

Bioinformatics Pipeline for Analyzing Space-Flown Mammary Gland

Soheil Abadifard¹, Maria Baloch¹, Joseph W. Guarnieri^{2,3}, Begum A. Mathyk^{4,5} and Hande K. McGinty^{1,*}

¹Department of Computer Science, Kansas State University, Manhattan, KS, USA

²Blue Marble Space Institute of Science, Seattle, WA, USA

³Center for Mitochondrial and Epigenomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA

⁴Department of Obstetrics and Gynecology, University of South Florida Morsani College of Medicine, Tampa, FL, USA

⁵Aerospace Science, Technology, Research and Applications (ASTRA) Center, Department of Mechanical and Aerospace Engineering, University of South Florida, FL, USA

Abstract

This study investigates how age affects the biological response of mammary gland tissue to spaceflight, focusing on potential risks for breast and ovarian cancers. We introduced a novel bioinformatics pipeline to analyze tissues from young and old mice which were sent to space. Through our analysis, we found that older mice had a significantly stronger molecular response to space flight elements. Key biological processes, such as signal transduction and gene regulation, were observed to be more active in the older group. The molecular changes in older mice showed a stronger association with genes linked to cancer, particularly ovarian and breast cancer. Our analysis points to conclusions that show how age is a critical factor in how the body responds to spaceflight. This study points to specific biological pathways that could be targeted for future cancer risk assessment and prevention for astronauts as well as assessing reproductive system risks for space missions.

Keywords

Bioinformatics, Gene Ontology, Drug Target Ontology (DTO), Differentially Expressed Genes, Space Radiation, Astronaut Health, Cancer Risk Assessment, NASA OSDR, Aging

1. Introduction

Breast cancer is one of the most common cancers among women. Exposure to ionizing radiation (IR) is an established risk factor for breast cancer development [1]. Space exploration and habitation aboard spacecraft pose several fundamental hazards to astronaut health, most notably microgravity, radiation, and enclosed spaces [2]. Space travel presents significant challenges, placing astronauts at an increased risk of developing a broad range of diseases and disorders [3]. A major concern for deep-space missions is space radiation, which is inevitably encountered during spaceflight. Therefore, female astronauts may face an increased risk of radiation exposure during long-duration deep space missions, where radiation exposure is increased significantly compared to Earth. The risk to breast tissue is highly relevant, given the fact that ionizing radiation is a risk factor for breast cancer. [1]. This biological risk is reflected in data derived from astronauts, where disease network analysis identifies breast cancer as one of the top disease/disorder conditions associated with space travel [3]. Therefore, investigations focused on characterizing the impact of spaceflight on sex-specific pathologies like breast cancer by analyzing the mammary gland tissues of model organisms are important for addressing female astronauts' health [2].

While the physiological hazards of spaceflight—including radiation and microgravity—are known to impact health across multiple systems, research characterizing these impacts on female astronauts needs further investigation [2]. Breast cancer is a significant health concern, yet most high-level

K-CAP 2025 Companion Volume

*Corresponding author.

✉ abadifard@ksu.edu (S. Abadifard); fmaria@ksu.edu (M. Baloch); guarnierijchop@gmail.com (J. W. Guarnieri); abegum@usf.edu (B. A. Mathyk); hande@ksu.edu (H. K. McGinty)

ORCID 0000-0002-2980-4251 (S. Abadifard); 0009-0002-6470-9361 (M. Baloch); 0000-0001-9832-6284 (B. A. Mathyk); 0000-0002-9025-5538 (H. K. McGinty)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

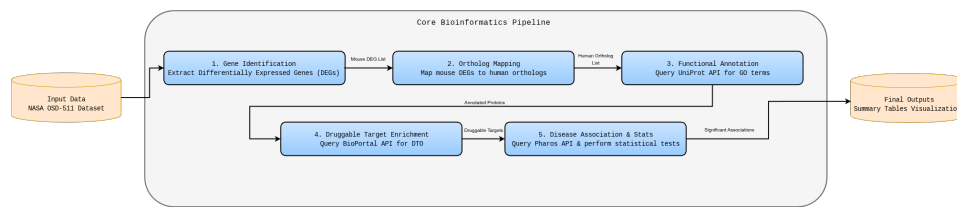


Figure 1: Schematic of Our Multi-layered Novel Bioinformatics Pipeline

transcriptomic analyses have centered on general metabolic or structural tissues, such as muscle or liver, rather than sex-specific pathologies [2]. Addressing this disparity requires a targeted focus on the molecular mechanisms within tissues like the mammary gland.

