University of São Paulo
Luiz de Queiroz College of Agriculture

Identification of eQTL from porcine muscle and liver mRNA sequencing

Felipe André Oliveira Freitas

Dissertation presented for obtaining the title of Master of Science. Area of concentration: Animal Science and Pastures

Piracicaba
2023

Felipe André Oliveira Freitas
Bachelor in Animal Science

Identification of eQTL from porcine muscle and liver mRNA sequencing

Advisor:
Associate Professor **ALINE SILVA MELLO CESAR**

Dissertation presented to obtain the title of Master of Science. Area of concentration: Animal Science and Pastures

Piracicaba
2023

RESUMO

Identificação de eQTL a partir do sequenciamento de RNAm de músculo e fígado de suínos

Este projeto teve como objetivo identificar polimorfismos de nucleotídeo único (SNP) localizados em regiões codificantes do DNA, identificados a partir do sequenciamento completo do mRNA, e associá-los ao nível de expressão gênica em músculo e fígado de suínos. Um total de 72 suínos machos imunocastrados e geneticamente magros foram usados em um estudo de 98 dias para avaliar características de carcaça e qualidade da carne do músculo *Longissimus lumborum* (LL). Os animais foram agrupados por peso corporal inicial (PC; 28,44 ± 2,95 kg) e distribuídos em um dos quatro tratamentos, com seis baias replicadas por tratamento e três porcos por baia. Os animais foram abatidos com peso vivo médio de 132,7±10,9 kg. Posteriormente, amostras de tecido muscular e hepático foram coletadas para extração e sequenciamento de mRNA. Posteriormente, uma análise de associação de eQTLs com características de desempenho (por exemplo, peso vivo, rendimento de carcaça, gordura intramuscular, área de olho de lombo e espessura de gordura) foi realizada com base em diferentes métodos estatísticos. Não houve eQTLs significativos (FDR<0,01) para SNPs do painel SNP de 50K. A contagem única de eQTL sem filtro para LD associados a genes variou de 2.066 a 2.247 para cis-eQTLs e 43 a 379 para trans-eQTLs. Para eQTLs com redução significativa (r^2 0,7, FDR<0,01), o número de SNPs únicos variou de 223 a 612 no cis-eQTL e de 29 a 403 no trans-eQTL. A contagem significativa de um único gene (FDR<0,01) nos eQTLs não podados variou de 159 a 304 em cis-eQTL e de 8 a 1.965 em trans-eQTL. A contagem de um único gene variou de 109 a 185 em cis-eQTL e de 6 a 5.993 em trans-eQTL. O tipo predominante de QTL anotado com os marcadores SNPs dos eQTLs significativos foi "Carne e Carcaça", seguido de "Saúde", enquanto o tipo de QTL que teve o menor percentual anotado foi "Carne e Carcaça eQTL" para cis e trans-eQTLs do sequenciamento muscular e todos os SNPs com e sem poda. Em resumo, este projeto contribuiu para o avanço do conhecimento e desenvolvimento de ferramentas práticas na área de genômica funcional da qualidade da carne suína e utilização de nutrientes em suínos.

Palavras-chave: *Sus scrofa*; Expressão gênica; Loci de característica quantitativa; SNP

ABSTRACT

Identification of eQTL from porcine muscle and liver mRNA sequencing

This project aimed to identify single nucleotide polymorphisms (SNP) located in coding regions of DNA, identified from complete mRNA sequencing, and associate them with the level of gene expression in muscle and liver of pigs. A total of 72 genetically lean, immunocastrated male pigs were used in a 98-day study to evaluate carcass traits and meat quality of Longissimus lumborum (LL) muscle. Animals were grouped by initial body weight (BW; 28.44 ± 2.95 kg) and assigned to one of four treatments, with six replicate pens per treatment and three pigs per pen. The animals were slaughtered with an average live weight of 132.7±10.9 kg. Subsequently, muscle and liver tissue samples were collected for mRNA extraction and sequencing. Subsequently, an association analysis of eQTLs with performance traits (eg live weight, carcass yield, intramuscular fat, loin eye area and fat thickness) was performed based on different statistical methods. There were no significant eQTLs (FDR<0.01) for SNPs from the 50K SNP panel. The unfiltered single eQTL count for gene-associated LD ranged from 2066 to 2247 for cis-eQTLs and 43 to 379 for trans-eQTLs. For significantly reduced eQTLs (r^2 0.7, FDR<0.01), the number of unique SNPs ranged from 223 to 612 in cis-eQTL and from 29 to 403 in trans-eQTL. The significant single gene count (FDR<0.01) in the unpruned eQTLs ranged from 159 to 304 in cis-eQTL and from 8 to 1965 in trans-eQTL. Single gene counts ranged from 109 to 185 in cis-eQTL and from 6 to 5993 in trans-eQTL. The predominant type of QTL annotated with the SNPs markers of the significant eQTLs was "Meat and Carcass", followed by "Health", while the type of QTL that had the lowest percentage annotated was "Meat and Carcass eQTL" for cis and trans-eQTLs of muscle sequencing and all SNPs with and without pruning. In summary, this project contributed to the advancement of knowledge and development of practical tools in the area of functional genomics of pork quality and nutrient utilization in pigs.

Keywords: *Sus scrofa;* Gene expression; Quantitative trait loci; SNP

1. INTRODUCTION

Pig production is an activity of great importance for the Brazilian economy (Krabbe, Filho, Miele, & Martins, 2008). Despite having been introduced in the country hundreds of years ago, the pork industry only started to develop significantly in recent decades (MIELE, SANTOS FILHO, MARTINS, & SANDI, 2011). Since then, pork production in Brazil has grown dramatically, becoming the fifth largest producer in the world, behind only China, the United States, the European Union and Russia in 2022 (USDA, 2023).

Brazil has one of the largest pig production systems in the world, with different levels of production, from small-scale pig farming to large-scale production in industrial farms (FISCHER et al., 2019; Krabbe et al., 2008). The main challenges faced by the Brazilian pork industry include reducing production costs, improving product quality, adopting more sustainable practices and adopting measures to reduce animal health and welfare problems (FISCHER et al., 2019; Krabbe et al., 2008). To face these challenges, the sector has sought to develop technologies that can increase productivity and product quality, in addition to improving animal welfare (FISCHER et al., 2019).

Pork production is highly intensive and involves high levels of technology (Dong, Moritaka, Liu, & Fukuda, 2020; FISCHER et al., 2019). The modernization of pig farms has brought significant advances to production, allowing pig farmers to produce higher quality meat at lower prices (Dong et al., 2020; FISCHER et al., 2019).The pork production sector also faces some challenges. The cost of producing pigs, especially for large producers, is significant due to the high consumption of food, feed and water (Alves et al., 2022; FISCHER et al., 2019), in addition to the effects of climate change (Ardlie et al., 2015; Hörtenhuber et al., 2020; Rauw et al., 2020; Renaudeau & Dourmad, 2022).

Concerning production efficiency, the use of animal genetic improvement is common, and pig farming has been an area in which this tool has been successfully applied (Knap & Kause, 2018; Merks, Mathur, & Knol, 2011). Pig breeders have sought to improve animal production characteristics and pork production, improve pork quality, and reduce production cost (Knap & Kause, 2018; Merks et al., 2011). In addition, animal genetic improvement in pig farming is also applied to increase the resistance of animals to diseases and environmental conditions (FISCHER et al., 2019; Knap & Kause, 2018; Merks et al., 2011).

The principles of breeding start with artificial selection, quantitative genetic improvement, and kin selection (Eler, 2017). Artificial selection involves the selection of genetically superior animals for breeding with the aim of improving desired traits (Eler, 2017). Quantitative genetic improvement involves the use of mathematical models to evaluate the effects of genes and select animals with the desired traits (Eler, 2017).

The advancement of sequencing and genotyping technologies, associated with computational development, which, in turn, combined with bioinformatics tools, as well as the publication of genome sequencing in species of zootechnical interest, opened a new era in genetics, livestock, in agriculture and even in human medicine (Depristo et al., 2011; Ellen et al., 2019; Gadea, Coy, Matás, Romar, & Cánovas, 2020; Wu & Bazer, 2019).

Compared to traditional data used in previous evaluations, thousands of genetic markers formed mainly by single nucleotide polymorphisms (SNPs) provide the possibility of predicting the genetic value through high-density panels (Matukumalli, Lawley, Schnabel, Taylor, & Allan, 2009). Thus, the best genomic unbiased linear prediction (GBLUP) was developed based on the principle that quantitative traits are controlled by many markers such as SNPs (Misztal et al., 2017). The estimated genome value (GEBV) is the sum of the effects of dense genetic markers or their haplotypes along the genome, which was predicted and applied to the selection program (Botelho et

al., 2020). Thus, genomic selection involving GEBV-based selection decisions has revolutionized the livestock industry. The main benefits of genome selection are the possibility of decreasing the generation interval, which is inversely proportional to the genetic gain, in addition to reducing the cost of progeny testing, as the GEBV can be obtained early in life (Ellen et al., 2019; Misztal et al., 2017).

Since the 1980s, when the first studies involving molecular markers in pig farming began (Chardon et al., 1985), the development of molecular genetic technology has allowed the integration of molecular markers to genetic improvement (Yang, Fu, Khan, Zeng, & Fu, 2013).

Single nucleotide polymorphism (SNP) is the variation of a nucleotide base, which can be a Transition, that occurs when a purine is replaced by another purine, or when a pyrimidine base is replaced by another pyrimidine base (Cytosine/ Thymine or Guanine/Adenine, respectively) (Helyar et al., 2011; Kim & Misra, 2007). It can also be a Transversion, when a purine base is replaced by a pyrimidine base or vice versa (C/G, C/A, T/A and T/G) (Helyar et al., 2011; Kim & Misra, 2007). SNPs are the main types of DNA polymorphisms used for studies of genetic variation. They are present throughout the genome, mainly in the intronic region, that is, non-coding regions (Hiremath et al., 2012). It also appears in gene coding sequences (exon) or in non-genetic coding regions (exon-intron splicing site) (Hiremath et al., 2012). The SNPs in the coding regions can be divided into synonyms, when there is no change in the protein, and non-synonyms, when the protein is changed (Zhao et al., 2019).

The SNPs that occur in the coding region and regulatory sequence, respectively, can have a considerable impact on protein function and gene expression (Zhao et al., 2019). This change in a nucleotide base can result in a change in the codon encoding an amino acid, thus altering protein synthesis. Or even this SNP when present in promoter regions of a gene or 3' non-Transcribed region (3'UTR) can change the level of expression and affect post-transcriptional regulations, respectively (Albert & Kruglyak, 2015; Buckingham & Relaix, 2015; Cesar et al., 2015; Gaffney, 2013; Siriluck Ponsuksili et al., 2015). Thus, these SNPs may be causing phenotypic differences in different individuals of a population, that is, be associated with characteristics of zootechnical interest, and animal and human health (pork consumer). Because the SNPs are numerous and wide-ranging in the genome, they are considered ideal for characterizing the genetic architecture and identifying functional genes for traits of economic interest. (Zhao et al., 2019).

Pig genotyping chips vary in the amount of genetic markers, as well as in cost (Badke, Bates, Ernst, Fix, & Steibel, 2014; Boison et al., 2015; Deng et al., 2022; Ferenčaković, Sölkner, & Curik, 2013; Guelfi et al., 2020). The simplest chips have between 40 and 120 thousand genetic markers, while the broader ones have between 500 and 800 thousand markers (Badke et al., 2014; Boison et al., 2015; Deng et al., 2022). In addition, some chips also offer a set of additional features, such as variable association analysis, SNP detection of interest and haplotype association analysis (Amaral, Megens, Crooijmans, Heuven, & Groenen, 2008). Some chips also offer features to improve genotyping accuracy, such as correction of genotyping errors and detection of genes of interest (Badke et al., 2014).

Although there is no "best" genotyping chip for pigs, as each chip has its own features and characteristics (Badke et al., 2014). The right chip depends on the user's needs and how the results will be used. For example, if the user wants to study genetic variability among pigs, a chip with more genetic markers will be more suitable (Badke et al., 2014; Boison et al., 2015). On the other hand, if you want to study the associations between a gene of interest and an inherited disease, a chip with advanced features for association analysis is the best choice.

