International Journal of Animal and Veterinary Advances 4(4): 292-302, 2012
ISSN: 2041-2894

Submitted: June 19, 2012 Accepted: July 09, 2012 Published: August 20, 2012

Ubiquitous Expression of Genes in tissues of Goat (*Capra hircus*) Using RNA-seq

[1]Upasna Sharma, [2]Priyanka Banerjee, [2]Jyoti Joshi and [2]Ramesh Kumar vijh
[1]Singhania University, Pacheri Bari, Jhunjhunu-313515, Rajasthan, India
[2]National Bureau of Animal Genetic Resources (ICAR), Karnal-132001 India

Abstract: Since very little information is available on goat transcriptome (only few sequenced genes and ESTs are available in the database), our main aim was to analyze the goat transcriptome for identification of ubiquitous genes through RNA-Seq approach, expression of genes across tissues and analyse the functional pathways which these highly expressed genes follow. RNA-Seq analysis was carried out on 10 tissues of Osmanabadi goats. The data was 2x76 base pair paired end reads generated using Illumina Genome Analyser GAIIx. More than 40 million reads per tissue were generated. The data was mapped using CLC Genomics Workbench on Ensembl cattle (Btau_4.0) genes version 62 downloaded from BioMart. The mapping statistics was discerned for each of the mapped gene for all the 10 tissues. There was a great deal of homology between the genes of cattle and goats and this was expected as both are ruminants and phylogenetically close to one another. The expression profile for genes of each of the 10 tissues was calculated in terms of RPKM values. The differential expression of genes among the different tissues was carried out using 3 algorithms. The genes common among the 3 algorithms were associated with GO IDs and the functional annotation was carried out by estimating the GO term occurrences using CateGOrizer web server. The study revealed that more than 75% genes are ubiquitously expressed (expressed in all tissues) with very limited number of tissue specific genes based on expression values (expressed in only one tissue). These genes are mostly related with the specific functions performed by various tissues.

Keywords: *Capra hircus*, differential gene expression, gene ontology, RNA-Seq, transcriptome

INTRODUCTION

The domestic goat *Capra hircus* is an important livestock species in India and other developing countries. Since it provides a good source of meat, milk, fibre and skin, it is popularly known as the "poor man's cow" (MacHugh and Bradley, 2001). Goats have fulfilled agricultural, economic, cultural and even religious roles from very early times in human civilization. These are the most adaptable and geographically widespread livestock species ranging from the high altitude of the Himalayas to the deserts of Rajasthan and hot humid coastal areas of India. The goats being mammals maintain their core body temperature utilizing various physiological mechanisms. As goats have preponderance in the developing world, very little information is available on its molecular architecture and genetic basis of adaptation of the species to extreme climatic conditions. It is imperative to understand the molecular mechanism of adaptation using high throughput sequencing technologies and intensive computational methodologies to analyze the enormous data.

Regulation of gene expression is fundamental to link genotypes with phenotypes. The synthesis and maturation of RNAs are tightly controlled and they shape complex gene expression networks that ultimately drive biological processes. The ability to analyze entire gene expression programs has opened new horizons for our understanding of global processes regulating gene expression. Similarly, with the increased realisation that RNAs transcribed from non-coding portions of genomes are playing fundamental roles, genome-wide approaches have provided valuable insights into various aspects of transcriptome. Understanding the transcriptome is essential for interpreting the functional elements of the genome and revealing the molecular constituents of cells and tissues and also for understanding various development and disease aspects. Recently, the development of novel high-throughput DNA sequencing technology has provided a new method for both mapping and quantifying genes. This method, termed RNA-Seq (RNA sequencing), has clear advantages over existing approaches and is expected to revolutionize the manner in which eukaryotic transcriptomes are analysed. In RNA-Seq, all RNAs of a sample (or, more often, $polyA^+$ RNAs) are randomly fragmented, reverse transcribed, ligated to adapters and then these fragments are sequenced. Gene expression levels can then be estimated from the number of sequence reads deriving

Corresponding Author: Ramesh Kumar Vijh, National Bureau of Animal Genetic Resources (ICAR), Karnal-132001 India

from each gene (Wang *et al*., 2009). Compared to microarrays, RNA-Seq is more sensitive, both in terms of detection of lowly expressed and differentially expressed genes (Wang *et al*., 2009; Marioni *et al*., 2008). The greater accuracy and coverage of the expressed transcriptome makes this method suitable for addressing global features of transcriptome. Very little information is available on goat transcriptome (only few sequenced genes and ESTs are available in the database), we applied the RNA-Seq approach to analyze the goat transcriptome for identification of ubiquitous genes, expression of genes across tissues and analyse the functional pathways which these highly expressed genes follow. We analysed 10 tissues from Osmanabadi goat using Illumina GAIIx next generation sequencer.

MATERIALS AND METHODS

The tissues samples of Osmanabadi goat were selected for the study. Osmanabadi breed is spread over the greater part of the central peninsular region, comprising the semi-arid areas of Maharashtra, Andhra Pradesh and Karnataka. Osmanabadi breed shows a very efficient reproductive performance and resistance against diseases not only in well managed semi stall feeding system but also in severe drought conditions. 10 tissue samples-liver, muscle, skin, heart, hypothalamus, kidney, blood, spleen, mammary gland and lungs were used for transcriptome analysis using Next Generation Sequencing. A fraction of the tissue samples was used for RNA isolation. Total RNA was isolated following the standard protocols of RNeasy kit (Qiagen). We quantified the RNA with the Agilent 2100 Bioanalyzer (Agilent, Foster city, USA) which provided ng RNA/µL values, an accurate molecular weight and the calculated molarity for each peak. The integrity of the RNA was checked and the samples with RIN more than 8.5 were only utilised for cluster and data generation. Standard Illumina kit was utilised for DNA library preparation which facilitated reading both the forward and reverse template strands of each cluster during one paired-end read. In addition to sequence information, both reads contain long range positional information, allowing for highly precise alignment of reads. The unique paired-end sequencing protocol allowed us the length of the insert (200-300 bp) generating high quality, alignable sequence data. A typical paired-end run could achieve 2×76 bp reads and up to 40-60 million reads of the transcriptome data. The image analysis, base calling and quality score calibration were processed using the Illumina Pipeline Software v 1.4.1 according to the manufacturer's instructions. Reads were exported in the FASTQ format and used for further analysis. The alignment of RNA-Seq reads was carried out using cattle Btau_4.0 Ensembl genes downloaded from BioMart. The RNA-Seq reads were mapped using the functions of CLC Genomics Workbench version 4.0.2. The number of reads mapped only on exons, introns, exon-intron junction etc was recorded. The results were classified on the feature (Ensembl ID) and expression of each gene in terms of RPKM values (Reads per kilobase of exon model per million mapped reads) was recorded. The gene expression values in terms of RPKM obtained from CLC Genomics Workbench were utilised to find out differential gene expression analysis. For each set of analysis each of the tissue was taken as treatment (one at a time) and compared to rest of the 9 tissues of Osmanabadi goats. For differential expression profiling we utilised DEB Webserver -(http://www.ijcb.org/ DEB/php/onlinetool.php): (Qiang and Fahong, 2011) and the analysis was carried out using 3 algorithms of edgeR, baySeq and DESeq keeping the False Discovery Rate of 10% (default value). The genes showing differential expression for each tissue (versus the rest of 9 tissues) were downloaded from the server as result file. There were large differences in the number of genes showing differential expression using the 3 algorithms. The number of genes differentially expressed in the 3 algorithms however contained few common genes and to visualise this, the Venn diagrams were produced for graphical representation of the results using the Google chart wizard available at (https://developers.google.com/chart/image/docs/chart wizard): The genes differentially expressed in each of the tissue were mapped to the GO terms. The GO terms were categorized into distinct classes using the web server (Categoriser - http://www.animalgenome.org/ bioinfo/ tools/catego/): (Zhi-Liang *et al*., 2008). The RNA-Seq analysis was done in National Bureau of Animal Genetic Resources, Karnal and Haryana, India.

