GENE EXPRESSION: AN OVERVIEW OF METHODS AND APPLICATIONS FOR CANCER RESEARCH

Marina Mitie de Souza Monobe[1]
Rodrigo Costa da Silva[2]

ABSTRACT

Gene expression is the study of how the genotype gives rise to the phenotype by investigating the amount of transcribed mRNA in a biological system. A lot of methods have been standardized to identify the variation in gene expression, including subtractive hybridization, differential display, serial analysis of gene expression, microarray hybridization, and RNA-seq sequencing. Most of techniques have been focused in cancer research and diagnosis, producing a huge amount of data, which allowed to understand the cancer progression and pathways, discover and evaluate new treatment interventions, new molecular tools for diagnosis and prognosis, and analyze the survival time in human and animal patients. In this way, gene expression techniques brought new important perspectives for the medical and veterinary fields, and further researches focusing oncology will provide much more knowledge concerning the pathways and interaction of healthy and tumor cells, improving the perspectives of the daily interventions by the oncologists and clinicians.

Keywords: gene sequencing, gene expression, molecular techniques, genotype, veterinary oncology.

EXPRESSÃO GÊNICA: UMA VISÃO GERAL DOS MÉTODOS E APLICAÇÕES NA PESQUISA DO CÂNCER

RESUMO

A expressão genética é o estudo de como o genótipo dá origem ao fenótipo a partir da investigação da quantidade de RNAm transcrito em um sistema biológico. Vários métodos já foram padronizados para identificar variações na expressão gênica, dentre eles a hibridização subtrativa, "differential display", análise em série da expressão genética, hibridização de microarranjo, e sequenciamento por RNA-seq. A maioria das técnicas tem focado na pesquisa e diagnóstico do câncer, gerando enorme quantidade de dados, o que permitiu compreender a progressão do câncer e suas vias, descobrir e analisar novas intervenções terapêuticas, novas ferramentas moleculares para o diagnóstico e prognóstico, e analisar o tempo de sobrevivência em pacientes humanos e animais. Desta forma, as diferentes técnicas de expressão gênica trouxeram novas e importantes perspectivas para a área médica e veterinária, e novas pesquisas focadas em oncologia fornecerão muito mais conhecimento sobre as vias e interações entre células saudáveis e tumorais, melhorando as perspectivas das intervenções diárias pelos oncologistas e clínicos.

Palavras-chave: sequência gênica, expressão gênica, técnicas moleculares, genótipo, oncologia veterinária.

[1] D.V.M. PhD Candidate, Department of Pathobiology and Population Medicine, College of Veterinary Medicine, Mississippi State University, Mississippi State USA. Contato principal para correspondência.
[2] D.V.M., PhD, Postdoctoral Associate Department of Pathobiology and Population Medicine, College of Veterinary Medicine, Mississippi State University, Mississippi State.

EXPRESIÓN GÉNICA: UNA VISIÓN GENERAL DE LOS MÉTODOS Y APLICACIONES EN LA PESQUISA DE CÁNCER

RESUMÉN

La expresión génica es el estudio de cómo el genotipo da lugar al fenotipo mediante la investigación de la cantidad de RNAm transcrito en un sistema biológico. Una gran cantidad de métodos fueron estandarizados para identificar variaciones en la expresión génica, incluyendo la hibridación sustractiva, "differential display", análisis en serie de la expresión génica, la hibridación de microarrays, y la secuenciación por RNA-seq. La mayoría de las técnicas se han centrado en la investigación y diagnóstico del cáncer, produciendo una gran cantidad de datos, lo que permitió a entender la progresión del cáncer y las vías, descubrir y evaluar nuevas intervenciones de tratamiento, nuevas herramientas moleculares para el diagnóstico y el pronóstico, y analizar el tiempo de sobrevivencia en pacientes humanos y animales. De esta manera, las técnicas de expresión génica trajeron nuevas perspectivas importantes para el campo de la medicina veterinaria, y nuevas investigaciones centradas en oncología proporcionarán mucho más conocimiento acerca de las vías y la interacción de las células sanas y tumorales, mejorando las intervenciones diarias por los oncólogos y los clínicos.

Palabras clave: secuencia génica, expresión génica, técnicas moleculares, genotipo, oncología veterinaria.

INTRODUCTION

Gene expression is a highly regulated mechanism that controls the function and adaptability of all living cells in which a gene information is converted to a functional gene product that can be either a protein or non-coding genes such as transfer ribonucleic acid (tRNA) or small nuclear RNA (snRNA) (1,2). The control of gene expression is the main point on understanding the interaction genotype: phenotype (3), because this process is used for all known life. Protein formation involves four main steps: transcription, RNA splicing, translation and posttranslational modification. During transcription, a single strand mRNA is complementary copied from one strand of a template DNA by the RNA polymerase. While transcription occurs, some modifications on the mRNA, i.e., RNA splicing, in which introns are removed from the sequence. After that, the new mRNA is used as a template to assemble a chain of amino acids to form the protein (4). Gene expression studies the amount of transcribed mRNA in a biological system and its importance to the cell function.

Sequencing techniques that lead to the knowledge of the entire genome sequences have been stimulating researchers to define their function (3). As genes with related functions are usually regulated together, techniques that evaluate global gene expression provide an important way for the initial identification and clustering of these genes. The functional gene sequence clustering provides enough information to direct additional experiments at defining the precise function of certain gene product.

The analysis of the gene expression can be done by targeting a single specific gene, also known as low-to-mid plex(reporter gene, Northern, Western, fluorescent "in situ hybridization – FISH, slot and dot blotting, semiquantitative and quantitative reverse transcriptase polymerase chain reaction – RT-PCR, and nuclease protection assays),or all genes that differ in expression among different experimental samples, also known as higher plex (subtractive hybridization, differential display, serial analysis of gene expression, microarray hybridization, RNA-seq sequencing), or that differ between healthy and cancer patients (5).Gene expression and pathway analysis allows identify several biological

processes and molecular functions that may be involved also with infectious (i.e., *Mycobacterium bovis*-infected bovine leukocytes) (6-8), and non-infectious (i.e., exercise-induced stress in equine, and canine and bovine cancers) (9-12).

