TRANSLATIONAL RESEARCH IN VETERINARY SCIENCE
Vol 2, No 1, 2019
DOI: http://dx.doi.org/10.12775/TRVS.2019.004
ISSN 2657-4489 (online)

RNA-seq based SNP discovery in *gluteus medius* muscle of Polish Landrace pigs

Mariusz Pierzchała[*1], Magdalena Ogłuszka[1], Dorota Goluch[1], Ewa Poławska[1], Taduesz Blicharski[1], Agnieszka Roszczyk[1], Agata Nawrocka[1], Paweł Urbański[1], Kamila Stepanow[1], Aleksandra Ciepłoch[1], Mateusz Sachajko[2], Edyta Juszczuk-Kubiak[#], Adrian Szczepański[#], Magdalena Buszewska-Forajta[#], Joanna Szczepanek[3], Chandra Shekhar Pareek[2,3]

[1] Institute of Genetics and Animal Breeding, Polish Academy of Sciences, Jastrzębiec, Poland
[2] Centre of Veterinary Sciences, Inter-University Centre of Veterinary Medicine, Nicolaus Copernicus University, Toruń, Poland
[3] Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University, Toruń, Poland
[#] Voluntary Authors

***Corresponding Author:**
Prof dr hab Mariusz Pierzchała, Head, Department of Genomics and Biodiversity, Institute of Genetics and Animal Breeding, Polish Academy of Sciences, ul. Postępu 36A, Jastrzębiec, 05-552 Magdalenka, Poland
Email: m.pierzchala@ighz.pl

Abstract

Background: Single nucleotide polymorphisms (SNPs) are the well-known molecular markers in genetics and breeding studies applied to veterinary sciences and livestock production. Advancement of next generation sequencing (NGS) provides a high-throughput means of potential putative SNP discovery. The aim of the study was to identify the putative genetic variants in *gluteus medius* muscle transcriptome of Polish Landrace pigs.

Methods. RNA-seq based NGS experiment was performed on Polish Landrace pigs fed with omega-6 and omega-3 polyunsaturated fatty acids (and normal diets. Isolation of total RNA from *gluteus medius* muscle was performed PUFAs dietary of Polish Landrace pigs. The RNA-seq libraries were constructed by mRNA enrichment, mRNA fragmentation, second strand cDNA synthesis, adaptor ligation, size selection and PCR amplification using the illumina TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego CA, USA), followed by NGS sequencing on MiSeq illumina platform. The quality control of raw RNA-seq data was performed using the Trimmomatic and FastQC tools. High QC paired-end RNA-seq data of *gluteus medius* muscle transcriptome were mapped to the reference genome *Sus scrofa* v.10.2. Finally, the SNPs discovery was performed using GATK and SAMtools bioinformatics SNPs caller tools.

Results: The Fastq RNA-seq data generated from two pooled paired-end libraries (151bp) of *gluteus medius* muscle tissue of Polish Landrace pigs were submitted to NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra). Study identified a total of 50.5 million paired-end reads (32.5 million low PUFAs dietary group and 18 million reads high PUFAs dietary group) of *gluteus medius* muscle transcriptome of Polish Landrace pigs. SNP discovery identified a total of 35436 homozygous and 28644 heterozygous cSNPs in *gluteus medius* muscle transcriptomes representing both dietary groups of Polish Landrace pig. Moreover, a total of 25187 and 5488 cSNP were identified as synonymous SNPs, and 18005 and 4780 cSNP were identified as nonsynonymous SNPs. Finally, single nucleotide variation (SNV) representing substitutions of all four possibilities (A,T,G,C) were identified ranging 2935 to 3227 SNVs (high PUFAs) and 3528 to 3882 SNVs (low PUFAs) for the heterozygous cSNPs and 2712 to 4058 (high PUFAs) and 4169 to 5692 SNVs (low PUFAs) for the heterozygous SNPs in *gluteus medius* muscle transcriptomes of Polish Landrace pigs.

