



Data set for transcriptome analysis of pituitary gland in cattle breeds

Herudzińska M.¹, Sachajko M.¹, Jaskowski J.M.¹, Szczepanek J.²,
Buszewska-Forajta M.[‡], Fang Y.³, Kumar D.³, Pareek C.S.^{1,2*}

¹ Institute of Veterinary Medicine, Faculty of Biological and Veterinary Sciences,
Nicolaus Copernicus University, Torun, Poland.

² Division of Functional Genomics in Biological and Biomedical research, Centre for
Modern Interdisciplinary Technologies, Nicolaus Copernicus University, Torun, Poland

³ Waksman Institute of Microbiology, Rutgers, The State University
of New Jersey, Piscataway, NJ 08 854, USA.

[‡] Voluntary author, Institute of Veterinary Medicine, Faculty of Biological
and Veterinary Sciences, Nicolaus Copernicus University, Torun, Poland.

Corresponding author(s):

Chandra S. Pareek, Institute of Veterinary Medicine, Faculty of Biological and Veterinary
Sciences, Nicolaus Copernicus University, Torun, Poland. Email: pareekcs@umk.pl

Abstract

Transcriptome data presented in this article is associated with the research article entitled
“Single nucleotide polymorphism discovery in bovine pituitary gland using RNA-seq tech-
nology” published in PLOS One [1]. Herein, we provide raw and analysed RNA-seq data
of pituitary gland tissues from three cattle breeds, viz., Polish-HF, Polish Red and Hereford

cattle breeds. Bioinformatics pipelines of high quality RNA-seq data includes the FastQC tools for quality controls, Trimmomatic cutadapt tools for trimming RNA-seq data, and BWA version 0.7.5-r404 for mapping and alignment to the *Bos taurus* reference genome, SAMtools for SNPs identifications in bovine pituitary gland transcriptome. Raw FASTq files for the RNA-seq libraries of bovine pituitary gland were deposited in the NCBI Sequence Read Archive (SRA) and have been assigned BioProject accession PRJNA312148.

Keywords: RNA-seq; cattle; pituitary gland; breeds; NGS; SNPs; SAMtools; BWA; FastQC; SRA; NCBI.

I. Omics database specifications table

Subject/category	Veterinary science, Animal science
Type / source of Omics data	Bovine pituitary gland transcriptome, linked to the external NCBI resources
How data were acquired, equipment, and technology	The transcriptome data were acquired using illumina NGS Next-seq 500 Sequencer.
Omics data format	Raw data (FASTq); Analysed data (SNPs); Filtered data (SNPs).
Omics experimental factors	Experiment animals: A total of 18 young bulls aged 6, 9, and 12 month from Polish-HF, Polish Red and Hereford cattle breeds.
Omics experimental features	Comparisons of three cattle breeds to identify SNPs and validation of identified SNPs expressed in bovine pituitary gland transcriptome.
Omics data source location	Waksman institute of Microbiology, Rutgers, The state university of New jersey, Piscataway, NJ 08 854, USA. Institute of veterinary Medicine, Faculty of Biological and Veterinary Science, Nicolaus Copernicus University, Torun, Poland.
Omics data accessibility to articles and at public repository	Accessibility with the article: The provided data accompanying the manuscript described the SNPs identification in bovine pituitary gland transcriptome of cattle breeds [1]. Accessibility public repository name: NCBI resources. Data identification number: SRA: PRJNA312148. Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA312148
Literature review [Not mandatory]: short description and citation any related articles.	[2] Zhao <i>et al.</i> 2019. [3] Djari <i>et al.</i> 2013. [4] Stothard <i>et al.</i> 2011. [5] Cánovas <i>et al.</i> 2010.

II. Importance and significance of the Omics data to research community

The identified putative SNPs could serve as potential trait-associated markers for bovine postnatal body growth trait. Published transcriptomic data can be useful to the bovine metabolome and hormonome research community and can be integrated to global and international consortium database, viz., bovine gene atlas (omicsX: <https://omictools.com/bovine-gene-atlas-tool>), functional annotation of animal genomes [FAANG: <https://www.animalgenome.org/community/FAANG/>].

III. Experimental Design, Materials, and Methods

Sample collection: Bovine pituitary gland tissues samples from 18 young bulls aged between 6 to 12 months in a panel of three selected cattle breeds: Polish HF, Polish Red, and Hereford were collected. All experimental animals were reared at Institute of Genetics and Animal Breeding, Jastrzębiec, Poland in a closed herd, and providing uniform feeding and environmental conditions.

Experimental procedure: Total RNA from all pituitary gland samples were extracted by TRIzol reagent. The mRNA was isolated by using the Dynabeads[®] mRNA Direct[™] kit (Thermo Fisher). The dUTP directional mRNA libraries preparation was performed by chemical hydrolysis, converted to first strand cDNA with random hexamers, and second strand synthesized with dUTP according to the NEBNext Ultra Directional RNA library preparation Kit for Illumina (New England Bio Labs). The cDNA fragments were end-repaired, A-tailed, and ligated to the TruSeq y-tail single indexes from Illumina TruSeq DNA kit. The indexed libraries were cut with USER enzyme, and PCR amplified for 12 cycles, followed by quantitation of libraries using qPCR according to the Illumina Sequencing Library qPCR Quantification Guide (Kapa Biosciences). The Omics

data with 156x156 bp paired-end sequence reads were generated using the Illumina NextSeq 500 platform High Output/300 cycle kits from Illumina.

Raw Omics data: Using cutadapt software, the raw data (FASTq) were obtained by removing the adaptor sequences (<https://cutadapt.readthedocs.io/en/stable/>) with minimum overlap length was set to 10 and error rate was set to 0.05. The trimming of the low quality bases at 3'- end were performed by using sequence quality control (FastQC).

Processed and filtered Omics data: The alignment of the paired-end reads on the *Bos taurus* reference genome Ensembl75_UMD3-1.1 plus the Chromosome Y from Btau_4.6.1 assembly was done by using BWA version 0.7.5-r404 [6]. The HT-Seq framework, version 0.5.3p9 (<https://pypi.python.org/pypi/HTSeq/0.5.3p7>), was used to count the aligned reads using the STAR BWA tools [7]. The SNPs data was processed by using SAMtools mPileUp package to call SNPs and indels [8] and filtered by using Microsoft Office Excel.

Acknowledgments

This research paper was financially supported by National Science Centre, Krakow, Poland, Project No. 2012/05/B/NZ2/01629, entitled “Analiza transkryptomów genomu *Bos taurus* przy zastosowaniu technologii sekwencjonowania kolejnej generacji”.

References

- [1] Pareek C.S., Smoczyński R., Kadarmideen H.N., Dziuba P., Błaszczuk 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.
- [2] Zhao Y., Wang K., Wang W.L., Yin T.T., Dong W.Q., Xu C.J. A high-throughput SNP discovery strategy for RNA-seq data. *BMC Genomics*. 2019;20:160.
- [3] Djari A., Esquerré D., Weiss B., Martins F., Meersseman C., Boussaha M., Klopp C., Rocha D. Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics*. 2013;14:307.
- [4] Stothard P., Choi J.W., Basu U., Sumner-Thomson J.M., Meng Y., Liao X., Moore S.S. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics*. 2011;12:559.
- [5] Cánovas A., Rincon G., Islas-Trejo A., Wickramasinghe S., Medrano J.F. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm Genome*. 2010;21:592–8.
- [6] Li H., Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
- [7] Anders S., Pyl P.T., Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 2015;1:166–169.
- [8] Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., et al. 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.