In this way, the present paper was aimed to review the most important methods for the gene expression analysis in different experimental samples, highlighting the advantages and disadvantages of each one, and provide some applications in cancer research.

Methods of gene expression analyzes.

Subtractive hybridization

The subtractive hybridization method was first described in 1983 by Sargent and Dawig to create a cDNA library (13)and generated probes (14) of different expressed genes. The principle of this methodology is to differentiate the expressed genes by hybridizing cDNA from one sample (tester sample) to an excess of mRNA from a second or control sample (driver). The transcripts that are expressed in both samples form a mRNA/cDNA hybrid molecule, whereas the sequence expressed only in the tester maintains as a single strand cDNA. Then, double and single strands are separated using hydroxylapatite chromatography, and the eluted single strands can be cloned, sequenced or used directly as probes for screening libraries. The problems with this method include the requirement of a large amount of mRNA (i.e. 270 µg for the first described protocol against 2µg of mRNA usually used on microarrays' protocol), and the observation of errors on identifying rare transcripts. After that, this protocol was updated along the time, and a similar protocol known as suppression subtractive hybridization (SSH) was created.

SSH differentiates from the original protocol by the selective amplification of differentially expressed transcripts and elimination of the separation single-double strand step (15). This protocol is based on the extraction of mRNA from two experimental samples (tester and driver) and performance of a RT-PCR, which will result in cDNA formation. Then the cDNA testers are bonding with two different adapter oligonucleotides (1 and 2) and until the second hybridization, the reactions of the tester-adapter 1and 2 are separated done. A first subtractive hybridization combine the excess of cDNA drivers with the testers in two separate reactions in which denaturation occurs and single stranded molecules can re-hybridize. One reaction involves the tester 1 and the other the tester 2. After the first hybridization, four possible molecules may result: single strand tester, double strand tester, hybrid tester-driver transcript and double strand driver. The second hybridization mixes all products from the two types of adapters. Then, the result can be a double strand expressing the two types of adapters, double strand with the same type of adapter, single strand tester 1 or 2, double strand adapter 1 or 2 with a driver, or a single strand driver. A PCR is performed using primers that recognize the adapter 1 (forward primer) and 2 (reverse primer).This reaction will result in the amplification of the double strand cDNA containing the two types of adapters (Figure 1).

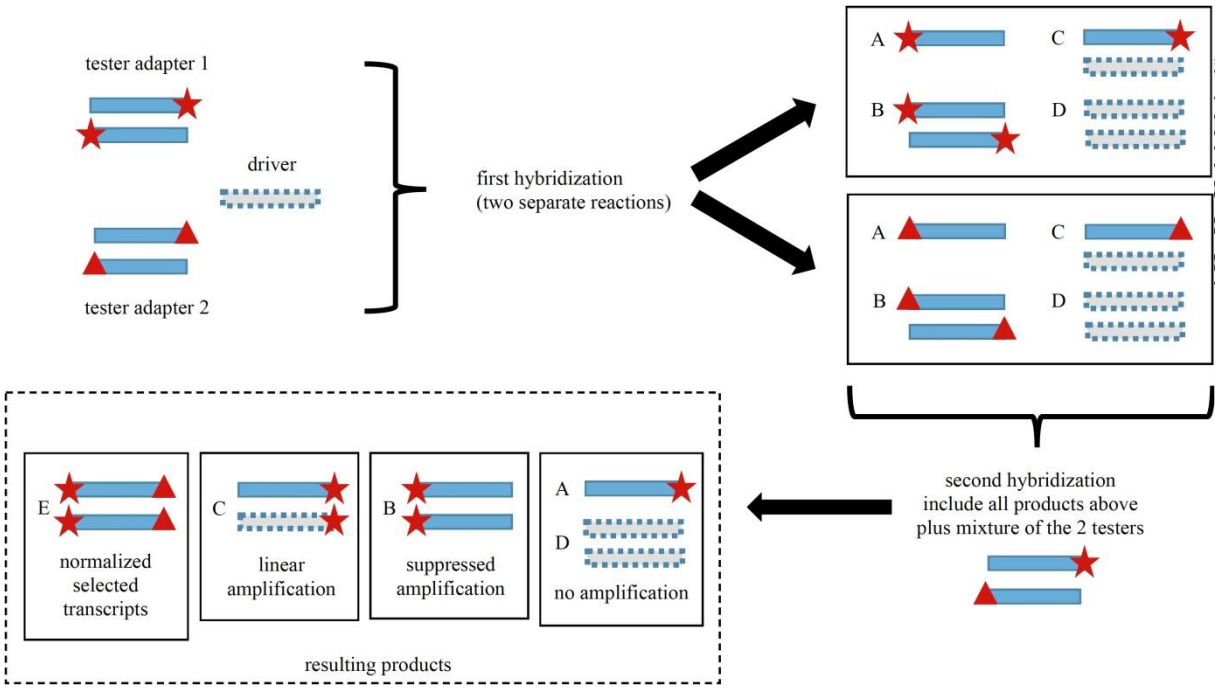

Figure 1. Suppression Subtractive Hybridization (SSH). This method includes two reaction steps. First step is carried out with two separated reactions, each one with an excess of driver and one type of Tester Adapter. Each reaction will result four types of products: (A) single stranded tester transcripts (B) double stranded tester transcripts (B) hybrids of tester and driver transcripts (D) double stranded driver. The second hybridization step includes products generated from the first step plus additional denatured driver cDNA. Possible resulting products include same type ones from the first step plus (E) a double stranded driver formed by the two types of adapters. A final PCR reaction is done to selectively amplify the amount of (E) drivers. Figure adapted from Moody (5).

The advantage using SSH is that it does not requiree prior knowledge of the transcript sequence to isolate it and does not require specialized equipment or analysis only industrialized kits. However, the limitation is that it can only apply for a matched pair treatment comparison and it is not a quantitative method (5).

