

## Research Campaign

# Sensory Transcriptomics: Transcriptomics as Versatile On-site Sensors During Spaceflight, Moon, and Mars Missions

Mohammad Shahbazi<sup>1\*</sup>, Richard Barker<sup>2</sup>, Afshin Beheshti<sup>3,4</sup>, Lauren M. Sanders<sup>5</sup>,  
Lindsay A. Rutter<sup>6</sup>

1. Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA.
2. Department of Botany, University of Wisconsin, Madison, WI 53706, USA.
3. KBR, Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA.
4. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
5. Blue Marble Space Institute of Science, Space Biosciences Division, NASA Ames Research Center, Moffett Field, CA 94035, USA.
6. Department of Genome Biology, University of Tsukuba, Ibaraki, 305-8575, Japan.

\* Primary author

[mshahbazi@uchicago.edu](mailto:mshahbazi@uchicago.edu)

We would like to thank Dr. Concetta Dipace from Oxford Nanopore Technologies for the information provided regarding their platforms.

### **Key words**

sensory omics, sensory transcriptomics, spaceflight sensors, Moon to Mars, light transcriptomics, artificial intelligence, machine learning

### **Abbreviations**

AI, artificial intelligence; ML, machine learning; SOP, standard operating procedure

## Introduction

Rather than analysis of single molecules or pathways, omics platforms analyze entire classes of biomolecules. RNA is at the center of orchestrating the molecular response of all living cells to their environment. Hence, the omics analysis of RNA (transcriptomics) can provide a holistic view on how the living organism perceives and reacts to the environment. This holistic view places transcriptomics in a unique position to serve as a complex sensor for monitoring living organisms during spaceflight missions, unparalleled in scope by the traditional sensors. Furthermore, the universal role of RNA makes transcriptomic sensors highly cross-compatible across species and life domains. This cross-compatibility means that the same platform with modular modifications could monitor all types of biological components of a spaceflight mission. This monitoring includes tracking crews' health for decision making on corrective medical steps or course of treatment<sup>1,2</sup>. Furthermore, there are a wide variety of transcripts that could be used to monitor the behavior of organisms of interest, from ripening of fruit<sup>3</sup> to response to environmental stress<sup>4,5</sup>. Therefore, transcriptomics data could provide critical sensory information for the quality control, management, or harvest of the bioproducts from the plant, animal, and microbial components of biological life support systems. The sensory information would be obtained via screening for affected biological pathways through gene ontology<sup>6</sup> or expression of signature transcripts.

Incorporating an artificial intelligence (AI) component would enable the assignment of standard operating procedures (SOPs) based on the transcriptomics outputs. The AI could improve the transcriptional sensory analyses and SOP assignments based on feedback from the crew or environmental sensors. The AI-based SOP assignment could be expanded to also incorporate data from other sensors during the decision-making process.

Here we propose a ten-year Research Campaign for developing both hardware and software infrastructure that would enable the use of transcriptomics as an on-site and cross-species sensory platform during spaceflight (Fig. 1). This sensory platform could provide detailed information for decision-making during spaceflight without the need for sample or data return. The independence from sample or data return would be of particular interest during Moon, Mars, and deep-space missions.

## Targets

### 1. Overall design adjustments and “light transcriptomics”

While "classic" transcriptomics designs with multiple replicates could be feasible during spaceflight, less demanding conditions like single sample methods<sup>7,8</sup> could be used in the sensory application of transcriptomics. These "light-transcriptomics" adaptations could significantly reduce the need for reagents and computational power.

### 2. Targets for sample preparation and sequencing

#### 2.1 Modular design for sample preparation pipeline

A modular design would be required for the cross-species application of transcriptomics sensors. At the hardware level, the modular design would be needed for sample preparation steps highlighted in Fig. 1. Due to the shared molecular principles involved in each sample preparation step, current platforms allow for a great deal of flexibility and cross-compatibility between organisms and tissues. Nevertheless, unification of these protocols into a single modular system would be the desired target and would enhance the transition to automation.

It is important to note the recent developments in direct RNA sequencing<sup>9</sup> and its potential to simplify the sample preparation steps. However, direct RNA sequencing currently relies on high concentrations of input RNA, and further developments are needed before it can be incorporated in pipelines with limited samples.

