

A Direction-Aware Meta-Analytic Framework for Aggregating Transcriptomic Evidence of Spaceflight Responses Across Species

Sulthan Rafi Abyan Hidayat

sulthanrafiabyanhidayat@gmail.com

Indonesia Defense University

Diah Ayu NurFadilah

Institut Teknologi Sepuluh Nopember

Isna Arofatun Nikmah

Indonesia Defense University

Rizky Dwi Satrio

Indonesia Defense University

Muhammad Rasyad Lubis

Institut Teknologi Sepuluh Nopember

Muhammad Rizky Ramadhan

Indonesia Defense University

Dian Yoel Panusunan Pasaribu

Indonesia Defense University

Ekelia Tian Permana Sitepu

Indonesia Defense University

Research Article

Keywords: spaceflight biology, transcriptomics, meta-analysis, Open Science Data Repository, p-value combination, microgravity

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11146490/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Small sample sizes in individual spaceflight missions make it difficult to establish whether transcriptomic responses to spaceflight are directionally consistent across independent studies, and previous meta-analyses have generally been limited to a single tissue, species or kingdom. We curated 121 transcriptomic datasets from NASA's Open Science Data Repository, spanning seven taxonomic groups across three kingdoms, and applied a directional (one-sided) Fisher p-value combination per species, embedding effect direction directly into each gene's significance test. A majority of genes tested reached significance after false discovery rate correction; however, a sign-permutation null check showed a similar proportion of genes reached significance when direction was randomly reassigned, indicating this directional test is, on this dataset, close to equivalent to a test of two-sided differential expression, with `directional_concordance` the more appropriate metric for directional consistency itself. A further, more substantial limitation: excluding accession-gene combinations showing a within-accession sign-cancellation signature reduced the significant-gene count by 12.6 per cent overall, and by 33.0 and 25.9 per cent in *Caenorhabditis elegans* and *Drosophila melanogaster* respectively, indicating unmodelled dependence among within-accession contrasts materially affects results in these species. Pathway enrichment and ortholog mapping recapitulated several previously reported spaceflight-response categories, including mitochondrial/oxidative phosphorylation signalling in *Drosophila melanogaster* and innate immunity/cytokine signalling in *Homo sapiens*, but identified no candidate gene shared across more than two species. These findings should be read primarily as a reproducible framework for aggregating cross-study transcriptomic evidence, together with an empirical account of its limits, rather than as confirmed evidence of conserved directional spaceflight responses.

1. Introduction

Understanding the mechanisms by which living systems respond to spaceflight is a crucial aspect of long-term human exploration of the Moon and Mars, including musculoskeletal deconditioning (Comfort et al. 2021), immune dysregulation (Crucian et al. 2018), and cardiovascular remodelling (Hughson et al. 2018). Over the past two decades, spaceflight experiments profiling gene expression across organisms from bacteria to rodents and humans have yielded a wealth of transcriptomic, proteomic and metabolomic 'omics' datasets from missions to the International Space Station (ISS). A key limitation of individual spaceflight studies is small sample size, driven by the cost and logistics of spaceflight experiments; most published analyses report findings from a single mission and a single tissue, making it difficult to establish a generalisable biological response to spaceflight (Lai Polo et al. 2020). NASA established GeneLab, later integrated into the broader NASA Open Science Data Repository (OSDR) in 2021 (Berrios et al. 2021), to address this fragmentation; by 2026 the OSDR had curated nearly 600 studies and over 1,000 datasets across 45 species and more than 80 assay types (OSDR web portal, accessed 2026), creating an opportunity for cross-study, cross-species synthesis.

1.1 Related Work

Several studies have analysed GeneLab data with narrower scope. da Silveira et al. (2020) combined GeneLab and astronaut biomedical data to identify mitochondrial stress as a central hub of the spaceflight response, but summarised effect magnitude at the tissue level rather than producing gene-level cross-study statistics, using data available only up to around 2019. Formal meta-analyses have otherwise been confined to single kingdoms: Morrison and Nicholson (2018) combined eight bacterial transcriptomes via p-value-based meta-analysis and found no single 'spaceflight-responsive' gene set shared across species and strains; Barker et al. (2023) applied a similar single-pipeline meta-analysis restricted to *Arabidopsis thaliana* across 15 experiments, building on their earlier TOAST tool (Barker et al. 2020). Barker et al. (2019), in a NASA Technical Reports Server entry that does not appear to have undergone conventional peer review, performed a cross-kingdom exploration spanning plants, animals and humans using enrichment and regulatory-network approaches rather than formal p-value combination, reporting evidence for conserved oxidative stress, DNA-damage and transcription-factor signalling; despite its non-peer-reviewed status, it remains the closest existing conceptual precedent for a cross-kingdom conserved spaceflight response. TOAST and its extension, the GeneLab X-Species Transcriptional Explorer (Barker et al. 2020), instead provide dashboard-based exploration tools that let users manually set significance thresholds across studies and species, valuable for hypothesis generation but not for computing reproducible, threshold-independent cross-study significance statistics.

Although cross-kingdom responses to spaceflight have previously been proposed through enrichment-based analyses (Barker et al. 2019), and formal GeneLab meta-analyses have been conducted within individual kingdoms such as bacteria (Morrison and Nicholson 2018) and plants (Barker et al. 2023), to our knowledge no peer-reviewed study has systematically applied a formal, reproducible, FDR-controlled p-value combination framework to combine gene-level directional evidence across multiple species spanning multiple kingdoms within the OSDR archive.

Several additional efforts published between 2024 and 2026 are adjacent to this work but do not occupy the same analytical space: GLARE (Seo et al. 2025) applies machine learning to a single *Arabidopsis* dataset rather than a formal cross-kingdom statistical framework; a separate OSDR reanalysis (Sakharkar et al. 2024) proposed systemic genome correlation loss across ten datasets using a correlation/entropy-based approach restricted to human and other mammalian tissues; the Inspiration4 single-cell multi-omics study (Kim et al. 2024) confirmed a shared spaceflight signature across human and mouse datasets using enrichment-based methods, confined to animals and humans; a cross-species comparison of spaceflight-associated sarcopenia (Ball et al. 2026) linked conserved neuromuscular and mitochondrial pathways between mouse and human via ortholog mapping, again without spanning multiple kingdoms; and a series of studies from Sun and colleagues (Zhang et al. 2025) applied combination-based methods to *Caenorhabditis elegans* ISS datasets but remained single-species. To our knowledge, none of these studies combine a formal p-value combination framework with a multi-kingdom scope, suggesting that this specific analytical niche has not yet been addressed in the peer-reviewed literature.

1.2 Study Aims

This study is positioned primarily as a resource and methodology paper rather than a biological discovery paper (Fig. 1): its main products are a reproducible, publicly released gene-level evidence-aggregation pipeline and dataset (Methods 2.5), and an empirical account of what that pipeline's significance calls do and do not demonstrate (Results 3.2.1, Discussion 4.3), rather than a claim of new spaceflight biology. To our knowledge, we found no peer-reviewed study using a reproducible gene-level p-value aggregation framework across multiple species and kingdoms within the OSDR spaceflight transcriptomic archive, and no claim is made that the 121 datasets analysed represent a novel biological finding in themselves. This study extends previous GeneLab meta-analysis efforts by providing a formal statistical framework for combining directional evidence at the gene level across multiple species and kingdoms using a directional combined p-value approach, with directional concordance reported separately as a companion descriptive metric rather than as the primary test statistic. Currently, to our knowledge, no single peer-reviewed study has produced a single, reproducible, cross-species statistic that aggregates directional evidence for a gene's expression change across various independent studies. Therefore, a formal combined p-value approach that accounts for this direction of change is compared with the magnitude-based and single-taxon approaches previously used to identify biological pathways. These biological pathways are known to be associated with spaceflight (da Silveira et al. 2020) and serve as a key contextualisation step to help assess whether the resulting gene rankings are biologically meaningful and not merely statistical artefacts arising from the volume of the dataset. Specifically, we aim to: (1) Curate the OSDR transcriptomic dataset by systematically comparing spaceflight with ground-based controls across networks and species; (2) Apply a formal combined p-value meta-analysis to derive quantitative gene-level statistics of aggregate directional evidence across studies, reported alongside directional concordance as a separate descriptive measure of cross-study agreement; (3) Contextualise the resulting gene rankings against known biological pathways involved in the spaceflight response; and (4) Release the resulting gene-level directional Fisher statistics as a public resource to serve as a reference for the space biology community.

In summary, the study makes three contributions. First, it provides, to our knowledge, the first reproducible gene-level directional p-value combination applied across multiple species and kingdoms within the OSDR archive, released together with per-gene statistics and companion metrics (Methods 2.5). Second, it provides an empirical account of the limits of that framework on real spaceflight data: a sign-permutation null check shows that the directional test closely tracks two-sided differential expression on this dataset, and a within-accession sign-mismatch check shows that unmodelled dependence among merged contrasts materially affects results in some species (Results 3.2 and 3.2.1). We report these as findings in their own right, because they bear directly on how any future cross-study aggregation of OSDR transcriptomes should be designed and interpreted. Third, it shows that, where shared spaceflight-associated signals are recovered, they appear mainly at the level of pathway categories rather than of shared orthologous genes (Results 3.3 to 3.5; Discussion 4.2).

2. Methods

2.1 Data Source and Inclusion Criteria

Several related terms are used throughout this manuscript to describe different levels of the OSDR data hierarchy, and we define them explicitly here to avoid ambiguity. An OSDR study is a self-contained research investigation registered in the repository, identified by an accession number (formerly 'GLDS-', now 'OSD-' prefixed; the two numbering schemes are numerically equivalent for the accessions used in this study, for example GLDS-168 and OSD-168 refer to the same study). Each study may contain one or more assays (RNA sequencing or microarray experiments), and each assay defines one or more contrasts, a specific spaceflight-versus-ground-control comparison, potentially further subdivided by tissue, timepoint, or genotype within the same study. Each contrast is itself computed from multiple biological replicates (individual specimens or samples) contributing to the spaceflight and ground-control groups being compared. In this manuscript: 'study' refers to the top-level OSDR registration (634 identified, Methods 2.1, Fig. 2); 'dataset' and 'accession' are used interchangeably to refer to one OSDR study that passed inclusion screening and was retained for analysis (203 meeting inclusion criteria, 121 ultimately analysed; Fig. 2); and 'contrast' or 'sub-study' refers to the individual spaceflight-versus-ground-control comparison(s) within one dataset, which are combined within a dataset before the cross-dataset meta-analysis (Methods 2.2, 2.3). The independent unit of evidence combined by the directional Fisher meta-analysis (Methods 2.3) is therefore the dataset (equivalently, accession), one combined p-value per gene per dataset, itself already aggregated across that dataset's constituent contrasts and biological replicates, not the study, contrast, or individual replicate directly.

The datasets were retrieved programmatically from the NASA Open Science Data Repository (OSDR) via the public REST API provided by the repository (NASA OSDR API documentation, OSDR REST API). Studies were included if they met four criteria: (i) the assay included transcriptomic data in the form of RNA sequencing or microarrays; (ii) a comparison was available between spaceflight (FLT) and ground control (GC) samples; (iii) processed differential expression tables are available, containing gene identifiers, $\log_2$ fold-change, and both raw and study-level adjusted p-values (columns labelled *P. value ...* and *Adj. p. value ...* respectively in the GeneLab differential expression output), and generated using the documented standard GeneLab pipeline (GL-DPPD; Overbey et al. 2021); and (iv) the OSDR study record contains descriptive metadata on the organism, tissue, and flight duration, and, where available, mission-related information in free-text or non-standardised fields; this criterion did not require a machine-readable mission identifier, which is addressed separately in Methods 2.3. Only the raw, unadjusted *P. value ...* column was extracted and carried forward into the within-dataset and cross-dataset combination steps (Methods 2.2, 2.3); the study-level adjusted p-value was retained in the intermediate per-contrast tables for reference but was not used in any combination step, so that multiple-testing correction is applied exactly once, across genes, after the cross-dataset directional Fisher combination (Methods 2.3), rather than being compounded with any correction already applied within individual studies.

Studies were excluded if they did not provide processed differential expression tables or if the FLT/GC comparison could not be clearly identified from the metadata. The dataset selection process is shown in Fig. 2 as an OSDR dataset curation workflow, from all identified studies to the datasets meeting the criteria for analysis.

2.2 Data Standardisation and Quality Control

Gene identifiers were harmonised into a single common namespace corresponding to the species under study, such as ENSEMBL gene IDs for mice, rats and humans; TAIR identifiers for *Arabidopsis thaliana*; FlyBase identifiers for *Drosophila melanogaster*; and WormBase identifiers for *Caenorhabditis elegans*. Meta-analyses were carried out separately for each species using these original identifiers. Subsequently, formal cross-species ortholog mapping was carried out in the second stage and applied only to the top significant genes generated using the g:Orth service from g:Profiler, rather than to the entire gene set during the standardisation stage. The direction of effect was then recorded based on up- or down-regulation in flight compared to ground controls for each gene in each study prior to the meta-analysis process. Established practices were followed in the RNA-seq meta-analysis (Rau et al. 2014). Rather than excluding genes with conflicting directions across studies prior to testing, effect direction was incorporated directly into the significance test itself, as described in Methods 2.3, so that no gene was discarded from consideration on the basis of a pre-test directional criterion.

2.2.1 Within-Accession Contrast Aggregation

Some OSDR accessions contain more than one spaceflight-versus-ground-control contrast, for example separate comparisons for different tissues, timepoints, or genotypes profiled within the same study (Methods 2.1). Because the independent unit of evidence in this study is the accession (dataset), not the individual contrast (Methods 2.1), all constituent contrasts within an accession were combined into a single p-value and a single mean log₂ fold-change per gene, per accession, before that accession's result was carried forward into the cross-dataset directional Fisher combination (Methods 2.3).

For each gene within an accession, the raw two-sided p-values from all of that accession's constituent contrasts in which the gene appeared were combined using unweighted Fisher's method: $F = -2 \times \sum(\ln(p_c))$ over all contributing contrasts c . Under independence, F follows a chi-squared distribution with $2 \times (\text{number of contrasts})$ degrees of freedom; because contrasts within an accession may be correlated (for example, when drawn from the same tissue, timepoint, or genotype series), this theoretical calibration may not strictly hold for every accession-level combination, and the resulting accession-level p-values should be interpreted subject to the within-accession dependence sensitivity analyses reported in Discussion 4.3. This combination does not distinguish between contrasts arising from different tissues, timepoints, or genotypes; all contrasts within an accession containing a given gene contribute equally and are not weighted by sample size, tissue type, or any other factor. The direction of effect for the accession-level result was taken as the unweighted mean of the log₂ fold-change values across all contributing contrasts; because this is a simple mean, contrasts reporting opposite directions of change for the same gene partially offset one another in the resulting mean, rather than being resolved by a majority vote or an explicit conflict-resolution rule. We did not further stratify or weight contrasts by tissue, timepoint, or genotype, and we did not model within-accession contrasts as non-independent replicates; this is a simplification, and its implications are discussed in Discussion 4.3.

Exact duplicate rows (identical gene ID, log2 fold-change, and p-value) within an accession were removed prior to combination, as such duplicates can arise from redundant entries in the source differential expression tables and would otherwise cause a single sub-study result to be double-counted. A gene that was not reported in a particular contrast within an accession (for example, because it was filtered out by that contrast's own analysis pipeline, such as a low-count filter) did not contribute a value for that contrast; the accession-level combination therefore used only the contrasts in which the gene was actually reported, rather than imputing a value for missing contrasts. Consequently, the number of contributing sub-studies per gene per accession ($n_{\text{substudies}}$, retained in the released data; Methods 2.5) can vary across genes within the same accession.

2.3 P-Value Combination Meta-Analysis

For each gene (g) observed across (S) datasets, the p-values from each study ($(p_{\{g1\}}, p_{\{g2\}}, \dots, p_{\{gS\}})$) are combined using Fisher's method. The test statistic is calculated using the formula

$$[F_g = -2 \sum_{s=1}^S \ln(p_{\{gs\}}),]$$

which follows a chi-squared distribution with $(2S)$ degrees of freedom, based on the null hypothesis that there is no effect across all analysed datasets (Fisher 1932).

