Title
SNP discovery in the bovine milk transcriptome using RNA-Seq technology

Journal
Mammalian Genome, 21(11)

ISSN
1432-1777

Authors
Cánovas, Angela
Rincon, Gonzalo
Islas-Trejo, Alma
et al.

Publication Date
2010-12-01

DOI
10.1007/s00335-010-9297-z

Peer reviewed

Mamm Genome (2010) 21:592–598
DOI 10.1007/s00335-010-9297-z

SNP discovery in the bovine milk transcriptome using RNA-Seq technology

Angela Cánovas · Gonzalo Rincon · Alma Islas-Trejo · Saumya Wickramasinghe · Juan F. Medrano

Received: 18 August 2010 / Accepted: 14 October 2010 / Published online: 6 November 2010

Abstract High-throughput sequencing of RNA (RNA-Seq) was developed primarily to analyze global gene expression in different tissues. However, it also is an efficient way to discover coding SNPs. The objective of this study was to perform a SNP discovery analysis in the milk transcriptome using RNA-Seq. Seven milk samples from Holstein cows were analyzed by sequencing cDNAs using the Illumina Genome Analyzer system. We detected 19,175 genes expressed in milk samples corresponding to approximately 70% of the total number of genes analyzed. The SNP detection analysis revealed 100,734 SNPs in Holstein samples, and a large number of those corresponded to differences between the Holstein breed and the Hereford bovine genome assembly Btau4.0. The number of polymorphic SNPs within Holstein cows was 33,045. The accuracy of RNA-Seq SNP discovery was tested by comparing SNPs detected in a set of 42 candidate genes expressed in milk that had been resequenced earlier using Sanger sequencing technology. Seventy of 86 SNPs were detected using both RNA-Seq and Sanger sequencing technologies. The KASPar Genotyping System was used to validate unique SNPs found by RNA-Seq but not observed by Sanger technology. Our results confirm that analyzing the transcriptome using RNA-Seq technology is an efficient and cost-effective method to identify SNPs in transcribed regions. This study creates guidelines to maximize the accuracy of SNP discovery and prevention of false-positive SNP detection, and provides more than 33,000 SNPs located in coding regions of genes expressed during lactation that can be used to develop genotyping platforms to perform marker-trait association studies in Holstein cattle.

A. Cánovas
IRTA, Genètica i Millora Animal, 191 Alcalde Rovira Roure Av, 25198 Lleida, Spain

G. Rincon · A. Islas-Trejo · S. Wickramasinghe · J. F. Medrano (✉)
Department of Animal Science, University of California-Davis, One Shields Ave, Davis 95616, CA, USA
e-mail: jfmedrano@ucdavis.edu

Introduction

Next-generation sequencing technologies have provided unprecedented opportunities for high-throughput functional genomic research, including gene expression profiling, genome annotation, small ncRNA discovery, and profiling and detection of aberrant transcription (Bentley 2006; Morozova and Marra 2008). Among these approaches, RNA sequencing (RNA-Seq) is a powerful new method for mapping and quantifying transcriptomes developed to analyze global gene expression in different tissues. Recently, this technique has also been used as an efficient and cost-effective method to systematically identify SNPs in transcribed regions in different species (Chepelev et al. 2009; Cirulli et al. 2010; Cloonan et al. 2008; Morin et al. 2008).

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 would not have been economically or logistically practical before (Marguerat et al. 2008). Taking this into account, we applied this novel approach to identify SNPs in the expressed coding regions of the bovine milk transcriptome.

The majority of gene expression analyses in the bovine mammary gland have been developed using a biopsy sample (Boutinaud and Jammes 2002; Finucane et al. 2008). An alternative sampling procedure has been proposed by

isolating mRNA directly from somatic cells that are naturally released into milk during lactation (Boutinaud et al. 2002). Recently, Medrano et al. (2010), using the RNA-Seq technique, compared the milk and mammary gland transcriptomes and showed extensive similarities of gene expression in both tissues.