SSH has been widely used in cancer research. For example, one performed research using normal cells and colorectal cancer cells found out that 37 different genes are overexpressed in this type of cancer and some of them may provide novel points for therapeutic intervention (16). Another research using hormone-responsive and hormone-nonresponsive breast carcinoma proved that a specific set of genes is responsible for regulating the estrogen receptors and causing the hormone-responsive breast carcinoma. These genes are not expressed in the other type of breast cancer neither in normal mammary cells, which means that its expression can help providing an accurate diagnosis and prognosis for these patients (17).

Differential display

RNA fingerprinting technique includes two different protocols: differential display and arbitrary primed PCR (RAP-PCR). Both protocols are based on PCR amplification of random subsets of genes from two or more samples that will lead to a different pattern of PCR product bands visualized during electrophoresis. This "unique" transcript pattern is called RNA fingerprint (18).

These methods include RNA extraction, RT-PCR forming the cDNA and a main PCR reaction that include one same forward primer and 12 to 20 different reverse primers. This

PCR reaction results in more than 240 combinations of sequences that are enough to represent all the mRNA present in the original RNA sample. After this reaction, the PCR products are labeled and an electrophoresis on polyacrylamide gel is performed (Figure 2). The differentially expressed gene is represented by comparing the presence and/or intensity of the bands in the different samples (19). These PCR products might be selected, purified from the gel and reamplified by PCR to be used in sequencing and other techniques.

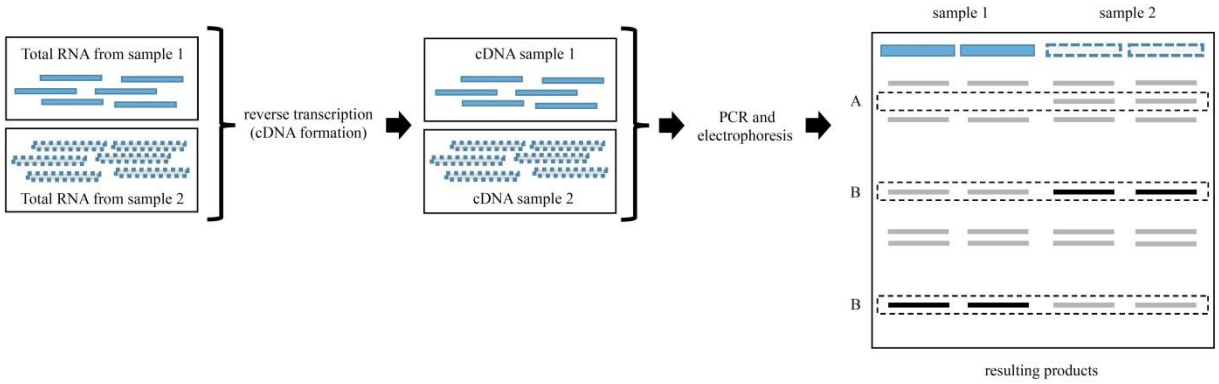

Figure 2. Differential display (RNA fingerprint method). Technique includes extraction of total RNA from two or more samples that are reverse transcribed to cDNA. Then, random cDNAs are amplified in duplicate by PCR using different combinations of forward and reverse primers. The amplified products are visualized by gel of electrophoresis and each sample will have a "unique" pattern of transcripts' products. Some fragments might appear only in one sample but not in other (A), some might have higher or lower intensity expression comparing to the other sample (B). Adapted from Moody (5).

The main difference between differential display and RAP-PCR is that differential display use anchored primer, which recognizes the Poly-A tail on the mRNAs while the RAP-PCR use arbitrary specific sequence primer with approximately 10 base pairs (19, 20).

The advantages include comparison of multiple samples and identification of up and down regulated genes. However, both protocols are not quantitative and are time consuming. Several false positive have also been described (21).

At cancer research, many researches in prostate cancer found out a highly specific gene (differential display code 3 gene, DDC3) that has been used as a marker for the diagnosis of this cancer and as a target for gene therapy (22-24). In gene therapy, studies focusing the regulation of DDC3 have observed better results in chemotherapy and regression of the cancer (23). Other different studies at the cancer field involve the identification of expression of specific genes after some kind of treatment. Murata et al. (25) analyzed the expression of the PHLDA1 gene in oral cancer cell lines in humans. They observed the overexpression of this gene in oral cancer and verified its significant suppression by hydrogen-sulfide oral therapy. Thus, the use of hydrogen-sulfide can prevent the oral cancer and help during its therapy in humans. In the same way, Finocchiaro et al. (26) have used a gene target therapy based on the high expression of interferon-βgenes at the melanoma cancer cells in dogs.

Expressed Sequence Tag sequencing

The conception of expressed sequence tag sequencing (EST) was first mentioned in 1991 (27). The principle includes the random selection of clones from a cDNA library and the performance of a single sequence reaction to produce 300-500 bp of sequence per clone (Figure 3). The differentially expressed genes are identified by the different number of times it appears at the EST library when compared with a control sample. Higher numbers of

counted times in an EST library means higher gene expression (28). EST method usually needs to be combined with other different gene expression methods because it is not an accurate qualitative and quantitative data. It is often generated from normalized cDNA libraries.

