muscle tissue development	0.0085	SOX6
GO:0048743	positive regulation of skeletal muscle fiber development	0.0085	MYOD1
GO:1905208	negative regulation of cardiocyte differentiation	0.0085	SOX6
Cellular Component GO Terms			
GO ID	**Description**	***p*-value**	**Significant Associated Gene(s)**
GO:0005887	integral component of plasma membrane	7.9726e-4	GPRC6A;CALHM4;CALHM5; CALHM6;KCNJ11
GO:0031226	intrinsic component of plasma membrane	9.6402e-4	GPRC6A;CALHM4;CALHM5; CALHM6;KCNJ11
GO:0044459	plasma membrane part	0.0017	GPRC6A;CALHM4;CALHM5; CALHM6;FRK;KCNJ11
GO:0071944	cell periphery	0.0073	GPRC6A;CALHM4;CALHM5; CALHM6;COL10A1;FRK;KCNJ11
GO:0030008	TRAPP complex	0.0116	TRAPPC3L
GO:0030315	T-tubule	0.0132	KCNJ11
GO:0005886	plasma membrane	0.0276	GPRC6A;CALHM4;CALHM5; CALHM6;FRK;KCNJ11
GO:0005801	cis-Golgi network	0.0376	TRAPPC3L
GO:0022627	cytosolic small ribosomal subunit	0.0440	RPS13
GO:0005844	polysome	0.0488	RWDD1
Molecular Function GO Terms			
GO ID	**Description**	***p*-value**	**Significant Associated Gene(s)**
GO:0008146	sulfotransferase activity	0.0012	SOT3A1L;HS3ST5
GO:0016782	transferase activity, transferring sulfur-containing groups	0.0019	SOT3A1L;HS3ST5
GO:0008199	ferric iron binding	0.0091	FTL
GO:0034483	heparan sulfate sulfotransferase activity	0.0109	HS3ST5
GO:0070181	small ribosomal subunit rRNA binding	0.0127	RPS13
GO:0016722	oxidoreductase activity, oxidizing metal ions	0.0182	FTL
GO:0008198	ferrous iron binding	0.0182	FTL
GO:0030506	ankyrin binding	0.0182	KCNJ11
GO:0008253	5'-nucleotidase activity	0.0182	NT5DC1
GO:0005242	inward rectifier potassium channel activity	0.0182	KCNJ11

Complementary, STRING databases were used to search for enriched pathways and protein domains on genes annotated within a 1-Mb window of significant CNV segments (Figure 4, Table 6). In figure 4, the larger network, in the middle, and smaller networks, on the right and left extremes, both relate to cell differentiation and muscle functioning. In table 6, considering the enriched pathways for CNV candidate genes related to performance traits, terms associated with potassium channels and regulation of insulin secretion can be observed. In addition, regarding protein domains, we can observe an enriched cluster for calcium homeostasis modulator family, consisting of three members of the FAM26 gene family. Furthermore, 78 publications significantly enriched in STRING containing gene interactions present in the network are presented in APPENDIX E.

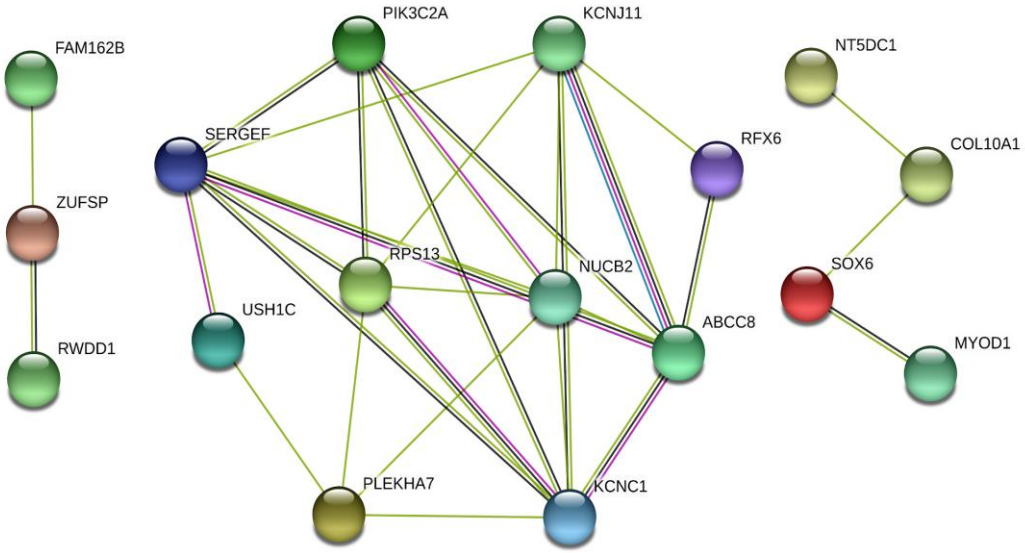


Figure 4. Confidence view of the network created by the STRING software. Nodes represent proteins produced by a single protein-coding gene locus. Edges represent protein-protein associations. Line colors indicate types of interaction evidence: known interactions from curated databases (cyan) or experimentally determined (pink); predicted interactions from gene neighborhood (green); and other sorts of interactions such as co-expression (black).

Table 6. STRING enriched pathways and protein domains for CNV candidate genes related to performance traits.

Reactome Pathways				
#term ID	**Term Description**	**FDR[1]**	**Matching proteins IDs[2]**	**Matching proteins labels[3]**
GGA-1296025	ATP sensitive Potassium channels	0.0043	ENSGALP00000009950, ENSGALP00000032081	ABCC8, KCNJ11
GGA-1296071	Potassium Channels	0.0145	ENSGALP00000009950, ENSGALP00000010023, ENSGALP00000032081	ABCC8, KCNC1, KCNJ11
GGA-1296065	Inwardly rectifying K+ channels	0.0214	ENSGALP00000009950, ENSGALP00000032081	ABCC8, KCNJ11
GGA-422356	Regulation of insulin secretion	0.0384	ENSGALP00000009950, ENSGALP00000032081	ABCC8, KCNJ11
PFAM Protein Domains				
#term ID	**Term Description**	**FDR**	**Matching proteins IDs**	**Matching proteins labels**
PF14798	Calcium homeostasis modulator	3.35e-05	ENSGALP00000024076, ENSGALP00000024082, ENSGALP00000024083	FAM26D, FAM26E, FAM26F
INTERPRO Protein Domains and Features				
#term ID	**Term Description**	**FDR**	**Matching proteins IDs**	**Matching proteins labels**
IPR029569	Calcium homeostasis modulator family	6.17e-05	ENSGALP00000024076, ENSGALP00000024082, ENSGALP00000024083	FAM26D, FAM26E, FAM26F

[1] False Discovery Rate
[2] matching proteins IDs in the network
[3] matching proteins labels in the network

2.4. Discussion

To investigate the effect of CNVs on production-related traits in broilers, we analyzed a Brazilian population, selected for body weight, carcass and cuts yield, feed conversion, fertility, chick viability and reduced abdominal fat. In addition, the known family structure of this population allowed the identification of family-based CNVs.

CNVs are significant sources of genetic variation [8] and have been associated with disease, abnormal development, physical appearance as well as many other economic traits in livestock animals [100, 109-111]. CNV mapping can be based on different reference genome assemblies, populations and platforms. Hence, variability of CNV breakpoints (i.e., genomic coordinates) can happen due to different biological and technical influences [17, 112]. Therefore, CNV comparison among studies is not prosaic, even in the same species, and, as a consequence, different approaches may be complementary to each other [41, 74, 79, 80].

In our population, copy number gains were more abundant than losses. Likewise, Yi et al. [79], Gorla et al. [41] and Sohrabi et al. [113] reported more gains than losses and mixed regions in chicken populations. One reason is that, duplications are more likely to be conserved than deletions because deletion regions are relatively gene-poor and therefore these regions are prone to purifying selection [114]. Nonetheless, deletion polymorphisms might have a significant role in the genetics of complex traits, even though not directly observed in several gene mapping studies [114].

