Methods 59 (2013) 89–100

Contents lists available at SciVerse ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Review Article

Application of next generation qPCR and sequencing platforms to mRNA biomarker analysis

Alison S. Devonshire [a], Rebecca Sanders [a], Timothy M. Wilkes [a], Martin S. Taylor [b], Carole A. Foy [a], Jim F. Huggett [a,*]

[a] Molecular and Cell Biology, LGC Limited, Queens Road, Teddington, Middlesex TW11 0LY, UK
[b] MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine at the University Of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

ARTICLE INFO

Article history:
Available online 24 July 2012
Communicated by Michael W. Pfaffl

Keywords:
Next generation sequencing
NGS
Digital PCR
mRNA biomarker
Nanofluidic PCR
RNA-Seq

ABSTRACT

Recent years have seen the emergence of new high-throughput PCR and sequencing platforms with the potential to bring analysis of transcriptional biomarkers to a broader range of clinical applications and to provide increasing depth to our understanding of the transcriptome. We present an overview of how to process clinical samples for RNA biomarker analysis in terms of RNA extraction and mRNA enrichment, and guidelines for sample analysis by RT–qPCR and digital PCR using nanofluidic real-time PCR platforms. The options for quantitative gene expression profiling and whole transcriptome sequencing by next generation sequencing are reviewed alongside the bioinformatic considerations for these approaches. Considering the diverse technologies now available for transcriptome analysis, methods for standardising measurements between platforms will be paramount if their diagnostic impact is to be maximised. Therefore, the use of RNA standards and other reference materials is also discussed.

1. Introduction

Changes in the expression of multiple genes are implicated in complex diseases such as breast cancer, type 2 diabetes mellitus and cardiovascular disease [1,2]. *In vitro* diagnostic multi-variate index assays (IVDMIAs) utilising gene expression measurements, such as OncotypeDx tests which predict cancer recurrence, have emerged in recent years [3,4]. The pipeline for RNA biomarker panel development involves screening the transcriptome for genes whose expression is associated positively or negatively with disease pathology. Multiple stages of potential marker refinement are required in order to define the best predictors of clinical outcomes coupled with expanded patient cohorts. For example, in the development of Oncotype Dx Colon Cancer Assay, 761 gene candidates were narrowed down to a panel of seven biomarkers and five reference genes, with over 3,000 patient samples screened [5,6].

The DNA microarray is a well-established technique, which has been used to screen for multiple potential gene expression biomarkers and drug targets, and microarray gene expression data continues to be a useful source for mining of potential biomarkers. However, DNA microarrays utilise probes containing known cDNA sequences and therefore do not enable the discovery of novel transcripts and sequence variants [7]. Additionally, limitations in microarray dynamic range make this platform less sensitive in the detection of transcripts of low abundance [8]. Recent technological innovations in the fields of DNA sequencing and PCR address these issues and provide an unprecedented level of information for the discovery and validation of novel RNA biomarkers [9,10].

Next generation sequencing (NGS – also referred to as second generation sequencing) platforms share the common technological feature of being capable of massively parallel sequencing on clonally amplified or single cDNA molecules. This design defines a major shift from "first generation" Sanger sequencing, which was based on the electrophoretic separation of chain-termination products, prepared in individual sequencing reactions. NGS technologies offer the possibility of hypothesis-neutral discovery of novel transcripts and isoforms in a fraction of the time required for genome-wide analysis performed by Sanger sequencing [11,12]. However, multiple template preparation stages, diverse sequencing chemistries and complex data processing of NGS experiments may impact on the verification of *bona fide* RNA biomarkers [13] (Section 4).

Reverse Transcription quantitative PCR (RT–qPCR) technology is central to biomarker validation where potential markers need to be measured with greater accuracy and precision in larger sample sets. A new generation of nanofluidic qPCR platforms has also emerged over recent years which can be used for the simultaneous screening of patient samples for the expression of 10s-100s of candidate biomarkers or enumeration of single copies of cDNA by

* Corresponding author. Fax: +44 020 8943 2767.
E-mail address: jim.huggett@lgcgroup.com (J.F. Huggett).

1046-2023/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ymeth.2012.07.021

digital PCR (dPCR) (Section 3). Considering the current use of qPCR for molecular microbiological testing in the clinical laboratory, such high-throughput RT–qPCR devices are also likely to be at the forefront of transcript-based diagnostics in the near-future.

The translation of gene expression biomarkers from validated panel to diagnostic test requires assurance of the accuracy and robustness of the developed multi-parametric assay through the use of QC materials and establishment of QA schemes. Potential means of standardisation of multi-parametric RNA biomarker measurements through the use of reference standards are also addressed (Section 5).

This article aims to summarise how this next generation of PCR and sequencing platforms can be applied to different stages of RNA biomarker analysis while highlighting key methodological differences between the varying approaches.

2. RNA as an analyte

2.1. RNA extraction

In order to investigate messenger RNA (mRNA) expression and biomarker profiles, mRNA first needs to be successfully extracted from source material. The variety of biological samples available for molecular analyses has given rise to a multitude of extraction methods, which may confer particular advantage in terms of yield and integrity when utilised for specific sample types. Current approaches include acid phenol/chloroform, silica-column and bead-based extraction methods. Generally, total RNA will be prepared from sample extractions, the majority of which will comprise ribosomal RNA populations [14].