Another significant and largely unaddressed knowledge gap concerns how age modulates this biological response. Previous analyses of human astronaut data are often limited by a lack of diverse representation; for instance, one major study (E-GEOD-74708) included only two female astronauts. This lack of data emphasizes the urgent need for targeted model organism studies focused on sex and age differences [3].

This investigation operates on the premise that the cellular strategy for adapting to spaceflight stressors changes qualitatively with age, rather than simply involving a generalized decline in resilience [4]. Therefore, a systematic, age-dependent analysis of the mammary tissue’s response is necessary to address this gap and inform countermeasures for future long-duration missions.

2. Study Aim and Methodology

This study investigated age-dependent spaceflight effects on the mammary gland by analyzing public transcriptomic data from the NASA OSD-511 dataset [3, 5]. The data are sourced from female mice participating in the Rodent Research-8 (RR-8) mission, comprising tissues collected from young (10–12 weeks) and old (32 weeks) cohorts following a 40-day spaceflight.

To analyze the functional and pathological significance of gene expression changes, we developed a novel specialized bioinformatics pipeline as illustrated in Figure 1. This methodology moves beyond traditional analysis and introduces knowledge graph-directed analysis to determine how the cellular *strategy* for adapting to space stressors changes qualitatively with age. The key steps of our novel pipeline include:

Data Ingestion and Gene Identification: The pipeline begins by programmatically downloading all study files from the NASA OSDR, replicating the repository’s directory structure locally. Gene identifiers are then extracted from the curated Differentially Expressed Gene (DEG) lists for both the young and old space-flown mice to create a master list for annotation.

Ortholog Mapping for Clinical Relevance: The identified mouse DEGs are systematically mapped to their human orthologs. This crucial step ensures that all subsequent functional, pathway, and disease analyses are directly relevant to human physiology and pathology.

Protein and Functional Annotation Using the human ortholog list, the pipeline queries the UniProt [6] REST API for functional annotation. This process retrieves detailed protein information, and Gene Ontology (GO) [7] terms across its three main branches: Biological Process, Molecular Function, and Cellular Component. Although automated methods exist [8] to ensure high-quality data, this step prioritizes manually reviewed Swiss-Prot entries.

Druggable Target and Ontology Enrichment: The annotated protein list is subsequently enriched by mapping targets to the Drug Target Ontology (DTO) [9, 10, 11] via the BioPortal API [12] to identify and classify potential druggable protein families.

Disease Association and Statistical Analysis: To establish links between molecular targets and human diseases, the pipeline queries the Pharos [13] GraphQL API. Downstream computational analysis includes frequency counts of GO and DTO terms and statistical comparisons of disease associations

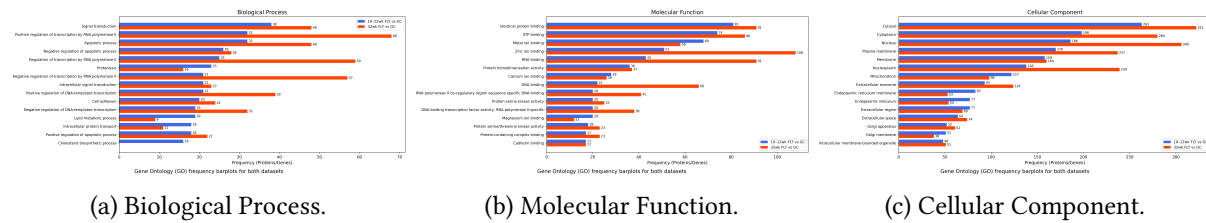


Figure 2: Age-Dependent Shifts in Gene Ontology (GO) Enrichment.

(a) Biological Process analysis shows a shift from lipid metabolism in young mice to transcriptional regulation in old mice. (b) Molecular Function reveals a massive spike in zinc ion binding in the older cohort. (c) Cellular Component analysis confirms the nuclear localization of the stress response in older mice compared to the mitochondrial focus in the young group.

