Variant discovery in the sheepmeat odour and flavour in javanese fat tailed sheep using RNA sequencing

Variant discovery in the sheepmeat odour and flavour in javanese fat tailed sheep using RNA sequencing

M A M Abuzahra[1], Jakaria[1], K Listyarini[1], A Furqon[1], C Sumantri[1], M J Uddin[2] and A Gunawan[1]

[1]Department of Animal Production and Technology, Faculty of Animal Science, Bogor Agricultural University, 16680 Bogor, Indonesia.
[2]School of Veterinary Science, The University of Queensland, Gatton Campus, Gatton Qld 4343, Australia.

E-mail: *aagun4780@gmail.com*

Abstract. High-throughput RNA sequencing (RNA-Seq) reveals new challenges for the detection of transcriptome variants (SNPs) in different tissues and species. The aims of this study was to characterize a SNP discovery analysis in the sheep meat odour and flavour transcriptome using RNA-Seq. Six liver samples from divergent sheep meat odour and flavour were analyzed using the Illumina Genome Hiseq 2500 Analyzer. The SNP detection analysis revealed 142 SNPs in sheep meat samples, and a large number of those corresponded to differences between high and low sheep meat odour and flavour ovis genome assembly OAR v4.0.Among them, about 90.4% of genes had multiple polymorphisms within 12 genes (JAML, ANGPTL8, LOC101103463, SEPW1, SCN5A, LOC101113036, DOCK6, GTSE1, KIF12, KCTD17, KANK2, CYP2A6). Several of the SNPs (JAML, CYP2A6, SEPW1, and KIF12) found in this study could be included as suitable markers in genotyping platforms to perform association analyses in commercial populations and apply genomic selection protocols in the sheep meat production.

1. Introduction

Next generation technologies of modern time for sequencing has been provided new opportunities for maximum output like genome annotation, functional genomic research dividing gene expression, detection and profiling of aberrant transcription and discovery of ncRNA [1]. It may be used in entire transcriptome to determine the changes in gene expression. RNA sequence also provides positive results to identify the splicing events, different transcripts of different family isoforms as well as polymorphism [2].

RNA-Seq has been widely used to detect single nucleotide polymorphisms (SNPs), alternative splicing (AS) events differentially expressed genes (DEGs) between two gene expression patterns and insertion or deletion (InDels). In addition, RNA-sequence helped to approach and describe previously inaccessible complexities in the transcriptome, such as isoforms, allele-specific expression and novel promoters [3]. Even in recent times the use of next-generation sequencing platforms to determine and to express the sheep transcriptomes is not too much developed and is very limited, and globally there is no authentic publication regarding analysis of sheep liver transcriptome by using RNA sequence. So to approach and for better understanding of fat deposition and metabolism in sheep, data of liver transcriptomes from Javanese fat tailed sheep was analyzed.

Among these different techniques, RNA sequencing (RNA-Seq) is a reliable and powerful new approach to quantify and for the mapping of transcriptomes used for the analysis of global gene expression in different tissues. In present, this approach efficiently works and also very cheap technique systematically used for the identification of SNPs in transcribed regions in different species [4]. RNA-Seq generates sequences on a very large scale at a fraction of the cost required for traditional Sanger sequencing, allowing the application of sequencing approaches to biological questions that were logically, logistically and economically difficult to practicalized [5].

The aim of this study was to characterize the SNPs discovery analysis in the sheep meat odour and flavour in Javanese fat tailed sheep using RNA-Seq. the transcriptome profile, polymorphisms and alternative splicing analysis using RNA deep sequencing could reveal potential candidate genes affecting odour and flavour of sheep meat in Javanese fat tailed sheep. It is exhibited that in future these polymorphisms could be used as markers for sheep meat related with odour and flavour trait.

2. MATERIAL AND METHODS

2.1 Animals and samples

A total of six liver samples from divergent of sheep meat flavour and odour in Javanese Fat Tailed Sheep were analyzed by using the Illumina Genome Hiseq 2500 Analyzer. All lambs were healthy and development was consistent with age. Management and feed rations were the same. Sheeps were slaughtered after 10 months following the "Guidelines on Ethical Treatment of Experimental Animals, which was formulated by the Ministry of agriculture Indonesia. Liver tissues were collected and frozen in liquid nitrogen within 10 min of slaughter. Subsequently, liver tissues were stored at -80°C until RNA isolation.

