University of São Paulo
“Luiz de Queiroz” College of Agriculture

Inherited copy number variation in the chicken genome and association with
breast muscle traits

Thaís Fernanda Godoy

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

Piracicaba
2018

Thaís Fernanda Godoy
Animal Scientist

**Inherited copy number variation in the chicken genome and association with breast
muscle traits**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. LUIZ LEHMANN COUTINHO

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

Piracicaba
2018

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

Godoy, Thaís Fernanda

Inherited copy number variation in the chicken genome and association with breast muscle / Thaís Fernanda Godoy. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2018.

68 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. De novo 2. Associação genômica 3. CNVs Herdados 4. *Gallus gallus* 5. Genotipagem 6. PennCNV I. Título

ACKNOWLEDGMENTS

First of all, thank God for giving me life, this opportunity and for renewed my strength in all difficult moments.

I would like to acknowledge the GLORIOUS College of Agriculture “Luiz de Queiroz” from São Paulo University (ESALQ/USP). Ten years ago, this college received me with great affection and taught me much more than I thought to learn. I will be eternally grateful!

I would like to thank Prof. Dr. Luiz Lehmann Coutinho for the opportunity to be part of Animal Biotechnology Laboratory group, for his trust, friendship, attention, advice, support, and comprehension during all these years.

Thanks to Embrapa Suínos e Aves and, mostly Dra. Mônica Corrêa Ledur, for the partnership and suggestions during the Doctorate's, giving scientific collaboration to the project.

Thanks Coordination for the Improvement of Higher Education Personnel (CAPES) for my first Doctorate's scholarship. and to São Paulo State Research Foundation (FAPESP) for the scholarships (Process 2015/00616-7) at ESALQ and Research Internship Abroad (Process 2016/05769-9) at Wageningen University (Netherlands).

Thanks to all members of the Wageningen University & Research, particularly to Prof. Dr. Martien A.M. Groenen and Prof. Dr. Richard P.M.A. Crooijmans, for helping me with their knowledge, critical ideas, and attention.

Thanks to Vinicius Henrique da Silva for helping me to develop this project and for all support.

Thanks to my friends from the Animal Biotechnology Laboratory and ESALQ: Aline, Ariana, Bruna, Clarissa, Fábio, Francisco, Gabi, Gabriel, Karina, Mayara, Mirele, Natalia,

Octávio, Paula, Polyana, Priscila Trevisoli, Ribamar, Rosi, Shilton, Tassia for all help, advice, and friendship during this period.

To the lab technicians and collaborators: Horácio, Jorge, Marcela, Mariane, Nirlei, Pilar, Priscila Villela, Rachel, Ricardo, Sônia.

Thanks to Animal Science Department for all support, especially to Prof. Dr. Gerson Barreto Mourão and his team.

Special thanks to my mom (Isabete) and sister (Ana Flávia) for all help, especially in bad moments. Thanks for love and support me all the time.

Thanks to my boyfriend Jeffrey for making me feel so special and for supporting me on this journey. Thanks to all his family for receiving me with so much love.

Thanks to my four-legged friends: Nicole (*in memoriam*), Nayara (*in memoriam*), Nina (*in memoriam*), Binho e Theodoro. Thank you for unconditional love and for true friendship.

Thanks for who have contributed in some way to this work.

Thaís Fernanda Godoy

SUMMARY

RESUMO	6
ABSTRACT	7
1. INTRODUCTION	9
2. LITERATURE REVIEW	11
2.1. Identification of CNV	11
2.2. Copy number variations in the chicken genome	12
3. OBJECTIVES	15
3.1. General Objectives	15
3.2. Specific Objectives.....	15
4. Material and methods.....	17
4.1. Populations, genetic background and family structure.....	17
4.2. DNA extraction, genotyping and quality control	17
4.3. CNV calling and inheritance assignment	18
4.4. Confidence of de novo CNVs	18
4.5. CNV validation by NGS data	19
4.6. QTL overlap and gene enrichment	19
4.7. Phenotype measurement.....	20
4.8. Genome-wide association study of inherited CNV state with breast muscle traits	20
5. RESULTS	23
5.1. CNV identification and distribution.....	23
5.2. Confidence of de novo CNVs	24
5.3. CNV validation by NGS data	25
5.4. A genome-wide map of CNV regions	26
5.5. QTL analysis and gene content of CNVRs	32
5.6. CNV association analysis with breast muscle traits.....	33
6. DISCUSSION	39
6.1. Copy number variation in the chicken genome	39
6.2. GWAS of inherited CNV state with breast muscle traits	42
7. CONCLUSION	45
REFERENCES.....	47
ANNEX.....	61

RESUMO

Variação de número de cópias herdadas no genoma da galinha e associação com características de músculo de peito

A variação de número de cópias (CNV) é um polimorfismo importante que está associado a uma ampla gama de características em seres humanos, espécies selvagens e domésticas. Em frango, que é uma importante fonte de proteína e considerado um modelo biológico, CNVs foram associados a vários fenótipos e passos evolutivos. No entanto, nenhum estudo foi realizado para a identificação e caracterização da herança da CNV no genoma da galinha. Identificamos as CNVs no genoma da galinha usando duas populações experimentais e com pedigree conhecido: uma população de frangos de corte e uma F₂. Em 826 frangos de corte, identificamos 25.819 CNVs (4.299 deleções e 21.520 duplicações), dos quais 21.077 foram herdados, 201 não foram herdados e 4.541 foram CNVs denominados de novo. Em 514 animais F₂, identificamos 21.796 CNVs (2.254 deleções e 19.543 duplicações) das quais 18.230 foram herdadas, 587 não foram herdadas e 2.979 foram de novo CNVs. Após a etapa de filtragem nos de novo CNVs, apenas 220 (4,84%) e 430 (14,43%) permaneceram nas populações de frango de corte e F₂, respectivamente. Um total de 33,11% (50 de 151) das CNV identificadas por dados de genotipagem em dez animais foram validados por dados de sequenciamento. Dos validados, 64% tinham mais de 80% do tamanho (pb) validados. Um total de 59% e 48,8% foram classificados como novas regiões de CNVs (CNVRs) nas populações de frango de corte e F₂, respectivamente. Considerando os *p-values* corrigidos por Bonferroni para testes múltiplos e estatisticamente significativos ($\leq 0,01$), encontramos dois segmentos de CNV significativamente associados ao peso do peito, um ao rendimento de peso de peito, seis ao peso de carne de peito, 18 ao rendimento de carne de peito, quatro ao peso de filé de peito e dois ao rendimento do filé de peito. Esses segmentos de CNV significativamente associados estão sobrepostos com 181 genes codificadores de proteínas. O CNVseg 300, que foi associado a todas as características e abrange seis CNVRs, foram sobrepostos a um total de 26 genes codificadores de proteínas. Entre estes genes, o gene *MYL1* (*Myosin Light Chain 1*) é expresso nas fibras rápidas do músculo esquelético, e os genes *MLPH* (*Melanophilin*), *PRLH* (*Prolactin Releasing Hormone*) e *RAB17* (*Member RAS Oncogene Family*), que foram anteriormente associados ao fenótipo de cor azul acinzentado de penas e à regulação da homeotermia e do metabolismo. O presente estudo melhora o conhecimento sobre CNVs no genoma de frango, especialmente sobre a distribuição de CNV herdadas, não herdadas e *de novo*, em duas populações experimentais de frango. Além disso, a associação genômica foi a primeira realizada na população de frangos de corte com características do músculo do peito, que são muito importantes para a avicultura. Os resultados do GWAS nos permitem compreender a provável relação entre alguns genes e CNVRs que foram significativamente associados às características do músculo do peito.

Palavras-chave: De novo; Associação genômica; CNVs herdados; *Gallus gallus*; Genotipagem; PennCNV

ABSTRACT

Inherited copy number variation in the chicken genome and association with breast muscle traits

Copy number variation (CNV) is an important polymorphism that is associated with a wide range of traits in human, wild and livestock species. In chicken, an important source of animal protein and a developmental model organism, CNV is associated with several phenotypes and evolutionary footprints. However, identification and characterization of CNV inheritance on chicken genome lacks further investigation. We screened CNVs in chicken using two distinct populations with known pedigree. In 826 broilers we identified 25,819 CNVs (4,299 deletions and 21,520 duplications) of which 21,077 were inherited, 201 showed no inheritance and 4,541 were classified as de novo CNVs. In 514 F₂ animals (layer and broiler cross) we identified 21,796 CNVs (2,254 deletions and 19,543 duplications) of which 18,230 were inherited, 587 not inherited and 2,979 were classified as de novo CNVs. After a strict filtering step to remove potential false positives and negative CNVs, only 220 (4.84%) and 430 (14.43%) de novo CNVs remained in the broiler and F₂ populations, respectively. A total of 33.11% (50 out of 151) of the inherited CNVs identified in ten animals were validated by sequencing data. From the validated CNVs, 64% had more than 80% of their size (bp) validated. A total of 59% and 48.8% were classified as novel CNVs regions (CNVRs) in the broiler and F₂, respectively. Considering the Bonferroni-corrected p-values for multiple testing and statistically significant p-values ≤ 0.01 , we found two CNV segments significantly associated with breast weight, one with breast weight yield, six with breast meat weight, 18 CNV segments with breast meat yield, four with breast filet weight and two with breast yield. These CNV segments that were significantly associated overlapped with 181 protein-coding genes. The CNVseg 300, that was associated with all traits and encompass six CNVRs, overlapped a total of 26 protein-coding genes. Among these genes, the gene *MYL1* (*Myosin Light Chain 1*) is expressed in the fast skeletal muscle fibers, and the genes *MLPH* (*Melanophilin*), *PRLH* (*Prolactin Releasing Hormone*) and *RAB17* (*Member RAS Oncogene Family*), that were associated with the lavender phenotype (feather blue-grey color) and regulation of homeothermy and the metabolism. The present study improves our knowledge about CNV in the chicken genome and provides insight in the distribution and of different classes of CNVs, i.e. inherited and de novo CNVs, in two experimental chicken populations. In addition, the genome-wide association analyses were the first performed on broiler population with breast muscle traits, that are important characteristics for poultry production. The GWAS results allow us to understand the probably relationship between some genes and CNVRs that are significantly associated with breast muscle traits.

Keywords: De novo; Genome-wide association; Inherited CNVs; *Gallus gallus*; Genotyping; PennCNV

1. INTRODUCTION

Brazil is the largest exporter and the second largest producer of chicken (*Gallus gallus*) meat in the world, with 12.9 million tons produced and 4.3 million ton exported in 2016 (“ABPA - Brazilian Association of Animal Protein - 2017 Annual Report”, 2017). Besides the chicken production impacts for both Brazilian and global economy, the chicken is also an important biological model in several areas, including cell biology and genetics (ELLEGRÉN, 2005; STERN, 2005; WOLPERT, 2004).

Valuable information for genetic studies in domestic chicken were developed in the last years, as the reference genome (HILLIER et al., 2004) and the consensus linkage map of the chicken genome (GROENEN et al., 2000, 2009). These tools allowed a more precise identification and characterization of genomic features e.g. genes and polymorphisms.

Several types of polymorphisms are known as the single nucleotide polymorphism (SNP), the main marker used and defined as a change in one base pair in DNA, with a minimum frequency of 1% in a population (KWOK; GU, 1999). The SNPs have been used as molecular markers, generating highly saturated and polymorphic genetic maps (KUMAR; BANKS; CLOUTIER, 2012).

However, a class of structural polymorphism called copy number variation (CNV) has been extensively studied. The CNVs are large regions of the genome, greater than 1.000 base pairs (1 kb), that differ in number of copies due to duplication or deletion events. It can also be called copy number polymorphisms (CNPs) when present in more than 1% of the studied population (LOCKE et al., 2006; SCHERER et al., 2007; SCHRIDER; HAHN, 2010).

Since CNV can encompass part or all of a gene, or be a genomic segment containing several genes (ZHANG et al., 2009), they can affect gene expression by disruption of genes, dosage alteration or positional effects (BECKMANN; ESTIVILL; ANTONARAKIS, 2007). Therefore, CNVs have been associated with Mendelian traits and common complex traits (STANKIEWICZ; LUPSKI, 2010).

Most CNVs follow Mendelian inheritance in humans (LOCKE et al., 2006), but de novo mutations can be also functionally important and have been associated with a wide range of diseases (LEPPA et al., 2016; LEVY et al., 2011; VELTMAN; BRUNNER, 2012). Besides this, studies have been shown that CNVs have played an important role in the evolution of genes

and genomes (CARVALHO; ZHANG; LUPSKI, 2010), because they are also susceptible to evolutionary pressures such as selection as well as genetic drift (NGUYEN et al., 2008).

