The Capstone Committee for Christopher Haas certifies that this is the approved version of the following capstone:

OmicsSpace: A Proposed Omics Data Plan for NASA Human Research and Clinical Investigations

Committee:

Edward Powers, MD

Erik Antonsen, MD, PhD, MS

Susana Zanello, PhD

Dean, Graduate School

OmicsSpace: A Proposed Omics Data Plan for NASA Human Research and Clinical Investigations

by

Christopher T. Haas, MD, BS

Capstone

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Public Health

The University of Texas Medical Branch

July 2018

Dedication

To my family, friends and faculty who have supported me along the way.

Acknowledgements

I would like to acknowledge my capstone committee, Dr. Edward Powers, Dr. Susanna Zanello, and Dr. Erik Antonsen for their guidance. I would also like to acknowledge support and advisement from numerous individuals at NASA Johnson Space Center and the University of Texas Medical Branch, School of Public Health

OmicsSpace: A Proposed Omics Data Plan for NASA Human Research and Clinical Investigations

Publication No._____

Christopher T. Haas, MPH

The University of Texas Medical Branch, 2018

Supervisor: Dr. Cara Pennel, DrPH, MPH

Abstract: Genetic mutations can be introduced during one's lifetime through errors in DNA processing or as the result of environmental exposures. Astronauts can experience numerous environmental changes during spaceflight including exposure to solar and cosmic radiation, circadian rhythm shifts, body fluid shifts, altered levels of carbon dioxide and oxygen, and changes in ambient pressure. It is unknown at this point as to whether spaceflight can cause permanent changes to our genes and how they are expressed. The answer likely depends on each astronaut's genetic background and what mission(s) the astronaut has flown. The field of omics allows us to look closely at one's DNA sequence as well as the modification of gene expression and downstream metabolic pathways. Astronaut omics analyses may take us one step closer to understanding how the blueprint for our cellular machinery and the translation of this blueprint can change over time. Through advancement of omics technologies and analytical methods, we are becoming more proficient at customizing treatment for select diseases and may even have opportunities to customize preventative measures in the future. I plan to propose a program plan for how NASA could capture operationally relevant omics data from astronaut experiments over time. The program plan will contain recommendations on how to scope a genetic repository in the future at NASA that can be used for future research, occupational surveillance to help identify clinical issues related to the spaceflight environment, and management of astronaut health.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

DNA	Deoxyribonucleic acid
GSBS	Graduate School of Biomedical Science
HRP	Human Research Program
LSAH	Lifetime Surveillance of Astronaut Health
NASA JSC	NASA Johnson Space Center
NPD	NASA Policy Directive
RNA	Ribonucleic acid
TDC	Thesis and Dissertation Coordinator
UTMB	University of Texas Medical Branch

Chapter 1 Introduction

"Omics" as a term can be interpreted as the suite of studies that investigates how our genetic information is translated into high-level functions of the human body. Genomics is the field that looks at the sequence of nucleotide bases within our DNA or genes – the code in our cells that serves as our master blueprint. Epigenomics studies slight transmittable modifications to DNA such as the creation of methyl groups as a means of turning genes on and off. Transcriptomics is the field that looks at how and when our DNA is transcribed into RNA, which is the messenger molecule that initiates the production of proteins. Proteomics looks at what proteins and how much protein is present in a biological pathway, and metabolomics looks at interactions between proteins in a biological pathway. The use of more than one of these fields to explain a biological pathway has been coined as another term, integrative omics (Karczewski & Snyder, 2018).

Regardless of what field of omics is being considered, omics is advancing rapidly with the introduction of new lab tools and techniques and data analytics software. With a greater understanding of how individuals' biological pathways differ, researchers and clinicians have been able to customize medical treatments and offer "personalized medicine" in the hospital and clinic.

While various institutions across the United States are creating departments for omics and personalized medicine efforts, NASA is at the initial stages of human omics investigations. NASA has started to analyze collections of omics data from experiments on plants, animals and microbes in an effort called GeneLab. As part of this effort, the GeneLab working group is defining standard omics data analysis pathways (Caldwell, 2018). A distinct set of human omics investigations were included in the NASA twin study (https://www.nasa.gov/twins-study/research). However, NASA has not yet created an

overarching framework for omics across the astronaut population or created personalized medicine methodology for individual astronauts.

SPECIFIC AIMS

The main question that I plan to address through my capstone is the following: How can NASA best establish a program to collect, store, and analyze astronaut omics data, and what are the best uses of omics data for the astronaut population in the future?

This proposed program plan is applicable to current NASA astronauts and retired NASA astronauts. NASA currently monitors astronaut health over time through the Lifetime Surveillance of Astronaut Health (LSAH) program. Under this program, retired astronauts can elect to participate in annual medical monitoring at NASA JSC. These annual screens include basic physical exams, basic lab tests, and eye and hearing assessments. In the past, NASA had not been directed to provide ongoing care to retired astronauts. However, in 2017, congress passed the TREAT ("To Research, Evaluate, Assess, and Treat Astronauts") act, which mandates that retired astronauts can continue to seek healthcare from NASA for any ongoing work-related medical issues (Lewis, 2017b). The TREAT act will likely provide NASA with increased longitudinal monitoring of the astronaut population and create additional opportunities for the collection of omics data over time in both active and retired astronauts. The knowledge gained from omics programs in the astronaut population can potentially be used to optimize astronaut health before, during, and after spaceflight.