In the present study, significant CNV segments associated with performance traits on chromosome 3, for body weight at 35, 41 and 42 days and body weight gain from 35-41 days, and on chromosome 5, for birth weight were identified. Given that these traits are not independent, and genetic correlations between performance traits have been widely reported in chickens [115-119], it is expected that certain CNV regions may be concomitantly associated with more than one trait, especially body weight measured in different ages (Figure 1).

In the qPCR validation, we systematically assessed the overall agreement rate of the significant CNV segments detected *in silico* with qPCR results. The validation results indicated that all CNV segments were confirmed in at least one qPCR assay, consequently all CNVs may be real. Our results indicated that there is a small discrepancy (7.41%) between qPCR and PennCNV callings, since CNV's exact genomic coordinates may vary and therefore, influence the hybridization of the qPCR primers and the amplification efficiency.

We identified one overlap of the significant CNV segment associated with body weight gain with a previously mapped QTL for residual feed intake (RFI). RFI is defined as the difference between actual feed intake and predicted feed intake based on energy requirements for body weight gain and maintenance [120]. Moreover, we found genomic windows defined by significant CNV segments overlapping published QTLs for growth-related traits in the Embrapa F2 Chicken Resource Population [60]. Many studies, conducted with different chicken lines, have successfully identified QTLs and genes associated with economically important traits [121]. Given that QTLs and genes underlie functional regions of the genome, they may not be prone to structural rearrangements and thus not expected to be subject to CNVs [51]. Therefore, QTLs and genes located inside or nearby CNVs are of special interest.

Noticeably, SNP-based studies [103, 120, 122] have identified many more QTLs associated with the traits analyzed in our study than the CNV-based approach applied here. Indeed, Pértille et al. [103] identified 88 QTLs associated with feed conversion, feed intake, birth weight, and body weight at 35 and 41 days of age in the Embrapa F2 Chicken Resource Population. Mebratie et al. [120] and Moreira et al. [122] identified, respectively, 11 and 19 QTLs associated with body weight traits in a commercial broiler chicken population and in the Embrapa F2 Chicken Resource Population. This difference in QTL mapping is expected since CNVs are more frequently associated with deleterious effects than favorable ones, and this is not the case of SNPs, at least those included in the SNP arrays [123]. In addition, since known QTLs were (mostly) mapped using microsatellite markers and SNPs, they will not necessarily capture the same effect as CNVs. If associated CNVs do not overlap with QTLs previously found in other studies, that could occur because specific CNV probes can be excluded from a SNP-GWAS due to Hardy-Weinberg equilibrium deviation or rigorous multiple testing corrections [51].

CNVs that comprise functional genes may induce phenotypic variation by altering gene structure, dosage and regulation, as a consequence of natural evolutionary processes [19], such as genetic drift [124] or artificial selection. We identified 32 genes annotated within a 1-Mb window of significant CNV segments associated with birth weight, body weight at 35, 41 and 42 days and body weight gain from 35-41 days.

Note that, regarding birth weight and body weight at 35, 41 and 42 days, copy number increase of each respective significant CNV segment was observed in animals with lower body weight at these different ages (Figure 2). Conversely, considering body weight gain, copy number increase of the respective significant CNV segment was positively associated with the phenotype (Figure 2). Since these CNV segments are located in proximity of several genes (Table 4) and, as it has been shown that the expression of a gene may be affected by their presence [112], CNVs may act as important modulators of gene expression. If inserted in an intronic or intergenic region, they could regulate enhancer activity, whereas if located in a promoter region or in the 3' UTR region of a gene, they may

modify binding sites to transcription factors or miRNAs, respectively. In this way, CNVs may have potential to calibrate core regulators located in their proximity, propagating such effects to genome-wide gene expression [125] and account for differences manifested at the phenotype level (Figure 2). Near to the CNV segment associated with birth weight (GGA5: 12059966-12062666), we identified *KCNJ11*, *MyoD1*, *PIK3C2A* and *SOX6* genes, which might have important effects on chicken growth and development regulation.

The *KCNJ11* gene, found to be 2,217 base pairs downstream of the significant CNV segment cited above, is known to regulate insulin secretion [126]. A glucose metabolism disorder is usually linked as a cause of reduced development of chicken muscle tissue under stress, especially in broilers [127]. This gene was enriched to the ATP-sensitive potassium channel (KATP) pathway. KATP subunits are, among other genes, encoded by *KCNJ11*. KATP channels have a high activity in rat fast-twitch fibers, distinguished by raised muscle strength, and a low activity in slow-twitch fibers, characterized by weakness, fragility and lowered muscle strength [128]. *KCNJ11* gene knockout mice present reduced glycogen, slender phenotype and weakness [128]. It has been reported that the effect of the *KCNJ11* gene on muscle may occur due to alterations in the KATP channel activity, which, in turn, affects the potassium flow inside the cell, settling the membrane potential needed for muscle activity [129]. For this reason, this gene may promote early growth and muscle development in chickens. In fact, it was found to be highly expressed in the muscle tissue of one-week-old chicks [126], which supports our findings that this gene, closely located to a significantly associated CNV segment, can play a role in the regulation of birth weight. In addition, a novel 163-bp indel in the downstream region of this gene was significantly associated with growth traits in chickens [130]. In the same study, synteny analyses found that *KCNJ11* maintains a close connection with its neighboring genes. It is interesting to note that one of these genes is the Myogenic differentiation 1 (*MyoD1*).

The myogenic regulatory factors are a family of vertebrate proteins (*MyoD*, *Myf5*, *Mrf4* and *Myog*) that are robust transcription factors for muscle genes [131]. The *MyoD1* gene can promote myoblast differentiation and have relevant effects on muscle development [132, 133]. Previous study in quail lines revealed that a delay in *MyoD1* expression is associated with increased body weight and muscle mass [134]. A very high degree of synteny is maintained between *MyoD1* containing regions of human chromosome 11 and chicken chromosome 5, comprising *ABCC8*, *KCNJ11*, *PIK3C2A*, *RPS13*, *SERGEF*, *NUCB2* and *PLEKHA7* genes [135]. Mutations in *PIK3C2A* gene were discovered to cause a growth-related genetic syndrome in humans, consisting of dysmorphic features, short stature and skeletal abnormalities [136]. This gene has been attributed to biological functions such as glucose transport, Akt pathway activation, endosomal trafficking, phagosome maturation, mitotic spindle organization, exocytosis and autophagy [137-143].

MyoD1 was enriched in biological processes associated with the *SOX6*, a gene related with muscle physiology, such as regulation of striated muscle cell differentiation and development, regulation of muscle cell differentiation and regulation of muscle organ development. The expression level of the *SOX6* gene was positively associated with CNV and increased during skeletal muscle cell differentiation, by upregulating expression levels of muscle-growth-related genes in chickens as well as in other animal species [144].

We found genes nearby significant CNV segments associated with body weight at 35, 41 and 42 days (*LPIN1* and *TRIB2*) and body weight gain (*GPRC6A* and *NT5DC1*) that may be of special importance and have potential effects on chicken growth. A significant association was found between a variant in the 3' UTR of chicken *LPIN1* gene and breast muscle fiber diameter [145], suggesting that this gene has a potential effect on muscle fiber development. The *TRIB2* gene, a novel regulator of thymocyte cellular proliferation, was found to be involved in reproduction and growth in White Leghorn chickens, and consequently might represent footprints of the selection

process [146]. The *GPRC6A* gene was found to have functions related to testis growth and development in broilers [147]. In addition, another interesting gene is the *NT5DC1*, previously related to muscle tissue, angiogenesis and amino acid metabolism [148]. We found an enriched cluster for calcium homeostasis modulator (*CALHM*) gene family, which included three members: *FAM26D*, *FAM26E* and *FAM26F*. Even though *CALHM*s have been classified as pore-forming subunits of plasma membrane ion channels, questions about their function remain unanswered, hence their role needs to be ascertained on further investigations [149].

