Linking Retinal Genes to Oxidative Stress in Space with Machine Learning

Anish Kelam

Innovation Academy, Alpharetta GA, United States

Background

Spaceflight Associated Neuro-Ocular Syndrome (SANS)

SANS is a complex condition affecting astronauts during and after extended space missions, causing neuro-ophthalmologic changes like optic disc edema, retinal and choroidal folds, globe flattening, and hyperopic refractive shifts. These changes pose significant risks to vision and long-term ocular health, potentially threatening mission success and astronaut well-being. Multiple hypotheses, including elevated intracranial pressure, compartmentalized cerebrospinal fluid in the optic nerve sheath, and glymphatic dysfunction, have been proposed. However, none have been fully accepted due to their inability to explain the condition's asymmetry, individual variability, or persistence post-flight (Lee et al., 2018; Martin Paez et al., 2020).

Role of Oxidative Stress in SANS

An underexplored aspect of SANS is the role of oxidative stress, characterized by an overproduction of reactive oxygen species and insufficient antioxidant defenses. Microgravity significantly increases oxidative stress in the retina, leading to cellular damage, impaired repair mechanisms, and blood-retinal barrier breakdown (Martin Paez et al., 2020; Ong et al., 2023). 4-Hydroxynonenal (HNE), a lipid peroxidation byproduct, serves as a biomarker for measuring oxidative stress and has been linked to SANS in literature. Genetic predispositions, including variants in methionine synthase reductase and serine hydroxymethyltransferase, appear to influence susceptibility to SANS. Despite these findings, genetic and molecular mechanisms of SANS remain poorly understood, highlighting the need for further investigation (Lee et al., 2018).

Causal Research and Inference Search Platform (CRISP)

The CRISP tool (Budd et al., 2021) offers a promising approach to uncover genetic factors linked to ocular problems observed to be caused by SANS. Designed by NASA, CRISP uses machine-learning-based causal inference models and was built to analyze low sample and high-dimensional datasets, like those usually produced during spaceflight experiments. This platform employs multiple machine learning models that work together to identify a single set of features that are most predictive of specific phenotypes. This allows for the integration of genetic, transcriptomic, and phenotypic data to isolate genes linked to oxidative stress in the retina. By identifying genetic markers and molecular mechanisms, CRISP provides a pipeline for developing targeted interventions to address SANS and enhance astronaut health during long-term missions (Casaletto et al., 2023).

Impact on Space Medicine

As humanity prepares for extended space missions, such as those to Mars, understanding the genetic and molecular mechanisms underlying SANS is critical. Insights into these mechanisms will enable the development of preventive strategies, targeted treatments, and improved astronaut selection processes, improving overall mission safety and success. Moreover, research on oxidative stress and retinal damage in SANS could inform treatments for Earth-based conditions like idiopathic intracranial hypertension and age-related macular degeneration, offering significant contributions to both space and ocular medicine.

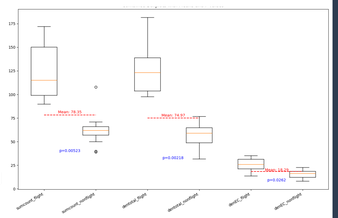
Purpose

The purpose of this study is to identify genetic and molecular targets linked to oxidative stress associated with SANS while also addressing significant gaps in the understanding of the role of oxidative stress in the cause and development of SANS. This will be achieved by utilizing CRISP to identify genes most casually related to oxidative stress and then analyzing pathways and gene sets associated with these genes to validate findings and identify associated molecular mechanisms. This study will use publicly available data on the NASA Open Science Data Repository that was collected from retinal tissues obtained from mice flown in space and stationed on Earth. This data includes RNA sequencing data that contains thousands of gene and gene expression levels and measurements from images of retinal tissue stained with the HNE biomarker (NASA, 2019; NASA, 2023).