In the present study we performed a SNP discovery analysis in milk transcriptome using RNA-Seq technology. For this purpose, seven milk samples from Holstein cows at different stages of lactation were analyzed by sequencing cDNA libraries using an Illumina GAII analyzer (Illumina, San Diego, CA) system. To evaluate the accuracy of SNPs detected with RNA-Seq, a comparison was made with SNPs detected in a set of 42 candidate genes expressed in milk that had been resequenced previously using Sanger sequencing technology. SNPs that were observed with only one technique were validated by the KASPar SNP Genotyping System.

Materials and methods

RNA-Seq library preparation

Seven milk samples were obtained from Holstein cows at two stages of lactation (day 15 and day 250). Milk samples were collected in 50-ml tubes 3 h after milking, kept on ice, and processed immediately for RNA extraction. Samples were centrifuged at $2000\times g$ for 10 min to obtain a pellet of cells. Total RNA was purified following a Trizol protocol (Invitrogen, Carlsbad, CA), and mRNA was isolated and purified using an RNA-Seq sample preparation kit (Illumina). mRNA was fragmented and first- and second-strand cDNA were synthesized. After adapters were ligated to the ends of double-stranded cDNA, a 300-bp fragment size was selected by gel excision and each sample was individually sequenced on an Illumina GAII analyzer.

RNA-Seq analysis and SNP detection

Short sequence reads (36-40 bp) were assembled and mapped to the annotated bovine reference genome Btau4.0 (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 seven samples were pooled to perform the RNA-Seq and SNP discovery analyses. We applied stringent criteria 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. Reads were then classified as uniquely mapped reads and nonspecifically mapped reads as shown in Table 1. 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 (4) SNPs located in read ends (last three bases) were not considered in the analysis due to possible sequencing errors.

Sanger resequencing of target genes and SNP detection

Resequencing was performed in a DNA resource population specifically developed for SNP discovery as described by Rincon et al. (2007). This population consisted of eight Holstein animals that were unrelated at least three generations back in their pedigrees. Genomic sequences for 42 candidate genes that were expressed in milk samples were obtained from the Btau4.0 assembly and resequenced using Sanger sequencing technology. Exons and conserved noncoding regions were identified using multiple-species genome alignments with Genome VISTA (Couronne et al. 2003). Coding regions and the conserved noncoding regions of each gene were resequenced at SeqWright DNA Technology Services (Houston, TX) using Sanger sequencing technology. SNPs were analyzed using CodonCode aligner software (http://www.codoncode.com); gene sequences and SNPs were assembled and annotated in Vector NTI advance 10.1.1 software (Invitrogen, Carlsbad, CA).

SNP validation by the KASPar SNP genotying system

The KASPar SNP Genotyping System (KBiociences, Herts, UK) was used to validate SNPs detected by RNA-Seq and not detected by Sanger sequencing. For this purpose, 15 bovine DNA samples (8 cows used for Sanger resequencing and 7 cows used for RNA-Seq) were selected. Genomic DNA was extracted from 5 ml of cow's blood following the protocol of the Gentra Puregene blood kit (Qiagen, Valencia, CA). KASPar assay primers (Table 2) were designed using the Primer Picker software available at http://www.kbioscience.co.uk/primer-picker.htm (KBiociences). Genotyping assays were carried out with a 7500 Fast Real Time instrument (Applied Biosystems, Foster City, CA) in a final volume of 8 µl containing 4× Reaction Mix (KBiociences), 120 nM of each allele-specific primers and 300 nM of common primer, 2.2 mM of $MgCl_2$, and 2 mM KTaq polymerase (KBiociences). The following thermal profile was used for all reactions: 15 min at 94°C; 20 cycles of 10 s at 94°C, 5 s at 57°C, and 10 s at 72°C; and 18 cycles of 10 s at 94°C, 20 s at 57°C, and 40 s at 72°C.