Chips with 50,000 (50k) SNPs are used for pig genotyping because they provide a considerable amount of genetic detail about a pig's traits. These chips contain about 50,000 genetic markers, which are used to provide information about genetic variability and inheritance trends of specific traits (Ferenčaković et al., 2013). This

genotyping has been used to analyze genetic variability and to identify genes associated with productive performance and behavior in pigs (Badke et al., 2014; Boison et al., 2015; Deng et al., 2022; Dixon et al., 2007; Kim & Misra, 2007). In addition, it can also be used for predicting response to selection and disease prevention (MERKS; MATHUR; KNOL, 2011).

In general, 50k genotyping in pigs can be used successfully for the identification of genetic variants associated with production and meat quality traits (Merks et al., 2011). In addition, it can also be used for selection response prediction and disease prevention (Ferenčaković et al., 2013). However, the costs involved in this type of analysis can be high and the process can be complex.

The Genomic Wide Association Study (GWAS) in swine has been widely applied to identify loci and genetic variants associated with important zootechnical traits and disease development (Ellen et al., 2019; Visscher et al., 2017). GWAS studies in pigs have allowed the discovery of loci related to meat production, disease resistance, immune response, meat quality, body size, reproductive traits and other traits (Visscher et al., 2017). It has also been possible to use the results of GWAS studies to select pigs with desirable traits, to predict disease susceptibility and to improve disease resistance (Ellen et al., 2019; Visscher et al., 2017; Võsa et al., 2018).

Linkage disequilibrium (LD) is a phenomenon that occurs when there are many alleles linked to the same genetic locus in a population (Amaral et al., 2008; Arcos-Burgos & Muenke, 2002; Slatkin, 1994; Weir, 1979). This inequality in the distribution of alleles can lead to an association between loci that are closer genetically than expected by chance (Amaral et al., 2008; Pérez O'Brien et al., 2014; Slatkin, 1994; Weir, 1979). For example, if two highly correlated alleles are present together in a population, this can lead to an LD. Therefore, LD can have significant effects on genetic variability and consequently on trait inheritance (Pérez O'Brien et al., 2014; Weir, 1979).

The identification of eQTL (Expression Quantitative Tracer Locus) is a process to discover the genetic locus or loci of a gene that are associated with gene expression (Cesar et al., 2018; Gibson, Powell, & Marigorta, 2015; Gilad, Rifkin, & Pritchard, 2008). This approach can be used to discover the loci that are involved in the expression of a gene in a population (Cesar et al., 2018) which in turn are related to traits of interest. Thus, this information can be used to improve animal selection for economic traits such as production and meat quality (Gilad et al., 2008).

Ribonucleic Nucleotide Acid (RNA-seq) sequencing using next-generation sequencing technologies (NGS) allows quantifying levels of (mRNA) in an organism to estimate gene expression profiles at a given time (Carrillo et al., 2016; Dobin et al., 2013; Grabherr et al., 2011). RNA-seq allows the measurement of gene expression on a large scale, which is of great importance for advancing our knowledge about the functional genomics of fatty acid metabolism and related features (Grabherr et al., 2011; S. Ponsuksili et al., 2010; Siriluck Ponsuksili et al., 2015).

Among sequencing technologies, RNA-seq has become one of the most representative high-throughput technologies due to its high accuracy and cost-effectiveness. There are several advantages to using RNA sequence data for polymorphism analysis. In addition to being able to find thousands of candidate SNPs, it is possible to detect the effects of polymorphisms on the expression levels of functional genes, at a reasonable cost (Zhao et al., 2019). It is also possible to locate the variation of the coding region related to the phenotypic characteristics of the animals, serving as an increment to predict the phenotype through the genotype (YU et al., 2014). In addition, it is useful for research such as gene characterization, quantification of gene expression and analysis of the post-Translation process (Quinn et al., 2013).

The development of the new generation, mainly increasing the read length, improved the quality of the original sequencing data, reducing sequencing and assembly errors (You et al., 2012). Thus, greater read length can

produce higher quality raw data and affect further analysis (Chaisson, Brinza, & Pevzner, 2009; Chhangawala, Rudy, Mason, & Rosenfeld, 2015).

Choosing the appropriate assembler is also crucial for SNP detection (Zhao et al., 2019). The SNP result is affected by the SNP calling program applied, each of the different tools (GeMS, SAMtools and GATK) to call SNPs individually from the sequencing data, resulting in a unique SNP accuracy rate in the search (You et al., 2012).

There are several NGS (Next Generation Sequencing) platforms, such as Illumina Genome Analyzer, Roche/454 FLX and ABI SOLiD, which differ in technical specifications, resulting in differences in sensitivity, precision, repeatability and throughput (Harismendy et al., 2009), which means that sequencing data obtained from different platforms has different limitations.

When evaluating third-generation SNP and RNA-seq sequencing, it is important to evaluate the performance of data analysis algorithms. Analysis algorithms must be able to detect variants with high precision and specificity, and must be able to provide a high degree of confidence in the results (Harismendy et al., 2009). Furthermore, it is important to consider the cost of processing the data, as well as, the time required to complete the process (Grabherr et al., 2011). Analysis algorithms must also be able to handle large volumes of data efficiently (Dobin et al., 2013; Harismendy et al., 2009). Finally, it is important to assess the quality of the data obtained, as well as the accuracy and consistency of the performed analyzes.

Among the control mechanisms of gene expression are (Bruce Alberts et al., 2017): alterations in the level of transcription of genes, there is, SNPs can alter the level of transcription of a gene, increasing or decreasing the amount of mRNA produced from a gene. This happens when SNPs are located in regions that regulate tanscription, such as promoters or regulatory elements. Alteration of gene regulation by environmental factors: SNPs can also affect how a gene is regulated by environmental factors. For example, they can change the way a gene is affected by stressors such as heat, light or food. Changing the way genes are processed: SNPs can change the way a gene is processed, such as alternative splicing or changes in the structure of the protein encoded by the gene. Changing the structure of the protein encoded by the gene: SNPs can influence how the protein encoded by the gene is formed, which can affect its activity and function. Changing linkage between genes: SNPs can affect how genes are linked to each other, thus changing how genes are expressed. Changing the stability of mRNAs: SNPs can also affect the stability of mRNAs, thus changing the level of expression of genes. Alteration of alternative splicing activity: SNPs can alter alternative splicing activity, which is the process by which different versions of a gene are produced from the same gene. Alteration of promoter activity: SNPs can alter the activity of promoters, which are regions that regulate the level of transcription of a gene. Altering the activity of transcriptional regulators: SNPs can also alter the activity of transcriptional regulators, which are proteins that regulate the level of transcription of a gene. Altering epigenetic silencing activity: SNPs can alter epigenetic silencing activity, which is the process by which certain genes are silenced through epigenetic modifications.

The Matrix eQTL package is simpler to use than these algorithms, as it does not require advanced technical knowledge to configure or implement the algorithm (Shabalin, 2012). It has a linear and anova model option, and the linear regression algorithm examines the eQTL data to find variations in gene expression that are significantly associated with a given genotype and corrects the p-values for multiple tests by the FDR method (Shabalin, 2012). Furthermore, it is an open-source tool, allowing anyone to use it for free. However, it may have disadvantages compared to other algorithms, such as its low accuracy compared to other eQTL identification algorithms, such as FastQTL (Nodzak, 2020). It cannot be used to perform more advanced analysis, such as interaction analysis. And it is also limited regarding the number of data that can be analyzed at the same time.

Despite its limitations, due to the amount of data, the Matrix eQTL is widely recommended due to the optimization of its algorithm according to the computational input (Ardlie et al., 2015; Nodzak, 2020; Shabalin, 2012).

In addition to the Matrix eQTL package there are other eQTL identification algorithms such as Merlin 1.1.2, GridQTL 3.3.0, QTLMap 0.9.7, Pseudomarker 2.04, snpMatrix 2.4, eMap 1.2, R/QTL, MapQTL 6, Second-Generation PLINK: PLINK 1.9 Beta and FastQTL 2.184 (Nodzak, 2020). These algorithms use different approaches to identify genes associated with gene expression traits. Since some programs have more complex models and demand greater computational input, others, such as Matrix-eQTL, have an optimized algorithm for large-scale data calculations (Nodzak, 2020).

The annotation and functional enrichment of genes and variants is an important technique for the identify new genes, variants and biological processes that affect the health and well-being of pigs (Subramanian et al., 2005). This technique involves the annotation of genome data, such as the identification of new genes and variants, the mapping of gene locations and the identification of functional patterns for DNA/RNA sequences, which allows the understanding of the biological implications of genes (Subramanian et al., 2005). Functional enrichment also allows the identification of functional variants, which may be related to genetic diseases, which favors the characterization of risk factors, as well as, the identification of new approaches for treatment (Subramanian et al., 2005).

Gene Ontology (GO) is a gene classification system that describes the biological function of genes and gene products. GO uses a hierarchical ontology to classify genes according to their function, process and cellular location (Ashburner et al., 2000; Subramanian et al., 2005). The ontology is a hierarchical structure of terms that describe relationships between various levels of abstraction. For example, "Metabolism" is a higher-level term that contains more specific terms such as "Citric Acid Cycle" and "Glycolysis" (Ashburner et al., 2000).

Metabolic pathways are specific sequences of chemical reactions that transform a set of substrates (or reagents) into a set of products (Kanehisa & Goto, 2000). Metabolic pathways are essential for maintaining homeostasis and metabolism in complex organisms, as they allow substrates to be converted into products necessary for survival (Kanehisa & Goto, 2000). Some of the main metabolic pathways include glycolysis, the citric acid cycle, oxidative phosphorylation, gluconeogenesis and amino acid biosynthesis (Kanehisa & Goto, 2000).

The use of eQTL to detect GO terms and metabolic pathways is an approach used to identify the functioning mechanism of complex traits (Conesa et al., 2016). This one method is based on analyzing gene expression data to identify genetic loci that are associated with specific phenotypic variations. This approach is extremely useful to identify candidate biological mechanisms and to better understand how genes may be involved in the genesis of complex diseases (Aguet et al., 2020; Conesa et al., 2016).

REFERENCIAS

Aguet, F., Barbeira, A. N., Bonazzola, R., Brown, A., Castel, S. E., Jo, B., … Volpi, S. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*(6509), 1318–1330. doi: 10.1126/SCIENCE.AAZ1776

Albert, F. W., & Kruglyak, L. (2015, April 19). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, Vol. 16, pp. 197–212. Nature Publishing Group. doi: 10.1038/nrg3891

Alves, L. K. S. ;, Gameiro, A. H. ;, Schinckel, A. P. ;, Garbossa, C. A. P., Schwarz, T., Kannan, L., … Garbossa, P. (2022). Development of a Swine Production Cost Calculation Model. *Animals 2022, Vol. 12, Page 2229*, *12*(17), 2229. doi: 10.3390/ANI12172229

Amaral, A. J., Megens, H. J., Crooijmans, R. P. M. A., Heuven, H. C. M., & Groenen, M. A. M. (2008). Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics*, *179*(1), 569–579. doi: 10.1534/GENETICS.107.084277

Arcos-Burgos, M., & Muenke, M. (2002). Genetics of population isolates. *Clinical Genetics*, *61*(4), 233–247. doi: 10.1034/J.1399-0004.2002.610401.X

Ardlie, K. G., DeLuca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., … Lockhart. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*, *348*(6235), 648–660. doi: 10.1126/science.1262110

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. doi: 10.1038/75556

Badke, Y. M., Bates, R. O., Ernst, C. W., Fix, J., & Steibel, J. P. (2014). Accuracy of estimation of genomic breeding values in pigs using low-density genotypes and imputation. *G3: Genes, Genomes, Genetics*, *4*(4), 623–631. doi: 10.1534/G3.114.010504

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(Database issue). doi: 10.1093/NAR/GKS1195

Boison, S. A., Santos, D. J. A., Utsunomiya, A. H. T., Carvalheiro, R., Neves, H. H. R., O'Brien, A. M. P., … da Silva, M. V. G. B. (2015). Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (Bos indicus) dairy cattle: Comparison of commercially available SNP chips. *Journal of Dairy Science*, *98*(7), 4969–4989. doi: 10.3168/JDS.2014-9213

Botelho, M. E., Lopes, M. S., Mathur, P. K., Knol, E. F., Guimarães, S. E. F., Marques, D. B. D., … Veroneze, R. (2020). Applying an association weight matrix in weighted genomic prediction of boar taint compounds. *Journal of Animal Breeding and Genetics*, (April), 1–12. doi: 10.1111/jbg.12528

Brookes, A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177–186. doi: 10.1016/S0378-1119(99)00219-X

Bruce Alberts, Alexander Johnson, Julian Lewis, Morgan, D., Martin Raff, Keith Roberts, … Tim Hunt. (2017). *Tradução :Biologia Molecular da Célula* (6ª).

