

Comprehensive molecular screening: from the RT-PCR to the RNA-seq

Carlota Costa¹, Ana Giménez-Capitán¹, Niki Karachaliou¹, Rafael Rosell^{1,2}

¹Breakthrough Cancer Research Unit, Pangaea Biotech S.L, USP Dexeus University Institute, Barcelona, Spain; ²Catalan Institute of Oncology, Hospital Germans Trias i Pujol, Ctra. de Can Ruti s/n. 08916 Badalona, Spain

Corresponding to: Carlota Costa, PhD. PangaeaBiotech/USP Instituto Universitario Dexeus, Laboratorio de Oncología, c/Sabino de Arana 5-19, Spain. Email: ccosta@pangaeabiotech.com.

Abstract: Up to now, the analysis of the mRNA expression in tumoral and non-tumoral has been conducted via RT-PCR. It is considered to be the gold standard for measuring the number of copies of specific cDNA targets. The application of RT-PCR has demonstrated that levels of RNA transcripts stratify patients and predict outcomes in a variety of diseases, providing the basis for several important clinical tests. However, the inherent variability in the quality of any quantitative PCR data makes it difficult to replicate and the analysis is time consuming in the laboratory for the analysis of one gene. Moreover, comparing expression levels across different experiments is often difficult and can require complicated normalization methods.

Many techniques have been developed over the years but without good clinical applications. A new, simple and effective way to measure transcriptome composition and to discover new exons or genes is by the RNA-seq. Some advantages of this technique are high reproducibility, the large dynamic range, requirement of less sample RNA, and the ability to detect novel transcripts, alternative splicing, even in the absence of a sequenced genome.

However, this RNA-Seq technique will not likely replace current RT-PCR methods, but will be complementary depending on the needs and the resources of the clinic and the laboratory as the results of the RNA-Seq will identify those genes that need to then be examined using RT-PCR methods.

The application of the two complementary technologies in the routine analysis of cancer laboratories would be useful in characterizing patients and would assist oncologists in making clinical decisions, as it allows us to identify all molecular characteristics of the tumor.

Key Words: Molecular screening; RT-PCR; RNA-seq; mRNA expression



Submitted Jan 18, 2013. Accepted for publication Feb 19, 2013.

doi: 10.3978/j.issn.2218-6751.2013.02.05

Scan to your mobile device or view this article at: <http://www.tlcr.org/article/view/1029/1680>

The analysis of the mRNA expression in tumoral and non-tumoral samples is becoming increasingly important in the diagnosis and treatment of cancer patients. Until the end of the 1990s, Northern blot had been used extensively for RNA quantification (1,2). However, the discovery of the reverse transcriptase enzyme led to the development of reverse transcription polymerase chain reaction (RT-PCR) technique which has since displaced Northern blot as the method of choice for RNA detection and quantification (3).

This enzyme allows the transformation of the mRNA in cDNA which allows it to be quantified. In 2000, Bustin *et al.* (3) describes a new technology based on this concept. RT-PCR is used to qualitatively detect gene expression through the creation of complementary DNA (cDNA) transcripts from mRNA, and quantitatively measure the amplification of cDNA using fluorescent probes (3). The process of RT-PCR consists of three steps: reverse transcriptase-based conversion of RNA to cDNA, the amplification of cDNA by PCR, and the

detection and quantification of amplified products—referred as amplicons (4).

RT-PCR can be utilized to quantify mRNA in both relative (5) and absolute terms (6). It has become one of the most widely used methods of gene quantification as it has a large dynamic range, with high sensitivity and can be highly sequence-specific. The scope of the RT-PCR technique makes it applicable across a wide range of experimental conditions, and allows experimental comparisons between normal and abnormal tissues (4).

Until now, quantitative RT-PCR assay is considered to be the gold standard for measuring the number of copies of specific cDNA targets. The extensive use of this technique has resulted in the development of various protocols that enable the generation of quantitative data using fresh, frozen, or archived FFPE samples, whole-tissue biopsies, microdissected samples, single cells, tissue cultured cells, total or mRNA, different cDNA, assays of distinct efficiencies, sensitivities, and robustness, diverse detection chemistry, reaction conditions, thermal cyclers, individual analysis and reporting methods (7).

As a result, while there are numerous publications describing the technique, the inherent variability in the quality of any quantitative PCR data makes it difficult to replicate (8). For this reason, the international consortium of academic scientists has published the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines to standardize a protocol. The MIQE guidelines describe the minimum information required to evaluate RT-PCR experiments (9).