#### 2.2 Automation of sample preparation pipeline

Sample preparation for transcriptomics is multistep, sensitive, and complex. Automation would remove the requirement for crew time and training for sample preparation. Furthermore, user errors and variations during sample preparations could chip away significant power from the transcriptomics analyses. Automation will improve the analysis power by reducing user errors and limiting variations. The feasibility of automation of sample preparation steps has been demonstrated<sup>10-14</sup>, and adaptation of similar systems for spaceflight conditions for partial or complete automation is desired.

#### 2.3 Adaptation of non-toxic reagents and reusable material

There has been a traditional reliance on RNA extraction methods based on toxic organic solvents in laboratory conditions. Adaptation of non-toxic RNA extraction methods<sup>15-17</sup> alongside full automation of the sample preparation pipeline could address the associated risks regarding reagent toxicity and would allow for recycling of water, salt, and organic compounds post-analysis.

There is also currently a reliance on single-use plastic consumables for the sample preparation procedures. However, this is primarily due to convenience rather than necessity. It is conceivable that reusable or 3D printed recyclable material could replace single-use consumables during the long missions.

## 2.4 Adaptation of amplification strategies for limited samples

Techniques for amplification of cDNA before sequencing could be implemented when biological resources for sampling are limited. For instance, the adaptation of cDNA amplification techniques used for sequencing small sample quantities and single cells could be considered<sup>18,19</sup>.

# 3. Targets for transcriptional analysis and SOP assignment

## 3.1 Modular design and automation of data analysis pipeline

The modular design at the analysis level would consist of preloading the analysis pipelines (highlighted in Fig. 1) for each target organism. Traditional analysis of transcriptomics data requires training in bioinformatics and molecular biology. However, automatic analysis pipelines of transcriptomics data are currently available<sup>20,21</sup>. Implementing fully automated analysis pipelines would minimize the crews' need for training and involvement in the process and would increase the consistency of analysis across samples.

## 3.2 Model development for rapid SOP assignment and forecasting

Design variables would significantly affect the expected timelines of sensory transcriptomics from sampling to SOP assignment. Nevertheless, depending on the required depth of sequencing, we estimate the timeline to generally be within the first 24 or 48 hours from sampling. During this time window, circumstances that require immediate decisions can utilize models that incorporate profiles of other sensory inputs alongside archival transcriptomics data to assign immediate SOPs. Furthermore, in the case of rapidly evolving events, models can also be engineered to account for the elapsed time using forecasting to issue time-adjusted SOPs post-analysis.

## 3.3 Implementation of machine learning to improve sensory readout and SOP assignment

Machine learning (ML) can be used to improve both sensory readout and SOP assignment, using feedback from other sensors or crew. ML techniques have been used to categorize medical samples based on their transcriptomic profiles<sup>8,22</sup>. Similar approaches can be employed to identify affected biological pathways using the archival transcriptomic profiles from the mission or simulated experiments.

### 3.4 Development of knowledge bridge between Illumina and Nanopore transcriptomics platforms

To this date, most of the RNA sequencing data from spaceflight and simulated conditions have been generated using Illumina platforms<sup>23</sup>. On the other hand, current sequencing platforms used during spaceflight are Nanopore systems<sup>24</sup> due to their portability and compact size. While the data from Nanopore platforms could be directly used through gene annotation to generate sensory readouts, the development of knowledge bridges would allow detailed interpretation of transcriptomics data during spaceflight using the insights from the archival data.

### 3.5 Data storage adaptations

The data generated from sequencing platforms can be fully analyzed on-site with the current technology. Nevertheless, accumulative storage of raw data over long missions might pose a storage challenge that might require the development of storage strategies. For instance, generated data could be converted and stored as summary tables of transcript expression for future use. When communication of sensory transcriptomics findings across missions or with base is desired, employment of decentralized ML techniques such as federated learning and swarming could allow for the transfer of summarized parameters only. Federated learning and swarming allow for ML models to be trained on separate datasets without the need to share or download the data to a central location<sup>25-27</sup>.

