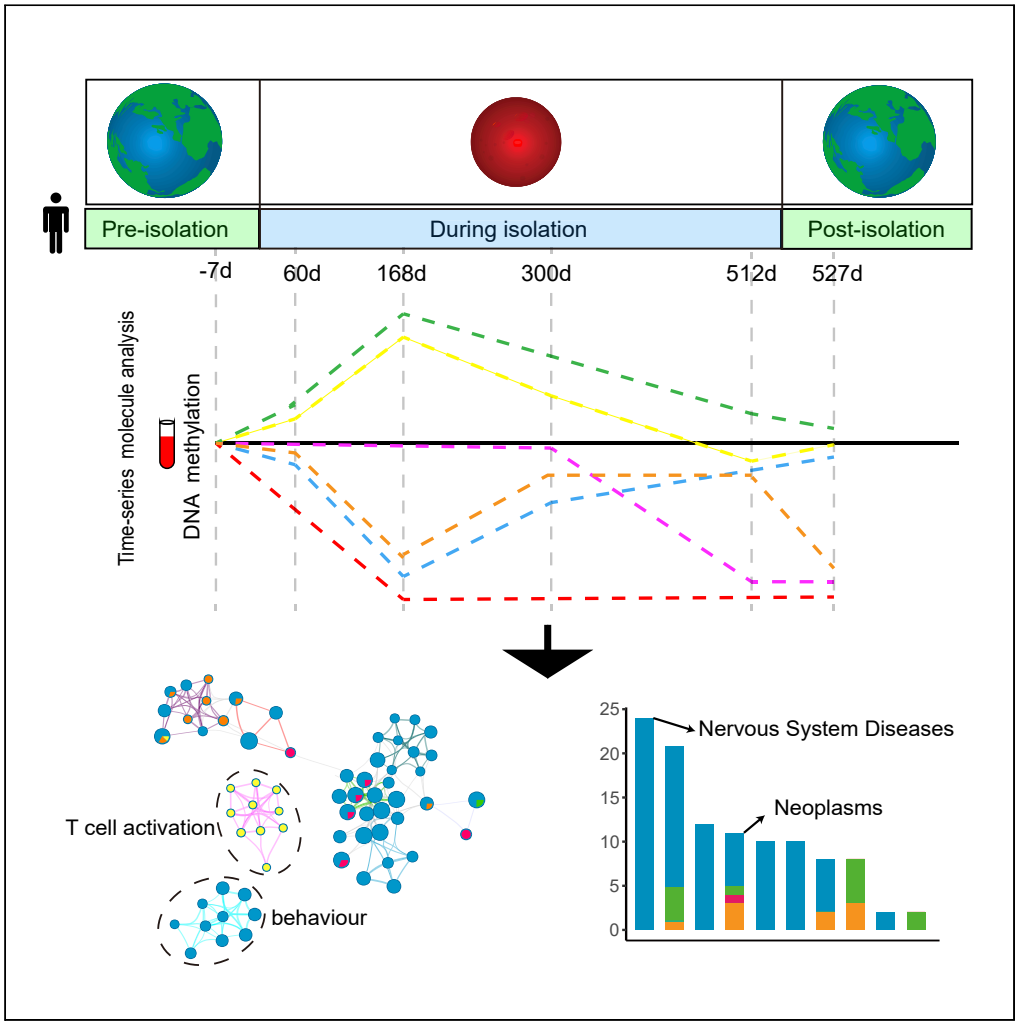
iScience

CellPress
OPEN ACCESS

Article

DNA methylation dynamics associated with long-term isolation of simulated space travel



Fei Hou, Xu Zhou, Shunheng Zhou, ..., Jicun Zhu, Xinyu Cao, Wei Jiang

weijiang@nuaa.edu.cn

Highlights

Six dynamic patterns of DNA methylation were identified during long-term isolation

Most of significantly fluctuating methylation sites recovered in post-isolation

Six patterns showed strong functional specificity

Genes with decreased methylation levels might be associated with tumor

Hou et al., iScience 25, 104493
June 17, 2022
https://doi.org/10.1016/j.isci.2022.104493

iScience

CellPress
OPEN ACCESS

Article

DNA methylation dynamics associated with long-term isolation of simulated space travel

Fei Hou,[1] Xu Zhou,[1] Shunheng Zhou,[1] Haizhou Liu,[1] Yu-e Huang,[1] Mengqin Yuan,[1] Jicun Zhu,[1] Xinyu Cao,[1] and Wei Jiang[1,2,*]

ABSTRACT

Long-term isolation is one of the risk factors that astronauts will encounter in spaceflight. At present, few researches have explored DNA methylation dynamics during long-term isolation. In this study, using time series DNA methylation data from "Mars-500" mission, we conducted a multi-step analysis to investigate the characteristics and dynamic patterns of DNA methylation as well as their functional insights during long-term isolation. The results showed that genome-wide methylation changes were minimal. In the six identified DNA methylation dynamic patterns, most of significantly fluctuating CpG sites could be returned to the baseline in post-isolation, and the remaining sites persistently decreased during isolation. Next, functional enrichment analysis of genes with each pattern revealed strong functional specificity. Some patterns were also significantly associated with nervous system diseases, digestive system diseases and cancers. In conclusion, the DNA methylation dynamics during long-term isolation have great functional significance, and might be helpful for protection of astronaut health.

INTRODUCTION

Human in long-term spaceflight will encounter five hazards, including radiation, isolation and confinement, distance from earth, gravity, and hostile and closed environments (Afshinnekoo et al., 2020), where long-term isolation has been reported to affect astronauts health in multiple aspects, including circadian rhythm disorders (Vigo et al., 2013), impaired neuromuscular performance (Belavý et al., 2013), cognitive impairment (Reed et al., 2001), impaired hippocampal functions (Leser and Wagner, 2015), increased anxiety (Wallace et al., 2009), and immune dysregulation (Crucian et al., 2014). To better deal with these potential health risks in future spaceflight missions, there are numerous studies analyzing the changes at the molecular and cellular levels during long-term isolation (Yi et al., 2014a, 2016; Strewe et al., 2015). DNA methylation, an important epigenetic modification regulating the gene expression, has also been reported to be affected by long-term isolation or spaceflight. In NASA Twins Study, the researchers discovered changes of genome-wide mean methylation levels (MML) in spaceflight, and found that genes with local methylation changes in promoters were enriched in completely different functions compared with ground control (Garrett-Bakelman et al., 2019).

The aberrant DNA methylation has been confirmed to be related to many human diseases (Portela and Esteller, 2010). Therefore, analyzing the changes of DNA methylation in long-term isolation is of great significance to protect the health of astronauts. Similar to NASA's Twins Study' conclusion, there are some researches revealing alteration of DNA methylation in response to long-term isolation or spaceflight (Ogneva et al., 2018; Ou et al., 2009; Zhou et al., 2019a). However, previous studies only focused on the differences of DNA methylation levels between isolation group and control group, the dynamic process of DNA methylation variation during long-term isolation has not been characterized.

Therefore, in the present study, we sought to investigate the DNA methylation dynamics associated with long-term isolation and to explore their functional significance. Based on 450k methylation data in bloods of six crews across six time points from "Mars-500" mission (Xiong et al., 2015), we identified six dynamic patterns of DNA methylation through a multi-step time-series analytic strategy, which included "Dec-Inc" (DNA methylation levels were firstly decreased and then increased during long-term isolation), "Inc-Dec" (DNA

[1]Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

[2]Lead contact

*Correspondence: weijiang@nuaa.edu.cn

https://doi.org/10.1016/j.isci.2022.104493

methylation levels were firstly increased and then decreased during long-term isolation), "Inc-Dec-Inc" (DNA methylation levels were firstly increased, then decreased, and then increased during long-term isolation), "Ear-Dec" (DNA methylation levels were decreased in the early stage of long-term isolation), "Mid-Dec" (DNA methylation levels were decreased in the middle stage of long-term isolation) and "Dec-Inc-Dec" (DNA methylation levels were firstly decreased, then increased, and then decreased during long-term isolation). In the six patterns, 73% CpG sites could return to the baseline in post-isolation, and remaining sites decreased compared to baseline. Pathway enrichment analysis of six patterns uncovered strong functional specificity, such as Dec-Inc genes were significantly enriched in brain development, Dec-Inc-Dec genes were significantly enriched in activation-induced cell death of T cells. In addition, we also found that several patterns were associated with some common diseases, such as nervous system diseases, digestive system diseases, and cancers. Taken together, our study characterized dynamic patterns of DNA methylation during long-term isolation, and revealed their important functional significance, which will provide new aspects for protection of astronaut health in future long-term spaceflight.

