

LETTERS

Raman spectroscopy-based detection of RNA viruses in saliva: A preliminary report

Sanket Desai^{1,5} | Saket V. Mishra^{2,5} | Asim Joshi^{1,5} | Debashmita Sarkar^{2,5} |
Arti Hole³ | Rohit Mishra^{1,5} | Shilpee Dutt^{2,5} | Murali K. Chilakapati^{3,5} |
Sudeep Gupta^{4,5} | Amit Dutt^{1,5,6*} 

¹Integrated Cancer Genomics Laboratory, Navi Mumbai, India

²Shilpee Dutt Laboratory, Navi Mumbai, India

³Chilkapati Laboratory, Navi Mumbai, India

⁴Department of Medical Oncology, Advanced Centre for Treatment, Research, and Education in Cancer, Navi Mumbai, India

⁵Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai, India

⁶Adjunct Faculty, Institute of Advanced Virology, Kerala State Council for Science, Technology and Environment, Govt of Kerala, Thonnakkal, Kerala, 695317, India

***Correspondence**

Dr Amit Dutt, Integrated Cancer Genomics Laboratory, Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Maharashtra, Navi Mumbai 410210, India. Email: adutt@actrec.gov.in

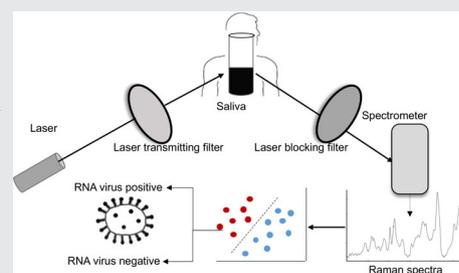
Abstract

Several non-invasive Raman spectroscopy-based assays have been reported for rapid and sensitive detection of pathogens. We developed a novel statistical model for the detection of RNA viruses in saliva,

based on an unbiased selection of a set of 65 Raman spectral features that mostly attribute to the RNA moieties, with a prediction accuracy of 91.6% (92.5% sensitivity and 88.8% specificity). Furthermore, to minimize variability and automate the downstream analysis of the Raman spectra, we developed a GUI-based analytical tool “RNA Virus Detector (RVD).” This conceptual framework to detect RNA viruses in saliva could form the basis for field application of Raman Spectroscopy in managing viral outbreaks, such as the ongoing COVID-19 pandemic. (<http://www.actrec.gov.in/pi-webpages/AmitDutt/RVD/RVD.html>).

KEYWORDS

GUI-based automated computational analysis, linear discriminant analysis, principal component analysis, Raman spectroscopy, RNA virus



1 | INTRODUCTION

Early identification of SARS-CoV-2 in infected patients is central to controlling COVID-19 pandemic with no available treatment option or vaccine [1]. The current trend for prescribing diagnostic assays for those expressing salient symptoms of the disease may be suboptimal in impeding the transmission rate [2, 3]. Raman

spectroscopy, based on the principle of inelastic scattering of light [4], is known to detect HIV, dengue, and various biomarkers and could potentially address practical concerns related to mass testing as a point-of-care device [5–8]. A recent study described a comparable sensitivity of SARS-CoV-2 detection in the saliva of an asymptomatic COVID-19 patient [9], consistent with an accelerated emergency use authorization approval to a saliva-based diagnostic kit by the US FDA [10]. Saliva, thus, offers an attractive, quick, non-invasive alternative to nasopharyngeal swabs for SARS-CoV-2 detection.

Sanket Desai, Saket V. Mishra and Asim Joshi contributed equally to this study.

2 | EXPERIMENTAL

We set to record Raman spectra to identify the presence of RNA virions, generated using pLL 3.7 lentivirus vector system (based on HIV-1), with 7.05×10^7 TU/mL stock titer spiked in water and human saliva at 1:10 and 1:1000 dilutions. The infectivity of the viral particles was simultaneously confirmed by the observation of GFP expression in HEK293FT cells (Figure S1). Reported spectral data from saliva of individuals not known to have viral infection was used as a negative control [11].