In summary, from SNP-chip data of a broiler population, we identified novel structural variation regions in the genome that, based on gene enrichment and literature information, harbor potential candidate genes, with important roles in a wide range of biological, cellular, and molecular processes, linked with muscle differentiation, growth, and development. Our findings reveal that alterations in copy number within or nearby these genes could result in phenotypic variation, thus contributing to a better understanding of performance regulation in chickens.

2.5. Conclusions

This study reports structural variations along the chicken genome associated with five complex traits of economic interest in a broiler population using a probe-level based CNV association approach. We identified CNV segments significantly associated with birth weight, body weight at 35, 41 and 42 days and body weight gain, spanning genes that play known key roles in a wide spectrum of molecular and biological processes linked with chicken growth, muscle differentiation and cellular processes regulation. Our results provide substantial information about the potential CNV impacts on animal production, growth, development, and performance-related traits, laying a foundation for incorporating CNVs into the future poultry breeding programs and contributing to expand scientific research on genetics, particularly on structural variations involved in animal biology and physiology.

REFERENCES

1. Wolpert L: Much more from the chicken's egg than breakfast--a wonderful model system. *Mechanisms of development* 2004, 121(9).
2. Ellegren H: The avian genome uncovered. *Trends in ecology & evolution* 2005, 20(4).
3. Stern C: The chick; a great model system becomes even greater. *Developmental cell* 2005, 8(1).
4. USDA: Livestock and Poultry: World Markets and Trade. In.; 2020.
5. Dodgson J, Delany M, Cheng H: Poultry genome sequences: progress and outstanding challenges. *Cytogenetic and genome research* 2011, 134(1).
6. Hillier L, Miller W, Birney E, Warren W, Hardison R, Ponting C, Bork P, Burt D, Groenen M, Delany M *et al*: Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 2004, 432(7018):695-716.
7. Hu Z, Park C, Wu X, Reecy J: Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic acids research* 2013, 41(Database issue).
8. Hastings P, Lupski J, Rosenberg S, Ira G: Mechanisms of change in gene copy number. *Nature reviews Genetics* 2009, 10(8).

9. Di Gerlando R, Sutera A, Mastrangelo S, Tolone M, Portolano B, Sottile G, Bagnato A, Strillacci M, Sardina M: Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PloS one* 2019, 14(4).
10. Jiang R, Cheng J, Cao X, Ma Y, Chaogetu B, Huang Y, Lan X, Lei C, Hu L, Chen H: Copy Number Variation of the SHE Gene in Sheep and Its Association with Economic Traits. *Animals : an open access journal from MDPI* 2019, 9(8).
11. Li W, Liu D, Tang S, Li D, Han R, Tian Y, Li H, Li G, Li W, Liu X *et al*: A multiallelic indel in the promoter region of the Cyclin-dependent kinase inhibitor 3 gene is significantly associated with body weight and carcass traits in chickens. *Poultry science* 2019, 98(2).
12. Liu M, Fang L, Liu S, Pan M, Seroussi E, Cole J, Ma L, Chen H, Liu G: Array CGH-based detection of CNV regions and their potential association with reproduction and other economic traits in Holsteins. *BMC genomics* 2019, 20(1).
13. Hao D, Wang X, Thomsen B, Kadarmideen H, Wang X, Lan X, Huang Y, Qi X, Chen H: Copy Number Variations and Expression Levels of Guanylate-Binding Protein 6 Gene Associated with Growth Traits of Chinese Cattle. *Animals : an open access journal from MDPI* 2020, 10(4).
14. Jing Z, Wang X, Cheng Y, Wei C, Hou D, Li T, Li W, Han R, Li H, Sun G *et al*: Detection of CNV in the SH3RF2 gene and its effects on growth and carcass traits in chickens. *BMC genetics* 2020, 21(1).
15. Zheng X, Zhao P, Yang K, Ning C, Wang H, Zhou L, Liu J: CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. *Journal of animal science and biotechnology* 2020, 11.
16. Clop A, Vidal O, Amills M: Copy number variation in the genomes of domestic animals. *Animal genetics* 2012, 43(5).
17. Redon R, Ishikawa S, Fitch K, Feuk L, Perry G, Andrews T, Fiegler H, Shapero M, Carson A, Chen W *et al*: Global variation in copy number in the human genome. *Nature* 2006, 444(7118).
18. McCarroll S, Altshuler D: Copy-number variation and association studies of human disease. *Nature genetics* 2007, 39(7 Suppl).
19. Zhang F, Gu W, Hurles M, Lupski J: Copy number variation in human health, disease, and evolution. *Annual review of genomics and human genetics* 2009, 10.
20. Zhang L, Jia S, Yang M, Xu Y, Li C, Sun J, Huang Y, Lan X, Lei C, Zhou Y *et al*: Detection of copy number variations and their effects in Chinese bulls. *BMC genomics* 2014, 15(1).
21. Carter N: Methods and strategies for analyzing copy number variation using DNA microarrays. *Nature genetics* 2007, 39(7 Suppl).
22. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo W, Chen C, Zhai Y *et al*: High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature genetics* 1998, 20(2).
23. Yau C, Holmes C: CNV discovery using SNP genotyping arrays. *Cytogenetic and genome research* 2008, 123(1-4).
24. Zhou B, Ho S, Zhang X, Pattni R, Haraksingh R, Urban A: Whole-genome sequencing analysis of CNV using low-coverage and paired-end strategies is efficient and outperforms array-based CNV analysis. *Journal of medical genetics* 2018, 55(11).
25. Bai H, Sun Y, Liu N, Liu Y, Xue F, Li Y, Xu S, Ni A, Ye J, Chen Y *et al*: Genome-wide detection of CNVs associated with beak deformity in chickens using high-density 600K SNP arrays. *Animal genetics* 2018, 49(3).

26. Kim J, Santure A, Barton H, Quinn J, Cole E, Visser M, Sheldon B, Groenen M, van Oers K, Slate J: A high-density SNP chip for genotyping great tit (Parus major) populations and its application to studying the genetic architecture of exploration behaviour. *Molecular ecology resources* 2018, 18(4).
27. Cozzi M, Martinez-Ruiz C, Roman-Ponce S, Murillo V, Utrera A, Montaño-Bermúdez M, Martinez-Velazquez G, Strillaccia M: Copy number variants reveal genomic diversity in a Mexican Creole cattle population. *Livestock Science* 2019, 229:194-202.
28. Wang Y, Zhang T, Wang C: Detection and analysis of genome-wide copy number variation in the pig genome using an 80 K SNP Beadchip. *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie* 2020, 137(2).
29. Lee Y, Bosse M, Mullaart E, Groenen M, Veerkamp R, Bouwman A: Functional and population genetic features of copy number variations in two dairy cattle populations. *BMC genomics* 2020, 21(1).
30. Liu M, Woodward-Greene J, Kang X, Pan M, Rosen B, Van Tassell C, Chen H, Liu G: Genome-wide CNV analysis revealed variants associated with growth traits in African indigenous goats. *Genomics* 2020, 112(2).
31. Wang Z, Guo J, Guo Y, Yang Y, Teng T, Yu Q, Wang T, Zhou M, Zhu Q, Wang W *et al*: Genome-Wide Detection of CNVs and Association With Body Weight in Sheep Based on 600K SNP Arrays. *Frontiers in genetics* 2020, 11.
32. Wang X, Byers S: Copy Number Variation in Chickens: A Review and Future Prospects. *Microarrays (Basel, Switzerland)* 2014, 3(1).
33. Peiffer D, Le J, Steemers F, Chang W, Jenniges T, Garcia F, Haden K, Li J, Shaw C, Belmont J *et al*: High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome research* 2006, 16(9).
34. Komura D, Shen F, Ishikawa S, Fitch K, Chen W, Zhang J, Liu G, Ihara S, Nakamura H, Hurles M *et al*: Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome research* 2006, 16(12).
35. Zhang D, Qian Y, Akula N, Alliey-Rodriguez N, Tang J, Gershon E, Liu C: Accuracy of CNV Detection from GWAS Data. *PloS one* 2011, 6(1).
36. Kim Y, Moon S, Hwang M, Kim D, Oh J, Kim Y, Han B, Lee J, Kim B: Gene-based copy number variation study reveals a microdeletion at 12q24 that influences height in the Korean population. *Genomics* 2013, 101(2).
37. Colella S, Yau C, Taylor J, Mirza G, Butler H, Clouston P, Bassett A, Seller A, Holmes C, Ragoussis J: QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. *Nucleic acids research* 2007, 35(6).
38. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant S, Hakonarson H, Bucan M: PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research* 2007, 17(11).
39. Zhang X, Du R, Li S, Zhang F, Jin L, Wang H: Evaluation of copy number variation detection for a SNP array platform. *BMC bioinformatics* 2014, 15.
40. Rao Y, Li J, Zhang R, Lin X, Xu J, Xie L, Xu Z, Wang L, Gan J, Xie X *et al*: Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. *Poultry science* 2016, 95(8).
41. Gorla E, Cozzi M, Román-Ponce S, Ruiz López F, Vega-Murillo V, Cerolini S, Bagnato A, Strillacci M: Genomic variability in Mexican chicken population using copy number variants. *BMC genetics* 2017, 18(1).