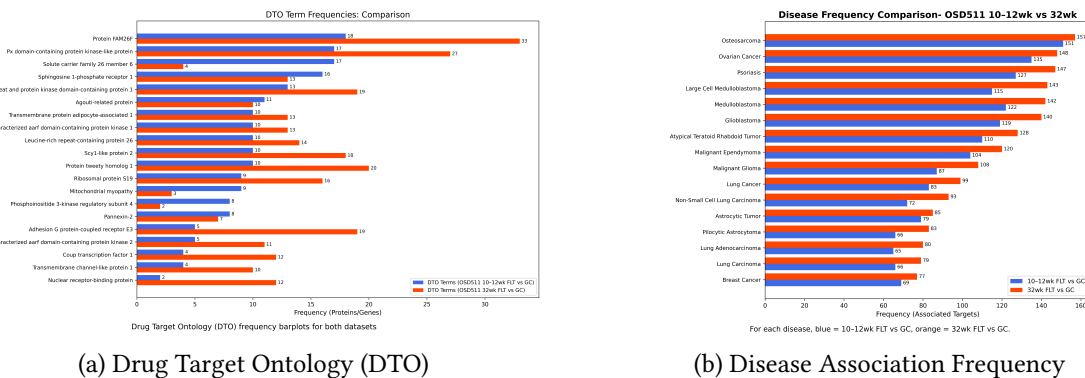


Figure 3: Qualitative Shift in Druggable Targets and Disease Risks.

(a) DTO mapping reveals a transition from transport-based targets (Solute Carriers) in young mice to adhesion-based targets (Protein FAM26F) in old mice. (b) This molecular shift correlates with an increased frequency of associations with sex-specific pathologies, including ovarian and breast cancer, in the older cohort (orange bars).

between the two age cohorts using Fisher's exact test with False Discovery Rate (FDR) correction.

This novel, comprehensive, multi-layered approach aims to provide necessary insights for identifying potential countermeasures relevant to sex-specific pathologies, such as breast and ovarian cancer, associated with space travel.

3. Findings and Discussion

Gene Ontology (GO) analysis of Biological Processes revealed a fundamental divergence in adaptive strategies (Figure 2a). The young cohort prioritized metabolic maintenance, exhibiting higher gene counts in *Lipid metabolic process* and *Proteolysis* compared to the old cohort. This metabolic focus was corroborated by Cellular Component analysis, where young mice showed greater enrichment in the *Mitochondrion* (122 proteins vs. 98 in old) (Figure 2c).

In contrast, the older cohort exhibited a "transcriptional overdrive." The most significantly enriched processes in older mice were *Positive regulation of transcription by RNA polymerase II* (68 proteins) and *Apoptotic process*, both of which substantially exceeded levels in young mice. This focus on genomic regulation was reflected in the Cellular Component data, where the *Nucleus* was the dominant compartment for older mice (306 proteins) compared to the young group (166 proteins). Furthermore, Molecular Function analysis highlighted a massive expansion in regulatory binding capacity in older mice (Figure 2b), particularly *Zinc ion binding* (108 vs. 51) and *RNA binding* (91 vs. 43), suggesting a heavy reliance on zinc-finger transcription factors to mediate the stress response.

Mapping the proteome to the Drug Target Ontology (DTO) confirmed a qualitative shift in the therapeutic landscape (Figure 3a). The response in young mice was characterized by transport and signaling targets, specifically *Solute carrier family 26 member 6* (17 vs. 4 in old) and *Sphingosine 1-phosphate receptor 1*. Conversely, the older cohort shifted toward targets involved in cell adhesion and

structural integrity. *Protein FAM26F* was the top druggable target in older mice (33 proteins vs. 18 in young), followed by a significant emergence of *Adhesion G protein coupled receptor E3* (19 vs. 5). This indicates that while young mice modulate transport efficiency to adapt, older mice potentially upregulate adhesion markers in response to tissue stress.

The distinct molecular shifts in older mice correlated with an increased frequency of disease associations. Older mice exhibited higher gene counts linked to pathologies across the board (Figure 3b). Notably, sex-specific pathologies saw elevated risk profiles: associations with *Ovarian Cancer* increased from 135 in young mice to 148 in old mice, and *Breast Cancer* targets rose from 69 to 77. This suggests that the "transcriptional overdrive" observed in the older cohort activates a gene network that overlaps significantly with carcinogenic pathways.