RESULTS

In present study, next generation sequence data on 10 tissues of Osmanabadi goat were utilised for transcriptome analysis. We generated more than 40 million paired end reads of 2x76 bp per tissue. The number of reads generated using Illumina next generation sequencer, number of reads that mapped onto the annotated genes (Fig. 1), the number of reads mapped on Exon-Exon junctions, Exon-Intron junctions, on exons, only on introns were obtained as mapping statistics and shown in Table 1. The mapping was also categorised as mapped reads (uniquely and

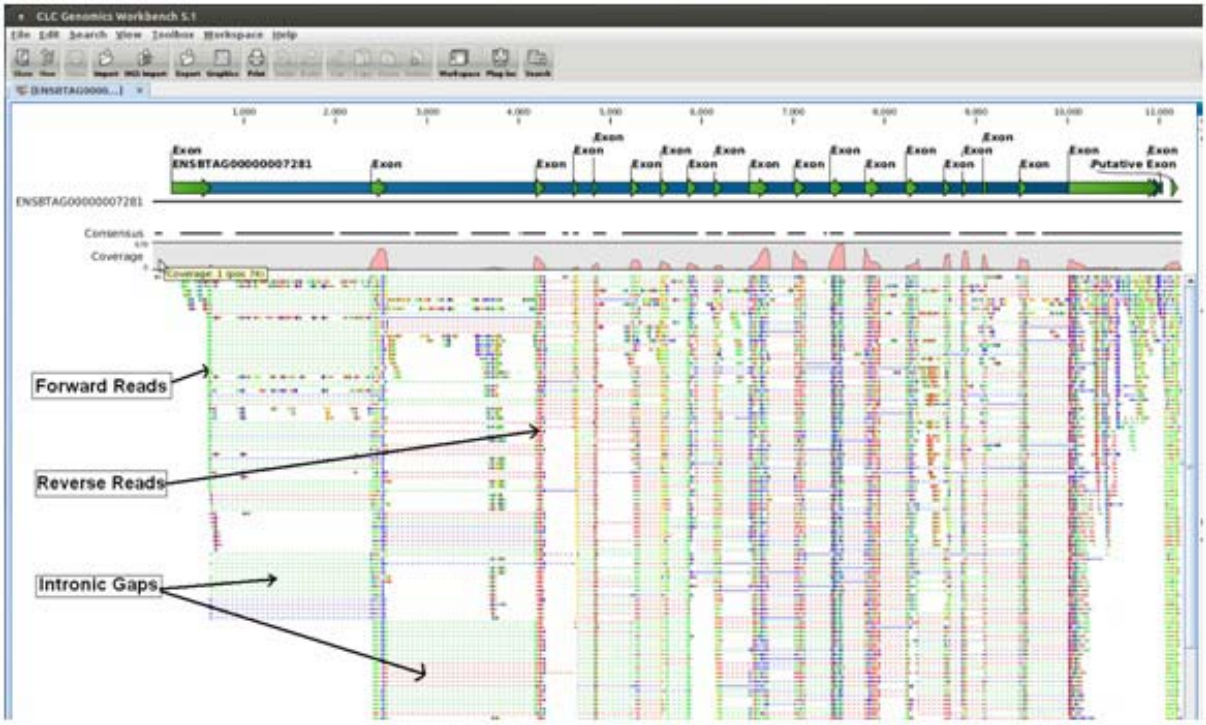


Fig. 1: A representative screenshot of mapping of goat paired end reads with a gene from *Bos taurus* Ensembl genes release 62 data

non-specifically) and unmapped reads. Most of the uniquely mapped reads correspond to total exon reads whereas a small fraction corresponded to total intron reads (Table 1). Intron reads are expressed regions that are not annotated as exons in Btau_4.0. A representative map of the reads on a gene with 21 exons and one putative exon has been shown in Fig. 1. The mapping has been carried out at the exons represented as green arrows at the reference gene (top blue bar). The pink peaks exhibit the coverage at the particular position where the reads have been mapped. The green lines below the mapped areas show forward strand reads and the red ones are reverse strand reads, the blue ones are the paired end reads. The putative exon are identified and shown on the gene reference. Exon-exon reads are mapped across 2 exons bypassing the intron. The intronic gaps are shown as faded green for forward and faded red lines for reverse reads.

Different levels of RPKM values (reads per kilobase per million mapped reads) (Mortazavi *et al*., 2008) were utilised to find out the number of genes expressed in each of the tissue under study. Various RPKM threshold values revealed the number of genes being expressed in each tissue on the basis of their expression. The Table 2 provides the number of expressed genes in a tissue at different levels of expression in 10 class intervals. The genes that had expression values of 1000 or more were categorised as first class interval and those between 500 to 999.9 formed second class intervals and so on. The expression profile for all the tissues of Osmanabadi goat expressed in RPKM classes have been tabulated as Table 2. The number of genes expressed in a given tissues ranged from 6700 to more than 16000 genes. The maximum number of highly expressed genes was observed in fibroblast (16744) while minimum genes were expressed in hypothalamus tissue (6787). Similar results of 11000 to 13000 genes which constitute roughly to 60-70% of the Ensembl protein coding genes in mouse and rat have been reported using RNA-Seq analysis (Ramskold *et al*., 2009). We mapped the top 10 highly expressed genes to broad Gene Ontology terms (GO). The analysis provided a perspective on the functional priorities of cells in each tissue, since allocation of a large fraction of poly A^+ RNA content in a cell (and likely of translational capacity) to one functional category represents a major investment of cellular resources. The 10 most expressed genes based on RPKM values were selected and mapped on GO IDs using the web server g: Profiler (http://biit.cs.ut.ee/gprofiler/) and investigated for the processes they were involved in. The results revealed that for blood the first GO term signified '*De Novo*' protein folding, '*De novo*' post translational protein folding, cellular protein metabolic process among biological processes, multi-vesicular body, non membrane

