Vaccine xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Vaccine

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

Focused human gene expression profiling using dual-color reverse transcriptase multiplex ligation-dependent probe amplification (dcRT-MLPA)

Mariëlle C. Haks[a], Jelle J. Goeman[b], Cecile Magis-Escurra[a,1], Tom H.M. Ottenhoff[a,*]

[a] Department of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands
[b] Biostatistics, Radboud Institute for Molecular Life Science, Radboud University, Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Available online xxx

Keywords:
dcRT-MLPA
Gene expression profiling
Host biomarkers

ABSTRACT

To investigate the human immune response to newly developed or existing vaccines, or during infection/disease on a population scale, we have recently developed a dual-color Reverse Transcriptase Multiplex Ligation-dependent Probe Amplification (dcRT-MLPA) assay, which can rapidly profile mRNA expression of multiple host genes. dcRT-MLPA has a dynamic range and sensitivity comparable to real-time QPCR and RNA-Sequencing. Since this assay is high-throughput, it is an exceptionally suitable technique for monitoring host biomarkers in semi-large scale human cohorts, such as cross sectional studies with multiple groups, or longitudinal studies with multiple time points. Multicomponent host biomarker signatures with excellent predictive values can easily be identified using lasso regression analysis, while exploring additional data adjustment methods like RUV-2 may further optimize the identification of informative host biomarker signatures. dcRT-MLPA also allows comparisons of gene expression patterns across different human populations to explore the impact of geographical diversity on for example vaccine induced responses. The use of dcRT-MLPA is not limited to peripheral blood but can be adapted to analyze host biomarkers derived from any tissue or body fluids, further demonstrating the versatility of the dcRT-MLPA platform. Several examples will be given and discussed.

1. Identification of host biomarkers by different gene expression profiling platforms

Biomarkers are defined as 'characteristics that are objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention' [1]. Host biomarkers can be powerful 'surrogate endpoints' to classify disease status, disease activity, disease progression, as well as prognosis and the effects of interventions (e.g. drugs, vaccines). Host biomarkers can be determined at the cellular-, protein-, metabolic, or transcript level, and can be analyzed in any tissue or body fluid, but peripheral blood is the most commonly used source in clinical practice.

Measurements of quantitative changes in RNA expression levels, to identify host biomarkers at the transcriptomic level, are currently primarily being analyzed using assays at the extreme ends of the RNA expression profiling spectrum: either genome-wide (microarray or RNA-Sequencing) or single gene (real-time QPCR) screening methods. Since genome-wide approaches are by definition hypothesis-free, they are the methods of choice to identify novel host biomarkers and biomarker signatures in small sample sets. However, because host biomarker signatures useful in classifying disease status, disease activity, disease progression, prognosis and the effects of interventions, will most likely encompass restricted sets of multiple gene transcripts, neither genome-wide nor single gene approaches will be ideally suited to achieve this. Furthermore, microarray and RNA-Sequencing assays are technically challenging, require substantial amounts of RNA, are costly and are complicated by complex data analysis, further limiting their applicability as a tool to routinely screen for host biomarkers in large human cohorts (Fig. 1).

To monitor and validate multicomponent host biomarker signatures on a population scale, techniques have been developed that combine sets of markers at the transcriptomic level. Fluidigm, for example, provides high-throughput real-time QPCR based on proprietary microfluidic chips [2]. These gene expression assays allow

* Corresponding author.
E-mail address: t.h.m.ottenhoff@lumc.nl (T.H.M. Ottenhoff).
[1] Present address: Department of Pulmonary Diseases, Radboud University Medical Centre Nijmegen, University Centre of Chronic Diseases Dekkerswald, Groesbeek, The Netherlands.