42. Dupuis M, Zhang Z, Durkin K, Charlier C, Lekeux P, Georges M: Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. *Animal genetics* 2013, 44(2).
43. Chen C, Qiao R, Wei R, Guo Y, Ai H, Ma J, Ren J, Huang L: A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. *BMC genomics* 2012, 13.
44. Wang J, Wang H, Jiang J, Kang H, Feng X, Zhang Q, Liu J: Identification of genome-wide copy number variations among diverse pig breeds using SNP genotyping arrays. *PloS one* 2013, 8(7).
45. Wang Y, Tang Z, Sun Y, Wang H, Wang C, Yu S, Liu J, Zhang Y, Fan B, Li K *et al*: Analysis of genome-wide copy number variations in Chinese indigenous and western pig breeds by 60 K SNP genotyping arrays. *PloS one* 2014, 9(9).
46. Schiavo G, Dolezal M, Scotti E, Bertolini F, Calò D, Galimberti G, Russo V, Fontanesi L: Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. *Animal genetics* 2014, 45(5).
47. Dong K, Pu Y, Yao N, Shu G, Liu X, He X, Zhao Q, Guan W, Ma Y: Copy number variation detection using SNP genotyping arrays in three Chinese pig breeds. *Animal genetics* 2015, 46(2).
48. Long Y, Su Y, Ai H, Zhang Z, Yang B, Ruan G, Xiao S, Liao X, Ren J, Huang L *et al*: A genome-wide association study of copy number variations with umbilical hernia in swine. *Animal genetics* 2016, 47(3).
49. McDaneld T, Kuehn L, Thomas M, Pollak E, Keele J: Deletion on chromosome 5 associated with decreased reproductive efficiency in female cattle. *Journal of animal science* 2014, 92(4).
50. Wang M, Dzama K, Hefer C, Muchadeyi F: Genomic population structure and prevalence of copy number variations in South African Nguni cattle. *BMC genomics* 2015, 16.
51. da Silva V, Regitano L, Geistlinger L, Pértille F, Giachetto P, Brassaloti R, Morosini N, Zimmer R, Coutinho L: Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. *PloS one* 2016, 11(6).
52. Santana M, Junior G, Cesar A, Freua M, da Costa Gomes R, da Luz E Silva S, Leme P, Fukumasu H, Carvalho M, Ventura R *et al*: Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *Journal of applied genetics* 2016, 57(4).
53. Durán Aguilar M, Román Ponce S, Ruiz López F, González Padilla E, Vásquez Peláez C, Bagnato A, Strillacci M: Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie* 2017, 134(1).
54. Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F *et al*: Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC genomics* 2013, 14.
55. Leppa V, Kravitz S, Martin C, Andrieux J, Le Caignec C, Martin-Coignard D, DyBuncio C, Sanders S, Lowe J, Cantor R *et al*: Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. *American journal of human genetics* 2016, 99(3).
56. Bujakowska K, Fernandez-Godino R, Place E, Consugar M, Navarro-Gomez D, White J, Bedoukian E, Zhu X, Xie H, Gai X *et al*: Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations. *Genetics in medicine : official journal of the American College of Medical Genetics* 2017, 19(6).

57. Bestetti I, Castronovo C, Sironi A, Caslini C, Sala C, Rossetti R, Crippa M, Ferrari I, Pistocchi A, Toniolo D *et al*: High-resolution array-CGH analysis on 46,XX patients affected by early onset primary ovarian insufficiency discloses new genes involved in ovarian function. *Human reproduction (Oxford, England)* 2019, 34(3).
58. de Koning D, Haley C, Windsor D, Hocking P, Griffin H, Morris A, Vincent J, Burt D: Segregation of QTL for production traits in commercial meat-type chickens. *Genetical research* 2004, 83(3).
59. Navarro P, Visscher P, Knott S, Burt D, Hocking P, Haley C: Mapping of quantitative trait loci affecting organ weights and blood variables in a broiler layer cross. *British poultry science* 2005, 46(4).
60. Nones K, Ledur M, Ruy D, EE B, Melo C, Moura A, Zanella E, Burt D, Coutinho L: Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. *Animal genetics* 2006, 37(2).
61. Zhou H, Evock-Clover C, McMurtry J, Ashwell C, Lamont S: Genome-wide linkage analysis to identify chromosomal regions affecting phenotypic traits in the chicken. IV. Metabolic traits. *Poultry science* 2007, 86(2).
62. Ambo M, Moura A, Ledur M, Pinto L, Baron E, Ruy D, Nones K, Campos R, Boschiero C, Burt D *et al*: Quantitative trait loci for performance traits in a broiler x layer cross. *Animal genetics* 2009, 40(2).
63. Campos R, Nones K, Ledur M, Moura A, Pinto L, Ambo M, Boschiero C, Ruy D, Baron E, Ninov K *et al*: Quantitative trait loci associated with fatness in a broiler–layer cross. *Animal Genetics* 2009, 40:729–736.
64. Baron E, Moura A, Ledur M, Pinto L, Boschiero C, Ruy D, Nones K, Zanella E, Rosário M, Burt D *et al*: QTL for percentage of carcass and carcass parts in a broiler x layer cross. *Animal genetics* 2011, 42(2).
65. Nassar M, Goraga Z, Brockmann G: Quantitative trait loci segregating in crosses between New Hampshire and White Leghorn chicken lines: II. Muscle weight and carcass composition. *Animal genetics* 2012, 43(6).
66. Boschiero C, Jorge E, Ninov K, Nones K, do Rosário M, Coutinho L, Ledur M, Burt D, Moura A: Association of IGF1 and KDM5A polymorphisms with performance, fatness and carcass traits in chickens. *Journal of applied genetics* 2013, 54(1).
67. Felício A, Boschiero C, Balieiro J, Ledur M, Ferraz J, Michelan Filho T, Moura A, Coutinho 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* 2013, 12(1).
68. Mignon-Grasteau S, Rideau N, Gabriel I, Chantry-Darmon C, Boscher M, Sellier N, Chabault M, Le Bihan-Duval E, Narcy A: Detection of QTL controlling feed efficiency and excretion in chickens fed a wheat-based diet. *Genetics, selection, evolution : GSE* 2015, 47.
69. Wang W, Zhang T, Wang J, Zhang G, Wang Y, Zhang Y, Zhang J, Li G, Xue Q, Han K *et al*: Genome-wide association study of 8 carcass traits in Jinghai Yellow chickens using specific-locus amplified fragment sequencing technology. *Poultry science* 2016, 95(3).
70. Grupioni N, Stafuzza N, Carvajal A, Ibelli A, Peixoto J, Ledur M, Munari D: Association of RUNX2 and TNFSF11 genes with production traits in a paternal broiler line. *Genetics and molecular research : GMR* 2017, 16(1).
71. Moreira G, Boschiero C, Cesar A, Reecy J, Godoy T, Trevisoli P, Cantão M, Ledur M, Ibelli A, Peixoto J *et al*: A genome-wide association study reveals novel genomic regions and positional candidate genes for fat deposition in broiler chickens. *BMC genomics* 2018, 19(1).
72. Wang X, Nahashon S, Feaster T, Bohannon-Stewart A, Adefope N: An initial map of chromosomal segmental copy number variations in the chicken. *BMC genomics* 2010, 11.
73. Wang Y, Gu X, Feng C, Song C, Hu X, Li N: A genome-wide survey of copy number variation regions in various chicken breeds by array comparative genomic hybridization method. *Animal genetics* 2012, 43(3).