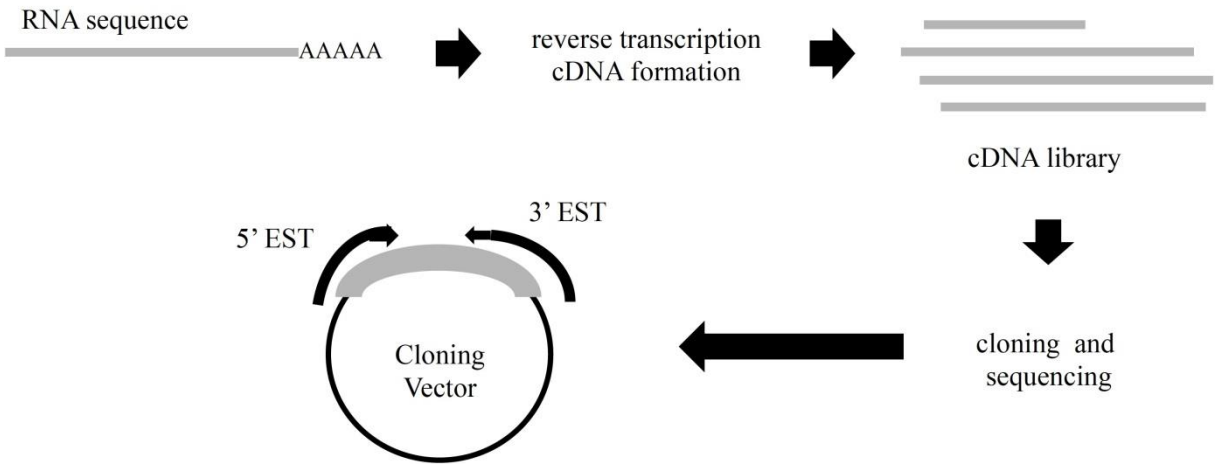

Figure 3. Expressed Sequence Tag sequencing (EST). EST is a short sub-sequence of cDNA selected from a cDNA library with 300-500 bp length. Different expressed genes are identified by the amount of sequences produced by the clones when compared with a control sample. Adapted from Moody (5).

Cancer research rarely uses EST only. However, the development of EST databases and normalized cDNA library may increase the use of this technique combined with other ones. The first description of the profile gene expression in breast cancer using EST combined with differential display was in 1994 (29). They concluded that the both methods combined still have great limitations mainly caused by the high heterogeneity of normal and cancer tissue leading to a large numbers of different ETS libraries. Moreover, the contamination of breast cancer tissue with normal breast tissue during the procedure also leads to misidentification of the specific expressed genes in cancer tissue. Therefore, the use of ETS remains uncommon especially in cancer research area.

Serial Analysis of Gene Expression (SAGE)

Although SAGE was first described in 1995 (30), it has been used recently in the cancer research combined with a new EST generation. This combined technique also known as Tag-seq allows the identification of more genes when compared with the SAGE protocol only. The researchers tested this method in breast cancer and concluded it is an accurate, sensitive and reproducible method to produce gene expression profile in cancers in general (31). Besides the cancer area, this method is also relevant to study gene expression profiling in other disease pathways.

Actually, SAGE is an accelerate version of EST sequencing in which short sequence tags (SAGE tags) are used to identify a gene transcript (Figure 4). The protocol includes RNA extraction and RT-PCR performance using a biotin-tagged primer. The resulted cDNA is digested by a specific endonuclease known as *Nla*III, which recognizes 4 (four) base pairs (bp) sequence CATG. The digested fragments are purified and then two linkers, A and B, are added in separate reactions. These linkers are complementary sequence of the PCR primers and serve as recognition sequence for another endonuclease *Fok*I. So, a second digestion step is done. The *Fok*I recognizes the linkers and cleaves the cDNA at 9 (nine) bp downstream. The products are purified again, ligated together and amplified by PCR. A final digestion step using the *Nla*III is performed to cleave away the linkers, producing ditags with 9bp from two

different transcripts. Multiple ditags are ligated together, cloned and sequenced to produce the SAGE sequence data (5).

In sum, SAGE tag is a nine base sequence locate specifically downstream from the last endonuclease recognition site of a transcript. Once each unique SAGE tag represents a unique transcript, it allows an overview of all genes expressed in the original RNA sample. The differences in gene expression can be identified by comparing the amount of specific SAGE tags in different libraries (32).

The advantages of SAGE are it is an accurate, quantitative and cumulative method. The disadvantages are that it requires specific equipment and misidentification of tags may occur. Misidentification includes from producing sequence errors to misidentification of the region in a database. Moreover, depending on the specific enzyme used in the protocol, it may fail to identify the transcripts (33).

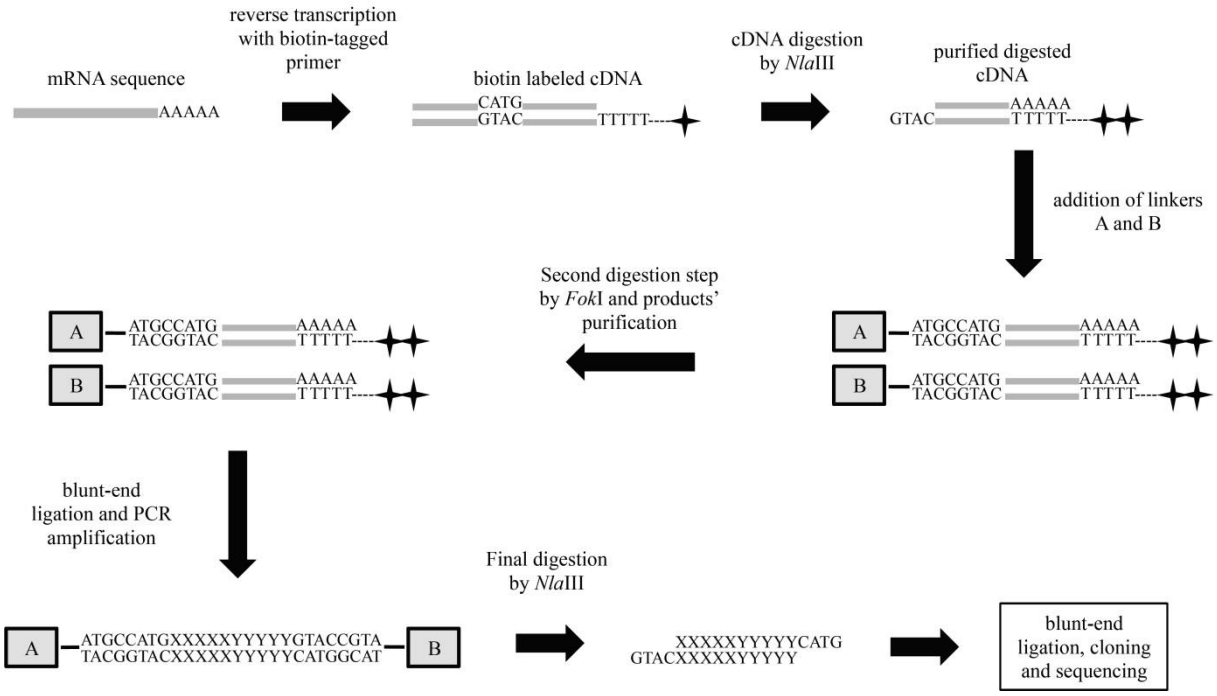

Figure 4. Serial Analysis of Gene Expression (SAGE). Adapted from Moody (5).

