

CORRESPONDENCE

Open Access



# Omission of non-poly(A) viral transcripts from the tissue level atlas of the healthy human virome

Signe Altmäe<sup>1,2,3\*</sup> , Nerea M. Molina<sup>1,2</sup> and Alberto Sola-Leyva<sup>1,2</sup>

## Abstract

A recent paper in *BMC Biology* entitled “A tissue level atlas of the healthy human virome” by Kumata et al. describes a meta-transcriptomic analysis of RNA-sequencing datasets from the Genotype-Tissue Expression (GTEx) Project. Using a workflow that maps the GTEx sequences to the human genome, then screens unmapped sequences to detect viral transcripts, the authors present a quantitative analysis of the presence of different viruses in the non-diseased tissues of over 500 individuals and assess the impact of these viruses on host gene expression. Here we draw attention to an issue not acknowledged in this study. Namely, by relying solely on GTEx datasets, which are enriched for transcripts with poly(A) tails, the analysis will have missed non-poly(A) viral transcripts, rendering this tissue level atlas of the virome incomplete.

A commentary on Kumata et al. (*BMC Biol* 18:55, 2020).

**Keywords:** GTEx, Poly(A) tail, RNA-seq, Virus, Virome

Viruses are obligate parasites and require a living cell to complete their life cycles. Like mRNAs in the eukaryotic host cell, RNAs of many DNA and RNA viruses generate polyadenylated transcripts (i.e., transcripts containing 3' poly(A) tails) that are synthesized post-transcriptionally [1], and in some RNA viruses also by direct transcription from poly(U) sequence on the stretched template strand [2, 3]. The viral poly(A) tails are important for regulating RNA stability and translation initiation, mimicking roles of the stable poly(A) tails in eukaryotic mRNA [4].

Many viruses, however, generate transcripts without poly(A) tails, a feature that has been maintained over

evolution, especially in positive-strand RNA viruses as for instance are dengue virus, West Nile virus, Japanese encephalitis virus, yellow fever virus, Zika virus, bovine viral diarrhoea virus, and hepatitis C virus in the *Flaviviridae* family [4–6]. Other important examples of non-poly(A) viral RNA transcripts are adenovirus-encoded non-coding RNA viral-associated RNAs and herpesvirus EBV-encoded non-coding small RNAs (EBERs) (the gold standard clinic markers for detection of EBV latent infection in specimens) [7]. Viral-encoded non-poly(A) RNAs have an important role in different physiological conditions and illnesses, including viral life cycle and function, and host cell immune evasion and transformation [8].

Next-generation sequencing offers high sensitivity, specificity, and reproducibility in detection of low levels of transcripts thereby serving as a sensitive and reliable tool to qualify and quantify viruses at DNA and RNA levels [9]. Nevertheless, depending on the exact sequencing protocol of choice, the non-polyadenylated viral RNA sequences could be detected or discarded (Fig. 1).

This comment refers to the article available at <https://doi.org/10.1186/s12915-020-00785-5>.

\* Correspondence: [signealtmae@ugr.es](mailto:signealtmae@ugr.es);

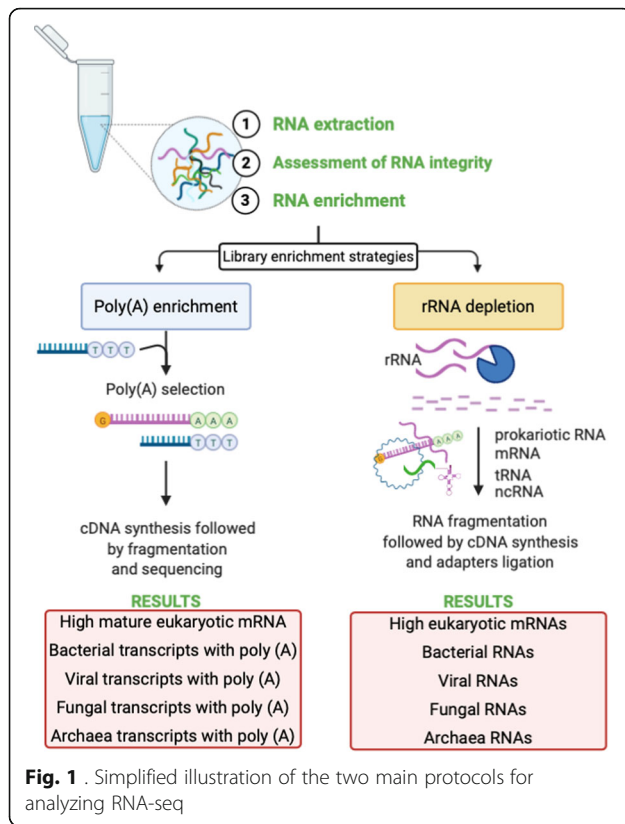
<sup>1</sup>Department of Biochemistry and Molecular Biology, Faculty of Sciences, University of Granada, 18071 Granada, Spain