Conclusions. Study concluded that identification of cSNPs dataset representing the *gluteus medius* muscle transcriptome of Polish Landrace pigs fed with a control diet (low) and pigs fed with a PUFAs diet (high) may be helpful to develop a new set of genetic markers specific to Polish Landrace pig breed. Such cSNP markers eventually can be utilized in genome-wide association studies (GWAS) and to finally implement on marker assisted selection (MAS) and genomics selection (GS) program in active breeding population of Polish Landrace pigs in Poland.

Keywords: Single nucleotide variations; SNPs; genetic markers; RNA-seq; illumina; paired-end read; GATK; SAMtools; mapping; bioinformatics; NGS; transcriptome; *gluteus medius* muscle; Landrace; pig; Omega-6 and omega-3 polyunsaturated fatty acids.

Background

With the rapid advancement in high-throughput (HT) next-generation genome sequencing (NGS), numerous studies were performed in last decade, to discovery the *de novo* and reference-based Single nucleotide polymorphisms (SNPs) discovery in cattle and pig domesticated animals [1-9]. The SNPs are generally the single nucleotide variations (SNVs) caused either by transitions (C/T or G/A) or transversions (C/G, C/A, or T/A, T/G), in the same position between individual genomic DNA sequences [10-11]. In transcriptome analysis studies, SNP is the dominant type of DNA markers for genetic variation studies of any genomes [12-13], representing intergenic region (regions between genes), coding region of genes (exons), or non-coding regions of genes (introns, 5'UTR, 3'UTR, or exon-intron splicing sites) [14]. SNPs in the coding region (cSNPs) can be categorized into two types: synonymous cSNPs (affecting coding region) and nonsynonymous cSNPs (affecting protein sequence). It has been proved that considerable effects on protein function and gene expression can be caused by the both synonymous and nonsynonymous cSNPs, particularly at the regulatory coding regions. That is why, cSNP as a genetic marker has great potential in genetics and breeding studies in veterinary sciences [3-4]. Due to the high density, scalability and genome-wide distribution, SNPs are considered as ideal genetic markers to characterize the genetic resources and functional genes associated with economic traits [15]. In animal genetics for instance, SNPs have been widely used to identify the trait-associated variation within the domesticated animals [7-9], and based on allele-specific assays and to discover genes linked to complex genetic traits in human [16-17].

The availability of NGS technologies also provides a convenient approach to discover all SNPs and obtain relevant information on genomic position and genotyping in a single step. Data from reliable large-scale sequencing, especially in domesticated animals could improve the cost-effectiveness and efficiency of detection of abundant SNPs. In NGS studies, several methods have been used for initial SNPs discovery in a high-

throughput manner, such as whole-genome sequencing, exome capture, RNA sequencing, methylated DNA sequencing, and restriction enzyme (RE) digestion [18-19]. In our study, we utilized the transcriptome sequencing using the RNA-seq NGS method. This method has become one of the most effective HT-NGS technique with high accuracy and cost-effective to discover wide range of novel genomics information including SNPs variations. The potential advantages to carry out SNP analysis using RNA-seq data includes i) identification of thousands of cSNPs and the expression levels of functional genes with sequence variations at a reasonable cost; ii) the location of variations in coding regions associated with the economic trait of domesticated animal's traits can be identified and the phenotypes can be predicted according to genotypes [20]. Moreover, RNA-seq is also a capable to perform the transcriptome studies such as gene characterization, gene expression quantification as well as post translational process analysis [21]. In this presented paper, we have performed the transcriptome sequencing of the *gluteus medius* muscle of Polish Landrace pigs and utilized the MiSeq illumina NGS platform known for its good sequencing coverage, read quality and SNP detection [3-4].