Methodology

1. HNE Phenotype Statistical Analysis:

- Compare spaceflight and ground control groups for each HNE metric area from the OSD-557 database entry by using box plots
- Identify metric areas with significant differences using t-tests ($p<0.05$)



2. RNA-Seq Filtering:

- Import raw RNA-seq data from the OSD-255 database entry
- Intersect RNA-seq data with HNE phenotype data to ensure data alignment with samples
- Apply RNA-seq filtering methods:

Filter	Description	Number of genes removed by filter	Number of genes after applying filter
-	Initial size of data set	-	56,840
protein-coding	Filter out transcripts that do not code for proteins	35,173	21,667
Low counts	Filter out genes with expression less than 50 counts in 80% or more of the samples	10,317	11,350
ENSEMBL mapping	Filter out transcripts which do not have a corresponding gene ID in ENSEMBL	3	11,347

3. Phenotype Data Binarization:

- Calculate the mean expression levels for each phenotype metric
- Convert phenotype data to binary format:
  - Assign 0 for values below the mean.
  - Assign 1 for values above the mean.

4. Normalization:

Transformation	Description	Reference using transformation for transcriptomic data
Log scale	Scales values to their log base 2	Quinn et al [26]
Z-score	Scales values to their number of standard deviations from the mean	Zwiener et al [27]
Square root	Scales values to their square roots	Zhang et al [28]
Median of ratios	Scales values to account for sequencing depth, gene length, and outliers	Robinson et al [29]

5. CRISP Experiment:

- Configure CRISP tool settings
- Run the CRISP experiment with modified Streamlit code to visualize data and identify predictive genes

6. Pathway Analysis:

- Perform pathway analysis using the ShinyGO tool by submitting the 20 genes identified by CRISP and all genes input into the CRISP experiment as background genes to identify any Gene Ontology biological pathways that overlap CRISP genes to the background genes (Ge et al., 2019)

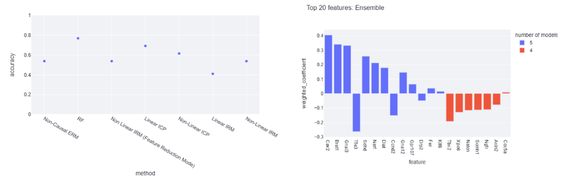
7. Gene Set Enrichment Analysis

- Perform gene set enrichment analysis by using the Human Molecular Signatures Database ontology gene sets C5 collection and identify the top 10 gene ontologies from all ontology gene sets that have a significant overlap with the CRISP gene set in the mouse genome (Subramanian et al., 2005)

Results

20 Genes Most Linked to Oxidative Stress

The CRISP experiment effectively identified the top 20 retinal genes most predictive of the oxidative stress phenotype associated with SANS. A graph of test accuracies showed robust predictive accuracy across the ensemble models, confirming the ability of CRISP in identifying causal genetic links. The identified genes exhibited a high degree of concordance among the models used in the ensemble, further validating the results. Additionally, the direction of the direction of the genes' corresponding bar indicates whether higher or lower expression impacts oxidative stress.



Pathway Analysis Results

Pathway analysis allowed for the identification of the four pathways below sorted by enrichment false discovery rate with a cutoff of <0.05 (Ge et al., 2019). The identified pathways highlight potential mechanisms by which retinal tissues respond to oxidative stress and lipid peroxidation induced by spaceflight conditions. Muscle cell and striated muscle cell proliferation pathways, although primarily associated with muscle tissues, may reflect similar cellular responses in the retina, as both tissues are metabolically active and susceptible to oxidative damage. In retinal cells, such as photoreceptors and those in the inner nuclear and ganglion cell layers, lipid peroxidation could activate reparative or compensatory proliferation pathways to replace or repair damaged cells. The broader pathways of cell population proliferation and its regulation suggest attempts to maintain homeostasis by balancing cell death and renewal in response to oxidative injury. Together, these pathways indicate the retina's effort to mitigate damage caused by oxidative stress while preserving its structure and function in harsh spaceflight conditions.

