TRANSLATIONAL RESEARCH IN VETERINARY SCIENCE
Vol 2, No 2, 2019
DOI: http://dx.doi.org/10.12775/TRVS.2019.009
ISSN 2657-4489 (online)

Data set for transcriptome analysis of liver in cattle breeds

Sachajko M.[1], Herudzińska 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 analysis using high-throughput next-generation sequencing (HT-NGS) technology provides the capability to understand global gene expression variations through a wide range of tissue samples in domesticated animals. We provide raw and analysed data

for transcriptomic analysis of liver tissues from Polish-HF, Polish Red and Hereford cattle breeds, obtained by RNA-seq. High quality sequencing data have been analysed using our bioinformatics pipeline which consists of FastQC for quality controls, Trimmomatic for trimming, and BWA version 0.7.5-r404 for alignment to the Bos taurus reference genome, SAMtools for SNPs identifications, and differentially expressed genes (DEGs) identification using DEseq and edgeR pipelines after adjustment for false-discovery rate (FDR) with adjusted two sided p values <0.01 and the trimmed mean of M values (TMM) normalisation method. The data accompanying the published manuscript describing the SNPs and DEGs identification in the bovine liver transcriptome of cattle breeds. Raw FASTq files for the RNA-seq libraries are deposited in the NCBI Sequence Read Archive (SRA) and have been assigned BioProject accession PRJNA312148. Raw and processed RNA-seq data were deposited and made publicly available on the Gene Expression Omnibus (GEO; GSE114233).

Keywords: RNA-seq; cattle; liver; breeds; NGS; SNPs; DEGs; DEseq; EdgeR; FDR; SAMtools; BWA; FASTq; SRA; NCBI; GEO.

I. Omics database specifications table

Subject/category	Veterinary science, Animal science
Type / source of Omics data	Bovine liver 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, DEGs), and Filtered data (SNPs, DEGs).
Omics experimental factors	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 DEGs.
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.

Subject/category	Veterinary science, Animal science
Omics data accessibility to articles and at public repository	**Accessibility with the article:** The provided data accompanying the manuscript described the SNPs identification [1] and DEGs identification [2] in bovine liver transcriptome of cattle breeds. **Accessibility public repository:** Repository name: NCBI resources. Data identification number: SRA: PRJNA312148. Data identification number: GEO: GSE114233). **Direct URL to data:** https://www.ncbi.nlm.nih.gov/bioproject/PRJNA312148 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114233
Literature review [Not mandatory]: short description and citation any related articles.	[3]. Lisowski *et al*. 2014. [4]. Salleh *et al*. 2017. [5]. Salleh *et al*. 2018.

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

The identified putative SNPs and DEGs could serve as potential trait-associated markers for bovine postnatal body growth trait. Published transcriptomic data can be useful to the bovine 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.animal-genome.org/community/FAANG/].

III. Experimental Design, Materials, and Methods

Sample collection: Liver tissues samples were collected from 18 young bulls aged between 6 to 12 months in a panel of three selected cattle breeds: Polish HF, Polish Red, and Hereford. All experimental animals were reared at Institute of Genetics and Animal Breeding, Jastrzębiec, Po-

land in a closed herd, and providing uniform feeding and environmental conditions.

Experimental procedure: Total RNA from all liver 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: The FASTq raw data were obtained after removal of the adaptor sequences using cutadapt software (https://cutadapt.readthedocs.io/en/stable/) with minimum overlap length was set to 10 and error rate was set to 0.05, followed by trimming of the low quality bases at 3'– end using sequence quality control (FastQC).

Processed and filtered Omics data: The paired-end reads were mapped/aligned to the Bos taurus reference genome Ensembl75_UMD3-1.1 plus the Chromosome Y from Btau_4.6.1 assembly, by using BWA version 0.7.5-r404 [6] and the HT-Seq framework, version 0.5.3p9 (https://pypi.python.org/pypi/HTSeq/0.5.3p7), to count the mapped/aligned reads using the STAR BWA tools [7]. The SNPs processed and filtered data was obtained by using SAMtools mPileUp package to call SNPs and indels [8] and by using Microsoft Office Excel. The DEGs processed and filtered data was obtained by using DE-seq [9] and EdgeR [10] pipelines and 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., Błaszczyk P., Dziuba P., Czarnik U., Fraser L., Sobiech P., Pierzchała M., Feng Y., Kadarmideen H.N., Kumar D. Single nucleotide polymorphism discovery in bovine liver using RNA-seq technology. PLoS One. 2017;12:e0172687.

[2] Pareek C.S., Sachajko M., Jaskowski J.M., Herudzinska M., Skowronski M., Domagalski K., Szczepanek J., Czarnik U., Sobiech P., Wysocka D., Pierzchala M., Polawska E., Stepanow K., Ogłuszka M., Juszczuk-Kubiak E., Feng Y., Kumar D. Comparative Analysis of the Liver Transcriptome among Cattle Breeds Using RNA-seq. Vet Sci. 2019;6:36.

[3] Lisowski P., Kościuczuk E.M., Gościk J., Pierzchała M., Rowińska B., Zwierzchowski L. Hepatic transcriptome profiling identifies differences in expression of genes associated with changes in metabolism and postnatal growth between Hereford and Holstein-Friesian bulls. Anim Genet. 2014;45:288–92.

[4] Salleh M.S., Mazzoni G., Höglund J.K., Olijhoek D.W., Lund P., Løvendahl P., Kadarmideen H.N. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics. 2017;18:258.

[5] Salleh S.M., Mazzoni G., Løvendahl P., Kadarmideen H.N. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. BMC Bioinformatics. 2018;19:513.

[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.
[9] Anders S., Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.
[10] Robinson M.D., McCarthy D.J., Smyth G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–140.