The application of RT-PCR has demonstrated that levels of RNA transcripts stratify patients and predict outcomes in a variety of diseases, providing the basis for several important clinical tests (10-12). An example is a 21-gene RT-PCR-based test, which interrogates tumor RNA to predict recurrence risk and magnitude of chemotherapy benefit in early estrogen receptor positive (ER+) breast cancer. This test is now used to guide treatment decisions for about half of ER+ breast cancer patients in the USA (13-16).

The problems with this technique are the time consumed in the laboratory for the analysis of one gene and the variability of the results. Moreover, comparing expression levels across different experiments is often difficult and can require complicated normalization methods.

To address the first problem, the scientific community developed new systems to analyze many genes at the same time. Several technologies are available in the market, such as Serial Analysis of Gene (SAGE), high-density

oligo microarrays gene expression (SAGE), cap analysis of gene expression (CAGE) and massively parallel signature sequencing (MPSS) (17-19). With these techniques it is possible to obtain the mRNA expression data for hundreds of genes in one experiment.

However, these new techniques have their own limitations. In the case of the methods based on hybridization, the main problems are: the limit in the detection of RNA splice patterns and previously unmapped genes (it is mandatory to know the genome sequence); the high background levels due to cross-hybridization; and the limited dynamic range of detection (20). In the techniques based on Sanger sequencing technology, which use cDNA sequencing, only a portion of the transcript is analyzed and isoforms are generally indistinguishable. These disadvantages limit the use of these methods in the translational investigation and clinical daily practice (21).

One simple and effective way to measure transcriptome composition and to discover new exons or genes is by the direct ultra-high-throughput sequencing of cDNA. This method, termed RNASeq, also called “Whole Transcriptome Shotgun Sequencing” (“WTSS”), has clear advantages over existing approaches and allows the mapping and quantification of transcriptomes. It is expected to revolutionize the field of transcriptome analysis (21).

RNA-seq refers to the use of high-throughput sequencing technologies to sequence cDNA in order to obtain information about RNA content (22). A single experiment can provide information about gene expression, novel transcripts, novel isoforms, alternative splice sites, allele-specific expression, cSNPs, and rare transcripts. Ideally, it should be able to directly identify and quantify all RNAs, small or large (23-26).

The sample analysis can be done with a variety of platforms, for example, the Illumina Genome Analyzer platform (27), the ABI Solid Sequencing (28), or the Life Science’s 454 Sequencing (29) (*Figure 1*).

RNA-Seq is not a mature technology. It is undergoing rapid evolution in the biochemistry of sample preparation; in sequencing platforms; in computational pipelines; and in the subsequent analysis methods that include statistical treatments and transcript model building. The ENCODE Consortium develops guidelines which include the most important aspects to bear in mind in study design (30).

The main steps of the technique are: isolate the RNAs from a sample, convert them to cDNA fragments using RT, then a high-throughput sequencer is used to generate millions of reads from the cDNA fragments, reads are

