



Do results obtained with RNA-sequencing require independent verification?

Measuring the expression of genes on a genome-wide scale has become an essential part of many biofilm studies. Historically this was done using microarrays (e.g. Refs. [1,2]), but currently 'next-generation RNA sequencing' (RNA-seq) (mostly using the Illumina sequencing technology) has become the method of choice for transcriptome studies (e.g. Refs. [3–5]). Besides these techniques that provide information about gene expression at the genome-scale level (i.e. quantify the expression level of all genes), other approaches can be used to measure the expression levels of a smaller subset of genes. This includes quantitative real-time PCR (qPCR) (e.g. Refs. [6,7]) and the construction of translational fusion reporters in which the gene coding for the transcript of interest is coupled to a reporter gene like eGFP (e.g. Ref. [8]) or *lacZ* (e.g. Ref. [9]). Historically the latter approaches (most often qPCR) have been used to confirm data obtained in large-scale transcriptomics studies, but whether this is necessary and/or provides an added value is not always clear. Authors, reviewers and editors often struggle with this question and the aim of this editorial is to provide a balanced overview of the issue and provide some guidance.

The main question in this debate comes down to: how reliable is RNA-seq to identify differentially expressed genes and to estimate how much their expression differs between different conditions? And is qPCR needed to validate such expression differences? The focus on validation of genome-scale expression studies likely stems from prior work with microarrays. While microarrays allowed to carry out gene expression studies on a scale not seen before, and despite their overall high level of performance, some concerns were raised about reproducibility and bias (e.g. Refs. [10,11]). Because of this, many researchers felt the need to validate microarray results with qPCR. However, RNA-seq does not suffer from the same issues as (some) microarrays did and there are a number of studies that have specifically addressed the correlation between results obtained with RNA-seq and qPCR. A comprehensive analysis was published by Everaert et al. [12], in which five RNA-seq analysis pipelines are compared to wet-lab qPCR results and this for >18.000 protein-coding genes. While this study is based on RNA samples from human origin, there is nothing that suggests the outcome of this study would be different for studies with microorganisms. One of the main conclusions from this study is that depending on the analysis workflow 15–20% of genes are considered as 'non-concordant' when results obtained with RNA-seq are compared to results obtained with qPCR (with 'non-concordant' defined as both approaches yielding differential expression in opposing directions, or one of the methods showing differential expression while the other does not). However, of the genes showing non-concordant results, 93% show a fold change lower than 2 and approx. 80% show a fold change lower than 1.5. In addition, of the non-concordant genes with a fold change > 2, the vast majority are

expressed at very low levels. Overall, the conclusion was that there appears to be a very small fraction (approx. 1.8%) of genes that are severely non-concordant, and these genes are typically lower expressed and shorter. Examples of other studies that show good correlations between results obtained with qPCR and with RNA-seq include [13–16]. A more general reflection on the value of validation in genome-scale studies can be found in Ref. [17].

A second important aspect in this discussion is feasibility. It is not *a priori* known for which genes RNA-seq potentially yields non-concordant results in a particular study set up and as such it could be suggested to determine expression levels of *all* genes with qPCR or, alternatively, randomly select some genes for follow-up with qPCR. The former option is obviously not realistic in terms of cost and workload (and defeats the purpose of doing RNA-seq in the first place). The latter option could be an alternative, but how many genes need to be confirmed with another approach? As some genes are concordant and others are non-concordant, obtaining concordant results for a random selection of genes is no guarantee that other genes have been correctly identified as differentially expressed by RNA-seq and seems unlikely to provide much added value in most cases.

If all experimental steps and data analyses are carried out according to the state-of-the-art, results from RNA-seq are expected to be reliable and if they are based on a sufficient number of biological replicates, the added value of validating them with qPCR (or any other approach) is likely to be low. However, the situation is different when an entire story is based on differential expression of only a few genes, especially if expression levels of these genes are low and/or differences in expression are small. In such a case, orthogonal method validation (e.g. by qPCR or reporter fusions) seems appropriate, as one wants to make sure that the observed differences in expression for these genes on which the story is based are real and can be independently verified. In addition, qPCR would be valuable to measure expression of selected genes in *additional* samples. E.g. when RNA-seq identifies differential expression of gene X in a particular strain and/or condition, qPCR could be used to confirm this differential expression in additional strains and/or conditions.

While not the main topic of this editorial, I would like to point out that it is important to follow the minimum information guidelines that have been developed for different techniques and biological experiments; an overview of these can be found at <https://fairsharing.org/collectio n/MIBBI>. Of particular relevance in this context are the MIQE guidelines for qPCR (<https://fairsharing.org/FAIRsharing.mxz4jy>) [18] and the MINSEQE guidelines for high-throughput sequencing (<https://fairsh aring.org/FAIRsharing.a55z32>). In addition, it is worth emphasizing that also for biofilm experiments such minimal guidelines are available (MIABIE, <https://fairsharing.org/FAIRsharing.6mk8xz>) [19] and that

<https://doi.org/10.1016/j.biofilm.2021.100043>

Received 7 January 2021; Accepted 7 January 2021

2590-2075/© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

there is a specific minimum information guideline for biofilm experiments in microtiter plates [20].

In conclusion, the data available suggest that RNA-seq methods and data analysis approaches are robust enough to not always require validation by qPCR and/or other approaches, although there are situations where this may be of added value. While this editorial by no means presents a comprehensive overview of this topic, the hope is that it will provide some guidance to scientists struggling with the question whether RNA-seq data obtained in biofilm studies need independent verification.

Acknowledgement

The idea for this editorial came after a tweet from Allan McNally (Birmingham, UK). I would like to thank Andrea Sass and Jo Vandesompele (both from Gent, Belgium), and my co-senior editors (Akos Kovacs, Birthe Kjellerup and Darla Goeres) for helpful discussions.

References

- [1] An D, Parsek MR. The promise and peril of transcriptional profiling in biofilm communities. *Curr Opin Microbiol* 2007;10(3):292–6.
- [2] Lazazzera BA. Lessons from DNA microarray analysis: the gene expression profile of biofilms. *Curr Opin Microbiol* 2005;8(2):222–7.
- [3] Cornforth DM, et al. *Pseudomonas aeruginosa* transcriptome during human infection. In: *Proceedings of the national academy of sciences of the United States of America*; 2018. p. 115 (22): p. E5125–E5134.
- [4] Wille J, et al. Does the mode of dispersion determine the properties of dispersed *Pseudomonas aeruginosa* biofilm cells? *Int J Antimicrob Agents* 2020;56(6): 106194.
- [5] Valli RXE, Lyng M, Kirkpatrick CL. There is no hiding if you Seq: recent breakthroughs in *Pseudomonas aeruginosa* research revealed by genomic and transcriptomic next-generation sequencing. *Journal of medical microbiology* 2020; 69(2):162–75.
- [6] Suzuki N, Yoshida A, Nakano Y. Quantitative analysis of multi-species oral biofilms by TaqMan Real-Time PCR. *Clin Med Res* 2005;3(3):176–85.
- [7] Nailis H, et al. Real-time PCR expression profiling of genes encoding potential virulence factors in *Candida albicans* biofilms: identification of model-dependent and -independent gene expression. *BMC Microbiol* 2010;10:114.
- [8] Van Acker H, et al. The role of small proteins in *Burkholderia cenocepacia* J2315 biofilm formation, persistence and intracellular growth. *Biofilms* 2019;1:100001.
- [9] Irigul-Sonmez O, et al. In *Bacillus subtilis* LutR is part of the global complex regulatory network governing the adaptation to the transition from exponential growth to stationary phase. *Microbiology (Reading, England)* 2014;160(Pt 2):243–60.
- [10] Zhang L, Yoder SJ, Enkemann SA. Identical probes on different high-density oligonucleotide microarrays can produce different measurements of gene expression. *BMC Genom* 2006;7:153.
- [11] Balazsi G, Oltvai ZN. A pitfall in series of microarrays: the position of probes affects the cross-correlation of gene expression profiles. *Methods Mol Biol* 2007;377: 153–62.
- [12] Everaert C, et al. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci Rep* 2017;7(1):1559.
- [13] Shi Y, He M. Differential gene expression identified by RNA-Seq and qPCR in two sizes of pearl oyster (*Pinctada fucata*). *Gene* 2014;538(2):313–22.
- [14] Wu AR, et al. Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods* 2014;11(1):41–6.
- [15] Griffith M, et al. Alternative expression analysis by RNA sequencing. *Nat Methods* 2010;7(10):843–7.
- [16] Asmann YW, et al. 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genom* 2009;10: 531.
- [17] Hughes TR. 'Validation' in genome-scale research. *J Biol* 2009;8(1):3.
- [18] Bustin SA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 2009;55(4):611–22.
- [19] Lourenco A, et al. Minimum information about a biofilm experiment (MIABiE): standards for reporting experiments and data on sessile microbial communities living at interfaces. *Pathogens Dis* 2014;70(3):250–6.
- [20] Allkja J, et al. Minimum information guideline for spectrophotometric and fluorometric methods to assess biofilm formation in microplates. *Biofilms* 2020:2.

Tom Coenye

Laboratory of Pharmaceutical Microbiology, Ghent University, Ghent,
Belgium

E-mail address: tom.coenye@ugent.be.