Formalin-fixed paraffin-embedded (FFPE) tissues provide a useful historical source of disease specimens for screening of potential biomarkers [15], however FFPE sections are challenging samples for RNA extraction, due to RNA degradation, cross-linking of RNA to proteins and modification of bases [16]. FFPE RNA extraction methods require deparaffinisation and extended lysis treatment at elevated temperatures; developments in automation of these steps offer potential for high-throughput screening of FFPE samples [17]. FFPE material is amenable to mRNA analysis using established methods like RT–qPCR and microarrays [18–20] and methods for 3'-end digital gene expression profiling by NGS have been developed [21]. Assuming the problems associated with RNA quality do not cause too great a challenge, FFPE samples will provide a valuable source of material for identifying mRNA biomarkers using next generation PCR and sequencing platforms.

While working with FFPE material offers a number of unique challenges, sample sourcing must also be considered when designing gene expression studies to investigate potential biomarkers from 'fresh' material. Some clinical sources such as tissue biopsies are difficult and intrusive to obtain, or may be particularly difficult from which to extract nucleic acid material (e.g. bone) [22–24]. This may lead to great variability in extraction efficiency (yield and quality), particularly when tissue-specific extraction methods are not employed. Consequently, less invasive yet easily handled sources of biological samples, such as blood, urine and buccal swabs, are a popular focus for the development of diagnostic tools.

It is similarly important to take into account tissue variability when planning to obtain samples. Gene expression profiles differ not only between different tissue types, but also between different cell types within the same sample. Furthermore, gene expression can be cyclical and may be influenced by many different genetic and environmental factors; including stress, satiety, nutrition, diurnal fluctuation, exercise, cellular proliferation, disease state and by mitogenic stimuli (e.g. growth factors) [25–31]. When conducting specific gene expression studies it is therefore important to ensure like-for-like samples are used in comparative studies and where possible, only the specific cell-type of interest is collected (for example, separating cell populations by centrifugation, using primary cell culture or laser microdissection) [14,32–35].

It is well known that RNA is more labile than DNA, and as such, precautions must be made in order to achieve the most reliable results. It is recommended that samples be collected in a buffer/preserving reagent suitable for safeguarding RNA against degradation or, alternatively, samples may be snap frozen using liquid nitrogen. The selection of an appropriate reagent may be heavily influenced by the intended down-stream applications (see Section 2.2). Moreover, specific RNA-handling procedures should be applied to reduce the risk of RNase activity. During collection of multiple samples, appropriate fixatives should be employed. Depending on storage buffers/fixatives, samples are usually stored at −20 °C or below until required, then thawed and maintained on ice during the extraction process. Where appropriate, purified RNA should be diluted in a solution designed to maintain RNA integrity, which is free of RNases.

2.2. Inhibition

Extracted RNA samples may be compromised due to the co-extraction of sample components (such as DNA, proteins, bile salts or haeme) or carry-over of chemicals used in sample stabilisation (such as EDTA or heparin) or extraction process (such as chloroform or ethanol) [14,36–39]. Every effort should be made to eliminate these constituents from the final RNA sample. DNA contamination may be further reduced by the application of DNase enzymes. However, no enzymatic reaction can be assumed 100% efficient and as such the presence of Genomic DNA (gDNA) should be monitored and accounted for, otherwise measurement bias may be introduced. Furthermore, if contaminating elements including PCR inhibitors are at reasonably low quantities in the extracted RNA, sample dilution may minimise or effectively eliminate their effect on target measurement.

For clinical application of mRNA biomarker-based diagnostics, thorough characterisation of assay performance should be performed [40] and standards for calibration and QC developed (Section 5). In this context, it is important to remember the influence of matrix effects when choosing an appropriate reference material. Where external standards may be applied for quantification purposes, these must be appropriate to the chosen target and analysed in background material that sufficiently mimics the sample matrix. Ideally, selected external standards should possess similar responses to matrix effects as experienced by the target, and must be spiked into target samples to ensure equal matrices [41].

Matrix-associated inhibition of qPCR may be detected by several different means. The simplest way is to measure samples in serial dilution and monitor linearity of amplification. Reversible inhibition, which may usually be observed at higher concentrations, will materialise as an increase in quantification cycle (C_q) and a decrease in correlation coefficient (R^2) when C_q is plotted against $\log_{10}$ RNA quantity. The SPUD assay has been developed to more accurately determine the extent of qPCR inhibition by measuring an external spike-in from potato (*Solanum tuberosum*) in control (water) vs. target cDNA samples. Analysis of C_q and assay efficiency between control and target samples for the SPUD assay indicates the extent of matrix inhibition [42].

Inhibition of the RT reaction is typically less readily quantified in the course of an RT–qPCR experiment, a factor that is of concern particularly when performing two-step RT–qPCR, where the RT reaction usually contains a higher concentration of both RNA and co-purified inhibitors. Defining the matrix impact on the RT step should be paramount as this reaction is a key component of both RT–qPCR and the majority of current RNA-seq methodologies.