Some studies was performed to identified CNVs on chicken genome using different methodologies and breeds (ABERNATHY et al., 2014; CROOIJMANS et al., 2013; FAN et al., 2013; GORLA et al., 2017; LUO et al., 2013; TIAN et al., 2013; WANG et al., 2010, 2012; YAN et al., 2015; YI et al., 2015, 2014). However, to the best of our knowledge, the CNV inheritance has not yet clearly elucidated. The characterization of CNV inheritance in the chicken genome would assist studies looking for their phenotypic association, especially economic traits for the poultry industry, as well as in the understanding of the evolutionary dynamics of these variants.

In this study we performed a genome-wide detection, characterization and functional annotation of inherited and de novo CNVs in the chicken genome using two experimental chicken populations with family information included. Thereafter, we validated the identified CNVs with the CNVs identified by Next Generation Sequence (NGS). Finally, we also performed a genome-wide association between inherited CNVs and breast muscle traits to understand the role of our detected CNVs in muscle deposition regulation in chickens.

2. LITERATURE REVIEW

2.1. Identification of CNV

The variability in the genome ranges from single nucleotide changes to complex rearrangements in its structure (YALCIN et al., 2012; ZHAO et al., 2016). Copy number variation is the major source of structural variability, and it can arise both meiotically and somatically, from different mutational mechanisms, including DNA recombination, replication and repair processes (CARVALHO; LUPSKI, 2016). However, there is some uncertainty regarding the emergence of CNVs, if these variants result from ancestral mutations and segregate on ancestral haplotypes or if they are multiple mutational events that have occurred on different haplotype backgrounds (LOCKE et al., 2006). In case that CNVs emergence from an ancestral mutations and are also in linkage disequilibrium (LD) with SNPs, the SNPs may be used as markers for genome-wide association studies of CNVs (LOCKE et al., 2006).

Different technologies can be used to identify CNVs in the genome, such as aCGH - comparative genome hybridization array (PINKEL et al., 1998), next generation sequencing (NGS) (KOBOLDT et al., 2012) or SNPs genotyping chip (YAU; HOLMES, 2009).

The aCGH methodology consists of competitive hybridization on DNA arrays with differentially labeled fragments. The number of copies can be identified by the difference in fluorescence intensity compared to the reference genome (PINKEL et al., 1998). The greatest limitation of this method is the low resolution to identify a subtle difference in the number of copies of a genomic fragment, especially when genomes with big number of copies are compared (CLOP; VIDAL; AMILLS, 2012).

Next generation of sequencing (NGS) allows us to identify CNVs of smaller sizes, to define the exact location of CNVs in the genome (WU et al., 2013) and to distinguish homozygous and heterozygous (WANG; BYERS, 2014). However, the method is still relatively expensive, time-consuming and with a high level of false positives (WANG; BYERS, 2014).

The SNP chip technology allows identifying CNVs because of abnormal hybridization that occurs for SNP probes located in the region of CNV (WANG; BYERS, 2014). This method generates large-scale data and distinct algorithms have been proposed for CNV detection

using signal variability (COLELLA et al., 2007) and heterozygosity level (WANG et al., 2007) in those platforms.

Noteworthy, different algorithms can display variable sensitivity, consistency and reproducibility (ZHANG et al., 2014b). Nevertheless, PennCNV (WANG et al., 2007) was widely used in previous studies in different species, such as in horse (DUPUIS et al., 2013), pig (CHEN et al., 2012; DONG et al., 2015; LONG et al., 2016; SCHIAVO et al., 2014; WANG et al., 2013, 2014), cattle (DA SILVA et al., 2016; DE ALMEIDA SANTANA et al., 2016; DURÁN AGUILAR et al., 2017; MCDANELD et al., 2014; WANG et al., 2015), sheep (LIU et al., 2013), including chicken (GORLA et al., 2017; JIA et al., 2013; RAO et al., 2016; ZHANG et al., 2014a; ZHOU et al., 2014), and display a better consistence in comparison with other algorithms (ZHANG et al., 2014b). Moreover, PennCNV is designed for CNV calling using parents-offspring trios (WANG et al., 2008), which allows higher confidence for detection of inherited and de novo CNVs in populations with known family structure.

Some studies have been performed on human genome to understand the difference between the effect of inherited and de novo CNVs in diseases. De novo mutations are usually thought to be more deleterious, whereas inherited CNVs are generally considered benign (NOWAKOWSKA, 2017). However, it is well known that both types of CNVs can happen in a normal individual as well as in individuals with abnormal phenotypes (REDON et al., 2006; VALSESIA et al., 2012)

However, CNV detection and subsequent characterization into inherited and de novo can be challenging given technical limitations (LI; OLIVIER, 2013; TEO et al., 2012; ZHAO et al., 2012). Moreover, their mutation rates are poorly understood (VELTMAN; BRUNNER, 2012) and can vary depending of local architecture in the genome (CARVALHO; LUPSKI, 2016).

2.2. Copy number variations in the chicken genome

The chicken was the first livestock species whose genome was sequenced (HILLIER et al., 2004), allowing more precise identification and characterization of genetic variants. Copy number variation have been identified in the chicken genome by all three methods mentioned before: aCGH, NGS and SNP chip technology.

The development of an aCGH array in chicken allowed the creation of the first genomic CNV map (WANG et al., 2010). Comparative maps had been performed between chicken and

turkey (GRIFFIN et al., 2008) and duck (SKINNER et al., 2009). Other studies also used aCGH technique to identify CNVs (ABERNATHY et al., 2014; CROOIJMANS et al., 2013; LUO et al., 2013; WANG et al., 2012). NGS has also been used for the identification of CNVs in the chicken genome (FAN et al., 2013; KERSTENS et al., 2011; YI et al., 2014).

CNV identification studies from SNP chips in chickens were performed using the Illumina 60K BeadChip developed by Groenen et al. (2011) (JIA et al., 2013; ZHANG et al., 2014b) and using the high density SNP chip available for the domestic chicken (Affymetrix 600K) developed by Kranis et al (2013) (GORLA et al., 2017; YI et al., 2015).

Copy number variations can induce phenotypic variation due to their effect on gene expression, which can be caused by disruption of genes, dosage alteration or positional effects (BECKMANN; ESTIVILL; ANTONARAKIS, 2007). CNVs have been associated with several important human diseases, such as cancer (KOHLENER et al., 2012; LEE et al., 2012), diabetes (KINGERY et al., 2012), obesity (WHEELER et al., 2013; YANG et al., 2013) and mental diseases (MCCARTHY et al., 2009; ROVELET-LECRUX et al., 2006; VINCENT et al., 2012).

Different phenotypes in chicken are associated with CNVs, including late feathering (ELFERINK et al., 2008), pea-comb phenotype (WRIGHT et al., 2009), dark brown plumage color (GUNNARSSON et al., 2011), dermal hyperpigmentation (DORSHORST; OKIMOTO; ASHWELL, 2010).

However, CNV association studies with quantitative traits are rare in animal production, it is more common in studies in humans and rats. In chickens, some studies have been performed to identify CNVs related with Marek's disease resistance and susceptibility (LUO et al., 2013; XU et al., 2017; YAN et al., 2015). Moreover, host-defense gene families experienced distinct CNV evolution among chicken breeds (LEE et al., 2016), revealing genes of potential importance for immune responses. These variants revealed genes underlying phenotypes (GU et al., 2011; HUANG et al., 2009; XU; NIE; ZHANG, 2013) and footprints of natural and artificial selection (FLINK et al., 2014; QANBARI et al., 2015).

An important quantitative trait in chicken for the poultry industry is related to breast muscle development due the consumer demand for meat with low fat and high protein levels (VELLEMAN, 2015). It is well known that both absolute and relative breast muscle weights are in linear positive correlations with the body weight of broilers (XIAO et al., 2017). Moreover, muscle growth or myogenesis is a complex but precisely regulated

developmental process, where transcription factors are positively involved in the regulation of skeletal muscle development (PERRY; RUDNICKI, 2000). In this context, we can note the importance of studies in chickens to identify CNVs, to understand their inheritance on genome, and associate them with breast muscle, an important trait for the poultry industry.

3. OBJECTIVES

3.1. General Objectives

To identify, characterize and validate CNVs in the genome of two experimental populations of chickens and perform a genome-wide association study of CNVs with quantitative breast muscle traits.

3.2. Specific Objectives

- 1) To identify and classify CNVs in the whole chicken genome in two experimental populations (a broiler and F₂ population) using PennCNV.
- 2) To validate the identified CNV with the CNVs identified by NGS.
- 3) To concatenate the CNVs into CNV regions (CNVRs).
- 4) To compare the CNVRs with previous studies in chicken.
- 5) To investigate possible overlap of CNVRs with QTLs and genes related to metabolic pathways of interest.
- 6) To identify CNVRs significantly associated with breast muscle traits in the broiler population.

4. MATERIAL AND METHODS

4.1. Populations, genetic background and family structure

We genotyped 1,340 chickens of which 828 were from the broiler population (called TT reference), and 515 from the F₂ population. Both experimental populations were developed in the Embrapa Swine and Poultry National Research Centre.

The Brazilian TT Reference population was developed from the expansion of a paternal broiler line TT, developed and maintained under multi-trait selection, including body weight, feed conversion, carcass and cut yield, reduced abdominal fat, and fertility, hatchability of fertile eggs and chick viability. The F₂ population was created for QTL mapping studies by reciprocal crosses between a broiler (called TT) and layer line (called CC). The layer line is a White Leghorn line, and the ancestral of the broiler line was originated by crosses between White Plymouth Rock, New Hampshire and White Cornish. For both populations pedigree information (father, mother and offspring) is available.

This study was conducted with the approval of the Ethics Committee for Animal Use (CEUA) from the Embrapa Swine and Poultry National Research Center under the protocol number 011/2010.

4.2. DNA extraction, genotyping and quality control

Genomic DNA was extracted from blood using PureLink® Genomic DNA (Invitrogen, Carlsbad, CA, USA) kit and quantified using Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). After extraction, DNA integrity was evaluated on agarose gel (1%). The samples were diluted following Affymetrix protocol to perform the genotyping analysis using 600 K Affymetrix Axiom Genotyping Array (Affymetrix, Inc. Santa Clara, CA, USA) by GeneTitan® equipment (Affymetrix) according to the manufacturer's protocol. This SNP genotyping array contains SNPs segregating in different chicken lines as described in more details by Kranis et al. (2013).

We performed the quality control of raw intensity files using Axiom™ Analysis Suite (1.0) program. Animals with Dish quality control (DQC) <0.82 and sample call rates <95% were

excluded for the further analysis. All subsequent analysis were performed using the Galgal5.0 assembly.

4.3. CNV calling and inheritance assignment

The Log R Ratio (LRR) and B Allele Frequency (BAF) values were generated by Axiom™ CNV Tool software, and the LRR of each sample was corrected for genomic wave through the estimated GC content in genomic windows of 1 Mb around each marker (DISKIN et al., 2008) using the `-gcmodel`. The population frequency of the B allele (PFB) file was calculated based on the BAF of each marker in the animals.

To optimize the Hidden Markov Model (HMM) parameters, we trained the initial file HMM using the `-train` option to create a specific HMM file, which was used in the subsequent step. The joint option was used to identify CNVs in a single step for three individuals in a family (father, mother and offspring). This option has a good performance to detect correct CNV boundaries and for reducing false negative rates on small CNV calls (<10 SNPs) (WANG et al., 2008). The CNV calling was focused on chromosomes 1–33 given that PennCNV results for the sex chromosomes were unreliable and difficult to interpret.

In the filtering step, two types of CNVs were removed from the result: (i) CNVs smaller than 1 kb and (ii) duplicated CNVs (same event in the same parental). The duplicated CNVs occur because some families were half sib families, and therefore some fathers and mothers were used more than one time in PennCNV software for CNV detection. The CNV in the offspring that was identified in at least one of the parents was defined as inherited. On the other hand, if it was not identified in the parents was defined as de novo. If the CNV was only detected in one or both of parents but it was not in the offspring, it was classified as not inherited.

4.4. Confidence of de novo CNVs

To increase the reliability of the identified de novo CNVs, we achieved the confidence of de novo events using SNP genotypes of the probes that overlapped each CNV call, assuming a Mendelian inheritance for those SNP alleles ('`infer_snp_allele.pl`' function previously described in PennCNV software: <http://penncnv.openbioinformatics.org/en/latest/user-guide/denovo/>).

As each CNV may have an origin (paternal or maternal homologous chromosome), in a real de novo CNV, the genotype of the SNPs supporting it should sign to this origin. In this step, only the de novo CNVs, which parental have two copies on homologous chromosomes, and the offspring has a partial deletion (one copy) or partial duplication (three copies) were analyzed.

Thereafter, apart from p-values provided by the PennCNV software, we proposed an index to reflect how ambiguous each of these de novo CNVs was. We used the origin assignment for each SNP provided by PennCNV to estimate the ratio between parental-specific ($|P-M|$) and all SNPs with known origin ($P+M$). If the assignment ratio ($|P-M|/P+M$) was equal to one, all assigned SNPs within this CNV pointed to a unique parental origin. Otherwise, as this ratio decreases, the CNV is more ambiguous until it reaches zero, which means that 50% of the SNP markers signed to maternal and 50% signed to paternal origin.