Table 1: Summary statistics of RNA-Seq reads mapping for 10 tissues of Osmanabadi goat to the Ensembl genes using RNA-Seq of CLC genomics workbench version 4.0.2

			Uniquely mapped reads		Non specifically mapped reads		Mapped reads	
SNo.	Tissue		Number of reads	Percentage	Number of reads	Percentage	Number of reads	Percentage
1	Blood	Total exon reads	7,94,710	0.94	46,798	0.06	8,41,508	78.41
		Exon-exon reads[a]	3,96,064	0.96	18,342	0.04	4,14,406	38.61
		Total intron reads	1,89,497	0.82	42,225	0.18	2,31,722	21.59
		Exon – intron reads[b]	45,797	0.93	3,637	0.07	49,434	4.61
		Total gene reads	9,84,207	0.92	89,023	0.08	10,73,230	100
2	Fibroblast	Total exon reads	96,89,681	0.92	8,33,928	0.08	1,05,23,609	88.58
		Exon-exon reads[a]	34,37,120	0.95	1,98,812	0.05	36,35,932	30.6
		Total intron reads	9,39,059	0.69	4,17,742	0.31	13,56,801	11.42
		Exon – intron reads[b]	1,67,142	0.82	36,521	0.18	2,03,663	1.71
		Total gene reads	1,06,28,740	0.89	12,51,670	0.11	1,18,80,410	100
3	Heart	Total exon reads	90,49,581	0.94	5,98,624	0.06	96,48,205	82.25
		Exon-exon reads[a]	30,59,145	0.95	1,56,926	0.05	32,16,071	27.42
		Total intron reads	16,23,651	0.78	4,58,176	0.22	20,81,827	17.75
		Exon – intron reads[b]	1,08,158	0.93	8,064	0.07	1,16,222	0.99
		Total gene reads	1,06,73,232	0.91	10,56,800	0.09	1,17,30,032	100
4	Hypoth-alamus	Total exon reads	2,42,584	0.93	17,094	0.07	2,59,678	78.31
		Exon-exon reads[a]	50,357	0.93	3,659	0.07	54,016	16.29
		Total intron reads	58,319	0.81	13,609	0.19	71,928	21.69
		Exon – intron reads[b]	5,900	0.94	363	0.06	6,263	1.89
		Total gene reads	3,00,903	0.91	30,703	0.09	3,31,606	100
5	Kidney	Total exon reads	3,05,441	0.92	27,881	0.08	3,33,322	86.79
		Exon-exon reads[a]	78,072	0.94	4,695	0.06	82,767	21.55
		Total intron reads	32,245	0.64	18,492	0.36	50,737	13.21
		Exon – intron reads[b]	11,513	0.87	1,682	0.13	13,195	3.44
		Total gene reads	3,37,686	0.88	46,373	0.12	3,84,059	100
6	Liver	Total exon reads	5,77,487	0.91	54,135	0.09	6,31,622	90.79
		Exon-exon reads[a]	1,36,943	0.94	8,570	0.06	1,45,513	20.92
		Total intron reads	40,609	0.63	23,491	0.37	64,100	9.21
		Exon – intron reads[b]	12,041	0.9	1,327	0.1	13,368	1.92
		Total gene reads	6,18,096	0.89	77,626	0.11	6,95,722	100
7	Lungs	Total exon reads	3,31,502	0.9	35,265	0.1	3,66,767	85.57
		Exon-exon reads[a]	59,345	0.91	5,582	0.09	64,927	15.15
		Total intron reads	39,965	0.65	21,893	0.35	61,858	14.43
		Exon – intron reads[b]	12,944	0.91	1,208	0.09	14,152	3.3
		Total gene reads	3,71,467	0.87	57,158	0.13	4,28,625	100
8	Mammary glands	Total exon reads	89,28,603	0.92	8,14,750	0.08	97,43,353	79.79
		Exon-exon reads[a]	29,30,651	0.94	1,92,177	0.06	31,22,828	25.57
		Total intron reads	18,73,504	0.76	5,94,932	0.24	24,68,436	20.21
		Exon – intron reads[b]	2,94,829	0.89	36,774	0.11	3,31,603	2.72
		Total gene reads	1,08,02,107	0.88	14,09,682	0.12	1,22,11,789	100
9	Muscle	Total exon reads	1,22,42,600	0.93	9,77,948	0.07	1,32,20,548	87.58
		Exon-exon reads[a]	42,82,779	0.95	2,42,092	0.05	45,24,871	29.97
		Total intron reads	13,54,528	0.72	5,20,789	0.28	18,75,317	12.42
		Exon – intron reads[b]	2,25,641	0.85	40,578	0.15	2,66,219	1.76
		Total gene reads	1,35,97,128	0.9	14,98,737	0.1	1,50,95,865	100
10	Spleen	Total exon reads	14,62,068	0.92	1,19,490	0.08	15,81,558	90.69
		Exon-exon reads[a]	6,88,497	0.95	37,895	0.05	7,26,392	41.65
		Total intron reads	1,20,084	0.74	42,230	0.26	1,62,314	9.31
		Exon – intron reads[b]	49,316	0.86	7,732	0.14	57,048	3.27
		Total gene reads	15,82,152	0.91	1,61,720	0.09	17,43,872	100

a Exon-exon reads reads mapping to two contiguous exons. Number is included in total exon reads
b Exon-intron reads reads mapping an exon and a contiguous intron. Number is included in total intron reads

Table 2: RPKM details of Osmanabadi breed of goat

Tissues of	No. of Expressed	RPKM Value									
Osmanabadi	genes	>1000	500-999.9	150-499.9	100-149.9	50-99.9	20-49.9	10-19.9	5-9.99	1.0-4.9	0.5-0.99
Blood	12569	47	76	335	294	867	2060	2338	2043	3102	824
Fibroblast	16744	53	84	282	247	750	2190	2401	2654	4119	3084
Heart	16712	51	72	284	229	632	1810	2147	2668	4162	1157
Hypothalamus	6787	88	48	360	230	731	1650	1683	1488	507	2
Kidney	12760	41	88	466	310	970	2324	2497	2347	3363	318
Liver	12571	67	91	372	248	637	1554	1809	2189	4319	927
Lungs	13457	53	80	357	293	961	2479	2640	2592	3584	352
Mammary gland	16384	56	57	281	273	905	2939	2914	2581	3365	1000
Muscle	15442	60	66	254	203	574	1832	2296	2651	4248	1091
Spleen	14491	93	52	369	315	994	2508	2434	2254	3477	990

bounded organelle and ribo-nucleoprotein complex in cellular component structural molecule activity and structural constituent of ribosome among molecular functions. For fibroblast, the highly expressed genes mapped on GOs associated with cellular components responsible for contractile fibre, myofibril and cytoskeleton. For heart, 10 most expressed genes mapped to 59 GO IDs of which 3 represented molecular

