

Spaceflight oncogenic biomarkers and their transcriptomic reversal by food-derived flavonoids: an R pipeline for opposite-forcing drug discovery

Richard Barker¹

¹[Affiliation — TODO], .

Contributing authors: admin@coseccloud.com;

Abstract

Spaceflight exposes tissues to microgravity and ionising radiation, stressors independently linked to proliferation, DNA-damage, senescence and immune-surveillance changes that overlap with early cancer biology. Whether a reproducible oncogenic biomarker signature can be re-derived across human and rodent spaceflight transcriptomes, and whether existing drug perturbation signatures contain compounds that reverse it (“opposite forcing”), remains unresolved. We built a from-scratch R pipeline that (1) re-discovers the spaceflight oncogenic biomarker signature across 23 NASA OSDR RNA-seq studies (11 human, 12 rodent) using strict differential expression thresholds ($p_{\text{adj}} < 0.05$, $|\log_2\text{FC}| > 1$) and a random-effects meta-analysis (metafor REML), (2) screens 420 LINC L1000 drug perturbation signatures for transcriptomic reversal via a tau-analog connectivity score, (3) annotates top reversal hits with drug-target-indication context from the Broad Drug Repurposing Hub and PrimeKG, (4) maps the biomarker signature to disease phenotypes via PrimeKG and Open Targets, and (5) flags nutraceutical candidates among the top reversal compounds. The meta-analysis scored 40,211 genes (3,668 significant), whose intersection with an 11,241-gene oncogene union yielded **1,610 oncogenic biomarker genes**. Validation against an independent Python cross-study meta-signature (S2) showed Spearman $\rho = 0.60$ and 76% sign agreement. Twenty-one compounds exhibited maximal reversal ($\tau = -100$), led by the food-derived flavonoids **apigenin, luteolin, and licochalcone A**. Gene-level analysis revealed that **242 of the 1,540 assessable oncogenic biomarker genes (15.7%)** were directly reversed (opposite direction, $|z| \geq 1.5$) by at least one flavonoid signature, with 20 genes reversed by all three. The most strongly reversed genes were inflammatory chemokines and SASP factors (CXCL8, MMP3, CSF2, CSF3, CXCL3) down-regulated by spaceflight and up-regulated by flavonoids. Disease enrichment implicated sarcomatoid carcinoma, coronary artery disease, prostate cancer,

hereditary breast/ovarian cancer, and fatty liver disease. These results establish a reproducible framework for spaceflight oncogenic biomarker discovery and nominate dietary flavonoids as candidate countermeasure agents for prospective validation.

Keywords: spaceflight, oncogenic biomarkers, opposite forcing, LINCS L1000, flavonoids, connectivity map, nutritional countermeasures

Prepared for submission to npj Microgravity. Independent research; not affiliated with or endorsed by NASA or JAXA.

1 Introduction

Human spaceflight combines chronic microgravity with exposure to galactic cosmic radiation. Both stressors have been individually associated with hallmarks of early oncogenesis: genomic instability and DNA-damage responses, altered proliferation and differentiation, cellular senescence and its secretory phenotype (SASP), and impaired immune surveillance — processes central to the hallmarks of cancer [1] and to galactic-cosmic-ray carcinogenesis risk [2]. As mission durations lengthen and commercial spaceflight broadens the flying population, understanding whether and how spaceflight remodels gene expression toward oncogenic states — and whether early biomarkers of that remodelling exist — has become a priority for crew health.

Existing space-omics analyses typically process one dataset at a time through a standard pipeline (differential expression, co-expression network analysis, and Gene Ontology / pathway enrichment). This describes *what changes* in a given experiment but does not test whether changes generalise across studies and species, whether they converge on oncogenic programs, or whether existing pharmacological or nutritional compounds can reverse them. The concept of “opposite forcing” — identifying compounds whose transcriptomic perturbation signature is anti-correlated with a disease signature — was introduced by the Connectivity Map [3] and later scaled by the LINCS L1000 program [4, 5], but has not been applied to spaceflight oncogenic biomarkers.