Microarray hybridization, RNA/DNA microarray

Hybridization-based techniques like microarray rely on and are limited to the transcripts bound to the array slides (3). The microarray technology was first described in 1995 (34). However, increasing improvements and changes on microarrays techniques and equipment still makes it one of the most common and newly technique in a variety of fields (35). Even that, microarrays are only as good as the bioinformatic data available for the model organism's genome and transcriptome,the complete set of transcripts in a cell (36).

Nowadays, different types of microarrays protocols are available such as protein microarray, RNA microarray, DNA microarray and Glycan microarray. The difference among them is what they are measuring. In other words, if the microarray is used to measure the amount of RNA it is called RNA microarray (35).

The principle is the hybridization of two DNA or mRNA, for RNA microarray, sequences that have complementary base pair. The first step includes purificationof the samples and obtainment of all DNA or cDNA. After that, these sequences are fluorescently or radioactively labeled and added in a microarray containing sequence fragments representing a

great variety of genes. Hybridization between the DNA/cDNA samples and the sequences in the microarray occurs, followed by washing off the non-specific bonding sequences. Once hybridized, the sequences release a fluorescent signal, which is captured by the equipment (Figure 5). The hybridization signal is proportional to the abundance of that mRNA transcript or DNA sequence from the original sample (37).

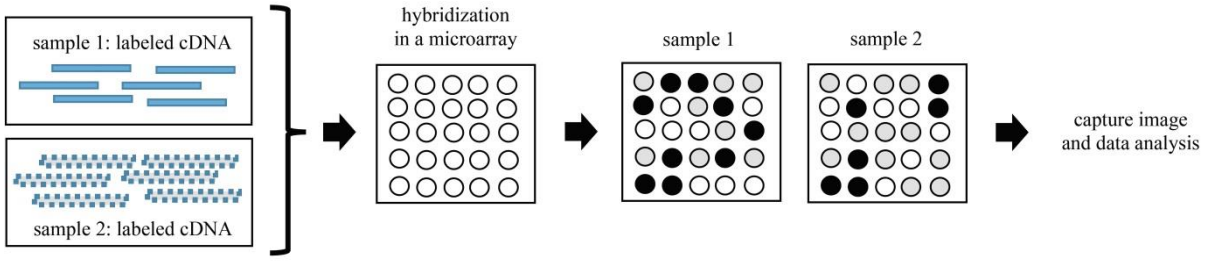

Figure 5. Microarray hybridization. The method includes reverse transcription of mRNA from two samples (i.e., control and experimental sample) into cDNA and labeling them. All cDNAs are hybridized in a microarray containing DNA fragments of different specific genes. Each sample generates a signal when hybridization occurs. The intensity of the signal is proportional to the abundance of the initial mRNA used. Signals from different samples are compared to identify differences in gene expression. Adapted from Moody (5).

Three different types of microarrays are commercial available: microarrays that include oligonucleotide chips made by the synthesis of short oligos directly on slide glass (Affymetrix Technology), microarrays made by spotting pre-synthesized oligos into glass slides or nylon membranes and cDNA microarrays made by spotting amplified PCR products from cDNA library clones onto glass slides or nylon membranes (5).

The selection of gene fragments also known as probes, which are spotted on the array, is one of the most important part. Even though it is possible to select more than 3,000-10,000 sequences, it might be possible that these sequences are not the best representation for a type of experiment (5). Each experiment should have its own cDNA/DNA library and the use of controls is required.

The analysis of the array data is divided in three main steps: identification and quantification of the hybridization signal, visualization of the data and clustering techniques. The clustering step include the organization of the genes in groups with similar expression patterns and it can be done by multiple available software (38).

Microarray technology is considered an accurate, sensitive, quantitative and qualitative method. Although it requires technological and specific equipment with high cost, it is main actual method used in gene expression research (35).

Cancer research has been widely using microarrays to generate gene expression profiling. In breast cancer, the discovery of the *DUSP4* gene leaded to the use of its expressed transcript as a molecular marker for the diagnosis of the negative breast cancer, an aggressive type of breast cancer. This gene is responsible for the resistance of the cancer cells to chemotherapy drugs (39). Another interesting research using this method was in the lung cancer area. By analyzing the different gene expression profiling between nasal cells from patients with lung cancer and healthy patients, the researchers found out that non-neoplastic cells from nasal tissue from patients with lung cancer express the same specific biomarkers of the lung cancer cells. Thus, they suggested a new non-invasive method of diagnosis of lung cancer once the usual diagnosis is performed by collecting samples directly from the cancer cells (40,41). Microarrays have added important results in the cancer area. Due that, this method has been being modified, improved and widely used in veterinary medicine, mainly in focusing canine mammary carcinoma (42-44) and osteosarcoma (45). Fowles et al. (45) observed that microarray gene expression based models built in an intra- or interspecies

manner can successfully predict the drug response in canine osteosarcoma, which may improve outcome in dogs and serve as pre-clinical validation for similar methods in human cancer research.

In various types of cancer, i.e. canine mammary carcinoma, different intracellular signaling pathways are perturbed. Thus, patients with pathologically the same type of cancer often have different genetic defects in their tumors, which also result in different responses to the anticancer treatment (43). To solve this gap, Pawlowski et al. (43) suggested that both molecular and pathological characterization of the tumor can complement one another. Also, microarray gene expression can discriminate metastatic canine mammary carcinomas from non-metastatic ones, as observed by Klopfleisch et al. (42), in which 265/1,011 (26.21%) significantly differentially expressed genes in metastatic, but not in non-metastatic canine carcinoma, are related to human breast cancer and, vice versa, parts of a human prognostic gene signature were identified in the expression profiles of the metastatic canine tumors, which allows the identification of prognostic molecular signatures for potential therapeutic targets.