Buckingham, M., & Relaix, F. (2015). PAX3 and PAX7 as upstream regulators of myogenesis. *Seminars in Cell and Developmental Biology*, *44*, 115–125. doi: 10.1016/j.semcdb.2015.09.017

Carrillo, J. A., He, Y., Li, Y., Liu, J., Erdman, R. A., Sonstegard, T. S., & Song, J. (2016). Integrated metabolomic and transcriptome analyses reveal finishing forage affects metabolic

pathways related to beef quality and animal welfare. *Scientific Reports*, *6*(April), 1–16. doi: 10.1038/srep25948

Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E., Fritz-Waters, E. R., Lanna, D. P. D., Gasparin, G., … Coutinho, L. L. (2015). Putative regulatory factors associated with intramuscular fat content. *PLoS ONE*, *10*(6), e0128350. doi: 10.1371/journal.pone.0128350

Cesar, A. S. M., Regitano, L. C. A., Reecy, J. M., Poleti, M. D., Oliveira, P. S. N., de Oliveira, G. B., … Coutinho, L. L. (2018). Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. *BMC Genomics*, *19*(1), 499. doi: 10.1186/s12864-018-4871-y

Chaisson, M. J., Brinza, D., & Pevzner, P. A. (2009). De novo fragment assembly with short mate-paired reads: Does the read length matter? *Genome Research*, *19*(2), 336–346. doi: 10.1101/gr.079053.108

Chardon, P., Vaiman, M., Kirszenbaum, M., Geffrotin, C., Renard, C., & Cohen, D. (1985). Restriction Fragment Length Polymorphism of the Major Histocompatibility Complex of the Pig. In *Immunogenetics* (Vol. 21).

Chhangawala, S., Rudy, G., Mason, C. E., & Rosenfeld, J. A. (2015). The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biology*, *16*(1), 131. doi: 10.1186/s13059-015-0697-y

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., … Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1). doi: 10.1186/S13059-016-0881-8

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. doi: 10.1093/NAR/GKAB1049

Deng, T., Zhang, P., Garrick, D., Gao, H., Wang, L., & Zhao, F. (2022). Comparison of Genotype Imputation for SNP Array and Low-Coverage Whole-Genome Sequencing Data. *Frontiers in Genetics*, *12*, 2652. doi: 10.3389/FGENE.2021.704118/BIBTEX

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. v., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–501. doi: 10.1038/ng.806

Dixon, A. L., Liang, L., Moffatt, M. F., Chen, W., Heath, S., Wong, K. C. C., … Cookson, W. O. C. (2007). A genome-wide association study of global gene expression. *Nature Genetics*, *39*(10), 1202–1207. doi: 10.1038/ng2109

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. doi: 10.1093/bioinformatics/bts635

Dong, D. D., Moritaka, M., Liu, R., & Fukuda, S. (2020). Restructuring toward a modernized agro-food value chain through vertical integration and contract farming: the swine-to-pork industry in Vietnam. *Journal of Agribusiness in Developing and Emerging Economies*, *10*(5), 493–510. doi: 10.1108/JADEE-07-2019-0097

Eler, J. P. (2017). *Teorias e métodos em melhoramento genético animal: bases do melhoramento genético animal*. Universidade de São Paulo. Faculdade de Zootecnia, Engenharia de Alimentos. doi: 10.11606/9788566404128

Ellen, E. D., van der Sluis, M., Siegford, J., Guzhva, O., Toscano, M. J., Bennewitz, J., … Bas Rodenburg, T. (2019). Review of sensor technologies in animal breeding: Phenotyping behaviors of laying hens to select against feather pecking. *Animals*, *9*(3). doi: 10.3390/ani9030108

Ferenčaković, M., Sölkner, J., & Curik, I. (2013). Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genetics Selection Evolution*, *45*(1), 1–9. doi: 10.1186/1297-9686-45-42/FIGURES/4

FISCHER, A., BARROS, E., MACHADO, I., MIELE, M., PINHEIRO, R., de MEDEIROS, S. A. F., & GUTH, T. (2019). Novos Caminhos Na Suinocultura. *Sebrae, ABCS*. Retrieved from https://ainfo.cnptia.embrapa.br/digital/bitstream/item/204600/1/final9061.pdf

Gadea, J., Coy, P., Matás, C., Romar, R., & Cánovas, S. (2020). Reproductive technologies in swine. *Reproductive Technologies in Animals*, 67–79. doi: 10.1016/b978-0-12-817107-3.00005-9

Gaffney, D. J. (2013). Global Properties and Functional Complexity of Human Gene Regulatory Variation. *PLoS Genetics*, *9*(5), e1003501. doi: 10.1371/journal.pgen.1003501

Gibson, G., Powell, J. E., & Marigorta, U. M. (2015, June 24). Expression quantitative trait locus analysis for translational medicine. *Genome Medicine*, Vol. 7, p. 60. BioMed Central Ltd. doi: 10.1186/s13073-015-0186-7

Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics*, *24*(8), 408–415. doi: 10.1016/J.TIG.2008.06.001

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. doi: 10.1038/NBT.1883

Guelfi, S., D'Sa, K., Botía, J. A., Vandrovcova, J., Reynolds, R. H., Zhang, D., … Ryten, M. (2020). Regulatory sites for splicing in human basal ganglia are enriched for disease-relevant information. *Nature Communications*, *11*(1), 1041. doi: 10.1038/s41467-020-14483-x

Harismendy, O., Ng, P. C., Strausberg, R. L., Wang, X., Stockwell, T. B., Beeson, K. Y., … Frazer, K. A. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*, *10*(3), 32. doi: 10.1186/gb-2009-10-3-r32

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., … Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, *11*(SUPPL. 1), 123–136. doi: 10.1111/j.1755-0998.2010.02943.x

Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., … Varshney, R. K. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal*, *10*(6), 716. doi: 10.1111/J.1467-7652.2012.00710.X

Hörtenhuber, S. J., Schauberger, G., Mikovits, C., Schönhart, M., Baumgartner, J., Niebuhr, K., … Zollitsch, W. (2020). The Effect of Climate Change-Induced Temperature Increase on Performance and Environmental Impact of Intensive Pig Production Systems. *Sustainability 2020, Vol. 12, Page 9442*, *12*(22), 9442. doi: 10.3390/SU12229442

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. doi: 10.1093/NAR/28.1.27

Kim, S., & Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annual Review of Biomedical Engineering*, *9*, 289–320. doi: 10.1146/ANNUREV.BIOENG.9.060906.152037

Knap, P. W., & Kause, A. (2018). Phenotyping for genetic improvement of feed efficiency in fish: Lessons from pig breeding. *Frontiers in Genetics*, *9*(MAY), 184. doi: 10.3389/FGENE.2018.00184/BIBTEX

Krabbe, E. L., Filho, J. I. D. S., Miele, M., & Martins, F. M. (2008). Cadeias produtivas de suínos e aves. *Tópicos Atuais Na Produção de Suínos e Aves*, 9–32. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., & Allan, M. F. (2009). Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE*, *4*(4), 5350. doi: 10.1371/journal.pone.0005350

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., … Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 1–14. doi: 10.1186/S13059-016-0974-4/TABLES/8

Merks, J. W. M., Mathur, P. K., & Knol, E. F. (2011). New phenotypes for new breeding goals in pigs. *Animal*, *6*(4), 535–543. doi: 10.1017/S1751731111002266

MIELE, M., SANTOS FILHO, J. I. dos, MARTINS, F. M., & SANDI, A. J. (2011). O desenvolvimento da suinocultura brasileira nos últimos 35 anos. *Embrapa Suínos e Aves*, *1975*, 85–102. Retrieved from http://www.alice.cnptia.embrapa.br/handle/doc/907870

Milano, I., Babbucci, M., Panitz, F., Ogden, R., Nielsen, R. O., Taylor, M. I., … Bargelloni, L. (2011). Novel tools for conservation genomics: Comparing two high-throughput approaches for SNP discovery in the transcriptome of the european hake. *PLoS ONE*, *6*(11), 28008. doi: 10.1371/journal.pone.0028008

Misztal, I., Bradford, H. L., Lourenco, D. A. L., Tsuruta, S., Masuda, Y., & Legarra, A. (2017). Studies on inflation of GEBV in single-step GBLUP for type. *Interbull Bulletin*, *0*(51).

Nodzak, C. (2020). Introductory methods for EQTL analyses. *Methods in Molecular Biology*, *2082*, 3–14. doi: 10.1007/978-1-0716-0026-9_1/TABLES/2

Pérez O'Brien, A. M., Mészáros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, J. F., van Tassell, C. P., … Sölkner, J. (2014). Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livestock Science*, *166*(1), 121–132. doi: 10.1016/J.LIVSCI.2014.05.007

Ponsuksili, S., Murani, E., Phatsara, C., Schwerin, M., Schellander, K., & Wimmers, K. (2010). Expression quantitative trait loci analysis of genes in porcine muscle by quantitative real-time RT-PCR compared to microarray data. *Heredity 2010 105:3*, *105*(3), 309–317. doi: 10.1038/hdy.2010.5

Ponsuksili, Siriluck, Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M., & Wimmers, K. (2015). Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. *PLoS ONE*, *10*(4), e0123678. doi: 10.1371/journal.pone.0123678

Quinn, E. M., Cormican, P., Kenny, E. M., Hill, M., Anney, R., Gill, M., … Morris, D. W. (2013). Development of Strategies for SNP Detection in RNA-Seq Data: Application to Lymphoblastoid Cell Lines and Evaluation Using 1000 Genomes Data. *PLoS ONE*, *8*(3). doi: 10.1371/journal.pone.0058815

Rauw, W. M., Rydhmer, L., Kyriazakis, I., Øverland, M., Gilbert, H., Dekkers, J. C. M., … Gomez-Raya, L. (2020). Prospects for sustainability of pig production in relation to climate change and novel feed resources. *Journal of the Science of Food and Agriculture*, *100*(9), 3575–3586. doi: 10.1002/JSFA.10338

Renaudeau, D., & Dourmad, J. Y. (2022). Review: Future consequences of climate change for European Union pig production. *Animal*, *16*, 100372. doi: 10.1016/J.ANIMAL.2021.100372

Shabalin, A. A. (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, *28*(10), 1353–1358. doi: 10.1093/bioinformatics/bts163

Slatkin, M. (1994). Linkage disequilibrium in growing and stable populations. *Genetics*, *137*(1), 331–336. doi: 10.1093/GENETICS/137.1.331

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. doi: 10.1073/PNAS.0506580102

USDA. (2023). Livestock and poultry: world markets and trade. *United States Department of Agriculture and Foreign Agricultural Service*, 16. Retrieved from https://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf

Vaughn, S. E. (2012). Review of the Third Edition of the Guide for the Care and Use of Agricultural Animals in Research and Teaching. *Journal of the American Association for Laboratory Animal Science : JAALAS*, *51*(3), 298. Retrieved from /pmc/articles/PMC3358976/

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. doi: 10.1016/J.AJHG.2017.06.005

Võsa, U., Claringbould, A., Westra, H. J., Bonder, M. J., Deelen, P., Zeng, B., … Franke, L. (2018, October 19). Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *BioRxiv*. bioRxiv. doi: 10.1101/447367

Weir, B. S. (1979). Inferences about Linkage Disequilibrium. *Biometrics*, *35*(1), 235. doi: 10.2307/2529947

Wu, G., & Bazer, F. W. (2019). Application of new biotechnologies for improvements in swine nutrition and pork production. *Journal of Animal Science and Biotechnology*, *10*(1), 1–16. doi: 10.1186/s40104-019-0337-6

Yang, L., Fu, S., Khan, M. A., Zeng, W., & Fu, J. (2013). Molecular cloning and development of RAPD-SCAR markers for Dimocarpus longan variety authentication. *SpringerPlus*, *2*(1), 1–8. doi: 10.1186/2193-1801-2-501

You, N., Murillo, G., Su, X., Zeng, X., Xu, J., Ning, K., … Bateman, A. (2012). *SNP calling using genotype model selection on high-throughput sequencing data*. *28*(5), 643–650. doi: 10.1093/bioinformatics/bts001

Zhao, Y., Wang, K., Wang, W. L., Yin, T. T., Dong, W. Q., & Xu, C. J. (2019). A high-throughput SNP discovery strategy for RNA-seq data. *BMC Genomics*, *20*(1), 1–10. doi: 10.1186/s12864-019-5533-4

2. IDENTIFICATION OF DELETERIOUS SINGLE NUCLEOTIDE POLYMORPHISMS IN DIFFERENT TISSUES OF PIGS ASSOCIATED WITH BLOOD BIOCHEMICAL PARAMETERS

2.1. Introduction

The pig is a monogastric species, being responsible for the production of one of the most consumed meats worldwide (MENDONÇA, 2017; USDA, 2023), representing about 33% of the meat consumed in the world, and being an important source of animal protein (OECD, 2023; WHITTON et al., 2021). According to FAO data (2023), world production of pork was approximately 120 million tons in 2020. The, *Sus scrofa species* is of great relevance not only in meat production, but for human health, as it is used as an animal model for scientific studies in humans (PAN et al., 2021).