74. Crooijmans R, Fife M, Fitzgerald T, Strickland S, Cheng H, Kaiser P, Redon R, Groenen M: Large scale variation in DNA copy number in chicken breeds. *BMC genomics* 2013, 14.
75. Fan W, Ng C, Chen C, Lu M, Chen Y, Liu C, Wu S, Chen C, Chen J, Mao C *et al*: Genome-wide patterns of genetic variation in two domestic chickens. *Genome biology and evolution* 2013, 5(7).
76. Luo J, Yu Y, Mitra A, Chang S, Zhang H, Liu G, Yang N, Song J: Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. *G3 (Bethesda, Md)* 2013, 3(2).
77. Tian M, Wang Y, Gu X, Feng C, Fang S, Hu X, Li N: Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. *BMC genomics* 2013, 14.
78. Abernathy J, Li X, Jia X, Chou W, Lamont S, Crooijmans R, Zhou H: Copy number variation in Fayoumi and Leghorn chickens analyzed using array comparative genomic hybridization. *Animal genetics* 2014, 45(3).
79. Yi G, Qu L, Liu J, Yan Y, Xu G, Yang N: Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. *BMC genomics* 2014, 15(1).
80. Yi G, Qu L, Chen S, Xu G, Yang N: Genome-wide copy number profiling using high-density SNP array in chickens. *Animal genetics* 2015, 46(2).
81. Yan Y, Yang N, Cheng H, Song J, Qu L: Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek's disease. *BMC genomics* 2015, 16.
82. Elferink M, Vallée A, Jungerius A, Crooijmans R, Groenen M: Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC genomics* 2008, 9.
83. Wright D, Boije H, Meadows J, Bed'hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin C, Imsland F, Hallböök F *et al*: Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. *PLoS genetics* 2009, 5(6).
84. Dorshorst B, Okimoto R, Ashwell C: Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the Silkie chicken. *The Journal of heredity* 2010, 101(3).
85. Gunnarsson U, Kerje S, Bed'hom B, Sahlqvist A, Ekwall O, Tixier-Boichard M, Kämpe O, Andersson L: The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. *Pigment cell & melanoma research* 2011, 24(2).
86. Xu L, He Y, Ding Y, Sun G, Carrillo J, Li Y, Ghaly M, Ma L, Zhang H, Liu G *et al*: Characterization of Copy Number Variation's Potential Role in Marek's Disease. *International journal of molecular sciences* 2017, 18(5).
87. Marchesi J, Buzanskas M, Cantão M, Ibelli A, Peixoto J, Joaquim L, Moreira G, Godoy T, Sbardella A, Figueiredo E *et al*: Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal : an international journal of animal bioscience* 2018, 12(6).
88. Kranis A, Gheyas A, Boschiero C, Turner F, Yu L, Smith S, Talbot R, Pirani A, Brew F, Kaiser P *et al*: Development of a high density 600K SNP genotyping array for chicken. *BMC genomics* 2013, 14.
89. Diskin S, Li M, Hou C, Yang S, Glessner J, Hakonarson H, Bucan M, Maris J, Wang K: Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic acids research* 2008, 36(19).
90. da Silva V, Ramos M, Groenen M, Crooijmans R, Johansson A, Regitano L, Coutinho L, Zimmer R, Waldron L, Geistlinger L: CNVRanger: association analysis of CNVs with gene expression and quantitative phenotypes. *Bioinformatics (Oxford, England)* 2019, 36(3).
91. Benjamini Y, Hochberg Y: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 1995, 57:289-300.

92. Duggal P, Gillanders E, Holmes T, Bailey-Wilson J: Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC genomics* 2008, 9.
93. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen J: Primer3Plus, an enhanced web interface to Primer3. *Nucleic acids research* 2007, 35(Web Server issue).
94. Sherry S, Ward M, Kholodov M, Baker J, Phan L, Smigielski E, Sirotkin K: dbSNP: the NCBI database of genetic variation. *Nucleic acids research* 2001, 29(1).
95. Kinsella R, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A *et al*: Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database : the journal of biological databases and curation* 2011, 2011.
96. Ramakers C, Ruijter J, Deprez R, Moorman A: Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience letters* 2003, 339(1).
97. Livak K, Schmittgen T: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods (San Diego, Calif)* 2001, 25(4).
98. D'haene B, Vandesompele J, Hellemans J: Accurate and objective copy number profiling using real-time quantitative PCR. *Methods (San Diego, Calif)* 2010, 50(4).
99. Weaver S, Dube S, Mir A, Qin J, Sun G, Ramakrishnan R, Jones R, Livak K: Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods (San Diego, Calif)* 2010, 50(4).
100. Han R, Yang P, Tian Y, Wang D, Zhang Z, Wang L, Li Z, Jiang R, Kang X: Identification and functional characterization of copy number variations in diverse chicken breeds. *BMC genomics* 2014, 15(1).
101. Lee M, Bornelöv S, Andersson L, Lamont S, Chen J, Womack J: Duplication of chicken defensin7 gene generated by gene conversion and homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America* 2016, 113(48).
102. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V: Software for computing and annotating genomic ranges. *PLoS computational biology* 2013, 9(8).
103. Pértille F, Moreira G, Zanella R, Nunes J, Boschiero C, Rovadoscki G, Mourão G, Ledur M, Coutinho L: Genome-wide association study for performance traits in chickens using genotype by sequencing approach. *Scientific reports* 2017, 7.
104. Flicek P, Ahmed I, Amode M, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S *et al*: Ensembl 2013. *Nucleic acids research* 2013, 41(Database issue).
105. Zhang B, Kirov S, Snoddy J: WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic acids research* 2005, 33(Web Server issue).
106. Liao Y, Wang J, Jaehnig E, Shi Z, Zhang B: WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic acids research* 2019, 47(W1).
107. Szklarczyk D, Gable A, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva N, Morris J, Bork P *et al*: STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* 2019, 47(D1).
108. Itsara A, Cooper G, Baker C, Girirajan S, Li J, Absher D, Krauss R, Myers R, Ridker P, Chasman D *et al*: Population analysis of large copy number variants and hotspots of human genetic disease. *American journal of human genetics* 2009, 84(2).