RESULTS

Global methylation characteristics during long-term isolation

We obtained 36 DNA methylation profiles of six crews spanning six time points from "Mars-500" mission, where DNA methylation levels of these samples were detected by the Illumina Infinium HumanMethylation450K BeadChip. After excluding low-quality probes, 373,309 probes were retained for the following analysis. Global DNA methylation levels of each sample obeyed typical bimodal distribution (Figure 1A). To visualize the global methylation differences between samples during the mission, we first performed t-Distributed Stochastic Neighbor Embedding (t-SNE) for all 36 samples based on the beta values. Specifically, we found that samples from the same crew tended to cluster together, which indicated there was smaller variation between time points compared to crews during long-term isolation (Figure 1B). Here we paid more attention to the impact of long-term isolation on DNA methylation, so we corrected genetic background differences between crews using "combat" (Johnson et al., 2007). As expected, the differences between crews were adjusted, and the samples were clustered together by time point (Figure 1C). The continuous two time points showed greater methylation similarity such as -7 days and 60 days, 168 days and 300 days, as well as 512 days and 527 days (Figure 1C), which illustrated that global DNA methylation experienced dynamic alteration during long-term isolation.

We next investigated the overall methylation changes among six time points by averaging the samples from the same time point. A subtle increase of median methylation level was observed from -7 days and 60 days to 168 days and 300 days, and then returned to baseline (pre-isolation: -7 days) at near the end of mission (512 days and 527 days) (Figure 1D). Furthermore, we found that most of the CpG sites underwent minor changes during long-term isolation, and only ~1% (n = 4,892) sites changed more than 0.1 (Figure 1E). Taken together, global DNA methylation changes were small during long-term isolation, but dynamic alterations between different time points can be observed.

Six DNA methylation dynamic patterns during long-term isolation

Next, a multi-step analytic strategy was performed to identify the DNA methylation dynamic patterns during long-term isolation. The multi-step strategy included filtration of inconsistent fluctuating CpGs among six crews, filtration of CpGs with small changes among six time points, identification of differentially methylated CpGs and clustering of significantly fluctuating CpGs. Table 1 presented the complete process of the multi-step analysis and the numbers of sites retained in each step.

In total, we identified 22 bases (clusters) with highly similar methylation trends, including 3,755 sites, through the aforementioned four-step analysis (Figure S2, Table S1). The similarity between bases was illustrated by a topology-preserving way (self-organizing map (SOM) grid map), where the more similar the bases were, the closer they were in SOM grid map (Figure 2A). Therefore, we further grouped the bases according to their similarity. As a consequence, we grouped all 22 bases into six DNA methylation patterns named "Dec-Inc", "Inc-Dec", "Inc-Dec-Inc", "Ear-Dec", "Mid-Dec" and "Dec-Inc-Dec", respectively (Figure 2B). Specifically, the Dec-Inc included 11 bases (base 1, 2, 3, 8, 9, 11, 12, 16, 17, 18, and 19) (Figures 2C and S2), which were close to each other in SOM grid map (Figure 2A). Similarly, the Inc-Dec included four bases (base 6, 7, 13, and 14) (Figure 2C). The Inc-Dec-Inc included one base (base 22) (Figure 2C). The Ear-Dec included two bases (base 15, 20) (Figure 2C). The Mid-Dec included one base (base 21) (Figure 2C). Finally, the Dec-Inc-Dec included three bases (base 4, 5, and 10) (Figures 2C and S2). Furthermore, the six patterns were also grouped into two classes termed "Recovering", which means

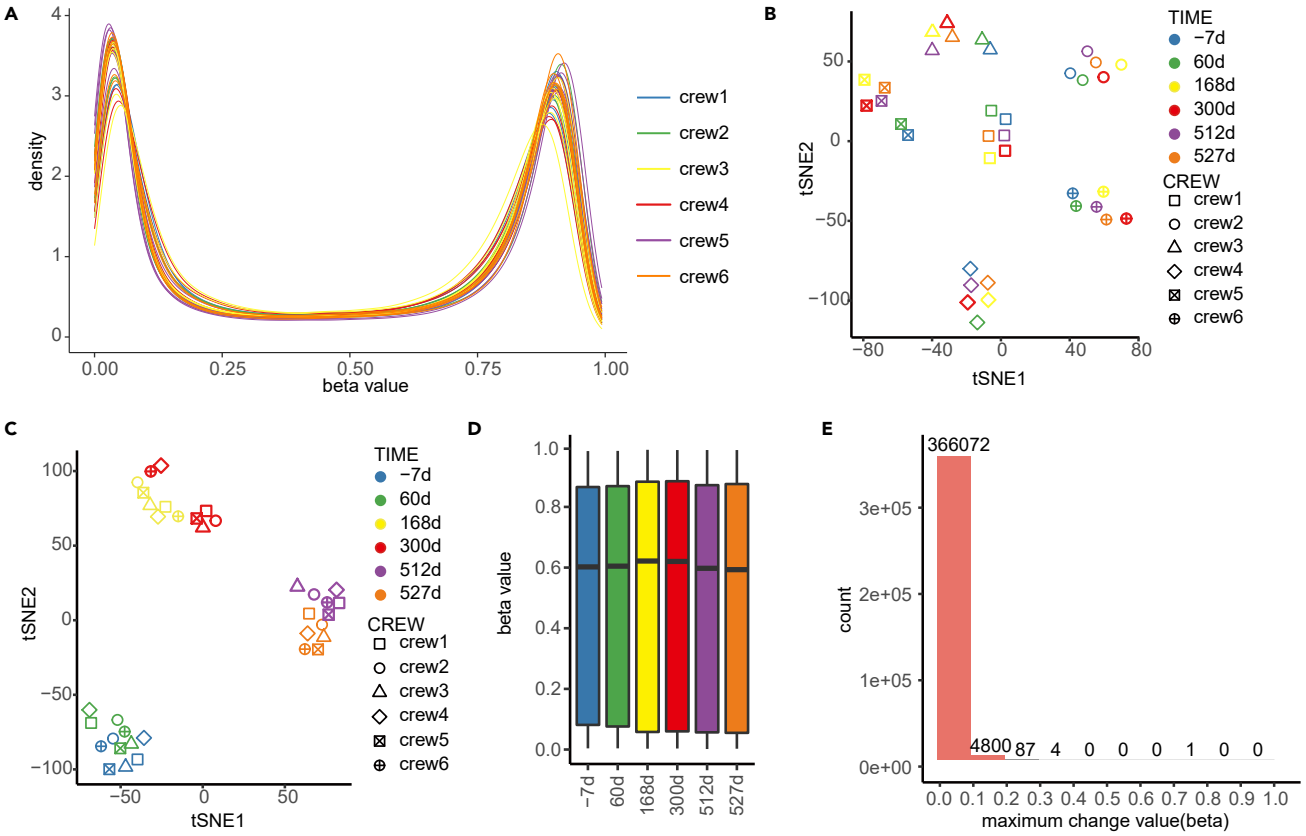


Figure 1. Global characteristics of DNA methylation in the mission

(A) Density plots representing the distribution of the beta values in all 36 samples, where samples are colored by crews.

(B) The visualization of DNA methylation profiles using t-SNE method. Each dot represents a sample colored by time points and shaped by crews.