For studies of genetic variation, the single nucleotide polymorphism (SNP) stands out among all types of DNA polymorphisms. The SNP can be found throughout the genome, mainly in intronic regions, i.e., non-coding regions, but also located in the exons, i.e., coding regions. (ZHAO et al., 2019). The SNP is a variation of bases in DNA, such as transitions and transversions. Transitions occur between the exchange of purine bases (A/G) or between pyrimidine bases (C/T). Transversions occur when there is substitution between purine bases for pyrimidines, or vice-versa (A/T, G/C, T/A and C/G) (TURCHETTO-ZOLET et al., 2017). In the coding regions, the SNP can cause alterations in the protein structure and, therefore, in its function, being able to trigger diseases, but also, being able to be used as a molecular marker in genetics (HELYAR et al., 2011).

The Genome Analysis Toolkit (GATK) is a widely used tool for analyzing genome and exome sequencing data. Nowadays, the GATK has been used to identify SNPs in germline DNA and RNA from New Generation Sequencing (NGS) data for discover new variants and genotyping (LIU; SHEN; BAO, 2022). These variants can be SNP, small insertions/deletions (InDels) or larger structure variants such as copy number variations (CNV) (MIELCZAREK; SZYDA, 2016).

Advances in sequencing and computational methods have enabled faster and more accurate identification of genetic variants in human populations. (MONSU; COMIN, 2021). DNA sequencing (DNA-Seq) is the gold standard for SNP detection; however, RNA sequencing (RNA-Seq) has several advantages such as RNA editing analysis resulting in nucleotide changes observed at the transcriptome level, and provides several and numerous sets of SNPs.

NGS is very sensitive to errors and relies on bioinformatics tools such as the alignment of small reads to the reference genome and therefore SNP detection. This is why reliance on alignment accuracy is important, as incorrectly aligned readings can lead to errors in the SNP call. Alignment is more difficult in regions with high levels of diversity between the reference genome and the sequencing genome, however diversity can be improved with paired reads (HELYAR et al., 2011; NIELSEN et al., 2011).

Using the RNA-seq technique, it is possible to detect new variants related to the tissue-specific Transcriptome. it is also possible to identify SNPs that may be related to different characteristics of zootechnical interest, such as blood parameters, which, in turn, is a technique considered to be less invasive and that can provide information on polymorphisms associated with health parameters.

Thus, the objective of the present study was to identify single nucleotide polymorphisms (SNP) in the Transcriptome of muscle, brain and liver of Large White pigs (*Sus scrofa*) and, later, to verify the association of deleterious SNPs of the three tissues with the blood parameters of the blood, Glucose (mg/dL), Aspartate amino Transferase (U/L), Total proteins (g/dL), Albumin (g/dL), Globulin (g/dL), Triglycerides (mg/dL), Cholesterol (mg/dL), HDL (mg/dL), LDL (mg/dL) and VLDL (mg/dL).

2.2. Conclusion

In this study with large white pigs (*Sus scrofa*), using the RNA-seq technique, it was possible to identify variants with different types of sequence, among which are the 3'UTR region, Missense, Downstream gene and Upstream gene, between others. In addition, several classified variants have been identified in brain tissue, liver tissue, and skeletal muscle.

Deleterious variants were also identified in all tissues, of which two were associated with biochemical parameter triglyceride of the blood of the of 71 Large White pigs. One of the variants is related to the biological process of morphogenesis, and the cellular component of the extracellular matrix. While the other is related to von Willebrand factor, an essential glycoprotein for hemostasis. The results obtained reaffirm that the RNA-seq technique is an important tool in the detection of new variants and that the data generated from it can be used to improve the understanding of the genetic architecture of organisms, in addition to providing data for future research.

REFERENCES

AKEN, B. L. et al. The Ensembl gene annotation system. Database: the journal of biological databases and curation, v. 2016, 2016.

ALMEIDA, V. V. et al. Effects of increasing dietary oil inclusion from different sources on growth performance, carcass and meat quality traits, and fatty acid profile in genetically lean immunocastrated male pigs. Livestock Science, v. 248, p. 104515, 1 jun. 2021.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), v. 57, n. 1, p. 289–300, 1 jan. 1995.

BROWN, J.; PIRRUNG, M.; MCCUE, L. A. FQC Dashboard: Integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics, v. 33, n. 19, p. 3137–3139, 1 out. 2017.

COLOMBATTI, A.; BONALDO, P.; DOLIANA, R. Type A Modules: Interacting Domains Found in Several Non-Fibrillar Collagens and in Other Extracellular Matrix Proteins. Matrix, v. 13, n. 4, p. 297–306, 1 jul. 1993.

CONSORTIUM, T. U. et al. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research, v. 51, n. D1, p. D523–D531, 6 jan. 2023.

CUNNINGHAM, F. et al. Ensembl 2022. Nucleic Acids Research, v. 50, n. D1, p. D988–D995, 7 jan. 2022a.

CUNNINGHAM, F. et al. Ensembl 2022. Nucleic Acids Research, v. 50, n. D1, p. D988–D995, 7 jan. 2022b.

DANECEK, P. et al. The variant call format and VCFtools. Bioinformatics, v. 27, n. 15, p. 2156–2158, 2011.

DANECEK, P. et al. Twelve years of SAMtools and BCFtools. GigaScience, v. 10, n. 2, 1 fev. 2021.

DOBIN, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, v. 29, n. 1, p. 15–21, jan. 2013.

FANALLI, S. L. et al. Effect of dietary soybean oil inclusion on liver-related Transcription factors in a pig model for metabolic diseases. Scientific Reports 2022 12:1, v. 12, n. 1, p. 1–10, 20 jun. 2022a.

FANALLI, S. L. et al. Differential Gene Expression Associated with Soybean Oil Level in the Diet of Pigs. Animals, v. 12, n. 13, p. 1632, 1 jul. 2022b.

FANALLI, S. L. et al. RNA-seq Transcriptome profiling of pigs' liver in response to diet with different sources of fatty acids. Frontiers in Genetics, v. 14, p. 54, 25 jan. 2023a.

FANALLI, S. L. et al. Transcriptome profile of skeletal muscle using different sources of dietary fatty acids in male pigs. Functional & Integrative Genomics 2023 23:1, v. 23, n. 1, p. 1–19, 3 mar. 2023b.

FAO. Crops and livestock products. Disponível em: <https://www.fao.org/faostat/en/#data/QCL/visualize>. Acesso em: 25 fev. 2023.

FONSECA, P. A. S. et al. GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience, v. 9, n. 12, 1 dez. 2020.

FRANKE, K. R.; CROWGEY, E. L. Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. Genomics & Informatics, v. 18, n. 1, 1 mar. 2020.

FRANTZ, C.; STEWART, K. M.; WEAVER, V. M. The extracellular matrix at a glance. Journal of Cell Science, v. 123, n. 24, p. 4195–4200, 15 dez. 2010.

FUJII, J. et al. Identification of a Mutation in Porcine Ryanodine Receptor Associated with Malignant Hyperthermia. Science, v. 253, n. 5018, p. 448–451, 1991.

GAUDET, P. et al. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Briefings in bioinformatics, v. 12, n. 5, p. 449–462, set. 2011.

GRETCHEN HILL et al. Guide for the Care and Use of Agricultural Animals in Research and Teaching. 4. ed. [s.l: s.n.].

HARMEGNIES, N. et al. Results of a whole-genome quantitative trait locus scan for growth, carcass composition and meat quality in a porcine four-way cross. Animal Genetics, v. 37, n. 6, p. 543–553, 1 dez. 2006.

HELYAR, S. J. et al. Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. Molecular Ecology Resources, v. 11, n. SUPPL. 1, p. 123–136, 2011.

HUANG, Q. Q. et al. Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Research, v. 46, n. 22, p. e133, 12 dez. 2018.

KRUEGER, F. et al. FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. 2 fev. 2023.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics, v. 25, n. 16, p. 2078–2079, ago. 2009.

LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, v. 27, n. 21, p. 2987, 11 nov. 2011.

LIAO, Y. et al. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Research, v. 47, n. W1, p. W199–W205, 2 jul. 2019.

LIU, J.; SHEN, Q.; BAO, H. Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. PLOS ONE, v. 17, n. 1, p. e0262574, 1 jan. 2022.

MARTÍNEZ-MONTES, A. M. et al. Using RNA-Seq SNP data to reveal potential causal mutations related to pig production traits and RNA editing. Animal Genetics, v. 48, n. 2, p. 151–165, 1 abr. 2017.

MCKENNA, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, v. 20, n. 9, p. 1297–1303, 1 set. 2010.

MCLAREN, W. et al. The Ensembl Variant Effect Predictor. Genome Biology, v. 17, n. 1, p. 1–14, 6 jun. 2016.

MENDONÇA, T. Exportações brasileiras de carne suína Medidas técnicas, sanitárias e fitossanitárias. XXII Congresso de Iniciação Científica da Unesp, v. 26, p. 124–141, 2017.

MIELCZAREK, M.; SZYDA, J. Review of alignment and SNP calling algorithms for next-generation sequencing data. Journal of applied genetics, v. 57, n. 1, p. 71–79, 1 fev. 2016.

NIELSEN, R. et al. Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics 2011 12:6, v. 12, n. 6, p. 443–451, 18 maio 2011.

OCZKOWICZ, M. et al. Variant calling from RNA-seq data of the brain Transcriptome of pigs and its application for allele-specific expression and imprinting analysis. Gene, v. 641, p. 367–375, 30 jan. 2018.

OECD. Meat consumption. Disponível em: <https://doi.org/10.1787/44db9980-en>. Acesso em: 25 fev. 2023.

PAN, Z. et al. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. Nature Communications 2021 12:1, v. 12, n. 1, p. 1–15, 6 out. 2021.

PERKINS, S. J. et al. The Secondary Structure of the von Willebrand Factor type A Domain in Factor B of Human Complement by Fourier Transform Infrared Spectroscopy: Its Occurrence in Collagen Types VI, VII, XII and XIV, the Integrins and Other Proteins by Averaged Structure Predictions. Journal of Molecular Biology, v. 238, n. 1, p. 104–119, 21 abr. 1994.

REINER, G. et al. Quantitative trait loci for red blood cell traits in swine. Animal Genetics, v. 38, n. 5, p. 447–452, 1 out. 2007.

RUGGERI, Z. M.; WARE, J. von Willebrand factor. The FASEB Journal, v. 7, n. 2, p. 308–316, 1 fev. 1993.

STEVENSON, K. R.; COOLON, J. D.; WITTKOPP, P. J. Sources of bias in measures of allele-specific expression derived from RNA-seq data aligned to a single reference genome. BMC Genomics, v. 14, n. 1, p. 1–13, 7 ago. 2013.