2.2 Library construction and sequencing

The ideal method for transcriptomics should be able to directly identify and quantify all small or large RNAs. RNA-Seq provides enough explanation and description of identification of splicing events and different family isoforms of transcripts, transcriptome profiling and polymorphisms [6]. Deep RNA sequencing approach was used to find polymorphism, differential expression and alternative splicing detection. To obtain this purpose, full-length cDNA was obtained from 1 μg of RNA, with the SMART cDNA Library Construction Kit (Clontech, USA), according to the manufacturer's instructions. Libraries of amplified RNA for each sample were prepared following the Illumina mRNA-Seq protocol. The library preparations were sequenced on an Illumina HiSeq 2500 as single- reads to 100 bp using 1 lane per sample on the same flow-cell (first sequencing run) at Macrogen, South Korea. All sequences were analysed using the CASAVA v1.7 (Illumina, USA).

2.3 RNA-Seq analysis and SNP detection

Short sequence reads (36-40 bp) were assembled and mapped to the annotated bovine reference genome B tau4.0 (http://www.ncbi.nlm.nih.gov/genome/guide/cow/index.html) using CLC Genomics Workbench software (CLC Bio, Aarhus, Denmark). Sequencing reads for each of the six samples were pooled to perform the RNA-Seq and SNP discovery analyses. Stringent criteria was applied in order to reduce the rate of detection of false-positive SNPs. For the assembly procedure, the sequences were mapped to the consensus genome accounting for a maximum of two gaps or mismatches in each sequence. SNP detection was performed using the following quality and significance filters: (1) the minimum average quality of surrounding bases and minimum quality of the central base were set as 15 and 20 quality score units, respectively; (2) minimum coverage was set at ten reads; (3) minimum variant frequency or count was set at 20% or two read counts per SNP; and SNPs located in read ends (last three bases) were not considered in the analysis due to possible sequencing errors.

2.4 Gene Variation Analysis

In this analysis, SNP calling was performed on the mapping files generated by TopHat algorithm using samtools mpileup command and associated algorithms. From the variants so generated, only those

variants with a minimum Root Mean Square (RMS) mapping quality of 20 and a minimum read depth of 100 were selected for further analysis. In the final step, the selected variants were cross- checked against dbSNP database to identify mutations that are already studied. In the next step, to understand whether these identified polymorphisms segregate either in only one sample group (high odour and flavour or low odour and flavour group) or in both groups (high and low odour and flavour group), we calculated the read/coverage depth of these polymorphisms in all the samples . The identified SNPs were furthermore classified as synonymous or non-synonymous using the GeneWise software (http:// www.ebi.ac.uk/Tools/psa/genewise/last accessed 21.03.2013) by comparing between protein sequence and nucleotides incorporated SNP position.

3. RESULTS AND DISCUSSION

3.1 Variant Discovery in Javanese fat tailed Sheep

In addition to the quantification process of transcriptome, RNA-Seq approach provides valuable information regarding gene polymorphisms which could be directly correlated with the relevant phenotype [2] .The method of variant detection applied to sequence javanese fat tailed sheep was the pooled samples analyses, and it revealed 142 SNPs in 37 genes.

In the results observation that most of the SNPs were in intergenic region with the maximum number of 142 SNPs (figure 1), among them, about 90.4% of genes had multiple polymorphisms (figure 2 A). It should also be noted that the *O. aries* reference genome was used in the present study; therefore, development of a *O. aries* reference genome could greatly enhance results from SNP discovery studies.