Table 1 Summary of mapping all the RNA-Seq reads to the reference genome (Btau4.0) obtained from seven pooled milk samples

	Uniquely mapped reads		Nonspecifically mapped reads		Mapped reads	
	No. of reads	%	No. of reads	%	No. of reads	%
Total exon reads	59888107	83	12480476	17	72368583	87.5
Exon-exon reads[a]	10961868	89	1313183	11	12275051	
Total intron reads	9100877	88	1271041	12	10371918	12.5
Exon-intron reads[b]	1475160	90	166238	10	1641398	
Total gene reads	68988984	83	13751517	17	82740501	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 KASPar primers used to validate SNP detected by RNA-Seq

Primer name	Sequence 5′ → 3′	Position[a]
DDIT3_60417924_A	GAAGGTCGGAGTCAACGGATTGGACTTCAGCCTTTAATATTGGAGAAA	I2
DDIT3_60417924_T	GAAGGTGACCAAGTTCATGCTGGACTTCAGCCTTTAATATTGGAGAAT	
DDIT3_60417924	CCATGGGATTTTCCAGGCAAGAGTA	I2
INSIG2_73132468_C	GAAGGTCGGAGTCAACGGATTAAGCACTCTTATAGTCTGCATGACG	I8
INSIG2_73132468_T	GAAGGTGACCAAGTTCATGCTAAAGCACTCTTATAGTCTGCATGACA	
INSIG2_73132468	ATATCGTATCACAGTGTTGATGTGCCAAA	I8
STAT5A_43749704_G	GAAGGTGACCAAGTTCATGCTCGAGCACCGGGTCAGGGC	I20
STAT5A_43749704_A	GAAGGTCGGAGTCAACGGATTCGAGCACCGGGTCAGGGT	
STAT5A_43749704	GCAGGCCAGCTCCCTCTGATA	E20
STAT5A_43746587_C	GAAGGTCGGAGTCAACGGATTCGCCTGGAAGTTTGACTCTCC	E15
STAT5A_43746587_G	GAAGGTGACCAAGTTCATGCTCGCCTGGAAGTTTGACTCTCG	
STAT5A_43746587	GGAGTGTGGGCAATGCAGGGAA	I16
STAT5A_43741732_T	GAAGGTGACCAAGTTCATGCTCTCCGCCAACTTCTCACACCT	I18
STAT5A_43741732_A	GAAGGTCGGAGTCAACGGATTCTCCGCCAACTTCTCACACCA	
STAT5A_43741732	GGCCCTGGGGCTCGGGTT	E18

[a] Position according to exon/intron distribution in bovine gene

Results and discussion

Detecting genetic variation in pooled milk transcriptome reads by RNA-Seq

RNA-Seq analysis included 118 million reads, ranging from 36 to 40 bp in size, that were assembled and mapped to the annotated NCBI bovine whole-genome assembly (27,368 genes). An average of 17 million short-sequence reads was obtained for each individual sample. The median coverage for the exons was 38×. The analysis revealed that 82.7 million reads (~70%) were categorized as mapped reads (68.9 million were uniquely mapped reads and 13.7 million were nonspecifically mapped reads), while 35 million were unmapped reads (Table 1). Most of the uniquely mapped reads corresponded to total exon reads (87.5%), whereas a small fraction corresponded to total intron reads (12.5%; Table 1). Intron reads are expressed regions that are not annotated as exons in Btau4.0.