(C) Similar to B, DNA methylation profiles were adjusted by combat.

(D) Boxplots representing the distribution of the beta values in different time points.

(E) The distribution of changes in beta values between time points. X axis represents the maximum changes of beta values calculated by maximum minus minimum in six time points for a specific CpG site.

that the methylation level in post-isolation could return to the baseline, and "Decreasing", which means that methylation level persistently decreased during isolation (Figure 2B). Intuitively, we found that most significantly fluctuating methylated sites (73%, n = 2,732) were classified as "Recovering" (Figure 2B), which is consistent with the conclusion of NASA Twins Study that genome-wide mean methylation levels (MML) inflight were subtly different from preflight, and returned to baseline postflight (Garrett-Bakelman et al., 2019). Notably, there were still 1,023 sites classified as "Decreasing" (Figure 2B). Besides, no increasing patterns were found during long-term isolation (Figure 2B). In addition, we noticed that there were two complex patterns with double turning points: Inc-Dec-Inc and Dec-Inc-Dec, where the Inc-Dec-Inc was divided into "Recovering" and the Dec-Inc-Dec was divided into "Decreasing" (Figure 2B). Although divided into different classes, they all changed at the beginning of isolation, then recovered during isolation, and finally changed again in post-isolation (Figure 2B), which indicated the two patterns may be vulnerable to the alteration of environment. Taken together, we strikingly identified six DNA methylation dynamic patterns during long-term isolation, which reflected the complex impact of long-term isolation on DNA methylation and provided us with a great support to deeply study the harm caused by long-isolation from the perspective of distinct methylation dynamic patterns.

Functional analysis of DNA methylation patterns

In terms of the six DNA methylation patterns, we further investigate their biological function. Previous studies have shown that aberrant DNA methylation located in the promoter and CpG Island played an important role in gene expression regulation (Moore et al., 2013). We noticed that a large proportion

Table 1. The number of CpG sites retained in each step of the multi-step analysis

Step	Method	Cutoff	Number of probes retained
1	ANOVA	FDR < 0.05	288,424
2	Maximum $\Delta\beta$	$\Delta\beta > 0.1$	4,196
3	EDGE	FDR < 0.05	3,777
4	SOM-SVD	FDR < 0.05	3,755

(44%) of CpG sites from the six patterns fell into the gene promoter region, whereas only 17% of the sites were in the CpG island region (Figure 3A). Moreover, almost all patterns were not significantly enriched in promoter or CpG island region, except for Ear-Dec enriched in the TSS1500 region (hypergeometric test, p value = 0.00046, Figure 3B). These observations illustrated that most methylation dynamic patterns did not specifically appear in the functional regions (promoter and CpG island).

Although the dynamic patterns were overall not enriched in functional regions, we hypothesized that local methylation changes at promoters of some genes can affect their expression. We thus identified the pattern genes for each pattern according to whether sites fell in the promoter region of genes (Table S2). Notably, only protein-coding genes specific for each pattern were retained for subsequent analysis (Figure S3). Pathway enrichment analysis of pattern genes was performed using the Metascape (Zhou et al., 2019b) (Table S3). Table 2 presents the numbers of genes and enriched functional terms for each pattern. Significantly, pathways enriched in different patterns were obviously specific (Figure 3C, Table S3), which indicated that different patterns may play different roles in maintaining the normal biological function responding to long-term isolation.

Because of the functional redundancy, we used the Metascape to cluster the enriched terms and constructed a functional network (Figure 3D). Consistent with the above conclusion, the specificity of patterns was also observed in the functional network (Figure 3D). Particularly, Dec-Inc genes were specifically enriched in some basic biological processes such as regulation of cell morphogenesis, developmental growth and segment specification (Figure 3D). Inc-Dec-Inc and Dec-Inc-Dec genes were specifically enriched in T cell activation and activation-induced cell death of T cells, respectively (Figure 3D). In particular, some of significant enriched pathways had been proved to be associated with long-term spaceflight or isolation, including brain development (Demertzi et al., 2016), behavior dysfunction (Doarn et al., 2019), immune dysregulation (Crucian et al., 2020; Garrett-Bakelman et al., 2019; Yi et al., 2014b). In addition, consistent with the result of BP of GO, KEGG pathway enrichment analysis also showed that Dec-Inc, Ear-Dec, Inc-Dec-Inc and Dec-Inc-Dec genes were enriched in obviously different pathways (Figure S4; Table S3). Taken together, pathway enrichment analysis revealed that most methylation patterns functions were uniform within the patterns but were specific among patterns, and some of these functions had been reported to be associated with long-term spaceflight and isolation, which provided a strong support for the objective existence of the patterns and a potential value for developing countermeasures in response to long-term isolation in spaceflight.

Association between methylation patterns and human diseases

To further functionally validate the patterns, we thus explored the association between the patterns and human diseases. We downloaded the human disease genes from DisGeNET (Piñero et al., 2017), and applied the hypergeometric test to determine whether pattern genes were specifically enriched in disease genes. As a consequence, Dec-Inc genes were enriched in the largest number of diseases (49 diseases), and other pattern genes (Inc-Dec, Ear-Dec and Dec-Inc-Dec) were enriched in a total 19 diseases (Table S4). Some diseases had been reported to be related to long-term spaceflight or isolation, such as Cardiovascular Diseases (Marshall-Goebel et al., 2019), Neoplasms (Nwanaji-Enwerem et al., 2020), Mental Disorders (Liang et al., 2019), Behavior and Behavior Mechanisms (Basner et al., 2014), Nutritional and Metabolic Diseases (Liang et al., 2019).

In particular, we found that Dec-Inc genes were mainly enriched in Nervous System Diseases (24/49) (Figure 4A; Table S4). Given the characteristics of Dec-Inc that the methylation level decreased and then returned to the baseline during the mission, we can speculate that the body's nervous system cannot be