3 | RESULTS AND DISCUSSION

A set of 15 spectra were recorded from each sample, generating a total of 201 spectra, including 131 from viral negative samples, 54 from viral positive samples spiked with

lentiviral RNA particles and 16 viral positive samples with enzymatic treatments (Table S1). Pre-processed spectra post-base-line correction and area normalization interpolated in 600 to 1800 cm^{-1} region were used as inputs for downstream analysis. Principal component analysis (PCA) of all spectra revealed homogenous clustering of the two groups, with PC1 and PC2 accounting for 76.6% variance of the total sample set (Figure S2). Next, an unbiased prediction model was generated using linear discriminant analysis (LDA) based on spectral features with an SD of greater than 0.404 (marking the top quantile—308 features of 1200), and samples divided into training ($n = 149$) and test ($n = 36$) sets, showed a prediction accuracy of 88%.

To obtain optimal number of features and refine the model, we used varying number top features from the principal components (PC1 and PC2) to build the model and computed prediction accuracy for each of them. Top 10, 25, 50, 75 and 100 features were iteratively used to build the

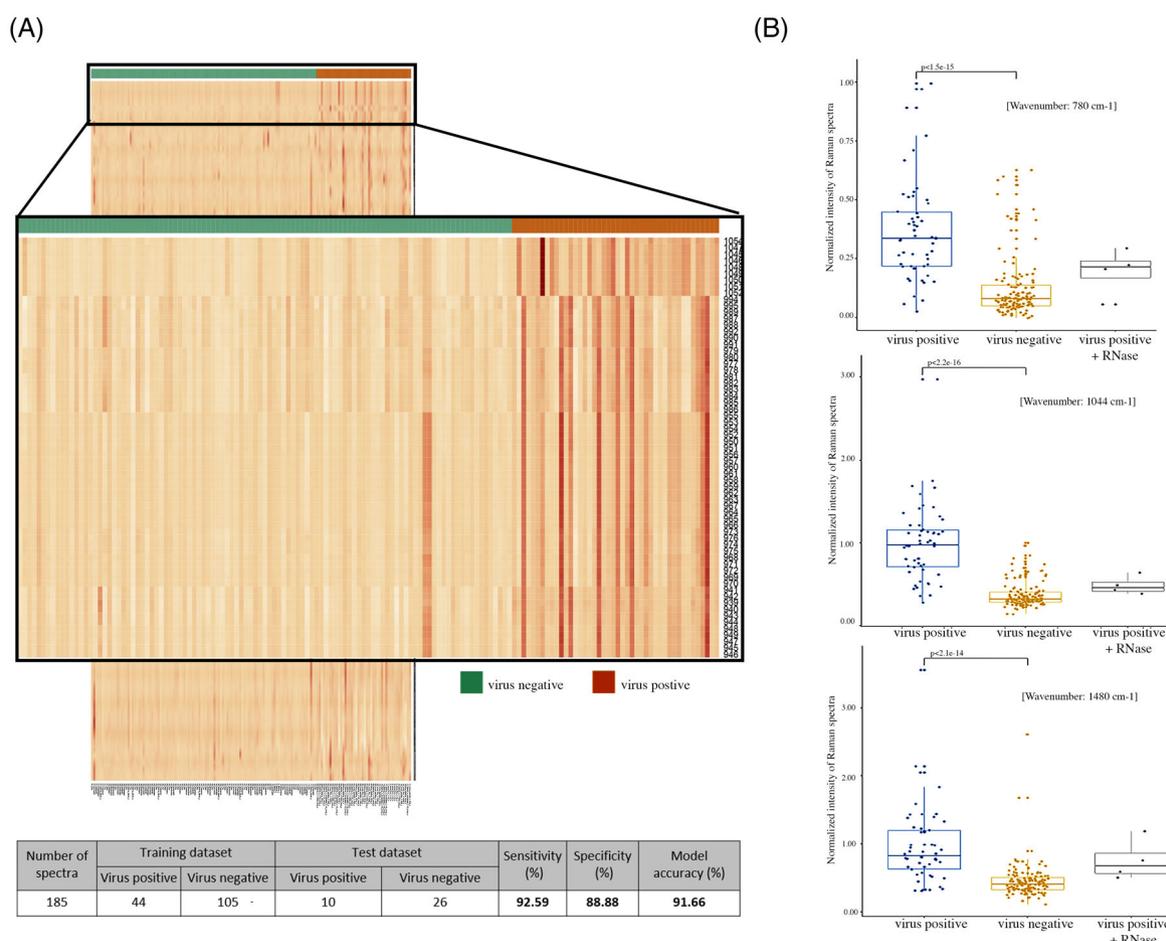


FIGURE 1 A Raman spectroscopy-based model to detect RNA viruses. A, Heatmap representation of the 1200 features from the virus positive and negative training sample set of Raman spectra with the zoomed-out version showing the peak intensities for the features selected ($n = 65$) for model construction. The table shows the overall sample statistics and summary of the performance of the 65-feature linear discriminant analysis-based model for prediction of viral presence in the human saliva samples. B, Comparison of Raman spectra peaks between virus positive, virus negative and RNase-treated viral positive sample for RNA-specific constituents: the nitrogenous uracil base, ribose-phosphate and A/G ring at 780, 1044 and 1480 cm^{-1} , respectively by Wilcoxon rank-sum test

model, wherein we observed that the model based on top 50 features from PC1 and PC2 (unique 65), achieved the highest accuracy (Figure S3). These 65 features (within the range of 939-1054 cm^{-1}) were adequate to effectively discern the viral positive and negative samples, as represented in the form of heatmap, increasing the model prediction accuracy to 91.6% (92.5% sensitivity and 88.8% specificity) (Figure 1A and Figure S2), which may be subject to a minor variability due to saliva-specific spectral interferences.