Enrichment FDR	nGenes	Pathway Genes	Fold Enrichment	Pathways (click for details)
1.2E-02	5	258	16.4	Muscle cell proliferation
4.2E-02	9	2102	4.3	Cell population proliferation
4.2E-02	3	78	29.5	Striated muscle cell proliferation
4.2E-02	8	1765	4.7	Reg. of cell population proliferation

Gene Set Enrichment Analysis Results

Gene set enrichment analysis identified the top 10 ontology gene sets below sorted by the false discovery rate adjusted significance q-value (Subramanian et al., 2005). The identified gene sets associated with oxidative stress are connected to structural and functional abnormalities in areas affected by SANS. Pathways such as abnormality of the orbital region and abnormal facial shape suggest that increased oxidative stress may contribute to structural changes in the orbital and craniofacial regions, leading to further neuro-ocular dysfunction during spaceflight. The enrichment of neoplasm and renal cell carcinoma points to broader implications of oxidative stress in promoting cellular and systemic pathologies, potentially affecting other organs as seen here. Furthermore, the presence of the catalytic complex indicates the role of protein complexes with catalytic activity in mediating oxidative processes, which may amplify lipid peroxidation and oxidative stress in retinal tissues.

Gene Set Name [# Genes (K)]	Description	# Genes in Overlap (k)	k/K	p-value	FDRq-value
HP_ABNORMALITY_OF_THE_ORBITAL_REGION [1623]	Abnormality of the orbital region	8		3.66 e-7	5.9 e-3
HP_MACROCEPHALY [440]	Macrocephaly	5		1.57 e-6	1.01 e-2
HP_ABNORMAL_EYELID_MORPHOLOGY [1369]	Abnormal eyelid morphology	7		1.88 e-6	1.01 e-2
HP_NEOPLASM_OF_THE_GENITOURINARY_TRACT [236]	Neoplasm of the genitourinary tract	4		4.16 e-6	1.38 e-2
HP_NEOPLASM [979]	Neoplasm	6		4.28 e-6	1.38 e-2
HP_RENAL_CELL_CARCINOMA [72]	Renal cell carcinoma	3		5.18 e-6	1.39 e-2
HP_ABNORMAL_FACIAL_SHAPE [1084]	Abnormal facial shape	6		7.67 e-6	1.55 e-2
GOCC_CATALYTIC_COMPLEX [1739]	A protein complex which is capable of catalytic activity. [GOC:bhm, GOC:TermGenie, PMID:8077207]	7		9.08 e-6	1.55 e-2
HP_ABNORMAL_PALATE_MORPHOLOGY [1152]	Abnormal palate morphology	6		1.08 e-5	1.55 e-2
HP_LANGUAGE_IMPAIRMENT [1156]	Language impairment	6		1.11 e-5	1.55 e-2

Conclusion

Causal Machine Learning in Space Biology

The application of causal inferencing machine learning techniques in this study further validates their use in the field of space biology. This subfield of space biology is currently heavily underdeveloped, so the use of AI based tools that successfully used casual inferencing greatly expands the scope of this subfield to a completely new bodily system. This also showcases the potential of using AI to better understand genetic and molecular mechanisms for complex biological functions affected by the extreme environment of spaceflight, as it identified new findings that other methods have not been able to do.

Potential Treatments

The identification of validated genetic and molecular mechanisms that are linked to a disease is an essential step for therapeutic development. With the discovery of these potential targets, medical researchers can focus on exploring these high priority genes and their associated pathways to develop drugs, gain a better understanding of potential diagnostic tools, and even lay the foundation for the development of personalized medicine based on the genetic information of astronauts suffering from the effects SANS. This also saves a large amount of time and resources for pharmaceutical companies that is spent on researching and testing for target genes and mechanisms associated with a disease, which is especially difficult with limitations with data collection in spaceflight.