SIGNIFICANCE

Omics data and its application to clinical medicine is evolving at a rapid pace. Numerous groups across the country such as the National Institute of Health and the

National Cancer Institute are working on ways to obtain data that can be used in a clinically meaningful way. Table 1 is a sampling of current omics projects in the United States:

Organization	Program Title	Description	Webpage
National Institutes of Health	The Precision Medicine Initiative	Collection and analysis of health records, medications, environmental exposure of 1 million US citizens.	https://allofus.nih.gov
National Cancer Institute	Cancer Genomics Research	Standardized omics data set for 14,500 cancer patients.	www.cancer.gov/research/areas/genomics
St. Jude's Hospital	Pharmacogenomics Project	Open database with sequencing data of over 800 pediatric cancer patients	https://www.stjude.org/research/pediatric-cancer-genome-project.html
Veterans Affairs	Million Veterans Project	Collection of genetic, lifestyle, military experience, and health records of 1 million veterans	https://www.research.va.gov/mvp/

Table 1: Sample of Current Omics Data Projects in the United States

As seen in Table 1, genomic programs are becoming ambitious in size and scope. The private sector is also contributing to available omics data. 23andMe for example has now published over 100 related omics studies on a wide range of disease from allergies to cancer from voluntarily obtained patient data (23andMe publications.). Personalized medicine will likely see significant advances in the future, and the early programs listed above should be studied to determine best practices for population genomics.

The same is true for omics data at NASA. While the GeneLab program does not currently capture astronaut data, GeneLab omics concepts may be transferrable to future astronaut omics initiatives. Astronaut omics data could be used for two main purposes.

First, astronaut omics data can be used to help answer research questions regarding space physiology. Researchers can use omics data to look for trends in astronaut gene mutations or expression that could help explain how human physiology changes in spaceflight. Second, as an astronaut omics dataset matures, there will likely be opportunities to either prevent or treat astronaut health issues using information garnered from the omics data set.

Chapter 2 Background and Literature Review

The following chapter discusses the organizational structure of NASA as it relates to human health and a proposed omics program. The current NASA position towards astronaut genetic data is also summarized. This chapter ends with a detailed review of key considerations for omics programs as derived from current omics literature.

ORGANIZATION DESCRIPTION

As an organization, NASA has divided human health management across functional tiers. The overarching entity is the Human Health and Performance Directorate, which is subsequently broken out into divisions and then branches. Figure 1 below provides an illustration of the Human Health and Performance Directorate and provides a potential function of various organizational entities as it relates to an omics program.

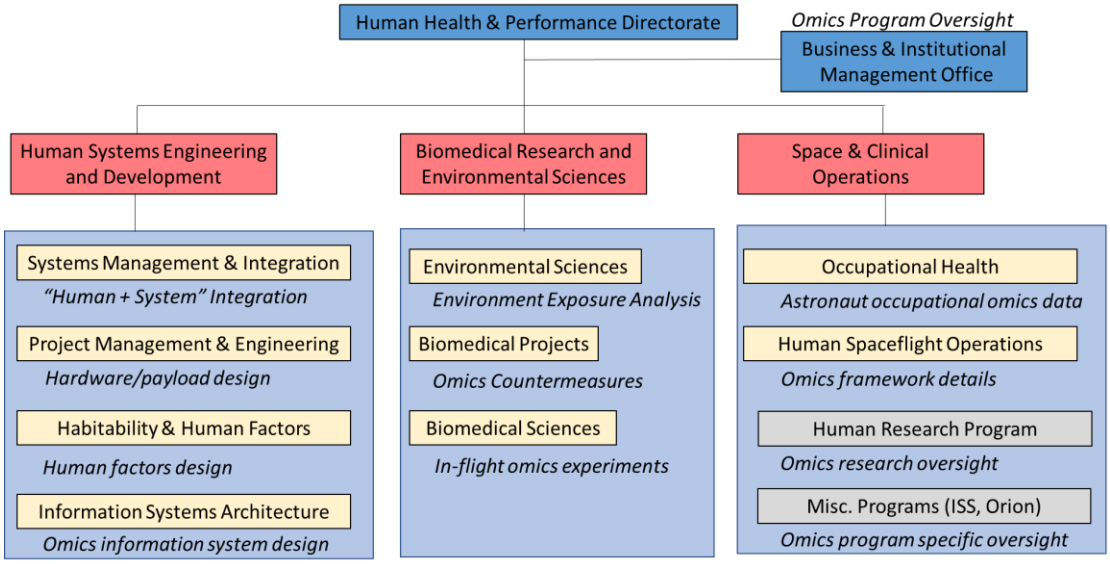

Figure 1: Notional Relationship between NASA Omics Program and the Human Health and Performance Directorate

(Adapted from https://www.nasa.gov/hhp/org)

As can be seen in Figure 1, the Human Health and Performance Directorate covers many aspects of health management within NASA. The omics roles presented here for each branch are notional, and it is important to note that programs at NASA can be spread across several divisions and branches. The management/business oversight of the directorate is handled by the Business & Institutional Management Office, and this would be the likely office to develop an omics program business development plan.

The Space and Clinical Operations Division covers the clinical management of astronaut health and human research of astronauts. The Occupational Health Branch would determine if any preventive actions are needed based on both publicly available omics data sets and from astronaut omics data. The Human Spaceflight Operations Branch, the organization where flight surgeons reside, will need to determine how omics information can and cannot be used to personalize healthcare for each astronaut.