TURCHETTO-ZOLET, A. C. et al. Marcadores Moleculares na Era genômica: Metodologias e Aplicações. [s.l: s.n.].

USDA. Livestock and poultry: world markets and trade. United States Department of Agriculture and Foreign Agricultural Service, p. 16, 2023.

VAN DER AUWERA, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.], v. 11, n. 1110, p. 11.10.1, 2013.

WARR, A. et al. An improved pig reference genome sequence to enable pig genetics and genomics research. bioRxiv, p. 668921, 24 out. 2019.

WATTRANG, E. et al. Confirmation of QTL on porcine chromosomes 1 and 8 influencing leukocyte numbers, haematological parameters and leukocyte function. Animal Genetics, v. 36, n. 4, p. 337–345, 1 ago. 2005.

WHITTON, C. et al. Are We Approaching Peak Meat Consumption? Analysis of Meat Consumption from 2000 to 2019 in 35 Countries and Its Relationship to Gross Domestic Product. Animals: an open access journal from MDPI, v. 11, n. 12, 1 dez. 2021.

YANG, J. et al. GCTA: A tool for genome-wide complex trait analysis. American Journal of Human Genetics, v. 88, n. 1, p. 76–82, 2011.

YE, M. et al. Transcriptomic analysis of chicken Myozenin 3 regulation reveals its potential role in cell proliferation. PloS one, v. 12, n. 12, 1 dez. 2017.

ZHAO, Y. et al. A high-throughput SNP discovery strategy for RNA-seq data. BMC Genomics, v. 20, n. 1, p. 1–10, 2019.

3. EFFECT OF DIFFERENT SET OF SNPS ON EQTL IDENTIFICATION ASSOCIATED WITH ECONOMIC TRAITS IN PIG PRODUCTION

3.1. Background

Pigs play an important role in the animal protein production scenario, with their meat being one of the most consumed worldwide (USDA, 2023). Thus, studies determining the impacts of genome variants on gene expression and phenotypes, related to production traits, such as feed efficiency, carcass yield, live weight, and composition are of great importance (Delpuech et al., 2021) to improve production efficiency in a sustainable way.

Genome-wide association studies (GWAS) based on single nucleotide polymorphisms (SNP) information and traits of economic interest focused on productive efficiency and meat quality have been extensively explored in recent years (Ellen et al., 2019; Ramayo-Caldas et al., 2019; Visscher et al., 2017; Võsa et al., 2018). Such studies allow the understanding of the genetic architecture involved with interest phenotypes. However, studies involving SNP within Transcriptome sequencing regions are recent and little explored.

The GWAS allow identifying genetic loci associated with performance and efficiency traits in the Longissimus dorsi muscle of pigs, such as carcass yield, live weight and composition, all traits related to meat quality (Kominakis et al., 2017; X. Liu et al., 2021a). Analyzing GWAS results, it is possible to determine which genes are implicated in a certain trait, as well as, the genetic variation that may contribute to the trait (X. Liu et al., 2021b). This information can then be used to identify candidate functional molecular markers for selecting animals with superior performance and meat quality (Visscher et al., 2017).

Among the SNPs from different regions, those present in coding regions are highly likely to change the level of global gene expression in the most diverse tissues present in the living organism. For example, a missense variant could result in the alteration of a codon that encodes a certain amino acid and, consequently, can lead to changes in protein synthesis and, consequently, in the functionality of these proteins in various tissues and physiological conditions of the organism (Moqa, Younas and Bashir, 2022; Zhao et al., 2019). Alternatively, when a SNP is present in promoter regions of a gene or 3' untranslated region (3'UTR), it can alter the level of expression and affect post-transcriptional regulations (Moqa, Younas, & Bashir, 2022). Thus, these mutations may be responsible for phenotypic differences among individuals in a population. In other words, SNPs may be associated with economically important traits such as animal performance and meat quality (Boison et al., 2015).

SNP mutations can be close (linked) and have a similar effect, that is, they can be in linkage disequilibrium (LD) with each other, that is, these variants can be inherited together and have the same effect on a phenotype, and can be represented then by only one of the variants (Arcos-Burgos & Muenke, 2002; Slatkin, 1994; Weir, 1979). In this case, it is common to use SNP pruning with the same effect or SNP tags in association analysis, where only one of the linked SNPs is kept (Arcos-Burgos & Muenke, 2002; Moqa et al., 2022; Slatkin, 1994; Wang et al., 2021; Zhang et al., 2022; al., 2022). In addition, according to Nyholt (2004), not performing pruning for the LD can lead to an overcorrection in inflated false positives, which can result in a decrease in the analysis power. However, there are few studies that address the identification of quantitative trait expression (eQTL) loci in pigs from SNPs data with LD pruning (Polizel et al., 2022a).

Skeletal muscle is associated with carcass traits, meat quality and is an important final product in pig production. However, studies involving data generated by sequencing skeletal muscle usually use only SNPs obtained from sequencing the skeletal muscle itself. Thus, it is necessary to understand the possible impacts of the

combination of variants derived from the sequencing of skeletal muscle, brain and liver tissues combined with the SNPs of medium density genotyping, on the level of gene expression in the skeletal muscle tissue of swine, which could help to elucidate possible comparative advantages and/or disadvantages of the combination or not of the SNPs in view of the results obtained. Then providing comparative information on the use of the combination of SNPs in the identification of cis and trans-eQTLs on the functional perspective of the analyzes of GO enrichment, metabolic pathways and enrichment of QTLs carried out later.

Therefore, we hypothesize that different sets of SNP data (scenarios) may affect the identification of eQTL associated with economic traits in pig production, which affects the expression level of genes associated with biological processes that could be related with the phenotypic variation. Based on it, our main objective was testing different sets of SNPs to identify eQTL associated with economic traits in pig production and biological processes related with these traits from a list of genes affected by identified eQTL. In this study, the evaluated phenotypes were slaughter weight in kg (SW), cold carcass yield in percentage (CCY), loin eye area in cm^2 measured by ultrasound (LEA), backfat thickness in cm^2 measured by ultrasound (BFT), muscle fat content in percentage (MFC).

3.2. Conclusions

We conclude that the different combinations of SNPs from the sequencing of muscle, brain, and liver tissues and from medium-density genotyping (50k) in large white pigs, resulted in different patterns of identification of eQTLs for the study of gene expression levels in skeletal muscle, which resulted in different functional enrichments for them. Furthermore, the combination of only the 50k genotyping data did not favor the identification of eQTLs different from those identified in the scenario where there were only SNPs from skeletal muscle sequencing.

It was also possible to observe that pruning for linkage disequilibrium removed the collinearity effects of the SNPs, which resulted in an improvement in the detection power of the analyzes in addition to reducing an overcorrection effect for multiple FDR tests. However, a larger number of samples is recommended to perform GWAS analyzes with the characteristics of slaughter weight, cold carcass yield in percentage, loin eye area measured by ultrasound, backfat thickness measured by ultrasound and muscle fat content in percentage.

This work can contribute to a better understanding of the genetic architecture of pigs, showing the impacts of using combinations of SNPs, the effect of pruning for LD, in addition to, indicating the genetic contexts in which the eQTLs are inserted through functional analysis.

REFERENCES

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., … Searle, S. M. J. (2016). The Ensembl gene annotation system. Database : The Journal of Biological Databases and Curation, 2016. doi: 10.1093/database/baw093

Almeida, V. v., Silva, J. P. M., Schinckel, A. P., Meira, A. N., Moreira, G. C. M., Gomes, J. D., … Cesar, A. S. M. (2021). Effects of increasing dietary oil inclusion from different sources on growth performance, carcass and meat quality traits, and fatty acid profile in genetically lean immunocastrated male pigs. Livestock Science, 248, 104515. doi: 10.1016/J.LIVSCI.2021.104515

Altmann, A., Weber, P., Bader, D., Preuß, M., Binder, E. B., & Müller-Myhsok, B. (2012). A beginners guide to SNP calling from high-Throughput DNA-sequencing data. Human Genetics, 131(10), 1541–1554. doi: 10.1007/s00439-012-1213-z

Arcos-Burgos, M., & Muenke, M. (2002). Genetics of population isolates. Clinical Genetics, 61(4), 233–247. doi: 10.1034/J.1399-0004.2002.610401.X

Ardlie, K. G., DeLuca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., … Lockhart. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235), 648–660. doi: 10.1126/science.1262110

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. doi: 10.1111/J.2517-6161.1995.TB02031.X

Boison, S. A., Santos, D. J. A., Utsunomiya, A. H. T., Carvalheiro, R., Neves, H. H. R., O'Brien, A. M. P., … da Silva, M. V. G. B. (2015). Strategies for single nucleotide polymorphism (SNP) genotyping to enhance genotype imputation in Gyr (Bos indicus) dairy cattle: Comparison of commercially available SNP chips. Journal of Dairy Science, 98(7), 4969–4989. doi: 10.3168/JDS.2014-9213

Bouwman, A. C., Visker, M. H. P. W., van Arendonk, J. A. M., & Bovenhuis, H. (2014). Fine mapping of a quantitative trait locus for bovine milk fat composition on Bos taurus autosome 19. Journal of Dairy Science, 97(2), 1139–1149. doi: 10.3168/jds.2013-7197

Bruce Alberts, Alexander Johnson, Julian Lewis, Morgan, D., Martin Raff, Keith Roberts, … Tim Hunt. (2017). Tradução :Biologia Molecular da Célula (6a).

Carmelo, V. A. O., & Kadarmideen, H. N. (2020). Genetic variations (eQTLs) in muscle transcriptome and mitochondrial genes, and trans-eQTL molecular pathways in feed efficiency from Danish breeding pigs. PLOS ONE, 15(9), e0239143. doi: 10.1371/JOURNAL.PONE.0239143

Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E., Fritz-Waters, E. R., Lanna, D. P. D., Gasparin, G., … Coutinho, L. L. (2015). Putative regulatory factors associated with intramuscular fat content. PLoS ONE, 10(6), e0128350. doi: 10.1371/journal.pone.0128350

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience, 4(1), 7. doi: 10.1186/S13742-015-0047-8/2707533

Chen, C., Yang, B., Zeng, Z., Yang, H., Liu, C., Ren, J., & Huang, L. (2013). Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. BMC Genomics, 14(1), 1–11. doi: 10.1186/1471-2164-14-848/TABLES/3

Chou, K. C. (2011). Some remarks on protein attribute prediction and pseudo amino acid composition. Journal of Theoretical Biology, 273(1), 236–247. doi: 10.1016/J.JTBI.2010.12.024

Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castelló, A., Fernández, A. I., & Folch, J. M. (2020). Identification of eQTLs associated with lipid metabolism in Longissimus dorsi muscle of pigs with different genetic backgrounds. Scientific Reports 2020 10:1, 10(1), 1–13. doi: 10.1038/s41598-020-67015-4

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., … Flicek, P. (2022). Ensembl 2022. Nucleic Acids Research, 50(D1), D988–D995. doi: 10.1093/NAR/GKAB1049

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., … Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. GigaScience, 10(2). doi: 10.1093/GIGASCIENCE/GIAB008

Delpuech, E., Aliakbari, A., Labrune, Y., Fève, K., Billon, Y., Gilbert, H., & Riquet, J. (2021). Identification of genomic regions affecting production traits in pigs divergently selected for feed efficiency. Genetics Selection Evolution, 53(1). doi: 10.1186/S12711-021-00642-1

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. doi: 10.1093/bioinformatics/bts635

Drag, M. H., Kogelman, L. J. A., Maribo, H., Meinert, L., Thomsen, P. D., & Kadarmideen, H. N. (2019). Characterization of eqtls associated with androstenone by rna sequencing in porcine testis. Physiological Genomics, 51(10), 488–499. doi: 10.1152/PHYSIOLGENOMICS.00125.2018/ASSET/IMAGES/LARGE/ZH700919437400007.JPEG

Ellen, E. D., van der Sluis, M., Siegford, J., Guzhva, O., Toscano, M. J., Bennewitz, J., … Bas Rodenburg, T. (2019). Review of sensor technologies in animal breeding: Phenotyping behaviors of laying hens to select against feather pecking. Animals, 9(3). doi: 10.3390/ani9030108