Table 3: List of 10 most expressed genes based on RPKM values for all the tissues of Osmanabadi goat with their Ensembl IDs

S.No	Ensembl IDs of Blood	Associated Gene Name	Ensembl IDs of Fibroblast	Associated Gene Name
1	ENSBTAG00000005654	TYB10_BOVIN	ENSBTAG00000013921	KCRM_BOVIN
2	ENSBTAG00000010156	TCTP_BOVIN	ENSBTAG00000010880	A5PJM2_BOVIN
3	ENSBTAG00000031814	SDS	ENSBTAG00000010156	TCTP_BOVIN
4	ENSBTAG00000015228	Q29630_BOVIN	ENSBTAG00000018369	MYL2
5	ENSBTAG00000001360	RS12_BOVIN	ENSBTAG00000011424	TPM2_BOVIN
6	ENSBTAG00000015285	RS8_BOVIN	ENSBTAG00000043561	Q6QTG9_BOVIN
7	ENSBTAG00000005142	RL37_BOVIN	ENSBTAG00000021218	MLRS_BOVIN
8	ENSBTAG00000013264	RS24_BOVIN	ENSBTAG00000011969	E1BEL7_BOVIN
9	ENSBTAG00000026199	Q9TTW4_BOVIN	ENSBTAG00000004965	NUP133
10	ENSBTAG00000025666	RPS29	ENSBTAG00000013103	CO1A1_BOVIN
	Ensembl IDs of Heart		Ensembl IDs of Hypo	
1	ENSBTAG00000018369	MYL2	ENSBTAG00000014534	EF1A1_BOVIN
2	ENSBTAG00000043561	Q6QTG9_BOVIN	ENSBTAG00000002648	RS18_BOVIN
3	ENSBTAG00000043550	CYB_BOVIN	ENSBTAG00000010799	MYL6_BOVIN
4	ENSBTAG00000043584	ATP6_BOVIN	ENSBTAG00000011184	FRIH_BOVIN
5	ENSBTAG00000005333	MYG_BOVIN	ENSBTAG00000013264	RS24_BOVIN
6	ENSBTAG00000006424	TNNI3_BOVIN	ENSBTAG00000014731	G3P_BOVIN
7	ENSBTAG00000009703	Q3SZ64_BOVIN	ENSBTAG00000043550	CYB_BOVIN
8	ENSBTAG00000008394	MYL3_BOVIN	ENSBTAG00000006977	MYPR_BOVIN
9	ENSBTAG00000005714	ACTC_BOVIN	ENSBTAG00000005211	RL4_BOVIN
10	ENSBTAG00000019766	Not available	ENSBTAG00000030368	Not available
	Ensembl IDs of Kidney		Ensembl IDs of Liver	
1	ENSBTAG00000043553	GPX3	ENSBTAG00000017121	ALB
2	ENSBTAG00000011184	FRIH_BOVIN	ENSBTAG00000001638	A5PJE3_BOVIN
3	ENSBTAG00000043561	Q6QTG9_BOVIN	ENSBTAG00000017280	Q693V9_BOVIN
4	ENSBTAG00000002559	MIOX_BOVIN	ENSBTAG00000022120	A6QPX7_BOVIN
5	ENSBTAG00000015358	A5PK73_BOVIN	ENSBTAG00000016151	Q3ZBS7_BOVIN
6	ENSBTAG00000015358	A5PK73_BOVIN	ENSBTAG00000007273	TF
7	ENSBTAG00000033326	DPEP1_BOVIN	ENSBTAG00000031376	Q0PH99_BOVIN
8	ENSBTAG00000009535	RS2_BOVIN	ENSBTAG00000000442	RET4_BOVIN
9	ENSBTAG00000002648	RS18_BOVIN	ENSBTAG00000010123	APOE
10	ENSBTAG00000014534	EF1A1_BOVIN	ENSBTAG00000000522	FETUA_BOVIN
11	ENSBTAG00000016648	Q3ZBX0_BOVIN		
	Ensembl IDs of Lungs		Ensembl IDs of Mammary Gland	
1	ENSBTAG00000012560	PSPC_BOVIN	ENSBTAG00000025666	RPS29
2	ENSBTAG00000011184	FRIH_BOVIN	ENSBTAG00000010156	TCTP_BOVIN
3	ENSBTAG00000010156	TCTP_BOVIN	ENSBTAG00000005574	CLUS_BOVIN
4	ENSBTAG00000023032	SFTPA_BOVIN	ENSBTAG00000015285	RS8_BOVIN
5	ENSBTAG00000014534	EF1A1_BOVIN	ENSBTAG00000018320	RLA1_BOVIN
6	ENSBTAG00000005654	TYB10_BOVIN	ENSBTAG00000005296	RL13A_BOVIN
7	ENSBTAG00000021230	PSPB_BOVIN	ENSBTAG00000014678	Q9TRB9_BOVIN
8	ENSBTAG00000010799	MYL6_BOVIN	ENSBTAG00000019701	RL31_BOVIN
9	ENSBTAG00000002648	RS18_BOVIN	ENSBTAG00000009535	RS2_BOVIN
10	ENSBTAG00000009535	RS2_BOVIN		
	Ensembl IDs of Muscle		Ensembl IDs of Spleen	
1	ENSBTAG00000013921	KCRM_BOVIN	ENSBTAG00000001219	Not available
2	ENSBTAG00000004965	NUP133	ENSBTAG00000031160	IGLL1
3	ENSBTAG00000018369	MYL2	ENSBTAG00000015285	RS8_BOVIN
4	ENSBTAG00000022158	TNNT3_BOVIN	ENSBTAG00000015228	Q29630_BOVIN
5	ENSBTAG00000011424	TPM2_BOVIN	ENSBTAG00000010156	TCTP_BOVIN
6	ENSBTAG00000010880	A5PJM2_BOVIN	ENSBTAG00000009535	RS2_BOVIN
7	ENSBTAG00000021218	MLRS_BOVIN	ENSBTAG00000002648	RS18_BOVIN
8	ENSBTAG00000006419	TNNT1_BOVIN	ENSBTAG00000025666	RPS29
9	ENSBTAG00000010156	TCTP_BOVIN	ENSBTAG00000005654	TYB10_BOVIN
10	ENSBTAG00000011969	E1BEL7_BOVIN	ENSBTAG00000017389	F1MDN4_BOVIN