Localization and predicted effect of polymorphisms discover in Javanese fat tailed

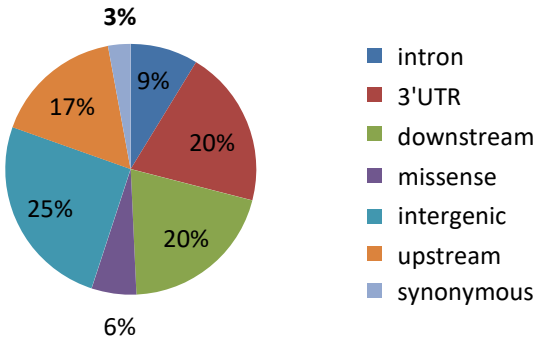

Figure 1. Descriptive statistics results from the Variant Effect Predict (VEP) tool f Ensembl using the 142 SNPs identified in Javanese Fat Tailed Sheep.

Because SNPs were detected in the RNA sequence that the current annotations described as being in introns, a transcript discovery analysis was conducted. Specifically, in Illumina Genome Hiseq 2500 assigned 37 SNPs to exons. Only 8 of these SNPs in the exons were non-synonymous (figure 1). Illumina Genome Hiseq 2500 classified some SNPs as splice variants. This means that polymorphisms occurred within the region of the splice site and potentially influenced the length of the RNA transcript [7]. SNPs located within the non-coding region (3'UTR) of the candidate gene may be located in a binding site for mRNA and/or the transcriptional machinery [8]. Moreover, SNPs in 3'UTR region can also influence mRNA stability [9]. RNA transcripts often contain a poly-A tail, which could have a role in gene expression as well as post-transcriptional stability [10]. Also have to take care that it should that SNPs located within an intron may also affect transcription factor binding sites or mRNA, transcription of other genes would be effected [8]. SNPs located within the upstream intergenic area of a gene could be within or can influence binding site for transcription factors [11].

Analysis of SNPs discovered in Javanese Fat Tailed Sheep samples revealed that 142 SNPs associated with sheepmeat odour and flavour within 37 genes (figure 1). The present study candidate 12 genes related with sheepmeat odour and flavour (JAML, LOC101119620, KIF12, KCTD17, ANGPTL8, LOC101103463, SEPW1, SCN5A, LOC101113036, DOCK6, GTSE1, KANK2, CYP2A6) (figure 2 B). SCN5A gene was the fewest numbers of SNPs with one polymorphism detection. The GTSE1 gene contained the largest number of mutations with 17 SNPs. The average number of SNPs per gene was 3 SNPs.