Fanalli, S. L., da Silva, B. P. M., Gomes, J. D., Ciconello, F. N., de Almeida, V. V., Freitas, F. A. O., … Cesar, A. S. M. (2022). Effect of dietary soybean oil inclusion on liver-related transcription factors in a pig model for metabolic diseases. Scientific Reports 2022 12:1, 12(1), 1–10. doi: 10.1038/s41598-022-14069-1

Fanalli, S. L., da Silva, B. P. M., Gomes, J. D., de Almeida, V. V., Freitas, F. A. O., Moreira, G. C. M., … Cesar, A. S. M. (2022). Differential Gene Expression Associated with Soybean Oil Level in the Diet of Pigs. Animals, 12(13), 1632. doi: 10.3390/ANI12131632/S1

Fanalli, S. L., Silva, B. P. M. da, Gomes, J. D., Durval, M. C., Almeida, V. V. de, Moreira, G. C. M., … Cesar, A. S. M. (2023). RNA-seq transcriptome profiling of pigs' liver in response to diet with different sources of fatty acids. Frontiers in Genetics, 14, 54. doi: 10.3389/FGENE.2023.1053021

Fonseca, P. A. S., Suárez-Vega, A., Marras, G., & Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience, 9(12). doi: 10.1093/GIGASCIENCE/GIAA149

Franke, K. R., & Crowgey, E. L. (2020). Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. Genomics & Informatics, 18(1). doi: 10.5808/GI.2020.18.1.E10

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. doi: 10.1038/NBT.1883

Hu, Z. L., Park, C. A., & Reecy, J. M. (2022). Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Research, 50(D1), D956–D961. doi: 10.1093/NAR/GKAB1116

Huang, Q. Q., Ritchie, S. C., Brozynska, M., & Inouye, M. (2018a). Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Research, 46(22), e133. doi: 10.1093/NAR/GKY780

Huang, Q. Q., Ritchie, S. C., Brozynska, M., & Inouye, M. (2018b). Power, false discovery rate and Winner's Curse in eQTL studies. Nucleic Acids Research, 46(22). doi: 10.1093/nar/gky780

Ibragimov, E., Pedersen, A. Ø., Xiao, L., Cirera, S., Fredholm, M., & Karlskov-Mortensen, P. (2022). Analysis of merged transcriptomic and genomic datasets to identify genes and pathways underlying residual feed intake in growing pigs. Scientific Reports, 12(1), 21946. doi: 10.1038/S41598-022-26496-1

Kang, M., & Gao, J. (2020). Integration of multi-omics data for expression quantitative trait loci (eQTL) analysis and eQTL epistasis. Methods in Molecular Biology, 2082, 157–171. doi: 10.1007/978-1-0716-0026-9_11/FIGURES/6

Kogelman, L. J. A., Zhernakova, D. v., Westra, H. J., Cirera, S., Fredholm, M., Franke, L., & Kadarmideen, H. N. (2015). An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. Genome Medicine, 7(1), 1–15. doi: 10.1186/S13073-015-0229-0

Kominakis, A., Hager-Theodorides, A. L., Zoidis, E., Saridaki, A., Antonakos, G., & Tsiamis, G. (2017). Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep. Genetics Selection Evolution, 49(1), 41. doi: 10.1186/s12711-017-0316-3

Krueger, F., James, F., Ewels, P., Afyounian, E., Weinstein, M., Schuster-Boeckler, B., … sclamons. (2023). FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. doi: 10.5281/ZENODO.7598955

Leal-Gutiérrez, J. D., Elzo, M. A., & Mateescu, R. G. (2020). Identification of eQTLs and sQTLs associated with meat quality in beef. BMC Genomics, 21(1), 1–15. doi: 10.1186/S12864-020-6520-5/TABLES/3

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987. doi: 10.1093/BIOINFORMATICS/BTR509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Research, 47(W1), W199–W205. doi: 10.1093/NAR/GKZ401

Liaubet, L., Lobjois, V., Faraut, T., Tircazes, A., Benne, F., Iannuccelli, N., … Cherel, P. (2011). Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: EQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. BMC Genomics, 12(1), 1–17. doi: 10.1186/1471-2164-12-548/FIGURES/6

Liu, H., Luo, X., Niu, L., Xiao, Y., Chen, L., Liu, J., … Yan, J. (2017). Distant eQTLs and Non-coding Sequences Play Critical Roles in Regulating Gene Expression and Quantitative Trait Variation in Maize. Molecular Plant, 10(3), 414–426. doi: 10.1016/J.MOLP.2016.06.016

Liu, J., Shen, Q., & Bao, H. (2022). Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. PLOS ONE, 17(1), e0262574. doi: 10.1371/JOURNAL.PONE.0262574

Liu, X., Zhang, J., Xiong, X., Chen, C., Xing, Y., Duan, Y., … Ma, J. (2021a). An Integrative Analysis of Transcriptome and GWAS Data to Identify Potential Candidate Genes Influencing Meat Quality Traits in Pigs. Frontiers in Genetics, 12, 1984. doi: 10.3389/FGENE.2021.748070/BIBTEX

Liu, X., Zhang, J., Xiong, X., Chen, C., Xing, Y., Duan, Y., … Ma, J. (2021b). An Integrative Analysis of Transcriptome and GWAS Data to Identify Potential Candidate Genes Influencing Meat Quality Traits in Pigs. Frontiers in Genetics, 12. doi: 10.3389/FGENE.2021.748070

Liu, Y., Liu, X., Zheng, Z., Ma, T., Liu, Y., Long, H., … Xu, X. (2020). Genome-wide analysis of expression QTL (eQTL) and allele-specific expression (ASE) in pig muscle identifies candidate genes for meat quality traits. Genetics Selection Evolution, 52(1), 1–11. doi: 10.1186/S12711-020-00579-X/TABLES/1

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297. doi: 10.1101/GR.107524.110

Mielczarek, M., & Szyda, J. (2016). Review of alignment and SNP calling algorithms for next-generation sequencing data. Journal of Applied Genetics, 57(1), 71–79. doi: 10.1007/S13353-015-0292-7

Moqa, R., Younas, I., & Bashir, M. (2022). Assessing effectiveness of many-objective evolutionary algorithms for selection of tag SNPs. PLOS ONE, 17(12), e0278560. doi: 10.1371/JOURNAL.PONE.0278560

Negro, S. S., Millet, E. J., Madur, D., Bauland, C., Combes, V., Welcker, C., … Nicolas, S. D. (2019). Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. BMC Plant Biology, 19(1), 1–22. doi: 10.1186/S12870-019-1926-4/TABLES/5

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. Nature Reviews Genetics 2011 12:6, 12(6), 443–451. doi: 10.1038/nrg2986

Nyholt, D. R. (2004). A Simple Correction for Multiple Testing for Single-Nucleotide Polymorphisms in Linkage Disequilibrium with Each Other. American Journal of Human Genetics, 74(4), 765–769. doi: 10.1086/383251

Polizel, G. H. G., Cesar, A. S. M., Cracco, R. C., Fernandes, A. C., Reginato, G. M., Xavier, P. L. P., … Santana, M. H. de A. (2022a). Identification of eQTLs and differential gene expression associated with fetal programming in beef cattle. Journal of Applied Genetics, 63(4), 717–728. doi: 10.1007/S13353-022-00711-1/TABLES/3

Polizel, G. H. G., Cesar, A. S. M., Cracco, R. C., Fernandes, A. C., Reginato, G. M., Xavier, P. L. P., … Santana, M. H. de A. (2022b). Identification of eQTLs and differential gene expression associated with fetal programming in beef cattle. Journal of Applied Genetics, 63(4), 717–728. doi: 10.1007/S13353-022-00711-1/TABLES/3

Ponsuksili, S., Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M., & Wimmers, K. (2015). Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. PLoS ONE, 10(4), e0123678. doi: 10.1371/journal.pone.0123678

Ponsuksili, S., Zebunke, M., Murani, E., Trakooljul, N., Krieter, J., Puppe, B., … Wimmers, K. (2015). Integrated Genome-wide association and hypothalamus eQTL studies indicate a link between the circadian rhythm-related gene PER1 and coping behavior. Scientific Reports 2015 5:1, 5(1), 1–14. doi: 10.1038/srep16264

Ramayo-Caldas, Y., Mármol-Sánchez, E., Ballester, M., Sánchez, J. P., González-Prendes, R., Amills, M., & Quintanilla, R. (2019). Integrating genome-wide co-association and gene expression to identify putative regulators and predictors of feed efficiency in pigs. Genetics Selection Evolution, 51(1). doi: 10.1186/S12711-019-0490-6

Raven, L. A., Cocks, B. G., Kemper, K. E., Chamberlain, A. J., Vander Jagt, C. J., Goddard, M. E., & Hayes, B. J. (2016). Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. Mammalian Genome, 27(1–2), 81–97. doi: 10.1007/s00335-015-9613-8

Revilla, M., Puig-Oliveras, A., Crespo-Piazuelo, D., Criado-Mesas, L., Castelló, A., Fernández, A. I., … Folch, J. M. (2018). Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions. Scientific Reports 2018 8:1, 8(1), 1–13. doi: 10.1038/s41598-018-20473-3

Shabalin, A. A. (2012). Matrix eQTL: Ultra-fast eQTL analysis via large matrix operations. Bioinformatics, 28(10), 1353–1358. doi: 10.1093/bioinformatics/bts163

Silva-Vignato, B., Cesar, A. S. M., Afonso, J., Moreira, G. C. M., Poleti, M. D., Petrini, J., … Coutinho, L. L. (2022). Integrative Analysis Between Genome-Wide Association Study and Expression Quantitative Trait Loci Reveals Bovine Muscle Gene Expression Regulatory Polymorphisms Associated with Intramuscular Fat and Backfat Thickness. Frontiers in Genetics, 13, 1811. doi: 10.3389/FGENE.2022.935238/BIBTEX

Slatkin, M. (1994). Linkage disequilibrium in growing and stable populations. Genetics, 137(1), 331–336. doi: 10.1093/GENETICS/137.1.331

USDA. (2023). Livestock and poultry: world markets and trade. United States Department of Agriculture and Foreign Agricultural Service, 16. Retrieved from https://apps.fas.usda.gov/psdonline/circulars/livestock_poultry.pdf

van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.], 11(1110), 11.10.1. doi: 10.1002/0471250953.BI1110S43

Vaughn, S. E. (2012). Review of the Third Edition of the Guide for the Care and Use of Agricultural Animals in Research and Teaching. Journal of the American Association for Laboratory Animal Science: JAALAS, 51(3), 298. Retrieved from /pmc/articles/PMC3358976/

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. American Journal of Human Genetics, 101(1), 5–22. doi: 10.1016/J.AJHG.2017.06.005

Võsa, U., Claringbould, A., Westra, H. J., Bonder, M. J., Deelen, P., Zeng, B., … Franke, L. (2018, October 19). Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. BioRxiv. bioRxiv. doi: 10.1101/447367

Wade, A. R., Duruflé, H., Sanchez, L., & Segura, V. (2022). eQTLs are key players in the integration of genomic and transcriptomic data for phenotype prediction. BMC Genomics, 23(1). doi: 10.1186/S12864-022-08690-7

Wang, S., Liu, G., Wang, X., Zhang, Y., He, S., & Zhang, Y. (2021). TagSNP-set selection for genotyping using integrated data. Future Generation Computer Systems, 115, 327–334. doi: 10.1016/J.FUTURE.2020.09.007

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., … Archibald, A. L. (2019). An improved pig reference genome sequence to enable pig genetics and genomics research. BioRxiv, 668921. doi: 10.1101/668921

Weir, B. S. (1979). Inferences about Linkage Disequilibrium. Biometrics, 35(1), 235. doi: 10.2307/2529947

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A Note on Exact Tests of Hardy-Weinberg Equilibrium. American Journal of Human Genetics, 76(5), 887. doi: 10.1086/429864

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. American Journal of Human Genetics, 88(1), 76–82. doi: 10.1016/j.ajhg.2010.11.011

You, N., Murillo, G., Su, X., Zeng, X., Xu, J., Ning, K., … Bateman, A. (2012). SNP calling using genotype model selection on high-throughput sequencing data. 28(5), 643–650. doi: 10.1093/bioinformatics/bts001

Zhang, Y., Wan, Q., Cheng, X., Lu, G., Wang, S., & He, S. (2022). A Tagging SNP Set Method Based on Network Community Partition of Linkage Disequilibrium and Node Centrality. Current Bioinformatics, 17(9), 825–834. doi: 10.2174/1574893617666220324155813

Zhao, Y., Wang, K., Wang, W. L., Yin, T. T., Dong, W. Q., & Xu, C. J. (2019). A high-throughput SNP discovery strategy for RNA-seq data. BMC Genomics, 20(1), 1–10. doi: 10.1186/s12864-019-5533-4

4. IDENTIFICATION OF EQTLS IN PIG LIVER

4.1. Introduction

Studying the transcriptome is of fundamental importance for a better understanding of the mechanisms of gene modulation. These studies, linked to functional enrichments, can reveal the genomic context in which the gene variants are inserted, gene modulation mechanisms and, consequently, their relationship with traits of interest for production, health, meat quality, and the level of gene expression.