Intriguingly, a majority of bands in this range are attributable to RNA moieties, especially the ribose sugars in the 900-1045 cm^{-1} range with a distinct peak at 1045 cm^{-1} and the phosphate group in 1080 to 1100 cm^{-1} range [12, 13]. Of note, biomolecules, such as DNA, RNA, proteins, lipids, from intact viruses are known to display unique Raman spectra vibrations based on their structural heterogeneity [14]. Thus, to ask if individual RNA moieties could determine the presence of RNA viruses, we performed an unbiased statistical analysis using Wilcoxon rank-sum test for RNA specific constituent, such as for the nitrogenous uracil base, ribose-phosphate and A/G ring at 780, 1044 and 1480 cm^{-1} , respectively [12, 13]. Interestingly, all the RNA specific distinct peaks could.

individually discern the viral positive and negative samples with high statistical confidence that was significantly reduced following RNase treatment (Figure 1B), but not with DNase or Proteinase treatment of the viral positive samples (Figure S4). Taken together, these results suggest a novel concept of using 65-feature spectra (majorly attributable to RNA moieties) obtained from intact virion to form the basis for the detection of RNA viruses in human sputum.

To minimize variability and automate the downstream 65-feature-based analysis of the Raman spectra we further developed a GUI-based analytical tool RNA Virus Detector, RVD (<http://www.actrec.gov.in/pi-webpages/AmitDutt/RVD/RVD.html>), an in silico automated tool to detect RNA virus from an individual or a group of samples in an unambiguous and reproducible manner using the raw SPC files recorded by Raman spectrometer as an input for analysis, with minimal requirement of third-party tools. The RVD accuracy based on the statistical model developed, however, may be limited by the spectral acquisition conditions such as excitation wavelength the laser power and the spectrograph employed, that may necessitate further improvisation for universal application.

4 | MATERIALS AND METHODS

4.1 | Sample collection

One sample of approximately 1 mL of unstimulated saliva of the principal investigator of this study was

collected in sterile micro centrifuge to be used as a medium for studying the viral spectra spiked in the saliva for this study. The tube and was immediately refrigerated at 4°C. The samples were then transferred to -20°C till further use.

4.2 | Transfection

1×10^6 HEK293FT cells were seeded per well on a 6-well plate. The following day, transfection of GFP-tagged lentiviral vector, pLL3.7 along with the packaging plasmids pPAX and pVSVG in 2:1:0.5 ratio using lipofectamine 3000 was performed according to manufacturer's instructions. The viral supernatant was collected 48- and 72-hours post-transfection, filtered using 0.45- μ m filter and subjected to ultracentrifugation at 30 000 rpm for 90 minutes at 4°C. The pellet was resuspended in 1 mL serum-free Dulbecco's Modified Eagle's medium (DMEM) media [15].

