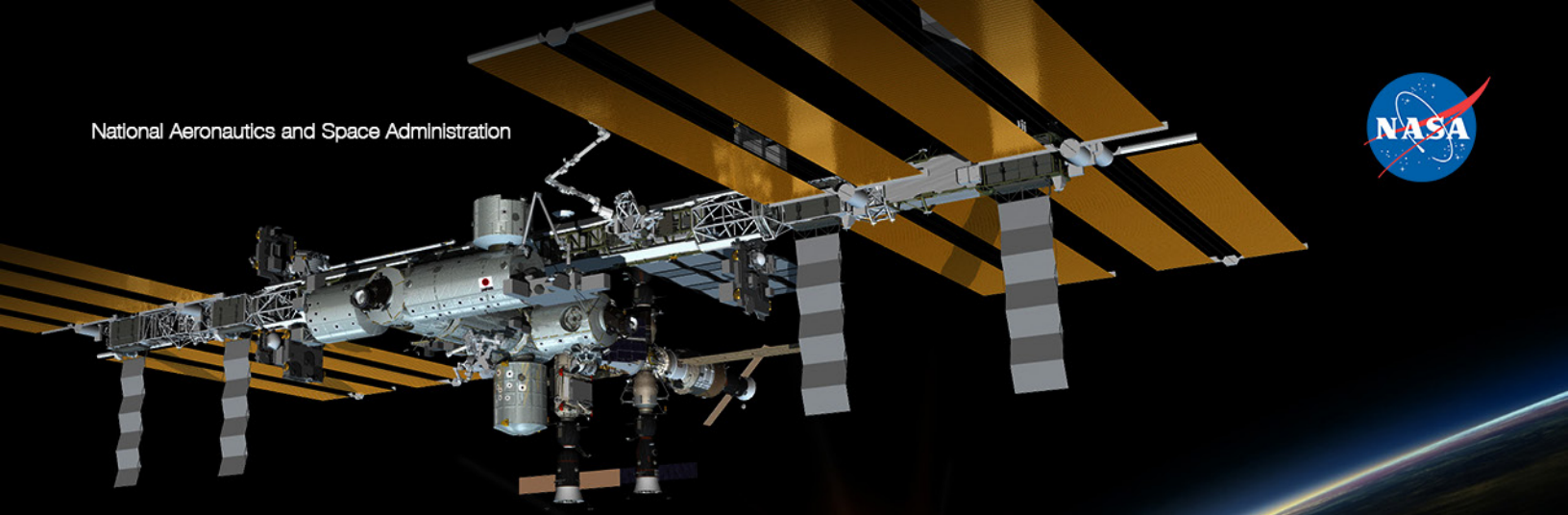
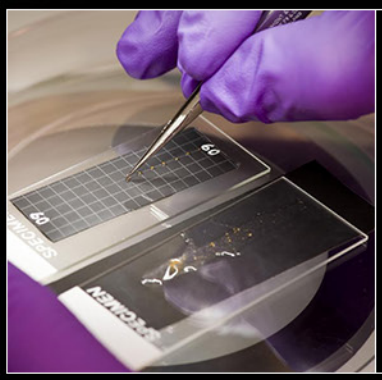
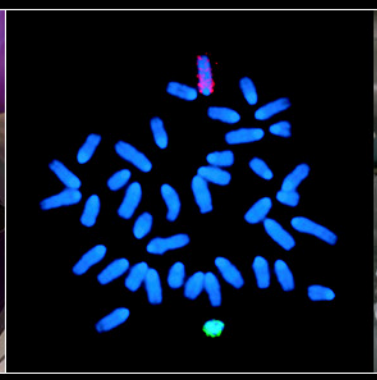
National Aeronautics and Space Administration

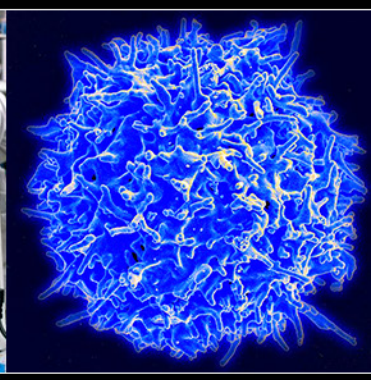


GeneLab

Expanding the Impact of Biological Research in Space







www.NASA.gov

Executive summary

Upon completion of the International Space Station (ISS), the National Aeronautics and Space Administration (NASA) has transitioned into a research utilization era focused on the advancement of science and technology development. In the early phases of this utilization era the numbers of space life science experiments have increased; however, the throughput of this research will ultimately be limited by logistics associated with spaceflight operations and crew-time availability. As a result, there is a significant demand from the scientific community for NASA-funded research opportunities on the ISS that cannot be met using traditional management tactics based on single Principal Investigator (PI)-led investigations. Given the factors that limit the scientific output required to address the significant biological problems required for human exploration beyond low-earth orbit, NASA will develop and implement a new multi-investigator approach based on high content bioinformatics analytics and with open science and data.

The 2011 National Research Council (NRC) Decadal Survey on NASA Life and Physical Sciences called for increased opportunities for multi-investigator spaceflight opportunities and greater use of genomic approaches to meet the needs of NASA researchers. To address these recommendations of the NRC Decadal Survey, the Space Life and Physical Sciences Research and Applications Division of NASA's Human Exploration and Operations Mission Directorate has initiated a transition to an Open Science architecture to increase research opportunities, and is developing the GeneLab Platform based on highly leveraged and integrated bioinformatics analytics.

Integration of various types of bioinformatics analytics such as genomic, transcriptomic, metabolomic and proteomic data, also known as "omics", requires a modern bioinformatics environment—consisting of a database, computational tools, and improved methods—that would facilitate sharing of unique life science data obtained from reference spaceflight experiments. These data sets will be made open to the scientific research community (academic and commercial) to encourage innovation and competition in the analysis and dissemination of the data. NASA will incentivize the use of ISS-derived scientific data by offering numerous grants to drive the research required to translate the GeneLab Platform data into information that will greatly increase the knowledge and discovery output from what would have traditionally been a single PI opportunity.

The GeneLab Platform will be based on an integrated multi-omics approach requiring the development of new bioinformatic technologies to deliver a next-generation system to approach spaceflight biological research. This will require that NASA work with outside partners to assure the GeneLab informatics system can interface with, and leverage other existing databases for genomic, transcriptomic, proteomic, metabolomics, and systems biology, ensuring the contemporary relevance of this platform. Ultimately the GeneLab Platform will enable community-driven reference experiments to generate standard, common, and open reference datasets to act as a powerful resource for scientific throughput and innovation. The platform will act as a repository to promote the development of new scientific hypotheses not previously conceived and novel experiments not previously envisioned. The data available to the scientific community from GeneLab reference experiments will promote new research and fuel ISS derivative research that will translate the limited number of flight opportunities into hundreds of investigations to advance knowledge and discovery. Based on the philosophy of open science, the GeneLab Platform approach will maximize the scientific return on investment and maximize the use of the ISS given the limited number of biological research opportunities in space. Open science will expand the number of researchers in the community, bringing new ideas and innovation to space biology research, while enabling discovery and advances for both NASA Exploration and Earth-based benefit.

D. Marshall Porterfield, PhD
Director, Space Life and Physical Sciences Division, NASA HQ

Table of Contents

Executive summary i

Introduction 1

Background 3

Technology Advances Enable GeneLab for Space Biology 3

Scientific Challenges from Decentralized Data 4

GeneLab: A Game Changing Approach to Space Life Sciences Research 5

Impact and Benefits of GeneLab 6

The GeneLab Goals 6

Goal 1: Develop an integrated repository and bioinformatics system for analysis and modeling 7

Goal 2: Enable the discovery and validation of molecular networks that are influenced by space conditions through ground-based and flight research using next generation omics technologies 9

Goal 3: Engage the broadest possible community of researchers, industry and the general public to foster innovation 13

Scientific Outreach 13

Training 13

Inform, engage, and inspire the public by sharing the GeneLab missions, challenges, and success 14

Goal 4: Strengthen international partnerships by leveraging existing capabilities and data sharing 14

The GeneLab Phased Approach 15

Strengths and Challenges 16

Space Life Science Research Strengths and Challenges 16

Data System Development Strengths and Challenges 18

Customers and Stakeholders 18

Performance Goals 20

Acknowledgements and Contributors 21

Introduction

The effects of gravity on biological systems have been recognized for hundreds of years. But only with the advent of spaceflight in the 1960s could biologists and physical scientists begin to understand these effects by exploring an environment where the constant acceleration of gravity could be controlled and effectively removed. In the earliest days of spaceflight, nothing was certain about the short- or long-term effects of the near absence of gravity, increased radiation, and other factors of the space environment. Although a good deal more is known today, much still remains to be discovered, explored, and learned about life in space.

At NASA, focused research on the effects of spaceflight on biological systems has helped elucidate the influence of radiation and microgravity on organisms from bacteria, to plants to humans. Particularly critical are the effects of microgravity (i.e., the near-absence of gravity), which are impossible to directly investigate on Earth. For example, the structure of cells depends on forces that confine the cell membrane within defined volumes. A change in gravity causes changes in the dynamic forces impinging on cells to alter their structure and, because function is dependent on structure, exerts profound effects on cell function.

By perturbing cell structure in microgravity, scientists can learn important facets of function not easily revealed under normal gravity. In fact, a basic premise of biological research is that inducing perturbations on systems is the most direct way to identify how the biological systems function. Significantly, biological experiments in space have already proven to be effective models for Earth-based human health–related issues, such as bone loss, aging, and immunology. These space experiments have also affected models of microbial virulence, developmental pathways in mammals, and growth mechanisms for plants. Therefore, new knowledge of gravitational effects in space not only benefits humans traveling into space, but also helps scientists understand the role genes play in human health and disease on Earth.

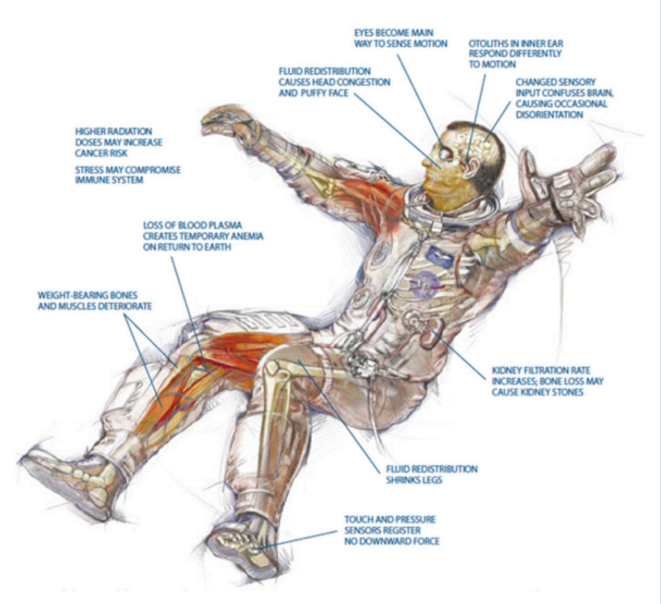
The fully operational International Space Station (ISS)—coupled with rapidly expanding and evolving technologies for biological investigations (high content and throughput)—is vastly increasing the opportunity to study and understand life in space. Through experiments conducted on the ISS, NASA is broadening scientific knowledge about the influence of gravity on living systems and leading efforts to develop advanced life science research capabilities. Specifically, NASA has begun the work to apply next generation bio-analytics for proteomics, transcriptomics, proteomics, and metabolomics to analyze spaceflight samples. These efforts will support a collaborative research environment for spaceflight studies that will closely resemble environments created by the U.S. Department of Energy (DOE), National Center for Biotechnology Information (NCBI), and other institutions in the study of cancer and environmental biology. To further enhance the collaborative research environment on board the ISS, NASA aims to collaborate with

Space Life Sciences Research Benefits Through the Years

From the earliest days of space exploration, NASA has engaged in life science research to better understand health and performance risks to humans. Continued research is still needed to understand the molecular underpinnings of the degenerative effects of weightlessness and ionizing radiation and to characterize the regulatory processes involved in responses to the stress of space travel. Addressing these safety considerations will not only enable development of countermeasures to protect astronauts, but will advance the understanding of fundamental biological processes to benefit health medicine on Earth.