In an earlier version of this pipeline, genes were first screened by a directional-concordance threshold (retaining genes for which at least 70 per cent of constituent datasets agreed in the direction of effect) before Fisher's method was applied to the same data. This procedure constitutes selection on the same data subsequently used for inference: because gene inclusion is conditional on an observed feature (effect direction) of the very p-values being combined, the resulting null distribution is not guaranteed to follow the standard chi-squared distribution, and the reported significance levels may not be correctly calibrated. We therefore replaced this two-step filter-then-test procedure with a single-step directional (one-sided) Fisher combination, in which the direction of effect is incorporated into the p-value itself prior to combination rather than used as a separate pre-test filter. For each gene observed in dataset (s), the two-sided p-value ($p_{\{gs\}}$) reported in the within-dataset combined differential expression table was converted to a one-sided p-value conditional on a candidate target direction (d in {up, down}): $[p_{\{gs\}}^{(d)} = p_{\{gs\}}/2 \text{ if sign of } \log_2\text{FC}_{\{gs\}} = d, \text{ else } 1-p_{\{gs\}}/2,]$ with $(p_{\{gs\}}^{(d)} = 0.5)$ assigned when $(\log_2\text{FC}_{\{gs\}}=0)$ (no directional evidence). The one-sided p-values were then combined across all (S) datasets in which the gene was observed using Fisher's method (Methods 2.3, above) separately for each candidate direction, yielding p_{up} and p_{down} . Because selecting the smaller of these two directional p-values as the gene's final p-value is itself a form of data-dependent selection, we applied a Bonferroni-style correction for this two-way choice: $p_{\text{final}} = \min(1, 2 \times \min(p_{\text{up}}, p_{\text{down}}))$. A Monte Carlo calibration check (null simulation with p-values drawn from Uniform(0,1) and log2FC sign assigned independently at random; 100,000 null replicates per K , generated with a fixed random seed for reproducibility, using six representative values of K , the number of datasets contributing to a gene, spanning 2 to 77 to match the range observed in this study) confirmed that $\min(p_{\text{up}}, p_{\text{down}})$ alone is not correctly calibrated under the null: the empirical false-positive rate at the nominal 0.05 threshold

ranged from 0.0991 to 0.1012 across the K values tested (95% confidence intervals approximately ± 0.001 to ± 0.002 , by the normal approximation to a binomial proportion), consistently around twofold inflated, whereas the corrected p_{final} was correctly calibrated at every K tested (empirical false-positive rate 0.0486 to 0.0501, with 95% confidence intervals overlapping the nominal 0.05 target in every case). All p-values, FDR-adjusted significance calls, and downstream analyses reported in this manuscript use this corrected p_{final} . The direction yielding the smaller of p_{up} and p_{down} was retained as the gene's assigned direction. Because direction is now embedded in the test statistic itself, every gene observed in at least two datasets was tested, with no pre-test exclusion based on directional agreement; the resulting concordance (the proportion of contributing datasets whose observed sign matches the assigned direction) is reported as a descriptive statistic alongside the test result rather than as a filtering criterion. Benjamini–Hochberg FDR correction (below) was applied across this full set of tested genes for each species. This change was made in response to two related concerns: standard Fisher's method combines two-sided p-values and is itself insensitive to effect direction: two datasets with identical p-values but opposite direction contribute identically to the combined statistic, so a separate directional filter was required to prevent this from generating spurious "consistent" genes; and, as noted above, applying that filter and then testing significance on the same data is a circular selection procedure. Embedding direction into the one-sided p-value resolves both issues simultaneously, without requiring a separate, arbitrarily chosen concordance threshold. We retain a threshold-based sensitivity comparison against the earlier v5 (filter-then-test) results in Results 3.2 and Discussion 4.3 for transparency, since the v5 outputs remain part of the publicly released data resource (Methods 2.5).

The Monte Carlo calibration described above establishes that the direction-selection correction is properly calibrated under an idealised null in which per-dataset p-values are drawn independently from Uniform(0,1). To assess calibration under the actual p-value structure this pipeline produces, which reflects whatever correlation and effect-size heterogeneity is already present in the real data, we additionally performed a sign-permutation null check: for each species, the real combined_pvalue for every accession-gene row was retained unchanged, while the sign of that row's log2 fold-change was randomly reassigned (independently per row, 10 permutations per species, seed 20260919), and the full directional Fisher combination, calibration correction and Benjamini-Hochberg procedure were then re-applied to the permuted data exactly as for the observed data. Because sign is reassigned at random, any true directional signal is destroyed while the real per-gene dataset counts and the real marginal distribution of combined_pvalue are preserved; a well-calibrated directional test should therefore call few or no genes significant under this null. Results of this check are reported in Results 3.2 and interpreted in Discussion 4.1 and 4.3.

As a supplementary analysis, we calculated the adaptive weighted Fisher (AW-Fisher) score for each gene (Li and Tseng 2011; Huo et al. 2020). The AW-Fisher results were used to identify the subset of the dataset that contributed most significantly to the combined signal. This information is included in the output table as supplementary information, but is not used in gene filtering or ranking. The combined p-value was corrected for multiple testing using the Benjamini–Hochberg false discovery rate (FDR) procedure (Benjamini and Hochberg 1995). All stages of curation, standardisation and meta-analysis

were carried out using Python. For very small p-values, calculations were performed using arbitrary-precision arithmetic with the mpmath library to prevent underflow in the float64 representation. The summary visualisations in Figs. 2–3 were created separately using R based on the output tables from Python; R was not used for the statistical calculations reported in this study. The dataset selection flowchart in Fig. 2 was created using draw.io in accordance with the number of datasets obtained in Results 3.1.

Fisher's method assumes that the p-values combined for a given gene are statistically independent. This assumption may be violated when multiple included datasets originate from the same spaceflight mission or animal cohort, since such datasets are not necessarily independent biological replicates. The OSDR metadata retrieved via the API did not include a populated mission identifier field for any of the 121 datasets analysed, precluding a direct, mission-based independence assessment; a study/publication identifier was similarly unavailable from the endpoints queried. In the absence of this metadata, we used dataset accession-number proximity as a proxy: datasets from the same species whose OSD accession numbers were separated by a gap of three or fewer were grouped into a candidate batch, on the grounds that OSDR accessions are assigned sequentially and closely spaced accessions from the same species frequently correspond to multiple tissue-specific datasets deposited from a single mission (for example, OSD-98 through OSD-105 span eight datasets across eight different mouse tissues). For every species with at least one candidate batch of two or more datasets, we repeated the directional Fisher analysis using a reduced dataset in which only one representative dataset per batch was retained (the dataset with the greatest number of within-dataset sub-studies, as a proxy for evidentiary depth), and compared the resulting significant-gene set against the full-dataset result using the Jaccard index. This sensitivity analysis was not applicable to *Homo sapiens*, *Bacillus subtilis* or *Saccharomyces cerevisiae*, for which no multi-dataset candidate batches were detected under this proxy. Results are reported in Results 3.2 and Discussion 4.3. As a complementary check on biological heterogeneity, we additionally examined the sensitivity of the *Mus musculus* result (the species with the largest number of datasets, 77, and the most complete tissue metadata, 55 of 77 datasets, recovered from sample-name parsing as described in the GeneLab abbreviation standard operating procedure) to the exclusion of individual tissue categories. Tissues were grouped into eight broad categories (Liver, Immune [thymus/spleen], Skin, Kidney, Brain/CNS, Muscle, Eye/Retina, and Other/Unknown) because most individual tissue types were represented by only one to three datasets, which is too few for a meaningful per-tissue leave-one-out analysis. For each category, we removed all constituent datasets, re-ran the directional Fisher analysis on the remaining datasets, and compared the resulting significant-gene set against the full 77-dataset baseline using the Jaccard index. This analysis was restricted to *Mus musculus* because no other species had sufficient tissue-metadata coverage or diversity to support it; this is reported as an explicit data limitation rather than a general claim that other species are free of tissue-driven heterogeneity. The directional Fisher statistic incorporates the observed direction of effect into the evidence-combination procedure itself, so that datasets whose direction opposes the tested hypothesis contribute correspondingly weaker one-sided evidence; directional concordance (the proportion of contributing datasets matching the assigned direction) is quantified separately as a descriptive measure (Methods 2.2.1) and is not itself part of the combined p-value. This approach differs from vote-counting, which

merely counts the number of studies with significant results, and from averaging effect sizes across studies.

The p-value pooling method was chosen because the OSDR/GeneLab differential expression tables used in this study do not uniformly provide standard errors for each gene. This information is required for effect-size-based meta-analyses, such as random-effects meta-analyses using Hedges' (g). RNA-Seq and microarray datasets can also produce different summary statistics, whilst the sample-level data required to recalculate standard errors is not consistently available across all curated outputs.

By combining the p-values already calculated by each study's respective pipeline, Fisher's method allows data from different platforms to be analysed within a single framework without requiring uniform estimates of effect size. This approach is also consistent with the study's objective, which is to rank genes by the strength of aggregate directional evidence across datasets, rather than estimating a single magnitude of change applicable to all tissues and experimental conditions.

2.4 Biological Contextualisation Against Previously Reported Spaceflight Responses

To assess the biological significance of the genes showing the most consistent scores, we compared the results of our analysis with pathways and gene clusters previously identified as relevant to living organisms' responses to spaceflight. Specifically, we used the following references: the mitochondrial stress and oxidative phosphorylation pathways identified by da Silveira et al. (2020), the immune and cardiovascular pathways in the space omics analysis proposed by Lee et al. (2025), and radiation response patterns in the dose-dependent spaceflight tissue analysis described by Zhang et al. (2024).

Biological pathway enrichment was performed for the most significant genes for each species using the g:Profiler service. This included the Gene Ontology (GO) Biological Processes, KEGG, and Reactome. g:Profiler uses the hypergeometric test for enrichment testing as well as the g_SCS multi-test correction (Kolberg et al. 2023) by default. To examine the results of the paper by da Silveira et al. (2020), we compared the pathway categories mentioned therein with the enrichment terms provided in our study. The comparison was based on the biological similarity of the terms rather than on gene overlap, as in the comparative study no list of genes was available; only the pathway categories were provided.

2.4.1 Gene Set Enrichment Analysis (Complete Ranked Gene List)

Because the top-N over-representation analysis (Methods 2.4) requires selecting a fixed-size gene list per species, and the resulting enrichment terms were shown to be sensitive to this cutoff (Results 3.4), we additionally ran pre-ranked gene set enrichment analysis (GSEA) using the complete ranked gene list per species, with no cutoff and no arbitrary gene-selection threshold. Because it avoids this cutoff dependency, GSEA is treated as at least as methodologically defensible as the top-500 over-representation analysis for biological interpretation, and the two are reported as complementary primary

enrichment analyses rather than as a primary analysis and a subordinate check. For each species, genes were ranked by the signed directional Fisher statistic (Methods 2.3), and GSEA prerank (gseapy v1.3.1, Python implementation of the Subramanian et al. (2005) algorithm) was run against the Enrichr GO_Biological_Process_2023 gene-set library, with permutation_num = 100, max_size = 500, min_size = 5, and a fixed random seed. For *Homo sapiens*, the analysis was additionally repeated against the KEGG_2021_Human and Reactome_2022 libraries. Because gseapy's Enrichr-based libraries use human gene symbols, gene identifiers for *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster* and *Caenorhabditis elegans* were first mapped to their human orthologue symbols via the mygene.info REST API (species-specific numerical NCBI taxonomy IDs; two-stage mapping through HomoloGene for *Caenorhabditis elegans*, given via WormBase gene identifiers) before ranking; *Arabidopsis thaliana* and *Bacillus subtilis* were excluded from this step, as no human-ortholog mapping consistent with the rest of the pipeline was available for these species at the time of analysis. Where a WBGene identifier corresponded to more than one row in the ranked list (arising from tied within-dataset scores combined into a single row; Methods 2.3), each constituent identifier was mapped independently and, where multiple identifiers converged on the same human orthologue symbol, the entry with the larger absolute ranking statistic was retained. This analysis is reported in Results 3.4 and interpreted alongside the top-500 over-representation result in Discussion 4.1; where the two analyses agree, this is taken as the stronger evidence, since GSEA does not depend on the top-N cutoff.

2.5 Public Resource Release

The results of the study are presented in a comprehensive table showing the directional Fisher combined statistic and p-value per gene, which has been made publicly available. This table includes the combined FDR-adjusted p-values, the direction of effect, the counts of contributor and concordant datasets, and the descriptive counts of the selected AW-Fisher subsets, as described in Section 2.3, together with a sign_mismatch_flag column identifying accession-gene rows matching the strong-p-value/weak-net-direction signature described in Discussion 4.3, so that readers can identify genes whose significance may be sensitive to this issue. Network and species coverage are also included as part of the supplementary data in this publication.

All of this data is archived on Zenodo (<https://doi.org/10.5281/zenodo.22676248>) and on GitHub (<https://github.com/sulthanrafiabyanhidayat-cloud/genelab-spaceflight-meta-analysis>) to serve as a data resource for the astrobiology community, in line with the FAIR data principles adopted by the OSDR (Berrios et al. 2021).

2.6 Use of AI-Assisted Tools

During the preparation of this work, the author(s) used Claude (Anthropic) to assist with organising content, refining language, and structuring sections of the manuscript. The author(s) reviewed and edited the output as needed and take full responsibility for the content of the published article.

3. Results

3.1 Dataset Landscape

Of the 634 studies available in the OSDR at the time of data collection, 203 met the inclusion criteria set out in Method 2.1, namely studies comparing spaceflight transcriptomics with ground controls using RNA-Seq or microarrays. OSDR is a dynamic, actively growing repository; the 634-study figure and the complete list of 121 accessions ultimately analysed (Methods 2.5) reflect a single frozen snapshot, retrieved via the OSDR REST API on 8 September 2026, rather than the repository's current holdings at the time of reading. Querying the same API endpoint on a later date would be expected to return a larger number of available studies as new datasets are deposited, and would not necessarily reproduce this exact figure. The full accession list used in this analysis is archived in full on Zenodo (Methods 2.5) as the definitive record of the frozen dataset underlying this study. Of these 203 studies, 128 datasets (63 per cent) provided processed differential expression files generated by the documented GeneLab pipeline. A further 75 datasets only included RSEM quantification results per sample (.genes.results) without differential expression tables; consequently, they could not be used directly in this analysis and were excluded from the subsequent stages (Fig. 2). Automatic extraction of the spaceflight versus ground control contrast was successfully carried out on 121 of the 128 datasets that contained differential expression tables (94.5 per cent). Seven datasets could not be used because the gene identifier column could not be recognised or there was no contrast that clearly showed a comparison between Space Flight and Ground Control.

The 121 datasets ultimately used in this meta-analysis were derived from 121 accessions in the GeneLab Data Set (GLDS). These datasets encompassed seven taxonomic groups from three kingdoms (Table 1), namely *Mus musculus* (77 datasets), *Homo sapiens* (12), *Arabidopsis thaliana* (15), *Drosophila melanogaster* (7), *Caenorhabditis elegans* (3), *Rattus norvegicus* (2) and *Bacillus subtilis* (2). Apart from these seven groups, there are two groups reported separately. OSD-554 could not be assigned to a specific species based on the available gene identifier patterns, and was therefore analysed as a single dataset. Meanwhile, the two *Saccharomyces cerevisiae* datasets yielded only six genes observed in both datasets, of which three reached significance under the directional Fisher test described in Methods 2.3. This group is included for the sake of completeness, but is not incorporated into the cross-species synthesis due to the very limited number of datasets.

Table 1
Coverage of datasets and genes by taxonomic group.

Species	Datasets	Genes tested (≥ 2 datasets)	Genes significant (FDR < 0.05)
<i>Mus musculus</i>	77	47,498	25,904
<i>Homo sapiens</i>	12	42,029	24,461
<i>Arabidopsis thaliana</i>	15	29,425	25,480
<i>Drosophila melanogaster</i>	7	18,680	16,585
<i>Caenorhabditis elegans</i>	3	20,380	16,162
<i>Rattus norvegicus</i>	2	13,013	343
<i>Bacillus subtilis</i>	2	4,424	2,512
^a <i>Saccharomyces cerevisiae</i>	2	6	3
^b Bacteria, OSD-554 (single)	1	1,736 (within-ds)	870 (within-ds)

^aIncluded for the sake of completeness alone; only six genes were observed across both datasets in this two-dataset cluster, of which three reached significance (FDR < 0.05) under the directional Fisher test, and this cluster was not incorporated into the cross-species synthesis across seven taxonomic groups given the very small number of genes available.