RPKM (reads per kilobase per million mapped reads) (Mortazavi et al. 2008) values were used to identify the total number of genes that were expressed in the milk transcriptome. A RPKM threshold value of 0.3 was established in order to balance the number of false positives and false negatives as described in Bentley et al. (2008) and Ramskold et al. (2009). A total of 13,807 genes were selected with RPKM threshold values greater than 0.3. For those genes with RPKM < 0.3, a detailed analysis was performed to determine the number of unique reads falling outside exon regions that can be representing either annotation errors or new exons not included in the current Btau4.0 genome. Using this strategy, 5368 expressed genes/regions were found with more than ten unique reads. We detected 19,175 expressed genes in milk samples

(70.06%) of the 27,368 total bovine annotated genes in Btau4.0 genome assembly. The SNP detection analysis revealed 100,734 SNPs in the seven Holstein samples. Of these SNPs, 67,689 (67.2%) were homozygous, corresponding to differences between Holsteins and the Hereford bovine whole-genome assembly Btau4.0. This is a large number of SNPs that are fixed in Holstein for a different allele to that found in the Hereford genome reference and requires further investigation. In some cases these SNPs may represent artifacts due to errors in the reference sequence or due to misalignment of the short reads to the reference (see subsection "Validation of unique SNP detected by RNA-Seq" below). It may also be possible that some of the Holstein fixed SNPs in fact correspond to variants with a very low frequency and a large number of cows will be needed to detect the common Hereford variant. A total of 33,045 (32.8%) SNPs were polymorphic within Holsteins. Allele frequencies for these heterozygous SNPs were obtained for the pooled samples by counting the number of reads representing each allele. In summary, 1,849 SNPs had an allele frequency of 80/20, 5,511 SNPs had an allele frequency of 70/30, 15,411 SNPs had an allele frequency of 60/40, and 10,274 SNPs had an allele frequency of 50/50. Figure 1 represents the total number of SNPs per gene mapped to the bovine Btau4.0 genome assembly. SNPs that are different between the Hereford consensus sequence and that of Holstein are shown in red, and SNPs that are polymorphic in Holstein samples are in blue. We observed that SNPs in expressed regions are distributed along the entire genome, but there are an increasing number of polymorphisms located in the extremes of the chromosomes' centromeric and telomeric regions. The pattern of SNP distribution in each chromosome is very similar between those that differentiate Holstein and Hereford and those SNPs that are polymorphic in Holsteins, suggesting that there are genomic regions that tend to accumulate a large number of SNPs. Interestingly, the collagen family genes in BTA1, BTA4, BTA12, and BTA19 showed the highest SNP count difference between the Holstein and the Hereford consensus sequences. A large amount of data was generated in this study; a detailed description of the SNPs is available from the authors upon request.

Accuracy of RNA-Seq technology for SNP detection

To analyze the accuracy of RNA-Seq technology for SNP detection, 42 genes highly expressed in milk and related to fatty acid synthesis and the growth hormone GH/IGF axis were resequenced using Sanger methodology. Nine genes did not show polymorphisms in exons by Sanger resequencing and were excluded from the SNP discovery and validation analyses: *IGF1*, *IGFBP3*, *IGFBP4*, *MBTPS1*, *MBTPS2*, *NR3C1*, *PIAS1*, *STAT2*, and *STAT4*. Eighty-six SNPs were detected in the remaining 33 candidate genes that exhibited variation in Holsteins. Seventy of 86 SNPs were also detected by RNA-Seq in 18 genes (Table 3). From the 16 SNPs that were not detected by RNA-Seq, 6 were located in exons that were not expressed in milk samples and therefore no sequencing reads were found.

It is important to note that the samples used for RNA-Seq were different from those sequenced by the Sanger method, so we were not expecting a 100% concordance of the results. However, it is noteworthy that despite the difference in sample composition in the analysis, only ten SNPs observed in Sanger were not detected in RNA-Seq.