function related to cytoskeletal, protein binding, actin binding and actin monomer binding. The 9 cellular component GOs related to contractile and myosin complex. The left 47 GOs related to system processes, blood circulation, heart contraction, muscle structure development, cardiovascular system development, morphogenesis, muscle tissue development etc. For kidney, only GO term related to molecular function of monosaccharide binding was identified. For liver, top 10 highly expressed genes related to response to wounding (biological process), lipid binding (molecular function) and extracellular region, its part and space along with fibrinogen complex (cellular component). For lungs, the 10 most abundantly expressed genes related to respiratory gaseous exchange. The highly expressed genes in mammary gland related to structural molecular activity (molecular function) and ribosome, ribo-nucleoprotein complex (cellular component) and gene expression, protein metabolism, biosynthetic process of macromolecules (biological processes). For

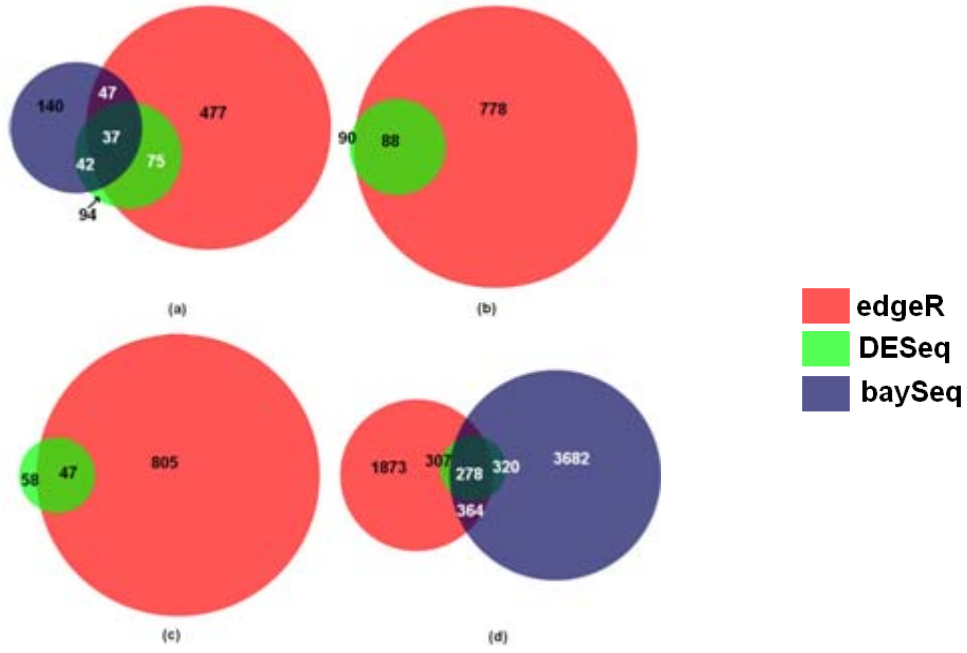
muscle, the highly expressed genes related to multi-cellular organismal movement, musculo-skeletal movement, skeletal muscle contraction (biological process) and cytoskeleton, contractile fiber part, contractile fiber, sarcomere (cellular component) and cytoskeletal protein binding (molecular function).

The high gene expression in top 10 genes of spleen related to translation, cellular component biogenesis, ribosome biogenesis (biological process) and multivesicular body, cytosolic part, ribosome, small ribosomal subunit (cellular component) and structural molecule activity, structural constituent of ribosome (molecular function). The list of 10 most expressed genes based on RPKM values for all the tissues of Osmanabadi goat with Ensembl IDs are shown in Table 3.

The expression values (RPKM) on 10 tissues were utilized for differential expression of genes among various tissues. Taken one at a time the tissue was compared with rest of the 9 tissues to find out the genes that were differentially expressed in a specific tissue. The normalization of data and significance of differential expression was tested using 3 algorithms, separately and a combination of 3 algorithms. The aim was ultimately to find out a set of genes that are differently expressed in each of the tissues, compared to rest of the 9 tissues. A total of 10 comparisons were made. We utilised the DEB server available at (http://www.ijbcb.org/DEB/php/onlinetool.php). The differentially expressed genes using the 3 algorithms had few genes identified as common to all the 3 algorithms at 10% False Discovery Rate (FDR). The results (Table 4) reveal that fibroblast, heart and muscle tissues did not show any differentially expressed gene using 3 algorithms while individual algorithms separately reveal differentially expressed genes. The largest numbers of genes differentially expressed were in hypothalamus, liver and kidney showed 278, 166 and 90 genes, respectively in the combination of the 3 algorithms. Figure 2 shows the Venn diagram

Table 4: Showing differentially expressed genes in specific tissue (Treatment) with rest of the nine tissues treated as controls

Tissues	edgeR	DESeq	baySeq	edger & DESeq	edgeR & baySeq	baySeq & DESeq	edger & DESeq & baySeq
Blood	477	94	140	75	47	42	37
Fibroblast	778	90	0	88	0	0	0
Heart	805	58	0	47	0	0	0
Hypothalamus	1873	357	3682	307	364	320	278
Kidney	1487	220	102	215	97	90	90
Liver	779	529	214	383	167	180	166
Lungs	210	34	36	33	24	10	10
Mammary gland	133	17	4	13	4	2	2
Muscle	719	0	0	0	0	0	0
Spleen	811	7	9	7	6	2	2