^b One dataset only; could not be classified to any resolvable species based on the gene identifier pattern and was treated as an intrastudy finding rather than being included in the meta-analysis across datasets. Kingdom level "Bacteria" assignment was made based on the OSDR study metadata information (organism/assay) available for the OSDR number 554, despite the gene symbol-based identifiers in the differential expression dataset not being able to be programmatically mapped to the species-specific namespace of the organism.

3.2 Genes with the Strongest Cross-Study Directional Evidence

For each taxonomic group, the p-values per gene were pooled across datasets using the directional (one-sided) Fisher combination described in Methods 2.3. Every gene observed in at least two datasets was tested; no gene was excluded from testing on the basis of directional agreement across datasets, in contrast to the two-step filter-then-test procedure used in an earlier version of this pipeline (Methods 2.3).

To directly test whether this directional significance reflects a genuine directional signal rather than the marginal strength of the underlying two-sided evidence alone, we performed a sign-permutation null

check using the actual combined_pvalue distribution this pipeline produced for every accession-gene row, with the sign of each row's log2 fold-change randomly reassigned (10 permutations per species, seed 20260919), breaking any true directional signal while preserving the real per-gene dataset counts and the real, non-idealised p-value structure already present in the data (Methods 2.3). Under this null, the number of genes called significant was close to the observed count in every species: for example, 54.3 per cent of tested genes reached $FDR < 0.05$ under random sign assignment in *Mus musculus*, compared with 54.5 per cent observed (a difference of 0.2 percentage points), and differences across all seven species ranged from -0.5 to $+1.9$ percentage points (Fig. 4). A supplementary check confirmed the mechanism: combining the same two-sided combined_pvalue values with direction ignored entirely gave 55.3 per cent significant in *Mus musculus*, statistically indistinguishable from both the directional-observed (54.5 per cent) and directional-null-permutation (54.3 per cent) results. Because selecting the smaller of two one-sided p-values derived from the same two-sided p-value (Methods 2.3) is, for a single gene, substantially determined by the two-sided p-value itself regardless of which direction is tested, the calibration correction applied to this two-way selection (which is itself correctly calibrated under an idealised independent-Uniform-p-value null; Methods 2.3) does not, on the real data, prevent directional significance from closely tracking two-sided significance. In practice, this means that $directional_fisher_fdr < 0.05$ in this study is close to synonymous with two-sided differential expression at the accession level being detectable at all across the contributing datasets, and provides comparatively little additional evidence, beyond that, that the gene's direction of change is itself the recurring signal. This finding is discussed further, together with its implications for how the headline count of 111,447 significant genes should be read, in Discussion 4.3.

For transparency, we retain a comparison against the earlier filter-then-test (v5) approach abandoned in Methods 2.3: 25,444 of the 25,450 genes significant under v5 (99.98 per cent) remained significant here, and 86,003 further genes were newly identified (Discussion 4.3; Table S1 in Online Resource 1 for full threshold-sensitivity detail).

As described in Methods 2.3, we assessed two potential sources of statistical instability in the pooled result: non-independence between datasets from the same species, and biological heterogeneity attributable to tissue of origin. Using dataset accession-number proximity as a proxy for shared-mission or shared-cohort origin (Methods 2.3), a representative-dataset sensitivity analysis was possible for five species with candidate multi-dataset batches. The overlap between the full-dataset and representative-only significant-gene sets, measured by the Jaccard index, was 0.989 for *Arabidopsis thaliana*, 0.983 for *Drosophila melanogaster*, 0.871 for *Mus musculus*, and 0.741 for *Caenorhabditis elegans* (Fig. 5A); the equivalent comparison could not be computed for *Rattus norvegicus*, whose two datasets were grouped into a single candidate batch, leaving only one representative dataset once reduced. No multi-dataset candidate batches were detected for *Homo sapiens*, *Bacillus subtilis* or *Saccharomyces cerevisiae* under this proxy, so the sensitivity analysis was not applicable to these species. For *Mus musculus*, the species with the most complete tissue metadata, a complementary leave-one-tissue-category-out analysis produced Jaccard values of 0.918–0.988 across eight broad tissue groupings, indicating that no single

tissue category disproportionately drove the species-level result (Fig. 5B). These findings are discussed further, together with their limitations, in Discussion 4.3.

Because the directional Fisher combined statistic scales with the number of contributing datasets ($F = -2\sum \ln(p)$ over n datasets), two genes with similarly small combined p-values can reflect very different underlying evidence: one gene observed in many datasets each contributing weak evidence, versus another observed in few datasets each contributing strong evidence. To let readers distinguish these cases, each gene's output record includes, alongside the combined statistic and p-value, the number of contributing datasets, the proportion concordant, the mean and median signed log₂ fold-change, the median absolute log₂ fold-change, a sign-disagreement heterogeneity measure (the proportion of contributing datasets whose direction differs from the gene's median direction), and a normalised Fisher statistic (the combined statistic divided by the number of contributing datasets, an approximate measure of average per-dataset evidence strength). None of these companion metrics is used to filter or re-rank genes; they are reported so that "strong evidence because many studies contribute" can be distinguished from "strong evidence because the effect is exceptionally consistent across a smaller number of studies".

Combining p-values from multiple datasets yields extremely small values. For a number of genes, the Fisher p-value is below 10^{-748} . Such values fall outside the range of standard double-precision arithmetic representation. Consequently, p-values were calculated using arbitrary-precision arithmetic to ensure that extremely small values do not become zero and do not assign the same rank to genes with the strongest statistical evidence. In this analysis, the pooled p-values were primarily used to rank genes by the strength of directional evidence across studies, rather than by directional consistency itself, which is reported separately via the `directional_concordance` metric.

In total, 175,449 genes observed in at least two datasets were tested across the seven taxonomic groups. Of these, 111,447 genes met the significance criteria following Benjamini–Hochberg false discovery rate (FDR) correction (calibration-corrected directional Fisher $FDR \leq 0.05$; Methods 2.3). The number of significant genes varied considerably across these groups. The proportion of tested genes reaching significance ranged from 2.6 per cent in *Rattus norvegicus* to 88.8 per cent in *Drosophila melanogaster*.

Of the total 111,447 significant genes, *Mus musculus* contributed 25,904 genes (23.2 per cent), *Arabidopsis thaliana* 25,480 (22.9 per cent), and *Homo sapiens* 24,461 (21.9 per cent). Next, *Caenorhabditis elegans* contributed 16,162 genes (14.5 per cent) and *Drosophila melanogaster* 16,585 genes (14.9 per cent). Smaller numbers were found in *Bacillus subtilis* (2,512 genes; 2.3 per cent) and *Rattus norvegicus* (343 genes; 0.3 per cent). These differences in numbers are related to the number of datasets available for each group and, for *Rattus norvegicus* specifically, to the very small number of qualifying datasets (two); they should be borne in mind when the total of 111,447 genes is used as an overall summary.

3.2.1 Sensitivity of Directional Significance to Within-Accession Directional Conflict

The within-accession contrast aggregation procedure (Methods 2.2.1) combines all contrasts within an accession using unweighted Fisher combination for the two-sided p-value, while assigning a single direction from the separately-computed mean log2 fold-change across the same contrasts. Because these are two different estimands, a very small combined p-value can in principle arise from contrasts that partially cancel in direction (Discussion 4.3). To quantify how often this occurs and how much it affects the reported results, we flagged accession-gene rows with more than one contributing contrast ($n_{\text{substudies}} > 1$), a combined p-value below 0.001, and a weak net direction ($|\text{mean log2 fold-change}| < 0.3$) as matching this signature, and re-ran the cross-accession directional Fisher combination with these rows excluded.

Of the 1,558,340 accession-gene rows with more than one contributing contrast, 177,784 (11.4 per cent) matched this signature. These rows were concentrated in accessions built from a larger number of merged contrasts: the median $n_{\text{substudies}}$ among flagged rows was 16 (interquartile range 8–36), compared with a median of 4 (interquartile range 4–12) among non-flagged multi-contrast rows. The prevalence of flagged rows varied considerably by species, from 0 per cent in *Rattus norvegicus* and *Bacillus subtilis* (no multi-contrast accessions), through 1.9 per cent in *Saccharomyces cerevisiae*, 5.6 per cent in *Mus musculus* and 7.8 per cent in *Homo sapiens*, to 19.5 per cent in *Drosophila melanogaster*, 24.8 per cent in *Arabidopsis thaliana* and 34.8 per cent in *Caenorhabditis elegans*.

Excluding flagged rows and re-running the full calibrated directional Fisher pipeline reduced the total significant-gene count from 111,447 to 97,374, a reduction of 14,073 genes (12.6 per cent). This reduction was again uneven across species: *Mus musculus*, *Homo sapiens*, *Bacillus subtilis* and *Rattus norvegicus* lost 0–10.0 per cent of their significant genes, whereas *Caenorhabditis elegans* and *Drosophila melanogaster* lost 33.0 per cent and 25.9 per cent respectively (Discussion 4.3 gives the full per-species breakdown). We treat this as a central limitation of the present directional estimator rather than a minor caveat: for a substantial minority of genes, particularly in *Caenorhabditis elegans* and *Drosophila melanogaster*, the reported directional significance depends on accession-level results whose assigned direction may not accurately represent the constituent contrasts, and the raw per-contrast data needed to resolve this directly were not retained separately from the within-accession combination step. The full gene-level output table released with this study (Methods 2.5) flags every row matching this signature so that readers can assess its impact on any individual gene of interest, and this sensitivity check is interpreted further, alongside the within-accession dependence checks above, in Discussion 4.3.

3.3 Cross-species Ortholog Overlap

To determine whether the pathway-level similarities observed in Results 3.4 were also present at the level of individual genes, the top 300 significant genes (directional Fisher FDR < 0.05, ranked by p-value; Methods 2.3, Results 3.2) from each of the five animal taxonomic groups were mapped to their orthologs in the other groups using g:Orth (Kolberg et al. 2023; Kuznetsov et al. 2023). Pairwise queries yielded between 151 and 455 orthologs, varying according to database coverage and the phylogenetic distance separating the two taxa involved; of these, between 1 and 8 orthologs per directional pair were themselves also present in the target species' own top-300 gene list, for a total of 88 raw source-gene-to-

target-ortholog matches across all ten unordered species combinations, of which 84 involve distinct target genes (Table 2); the difference of four reflects a small number of cases where two or three source genes in the same species converge on the same target ortholog (for example, two *Homo sapiens* genes, ENSG00000147119 and ENSG00000175040, both map to the same *Drosophila melanogaster* ortholog, FBgn0051637), compared with 68 gene-level pairs identified under the earlier v5 filter-then-test approach. As in the earlier analysis, all ten directional queries involving *Arabidopsis thaliana* returned no orthologs, despite *Arabidopsis* genes having been submitted correctly. This is not a biological finding; it reflects a known gap in g:Orth's architecture, where the Ensembl Plants database used for *Arabidopsis* is not linked to the main Ensembl database used for the animal genomes, so no bridge exists between the two. *Arabidopsis* was therefore left out of the pairwise comparison in Table 2 and Fig. 6; this says nothing about whether it shares orthologs with the animal groups, only that the tool couldn't check.

As a follow-up check on this limitation, we repeated the *Arabidopsis* query using the OMA (Orthologous Matrix) browser REST API (Kaleb et al. 2019), which explicitly covers plant and animal genomes within a single database and does not share g:Orth's Ensembl Plants/Ensembl animal architectural gap. Querying OMA for the same 300 top-ranked *Arabidopsis thaliana* genes returned at least one ortholog for 62 of the 300 genes queried; of the remaining 238 genes, 231 returned no match in OMA and 7 failed at the query stage itself (HTTP 404), yielding 556 ortholog relationships to the five animal species after resolving OMA's protein-level identifiers to Ensembl gene identifiers. Of these, seven were themselves present in the corresponding target species' own top-300 significant gene list: four with *Rattus norvegicus*, two with *Mus musculus*, and one with *Homo sapiens*; no hits were resolved for *Drosophila melanogaster* or *Caenorhabditis elegans*, for which OMA's cross-reference identifiers did not resolve cleanly to the FlyBase and WormBase identifiers used in this study's own gene lists, a further identifier-mapping limitation distinct from the Ensembl Plants gap that affected g:Orth. This confirms that *Arabidopsis* does share detectable orthologs with the top-ranked gene sets of at least three animal species in this study, at a similar order of magnitude to the pairwise hit counts observed among the animal species themselves (Table 2); the earlier absence of any signal under g:Orth was a tool limitation rather than evidence of a lack of shared orthology. Given the additional identifier-resolution step this analysis required, and its restriction to species with Ensembl gene identifiers available, we report it here as a targeted follow-up rather than a full replacement of Table 2 and Fig. 6, and recommend that a future revision of this pipeline adopt OMA (or a similarly cross-kingdom-capable resource such as Ensembl Compara, eggNOG, or OrthoDB) as the primary ortholog mapping tool for any comparison involving plant species.

Table 2

Pairwise gene-level ortholog overlap among the five animal taxonomic groups (top 300 significant genes per species; bidirectional counts summed).

Species pair	Matched ortholog pairs	BH-FDR adjusted p (min of both directions)
<i>Mus musculus</i> ↔ <i>Homo sapiens</i>	8	0.3142
<i>Mus musculus</i> ↔ <i>Rattus norvegicus</i>	14	0.0135
<i>Mus musculus</i> ↔ <i>Drosophila melanogaster</i>	9	0.3508
<i>Mus musculus</i> ↔ <i>Caenorhabditis elegans</i>	8	0.3652
<i>Homo sapiens</i> ↔ <i>Rattus norvegicus</i>	16	0.0101
<i>Homo sapiens</i> ↔ <i>Drosophila melanogaster</i>	9	0.4988
<i>Homo sapiens</i> ↔ <i>Caenorhabditis elegans</i>	6	0.7176
<i>Rattus norvegicus</i> ↔ <i>Drosophila melanogaster</i>	8	0.9444
<i>Rattus norvegicus</i> ↔ <i>Caenorhabditis elegans</i>	4	0.9936
<i>Drosophila melanogaster</i> ↔ <i>Caenorhabditis elegans</i>	2	0.9936

Taken together, the *Drosophila* mitochondrial signal and the pairwise cross-kingdom stress-response patterns point to two different kinds of follow-up work: the former is specific enough to prioritise as a candidate for future functional investigation now, while the latter needs gene-level and multi-omics confirmation before it can be treated as more than a shared pathway label. Both directions depend on the same resource this study is releasing, namely the per-gene directional Fisher statistics and companion metrics (Methods 2.3), that future analyses, including re-runs as OSDR grows, can build on directly.

Signals of oxidative phosphorylation and the electron transport chain observed in *Drosophila melanogaster* (Table 4) emerged independently of the statistical ranking process and were not specifically selected to test the hypothesis of mitochondrial dysfunction. These findings reinforce the results of previous studies that identify mitochondria as one of the key components of the biological response to spaceflight. As they were obtained through a different statistical approach and utilised a broader set of datasets, these results provide further support that mitochondrial dysfunction warrants prioritisation as a candidate target for future functional investigation. It should be stressed that the present evidence is limited to repeated transcriptomic association across datasets; no functional perturbation, knockdown, overexpression, phenotypic, protein-level, or pharmacological evidence is presented here, and the genes and pathways discussed below should accordingly be read as hypothesis-generating candidates for prioritisation rather than as validated countermeasure targets.

In contrast to the mitochondrial signals, the findings regarding cross-kingdom stress responses in this study remain at the level of biological pathways. The enrichment of terms such as ‘response to stress’, ‘response to chemical’ and ‘response to abiotic stimulus’ indicates the presence of recurring response patterns across several organisms; however, this is not yet sufficient to identify specific genes as targets for intervention. Therefore, further analysis is required to identify the genes contributing to these signals in each species and to assess whether these genes possess conserved orthologous relationships.

The ortholog mapping results described above do not reveal a small set of genes that universally appear across all analysed species. Under the earlier v5 filter-then-test analysis, one candidate conserved responder gene was identified: WBGene00001685 (*gpd-3*) in *Caenorhabditis elegans*, encoding a glyceraldehyde-3-phosphate dehydrogenase (GAPDH) family member, whose orthologs also appeared amongst the top-ranked genes in two other species. Under the present directional Fisher approach, this specific gene no longer appears within the top-300 significant gene list for *Caenorhabditis elegans*, and consequently does not appear in the ortholog mapping results at all. We therefore do not carry the *gpd-3*/GAPDH candidacy forward under the revised methodology; its earlier prominence appears to have been an artefact of the filter-then-test procedure rather than a signal that is robust to the choice of combination method, and the caveats previously raised against it (its use as a housekeeping normalisation control, and prior reports of GAPDH-family downregulation in spaceflight *Drosophila*) are correspondingly moot for this dataset.