The Human Research Program (HRP) is another important entity that will play a large role in an omics program at NASA. HRP is tasked with "predicting, assessing, and solving" spaceflight issues that astronauts encounter (Mars, 2018a). HRP thus serves as the foundation for human spaceflight research at NASA and would play a significant role in guiding astronaut omics research.

In the Human Systems Engineering and Development Division, an integral branch for an omics program would likely be the Information Systems Architecture Branch, which would play a role in defining the omics database and health record architecture. The Systems Management & Integration branch would help define how information learned from omics astronaut studies should be used in future vehicle design. The remaining two branches of the division, the Project Management & Engineering Branch, which manages development of crew health hardware, and the Habitability & Human Factors branch may not initially have large roles in an omics program.

The Biomedical Research and Environmental Sciences Division would work on flight-related aspects of an omics program. The Environment Sciences Branch, in charge

of spacecraft environmental monitoring and tracking, may need to look at how environment exposure data can be stored in a way that interface with omics software (to be discussed later in greater detail). The Biomedical Sciences Branch would be involved in development of hardware for in-flight omics data collection. The Biomedical Projects branch assists with the development of spaceflight countermeasures. However, the information needed for omics-based spaceflight countermeasures will likely require years of maturation of an omics spaceflight dataset (Lewis, 2017a).

From a programmatic standpoint, NASA has created an internal policy directive for the handling of genetic data within the astronaut population, NPD 7170.1, "Use of Human Research Genetic Testing." (*Use of human research genetic testing.*2018) The key points of the policy directive are that genetic data cannot be used by NASA for the purposes of astronaut selection or for the selection of astronauts to specific mission(s). NASA needs to maintain privacy of any genetic data and cannot release whole genomic data (an entirely sequenced gene set) or data that is personally identifiable without express consent of the astronaut. Furthermore, genetic information cannot be used to determine the type of training an astronaut can and cannot do. NPD 7170.1 dictates that omics research data and omics clinical data need to be maintained in separate databases, and that transfer of omics research data to a clinical database can only be done through an explicit request from an astronaut. The policy also defines NASA planned uses for omics data and defines when uses require a voluntary use. This information is summarized in Table 2 below. Astronauts need to provide voluntary consent when genetic data is obtained for mission design purposes. Conversely, there is no specific requirement for voluntary consent if the genetic data will be obtained for direct care of an astronaut. This delineation seems reasonable so long as astronauts are informed in the clinic prior to the collection of any omics related lab work.

Genetic Data Use	Voluntary Data Collection?
Medical risk identification for space exploration	Yes
Development of engineering requirements for spacecraft	Yes
Mitigation of space hazards	Yes
Development and assessment of space health risk countermeasures	Yes
Occupational Surveillance	No
Tailoring of individual countermeasures	No
Informing Clinical Care	No

Table 2: NPD 7170.1, NASA Policy Directive, Use of Human Research Genetic Testing: Agency Genetic Data Uses and Consent Requirements.

Note: genetic data use terminology taken directly from NPD 7170.1 (*Use of human research genetic testing.*2018)(Use of human research genetic testing.2018)(Use of human research genetic testing.2018).

As mentioned previously, NASA does not yet have a common framework for the collection and storage of omics data. Omics data sets have been collected in the past on NASA astronauts for specific studies, and storage of this data has been managed at the principal investigator's lab (whether at NASA or elsewhere). This is the case for the NASA twin study, where multiple investigators have performed experiments in genomics, transcriptomics, epigenomics, proteomics, metabolomics, and integrative omics (Mars, 2018b). The results of the NASA twin studies are pending, and an initial summary publication is expected later this year. Information gleaned from these omics studies could help scope what types of omics analyses will be most needed in the future.

SCIENTIFIC BACKGROUND AND RATIONAL

Omics technologies have advanced, and costs are decreasing. While technology has increased our ability to detect changes, it also becomes more challenging to define which changes are clinically significant. This section describes how large data sets are being refined and structured to translate study results to clinical applications.

"Big data" as applied to omics can be interpreted two different ways: analyzing an individual's genes for variants or analyzing electronic health records for a diagnosis with

the same variant(s). In both instances, copious amounts of data are being assessed. For the individual, thousands of genes or proteins may be included in an analysis. To analyze exposures and disease across individuals, a researcher may be looking at thousands to millions of health records. Both require complex algorithms and advanced computational biology technologies. For NASA, it will be important that a framework is established for how an astronaut's omics data is created so that there is consistency across astronauts for comparison. NASA will also need to determine what public databases will be utilized to discover new variants of interest for spaceflight studies, and how decisions will be made on those variants that are clinically useful for prevention and treatment of astronaut conditions.

Variant identification can be challenging in omics studies. While omics technologies have advanced, omics analysis software is often lagging. Rigorous software testing and documentation standards, which is often lacking, needs to be incorporated to improve the validity of omics analysis software (Doig, Papenfuss, & Fox, 2015).

Naming schemes are important when it comes to variant identification. First, the American College of Medical Genetics and Genomics (ACMGG) and the Association for Molecular Pathology (AMP) recommend using the term variant over mutation to avoid any suggestion of pathogenicity. Second, the ACMGG/AMP recommend classifying variants according to the following scheme: pathogenic variant, likely pathogenic variant, likely benign variant, benign variant, or variant of uncertain significance, respectively (Richards et al., 2015). These labels should be applied for each variant within a variant database.