4.3 | Transduction

1.5×10^6 HEK293FT cells were seeded in 60 mm tissue culture petri-dish. The next day, transduction of the viral particles were performed using two dilutions (1:10; 1:1000) in plain DMEM using 8 $\mu\text{g}/\text{mL}$ polybrene in HEK293FT cells. GFP-positive cells were observed 48 hours post-transduction [15].

4.4 | Sample preparation for Raman spectroscopy

Concentrate viral particle stock of 7.05×10^7 TU/mL were diluted in 1:10 and 1:1000 ratios using two different diluents- sterile ultrapure water and human saliva to mimic the patient condition. Plain media in ultrapure water and saliva without viral particles were taken as negative controls. Besides, to check the specific biomolecule (DNA, RNA or proteins) which help in differentiation of viral particles, the samples were treated with either DNase or RNase or Proteinase K, respectively.

4.5 | Raman spectroscopy

Viral samples were subjected to Raman Spectroscopy by loading 20 μL on calcium fluoride (CaF_2) window. Raman spectra were recorded using a WITec Raman alpha300R (WITec, GmbH, Germany) confocal microscope. In the 500 to 3800 cm^{-1} range, Raman spectra (ex 532 nm,

8 mW, 100X Zeiss 1.46 NA) were acquired using 50 μm fiber and directed to a 300 mm spectrograph equipped with 600 g/mm grating and thermo-cooled CCD. Spectra were integrated for 10 seconds and averaged over 10 accumulations. On average, 15 spectra were recorded from each sample at a constant temperature of 20°C. Pre-processed spectra, that is, base-line corrected, area normalized and interpolated in 600 to 1800 cm^{-1} region were used as inputs for further data analysis [6, 16].

4.6 | Spectra analysis, virus prediction model and tool development

Raman spectroscopy files (SPC) were converted to the CSV format using SPC python package (<https://github.com/rohanisaac/spc>), data normalization and statistical analysis were performed in the R programming environment (<https://www.r-project.org/>). Data pre-processing and linear discriminant analysis was performed using the caret (<https://topepo.github.io/caret/>) and MASS (<https://cran.r-project.org/web/packages/MASS/>) package, respectively. Caret package was also used to perform k-fold cross-validation to check for the effect of data-splitting on model prediction accuracy. Mean accuracy over 10-folds, was found to be 93% (min: 84.21%, max: 100%, median: 94.59%, Kappa statistic 0.847). RVD was developed in python and R, and python Tkinter (<https://docs.python.org/3/library/tkinter.html>) library was used for development of the GUI.

5 | CONCLUSION

We describe a string proof of concept method using Raman spectroscopy to detect the RNA viruses in saliva with a detection sensitivity of 7.05×10^7 TU/mL viral particles. This provides the essential framework for field application of Raman Spectroscopy-based RVD in monitoring and responding to the COVID-19 pandemic, wherein 10^6 – 10^{11} viral RNA copies/mL of saliva is reported [17, 18]. The positive tests could be validated by follow-up testing with molecular biology laboratory-based diagnosis to confirm the status of SARS-CoV-2, while epidemiological containment measures are implemented for the individual.

ACKNOWLEDGMENTS

Authors thank all members of the Dutt laboratory for critically reviewing the manuscript.

CONFLICT OF INTEREST

The authors declare no financial or commercial conflict of interest.

AUTHOR CONTRIBUTIONS

Sanket Desai, Saket V. Mishra, Asim Joshi, Murali K. Chilakapati and Amit Dutt designed research. Sanket Desai, Saket V. Mishra, Asim Joshi and Arti Hole performed research. Shilpee Dutt, Saket V. Mishra, Asim Joshi, Murali K. Chilakapati, Shilpee Dutt, Sudeep Gupta and Amit Dutt analyzed data. Sanket Desai developed the RVD. Rohit Mishra developed the GUI. Shilpee Dutt, Debashmita Sarkar and Amit Dutt wrote the paper.