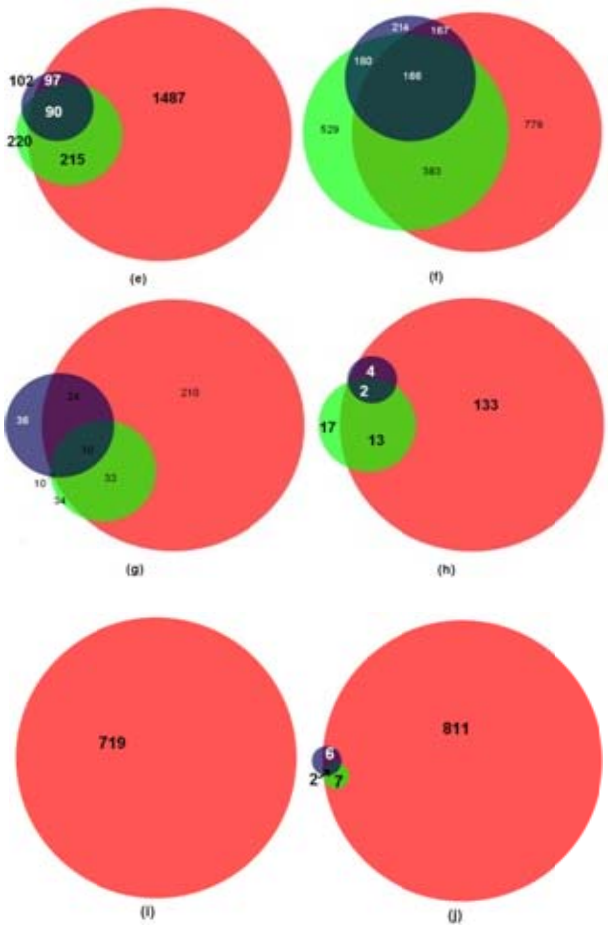


Fig. 2: Venn diagram showing differentially expressed genes for ten tissues of Osmanabadi goat (a) blood, (b) fibroblast, (c) heart, (d) hypothalamus, (e) kidney, (f) liver, (g) lungs, (h) mammary gland, (i) muscle and (j) spleen

representing the expression of genes utilising the 3 algorithms for each of the tissues of Osmanabadi goat. The differentially expressed genes common to 3 algorithms were mapped on to Gene Ontology terms by downloading associated GO terms of the genes from Ensembl release 62 of Btau_4.0. We utilised the web based program to batch analyse Gene Ontology Classification categories. The web-based tool utilized was available at (http://www.animalgenome.org/bioinfo/tools/catego/). The output was the definitions of the Biological Processes (BP), Cellular Components (CC) and Molecular Functions (MF). We subsequently extracted the above mentioned 3 larger categories (Table 5 and Fig. 3).

The rest of the GO term occurrences have been depicted as pie charts for the 6 tissues which exhibited differential gene expression (Fig. 4). The GO terms occurrence count by category for the 3 processes (biological processes, cellular component and moleular function) were 34%. The balance GO terms

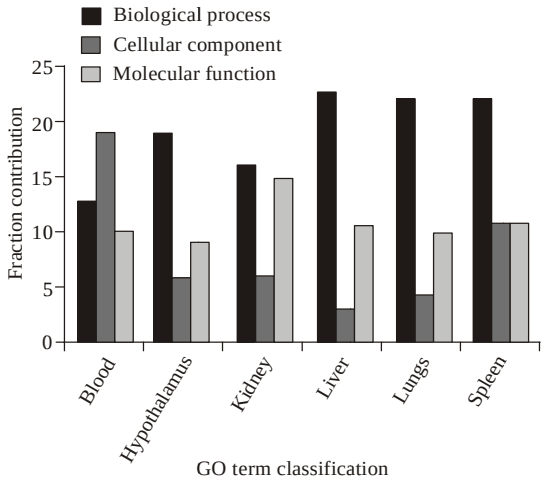
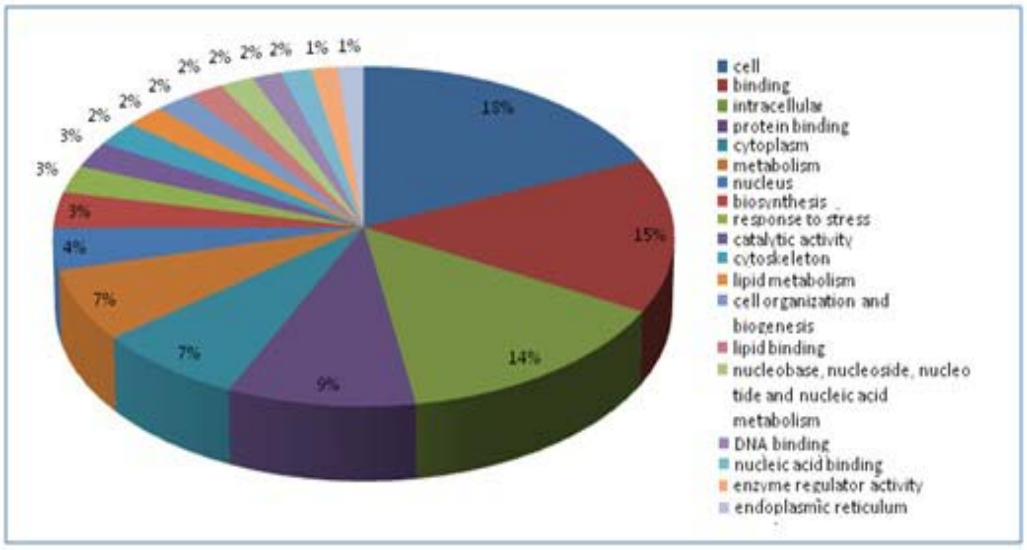


Fig. 3: A representative GO terms occurrence count by category for the biological process, molecular function and cellular component for the six tissues in Osmanabadi goat

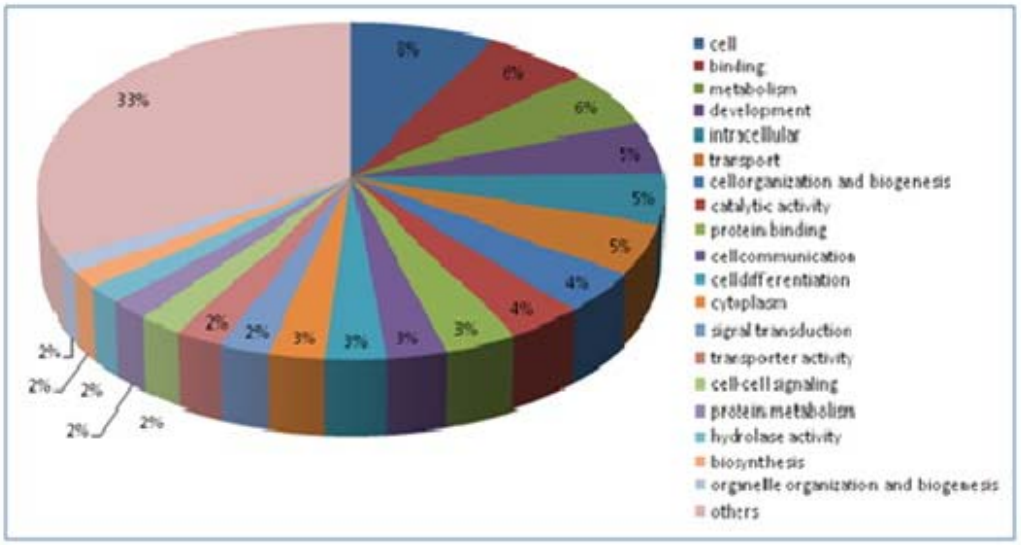
occurrence count of approx 66 % belong to metabolism, cell, binding, transport, intracellular catalytic activity development and transporter activity. The less frequent GO term occurrences were clubbed into miscellaneous "other" category. Thus the combined analysis of the genes expressed differentially in various tissues and their mapping to the GO terms and its analysis revealed that 30-45% GO term occurrences were for 3 major processes of biological process, cellular component and molecular function. The tissue wise GO term occurrence counts contributing to the 3 processes are depicted in the Table 5. It is to mention that fibroblast, heart and muscle of Osmanabadi goat do not reveal differentially expressed genes common to the 3 algorithms used for differential gene expression analysis and thus were not taken up for further study.