109. Moro C, Cornette R, Vieaud A, Bruneau N, Gourichon D, Bed'hom B, Tixier-Boichard M: Quantitative effect of a CNV on a morphological trait in chickens. *PloS one* 2015, 10(3).
110. Gao Y, Jiang J, Yang S, Hou Y, Liu G, Zhang S, Zhang Q, Sun D: CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *BMC genomics* 2017, 18(1).
111. Zheng L, Xu J, Li J, Wang D, An Q, Xu L, Ma Y, Wang J, Peng S, Lei C *et al*: Distribution and association study in copy number variation of KCNJ12 gene across four Chinese cattle populations. *Gene* 2019, 689.
112. Cahan P, Li Y, Izumi M, Graubert T: The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. *Nature genetics* 2009, 41(4).
113. Sohrabi S, Mohammadabadi M, Wu D, Esmailizadeh A: Detection of breed-specific copy number variations in domestic chicken genome. *Genome* 2018, 61(1).
114. Conrad D, Andrews T, Carter N, Hurles M, Pritchard J: A high-resolution survey of deletion polymorphism in the human genome. *Nature genetics* 2006, 38(1).
115. Gaya L, Ferraz J, Rezende F, Mourão G, Mattos E, Eler J, Michelan Filho T: Heritability and genetic correlation estimates for performance and carcass and body composition traits in a male broiler line. *Poultry science* 2006, 85(5).
116. Ige A: Relationship Between Body Weight and Growth Traits of Crossbred Fulani Ecotype Chicken in Derived Savannah Zone of Nigeria. *International Journal of Applied Agricultural and Apicultural Research* 2013, 9(1-2):157-166.
117. Momen M, Mehrgardi A, Sheikhy A, Esmailizadeh A, Fozi M, Kranis A, Valente B, Rosa G, Gianola D: A predictive assessment of genetic correlations between traits in chickens using markers. *Genetics, selection, evolution : GSE* 2017, 49(1).
118. Manjula P, Park H, Seo D, Choi N, Jin S, Ahn S, Heo K, Kang B, Lee J: Estimation of heritability and genetic correlation of body weight gain and growth curve parameters in Korean native chicken. *Asian-Australasian journal of animal sciences* 2018, 31(1).
119. Bungsrisawat P, Tumwasorn S, Loongyai W, Nakthong S, Sopannarath P: Genetic parameters of some carcass and meat quality traits in Betong chicken (KU line). *Agriculture and Natural Resources* 2018, 52(3):274-279.
120. Mebratie W, Reyer H, Wimmers K, Bovenhuis H, Jensen J: Genome wide association study of body weight and feed efficiency traits in a commercial broiler chicken population, a re-visitation. *Scientific reports* 2019, 9(1).
121. Abasht B, Dekkers J, Lamont S: Review of quantitative trait loci identified in the chicken. *Poultry science* 2006, 85(12).
122. Moreira G, Poleti M, Pértille F, Boschiero C, Cesar A, Godoy T, Ledur M, Reecy J, Garrick D, Coutinho L: Unraveling genomic associations with feed efficiency and body weight traits in chickens through an integrative approach. *BMC genetics* 2019, 20(1).
123. Rice A, McLysaght A: Dosage-sensitive genes in evolution and disease. *BMC biology* 2017, 15(1).
124. Hou Y, Liu G, Bickhart D, Cardone M, Wang K, Kim E, Matukumalli L, Ventura M, Song J, VanRaden P *et al*: Genomic characteristics of cattle copy number variations. *BMC genomics* 2011, 12.
125. Geistlinger L, da Silva V, Cesar A, Tizioto P, Waldron L, Zimmer R, Regitano L, Coutinho L: Widespread modulation of gene expression by copy number variation in skeletal muscle. *Scientific reports* 2018, 8(1).
126. Wang X, Wei C, Zhang Z, Liu D, Guo Y, Sun G, Wang Y, Li H, Tian Y, Kang X *et al*: Association of growth traits with a structural variation downstream of the KCNJ11 gene: a large population-based study in chickens. *British poultry science* 2020, 61(3).

127. Liu W, Zhao J: Insights into the molecular mechanism of glucose metabolism regulation under stress in chicken skeletal muscle tissues. *Saudi journal of biological sciences* 2014, 21(3).

128. Tricarico D, Selvaggi M, Passantino G, De Palo P, Dario C, Centoducati P, Tateo A, Curci A, Maqoud F, Mele A *et al*: ATP Sensitive Potassium Channels in the Skeletal Muscle Function: Involvement of the KCNJ11(Kir6.2) Gene in the Determination of Mechanical Warner Bratzer Shear Force. *Frontiers in physiology* 2016, 7.

129. Alekseev A, Reyes S, Yamada S, Hodgson-Zingman D, Sattiraju S, Zhu Z, Sierra A, Gerbin M, Coetzee W, Goldhamer D *et al*: Sarcolemmal ATP-sensitive K(+) channels control energy expenditure determining body weight. *Cell metabolism* 2010, 11(1).

130. Wang X, Wang X, Chen B, Guo Y, Tang H, Li D, Liu D, Wang Y, Li G, Kang X *et al*: Association of a new 99-bp indel of the CEL gene promoter region with phenotypic traits in chickens. *Scientific reports* 2020, 10(1).

131. Pownall M, Gustafsson M, Emerson C: Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annual review of cell and developmental biology* 2002, 18.

132. Ling B, Bharathy N, Chung T, Kok W, Li S, Tan Y, Rao V, Gopinadhan S, Sartorelli V, Walsh M *et al*: Lysine methyltransferase G9a methylates the transcription factor MyoD and regulates skeletal muscle differentiation. *Proceedings of the National Academy of Sciences of the United States of America* 2012, 109(3).

133. Zhang X, Patel S, McCarthy J, Rabchevsky A, Goldhamer D, Esser K: A non-canonical E-box within the MyoD core enhancer is necessary for circadian expression in skeletal muscle. *Nucleic acids research* 2012, 40(8).

134. Coutinho L, Morris J, Marks H, Buhr R, Ivarie R: Delayed somite formation in a quail line exhibiting myofiber hyperplasia is accompanied by delayed expression of myogenic regulatory factors and myosin heavy chain. *Development (Cambridge, England)* 1993, 117(2).

135. Macqueen D, Johnston I: An update on MyoD evolution in teleosts and a proposed consensus nomenclature to accommodate the tetraploidization of different vertebrate genomes. *PloS one* 2008, 3(2).

136. Tiosano D, Baris H, Chen A, Hitzert M, Schueler M, Gulluni F, Wiesener A, Bergua A, Mory A, Copeland B *et al*: Mutations in PIK3C2A cause syndromic short stature, skeletal abnormalities, and cataracts associated with ciliary dysfunction. *PLoS genetics* 2019, 15(4).

137. Behrends C, Sowa M, Gygi S, Harper J: Network organization of the human autophagy system. *Nature* 2010, 466(7302).

138. Krag C, Malmberg E, Salcini A: PI3KC2α, a class II PI3K, is required for dynamin-independent internalization pathways. *Journal of cell science* 2010, 123(Pt 24).

139. Falasca M, Maffucci T: Regulation and cellular functions of class II phosphoinositide 3-kinases. *The Biochemical journal* 2012, 443(3).

140. Devereaux K, Dall'Armi C, Alcazar-Roman A, Ogasawara Y, Zhou X, Wang F, Yamamoto A, De Camilli P, Di Paolo G: Regulation of mammalian autophagy by class II and III PI 3-kinases through PI3P synthesis. *PloS one* 2013, 8(10).

141. Franco I, Gulluni F, Campa C, Costa C, Margaria J, Ciraolo E, Martini M, Monteyne D, De Luca E, Germena G *et al*: PI3K class II α controls spatially restricted endosomal PtdIns3P and Rab11 activation to promote primary cilium function. *Developmental cell* 2014, 28(6).