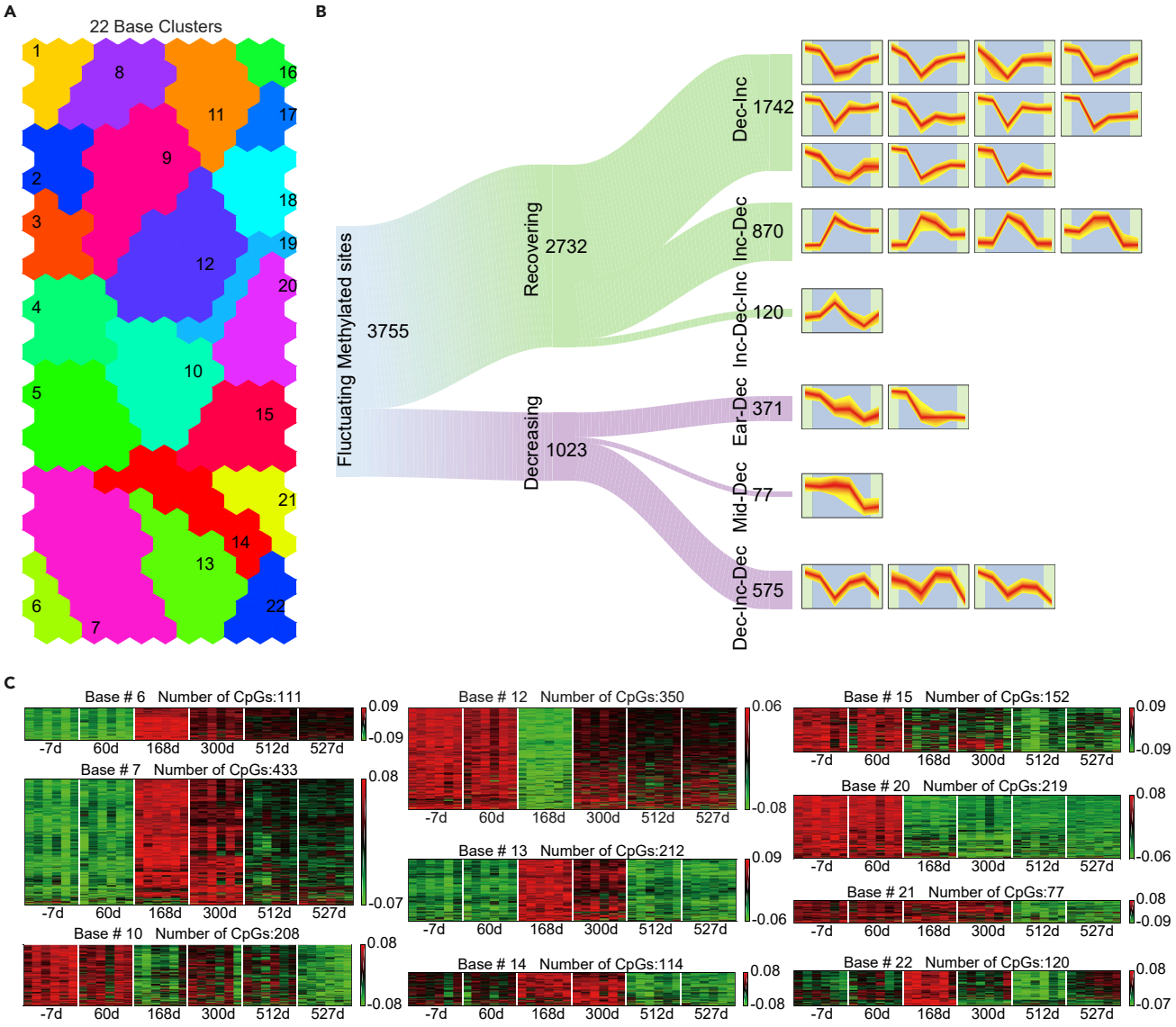


Figure 2. Six DNA methylation dynamic patterns during long-term isolation

(A) SOM grid map reflecting the 22 bases. SOM-SVD first clustered CpG sites into neurons in the SOM grid map, and then clustered neurons into bases using a region growing procedure. Seed neurons are marked with the numbers representing the index of each base.

(B) The division of six DNA methylation dynamic patterns. Each rectangular box represents a sketch map of a base cluster, where 5th and 95th percentiles of abundance are indicated (the region with gradient from yellow to red). The pre-isolation and post-isolation periods are represented by light green shaded regions; the light blue shaded regions indicate the "during isolation" period. The sankey diagram illustrates the details of pattern division, where numbers represent the counts of CpG sites from each part.

(C) Partial heatmaps illustrating the CpG sites methylation clusters in (A). Color-bars correspond to the beta values after mean centralization by row.

See also Figure S2.

irreversibly damaged though indeed affected by long-term isolation. Moreover, Inc-Dec genes were mainly enriched in Digestive System Diseases (4/7) (Figure 4A; Table S4). Similar to Dec-Inc, Inc-Dec also recovered after the mission, which meant that the long-term isolation might influence the digestive function of the body but cannot cause irreversible harm to it. In addition, two "Decreasing" patterns including Ear-Dec and Dec-Inc-Dec were also found be associated with some human diseases, where Ear-Dec genes were enriched in only neoplasms (Squamous cell carcinoma of the head and neck), and Dec-Inc-Dec genes showed diverse enrichment in multiple types of diseases including Digestive System Diseases (3/8), Neoplasms (3/8) and mental Disorders (2/8) (Figure 4A; Table S4). As we knew, promoter

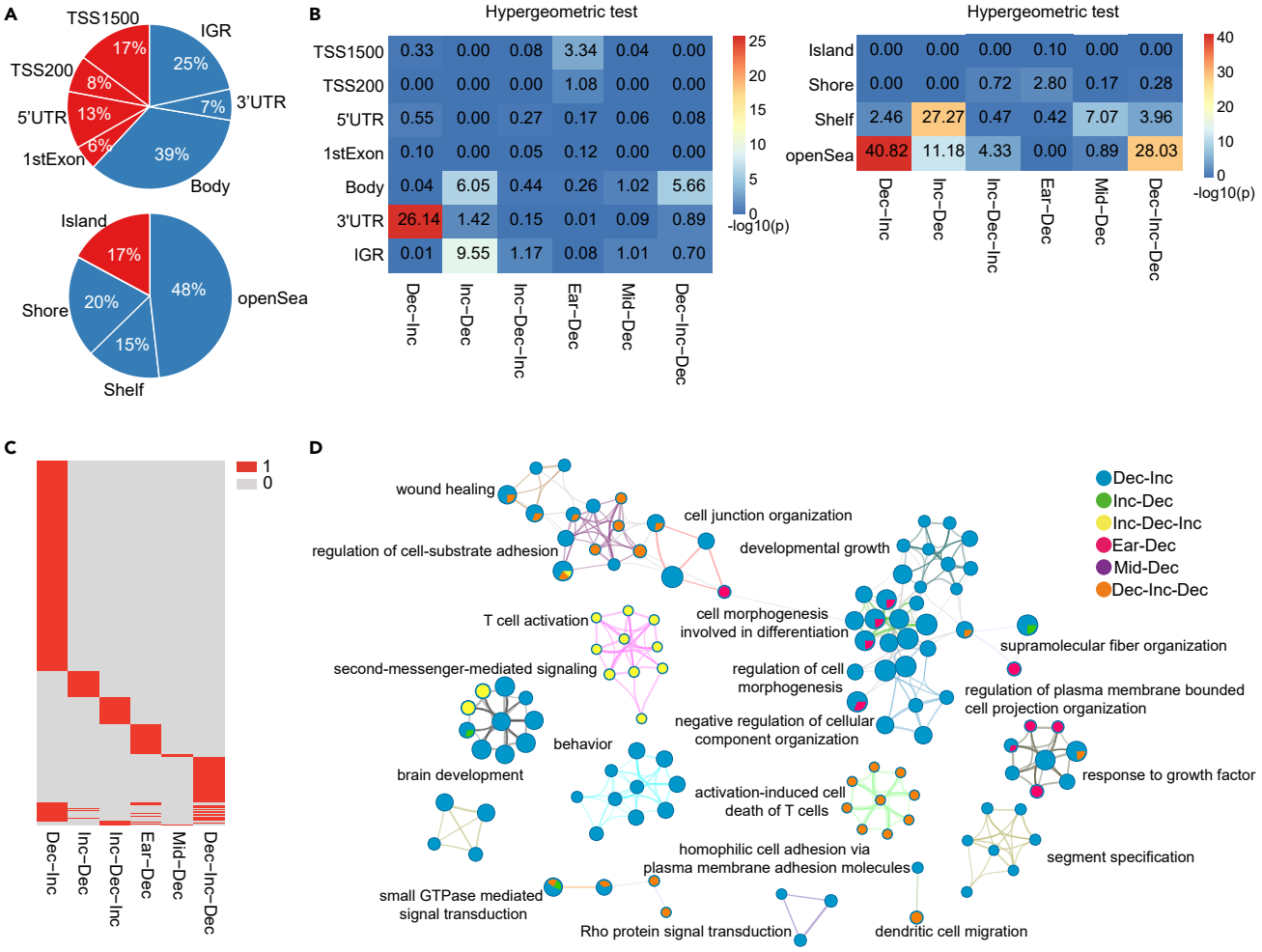


Figure 3. Functional analysis of DNA methylation pattern

(A) Pie charts depicting the proportion of significantly fluctuating methylated CpG sites in each of gene regions (top) and CpG island regions (bottom).

(B) Heatmaps representing –log10 p value of hypergeometric test. Hypergeometric test was performed for each pattern in each gene region and each CpG island region.

(C) Heatmap representing the result of pathway enrichment analysis for the GO biological process. "1" represents that the pattern genes were enriched in the pathway, and "0" means that pattern genes were not enriched.