Future Work

This pipeline can also be adapted to further analyze SANS related effects using other available data on the NASA Open Science Data Repository, including micro-CT imaging to explore structural changes in tissue layers, peanut agglutinin phenotyping for cone photoreceptor degeneration, TUNEL assays for apoptotic DNA fragmentation, PECAM data for vascular integrity, and tonometry metrics for intraocular pressure changes. Future studies could use this framework to identify genetic and molecular targets associated with specific ophthalmological dysfunctions observed in the data, paving the way for tailored interventions for other aspects of SANS related complications. Additional work incorporating other omics data, such as metabolomics and proteomics, could produce higher quality and a more thorough understandings of the molecular aspects of SANS, enabling the development of truly comprehensive countermeasures to protect astronaut ocular health. This research not only fills critical knowledge gaps in SANS pathology but also establishes a scalable pipeline for investigating the genetic foundation of other harmful effects on the eye and other organ systems due to spaceflight.

References

Budd, S., Blaas, A., Hoarfrost, A., Khezeli, K., D'Silva, K., Soboczenski, F., Mackintosh, G., Chia, N., & Kalantari, J. (2021). Prototyping CRISP: A Causal Relation and Inference Search Platform applied to Colorectal Cancer Data. *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*. https://doi.org/10.1109/lifetech52111.2021.9391819

Casaletto, J., Scott, R., Myrick, M., Mackintosh, G., Hamed Chok, Saravia-Butler, A., Hoarfrost, A., Galazka, J., Sanders, L., & Costes, S. (2023). Analyzing the Relationship Between Gene Expression and Phenotype in Space-Flown Mice Using a Causal Inference Machine Learning Ensemble. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-2332064/v1

Ge, S. X., Jung, D., & Yao, R. (2019). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Lee, A. G., Mader, T. H., Gibson, C. R., Brunstetter, T. J., & Tarver, W. J. (2018). Space flight-associated neuro-ocular syndrome (SANS). *Eye*, *32*(7), 1164–1167. https://doi.org/10.1038/s41433-018-0070-y

Martin Paez, Y., Mudie, L. I., & Subramanian, P. S. (2020). Spaceflight Associated Neuro-Ocular Syndrome (SANS): A Systematic Review and Future Directions. *Eye and Brain*, *Volume 12*, 105–117. https://doi.org/10.2147/eb.s234076

NASA. (2019). *Spaceflight influences gene expression, photoreceptor integrity, and oxidative stress-related damage in the murine retina*. NASA OSDR. https://doi.org/10.26030/mebr-1747

NASA. (2023). *Spaceflight influences gene expression, photoreceptor integrity, and oxidative stress related damage in the murine retina (RR-9)*. NASA OSDR. https://doi.org/10.26030/yv31-1a54

Ong, J., Mader, T. H., Gibson, C. R., Mason, S. S., & Lee, A. G. (2023). Spaceflight associated neuro-ocular syndrome (SANS): an update on potential microgravity-based pathophysiology and mitigation development. *Eye (London, England)*, *37*(12), 2409–2415. https://doi.org/10.1038/s41433-023-02522-y

Quinn, T. P., Crowley, T. M., & Richardson, M. F. (2018). Benchmarking differential expression analysis tools for RNA-Seq: normalization-based vs. log-ratio transformation-based methods. *BMC Bioinformatics*, *19*(1). https://doi.org/10.1186/s12859-018-2261-8

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), R25. https://doi.org/10.1186/gb-2010-11-3-r25

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Zhang, Z., Yu, D., Seo, M., Hersh, C. P., Weiss, S. T., & Qiu, W. (2019). Novel Data Transformations for RNA-seq Differential Expression Analysis. *Scientific Reports*, *9*(1), 4820. https://doi.org/10.1038/s41598-019-41315-w

Zwiener, I., Frisch, B., & Binder, H. (2014). Transforming RNA-Seq Data to Improve the Performance of Prognostic Gene Signatures. *PLoS ONE*, *9*(1), e85150. https://doi.org/10.1371/journal.pone.0085150