Rao et al. (44) used cDNA microarray to characterize the canine mammary carcinoma cell lines by the expression profiling, and concluded that the identification of major differences observed in the signaling pathwaysallowsits use to develop new therapiesdirected towards specific pathways.

RNA-Sequencing (RNA-Seq) or Whole Transcriptome Shotgun Sequencing (WTSS)

RNA-Seq is a powerful and recent approach to the analysis of the transcriptomethat uses the capabilities of next-generation sequencing (NGS) to detect and quantify the interest RNA from a genome at a moment (46-48). It providesa more precise measurement of levels of transcripts and their isoforms,and a better transcriptome analysis of the genome with higher resolution than actual methods (48,49).

RNA-Seq detects annotated transcripts as microarrays also do, besides novel sequences and splice variants (36),and provides additional tools by the generation of large amount of data, with an increased base coverage of whole DNA sequences in human, and animal genome (6,9,10,50-52). This technique looks at different populations of RNA to include total RNA, small RNA, such as miRNA, tRNA, and ribosomal profiling (53), also known as key players in cancer development and progression (54).

The basic components of the RNA-Seq technology in the clinical context include 5 (five) major steps: RNA extraction from clinical samples, NGS library preparation, sequencing, data analysis, and data interpretation/reporting (54). It allows the increased base coverage of a DNA sequence, higher sample throughput, and the use of data to characterize exon junctions, detect non-coding RNA(7), alternative gene spliced transcripts, post-transcriptional modifications, gene fusion, mutations/SNPs, and changes in gene expression(Figure 6) (55). Furthermore, existing data sets can be re-evaluated as new sequences (56).

RNA-Seq has a more wide coverage than microarrays, once arrays target the identification of known common alleles that represent approximately 500,000 to 2,000,000 SNPs of the more than 10,000,000 in the genome (57). Because the libraries are not usually available to detect and evaluate rare allele variant transcripts (58), and the arrays are only as good as the SNP databases they are designed from, it limitsitsuse for research purposes (59).

The sample processing method for RNA-Seq involves fragmenting transcripts, and more fragments will be available for sequencing as longer is the transcript (60). This length bias is one limitation not observed for microarrays, as well as the proportional expression levels of the degree of hybridization to probes. Although microarrays are reliable and more cost effective than RNA-Seq for gene expression profiling in model organisms, it is still more

routinely used. Nowadays the techniques can be complementary used to each other (3). Additionally, biases in both methods exist for higher abundance transcripts and underscore the need for validation of results, which can be validated by other molecular techniques, i.e., quantitative PCR (qPCR),or proteomics (61).

Microarrays have been used a lot for clinical applications, and probably will be approved earlier than RNA-Seq for diagnostic use, mainly focusing the application on bioinformatic data (3). As cancer is characterized by genetic lesions directly affecting genes, i.e., point mutation, insertion, deletion, translocation, exon-skipping and gene fusion, RNA-Seq can detectthese mutant genes in lesions, and measure the transcript abundance (54,62), as well as gene fusions,which can result from many types of genetic lesions, e.g., translocation, deletion and inversion;it allows the identification of biomarkers for cancer risks, subtypes and stages of progression (54), discover cancer specific isoforms, global transcriptomic alterations and post-transcriptional changes on large scale (63), reflecting on the cancer diagnosis,and prognosis of the human (54), and animal disease (11).

In dogs, the breed-specific cancer susceptibility combined with recent advances in high-throughput sequencing technologies (e.g., RNA-Seq) allows the use of pet dog populations to find often subtle mutations that promote cancer susceptibility and progression (10). A good example of theapplicability of this technology on cancer field is in canine osteosarcoma, the best example of breed-enriched cancer, with an incidence rate of about 52 per 100,000, keeping dogs as a good model for human osteosarcoma studies (51,52). Not only osteosarcoma, but also histiocytic sarcoma and bladder cancer can use dogs as a model for RNA-Seq analysis. In this way, RNA-Seq can provide knowledge to expand the existing areas of research, and detect the misexpression of many genes between human and animal, affected and normal tissues (64).

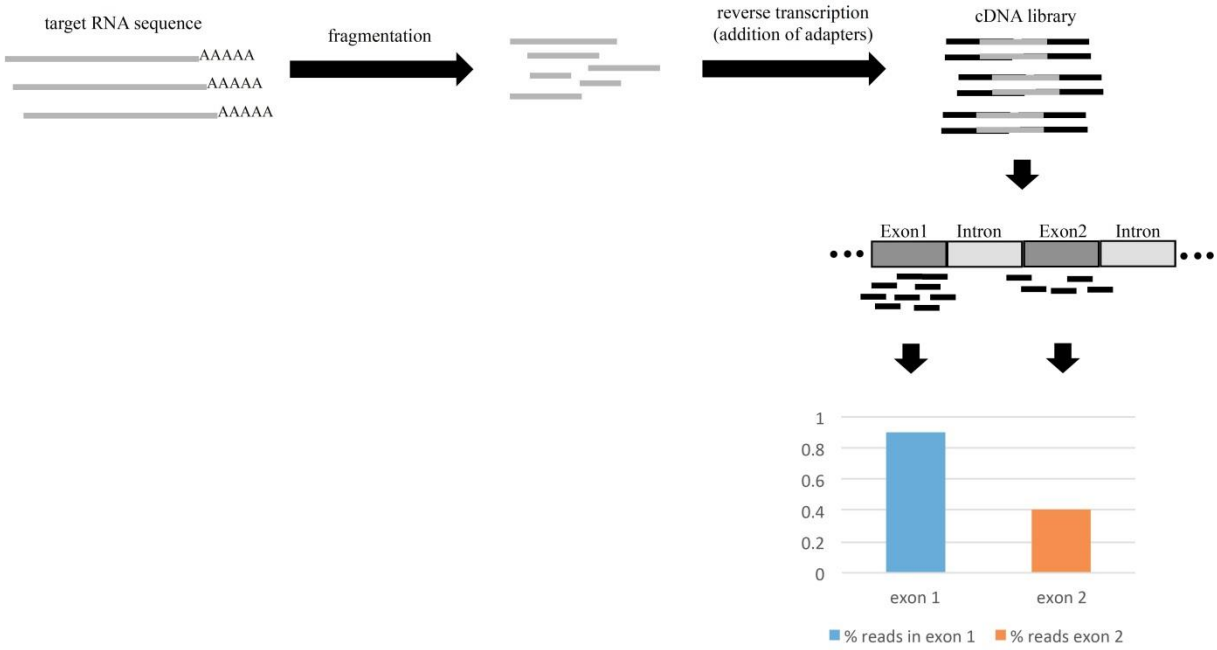

Figure 6. RNA-Seq method.