http://dx.doi.org/10.1016/j.vaccine.2015.04.054
0264-410X/

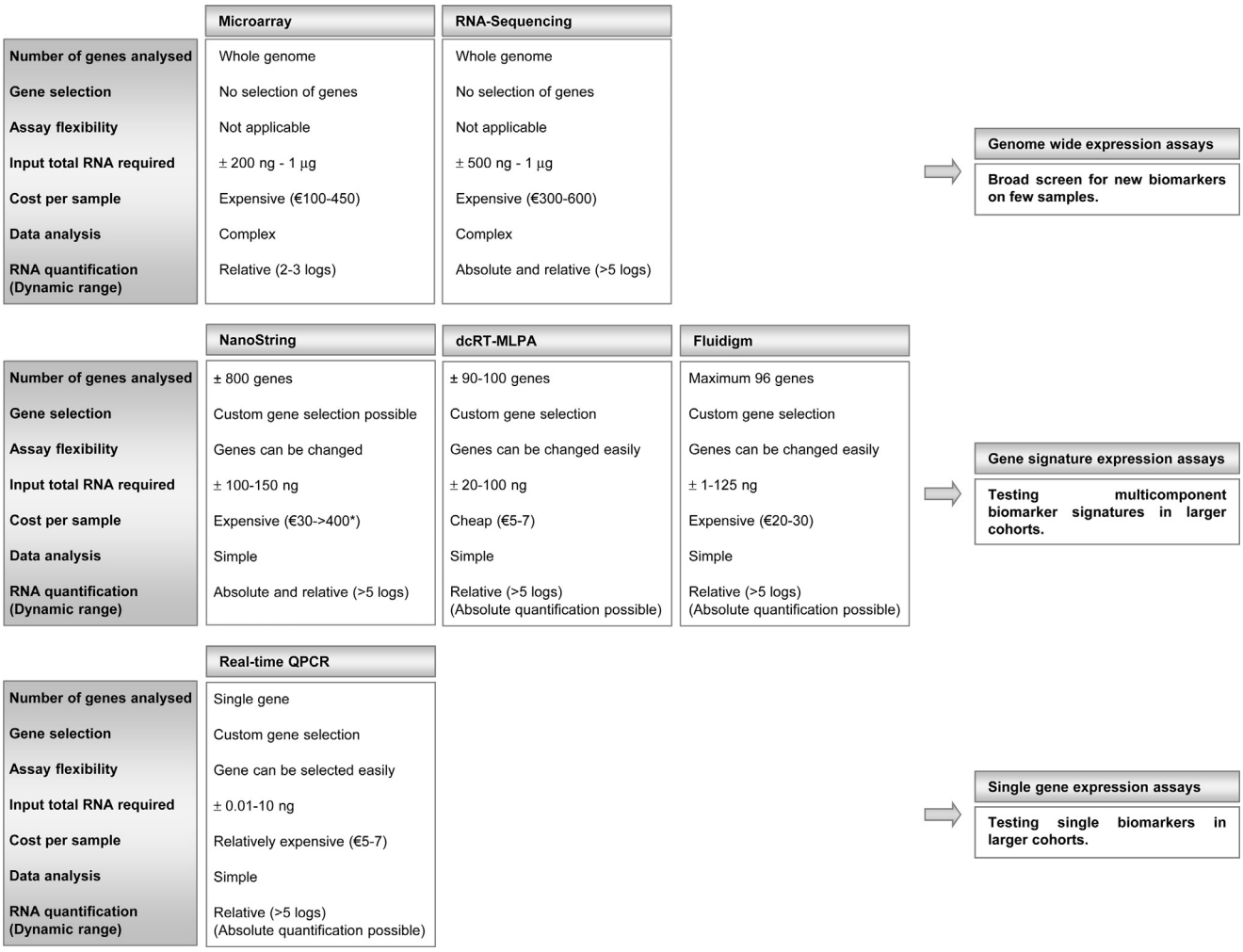

Fig. 1. *Spectrum of methods to determine gene expression profiles.* Features and characteristics of different methods used to quantify the abundance of RNA transcripts. (* Assay pricing is complicated and highly variable and depends on the number of assays ordered, the use of custom versus pre-designed gene panels, and the number of genes within a panel).

testing of either 48 genes in a 48-well format or 96 genes in a 96-well format. NanoString's nCounter gene expression assay is a novel technology using color-coded molecular barcodes and microscopic imaging to detect and count up to 800 unique transcripts in a single reaction. Although it does not require the conversion of mRNA to cDNA or amplification of the resulting cDNA, its sensitivity rivals PCR-based methods [3]. Both Fluidigm and nCounter are commercially available but are still costly and require specialized equipment not routinely available in many laboratories, especially not in resource-poor settings. Alternatively, dual-color reverse-transcriptase multiplex ligation-dependent probe amplification (dcRT-MLPA) is a robust technique that can rapidly and accurately profile RNA expression of as many as ~90–100 host genes in a single reaction [4]. As the assay is reliable, inexpensive, user friendly, high-throughput (96-well format), and only requires 100 ng RNA per sample, it is exceptionally suited to determine host biomarker signatures in larger sample sets, such as in population and/or longitudinal studies. Furthermore, genes of interest can be selected on a tailor-made basis and are interchangeable, adding flexibility to the selection of host biomarkers being analyzed (Fig. 1).

2. Principle of the dcRT-MLPA technique

dcRT-MLPA is based on the principle that for each target specific sequence, a specific RT primer is designed that is located immediately downstream of the probe target sequence. Following reverse transcription, two half-probes are hybridized directly adjacent to a target sequence and ligated together. DcRT-MLPA is designed such that all ligated products are amplified with the use of only two fluorophore-conjugated PCR primer pairs while the length of each amplification product is unique. The PCR amplification step ensures assay sensitivity, which is an essential prerequisite for the relative quantification of scarcely expressed genes. Amplified products are subsequently size-separated using capillary electrophoresis followed by analysis of trace data using the GeneMapper software package and exporting category tables quantifying the fluorescence-intensity of each amplification product (peak area) for further analysis (described in more detail below) (Fig. 2).

3. Comparing data generated with dcRT-MLPA to microarray, RNA-sequencing, and real-time QPCR

As both dcRT-MLPA and real-time QPCR are based on amplification of target products by PCR, dcRT-MLPA and real-time QPCR have very similar dynamic range (4–7 log) and sensitivity [4]. To directly compare gene expression profiles using dcRT-MLPA and real-time QPCR, the same RNA samples were profiled using both techniques. Data illustrated in Fig. 3A clearly show that the results obtained with these two RNA expression-profiling techniques are

Please cite this article in press as: Haks MC, et al. Focused human gene expression profiling using dual-color reverse transcriptase multiplex ligation-dependent probe amplification (dcRT-MLPA). Vaccine (2015), http://dx.doi.org/10.1016/j.vaccine.2015.04.054