Nov 19, 2019

Next Generation Sequencing and RNA-Seq Characterization of adipose tissue in the Nile crocodile

PLOS One

Odunayo Azeez[1,2]

[1]Azeez O. I., [Department of Veterinary Physiology and Biochemistry, University of Ibadan, Ibadan, Nigeria, [2]Department of Veterinary Anatomy and Physiology, Faculty of Veterinary Sciences, University of Pretoria, Onderstepoort 0110, Pretoria, South Africa.], Myburgh J. G., [Paraclinical Science Dept., Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, Pretoria, South Africa.], Bosman A. M., [Department of Veterinary Tropical Diseases, Faculty of Veterinary Sciences, University of Pretoria, Onderstepoort 0110, Pretoria, South Africa], Featherston J., [Agricultural Research Council, Onderstepoort Veterinary Institute Sibeko-Matjilla, K. Pb., Oosthuizen, M. Cb and J. P Chamunorwaa a Anatomy and Physiology Dept., Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, Pretoria, South Africa, email: odunayoazeez@gmail.com. b Veterinary Tropical Diseases Dept., Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, Pretoria, South Africa. c Paraclinical Science Dept., Faculty of Veterinary Science, University of Pretoria, Onderstepoort 0110, Pretoria, South Africa. d Dept. of Veterinary Physiology and Biochemistry, Faculty of Veterinary Medicine, University of Ibadan, Ibadan, Nigeria. e Biotechnology Platform, Agricultural Research Council, Onderstepoort 0110, Pretoria, South Africa.

1 Works for me dx.doi.org/10.17504/protocols.io.69thh6n

Odunayo I. Azeez
University of Ibadan, Ibadan, Nigeria.

ABSTRACT

Next Generation sequencing by RNA-Seq of the adipose tissue in health and in pansteatitis was carried out in the Nile crocodile, with the aim of determining the role of adipose tissue in pansteatitis - a generalized inflammation of the adipose tissue that have been incriminated in the crocodile die offs amon the Nile crocodiles along Olifant River, Kruger National Park, Mpumalanga province in South Africa.

Healthy adipose tissue consisting of fifty samples was collected from the subcutaneous, visceral, intermuscular adipose tissues and the abdominal fat body of ten 4 years old juvenile crocodiles from a local crocodile farm in Pretoria, South Africa. Ten pansteatitis samples were collected from visceral and intermuscular adipose tissues of five crocodiles that were dying of pansteatitis.

RNA sequencing (RNA-Seq) using Next Generation Sequencing (NGS) and *de novo* assembly of the adipose transcriptome, followed by differential gene expression analysis.

EXTERNAL LINK

https://doi.org/10.1371/journal.pone.0225073

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Azeez OI, Myburgh JG, Bosman A, Featherston J, Sibeko-Matjilla KP, Oosthuizen MC, Chamunorwa JP (2019) Next generation sequencing and RNA-seq characterization of adipose tissue in the Nile crocodile (Crocodylus niloticus) in South Africa: Possible mechanism(s) of pathogenesis and pathophysiology of pansteatitis. PLoS ONE 14(11): e0225073. doi: 10.1371/journal.pone.0225073

1 Total RNA Extraction for Next-Generation Sequencing Total RNA for NGS was extracted from normal adipose tissue, pansteatitis samples and the liver using TRIzol, a phenol and guanidium isothiocyanate based RNA extraction and purification protocol according to Rio et al. (2010).

2 Extraction involved homogenization of 100 mg of each sample in 900 µl of Trizol homogenizing buffer. The samples were homogenized on Precelleys 24® homogenizer, the homogenate was transferred into fresh 1.5 ml centrifuge tubes from which 200 µl were pipetted into another tube.

3 To the 200 µl of the homogenate was added 800 µl of Trizol reagent and mixed. Then, 200 µl of chloroform was added to the mixture and allowed to stand at room temperature for 10 mins for phase separation. This was then centrifuged at 12,000 xg for 10 min at 4 oC.

4 The clear supernatant was transferred into another tube and 200 µl of Isopropanol added and stored at -20 oC for 12 hours.

5 The mixtures were then centrifuged at 12,000 xg for 15 mins at 4 oC and the supernatant including isopropanol was removed.

6 Then, 500µl of 70% ethanol was added to wash the RNA before it was centrifuged again at 16,000 g for 5-10 mins after which ethanol was gently removed and white patched of RNA was seen on the side of the tube.

7 The tubes were left in the hood for 10 mins to allow ethanol to evaporate and RNA dissolved in 20 µl of RNAse free water.