Table 5: Major processes associated with differentially expressed genes among the goat tissues based on GO term occurrence count

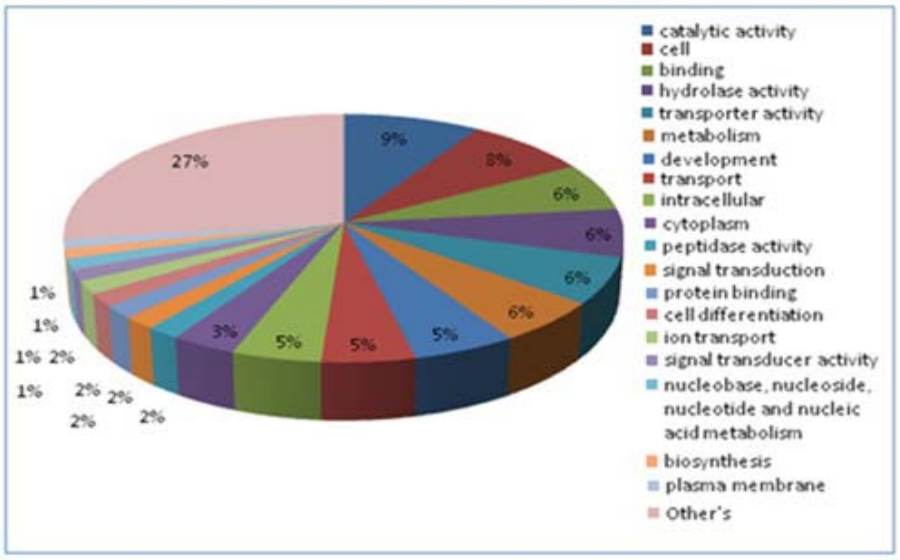
Definitions	Blood	Hypothalamus	Kidney	Liver	Lungs	Spleen
Biological process	12.74%	18.99%	16.24%	22.74%	22.18%	22.22%
Cellular component	19.07%	5.91%	5.92%	3.32%	4.44%	11.11%
Molecular function	9.94%	9.12%	14.87%	10.75%	10.08%	11.11%



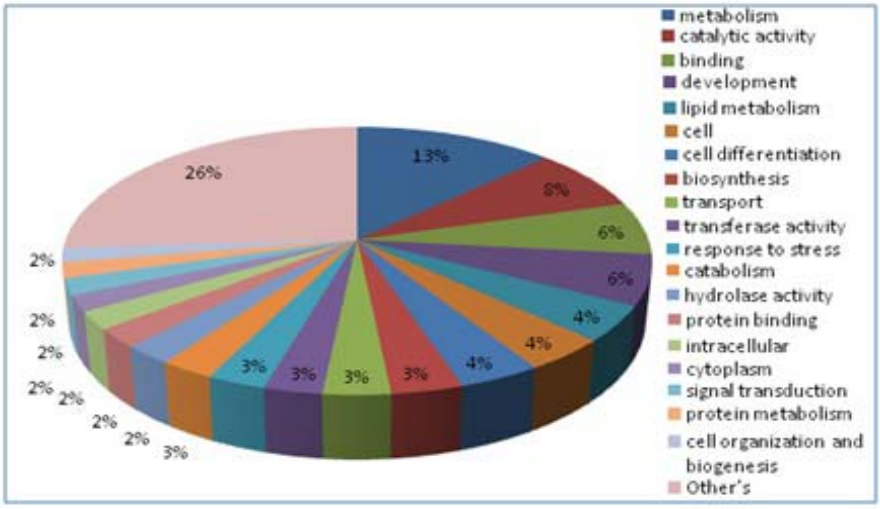
(a) blood



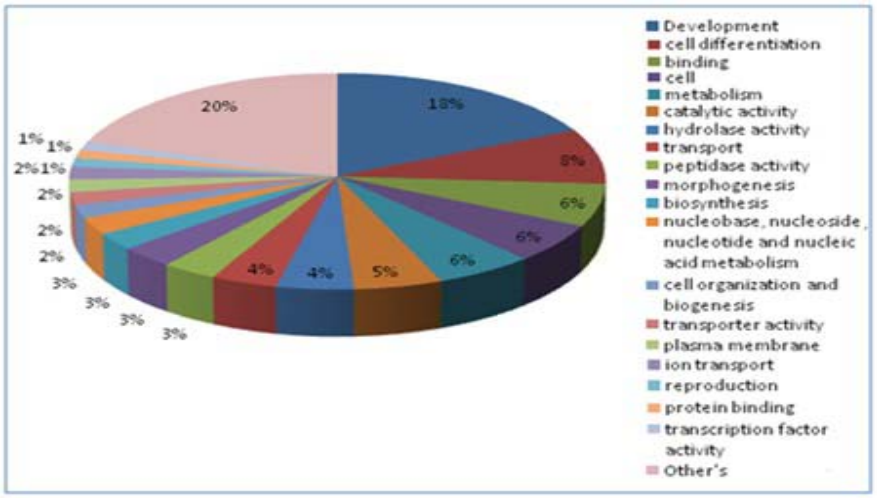
(b) hypothalamus



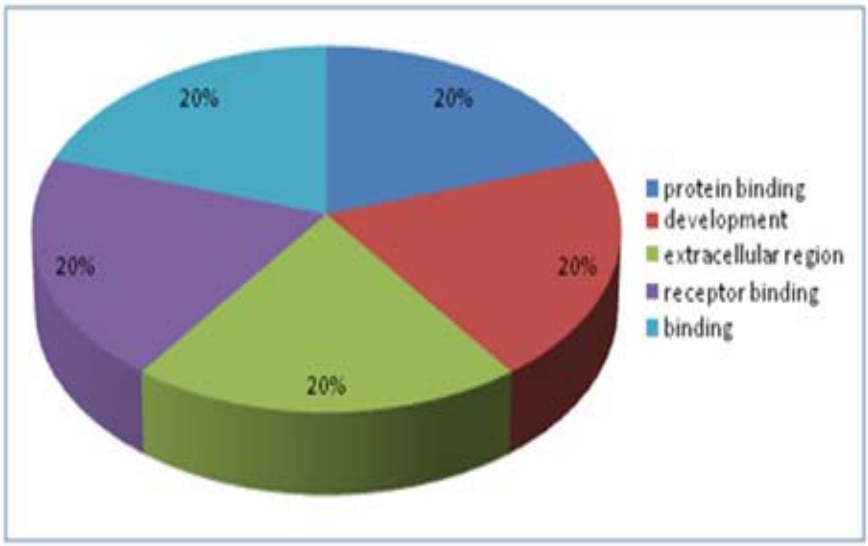
(c) kidney



(d) liver



(e) lungs



(f) Spleen

Fig. 4: Pie Chart showing GO terms occurrence count by category (other than BP, CC and MF) for (a) blood, (b) hypothalamus, (c) kidney, (d) liver, (e) lungs and (f) spleen tissue in Osmanabadi goat