More broadly, no gene in the ortholog mapping results has an ortholog present in the top-300 significant gene list of more than one other species (archived as candidate_universal_responder_genes_v6.csv on Zenodo, file name retained from the internal pipeline version at release); the 88 raw source-gene-to-target-ortholog matches (84 involving distinct target genes; Table 2) are each specific to a single pairwise comparison. The cross-kingdom stress-response signal detected in this study therefore remains a pathway-level convergence rather than a shared-gene signature, and this study does not propose a single candidate cross-species responder gene under the present methodology.

Raw overlap counts alone do not establish that these matches exceed chance. A hypergeometric enrichment test across all 20 directional source-target comparisons (BH-FDR corrected) found that only two comparisons, both involving *Rattus norvegicus* and another mammal (*Rattus norvegicus*-*Mus musculus*, FDR = 0.0135; *Rattus norvegicus*-*Homo sapiens*, FDR = 0.0101), exceeded chance-level overlap; the remaining 18, including all comparisons involving *Drosophila melanogaster* or *Caenorhabditis elegans*, were consistent with chance. This pattern was robust to the choice of 300 genes per species: repeating the analysis at cutoffs of 100, 200 and 500 genes (Table S2 in Online Resource 1) showed the same two mammalian pairs were the first to reach significance and remained significant at every larger cutoff tested, while non-mammalian comparisons did not reach significance at any cutoff below 500 genes. The 88 raw source-gene-to-target-ortholog matches (84 involving distinct target genes; Table 2) should therefore be read as a candidate list for future investigation, with the mammalian overlap – consistent with the closer evolutionary relationship among these taxa – the only signal statistically distinguishable

from chance, rather than as evidence of widespread conserved spaceflight-responsive orthologs across the five animal taxa.

Furthermore, the orthologous overlaps identified between the *Homo sapiens*–*Rattus norvegicus* (16 matched pairs) and *Mus musculus*–*Rattus norvegicus* (14 matched pairs) comparisons (Table 2) provide a number of potential candidates for further investigation. Although these genes do not appear in three or more species simultaneously, their presence in two distinct taxonomic groups indicates recurring response patterns across independent datasets. Therefore, the orthologous gene pairs listed in the supplementary data (archived as `cross_species_ortholog_overlap_gene_pairs_v6.csv` on Zenodo, file name retained from the internal pipeline version at release) may be considered as an initial candidate list for prioritisation in future gene-level and functional follow-up work, rather than as validated countermeasure targets.

Table 3 presents the highest-ranked gene in each taxonomic group based on Fisher’s combined statistics. Gene selection was performed using the $fisher_{pvalue_ise}$ value because, at the extreme ends of the distribution, some genes had identical $fisher_{fdr}$ values due to the limitations of `float64` numerical representation. Consequently, the $fisher_{fdr}$ value was not used to distinguish gene rankings under these conditions.

Gene symbols and names were obtained from species-specific curated databases, including NCBI Gene, Ensembl, FlyBase, WormBase, TAIR, and Wikidata cross-references accessed in 2026. The three highest-ranked genes, found in *Drosophila melanogaster*, *Caenorhabditis elegans* and *Bacillus subtilis*, do not yet have a complete functional characterisation and can therefore only be identified based on their locus ID or available annotations. This does not imply that these genes are of lesser biological importance, but rather reflects the limitations of the knowledge available at the time the research was conducted. Six of the seven top-ranked genes are identical to those identified under the earlier v5 filter-then-test procedure; the exception is *Mus musculus*, where the top-ranked gene under the directional Fisher approach is *Gvin1* (GTPase, very large interferon inducible 1), an interferon-response gene, rather than the ribosomal protein *Rpl26* identified previously. This change is consistent with the broader finding (Results 3.2) that *Mus musculus* is the species most affected by the transition from the filter-then-test to the directional approach. All directional Fisher statistics and p-values, including high-precision values for each species, are provided in the dataset published via Zenodo (Method 2.5).

Table 3
Top-ranked candidate gene by Fisher statistic in each taxonomic group.

Species	Top Gene (ID)	Symbol	Gene Name	Direction	Fisher p-value
<i>Mus musculus</i>	ENSMUSG00000045868	<i>Gvin1</i>	GTPase, very large interferon inducible 1	up	$< 1 \times 10^{-300}$
<i>Homo sapiens</i>	ENSG00000095739	<i>BAMBI</i>	BMP and activin membrane bound inhibitor	down	$< 1 \times 10^{-300}$
<i>Rattus norvegicus</i>	ENSRNOG00000057078	<i>Ddit4</i>	DNA damage inducible transcript 4	up	$< 1 \times 10^{-300}$
<i>Arabidopsis thaliana</i>	AT1G62660	<i>VI1</i>	vacuolar invertase 1	down	$< 1 \times 10^{-300}$
<i>Drosophila melanogaster</i>	FBgn0031490	<i>CG17264</i>	uncharacterised (annotation ID only)	up	$< 1 \times 10^{-300}$
<i>Caenorhabditis elegans</i>	WBGene00020670	<i>T22B2.6</i>	uncharacterised (sequence name only)	down	7.00×10^{-120}
<i>Bacillus subtilis</i>	BSU_15960	N/A	conserved exported protein of unknown function	up	$< 1 \times 10^{-300}$

The top-ranked single gene per species according to Fisher combined-statistic rank (Methods 2.3). For six of the seven species, the exact directional Fisher p-value falls far beyond the smallest value representable in standard double-precision (float64) arithmetic; these are reported as " $< 1 \times 10^{-300}$ " rather than as a specific extreme value, because the difference between, for example, 10^{-305} and 10^{-307} carries no additional biological meaning and depends on arbitrary-precision computation (Methods 2.3) rather than on any property of the gene itself. Gene selection and ranking among genes at this floor were nonetheless based on the exact arbitrary-precision p-value (Methods 2.3), not on this rounded display value. As discussed in Results 3.2 and Discussion 4.1, an extreme directional Fisher p-value of this kind should be read as strong evidence that a gene is differentially expressed at all across the contributing datasets, rather than as specific confirmation that its direction of change is itself unusually consistent. Gene symbols and names have been obtained from species-specific databases (NCBI Gene, Ensembl, FlyBase, WormBase, TAIR, and Wikidata referencing these sources, accessed 2026); three of the seven top genes (*Drosophila melanogaster*, *Caenorhabditis elegans*, and *Bacillus subtilis*) are still uncharacterised in these databases and are provided here by loci or annotation ID only, without a gene name, but not due to lack of biological significance, only due to having not been experimentally characterised as of the date.

3.4 Pathway Enrichment of Significant Genes

For each species, we selected the top 500 significant genes (directional Fisher FDR < 0.05, ranked by p-value; Methods 2.3), or all significant genes if there were fewer than 500. These genes were tested for over-representation in the Gene Ontology Biological Process (GO), KEGG and Reactome gene sets using the g:Profiler web service, against an explicit background universe consisting of every gene tested in the directional Fisher meta-analysis for that species (that is, every gene observed in at least two datasets; Methods 2.3, Results 3.2), rather than the full species genome or the g:Profiler default background. Correction for multiple testing was performed using the g_SCS method, with a significance threshold of $p_{\text{adj}} < 0.05$.

The number 500 was chosen as a practical threshold to ensure the analysis remained focused on genes with strong statistical evidence, whilst not unduly limiting the scope of identifiable biological pathways. Because this threshold was not derived from any formal criterion, we additionally ran the same enrichment analysis at cutoffs of 100, 250 and 1,000 genes per species, using the same explicit background universe, to assess how sensitive the enrichment results are to this choice. The number of enriched terms increased markedly and near-monotonically with cutoff size in most species, and this relationship differed substantially between species: *Homo sapiens* and *Mus musculus* yielded a substantial number of enriched terms even at the smallest cutoff of 100 genes (90 and 91 terms respectively), whereas *Caenorhabditis elegans* and *Drosophila melanogaster* yielded no significantly enriched terms at all at a cutoff of 100 genes, only reaching double-digit term counts once the cutoff reached 500 genes or more. This indicates that the choice of cutoff is not a neutral analytical convenience for every species, and that conclusions drawn from a single fixed cutoff should be interpreted with this species-dependent sensitivity in mind. The top-500 cutoff is retained as the primary reported result below for comparability with the species-contribution ranking in Results 3.2, and because it provided the most stable balance across species of the four cutoffs tested.

The complete-ranked-list GSEA (Methods 2.4.1) broadly recapitulated the mitochondrial and translational signal identified by the top-500 over-representation analysis above, while also revealing that this signal is not uniform across species. Suppression of translation-related gene sets (GO:0006412, Translation) reached FDR < 0.05 independently in *Mus musculus* (normalised enrichment score, NES=-3.17), *Rattus norvegicus* (NES=-2.35) and *Drosophila melanogaster* (NES=-3.97), all with the same negative (down-regulated in spaceflight) direction. Oxidative phosphorylation (GO:0006119) was most strongly suppressed in *Mus musculus* (NES=-3.30, FDR < 0.001), with the same negative direction, though not reaching FDR < 0.05, in *Rattus norvegicus* (NES=-1.66, FDR = 0.216) and *Drosophila melanogaster* (NES=-1.90, FDR = 0.092). Of the 3,668, 4,726 and 1,059 GO Biological Process terms tested in these three species respectively, 48, 27 and 32 reached FDR < 0.05.

In *Homo sapiens*, GSEA against all three gene-set libraries converged instead on a distinct theme: striated and cardiac muscle contraction was the dominant up-regulated signature (GO Biological Process NES = 3.20; KEGG NES = 2.68; Reactome NES = 2.84; all FDR < 0.001), while down-regulated terms centred on

antiviral and interferon signalling (GO:0140546, Defense Response to Symbiont, NES=-2.61, FDR < 0.001) rather than mitochondrial suppression; 200, 48 and 94 terms reached FDR < 0.05 in the GO Biological Process, KEGG and Reactome libraries respectively, out of 4,012, 313 and 1,507 tested. *Caenorhabditis elegans*, analysed via the human-ortholog mapping described in Methods 2.4.1, had only 3,724 of 20,872 WBGene identifiers (17.8 per cent) resolved to a human orthologue symbol through HomoloGene; because fewer than one-fifth of the species' tested genes contributed to this GSEA, the *Caenorhabditis elegans* result should be interpreted as a partial, orthology-limited representation of the transcriptome rather than a complete species-level enrichment analysis, and compared with the other species' GSEA results with this limitation in mind. Within this reduced gene set, *Caenorhabditis elegans* showed the opposite trend for oxidative phosphorylation to the mammalian species (NES = + 2.01, FDR = 0.068, not significant at FDR < 0.05); its only term reaching FDR < 0.05 was suppression of the MAPK cascade (GO:0000165, NES=-2.00, FDR = 0.020), out of 1,909 terms tested, of which 12 reached FDR < 0.25. *Arabidopsis thaliana* showed no term reaching FDR < 0.05 out of 516 tested (one at FDR < 0.25), consistent with its more limited enriched-term count under the top-500 over-representation analysis above.

Taken together, the translation/ribosome-suppression signal is the most consistent finding across the complete-ranked-list GSEA, reaching FDR < 0.05 in three of the five animal species tested (*Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*) with no cutoff applied, and is therefore unlikely to be an artefact of the top-500 gene-list cutoff used in the over-representation analysis above. The oxidative phosphorylation signal, by contrast, reached significance only in *Mus musculus* and trended in the opposite direction in *Caenorhabditis elegans*, indicating that this particular signal is less robust to the choice of enrichment method than the translation signal, a distinction not visible from the top-500 over-representation analysis alone. *Homo sapiens* did not recapitulate either mitochondrial signal under either method, showing a muscle-contraction and antiviral-signalling profile instead in both analyses; this cross-method consistency within *Homo sapiens* strengthens the interpretation offered in Discussion 4.1 that this profile reflects the tissue composition of the curated *Homo sapiens* dataset pool rather than an artefact of either enrichment method individually.

This analysis was carried out on six of the seven taxonomic groups. *Bacillus subtilis* was not analysed using g:Profiler as the service does not yet support bacterial organisms. Consequently, the limitation regarding this group stems from the analytical service's support, rather than from the availability or quality of the dataset.

Of the six species analysed, 867 terms were identified as showing significant enrichment. This figure comprises 646 terms in *Homo sapiens*, 139 in *Mus musculus*, 31 in *Drosophila melanogaster*, 23 in *Caenorhabditis elegans*, 16 in *Arabidopsis thaliana* (all five at the top-500 cutoff), and 12 in *Rattus norvegicus* (at the v2 calibration-corrected gene list of 343 significant genes; Methods 2.3, Results 3.2).

3.5 Biological Contextualisation Against Previously Reported Spaceflight-Responsive Pathways

da Silveira et al. (2020) reported five categories of biological pathways that showed enrichment in the multi-omics analysis of GeneLab data and biomedical data from astronauts, namely mitochondrial processes, innate immunity, chronic inflammation, the cell cycle, and circadian rhythms. That study reported pathway categories, rather than a list of genes that could be used for direct comparison, and (as noted in Discussion 4.3) some of the datasets analysed here overlap with those used in that prior study. Therefore, the comparison performed in this study is more accurately described as biological contextualisation, or partial replication, against previously reported categories, rather than as an independent validation: the enrichment results obtained in Section 3.4 were compared against the categories reported by da Silveira et al. (2020), using identical terms or those with close biological relevance as the basis for comparison, rather than direct gene overlap.

Under the directional Fisher approach (Methods 2.3, Results 3.2), the mitochondrial processes and oxidative phosphorylation signal reported previously remained present in *Drosophila melanogaster*, though with a reduced set of directly associated terms compared with the earlier filter-then-test analysis: four terms were identified at the top-500 cutoff, including the KEGG 'Oxidative phosphorylation' term ($p = 1.0 \times 10^{-6}$; 19 genes) and three GO Biological Process terms relating to aerobic respiration and electron transport (Table 4). As in the earlier analysis, this signal did not appear amongst the top-500 genes for *Mus musculus*, *Homo sapiens* or *Rattus norvegicus*. This remains consistent with a gene-selection method that prioritises the strength of aggregate directional evidence over the magnitude of expression change, and with the possibility that a 500-gene query depth does not capture every biologically relevant term in species with larger, more heterogeneous dataset collections (Discussion 4.3).

Unlike the earlier filter-then-test analysis, in which a weak innate immunity signal was identified only in *Mus musculus* (GO 'immune response', $p = 0.012$; GO 'immune system process', $p = 0.018$), the directional Fisher top-500 gene set for *Homo sapiens* showed a substantially stronger innate immunity signal, including the Reactome term 'Innate Immune System' ($p = 9.6 \times 10^{-11}$; 47 genes) and GO 'immune system process' ($p = 4.4 \times 10^{-15}$; 93 genes). This signal was not detected amongst the top-500 genes for any other species.

A specific GO term for the inflammatory response (GO:0006954) was, as in the earlier analysis, not significantly enriched in any species at the top-500 cutoff, and this study therefore does not report a direct replication of that specific term. However, the *Homo sapiens* top-500 gene set showed strong enrichment for a cluster of closely related cytokine- and NF- κ B-signalling terms not previously detected, including the Reactome term 'Cytokine Signaling in Immune system' ($p = 7.8 \times 10^{-10}$; 39 genes), GO 'response to cytokine' ($p = 6.0 \times 10^{-7}$; 40 genes), and GO 'canonical NF-kappaB signal transduction' ($p = 3.0 \times 10^{-3}$; 17 genes), amongst seven further related terms. Cytokine and NF- κ B signalling are core mechanistic components of the chronic inflammation category reported by da Silveira et al. (2020), even though the specific 'inflammatory response' term itself was not enriched; we therefore treat this as partial, mechanism-level support for the chronic inflammation category in *Homo sapiens*, rather than as a direct replication of the same GO term reported previously, and note this distinction explicitly to avoid overstating the finding.