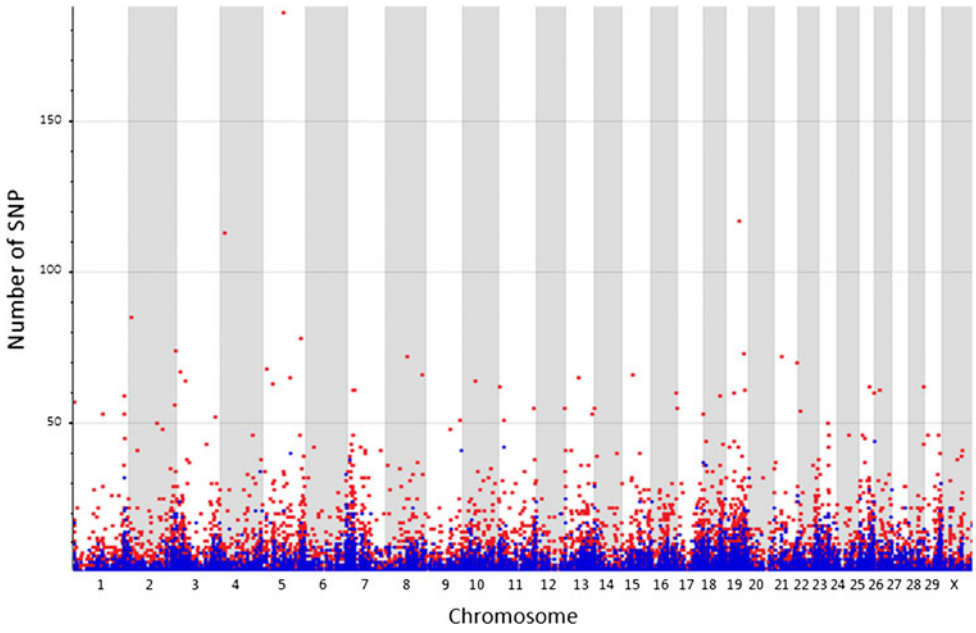

Fig. 1 Total number of SNPs per gene expressed in milk cells mapped to the Btau4.0 genome assembly. *Red dots* represent the number of SNPs per gene in coding regions that are different between Hereford and Holstein. *Blue dots* represent the SNPs per gene that are polymorphic in Holsteins

Table 3 List of 70 SNPs in 18 genes validated in coding regions using RNA-Seq and Sanger sequencing

Gene	BTA	SNP location[a]	Allele variation	Frequency
ADCY4 (Adenylate cyclase 4) ENSBTAG00000018419	10	21091039	A/T	50/50
		21093914	A/T	60/40
		21095171	A/T	60/40
		21096245	C/G	60/40
		21096645	T/A	60/40
CISH (Cytokine-inducible SH2-containing protein) ENSBTAG00000022622	22	50657856	G/T	50/50
		50659783	T/C	60/40
		50659810	C/T	60/40
DDIT3 (DNA damage-inducible transcript 3) ENSBTAG00000031544	5	60415461	C/G	60/40
		60416148	G/T	60/40
		60418030	G/A	60/40
FURIN (Trans Golgi network protease furin) ENSBTAG00000002939	21	21527200	C/G	50/50
		21528000	A/T	60/40
		21528001	T/G	60/40
		21528082	C/T	50/50
		21532215	C/A	60/40
IGF1R (Insulin-like growth factor 1 receptor) ENSBTAG00000021527	21	6862322	T/C	60/40
		6866036	T/C	60/40
		6868728	C/A	60/40
		6869964	A/C	50/50
IGFBP6 (Insulin-like growth factor-binding protein 6) ENSBTAG00000021467	5	29836177	A/T	50/50
INSIG1 (Insulin-induced gene 1) ENSBTAG00000001592	4	121362562	T/A	60/40
		121363743	C/A	60/40
		121363744	T/G	60/40
		121364402	A/C	50/50
		121364403	T/C	70/30
		121364574	T/G	60/40
		121364685	G/A	50/50
		121364911	A/G	60/40
		121365407	G/A	60/40
		121365607	T/C	60/40
		121365773	G/A	60/40
		121365941	G/A	60/40
		121366597	A/G	50/50
		121369136	G/A	60/40
		121369282	A/G	60/40
		121369843	G/A	60/40
		121370020	T/A	60/40
		121370021	T/A	50/50
		121371117	C/G	50/50
INSIG2 (Insulin-induced gene 2) ENSBTAG00000002112	2	73130467	C/T	50/50
NMI (N-myc (and STAT) interactor) ENSBTAG00000016219	2	47179315	C/G	60/40
PAPPA (Pregnancy-associated plasma protein-A) ENSBTAG00000004010	8	110792768	G/A	60/40
		110968532	C/T	60/40