(D) Functional network of enriched terms for six patterns. Metascape clusters all enriched terms by their similarity and showed the 20 top-score clusters. Each node represents a functional term colored by patterns and edges are colored by clusters. In addition, each node is a pie chart, where the size of the pie itself represents the number of total genes from all patterns under the term, and the size of the slices represents the percentage of genes from the corresponding pattern under the term. Only one label representing the most significant enriched term is displayed per cluster.

See also Figure S4.

methylation was often associated with gene expression repression. The two "Decreasing" patterns thus might imply an up-regulation of the diseased-associated genes, which might be a potential risk of developing the diseases. Therefore, the associations between two "Decreasing" patterns and related diseases deserve further study by researchers in future to ensure the physical and mental health of astronauts during long-term spaceflight.

Notably, Neoplasms was only disease type associated with two "Decreasing patterns", where Ear-Dec genes were enriched in "Squamous cell carcinoma of the head and neck", and Dec-Inc-Dec genes were enriched in three "Acute Myeloid Leukemia (AML)"-related diseases (Figure 4A; Table S4). Thus, we further explored the role of the two patterns in tumorigenesis to figure out the relationship between long-term isolation and cancer. We downloaded the oncogenes and tumor suppressor genes from ONGene (Liu et al., 2017) and TSGene2.0 (Zhao et al., 2016), respectively. We found that Dec-Inc-Dec genes

Table 2. The number of CpG sites, genes with CpG sites in promoter, and enriched GO terms for each DNA methylation dynamic pattern

	Dec-Inc	Inc-Dec	Inc-Dec-Inc	Ear-Dec	Mid-Dec	Dec-Inc-Dec
CpG sites	1,742	870	120	371	77	575
Genes (promoter)	673	162	34	172	17	150
Pathways (GO_BP)	267	33	38	43	4	67

(p value = 7×10^{-4}) and the union of the two patterns (p value = 0.045) were significantly enriched in tumor suppressor genes (Table 3), which indicated that long-term isolation might increase the expression of tumor suppressor genes, and thus may have an inhibitory effect on tumors. Consistent with the inference, Nwanaji-Enwerem et al. found that "epiTOC2" decreased during the "Mars-500" mission (Nwanaji-Enwerem et al., 2020). The "epiTOC2" denotes the estimation of the mitotic age, and the decrease in "epiTOC2" indicates the reduced risk of cancers (Teschendorff, 2020).

To validate the inference, we further compared the expression of the two patterns between tumors and healthy controls. Only "AML" was considered here as three of the four enriched neoplasms were associated with AML (Figure 4A). We downloaded two gene expression profiles of AML samples and their healthy controls from public datasets (ArrayExpress: E-MTAB-220, GEO: GSE30029). We considered Ear-Dec genes, Dec-Inc-Dec genes and their union as three functional gene sets, and used single-sample gene set enrichment analysis (ssGSEA) to evaluate the activities of these three gene sets in AML samples and controls (Barbie et al., 2009). As a result, except for Ear-Dec in E-MTAB-220, both the two datasets showed that the activities of the three gene sets in healthy samples were significantly higher than the activities in AML samples (Figure 4B). Combined with the above findings, it can be concluded that long-term isolation might up-regulate the overall expression of genes with relatively low activities in tumor samples and make their activities closer to the healthy state, therefore inhibiting cancer.

DISCUSSION

In this study, using data from the "Mars-500", we firstly characterized the global methylation landscape of six crews during long-term isolation. The result showed that genome-wide DNA methylation changes were minimal, whereas loci-specific changes in DNA methylation could be observed during long-term isolation (Figures 1D and 1E), which were consistent with the conclusion in NASA Twins Study. Next, we identified six dynamic patterns of DNA methylation during long-term isolation. The six patterns showed both transient ("Recovering") and persistent ("Decreasing") changes during long-term isolation (Figure 2B), where "Recovering" patterns suggested that most of significantly fluctuating CpG sites can gradually adapt to isolation and finally return to normal levels (Figure 2B). Furthermore, pathway enrichment analysis of genes in each pattern revealed strong functional specificity (Figure 3C), which illustrated the identified patterns might play diverse roles in response to long-term isolation. In addition, disease genes enrichment analysis revealed some patterns were significantly associated with nervous system diseases, digestive system diseases and cancers (Figure 4). Particularly, at present, the discussion on the relationship between long-term isolation and cancer is still controversial (Nwanaji-Enwerem et al., 2020).

We made a comprehensive comparison between the current study and the previous study based on DNA methylation data from Mars-500 (Liang et al., 2019). Specifically, the previous study mainly investigated the association between DNA methylation and phenotypes (glucose and mood state) during Mars-500 mission, and identified DNA methylation sites with the same trend as the two phenotypes (phenotype-synchronized methylation sites) during the mission. They confirmed that these phenotype-synchronized epigenetic features could reflect the effects of long-term isolation on the human body. For example, functional annotation for the most glucose-synchronized methylation sites showed that type II diabetes mellitus pathway was significantly enriched. However, the current study mainly identified dynamic patterns of DNA methylation, and revealed important functional significance of the patterns. Nonetheless, some findings and analysis methods in the current study are comparable with the previous study. Specifically, the previous study observed great individual differences on DNA methylation and used personalized analysis to solve the problem. The current study also found the individual difference and corrected the difference (Figures 1B