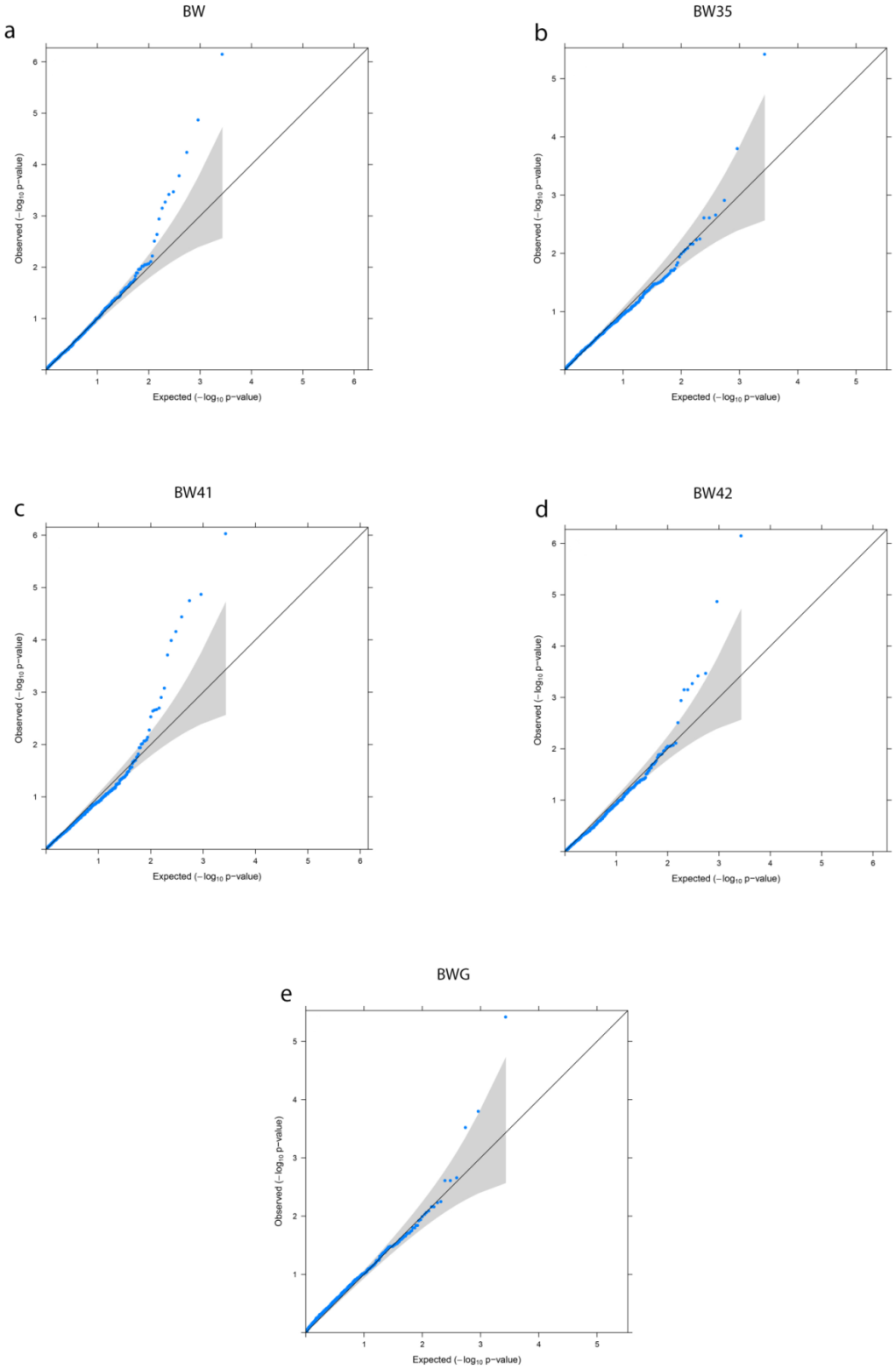
142. Campa C, Franco I, Hirsch E: PI3K-C2α: One enzyme for two products coupling vesicle trafficking and signal transduction. *FEBS letters* 2015, 589(14).

143. Gulluni F, Martini M, De Santis M, Campa C, Ghigo A, Margaria J, Ciraolo E, Franco I, Ala U, Annaratone L *et al*: Mitotic Spindle Assembly and Genomic Stability in Breast Cancer Require PI3K-C2α Scaffolding Function. *Cancer cell* 2017, 32(4).
144. Lin S, Lin X, Zhang Z, Jiang M, Rao Y, Nie Q, Zhang X: Copy Number Variation in SOX6 Contributes to Chicken Muscle Development. *Genes* 2018, 9(1).
145. Zhang S, Li S, Chen W, Lu W, Huang Y: A single-nucleotide polymorphism in the 3' untranslated region of the LPIN1 gene and association analysis with performance traits in chicken. *British poultry science* 2013, 54(3).
146. Zhao Q, Liao R, Sun H, Zhang Z, Wang Q, Yang C, Zhang X, Pan Y: Identifying Genetic Differences Between Dongxiang Blue-Shelled and White Leghorn Chickens Using Sequencing Data. *G3 (Bethesda, Md)* 2018, 8(2).
147. Zhang H, Na W, Zhang H, Wang N, Du Z, Wang S, Wang Z, Zhang Z, Li H: TCF21 is related to testis growth and development in broiler chickens. *Genetics, selection, evolution : GSE* 2017, 49(1).
148. Marchesi J, Ibelli A, Peixoto J, Cantão M, Pandolfi J, Marciano C, Zanella R, Settles M, Coutinho L, Ledur M: Whole transcriptome analysis of the pectoralis major muscle reveals molecular mechanisms involved with white striping in broiler chickens. *Poultry science* 2019, 98(2).
149. Foskett J: Structures of CALHM channels revealed. *Nature structural & molecular biology* 2020, 27(3).

APPENDIX

APPENDIX A.

QQ-plots show the relation of normal theoretical quantiles of the probability distributions between expected (X-axis) and observed (Y-axis) p-values from (a) birth weight, (b) body weight at 35 days, (c) body weight at 41 days, (d) body weight at 42 days and (e) body weight gain.



APPENDIX B. Information of the CNV segments validated by qPCR and the primers used for qPCR.

CNVSegment_ID	GGA	Start (bp)	End (bp)	Length (bp)	Type	Primer_ID	Forward (5'-3')	Reverse (5'-3')	Amplication Length (bp)	Left-Right Position (bp) - Forward	Left-Right Position (bp) - Reverse
GGA3_64Mb	3	64169030	64171297	2268	deletion	1	AATGCAGTGAGCTACGAGAAGA	TCCTTCTTGCACAGACTACACA	91	64169267 - 64169288	64169336 - 64169357
						2	GTGGGAAAATAGATGGATGAGG	CAGGGTAAAGACAGGACAAACA	94	64170053 - 64170074	64170125 - 64170146
						3	TGCAGTTTTGACAGGTGCTT	CCTGCAACAGTTCAAGAGAGAA	91	64170717 - 64170736	64170786 - 64170807
GGA3_97Mb	3	97801202	97809208	8007	deletion	4	GAAGACAAACACACAGGGTGAG	GGCAGACATAGAACAGCTTCAG	100	97802574 - 97802595	97802652 - 97802673
						5	CAAATGATGATTGCCTCCAG	AAGCAAGTAAAGGGAAGGTGAG	87	97805528 - 97805547	97805593 - 97805614
						6	AGGGCCTGATGTTTATTTGG	CCAAGTGATGAGCACAGTATCA	100	97808547 - 97808566	97808625 - 97808646
GGA5_12Mb	5	12059966	12063208	3243	deletion	7	CAGTGTGCTGTGCTGACTTCTA	TGCAGAAGACATTCTTGAGAGG	93	12060133 - 12060154	12060204 - 12060225
						8	AACTAGCCACAAGAACCAGTCA	GTGGGTAAATACTGCTGTGTGC	94	12061132 - 12061153	12061204 - 12061225
						9	TGAAAGGGTCCTCCAAAATA	TTATTCTTCCCCCACTTTCAAC	91	12062215 - 12062234	12062284 - 12062305
PCCA	1						CAGACACACAGAGCCCATCTCT	TGGAGCAGTGGTGGCTGTT	65	143245276	143245340

APPENDIX C. Validation rates of 9 samples.

CNVSegment_ID: GGA3_64Mb			
Primer_ID	**sample11C2**	**sample11C3**	**sample11D2**
1	yes	yes	yes
2	yes	yes	yes
3	yes	yes	yes

CNVSegment_ID: GGA3_97Mb			
Primer_ID	**sample5D2**	**sample10C9**	**sample11D1**
4	yes	yes	yes
5	yes	yes	yes
6	yes	yes	yes

CNVSegment_ID: GGA5_12Mb			
Primer_ID	**sample2F4**	**sample4C6**	**sample11D12**
7	no	yes	yes
8	no	yes	yes
9	yes	yes	yes
Confirmed	7	9	9
Totals	9	9	9
Validation Rate	0.7778	1.0000	1.0000
Average			**0.9259**

APPENDIX D. List with the 32 genes annotated within a 1-Mb window of significant CNV segments.