Under the v2 (calibration-corrected) results, the number of FDR-significant genes for *Rattus norvegicus* fell from 555 to 343 (Methods 2.3, Results 3.2), and the top-gene list used for enrichment (now 343 genes, all available FDR-significant genes rather than a fixed top-500 cutoff) no longer includes the KEGG ‘Cell cycle’ term that was identified under the uncorrected v1 approach ($p = 0.044$; 14 genes), which was already weak and did not survive this reduction in gene-list size. A related term, KEGG ‘Cellular senescence’ ($p = 0.014$; 12 genes), remains significant and shares substantial biological overlap with cell-cycle regulation, so we treat this as weak, indirect support for the cell cycle category rather than a direct replication of the same term reported previously. Circadian rhythms were not identified in any species at the applicable cutoff, as in the earlier analysis.

Table 4

Mitochondria- and oxidative phosphorylation-related terms that were enriched in *Drosophila melanogaster*.

Term	Source	p-value	Genes
Oxidative phosphorylation	KEGG	1.0×10^{-6}	19
Aerobic respiration	GO:BP	4.6×10^{-3}	19
Cellular respiration	GO:BP	6.7×10^{-3}	19
Aerobic electron transport chain	GO:BP	2.5×10^{-2}	12

Beyond these five categories, we examined whether any ‘response to X’ GO Biological Process term was shared across multiple species at the top-500 cutoff. Under the earlier filter-then-test analysis, four such terms were shared identically across three species from two kingdoms (*Arabidopsis thaliana*, *Homo sapiens* and *Rattus norvegicus*). Under the directional Fisher approach, this specific three-species pattern was not reproduced: no ‘response to X’ term was shared by all three of these species, and *Rattus norvegicus* did not show significant enrichment for any stress-response term at the top-500 cutoff. Instead, a more limited, pairwise pattern emerged, still spanning two kingdoms: *Homo sapiens* and *Arabidopsis thaliana* shared four overlapping ‘response to X’ terms (response to chemical, response to abiotic stimulus, response to hormone, and response to endogenous stimulus), each independently significant in both species though with a much stronger signal in *Homo sapiens* (Table 5), and *Homo sapiens* and *Mus musculus* independently shared enrichment for ‘response to stress’ itself ($p = 6.1 \times 10^{-25}$ and $p = 4.1 \times 10^{-2}$ respectively).

No single ‘response to X’ term was therefore shared across three or more species under the present directional approach, in contrast to the earlier filter-then-test analysis. The pairwise *Homo sapiens*–*Arabidopsis thaliana* and *Homo sapiens*–*Mus musculus* overlaps that remain still span two kingdoms (Animalia and Plantae) and are consistent with a general stress-response programme rather than a single shared mechanism, but this study no longer reports a three-species signature for this category, and we do not carry the earlier three-species framing forward. This revision, and its implications for the cross-kingdom claim made in the Introduction, are discussed in Discussion 4.2 and 4.3.

Table 5
Stress response terms shared between *Arabidopsis thaliana* and *Homo sapiens*.

Term	<i>Arabidopsis thaliana</i>	<i>Homo sapiens</i>
response to chemical	7.8×10 ⁻⁴ (86)	4.2×10 ⁻¹⁹ (125)
response to abiotic stimulus	3.3×10 ⁻³ (66)	3.7×10 ⁻¹⁸ (62)
response to hormone	3.0×10 ⁻² (51)	6.8×10 ⁻⁹ (42)
response to endogenous stimulus	3.1×10 ⁻² (51)	7.0×10 ⁻²³ (79)

In *Homo sapiens*, the enrichment results are predominantly characterised by general terms relating to development and regulation. For example, ‘anatomical structure development’ ($p = 2.6 \times 10^{-11}$) and ‘regulation of cell communication’ ($p = 4.2 \times 10^{-11}$). This pattern is consistent with the nature of the 12 human datasets used, which originate from a variety of cell types, tissues and experimental conditions. This heterogeneity must be taken into account when comparing the results for *Homo sapiens* with those for other species.

4. Discussion

4.1 Summary of Principal Findings

Consistent with its positioning as a resource and methodology paper (Study Aims), the primary contribution of this study lies in its scale, transparency and reproducibility: 121 curated datasets spanning seven taxonomic groups across three kingdoms, a directional (one-sided) Fisher combination with a calibration correction for the direction-selection step (Methods 2.3) that tests every gene observed in at least two datasets without a pre-test directional filter, a fully reproducible pipeline released as a public resource (Methods 2.5), and an explicit empirical account, via the sign-permutation null check above, of what the resulting 111,447 FDR-significant genes do and do not demonstrate. Enrichment analysis yielded direct or indirect support for three of the five biological pathway categories previously reported by da Silveira et al. (2020), with innate immunity representing an additional related category and circadian-rhythm enrichment not recovered; this provides biological contextualisation and partial replication of previously reported spaceflight-response categories at varying levels of specificity, independent of any claim to methodological primacy. This is not presented as fully independent external validation, given the partial dataset overlap with that prior study noted in Discussion 4.3.

This study conducted a p-value-based meta-analysis of 121 GeneLab/OSDR datasets comparing spaceflight and ground control conditions. These datasets encompassed seven groups of organisms from three kingdoms. The number of datasets analysed is greater than in some previous GeneLab meta-analyses, which generally focused on a single taxon, although differences in the scope of the studies mean that direct comparisons are not entirely equivalent.

This large number of significant genes should not be read as a claim that 111,447 genes are each individually, biologically important to the spaceflight response. Fisher's method, by construction, tests whether the collection of per-dataset directional p-values provides aggregate statistical evidence for a specified direction, not whether a majority of contributing datasets individually agree on that direction. Because it is an aggregate test, a small number of datasets with very strong directional evidence can, in principle, produce a small combined p-value even when most contributing datasets show weak or opposing evidence; this is a property of Fisher's method under any use, not an artefact specific to this pipeline. For this reason, FDR-significance under this test does not by itself establish that a gene's direction of change is representative of most contributing datasets, and should not be interpreted as such. The directional_concordance metric reported alongside each gene (Methods 2.3, Section 2.2.1) is the appropriate quantity for that question, and low concordance despite a small combined p-value should be read as aggregate evidence concentrated in a subset of datasets rather than as broad agreement. FDR-significance in this study therefore answers the question 'does the pooled directional evidence across datasets favour one direction more than chance would predict?' rather than 'how large is this gene's spaceflight response?', 'is this gene of comparable biological importance to every other significant gene?', or 'do most contributing datasets agree on this gene's direction?' For this reason, we regard the rank ordering of genes by combined evidence strength (Methods 2.3; Table 3; the companion metrics accompanying each gene, including median effect size and per-dataset contribution) as more informative for prioritising candidates than the binary significant/not-significant classification alone, and we discourage treating the total count of significant genes as a headline measure of biological effect.

The sign-permutation null check reported in Results 3.2 substantiates this concern empirically: directional significance under this pipeline tracks two-sided differential-expression significance too closely to serve as strong evidence of directionally-consistent recurrence on its own. The headline count of 111,447 significant genes is therefore better read as an index of detectable two-sided differential expression pooled across datasets than as a direct count of genes with confirmed directional recurrence; directional_concordance and the companion metrics (Methods 2.3, Table 3) remain the appropriate resource for that latter question, and readers prioritising candidate genes should weight the total significant-gene count accordingly.

Biological pathway enrichment analysis of the significant genes recovered several patterns previously reported by da Silveira et al. (2020), though not always in the same species as under the earlier v5 procedure. The categories recovered were mitochondrial processes and oxidative phosphorylation, dominant in *Drosophila melanogaster*; innate immunity and a cluster of cytokine- and NF- κ B-signalling terms providing partial, mechanism-level support for the chronic inflammation category, both concentrated in *Homo sapiens*; and weak, indirect support for the cell cycle category (via a related senescence pathway term) in *Rattus norvegicus* (Results 3.5). Circadian rhythms were not identified in any species under either procedure.

The complete-ranked-list GSEA (Methods 2.4.1; Results 3.4) refines this picture in an important way: the mitochondrial and oxidative phosphorylation signal is not a single, uniformly-conserved cross-species

finding but two partially dissociable components. Suppression of translation-related gene sets was the more robust of the two, reaching significance under GSEA, with no gene-list cutoff applied, in three of the five animal species tested (*Mus musculus*, *Rattus norvegicus* and *Drosophila melanogaster*), which strengthens confidence that this component of the signal is not an artefact of the top-500 cutoff used in the primary over-representation analysis. Oxidative phosphorylation itself, by contrast, reached GSEA significance only in *Mus musculus*, and trended in the opposite direction, though not significantly, in *Caenorhabditis elegans*; the strong *Drosophila melanogaster* oxidative phosphorylation and electron transport chain signal reported above (Table 4) therefore rests more heavily on the top-500 over-representation result than on the cutoff-free GSEA check, and should be read with this distinction in mind. *Homo sapiens*, in both the top-500 over-representation analysis and the complete-ranked-list GSEA, showed no evidence of this mitochondrial signal at all, converging instead on up-regulated striated and cardiac muscle contraction pathways and down-regulated antiviral and interferon signalling. The convergence of the two enrichment procedures, which use different gene-selection rules (a fixed top-N cutoff versus a cutoff-free rank-based test), reduces concern that this profile is solely a consequence of the top-500 cutoff, although both analyses remain dependent on the same underlying *Homo sapiens* dataset pool and are therefore not fully independent checks. The profile more plausibly reflects the tissue composition of that dataset pool, which is not restricted to a single tissue type (Discussion 4.3), than a spaceflight response fundamentally distinct in kind from the other species studied.

Where the recovered signals can be linked to known spaceflight physiology, the links are consistent with, but do not by themselves establish, previously described system-level effects. The down-regulation of antiviral and interferon signalling in the *Homo sapiens* dataset pool is directionally compatible with the immune dysregulation reported in astronauts, which includes reactivation of latent herpesviruses (Crucian et al. 2018), and the up-regulation of striated and cardiac muscle contraction pathways overlaps with the cardiovascular system, one of the systems most consistently affected by spaceflight (Hughson et al. 2018). Because the *Homo sapiens* datasets combine several tissues and experimental systems (Discussion 4.3), we do not interpret these directions as evidence of a specific immune or cardiac mechanism. The suppression of translation-related gene sets in three animal species under the cutoff-free GSEA is the most consistent multi-species signal in this study; because translation was not among the pathway categories pre-specified for comparison with prior literature (Methods 2.4), we report it as a candidate for targeted follow-up rather than interpreting it further here.

In addition to previously reported findings, this study identified enrichment of general stress-response terms across species from different kingdoms, though the specific pattern differs from that reported under the earlier v5 procedure. Terms such as ‘response to chemical’, ‘response to abiotic stimulus’, ‘response to hormone’ and ‘response to endogenous stimulus’ were shared between *Arabidopsis thaliana* and *Homo sapiens*, and ‘response to stress’ itself was independently enriched in *Homo sapiens* and *Mus musculus* (Results 3.4). No single term was shared across three or more species under the present approach. This pairwise cross-kingdom overlap indicates that broad, stress-associated biological processes recur across some taxonomic groups in the curated datasets, though the terms involved (‘response to chemical’, ‘response to stress’, and similar broad Gene Ontology categories) are general

enough to be enriched under many kinds of perturbation, and the specificity of this pairwise overlap is insufficient to establish a conserved, spaceflight-specific mechanism. To date, such cross-kingdom patterns have not been extensively discussed in the GeneLab literature and are reported here as an exploratory observation warranting more mechanistically specific follow-up, rather than as evidence of a general biological response to the space environment.

Nevertheless, the terms identified represent broad Gene Ontology categories that can be triggered by various types of cellular stress. Consequently, the strength of the evidence is not on a par with that of more mechanistically specific biological pathways. To assess the possible underlying mechanisms, all significant enrichment results at the top-500 cutoff (867 terms; Results 3.4) were subsequently re-examined, focusing on pathways frequently associated with adaptation to the space environment, such as ROS-based oxidative stress, the heat shock protein and unfolded protein responses, and DNA damage response and repair.

Unlike the earlier v5 procedure, in which these specific categories were found only weakly and in *Arabidopsis thaliana* and *Rattus norvegicus* respectively (or not found at all), the directional Fisher top-500 gene set for *Homo sapiens* showed enrichment across all four mechanism categories examined: 'response to oxidative stress' (GO:BP, $p = 2.0 \times 10^{-5}$; 22 genes) and 'cellular response to oxidative stress' ($p = 1.4 \times 10^{-3}$; 15 genes); the chaperone-related Reactome terms 'XBP1(S) activates chaperone genes' ($p = 1.2 \times 10^{-3}$; 7 genes) and 'IRE1alpha activates chaperones' ($p = 1.6 \times 10^{-3}$; 7 genes); the Reactome term 'Unfolded Protein Response (UPR)' ($p = 1.4 \times 10^{-2}$; 8 genes); and the GO term 'intrinsic apoptotic signalling pathway in response to DNA damage' ($p = 3.1 \times 10^{-2}$; 9 genes). None of these four mechanism categories was enriched in any other species at the top-500 cutoff. Consequently, the cross-kingdom stress-response signal detected in this study (Section 4.1) remains a broad, pathway-level convergence, while the more mechanistically specific sub-categories examined here are concentrated within a single species rather than being shared across the two-kingdom pattern. Distinguishing which mechanisms are shared across species still requires further analysis at the level of individual orthologous genes (Results 3.3).

4.2 Comparison with Prior Work

The findings of this study share several similarities and differences with previous studies. Da Silveira et al. (2020) identified mitochondrial stress as the primary driver of the biological response to spaceflight, based on the average number of differentially expressed molecules across various mammalian tissues. The approach used in this study differs in three key aspects. Firstly, this study utilised the combined Fisher's p-value method, such that each gene was evaluated based on the strength of aggregate directional evidence across various datasets, rather than on the number of differentially expressed molecules. Secondly, the datasets used were sourced from the 2026 OSDR archive, thereby incorporating additional data spanning approximately five to six years following the publication of the study by da Silveira et al. Thirdly, the scope of the organisms analysed is broader, as it involves three kingdoms, not just mammals and human cell lines.

These methodological differences result in patterns of findings that are not entirely the same. In this study, a prominent mitochondrial and oxidative phosphorylation enrichment signal was found within the *Drosophila melanogaster* datasets analysed (Results 3.5), whilst the study by da Silveira et al. highlighted mammals as the primary source of these signals; given the limited number of species and datasets analysed on either side, this should not be read as evidence that the mitochondrial response is dominant or universal across spaceflight biology generally. Furthermore, this study independently recovered a pairwise stress-response pattern shared between *Homo sapiens* and *Arabidopsis thaliana*, and separately between *Homo sapiens* and *Mus musculus* (Results 3.4). Similar broad stress-related themes, including oxidative stress and DNA-damage responses, have previously been reported by Barker et al. (2019) and da Silveira et al. (2020). The present study differs by providing a formal p-value-based statistical framework for combining directional transcriptomic evidence underlying these observations.

However, this shift is likely influenced more by the statistical characteristics of the data than by underlying biological differences. Although represented by only seven datasets, 88.8 per cent of the genes tested in *Drosophila melanogaster* reached statistical significance under the directional Fisher approach. In contrast, in *Mus musculus*, which had 77 datasets, the proportion of tested genes reaching significance was 54.5 per cent (Table 1; Fig. 3). As the sign-permutation null check in Results 3.2 shows, this pattern is not itself evidence that direction is more uniform in species with fewer datasets; rather, because the directional test closely tracks two-sided significance on this dataset (Results 3.2, Discussion 4.1), a higher proportion of significant genes in a species with fewer datasets more plausibly reflects a lower effective multiple-testing burden and less averaging-out of two-sided signal across studies, rather than a stronger directional signal specifically. In large and heterogeneous datasets, inter-study variation can reduce the strength of the pooled two-sided signal even though an underlying biological effect remains. The curated datasets originated from different tissues, hardware platforms, missions, flight durations and assay technologies. Previous GeneLab studies have shown that these factors may explain a substantial proportion of transcriptomic variation. Consequently, some observed inconsistencies may reflect biological heterogeneity rather than statistical noise.

Therefore, the absence of mammalian mitochondrial signals at the top of the rankings cannot be directly interpreted to mean that mitochondrial dysfunction is not an important response to spaceflight. These results are more accurately viewed as a consequence of an approach that emphasises aggregate directional evidence across studies rather than agreement among a majority of studies. In other words, the approach used by da Silveira et al. (2020) and the approach used in this study capture different aspects of the same biological phenomenon. Both are better regarded as complementary rather than contradictory.