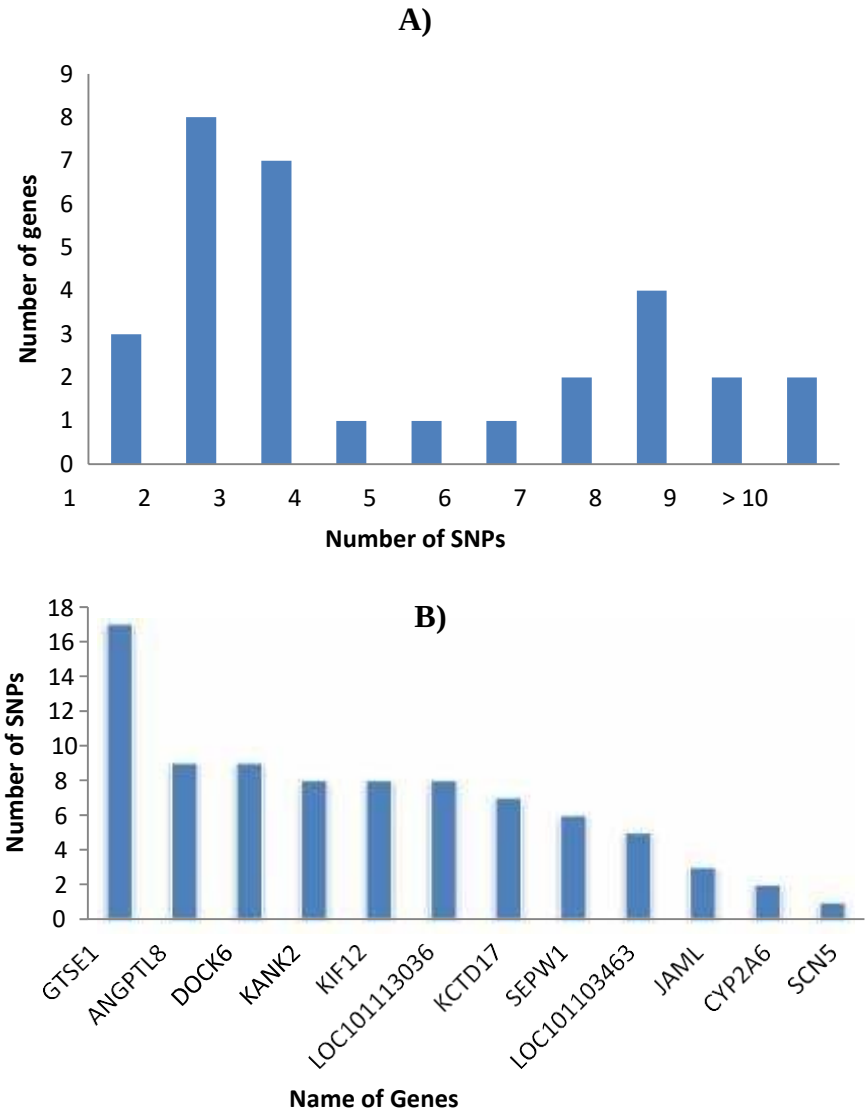

Figure 2. Distribution of the number of SNPs detected in the DEGs. The distribution of the number of SNPs occurred in each gene (A); numbers of SNPs in the genes selected for the association validation (B).

After SNP discovery executed in javanese fat tailed sheep, a comparative analysis was carried out Furthermore, the number of mutations could depend on the gene length, and the GTSE1 gene was the longest gene among this studies. Gene KIF12 is a gene encoding a novel kinesin, mapped beneath the 34 cM QTL peak and has expression level variants and strain-specific sequences that were associated with renal disease severity in affected mice [12]. ANGPTL8 (also called betatrophin) is a major and an important influencing factor in triglyceride levels and adipose tissue mass as well as in replenishing the

adipose tissue triglyceride store [13]. Dock6 gene, Expression and the cellular phenotype of fibroblasts of individual further confirms the role of DOCK6 in Adams-Oliver syndrome (AOS) pathogenesis, which appears to converge with that reported for ARHGAP31 in perturbation of the actin cytoskeleton through inactivation of Cdc42 and Rac1 [14]. Gadea and Blangy [15] had identified Kank2 proteins as talin activators that decrease the grip between the integrin–talin complex and actomyosin to regulate cell migration velocity.

Among the 12 genes that contained SNPs in Javanese Fat Tailed Sheep in the present study, CYP2A6 was associated with boar trait in a gene expression study by [16]. Nonetheless, the results of our study may be useful for the design of a focused SNP panel for use in genotype-to-reproductive phenotype association analyses in a multi-breed setting. The focused (i.e. functional) SNP panel is an approach that can reduce costs associated with genotyping [17]. Biologically relevant polymorphisms are used to design a focused or low-density panel of selected SNPs [17]. This is the first study used candidate genes to be determined as associated with sheepmeat odour and flavour in javanese fat tailed sheep, the SNPs discovered should be useful to reduce sheepmeat odour and flavour.

4. Conclusions

In this study for the very first time NGS technology has been used to analyze the expression profiles of sheepmeat odour and flavour in javanese fat tailed sheep by using RNA Sequencing. Several of the SNPs (JAML, CYP2A6, SEPW1, and KIF12) were founded in this study could be included as suitable markers in genotyping platforms to perform association analyses in commercial populations and apply genomic selection protocols in the sheepmeat production. This transcriptome, polymorphisms and alternative splicing analysis using RNA deep sequencing revealed potential candidate genes affecting odour and flavour of sheepmeat in Javanese fat tailed sheep. It is exhibited that in future these polymorphisms could be used as markers for sheepmeat related with odour and flavour traits. However, further research and evidences are required to confirm the effect of these genetic markers in other sheepmeat populations.