Gene stable ID	Gene name	Description	Chr.	Gene start (bp)	Gene end (bp)
ENSGALG00000006074	SOX6	SRY-box 6	5	11366630	11601652
ENSGALG00000006077	C5H11orf58	chromosome 5 C11orf58 homolog	5	11765074	11770877
ENSGALG00000006096	RPS13	ribosomal protein S13	5	11951870	11956268
ENSGALG00000006121	PIK3C2A	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 alpha	5	11958466	12018064
ENSGALG00000006147	NUCB2	nucleobindin 2	5	12022253	12051047
ENSGALG00000006172	ABCC8	ATP binding cassette subfamily C member 8	5	12070315	12134884
ENSGALG00000006192	USH1C	USH1 protein network component harmonin	5	12148432	12193954
ENSGALG00000006216	MYOD1	myogenic differentiation 1	5	12395624	12398842
ENSGALG00000006220	KCNC1	potassium voltage-gated channel subfamily C member 1	5	12414763	12510791
ENSGALG00000006231	SERGEF	secretion regulating guanine nucleotide exchange factor	5	12523791	12654177
ENSGALG00000014918	RFX6	regulatory factor X6	3	63709871	63742912
ENSGALG00000014925	GPRC6A	G protein-coupled receptor class C group 6 member A	3	63761182	63771757
ENSGALG00000014937	KPNA5	karyopherin subunit alpha 5	3	63782064	63799496
ENSGALG00000014940	ZUP1	zinc finger with UFM1 specific peptidase domain	3	63799081	63805160
ENSGALG00000014950	SOT3A1L	sulfotransferase family 3A member 1-like	3	63813421	63820420
ENSGALG00000014953	RWDD1	RWD domain containing 1	3	63823638	63834723
ENSGALG00000014955	FAM26D	family with sequence similarity 26 member D	3	63835589	63839410
ENSGALG00000014962	CALHM6	family with sequence similarity 26 member F	3	63868068	63871243
ENSGALG00000014963	DSE	dermatan sulfate epimerase	3	63884722	63905439
ENSGALG00000014964	NT5DC1	5'-nucleotidase domain containing 1	3	63927963	64048903
ENSGALG00000014965	COL10A1	collagen type X alpha 1	3	64024537	64031213
ENSGALG00000014979	FRK	fyn related Src family tyrosine kinase	3	64064317	64109976
ENSGALG00000016455	GREB1	growth regulating estrogen receptor binding 1	3	97305446	97393711
ENSGALG00000016456	LPIN1	lipin 1	3	97402679	97470693
ENSGALG00000016457	TRIB2	tribbles pseudokinase 2	3	97859114	97880210
ENSGALG00000019941	FAM162B	family with sequence similarity 162 member B	3	63775429	63778379
ENSGALG00000020505	KCNJ11	potassium inwardly rectifying channel subfamily J member 11	5	12064883	12066076
ENSGALG00000026594	HS3ST5	heparan sulfate-glucosamine 3-sulfotransferase 5	3	64588327	64784997
ENSGALG00000028539	TRAPPC3L	trafficking protein particle complex 3 like	3	63834724	63863461
ENSGALG00000028696	FTL	ferritin light polypeptide	5	12387716	12391494
ENSGALG00000029679	PLEKHA7	pleckstrin homology domain containing A7	5	11778632	11840720
ENSGALG00000038162	FAM26E	family with sequence similarity 26 member E	3	63853190	63861087

APPENDIX E. 78 reference publications significantly enriched in the STRING network.

#term ID	FDR[1]	matching proteins in network (labels)
PMID:18253507	6.48e-12	ABCC8,KCNC1,KCNJ11,NUCB2,PIK3C2A,RPS13,SERGEF
PMID:30984245	6.61e-05	ABCC8,MYOD1,NUCB2,PIK3C2A
PMID:21544516	0.0023	ABCC8,KCNJ11,RFX6
PMID:22701481	0.0076	ABCC8,KCNC1,KCNJ11
PMID:23266642	0.0098	LPIN1,MYOD1
PMID:26901059	0.0098	ABCC8,KCNJ11,RFX6
PMID:30857219	0.0098	KCNC1,NUCB2,USH1C
PMID:25051960	0.0103	ABCC8,KCNJ11,RFX6
PMID:28389584	0.0103	ABCC8,KCNJ11
PMID:31292350	0.0103	ABCC8,KCNJ11
PMID:18723823	0.0113	ABCC8,KCNJ11
PMID:26226329	0.0113	ABCC8,KCNJ11
PMID:15647111	0.0133	ABCC8,KCNJ11
PMID:19437544	0.0133	COL10A1,MYOD1,SOX6
PMID:22662242	0.0133	GPRC6A,RFX6
PMID:24409153	0.0133	ABCC8,KCNJ11
PMID:25501044	0.0133	NUCB2,PIK3C2A
PMID:30655517	0.0133	ABCC8,KCNJ11
PMID:31988092	0.0133	GREB1,MYOD1
PMID:19403831	0.0148	ABCC8,KCNJ11
PMID:21049026	0.0148	ABCC8,KCNJ11
PMID:21984445	0.0148	ABCC8,KCNJ11
PMID:27956473	0.0148	ABCC8,KCNJ11
PMID:15569391	0.0154	KCNC1,MYOD1
PMID:20376350	0.0154	MYOD1,SOX6
PMID:21035764	0.0154	ABCC8,KCNJ11
PMID:21615117	0.0154	ABCC8,KCNJ11
PMID:23173756	0.0154	ABCC8,KCNJ11
PMID:23346115	0.0154	ABCC8,KCNJ11
PMID:23691027	0.0154	ABCC8,KCNJ11
PMID:23974906	0.0154	ABCC8,KCNJ11
PMID:24763754	0.0154	MYOD1,RPS13
PMID:28417917	0.0154	COL10A1,SOX6
PMID:20676705	0.0159	COL10A1,SOX6
PMID:25564733	0.0159	ABCC8,KCNJ11
PMID:23610594	0.0177	FRK,MYOD1
PMID:20465544	0.0201	ABCC8,KCNJ11
PMID:26243583	0.0201	KCNJ11,MYOD1
PMID:27246103	0.0201	ABCC8,KCNJ11
PMID:25397325	0.0214	ABCC8,KCNJ11
PMID:29670283	0.0214	ABCC8,KCNJ11
PMID:19272164	0.0217	COL10A1,MYOD1,SOX6

PMID:24831221	0.0227	ABCC8,KCNJ11
PMID:27926480	0.0227	C5H11ORF58,MYOD1
PMID:28877214	0.0227	ABCC8,KCNJ11
PMID:30518151	0.0227	MYOD1,SOX6
PMID:31159511	0.0227	MYOD1,SOX6
PMID:15492776	0.0231	COL10A1,SOX6
PMID:17051351	0.0231	MYOD1,NT5DC1
PMID:23064164	0.0231	ABCC8,KCNJ11
PMID:25909383	0.0231	ABCC8,RFX6
PMID:26707643	0.0231	MYOD1,SOX6
PMID:26806292	0.0231	COL10A1,SOX6
PMID:29342086	0.0231	MYOD1,SOX6
PMID:29518216	0.0231	RFX6,SOX6
PMID:16504022	0.0248	COL10A1,SOX6
PMID:19306868	0.0248	COL10A1,SOX6
PMID:23118920	0.0248	ABCC8,KCNJ11
PMID:27434733	0.0248	KPNA5,MYOD1
PMID:21423376	0.0257	ABCC8,KCNJ11
PMID:21529371	0.0257	GREB1,MYOD1
PMID:20213696	0.0274	COL10A1,SOX6
PMID:30303066	0.0274	ABCC8,KCNJ11
PMID:19334283	0.0319	ABCC8,KCNJ11
PMID:22428055	0.0319	COL10A1,SOX6
PMID:30467957	0.0319	ABCC8,KCNJ11
PMID:24421874	0.0332	COL10A1,SOX6
PMID:25802528	0.0332	MYOD1,SOX6
PMID:25210496	0.0350	MYOD1,SOX6
PMID:21040371	0.0374	MYOD1,SOX6
PMID:22150363	0.0397	ABCC8,KCNJ11
PMID:24664432	0.0397	MYOD1,SOX6
PMID:22676585	0.0446	FTL,RPS13
PMID:28642850	0.0471	COL10A1,SOX6
PMID:30760930	0.0471	ABCC8,USH1C
PMID:29616080	0.0489	MYOD1,SOX6
PMID:30836719	0.0489	LPIN1,MYOD1
PMID:32110997	0.0489	LPIN1,MYOD1

[1]false discovery rate