A second point of comparison concerns the level at which cross-species convergence appears. Morrison and Nicholson (2018), combining eight bacterial transcriptomes, found no response shared across all datasets and attributed most of the variation to differences between the experiments themselves, whereas Barker et al. (2023), within a single plant species, recovered spaceflight-associated changes in cell-wall processes and oxidative stress. The present results sit between these two outcomes. At the

gene level, significant ortholog overlap among top-ranked genes was confined to two mammalian pairs, both involving *Rattus norvegicus* (Table 2); no pair involving *Drosophila melanogaster* or *Caenorhabditis elegans* reached significance, and no candidate gene was shared across more than two species. At the pathway level, by contrast, broad stress-response, mitochondrial and translation-related categories recurred in more than one species (Results 3.4; Discussion 4.1). Taken together, these observations suggest that, in the OSDR data currently available, shared spaceflight-associated signals across distant taxa are more readily detected as convergence of pathway categories than as recurrence of the same orthologous genes. This is consistent with the conclusion of Morrison and Nicholson (2018) for bacteria, extended here to a multi-kingdom setting, although the limited ortholog-mapping coverage for *Arabidopsis thaliana* (Results 3.3) and the uneven number of datasets per species (Discussion 4.3) mean that gene-level convergence cannot be ruled out.

The effect of the number of datasets also needs to be understood with care. The Benjamini–Hochberg false discovery rate correction in this study was applied separately for each species. Consequently, the significance status of genes in *Mus musculus* is not influenced by the number of datasets available for *Drosophila melanogaster* or other species. The resulting gene rankings remain valid for comparing genes within the same species. However, the sensitivity analysis in Discussion 4.3 shows that in species with a large number of datasets, such as *Mus musculus*, the earlier v5 list of genes declared significant was sensitive to the concordance threshold used, whereas the present directional Fisher approach removes this threshold from the primary inferential procedure entirely.

4.3 Limitations

This study has several limitations that should be borne in mind when interpreting the results. Firstly, the number of datasets is uneven across taxonomic groups. *Mus musculus* has a far greater number of datasets than other species, whilst *Rattus norvegicus* and *Saccharomyces cerevisiae* are represented by only two datasets each. This imbalance affects the statistical power of Fisher’s method, meaning that the proportion of significant genes across species cannot be directly compared as a measure of the strength of the biological response to spaceflight. Therefore, differences in the number of significant genes are more accurately viewed as a consequence of data availability rather than underlying biological differences.

Secondly, 75 of the 203 studies meeting the inclusion criteria could not be included in the analysis as they provided only RSEM quantification results without the calculated differential expression tables. Consequently, some potentially relevant information could not be utilised. A re-analysis using raw count data has the potential to broaden the scope of the available evidence.

Thirdly, the harmonisation of gene identifiers was carried out automatically based on standard identifier formats such as Ensembl, TAIR, FlyBase and WormBase. This approach could not be applied to one dataset (OSD-554), which provided only gene symbols without sufficient information to identify the species with certainty. That dataset was ultimately analysed separately and was not integrated into the main meta-analysis.

Fourthly, biological contextualisation using g:Profiler could not be performed for *Bacillus subtilis* as the platform does not yet support the analysis of biological pathway enrichment in bacterial organisms. Consequently, the biological interpretation for this group remains limited, even though the proportion of significantly enriched genes is relatively high compared to the other species analysed.

Fifthly, conclusions regarding cross-kingdom stress responses are still based on similarities in biological pathway enrichment patterns, rather than on the identification of the same orthologous genes across all species. Thus, the results obtained indicate a convergence of biological functions, but cannot yet be used as evidence that the same genes are responsible for these responses in every organism.

Sixthly, the earlier filter-then-test (v5) pipeline was threshold-sensitive (Methods 2.3, Table S1 in Online Resource 1); this is resolved in the present directional Fisher approach, which requires no concordance threshold.

Seventh, Fisher's method assumes that all combined datasets are statistically independent. This assumption is unlikely to be consistently met, as some datasets within the OSDR may originate from the same spaceflight mission or animal cohort. A machine-readable mission identifier was not available from the OSDR API for any of the 121 datasets analysed (Methods 2.3), so this could not be assessed directly. Using dataset accession-number proximity as a proxy, we conducted a representative-dataset sensitivity analysis for the five species with candidate multi-dataset batches (Methods 2.3, Results 3.2): the overlap between the full-dataset and representative-only significant-gene sets was high for *Arabidopsis thaliana* (Jaccard 0.989) and *Drosophila melanogaster* (0.983), moderate for *Mus musculus* (0.871), and lower for *Caenorhabditis elegans* (0.741, though based on only three datasets). This indicates that non-independence, where it exists, has a measurable but generally modest effect on the significant-gene lists reported here, with *Mus musculus* the species most affected given its large number of datasets. Because this proxy is not a substitute for genuine mission or cohort metadata, we regard this as a partial mitigation rather than a full resolution of the independence assumption. This analysis should be interpreted as a sensitivity analysis rather than a formal test of independence: a high Jaccard overlap indicates that the reported significant-gene set is not highly sensitive to the specific candidate batches identified under this accession-proximity proxy, but it does not, and cannot, establish that the datasets contributing to any given gene's combined p-value are statistically independent in the sense Fisher's method assumes. If machine-readable mission metadata becomes available in future OSDR releases, this analysis should be repeated using it directly. This cross-accession independence check is complementary to, and distinct from, the within-accession dependence sensitivity analysis reported below (Ninthly), which addresses potential correlation among the multiple contrasts merged within a single accession rather than correlation among accessions themselves.

Eighthly, the *Homo sapiens* dataset used encompasses a variety of different cell types, tissues and experimental systems. Consequently, the results obtained for humans reflect general patterns observed across various human biological systems rather than the specific response of a particular tissue or organ.

Ninthly, the within-accession contrast aggregation procedure (Methods 2.2.1) combines all contrasts within an accession, including those from different tissues, timepoints, or genotypes where present, using unweighted Fisher combination for p-values and a simple mean for log2 fold-change, without modelling these contrasts as potentially non-independent or weighting them by sample size or biological relevance. Where an accession contains contrasts with opposing directions of effect for the same gene, the resulting mean log2 fold-change reflects a partial offset between them rather than an explicit resolution of the conflict. This is a simplification adopted for tractability across 121 heterogeneously structured accessions, and it may understate genuine tissue-, timepoint-, or genotype-specific effects that go in different directions within the same accession; a substudy-weighted or tissue-stratified aggregation procedure could be explored in future work. Relatedly, the Monte Carlo calibration reported in Methods 2.3 addresses the two-direction selection step under an independent-null simulation; it does not constitute a calibration of the full hierarchical Fisher procedure (within-accession contrast combination followed by cross-accession combination) under correlated contrasts or shared-mission datasets, both of which remain only partially addressed by the proxy-based sensitivity analyses described above.

As a direct empirical check on this concern, we re-ran the directional Fisher combination under two alternative treatments of within-accession contrast merging, using $n_{\text{substudies}}$ (the number of raw contrasts merged into each accession-level combined_pvalue; Methods 2.2.1) as an observable proxy for the risk of within-accession non-independence. In a capped-weight analysis, each accession's contribution to the cross-accession Fisher combination was down-weighted by $1/\sqrt{n_{\text{substudies}}}$, discounting evidence built from many potentially-correlated contrasts rather than treating it as full-strength independent evidence; in an exclude-high-substudy analysis, accession-gene combinations with $n_{\text{substudies}}$ above 20 were dropped entirely before re-combining. Compared with the primary (unweighted) result, the capped-weight analysis retained a significant-gene set with Jaccard overlap of 0.94-1.00 in five of seven species, 0.84–0.85 in *Homo sapiens* and *Mus musculus*, and 0.00 in the six-gene *Saccharomyces cerevisiae* cluster (too small for a meaningful overlap statistic); the exclude-high-substudy analysis showed a similar pattern (Jaccard 0.89-1.00) except in *Drosophila melanogaster*, where excluding the highest-substudy accession-gene combinations reduced overlap to 0.75. This indicates that within-accession dependence has a measurable but generally modest effect on the reported significant-gene sets for most species, with *Homo sapiens*, *Mus musculus* and *Drosophila melanogaster* the most sensitive to how heavily-merged accessions are treated. As with the representative-dataset and leave-one-tissue-out analyses above, this remains a sensitivity analysis rather than a formal correction for non-independence: it does not use the raw per-contrast p-values and log2 fold-changes (which were not retained separately from the within-accession combination step in the released pipeline), and $n_{\text{substudies}}$ is an observable proxy for merge count rather than a direct measure of statistical dependence among contrasts.

The within-accession sign-mismatch check reported as a central result in Results 3.2.1 (11.4 per cent of multi-contrast accession-gene rows flagged; total significant genes falling from 111,447 to 97,374, disproportionately in *Caenorhabditis elegans* and *Drosophila melanogaster*) is, we regard, the most consequential limitation identified in this study, rather than a minor caveat. Because the raw per-contrast

p-values and log2 fold-changes were not retained separately from the within-accession combination step (Ninthly, above), we cannot recompute the theoretically correct contrast-level directional combination for the affected genes; the exclusion analysis in Results 3.2.1 should be read as a lower-bound sensitivity check using an imperfect proxy, not as a corrected result.

Tenthly, the biological pathway comparison against da Silveira et al. (2020) (Results 3.5) is not fully independent of the source data. Cross-referencing the GeneLab dataset identifiers listed in the key resources table of da Silveira et al. (2020) against the 121 datasets curated for this study shows that 15 datasets overlap: 14 of the 77 *Mus musculus* datasets (18.2 per cent; from the Rodent Research-1 and Rodent Research-3 missions) and 1 of the 12 *Homo sapiens* datasets (8.3 per cent; a human umbilical vein endothelial cell dataset). No overlap was found for *Drosophila melanogaster*, *Rattus norvegicus*, *Arabidopsis thaliana*, *Caenorhabditis elegans* or *Bacillus subtilis*. This means the mitochondrial and oxidative phosphorylation signal in *Drosophila melanogaster* (Results 3.5), the species providing the strongest single-pathway replication in this study, is drawn from datasets entirely independent of da Silveira et al. (2020). By contrast, the innate immunity and cytokine/NF-κB signalling signal identified in *Homo sapiens* (Results 3.5) cannot be considered fully independent, given that one of the twelve contributing datasets was also analysed in that prior study; the effect of this single dataset on the overall *Homo sapiens* result was not separately assessed. The overlap in *Mus musculus* is more substantial (18.2 per cent of datasets) but did not, under either the v5 or the present directional procedure, produce a top-ranked signal matching da Silveira et al.'s categories (Results 3.5). We did not conduct a validation restricted to non-overlapping datasets; this comparison should be reported as biological contextualisation against partially overlapping prior literature rather than independent validation, and a dataset-overlap table is provided as a direction for future work (Discussion 4.4).

4.4 Future Directions

There are several areas for further development arising from this study. One of these is a comprehensive cross-kingdom orthology analysis using a single harmonised orthology framework capable of resolving identifiers across all seven taxonomic groups, extending the present pairwise g:Orth-based animal comparison and the targeted OMA follow-up for *Arabidopsis thaliana* (Results 3.3) into one consistent cross-kingdom comparison. This would help determine more definitively whether the similarities in biological pathways found across kingdoms also stem from the same orthologous genes, or arise from different genes that nevertheless function in similar pathways.

The analysis could also be expanded by utilising proteomic and metabolomic data available in the OSDR. Data at the RNA level do not always reflect changes occurring at the protein or metabolite levels. Therefore, integrating multiple types of data could help to examine the spaceflight response from a broader perspective. A similar approach was employed by Hauserman et al. (2024) in their multi-omics analysis of *Staphylococcus aureus* and can be compared with the multi-omics framework used by da Silveira et al. (2020).

A total of 75 datasets that could not be utilised in this analysis may still serve as data sources for future research. These datasets provide only sample-level data without differential expression results; consequently, they cannot yet be directly incorporated into the pipeline used. Should raw or intermediate data become available, these datasets could be reanalysed. This would be particularly useful for providing further evidence for taxonomic groups that currently have only a limited number of datasets.

Specifically for *Bacillus subtilis* and OSD-554, enrichment analysis using databases that support these bacterial organisms is still required. In this study, the limited availability of organism-specific data meant that validation of the enrichment results for these groups could not yet be carried out. The use of STRING-db, KEGG or other bacterial annotation sources could be employed in the next stage to determine whether the identified genes exhibit consistent patterns of biological function.

The statistical limitations identified in this study (Discussion 4.3) also point to specific methodological directions. A random-effects meta-analysis framework, in which between-dataset heterogeneity is modelled explicitly rather than assumed away by a fixed-effect Fisher combination, could accommodate the tissue- and mission-driven variability documented here without requiring the two-step filter-then-test procedure this study set out to avoid. Relatedly, a mission-level hierarchical model, nesting contrasts within accessions within missions, would let dependence among within-accession contrasts (Results 3.2.1) be estimated directly from the data rather than assessed only through post-hoc sensitivity checks, provided machine-readable mission metadata becomes available from OSDR. For the cross-species analysis, a phylogenetically-informed orthology framework, weighting or interpreting ortholog overlap in light of the evolutionary distance separating each species pair rather than treating all ten pairwise comparisons as exchangeable, would make better use of the pattern already visible in Table 2, where the two FDR-significant comparisons both involve *Rattus norvegicus* and the other mammals. Finally, this study's reliance on p-value combination, chosen because standard errors are not uniformly available across the curated differential expression tables (Methods 2.3), means that genes are ranked by evidential strength rather than by a comparable effect size; an effect-size integration approach, whether through re-analysis of raw counts where available or through standardised mean differences where they are not, would complement the present directional Fisher statistics with a magnitude-based ranking that this study cannot currently provide.

5. Conclusion

This study conducted a directional, one-sided p-value meta-analysis of 121 spaceflight transcriptomic datasets covering seven taxonomic groups from three kingdoms. By embedding the direction of expression change directly into the significance test itself, rather than applying it as a separate pre-test filter, and combining statistical evidence across datasets using Fisher's method (Methods 2.3), this study produced a ranking of genes based on the strength of aggregate directional evidence for their response to spaceflight, with directional concordance reported alongside as a separate descriptive measure. A sign-permutation null check indicated that, on this dataset, directional significance closely tracks two-sided differential-expression significance (Results 3.2, Discussion 4.1), so the headline count of significant

genes is best read as an index of detectable differential expression pooled across datasets rather than as direct confirmation of directionally-consistent recurrence at the individual-gene level; `directional_concordance` and the companion per-gene metrics remain the appropriate resource for that latter question. This approach offers an alternative to previous meta-analysis methods, which predominantly involved normalising the number of genes or focusing on a single taxonomic group. All analysis results, including the data used, the dataset inclusion and exclusion criteria, and the analysis pipeline, have been made available via Zenodo and GitHub as resources for further analysis.

Biological pathway enrichment analysis recapitulated several previously reported pathway categories from da Silveira et al. (2020) at varying levels of specificity, ranging from direct enrichment of mitochondrial processes to weaker, mechanistically related signals for inflammation and cell-cycle regulation, though not always in the same species reported previously. These included mitochondrial processes and oxidative phosphorylation in *Drosophila melanogaster*; innate immunity and a cluster of cytokine- and NF- κ B-signalling terms providing partial support for the chronic inflammation category, both in *Homo sapiens*; and weak, indirect support for the cell cycle category (via a related senescence pathway term) in *Rattus norvegicus*. Furthermore, this study identified pairwise enrichment of stress-response terms shared between *Arabidopsis thaliana* and *Homo sapiens*, and independently between *Homo sapiens* and *Mus musculus*, rather than a single term shared across three or more species. These findings suggest the existence of transcriptional response patterns to stress conditions that arise in organisms from different kingdoms, albeit more narrowly than initially observed under an earlier, less statistically rigorous version of this pipeline. However, as the terms identified are general in nature, these results cannot yet be used to conclude that there is a single, specific molecular mechanism common to all species.

Ortholog analysis of the top 300 significant genes per species revealed 88 raw source-gene-to-target-ortholog matches (84 involving distinct target genes; Table 2) amongst the five animal species. No single gene was found to have an ortholog present in the top-ranked gene list of more than one other species, and this study does not propose a candidate cross-species responder gene under the present methodology (Results 3.3). A targeted follow-up analysis using the OMA orthology database identified a limited number of *Arabidopsis thaliana*-animal ortholog relationships (Results 3.3); however, identifier-resolution and cross-database limitations prevented a harmonised ortholog comparison across all taxa using the same pairwise framework applied to the five animal species. Consequently, the evidence for cross-kingdom responses in this study remains stronger at the biological pathway level than at the orthologous gene level.