Declaration on Generative AI

During the preparation of this study, the author(s) used ChatGPT-Pro and Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

- [1] K. Kumar, B.-H. Moon, K. Datta, J. Fornace, Albert J., S. Suman, Simulated galactic cosmic radiation (gcr)-induced expression of spp1 coincide with mammary ductal cell proliferation and preneoplastic changes in apc mouse, *Life Sciences in Space Research* 36 (2023) 116–122. doi:10.1016/j.lssr.2022.09.006.
- [2] C. Arnold, J. Casaletto, P. Heller, Spaceflight disrupts gene expression of estrogen signaling in rodent mammary tissue, *Medical Research Archives* 12 (2024). doi:10.18103/mra.v12i3.5220.
- [3] A. Sakharkar, J. Yang, Designing a novel monitoring approach for the effects of space travel on astronauts' health, *Life* 13 (2023) 576. doi:10.3390/life13020576.
- [4] J. A. Casaletto, T. Zhao, J. Yeung, A. Ansari, A. Raj, A. Mishra, A. Fry, K. Sun, A. Lee, M. Cline, L. M. Sanders, S. V. Costes, Machine learning ensemble identifies distinct age-related response to spaceflight in mammary tissue, *bioRxiv* (2025). doi:10.1101/2025.02.17.638732.
- [5] S. G. Gebre, R. T. Scott, A. M. Saravia-Butler, D. K. Lopez, L. M. Sanders, S. V. Costes, Nasa open science data repository: open science for life in space, *Nucleic Acids Research* 53 (2024) D1697–D1710. doi:10.1093/nar/gkae1116.
- [6] A. Bateman, Uniprot: the universal protein knowledgebase in 2025, *Nucleic Acids Research* 53 (2024) D609–D617. doi:10.1093/nar/gkae1010.
- [7] S. A. Aleksander, The gene ontology knowledgebase in 2026, *Nucleic Acids Research* 54 (2025) D1779–D1792. doi:10.1093/nar/gkaf1292.
- [8] A. Singh Dalal, H. McGinty, Echo-llm evidence-checked hierarchical ontology, *Open Conference Proceedings* 8 (2025). doi:10.52825/ocp.v8i.3173.
- [9] H. K. McGinty, Knowledge Acquisition and Representation Methodology (KNARM) and Its Applications, Ph.D. thesis, University of Miami, 2018.
- [10] H. Küçük McGinty, U. Visser, S. Schürer, How to develop a drug target ontology: Knowledge acquisition and representation methodology (knarm), in: *Bioinformatics and Drug Discovery*, Springer, 2019, pp. 49–69. doi:10.1007/978-1-4939-9089-4_4.
- [11] Y. Lin, S. Mehta, H. k. McGinty, J. P. Turner, D. Vidovic, M. Forlin, A. Koleti, D.-T. Nguyen, L. J. Jensen, R. Guha, S. L. Mathias, O. Ursu, V. Stathias, J. Duan, N. Nabizadeh, C. Chung, C. Mader, U. Visser, J. J. Yang, C. G. Bologa, T. I. Oprea, S. C. Schürer, Drug target ontology to classify and integrate drug discovery data, *Journal of Biomedical Semantics* 8 (2017). doi:10.1186/s13326-017-0161-x.
- [12] J. Vendetti, N. L. Harris, M. V. Dorf, A. Skrenchuk, J. H. Caufield, R. S. Gonçalves, J. B. Graybeal, H. Hegde, T. Redmond, C. J. Mungall, M. A. Musen, Biportal: an open community resource for sharing, searching, and utilizing biomedical ontologies, *Nucleic Acids Research* 53 (2025) W84–W94. doi:10.1093/nar/gkaf402.
- [13] K. J. Kelleher, T. K. Sheils, S. L. Mathias, J. J. Yang, V. T. Metzger, V. B. Sramshetty, D.-T. Nguyen, L. J. Jensen, D. Vidović, S. C. Schürer, J. Holmes, K. R. Sharma, A. Pillai, C. G. Bologa, J. S. Edwards, E. A. Mathé, T. I. Oprea, Pharos 2023: an integrated resource for the understudied human proteome, *Nucleic Acids Research* 51 (2022) D1405–D1416. doi:10.1093/nar/gkac1033.