Astronaut Buzz Aldrin descending the ladder of the Eagle lunar lander on the surface of the moon

Effects of Microgravity on the Human Body



Space biology research over the past 50 years has established the myriad effects of the microgravity environment on biological systems: deterioration of muscle tissue and loss of bone; fluid shifts, changes in cardiovascular function; lowered resistance to disease and impaired wound healing; and such central nervous system consequences as changes in visual acuity, sensory deficits, and the vestibular system, resulting in disorientation. The modern analytical techniques employed in genomics, transcriptomics, proteomics, and metabolomics provide an opportunity to understand the effects of spaceflight on these biological systems at the molecular level. The GeneLab Platform will expand participation of the life sciences research community in working to understand the mechanisms that drive these changes at a systems biology level and will accelerate the development of effective countermeasures.

Figure from the article by Ronald J. White, "Weightlessness and the Human Body," Scientific American, Sept. 1998, pp. 58-63.

all NASA and non-NASA research entities. As the ISS was nearing completion as a functioning research laboratory, Congress recognized this as an opportunity to diversify and increase users by designating the U.S. portion of the station as the nation's newest laboratory – the U.S. National Lab. While NASA still maintains the functions of the station, funds and implements NASA research and operations, this designation opened the door for quick access to ISS research for other federal agencies, academic and commercial institutions. To maximize use of the ISS as a research platform and enable utilization by non-traditional users, the Center for the Advancement of Science in Space (CASIS) was chosen in 2011 to manage the non-NASA activities on board the ISS. NASA and CASIS work synergistically to ensure the greatest possible scientific breath and reach of activities on orbit. Utilizing systems biology approaches and high throughput technologies will greatly enhance the collaboration between NASA and CASIS and increase the translational relevance of fundamental research.

NASA's next goal is to democratize access to the large amounts of data produced from next-generation life science investigations aboard ISS. Such access would enable investigators to participate in a transformative collaborative effort for interpreting the effects of spaceflight at the systems level and move towards predictive modeling. This model of open-access science promises to rapidly increase information-sharing and the rate of experimental iteration, leading to a faster pace of scientific discovery and the application of new knowledge to advance technology, medicine or, for NASA, space exploration. The impact and benefits of collaborative and predictive biology, as

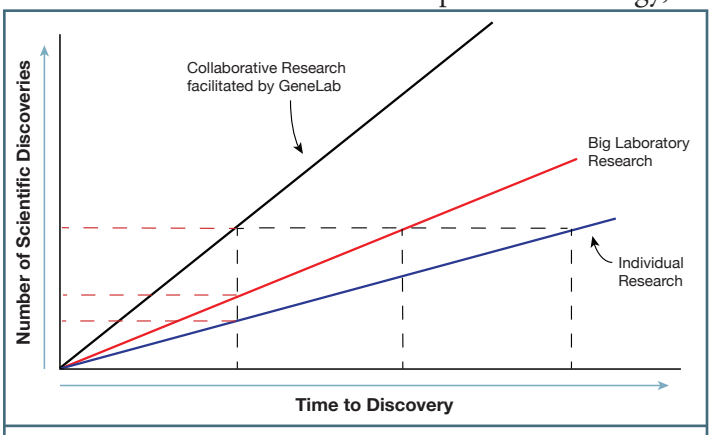


Figure 1: Faster Rate to Scientific Discovery Through Collaborative and Predictive Biology

The collaborative spirit and the GeneLab-enabled integration of experimental data from ground and flight studies will provide a powerful resource to accelerate the scientific advancements needed to improve knowledge translation to achieve predictive biology. Building on the wealth of data being generated across many laboratories, GeneLab will greatly accelerate the pace of scientific discovery over that which can be achieved based on isolated researchers focused on a limited scope of hypothesis-driven experimental design. The GeneLab Platform will act as a repository of space-derived biological observations promoting the development of new hypothesis not previously conceived and novel experiments not previously envisioned.

compared to those of traditional research driven by a single PI or single laboratory, are depicted in Figure 1.

Despite recent advances in space biotechnology and the increased availability of flight opportunities with the ISS, challenges to space biology research persist. When compared to typical ground-based laboratories and research, which can be as large as a campus and employ hundreds of staff to perform many experiments every day, the limited size of the ISS and limited access to and from space by way of supply missions limits the number of studies and experiments that can be performed at a given time. There is also a limit to crew availability and training for running biological protocols. Procedures that would be simple to perform in a laboratory on Earth can take much more time to perform in the microgravity of space, where everything must be pinned down or attached to a surface, and liquids must always remain in closed containers. Automated systems are utilized but it is very technically challenging to develop hardware and systems to support biological research which must function independently of gravity. There are also limitations to the ways biological samples can be processed, preserved, and stored.

All of these challenges add to the cost of experiments, making biological science in space an expensive endeavor. Significant care and effort must be taken to ensure every spaceflight experiment has strong scientific value and a very high likelihood of success. Moreover, NASA must maximize the value from every spaceflight experiment, as well as expand the breadth of research and the number of researchers who can benefit from every rare and precious spaceflight opportunity.

NASA's GeneLab Platform aims to optimize use of the ISS for biological science in space and greatly increase the number of scientists who can participate in spaceflight biological research. This science platform will encompass spaceflight experiments, sample collection, and advanced omics analytics, including an informatics infrastructure on Earth to organize the data and provide open access of the ISS research data (both spaceflight and analogous ground studies). By capturing more data from spaceflight samples and allowing broader open access to that data, GeneLab will increase the return on investment (ROI) for research conducted on the ISS and heighten the scientific impact of each experiment. The resulting ability for comparative analyses will help scientists around the world take a major leap forward in understanding the effect of microgravity, radiation, and other aspects of the space environment on model organisms. These efforts will speed the process of scientific sharing, iteration, and discovery to advance NASA missions and improve life on Earth. Partnership with CASIS on this platform will help increase the number of scientists reached while ensuring that application to Earth-based research, translation, and benefits are realized while enabling NASA's human exploration mission.

In addition, GeneLab will facilitate greater discovery by collecting baseline reference data that can be mined. It is understood that this effort by researchers around the world will result in new hypothesis-driven science that will lead to new, high-impact, scientific investigations in space.

Background

Technology Advances Enable GeneLab for Space Biology

The sequencing and annotation of the human, mouse, yeast, bacteria, fruit fly, and other genomes has significantly advanced our understanding of the genetic basis of biology and imparted a foundation for the new omics fields, which are now driving a new generation of systems level biology. These capabilities have also begun to foster the development of molecular medicine and diagnostics. For example, genomic studies can identify polymorphisms or mutations in genes that may predispose people to certain health risks or disease. This information can be used to preempt disorders before they manifest by prompting the design of lifestyle changes or therapies aimed at diminishing the onset of adverse effects of unhealthy genetics. Similarly, changes in gravity or exposure to radiation can prompt genetic changes that predispose organisms to predictable systematic changes. Information gained from such perturbations can prove valuable in the further growth and refinement of pathways involved in the development of countermeasures for human exploration and cures for human disease or pathogenicity here on earth.

Much of the predictive capabilities of genomic information are based on informatics analyses of different datasets. As in most computational systems, the more data available, the better the underlying models that give rise to predictive capabilities. This simple fact has ignited an explosion in genomic analysis of many organisms, including humans. Significantly, data garnered from lower organisms such as yeast can apply directly to health predictions in humans, and studies on gene and protein

expression in disease models in rodents can be used to design new drugs to treat human afflictions.

In addition, the enormous amount of data derived from the Human Genome Project and subsequent genome projects conducted over the last 10 years has provided the foundation for the development of novel bioinformatics technologies. These technologies enable integration of genetic data across most fields of science, from chemistry to medicine, to potentiate the growth of innovation and scientific discovery. Scientists worldwide in diverse disciplines, including NASA scientists, are using these resources for their everyday research. Researchers can now query and analyze data in multiple databases on the genomics, proteomics, metabolomics, and systems biology of a wide range of organisms and link the available data to their own experimentation. This ability facilitates the development of new hypotheses and the design of novel studies to enrich research programs and make unanticipated scientific discoveries that may relate to space or terrestrial biology.

Scientific Challenges from Decentralized Data

One problem facing the scientific community today is that enormous biological data sets are placed in multiple public and private databases that have not been designed for integration, hindering cross-talk of valuable experimental biology resources. Furthermore, the experimental design of studies used to gather the different datasets may vary from one scientist to another, precluding easy comparison of data across databases.

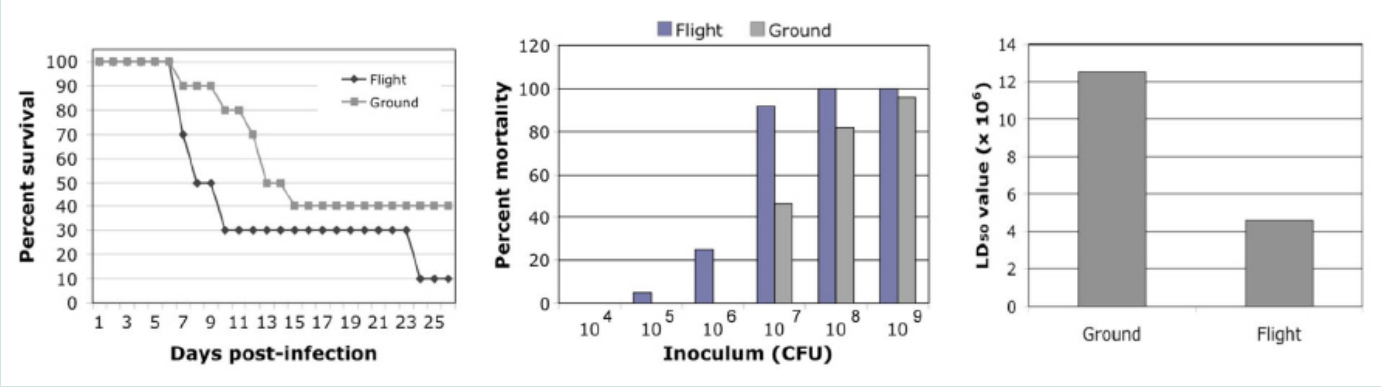
Space Research to Understand Human Health

Researchers have been intrigued by the reported 3–7 fold increases in *Salmonella* bacteria grown onboard the space shuttle compared to cultures grown on Earth under otherwise identical conditions. Understanding the molecular basis for this increased virulence could help safeguard astronauts from disease and lead to new treatments for bacterial infection.

To investigate findings from earlier Earth-based microgravity research, scientists conducted an experiment (Microbe) on the space shuttle (STS-115) using *Salmonella typhimurium* to perform molecular genetic and phenotypic analyses of the response of the bacteria in space. Analysis identified 167 genes that were differentially expressed by two-fold or more compared to Earth controls, and discovered 64 of 167 of these genes belonged to the Hfq regulon, suggesting that the Hfq could be a global regulator of the bacterial response to the space flight environment. Hfq is an RNA chaperone; facilitates mRNA translation in response to envelope stress, environmental stress, or changes in metabolite concentration; and is involved in promoting virulence in *Salmonella typhimurium*.

Earth experiments that infected vivarium caged mice with the flight *Salmonella* provided strong supporting evidence for Hfq participation in the spaceflight response of *S. typhimurium*. The mice infected with the flight bacteria had a shorter time to death and greater mortality rate at lower infection doses compared to ground controls, demonstrating that changes in gene expression due to spaceflight could affect the phenotypic behavior of bacteria.

Wilson et al (2007) *Proc Natl Acad Sci USA* 104(41):16299-16304
Wilson et al (2008) *PLoS One* 3(12): e3923



Infection of vivarium caged mice with flight or ground control S. typhimurium. From left to right: mouse survival, mouse mortality versus inoculation, and the 50% lethal dose (colony forming unit)

Chang et. al. (2012) J. Leukocyte Biology 92: 1133-1145

The lack of integrated standardized databases is a particular problem for NASA researchers utilizing the ISS and studying the effects of spaceflight on genomics, transcriptomics, proteomics, or metabolomics on different organisms. To understand how gravity changes biology, these NASA scientists need information that is complete, comparable, and easily obtained from relevant studies of similar organisms performed on Earth to compare to results obtained in space. Similarly, Earth-based scientists need access to data from spaceflights to study how definable perturbations in the environment, such as lack of gravity or chronic exposure to radiation, can affect the biology of well-studied organisms. For both sets of scientists, the inability to integrate databases is an immense roadblock.

GeneLab: A Game Changing Approach to Space Life Sciences Research

With more scientists using the ISS for their research and with continuous improvements in bioanalytic technology, NASA-funded researchers are generating nucleic acid sequencing data and RNA, protein, and metabolite expression data in increasing quantities. This results in a greater need for a common facility for bioinformatics system storage and informatics capabilities, as well as improved methods to share this unique data with the broader life sciences community. In fact, the 2011 National Research Council (NRC) Decadal Survey on NASA Life and Physical Sciences, conducted to help NASA identify and prioritize future research areas, called for greater informatics technologies to meet the growing needs of NASA researchers. The 2011 Decadal Survey pointed out a strong desire by researchers for a systems biology approach, regardless of their area of scientific interest. As examples, the following quotes from the 2011 Decadal Survey highlight the thoughts of NASA researchers on generating an open-access repository for biological information:

> "A strategy that would benefit all three of these research approaches would be the creation of robust databases that could be used by extramural scientists to address research questions. The databases could be populated retrospectively, with currently archived data from NASA-sponsored projects in the [Life Sciences Data Archive] (LSDA), archived data from flight medicine, and available long-term follow-up health data such as the Longitudinal Study of Astronaut Health (LSAH), with plans to expand the databases prospectively…" (12-12,13)

> "Because spaceflight represents an abrupt change in physical activity level, a systematic evaluation of genomic changes during spaceflight will advance the understanding of gene-environment interactions that influence chronic disease risk." (7-19)

> "…Because of the limited number of humans who undergo exposure to the space environment, maintaining an extensive and well-organized database and managing it as a resource to be shared with the scientific community has long been viewed as an essential step for scientific discovery." (12-12,13)

To address the needs identified in these reports, GeneLab envisions creating a centralized and integrated infrastructure for depositing, accessing, analyzing, modeling, and testing diverse datasets related to spaceflight studies. This comprehensive bioinformatics system will transform current spaceflight research analysis, which is performed in relative isolation from similar Earth-based work, to a higher level of creative science by facilitating the ability of the entire research community to work together.

This unique platform will be designed for integration with the many databases containing results of biological studies done on Earth to allow cross-talk of data and researchers. Such integration is particularly important because the effect of spaceflight on biology can only be understood if space research can be framed by the biology on Earth. Thus, a GeneLab Platform that integrates with informatics systems already in operation for Earth-based studies will help synergize space- and Earth-based research, increasing understanding of human physiology and medical science.

Providing researchers around the world the ability to store, share, analyze, and compare data from all biological, chemical, and physical studies (RNA and protein expression, epigenetic, metabolite data, etc.) will be an extremely powerful step in the quest to better understand the effects of spaceflight on living systems. These types of studies will help answer underlying questions of general and developmental biology in relation to normal and microgravity environments and how organisms adapt to the microgravity environment.

NASA's Space Life and Physical Sciences is charged with advancing fundamental knowledge of life in space by supporting scientific investigations on Earth and in space that will enable NASA's Human Exploration missions and benefit life on Earth. The ISS Program manages and operates the ISS and has the mission to make maximum use of the ISS for fundamental and applied research in

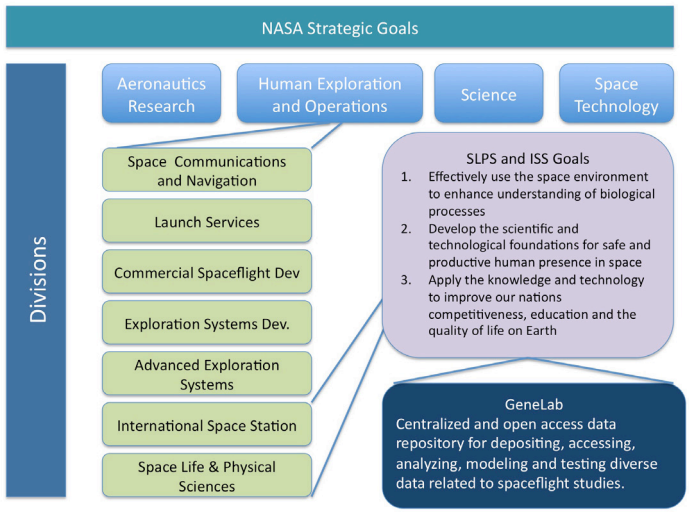


Figure 2. GeneLab Science Platform supports goals of NASA's Space Life and Physical Sciences Division and ISS Program. These NASA Programs are part of the Human Exploration and Operations Mission, one of the four major missions that help NASA meet its strategic goals.

space. The GeneLab Science Platform supports the goals of these two important NASA elements (Figure 2).

Impact and Benefits of GeneLab

The GeneLab Platform will greatly increase the availability and widespread use of NASA Life Science research results and maximize utilization of the few and costly opportunities to conduct experiments on ISS. The GeneLab Integrated Omics Informatics system will create a valuable resource that enables a number of NASA programs to better achieve their goals. The Human Research Program (HRP) and Space Life and Physical Sciences Research and Applications Division will benefit directly; but we also anticipate that the GeneLab Platform will provide significant opportunities to non-NASA users and the ISS National Laboratory CASIS. Both the technology to develop an integrated omics informatics system, and the data contained in this system will provide a wealth of new opportunities that can be translated and used outside NASA for Earth-based scientific and commercial benefit, making the ISS-derived research much more relevant to Earth applications. A list of the major benefits of the GeneLab Platform follows:

- **Maximize Return On Investment (ROI) for Life Sciences Flight Experiments and ISS Utilization:** The open-access, systems-biology spaceflight experiments of GeneLab will provide foundational science that maximizes ROI for rare and costly spaceflight opportunities by gathering knowledge at all levels, from molecules to entire pathways. By allowing open-access to all data generated, NASA will help to build the scientific ROI by allowing independent researchers to build next generation experiments with data they would not have otherwise. This in turn limits the number of parallel research studies needed for biological discoveries thereby decreasing the overall spend.
- **Maximize Use of Modern Bioanalytical Tools and Techniques:** Multiple omics datasets and an integrated data system enable scientists to interrogate ISS-derived samples using state-of-the-art high throughput genomics, proteomics, metabolomics, and bioinformatics tools.
- **Create a PI-Multiplier Effect for Space Biology:** The open-access data system greatly expands the community of researchers using ISS-derived data for investigations. Thus, ISS research investments will yield hundreds of follow-up investigations and generate next-generation hypothesis-driven research, as well as next-generation applied translation research.
- **Speed the Pathway to Discovery and Application:** The size and breadth of data and tools from GeneLab will allow researchers to discover emergent properties in the data for identifying and understanding pathways/macromolecules influenced by space stressors that can be targeted for pharmaceutical development.
- **Leverage both NASA and External Partner Strengths:** GeneLab brings together NASA's strengths in life sciences and big data analysis with commercial, government, and international partners through a scaled and iterative approach that capitalizes on existing databases, analytical tools, and biotech capabilities.

The GeneLab Goals

The knowledge gained by studying organisms from today's space biology experiments can lead to discoveries in fundamental research, which will drive the development of the next generation of spaceflight experiments. The GeneLab Platform will increase this iterative process and amplify scientific understanding, moving life sciences researchers to conclusions and solutions faster. The production of large datasets from GeneLab spaceflight experiments will raise

the ROI for ISS research by greatly increasing the number of investigations and the development of translational applications. The success of the GeneLab Platform will rely largely on its ability to meet the dynamic information needs of NASA-funded researchers, as well as of the larger, non-NASA scientific community and CASIS commercial partners. If successful, large and small biological research efforts will be greatly improved by their access to dramatically more data and robust analytical and modeling tools than previously available. Scientists will be able to integrate knowledge from their own research and also draw upon data generated from the entire research community. The open community science facilitated by GeneLab will thus advance systems biology research and accelerate the pace toward predictive medicine.

The GeneLab Platform has four goals:

1. Develop an integrated repository and bioinformatics data system for analysis and modeling
2. Enable the discovery and validation of molecular networks that are influenced by space conditions through ground-based and flight research using next-generation omics technologies
3. Engage the broadest possible community of researchers, industry, and the general public to foster innovation
4. Strengthen international partnerships by leveraging existing capabilities and data sharing

These goals focus on building the infrastructure for scientific exploration and analysis, providing open access to data, collaboration, and knowledge sharing, and enhancing public awareness for such a site to encourage collaborative innovation. Each goal is described in detail below.

Goal 1: Develop an integrated repository and bioinformatics system for analysis and modeling

A GeneLab Data System will be developed to support an open-access data infrastructure for systems biology research relating to spaceflight. Later in the development and implementation phase, the integrated platform will also include an open-access, open-architecture software element for developing analytic tools. As a public and scientific community resource, GeneLab will provide the data, computational backdrop, and tools for integrating and analyzing large, diverse datasets generated by the scientific community to advance predictive understanding, manipulation, and design of biological models. The purpose of GeneLab is to enable users to integrate a wide spectrum of omics data, models development and refinement, and information to accelerate discovery. Powerful tools either within or accessed by GeneLab will enable users to analyze and simulate data to predict biological behavior, generate and test hypotheses, design new biological functions, and propose new experiments. The GeneLab Platform will differ from current informatics efforts by migrating research from multiple laboratories and projects into a single user-friendly computational environment focused on spaceflight research, along with reference experiments and corresponding ground controls.

The GeneLab Platform is uniquely capable of accomplishing this goal because it has access to all space-related research in both the public domain and through unpublished data. As such, it can facilitate the flow of data from spaceflight biological research into a common bioinformatics-system. Furthermore, the GeneLab Platform is predicated on the ability to select and implement biological research conducted in space through direct engagement of the stakeholders in the research community. As a consequence,

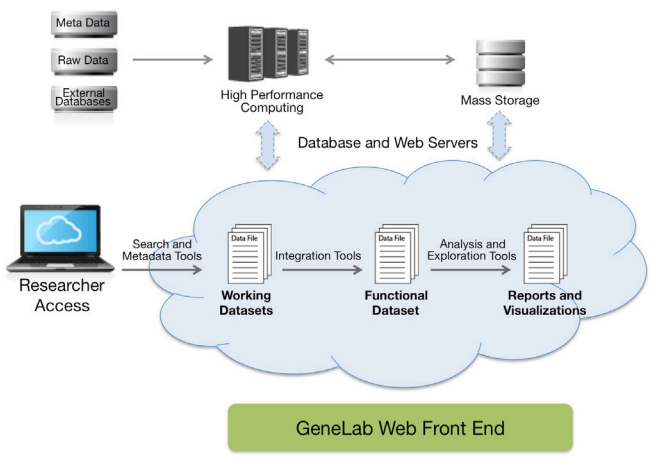


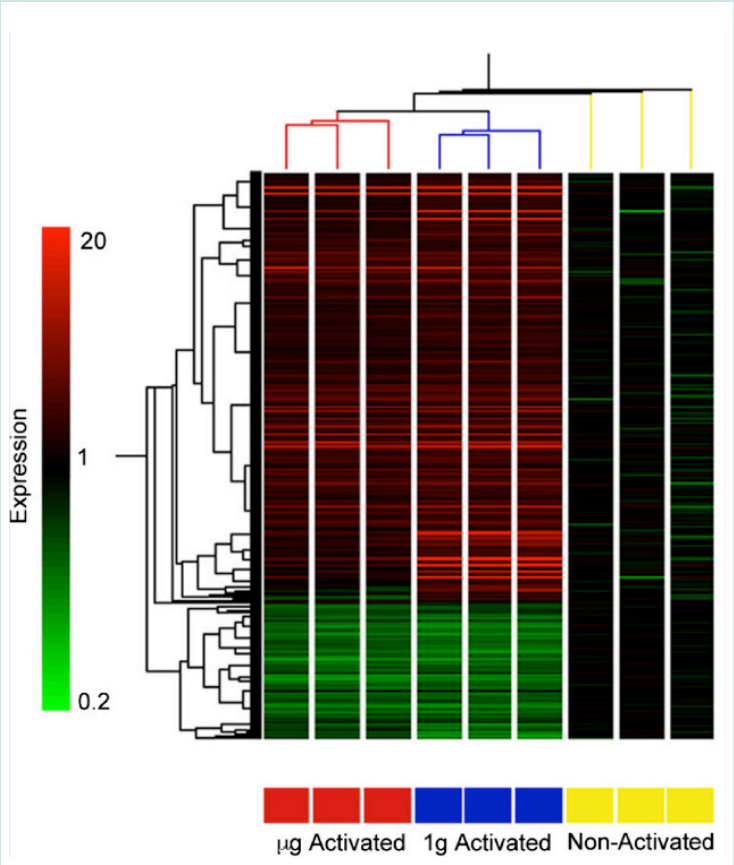
Figure 3: GeneLab Data System Architecture

Raw data is processed into working formats and retained for future algorithmic improvements. Datasets relevant to space biology from external databases and direct user submissions are incorporated into the knowledgebase to supplement GeneLab-generated data. Detailed experimental metadata conforming to reporting standards enables users of the web application to search and explore dataset relationships. Community-developed workflow tools take selected datasets and integrate them for statistical analysis and feature exploration. All of these processes rely on an underlying infrastructure of high performance computing, mass storage, servers, and high bandwidth data transfer.

The Role of Immediate Early Gene Expression in Loss of T Cell Activation in Microgravity

The microgravity environment causes impairment in T cell activation. Several spaceflight studies discovered that peripheral blood lymphocytes stimulated with Conclanavalin A (Con A) produced significantly lower IL-2, IL-2R, and IFN-γ. Hughes-Fulford and colleagues further examined this discovery by reducing Con A exposure time to 4 hours; they then identified the down regulation of 99 genes in simulated microgravity compared to 1g controls. Analyses of the expressed genes suggested the possible impairment of specific transcription factors, such as NF-κB, CREB, Ets-like protein-1, AP-1, and STAT, in turn suggesting that the affected factors might be associated with the early immediate gene expression cascade and pathways.

In their LEUKIN experiment on ISS, researchers reduced the treatment time of blood donor–isolated human T cells with Con A and anti-CD28 to 1.5 hours, and added RNA later to preserve the RNA. Microarray analysis of the RNA showed that 617 genes were differentially expressed compared to the on-orbit 1g controls. Further analysis found that 47 of the 617 genes had a two-fold or lower expression in microgravity compared to the on-orbit1g control and determined that the Rel/ NF-κB signaling was the key pathway inhibited by microgravity. Finally, a pathway analysis conducted on the 47 genes showed that they were all associated with the TNF, suggesting that TNF may be a key gravity-dependent effector function that is inhibited in microgravity.



Differentially regulated genes 1.5 hours after T cell activation in spaceflight microgravity and on-orbit 1g controls

After activation, the different gene expression pattern seen for the 3 different donors (columns identified by yellow boxes) becomes identical within a gravity treatment group (blue box columns: on-orbit 1g control; red box columns: microgravity). The experiment was run in ESA Kubik, which provides a centrifuge for an on-orbit 1g control and static holders for microgravity treatment.

The Rel/NF-κB pathway and transcription of immediate early genes in T cell activation are inhibited by microgravity. Chang et al. (2012) J. Leukocyte Biology 92(6): 1133-1145

it can perform standard reference experiments that will support a larger number of investigator research derived by using the GeneLab Platform. In the transition to full implementation, NASA will work directly with the scientific community to develop standardized experimental methods and data analysis systems to facilitate the integration of space- and Earth-related research findings. At the same time, GeneLab can publicize the availability of GeneLab data and GeneLab's unique bioinformatics technologies. A graphic representation of the GeneLab Data System architecture showing how researchers interact with GeneLab is depicted in Figure 3.

Some elements to be included within the GeneLab Data System follow:

- Raw and analyzed data generated by NASA-funded researchers, including sets from genomic, transcriptomic, proteomic, and metabolomic data collection
- Full GeneLab reference experiment data enabling comparisons of organisms exposed to the space environment with those remaining on earth
- Incorporation of a laboratory information management system (LIMS) to gather and store experiment details as metadata
- Intuitive and efficient user interface
- Clear data policy following the existing standards set

- Infrastructure to provide the ability for computational biologists to build upon existing analysis tools or develop new ones
- Community and collaboration portal for researchers to discuss and build hypotheses and experimental designs faster than if each researcher remained separate, as well as the opportunity for group development of software tools and algorithms for testing new hypotheses
- Linking of publications using GeneLab back to the website
- Metric, analytic, and feedback tools to gather public input for GeneLab Platform improvements and to aid in development and refining the next-generation GeneLab

Goal 2: Enable the discovery and validation of molecular networks that are influenced by space conditions through ground-based and flight research using next generation omics technologies

NASA's GeneLab Platform will provide a new community resource for predictive biology as it relates to the effects of space. It will combine diverse bioanalytical tools for genomic, transcriptomic, proteomic, and metabolomic (integrated omics) analysis and apply these to a wide spectrum of model species. The resulting datasets will then be housed in the GeneLab Platform, which will be based on next-generation computational tools that can analyze and simulate data to help visualize complex signaling networks, generate and test hypotheses, identify new biological mechanisms, and develop protocols for

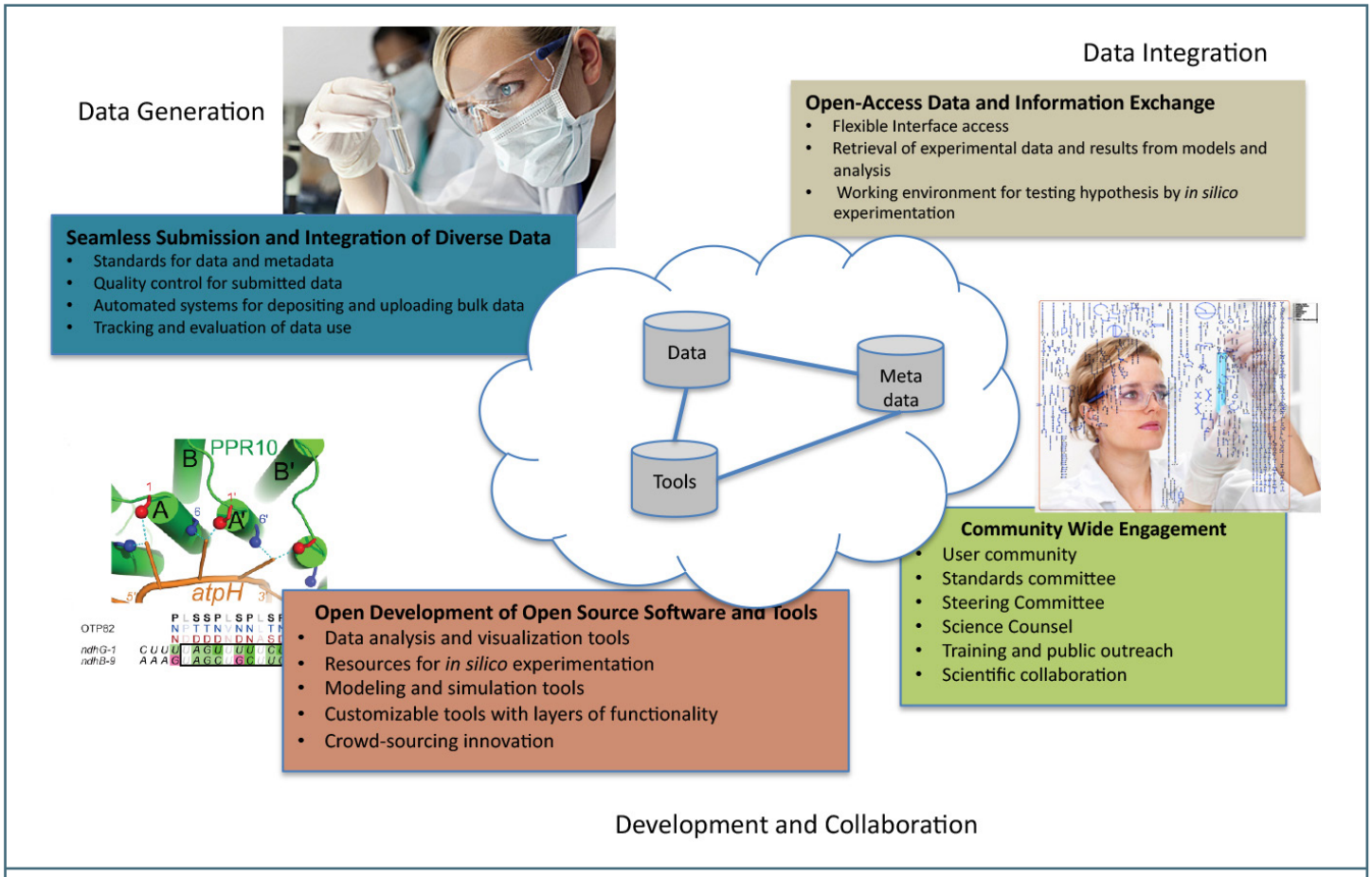


Figure 4: GeneLab's Cycle of Innovation Through an Open-Access Model

The GeneLab Platform will acquire space-flown samples, process the samples for data generation, and employ an open-access model to expand the scientific audience. Full data integration, community engagement, and availability of open source software in a single web interface will augment scientific research. This model will enable formation of novel hypotheses and follow-on space grants.

new experiments. The overarching objective is to provide an informatics platform that supports a systems biology approach and predictive medicine in a framework that does not require users to learn separate systems to formulate and answer questions spanning a variety of topics in life sciences research. To achieve this, researchers must be able to derive new observations from standard spaceflight reference experiments (as compared to ground controls) to conduct derivative ground-based experiments to translate the spaceflight data into knowledge (Figure 4).

Moving forward, GeneLab must consider multiple factors:

- Database size and accessibility
- How users from various scientific backgrounds plan to use GeneLab
- How to integrate new data with existing data

As a first step, GeneLab will utilize existing data held in public databases and engage currently funded NASA investigators selected to conduct experiments on ISS via the peer-review process. A requirement of funding will be inclusion of data from the biological studies in space in the GeneLab Platform upon publication of their results. Investigators can propose how, using their experimental system, the scientific community will benefit from the data received. The goal is to use currently planned missions to generate data for GeneLab, whether for testing protocols or for use as baseline data for other researchers. As the GeneLab Platform develops, other means of gaining data will present themselves, such as GeneLab-derived investigator flights, and GeneLab-dedicated flights proposed by the outside scientific community. The GeneLab data system will also be partitioned with limited data access to support DOD and commercial flights through the CASIS program where data security and intellectual property issues present themselves. The 5–10 year flight plan will be mapped accordingly and adjusted as needed to ensure flexibility for the program to benefit from opportunities as they arise. As the platform matures, GeneLab will have tested the entire process, from experiment and protocol setup to data generation and analysis, and compared the results obtained to previous results for the same purpose and organism. In this way, GeneLab will be able to effectively adjust methods to ensure more desirable outcomes for the researchers.

To maximize the value of data from life science experiments performed in space, and to take advantage of the remaining ISS research window, the GeneLab Platform will apply an iterative approach to the performance of spaceflight experiments, generating, saving, and sharing the data derived from these life sciences experiments in space (Figure 5).

The process for a dedicated GeneLab experiment begins with a competitively selected self-forming team of scientists. This team, termed the Science Definition Team (SDT), proposes an experimental design for a particular reference organism that will provide the broadest dataset for the greatest scientific return. The organism or relevant biological area selected will be dependent on the requirements for the specific mission. The GeneLab team will work with the SDT to develop and prepare the proposed experiment for flight, and the GeneLab team will work with the NASA spaceflight payload team to execute the successful GeneLab spaceflight experiment. After the launch of the experiment into microgravity, where crews implement the protocol and harvest tissues, the samples will be preserved for return from the ISS. The samples are then sent back to NASA for processing. For collaborative missions, GeneLab will support an existing space biology PI, on a non-interference basis, in exchange for sample tissues or cultures from that PI's experiment. More details of types of missions are described below. If samples are processed by NASA, the tissues will be prepared and DNA, RNA, protein, or other metabolites will be extracted. The prepared samples will then be sent to qualified omics centers for the generation of sequence, transcript, or protein data. Upon receipt of the data generated from experiment samples, the investigator or SDT will be afforded a period of time for data analysis and modeling. NASA will conduct data quality control tasks and format raw data returned from the omics centers into standardized, annotated information sets that can be readily searched and linked to spaceflight metadata. Once prepared, the biological datasets, as well as any analysis completed, will be made public through the GeneLab Data System web-based portal.

The process of data preparation and public release will be expedited to allow public access to the unique spaceflight results at the earliest reasonable opportunity. Open access to the data is expected to greatly expand the scientific audience and engage new PIs in the formation of novel hypotheses. Both new and original investigators can then submit informed proposals for subsequent research grants. These follow-on explorations will in turn generate new data that, when widely shared, will continue to drive a cycle of accelerated innovation.

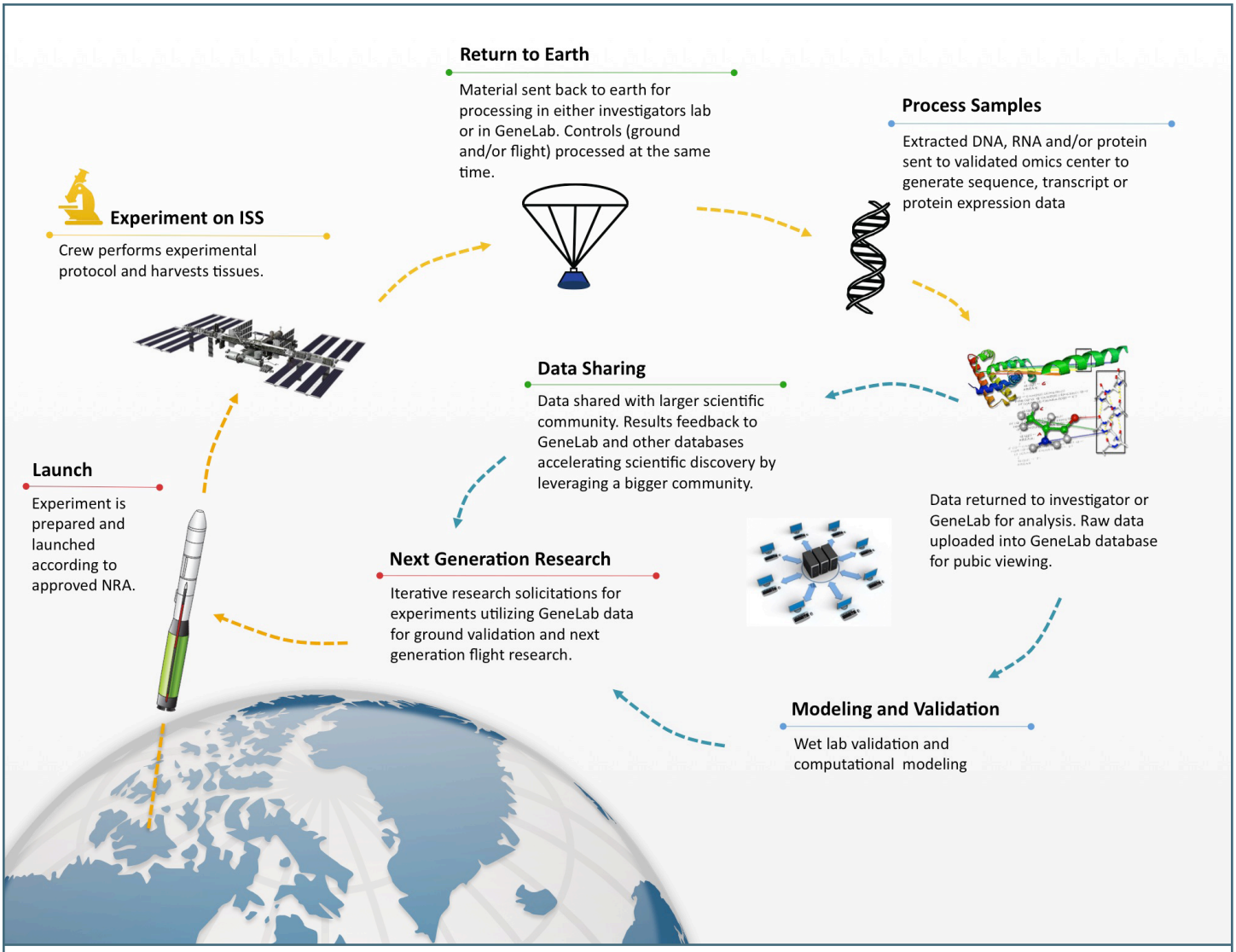


Figure 5. Leveraging space biology experiments to drive an ongoing cycle of innovation

To maximize the value of data from life science experiments performed in space, and to make the most advantageous use of the remaining ISS research window, GeneLab will employ an iterative approach performing spaceflight experiments, generating, saving, and sharing the datasets derived from biological research in space.

GeneLab will perform spaceflight experiments in two ways:

Fully Dedicated GeneLab Mission – For this mission, NASA will competitively select a Science Definition Team (SDT) to develop the requirements for the mission and determine the complete experimental protocol for a particular reference organism that will enable the greatest science return. NASA will execute the dedicated GeneLab flight experiment, with guidance from the SDT, and will be responsible for all extraction and data analysis after the flight. The SDT may also be involved in the post-flight analysis and submission of data to the GeneLab bioinformatic-system. All data that is generated from a dedicated GeneLab mission is expediently released to the public as soon as is reasonably possible.

Collaborative Mission – This type of mission is defined as one where GeneLab either negotiates or partners with the PI to obtain specimens for specific use by GeneLab. The provision of the samples to GeneLab does not interfere in any way with the flight experiment of the PI. Either GeneLab or the PI may perform the sample processing and/or analysis, depending on negotiation and feasibility. Data generated from the shared samples will be submitted into the GeneLab bioinformatic-system. The timing of the GeneLab data submission and release for general science community use will be consistent with allowing the PI time to publish the data from the primary experiment. Omics data from the PI experiment will be submitted to the bioinformatics-system based NRA grant and

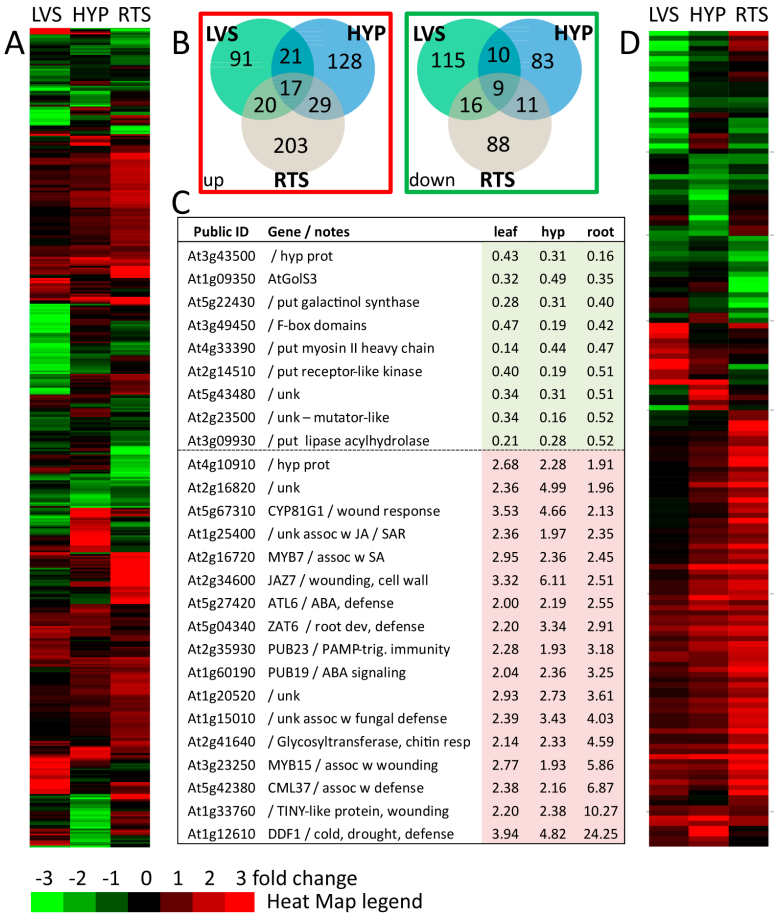
publication requirements. There are two types of collaborative missions:

- **Specimen Sharing** – GeneLab obtains samples from the existing lot of specimens of the PI space flight experiment. Samples may be unused specimens or extraction products. The specimen-sharing plan is approved through a review process conducted by NASA SLPS Division.
- **Specimen Augmentation** – GeneLab identifies a NRA or externally funded spaceflight experiment for supplemental funding to augment the quantity and/or type of existing specimens. The augmentation plan is approved through a review process conducted by NASA SLPS Division. The augmentation to the PI experiment does not in anyway change the peer reviewed and approved science plan, objectives, and specific aims of the grant. Addition of hardware/samples will provide GeneLab and the PI with more specimens to strengthen statistical significance or increase the scientific data return. Samples for GeneLab use are obtained from the additional GeneLab-funded set of specimens.

Initially, GeneLab will focus on collaborative missions to allow time for proof of concept and process revision. However, GeneLab will quickly move to fully dedicated missions beginning as early as 2016, in order to maximize

Organ-specific remodeling of the Arabidopsis transcriptome in response to spaceflight

The The transcriptome of *Arabidopsis thaliana* demonstrated organ-specific changes in response to spaceflight, with 480 genes showing significant changes in expression in spaceflight plants compared with ground controls by at least 1.9-fold, and 58 by more than 7-fold. Leaves, hypocotyls, and roots each displayed unique patterns of response, yet many gene functions within the responses are related. Particularly represented across the dataset were genes associated with cell architecture and hormone signaling, processes that would not be anticipated to be altered in microgravity yet may correlate with morphological changes observed in spaceflight plants. As examples, differential expression of genes involved with touch, cell wall remodeling, root hairs, and cell expansion may correlate with spaceflight-associated root skewing, while differential expression of auxin-related and other gravity-signaling genes seemingly correlates with the microgravity of spaceflight. Although functionally related genes were differentially represented in leaves, hypocotyls, and roots, the expression of individual genes varied substantially across organ types, indicating that there is no single response to spaceflight. Rather, each organ employed its own response tactics within a shared strategy, largely involving cell wall architecture.



C

Public ID	Gene / notes	leaf	hyp	root
At3g43500	/ hyp prot	0.43	0.31	0.16
At1g09350	AtGolS3	0.32	0.49	0.35
At5g22430	/ put galactinol synthase	0.28	0.31	0.40
At3g49450	/ F-box domains	0.47	0.19	0.42
At4g33390	/ put myosin II heavy chain	0.14	0.44	0.47
At2g14510	/ put receptor-like kinase	0.40	0.19	0.51
At5g43480	/ unk	0.34	0.31	0.51
At2g23500	/ unk – mutator-like	0.34	0.16	0.52
At3g09930	/ put lipase acylhydrolase	0.21	0.28	0.52
At4g10910	/ hyp prot	2.68	2.28	1.91
At2g16820	/ unk	2.36	4.99	1.96
At5g67310	CYP81G1 / wound response	3.53	4.66	2.13
At1g25400	/ unk assoc w JA / SAR	2.36	1.97	2.35
At2g16720	MYB7 / assoc w SA	2.95	2.36	2.45
At2g34600	JAZ7 / wounding, cell wall	3.32	6.11	2.51
At5g27420	ATL6 / ABA, defense	2.00	2.19	2.55
At5g04340	ZAT6 / root dev, defense	2.20	3.34	2.91
At2g35930	PUB23 / PAMP-trig. immunity	2.28	1.93	3.18
At1g60190	PUB19 / ABA signaling	2.04	2.36	3.25
At1g20520	/ unk	2.93	2.73	3.61
At1g15010	/ unk assoc w fungal defense	2.39	3.43	4.03
At2g41640	/ Glycosyltransferase, chitin resp	2.14	2.33	4.59
At3g23250	MYB15 / assoc w wounding	2.77	1.93	5.86
At5g42380	CML37 / assoc w defense	2.38	2.16	6.87
At1g33760	/ TINY-like protein, wounding	2.20	2.38	10.27
At1g12610	DDF1 / cold, drought, defense	3.94	4.82	24.25

Differentially expressed genes in response to the spaceflight environment. (A) The hierarchical clustering of all 480 genes with statistically significant ($p<0.01$) differential expression in the spaceflight environment by at least 1.9-fold, in at least one of the three organs (LVS – Leaves; HYP – Hypocotyls; RTS – Roots). (B) A pair of Venn diagrams that illustrates the organ-specific gene expression patterns (C) A list of the 26 genes coordinately up or down regulated in all organs. (D) A hierarchical clustering of 158 genes from (A), which have an association with cell wall remodeling and cell expansion, pathogen or wounding responses, and growth hormone signal transduction.

Paul et al. 2013 BMC Plant Biology 2013, 13:112

ROI on each GeneLab flight mission and to maximize scientific community involvement in the development and execution of these dedicated missions. After the initiation of regular dedicated flight missions for GeneLab, collaborative missions will continue to be an important part of GeneLab, leveraging with others whenever the opportunity for mutual benefit arises. For the dedicated missions, GeneLab will partner with the appropriate project team who will conduct payload implementation and operations. In most cases, the partner will be the project team who is providing the hardware or on-orbit facility, such as Bioserve, for the GeneLab specimens.

To make the GeneLab data system most valuable for systems biology and enable the process of discovery, a set of process and community standards must be adopted. The field of systems biology presents a number of challenges for achieving the GeneLab goals. The key issues involve data integration and sharing between the different simulation software systems. While systems-level investigations rely on a wide range of experimental data types provided by an ever-growing number of public databases and resources, the datasets are often not standardized or sufficiently annotated to facilitate data integration for systems-level analysis. For this, GeneLab created standards will include the following:

- **Community-driven standards for data** – GeneLab will require submitted datasets to adopt more frequently used formats for RNA expression, DNA sequence, protein, and metabolite expression/profiles (e.g., MIAME, MIAPE, MIGS, PSI MI, etc.) and ask researchers to follow the minimum information specifications as they are defined by the global systems biology community.
- **Protocol standardization for sample preservation and assay conditions** – GeneLab will require the metadata for each dataset submitted. That is, experimental data will have full descriptions of experimental conditions including details on sample preservation, sample handling, processing, and service centers used.

Goal 3: Engage the broadest possible community of researchers, industry and the general public to foster innovation

It is critically important for GeneLab to provide scientific outreach and training on the use of the database, computational tools, and new technologies for systems biology to the end users. It is also important to share goals and missions with the public, including educators and students, to foster innovation. The initial focus will be on engaging the scientific community through a GeneLab website and meetings, training, and publications. Eventually, a broad public outreach program will help expose the greater public to the technology advances made by NASA through the GeneLab Platform.

Scientific Outreach

Scientific outreach and engagement is critical to the success of GeneLab. In addition to fostering greater use of spaceflight data to advance discovery, scientific outreach can help expose researchers and their students to exciting aspects of the GeneLab Platform and NASA missions, spark interest and enthusiasm, and encourage communities to support the NASA biological sciences directives. Outreach will be facilitated by the development of the GeneLab bioinformatics system, which will provide scientists from around the world first-hand access to unique scientific discoveries in space and thereby open their scientific thinking and explorations in manners not previously anticipated.

Scientific outreach can be accomplished on several levels, from student engagement to scientist participation within the platform. GeneLab will use a variety of engagement techniques, including providing online trainings and webinars and presenting results at scientific meetings and workshops. Further, NASA-funded university scientists will be encouraged to share their experiences and research from GeneLab at relevant computational, organism-specific, or systems biology meetings and workshops globally. GeneLab may also develop roadshows to demonstrate the tremendous value of the platform for advancing scientific discovery and use the online chat groups of social media, such as Facebook and Twitter, to familiarize scientists with the capabilities and advantages of the unique bioinformatics available from GeneLab resources.

Training

Training scientists on the benefits of the research and mechanism for maximizing the GeneLab environment for their research objectives will be an instrumental part of outreach. Several types of training may be offered, either as in-person sessions or as web-based resources:

- **User manuals** – Online procedural documents describing each of the services, with detailed information about functionality and usage
- **Tutorials** – Online training resources focused on

using specific aspects of GeneLab, with written walkthroughs and video tutorials

- **Workshops** – In-person sessions focused on scientific problems that guide users through GeneLab use
- **Webinars** – Web-based seminars 1-2 hours in length that offer instruction on GeneLab computational tools, services, and standards; could be held several times a month

Inform, engage, and inspire the public by sharing the GeneLab missions, challenges, and success

As noted above, sharing GeneLab with scientists all over the world through conferences, workshops, publications, and meetings is mandatory for mission success. In addition, sharing GeneLab with the general public and students in particular will help increase awareness of the NASA mission while providing real world examples of cutting edge science in the classroom. GeneLab will seek to leverage existing educational and public outreach programs and activities to share GeneLab activities with a broader audience. NASA places particular emphasis on Science, Technology, Engineering, and Math (STEM) engagement activities at the K-12 and college levels to engage Americans in the NASA mission, and to attract and retain students in STEM disciplines, thereby strengthening the NASA and the Nation's future workforce. These broader Agency goals dovetail with the goals of GeneLab to engage and inspire while fostering innovation.

Fostering innovation is also an important part of the GeneLab Platform. This will be accomplished to some degree through the scientific and public engagement described above, possibly augmented by other creative ways to foster innovation, such as crowdsourcing. Crowdsourcing, often done online, involves obtaining needed services, ideas, or content by soliciting contributions from a large group of participants. NASA is already using this technique in many of its programs. For example, for the agency's Asteroid Grand Challenge, NASA partners with the public in developing software solutions to enhance detection of near-Earth objects using agency-funded data. Another program, Disk Detective, is perhaps the largest example of NASA crowdsourcing. With the primary objective of producing publishable scientific results, this program invites the public to help astronomers discover embryonic planetary systems hidden among data from the agency's Wide-field Infrared Survey Explorer (WISE) mission through the website, DiskDetective.org. GeneLab will mirror these efforts by partnering with the community through contests or through other open methods to develop software, design experiments, or generate hypotheses for biological systems. In this way, GeneLab will introduce advances in systems and computational biology that will lead to future discoveries in biology and human health.

Goal 4: Strengthen international partnerships by leveraging existing capabilities and data sharing

Numerous research organizations are engaged in activities that are similar to and synergistic with GeneLab both in the U.S. and around the globe. In the U.S., the DOE and the National Institutes of Health (NIH) have funded large-scale omics data-sharing in various research fields, and research and technology advances are occurring rapidly. To optimize NASA's limited GeneLab budget and leverage existing systems and resources, NASA will partner with other research programs to learn from their advances to keep up with the state of the art without recreating unnecessary or redundant infrastructure. Numerous NASA programs engaged in biological research, including the HRP, Space Biology, Synthetic Biology, Space Technologies and Astrobiology can contribute to and benefit from the GeneLab datasets. NASA will use GeneLab to establish more intra-agency collaborations and enhance data-sharing between these programs.

Following past precedent of enhancing progress through collaborations between NASA and international partners, especially with projects touching life science research, GeneLab will initiate international partnerships to extend this ISS Science Platform far beyond national boundaries. Throughout the 35+ year history of international cooperation in life space research, NASA and international partners have realized greater successes by sharing technologies, resources, and approaches. The full potential of these efforts will only be realized if the data are openly accessible to all, providing an opportunity to leverage investments and expertise internationally. International partnerships will facilitate omics data management and mining through standardizing operations and leveraging existing hardware. Key areas for GeneLab to engage with the International partners include:

- Collaborate with other federal agencies on existing bioinformatics programs to create a data environment with relevance to ground-based research.

- Collaborate with international partners to leverage existing space hardware and flight experiments. These partners could include the Russian Federal Space Agency (Roscosmos), Japan Aerospace Exploration Agency, European Space Agency, and Canadian Space Agency.
- Providing a platform with standardized data collection approaches to increase cooperation and relevancy of data sets on international space flight activities
- Support international partnership efforts to further the understanding of the space effects on living systems
- Capitalize on international launch and flight systems to enable the greatest diversity of research endeavors

The GeneLab Phased Approach

NASA's GeneLab Science Platform will follow a phased approach to achieve its four key goals. Each platform phase will be marked by increasing complexity and include a final review to allow for any adjustments needed for successful implementation and maximum benefit to the scientific community (Figure 6). The final phase of the platform will enable full implementation of GeneLab. These phases are as follows:

- **Phase I. Proof of Concept** – The team will stand-up the GeneLab Platform and demonstrate the proof of concept for GeneLab. The early platform will consist of a public website and a searchable database for investigators to browse through existing Space-flight data or data generated from early ground-based studies This phase will establish the requirements, metrics, and implementation plans for the platform. In addition, protocols will be developed, and a limited set of samples and data will be generated and used to demonstrate the prototype bioinformatics system(s). The completion of Phase I will include an evaluation of the GeneLab proof of concept and decision-point to make adjustments before moving into the second phase.
- **Phase II. Data Acquisition** – As the platform grows, links will be made to existing data search sites

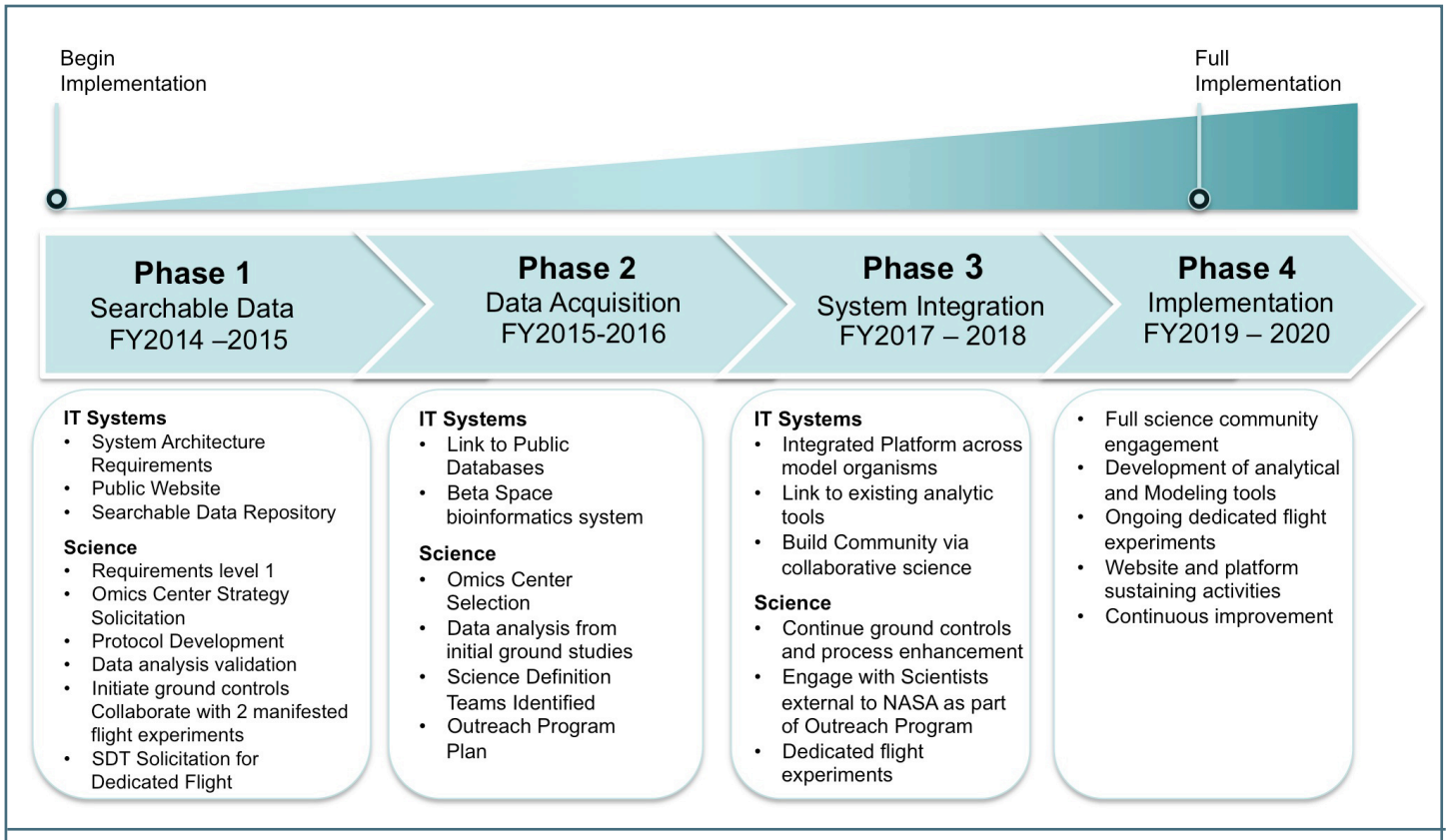


Figure 6. The GeneLab Project Phased Approach

The GeneLab project will follow four distinct phases of implementation. Each phase of the project will be marked by increasing complexity and increasing value and utilization of the GeneLab data system. The final phase of the project will enable full implementation of GeneLab.

and repositories such as ArrayExpress, GenBank, EMBL, and the Cancer Genome Hub. With these links, users will be able to find spaceflight data using popular existing data search sites enabling more scientific linkages. This phase will continue to build up GeneLab capabilities, including data system capability and enhanced datasets from a number of selected GeneLab spaceflight and ground-based analogues experiments to populate the data system with GeneLab datasets. GeneLab will also begin to assess data curation capabilities to further benefit the scientific community. GeneLab will also define an Outreach Plan to further enable both public and scientific community engagement. After Phase II, the GeneLab Platform will be reviewed by both NASA and external stakeholders to allow for any needed changes or course corrections prior to full implementation of the platform.

- **Phase III. Integrated Omic Capability** – This phase will initiate a platform capable of integrating various data types across model organisms. It will also see the beginning of collaborative science where investigators can partner through the website on modeling and experimental ideas. The Outreach Program will be fully implemented, and annually dedicated spaceflight experiments and ground-based analogue studies will be running with the datasets for those experiments posted to the GeneLab data system. This phase will expand the data system capabilities to include tools for data analysis, and maintain the data system as it grows.
- **Phase IV. Full Implementation and Continuous Improvement** – This phase will continue to develop the data system, expand capabilities, add more tools for data analysis, and maintain the data system as it grows. Phase IV will also include full science community engagement with regular solicitations for data mining and publication of results from the GeneLab data. Phase IV will also have regular reporting requirements to demonstrate the value and impact of the GeneLab Platform to NASA, the scientific community, other external stakeholders, and the general public. The full implementation phase of GeneLab will have regular, nominally annual, dedicated spaceflight missions as well as a number of collaborative spaceflight missions and ground-based analogue experiments each year. The full implementation phase of GeneLab will have annual NASA reviews for progress and impact as well as external reviews every 3 years to assure the continued improvement of the GeneLab Platform for many years to come.

Strengths and Challenges

This initiative must take into account both the strengths residing within NASA and the challenges that must be addressed to be successful. NASA will utilize its experience with performing life science experiments in space and big data enterprises for major missions during the implementation of GeneLab. Nonetheless, gaps still exist. Because GeneLab is a new focused area of research, NASA will need to augment its experience and capabilities to ensure the highest return of science for use by the research community at large. The biggest challenges facing GeneLab are implementing a systems biology approach, expanding the space biology community, and generating an omic-based bioinformatics system utilizing computational biology expertise. Further, NASA must apply its strengths and long history with space biology research to mitigate limitations to flight experiments and ensure the highest quality and quantity of samples collected for GeneLab.

Space Life Science Research Strengths and Challenges

NASA has been conducting space life science experiments for over 50 years. As a result, NASA and its partners have unique capabilities, including vehicles, technology, collaborations, expertise, and laboratories, for studying biological systems in space. This capability includes access to the ISS laboratory and other free flyer capsules. In addition, NASA and its partners have developed a suite of hardware to support a variety of model organisms for research on cells, drosophila, plants, and rodents. Even the crew of the ISS is utilized for experimentation in space. NASA has supported hundreds of experiments, optimizing the experimental design to maximize the scientific return within the limitations of the ISS. If needed, NASA can call upon its specialized engineers to build scientific equipment required for the success of the GeneLab Platform. This experience underscores NASA's ability to perform the highest quality research possible with the resources available in space.

There are also a number of challenges to performing high-value, high-impact, biological research in space. These challenges, inherent to all space biology research, may limit the success of the GeneLab Platform if they are

Table 1.

Limitation on spaceflight experiments affecting GeneLab	NASA mitigation strategy
Frequency of access to space is limited by ascent, in-flight, and descent resources that are needed to conduct research	• Prioritize GeneLab flight experiments and maximize data collection through Earth-based analog experiments to increase ability to compare and analyze limited flight data with ground controls and space analogs
Limited access and the challenges of performing experiments in space results in limited sample quantities	• Complete thorough ground protocol development to maximize the success of sample data acquisition upon recovery of the samples • Work with the PIs and SDTs to maximize experimental designs toward recovery of high quality sample in sufficient quantity to enable the high throughput analysis required
Due to safety restrictions imposed on human flight missions and current preservation technology, preservation methods for ensuring high quality omics analysis are limited	• Research alternate preservation methods to improve sample collection methods for the space laboratory • Continue ongoing research to develop a rapid freeze technology to more closely mimic the preferred method of preservation in terrestrial laboratories

Table 2.

Gap	Mitigation strategy
Need for in-house computational biology expertise to develop a database compatible with existing life science databases	• Work with partners and contractors to identify qualified staff to lead the science definition and database development • Investigate possible collaborations with other agencies currently implementing similar systems, CASIS, and the Steering Committee to ensure required skills are identified and brought to the team
Need to keep pace with the rapid advances in omics technology and research	• Continue to refine and improve protocols, standards and data acquisition per the best technology available • Partner with researchers and procure services from reputable organizations implementing the latest technology and process • As necessary, buy critical equipment and services to ensure the highest quality of sample processing and data acquisition
Lack of data compatibility standards adopted by the omics-based research community or technology developers	• Develop strong criteria for the data to be included in the GeneLab bioinformatics system • Maximize the acceptable formats to increase the likelihood of integration and compatibility with data from other databases
Need for high quality data to ensure GeneLab success	• Ensure that all data submitted to GeneLab includes raw, annotated, and metadata and has passed high quality control standards set by the Platform • Curate date to ensure it meets the standard • Continually update standards to reflect technology advancements • Integrate the GeneLab data source with most of the major genomics, proteomics, and metabolomics databases in common use, developing user-friendly interfaces to allow the databases to interact
GeneLab for human and commercial data requires special consideration and bioinformatics system architecture	• Remain committed to making the bulk of the data open to any researcher, while addressing concerns from the commercial sector regarding proprietary data and protecting the personal identification of any human data • Overcome any potential legal hurdles or issues of national security in using NASA-sponsored databases for the purpose of bioterrorism • In developing a data policy, investigate the related policies, laws, and security issues and identify technical and policy-based solutions to protect the data and user community

not addressed. Table 1 outlines how NASA will apply its capabilities and experience to mitigate current limitations on flight research.

Because of these capabilities and gap mitigation strategies, NASA is situated to work with Earth-based scientists in the design and implementation of space-bound experiments that complement similar ongoing life science research conducted on earth.

Data System Development Strengths and Challenges

NASA is uniquely positioned to develop bioinformatics systems to analyze biological data from space-based experiments in a manner compatible with databases and informatics systems already in play for the study of organism genomics, proteomics, and metabolomics. In addition to conducting research in space, NASA has been using big data for many applications and has experience with high performance computing, archiving of large datasets, and data transfer. Major relevant capabilities and experience follow:

- Through the NASA Center for Climate Simulation, modeling of the Earth ozone layer, aerosol, weather patterns, and temperature projections from satellite data
- Solar flare modeling with Interface Region Imaging Spectrograph (IRIS) mission data
- Fluid dynamic studies for aircraft design
- Simulations of galaxy formation and evolution
- NASA Earth Exchange studies of climate and environmental change using satellite imagery and applying cloud computing scalability
- The Square Kilometer Array planned generation of data at 700 gigabytes per second
- The NASA High End Computing Capability (HECC) maintains some of the largest and fastest supercomputers in the world, delivering services to NASA's aeronautics, exploration, science, and space technology missions

To augment experience and expertise with big data and computing systems, NASA is increasing its knowledge in omics research; but the Agency currently has gaps in knowledge related to omics and bioinformatics that are challenges to GeneLab success. Recognizing these challenges, the GeneLab Project will implement the following strategies to mitigate gaps with the aim of ensuring mission success (Table 2):

Customers and Stakeholders

The data and information generated by GeneLab and made available to the public will have many broad-reaching users and applications. This large customer base includes many NASA programs and projects, as well as external commercial, academic, and government organizations. The main customers for GeneLab data and information will be from the life sciences and omics research communities, including the following:

- **Traditional Space Biology Science Community:** NASA researchers in rodent, plant, microbial, cell biology, and other fields will use GeneLab to study and understand the fundamental processes of biological systems in space to address the high priority recommendations of the 2011 NRC Decadal Survey.
- **NASA Human Research Program (HRP):** GeneLab is a potential host for the One Year Twins Study Data – Pilot Study for Human Omics Data. HRP researchers will use GeneLab to help close gaps in knowledge related to the risks to human health in space and develop more effective countermeasures to ameliorate the detrimental effects of spaceflight on human health and performance.
- **Center for the Advancement of Science In Space (CASIS):** GeneLab tools will provide an opportunity for data mining from commercial interests to identify potential targets for drug development, personalized medicine, and systems biology of model organisms and humans. Emergent properties from GeneLab analyses will help inform commercial companies that may be interested in using the space environment for their own applications. These applications strengthen the NASA CASIS Mission to support commercial research and utilization of the ISS National Laboratory.
- **Non-Traditional Genomics and Systems Biology Communities:** GeneLab data will benefit pharmaceutical and commercial interests wherever those Earth-based research interests intersect with the environmental and gravitational effects of spaceflight. The multi-channel omics approach of GeneLab will also demonstrate analytics and data processing possibilities that have broader value and benefit beyond the analysis of spaceflight data alone.
- **The General Public:** Anyone with internet access and interest will be able to access GeneLab data and freely explore space biology and ISS research results for themselves.

NASA Stakeholders

The major NASA stakeholders of the GeneLab Science Platform include the organizations within the Agency that sponsor—and benefit most—from the GeneLab science, spaceflight experiments, and omics data system. The GeneLab stakeholders will oversee the development and implementation of the platform. The Director of the Space Life and Physical Sciences Research Application Division (SLPSRA) at NASA Headquarters will bear primary responsibility for leading and executing the GeneLab Platform, supported by the other major NASA stakeholders for GeneLab:

- Human Exploration and Operations (HEO), NASA HQ
- ISS Program, NASA JSC
- ISS Payloads Office, NASA JSC
- Space Biology Project, NASA ARC
- Utilization and Life Sciences Division, Kennedy Space Center

The International Space Station Program Office (ISSPO) at the Johnson Space Center will support the integration and operations of research activities, development of a bioinformatics system and analysis of space flight samples. GeneLab will be implemented by science and technical staff at NASA Ames Research Center, the NASA Jet Propulsion Laboratory, NASA Johnson Space Center, and NASA Kennedy Space Center.

External Stakeholders

NASA will engage a variety of external scientists, technologists, industry experts, and representatives of other government agencies in several ways to create a network of external stakeholders for GeneLab. These external stakeholders will offer guidance and direction to help ensure achievement of the GeneLab Project goals. The external stakeholders are described below.

Steering Committee

The GeneLab Steering Committee includes scientific advisory members who are senior leaders from various scientific fields. The Council Members will act as advisors and collaborators to the GeneLab Platform on matters and activities related to achievement of the GeneLab Project goals and objectives.

The role of the Steering Committee is to evaluate the appropriateness and feasibility of the proposed research projects for the stated goals and objectives. They will assess the strategy for relevance to space life science and the omics industry and when appropriate, identify and help develop strategic partnerships for collaboration opportunities. The Committee will make recommendations to help build advocacy within NASA and the omics industry and to mitigate the project risks.

These recommendations or changes will be communicated to the GeneLab Platform via the Chair or Vice Chair of the Steering Committee. The Steering Committee will meet as needed. Other GeneLab Managers and Fellows may also participate in the Steering Committee, if appropriate.

Science Definition Teams

Experts from academia, industry, and government organizations, selected through a competitive process, will be invited to participate in SDTs to define and execute specific GeneLab experiment as required. These organism-specific teams will consult with subject matter experts to assist with the development and execution of experiments as well as to establish end user requirements for the bioinformatics system including such duties as:

- Identify organism-unique priorities and requirements
- Assist with development of the GeneLab ground study requirements and acceptance criteria
- Input and evaluate specimen selection and processing requirements for on-orbit collection and preservation
- Review the GeneLab science standards, process, and procedures for sample preparation for model-specific sample preparation as applicable
- Provide input to science success criteria

Customer and Stakeholder Advocacy

Stakeholder and customer advocacy is maintained by demonstrating progress in achieving the objectives and deliverables of the platform, on schedule and within budget. Frequent communication will ensure GeneLab products remain relevant to objectives outlined in this plan, as follows:

- Direct communication between the project leadership and the GeneLab Platform elements at multiple NASA centers; ongoing transition of end-item platform deliverables
- Scientific and technical exchange activities, including annual technical interchange meetings
- Meetings, workshops, professional conferences, and reporting in scientific literature as needed

- Education and training programs and activities
- Monthly and quarterly reporting to appropriate stakeholders
- Annual Steering Committee meetings

Performance Goals

NASA is developing the GeneLab Platform to increase the ROI for biological experiments conducted in space. Success for GeneLab will be assessed upon reaching full implementation of the stated objectives and measured in three categories:

- Community Adoption
- Applied Research
- Increased Publications and References

Throughout the GeneLab development process leading to full implementation, regular management reviews will be held to assess the platform's progress against the defined milestones outlined in the Phased Platform Approach. These reviews will evaluate the programmatic and scientific advances toward the stated goals of the platform.

The three categories to be assessed are described in detail below:

- **Community Adoption** – Use by the scientific community is a strong indicator that the system contains valuable data and tools to further research endeavors. The GeneLab Platform will make monthly measurements of the number of users, how often the bioinformatics system is queried, the number of analyses performed within the GeneLab system, and dataset downloads. Additional analyses will identify which communities are using GeneLab and how to focus community outreach efforts. After the first year of full implementation, it is expected that GeneLab will have 200 users, that 50 analyses will have been completed within GeneLab, and that at least 50 datasets will have been downloaded for analysis outside of the GeneLab system.
- **Applied Research** – Providing high throughput, broad-spectrum datasets from GeneLab-sponsored experiments is required to meet the goals of the GeneLab Platform. Researchers will be required to submit their data and metadata to GeneLab for all omics-based ground and flight experiments funded by NASA. While valuable, the scope of these dataset is focused on the specific area of research proposed. To maximize the data in the system, GeneLab will conduct experiments in all model organisms and obtain multi-platform data for each species. By full implementation, GeneLab will have the ability to conduct pilot studies and/or space analog ground studies, collaborate with a manifested experiment and GeneLab reference experiments for dataset generation each year as the budget permits.
- **Increased Publications and References** – Presently, Space Biology expects a minimum of one publication per investigator/experiment flown in space. As stated previously, data mining in GeneLab (from flight and ground experiments) is expected to increase publications resulting from spaceflight research. It is also expected to contribute to the development of new hypothesis-driven research. As customary, GeneLab will request that any user who is publishing a reference to GeneLab use a standard reference line to aid in publication searches. GeneLab will track publication rates against historical publication trends for grant-funded research in space. GeneLab will be considered successful when those rates increase each year. In addition, the GeneLab Project will track references to GeneLab data use in future NASA Research Announcement proposals submitted. The cyclical nature of the program will be demonstrated when new proposals reference GeneLab.

Acknowledgements and Contributors

Rick Chen, Intrynsix

Homer Fogle, Dynamac

Kristina Gibbs, Lockheed Martin

Masood Hadi, Ph.D., NASA Ames Research Center

Jimmy Jung, Intrynsix

Elizabeth Keller, Dynamac

Holly Larsen, Larsen Communications

Nicki Rayl, NASA Headquarters

Terry Reisine, Ph.D., LZ Technologies

Kevin Sato, Ph.D., Lockheed Martin

Jeff Smith, Ph.D., NASA Ames Research Center

Sidney Sun, NASA Ames Research Center

Terri Thompson, Ph.D., Critical Mass Consulting Services, LLC.

Linda Timucin, GHG

Jon Welch, AEWood & Associates