8 The optical density of the total RNA was measured on Biotek Nanodrop Powerwave XS 2 Microplate Spectrophotometer to determine the concentration and quality. Only samples with total RNA concentration of 1000 ng/µl and above were used for further downstream analysis (O'Neil et al., 2013).

9 RNA Quality Assurance Before sample preparation, total RNA quality was assessed using Biotek Nanodrop Powerwave XS 2 Microplate Spectrophotometer at the wavelength of 260 and 280 nm, 260/280 ratios ranging from 1.065 to 1.661 at concentrations of 1000 ng/µl and above. Integrity and quality of selected samples were then evaluated on Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA), which measures only RNA concentration directly using target-specific fluorescent dye. Samples with RNA concentration greater than 1 µg/ µL without genomic DNA contaminants were considered adequate for the downstream analysis according to Pfaffl et al. (2008).

10 Library Preparation Total RNA from 11 samples consisting of the healthy liver (1), abdominal fat body (1), visceral (1), subcutaneous (1), intermuscular (1) adipose tissue as well as pansteatitic visceral (3) and intermuscular (3) adipose tissues were used for the library preparation. Complete transcriptome sequencing was carried out at the Biotechnology Platform, Agricultural Research Council, Onderstepoort, Pretoria. 11. Library preparation for Ribo-zero depletion of ribosomal RNA and cDNA synthesis was done using Illumina Truseq Stranded Total RNA with Ribo-Zero Gold, high thorough-put configuration kit. The protocol, briefly described, removes ribosomal RNA (rRNA) using biotinylated target specific oligos combined with Ribo-Zero beads.

11 Following purification, the RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first-strand cDNA synthesis using DNA polymerase I and RNAse H.

12 Subsequently, A nucleotide base added to the cDNA and adapter ligation, the enriched PCR products were then used to create the final cDNA library for the sequencing of the entire transcriptome, according to O'Neil et al. (2013).

13 Ribosomal RNA (rRNA) depletion is an extensive protocol using the Illumina rRNA removal mix gold. It started by measurement of 1 µl of the total RNA at 1 µg/µl diluted with nuclease free water up to 10 µl on a 0.3 ml, 96 well PCR plate.

14 Five µl of rRNA binding buffer was added followed by 5 µl of rRNA removal mix gold. This was then incubated in the thermocycler, set at 68 oC for 5 min after which it was removed and allowed to stand at room temperature for 1 min.

15 The mixture was then transferred into a fresh plate in which 35 µl of rRNA removal bead had been added and mixed thoroughly. After allowing the content to stand at room temperature for 1 min on the magnetic plate, the supernatant from each well was transferred into a fresh PCR plate and placed on the magnetic plate again.

16 The previous step was repeated until all beads were removed. This was followed by addition of 99 µl of RNAClean XP beads and incubated at room temperature for 15 min after which it was placed on the magnetic plate for 5 min to allow the beads to attach to the side of the well.

17 The supernatant was discarded and the bead in the RNA Clean Plate (RCP) plate washed with 200 µl of 70% ethanol, allowed to incubate for 1 min at room temperature and the supernatant removed each from well.

18 The RCP plate was then allowed to stand on the magnetic plate for 15 min to dry.

19 Then 11 µl of the elution buffer was added to each well on the plate and mixed thoroughly by pipetting. The plate was further incubated at room temperature for 2 mins before it was transferred again to the magnetic plate at room temperature for 5 mins.

20 From the RCP plate was transferred 8.5 µl of the supernatant into a new 96 well PCR plate tagged Depleted RNA Fragment Plate (DFP) after which 8.5 µl each of the elute, primer and fragment high mix buffers were added and mixed.

21 The new plate and the content were then incubated in the thermocycler at 94oC for 8 min and 4oC hold before it was centrifuged briefly.

22 The first-strand cDNA was then synthesized from the rRNA depleted total RNA using Actinomycin D (Act D) mix prepared by adding 50 µl of Superscript II to it at the ratio of 1 µl of superscript II to 9 µl of Act D.

23 To the DFP plate was then added 8 µl of the first strand mix and incubated using the following Illumina specified condition with the lid preheated to 100oC: 25oC for 10 mins, 42oC for 15 min, 70oC for 15 mins and hold at 4oC.

24 The second strand cDNA was synthesized immediately to remove and replace the RNA template with a strand with the dUTP in place of dTTP to form a double-stranded cDNA.

25 In the preparation of the second strand cDNA, 5 µl of resuspension buffer was added to each well of the DFP plate (that contains the first-strand cDNA and the RNA), followed by 20 µl of the second strand marking master mix, and mixed thoroughly. The mixture was incubated at 16oC for 1 hour.