4.5. CNV validation by NGS data

Ten parental chickens from the F_2 population were sequenced individually using a paired end protocol by HiSeq1000 or HiscanSQ sequencer (Illumina) (Boschiero et al. – unpublished results) and CNVs were identified by CNVnator (ABYZOV et al., 2011). This data was used to validate the CNVs identified in this study. The validation was carried out using R script based on the SubsetByOverlaps function from Bioconductor package GenomicRanges (version 1.24.3) (LAWRENCE et al., 2013) to investigate possible overlaps of the CNVs identified by two different platforms (NGS x SNP genotyping array).

4.6. QTL overlap and gene enrichment

For each population, CNVs regions (CNVRs) were determined by aggregating overlapping CNVs identified in different chickens using the reduce function from Bioconductor package GenomicRanges (version 1.24.3) (LAWRENCE et al., 2013). These regions were separately inferred for the three types of CNVs (inherited, not inherited and de novo), after the filtering step in the item Confidence of de novo.

The CNV regions identified in the present study were compared with 11 previous studies in chicken. Most of the previous studies were performed on galGal3.0 (ABERNATHY et al.,

2014; CROOIJMANS et al., 2013; FAN et al., 2013; LUO et al., 2013; TIAN et al., 2013; WANG et al., 2010, 2012). A few recent studies were performed on galGal4.0 (GORLA et al., 2017; YAN et al., 2015; YI et al., 2014, 2015), but to date only our study was performed on the most recent version of the genome (galGal5.0). Therefore, we converted the CNVRs identified in galGal3.0 in some studies to galGal4.0 by CrossMap tool (version 0.2.5) (ZHAO et al., 2014) and we repeated the entire CNV calling analyses performed in our study using the galGal4.0. After these steps, it was possible to compare the CNVRs in the same version of the genome (galGal4.0).

We also investigated the overlaps between inherited CNVRs and reported QTL regions collected from (i) chicken QTL database (HU et al., 2013), and (ii) QTLs previously mapped in the same F_2 population evaluated in this study (AMBO et al., 2009; BARON et al., 2011; BOSCHIERO et al., 2013; CAMPOS et al., 2009; FELÍCIO et al., 2013; NONES et al., 2006, 2012). Overlap analysis was carried out using R script, and considering at least one overlapped base pair.

Genes overlapping inherited CNVRs were subjected to an enrichment analysis with DAVID (HUANG; SHERMAN; LEMPICKI, 2008, 2009) to identify metabolic pathways that were predominantly associated with CNVRs. The enrichment analysis was performed with CNVRs that occurred in >1% of the population, defined as polymorphic CNVRs.

4.7. Phenotype measurement

Chickens from the Brazilian broiler TT Reference Population were slaughtered at 42 days of age after six hours of fasting. In this step, the blood of each chicken was collected for DNA extraction. The body weight at 42 days of age (BW42), breast weight (BW), breast meat weight (BMW), breast fillet weight (BFW) were recorded. Then, the breast yield (BY), breast meat yield (BMY) and breast fillet yield were estimated based on BW42. More details about the rearing conditions and phenotypes measuring are available in (VENTURINI et al., 2015).

4.8. Genome-wide association study of inherited CNV state with breast muscle traits

Using the genotypic information, we calculated the genomic estimated breeding values (GEBV) for each trait (BW, BY, BMW, BMY, BFW and BFY). The estimation of variance

components (genetic variance, residual variance and total variance) was performed using Bayes C model in GenSel software (GARRICK; FERNANDO, 2013). These variance components were used as priors in Bayes B model to estimate GEBV.

Association analysis between inherited CNVs, detected in 826 chickens from broiler population, and GEBVs was performed using PLINK (PURCELL et al., 2007). The analysis was carried out as described by Silva et al. (2016): (i) Probes within inherited CNVs were assigned to the corresponding number of copies (0n, 1n, 2n, 3n, or 4n); (ii) All genotyped probes that deviated from the normal state (2n) in $\geq 1\%$ of the population were selected; (iii) Genome-wide association analysis with the GEBV of each trait performed using PLINK 1.07; (iv) Subsequent probes with identical genotype in $\geq 95\%$ of our population were concatenated to CNV segments; (v) The probe with most significant p-value was selected as representative for the CNV segment; (vi) CNV segment p-values were corrected for multiple testing by Bonferroni correction.

5. RESULTS

5.1. CNV identification and distribution

The initial quality filtering by the Axiom™ Analysis Suite (1.0) program excluded two individuals from the broiler population and one from the F₂ population, retaining 826 and 514 chickens from the respective populations. After the CNV calling using the `–joint` option from PennCNV and after filtering, we identified 25,819 autosomal CNVs, including 4,299 deletions and 21,520 duplications in the broiler population, and 21,796 CNVs in the F₂ population including 2,254 deletions and 19,543 duplications.

In the broiler population 21,077 CNVs were inherited (81.63%), 201 (0.78%) were not inherited and 4,541 (17.59%) were classified as de novo. In the F₂ population, 18,230 CNVs (83.63%) were inherited, 587 (2.7%) not inherited, and 2,979 (13.67%) were classified as de novo (Figure 1). For all three types of CNV (inherited, not inherited and de novo), in both populations, the gain state (duplication) was more common than loss (deletion) (Table 1). In the broiler population the average number of CNVs per animal was 31.25 CNVs, ranging in size from 1 kb to 4.2 Mb with mean length of 30.19 kb and a median length of 9.67 kb. In the F₂ population the average number was 23 CNVs, ranging in size from 1 kb to 8.5 Mb, with mean length of 43.92 kb and median of 12.39 kb.

Figure 1. Numbers of de novo, inherited and not inherited CNV in broiler and F₂ populations.

Table 1. Number of loss and gain for de novo, inherited and not inherited CNVs in the broiler and F₂ populations.

Population	Number Animals	De Novo		Inherited		Not Inherited	
		Loss	Gain	Loss	Gain	Loss	Gain
Broiler	826	480	4,061	3,800	17,277	19	182
F ₂	514	760	2,219	1,466	16,764	27	560

5.2. Confidence of de novo CNVs

In this analysis, only a specific category of CNVs was considered to assess the confidence of the de novo CNVs: paternal and maternal are 2n (two copies), and the offspring has one copy (1n) or 3 copies (3n). Therefore, from the 4,541 de novo CNVs initially identified in the broiler and 2,979 in the F₂ population, only 3,963 and 2,679 were analyzed, respectively.

Among these de novo CNVs, 3,507 (88.49%) in the broiler population and 2,188 (81.67%) in the F₂ population had at least one informative SNP, in other words, it was possible to assign them to one particular origin. The PennCNV software provided p-values to all de novo CNVs with at least one informative SNPs, from which a total of 914 (23.06%) in the broiler population and 1,095 (40.87%) in the F₂ population were significant (p-value ≤ 0.05).

To increase stringency and avoid ambiguous de novo CNVs, we selected those with 15 or more SNPs, as well as being highly unambiguous regarding parental origin ($|P-M|/P+M \geq 0.8$, see methods for details). This subset of de novo CNVs represented 220 (5.5%) de novo CNVs with informative SNPs in the broiler and 430 (16.05%) in the F₂ population (Figure 2). From this new subset, the total of 220 de novo CNVs in the broiler population were gain events and 393 were gain and 37 loss events in the F₂ population.

Figure 2. Relative ambiguity of the parental origin in de novo CNVs with distinct number of SNP probes. Origin ambiguity was measured based on the ratio between $|P-M|$ and $P+M$, where P and M represent the number of SNPs assigned to the paternal and maternal lines, respectively. The red line indicates the threshold used to select the subset with 15 or more SNPs and with $|P-M|/P+M \geq 0.8$. The reliability of these de novo CNVs is also reflected by the p-values provided by PennCNV. (A) Broiler population; (B) F_2 population.

5.3. CNV validation by NGS data

Using the PennCNV results from the SNP array genotyping data, we identified 151 CNVs in 10 parental chickens of the F_2 population. The same 10 individuals were resequenced with an average of sequencing depth of 10.93 X (Boschiero et al. – unpublished results). A total of 33.11% (50 out of 151) of the CNVs identified by genotyping data were validated. From the 10 chickens, nine had at least one CNV which was validated, and only in one chicken (TT6037) none of the CNVs were validated by sequencing (Annex A).

Among the 50 validated CNVs, 37 were inherited to offspring and 13 were identified only in these individuals. The majority of the validated CNVs, as well as those identified in these animals, were duplications (37 duplications and 13 deletions). Among these validated CNVs, 13 (CNV 16, CNV 26, CNV 29, CNV 38, CNV 67, CNV 71, CNV 87, CNV 91, CNV 103, CNV 119, CNV 132, CNV 133 and CNV 142) were validated but in different state by resequencing. In other words, a PennCNV detected duplication was validated by CNV deletion with sequencing and vice versa. This was more common in CNVs that were validated by more

than one CNV from sequencing data. The CNV 29 (duplication) was validated by three CNVs (two duplications and one deletion), the CNV 38 (deletion) was validated by four CNVs (three duplications and one deletion), the CNV 87 (duplication) was validated by one duplication and one deletion and the CNV 103 (duplication) was validated by two duplications and one deletion (Annex A).

From the validated CNVs, 32 (64%) had more than 80% of their size (bp) validated by NGS data whereas eight CNVs (16%) had less than 20% of the size validated (Figure 3).

Figure 3. Distribution of the percentage of the CNV size (bp) validated.

5.4. A genome-wide map of CNV regions

In both populations each type of CNV showed overlap among samples and these CNVs were summarized across all individual into CNV regions (CNVRs). In the broiler population, a total of 3,608 CNVRs were identified (3,240 inherited, 166 not inherited and 202 de novo), covering 128.46, 4.81 and 39.24 Mb of sequence length, respectively. In the F₂ population, 3,684 CNVRs were identified (2,716 inherited, 547 not inherited and 421 de novo), covering 51.05, 35.04 and 113.4 Mb of sequence length, respectively. In Figure 4 the genome wide CNVR map of the two populations is shown, divided in inherited, not inherited and de novo.

Figure 4. Distribution of the inherited, not inherited and de novo CNVRs across the chicken genome in both populations.

In the broiler population, we identified 630 polymorphic CNVRs (occur in more than 1% of the population). From these CNVRs, 629 were inherited and only one was not inherited. In the F₂ population, all the 837 polymorphic CNVRs were inherited. These regions represented 4.91% (60.5 Mb) and 1.92% (23.7 Mb) of the chicken genome in the broiler and in the F₂ populations, respectively.

In the broiler population, the inherited CNVRs comprise 10.44% of the chicken genome whereas not inherited and de novo represented 0.39% and 3.18%, respectively. A 0.67 Mb region showed overlap with all three types of CNVRs in different individuals. The genomic regions in common between inherited and de novo CNVRs had a length of 8.56 Mb, between inherited and not inherited the size of the overlap was 3.83 Mb and between not inherited and de novo the overlap was 0.72 Mb (Figure 5).

In the F₂ population, the inherited CNVRs represented 4.14% of the chicken genome and not inherited and de novo represented 2.84% and 9.21%, respectively. A 1.16 Mb region overlapped with all three types of CNVRs in different individuals of the F₂ population. The size of the overlap between inherited and de novo CNVRs was 6.47 Mb, between inherited and not inherited it was 8.8 Mb and between not inherited and de novo the overlap was 5.13 Mb (Figure 5).

Figure 5. Overlap between the three types of CNVRs in broiler and F₂ population.

The overlap between the CNVRs detected in our study and those identified in previous studies was carried out on genome build galGal4.0. Therefore, map locations of CNVRs previously identified using genome build galGal3.0 were converted to Galgal4.0. The conversion rates for these studies were: 81.66% in Wang et al. (2010), 100% in Wang et al. (2012), 93.92% in Crooijmans et al.(2013), 72.61% in Luo et al. (2013), 99.67% in Tian et al.(2013), 85.2% in Fan et al. (2013) and, 99.2% in Abernathy et al. (2014).

In the broiler population we obtained 3,835 CNVRs (3,585 inherited, 177 not inherited and 73 de novo), with a total of 66.79 Mb on the galGal4.0 (48.36 Mb, 4.75 Mb and 13.67 Mb from each type of CNVRs, respectively). In the F_2 , we identified 4,084 CNVRs with a total of 412.46 Mb, being 3,075 CNVRs inherited (59.49 Mb), 534 not inherited (35.65 Mb) and 475 de novo (317.32 Mb). For the overlapping analysis with previous studies, we did not take into consideration the overlaps that exist between the three types of CNVRs in each population. However, it is important to remember that there are regions that have been defined as different types of CNVR in different families, as it was shown in galGal5.0 (Figure 5).