Overall, the results indicate that responses to spaceflight comprise several components that may consistently emerge across different organisms, but the form of the response at the gene level is not always the same. Innate immunity and cytokine/NF- κ B signalling in *Homo sapiens*, and translation suppression recurring across *Mus musculus*, *Rattus norvegicus* and *Drosophila melanogaster* under the complete-ranked-list GSEA (Results 3.4), provide candidate pathways for future functional investigation; the *Drosophila melanogaster* oxidative phosphorylation signal, in particular, was supported by the top-500 over-representation analysis but not replicated under the cutoff-free GSEA (Discussion 4.1), so this signal

should be treated with more caution than the others until further evidence accumulates. Cross-kingdom stress response patterns still require further validation at the gene and multi-omics levels. These results may form the basis for future research that integrates orthologous, proteomic, metabolomic and experimental provenance data to obtain a more robust understanding of the biological response mechanisms to the spaceflight environment.

Our findings should be viewed primarily as a statistical formalisation and extension of previously proposed conserved spaceflight-response hypotheses rather than as evidence for an entirely new biological phenomenon.

Declarations

Ethics Approval

Not applicable. This study analysed only publicly available data from the NASA Open Science Data Repository and involved no new experiments on human participants or animals.

Competing Interests

The authors declare no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

S.R.A.H. and D.A.N. conceived and designed the study, curated and analysed the data, implemented the meta-analysis and validation pipeline, and interpreted the results. S.R.A.H. and D.A.N. wrote the original draft of the manuscript. I.A.N., R.D.S., M.R.L., M.R.R., D.Y.P.P. and E.T.P.S. contributed to writing, review, and editing of the manuscript. All authors read and approved the final manuscript.

Acknowledgement

The authors thank NASA and the U.S. Government for making the Open Science Data Repository (OSDR) and its GeneLab spaceflight transcriptomic datasets openly available, without which this meta-analysis would not have been possible. The authors also thank Indonesia Defense University (Universitas Pertahanan Indonesia) and Sepuluh Nopember Institute of Technology (ITS) for their institutional support.

Data Availability

All raw and processed datasets analysed in this study are publicly available through the NASA Open Science Data Repository (<https://osdr.nasa.gov>). The complete analysis pipeline (all curation, meta-analysis, and validation scripts) is available at <https://github.com/sulthanrafiabyanhidayat-cloud/genelab-spaceflight-meta-analysis>. The complete per-gene directional Fisher statistics, p-values, and companion metrics (Methods 2.3) for all species are additionally archived on Zenodo with a permanent DOI: <https://doi.org/10.5281/zenodo.22676248>.

References

1. Ball, B.K., Khan, H.F., Park, J.H., Jayant, K., Chan, D.D., Brubaker, D.K.: Integrated cross-species translation and biophysical multi-scale modeling links molecular signatures and locomotory phenotypes in spaceflight-induced sarcopenia. *npj Microgravity*. **12**, 13 (2026)
2. Barker, R., Weitz, S., do Silva, W., McDonald, J.T., Reinsch, S., Vandenbrink, J., Gilroy, S., Beheshti, A., McKenna, M.: Cross kingdom analysis of data within the GeneLab repository identifies a potential conserved response of life to the stress associated with spaceflight. NASA Technical Reports Server, NTRS 20190033196 (2019)
3. Barker, R.J., Lombardino, J., Rasmussen, K., Gilroy, S.: TOAST: a discovery environment to explore multiple plant biology spaceflight experiments. *Front. Plant Sci.* **11**, 147 (2020)
4. Barker, R., et al.: Meta-analysis of the space flight and microgravity response of the *Arabidopsis* plant transcriptome. *npj Microgravity*. **9**, 21 (2023)
5. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B.* **57**, 289–300 (1995)
6. Berrios, D.C., Galazka, J., Grigorev, K., Gebre, S., Costes, S.V.: NASA GeneLab: interfaces for the exploration of space omics data. *Nucleic Acids Res.* **49**, D1515–D1522 (2021)
7. Comfort, P., et al.: Effects of spaceflight on musculoskeletal health: a systematic review and meta-analysis, considerations for interplanetary travel. *Sports Med.* **51**, 2097–2114 (2021)
8. Crucian, B.E., et al.: Immune system dysregulation during spaceflight: potential countermeasures for deep space exploration missions. *Front. Immunol.* **9**, 1437 (2018)
9. da Silva, W.A., et al.: Comprehensive multi-omics analysis reveals mitochondrial stress as a central biological hub for spaceflight impact. *Cell*. **183**, 1185–1201 (2020)
10. Fisher, R.A.: *Statistical Methods for Research Workers*, 4th edn. Oliver and Boyd, Edinburgh (1932)
11. Hauserman, M.R., Ferraro, M.J., Carroll, R.K., Rice, K.C.: Altered quorum sensing and physiology of *Staphylococcus aureus* during spaceflight detected by multi-omics data analysis. *npj Microgravity*. **10**, 2 (2024)
12. Hughson, R.L., Helm, A., Durante, M.: Heart in space: effect of the extraterrestrial environment on the cardiovascular system. *Nat. Reviews Cardiol.* **15**, 167–180 (2018)

13. Huo, Z., Tang, S., Park, Y., Tseng, G.: P-value evaluation, variability index and biomarker categorization for adaptively weighted Fisher's meta-analysis method in omics applications. *Bioinformatics*. **36**, 524–532 (2020)
14. Kaleb, K., Vesztry, A.W., Altenhoff, A., Dessimoz, C.: Expanding the Orthologous Matrix (OMA) programmatic interfaces: REST API and the OmaDB packages for R and Python. *F1000Research* **8**, 42 (2019)
15. Kim, J., Tierney, B.T., Overbey, E.G., et al.: Single-cell multi-ome and immune profiles of the Inspiration4 crew reveal conserved, cell-type, and sex-specific responses to spaceflight. *Nat. Commun.* **15**, 4954 (2024)
16. Kolberg, L., et al.: g:Profiler: interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023)
17. Kuznetsov, D., et al.: OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* **51**, D445–D451 (2023)
18. Lai Polo, S.-H., et al.: RNAseq analysis of rodent spaceflight experiments is confounded by sample collection techniques. *iScience*. **23**, 101733 (2020)
19. Lee, R., et al.: Genomic analysis of cardiovascular diseases utilizing space omics and medical atlas. *Genes*. **16**, 996 (2025)
20. Li, J., Tseng, G.C.: An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies. *Annals Appl. Stat.* **5**, 994–1019 (2011)
21. Morrison, M.D., Nicholson, W.L.: Meta-analysis of data from spaceflight transcriptome experiments does not support the idea of a common bacterial spaceflight response. *Sci. Rep.* **8**, 14403 (2018)
22. NASA Open Science Data Repository (OSDR): (2026). <https://osdr.nasa.gov> Accessed 8 September 2026
23. Overbey, E.G., et al.: NASA GeneLab RNA-seq consensus pipeline: standardized processing of short-read RNA-seq data. *iScience*. **24**, 102361 (2021)
24. Rau, A., Marot, G., Jaffrezic, F.: Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinform.* **15**, 91 (2014)
25. Sakharkar, A., Berliner, A.J., Lukong, K.E., Sanders, L.M., Costes, S.V., Yang, J., Taghibiglou, C., Mason, C.E.: Systemic genome correlation loss as a central characteristic of spaceflight. (2024). *bioRxiv* 2024.01.24.577100
26. Seo, D., Strickland, H.F., Zhou, M., Barker, R., Ferl, R.J., Paul, A.-L., Gilroy, S.: GLARE: discovering hidden patterns in spaceflight transcriptome using representation learning. *npj Microgravity*. **11**, 76 (2025)
27. Subramanian, A., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005)
28. WormBase: *gpd-3* (gene) – *Caenorhabditis elegans*, WBGene00001685. (2026). https://wormbase.org/species/c_elegans/gene/WBGene00001685 Accessed 2026. See also

BioGRID: gpd-2 (K10B3.8) result summary. <https://thebiogrid.org/532822/table/Caenorhabditis-elegans/gpd-2.html> (2026). Accessed 2026

29. Zhang, G., Zhao, L., Li, Z., Sun, Y.: Integrated spaceflight transcriptomic analyses and simulated space experiments reveal key molecular features and functional changes driven by space stressors in space-flown *C. elegans*. *Life Sci. Space Res.* **44**, 10–22 (2025)

30. Zhang, Y., Zhao, L., Sun, Y.: Using single-sample networks to identify the contrasting patterns of gene interactions and reveal the radiation dose-dependent effects in multiple tissues of spaceflight mice. *npj Microgravity.* **10**, 45 (2024)

Figures

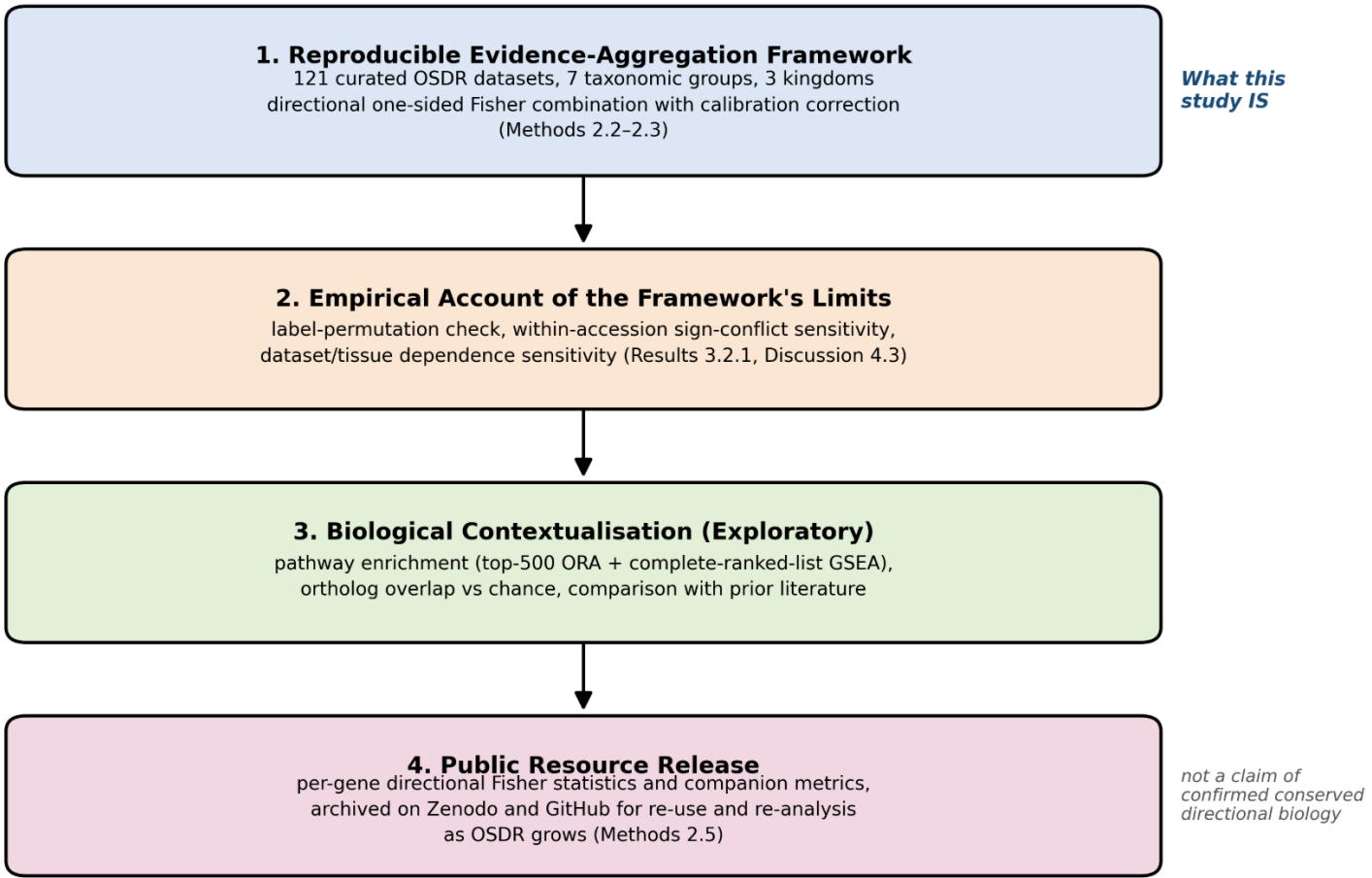
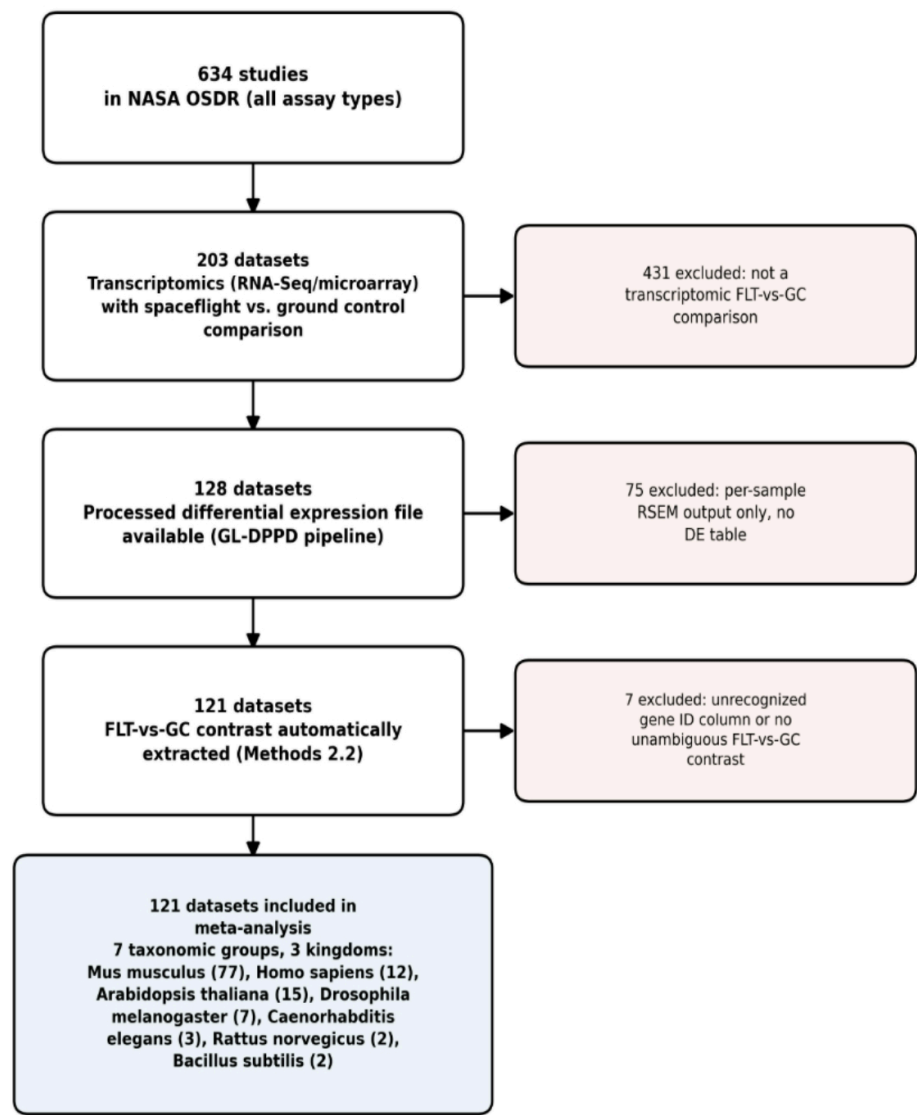


Figure 1

Four layers of contribution in this study, from the reproducible evidence-aggregation framework through to the released public resource. This study is positioned as a resource and methodology paper (Study Aims): layers 1 and 2 are the primary, statistically-grounded contributions, layer 3 is exploratory biological

contextualisation subject to the limitations discussed throughout, and layer 4 is the lasting output intended for re-use as OSDR grows



(One further dataset, OSD-554, reported separately as a single-dataset result; the two *Saccharomyces cerevisiae* datasets shared 6 genes, of which 3 were significant, and were excluded from the core cross-species synthesis)