Table 6: GO term occurrences count other than Biological processes, cellular component and molecular function in different tissues of goat

	GO term occurrence count by category		GO term occurrence count by category		GO term occurrence count by category
Blood	Fraction Contribution	Hypothalamus	Fraction Contribution	Kidney	Fraction Contribution
cell	8.89%	cell	5.25%	catalytic activity	5.61%
binding	7.09%	binding	3.91%	cell	5.01%
intracellular	6.94%	metabolism	3.78%	binding	3.95%
protein binding	4.50%	development	3.33%	hydrolase activity	3.95%
cytoplasm	3.57%	intracellular	3.20%	transporter activity	3.79%
Liver		Lungs		Spleen	
metabolism	8.08%	Development	11.29%	protein binding	11.11%
catalytic activity	4.82%	Cell differentiation	5.24%	development	11.11%
binding	4.04%	binding	4.03%	extracellular region	11.11%
development	3.84%	Cell	3.63%	receptor binding	11.11%
lipid metabolism	2.74%	metabolism	3.63%	binding	11.11%

The 2 genes differentially expressed in mammary gland could not be associated with GO terms. The fraction contribution of differentially expressed genes for 6 tissues (excluding biological processes, cellular component and molecular functions) is given in Table 6. The analysis revealed that the expressed genes may be tissue specific like lipid metabolism, catalytic activity of liver. Similarly for rest of the 5 tissues specific functions carried out by the gene expressed in specific tissues are reflected in the GO term occurrences.

DISCUSSION

RNA-Seq permits the discovery of novel transcripts. These may include mRNA molecules that encode fusion proteins in tumor cell populations or, as can be detected using sequence analysis, alternate splice forms (Ozsolak and Milos, 2011; Mortazavi *et al*., 2008). Known genes are relatively simple to identify and quantify in a normalized manner in reads per kilobase per million (RPKM), values, which reflect actual RNA levels (Mortazavi *et al*., 2008; Pepke *et al*., 2009). Exon sequences and known splice sites in reference sequences can be applied to RNA-Seq reads. The exon models used in a particular experiment however can influence RPKM values, depending on the presence of novel transcripts or transcripts that do not match reference sequences (Pepke *et al*., 2009). We utilized the CLC Genomics Workbench for the estimation of RPKM values and utilized the curated database downloaded from Ensembl. It was possible for us to reach fairly accurate level of identification of ubiquitous genes. The number of genes expressed in a tissue widely ranged with up to more than 16500 genes being expressed leading to a generalized conclusion

that most of the genes (>75%) are ubiquitously expressed. These genes are largely the ones which are poly A^+ RNA molecule from most of the tissues. The largest number of genes expressed has RPKM value of below 50. In most of the tissues the number of genes expressed at RPKM of >1000 were very few (<100). The number of genes expressed at different RPKM as a proportion to the total number showed a specific trend in all the tissues investigated. Thus in this experiment, we could investigate the total number of genes that were expressed in each of the tissue, the expression profile of each gene and the level of expression in terms of RPKM for each of the 10 tissues studied. A large number of genes that were expressed in all or most of the tissues were categorised as ubiquitous genes. The top 10 genes with highest level of expression were specific and revealed the highest investment of cellular machinery for the tissue specific tasks the genes carry out.

CONCLUSION

Our results demonstrated that RNA-Seq can be successfully used for gene identification and transcript profiling in goat. There is a great deal of homology between the genes of cattle and goats and this was expected as both are ruminants and phylogenetically close to one another. Next generation sequencing can provide highly accurate sequences and can quantify a large number of genes simultaneously in a given tissue. RNA-Seq analysis also provided a comprehensive view of the participation of several multigene families, identifying which members are expressed. The different genes had different expression levels in different tissue and a large number of genes were ubiquitous in nature. The edgeR, DESeq and baySeq algorithms identified varied number of genes as differentially expressed. There were however common genes among 3 algorithms. Usually the maximum numbers of genes were identified as differentially expressed by edgeR followed by DESeq and then baySeq. These differentially expressed genes were associated with Gene Ontology terms for their functional annotation. Therefore, the RNA-Seq method combined with the appropriate bioinformatic tools provides a new approach to study gene expression dynamics on a global scale allowing specific candidate genes to be highlighted for further functional analysis.

ACKNOWLEDGMENT

The study was funded by Indian Council of Agricultural Research grant C 30033 under National Agricultural Innovative Project Component IV, Basic and Strategic Research which is gratefully acknowledged.

REFERENCES

MacHugh, D.E. and D.G. Bradley, 2001. Livestock genetic origins: Goats buck the trend. Proc. Natl. Acad. Sci. USA., 98: 5382-5384.

Marioni, J.C., C.E. Mason, S.M. Mane, M. Stephens and Y. Gilad, 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res., 18: 1509-1517.

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer and B. Wold, 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Method., 5: 621-628.

Ozsolak, F. and P.M. Milos, 2011. RNAsequencing: advances, challenge sand opportunities. Nature Rev. Genet., 12: 87-98.

Pepke, S., B. Wold and A. Mortazavi, 2009. Computation for ChIP-seq and RNA-Seq studies. Nature Methods., 6: 22-32.

Qiang, Y.J. and Y. Fahong, 2011. DEB: A web interface for RNA-seq digital gene expression analysis. Bioinform., 7(1): 44-45.

Ramskold, D., E.T. Wang, C.B. Burge and R. Sandberg, 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol., 5(12): e1000598.

Wang, Z., M. Gerstein and M. Snyder, 2009. RNA-Seq: A revolutionary tool for transcriptomics. Nat Rev. Genet., 10: 57-63.

Zhi-Liang, H., J. Bao and J.M. Reecy, 2008. CateGOrizer: A web-based program to batch analyze gene ontology classification categories. Online J. Bioinform., 9(2): 108-112.