So how is it determined whether a variant is of clinical utility? Matching genotype and phenotype can be challenging. As is often the case in Whole Genome Sequencing studies, incidental variants without clinical significance are often found. First, the comparison of omics data between organizations can be difficult. Different processing algorithms in whole genome sequencing software can lead to conflicting results between studies, and furthermore clinical lab tests between centers can differ (Andrews & Luikart,

2014; Kohane, 2015). Second, there is also the challenge of determining a therapeutic association between genetic markers and therapy. As pointed out by Ionnadis and Khourry (2018), in a recent National Cancer Institute effort to match therapy to molecular markers in tumors, only 2.5% of patients could be matched to an individualized treatment plan. This demonstrates that identification of a tumor marker does not necessarily prove causality or guarantee the existence a successful therapy especially with cancer cells that can mutate frequently, (Doig et al., 2015; Ioannidis & Khoury, 2018).

Innovative programs are being developed in the omics community to match genotype and phenotypes. For these programs to be successful, there needs to be a standard way for computers to read phenotype data. An example organization that is helping to define phenotype standard nomenclature is the Human Phenotype Ontology (van Karnebeek et al., 2018). Ideally, electronical medical record companies and researcher organizations including NASA should adhere to standards set forth in such organizations.

It is certain that over time the scientific and medical community will increase efficiency at generating clinically useful information from omics studies. Beyond the challenges of identifying pathogenic variants, there are also inherent challenges in storing the data. From a clinical perspective in the US, EMRs frequently do not store sequencing data in a standardized database but instead are stored off in document form (Kuehn, 2017). Fortunately, there are clinical databases that are being generated with variant information. The Human Genome Variation Society is the acknowledged group that sets standards for mutation database nomenclature (Doig et al., 2015). ClinVar is a database that annotates gene variations. At present, there are over 400,000 unique variants with identified interpretations. This means that the variant has been correlated with a downstream function of the gene but does not necessarily mean that there is a corresponding therapy. In fact, only 23 variants are currently associated with clinical practice guidelines, but this will certainly grow over time as the omics field expands (He, Ge, & He, 2017).

Large-scale omics studies, like the NIH-funded Precision Medicine Initiative (PMI) mentioned in Table 1, are only likely to benefit the greater body of omics knowledge by following accepted data standards. PMI is looking to employ similar data frameworks from existing large-scale omics programs such as the Million Veterans Program and the UK Biobank. As the PMI program states, the hope is that maintaining data uniformity will provide "a global determinant of human health and disease." (Hudson, Lifton, & Patrick-Lake, 2015)

Omics programs can also be strengthened by storing original patient samples in a biorepository. The National Heart and Lung and Blood Institute (NHLBI), an NIH center, has been performing whole genome sequencing of patients with heart, lung, blood, or sleep disorders. The center has sequenced over 90,000 patients to date. To support their efforts, the NHLBI has created an open access biorepository consisting of blood, urine, DNA, and RNA. Samples are stored with identifying information removed, and the samples can be requested by investigators for future studies ((National heart and lung and blood institute: Precision medicine initiative.; *The NHLBI biorepository: Guide to building biospecimen collections for study and future research use.*2016)(National heart and lung and blood institute: Precision medicine initiative.; The NHLBI biorepository: Guide to building biospecimen collections for study and future research use.2016).

Omics studies for cancer patients have been at the forefront for developing individualized treatment regimens. Researchers and clinicians are also having success with non-cancer diagnoses as well. Programs like the NIH-funded "Integrating Genomics in Practice Network" or IGNiTE program are developing pilot programs for personalized medicine. For example, IGNiTE is piloting the use of genomics in the treatment of hypertension and in pharmacogenomics in primary care settings (Kuehn, 2017). Hoffman et al led a study looking at genetic markers of dyslipidemia in nearly 95,000 patients from Kaiser Permanente. Researchers identified 177 variants associated with LDL and triglyceride pathways (including 15 mutations newly identified within the study) that were

predictive of what age patients started medication for high cholesterol. The identified variants are mainly tied to expression of proteins that deal with processing of cholesterol in the liver (Hoffmann et al., 2018).

In a study by Mates et al, researchers analyzed genes related to sudden cardiac death in patients and found a 4-5% prevalence of related variants in patients who experienced arrythmias (which is the leading cause of sudden cardiac death in those under age 35), long QT syndrome, or dilated cardiomyopathy. The authors suggested that the variant prevalence was high enough to consider genetic profiling of such patients (Mates et al., 2018). While these cardiac abnormalities are not likely to be seen in a healthy astronaut population, this study demonstrates how targeted studies can be used to detect variants for a disease.

As a pharmacogenomics example, whole genome sequencing is now providing an alternative to identifying and treating bacteria with resistance to first-line drugs. WGS has helped to establish thousands of gene mutations that can confer antibiotic resistance. The current challenge is how to make clinical correlations with the identified mutations. For common pathogens such as S. Aureus and E. coli, research has shown that WGS can identify drug resistance susceptibility with 95% sensitivity and specificity (Punina, Makridakis, Remnev, & Topunov, 2015).

Chapter 3 Methods

In terms of the literature review, PubMed was searched using the following terms: bioinformatics, translational genetics, precision medicine, genomic data analysis, million veterans program. Articles were selected for review if they discussed considerations for the establishment of an omics programs. Articles were also selected for review when the study looked for genetic susceptibilities in individuals prior to the development of disease. Given the rapidly accelerating pace of omics technology, the literature search was limited to those articles published within the past 5 years.