Results

Raw RNA-seq data of high PUFAs and low PUFAs dietary group: Based on the two pooled total RNA samples (high PUFAs and low PUFAs dietary groups) of *gluteus medius* muscle of Polish Landrace pigs, TruSeq-type libraries were paired-end sequenced using the Illumina MiSeq genomic sequencing device. The obtained raw RNA-seq data in the form of two *.fastq* format representing high PUFAs and low PUFAs dietary group of Polish Landrace pigs were submitted to NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra).

Transcriptome mapping of *gluteus medius* muscle of Polish Landrace pig to reference genome: Transcriptome sequencing of two pooled samples of *gluteus medius* muscle of Polish Landrace pigs were performed by

Illumina MiSeq™ NGS platform and paired-end read lengths of 151bp were sequenced. For each pooled sample of high PUFAs and low PUFAs dietary group of Polish Landrace pigs, the processed *fastq* data were mapped to the Sus scrofa genome in version 10.2 (https://may2017.archive.ensembl.org/Sus_scrofa/Info/Index). A summary of the paired-end fragments counts of *gluteus medius* muscle transcriptome of Polish Landrace pigs is shown in Table 1. A total of 50 million paired-end reads were identified. Obtained results showed paired-end fragments as well as mapping reads were as much as twice higher in the low PUFAs dietary group (32.5 million reads) in comparison to high PUFAs dietary group (approximately 18 million reads) of Polish Landrace pig. However, proportion of counted (63-68%) and uncounted (32-37%) fragments were equally distributed in both high and low PUFAs dietary groups of Polish Landrace pigs (Table 1).

Table 1. Paired-end fragments count and read count of gluteus medius muscle transcriptome of Polish Landrace pigs

Mapping parameters	High PUFAs group		Low PUFAs group	
	Paired-end fragments	%	Paired-end fragments	%
Counted fragments	5,743,574	64.08	10,296,956	63.24
Uncounted fragments	3,219,000	35.92	5,985,140	36.76
Total fragments	8,962,574	100.00	16,282,096	100.00
Reads mapped in pairs	11,487,148	64.08	20,593,912	63.24
Reads mapped in broken pairs	593,513	3.31	990,620	3.04
Reads not mapped	5,844,487	32.60	10,979,660	33.72
Total	17,925,148	100.00	32,564,192	100.00

Moreover, a detailed mapping statistics of *gluteus medius* muscle transcriptome of Polish Landrace pig is analyzed and presented in Table 2. Results revealed that *gluteus medius* muscle transcriptome of Polish

Landrace pig was successfully mapped to the Exon-exon and Exon-intron boundaries, total exon, total intron and total genes of the *sus scrofa* reference genome (Table 2).

Table 2. Detailed mapping statistics of the gluteus medius muscle transcriptome of Polish Landrace pig

RNA pools groups	High PUFAs group				Low PUFAs group			
	Mapped	% of total mapped	Uniquely mapped	Non-specifically mapped	Mapped	% of total mapped	Uniquely mapped	Non-specifically mapped
Exon--exon	3,286,851	57.23	3,124,464	162,387	5,913,311	57.43	5,626,532	286,779
Exon--in-tron	44,592	0.78	40,728	3,864	81,644	0.79	73,646	7,998
Total exon	5,511,985	95.97	5,139,879	372,106	9,868,048	95.83	9,228,075	639,973
Total intron	231,589	4.03	203,143	28,446	428,908	4.17	379,343	49,565
Total genes	5,743,574	100.00	5,343,022	400,552	10,296,956	100.00	9,607,418	689,538

Data filtration using Microsoft Excel and SNP discoveries: The processed paired-end reads of 151bp were assembled into contigs to the reference sus scrofa genome using the GATK [29] and SAMtools [30] pipelines to detect the novel cSNP in *gluteus medius* muscle transcriptome of Polish Landrace pigs. The detailed obtained results are presented in two supplementary tables `) and further summarized in Table 3. The obtained identified cSNPs results were classified into two types, the first one was homozygous type cSNP (cSNPs that were presented between PUFAs dietary groups but not between two alleles of an individual dietary group of Polish Landrace pig), and heterozygous types cSNP (cSNPs that were presented between two alleles of an individual PUFUs dietary group of Polish Landrace pig). As shown in Table 3, the number of heterozygous cSNPs were lower (12965 and 15679, respectively) than homozygous