FINAL CONSIDERATIONS

The evaluation of gene expression is an effective way to identify important genes related to the regulation and progress of cancers in general. Several techniques and protocols were developed for this purpose and currently researches are modifying and improving techniques to facilitate the continuing discovery in this area. The application of gene

expression methods in cancer research involves from the study of different biological pathways involved in the cancer progression, development of different markers that can be used in the diagnosis and prognosis of specific cancers and the development of different protocols of treatments. Protocol treatments include the synthesis of specific targets, which can be combined with radioisotopes or chemotherapy drugs leading them to specific attack cancer cells, minimizing the patient's clinical signs; and gene therapy, which include the use of vectors to modify the expression of cancer genes. Therefore, gene expression methods are a powerful tool for the cancer research area.

REFERENCES

1. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Control of gene expression. In: Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. 5th ed. New York: Garland Science; 2007. p.1112-30.

2. Hoopes L. Introduction to the gene expression and regulation topic room. Nat Educ. 2008;1(1):160.

3. Mantione KJ, Kream RM, Kuzelova H, Ptacek R, Raboch J, Samuel JM, et al. Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. Med Sci Monit Basic Res. 2014;20:138-42.

4. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. Expression of the genome. In: Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R. Molecular biology of the gene. 7th ed. New York: Cold Spring Harbor Laboratory Press; 2014. p.423-608.

5. Moody DE. Genomics techniques: an overview of methods for the study of gene expression. J Anim Sci. 2001;79(E-suppl):E128-35.

6. McLoughlin KE, Nalpas NC, Rue-Albrecht K, Browne JA, Magee DA, Killick KE, et al. RNA-seq transcriptional profiling of peripheral blood leukocytes from cattle infected with *Mycobacterium bovis*. Front Immunol. 2014;5:396.

7. Arnvig KB, Comas I, Thomson NR, Houghton J, Boshoff HI, Croucher NJ, et al. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of *Mycobacterium tuberculosis*. PLoS Pathog. 2011;7(11):e1002342.

8. Qian F, Chung L, Zheng W, Bruno V, Alexander RP, Wang Z, et al. Identification of genes critical for resistance to infection by West Nile virus using RNA-Seq analysis. Viruses. 2013;5(7):1664-81.

9. Capomaccio S, Vitulo N, Verini-Supplizi A, Barcaccia G, Albiero A, D'Angelo M, et al. RNA sequencing of the exercise transcriptome in equine athletes. PLoS One. 2013;8(12):e83504.

10. Davis BW, Ostrander EA. Domestic dogs and cancer research: a breed-based genomics approach. ILAR J. 2014;55(1):59-68.

11. Patel AK, Bhatt VD, Tripathi AK, Sajnani MR, Jakhesara SJ, Koringa PG, et al. Identification of novel splice variants in horn cancer by RNA-Seq analysis in Zebu cattle. Genomics. 2013;101(1):57-63.

12. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, et al. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. Cancer Prev Res. 2011;4(6):803-17.

13. Sargent TD, Dawid IB. Differential gene expression in the gastrula of *Xenopus laevis*. Science. 1983;222(4620):135-9.

14. Davis MM, Cohen DI, Nielsen EA, Steinmetz M, Paul WE, Hood L. Cell-type-specific cDNA probes and the murine I region: the localization and orientation of Ad alpha. Proc Natl Acad Sci U S A. 1984;81(7):2194-8.

15. Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A. 1996;93(12):6025-30.

16. Hufton SE, Moerkerk PT, Brandwijk R, de Bruine AP, Arends JW, Hoogenboom HR. A profile of differentially expressed genes in primary colorectal cancer using suppression subtractive hybridization. FEBS Lett. 1999;463(1-2):77-82.

17. Kuang WW, Thompson DA, Hoch RV, Weigel RJ. Differential screening and suppression subtractive hybridization identified genes differentially expressed in an estrogen receptor-positive breast carcinoma cell line. Nucleic Acids Res. 1998;26(4):1116-23.

18. Parmigiani G, Garrett ES, Irizarry RA, Zeger SL. The analysis of gene expression data: an overview of methods and software. In: The analysis of gene expression data: methods and software. New York: Springer; 2003. p.1-45.

19. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science. 1992;257(5072):967-71.

20. Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Res. 1992;20(19):4965-70.

21. Janzen MA, Kuhlers DL, Jungst SB, Louis CF. ARPP-16 mRNA is up-regulated in the longissimus muscle of pigs possessing an elevated growth rate. J Anim Sci. 2000;78(6):1475-84.

22. Hu E, Wang D, Chen J, Tao X. Novel cyclotides from *Hedyotis diffusa* induce apoptosis and inhibit proliferation and migration of prostate cancer cells. Int J Clin Exp Med. 2015;8(3):4059-65.

23. Schalken JA, Hessels D, Verhaegh G. New targets for therapy in prostate cancer: differential display code 3 (DD3(PCA3)), a highly prostate cancer-specific gene. Urology. 2003;62(5 Suppl 1):34-43.

24. Tatsumi Y, Nakagawara A, inventors; Hisamitsu Pharmaceutical Co., Inc. Chiba-Prefecture, assignee. Cancer marker and therapeutic agent for cancer. United States patent US 008202690B2. 2012 Jun 19.

25. Murata T, Sato T, Kamoda T, Moriyama H, Kumazawa Y, Hanada N. Differential susceptibility to hydrogen sulfide-induced apoptosis between PHLDA1-overexpressing oral cancer cell lines and oral keratinocytes: role of PHLDA1 as an apoptosis suppressor. Exp Cell Res. 2014;320(2):247-57.