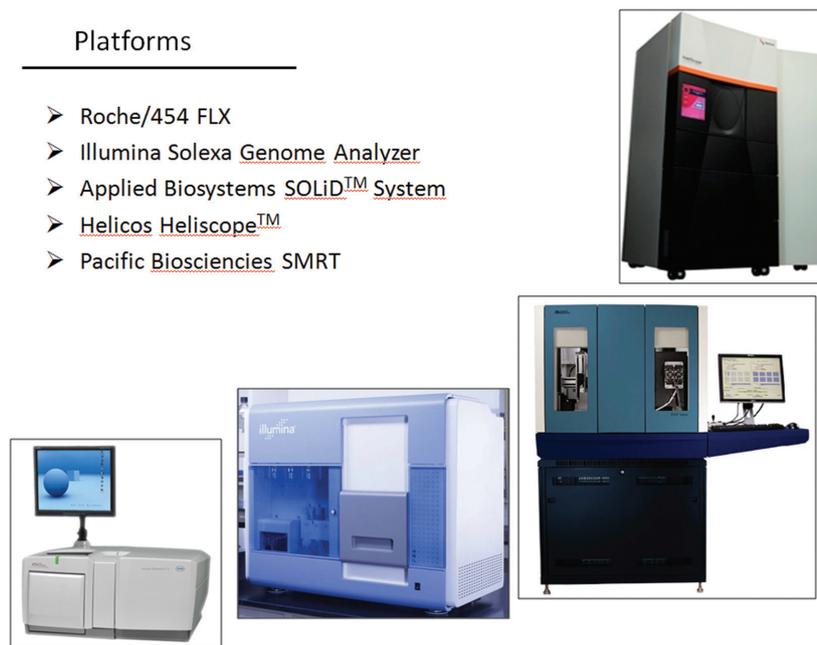


Figure 1 Different platforms for RNAseq analysis

mapped to a reference genome or transcript set with an alignment tool, and counts of reads mapped to each gene. For transcriptome sequencing, expression levels of different genes are determined by counting the number of reads mapped to the gene and then normalizing this read count by the length of the gene model and the total number of mapped reads in the sample (31).

In the case of RNA-Seq, the mapping of reads is only the first step in a complex data processing schema. However, it is essential that the results are analyzed by a specialized bioinformatician. This role requires an extensive background to be able to analyze the results of complex experiments with a large number of data sets. To address this need, software systems and guidelines have recently been developed for analyzing the data (32,33).

The primary advantages of this technique are high reproducibility, the large dynamic range, low background noise, requirement of less sample RNA, and the ability to detect novel transcripts, even in the absence of a sequenced genome (20).

Several recent examples of identification of novel mutations in tumor cell populations reveal the utility of RNA-Seq in disease classification. David Huntsman's group has applied RNA-Seq methodologies to identify mutations in gynecological tumors, novel mutations in FOXL2 in the previously ill-defined and treatment-

resistant granulosa cell tumor of the ovary, and in ARID1A in endometriosis-associated ovarian carcinomas (34,35). Diagnosis of granulosa cell tumors is difficult given the lack of knowledge of their pathogenesis and their relatively ambiguous histology, making a genetic variant associated with these tumors particularly valuable. Paired-end RNA-Seq analysis of several primary tumor samples identified mutations in the transcription factor FOXL2, specifically in tumor cell populations (34).

Another study used RNA-seq of cell lines derived from tumors to discover somatic mutations associated with human activated B-cell-like diffuse large B-cell lymphoma, identifying MYD88 as a candidate oncogenic mutation (36). That result was used to focus on studies of primary patient samples to rapidly identify mutations specific to this lymphoma subtype. These examples indicate the utility of RNA-Seq-based analysis of tumors to identify novel mutations.

For cancer expression profiling, RNA-seq has been applied to the expression profiling of melanoma. They identified 11 novel melanoma gene fusions and 12 novel read-through transcripts, providing an example of novel avenues for target discovery in cancers (37).

Another application of the RNA-Seq was the ability to detect transcripts from single cells [circulating tumor cells (CTCs)]. The characterization of these cells would

Table 1 Advantages and disadvantages of RT-PCR and RNAseq

| Technique | Advantages | Disadvantages |
|-----------|---|---|
| RT-PCR | <ul style="list-style-type: none"> ➤ High sensitivity ➤ High sequence-specific | <ul style="list-style-type: none"> ➤ Time consumed ➤ Analyze one gene each time ➤ Variability of the results depending the laboratory ➤ Necessary to know the gene sequence ➤ Require many amount of RNA |
| RNA-seq | <ul style="list-style-type: none"> ➤ A single experiment can provide information about all the genes (translocation, ...) ➤ High reproducibility ➤ Require low amount of RNA | <ul style="list-style-type: none"> ➤ Needs a bioinformatician for the analysis of the results ➤ High cost for one analysis |

provide a highly sensitive and effective method for potential diagnosis and monitoring of disease status (38).

As the cost of next-generation sequencing drops rapidly, RNA-Seq may replace current methods in transcriptome surveys of gene expression. Compared to the rest, RNA-Seq has several advantages, including the ability to simultaneously detect mutations, discovering alternative transcript and alternative splicing.

However, this RNA-Seq technique will not likely replace current RT-PCR methods, but will be complementary (Table 1). The results of the RNA-Seq will identify those genes that need to then be examined using RT-PCR methods, more so in those laboratories with resource constraints. The application of the two complementary technologies in the routine analysis of cancer laboratories would be useful in characterizing patients and would assist oncologists in making clinical decisions, as it allows us to identify all molecular characteristics of the tumor.

Acknowledgements

Disclosure: The authors declare no conflict of interest.