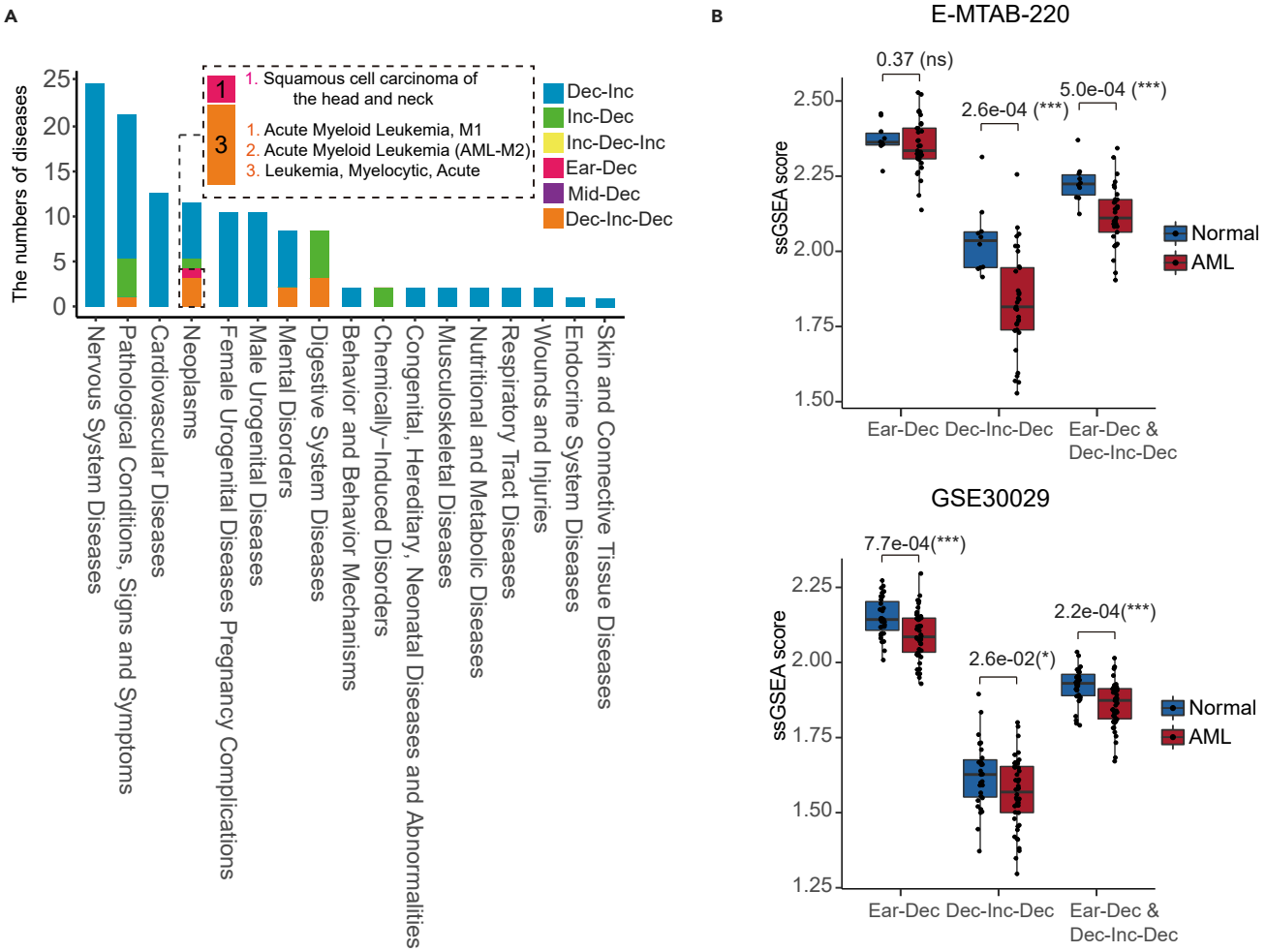


Figure 4. Association analysis between methylation patterns and human diseases

(A) Stacked bar graph showing the numbers of diseases associated with patterns for each disease type. Specific Neoplasms associated with "Ear-Dec" and "Dec-Inc-Dec" patterns were shown, where the numbers in the bar represented the counts of diseases.

(B) Comparison of ssGSEA-scores of genes with "Ear-Dec" and "Dec-Inc-Dec" and their union genes between AML and Normal samples in two public datasets (ArrayExpress: E-MTAB-220, GEO: GSE30029). T-test was performed to evaluate the significance of differences between different groups. "*" represents $p < 0.05$, "**" represents $p < 0.01$, "***" represents $p < 0.001$, and "ns" represents "not significant".

and 1C). Furthermore, consistent with the previous study, we also identified differentially methylated sites in the mission. However, the methods of identifying the differences were different. Previous study used "ANOVA" and "nparLD" to directly identify differentially methylated sites. However, before identifying the differences, the current study also considered the consistency among crews and the absolute changes of methylation levels (Table 1). Although the methods were somewhat distinct, some results were consistent. Differentially methylated sites were significantly enriched in the intergenic region (IGR) and non-CpG island region in the current study (Figure 3B), which is the same as the previous study. In addition, in the current study, differentially methylated genes were enriched in the pathway "platelet activation (hsa04611)" and the process "circadian rhythm (GO:0007623)" (Table S3), which were consistent with previous study.

In conclusion, using the time series DNA methylation data from the simulated spaceflight "Mars-500", we characterized the DNA methylation alterations and identified six methylation dynamic patterns during long-term isolation. Through functional analysis and associated analysis with diseases, we found different patterns played different roles in affecting the biological process or human health. Specially, the two vulnerable patterns (Inc-Dec-Inc and Dec-Inc-Dec) were significantly associated with immune response in response to long-term isolation, and another two "Decreasing" patterns (Ear-Dec and Dec-Inc-Dec)

Table 3. Enrichment of Ear-Dec and Dec-Inc-Dec genes in oncogenes and tumor suppressor genes

Pattern	FunSet_name	Pattern_gene	Total_gene	FunSet_gene	Overlap_gene	p value
Ear-Dec	Oncogene	172	19,687	698	6	0.57
Ear-Dec	TSG	172	19,687	1,014	6	0.88
Dec-Inc-Dec	Oncogene	150	19,687	698	8	0.16
Dec-Inc-Dec	TSG	150	19,687	1,014	18	7×10^{-4}
Ear-Dec & Dec-Inc-Dec	Oncogene	322	19,687	698	14	0.25
Ear-Dec & Dec-Inc-Dec	TSG	322	19,687	1,014	24	0.045

were found to be related to decreased risk of cancers, which were worthy of further study with more omics data and larger numbers of samples to provide protection for the health of astronauts in long-term space-flight in future.

Limitations of study

There are some limitations of the study. Given that DNA methylation was measured by Illumina HumanMethylation450 BeadChip in the "Mars-500" mission, whole-genome DNA methylation changes can't be characterized. Therefore, detective methods with higher genome coverage are necessary for exploring the comprehensive changes of DNA methylation during long-term isolation such as "WGBS". In addition, because of lack of the gene expression data, the regulation of the identified methylation pattern on gene expression cannot be effectively verified.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Data collection and pre-processing
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Global methylation characteristics analysis
  - The multi-step time series analysis
  - Identification of pattern genes
  - Functional enrichment analysis
  - Association analysis with diseases

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104493.

ACKNOWLEDGMENTS

This study was supported by the Foundation for the National Natural Science Foundation of China [61872183, 62172213].

AUTHOR CONTRIBUTIONS

W.J. conceived and designed the study. F.H. and X.Z. performed data analysis and drafted the manuscript. S.Z., H.L., Y-e.H., M.Y., J.Z., and X.C. provided scientific advice and contributed to the analysis. All authors revised and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 14, 2022
Revised: April 13, 2022
Accepted: May 20, 2022
Published: June 17, 2022

REFERENCES

Afshinnekoo, E., Scott, R.T., MacKay, M.J., Pariset, E., Cekanaviciute, E., Barker, R., Gilroy, S., Hassane, D., Smith, S.M., Zwart, S.R., et al. (2020). Fundamental biological features of spaceflight: advancing the field to enable deep-space exploration. Cell *183*, 1162–1184. https://doi.org/10.1016/j.cell.2020.10.050.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112. https://doi.org/10.1038/nature08460.

Basner, M., Dinges, D.F., Mollicone, D.J., Savelev, I., Ecker, A.J., Di Antonio, A., Jones, C.W., Hyder, E.C., Kan, K., Morukov, B.V., and Sutton, J.P. (2014). Psychological and behavioral changes during confinement in a 520-day simulated interplanetary mission to mars. PLoS One *9*, e93298. https://doi.org/10.1371/journal.pone.0093298.

Belavý, D.L., Gast, U., Daumer, M., Fomina, E., Rawer, R., Schießl, H., Schneider, S., Schubert, H., Soaz, C., and Felsenberg, D. (2013). Progressive adaptation in physical activity and neuromuscular performance during 520d confinement. PLoS One *8*, e60090. https://doi.org/10.1371/journal.pone.0060090.

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J.M., Delano, D., Zhang, L., Schroth, G.P., Gunderson, K.L., et al. (2011). High density DNA methylation array with single CpG site resolution. Genomics *98*, 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007.

Chen, Y.-a., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics *8*, 203–209. https://doi.org/10.4161/epi.23470.

Crucian, B., Simpson, R.J., Mehta, S., Stowe, R., Chouker, A., Hwang, S.-A., Actor, J.K., Salam, A.P., Pierson, D., and Sams, C. (2014). Terrestrial stress analogs for spaceflight associated immune system dysregulation. Brain Behav. Immun. *39*, 23–32. https://doi.org/10.1016/j.bbi.2014.01.011.

Crucian, B.E., Makedonas, G., Sams, C.F., Pierson, D.L., Simpson, R., Stowe, R.P., Smith, S.M., Zwart, S.R., Krieger, S.S., Rooney, B., et al. (2020). Countermeasures-based improvements in stress, immune system dysregulation and latent herpesvirus Reactivation onboard the international space station - Relevance for deep space missions and terrestrial medicine. Neurosci. Biobehav. Rev. *115*, 68–76. https://doi.org/10.1016/j.neubiorev.2020.05.007.

Demertzi, A., Van Ombergen, A., Tomilovskaya, E., Jeurissen, B., Pechenkova, E., Di Perri, C., Litvinova, L., Amico, E., Rumshiskaya, A., Rukavishnikov, I., et al. (2016). Cortical reorganization in an astronaut's brain after long-duration spaceflight. Brain Struct. Funct. *221*, 2873–2876. https://doi.org/10.1007/s00429-015-1054-3.