(14439 and 20997, respectively) in *gluteus medius* muscle transcriptomes of both dietary groups of Polish Landrace pig. Similarly, the identified synonymous and nonsynonymous SNPs were higher in the homozygous (synonymous: 10488, 14699 and nonsynonymous: 2251, 3237, respectively) than in heterozygous (synonymous: 8213, 9792 and nonsynonymous: 2138, 2642, respectively). Furthermore, single nucleotide variation (SNV) representing substitutions of all four possibilities (A,T,G,C) were ranged from 2935 to 3227 SNVs (high PUFAs) and 3528 to 3882 SNVs (low PUFAs) for the heterozygous cSNPs and 2712 to 4058 (high PUFAs) and 4169 to 5692 SNVs (low PUFAs) for the heterozygous SNPs in *gluteus medius* muscle transcriptomes of Polish Landrace pigs (Table 3).

Table 3. Summary of SNP discoveries statistics based on the large dataset results illustrated in supplementary files sTable S1 and sTable S2

cSNPs category	**cSNPs statistics**	**Identified cSNPs in high PUFAs group (sTable S1)**	**Identified cSNPs in low PUFAs group (sTable S2)**
Homozygous cSNPs	Insertion	829	1216
	Deletion	616	924
	SNV	12994	18857
	Total homozygous cSNP	14439	20997
	cSNP changing coding region (synonymous SNPs)	10488	14699
	cSNP changing amino acid (nonsynonymous SNPs)	2251	3237
	SNV: T,G,C>A	2963	4324
	SNV: A,T,G>C	3930	5692
	SNV: A,T,C>G	4058	5629
	SNV: A,G,C>T	2712	4169

Table 3. Summary of SNP discoveries statistics (continuation)

cSNPs category	cSNPs statistics	Identified cSNPs in high PUFAs group (sTable S1)	Identified cSNPs in low PUFAs group (sTable S2)
Heterozygous cSNPs	Insertion	428	609
	Deletion	371	551
	SNV	12166	14519
	Total heterozygous cSNP	12965	15679
	cSNP changing coding region (synonymous SNPs)	8213	9792
	cSNP changing amino acid (nonsynonymous SNPs)	2138	2642
	SNV: T,G,C>A	3120	3782
	SNV: A,T,G>C	3227	3703
	SNV: A,T,C>G	2935	3528
	SNV: A,G,C>T	3147	3882

Discussions

The objective of presented study was to investigate the porcine muscle transcriptome and to explore the genetic variability in terms of SNPs/SNVs in the Polish Landrace pigs fed with omega-6 and omega-3 fatty acid. Our results provide novel cSNP transcriptomic data (https://www.ncbi.nlm.nih.gov/sra) of Polish Landrace breed to explain the regulatory mechanisms of the muscle tissue's biological response to increased concentration of omega-6 and omega-3 fatty acids (PUFAs) diets [22]. Recent studies have reported that consumption of PUFAs and other fatty acids diets were genetically regulated the growth performance and immune system in pigs [23]. In our study, the mapping results identified a total of 50.5 million paired-end reads (18 million in high PUFAs and 32.5 million low PUFAs group) in *gluteus medius* muscle transcriptome of Polish Landrace pig and were successfully mapped to the Exon-exon and Exon-