References

1. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* 1977;74:5350-4.
2. Streit S, Michalski CW, Erkan M, et al. Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nat Protoc* 2009;4:37-43.
3. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000;25:169-93.
4. Jozefczuk J, Adjaye J. Quantitative real-time PCR-based analysis of gene expression. *Methods Enzymol* 2011;500:99-109.
5. Kang XP, Jiang T, Li YQ, et al. A duplex real-time RT-PCR assay for detecting H5N1 avian influenza virus and pandemic H1N1 influenza virus. *Virology* 2010;7:113.
6. Bustin SA, Benes V, Nolan T, et al. Quantitative real-time RT-PCR--a perspective. *J Mol Endocrinol* 2005;34:597-601.
7. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nat Protoc* 2006;1:1559-82.
8. Bustin SA. Why the need for qPCR publication guidelines?--The case for MIQE. *Methods* 2010;50:217-26.
9. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 2009;55:611-22.
10. Rosell R, Molina MA, Costa C, et al. Pretreatment EGFR T790M mutation and BRCA1 mRNA expression in erlotinib-treated advanced non-small-cell lung cancer patients with EGFR mutations. *Clin Cancer Res* 2011;17:1160-8.
11. Soda M, Takada S, Takeuchi K, et al. A mouse model for EML4-ALK-positive lung cancer. *Proc Natl Acad Sci U S A* 2008;105:19893-7.
12. Akiyama T, Dass CR, Choong PF. Bim-targeted cancer therapy: a link between drug action and underlying molecular changes. *Mol Cancer Ther* 2009;8:3173-80.
13. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817-26.
14. Paik S, Tang G, Shak S, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;24:3726-34.

15. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res* 2006;8:R25.
16. Gianni L, Zambetti M, Clark K, et al. Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol* 2005;23:7265-77.
17. Velculescu VE, Zhang L, Vogelstein B, et al. Serial analysis of gene expression. *Science* 1995;270:484-7.
18. Kodzius R, Kojima M, Nishiyori H, et al. CAGE: cap analysis of gene expression. *Nat Methods* 2006;3:211-22.
19. Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 2000;18:630-4.
20. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57-63.
21. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;5:621-8.
22. Morin R, Bainbridge M, Fejes A, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 2008;45:81-94.
23. Shah SP, Köbel M, Senz J, et al. Mutation of FOXL2 in granulosa-cell tumors of the ovary. *N Engl J Med* 2009;360:2719-29.
24. Wiegand KC, Shah SP, Al-Agha OM, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* 2010;363:1532-43.
25. Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* 2011;470:115-9.
26. Pepke S, Wold B, Mortazavi A. Computation for ChIP-seq and RNA-seq studies. *Nat Methods* 2009;6:S22-32.
27. Mortazavi A, Williams BA, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008;5:621-8.
28. Cloonan N, Forrest AR, Kolle G, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 2008;5:613-9.
29. Barbazuk WB, Emrich SJ, Chen HD, et al. SNP discovery via 454 transcriptome sequencing. *Plant J* 2007;51:910-8.
30. Standards, Guidelines and Best Practices for RNA-Seq. V1.0. 2011. The ENCODE Consortium.
31. Li B, Ruotti V, Stewart RM, et al. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* 2010;26:493-500.
32. Goncalves A, Tikhonov A, Brazma A, et al. A pipeline for RNA-seq data processing and quality assessment. *Bioinformatics* 2011;27:867-9.
33. Shah SP, Köbel M, Senz J, et al. Mutation of FOXL2 in granulosa-cell tumors of the ovary. *N Engl J Med* 2009;360:2719-29.
34. Knowles DG, Röder M, Merkel A, et al. Grape RNA-Seq analysis pipeline environment. *Bioinformatics* 2013;29:614-21.
35. Wiegand KC, Shah SP, Al-Agha OM, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med* 2010;363:1532-43.
36. Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* 2011;470:115-9.
37. Berger MF, Levin JZ, Vijayendran K, et al. Integrative analysis of the melanoma transcriptome. *Genome Res* 2010;20:413-27.
38. Helzer KT, Barnes HE, Day L, et al. Circulating tumor cells are transcriptionally similar to the primary tumor in a murine prostate model. *Cancer Res* 2009;69:7860-6.

Cite this article as: Costa C, Giménez-Capitán A, Karachaliou N, Rosell R. Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl Lung Cancer Res* 2013;2(2):87-91. doi: 10.3978/j.issn.2218-6751.2013.02.05