<sup>2</sup>Instituto de Investigación Biosanitaria ibs.GRANADA, 18014 Granada, Spain

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



The recent *BMC Biology* article by Kumata et al. presented the first tissue level atlas of the human virome by analyzing the RNA-seq data from the GTEx database [10]. GTEx uses oligo (dT) primers for obtaining poly(A)-enriched fraction in the initial RNA purification step, meaning that only the RNA transcripts with

poly(A) tail will be enriched and sequenced [11]. We believe that Kumata et al. study has overlooked this important aspect, and although the first comprehensive investigation of the human virome in somatic tissues was presented, an important part of the human virome was not detected. A recently published study comparing poly(A)-enriched RNA-seq and non-poly(A)-selected RNA-seq in the lung virome analysis from the same samples supports our concern, as in this study it was demonstrated clearly that poly(A)-enriched RNA-seq failed to detect several viruses [7]. Furthermore, Kumata et al. conclude that mainly DNA viruses shape the healthy human virome as most of the detected viruses in their study were DNA viruses, although they acknowledge the possibility that the detection sensitivity of RNA viruses could have been lower [10]. Indeed, especially RNA viruses lack poly(A) tail [5, 6], which could be one solid explanation why RNA viruses were under-detected and DNA viruses predominated in the study by Kumata et al.

Before other researchers are motivated to apply their meta-transcriptomic study approach [10] to other datasets with the aim of revealing the impact of viral infections on human health, we would like to highlight that the choice of sequencing protocol is crucial in obtaining and interpreting the study findings. In short, the recently presented tissue level atlas of the healthy human virome should be acknowledged as a partial tissue level atlas, and the comprehensive investigation should be completed with meta-transcriptome analysis of data generated using the total RNA extraction method in order to achieve a more complete view of the human virome.

## Response to “Omission of non-poly(A) viral transcripts from the tissue level atlas of the healthy human virome”

Ryuichi Kumata, Jumpei Ito and Kei Sato\*

Division of Systems Virology, Department of Infectious Disease Control, International Research Center for Infectious Diseases, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 1088639, Japan

Correspondence: ksato@ims.u-tokyo.ac.jp

In our recent paper in *BMC Biology* titled “A tissue level atlas of the healthy human virome” [10], we performed meta-transcriptomic analysis using the RNA-sequencing (RNA-Seq) dataset from the Genotype-Tissue Expression (GTEx) Project, which includes 8991 RNA-Seq data obtained from 51 somatic tissues from 547 individuals. In this study, we detected 39 viral species. Omission of

non-poly(A) viral transcripts from the tissue in at least one tissue and furthermore investigated associations between virus infection (e.g., hepatitis C virus and some human herpes viruses) and human gene expression [e.g., type I interferons (IFNs) and IFN-stimulated genes].

As described in the first sentence of the “Method” section of our paper [10], we used the pair-ended, poly(A)-

enriched RNA-Seq data provided by GTEx (version 7.p2). Altmäe et al. are correct to point out that because the dataset we used lacks the information of non-poly(A) RNAs, our study does not include the data of non-poly(A) viral transcripts. In this regard, some RNA viruses (e.g., Flaviviruses) potentially produce non-poly(A) RNAs. Therefore, in terms of the description of the human virome, particularly that of human RNA viruses, we cannot exclude the possibility that the results shown in our recent paper [10] are incomplete. If there were large datasets available that included the data of non-poly(A) RNAs, it would be possible to survey the bona fide human virome more efficiently and effectively, as suggested by Altmäe et al.

Some aspects of the technical limitations of our study were described in the “Discussion” and “Conclusion” sections of our study [10], and it was an oversight not to have also acknowledged the limitation of the available dataset. The GTEx project does not provide the dataset of non-poly(A) RNA-Seq and is not just specified for the investigation of the human virome. However, to our knowledge, the GTEx project provides public access to the biggest set of transcriptome data for non-diseased human tissues (i.e., 8991 RNA-Seq data obtained from 51 somatic tissues from 547 individuals) that exists to date. Using this dataset, we quantified viral “mRNA” transcripts [i.e., poly(A)-added viral RNAs], and to the best of our knowledge, this is the first and biggest investigation addressing the presence of viruses in a variety of human tissues. We recognize, however, that our meta-analysis shows the human tissue virome based on the data of viral “mRNA” transcripts and does not include non-poly(A)-added RNAs. The technology of next-generation sequence data analysis progresses rapidly, and the publicly available datasets are increasing day by day. The advanced investigation of the human virome using non-poly(A) RNA-Seq data is an intriguing prospect, and we look forward to seeing the results of such an endeavor in the future.

#### Abbreviations

EBERs: EBV-encoded non-coding small RNAs; EBV: Epstein-Barr virus; mRNA: Messenger RNA

#### Acknowledgements

Not applicable.

#### Authors' contributions

Signe Altmäe, Nerea M. Molina, and Alberto Sola-Leyva drafted, revised, and approved this correspondence.

#### Authors' information

Twitter:

Signe Altmäe, @SigneAltmae  
Nerea M. Molina, @ner3wis  
Alberto Sola-Leyva, @sola\_leyva

#### Funding

This work is supported by the Spanish Ministry of Economy, Industry and Competitiveness (MINECO) and European Regional Development Fund (FEDER): grants RYC-2016-21199 and ENDOR SAF2017-87526-R; FEDER/Junta de

Andalucía-Consejería de Economía y Conocimiento: MENDO (B-CTS-500-UGR18) and by the University of Granada Plan Propio de Investigación 2016 - Excellence actions: Unit of Excellence on Exercise and Health (UCEES) (SOMM17/6107/UGR). AS-L and NMM are funded by the Spanish Ministry of Science, Innovation and Universities: PRE2018-0854409 (AS-L) and FPU19/01638 (NMM). This work is part of a PhD thesis conducted in the Biomedicine Doctoral Studies of the University of Granada, Spain.

#### Availability of data and materials

Not applicable.

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Biochemistry and Molecular Biology, Faculty of Sciences, University of Granada, 18071 Granada, Spain. <sup>2</sup>Instituto de Investigación Biosanitaria ibs.GRANADA, 18014 Granada, Spain. <sup>3</sup>Competence Centre on Health Technologies, 50411 Tartu, Estonia.

Received: 17 August 2020 Accepted: 22 October 2020

Published online: 25 November 2020

#### References

- Carter J, Saunders V. *Virology: principles and applications*; 2013.
- Geng G, Yu C, Li X, Yuan X. Variable 3' polyadenylation of Wheat yellow mosaic virus and its novel effects on translation and replication. *Virology*. 2019;16:23.
- Cross ST, Michalski D, Miller MR, Wilusz J. RNA regulatory processes in RNA virus biology. *Wiley Interdiscip Rev RNA*. 2019;10:e1536.
- Barr JN, Fearn R. How RNA viruses maintain their genome integrity. *J Gen Virol*. 2010;91:1373–87.
- He M, Jiang Z, Li S, He P. Presence of poly(a) tails at the 3'-termini of some mRNAs of a double-stranded RNA virus, southern rice black-streaked dwarf virus. *Viruses*. 2015;7:1642–50.
- Gomila RC, Martin GW, Gehrke L. NF90 binds the dengue virus RNA 3' terminus and is a positive regulator of dengue virus replication. *Plos One*. 2011;6:e16687.
- Yin Q, Strong MJ, Zhuang Y, Flemington EK, Kaminski N, de Andrade JA, et al. Assessment of viral RNA in idiopathic pulmonary fibrosis using RNA-seq. *BMC Pulm Med*. 2020;20:81.
- Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, et al. Viral noncoding RNAs: more surprises. *Genes Dev*. 2015;29:567–84.
- Noell K, Kolls JK. Further defining the human virome using NGS: identification of Redondoviridae. *Cell Host Microbe*. 2019;25:634–5.
- Kumata R, Ito J, Takahashi K, Suzuki T, Sato K. A tissue level atlas of the healthy human virome. *BMC Biol*. 2020;18:55.
- Consortium Gte. The GTEx Project. <https://www.gtexportal.org/home/documentationPage#staticTextDataProduction>. Accessed 20 July 2020.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.