ORCID

Amit Dutt  <https://orcid.org/0000-0002-1119-4774>

REFERENCES

- [1] *Features, Evaluation and Treatment Coronavirus (COVID-19)—Mar20, 2020 StatPearls—NCBI Bookshelf.pdf*.
- [2] B. Udugama, P. Kadhiresan, H. N. Kozlowski, A. Malekjahani, M. Osborne, V. Y. C. Li, H. Chen, S. Mubareka, J. B. Gubbay, W. C. W. Chan, *ACS Nano* **2020**, *14*, 3822.
- [3] T. Yang, W. Yung-Chih, S. Ching-Fen, C. Chao-Min, *Diagnostics (Basel)* **2020**, *10*(3), 165.
- [4] C. V. Raman, K. S. Krishnan, *Nature* **1928**, *121*(3048), 501.
- [5] E. Ryzhikova, O. Kazakov, L. Halamkova, D. Celmins, P. Malone, E. Molho, E. A. Zimmerman, I. K. Lednev, *J. Biophotonics* **2015**, *8*(7), 584.
- [6] A. Sahu, K. Dalal, S. Naglot, P. Aggarwal, C. Murali Krishna, *PLoS One* **2013**, *8*(11), e78921.
- [7] A. Rehman, S. Anwar, S. Firdous, M. Ahmed, R. Rasheed, M. Nawaz, *Laser Physics* **2012**, *22*(6), 1085.
- [8] D. L. Farkas et al., Study of virus by Raman spectroscopy. in *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XI*, SPIE BIOS, San Francisco, CA **2013**.
- [9] A. L. Wyllie et al., *medRxiv* **2020**, 2020.04.16.20067835.
- [10] Administration., U.S.F.D, *Rutgers Clin Genom Laborat* **2020**, *1*. <https://www.fda.gov/media/136875/download>.
- [11] Gunjan Tyagi, A.D., Raviraj Deshpande, Vikram Gota, Murali Krishna Chilakapati. *Saliva Raman Spectroscopy: Explorations Of Sampling Methods*. in *18th European Conference on the Spectroscopy of Biological Molecules*. **2019**. Dublin, Ireland: ECSBM18.
- [12] B. Gong, J. H. Chen, R. Yajima, Y. Chen, E. Chase, D. M. Chadalavada, B. L. Golden, P. R. Carey, P. C. Bevilacqua, *Methods* **2009**, *49*(2), 101.
- [13] J. De Gelder, K. De Gussem, P. Vandenebeele, L. Moens, *J Raman Spectroscopy* **2007**, *38*(9), 1133.
- [14] P. Ruokola, E. Dadu, A. Kazmertsuk, H. Hakkanen, V. Marjomaki, J. A. Ihalainen, *J. Virol.* **2014**, *88*(15), 8504.
- [15] P. Iyer, S. V. Shrikhande, M. Ranjan, A. Joshi, N. Gardi, R. Prasad, B. Dharavath, R. Thorat, S. Salunkhe, B. Sahoo, P. Chandrani, H. Kore, B. Mohanty, V. Chaudhari, A. Choughule, D. Kawle, P. Chaudhari, A. Ingle, S. Banavali, P. Gera, M. R. Ramadwar, K. Prabhash, S. G. Barreto, S. Dutt, A. Dutt, *Int. J. Cancer* **2019**, *144*(8), 2008.
- [16] S. K. Patel, N. Rajora, S. Kumar, A. Sahu, S. K. Kochar, C. M. Krishna, S. Srivastava, *Anal. Chem.* **2019**, *91*(11), 7054.
- [17] Y. Pan, D. Zhang, P. Yang, L. L. M. Poon, Q. Wang, *Lancet Infect. Dis.* **2020**, *20*(4), 411.
- [18] Y. M. Bar-On et al., *eLife* **2020**, *9*, e57309.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Desai S, Mishra SV, Joshi A, et al. Raman spectroscopy-based detection of RNA viruses in saliva: A preliminary report. *J. Biophotonics*. 2020;13:e202000189. <https://doi.org/10.1002/jbio.202000189>