Table 3 continued

Gene	BTA	SNP location[a]	Allele variation	Frequency
SCAP (Sterol regulatory element-binding protein cleavage- activating protein) ENSBTAG00000015782	22	53546833	C/T	50/50
		53549091	C/G	50/50
		53552713	A/T	60/40
SOCS5 (Suppressor of cytokine signaling 5) ENSBTAG00000008987	11	30359073	T/A	60/40
		30360874	T/C	60/40
SREBF1 (Sterol regulatory element binding protein-1) ENSBTAG00000007884	19	35680267	G/A	60/40
		35680574	A/G	60/40
		35682842	A/G	60/40
		35683588	G/C	60/40
		35685082	A/T	60/40
SRPR (Signal recognition particle receptor subunit alpha) ENSBTAG00000014105	29	31200612	A/C	50/50
STAT1 (Signal transducer and activator of transcription 1) ENSBTAG00000007867	2	83382392	A/T	60/40
STAT3 (Signal transducer and activator of transcription 3) ENSBTAG00000021523	19	43780445	A/G	70/30
		43780740	C/A	60/40
STAT5A (Signal transducer and activator of transcription 5A) ENSBTAG00000009496	19	43729581	C/G	50/50
		43730210	G/A	60/40
		43730211	T/A	60/40
		43741509	G/A	60/40
		43743914	C/G	50/50
		43745596	C/G	60/40
		43748702	C/G	70/30
STAT5B (Signal transducer and activator of transcription 5B) ENSBTAG00000010125	19	43655236	A/C	50/50
STAT6 (Signal transducer and activator of transcription 6) ENSBTAG00000006335	5	60837392	A/G	60/40
		60837393	C/T	50/50
		60845709	G/T	60/40
		60845948	G/T	60/40

[a] SNP location is based on the bovine genome assembly Btau4.0

On the other hand, five SNPs were observed in three genes, *DDIT3* (DNA-damage-inducible transcript 3), *INSIG2* (insulin-induced gene 2), and *STAT5A* (signal transducer and activator of transcription 5A), in RNA-Seq that were not detected by Sanger (Table 4). These SNPs were further validated using the KASPar SNP Genotyping System.

Validation of unique SNPs detected by RNA-Seq

In order to confirm the presence of the five SNPs uniquely found by RNA-Seq, they were genotyped using the KASPar SNP Genotyping System. Three out the five SNPs were validated, as shown in Table 4. Two SNPs in the *STAT5A* gene that failed with the KASPar assay were further examined with a detailed analysis of the sequence reads containing the putative SNPs. We observed that the corresponding 40-bp sequence that mapped to a *STAT5A* region had a 99% homology with the *STAT5B* gene. In the SNP discovery analysis we set up a threshold of a maximum number of mismatches to two. With this mismatch

Table 4 Unique SNPs detected by RNA-Seq that were validated using the KASPar Genotyping System

Gene	Chromosome	SNP	Position	Frequency (%)	Confirmed[a]
DDIT3	5	T/A	60417924	50.0/50.0	Yes
INSIG2	2	T/C	73132468	50.0/50.0	Yes
STAT5A	19	G/A	43749704	54.5/45.5	Yes
STAT5A	19	G/C	43746587	70.6/29.4	No
STAT5A	19	T/A	43741732	66.7/33.3	No