Among the CNVRs identified in the broiler population, 1,571 (40.96%) had been reported in previous studies (1,422 inherited, 85 not inherited, and 64 de novo), and the remaining 2,264 (59.04%) were classified as novel CNVRs. In the F_2 population, 2,091 (51.2%) had been reported by previous studies (1,323 inherited, 336 not inherited and 432 de novo), and the remaining 1,993 (48.8%) were classified as novel CNVRs. The detailed information of the comparison of results is presented in Table 2, and in the Annex B (Table 5 and 6), the results are shown by each type of CNVRs (inherited, not inherited and de novo).

Table 2. Comparisons between autosomal CNVRs identified in this study (two populations) with previous chicken studies.

Study	Number CNVRs	Total length (Mb)	CNVRs overlapping with present study (Broiler Population)				CNVRs overlapping with present study (F2 Population)			
			Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)	Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)
Gorla et al (2017)	1216	47,05	311	38.65	5.64	11.99	625	51.40	22.87	48.61
Yi et al (2015)	221	5,6	30	13.57	0.49	8.75	91	41.18	2.66	47.05
Wang et al (2010) ¹	2462	10,94	470	19.09	1.44	13.16	1165	47.32	4.94	45.16
Wang et al (2012) ¹	823	2,78	351	42.65	0.81	29.14	494	60.02	1.92	69.06
Crooijmans et al (2013) ¹	10411	47,2	1570	15.08	4.43	9.39	4516	43.38	20.94	44.36
Luo et al (2013) ¹	528	1,06	190	35.98	0.26	24.53	334	63.27	0.78	73.58
Tian et al (2013) ¹	2369	8,54	676	28.54	1.31	15.34	1146	48.37	4.07	47.66
Abernathy et al (2014) ¹	2376	5,18	501	21.09	0.87	16.80	928	39.06	2.31	44.59
Fan et al (2013) ¹	2301	2,48	402	17.47	0.51	20.56	1194	51.89	1.43	57.66
Yi et al (2014)	8487	94,93	970	11.43	6.59	6.94	3481	41.02	39.38	41.48
Yan et al (2015)	5010	18,62	534	10.66	2.36	12.67	2005	40.02	8.56	46.08

¹ CNVRs coordinates were converted from galGal3 to galGal4.

5.5. QTL analysis and gene content of CNVRs

We found that 1,574 (48.48%) inherited CNVRs in the broiler population showed overlap by at least one base with QTLs in the Chicken QTLdb (HU et al., 2013). In most of the cases, the CNVRs completely overlapped with a QTL region. A total of 305 polymorphic inherited CNVRs overlapped with 81 QTLs, corresponding to a variety of traits that including features related to feather (feather-crested head, feathered feet, feather pecking), wattles (length and weight), comb (weight), egg (weight, egg shell color and strength), body and breast muscle weight and abdominal fat percentage (Figure 6).

In the F_2 population, 1,327 (48.85%) inherited CNVRs overlapped at least one base with 83 QTLs, although again most of them overlap completely. Regarding the inherited polymorphic CNVRs, 435 showed overlap with 73 QTLs. The majority of the traits were related to egg (weight, number, shell weight and strength), feather (feather-crested head and feathered feet), body and breast muscle weight and intramuscular fat (Figure 6).

Figure 6. QTL-traits that overlapped with polymorphic CNVRs in the broiler and F_2 populations.

The F₂ population evaluated here was created previously for QTL mapping studies (AMBO et al., 2009; BARON et al., 2011; BOSCHIERO et al., 2013; CAMPOS et al., 2009; FELÍCIO et al., 2013; NONES et al., 2006, 2012). Therefore, we compared the inherited CNVRs that overlapped with 2,879 QTLs mapped in the same population. A total of 2,105 inherited CNVRs, from 2,716 identified in the F₂ population, showed overlap with 1,081 QTLs from these studies. The CNVR 205 (1:62,633,402-62,645,805) was the most frequent inherited CNVR that overlapped with a region that was mapped for different traits, such as, egg and ovary weight, body weight in different ages and intramuscular fat (BARON et al., 2011; BOSCHIERO et al., 2013; NONES et al., 2006, 2012). From a total of 837 inherited polymorphic CNVRs, 653 overlapped with 938 QTLs mapped in the F₂ population. The CNVR 205 was the most frequent polymorphic CNV also.

The intersection analysis performed between the chicken gene database (Ensembl galGal5) and our polymorphic inherited CNVRs in both populations allowed the identification of 1,429 Ensembl genes ID's in the broiler population, corresponding to 877 genes with an official gene ID. Among these genes, 815 (92.93%) were protein-coding genes, 42 (4.78%) were microRNA (miRNAs), 5 (0.57%) miscellaneousRNA (miscRNA), 10 (1.14%) small nuclear RNA (snRNA), 4 (0.45%) small nucleolar RNA (snoRNAs) and one (0.11%) ribosomal RNA (rRNA). In the F₂ population, we identified 967 Ensembl genes ID, corresponding to 550 genes with an official gene ID. Among these genes, 507 (92.18%) were protein-coding, 17 (3.09%) miRNA, 12 (2.18%) snRNA, nine (1.63%) miscRNA, three (0.54%) snoRNA and two (0.36%) rRNA.

The GO analysis revealed 130 GO terms in the broiler population and 84 in the F₂ population. However, only one of them in both populations was statistically significant after Benjamini correction. In both populations, the KEGG pathway analysis did not indicate any gene that was overrepresented after the Benjamini correction.

5.6. CNV association analysis with breast muscle traits

We performed the genome-wide association of the inherited CNVs with six breast traits. We used the genomic estimated breeding values (GEBV) for each trait as the quantitative phenotype and we constructed CNV segments from subsequent probes with similar genotype (DA SILVA et al., 2016). CNV segments represent the fine-grained CNV

genotypes whereas CNVRs are coarse-grained and may join CNVs containing different CN states (see item 4. Material and Methods for details).

The 21,077 inherited CNVs were concatenated in 461 CNV segments, that were constructed from 1,897 probes that deviate from 2n state in $\geq 5\%$ of the population. The number of probes in each segment ranged from 1 to 8, with 6.57 probes on average.

Considering the Bonferroni-corrected p-values for multiple testing and statistically significant p-values ≤ 0.01 , we found two CNV segments significantly associated with breast weight. One CNV segment (CNVseg 52) was inside the inherited CNVR 287, and the other CNV segment (CNVseg 300) encompassed six inherited CNVRs (CNVR 2265, 2266, 2267, 2268, 2269 and 2270). We found that the CNVseg 300 mentioned above was also associated with breast weight yield (Figure 7).

Figure 7. Manhattan plots of associated CNVs state for BW and BY traits. Negative log 10-transformed P values from a genome-wide scan were plotted against genomic coordinates on 28 autosomal chromosomes.

We identified six significantly associated CNV segments for breast meat weight. Two segments (CNVseg 254 and 246) were inside the inherited CNVR 1765, three segments (CNVseg 297, 298 and 299) were inside the inherited CNVR 2246 and the last one is the CNVseg 300. Regarding the breast meat yield, a total of 18 CNV segments were significantly associated. Among them, 13 CNV segments (CNVseg 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177) were inside the inherited CNVR 1180, one CNV segment (CNVseg 178) was partly inside the inherited CNVRs 1180 and 1181. The CNVsegs 245 and 246 were inside the inherited CNVR 1765. The CNVseg 164 encompassed seven CNVRs (CNVR 1174, 1175, 1176, 1177, 1178, 1179, 1180), and the last one associated is the CNVseg 300, mentioned before (Figure 8).

Figure 8. Manhattan plots of associated CNVs state for BMW and BMY traits. Negative log 10-transformed P values from a genome-wide scan were plotted against genomic coordinates on 28 autosomal chromosomes.

We found four CNV segments significantly associated with breast file weight. The CNVseg 342 and 343 were inside the inherited CNVR 2504, the CNVseg 341 was partly inside two CNVRs (2503 and 2504) and the CNVseg 300. Regarding the breast file yield, the CNVseg 300 was again associated and the CNVseg 421, that was inside the inherited CNVR 3027, was associated (Figure 9). More information about the position of CNVsegs and CNVRs associated with each trait are available in the (ANNEX C).

Figure 9. Manhattan plots of associated CNVs state for BFW and BFY traits. Negative log 10-transformed P values from a genome-wide scan were plotted against genomic coordinates on 28 autosomal chromosomes.

The CNVseg 300, that encompassed six inherited CNVRs, was significantly associated with all traits. Figure 10 shows the CN state distribution of each individual in all traits studied. We investigated a genomic window 500 kb up and downstream from the boundaries of inherited CNVRs significantly associated.

Because it has been reported that the expression of genes can be affected for CNVRs around this distance (HENRICHSEN et al., 2009). We found a total of 181 protein-coding genes to be located in the vicinity of the segments (Table 3).

Figure 10. Distribution of CNV state for the individuals in one CNV segment (CNVseg 300) significantly associated for all six traits. Each dot represents an animal in the corresponding CNV state (1 or 2n) on the x-axis and the GEBV of each trait (BW-breast weight, BY- breast yield, BMW-breast meat weight, BMY-breast meat yield, BFW-breast filet weight and BFY- breast filet yield) on the y-axis. The legend on the top right displays the color code for the CN states.

Table 3. Protein-coding genes within and 500 kb up and downstream from the boundaries of inherited CNVRs significantly associated with all studied traits.

CNVRs Significantly Associated	CNVRs Position	Window Position	Protein-coding Genes	Genes
CNVR 1765	4:43906643-43913344	4:43406643-44413344	10	GALNTL6, GALNT7, HMGB2, SAP30, SCRG1, HAND2, FBXO8, CEP44, HPGD, GLRA3
CNVR 2246	6:30235695-30254967	6:29735695-30754967	17	EIF3A, SEC23IP, PLPP4, PRDX3, TIAL1, BAG3, GRK5, FAM45A, RGS10, INPP5F, CACUL1, SFXN4, NANOS1, WDR11, MCMBP
CNVR 2265	7:2106711-2118488	7:1606711-2618488	4	ZNF804A, DUSP19, NCKAP1, FRZB
CNVR 2266	7:2904174-2910410	7:2404174-3410410	4	ACADL, MYL1, LANCL1, ERBB4
CNVR 2267	7:3426693-3447393	7:2926693-3947393	3	ERBB4, IKZF2, SPAG16
CNVR 2268	7:3495801-3510856	7:2995801-4010856	3	ERBB4, IKZF2, SPAG16
CNVR 2269	7:3854947-3861248	7:3354947-4361248	6	ERBB4, IKZF2, SPAG16, VWC2L, BARD1, ABCA12
CNVR 2270	7:4551322-4593640	7:4051322-5093640	17	SPAG16, VWC2L, BARD1, ABCA12, ATIC, FN1, MREG, IFNL3, UBE2F, RAMP1, RAB17, PRLH, MLPH, COL6A3, COPS8, ACKR3, IQCA1
CNVR 1174	2:125166870-125178467	2:124666870-125678467	12	RIPK2, DECR1, C8orf88, SLC26A7, TMEM55A, OTUD6B, CALB1, RUNX1T1, OSGIN2, NBN, TMEM64, NECAB1
CNVR 1175	2:125835210-125844798	2:125335210-126344798	2	SLC26A7, RUNX1T1
CNVR 1176	2:126090326-126118538	2:125590326-126618538	5	PDP1, CDH17, RBM12B, GEM, RUNX1T1
CNVR 1177	2:126632301-126644089	2:126132301-127144089	12	TP53INP1, PDP1, CDH17, ESRP1, RBM12B, INTS8, GEM, DPY19L4, CCNE2, NDUFAF6, PLEKHF2, C8orf37
CNVR 1178	2:126936355-12696200	2:126436355-13196200	0	
CNVR 1179	2:127598517-127613971	2:127098517-128113971	10	RPL30, UQCRB, CPQ, MTERF3, RIDA, POP1, MTDH, PTDSS1, LAPTM4B, SDC2
CNVR 1180	2:127635080-127761687	2:127135080-128261687	13	RPL30, UQCRB, CPQ, MTERF3, RIDA, POP1, MTDH, NIPAL2, KCNS2, STK3, PTDSS1, LAPTM4B, SDC2
CNVR 1181	2:127933580-127945741	2:127433580-128445741	11	RPL30, CPQ, RIDA, POP1, MTDH, NIPAL2, KCNS2, STK3, LAPTM4B, SDC2, OSR2
CNVR 2266	7:2904174-2910410	7:2404174-3410410	4	ACADL, MYL1, LANCL1, ERBB4
CNVR 2503	9:15784391-15790362	9:15284391-16290362	36	ARMC9, B3GNT7, NCL, NMUR1, PTMA, PDE6D, COPS7B, DIS3L2, SLC12A9, PRSS56, CHRND, CHRNG, EIF4E2, CAPN10, ABCC5, EIF2B5, DVL3, AP2M1, ABCF3, VWA5B2, CAMK2N2, ECE2, PSMD2, EIF4G1, FAM131A, CLCN2, CHRD, POLR2H, EPHB3, SENP5, AHSB, FETUB, KNG1, EIF4A2, RFC4, MCF2L2
CNVR 2504	9:15809112-15820640	9:15309112-16320640	35	B3GNT7, NCL, NMUR1, PTMA, PDE6D, COPS7B, DIS3L2, SLC12A9, PRSS56, CHRND, CHRNG, EIF4E2, CAPN10, ABCC5, EIF2B5, DVL3, AP2M1, ABCF3, VWA5B2, CAMK2N2, ECE2, PSMD2, EIF4G1, FAM131A, CLCN2, CHRD, POLR2H, EPHB3, SENP5, AHSB, FETUB, KNG1, EIF4A2, RFC4, MCF2L2
CNVR 3027	19:6823309-6851703	19:6323309-7351703	34	ABR, TIMM22, SPECC1, ADORA2B, ZSWIM7, TTC19, NCOR1, TRPV2, EMC6, TAX1BP3, CTNS, TRPV1, TRPV3, ASPA, AKAP10, ULK2, SLC47A2, MRM3, GLOD4, GEMIN4, FAM57A, VPS53, RFLNB, RPH3AL, TEX14, RAD51C, PPM1E, TRIM37, GPD1, YPEL2, DHX40, CLTC, PTRH2, VMP1
CNVR 287	1:77752486-77761509	1:77252486-78261509	22	GNB3, CDCA3, USP5, LRRC23, ENO2, ATN1, C12orf57, PTPN6, PHB2, LPCAT3, C1S, RBP5, CLSTN3, PEX5, EPHA1, ZYX, FAM131B, CLCN1, CASP2, RAP1GAP, GSTK1, EPHB6