Additional genomic programs to research were identified through the literature and through subject matter experts (SMEs). Omics program information was obtained from website publications from the respective programs. An understanding of current NASA perspective on omics was gained through participation in the GeneLab working group and a steering committee for an Omics Conference being planned for Fall 2018 at NASA JSC.

NEEDS ASSESSMENT

It is important that NASA consider starting a formal omics program now so that there will be baseline astronaut omics data sets that can be used for future comparison. Since humans accumulate genetic variations over a lifetime from environmental exposures, it would be valuable to have genetic studies available for each astronaut before subsequent missions. Current variant information in public databases is largely derived from clinical populations that are less healthy than astronauts. While certain variants may have significance in the setting of disease, the same variant may not have clinical meaning when present in a healthy individual (Christensen, Dukhovny, Siebert, & Green, 2015). By collecting omics data on the healthy astronaut population prior to flight, we will have increased ability to interpret any future variants that are identified.

PROGRAM DESCRIPTION

An astronaut omics program at NASA will be distributed across the divisions and branches of the Human Health and Performance Directorate as described in Chapter 2, Figure 1. The program should be supported by creating an official omics team with permanent designees from HRP, Human Spaceflight Operations, Information Systems Architecture, and the Business and Institutional Management Office. NASA could also benefit from including a permanent position for an omics team consultant(s) from industry. The initial omics program team should ensure that the following questions are answered:

- What is the initial omics data set that should be captured on each astronaut?
- For each type of omics data (genome, transcriptome, etc), how often should omics data be collected on astronauts including during training and assigned missions?
- What software analysis packages will be used to analyze the omics data?
- What data architecture will be used to store variant information?
- What lab equipment will be used for omics analyses (internal NASA equipment or from external contractor)?
- How will NASA keep current on clinically actionable variants as defined by the greater genetics community?
- How should NASA define variants of interest for spaceflight?
- What variants should be used to guide preventive health measures?
- How will genetic counseling of results be provided? (This is required by NPD 7170.1)

The above questions fall into 3 main categories: astronaut omics data collection, software and database development, and determination on clinical uses of variant data. Although the last 4 questions are specific to variation in genetic structure, the same

questions could be applied to results of analyses in transcriptomics, proteomics, metabolomics, etc.

LOGIC MODEL

Figure 3 below provides a logic model for creating an astronaut omics program at NASA.

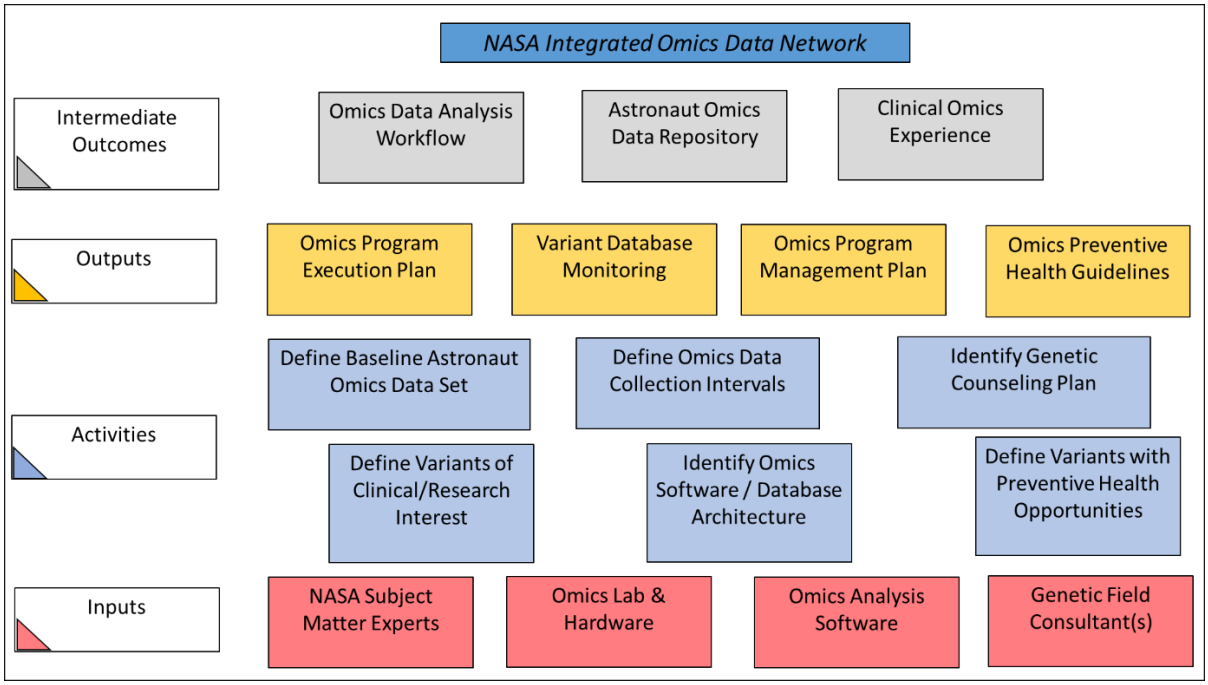

Figure 2: A Logic Model for a NASA Integrated Omics Data Network

The main inputs to the logic model are experts within NASA and from the genetics and bioinformatics fields. Additionally, omics analysis software represents not only the information technology but also bioinformatics personnel that will be needed to select or design omics analysis algorithms with sensitivities suitable for a healthy astronaut population. The activities above reflect determining answers for the questions stated in the above section. Outputs of the program startup phase will be a NASA-approved program and management plan, a methodology for monitoring and identifying clinically significant variants in the public domain and within NASA and plans for advising preventive measures

to astronauts for those markers that portend a higher chance of future disease development. Once the program is up and running, the intermediate outcomes will be a repository of astronaut omics data, preliminary methodologies for identifying variants of interest, and initial experience utilizing astronaut omics data on a case-by-case basis. Once the program matures, NASA will have developed a fully integrated omics data network for the astronaut population.