intron boundaries, total exon, total intron and total genes of the sus scrofa reference genome. A recent DNA-seq based study on the domestic Nero Siciliano male pig in Brazil [24], a total of 346.8 million paired reads were identified and after quality control 99.03% of the reads were successfully mapped to the reference genome. In this paper, we have identified large set of both homozygous and heterozygous cSNP dataset that were regulated in the both high PUFAs (12965 and 15679, respectively) and low PUFAs (14439 and 20997, respectively) dietary groups of Polish Landrace pigs. Study on a male domestic Nero Siciliano pig in Brazil compared to the last pig reference genome *Sus scrofa11.1*, identified over 11 million SNPs, and a total of 6,747 genetic variants at a rate of 1 variant every ~276 bases, and among these variants 1,132 were novel to the dbSNP151 database [24]. Among the strengths of the present study is that the scale of RNA-seq based NGS analysis allowed the cSNP/cSNV discovery in Polish Landrace, However, the limitation of lack of causal mutations within the coding region as well as animal acid substitutions should be considered in the interpretation of most of the finding results.

Conclusions

In our study, we have identified two cSNPs dataset representing the *gluteus medius* muscle transcriptome of Polish Landrace pigs fed with a control diet (low) and pigs fed with a PUFAs diet (high). Identification of cSNPs in *gluteus medius* muscle may be helpful to develop a set of candidate genetic markers specific to Polish Landrace pig breed, which eventually can be utilized in genome-wide association studies (GWAS), and to finally implement on marker assisted selection (MAS) and genomics selection (GS) program in active breeding population of Polish Landrace pigs in Poland.

Materials and Methods

Animals: The *gluteus medius* muscle samples of Polish Landrace pigs for RNA-seq based NGS experiment were selected based on fatty acid profiles in order to compare muscle transcriptomes characterized by low (control groups) and high (PUFAs enriched diet groups) omega-6/omega-3 fatty acids ratio in Polish Landrace pigs. The threshold value of < 7 for low and > 7 for high omega-6/omega-3 fatty acids ratio was used in this study [22]. After slaughtering the *gluteus medius* muscle tissues were collected at the Institute IGHZ, Jastrzębiec and were frozen in liquid nitrogen and stored at – 80 °C for further RNA extraction. The number of samples used in RNA-seq based NGS experiment were 6 each for high and low PUFAs group of Polish Landrace purebred.

RNA extraction: Total RNA was extracted from 20-30 mg of *gluteus medius* muscle of Polish Landrace purebred pigs. The total RNAs (n=12) were isolated by a guanidine protocol as described by Chomczynski and Sachi *et al*. [25] using RNeasy Lipid Tissue Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions and treated with DNA-free DNA Removal kit (Invitrogen, Life Technologies, CA, USA) to remove DNA contamination. The concentration and purity of RNA samples were assessed by Qubit (Life Technologies, CA, USA). RNA purity was estimated with the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) and integrity was evaluated by the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA). The high quality extracts (RIN ≥ 8) were used for cDNA libraries synthesis.

Library construction and transcriptome sequencing: The construction of cDNA libraries was performed following mRNA enrichment, mRNA fragmentation, second strand cDNA synthesis, adaptor ligation, size selection and PCR amplification using illumina TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego CA, USA) protocol according to the manufacturer's recommendations. Briefly, 1μg of total RNA was used as an input

for poly(A) selection for enrichment of messenger RNA (mRNA). The purified extracts of mRNA were pooled together within each group and equal amount of mRNA were used to ensure the most robust transcriptome. The synthesis of libraries for each dietary group experimental group combination were prepared in duplicate (technical replicates) represented as two for control (normal diet) groups and two for experimental (PUFAs diet) groups. The pooled cDNA libraries were loaded onto the flow cell channels of Illumina MiSeq platform to proceed the NGS sequencing of *gluteus medius* muscle transcriptome using standard Illumina sequencing kit protocols according to the manufacturer's recommendations. The RNA-seq experiment yielded the paired-end sequencing reads of 151bp.