6. DISCUSSION

6.1. Copy number variation in the chicken genome

CNVs play an essential role in genome evolution and have been implicated in a variety of genetic diseases. The consequences of CNVs in different species are becoming increasingly understood due to the availability of complete and good quality genome sequence assemblies, genome resequencing data, and high-throughput genotyping data, and powerful bioinformatics tools. However, the inheritance of CNVs is rarely studied. Our study is the first, in chicken, with family information (father, mother and offspring) allowing the identification of both inherited and de novo CNVs.

Our discoveries not only complement the current CNV databases but also emphasized the proportion of inherited CNVs and which regions and genes harbor these events in the chicken genome. We analyzed two different Brazilian chicken populations: (i) one composed of broilers that were generated by the expansion of a paternal line, maintained under multi-trait selection for body weight, feed conversion, reduced abdominal fat, carcass and cuts yield, fertility, hatchability of fertile eggs and chick viability and (ii) a broiler-layer F_2 population that was created for QTL mapping studies.

The known family structure allowed us to use the joint function in PennCNV software, which identifies CNVs from allele intensities and frequencies (i.e. LRR and BAF). This function provides the CNVs that are inherited from both or from only one parental of the offspring, and the one which is only present in the offspring (de novo). These results provide us more confidence for CNV inference, besides give us the same start and end point in the inherited CNVs (WANG et al., 2007).

With regard to de novo CNVs, the number initially identified in both populations was considered above the expected. Although some CNVs are inherited whereas others spontaneously arise de novo (THAPAR; COOPER, 2013), it is insufficient to use CNV calls to identify candidate de novo events due the relatively high false-negative rates (ITSARA et al., 2010). Because of this, we proposed an index for inheritance ambiguity based on results from a PennCNV function which detects de novo confidences (<http://penncnv.openbioinformatics.org/en/latest/user-guide/denovo/>). We used an index that considers the least ambiguous meiotic origin of this de novo CNV, because it is known

that one of the ways to arise de novo mutations is by germline events (ACUNA-HIDALGO; VELTMAN; HOISCHEN, 2016) and it is quite unlikely that the same de novo mutation arises from the paternal and maternal origin.

After the filtering, a total of 220 (4.8% from the de novo identified initially) remained as de novo CNV in the broiler population and 430 (14.43%) in the F₂ population. Although it has been already proven that de novo CNVs are enriched in individuals with complex neurodevelopmental disorders, such as intellectual disability, epilepsy, autism or schizophrenia (COE; GIRIRAJAN; EICHLER, 2012; IYER; GIRIRAJAN, 2015), association studies require a large cohort to obtain statistical power, therefore, it is a challenge with de novo CNVs because they are not frequent in a population (NOWAKOWSKA, 2017).

The validation of CNVs was performed using the CNVs identified by NGS data from the same animals. The apparent low overlap between CNVs detected by genotyping and sequencing data might be due to their different advantages and disadvantages about the size distribution and resolution for confident CNV calls in each technology (ZHAN et al., 2011). Array-based platforms have relatively limited power to detect small CNVs and they do not have a reliable break-point prediction. On the other hand, NGS-based methods are able to detect smaller CNVs with the efficiency of breakpoint prediction at single base resolution (LI; OLIVIER, 2013). CNVnator, in particular, is capable to detect small CNVs, break point position and copy number estimation (YAO et al., 2017).

Most of the validated CNVs (50 CNVs) had high validation percentage as 32 CNVs (61.5%) had between 80 and 100% of size validation (Figure 3). Beside this, a large proportion was validated by the same CNV state. The CNVs validated by a divergent state were more common in CNVs that were validated by more than one sequenced CNV. This is explained because the CNVnator detected smaller CNV better than PennCNV in the genotyping data. Therefore, a CNV identified by PennCNV is overlapped by different small CNVs detected by CNVnator.

We observed that CNVRs gain were more abundant than losses in the three types of CNVRs in both populations, in agreement with previous studies (JIA et al., 2013; WANG et al., 2012; YI et al., 2015). Conrad et al. (2006) reported that duplicated events are more likely to be maintained than deletions, because deletion CNVRs were relatively gene-poor and therefore these genes were subject to purifying selection. However, it is important to note that, although deletion polymorphisms are not directly observed in most current gene

mapping studies, they may have an important role in the genetics of complex traits (CONRAD et al., 2006).

With regard to the length of the CNVRs the populations are divergent. Although more inherited CNVRs were identified in both populations, those covered a larger region (128.46 Mb) from the genome in the broiler population than in the F₂ population, in which the region covered by de novo CNVRs was larger (113.4 Mb) than the other CNVRs (Figure 4). This might be explained by the higher genetic variability of the F₂ population relative to the broiler, allowing the emergence of more de novo CNVs.

We compared all CNVRs found on autosomes with those found in 11 previous studies. In the broiler population, 40.96% of CNVRs were overlapped by CNVs previously described (ABERNATHY et al., 2014; CROOIJMANS et al., 2013; FAN et al., 2013; LUO et al., 2013; TIAN et al., 2013; WANG et al., 2010, 2012; YAN et al., 2015; YI et al., 2014, 2015), whereas in the F₂ population, 51.2% were overlapped. Regarding the overlap with each of the previous studies, the broiler population had lower overlaps than F₂ (Table 2). This difference might be explained due to the broiler is a pure line while the F₂ is a layer-broiler cross. Therefore, the F₂ population is more likely to be in overlap with the populations of the different studies.

In both populations there was substantial number of novel CNVRs detected, which is in agreement with previous studies that compared the identified CNVRs with the other datasets (ABERNATHY et al., 2014; CROOIJMANS et al., 2013; FAN et al., 2013; GORLA et al., 2017; LUO et al., 2013; TIAN et al., 2013; WANG et al., 2010; YI et al., 2014, 2015). Therefore, no single approach is able to capture all CNVs and the different platforms may be complementary.

We found that roughly half of our CNVRs, in both populations, were overlapped with QTLs involved in economically important traits like body weight, breast muscle, abdominal fat, antibody responses and egg phenotypes. These results might give hint to CNVs which are associated with economic traits and are potentially important for future studies.

Regarding the information available in public databases, the genes annotated in the inherited and polymorphic CNVRs in both populations are not directly involved in any regulation of the economic and important traits. Because these CNVRs are inherited and polymorphic (> 1% population), means that they may have relatively remote origins or neutral evolutionary histories and seem to be subsequently inter-crossed in multiple breeds

(CLOP; VIDAL; AMILLS, 2012; YI et al., 2015). In addition, this result may be due to the fact that the chicken genome is not well annotated.

6.2. GWAS of inherited CNV state with breast muscle traits

We performed the genome-wide association analysis between CNV segments, with highly similar probes, and six breast muscle traits (BW, BY, BMW, BMY, BFW, BFY). Among the resulting significantly associated segments, we identified some CNV segments that encompassed more than one CNVR. These events occurred in the CNVseg 164 (CNVRs 1174, 1175, 1176, 1177, 1178, 1179 and 1180) that was associated with BY, CNVseg 178 (CNVRs 1180 and 1181) that was associated with BMY, CNVseg 300 (CNVRs 2265, 2266, 2267, 2268, 2269 and 2270) that was associated with all traits, and CNVseg 341 (CNVRs 2503 and 2504) that was associated with BFW. Linkage disequilibrium among the SNPs present in the CNV may explain these observations.

The CNVRs 2265, 2266, 2267, 2268, 2269 and 2270, that represent the CNVseg 300, were significantly associated with traits (Annex C). Venturini and collaborators. (2014) estimated in the same population the genetic correlation between breast fillet weight and breast meat weight was high and positive (0.85).

The CNVRs 2265 and 2266 were loss events and present in 3.14 and 1.8 % of population, respectively. The CNVRs 2267, 2268 and 2270 were gain events and present in 0.84, 0.48 and 8.23 % respectively. The CNVR 2269 was both events (gain and loss) and present in 0.48 % of population. These CNVRs overlapped with 26 protein-coding genes (*ZNF804A*, *DUSP19*, *NCKAP1*, *FRZB*, *ACADL*, *MYL1*, *LANCL1*, *ERBB4*, *IKZF2*, *SPAG16*, *VWC2L*, *BARD1*, *ABCA12*, *ATIC*, *FN1*, *MREG*, *IFNL3*, *UBE2F*, *RAMP1*, *RAB17*, *PRLH*, *MLPH*, *COL6A3*, *COPS8*, *ACKR3*, *IQCA1*) (Table 3). Moreover, the CNVR 2265 overlapped with glucose level QTL (DEMEURE et al., 2013), and the CNVRs 2269 and 2270 overlapped with comb weight QTL (SUN et al., 2015).

Among these genes, the gene *MYL1* (*Myosin Light Chain 1*) is expressed in early developmental stages in the mouse embryonic skeletal muscle (SCHIAFFINO et al., 2015). The splicing variants of the *MYL1* gene expressed in the fast skeletal muscle fibers are the MyLC1f and MyLC3f (SCHIAFFINO; REGGIANI, 2011). CNVRs presented in this gene may have an important role because alterations in protein expression patterns can reflected in

molecular and cellular modifications in skeletal muscle tissues (GELFI; VASSO; CERRETELLI, 2011).

The genes *MLPH* (*Melanophilin*), *PRLH* (*Prolactin Releasing Hormone*) and *RAB17* (*Member RAS Oncogene Family*) were associated with the lavender phenotype in chickens, due to two inversions and one deletion identified near them. This phenotype is caused by a dilution of both eumelanin and phaeomelanin in feathers that produces a blue-grey color on a wild-type feather chicken. Moreover, the chickens with lavender phenotype showed lighter body weight, lower body temperature and increased feed consumption and residual feed intake than wild-type chickens. This may indicate that complex mutation in these genes is altering the regulation of homeothermy and the metabolism (BED'HOM et al., 2012).

The *CLCN1* (*Chloride Voltage-Gated Channel 1*) protein-coding gene overlapped with the CNVR 287, a loss event, present in 3.87% of population and that was associated with breast weight (Figure 7). Mutations identified in exons 4 and 5 of the *CLCN1* gene in humans annulated the Myotonic congenital disease expression. This disease is characterized by damage muscle relaxation after forceful contraction (myotonic) (BRUGNONI et al., 2013). The CNVR 287 overlapped with body temperature QTL (VAN GOOR et al., 2015)

The *INPP5F* (*Inositol Polyphosphate-5-Phosphatase*) protein-coding gene overlapped with the CNVR 2246, a gain event, present in 13.92% of population, and that was associated with breast meat weight (Figure 8). In humans, the increase in the *INPP5F* gene expression could avoid the hypertrophy in the adult heart (TRIVEDI et al., 2007). In pigs, polymorphism in this gene had significant differences in average daily gain in Yorkshire pigs (ZHOU et al., 2009). The CNVR 2246 overlapped with breast muscle weight QTL (MIGNON-GRASTEAU et al., 2015).

The CNVR 1765, a loss event and present 3.38% of population, overlapped with ileum weight QTL (MIGNON-GRASTEAU et al., 2015). The CNVR 1174 (gain, frequency of 4.23%), CNVRs 1175, 1176, 1178, 1179 (gain, frequency of 0.24%), CNVR 1177 (gain, frequency of 0.6%), CNVR 1180 (both events, frequency of 20.9%) and CNVR 1181 (gain, frequency of 3.51%) overlapped with feather-crested head QTL (SUN et al., 2015). The CNVR 3027 (gain, frequency of 6.9%) overlapped with gizzard weight QTL (MIGNON-GRASTEAU et al., 2015).