Figure 2

OSDR dataset curation workflow illustrating how the dataset was selected from the NASA Open Science Data Repository (OSDR). Out of the 634 studies accessible in a frozen snapshot retrieved via the OSDR

REST API on 8 September 2026 (Methods 2.1), 203 satisfied the inclusion criteria described in Methods 2.1. Among those, 128 (63%) offered differential expression tables produced by the documented GeneLab pipeline, and for 121 of these 128 datasets (94.5%) the spaceflight-versus-ground-control contrast was successfully extracted; these 121 datasets were ultimately included in the cross-dataset meta-analysis

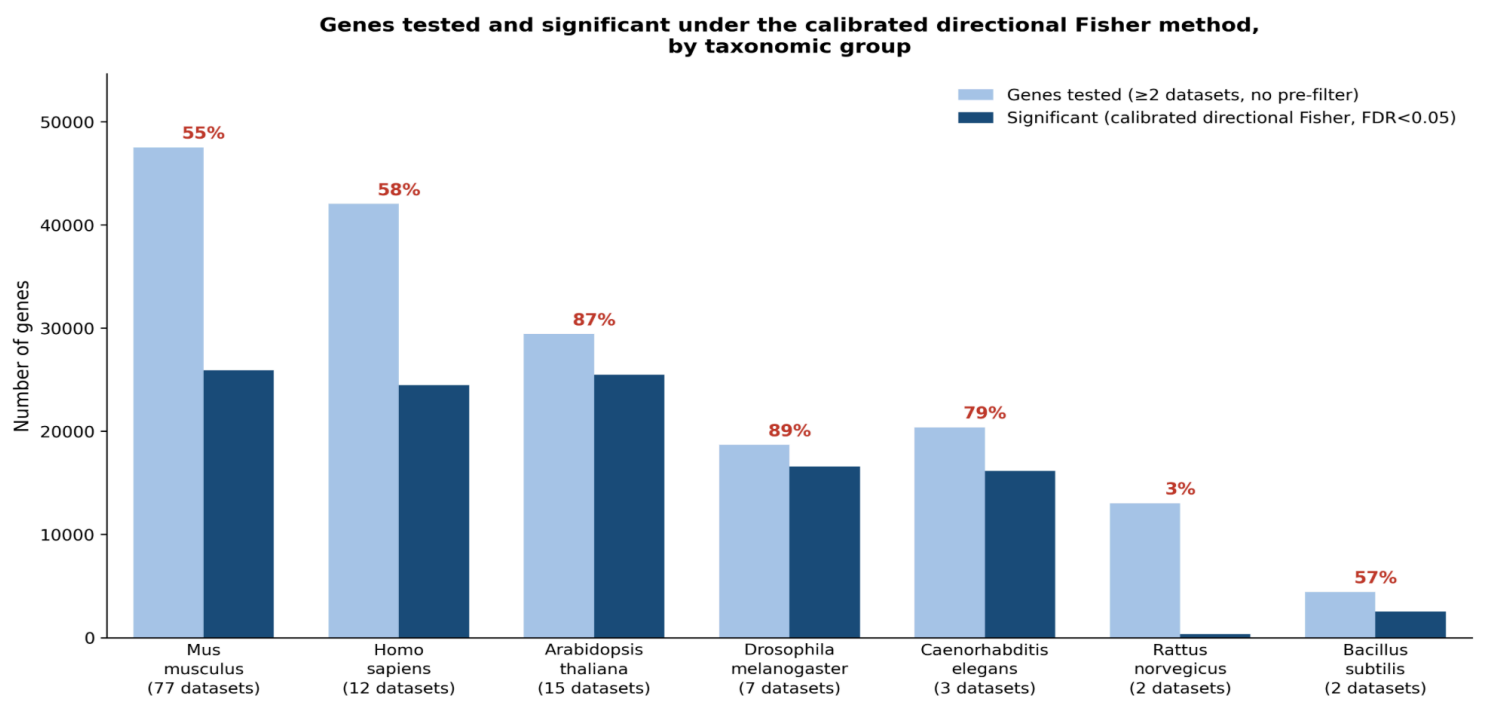


Figure 3

Directionally concordant genes that reached significance under the directional Fisher test by taxonomic group. The percentages listed above each data point represent the percent of tested genes that achieve significance, which illustrates the relationship between statistical power and number of datasets mentioned in Discussion 4.3 (compare *Mus musculus* with 77 datasets and 55% significant to *Drosophila melanogaster* with 7 datasets and 89% significant)

Directional significance closely tracks two-sided significance, not random-sign null rejection

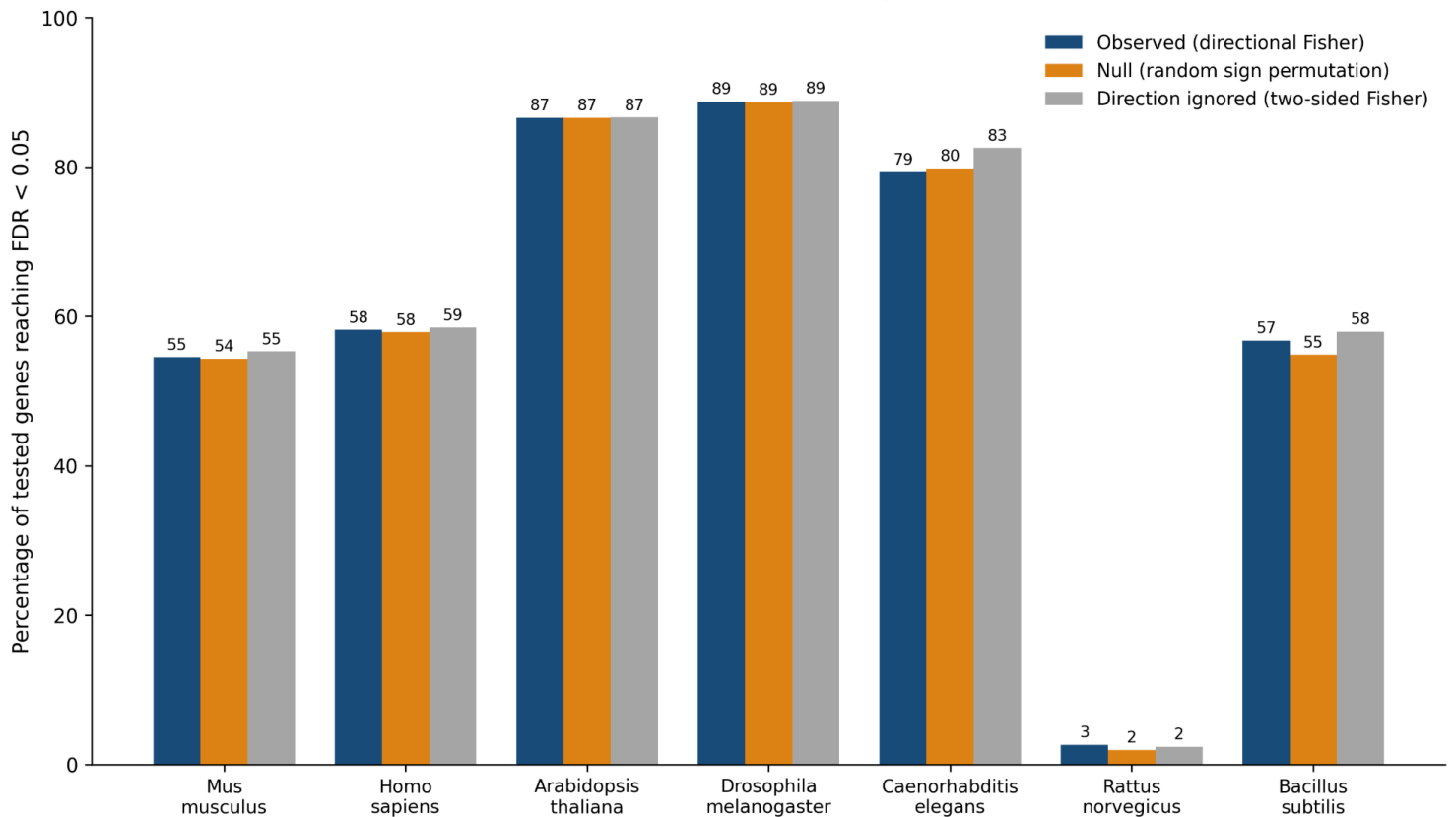


Figure 4

Percentage of tested genes reaching $FDR < 0.05$ under three conditions, by species: the observed directional Fisher result (Methods 2.3); the mean across ten sign-permutation null replicates in which each accession-gene row's sign was randomly reassigned, breaking any true directional signal while preserving the real combined_pvalue distribution and per-gene dataset counts (Results 3.2); and a two-sided Fisher combination with direction ignored entirely. The near-total overlap between all three bars in every species shows directly why a large proportion of the genome reaches significance under this procedure: significance is driven almost entirely by the strength of the underlying two-sided evidence pooled across datasets, not by the directional test's ability to distinguish genuine directional recurrence from random sign assignment. *Saccharomyces cerevisiae* (six genes total) is omitted as too small for a meaningful percentage

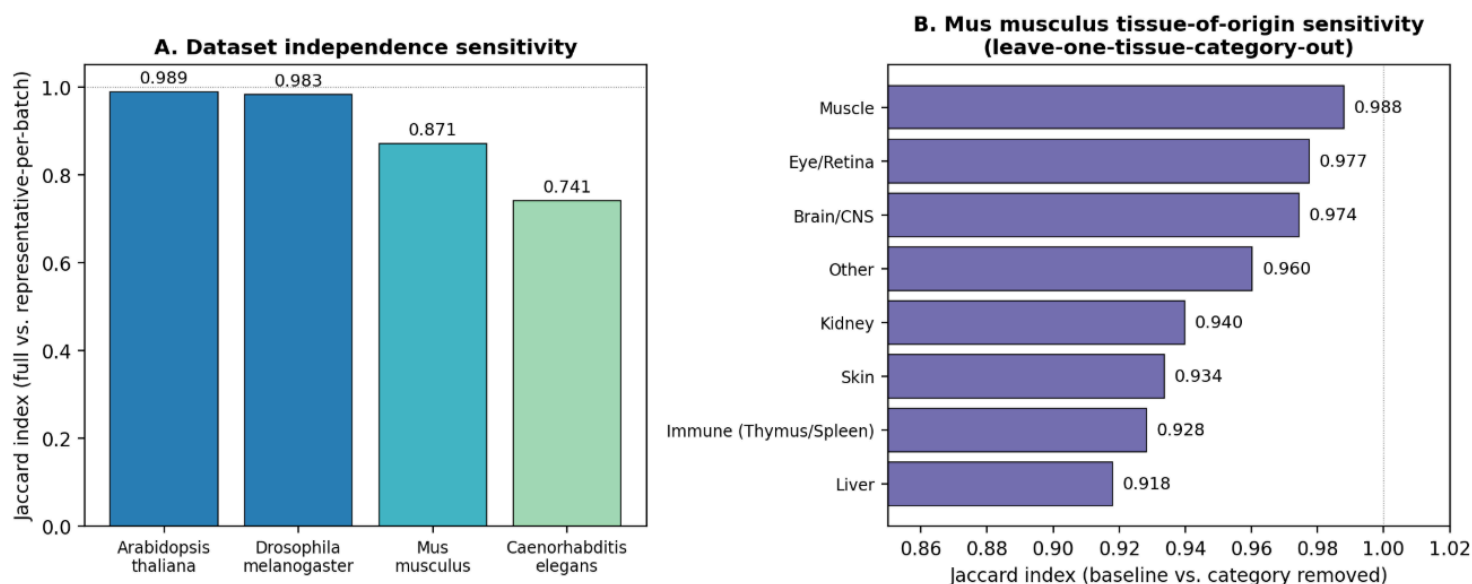


Figure 5

Sensitivity of gene prioritisation to candidate dataset dependence and tissue-of-origin heterogeneity (Methods 2.3). (A) Jaccard overlap between the full-dataset and representative-per-batch significant-gene sets, by species (Results 3.2); higher values indicate that the significant-gene set is less sensitive to the candidate multi-dataset batches identified via accession-number proximity. (B) Jaccard overlap between the full 77-dataset *Mus musculus* baseline and the significant-gene set obtained after removing all datasets from one tissue category at a time (leave-one-tissue-category-out; Results 3.2). This figure is intended as the central sensitivity-analysis result for this study's directional Fisher meta-analysis, replacing the earlier v5 filter-then-test concordance-threshold sensitivity analysis (Table S1 in Online Resource 1), which is retained as a historical record of the superseded procedure only

Pairwise gene-level ortholog overlap among top significant genes
Edge width and label = number of matched ortholog pairs; top 300 genes per species (Table 4)

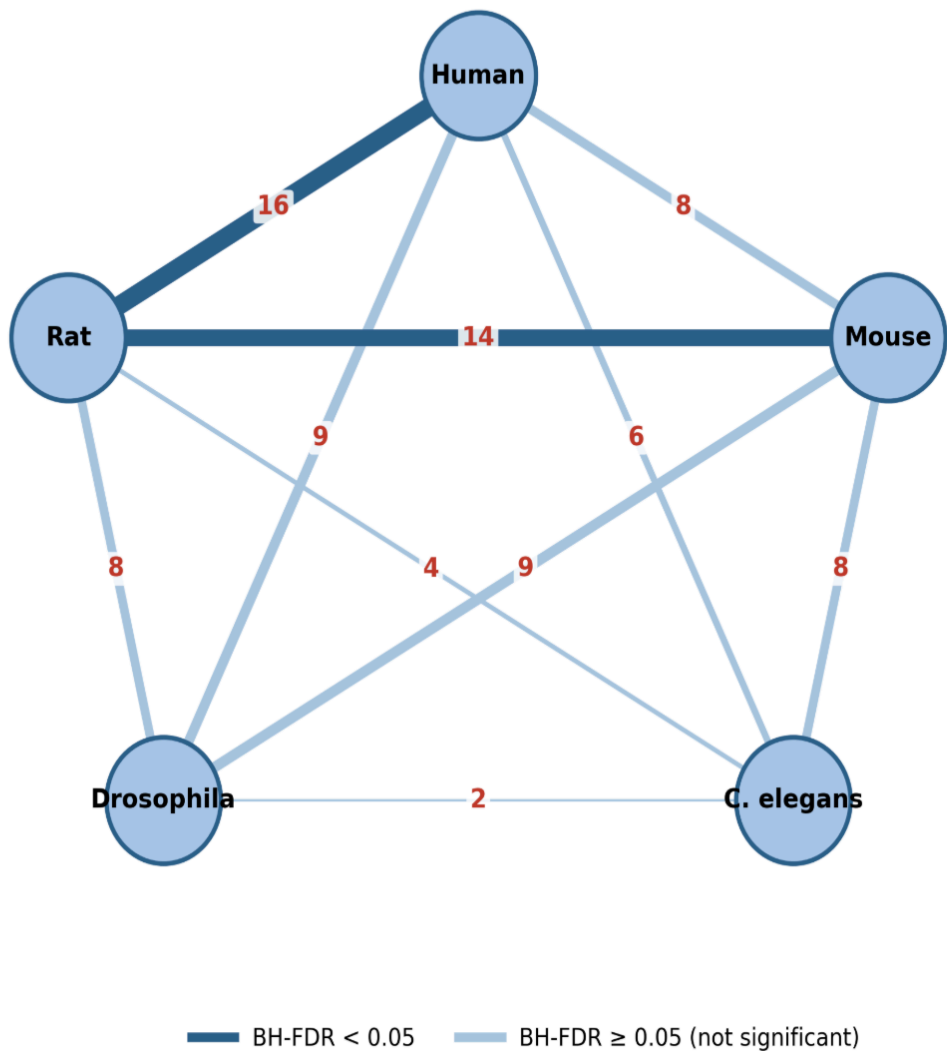


Figure 6
Gene-level ortholog overlap at the pair-wise level for the top 300 significant genes in the five animal taxa (Table 2). Edge thickness and label represent the number of matched ortholog pairs between each species pair; dark blue edges (*Mus musculus*–*Rattus norvegicus* and *Homo sapiens*–*Rattus norvegicus*) indicate the two comparisons surviving BH-FDR correction at FDR<0.05, while light blue edges indicate comparisons not reaching this threshold. No edges are included for *Arabidopsis thaliana* due to the g:Orth mapping limitation mentioned above

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ESM1.docx](#)