Doarn, C.R., Polk, J.D., and Shepanek, M. (2019). Health challenges including behavioral problems in long-duration spaceflight. Neurol. India *67*, S190. https://doi.org/10.4103/0028-3886.259116.

Fang, H., Du, Y., Xia, L., Li, J., Zhang, J., and Wang, K. (2011). A topology-preserving selection and clustering approach to multidimensional biological data. OMICS *15*, 483–494. https://doi.org/10.1089/omi.2010.0066.

Garrett-Bakelman, F.E., Darshi, M., Green, S.J., Gur, R.C., Lin, L., Macias, B.R., McKenna, M.J., Meydan, C., Mishra, T., Nasrini, J., et al. (2019). The NASA Twins Study: a multidimensional analysis of a year-long human spaceflight. Science *364*, eaau8650. https://doi.org/10.1126/science.aau8650.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037.

Leser, N., and Wagner, S. (2015). The effects of acute social isolation on long-term social recognition memory. Neurobiol. Learn. Mem. *124*, 97–103. https://doi.org/10.1016/j.nlm.2015.07.002.

Liang, F., Lv, K., Wang, Y., Yuan, Y., Lu, L., Feng, Q., Jing, X., Wang, H., Liu, C., Rayner, S., et al. (2019). Personalized epigenome Remodeling under biochemical and psychological changes during long-term isolation environment. Front. Physiol. *10*, 932. https://doi.org/10.3389/fphys.2019.00932.

Liu, Y., Sun, J., and Zhao, M. (2017). ONGene: a literature-based database for human oncogenes. J Genet Genomics *44*, 119–121. https://doi.org/10.1016/j.jgg.2016.12.004.

Marshall-Goebel, K., Laurie, S.S., Alferova, I.V., Arbeille, P., Auñón-Chancellor, S.M., Ebert, D.J., Lee, S.M.C., Macias, B.R., Martin, D.S., Pattarini, J.M., et al. (2019). Assessment of jugular venous blood flow stasis and thrombosis during spaceflight. JAMA Netw. Open *2*, e1915011. https://doi.org/10.1001/jamanetworkopen.2019.15011.

Moore, L.D., Le, T., and Fan, G. (2013). DNA methylation and its basic function. Neuropsychopharmacology *38*, 23–38. https://doi.org/10.1038/npp.2012.112.

Nwanaji-Enwerem, J.C., Nwanaji-Enwerem, U., Van Der Laan, L., Galazka, J.M., Redeker, N.S., and Cardenas, A. (2020). A longitudinal epigenetic aging and leukocyte analysis of simulated space travel: the mars-500 mission. Cell Rep. *33*, 108406. https://doi.org/10.1016/j.celrep.2020.108406.

Ogneva, I.V., Loktev, S.S., and Sychev, V.N. (2018). Cytoskeleton structure and total methylation of mouse cardiac and lung tissue during space flight. PLoS One *13*, e0192643. https://doi.org/10.1371/journal.pone.0192643.

Ou, X., Long, L., Zhang, Y., Xue, Y., Liu, J., Lin, X., and Liu, B. (2009). Spaceflight induces both transient and heritable alterations in DNA methylation and gene expression in rice (Oryza sativa L.). Mutat. Res. *662*, 44–53. https://doi.org/10.1016/j.mrfmmm.2008.12.004.

Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., and Furlong, L.I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. *45*, D833–D839. https://doi.org/10.1093/nar/gkw943.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat. Biotechnol. *28*, 1057–1068. https://doi.org/10.1038/nbt.1685.

Reed, H.L., Reedy, K.R., Palinkas, L.A., Van Do, N., Finney, N.S., Case, H.S., LeMar, H.J., Wright, J., and Thomas, J. (2001). Impairment in cognitive and exercise performance during prolonged antarctic residence: effect of thyroxine supplementation in the polar triiodothyronine syndrome. J. Clin. Endocrinol. Metab. *86*, 110–116. https://doi.org/10.1210/jc.86.1.110.

Storey, J.D., Xiao, W., Leek, J.T., Tompkins, R.G., and Davis, R.W. (2005). Significance analysis of time course microarray experiments. Proc. Natl. Acad. Sci. U S A *102*, 12837–12842. https://doi.org/10.1073/pnas.0504609102.

Strewe, C., Muckenthaler, F., Feuerecker, M., Yi, B., Rykova, M., Kaufmann, I., Nichiporuk, I., Vassilieva, G., Hörl, M., Matzel, S., Schelling, G., Thiel, M., Morukov, B., and Choukèr, A. (2015). Functional changes in neutrophils and psychoneuroendocrine responses during 105 days of confinement. J. Appl. Physiol. *118*, 1122–1127. https://doi.org/10.1152/japplphysiol.00755.2014.

Teschendorff, A.E. (2020). A comparison of epigenetic mitotic-like clocks for cancer risk prediction. Genome Med. *12*, 56. https://doi.org/10.1186/s13073-020-00752-3.

The European Space Agency (2020). Mars500: Study Overview. https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Mars500/Mars500_study_overview.

Van Der Maaten, L. (2014). Accelerating t-SNE using tree-based algorithms. J. Mach. Learn. Res. *15*, 3221–3245.

Van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. J. Mach. Learn. Res. *9*.

Vigo, D.E., Tuerlinckx, F., Ogrinz, B., Wan, L., Simonelli, G., Bersenev, E., Van den Bergh, O., and Aubert, A.E. (2013). Circadian rhythm of autonomic cardiovascular control during Mars500 simulated mission to Mars. Aviat Space Environ. Med. *84*, 1023–1028. https://doi.org/10.3357/asem.3612.2013.

Wallace, D.L., Han, M.-H., Graham, D.L., Green, T.A., Vialou, V., Iñiguez, S.D., Cao, J.-L., Kirk, A., Chakravarty, S., Kumar, A., et al. (2009). CREB regulation of nucleus accumbens excitability mediates social isolation-induced behavioral deficits. Nat. Neurosci. *12*, 200–209. https://doi.org/10.1038/nn.2257.

Xiong, J., Lyu, K., Liang, F., Wang, Y., Yuan, Y., Wang, H., Liu, C., Rayner, S., Ling, S., Yang, F., et al. (2015). Dynamic nature of epigenetic patterns observed during the Mars 520-d mission simulation. In Report of GeneLab Study, August 1, 2018. https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-140/.

Yi, B., Nichiporuk, I., Nicolas, M., Schneider, S., Feuerecker, M., Vassilieva, G., Thieme, D., Schelling, G., and Choukèr, A. (2016). Reductions in circulating endocannabinoid 2-arachidonoylglycerol levels in healthy human subjects exposed to chronic stressors. Prog. Neuropsychopharmacol. Biol. Psychiatry *67*, 92–97. https://doi.org/10.1016/j.pnpbp.2016.01.004.

Yi, B., Rykova, M., Feuerecker, M., Jäger, B., Ladinig, C., Basner, M., Hörl, M., Matzel, S., Kaufmann, I., Strewe, C., et al. (2014a). 520-d Isolation and confinement simulating a flight to Mars reveals heightened immune responses and alterations of leukocyte phenotype. Brain Behav. Immun. *40*, 203–210. https://doi.org/10.1016/j.bbi.2014.03.018.

Yi, B., Rykova, M., Feuerecker, M., Jäger, B., Ladinig, C., Basner, M., Hörl, M., Matzel, S., Kaufmann, I., Strewe, C., et al. (2014b). 520-d Isolation and confinement simulating a flight to Mars reveals heightened immune responses and alterations of leukocyte phenotype. Brain Behav. Immun. *40*, 203–210. https://doi.org/10.1016/j.bbi.2014.03.018.

Zhao, M., Kim, P., Mitra, R., Zhao, J., and Zhao, Z. (2016). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acids Res. *44*, D1023–D1031. https://doi.org/10.1093/nar/gkv1268.