Therefore, we identified several CNV regions significantly associated with six breast muscle traits. This result is a good starting point for future studies on structural variations related with breast muscle and skeletal muscle physiology in a broiler population.

7. CONCLUSION

In conclusion, we performed the first genome-wide detection, characterization and functional annotation of inherited and de novo CNVs in the new version of the chicken genome (Galgal5.0) in two different experimental chicken populations. Our results allow us to understand the rate and distribution of CNVRs that are inherited, not inherited and de novo in the chicken genome. Besides, allow a much better insight in the degree of false positives and negatives in CNV studies based of Affymetrix data.

Our results allow us to compare the CNVs regions with previous studies in chicken and to investigate the overlap of inherited CNVRs with QTLs and genes related to metabolic pathways of interest.

In addition, the genome-wide association analysis was the first performed on broiler population with breast muscle traits, that are important characteristics for poultry production. The GWAS results allow us to understand the probably relationship between some genes and CNVRs that were significantly associated with breast muscle traits.

REFERENCES

ABERNATHY, J. et al. Copy number variation in Fayoumi and Leghorn chickens analyzed using array comparative genomic hybridization. **Animal Genetics**, v. 45, n. 3, p. 400–411, 2014.

ABPA - Brazilian Association of Animal Protein - 2017 Anual Report. . [s.l: s.n.]. Disponível em: http://abpa-br.com.br/storage/files/final_abpa_relatorio_anual_2017_ingles_web.pdf.

ABYZOV, A. et al. CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. **Genome Research**, v. 21, p. 974–984, 2011.

ACUNA-HIDALGO, R.; VELTMAN, J. A.; HOISCHEN, A. New insights into the generation and role of de novo mutations in health and disease. **Genome Biology**, v. 17, n. 1, p. 241:260, 2016.

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

BARON, E. E. et al. QTL for percentage of carcass and carcass parts in a broiler x layer cross. **Animal Genetics**, v. 42, n. 2, p. 117–124, 2011.

BECKMANN, J. S.; ESTIVILL, X.; ANTONARAKIS, S. E. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. **Nature reviews. Genetics**, v. 8, n. 8, p. 639–646, 2007.

BED'HOM, B. et al. The lavender plumage colour in Japanese quail is associated with a complex mutation in the region of MLPH that is related to differences in growth, feed consumption and body temperature. **BMC Genomics**, v. 13, p. 442–452, 2012.

BOSCHIERO, C. et al. Association of IGF1 and KDM5A polymorphisms with performance, fatness and carcass traits in chickens. **Journal of Applied Genetics**, v. 54, n. 1, p. 103–112, 2013.

BRUGNONI, R. et al. A large cohort of myotonia congenita probands: Novel mutations and a high-frequency mutation region in exons 4 and 5 of the CLCN1 gene. **Journal of Human Genetics**, v. 58, n. 9, p. 581–587, 2013.

CAMPOS, R. L. R. et al. Quantitative trait loci associated with fatness in a broiler-layer cross. **Animal Genetics**, v. 40, n. 5, p. 729–736, 2009.

CARVALHO, C. M. B.; LUPSKI, J. R. Mechanisms underlying structural variant formation in genomic disorders. **Nature Reviews Genetics**, v. 17, p. 224–238, 2016.

CARVALHO, C. M. B.; ZHANG, F.; LUPSKI, J. R. Genomic disorders: A window into human gene and genome evolution. **Proceedings of the National Academy of Sciences**, v. 107, n. suppl_1, p. 1765–1771, 2010.

CHEN, C. et al. 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.

CLOP, A.; VIDAL, O.; AMILLS, M. Copy number variation in the genomes of domestic animals. **Animal Genetics**, v. 43, p. 503–517, 2012.

COE, B. P.; GIRIRAJAN, S.; EICHLER, E. E. The genetic variability and commonality of neurodevelopmental disease. **American Journal of Medical Genetics, Part C: Seminars in Medical Genetics**, v. 160 C, n. 2, p. 118–129, 2012.

COLELLA, S. et al. QuantiSNP: An objective bayes hidden-markov model to detect and accurately map copy number variation using SNP genotyping data. **Nucleic Acids Research**, 2007.

CONRAD, D. F. et al. A high-resolution survey of deletion polymorphism in the human genome. **Nature Genetics**, v. 38, n. 1, p. 75–81, 2006.

CROOIJMANS, R. P. M. A. et al. Large scale variation in DNA copy number in chicken breeds. **BMC Genomics**, v. 14, p. 398, 2013.

DA SILVA, V. H. et al. Genome-wide detection of CNVs and their association with meat tenderness in Nelore cattle. **PLoS ONE**, v. 11, n. 6, p. 1–22, 2016.

DE ALMEIDA SANTANA, M. H. 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**, v. 57, n. 4, p. 495–504, nov. 2016.

DEMEURE, O. et al. Genome-wide interval mapping using SNPs identifies new QTL for growth, body composition and several physiological variables in an F2 intercross between fat and lean chicken lines. **Genetics Selection Evolution**, v. 45, n. 1, p. 36, 2013.

DISKIN, S. J. et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. **Nucleic Acids Research**, v. 36, n. 19, p. e126, 2008.

DONG, K. et al. Copy number variation detection using SNP genotyping arrays in three Chinese pig breeds. **Animal Genetics**, v. 46, n. 2, p. 101–109, 2015.

DORSHORST, B.; OKIMOTO, R.; ASHWELL, C. Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the silkie chicken. **Journal of Heredity**, 2010.

DUPUIS, M. C. et al. Detection of copy number variants in the horse genome and examination of their association with recurrent laryngeal neuropathy. **Animal Genetics**, 2013.

DURÁN AGUILAR, M. et al. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. **Journal of Animal Breeding and Genetics**, v. 134, n. 1, p. 49–59, fev. 2017.

ELFERINK, M. G. et al. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. **BMC Genomics**, v. 9, n. 391, p. 1–9, 2008.

ELLEGREN, H. The avian genome uncovered. **Trends in Ecology and Evolution**, v. 20, n. 4, p. 180–186, 2005.

FAN, W. L. et al. Genome-wide patterns of genetic variation in two domestic chickens. **Genome Biology and Evolution**, v. 5, n. 7, p. 1376–1392, 2013.

FELÍCIO, A. M. et al. Polymorphisms in FGFBP1 and FGFBP2 genes associated with carcass and meat quality traits in chickens. **Genetics and**, v. 12, n. 1, p. 208–222, 2013.

FLINK, L. G. et al. Establishing the validity of domestication genes using DNA from ancient chickens. **PNAS**, v. 111, n. 17, p. 6184–6189, 2014.

GARRICK, D. J.; FERNANDO, R. L. Detection Study (GWAS) Using Genomic Prediction Methodology. In: GONDRO, C.; VAN DER WERF, J.; HAYES, B. (Eds.). . **Genome-Wide Association Studies and Genomic Prediction**. Humana Pre ed. [s.l.: s.n.]. v. 1019p. 275–298.

GELFI, C.; VASSO, M.; CERRETELLI, P. Diversity of human skeletal muscle in health and disease: Contribution of proteomics. **Journal of Proteomics**, v. 74, n. 6, p. 774–795, 2011.

GORLA, E. et al. Genomic variability in Mexican chicken population using copy number variants. **BMC Genetics**, v. 18, n. 1, p. 61, 2017.

GRIFFIN, D. K. et al. Whole genome comparative studies between chicken and turkey and their implications for avian genome evolution. **BMC Genomics**, 2008.

GROENEN, M. A. et al. The development and characterization of a 60K SNP chip for chicken. **BMC Genomics**, v. 12, p. 274:283, 2011.

GROENEN, M. A. M. et al. A consensus linkage map of the chicken genome. **Genome Research**, v. 10, n. 1, p. 137–147, 2000.

GROENEN, M. A. M. et al. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. **Genome Research**, v. 19, p. 510–519, 2009.

GU, X. et al. Genome-wide association study of body weight in chicken F2 resource population. **PLoS ONE**, v. 6, n. 7, 2011.

GUNNARSSON, U. et al. The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of SOX10. **Pigment Cell and Melanoma Research**, 2011.

HENRICHSEN, C. N. et al. Segmental copy number variation shapes tissue transcriptomes. **Nature Genetics**, v. 41, n. 4, p. 424–429, 2009.

HILLIER, L. W. et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution International Chicken Genome Sequencing Consortium*. **Nature**, v. 432, n. 7918, p. 695–716, 2004.

HU, Z. L. et al. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. **Nucleic Acids Research**, v. 41, n. D1, p. 871–879, 2013.

HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. **Nature Protocols**, v. 4, n. 1, p. 44–57, 2008.

HUANG, D. W.; SHERMAN, B. T.; LEMPICKI, R. A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. **Nucleic Acids Research**, v. 37, n. 1, p. 1–13, 2009.

HUANG, X. et al. High-throughput genotyping by whole-genome resequencing. p. 1068–1076, 2009.

ITSARA, A. et al. De novo rates and selection of large copy number variation De novo rates and selection of large copy number variation. **Genome Research**, v. 20, p. 1469–1481, 2010.

IYER, J.; GIRIRAJAN, S. Gene discovery and functional assessment of rare copy-number variants in neurodevelopmental disorders. **Briefings in Functional Genomics**, v. 14, n. 5, p. 315–328, 2015.

JIA, X. et al. Copy number variations identified in the chicken using a 60K SNP BeadChip. **Animal Genetics**, v. 44, n. 3, p. 276–284, 2013.

KERSTENS, H. H. et al. Structural variation in the chicken genome identified by paired-end next-generation DNA sequencing of reduced representation libraries. **BMC genomics**, v. 12, n. 1, p. 94, 2011.

KINGERY, S. E. et al. Gene Copy-Number Variations (CNVs) and Protein Levels of Complement C4A and C4B as Novel Biomarkers for Partial Disease Remissions in New-Onset Type 1 Diabetes Patients Suzanne. **Pediatra Diabetes**, v. 13, n. 5, p. 408–418, 2012.

KOBOLDT, D. C. et al. Massively Parallel Sequencing Approaches for Characterization of Structural Variation. **Methods Mol Biol**, v. 838, p. 369–384, 2012.

KOHLER, L. H. et al. FGFR1 expression and gene copy numbers in human lung cancer. **Virchows Archiv**, v. 461, n. 1, p. 49–57, 2012.

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

KUMAR, S.; BANKS, T. W.; CLOUTIER, S. SNP discovery through next-generation sequencing and its applications. **International Journal of Plant Genomics**, v. 2012, p. 15, 2012.

KWOK, P. Y.; GU, Z. Single nucleotide polymorphism libraries: Why and how are we building them. molecular medicine today. **Molecular Medicine Today**, v. 5, p. 538–543, 1999.

LAWRENCE, M. et al. Software for Computing and Annotating Genomic Ranges. **PLoS Comput Biol**, v. 9, n. 8, 2013.

LEE, H. E. et al. MET in gastric carcinomas: comparison between protein expression and gene copy number and impact on clinical outcome. **British Journal of Cancer**, v. 107, n. 2, p. 325–333, 2012.

LEE, M. O. et al. 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**, v. 113, n. 48, p. 13815–13820, nov. 2016.

LEPPA, V. M. et al. Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. **The American Journal of Human Genetics**, v. 99, p. 540–554, 2016.

LEVY, D. et al. Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. **Neuron**, v. 70, p. 886–897, 2011.

LI, W.; OLIVIER, M. Current analysis platforms and methods for detecting copy number variation. **Physiology Genomics Review**, v. 45, p. 1–16, 2013.

LIU, J. et al. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. **BMC genomics**, v. 14, n. 1, p. 229, 2013.

LOCKE, D. P. et al. Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome. **The American Journal of Human Genetics Am. J. Hum. Genet**, v. 7979, p. 275–290, 2006.

LONG, Y. et al. A genome-wide association study of copy number variations with umbilical hernia in swine. **Animal Genetics**, v. 47, n. 3, p. 298–305, jun. 2016.

LUO, J. et al. Genome-Wide Copy Number Variant Analysis in Inbred Chickens Lines With Different Susceptibility to Marek's Disease. **G3 Genes\Genomes\Genetics**, v. 3, n. 2, p. 217–223, 2013.

MCCARTHY, S. S. E. S. et al. Microduplications of 16p11. 2 are associated with schizophrenia. **Nature genetics**, v. 41, n. 11, p. 1223–1227, 2009.

MCDANELD, T. G. et al. Deletion on chromosome 5 associated with decreased reproductive efficiency in female cattle. **Journal of Animal Science**, v. 92, n. 4, p. 1378–1384, abr. 2014.

MIGNON-GRASTEAU, S. et al. Detection of QTL controlling feed efficiency and excretion in chickens fed a wheat-based diet. **Genetics Selection Evolution**, v. 47, n. 1, p. 74, 2015.