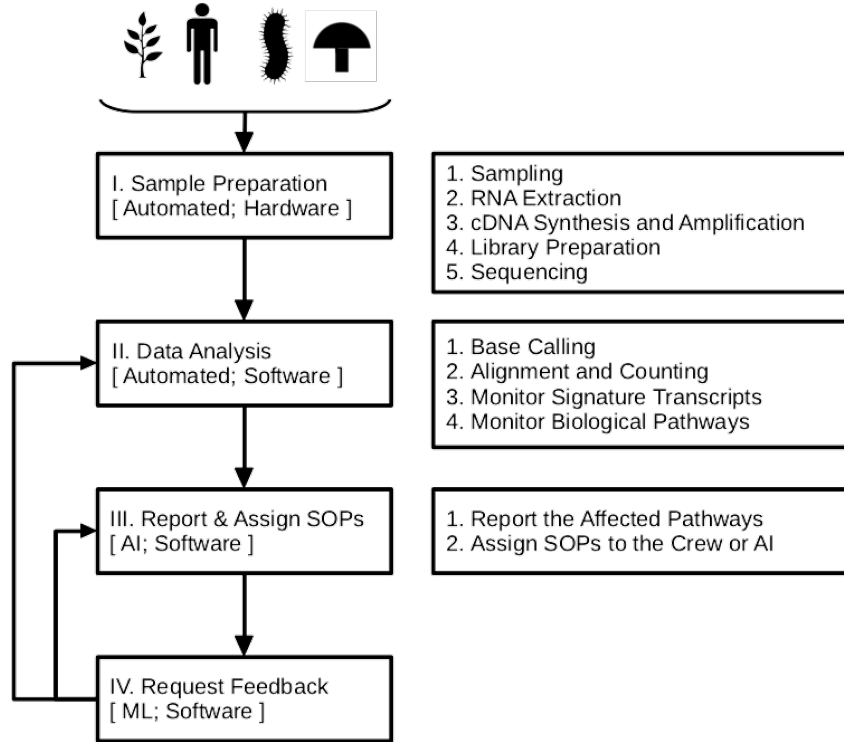
## Closing remarks

We would like to highlight two campaign papers focusing on the potential of AI and ML platforms in space biology<sup>28</sup> and space plant production<sup>29</sup>. These platforms have the potential to be unified or work in tandem with the AI/ML components presented in the context of sensory transcriptomics here.

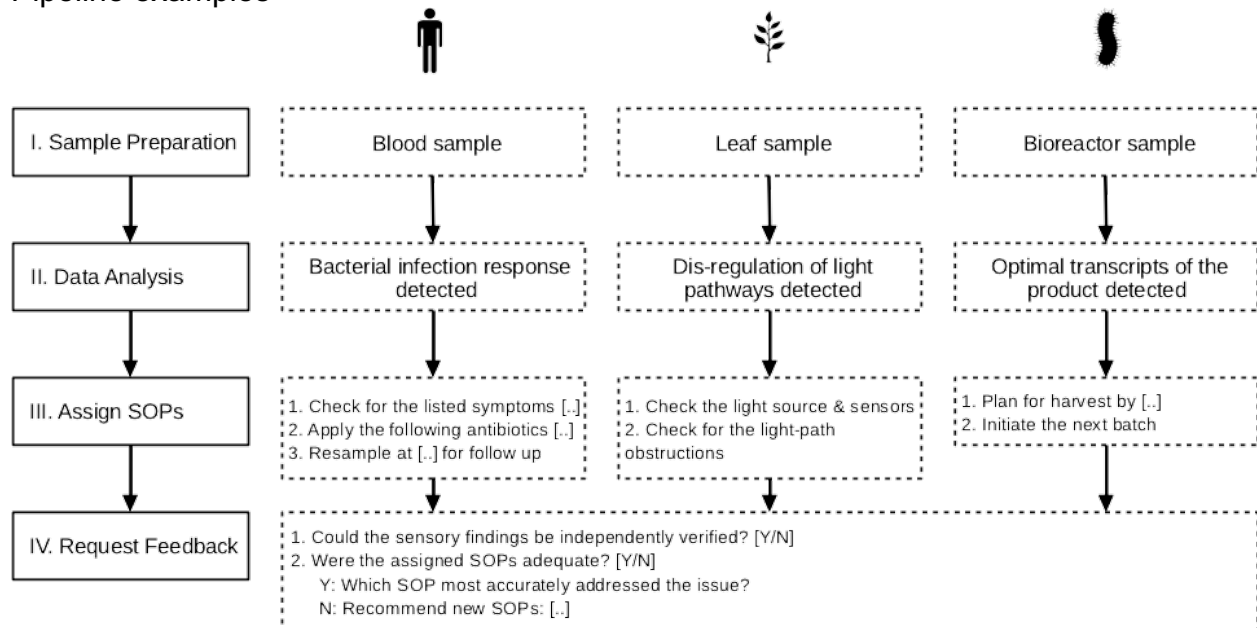
In addition, an important topical white paper has been recently submitted, with the guidelines for the improvement of transcriptomics experiments during spaceflight<sup>30</sup>. In particular, the authors have described the possibility of automation, on-site analysis of transcriptomic data, data size reduction strategies, and the use of ML in experiments during spaceflight missions. The similarities in the implementation strategies are the result of converging thoughts in the wake of the significant potential of transcriptomics in the future of space missions.