26 The DFP was thereafter cleaned up using AMPure XP beads by adding 90 µl of AMPure XP beads to each well of the DFP plate containing 50 µl of the double-stranded cDNA and incubated for 15 min at room temperature. The beads were bound to the sides of the wells by placing the plate on a magnetic plate for 5 mins as described earlier and the supernatant removed.

27 This was followed by washing with 200 µl of 80% ethanol twice and dried on the magnetic plate.

28 Resuspension buffer (17.5 µl) was then added to the plate, sealed and shaken on a microplate shaker at 1800 rpm for 2 mins and placed on the magnetic stand for 5 min.

29 Finally, 15 µl of the ds cDNA (supernatant) was pipetted into another plate for onward addition of Adenylate 3' adapter, adapter ligation and sequencing.

30 Adenylate 3’ end was incorporated by adding 12.5 µl of A tailing mix (containing 1 µl of A tailing control and 99 µl of resuspension buffer), shaken at 1800 rpm for 2 mins, centrifuged at 280 xg for 1 min and then incubated at 37 oC for 30 mins.

31 The indexed adapters were then ligated by adding 2.5 µl of the Ligation mix to each of the wells, mixed well and incubate at 30 oC for 10 min before 5 µl of the stop ligation buffer was added.

32 The plate was cleaned up using AMPure XP beads as described previously and washed with 80% ethanol.

33 Resuspension buffer was used after the ethanol had dried and 20 µl of the supernatant containing the ligated ds cDNA transferred into a fresh PCR plate.

34 The ligated ds DNA fragments with the adapters were then enriched using PCR with preheated lid option at 100 oC, 98 oC for 30 secs followed by 15 cycles of 98 oC for 10 secs, 60 oC for 30 secs, 72 oC for 30 secs, 72 oC for 5 mins and held at 4 oC.

35 Transcriptome Sequencing Transcriptome sequencing was done at the Biotechnology Platform, Agricultural Research Council (ARC), Onderstepoort Veterinary Institute, Onderstepoort, Pretoria.

36 The transcriptome was sequenced on Illumina Hiseq 2500 using Illumina Hiseq SBS v4 master mix kit for dual indexed paired-end sequencing. After thawing, the incorporation Reagent Master mix, (IRM), Universal Scan mix (USM), Cleavage Reagent mix (CRM) and the high salt buffer, as well as incorporation and cleavage buffers, were loaded onto the recommended position on the loading rack and the instrument set at 250 cycle run.

37 Read quality filtering and de novo transcriptome assembly. Hiseq fasta and fastq files of sequence reads from samples representing 11 transcriptomes (consisting of 6 pansteatitis, 4 normal adipose tissue and 1 liver sample) were downloaded from the mainframe server unto UNIX shell command machine onto MacBook Pro (16G RAM and 512G hard disk space) for analysis on the UNIX Shell Terminal.

38 Each read was trimmed and quality filtered using Trimmomatic-0.35 according to Bolger et al. (2014). Trimmomatic-0.35 cuts the Illumina adapters and other Illumina related sequences and trim the reads to a specified length to maintain the quality. It also removes low-quality reads and drops reads that are below 36 base pairs (http://www.usadellab.org/cms/?page=trimmomatic).

39 Trimming and quality filter was done using the command: java -jar trimmomatic-0.35.jar PE -phred33 input_forward.fq.gz input_reverse.fq.gz output_forward_paired.fq.gz output_forward_unpaired.fq.gz output_reverse_paired.fq.gz output_reverse_unpaired.fq.gz ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36.

40 Several choice algorithms have been developed for de novo transcriptome assemblies without the use of reference genome, including, Trinity, Bridger, Trans-AByss, SOAPde novo-Trans, Velvet and Oases; each with its peculiar merits and demerits ranging from accuracy, run time and computer programming requirements or CPU usage. But it should be noted that most of them utilize similar algorithm by constructing various possible k-mers in the reads from where various possible contigs are constructed in the form of deBruign graphs to produce all possible transcripts after removal of sequencing errors (Chang et al., 2015).

41 For the purpose of our study after trial of some of these algorithms, Bridger was used for the de novo transcriptome as it has the lowest CPU requirement and lowest runtime. Our attempt to use Trinity failed to produce the required transcriptome assembly after several weeks of runs on a 500GB RAM UNIX Shell multicore server.

42 Bridger builds splicing graphs for all genes encoded in the genome based on the given RNA-Seq data using a rigorous mathematical model to search for the minimal set of paths/transcripts that are supported by the RNA-Seq reads (Chang et al., 2015).

43 The following command was used to assemble the transcriptome on Bridger at 25 bp k-mer length: nohup perl ./Bridger.pl—seqType fq—kmer_length 25 --left ../CrocTR1.fastq—right ../CrocTR2.fastq—SS_lib_type RF—CPU 10 --output Croc &. Where TR1 and TR2 are the trimmed reads for each sample.