SNPs are the main types of DNA polymorphisms used for studies of genetic variation. They are present throughout the genome, mainly in the intronic region, that is, non-coding regions (Hiremath et al., 2012). It also appears in gene coding sequences (exon) or in non-genetic coding regions (exon-intron splicing site) (Hiremath et al., 2012). SNPs in the coding regions can be divided into two types, synonymous SNPs, when there is no protein change, and non-synonymous SNPs, when the protein is changed (Zhao et al., 2019).

According to Zhao et al., (2019), the SNPs occuring in coding regions and regulatory sequences, respectively, can have a considerable impact on protein function and gene expression. This change in a nucleotide base can result in a change in the codon encoding an amino acid, thus altering protein synthesis. Or even this SNP when present in promoter regions of a gene or 3' non-transcribed region (3'UTR) can change the expression level and affect post-transcriptional regulations, respectively, as pointed out in several studies (Albert & Kruglyak, 2015; Buckingham & Relaix, 2015; Cesar et al., 2015; Siriluck Ponsuksili, Zebunke, et al., 2015). Thus, these SNPs may be causing phenotypic differences in different individuals of a population, that is, be associated with characteristics of zootechnical interest and animal and human health (pork consumer). According to Helyar et al., (2011) and Zhao et al., (2019), because SNPs have a high occurrence and extensive genome distribution, in genetic research, they are considered ideal for the characterization of genetic structure and identification of functional genes associated with traits of economic relevance.

In this context, ribonucleic acid nucleotide sequencing (RNA-seq) using next-generation sequencing technologies (NGS) allows quantifying levels of mRNA in an organism, and to estimate gene expression profiles at a given time (Carrillo et al., 2016; Dobin et al., 2013; Grabherr et al., 2011). The RNA-seq technique allows the measurement of gene expression on a large scale, which is of great importance for the advancement of our knowledge on functional genomics (Grabherr et al., 2011; S. Ponsuksili et al., 2010; Siriluck Ponsuksili, Siengdee, et al., 2015), as in addition to gene expression levels, it is possible to identify SNPs belonging to the transcribed genes, which gives them the characteristics of putative functional candidate variants (Zhao et al., 2019).

According to Harismendy et al. (2009), there are several NGS (Next Generation Sequencing) platforms, such as Illumina Genome Analyzer, Roche/454 FLX and ABI SOLiD, which have different technical specifications such as sensitivity, precision, repeatability and throughput, which means that the sequencing data obtained of such platforms will have specific limitations linked to the technique used.

Among the techniques for analyzing gene expression, the identification of eQTL (Expression Quantitative Tracer Locus) is widely used (Bahcall, 2015; Cesar et al., 2018; Criado-Mesas et al., 2020; Drag et al., 2019; Gibson, Powell, & Marigorta, 2015; Gilad, Rifkin, & Pritchard, 2008; Nodzak, 2020; Powder, 2020). According to Cesar et al., (2018), this approach can be used to discover the loci involved in the expression of a gene in a population which, in turn, are related to traits of interest. In pigs, eQTL analyzes have been shown to be useful in identifying genetic variants associated with different meat quality traits such as tenderness, flavor and nutritional

value. As described by Schomberg et al., (2016) and White et al., (2018), pigs can also be used as a model for studying diseases in humans due to the compatibility of genetic and biological mechanisms.

Segundo Subramanian et al., (2005), anotação e enriquecimento funcional de suínos é uma técnica importante para a avaliação de novos genes, variantes e processos biológicos que resultam na saúde e no bem-estar dos suínos. Esta técnica envolve a anotação de dados do genoma, como a identificação de novos genes e variantes, o mapeamento de localizações gênicas e a identificação de padrões funcionais para sequências de DNA/RNA, o que permite entender as instruções biológicas das variantes gênicas (Ashburner et al., 2000; Carbon et al., 2021).

Among the types of functional enrichment, Gene Ontology (GO) is a gene classification system that describes the biological function of genes and gene products. GO uses a hierarchical ontology to classify genes according to their function, process and cellular location called GO domains (Ashburner et al., 2000; Carbon et al., 2021; Subramanian et al., 2005). The ontology is a hierarchical structure of terms that describe relationships between various levels of abstraction. For example, "Metabolism" is a higher-level term that contains more specific terms such as "Citric Acid Cycle" and "Glycolysis" (Ashburner et al., 2000). According to Bettembourg, Diot, & Dameron, (2015), From the enriched GO terms of a set of target genes, it is also possible to identify metabolic pathways, which are essential for the maintenance of homeostasis and metabolism in complex organisms, as they allow substrates to be converted into products necessary for survival (Kanehisa & Goto, 2000).

The use of eQTL to detect GO terms and metabolic pathways is an approach used to identify the functioning mechanism of complex traits (Conesa et al., 2016). This approach is extremely useful to identify candidate biological mechanisms and to better understand how genes can be involved with complex diseases (Aguet et al., 2020; Conesa et al., 2016). In this context, the identification of eQTL (Expression Quantitative Trait Loci) from liver sequencing is of great importance to understand how genetic variation can affect gene expression and the regulation of metabolic processes. The analysis of eQTLs allows the identification of genetic polymorphisms that can affect gene expression and, consequently, alter the metabolic response of an individual. In addition, the analysis of eQTLs in pig liver can also be used as a model to identify genes associated with diseases, such as some liver diseases.

There are still a few gaps in knowledge regarding the genetic architecture of the expression modulating controls, and also, there are few studies reported using the liver transcriptome of large white pigs (*Sus scrofa*), with data obtained from the liver transcriptome of pigs using the technique of RNA-seq. In this context, the study of eQTLs is of fundamental importance to understand the impact of single-type variants that affect gene expression and, consequently, production traits, meat quality and health of pigs. The need for studies of the complete transcriptome in pig liver is evident to fill in gaps in knowledge regarding the regulatory control mechanisms of gene expression levels, and in the functional context. Thus, the objective of this work was to analyze data from the porcine liver transcriptome obtained by RNA-seq to identify cis and trans-eQTLs, in order to elucidate the functional context through the study of metabolic pathways, GO terms, and QTLs.

4.2. Conclusion

Cis and trans-eQTLs were identified from the liver tissue of 71 pigs generating a relatively large amount of cis-eQTLs (8,025) compared to the amount of trans-eQTL (132). With these data, QTL enrichment analyzes were performed for the genomic coordinates of cis and trans-eQTLs, allowing a contextualization of the relationship with several characteristics, based on previous studies, which were enriched for types of QTLs such as production, health, reproduction, "meat and carcass quality" in addition to exterior. Furthermore, from the enrichment for the go terms,

and metabolic pathways, we identified significant GO terms (FDR < 0.05) only for cis-eQTL, belonging to the domains of biological processes and molecular function. The same happened with metabolic pathways.

Thus, our studies bring important information about genetic architecture related to the transcriptome of pigs, in which it was possible to observe an overlap of our findings with several QTL associated with traits of zootechnical interest. As well as, we observed functional enrichments for GO domains and metabolic pathways The data generated in this study can serve as a basis for several studies in future works. However, due to the large volume of data, it was not possible to exhaust all discoveries.

REFERENCES

Aguet, F., Barbeira, A. N., Bonazzola, R., Brown, A., Castel, S. E., Jo, B., … Volpi, S. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science, 369(6509), 1318–1330. doi: 10.1126/SCIENCE.AAZ1776

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., … Searle, S. M. J. (2016). The Ensembl gene annotation system. Database: The Journal of Biological Databases and Curation, 2016. doi: 10.1093/database/baw093

Albert, F. W., & Kruglyak, L. (2015, April 19). The role of regulatory variation in complex traits and disease. Nature Reviews Genetics, Vol. 16, pp. 197–212. Nature Publishing Group. doi: 10.1038/nrg3891

Almeida, V. v., Silva, J. P. M., Schinckel, A. P., Meira, A. N., Moreira, G. C. M., Gomes, J. D., … Cesar, A. S. M. (2021). Effects of increasing dietary oil inclusion from different sources on growth performance, carcass and meat quality traits, and fatty acid profile in genetically lean immunocastrated male pigs. Livestock Science, 248, 104515. doi: 10.1016/J.LIVSCI.2021.104515

Ardlie, K. G., DeLuca, D. S., Segrè, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., … Lockhart. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235), 648–660. doi: 10.1126/science.1262110

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics, 25(1), 25. doi: 10.1038/75556

Bahcall, O. G. (2015, July 1). GTEx pilot quantifies eQTL variation across tissues and individuals. Nature Reviews Genetics, Vol. 16, p. 375. Nature Publishing Group. doi: 10.1038/nrg3969

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. doi: 10.1111/J.2517-6161.1995.TB02031.X

Bettembourg, C., Diot, C., & Dameron, O. (2015). Optimal threshold determination for interpreting semantic similarity and particularity: Application to the comparison of gene sets and metabolic pathways using GO and ChEBI. PLoS ONE, 10(7). doi: 10.1371/JOURNAL.PONE.0133579

Bryois, J., Buil, A., Evans, D. M., Kemp, J. P., Montgomery, S. B., Conrad, D. F., … Dermitzakis, E. T. (2014). Cis and Trans Effects of Human Genomic Variants on Gene Expression. PLOS Genetics, 10(7), e1004461. doi: 10.1371/JOURNAL.PGEN.1004461

Buckingham, M., & Relaix, F. (2015). PAX3 and PAX7 as upstream regulators of myogenesis. Seminars in Cell and Developmental Biology, 44, 115–125. doi: 10.1016/j.semcdb.2015.09.017

Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., … Elser, J. (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Research, 49(D1), D325–D334. doi: 10.1093/NAR/GKAA1113

Carmelo, V. A. O., & Kadarmideen, H. N. (2020). Genetic variations (eQTLs) in muscle transcriptome and mitochondrial genes, and trans-eQTL molecular pathways in feed efficiency from Danish breeding pigs. PLOS ONE, 15(9), e0239143. doi: 10.1371/JOURNAL.PONE.0239143

Carrillo, J. A., He, Y., Li, Y., Liu, J., Erdman, R. A., Sonstegard, T. S., & Song, J. (2016). Integrated metabolomic and transcriptome analyses reveal finishing forage affects metabolic pathways related to beef quality and animal welfare. Scientific Reports, 6(April), 1–16. doi: 10.1038/srep25948

Cesar, A. S. M., Regitano, L. C. A., Koltes, J. E., Fritz-Waters, E. R., Lanna, D. P. D., Gasparin, G., … Coutinho, L. L. (2015). Putative regulatory factors associated with intramuscular fat content. PLoS ONE, 10(6), e0128350. doi: 10.1371/journal.pone.0128350

Cesar, A. S. M., Regitano, L. C. A., Reecy, J. M., Poleti, M. D., Oliveira, P. S. N., de Oliveira, G. B., … Coutinho, L. L. (2018). Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. BMC Genomics, 19(1), 499. doi: 10.1186/s12864-018-4871-y

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience, 4(1), 7. doi: 10.1186/S13742-015-0047-8/2707533

Chen, C., Yang, B., Zeng, Z., Yang, H., Liu, C., Ren, J., & Huang, L. (2013). Genetic dissection of blood lipid traits by integrating genome-wide association study and gene expression profiling in a porcine model. BMC Genomics, 14(1), 1–11. doi: 10.1186/1471-2164-14-848/TABLES/3

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., … Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology, 17(1). doi: 10.1186/S13059-016-0881-8