Acknowledgment

I would like to pay my special thanks to my supervisor honorable Dr Asep Gunawan for his guidance, encouragement, availability and support to perform this research and achieve main objectives. I'm also thankful to my lab fellows for their support and help during research, Authors greatly thankful to development ministry of Research, Technology, and higher Education year 2016 number: 079/SP2H/LT/DPRM/II/2016.

References

[1] Cánovas A, Rincon G, Islas-Trejo A, Wickramasinghe S and Medrano J F 2010 SNP discovery in the bovine milk transcriptome using RNA-Seq technology *Mamm. Genome*. **21** 592

[2] Gunawan A, Sahadevan S, Cinar M U, Neuhoff C, Große-Brinkhaus C, Frieden L, Tesfaye D, Tholen E, Looft C, Wondim D S and Hölker M 2013 Identification of the novel candidate genes and variants in boar liver tissues with divergent skatole levels using RNA deep sequencing *PloS One*. **8** e72298.

[3] Li C, Cai W, Zhou C, Yin H, Zhang Z, Loor J J, Sun D, Zhang Q, Liu J and Zhang S 2016 RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population *Sci. Rep.* **6** 26813

[4] Cirulli E T, Singh A, Shianna K V, Ge D, Smith J P, Maia J M, Heinzen E L, Goedert J J and Goldstein D B 2010 Screening the human exome: a comparison of whole genome and whole transcriptome sequencing *Genome Biol.* **11** R57

[5] Marguerat S, Wilhelm B T and Bähler J 2008 Next-generation sequencing: applications beyond genomes *Biochem. Soc. Trans.* **36** 1091-6

[6] Marguerat S and Bähler J 2010. RNA-seq: from technology to biology *Cell Mol Life Sci.* **67** 569-79

[7] Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S and Gil L 2014 Ensembl 2015 *Nucleic Acids Res.* **43** 662-9
[8] Jin Y and Lee C G 2013. Single nucleotide polymorphisms associated with microRNA regulation *Biomolecules* **3** 287-302
[9] Matoulkova E, Michalova E, Vojtesek B and Hrstka R 2012 The role of the 3'untranslated region in post-transcriptional regulation of protein expression in mammalian cells *RNA Biol*. **9** 563-76
[10] Kuersten S and Goodwin E B 2003 The power of the 3 [variant prime] UTR: translational control and development Nature reviews *Genetics* **1** 626
[11] Vinsky M, Islam K, Chen L and Li C 2013 Association analyses of a single nucleotide polymorphism in the promoter of OLR1 with growth, feed efficiency, fat deposition, and carcass merit traits in hybrid, Angus and Charolais beef cattle *Can. J. Anim. Sci.* **93** 193-7
[12] Mrug M, Li R, Cui X, Schoeb T R, Churchill G A and Guay-Woodford LM 2005 Kinesin family member 12 is a candidate polycystic kidney disease modifier in the cpk mouse *J. Am. Soc. Nephrol.* **16** 905-16
[13] Li Y and Teng C 2014 Angiopoietin-like proteins 3, 4 and 8: regulating lipid metabolism and providing new hope for metabolic syndrome *J. Drug Target.* **22** 679-87
[14] Shaheen R, Faqeih E, Sunker A, Morsy H, Al-Sheddi T, Shamseldin H E, Adly N, Hashem M and Alkuraya F S 2011 Recessive mutations in DOCK6, encoding the guanidine nucleotide exchange factor DOCK6, lead to abnormal actin cytoskeleton organization and Adams-Oliver syndrome *Am J Hum Genet*. **89** 328-33
[15] Gadea G and Blangy A 2014 Dock-family exchange factors in cell migration and disease *Eur. J. Cell Biol*. **93** 466-77
[16] Habier D, Fernando RL, Dekkers JC 2009 Genomic selection using low-density marker panels *Genetics* **182** 343-53
[17] Fortes M R, Porto-Neto L R, DeAtley K L, Reverter A, Thomas M G, Moore S S, Lehnert S A and Snelling W M 2014 *Genetic Markers in Transcription Factors of Differentially Expressed Genes Associated With Post-Partum Anoestrus Predict Pregnancy Outcome in An Independent Population Of Beef Cattle* 10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada