

# Integrating bioinformatic strategies in spatial life science research

Yangyang Hao<sup>†</sup>, Liang Lu<sup>†</sup>, Anna Liu<sup>†</sup>, Xue Lin<sup>†</sup>, Li Xiao, Xiaoyue Kong, Kai Li, Fengji Liang, Jianghui Xiong, Lina Qu, Yinghui Li and Jian Li 

Corresponding authors: Yinghui Li, The State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China. E-mail: yinghuidd@vip.sina.com; Jian Li, Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing, 210096, China. E-mail: jianli2014@seu.edu.cn

<sup>†</sup>Yangyang Hao, Liang Lu, Anna Liu and Xue Lin are co-first authors and equally contribute to this work.

## Abstract

As space exploration programs progress, manned space missions will become more frequent and farther away from Earth, putting a greater emphasis on astronaut health. Through the collaborative efforts of researchers from various countries, the effect of the space environment factors on living systems is gradually being uncovered. Although a large number of interconnected research findings have been produced, their connection seems to be confused, and many unknown effects are left to be discovered. Simultaneously, several valuable data resources have emerged, accumulating data measuring biological effects in space that can be used to further investigate the unknown biological adaptations. In this review, the previous findings and their correlations are sorted out to facilitate the understanding of biological adaptations to space and the design of countermeasures. The biological effect measurement methods/data types are also organized to provide references for experimental design and data analysis. To aid deeper exploration of the data resources, we summarized common characteristics of the data generated from longitudinal experiments, outlined challenges or caveats in data analysis and provided corresponding solutions by recommending bioinformatics strategies and available models/tools.

**Keywords:** space environment factors, disease, data characteristics, computational models, association prediction

## Introduction

Beginning in the early 1970s, a series of Soviet space stations, US Skylab stations and numerous space shuttles offered a basis for humans to live and experiment in space. The International Space Station (ISS), which was established through multinational cooperation from 1998 to 2011, will continue to serve as the Space Environment Research Laboratory until at least 2024 [1]. Additionally, China successfully launched a manned spacecraft in 2003 [2], and the China Space Station project is now progressing steadily. Furthermore, private companies such as SpaceX are developing new systems optimized for spacecraft landing

modes and other aspects [3]. These production advances have dramatically improved the reusability of space vehicles, resulting in a sharp decrease in launch costs, and will lead to the development of a new generation of space launch vehicle systems [4]. Consequently, the commercialization of Low Earth Orbit (LEO) travels and the acceleration of long-range exploration programs will be greatly advanced. The majority of early manned flights, including the ISS, orbited Earth in LEO, which is still shielded by Earth's magnetosphere [5]. As the scope of human exploration broadens, forthcoming missions to the Moon, Mars and beyond will expose the astronauts to more intense space radiation

**Yangyang Hao** is a PhD student at the Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing, China.

**Liang Lu** is an associate researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Anna Liu** is a master student at the Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing, China.

**Xue Lin** is an associate professor at the Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China. Her research focuses on bioinformatics, data mining and machine learning.

**Li Xiao** is an associate researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Xiaoyue Kong** is a master student at the Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing, China.

**Kai Li** is an assistant researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Fengji Liang** is an assistant researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Jianghui Xiong** is a researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Lina Qu** is a researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Yinghui Li** is a researcher at the State Key Laboratory of Space Medicine Fundamentals and Application, China Astronaut Research and Training Center, No. 26 Beiqing Road, Haidian District, Beijing, 100094, China.

**Jian Li** is a professor at the Key Laboratory of DGHD, MOE, School of Life Science and Technology, Southeast University, Nanjing, China. His research interests lie in bioinformatics, genomics and big data computing.

**Received:** July 6, 2022. **Revised:** August 15, 2022. **Accepted:** August 27, 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

and longer mission durations, meaning higher health hazards for them.

Environment factors affecting living systems in space include microgravity, radiation, confinement/isolation, distance from Earth [6], etc. They are inextricably related and often investigated independently due to research constraints, while multifactorial research does also exist. By simulating the effects of individual stressors, it has been found that they can lead to different physiological or psychological problems [7, 8]. Moreover, the effects of multiple environmental factors co-existing in the actual mission are not simply superimposed [9]. To protect the astronauts from these environment factors and complete space missions, there is a pressing need to understand what and how changes occur in the living systems, which will contribute to providing appropriate countermeasures to reduce the adverse effects. Furthermore, space life science research could provide insights into organismal health on Earth, such as muscle loss and osteoporosis in the elderly, as well as the impact of isolation on mental health [10]. The biological adaptation changes caused by the space environment that have been identified so far are complex and lack systematic collation, especially in terms of cascading relationships.

Numerous spaceflight biology studies have been dedicated to finding out organismal health threats in space [11]. They measured biological adaptation changes in the presence of space environment factors from multiple perspectives and provided data with multiple dimensions, including but not limited to multi-omics data. These data were incorporated into a variety of associated spaceflight biodata resource platforms [1], such as NASA's GeneLab database (<https://genelab.nasa.gov/>) [12] and the Life Sciences Data Archive (<https://lsda.jsc.nasa.gov/>). GeneLab is a comprehensive space-related omics database that provides access to data from experiments that explore the molecular response of terrestrial biology to the spaceflight environment. The Life Sciences Data Archive is a publicly accessible active archive of data from spaceflight, flight-analog and ground-based life sciences research investigations. In addition, Earth-based human space simulation research [13, 14] will continue to produce more experimental data, which are more affordable and accessible than space missions [6]. How leveraging these accumulated data resources to reveal more comprehensive patterns of biological adaptations is the challenge of the day, making it important to manage and integrate data across multiple platforms, followed by data analysis and interpretation to achieve biological understanding and provide countermeasures.

In this review, we summarized the multi-level adaptive changes that occur in the living systems in response to space environment factors and their intrinsic connections, including molecular, cellular and systemic changes at the physiological level, as well as psychological outcomes. We also revealed many unknown parts that remain to be complemented. Furthermore, we compiled accessible metrics at each level, including omics and phenotypic data, and outlined common challenges in data analysis. Accordingly, we proposed some optional bioinformatic strategies and assessed related models/tools to provide a reference framework for the analysis of space biological data (Figure 1).

## Biological adaptations to space environment factors

The effects of environmental factors such as radiation, microgravity, confinement/isolation and distance from Earth on organisms have now been explored in a number of ways, including single-factor and multi-factor studies [6, 9, 15]. We summarized the biological effects of these widely investigated environmental

factors in space. In addition to causing the global environmental shift, 'the distance from Earth' has been singled out for its significant impact on human psychology. These studies have shown that these factors may have both psychological effects, such as increased stress and mood disturbances, as well as physical health problems, such as altered musculoskeletal structure and function, sensory-motor impairment and cardiovascular dysfunction [7, 8]. And these factors' combined effects differ from their individual effects and require further investigation. There are numerous findings related to these effects that need to be systematically sorted out to reveal their intrinsic connection. Thus, we summarized the multi-level adaptive changes in the living systems in response to space environment factors and their intrinsic connections (Figure 2).

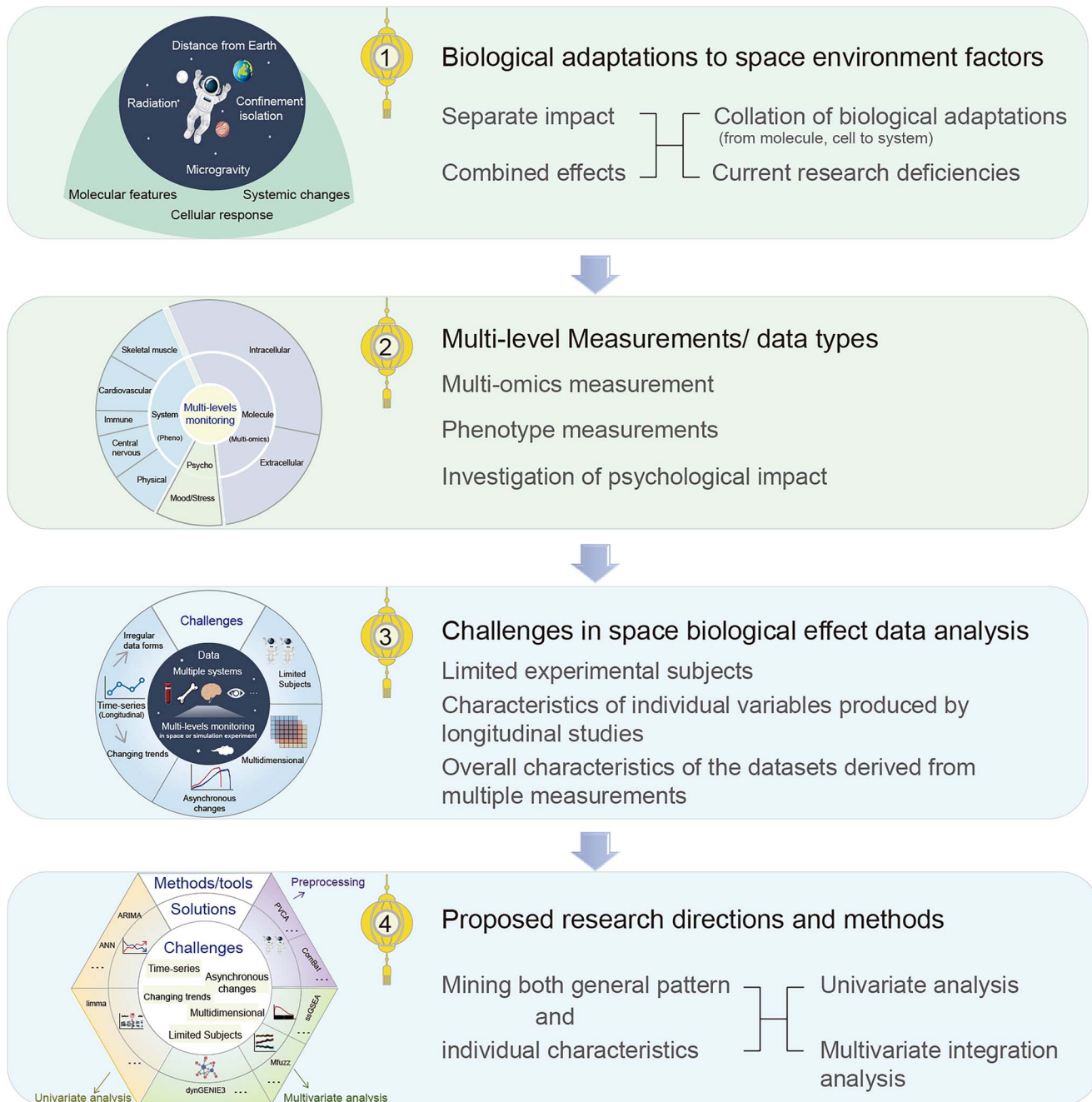
## Biological effects of microgravity

All life has evolved to form its present organismal structure under constant gravity on Earth. In the microgravity environment of space, the balance between cellular structure and external forces is disturbed, leading to extensive changes at the cellular and subcellular levels [16]. Studies on mice after space flight found significantly altered genes, i.e. Gridley *et al.* [17] reported that the expression of apoptosis-related genes, as well as genes involved in extracellular matrix proteins and stem cell signaling proteins in mouse lung cells, was significantly altered. In addition, Hammond *et al.* [15] reported that the expression of genes involved in apoptosis and cell death were significantly upregulated in mouse kidneys and liver. It was found that microgravity has different impacts on apoptosis of different cell types, mediated by different signal transduction processes [18]. Changes in signal transduction in microgravity-induced apoptosis have led to new insights into the underlying regulatory mechanisms of apoptosis. And cancer researchers have discovered a new direction for cancer therapy. In most (but not all) tumor cell lines, the ability of microgravity to trigger cell apoptosis has been proven [19, 20]. However, under some circumstances, apoptosis of some cancer cells can be reduced in microgravity environments [21–23]. Overall, the mechanisms and outcomes of microgravity affecting different tumor cell types vary and need to be further investigated.

In addition to the increased probability of apoptosis, cellular changes under microgravity exposure include differentiation, adhesion, migration and proliferation. By promoting apoptosis or other changes in various cell types, microgravity may affect multiple physiological systems in astronauts, including the musculoskeletal system [24], the cardiovascular system [25], the immune system [26], the digestive system [27] and the central nervous system [28, 29]. It has also been linked to eye problems (e.g. cataracts) after space missions [30, 31]. In conclusion, microgravity requires more investigation, given the significant effects on numerous aspects of living systems.

## Biological effects of radiation

Radiation exposure in spaceflight poses a major potential risk to astronauts' health in the long run. The main effect of radiation exposure is the damage to DNA, including base damage, single-strand breaks (SSBs), double-strand breaks (DSBs), chromosomal aberrations, micronuclei and genomic instability [32]. While SSBs can be repaired by excision repair [33], DSBs involve a more complex repair process. The repair process may be subject to misrepair, further causing cell cycle arrest, cell death, mutations and chromosomal rearrangements [34, 35]. The cellular responses to DNA damage differ depending on the cell type, cell cycle stage and degree of damage [32]. Damage at varying levels in cell types causes multi-system damage, including the central



**Figure 1.** The organizing framework of this review.

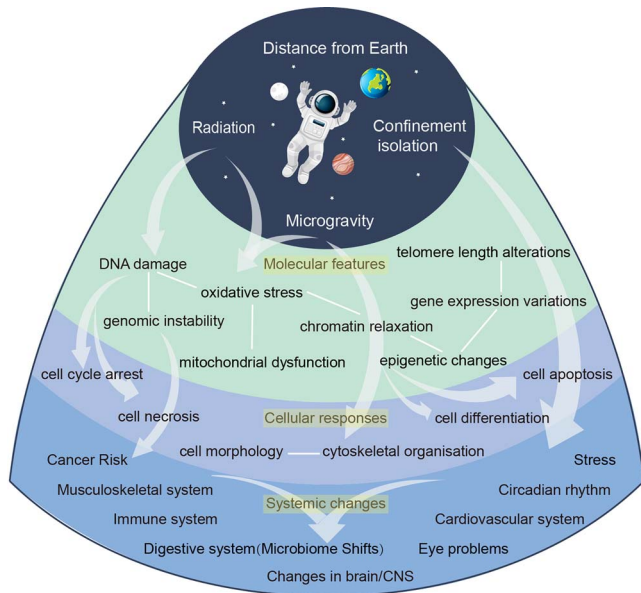
nervous system, musculoskeletal system, cardiovascular system [36, 37] and immune system [38]. The carcinogenic risk of space radiation is also a major health concern for astronauts because ionizing radiation-induced genomic instability is a driving factor for radiogenic carcinogenesis [39, 40]. The degree of carcinogenic risk varies by tissue type, radiation type and age at exposure. Single particle responses have been examined more widely, whereas the impacts of mixed radiation types are less clear and lack appropriate study support. Moreover, since outer space radiation occurs in a microgravity environment, it is unknown if clustered DNA damage occurs and is repaired under their dual action.

### Combined effects of multiple space environment factors

Biological effects in outer space are responses of organisms when they are exposed to multiple space environment factors simulta-

neously, while most studies only examine the effects of individual factors in a static environment. To fully comprehend the biological effects in space, it is necessary to accurately assess the combined effects of multiple factors. The performance of cells [41] and mouse models [9] exposed to radiation and microgravity simultaneously revealed that the dual effect posed a greater health risk than radiation alone. According to Xu *et al.* [9], heavy ion radiation-induced human B lymphocyte apoptosis increased in microgravity. We compiled a list of biological responses resulting from the combined effects of all environment factors in space, including the physiological changes and the psychological consequences.

Oxidative stress and redox imbalance are typical molecular features of spaceflight, induced by radiation and microgravity, which may also trigger DNA damage. And DNA damage is often correlated to apoptosis when there are defects in the DNA repair system [42]. At the physiological level, oxidative stress and redox



**Figure 2.** A collection of biological adaptations at various levels, including molecular features, cellular responses and systemic changes. The white arrows and lines represent cause–effect relationships or potential associations between different changes. CNS, central nervous system.

imbalance lead to dysregulation of the cardiovascular, immune, neurological and metabolic systems. Additionally, oxidative stress is closely associated with mitochondrial dysfunction. Mitochondrial dysfunction is characterized by a reduction in the expression of the mitochondrial oxidative phosphorylation (OXPHOS) gene encoded by nuclear DNA. Moreover, oxidative stress can induce epigenetic changes through chromatin relaxation and thus regulate gene expression. Dynamic alterations in telomere length have also been observed during spaceflight, which has been linked to age-related disorders including dementia, cardiovascular disease and cancer, all of which have the potential to influence astronaut health and performance during and after long-term missions [6]. The space environment can also cause a shift in the microbiome [43, 44]. Interactions between the microbiome and the host affect key human physiological processes, including inflammatory responses, metabolic functions, hormone levels, disease susceptibility and pathogenesis [45]. The gut microbiome, for example, is implicated in the pathogenesis of numerous digestive diseases [46].

Aside from the major molecular features listed above, there are a number of functional pathways that have been linked to spaceflight health. The NF- $\kappa$ B pathway, for example, has been linked to recognized spaceflight-related health hazards such as immunological dysfunction, bone loss, muscle atrophy, central nervous system dysfunction and space radiation dangers [47]. Accordingly, we suggest that the space environment induces a wide range of adaptive changes at the molecular level, and many are left to be discovered. Also, researches on the combined effects of multiple spatial environmental factors are still at a preliminary stage, and more studies for multi-factor situations are needed.

### Biological effects of confinement and isolation

In long-term confined/isolated environments, such as the Mars-500 mission [48] and the 180-day controlled ecological life support system (CELSS) experiment [14], many aspects of human health may be affected, including mental–emotional disturbances [49, 50], reduced muscle activity [13], changes in immune responses

[51], gut microbiota [52] and metabolism [44]. In addition, mood disorders such as anxiety brought on by long-term isolation are associated with abnormal bone metabolism [53]. Confinement/isolation also disrupts circadian rhythms [54], the disruption of which may affect mood, cognition and performance [55] and further lead to additional health disturbances.

Furthermore, prolonged isolation could trigger psychological stress, which might result in a shift in biological vulnerability to radiation danger. According to studies in which mice were subjected to both psychological stress and low linear energy transfer radiation, stress improved bone marrow radiation susceptibility in some susceptible animals, but it did not affect hematological toxicity or genotoxicity in wild-type mice [32]. The mechanisms of how psychological stress modulates radiation susceptibility have not yet been elucidated. Hence, more experiments are needed to produce additional data for further research.

Advances in technology will enable exploration at farther distances from Earth, where medical and surgical events will be limited, thus endangering the safety of astronauts. As the exploration mission becomes further away from Earth, the crew may experience communication delays. A Mars mission could cause communication delays of up to 20 min with Earth. And there will be many unknown environmental factors, such as higher doses of radiation and changes in the light and dark cycles [56]. So, astronauts will be more stressed as they travel further away from Earth. The exact impact is to be supported by the conduct of relevant studies.

### Multi-level measurements/data types

Multifaceted experiments were conducted to explore the effect of spatial environmental factors on organismal health, including neuroimaging, electrophysiology, biochemistry, systems biology and clinical questionnaires, thus producing large amounts of high-dimensional data. These data include but are not limited to the following: multi-omics measurements at the molecular level (epigenomics, transcriptomics, proteomics, metabolomics, microbiomics, etc.), systems level (biochemical index data, image data, electrophysiological data) and psychological level (stress surveys, mood). In this review, we have compiled various measurements and the biological issues they can reflect (Figure 3), which will assist in designing experiments to dissect the biological effects in space.

### Multi-omics measurement

Whether in spaceflight simulations or actual spaceflight experiments, space biologists around the world are increasingly reliant on omics approaches due to their ability to maximize the knowledge gained from rare spaceflight experiments [57]. We reviewed common omics measurements used in space biology, focusing on the biological issues they can reflect and the available detection platforms.

#### Epigenomics

Epigenomics can be used to detect space environment-induced reversible modifications at DNA or RNA level, such as DNA methylation, histone acetylation, RNA methylation, etc. Modifications like these perform critical regulatory roles in gene transcription and subsequent cellular functions [58]. They can also be used as biomarkers, for example, one of the earliest events in the DSB damage response is the phosphorylation of histone H2AX to produce  $\gamma$ -H2AX, which can be used as a sensitive tool for detecting DSB [59–61]. Space environment factors can trigger alterations in



before and after the modification [78]. Protein interactions can be discovered utilizing unbiased approaches (e.g. MS, yeast two-hybrid tests) or affinity purification methods (using antibodies or genetic tags). Affinity methods can also examine overall interactions between proteins and nucleic acids (e.g. ChIP-Seq).

### Metabolomics

Metabolomics simultaneously quantifies multiple small molecule metabolic function products in cells, including amino acids, fatty acids, carbohydrates and other small molecules. Metabolite levels and relative ratios reflect metabolic functions, and deviations from the normal range are usually associated with diseases. Small molecule abundance can be quantified using MS-based methods [79–82].

### Microbiome

The space environment, irregular diet and disrupted circadian rhythms may lead to changes in the ecosystem of the microbiome [83, 84], including the environmental microbiome [85], the skin microbiome, the oral microbiome [86] and the gut microbiome. The microbiome can be analyzed by amplifying and sequencing certain highly variable regions of bacterial 16S rRNA genes, or by birdshot metagenomics sequencing that sequences total DNA. Several analytical tools for NGS data targeting 16S or metagenomics analysis have been developed, such as QIIME (Quantitative Insights into Microbial Ecology) [87], which can be used to identify taxa associated with diseases or other phenotypes of interest [88].

### Phenotype measurements

Phenotypes are the observable characteristics or traits of an organism and can provide valuable explanations for the consequences of living system responses to space environments. Phenotypic data can be used to link genetics and phenotype. Phenomics is a field that deals with high-dimensional phenotypic data at the organismal scale and is an important complement to genomics. The current phenotypic number throughput is low, and technological advancements can reduce costs to enhance phenotype throughput [89]. For space response assessments, we compiled commonly used multi-system phenotypic metrics.

### Skeletal muscular system measurements

This includes both bone and skeletal muscle. Bone strength is reflected by measuring bone mineral density or bone mineral content [90], and changes in bone mass are interpreted using markers of bone status assessment (such as osteocalcin, OC; procollagen type I N-terminal propeptide, P1NP; procollagen type I C-terminal propeptide, P1CP; bone alkaline phosphatase, BAP; calcitonin, CT; osteoprotegerin, OPG; tartrate resistant acid phosphatase, TRAP). Skeletal muscle mass, function and muscle fiber changes are measured to assess maximum voluntary isometric contraction of the calf (mainly type I fibers) and maximum voluntary isometric force of the quadriceps/hamstrings (mainly type II fibers) [13]. Reliable non-invasive measurements of muscle function, such as muscle fiber type composition, muscle fiber size, cross-sectional area, etc., can be performed using surface electromyography [91].

### Cardiovascular system measurements

Cardiovascular function is reflected by measuring heart rate variability (HRV), cardiac and macrovascular morphology and function, and endothelial status [92]. HRV is recorded using a 24-h

EKG and autonomic activity is assessed by time- and frequency-domain indices of HRV analysis [93]. Left ventricular diastolic volume, output per beat, cardiac output, aortic velocity and myocardial thickness are estimated to characterize cardiac morphology and function. Carotid intima-media thickness, carotid artery dilatability and portal diameter are estimated to characterize the morphology and function of the great vessels.

### Immune system measurements

Immune cells, cytokines, chemokines, proinflammatory and regulatory proteins are all involved in immune regulation and induction of inflammation in the body. Absolute leukocyte counts and percentages of each type of leukocyte are measured in whole blood samples by a hematology analyzer, and peripheral blood immunophenotyping is performed by flow cytometry [51].

### Measurement of brain change

Numerous studies have revealed that spaceflight influences the brain's macrostructure as well as the microstructure and connectivity of brain tissue. Of these, the integrity of the central nervous system and the brain is the primary concern [94]. Cortical activity before and after exercise is recorded using electroencephalography (EEG) [95]. Neuronal and especially axonal integrity is assessed using diffusion tensor imaging [96]. Non-invasive ultrasound and lumbar puncture are used to assess intracranial pressure [29]. The cognitive abilities (Wechsler Memory Scale), visuospatial working memory (Corsi Cubes test) and spatial reasoning (Kohs Cubes test) of subjects are also measured [97].

### Sleep–wake cycle measurements

The duration of active arousal, sleep or wakeful rest is recorded using a wrist activity recorder [54]. Drowsiness and alertness are assessed using the Karolinska Sleepiness Scale and the Brief Psychomotor Vigilance Test.

### Investigation of psychological impact Stress levels

A study examining the relationship between stress and simulated flight performance assessed changes in stress awareness using the Stress Rating Questionnaire and evaluated crews' acute psychological stress state using heart rate and HRV [98]. In a Mars 105-day isolation experiment, stress levels were evaluated by tonic cortisol levels, which were measured using urinary free cortisol test-kit DKO018 Lot 1730 from DIAMETRA, Milan, Italy and the Perceived Stress Scale questionnaire. These researchers also recorded sleep EEG to investigate the relationships between stress and sleep during isolation [97].

### Emotional state

Subjects' emotional state is usually measured in the form of questionnaires and can be reflected by some hormone levels [50]. During the Mars 520-day mission, crewmembers completed a series of psychological measures including the Social Desirability Scale 17, Visual Analog Scales, Profile of Mood States—Short Form, Beck Depression Inventory and Conflict Questionnaire, which described the crews' subjective ratings of mood, psychological distress, health, stress, fatigue, sleep quality and workload [99]. In addition, levels of four plasma hormones, cortisol, 5-hydroxytryptamine, dopamine and norepinephrine were also collected and analyzed [49]. A test run with 105 days of isolation was performed prior to 520 days of isolation, and mood assessments were made using MoodMeter®, which included three

dimensions: perceived physical state (PEPS), psychological state (PSYCHO) and motivational state (MOT). Meanwhile, EEG data were recorded and correlation analysis revealed a significant relationship between mood data and electrocortical activity [50].

## Challenges in space biological effect data analysis

We highlighted the common characteristics of data generated from longitudinal experiments, which are also the major challenges faced in data analysis. For individual variables (e.g. the expression values of a gene at different time points), we considered the time-series properties of the environmental adaptation experimental design, as well as the range and trend of fluctuations. We believe that the fluctuation pattern of time series can reflect the process of biological adaptation to the environment. In addition, there also exist several obstacles to overcome in space life science data analysis, including but not limited to complex influencing factors, small sample size, high dimensionality as well as the heterogeneity of data and asynchronously changed features.

### Limited experimental subjects

Owing to the extraordinary expense of space launch payload delivery systems and the limitations of orbital platform capacity, the number of experimental replicates and variables in space flight is very limited. Despite the relatively inexpensive Earth-based experiments in support, scientific evidence is still restricted by the limited number of experimental subjects. Small replicate numbers constrain statistical power, in which case the impact of interindividual variability on statistical outcomes must be carefully evaluated. And it is necessary to carry out more experiments in space or on Earth for the advancement of the field of knowledge. Notably, each individual is usually sampled at multiple time points for various measurements in environmental adaptation experiments.

### Characteristics of individual variables produced by longitudinal studies

#### *Time-series experiments*

To detect adaptive changes in the living systems due to the space environment, the multi-level performance is usually tracked and measured before, midway and after the space flight, such as the Mars-500 mission [48], the 180-day CELSS experiment [14] and the NASA twin study [11]. The resulting measured data are time-sequenced, rather than the common case/control experimental design. Time-series experiments sample the same individual at different times and obtain multiple samples with strong autocorrelation between the measured values, more specifically the measured values at a certain time are correlated with the measured values over the previous period. In contrast, static experiments assume multiple samples are measured simultaneously and the resulting values are independent. As a result, conventional statistical analysis tools established for static data are inapplicable to time-series data analysis in operations like difference analysis, clustering analysis, missing value filling, etc. It is required to build or introduce more specialized analysis procedures.

### Changing trends within the normal range

The majority of biological adaptations induced by the space environment do not necessarily progress to a pathological state in a

short period, but rather show a pattern of progressive changes within the normal range [11]. However, these changes are still notable, given that these changes may break the threshold of the normal level in longer stays of future space travel missions [100]. On the other hand, effects within the normal range, although not pathological, can still cause stress in the body and thus increase the risk of pathogenesis [101].

### Overall characteristics of the datasets derived from multiple measurements

#### *Comprehensive and multidimensional data types*

Because biological function requires synergistic control at multiple levels, measurements of different systems at multiple levels yield multidimensional data, ranging from molecular to systemic. The Mars-500 [48], for example, measured not only multiple omics data (e.g. epigenomics, transcriptomics) but also various biochemical indexes (e.g. cortisol levels), as well as psychological assessments (e.g. mood), with a variety of data types (discrete, continuous). Organizing the multi-level datasets and extracting the information interactions between them is one of the challenges of such large studies.

#### *Asynchronous changes in different features*

Changes in organismal systems do not always happen simultaneously, and even alterations of two genes with regulatory links are not completely synchronized. Analyzing cascading changes at different levels might provide more information on causal associations. As a result, while evaluating the connection between distinct characteristics, the issue of time delay should be considered.

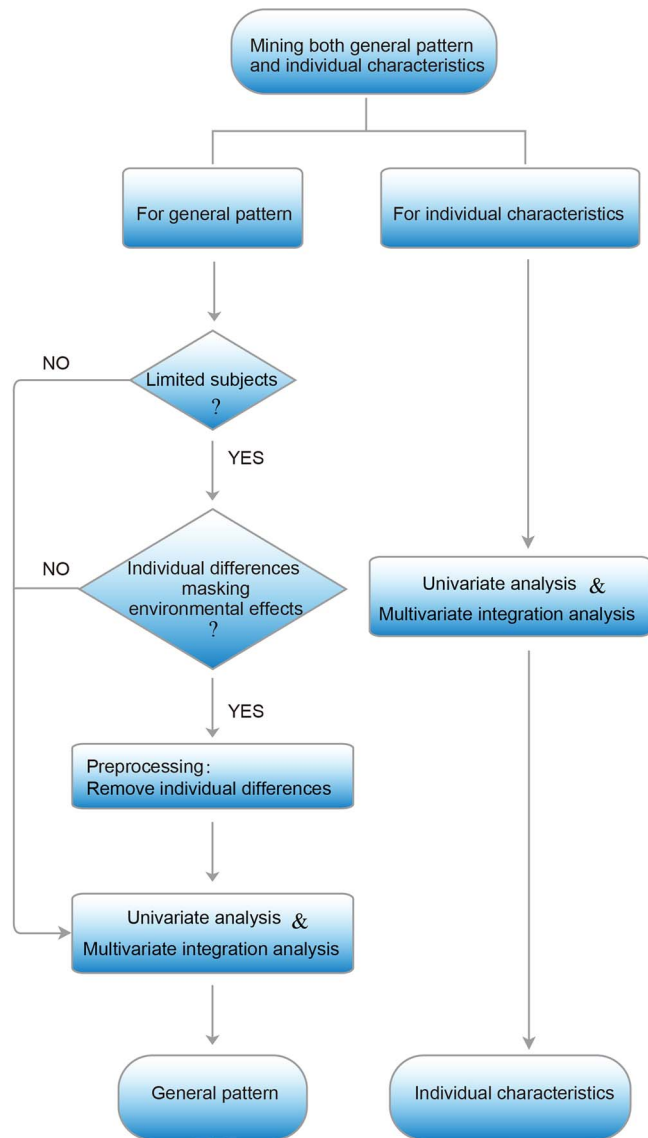
## Proposed research directions and methods

To provide solutions to the main challenges faced in longitudinal space experiment data analysis, we have compiled relevant bioinformatic strategies as well as available models/tools for data analysis. In cases where sample sizes are limited, it may be considered in determining whether individual differences mask environmental effects and should be pre-treated (Figure 4). Analyses can then be conducted including forecasting and difference analysis methods for univariate time series, the integration and classification of multiple variables and the identification of their regulating relationships. All of the above analysis methods have considered temporal attributes of variables from longitudinal space experiments.

### Mining both general pattern and individual characteristics

#### *Mining the general adaptation pattern of experimental subjects*

It is generally accepted that a larger sample size is more beneficial for mining commonalities between samples. However, due to the specificity of the spaceflight environment, the number of subjects that can participate in the experiment is limited, and each individual could be sampled several times during the process. In this case, the presence of individual differences must be carefully evaluated. We offer a perspective here by treating samples from various individuals as separate batches. The degree of interindividual differences could be assessed like batch effects, and if significant individual differences do exist, batch effect removal methods can be used to eliminate the effect of individual differences (Figure 4). We have compiled a list of common approaches for the evaluating and correcting batch effects (Figure 5).



**Figure 4.** A suggested process for mining both general pattern and individual characteristics.

Principal Variance Component Analysis (PVCA) [102] and Manifold Approximation and Projection (UMAP) [103] can be used for evaluation and visualization of batch effects. A commonly used algorithm to remove gene expression batch effects is the empirical Bayesian approach, based on which the ComBat method is more effective for small sample data [104, 105]. It can be implemented using the `combat` function of the `sva` package [106] in R. The Removing Unwanted Variation approach, which relies on negative control genes and duplicate samples to remove unwanted variance from microarray gene expression data, is more suitable for large-scale datasets [107]. BatchServer is a web server that includes autoComBat, a modified version of ComBat, as well as PVCA and UMAP, which can be used to evaluate, visualize and correct batch effects [108].

The presence of batch-correlated variation may skew analysis in two ways without batch-effect correction: false positives and false negatives. With batch-effect corrections, the results may skew according to the way how the batch effects are removed, e.g. the batch-group design, the completeness of the batch-effect

removal and appropriateness of the batch-effect removal. In a multi-category sample analysis, variations across samples can come from a variety of causes, but we are only interested in differences are the result of experimental factors. If additional non-experimental factors are causing significant batch effects, we may be unable to isolate the differences of interest. In these situations, it will be helpful to remove batch effects properly, while excessive batch effect correction may make slight differences significant, leading to false positive results. Therefore, it is necessary to conduct repeated tests to determine whether or not it is appropriate to remove the batch effect. And the degree of correction of batch effects by different methods should be compared to choose the most suitable treatment method.

### Explore individual adaptation patterns for each subject

It is of great importance to gain insights into general pattern; moreover, depicting individual characteristics matters, as health assessment and early warning will be highly personalized during spaceflight. To address this issue, both data accumulation and methods development would be crucial.

On the one hand, with the accumulation of spaceflight data, there will be a sufficient amount of cohort data as reference, it will be more easily and more directly to exact individual characteristics from general pattern of spaceflight cohort; thus, the limits of small sample size will be overcome eventually. However, it puts forward higher demands for the experiment design and data type consistency throughout sequential spaceflight missions. On the other hand, analysis methods aiming to model with insufficient data would help. In each specific analysis step, we mentioned some of the analysis methods applicable to small sample size data (Figure 5).

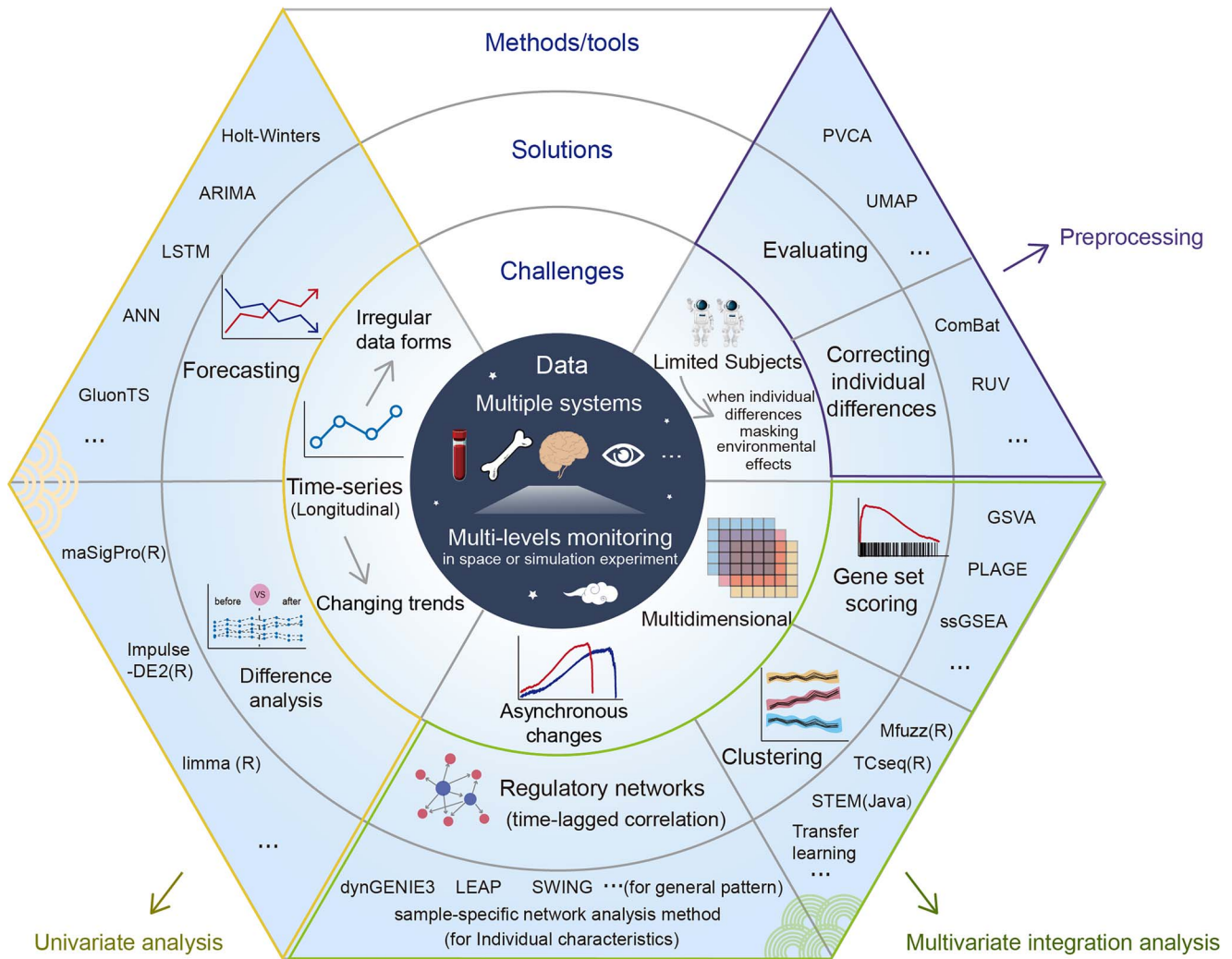
### Univariate analysis method

#### Forecasting methods on time-series biological data

Although many tools for analyzing biological datasets with time-series properties are available, irregular input data from time-series experiments often lead to inaccurate clustering results, such as missing values [109], unequal time intervals and an unequal number of time points in various features [110]. Time-series forecasting methods may be the solution to the above problems. Forecasting can estimate the values for missing data points and predict the performance of specific genes at future time points where experimental values are not available.

There are few studies dedicated to the prediction of time-series gene expression data, but many statistical and machine learning-based methods have been developed for time series forecasting in other fields. ARIMA (autoregressive integrated moving average) [111] and Holt-Winters (tri-exponential smoothing) [112] are two of the most popular and widely used statistical forecasting methods in various fields. ARIMA combines autoregressive model, moving average model and different methods to describe the autocorrelation between historical data to predict the future. It assumes that the future will repeat the historical trend, which requires the time series to be stationary [111]. The Holt-Winters model is suitable for non-stationary time series containing linear trends and periodic fluctuations, using the Exponential Smoothed Moving Average calculation method to allow the model parameters to gradually adapt to changes in the non-stationary series [112].





**Figure 5.** Challenges in biological effect measurement data analysis and proposed solutions, including forecasting and difference analysis methods for univariate time series, the integration and classification for multiple variables, and the evaluation of whether individual differences mask environmental effects and should be preprocessed.

In addition, there are time-series forecasting models based on deep learning, for example, Gluon Time Series (GluonTS) developed by Alexandrov *et al.* [113], a toolkit for probabilistic time-series modeling, focusing on deep learning-based models including different generative, discriminative and autoregressive models. And Long Short-Term Memory (LSTM) is an artificial recurrent neural network architecture model with the advantage of being relatively insensitive to gap length. Tripto *et al.* [114] evaluated Holt–Winters, ARIMA, LSTM, Artificial Neural Network (ANN) and GluonTS feedforward neural networks for forecasting time series in five sets of temporal gene expression profile data of different sizes, and found that ARIMA and ANN worked better.

#### Differential expression analysis of time-series data

Since the values of time-series data from longitudinal space experiments are probably not independent of each other, commonly used differential expression analysis methods such as *t*-tests are no longer applicable, and therefore tools for differential expression analysis dedicated to time-series data have arisen.

maSigPro (significant gene expression profile differences in time course microarray data) [115] is an R package for analyzing time-series data, supporting experiments with only time series as well as complex designs with both time series and grouping.

This R package fits the relationship between factors such as time, experimental conditions and gene expression based on a multiple linear regression model and then uses stepwise regression to find the best combination of independent variables. It can identify genes with significant expression changes by statistical procedures and cluster genes with significant expression changes over time. ImpulseDE2 [116] is another differential expression algorithm for time-course sequencing experiments that simulates temporal changes with a simple continuous function single pulse (impulse) model. ImpulseDE2 employs a noise model specific to count data from multiple batches and combines it with a likelihood ratio test, leading to a much faster and more accurate inference. It performs best when looking for differential genes in time-course data in some review articles [117]. In addition, R package limma [118] is widely used in differential expression analysis, which uses linear models to determine the size and direction of the changes in gene expression. Through borrowing information across genes, it has features that make the analyses stable even for experiments with a small number of samples. Additionally, it could handle time-series data with group information.

In summary, maSigPro is suitable for the case where samples are grouped (e.g. from male and female astronauts). The performance of ImpulseDE2 may be better when grouping is not

considered, while limma provides one more possible choice. In practice, given the poor robustness caused by small sample size, a validation among different kinds of methods could help.

## Multivariate integration analysis method

### Organization of multidimensional data types

Available sequencing technologies and computational methods allow people to obtain measurements of a wide range of analytes from the molecular to the macroscopic level. For example, at the cellular level, lymphocytes not only can be directly counted by cell sorting, but also the proportion of various lymphocytes can be estimated by computational methods based on tissue transcriptome sequencing data. A commonly used tool is the CIBERSORT method [119] which estimates the relative content of multiple immune cells by an inverse convolution algorithm. Ultimately, direct measurements or estimates of multi-level measurements can be obtained, including at molecular level, cellular level, tissue-organ level, system level, etc. The challenge is to figure out how to organize these datasets so that complex changes at various levels can be resolved.

### Data combination and scaling

One of the important reasons why different analyte measurements cannot be analyzed simultaneously is that they have different magnitudes. In the NASA twin study [11], to identify complex changes over time that occur across different analyte classes, different data types were combined and scaled for the subsequent analysis. In fact, the main focus of time-series analysis is usually on trends of change rather than specific measurements, so combining and analyzing the features at different levels by removing the magnitudes can make it easier to observe them at the same time. A simple operation is to normalize the data, which only changes the range of values without influencing their distribution.

### Clustering time-series biological data

While static experiments typically focus on common patterns among samples, and the most common analysis method is to cluster samples based on profiles, time-series experiments focus on patterns that change over time, necessitating the clustering of numerous time-series features. In the NASA twin study [11], the c-mean clustering analysis was performed to observe features with the same pattern of change. And several tools designed for clustering multiple time series according to patterns of variation are already available, which can be used in the analysis of multi-omics data or other biological datasets with time-series properties. A few commonly used tools are listed as follows:

R package Mfuzz (<http://mfuzz.sysbiolab.eu>) [120] is a clustering tool based on Fuzzy C-Means Clustering, which is a soft clustering algorithm with better noise-tolerance compared to hard clustering algorithm. It can be used to analyze transcriptomic and proteomic data with time-series properties to obtain temporal trends of gene or protein expression, and to cluster genes or proteins with similar expression patterns. TCseq package has similar functions to Mfuzz. It has more options for time-series clustering methods, including fuzzy c-mean clustering, hierarchical clustering, k-mean clustering, etc. Short Time-series Expression Miner (STEM) [121], a commonly used tool for clustering temporal expression patterns, is a Java program that can be used to cluster, compare and visualize gene expression data from short time series (typically within eight time points). STEM is based on a novel clustering algorithm. First, a unique and representative set of temporal expression sequences (pattern sequences) is selected,

and then other genes are individually assigned to the pattern group closest to that gene expression profile [122]. Also, STEM can perform functional enrichment analysis on gene sets with the same temporal expression pattern.

In addition to the above tools, machine learning algorithms can also be considered. Few-shot learning uses limited numbers of samples to build a model, the key step of which is to reduce parameter dimension and combine regularization with loss functions to resolve the overfitting problem. It can be performed through various tools like Torchmeta [123], Meta-Transfer Learning for Few-Shot Learning [124], LibFewShot [125], etc. Transfer learning reuses a pre-trained model on a different but related task. It develops rapidly in deep learning for the advantage of training with much less data, which quite fits the scenario of spaceflight. And transfer learning methods can be used for time-series classification [126]. It has been applied to solve a variety of biological problems, including but not limited to medical image analysis [127], drug discovery [128], cancer morbidity prediction [129], cancer classification [130], etc.

The above tool visualizes the multidimensional features after clustering, helping understand the dynamic patterns of these biological molecules over time. Based on the resulting clusters, some interesting sets of genes or other features from the graph can be identified, such as certain clustered groups of genes showing the expected trend of increasing or decreasing over time, or observing a clear inflection point at a certain time point, etc.

### Gene set scoring

In differential expression analyses, a high level of significance is usually selected and some subtle gene expression changes are ignored. Such subtle changes are generally considered insignificant, but assuming that a set of genes that perform similar functions are all slightly altered, it may result in significant changes in that function. Therefore, detecting overall differences in the activity of a functionally important gene set can compensate for subtle changes missed by single-gene differential analysis. Four unsupervised, single sample enrichment methods have been developed, Gene Set Variation Analysis [131], Pathway Level Analysis of Gene Expression [132], single sample GSEA [133] and the combined z-score [134]. For small datasets (the number of samples < 25), the singscore method may help. All genes are first sorted by expression level and then an enrichment score is calculated based on the position of the gene set in the overall sort, which can be used to assess the activity of each gene set in each sample. Once the activity of the gene sets is obtained, how they change over time can be analyzed. In addition, the idea of integrating features with similar meanings can be extended to analyze other types of high-dimensional biological data.

### Association prediction between different features

In addition to tracking trends in features over time, it is valuable to broadly predict associations between different analytes, such as transcriptional regulatory networks. In biology, constructing regulatory networks is a typical approach to addressing causes and correlation issues. Predicting regulatory relationships based on time-series expression data has unique advantages and challenges. Compared with static expression data, time-series expression data of the same size contain additional information due to temporal order, which can be utilized to develop regulatory networks. However, there are also some challenges. First, time-series expression data usually detect a few time points, which have a great impact on the estimation of model parameters. Another issue is the time difference between changes in multiple

**Table 1.** Available methods/tools for regulatory network prediction

Method	Type	Open source availability	Short summary	Link
LEAP	Correlation	Yes	LEAP infers gene regulatory networks based on gene co-expression relationships and considers possible lags in time.	<a href="https://cran.r-project.org/web/packages/LEAP/index.html">https://cran.r-project.org/web/packages/LEAP/index.html</a>
dynGENIE3	Regression	Yes	dynGENIE3 extends GENIE3 by considering changes in expression over time and building dynamic models based on ordinary differential equations.	<a href="https://github.com/vahuynh/dynGENIE3">https://github.com/vahuynh/dynGENIE3</a>
Inferelator	Regression	Yes	Inferelator infers gene regulatory networks by selecting the regulators whose levels are most predictive of gene expression based on a LASSO regression model.	<a href="https://github.com/baliga-lab/cMonkeyNwinf">https://github.com/baliga-lab/cMonkeyNwinf</a>
SWING	Granger causality	Yes	SWING is a gene regulatory network inference framework based on multivariate Granger causality and sliding window regression.	<a href="https://github.com/bagherilab/SWING">https://github.com/bagherilab/SWING</a>
DREM	Probabilistic graph model	Yes	DREM integrates time-series gene expression data and static or dynamic transcription factor–gene interaction data (e.g. ChIP-seq data) and produces as output a dynamic regulatory map.	<a href="http://sb.cs.cmu.edu/drem/">http://sb.cs.cmu.edu/drem/</a>

LEAP, lag-based expression association for pseudotime-series; dynGENIE3, dynamical GENIE3; SWING, sliding window inference for network generation; DREM, Dynamic Regulatory Events Miner.

dimensions, and even transcriptional regulation between genes has a delay problem. Many tools for gene regulatory network prediction based on time-series gene expression data are now available, and we exemplify some representative ones (Table 1).

One class of regulatory network prediction methods is based on correlation, for example, LEAP (lag-based expression association for pseudotime series) [135], which considers the time-lagged correlation of one gene before another and therefore predicts the directed regulatory relationship. The positive/negative coefficients represent the activation/repression regulation between genes. LEAP considers all possible time spans of lags to search for the maximum correlation for each gene pair to construct the regulatory network. There are also several regression-based methods available for determining dynamic interactions in time-series expression data. These include dynGENIE3 (dynamical GENIE3) [136], a modified method based on GENIE3 (gene network inference with ensemble of trees), which is a model-free method for inferring networks based on static expression data [137]. GENIE3 combines regression and random forest (RF) to determine the regulator for each target gene, providing excellent scalability and ease of use due to its non-parametric nature. The improved dynGENIE3 models changes in expression over time with ordinary differential equations (ODEs) and then learns the putative gene interactions using an RF regression framework. Another similar tool is Inferelator [138], which combines regression and ODE to reveal gene regulatory relationships. In addition, Granger causality test is also a time-series regulatory prediction method. Granger causality test is a statistical method for hypothesis testing, which is based on the autoregressive model in regression analysis and can be used to test whether there is a causal relationship between time series. SWING (sliding window inference for network generation) is a tool based on this statistical method [139]. Probabilistic graphical model is another widely used method for inferring interaction networks from time-series data. Based on this method, Dynamic Regulatory Events Miner (DREM) [140] integrates time-series gene expression data and static protein DNA interaction data (e.g. ChIP-Seq data) using input-output hidden Markov models to produce dynamic regulatory maps as output. Dynamic regulatory maps highlight the major divergence

events in the time-series expression data and the transcription factors that may be responsible for them.

All of these methods can predict gene regulatory networks based on temporal data, and fully consider the time delay. Correlation-based methods (e.g. LEAP) are the fastest and more suitable for large datasets, while regression-based (e.g. dynGENIE3, Inferelator, SWING) or probabilistic graphical models (e.g. DREM) are more computationally intensive but are expected to be more accurate. DREM method combines data on protein–DNA interactions and the predicted regulatory relationships are more reliable. The other methods that only consider expression value associations may be extended to predict associations between different measures (e.g. between gene expression and phenotype).

As health disorders in spaceflight are complex, it is crucial to uncover the underlying biological mechanism of each individual. Liu *et al.* [141] developed a sample-specific network analysis method to meet this demand, which implements personalized characterization of disorders.

## Future directions in spaceflight biology research

In the future, the costs and hazards of manned space flight may become more affordable to support burgeoning space tourism. The larger sample sizes and more diverse study populations will provide unprecedented opportunities for spaceflight biology research. Some humans may leave Earth and establish permanent bases and larger settlements on the Moon, Mars or elsewhere. Under longer exposure, slight changes in short-term space missions may develop health hazards [100]. When space migration programs become a reality, human populations in the new environment may have evolved to distinct genotypes, at which point an immigration genome project may even be conducted. For these ambitious frontiers, developments in the fields of space biology and aerospace medicine are crucial enablers. Furthermore, multi-omics, longitudinal profiling can capture the combined effects of multiple space environment factors as well as interactions between multiple levels, paving the way for a thorough examination of space biological adaptations.

There are still many mechanisms of space biological adaptations unknown, such as the mechanisms of telomere length dynamics and their long-term consequences. Moreover, some trends within the normal range that have been overlooked in previous studies may also be of interest. Additional studies and systematic research protocols will provide more comprehensive insights. And it still takes a lot of effort to transform detected data into interpretable results, and a systematic analysis process will speed up the process. This review clarifies each method has a specific range of applicability when compiling the known available data analysis methods. Appropriate methods must be chosen based on the data characteristics in a given study. Especially, determining whether interindividual differences should be removed requires careful assessment of the data distribution and considerations of whether this action would disrupt time trends from a single individual.

In bioinformatics, machine learning has become a popular and successful method for extracting knowledge from big data. While traditional machine learning relies on feature selection, deep learning overcomes these limitations to demonstrate advanced performance in bioinformatics problems [142], such as splice site discovery from DNA sequences [143], finger joint identification from X-ray images [144], error detection from EEG signals, etc. However, because most deep learning approaches require appropriate and balanced data to optimize numerous weight parameters in a neural network, they are usually not applicable to restricted and unbalanced data in bioinformatics [120]. This is due to the need to optimize a large number of weight parameters in neural networks. Biological studies usually contain small sample sizes that limit statistical power, in which case simple models with fewer parameters may be more suitable while more parameters may introduce additional errors and overfit. Deep learning is still making efforts to improve interpretability. Both the assessment of applicability of existing methods and the proposal of new improved methods are necessary processes to perform human spatial-omics analysis.

The study of biological adaptations has led to a deeper understanding of the needs of astronauts. In response to these needs, researchers have made many attempts to improve the quality of life of astronauts, which is the ultimate goal of future biological research in space. For example, space synthetic biology aims to leverage local resources to manufacture critical products for the crew. The Space Synthetic Biology (SynBio) project conducted at NASA's Ames Research Center in California's Silicon Valley is concentrating on developing in-space nutrient production methods and microbial biomanufacturing technologies that chemically convert carbon dioxide (CO<sub>2</sub>) and water into organic compounds for 'feeding' microbes to produce food, pharmaceuticals, plastics, etc.

Currently, most of the researches on space response studies are scattered across tissues or systems and lack consideration of temporality. The emergence of spatiotemporal molecular medicine promises to provide more comprehensive insights by integrating clinical spatialization, temporalization, phenomics and molecular multi-omics to present a four-dimensional dynamic picture of disease [125]. The perspective of spatialization encompasses genetics, population distribution and intra-individual location. The temporal perspective considers the disease's initiation and progression, clinical phenotype changes over time and patient response to treatment. When depicting overall body changes in space, it is essential to note that they were multisystemic related and duration time-dependent. Application of perspectives from spatio-temporal molecular medicine in space physiopathology

studies may provide a holistic and dynamic picture. Some aging system research programs that combine temporal, spatial (structural organization) and molecular processes [145] may also serve as references for studying temporal changes.

### Key Points

- The compilation of previous biological response investigation results not only contributes to refining the process of biological adaptation to the spaceflight environment but also reveals many parts to be complemented.
- The collation of multi-level measurements, data types and the biological functions they reflect can be referenced by researchers in designing biological experiments.
- A summary of common features of data generated from longitudinal biological experiments related to space environment factors suggests challenges or caveats in data analysis.
- This review provides strategies and models/tools to address the challenges in data analysis from a bioinformatics perspective for different analytical goals.

### Funding

Space Medical Experiment Project of China Manned Space Program (HYZHXM01004); State Key Laboratory of Space Medicine Fundamentals and Application (SMFA19A03, SMFA19C01, SMFA19B01); National Natural Science Foundation of China (31871322, 31900473).

### References

1. Warren LE. International Space Station open-source data. *Patterns* 2020;**1**:100172.
2. Pei W, Hu W, Chai Z, et al. Current status of space radiobiological studies in China. *Life Sci Space Res* 2019;**22**:1–7.
3. SpaceX's astronaut launch is a boost for the International Space Station. *Nature* 2020;**582**:8.
4. Di Battista R. Towards a unified eulerian modeling framework for two-phase flows: geometrical small scale phenomena and associated flexible computing strategies. *Institut Polytechnique de Paris* 2021. <https://www.theses.fr/2021IPPAX055>.
5. Hassler DM, Zeitlin C, Wimmer-Schweingruber RF, et al. Mars' surface radiation environment measured with the Mars Science Laboratory's Curiosity rover. *Science* 2014;**343**:1244797.
6. Afshinnekoo E, Scott RT, MacKay MJ, et al. Fundamental biological features of spaceflight: advancing the field to enable deep-space exploration. *Cell* 2020;**183**:1162–84.
7. Kandarpa K, Schneider V, Ganapathy K. Human health during space travel: an overview. *Neurol India* 2019;**67**:S176–s181.
8. Li Y-H, Qu L-N, Chen H-L. Space stress injury and related protective measures. *Sheng Li Ke Xue Jin Zhan* 2013;**44**:354–8.
9. Xu D, Zhao X, Li Y, et al. The combined effects of X-ray radiation and hindlimb suspension on bone loss. *J Radiat Res* 2014;**55**:720–5.
10. Ruyters G, Braun M, Stang KM. Success stories: incremental progress and scientific breakthroughs in life science research. *Breakthroughs in Space Life Science Research Springer* 2021;43–113.

11. Garrett-Bakelman FE, Darshi M, Green SJ, et al. The NASA Twins Study: a multidimensional analysis of a year-long human spaceflight. *Science* 2019;**364**:eaau8650.
12. Berrios DC, Galazka J, Grigorev K, et al. NASA GeneLab: interfaces for the exploration of space omics data. *Nucleic Acids Res* 2021;**49**:D1515–22.
13. Gaffney CJ, Fomina E, Babich D, et al. The effect of long-term confinement and the efficacy of exercise countermeasures on muscle strength during a simulated mission to Mars: data from the Mars500 study. *Sports Med Open* 2017;**3**:40.
14. Yuan M, Custaud M-A, Xu Z, et al. Multi-system adaptation to confinement during the 180-day controlled ecological life support system (CELSS) experiment. *Front Physiol* 2019;**10**:575.
15. Hammond TG, Allen PL, Birdsall HH. Effects of space flight on mouse liver versus kidney: gene pathway analyses. *Int J Mol Sci* 2018;**19**:4106.
16. Ingber D. How cells (might) sense microgravity. *FASEB J* 1999;**13**(Suppl):S3–15.
17. Gridley DS, Mao XW, Tian J, et al. Genetic and apoptotic changes in lungs of mice flown on the STS-135 mission in space. *In Vivo* 2015;**29**:423–33.
18. Prasad B, Grimm D, Strauch SM, et al. Influence of microgravity on apoptosis in cells, tissues, and other systems in vivo and in vitro. *Int J Mol Sci* 2020;**21**:9373.
19. Topal U, Zamur CJS. Microgravity, stem cells, and cancer: a new hope for cancer treatment. *Stem Cells Int* 2021;**2021**:5566872.
20. Nassef MZ, Melnik D, Kopp S, et al. Breast cancer cells in microgravity: new aspects for cancer research. *Int J Mol Sci* 2020;**21**:7345.
21. Ambrosini G, Adida C, Altieri DC. A novel anti-apoptosis gene, survivin, expressed in cancer and lymphoma. *Nat Med* 1997;**3**:917–21.
22. Masiello MG, Cucina A, Proietti S, et al. Phenotypic switch induced by simulated microgravity on MDA-MB-231 breast cancer cells. *Biomed Res Int* 2014;**2014**:652434.
23. Ma X, Pietsch J, Wehland M, et al. Differential gene expression profile and altered cytokine secretion of thyroid cancer cells in space. *FASEB J* 2014;**28**:813–35.
24. Grimm D, Grosse J, Wehland M, et al. The impact of microgravity on bone in humans. *Bone* 2016;**87**:44–56.
25. Shen M, Frishman WH. Effects of spaceflight on cardiovascular physiology and health. *Cardiol Rev* 2019;**27**:122–6.
26. Blaber E, Marcal H, Burns BP. Bioastronautics: the influence of microgravity on astronaut health. *Astrobiology* 2010;**10**:463–73.
27. Yang JQ, Jiang N, Li ZP, et al. The effects of microgravity on the digestive system and the new insights it brings to the life sciences. *Life Sci Space Res* 2020;**27**:74–82.
28. Panesar SS, Fernandez-Miranda JC, Kliot M, et al. Neurosurgery and manned spaceflight. *Neurosurgery* 2020;**86**:317–24.
29. Swinney CC, Allison Z. Spaceflight and neurosurgery: a comprehensive review of the relevant literature. *World Neurosurg* 2018;**109**:444–8.
30. Mader TH, Gibson CR, Pass AF, et al. Optic disc edema, globe flattening, choroidal folds, and hyperopic shifts observed in astronauts after long-duration space flight. *Ophthalmology* 2011;**118**:2058–69.
31. Philpott DE, Corbett R, Turnbull C, et al. Cosmic ray effects on the eyes of rats flown on Cosmos No. 782, experimental K-007. *Aviat Space Environ Med* 1978;**49**:19–28.
32. Furukawa S, Nagamatsu A, Neno M, et al. Space radiation biology for "living in space". *Biomed Res Int* 2020;**2020**:4703286.
33. Hall E, Giaccia AJP. *Radiobiology for the Radiologist*, 6th edn. *J Radiother Pract* 2006;**5**:237–237.
34. Sridharan DM, Asaithamby A, Bailey SM, et al. Understanding cancer development processes after HZE-particle exposure: roles of ROS, DNA damage repair and inflammation. *Radiat Res* 2015;**183**:1–26.
35. Costes SV, Chiolo I, Pluth JM, et al. Spatiotemporal characterization of ionizing radiation induced DNA damage foci and their relation to chromatin organization. *Mutat Res* 2010;**704**:78–87.
36. Willey JS, Britten RA, Blaber E, et al. The individual and combined effects of spaceflight radiation and microgravity on biologic systems and functional outcomes. *J Environ Sci Health C Toxicol Carcinog* 2021;**39**:129–79.
37. Davis CM, Allen AR, Bowles DE. Consequences of space radiation on the brain and cardiovascular system. *J Environ Sci Health C Toxicol Carcinog* 2021;**39**:180–218.
38. Todd P, Peca M, Fleshner M. Combined effects of space flight factors and radiation on humans. *Mutat Res* 1999;**430**:211–9.
39. Limoli CL, Ponnaiya B, Corcoran JJ, et al. Genomic instability induced by high and low LET ionizing radiation. *Adv Space Res* 2000;**25**:2107–17.
40. Imaoka T, Nishimura M, Daino K, et al. Risk of second cancer after ion beam radiotherapy: insights from animal carcinogenesis studies. *Int J Radiat Biol* 2019;**95**:1431–40.
41. Dang B, Yang Y, Zhang E, et al. Simulated microgravity increases heavy ion radiation-induced apoptosis in human B lymphoblasts. *Life Sci* 2014;**97**:123–8.
42. De Zio D, Cianfanelli V, Cecconi F. New insights into the link between DNA damage and apoptosis. *Antioxid Redox Signal* 2013;**19**:559–71.
43. Voorhies AA, Mark Ott C, Mehta S, et al. Study of the impact of long-duration space missions at the International Space Station on the astronaut microbiome. *Sci Rep* 2019;**9**:9911.
44. Feng Q, Lan X, Ji X, et al. Time series analysis of microbiome and metabolome at multiple body sites in steady long-term isolation confinement. *Gut* 2021;**70**:1409–12.
45. Siddiqui R, Akbar N, Khan NA. Gut microbiome and human health under the space environment. *J Appl Microbiol* 2021;**130**:14–24.
46. Ananthakrishnan AN, Singal AG, Chang LJCG, et al. The gut microbiome and digestive health—a new frontier. *Clin Gastroenterol Hepatol* 2019;**17**:215–7.
47. Zhang Y, Moreno-Villanueva M, Krieger S, et al. Transcriptomics, NF-kappaB pathway, and their potential spaceflight-related health consequences. *Int J Mol Sci* 2017;**18**:1811.
48. Liang F, Lv K, Wang Y, et al. Personalized epigenome remodeling under biochemical and psychological changes during long-term isolation environment. 2019;**10**:932.
49. Wang Y, Jing X, Lv K, et al. During the long way to Mars: effects of 520 days of confinement (Mars500) on the assessment of affective stimuli and stage alteration in mood and plasma hormone levels. *PLoS One* 2014;**9**:e87087.
50. Schneider S, Brummer V, Carnahan H, et al. Exercise as a countermeasure to psycho-physiological deconditioning during long-term confinement. *Behav Brain Res* 2010;**211**:208–14.
51. Yi B, Rykova M, Feuerecker M, et al. 520-d Isolation and confinement simulating a flight to Mars reveals heightened immune responses and alterations of leukocyte phenotype. *Brain Behav Immun* 2014;**40**:203–10.
52. Turrioni S, Rampelli S, Biagi E, et al. Temporal dynamics of the gut microbiota in people sharing a confined environment, a 520-day ground-based space simulation, MARS500. *Microbiome* 2017;**5**:39.

53. Yang F, Liu Y, Chen S, et al. A GABAergic neural circuit in the ventromedial hypothalamus mediates chronic stress-induced bone loss. *J Clin Invest* 2020;**130**:6539–54.
54. Basner M, Dinges DF, Mollicone D, et al. Mars 520-d mission simulation reveals protracted crew hypokinesia and alterations of sleep duration and timing. *Proc Natl Acad Sci U S A* 2013;**110**:2635–40.
55. Guo JH, Qu WM, Chen SG, et al. Keeping the right time in space: importance of circadian clock and sleep for physiology and performance of astronauts. *Mil Med Res* 2014;**1**:23.
56. Palinkas LA, Suedfeld P. Psychosocial issues in isolated and confined extreme environments. *Neurosci Biobehav Rev* 2021;**126**:413–29.
57. da Silveira WA, Fazelinia H, Rosenthal SB, et al. Comprehensive multi-omics analysis reveals mitochondrial stress as a central biological hub for spaceflight impact. *Cell* 2020;**183**:1185, e1120–201.
58. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* 2016;**352**:aad9780.
59. Chowdhury D, Keogh MC, Ishii H, et al. Gamma-H2AX dephosphorylation by protein phosphatase 2A facilitates DNA double-strand break repair. *Mol Cell* 2005;**20**:801–9.
60. Fernandez-Capetillo O, Lee A, Nussenzweig M, et al. H2AX: the histone guardian of the genome. *DNA Repair* 2004;**3**:959–67.
61. Rogakou EP, Pilch DR, Orr AH, et al. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 1998;**273**:5858–68.
62. Gut P, Verdin E. The nexus of chromatin regulation and intermediary metabolism. *Nature* 2013;**502**:489–98.
63. Roadmap Epigenomics C, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–30.
64. Nature EPCJ. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57.
65. Knoll M, Lodish HF, Sun L. Long non-coding RNAs as regulators of the endocrine system. *Nat Rev Endocrinol* 2015;**11**:151–60.
66. Alvarez-Dominguez JR, Bai Z, Xu D, et al. De novo reconstruction of adipose tissue transcriptomes reveals long non-coding RNA regulators of brown adipocyte development. *Cell Metab* 2015;**21**:764–76.
67. Schulze A, Downward J. Navigating gene expression using microarrays—a technology review. *Nat Cell Biol* 2001;**3**:E190–5.
68. Duggan DJ, Bittner M, Chen Y, et al. Expression profiling using cDNA microarrays. *Nat Genet* 1999;**21**:10–4.
69. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;**10**:57–63.
70. Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 2011;**12**:87–98.
71. Wang Y, Zhao Y, Bollas A, et al. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 2021;**39**:1348–65.
72. Lin B, Hui J, Mao H. Nanopore technology and its applications in gene sequencing. *Biosensors* 2021;**11**:214.
73. McIntyre ABR, Rizzardi L, Yu AM, et al. Nanopore sequencing in microgravity. *Microgravity* 2016;**2**:16035.
74. Selevsek N, Chang CY, Gillet LC, et al. Reproducible and consistent quantification of the *Saccharomyces cerevisiae* proteome by SWATH-mass spectrometry. *Mol Cell Proteomics* 2015;**14**:739–49.
75. Beck HC, Nielsen EC, Matthiesen R, et al. Quantitative proteomic analysis of post-translational modifications of human histones. *Mol Cell Proteomics* 2006;**5**:1314–25.
76. Mann M, Jensen ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 2003;**21**:255–61.
77. Wu R, Haas W, Dephoure N, et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods* 2011;**8**:677–83.
78. Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* 2010;**11**:427–39.
79. Dettmer K, Aronov PA, Hammock BD. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* 2007;**26**:51–78.
80. Patti GJ, Yanes O, Siuzdak G. Innovation: metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* 2012;**13**:263–9.
81. Steuer R. Review: on the analysis and interpretation of correlations in metabolomic data. *Brief Bioinform* 2006;**7**:151–8.
82. Madsen R, Lundstedt T, Trygg J. Chemometrics in metabolomics—a review in human disease diagnosis. *Anal Chim Acta* 2010;**659**:23–33.
83. Dong HS, Chen P, Yu YB, et al. Simulated manned Mars exploration: effects of dietary and diurnal cycle variations on the gut microbiome of crew members in a controlled ecological life support system. *PeerJ* 2019;**7**:e7762.
84. Senatore G, Mastroleo F, Leys N, et al. Effect of microgravity & space radiation on microbes. *Future Microbiol* 2018;**13**:831–47.
85. Milojevic T, Weckwerth W. Molecular mechanisms of microbial survivability in outer space: a systems biology approach. *Front Microbiol* 2020;**11**:923.
86. Bacci G, Mengoni A, Emiliani G, et al. Defining the resilience of the human salivary microbiota by a 520-day longitudinal study in a confined environment: the Mars500 mission. *Microbiome* 2021;**9**:152.
87. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.
88. Org E, Mehrabian M, Lusic AJ. Unraveling the environmental and genetic interactions in atherosclerosis: central role of the gut microbiota. *Atherosclerosis* 2015;**241**:387–99.
89. Houle D, Govindaraju DR, Omholt S. Phenomics: the next challenge. *Nat Rev Genet* 2010;**11**:855–66.
90. Swaffield TP, Neviasser AS, Lehnhardt K. Fracture risk in spaceflight and potential treatment options. *Aerosp Med Hum Perform* 2018;**89**:1060–7.
91. Meigal A, Fomina E. Electromyographic evaluation of countermeasures during the terrestrial simulation of interplanetary spaceflight in Mars500 project. *Pathophysiology* 2016;**23**:11–8.
92. Yuan M, Custaud MA, Xu Z, et al. Multi-system adaptation to confinement during the 180-day Controlled Ecological Life Support System (CELSS) experiment. *Front Physiol* 2019;**10**:575.
93. Vigo DE, Tuerlinckx F, Ogrinz B, et al. Circadian rhythm of autonomic cardiovascular control during Mars500 simulated mission to Mars. *Aviat Space Environ Med* 2013;**84**:1023–8.
94. Roy-O'Reilly M, Mulavara A, Williams T. A review of alterations to the brain during spaceflight and the potential relevance to crew in long-duration space exploration. *NPJ Microgravity* 2021;**7**:5.
95. Schneider S, Abeln V, Popova J, et al. The influence of exercise on prefrontal cortex activity and cognitive performance during a simulated space flight to Mars (MARS500). *Behav Brain Res* 2013;**236**:1–7.
96. Brem C, Lutz J, Vollmar C, et al. Changes of brain DTI in healthy human subjects after 520 days isolation and confinement on a simulated mission to Mars. *Life Sci Space Res* 2020;**24**:83–90.

97. Gemignani A, Piarulli A, Menicucci D, et al. How stressful are 105 days of isolation? Sleep EEG patterns and tonic cortisol in healthy volunteers simulating manned flight to Mars. *Int J Psychophysiol* 2014;**93**:211–9.
98. Dai J, Wang H, Yang L, et al. The effects of emotional trait factors on simulated flight performance under an acute psychological stress situation. *Int J Occup Saf Ergon* 2021;**12**:1–8.
99. Basner M, Dinges DF, Mollicone DJ, et al. Psychological and behavioral changes during confinement in a 520-day simulated interplanetary mission to Mars. *PLoS One* 2014;**9**: e93298.
100. Li Y, Wan Y, Qu L, et al. Needs and challenges of space medicine in China's follow-up manned space missions. *Manned Spaceflight* 2007;**1**:4–7.
101. Cagampang FR, Poore KR, Hanson MA. Developmental origins of the metabolic syndrome: body clocks and stress responses. *Brain Behav Immun* 2011;**25**:214–20.
102. Li J, Bushel PR, Chu TM, et al. Principal variance components analysis: estimating batch effects in microarray gene expression data. *Sources and Solutions* 2009;141–54.
103. McInnes L, Healy J, JAJPA M. Umap: uniform manifold approximation and projection for dimension reduction. 2018.
104. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2006;**8**:118–27.
105. Müller C, Schillert A, Röthmeier C, et al. Removing batch effects from longitudinal gene expression - quantile normalization plus ComBat as best approach for microarray transcriptome data. *PLoS One* 2016;**11**:e0156594.
106. Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**: 882–3.
107. Jacob L. RUV for normalization of expression array data. *Bioconductor* 2014;3.
108. Zhu T, Sun R, Zhang F, et al. BatchServer: a web server for batch effect evaluation, visualization, and correction. *J Proteome Res* 2020;**20**:1079–86.
109. Bar-Joseph Z, Gerber G, Gifford DK, et al. A new approach to analyzing gene expression time series data. In: *Proceedings of the sixth annual international conference on Computational biology*. RECOME 2002;39–48.
110. Rueda L, Bari A, Ngom A. Clustering time-series gene expression data with unequal time intervals. In: *Transactions on Computational Systems Biology X*. Springer 2008;**5410**:100–23.
111. Tseng F-M, Tzeng G-H. A fuzzy seasonal ARIMA model for forecasting. *Fuzzy Sets Syst* 2002;**126**:367–76.
112. Kalekar PSJKRsoiT. Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi school of information Technology* 2004;**4329008**:1–13.
113. Alexandrov A, Benidis K, Bohlke-Schneider M, et al. GluonTS: probabilistic and neural time series modeling in Python. *J Mach Learn Res* 2020;**21**:1–6.
114. Tripto NI, Kabir M, Bayzid MS, et al. Evaluation of classification and forecasting methods on time series gene expression data. *PLoS One* 2020;**15**:e0241686.
115. Conesa A, Nueda MJ, Ferrer A, et al. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 2006;**22**:1096–102.
116. Fischer DS, Theis FJ, Yosef NJNar. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res* 2018;**46**:e119–9.
117. Spies D, Renz PF, Beyer TA, et al. Comparative analysis of differential gene expression tools for RNA sequencing time course data. 2019;**20**:288–98.
118. Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47–7.
119. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;**12**:453–7.
120. Kumar L, M EF. Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2007;**2**:5–7.
121. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 2006;**7**: 191.
122. Ernst J, Nau GJ, Bar-Joseph Z. Clustering short time series gene expression data. *Bioinformatics* 2005;**21**(Suppl 1): i159–68.
123. Deleu T, Würfl T, Samiei M et al. Torchmeta: a meta-learning library for pytorch *arXiv preprint* 2019;arXiv:1909.06576.
124. He K, Zhang X, Ren S, et al. Computer vision and pattern recognition. *Int J Comput Math* 2016; **84**:1265–1266.
125. Li W, Dong C, Tian P, et al. LibFewShot: a comprehensive library for few-shot learning. *arXiv preprint* 2021;arXiv:2109.04898.
126. Fawaz HI, Forestier G, Weber J, et al. Transfer learning for time series classification. In: *2018 IEEE international conference on big data (Big Data)*. IEEE 2018;**6**:1367–76.
127. Yari Y, Nguyen TV, Nguyen HTJIA. Deep learning applied for histological diagnosis of breast cancer. *IEEE Access* 2020;**8**: 162432–48.
128. Turki T, Wei Z, Wang JTJIA. Transfer learning approaches to improve drug sensitivity prediction in multiple myeloma patients. *IEEE Access* 2017;**5**:7381–93.
129. Song Q, Zheng Y-J, Sheng W-G, et al. Tridirectional transfer learning for predicting gastric cancer morbidity. *IEEE Trans Neural Netw Learn Syst* 2020;**32**:561–74.
130. Singh R, Ahmed T, Kumar A, et al. Imbalanced breast cancer classification using transfer learning. *IEEE/ACM Trans Comput Biol Bioinform* 2020;**18**:83–93.
131. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 2013;**14**:7.
132. Tomfohr J, Lu J, Kepler TB. Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics* 2005;**6**:225.
133. Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 2009;**462**:108–12.
134. Lee E, Chuang H-Y, Kim J-W, et al. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol* 2008;**4**:e1000217.
135. Specht AT, JJB L. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics* 2017;**33**:764–6.
136. Huynh-Thu VA, PJSR G. dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci Rep* 2018;**8**:1–12.
137. Petralia F, Wang P, Yang J, et al. Integrative random forest for gene regulatory network inference. *Bioinformatics* 2015;**31**:i197–205.
138. Bonneau R, Reiss DJ, Shannon P, et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol* 2006;**7**:1–16.

139. Finkle JD, Wu JJ, Bagheri N, PotNAoS. Windowed Granger causal inference strategy improves discovery of gene regulatory networks. *Proc Natl Acad Sci U S A* 2018;**115**:2252–7.
140. Schulz MH, Devanny WE, Gitter A, et al. DREM 2.0: improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst Biol* 2012;**6**:1–9.
141. Liu X, Wang Y, Ji H, et al. Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res* 2016;**44**:e164–4.
142. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform* 2017;**18**:851–69.
143. Lee T, Yoon S. Boosted categorical restricted Boltzmann machine for computational prediction of splice junctions. In: *International conference on machine learning. Proc Int Conf Mach Learn* 2015;**37**:483–92.
144. Lee S, Choi M, Choi H-s, et al. FingerNet: Deep learning-based robust finger joint detection from radiographs. In: *2015 IEEE Biomedical Circuits and Systems Conference, Atlanta, GA, USA (BioCAS) IEEE* 2015;**3**:1–4.
145. Kriete A, Sokhansanj BA, Coppock DL, et al. Systems approaches to the networks of aging. *Ageing Res Rev* 2006;**5**:434–48.