Chapter 4 Results

As demonstrated in this plan, an omics data program is complex, but can be built up slowly over time. Stakeholders for the plan need to be identified early and buy-in achieved prior to program initiation. Stakeholders should also be educated on omics data advantages and limitations so that teams can weigh risk and benefits of program features along the way. Ioannidis and Khoury pointed out that results from big data studies may be as straightforward to interpret compared to more conventional statistical comparisons such as relative risk and odds ratios (Ioannidis & Khoury, 2018). It is therefore crucial to provide education to any scientific community including NASA so that the advantages and disadvantages of omics technology is appreciated and understood.

Implementation Plan

In the near-term, this program proposal should be briefed to the Spaceflight Operations Branch and to HRP. HRP management and the newly identified omics research coordinator should be included in the briefing. Additional stakeholders at the division and directorate level can be briefed once branch chiefs have provided inputs to the program plan.

The cost for initial program implementation will need to be determined, but key characteristics of the desired program will need to be determined first. NASA will need to determine what omics program activities will be performed in-house and which activities will be handled by outside entities. Furthermore, to save costs, it is possible that a subset of omics fields might be used as a pilot for the program. It might also be possible to develop technology to save preserved astronaut samples that can be used later for new omics analyses as they are introduced into the overall framework.

NASA should consider identifying variants of interest for spaceflight at program onset. Schmidt et al have provided recommendations in several areas for further potential omics investigations. These include investigations into genes and molecular pathways that will make astronauts more susceptible to performance or health decrements during or after spaceflight. Schmidt highlights potentially examining CYP450 pathways for drug metabolism differences amongst individuals, MTHFR gene mutations for its potential relationship to one-carbon metabolism deficiencies and visual changes observed in spaceflight, and several gene variants that could potentially increase the risk for kidney stone development (Schmidt & Goodwin, 2013; Schmidt, Goodwin, & Cuttino, 2017). The forthcoming results of the NASA twins study will also likely provide areas of interest for future omics investigations.

Program ethics will also need to be considered. As discussed earlier, the NASA policy directive on genetic research protects astronauts from career or mission determinations from omics investigations. Since omics data collection for individual astronaut care has not been identified as needing voluntary consent, there may be ethical concerns related to performing omics testing for reducing health risks. Specifically, any omics test on an astronaut should be performed thoughtfully as incidental findings can lead to an increase in follow-on medical testing and/or procedures.

As is the case for any human research study, investigators will need to obtain approval from the NASA IRB prior to initiating any studies that will generate or make use of existing astronaut omics data. As directed by NPD 7170.1, omics research data will need to be stored independently from clinical omics data. It would be of value to design the omics research database with the same schema as the clinical omics database to facilitate approved data transfers.

NASA should consider if astronaut samples used for omics should be saved into a biorepository. NASA in fact does already save blood and urine in certain cases, but NASA may want to consider expanding the biorepository in the future. The setup of a

biorepository should take into consideration what assays will be performed. Each assay would have a set of defined data collection parameters that could be included in the omics database. Study investigators should also define appropriate biorepository sample amounts in advance such that current, ancillary, and future studies will have access to samples.

EVALUATION PLAN

The main markers of success for the program should be based upon the discovery of actionable variants and corresponding reduction of astronaut health risks. For NASA's purposes, an actionable variant would be either one that mitigates undesirable clinical effects from prior exposure to the space environment or a variant that can be used to implement countermeasures during spaceflight.

The omics program should also be evaluated using several objective markers as well. The main marker would likely be time to implementation of program milestones such as completion of the astronaut omics repository. Publications related to discoveries from the omics program could also be tracked. Program costs should also be evaluated regularly.

Chapter 5 Proposal Summary

The omics field is expanding rapidly and offers the potential for a greater understanding of the human body and increasing opportunities for personalization of medical care. NASA should consider early adoption of an omics framework so that the organization is better positioned to support omics studies on astronauts in the future. Just as search engines became critical to finding the right knowledge, omics analysis tools should be integrated and customized for NASA workflows to find variants that could be impactful for astronaut health in the future.

EXPECTED OUTCOMES

As has been discussed throughout this proposal, one of the main expected outcomes of this program is that there will be standard methods by which NASA will collect, process and store astronaut omics data. An omics program will improve our understanding of how the space environment alters biologic pathways. We can then also determine if the space environment imparts any lasting effects on genes and gene expression. Omics can also help guide clinical care for astronauts in circumstances where a diagnosis uncertain. For example, a flight surgeon's decision on whether to treat or not treat an astronaut's benign lesion with unknown significance could be guided by an omics profile.

STRENGTHS AND LIMITATIONS

Strengths

NASA's small astronaut population can be an advantage to omics research. The cost to perform and manage omics studies on the astronaut population will be lower than

many clinical trials and research initiatives in industry. For example, as noted by He et al, the storage size required for one whole genome is currently 100 gigabyte (He et al., 2017). Given that the NASA astronaut population is low, NASA will be at an advantage for not having high data storage requirements in comparison to large studies with a million or more participants. Another strength of a NASA omics program is that the astronaut data library will be representative of a highly-selected and fit population. This will be a unique environment to analyze gene function within since most other omics studies will be focusing on unhealthy populations. Astronauts may also more commonly come to NASA with the sequenced genomes. Astronauts will likely be asking the agency about how their genes may be affected by spaceflight, and building up expertise now will help prepare NASA for this eventuality. Lastly, the agency can use omics results to help refine what medical supplies, equipment, and pharmacy will be necessary to specific crew members on specific missions.

Limitations

Due to the low number of individuals within the astronaut program, accumulation of knowledge and experience utilizing omics specifically for spaceflight mission applications would be slow. NASA could consider working with international and commercial partners to develop the omics framework to help increase numbers. Additionally, the ratio of expense to number of individuals benefiting from NASA omics technologies will be low at first, but as mentioned earlier the NASA omics program could contribute invaluable benefits to the omics community at large. Another limitation is that it may not be as easy to leverage cloud computing for a NASA omics program due to data security concerns for the astronauts. While NASA may not be able to store data on public omics cloud systems, NASA may still be able to take advantage of omics data analysis platforms serviced on the cloud so long as streamed data can be adequately encrypted.

SUSTAINABILITY PLAN

The omics program will be sustainable at NASA so long as stakeholder buy-in is maintained from the scientific community at NASA. Continuous education of NASA staff on the benefits and limitations of omics data sets will be important. NASA will need to have a formal data structure, but also have the ability to change the structure over time. The same argument is true for omics data processing software.

NASA astronauts will also continue to have an incentive to make use of the omics program if there is adequate counseling to the astronauts on what can and cannot be determined from their omics profiles. It will also be difficult for clinicians to keep up with new omics information without assistance from software programs (Kohane, 2015), and purchasing omics clinical decision-making software packages may be beneficial for researchers and clinicians at NASA.

PUBLIC HEALTH IMPLICATIONS

As mentioned above, a NASA omics program would be unique in that it would provide an omics baseline for a population that is healthier than the general population. At present, omics studies mainly obtain variant information from the general population that has a preponderance of active illness or chronic disease. A NASA omics program could help guide how omics can be used to prevent or minimize illness at an early stage in astronauts. The resulting body of knowledge would be directly applicable to informing how omics data can be used to improve the health of the general population as well.

RECOMMENDATIONS

The main recommendation of this program proposal is that NASA should now begin to develop an agency astronaut omics data framework. The program should be

created with representatives from all levels of the Health and Human Performance Directorate. The program proposal presented herein should be refined into a full program plan that is approved by NASA. An agency approved plan should contain: a list of roles and responsibilities across the agency, a list of program products to be developed and maintained at NASA, a program timeline, budget, and management plan, and a list of defined relationships to external entities.

The NASA 2018 strategic plan lists four strategic themes/objectives to "accomplish its Mission and contribute to U.S. pre-eminence in space exploration, science, technology development, and aeronautics—all to the benefit of the American economy" (*2018 strategic plan.*2018)(2018 strategic plan.2018)(2018 strategic plan.2018). These are:

- An "enduring purpose of scientific discovery"
- A "push to expand the boundaries of human presence in space"
- A "broad mandate to promote the technologies of tomorrow"
- "Enable…capabilities, workforce, and facilities…to achieve <the> Mission"

A NASA program in astronaut omics will satisfy all these objectives. NASA has an opportunity to further enable our exploration missions with the use of omics, and at the same time NASA will enable discovery in the future technologies of omics and precision medicine.

Appendix A: Notable Omics Related Websites

Antibiotics Resistance Genes Database: https://ardb.cbcb.umd.edu/

Clinical Pharmacogenomics Guidelines: https://cpicpgx.org/guidelines/

Consortium Exchange of Genotype/Phenotype Data: matchmakerexchange.org

Global Alliance for Genomics and Health: https://www.ga4gh.org/ (defines genomic data file format)

Human Genome Variation Society: http://www.hgvs.org/

The Human Metabolome Database: www.hmdb.ca

Human Phenotype Ontology: https://hpo.jax.org/app/

Matchmaker Exchange: www.matchmakerexchange.org (matches genotype and phenotype)

NCBI ClinVar Database Submitters Summary: https://www.ncbi.nlm.nih.gov/clinvar/submitters/)

Sequence Read Archive Database: https://www.ncbi.nlm.nih.gov/sra

References

2018 strategic plan. (2018). ().NASA. Retrieved from https://www.nasa.gov/sites/default/files/atoms/files/nasa_2018_strategic_plan.pdf

23andMe publications. Retrieved from https://www.23andme.com/en-int/publications/

Andrews, K. R., & Luikart, G. (2014). Recent novel approaches for population genomics data analysis. *Molecular Ecology, 23*(7), 1661-1667. doi:10.1111/mec.12686

Caldwell, S. (2018). GeneLab: Open science for life in space. Retrieved from https://genelab.nasa.gov/

Christensen, K. D., Dukhovny, D., Siebert, U., & Green, R. C. (2015). Assessing the costs and cost-effectiveness of genomic sequencing. *Journal of Personalized Medicine, 5*(4), 470-486. doi:10.3390/jpm5040470

Doig, K., Papenfuss, A. T., & Fox, S. (2015). Clinical cancer genomic analysis: Data engineering required. *Lancet Oncology, The, 16*(9), 1015-1017. doi:10.1016/S1470-2045(15)00195-3