26. Finocchiaro LM, Fondello C, Gil-Cardeza ML, Rossi UA, Villaverde MS, Riveros MD, et al. Cytokine-enhanced vaccine and interferon-beta plus suicide gene therapy as surgery adjuvant treatments for spontaneous canine melanoma. Hum Gene Ther. 2015;26(6):367-76.

27. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 1991;252(5013):1651-6.

28. Soares MB, Bonaldo MF, Jelene P, Su L, Lawton L, Efstratiadis A. Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci U S A. 1994;91(20):9228-32.

29. Watson MA, Fleming TP. Isolation of differentially expressed sequence tags from human breast cancer. Cancer Res. 1994;54(17):4598-602.

30. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science. 1995;270(5235):484-7.

31. Morrissy AS, Morin RD, Delaney A, Zeng T, McDonald H, Jones S, et al. Next-generation tag sequencing for cancer gene expression profiling. Genome Res. 2009;19(10):1825-35.

32. Datson NA, van der Perk-de Jong J, van den Berg MP, de Kloet ER, Vreugdenhil E. MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. Nucleic Acids Res. 1999;27(5):1300-7.

33. Hanriot L, Keime C, Gay N, Faure C, Dossat C, Wincker P, et al. A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. BMC Genomics. 2008;9:418.

34. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270(5235):467-70.

35. Sealfon SC, Chu TT. RNA and DNA microarrays. Methods Mol Biol. 2011;671:3-34.

36. Howard BE, Hu Q, Babaoglu AC, Chandra M, Borghi M, Tan X, et al. High-throughput RNA sequencing of pseudomonas-infected *Arabidopsis* reveals hidden transcriptome complexity and novel splice variants. PLoS One. 2013;8(10):e74183.

37. Churchill GA. Fundamentals of experimental design for cDNA microarrays. Nat Genet. 2002;32 Suppl:490-5.

38. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863-8.

39. Baglia ML, Cai Q, Zheng Y, Wu J, Su Y, Ye F, et al. Dual specificity phosphatase 4 gene expression in association with triple-negative breast cancer outcome. Breast Cancer Res Treat. 2014;148(1):211-20.

40. Perez-Rogers JF, Gerrein J, Anderlind C, Kusko RL, Campbell JD, Wang TW, et al., editors. Leveraging gene expression in the bronchial airway to develop a nasal biomarker for early detection of lung cancer. Proceedings of the American Thoracic Society 2014 International Conference; 2014; San Diego. San Diego; ATSJournals; 2014.

41. Walter K, Holcomb T, Januario T, Yauch RL, Du P, Bourgon R, et al. Discovery and development of DNA methylation-based biomarkers for lung cancer. Epigenomics. 2014;6(1):59-72.

42. Klopfleisch R, Lenze D, Hummel M, Gruber AD. Metastatic canine mammary carcinomas can be identified by a gene expression profile that partly overlaps with human breast cancer profiles. BMC Cancer. 2010;10:618.

43. Pawlowski KM, Maciejewski H, Dolka I, Mol JA, Motyl T, Krol M. Gene expression profiles in canine mammary carcinomas of various grades of malignancy. BMC Vet Res. 2013;9:78.

44. Rao NA, van Wolferen ME, van den Ham R, van Leenen D, Groot Koerkamp MJ, Holstege FC, et al. cDNA microarray profiles of canine mammary tumour cell lines reveal deregulated pathways pertaining to their phenotype. Anim Genet. 2008;39(4):333-45.

45. Fowles JS, Brown KC, Hess AM, Duval DL, Gustafson DL. Intra- and interspecies gene expression models for predicting drug response in canine osteosarcoma. BMC Bioinformatics. 2016;17(1):93.

46. Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques. 2008;45(1):81-94.

47. Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther. 2012;22(4):271-4.

48. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol. 2010; Suppl 89:1-13.

49. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.

50. Gnimpieba EZ, Chango A, Lushbough CM. RNA-Seq gene and transcript expression analysis using the BioExtract Server and iPlant Collaborative. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2014); 2014 Sept 20-23; Newport Beach, CA. Newport Beach, CA: Association for Computing Machinery; 2014. p.661-9.

51. Rankin KS, Starkey M, Lunec J, Gerrand CH, Murphy S, Biswas S. Of dogs and men: comparative biology as a tool for the discovery of novel biomarkers and drug development targets in osteosarcoma. Pediatr Blood Cancer. 2012;58(3):327-33.

52. Mueller F, Fuchs B, Kaser-Hotz B. Comparative biology of human and canine osteosarcoma. Anticancer Res. 2007;27(1A):155-64.

53. Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc. 2012;7(8):1534-50.

54. Wan M, Wang J, Gao X, Sklar J. RNA sequencing and its applications in cancer diagnosis and targeted therapy. North Am J Med Sci. 2014;7(4):156-62.

55. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009;458(7234):97-101.

56. Roberts A, Schaeffer L, Pachter L. Updating RNA-Seq analyses after re-annotation. Bioinformatics. 2013;29(13):1631-7.

57. Ravi I, Baunthiyal M, Saxena J. Advances in biotechnology. New Delhi: Springer; 2014.

58. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008;18(9):1509-17.

59. Siu H, Zhu Y, Jin L, Xiong M. Implication of next-generation sequencing on association studies. BMC Genomics. 2011;12:322.

60. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010;11(2):R14.

61. Fu X, Fu N, Guo S, Yan Z, Xu Y, Hu H, et al. Estimating accuracy of RNA-Seq and microarrays with proteomics. BMC Genomics. 2009;10:161.

62. Atak ZK, Gianfelici V, Hulselmans G, De Keersmaecker K, Devasia AG, Geerdens E, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. PLoS Genet. 2013;9(12):e1003997.

63. Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, et al. RNA sequencing of cancer reveals novel splicing alterations. Sci Rep. 2013;3:1689.

64. Märtson A, Kõks S, Reimann E, Prans E, Erm T, Maasalu K. Transcriptome analysis of osteosarcoma identifies suppression of wnt pathway and up-regulation of adiponectin as potential biomarker. Genom Discov. 2013;1(3):1-9.

Recebido em: 19/06/2015
Aceito em: 26/09/2016