Here we highlighted the transformative potential of sensory transcriptomics and provided the targets for its implementation as a universal sensor for biological components of the missions. This platform could be used as an integral part of the decision-making apparatus during spaceflight, Moon, and Mars missions.

## Pipeline outline



## Pipeline examples



**Figure 1. Sensory transcriptomics pipeline.**

AI, artificial intelligence; ML, machine learning; SOP, standard operating procedure; [...] customized entry.

## References

1. Marco-Puche G, Lois S, Benítez J, Trivino JC. RNA-Seq Perspectives to Improve Clinical Diagnosis. *Front Genet.* 2019;10:1152. [doi:10.3389/fgene.2019.01152](https://doi.org/10.3389/fgene.2019.01152)
2. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet.* May 2016;17(5):257-71. [doi:10.1038/nrg.2016.10](https://doi.org/10.1038/nrg.2016.10)
3. Zhu B, Yang Y, Li R, et al. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *J Exp Bot.* Aug 2015;66(15):4483-95. [doi:10.1093/jxb/erv203](https://doi.org/10.1093/jxb/erv203)
4. Beheshti A, Cekanaviciute E, Smith DJ, Costes SV. Global transcriptomic analysis suggests carbon dioxide as an environmental stressor in spaceflight: A systems biology GeneLab case study. *Sci Rep.* 03 08 2018;8(1):4191. [doi:10.1038/s41598-018-22613-1](https://doi.org/10.1038/s41598-018-22613-1)
5. Ding Y, Jian H, Wang T, et al. Screening of candidate gene responses to cadmium stress by RNA sequencing in oilseed rape (*Brassica napus* L.). *Environ Sci Pollut Res Int.* Nov 2018;25(32):32433-32446. [doi:10.1007/s11356-018-3227-0](https://doi.org/10.1007/s11356-018-3227-0)
6. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 01 08 2019;47(D1):D330-D338. [doi:10.1093/nar/gky1055](https://doi.org/10.1093/nar/gky1055)
7. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. *BMC Bioinformatics.* Nov 06 2018;19(1):404. [doi:10.1186/s12859-018-2435-4](https://doi.org/10.1186/s12859-018-2435-4)
8. Cascianelli S, Molineris I, Isella C, Masseroli M, Medico E. Machine learning for RNA sequencing-based intrinsic subtyping of breast cancer. *Sci Rep.* 08 21 2020;10(1):14071. [doi:10.1038/s41598-020-70832-2](https://doi.org/10.1038/s41598-020-70832-2)
9. Gleeson J, Leger A, Praver YDJ, et al. Accurate expression quantification from nanopore direct RNA sequencing with NanoCount. *Nucleic Acids Res.* Nov 25 2021;[doi:10.1093/nar/gkab1129](https://doi.org/10.1093/nar/gkab1129)
10. Bohmann K, Hennig G, Rogel U, et al. RNA extraction from archival formalin-fixed paraffin-embedded tissue: a comparison of manual, semiautomated, and fully automated purification methods. *Clin Chem.* Sep 2009;55(9):1719-27. [doi:10.1373/clinchem.2008.122572](https://doi.org/10.1373/clinchem.2008.122572)
11. Vuosku J, Jaakola L, Jokipii S, et al. Does extraction of DNA and RNA by magnetic fishing work for diverse plant species? *Mol Biotechnol.* Jul 2004;27(3):209-15. [doi:10.1385/MB:27:3:209](https://doi.org/10.1385/MB:27:3:209)
12. Hess JF, Kohl TA, Kotrová M, et al. Library preparation for next generation sequencing: A review of automation strategies. *Biotechnol Adv.* 2020 Jul - Aug 2020;41:107537. [doi:10.1016/j.biotechadv.2020.107537](https://doi.org/10.1016/j.biotechadv.2020.107537)