Zhou, M., Sng, N.J., LeFrois, C.E., Paul, A.-L., and Ferl, R.J. (2019a). Epigenomics in an extraterrestrial environment: organ-specific alteration of DNA methylation and gene expression elicited by spaceflight in Arabidopsis thaliana. BMC Genom. *20*, 205. https://doi.org/10.1186/s12864-019-5554-z.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019b). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.

STAR★METHODS

KEY RESOURCES TABLE

RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Normalized DNA methylation 450k data	GeneLab	GLDS-140, https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-140/
Gene-disease association data	DisGeNET	https://www.disgenet.org/
Oncogene	ONGene	http://ongene.bioinfo-minzhao.org/
Tumor suppressor gene	TSGene2.0	https://bioinfo.uth.edu/TSGene/
AML Gene expression data	GEO	GSE30029, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30029
	ArrayExpress	E-MTAB-220, https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-220/
Intermediate datasets	In this paper	https://doi.org/10.17632/nt4db4wwrs.1
Software and algorithms		
R version 4.0.0	CRAN	https://cran.r-project.org
R package Rtsne version 0.15	CRAN	https://cran.r-project.org/web/packages/Rtsne/index.html
R package sva version 3.36.0	Bioconductor	https://bioconductor.org/packages/release/bioc/html/sva.html
R package ggplot2 version 3.3.1	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
R package GSVA version 1.36.3	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/GSVA.html
EDGE	Storey et al., (2005)	https://genomics.princeton.edu/storeylab/edge/version1.html
SOM-SVD	Fang et al., (2011)	http://people.cs.bris.ac.uk/~hfang/TPSC/somsvd_howto.html
Metascape	Zhou et al., 2019b	https://metascape.org/gp/index.html
All original code	In this paper	https://doi.org/10.17632/nt4db4wwrs.1

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Wei Jiang (weijiang@nuaa.edu.cn).

Materials availability

This study did not generate unique reagents.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. Intermediate datasets used for conclusions were deposited on Mendeley Data. The DOI is listed in the key resources table.
- All original code has been deposited at Mendeley Data and is publicly available as of the data of publication. The DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mars-500 mission was a simulated manned flight to Mars. The mission was conducted by European Space Agency (ESA) and Russian Institute for Biomedical Problems (IBMP). The main purpose of the mission was to determine key psychological and physiological effects of being in such an isolation environment for such an extended period. Specifically, six male crew members (three Russian, two European, one Chinese) were sealed in an isolation chamber for 520 days (June 2010 – November 2011). During the isolation period the crew members had only personal contact with each other plus voice contact with a simulated control center and family. At the same time, the crew members during the isolation period were responsible for some

important scientific investigations including physiological investigations, biochemical, immunological, and biological investigations, etc. (The European Space Agency, 2020).

Before participating in Mars-500 mission, all six participants underwent a thorough physical examination and signed an informed consent form for the simulation experiment. All study protocols and experiments, including the collection of DNA methylation data, were approved by the IBMP committee on bioethics (Liang et al., 2019). In particular, ages for the six crew members were not provided in the publicly available metadata.

METHOD DETAILS

Data collection and pre-processing

Peripheral whole blood cells for this study were extracted from six crew members at six time points: -7d (7th day before Mars-500 mission start, pre-isolation), 60d, 168d, 300d and 512d (during isolation), 527d (7th day after Mars-500 mission end, post-isolation) (Figure S1). DNA methylation of the peripheral whole blood cells was measured by Illumina HumanMethylation450 BeadChip, and full descriptions about the measurement could be viewed in the original article (Liang et al., 2019). The normalized 450k methylation data were downloaded from GeneLab database (GLDS-140), where detailed data normalization flow could be viewed (Xiong et al., 2015). Methylation level for each CpG site was represented by beta value, calculated by dividing the methylated probe intensity by the sum of methylated probe intensity and unmethylated probe intensity (Bibikova et al., 2011). Low-quality probes were filtered out according to the following criteria: i) probes with detective p value > 0.01 in any sample, ii) probes overlapped with SNPs from HumanMethylation450 Manifest annotation file, iii) probes aligned to multi position of genome (Chen et al., 2013).

QUANTIFICATION AND STATISTICAL ANALYSIS

Global methylation characteristics analysis

We used R (v4.0.0) package "Rtsne" (v0.15) to performed t-SNE with the parameter perplexity 6 (Van Der Maaten, 2014; Van der Maaten and Hinton, 2008). "Combat" from R package "sva" (v3.36.0) was used to correct the heterogeneity of crews by assigning crews as different batch (Johnson et al., 2007). As for maximum changes of beta values, we first averaged the beta values of samples from the same time point, and then defined the maximum differences between any two time points as the maximum changes of beta values.

The multi-step time series analysis

A multi-step time-series analysis consisted of four steps. Firstly, to identify the consistently variable DNA methylation sites across six time points in six crews, probes were filtered out by ANOVA using R function "aov" (False Discovery Rate (FDR) < 0.05). The six time points were regarded as six groups in ANOVA analysis. Secondly, considering the technical error for the detection of beta value, CpG sites with maximum beta changes exceeding 0.1 were selected. Thirdly, EDGE was used to identify differentially methylated sites in time series (FDR < 0.05) (Storey et al., 2005). Finally, SOM-SVD was applied to further select the significantly fluctuating sites (FDR < 0.05) and cluster them into clusters. As a topology-preserving clustering method, SOM-SVD integrated SOM and singular value decomposition (SVD), complemented with SOM-based two-phase clustering procedure, identifying the clusters with highly similar trends. In order to meet the input of SOM-SVD (Fang et al., 2011), methylation value of each CpG in each sample was converted by subtracting the mean values of the CpG site in all samples, and then the matrix containing converted methylation values with CpG sites as rows and samples as columns was input to SOM-SVD.

Identification of pattern genes

According to 450k annotation file, each probe could be assigned one or multiple genes with specific genome region information. Here, we defined the promoter region as TSS1500, TSS200, 5'UTR and 1st exon. We retained the genes whose promoters covered the CpG sites with dynamic patterns. Furthermore, because one gene might have multiple dynamic sites, we filtered out genes with inconsistent dynamic patterns (Figure S3). Finally, only protein-coding genes of each pattern were considered as pattern genes (Figure S3).

Functional enrichment analysis

Functional enrichment analysis of patterns genes was performed by "Metascape", which is a web-based portal used to provide a comprehensive gene list annotation and functional enrichment analysis (Zhou et al., 2019b). Here functional sets were limited to "GO Biological Processes" and "KEGG Pathways". Default parameters were selected (min overlap = 3, *p* value cutoff = 0.01, min enrichment = 1.5). Moreover, since there was functional redundancy in the enriched terms, we also used Metascape to cluster the enriched terms and construct a functional network based on top 20 significant clusters (Zhou et al., 2019b).

Association analysis with diseases

Human diseases-associated genes were obtained from "Curated gene-disease associations" in DisGeNET, which contained 84,038 gene-disease associations involving 11,181 human diseases (Piñero et al., 2017). The diseases in DisGeNet were classified based on MeSH (Medical Subject Headings). Hypergeometric test was used to evaluate the significance of the association between methylation patterns and human diseases (p value < 0.05, count ≥ 3).

Oncogenes and tumor suppressor genes were downloaded from ONGene (Liu et al., 2017) and TSGene2.0 (Zhao et al., 2016), respectively, and only protein-coding oncogenes or tumor suppressor genes were retained. We obtained two sets of gene expression profiles of AML samples and healthy controls, where one set included 46 AML samples and 31 healthy controls from GEO (GSE30029), and another set included 33 AML samples and 10 healthy controls from ArrayExpress (E-MTAB-220). The ssGSEA was used to evaluate the activities of gene sets in samples, where the ssGSEA scores represented the activation scores of gene sets (Barbie et al., 2009). R package "GSVA" (v1.36.3) was used to perform the ssGESA. T-test was used to compare the difference of activities of the functional sets between AML samples and healthy controls. "*" represents p value < 0.05, "**" represents p value < 0.01, "***" represents p value < 0.001, and "ns" represents "not significant".