NGUYEN, D. et al. Reduced purifying selection prevails over positive selection in human copy number variant evolution. **Genome Research**, v. 18, p. 1711–1723, 2008.

NONES, K. et al. Mapping QTLs on chicken chromosome 1 for performance and carcass traits in a broiler x layer cross. **Animal Genetics**, v. 37, p. 95–100, 2006.

NONES, K. et al. Quantitative trait loci associated with chemical composition of the chicken carcass. **Animal Genetics**, v. 43, n. 5, p. 570–576, 2012.

NOWAKOWSKA, B. Clinical interpretation of copy number variants in the human genome. **Journal of Applied Genetics**, v. 58, p. 449–457, 2017.

PERRY, R. L. S.; RUDNICKI, M. A. MOLECULAR MECHANISMS REGULATING MYOGENIC DETERMINATION AND DIFFERENTIATION Robert. **Frontiers in Bioscience**, v. 5, p. D750–67, 2000.

PINKEL, D. et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. **Nature genetics**, v. 20, p. 207–211, 1998.

PURCELL, S. et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. **The American Journal of Human Genetics**, v. 81, n. 3, p. 559–575, 2007.

QANBARI, S. et al. Parallel Selection Revealed by Population Sequencing in Chicken. **Genome Biol Evol**, v. 7, n. 12, p. 3299–3306, 2015.

RAO, Y. S. et al. Copy number variation identification and analysis of the chicken genome using a 60K SNP BeadChip. **Poultry science**, v. 95, n. 8, p. 1750–6, 2016.

REDON, R. et al. Global variation in copy number in the human genome. **Nature**, v. 444, n. 7118, p. 444–454, 2006.

ROVELET-LECRUX, A. et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. **Nature Genetics**, v. 38, n. 1, p. 24–26, 2006.

SCHERER, S. W. et al. Challenges and standards in integrating surveys of structural variation. **Nature Genetics**, v. 39, n. 7s, p. S7–S15, 2007.

SCHIAFFINO, S. et al. Developmental myosins: expression patterns and functional significance. **Skeletal Muscle**, v. 5, n. 1, p. 22, 2015.

SCHIAFFINO, S.; REGGIANI, C. Fiber Types in Mammalian Skeletal Muscles. **Physiological Reviews**, v. 91, n. 4, p. 1447–1531, 2011.

SCHIAVO, G. et al. Copy number variants in Italian Large White pigs detected using high-density single nucleotide polymorphisms and their association with back fat thickness. **Animal Genetics**, v. 45, n. 5, p. 745–749, out. 2014.

SCHRIDER, D. R.; HAHN, M. W. Gene copy-number polymorphism in nature. **Proceedings of the Royal Society B: Biological Sciences**, v. 277, n. 1698, p. 3213–3221, 2010.

SKINNER, B. M. et al. Comparative genomics in chicken and Pekin duck using FISH mapping and microarray analysis. **BMC genomics**, v. 10, p. 357, 2009.

STANKIEWICZ, P.; LUPSKI, J. R. Structural Variation in the Human Genome and its Role in Disease. **Annual Review of Medicine**, v. 61, n. 1, p. 437–455, 2010.

STERN, C. D. The chick: A great model system becomes even greater. **Developmental Cell**, v. 8, n. 1, p. 9–17, 2005.

SUN, Y. et al. Genome-Wide Linkage Analysis Identifies Loci for Physical Appearance Traits in Chickens. **G3 Genes\Genomes\Genetics**, v. 5, n. 10, p. 2037–2041, 2015.

TEO, S. M. et al. Statistical challenges associated with detecting copy number variations with next-generation sequencing. **Bioinformatics**, v. 28, n. 21, p. 2711–2718, 2012.

THAPAR, A.; COOPER, M. Copy number variation: What is it and what has it told us about child psychiatric disorders? **Journal of the American Academy of Child and Adolescent Psychiatry**, v. 52, n. 8, p. 772–774, 2013.

TIAN, M. et al. Copy number variants in locally raised Chinese chicken genomes determined using array comparative genomic hybridization. **BMC genomics**, v. 14, n. 1, p. 262, 2013.

TRIVEDI, C. et al. Hdac2 regulates the cardiac hypertrophic response by modulating Gsk3 beta activity. **Nature Medicine**, v. 13, n. 3, p. 324–331, 2007.

VALSESIA, A. et al. Identification and validation of copy number variants using SNP genotyping arrays from a large clinical cohort. **BMC Genomics**, v. 13, n. 1, p. 241, 2012.

VAN GOOR, A. et al. Identification of quantitative trait loci for body temperature, body weight, breast yield, and digestibility in an advanced intercross line of chickens under heat stress. **Genetics Selection Evolution**, v. 47, n. 1, p. 96, 2015.

VELLEMAN, S. G. Relationship of Skeletal Muscle Development and Growth to Breast Muscle Myopathies: A Review. **Avian diseases**, v. 59, n. 4, p. 525–531, 2015.

VELTMAN, J. A.; BRUNNER, H. G. De novo mutations in human genetic disease. **Nature Review Genetics**, v. 13, p. 565–575, 2012.

VENTURINI, G. C. et al. Genetic and phenotypic parameters of carcass and organ traits of broiler chickens. **Genetics and Molecular Research**, v. 13, n. 4, p. 10294–10300, 2014.

VENTURINI, G. C. et al. Association between ACTA1 candidate gene and performance, organs and carcass traits in broilers. **Poultry Science**, v. 94, n. 12, p. 2863–2869, 2015.

VINCENT, A. K. et al. Identification of genomic deletions spanning the PCDH19 gene in two unrelated girls with intellectual disability and seizures. **Clinical Genetics**, v. 82, n. 6, p. 540–545, 2012.

WANG, J. et al. Identification of Genome-Wide Copy Number Variations among Diverse Pig Breeds Using SNP Genotyping Arrays. **PLoS ONE**, v. 8, n. 7, p. e68683, 2013.

WANG, K. et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v. 17, n. 11, p. 1665–1674, 2007.

WANG, K. et al. Modeling genetic inheritance of copy number variations. **Nucleic Acids Research**, v. 36, n. 21, 2008.

WANG, M. D. et al. Genomic population structure and prevalence of copy number variations in South African Nguni cattle. **BMC Genomics**, v. 16, p. 894–910, 2015.

WANG, X. et al. An initial map of chromosomal segmental copy number variations in the chicken. **BMC Genomics**, v. 11, p. 351, 2010.

WANG, X.; BYERS, S. Copy Number Variation in Chickens: A Review and Future Prospects. **Microarrays**, v. 3, n. 1, p. 24–38, 2014.

WANG, Y. et al. A genome-wide survey of copy number variation regions in various chicken breeds by array comparative genomic hybridization method. **Animal Genetics**, v. 43, n. 3, p. 282–289, 2012.

WANG, Y. et al. Analysis of Genome-Wide Copy Number Variations in Chinese Indigenous and Western Pig Breeds by 60 K SNP Genotyping Arrays. **PLoS ONE**, v. 9, n. 9, p. e106780, 2014.

WHEELER, E. et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. **Nature Genetics**, v. 45, n. 5, p. 513–517, 2013.

WOLPERT, L. Much more from the chicken's egg than breakfast - A wonderful model system. **Mechanisms of Development**, v. 121, n. 9, p. 1015–1017, 2004.

WRIGHT, D. et al. Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens. **PLoS Genetics**, v. 5, n. 6, p. 1–10, 2009.

WU, Y. et al. MATCHCLIP: Locate precise breakpoints for copy number variation using CIGAR string by matching soft clipped reads. **Frontiers in Genetics**, v. 4, n. AUG, p. 1–7, 2013.

XIAO, Y. et al. Association of growth rate with hormone levels and myogenic gene expression profile in broilers. **Journal of Animal Science and Biotechnology**, v. 8, n. 1, p. 43, 2017.

XU, L. et al. Characterization of Copy Number Variation's Potential Role in Marek's Disease. **International Journal of Molecular Sciences**, v. 18, n. 5, p. 1020, 2017.

XU, Z.; NIE, Q.; ZHANG, X. Overview of Genomic Insights into Chicken Growth Traits Based on Genome-Wide Association Study and microRNA Regulation. **Current genomics**, v. 14, n. 2, p. 137–46, 2013.

YALCIN, B. et al. The fine-scale architecture of structural variants in 17 mouse genomes. **Genome Biology**, v. 13, p. R18, 2012.

YAN, Y. et al. Genome-wide identification of copy number variations between two chicken lines that differ in genetic resistance to Marek's disease. **BMC Genomics**, v. 16, n. 1, p. 843, 2015.

YANG, T.-L. et al. Ethnic differentiation of copy number variation on chromosome 16p12.3 for association with obesity phenotypes in European and Chinese populations. **International Journal Obesity**, v. 37, n. 2, p. 188–190, 2013.

YAO, R. et al. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. **Molecular Cytogenetics**, v. 10, n. 1, p. 30, 2017.

YAU, C.; HOLMES, C. C. CNV discovery using SNP genotyping arrays. **Cytogenetic and Genome Research**, v. 123, n. 1–4, p. 307–312, 2009.

YI, G. et al. Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing. **BMC Genomics**, v. 15, n. 1, p. 962, 2014.

YI, G. et al. Genome-wide copy number profiling using high-density SNP array in chickens. **Animal Genetics**, v. 46, n. 2, p. 148–157, 2015.

ZHAN, B. et al. Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. **BMC Genomics**, v. 12, n. 1, p. 557, 2011.

ZHANG, F. et al. Copy Number Variation in Human Health, Disease, and Evolution. **Annu Rev Genomics Hum Genet**, p. 451–481, 2009.

ZHANG, H. et al. Detection of genome-wide copy number variations in two chicken lines divergently selected for abdominal fat content. **BMC genomics**, v. 15, p. 517, 2014a.

ZHANG, X. et al. Evaluation of copy number variation detection for a SNP array platform. **BMC bioinformatics**, v. 15, n. 1, p. 50, 2014b.

ZHAO, H. et al. CrossMap: A versatile tool for coordinate conversion between genome assemblies. **Bioinformatics**, v. 30, n. 7, p. 1006–1007, 2014.

ZHAO, M. et al. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. **BMC Bioinformatics**, v. 14, p. 1–16, 2012.

ZHAO, X. et al. Resolving complex structural genomic rearrangements using a randomized approach. **Genome Biology**, v. 17, n. 126, p. 1–13, 2016.

ZHOU, Q. Y. et al. Molecular characterization and association analysis with production traits of the porcine INPP5F Gene. **Molecular Biology Reports**, v. 36, n. 5, p. 1095–1098, 2009.

ZHOU, W. et al. A genome-wide detection of copy number variation using SNP genotyping arrays in Beijing-You chickens. **Genetica**, v. 142, n. 5, p. 441–450, 2014.

ANNEX

ANNEX A. Table 4

Table 4. Information about the CNVs identified by genotyping data that were validated by sequencing in the animals.

Animal	CNV	Chr	Start	End	Size (bp)	Type	Inherited	Position Sequenced CNV	Size of the Validated Region	% Validated
CC332	CNV 7	1	4123032	4142881	19850	duplication	Yes	1:4105101-4149800	19850	100
	CNV 11	2	1682872	1689661	6790	duplication	Yes	2:1683801-1690300	5861	86.32
	CNV 12	3	5427323	5449988	22666	duplication	Yes	3:5426401-5441500	21766	96.03
	CNV 13	3	5417840	5454230	36391	duplication	No	3:5442401-5456100	26930	74.00
								3:5426401-5441500		
								3:5442401-5456100		
	CNV 16	5	5796535	5807928	11394	duplication	Yes	5:5806301-6308700	1628	14.29
	CNV 19	9	9039971	9042949	2979	deletion	Yes	9:9039101-9044700	2979	100
	CNV 20	14	13727313	13815423	88111	total	No	9:9039101-9044700	1500	1.70
	CNV 21	19	2472566	2481908	9343	deletion	Yes	14:13771201-13772700	9343	100
	CNV 26	33	662744	673924	11181	duplication	Yes	19:2470701-2483700	3357	30.02
CC886	CNV 27	1	42379867	42404894	25028	duplication	No	33:658601-666100	15600	62,33
	CNV 29	1	1,8E+08	1,8E+08	20631	duplication	Yes	1:42380401-42396000	16631	80.61
	CNV 37	14	1,4E+07	1,4E+07	94686	deletion	No	1:182529101-182534000	2900	3.06
								1:182538001-182545400		
								1:182545401-182554100		
	CNV 38	14	1,4E+07	1,4E+07	53782	deletion	No	14:13725801-13727200	22700	42.21
								14:13771201-13772700		
								14:14288001-14291000		
								14:14306801-14318900		
								14:14323301-14328600		