He, K. Y., Ge, D., & He, M. M. (2017). Big data analytics for genomic medicine. *International Journal of Molecular Sciences, 18*(2), 412. doi:10.3390/ijms18020412

Hoffmann, T. J., Theusch, E., Haldar, T., Ranatunga, D. K., Jorgenson, E., Medina, M. W., . . . Risch, N. (2018). A large electronic-health-record-based genome-wide study of serum lipids. *Nature Genetics, 50*(3), 401-413. doi:10.1038/s41588-018-0064-5

Hudson, K., Lifton, R., & Patrick-Lake, B. e. a. (2015). *The precision medicine initiative cohort program – building a research foundation for 21st century medicine.* ().All of Us Research Program. Retrieved from https://www.nih.gov/allofus-research-program/pmi-working-group

Ioannidis, J. P. A., & Khoury, M. J. (2018). Evidence-based medicine and big genomic data. *Human Molecular Genetics, 27*(R1), R2-R7. doi:10.1093/hmg/ddy065

Karczewski, K. J., & Snyder, M. P. (2018). Integrative omics for health and disease. *Nature Reviews.Genetics, 19*(5), 299-310. doi:10.1038/nrg.2018.4

Kohane, I. S. (2015). Ten things we have to do to achieve precision medicine. *Science, 349*(6243), 37-38. doi:10.1126/science.aab1328

Kuehn, B. M. (2017). Pilot programs seek to integrate genomic data into practice. *Jama, 318*(5), 410-412. doi:10.1001/jama.2017.7181

Lewis, R. (2017a). Human health and performance. Retrieved from https://www.nasa.gov/hhp/org

Lewis, R. (2017b). TREAT astronauts act. Retrieved from www.nasa.gov/hhp/treat-act

Mars, K. (2018a). NASA human research program. Retrieved from https://www.nasa.gov/hrp

Mars, K. (2018b). Twins study; research. Retrieved from https://www.nasa.gov/twins-study/research

Mates, J., Mademont-Soler, I., Del Olmo, B., Ferrer-Costa, C., Coll, M., Pérez-Serra, A., . . . Brugada, R. (2018). Role of copy number variants in sudden cardiac death and related diseases: Genetic analysis and translation into clinical practice. *European Journal of Human Genetics : EJHG, 26*(7), 1014-1025. doi:10.1038/s41431-018-0119-1

National heart and lung and blood institute: Precision medicine initiative. Retrieved from https://www.nhlbi.nih.gov/science/precision-medicine-activities

The NHLBI biorepository: Guide to building biospecimen collections for study and future research use. (2016). (No. Version 1.0).National Heart Lung & Blood Institue. Retrieved from https://biolincc.nhlbi.nih.gov/nhlbi_biorepository_guide/

Punina, N. V., Makridakis, N. M., Remnev, M. A., & Topunov, A. F. (2015). Whole-genome sequencing targets drug-resistant bacterial infections. *Human Genomics, 9*(1), 19. doi:10.1186/s40246-015-0037-z

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genetics in Medicine : Official Journal of the American College of Medical Genetics, 17*(5), 405.

Schmidt, M. A., & Goodwin, T. J. (2013). Personalized medicine in human space flight: Using omics based analyses to develop individualized countermeasures that enhance

astronaut safety and performance. *Metabolomics, 9*(6), 1134-1156. doi:10.1007/s11306-013-0556-3

Schmidt, M. A., Goodwin, T., & Cuttino, M. (2017). Personalized medicine in space flight, part II : Personalized precision medicine approaches. (Third ed., pp. 673-693) doi:10.1016/B978-0-12-803506-1.00064-4

Use of human research genetic testing. (2018). (No. NPD 7170.1). NASA Online Directives Information System: NASA. Retrieved from https://nodis3.gsfc.nasa.gov/lib_docs.cfm?range=7

van Karnebeek, C. D. M., Wortmann, S. B., Tarailo-Graovac, M., Langeveld, M., Ferreira, C. R., van de Kamp, J. M., . . . Boycott, K. M. (2018). The role of the clinician in the multi-omics era: Are you ready? *Journal of Inherited Metabolic Disease, 41*(3), 571-582. doi:10.1007/s10545-017-0128-1

Vita

Christopher Haas, MD, is an aerospace medicine resident with the University of Texas Medical Branch/NASA. He was born in Warwick, RI in 1977 to Linda Haas and Dr. Klaus Haas. He attended high school in East Greenwich, RI and Holderness School in Plymouth, NH. He graduated with a B.S. in Aerospace Engineering from the University of Colorado, Boulder. He worked for Raytheon Company as a systems and test engineer from 2001 to 2010 prior to starting medical school at the University of Colorado-Denver Anschutz Medical Campus. He completed a residency in internal medicine at Baylor College of Medicine in Houston, TX in 2017.

Dr. Haas has been the recipient of numerous awards including the Whitehead, Adler, and University Physicians Merit Scholarships at the University of Colorado School of Medicine. He was a finalist for the Jeffrey Meyers Young Investigators award in aerospace medicine in 2017, and he recently received the Dr. Douglas Endowed Scholarship for aerospace medicine.

He is board certified in internal medicine, and he volunteers and teaches regularly at a student charity clinic in Galveston, TX.

Permanent address:
UTMB Aerospace Medicine Residency, 301 University Blvd, Galveston, TX 77550-1110
This capstone was typed by Christopher Haas