Criado-Mesas, L., Ballester, M., Crespo-Piazuelo, D., Castelló, A., Fernández, A. I., & Folch, J. M. (2020). Identification of eQTLs associated with lipid metabolism in Longissimus dorsi muscle of pigs with different genetic backgrounds. Scientific Reports 2020 10:1, 10(1), 1–13. doi: 10.1038/s41598-020-67015-4

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., … Flicek, P. (2022). Ensembl 2022. Nucleic Acids Research, 50(D1), D988–D995. doi: 10.1093/NAR/GKAB1049

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., … Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. GigaScience, 10(2). doi: 10.1093/GIGASCIENCE/GIAB008

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics, 29(1), 15–21. doi: 10.1093/bioinformatics/bts635

Drag, M. H., Kogelman, L. J. A., Maribo, H., Meinert, L., Thomsen, P. D., & Kadarmideen, H. N. (2019). Characterization of eqtls associated with androstenone by rna sequencing in porcine testis. Physiological Genomics, 51(10), 488–499. doi: 10.1152/PHYSIOLGENOMICS.00125.2018/ASSET/IMAGES/LARGE/ZH70091943740007.JPEG

Fanalli, S. L., da Silva, B. P. M., Gomes, J. D., Ciconello, F. N., de Almeida, V. V., Freitas, F. A. O., … Cesar, A. S. M. (2022). Effect of dietary soybean oil inclusion on liver-related transcription factors in a pig model for metabolic diseases. Scientific Reports 2022 12:1, 12(1), 1–10. doi: 10.1038/s41598-022-14069-1

Fanalli, S. L., da Silva, B. P. M., Gomes, J. D., de Almeida, V. V., Freitas, F. A. O., Moreira, G. C. M., … Cesar, A. S. M. (2022). Differential Gene Expression Associated with Soybean Oil Level in the Diet of Pigs. Animals, 12(13), 1632. doi: 10.3390/ANI12131632/S1

Fanalli, S. L., Silva, B. P. M. da, Gomes, J. D., Durval, M. C., Almeida, V. V. de, Moreira, G. C. M., … Cesar, A. S. M. (2023). RNA-seq transcriptome profiling of pigs' liver in response to diet with different sources of fatty acids. Frontiers in Genetics, 14, 54. doi: 10.3389/FGENE.2023.1053021

Fonseca, P. A. S., Suárez-Vega, A., Marras, G., & Cánovas, Á. (2020). GALLO: An R package for genomic annotation and integration of multiple data sources in livestock for positional candidate loci. GigaScience, 9(12). doi: 10.1093/GIGASCIENCE/GIAA149

Franke, K. R., & Crowgey, E. L. (2020). Accelerating next generation sequencing data analysis: an evaluation of optimized best practices for Genome Analysis Toolkit algorithms. Genomics & Informatics, 18(1). doi: 10.5808/GI.2020.18.1.E10

Fujii, J., Otsu, K., Zorzato, F., De Leon, S., Khanna, V. K., Weiler, J. E., … Maclennan, D. H. (1991). Identification of a Mutation in Porcine Ryanodine Receptor Associated with Malignant Hyperthermia. Science, 253(5018), 448–451. doi: 10.1126/SCIENCE.1862346

Gibson, G., Powell, J. E., & Marigorta, U. M. (2015, June 24). Expression quantitative trait locus analysis for translational medicine. Genome Medicine, Vol. 7, p. 60. BioMed Central Ltd. doi: 10.1186/s13073-015-0186-7

Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. Trends in Genetics, 24(8), 408–415. doi: 10.1016/J.TIG.2008.06.001

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644–652. doi: 10.1038/NBT.1883

Harismendy, O., Ng, P. C., Strausberg, R. L., Wang, X., Stockwell, T. B., Beeson, K. Y., … Frazer, K. A. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biology, 10(3), 32. doi: 10.1186/gb-2009-10-3-r32

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., … Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. Molecular Ecology Resources, 11(SUPPL. 1), 123–136. doi: 10.1111/j.1755-0998.2010.02943.x

Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., … Varshney, R. K. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnology Journal, 10(6), 716. doi: 10.1111/J.1467-7652.2012.00710.X

Hu, Z. L., Park, C. A., & Reecy, J. M. (2022). Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Research, 50(D1), D956–D961. doi: 10.1093/NAR/GKAB1116

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27–30. doi: 10.1093/NAR/28.1.27

Kogelman, L. J. A., Zhernakova, D. v., Westra, H. J., Cirera, S., Fredholm, M., Franke, L., & Kadarmideen, H. N. (2015). An integrative systems genetics approach reveals potential causal genes and pathways related to obesity. Genome Medicine, 7(1), 1–15. doi: 10.1186/S13073-015-0229-0

Krueger, F., James, F., Ewels, P., Afyounian, E., Weinstein, M., Schuster-Boeckler, B., … sclamons. (2023). FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. doi: 10.5281/ZENODO.7598955

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987. doi: 10.1093/BIOINFORMATICS/BTR509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. doi: 10.1093/bioinformatics/btp352

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Research, 47(W1), W199–W205. doi: 10.1093/NAR/GKZ401

Liaubet, L., Lobjois, V., Faraut, T., Tircazes, A., Benne, F., Iannuccelli, N., … Cherel, P. (2011). Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: EQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. BMC Genomics, 12(1), 1–17. doi: 10.1186/1471-2164-12-548/FIGURES/6

Liu, X., Zhang, J., Xiong, X., Chen, C., Xing, Y., Duan, Y., … Ma, J. (2021). An Integrative Analysis of Transcriptome and GWAS Data to Identify Potential Candidate Genes Influencing Meat Quality Traits in Pigs. Frontiers in Genetics, 12, 1984. doi: 10.3389/FGENE.2021.748070/BIBTEX

Liu, Y., Liu, X., Zheng, Z., Ma, T., Liu, Y., Long, H., … Xu, X. (2020). Genome-wide analysis of expression QTL (eQTL) and allele-specific expression (ASE) in pig muscle identifies candidate genes for meat quality traits. Genetics Selection Evolution, 52(1), 1–11. doi: 10.1186/S12711-020-00579-X/TABLES/1

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297. doi: 10.1101/GR.107524.110

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics (Oxford, England), 26(16), 2069–2070. doi: 10.1093/BIOINFORMATICS/BTQ330

Nodzak, C. (2020). Introductory methods for EQTL analyses. Methods in Molecular Biology, 2082, 3–14. doi: 10.1007/978-1-0716-0026-9_1/TABLES/2

Ponsuksili, S., Murani, E., Phatsara, C., Schwerin, M., Schellander, K., & Wimmers, K. (2010). Expression quantitative trait loci analysis of genes in porcine muscle by quantitative real-time RT-PCR compared to microarray data. Heredity 2010 105:3, 105(3), 309–317. doi: 10.1038/hdy.2010.5

Ponsuksili, Siriluck, Siengdee, P., Du, Y., Trakooljul, N., Murani, E., Schwerin, M., & Wimmers, K. (2015). Identification of common regulators of genes in co-expression networks affecting muscle and meat properties. PLoS ONE, 10(4), e0123678. doi: 10.1371/journal.pone.0123678

Ponsuksili, Siriluck, Zebunke, M., Murani, E., Trakooljul, N., Krieter, J., Puppe, B., … Wimmers, K. (2015). Integrated Genome-wide association and hypothalamus eQTL studies indicate a link between the circadian rhythm-related gene PER1 and coping behavior. Scientific Reports 2015 5:1, 5(1), 1–14. doi: 10.1038/srep16264

Powder, K. E. (2020). eQTL Analysis (Vol. 2082). Retrieved from http://link.springer.com/10.1007/978-1-0716-0026-9%0Ahttp://link.springer.com/10.1007/978-1-0716-0026-9_15

Quan, J., Ding, R., Wang, X., Yang, M., Yang, Y., Zheng, E., … Yang, J. (2017). Genome-wide association study reveals genetic loci and candidate genes for average daily gain in Duroc pigs. Asian-Australasian Journal of Animal Sciences, 31(4), 480–488. doi: 10.5713/AJAS.17.0356

Quinn, E. M., Cormican, P., Kenny, E. M., Hill, M., Anney, R., Gill, M., … Morris, D. W. (2013). Development of Strategies for SNP Detection in RNA-Seq Data: Application to Lymphoblastoid Cell Lines and Evaluation Using 1000 Genomes Data. PLoS ONE, 8(3). doi: 10.1371/journal.pone.0058815

Raven, L. A., Cocks, B. G., Kemper, K. E., Chamberlain, A. J., Vander Jagt, C. J., Goddard, M. E., & Hayes, B. J. (2016). Targeted imputation of sequence variants and gene expression profiling identifies twelve candidate genes associated with lactation volume, composition and calving interval in dairy cattle. Mammalian Genome, 27(1–2), 81–97. doi: 10.1007/s00335-015-9613-8

Reiner, G., Clemens, N., Fischer, R., Köhler, F., Berge, T., Hepp, S., & Willems, H. (2009). Mapping of quantitative trait loci for clinical–chemical traits in swine. Animal Genetics, 40(1), 57–64. doi: 10.1111/J.1365-2052.2008.01804.X

Revilla, M., Puig-Oliveras, A., Crespo-Piazuelo, D., Criado-Mesas, L., Castelló, A., Fernández, A. I., … Folch, J. M. (2018). Expression analysis of candidate genes for fatty acid composition in adipose tissue and identification of regulatory regions. Scientific Reports 2018 8:1, 8(1), 1–13. doi: 10.1038/s41598-018-20473-3

Santiago, H., Luiz, R., Teixeira, F., Juarez, A., Donzele, L., Gomes, P. C., … Euclides, R. F. (2011). Composição de Alimentos e Exigências Nutricionais 3 a Edição Editor: Horacio Santiago Rostagno.

Schomberg, D. T., Tellez, A., Meudt, J. J., Brady, D. A., Dillon, K. N., Arowolo, F. K., … Shanmuganayagam, D. (2016). Miniature Swine for Preclinical Modeling of Complexities of Human Disease for Translational Scientific Discovery and Accelerated Development of Therapies and Medical Devices. Toxicologic Pathology, 44(3), 299–314. doi: 10.1177/0192623315618292

Shabalin, A. A. (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. Bioinformatics, 28(10), 1353–1358. doi: 10.1093/bioinformatics/bts163

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 102(43), 15545–15550. doi: 10.1073/PNAS.0506580102

van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.], 11(1110), 11.10.1. doi: 10.1002/0471250953.BI1110S43

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., … Archibald, A. L. (2019). An improved pig reference genome sequence to enable pig genetics and genomics research. BioRxiv, 668921. doi: 10.1101/668921

White, K. A., Swier, V. J., Cain, J. T., Kohlmeyer, J. L., Meyerholz, D. K., Tanas, M. R., … Weimer, J. M. (2018). A porcine model of neurofibromatosis type 1 that mimics the human disease. JCI Insight, 3(12). doi: 10.1172/JCI.INSIGHT.120402

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A Note on Exact Tests of Hardy-Weinberg Equilibrium. American Journal of Human Genetics, 76(5), 887. doi: 10.1086/429864

Yoo, C. K., Cho, I. C., Lee, J. B., Jung, E. J., Lim, H. T., Han, S. H., … Park, H. B. (2012). QTL analysis of clinical-chemical traits in an F2 intercross between Landrace and Korean native pigs. Physiological Genomics, 44(13), 657–668. doi: 10.1152/PHYSIOLGENOMICS.00172.2011

Yu, Y., Wei, J., Zhang, X., Liu, J., Liu, C., Li, F., & Xiang, J. (2014). SNP Discovery in the Transcriptome of White Pacific Shrimp Litopenaeus vannamei by Next Generation Sequencing. PLoS ONE, 9(1), 87218. doi: 10.1371/journal.pone.0087218

Zhao, Y., Wang, K., Wang, W. L., Yin, T. T., Dong, W. Q., & Xu, C. J. (2019). A high-throughput SNP discovery strategy for RNA-seq data. BMC Genomics, 20(1), 1–10. doi: 10.1186/s12864-019-5533-4

Zhernakova, D. V., Deelen, P., Vermaat, M., Van Iterson, M., Van Galen, M., Arindrarto, W., … Franke, L. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. Nature Genetics, 49(1), 139–145. doi: 10.1038/ng.3737