13. Mayday MY, Khan LM, Chow ED, Zinter MS, DeRisi JL. Miniaturization and optimization of 384-well compatible RNA sequencing library preparation. *PLoS One*. 2019;14(1):e0206194. [doi:10.1371/journal.pone.0206194](https://doi.org/10.1371/journal.pone.0206194)
14. Haan T, McDougall S, Drown DM. Complete Genome Sequence of *Bacillus mycooides* TH26, Isolated from a Permafrost Thaw Gradient. *Microbiol Resour Announc*. Jun 13 2019;8(24)[doi:10.1128/MRA.00507-19](https://doi.org/10.1128/MRA.00507-19)
15. Oñate-Sánchez L, Vicente-Carbajosa J. DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC Res Notes*. Oct 20 2008;1:93. [doi:10.1186/1756-0500-1-93](https://doi.org/10.1186/1756-0500-1-93)
16. Drygin YF, Butenko KO, Gasanova TV. Environmentally friendly method of RNA isolation. *Anal Biochem*. 05 01 2021;620:114113. [doi:10.1016/j.ab.2021.114113](https://doi.org/10.1016/j.ab.2021.114113)
17. Brisco P, Sankbeil J, Kephart D. RNA Purification: A Rapid and Versatile Protocol for the Isolation of Total RNA. *Promega Notes*. 1997. p. p. 07.
18. Ziegenhain C, Vieth B, Parekh S, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell*. Feb 16 2017;65(4):631-643.e4. [doi:10.1016/j.molcel.2017.01.023](https://doi.org/10.1016/j.molcel.2017.01.023)
19. Pan X, Durrett RE, Zhu H, et al. Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc Natl Acad Sci U S A*. Jan 08 2013;110(2):594-9. [doi:10.1073/pnas.1217322109](https://doi.org/10.1073/pnas.1217322109)
20. Wolfien M, Rimbach C, Schmitz U, et al. TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation. *BMC Bioinformatics*. Jan 06 2016;17:21. [doi:10.1186/s12859-015-0873-9](https://doi.org/10.1186/s12859-015-0873-9)
21. Spinozzi G, Tini V, Adorni A, Falini B, Martelli MP. ARPIR: automatic RNA-Seq pipelines with interactive report. *BMC Bioinformatics*. Dec 21 2020;21(Suppl 19):574. [doi:10.1186/s12859-020-03846-2](https://doi.org/10.1186/s12859-020-03846-2)
22. Neff RA, Wang M, Vatansever S, et al. Molecular subtyping of Alzheimer's disease using RNA sequencing data reveals novel mechanisms and targets. *Sci Adv*. Jan 2021;7(2)[doi:10.1126/sciadv.abb5398](https://doi.org/10.1126/sciadv.abb5398)
23. Ray S, Gebre S, Fogle H, et al. GeneLab: Omics database for spaceflight experiments. *Bioinformatics*. 05 15 2019;35(10):1753-1759. [doi:10.1093/bioinformatics/bty884](https://doi.org/10.1093/bioinformatics/bty884)
24. Castro-Wallace SL, Chiu CY, John KK, et al. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. *Sci Rep*. 12 21 2017;7(1):18022. [doi:10.1038/s41598-017-18364-0](https://doi.org/10.1038/s41598-017-18364-0)
25. Chaterji S, Koo J, Li N, Meyer F, Grama A, Bagchi S. Federation in genomics pipelines: techniques and challenges. *Brief Bioinform*. 01 18 2019;20(1):235-244. [doi:10.1093/bib/bbx102](https://doi.org/10.1093/bib/bbx102)
26. Warnat-Herresthal S, Schultze H, Shastry KL, et al. Swarm Learning for decentralized and confidential clinical machine learning. *Nature*. 06 2021;594(7862):265-270. [doi:10.1038/s41586-021-03583-3](https://doi.org/10.1038/s41586-021-03583-3)



27. Li T, Sahu AK, Talwalkar A, Smith V. Federated Learning: Challenges, Methods, and Future Directions. IEEE; 2020. p. 50-60. [doi:10.1109/MSP.2020.2975749](https://doi.org/10.1109/MSP.2020.2975749)
28. Sanders LM, Scott RT, Costes SV. Machine Learning, Artificial Intelligence and Data Modeling for the Next Decade of Space Biology Research and Astronaut Health Support. 2021 (Campaign paper; in preparation).
29. Escobar C, Altaf N, Barker R, et al. Research Campaign - Artificial Intelligence for Autonomous Space Plant Production. 2021 (Campaign paper; in preparation).
30. Giacomello S, Klotz R, Scott RT, Das S, Borg J, Madrigal P. Topical: Taking space biology to the next era by ensuring standardization, applicability of in situ tissue analysis and computational predictions. 2021 (Topical paper, submitted).