Preprocessing of the RNA-seq data: After NGS sequencing, two RNA-seq raw data were obtained. Illumina pipeline were used for image acquisition and base-calling. Raw reads were first filtered and discarded for mistaken base calling. Raw reads were then processed by adaptor trimming and removing of low-quality bases. Very short reads (less than 20 nt) were also dropped. The preprocessing of the raw RNA-seq data was performed using Trimmomatic tool [26] and included trimming of adaptor sequences and read filtering based on Illumina quality score (QS) after cleaning and correcting the abundance of contaminating sequences, average read length, GC content, presence of adaptors or overrepresented k-mers. To ensure reliability of the resulting data, its characteristics were inspected using FastQC tool [27]. Finally, RNA-seq data were achieved as a FastQ files type containing processed raw reads with high quality score [28].

Transcriptome mapping and SNP detection: The raw RNA-seq reads imported from Illumina were uploaded into CLC Genomics workbench v. 6.0 (CLC Bio, Aarhus, Denmark). The reference genome *Sus scrofa* v.10.2 was downloaded from pig reference genome assembly website resources (https://www.ensembl.org/Sus_scrofa/Info/Index) and RNA-seq reads of 151bp in lengths were mapped onto the reference genome. Subsequently, were applied to perform the. After read alignment, SNP detec-

tion was performed using GATK [29] and SAMtools [30] filter parameters (read depth no less than 10, quality score no less than 20, consecutive single base errors no more than 3 in 35bp, FS 20.0 window 25) and Only SNPs after sequential depth standardization with quality value >2 were retained.

Acknowledgement

This study was funded by the project "BIOFOOD – innovative, functional products of animal origin" No. POIG.01.01.02-014-090/09 and was co-financed by the European Union from the European Regional Development Fund under Innovative Economy Operational Program 2007–2013.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files. The transcriptome sequencing raw Fastq data has been deposited in NCBI SRA web resources.

Supplement file 1:
https://box.pionier.net.pl/f/ca6288d74f134b9782af/?dl=1
Supplement file 2:
https://box.pionier.net.pl/f/fbd87dccd70143afacf3/?dl=1

References

1. Zwane AA, Schnabel RD, Hoff J, Choudhury A, Makgahlela ML, Maiwashe A, Van Marle-Koster E, Taylor JF. Genome-Wide SNP Discovery in Indigenous Cattle Breeds of South Africa. Front Genet. 2019; 10: 273.
2. Wang W, Gan J, Fang D, Tang H, Wang H, Yi J, Fu M. Genome-wide SNP discovery and evaluation of genetic diversity among six Chinese indigenous cattle breeds in Sichuan. PLoS One. 2018; 13:e0201534.

3. Pareek CS, Błaszczyk P, Dziuba P, Czarnik U, Fraser L, Sobiech P, Pierzchała M, Feng Y, Kadarmideen HN, Kumar D. Single nucleotide polymorphism discovery in bovine liver using RNA-seq technology. PLoS One. 2017; 12:e0172687.
4. Pareek CS, Smoczyński R, Kadarmideen HN, Dziuba P, Błaszczyk P, Sikora M, Walendzik P, Grzybowski T, Pierzchała M, Horbańczuk J, Szostak A, Ogluszka M, Zwierzchowski L, Czarnik U, Fraser L, Sobiech P, Wąsowicz K, Gelfand B, Feng Y, Kumar D. Single Nucleotide Polymorphism Discovery in Bovine Pituitary Gland Using RNA-Seq Technology. PLoS One. 2016; 11:e0161370.
5. Doran AG, Creevey CJ. Snpdat: easy and rapid annotation of results from de novo snp discovery projects for model and non-model organisms. BMC Bioinformatics. 2013; 14:45.
6. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498.
7. Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, Moore SS. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics. 2011; 12:559.
8. Zhang Q, Zhao S, Chen H, Zhang L, Zhang L, Li F, Wang X. SNP discovery and haplotype analysis in the bovine PRKAA2 gene. Mol Biol Rep. 2011; 38:1551–6.
9. Cánovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano JF. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm Genome. 2010; 21:592-598.
10. Brookes AJ. The essence of SNPs. Gene. 1999; 234:177–186.
11. Trick M, Long Y, Meng J, Bancroft I. Single nucleotide polymorphism (SNP) discovery in the polyploidy Brassica napus using Solexa transcriptome sequencing. Plant Biotechnol J. 2009; 7:334–346.
12. Jehan T, Lakhanpaul S. Single nucleotide polymorphism (SNP) – methods and applications in plant genetics: a review. Indian J Biotechnol. 2006; 5:435–459.
13. Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD, et al. Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. Plant Biotechnol J. 2012; 10:1–17.
14. Garrido-Cardenas JA, Mesa-Valle C, Manzano-Agugliaro F. Trends in plant research using molecular markers. Planta. 2018; 247:543–557.

15. Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW. Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol Ecol Resour. 2011; 11:1–8.
16. Suh Y, Vijg J. SNP discovery in associating genetic variation with human disease phenotypes. Mutat Res. 2005; 573:41-53.
17. Stouffer K, Nahorski M, Moreno P, Sarveswaran N, Menon D, Lee M, Geoffrey Woods C. Functional SNP allele discovery (fSNPd): an approach to find highly penetrant, environmental-triggered genotypes underlying complex human phenotypes. BMC Genomics. 2017; 18:944.
18. Kumar S, Banks TW, Cloutier S. SNP discovery through next-generation sequencing and its applications. Int J Plant Genomics. 2012; 2012:1–15.
19. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 2011; 12:499–510.
20. Pareek CS, Smoczynski R, Pierzchala M, Czarnik U, Tretyn A. From genotype to phenotype in bovine functional genomics. Brief Funct Genomics. 2011; 10:165-171.
21. Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, Morris DW. Development of strategies for SNP detection in RNA-Seq data: application to lymphoblastoid cell lines and evaluation using 1000 genomes data. PLoS One. 2013; 8: e58815.
22. Ogłuszka M, Szostak A, Te Pas MFW, Poławska E, Urbański P, Blicharski T, Pareek CS, Juszczuk-Kubiak E, Dunkelberger JR, Horbańczuk JO, Pierzchała M. A porcine *gluteus medius* muscle genome-wide transcriptome analysis: dietary effects of omega-6 and omega-3 fatty acids on biological mechanisms. Genes Nutr. 2017; 12: 4.
23. Huber LA, Hooda S, Fisher-Heffernan RE, Karrow NA, de Lange CFM. Effect of reducing the ratio of omega-6-to-omega-3 fatty acids in diets of low protein quality on nursery pig growth performance and immune response. J Anim Sci. 2018; 96: 4348-4359.
24. D'Alessandro E, Giosa D, Sapienza I, Giuffrè L, Cigliano RA, Romeo O, Zumbo A. Whole genome SNPs discovery in Nero Siciliano pig. Genet Mol Biol. 2019; pii: S1415-47572019005021102.
25. Chomczynski P., Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 1987; 162:156–159.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120.
27. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. Version 2. F1000Res. 2018; 7 :1338.
28. Pareek CS, Sachajko M, Szczepański A, Buszewska-Forajta M, Zarczynska K, Sobiech P, Juszczuk-Kubiak E, Shahzad Q, Lu YQ, Ogłuszka M, Poławska E, Pierzchała M. Quality control assessment of the RNA-Seq data generated from liver and pituitary transcriptome of Hereford bulls using StrandNGS software. Trans Res Vet Sci. 2019; 2: 9-22. DOI: http://dx.doi.org/10.12775/TRVS.2019.001.
29. do Valle ÍF, Giampieri E, Simonetti G, Padella A, Manfrini M, Ferrari A, Papayannidis C, Zironi I, Garonzi M, Bernardi S, Delledonne M, Martinelli G, Remondini D, Castellani G. Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole--exome sequencing data. BMC Bioinformatics. 2016; 17: 341.
30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup The sequence alignment/map format and SAMtools. Bioinformatics. 2009; 25: 2078–2079.