14:14329401-14331700										
CC37	CNV 42	2	57897208	57917998	20791	duplication	Yes	2:57903701-57910700	7000	33.67
	CNV 47	9	3931594	3958661	27068	duplication	No	9:3920501-3961300	27068	100
CC88	CNV 53	1	1,4E+08	1,4E+08	73792	duplication	Yes	1:140027101-140046900	27334	37.04
								1:140047101-140051300		
								1:140051701-140058700		
	CNV 54	1	140019847	140034818	14972	duplication	Yes	1:140027101-140046900	7718	51.55
	CNV 55	2	126258399	126276813	18415	duplication	Yes	2:126257501-126277100	18415	100
	CNV 57	4	37855831	37870766	14936	deletion	Yes	4:37851501-37874300	14936	100
	CNV 60	4	24996355	25009951	13597	duplication	Yes	4:24995201-25010500	13597	100
	CNV 62	9	9039971	9042949	2979	deletion	Yes	9:9039501-9044800	2979	100
	CNV 63	9	3929048	3958661	29614	total	Yes	9:3920401-3961400	29614	100
	CNV 64	10	13327188	13334157	6970	duplication	Yes	10:13326501-13335800	6970	100
CC241	CNV 66	14	1,4E+07	1,4E+07	94686	deletion	No	14:13725501-13727200	3200	3.38
								14:13771201-13772700		
	CNV 67	1	75534848	75662569	127722	deletion	No	1:75627301-75635200	7900	6.19
	CNV 68	2	1682872	1689661	6790	duplication	Yes	2:1681301-1690200	6790	100
	CNV 71	2	66707701	66712392	4692	duplication	Yes	2:66705701-66714000	4692	100
	CNV 73	4	37852626	37870766	18141	deletion	No	4:37851501-37874300	18141	100
	CNV 74	4	33901385	33913055	11671	duplication	Yes	4:33902901-33915500	10155	87.01
	CNV 85	13	4971418	4982046	10629	duplication	Yes	13:4970901-4983900	10629	100
	CNV 87	16	245192	269662	24471	duplication	Yes	16:245701-252400	23300	95.21
								16:252401-269000		
TT5661	CNV 88	20	2075550	2080473	4924	duplication	No	20:2072601-2083100	4924	100
						deletion	Yes			
	CNV 91	1	131513799	131629165	115367	total		1:131511601-131629600	115367	100
	CNV 96	8	16571222	16584900	13679	duplication	Yes	8:16556001-16585000	13679	100

	CNV 97	9	11879228	11882853	3626	deletion	Yes	9:11877801-11883000	3626	100
	CNV 100	15	10859252	10864455	5204	duplication	Yes	15:10862401-10866100	2055	39.49
	CNV 103	1	1,4E+08	1,4E+08	86173	duplication	Yes	1:139999601-140025200 1:140044801-140057700 1:140062401-140064500	40600	47.11
TT5921	CNV 104	1	147135266	147156793	21528	duplication	Yes	1:147133501-147157100	21528	100
	CNV 107	4	60856679	60869438	12760	duplication	Yes	4:60854101-60869400	12722	99.70
	CNV 109	5	20784969	20811465	26497	duplication	No	5:20792701-20808300	15600	58.87
	CNV 110	5	42694839	42894203	199365	duplication	No	5:42689101-42895100	199365	100
	CNV 119	1	71401698	71425678	23981	deletion total	Yes	1:71398301-71445600	23981	100
	CNV 132	3	105637524	105652306	14783	duplication	Yes	3:105643701-105647800	4100	27.73
TT6232	CNV 133	5	29255145	29262357	7213	duplication	Yes	5:29253901-29264400	7213	100
	CNV 136	12	1241822	1250601	8780	duplication	Yes	12:1219701-1261300	8780	100
	CNV 139	19	6825744	6846796	21053	duplication	Yes	19:6823701-6853800	21053	100
	CNV 142	22	4594855	4632164	37310	duplication	Yes	22:4620301-4621100	800	2.14
	CNV 146	1	71407197	71442752	35556	duplication	Yes	1:71400201-71449000	35556	100
TT5649	CNV 150	23	4044662	4061219	16558	duplication	No	23:4040901-4058500	13839	83.58
	CNV 151	28	2867712	2875665	7954	duplication	Yes	28:2866901-2870600	2889	36.32

ANNEX B.

Tables 5 and 6.

Table 5. Information about CNVRs from previous studies that overlapping with inherited, not inherited and de novo CNVRs from broiler population.

Study	Number CNVRs	Total length (Mb)	INHERITED				NOT INHERITED				DE NOVO			
			Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)	Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)	Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)
Gorla et al (2017)	1216	47.05	290	23.84	4.37	9.28	29	2.38	0.6	1.27	22	1.80	0.67	1.42
Yi et al (2015)	221	5.6	29	13.12	0.43	7.67	4	1.8	0.04	0.71	1	0.45	0.005	0.08
Wang et al (2010) ¹	2462	10.94	464	18.84	1.1	10.05	79	3.2	0.19	1.73	8	0.32	0.13	1.18
Wang et al (2012) ¹	823	2.78	349	42.4	0.63	22.66	54	6.56	0.16	5.75	2	0.24	0.02	0.71
Crooijmans et al (2013) ¹	10411	47.2	1435	13.78	3.34	7.07	191	1.83	0.46	0.97	145	1.39	0.63	1.33
Luo et al (2013) ¹	528	1.06	190	35.98	0.21	19.81	27	5.11	0.04	3.77	0	0	0	0
Tian et al (2013) ¹	2369	8.54	624	26.34	0.96	11.24	90	3.79	0.19	2.22	52	2.19	0.16	1.87
Abernathy et al (2014) ¹	2376	5.18	483	20.32	0.66	12.74	73	3.07	0.18	3.47	18	0.75	0.03	0.57
Fan et al (2013) ¹	2301	2.48	383	16.64	0.36	14.51	47	2.04	0.08	3.22	21	0.91	0.07	2.82
Yi et al (2014)	8487	94.93	864	10.18	5.09	5.36	77	0.90	0.41	0.43	111	1.30	1.09	1.14
Yan et al (2015)	5010	18.62	445	8.88	1.76	9.45	52	1.03	0.32	1.71	85	1.69	0.27	1.45

Table 6. Information about CNVRs from previous studies that overlapping with inherited, not inherited and de novo CNVRs from F₂ population.

Study	Number CNVRs	Total length (Mb)	INHERITED				NOT INHERITED				DE NOVO			
			Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)	Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. of length (%)	Number CNVRs	Pct. of number (%)	Total length (Mb)	Pct. len (%)
Gorla et al (2017)	1216	47.05	287	23.6	5.61	11.92	94	7.73	3.03	6.43	408	33.55	14.22	30.
Yi et al (2015)	221	5.6	26	11.76	0.38	6.78	15	6.78	0.39	6.96	72	32.57	1.88	33.
Wang et al (2010) ¹	2462	10.94	404	16.4	0.93	8.5	309	12.55	0.62	5.66	736	28.89	3.38	30.
Wang et al (2012) ¹	823	2.78	381	46.29	0.83	29.85	195	23.69	0.55	19.78	91	11.05	0.54	19.
Crooijmans et al (2013) ¹	10411	47.2	1486	14.27	4.09	8.66	764	7.33	2.84	6.01	2933	28.17	14.01	26.
Luo et al (2013) ¹	528	1.06	205	38.82	0.22	20.75	150	28.4	0.06	5.66	176	33.33	0.49	46.
Tian et al (2013) ¹	2369	8.54	639	26.97	0.12	1.4	465	19.62	1.04	12.17	407	17.18	1.8	21.
Abernathy et al (2014) ¹	2376	5.18	494	20.79	0.59	11.38	173	7.28	0.27	5.21	486	20.45	1.43	27.
Fan et al (2013) ¹	2301	2.48	539	23.42	0.37	14.91	284	12.34	0.22	8.87	709	30.81	0.84	33.
Yi et al (2014)	8487	94.93	902	10.62	5.81	6.12	414	4.87	3.98	4.19	2686	31.64	29.58	31.
Yan et al (2015)	5010	18.62	477	9.52	1.84	9.88	263	5.24	1.09	5.85	1530	30.53	5.63	30.

ANNEX C. Table 7

Table 7. Position of CNV segments and CNVR significantly associated (p-value < 0.01) with the six traits (BMW- breast meat weight, BMY – breast meat yield, BFW – breast fillet weight, BFY- breast fillet yield, BW – breast weight and BY- breast yield).

Traits	CNV Segment	Position CNV Segment			P-value	CNVR	Position CNVR		
		Chr	Start	End			Chr	Start	End
BMW	245	4	43907116	43909719	8.52E-05	CNVR1765	4	43906643	43913344
	246	4	43912451	43912451	8.52E-05	CNVR1765	4	43906643	43913344
	297	6	30252488	30252488	5.65E-03	CNVR2246	6	30235695	30254967
	298	6	30253689	30253689	5.65E-03	CNVR2246	6	30235695	30254967
	299	6	30254167	30254561	5.65E-03	CNVR2246	6	30235695	30254967
	300	7	2107030	4563632	6.15E-05	CNVR2265	7	2106711	2118488
						CNVR2266	7	2904174	2910410
						CNVR2267	7	3426693	3447393
						CNVR2268	7	3495801	3510856
						CNVR2269	7	3854947	3861248
BMV						CNVR2270	7	4551322	4593640
	164	2	125171579	127639092	5,94E-04	CNVR1174	2	125166870	125178467
						CNVR1175	2	125835210	125844798
						CNVR1176	2	126090326	126118538
						CNVR1177	2	126632301	126644089
						CNVR1178	2	126936355	12696200
						CNVR1179	2	127598517	127613971
						CNVR1180	2	127635080	127761687
	165	2	127642735	127642735	8,99E-04	CNVR1180	2	127635080	127761687
	166	2	127645279	127660504	1,82E-03	CNVR1180	2	127635080	127761687
	167	2	127661360	127662783	3,69E-03	CNVR1180	2	127635080	127761687

	168	2	127664539	127664539	3,69E-03	CNVR1180	2	127635080	127761687
	169	2	127665659	127670206	3,52E-03	CNVR1180	2	127635080	127761687
	170	2	127670509	127670509	3,52E-03	CNVR1180	2	127635080	127761687
	171	2	127678707	127687279	2,44E-03	CNVR1180	2	127635080	127761687
	172	2	127692685	127693079	2,44E-03	CNVR1180	2	127635080	127761687
	173	2	127693779	127693779	2,44E-03	CNVR1180	2	127635080	127761687
	174	2	127695707	127698714	3,69E-03	CNVR1180	2	127635080	127761687
	175	2	127700836	127701458	3,69E-03	CNVR1180	2	127635080	127761687
	176	2	127708209	127717774	7,17E-03	CNVR1180	2	127635080	127761687
	177	2	127718070	127718070	7,17E-03	CNVR1180	2	127635080	127761687
	178	2	127721507	127935737	1,64E-03	CNVR1180	2	127635080	127761687
						CNVR1181	2	127933580	127945741
	245	4	43907116	43909719	1,74E-03	CNVR1765	4	43906643	43913344
	246	4	43912451	43912451	1,74E-03	CNVR1765	4	43906643	43913344
	300	7	2107030	4563632	8,98E-04	CNVR2265	7	2106711	2118488
						CNVR2266	7	2904174	2910410
						CNVR2267	7	3426693	3447393
						CNVR2268	7	3495801	3510856
						CNVR2269	7	3854947	3861248
						CNVR2270	7	4551322	4593640
	300	7	2107030	4563632	1,70E-03	CNVR2265	7	2106711	2118488
						CNVR2266	7	2904174	2910410
						CNVR2267	7	3426693	3447393
BFW						CNVR2268	7	3495801	3510856
						CNVR2269	7	3854947	3861248
						CNVR2270	7	4551322	4593640
	341	9	15786795	15810346	2,40E-04	CNVR2503	9	15784391	15790362

	342	9	15810367	15810367	2,40E-04	CNVR2504	9	15809112	15820640
						CNVR2504	9	15809112	15820640
	343	9	15812469	15815453	2,69E-03	CNVR2504	9	15809112	15820640
BFY	300	7	2107030	4563632	1,02E-06	CNVR2265	7	2106711	2118488
						CNVR2266	7	2904174	2910410
						CNVR2267	7	3426693	3447393
						CNVR2268	7	3495801	3510856
						CNVR2269	7	3854947	3861248
						CNVR2270	7	4551322	4593640
	421	19	6824730	6830430	6,53E-03	CNVR3027	19	6823309	6851703
BW	52	1	77759543	77761354	7,15E-03	CNVR287	1	77752486	77761509
	300	7	2107030	4563632	2,58E-07	CNVR2265	7	2106711	2118488
						CNVR2266	7	2904174	2910410
						CNVR2267	7	3426693	3447393
						CNVR2268	7	3495801	3510856
						CNVR2269	7	3854947	3861248
						CNVR2270	7	4551322	4593640
BY	300	7	2107030	4563632	7,30E-06	CNVR2265	7	2106711	2118488
						CNVR2266	7	2904174	2910410
						CNVR2267	7	3426693	3447393
						CNVR2268	7	3495801	3510856
						CNVR2269	7	3854947	3861248
						CNVR2270	7	4551322	4593640