44 Transcriptome annotation All identified transcripts were blasted against Crocodylus porosus transcriptome and protein sequences identified on Uniprot (http://www.uniprot.org/), NCBI Non-redundant (NR) database and EMBL-EBI (http://www.ebi.ac.uk/uniprot) databases for protein ID. Gene ontology classification of genes was carried out using PANTHER GO, version 10 (Mi et al., 2016).

45 Open reading frame prediction Open reading frames (ORFs) of the transcripts were predicted directly by Bridger and incorporated into the output summary. These were however evaluated on Transrate v1.0.3 (Smith-Unna et al., 2016).

46 Conserved domains were identified using similarity searches against salt water crocodile (Crocodylus niloticus) genome database (http://gigadb.org/dataset/100127) downloaded on 08/01/2016 (Green et al., 2014).

47 Differential gene expression The reads (FASTQ files) were aligned with Tophat, while cufflinks was used to convert the reads to transcripts (Trapnell et al., 2012). RNA-Seq by Expectation-Maximization (RSEM) was used to determine the differential expression (http://deweylab.github.io/RSEM/) according to Li and Dewey (2011) without the use of any reference genome on UNIX Shell terminal server with 500Gb RAM and a 64 bit computer with 16 Gb RAM on the same Unix shell operating system (Mac OS X). RSEM computes the maximum likelihood abundance estimates (ML) from aligned reads and posterior mean estimate for each gene and isoforms at a 95% credibility interval. It helped to capture and differentiate any uncertainty due to both random sampling effects and read mapping ambiguity (Li et al., 2010, Li and Dewey, 2011).

48 RSEM works by preparing a set of reference transcripts or accept previously assembled reference from de novo assembly or any other aligner using the simple script: 'rsem-prepare-reference' before calculation of the expression abundance by re-aligning the reads to the reference transcript sequence top estimate abundance and the credibility intervals (CI, set at 95%) using the script: 'rsem-calculate-expression'.

49 For better and more rigorous alignment of the paired-end reads, before feeding into RSEM for differential expression estimates, reads were aligned using Tophat 2.1.1 (https://ccb.jhu.edu/software/tophat/index.shtml), a fast splice junction aligner algorithm that utilizes Bowtie 2.0 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml) as the aligner engine.

50 Gene expressed were blasted (blastx) on SwissProt and TrEMBL databases of Uniprot (http://www.uniprot.org/) as well as Non-redundant (NR) database of the NCBI with an E-value-cut off of 1e-5. Genes were classified using gene ontology (GO) on Panther (http://pantherdb.org/) (Mi et al., 2016).

51 A differential expression volcano plot was constructed from expressed genes by plotting log10P values against log2Fold change using ggplot2 in R statistical software (http://ggplot2.org/). Finally, the differentially expressed genes were uploaded unto Reactome pathways database (http://www.reactome.org/), an integrated database for pathway enrichment analysis), for determination of the specific pathway participation and the effects of the upregulated genes (Croft et al., 2014, Fabregat et al., 2016).

52

53 References

54 Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA Using TRIzol (TRI Reagent). Cold Spring Harbor Protocols. 2010;2010(6):pdb.prot5439. doi: 10.1101/pdb.prot5439.

55 O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA Depletion for Efficient Use of RNA-Seq Capacity. Current protocols in molecular biology. 2013:4.19. 1-4.. 8.

56 Pfaffl M, Fleige S, Riedmaier I. Validation of lab-on-chip capillary electrophoresis systems for total RNA quality and quantity control. Biotechnology & Biotechnological Equipment. 2008;22(3):829-34.

57 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014:btu170.

58 Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, et al. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. Genome Biol. 2015;16(1):30.

59 Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Research. 2016;44(D1):D336-D42. doi: 10.1093/nar/gkv1194.

60 Smith-Unna R, Boursnell C, Patro R, Hibberd J, Kelly S. TransRate: reference free quality assessment of de novo transcriptome assemblies. Genome research. 2016:gr. 196469.115.

61 Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, et al. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science. 2014;346(6215). doi: 10.1126/science.1254449.

62 Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols. 2012;7(3):562-78.

63 Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):1-16. doi: 10.1186/1471-2105-12-323.

64 Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics. 2010;26. doi: 10.1093/bioinformatics/btp692.

65 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

66 Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Research. 2014;42(Database issue):D472-D7. doi: 10.1093/nar/gkt1102. PubMed PMID: PMC3965010.

67 Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. Nucleic Acids Research. 2016;44(Database issue):D481-D7. doi: 10.1093/nar/gkv1351. PubMed PMID: PMC4702931.