DDIT3 DNA damage inducible transcript 3, *INSIG2* insulin-induced gene 2, *STAT5A* signal transducer and activator of transcription 5A

[a] SNP confirmed by KASPar SNP Genotyping System and Sanger resequencing

rate, reads that correspond to a given gene, like *STAT5B*, can be assigned to *STAT5A*. This was not a common situation for most of the genes studied in this analysis, but it could represent a problem in gene families with highly conserved domains when using short sequence reads. In a similar study, Cirulli et al. (2010) observed that some false-positive SNPs identified in cDNA arose from alignment of a read to the wrong gene and that in these cases the correct gene and the gene chosen for the alignment always had very similar sequences. This situation has also been observed in regions associated with sequence repeats (Morozova and Marra 2008). Although the short-read structure of next-generation sequencers has some potential problems with respect to sequence assembly, the result is a system that generates accurate data and large coverage of consensus sequence and SNP calling at very high throughput and low cost (Thomas et al. 2006).

Conclusion

We have demonstrated that analyzing the transcriptome using RNA-Seq technology is an efficient and cost-effective method to identify SNPs in transcribed regions. Stringent criteria have to be applied to maximize the accuracy and prevent false-positive SNP detection. This study provides a valuable resource of more than 33,000 SNPs located in coding regions of genes expressed during lactation that can be used for further gene variation analysis and association studies in Holstein cattle.

Acknowledgments This project was supported by a grant from Dairy Management Inc. and the California Dairy Research Foundation. We thank Charlie Nicolet of the UC Davis Genome Center for his excellent technical expertise to perform Illumina GAII sequencing. A. Cánovas received a predoctoral scholarship from INIA, Spain.

References

Bentley DR (2006) Whole-genome re-sequencing. Curr Opin Genet Dev 16:545–552

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J et al (2008) Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53–59

Boutinaud M, Jammes H (2002) Potential uses of milk epithelial cells: a review. Reprod Nutr Dev 42:133–147

Boutinaud M, Rulquin H, Keisler DH, Djiane J, Jammes H (2002) Use of somatic cells from goat milk for dynamic studies of gene expression in the mammary gland. J Anim Sci 80:1258–1269

Chepelev I, Wei G, Tang Q, Zhao K (2009) Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucl Acids Res 37:e106

Cirulli ET, Singh A, Shianna KV, Ge D, Smith JP et al (2010) Screening the human exome: a comparison of whole genome and whole transcriptome sequencing. Genome Biol 11:R57

Cloonan N, Forrest A, Kolle G, Gardiner B, Faulkner G et al (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods 5:613–619

Couronne O, Poliakov A, Bray N, Ishkhanov T, Ryaboy D et al (2003) Strategies and tools for whole-genome alignments. Genome Res 13:73–80

Finucane KA, McFadden TB, Bond JP, Kennelly JJ, Zhao FQ (2008) Onset of lactation in the bovine mammary gland: gene expression profiling indicates a strong inhibition of gene expression in cell proliferation. Funct Integr Genomics 8:251–264

Marguerat S, Wilhelm BT, Bahler J (2008) Next-generation sequencing: applications beyond genomes. Biochem Soc Trans 36:1091–1096

Medrano JF, Rincon G, Islas-Trejo A (2010) Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. In: 9th World congress on genetics applied to livestock production, Leipzig, Germany, August 1-6, 2010, Paper no 0852

Morin R, O'Connor M, Griffith M, Kuchenbauer F, Delaney A et al (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18:610–621

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255–264

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

Ramskold D, Wang ET, Burge CB, Sandberg R (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput Biol 5:e1000598

Rincon G, Thomas M, Medrano JF (2007) SNP identification in genes involved in GH-IGF1 signaling on BTA5. In: Plant & animal genomes XV conference, San Diego, CA, January 13-17, 2007, Abstract no P536

Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T et al (2006) Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med 12:852–855