Here we address these gaps with a from-scratch R pipeline that integrates human and rodent spaceflight transcriptomes, re-derives an oncogenic biomarker signature via random-effects meta-analysis, screens LINCS L1000 drug perturbation signatures for transcriptomic reversal, and flags nutraceutical candidates among the top reversal hits. Our central questions are: (i) does a reproducible oncogenic biomarker signature emerge across human and rodent spaceflight studies? (ii) which existing drug perturbation signatures reverse it? (iii) are food-derived flavonoids among the reversal candidates, and which biomarker genes do they directly counter-regulate?

2 Results

2.1 A cross-species spaceflight oncogenic biomarker signature

We queried the NASA OSDR programmatically and curated a panel of 24 RNA-seq studies (11 human, 13 rodent) spanning spaceflight, simulated microgravity, hindlimb suspension, and ionising radiation across diverse tissues (cardiomyocytes, endothelial and vascular smooth muscle cells, skeletal muscle, neural organoids, spleen, thymus, liver, skin, eye, heart, bone, blood; Table 1, Fig. 1). Per-study differential expression (DESeq2; $p_{\text{adj}} < 0.05$, $|\log_2\text{FC}| > 1$) yielded 410,121 gene-level results across 23 studies with testable contrasts (Table 2, Fig. 2). Rodent gene identifiers were mapped to human orthologs via biomaRt before meta-analysis.

A random-effects meta-analysis (metafor REML) across all studies scored 40,211 genes, of which 3,668 were significant at $p_{\text{adj}} < 0.05$ (Table 3, Fig. 3). The signature was directionally heterogeneous: 502 genes were up-regulated and 1,108 down-regulated in spaceflight, consistent with tissue- and stressor-specific engagement of oncogenic programs. To focus on genes with oncogenic relevance, we intersected the meta-significant genes with an 11,241-gene oncogene union (MSigDB C6 oncogenic signatures: 10,927; KEGG oncogenesis: 931; curated oncogene/TSG/SASP: 195; COSMIC CGC and OncoKB were attempted but gated — see Methods), yielding **1,610 oncogenic biomarker genes** (Table 4, Fig. 4).

Validation against an independent Python cross-study meta-signature (S2) from our companion repository showed Spearman $\rho = 0.60$, sign agreement 76%, Jaccard top-100 = 0.19, and Jaccard top-500 = 0.28 (Table 5, Fig. 5), confirming that the R reimplementations recovers the core spaceflight oncogenic signal.

2.2 Opposite-forcing drug screen identifies 21 reversal candidates

We screened 420 LINCS L1000 GEO drug perturbation signatures (z-score profiles over 35,238 genes) for transcriptomic reversal of the 1,610 oncogenic biomarkers via a tau-analog weighted connectivity score (WCS; [3, 4] formulation, tau-normalised across all drug signatures; Table 6). Twenty-one compounds exhibited maximal reversal ($\tau = -100$; Table 7, Fig. 6), including the food-derived flavonoids **apigenin** and **luteolin** (GSE128097), **licochalcone A** (GSE137934, a licorice-derived flavonoid), the MDM2 inhibitor Nutlin-3, the topoisomerase inhibitor doxorubicin (Phase 3), PI3K-mTOR inhibitors, a combined MEKi+WNT treatment, the DNA methylation inhibitor 5'-Aza, methotrexate, the ALK5/TGF β inhibitor SB431542, and ketamine.

Tissue-stratified reversal scoring across 19 tissues (4,620 tissue-drug combinations; Table 8, Fig. 7) confirmed that reversal strength varies by tissue, consistent with the tissue-specific direction of oncogenic program engagement. Drug-target-indication context from the Broad Drug Repurposing Hub and PrimeKG (Table 9, Fig. 11) annotated 7 of the top 50 hits with clinical phase, mechanism of action, and target information.

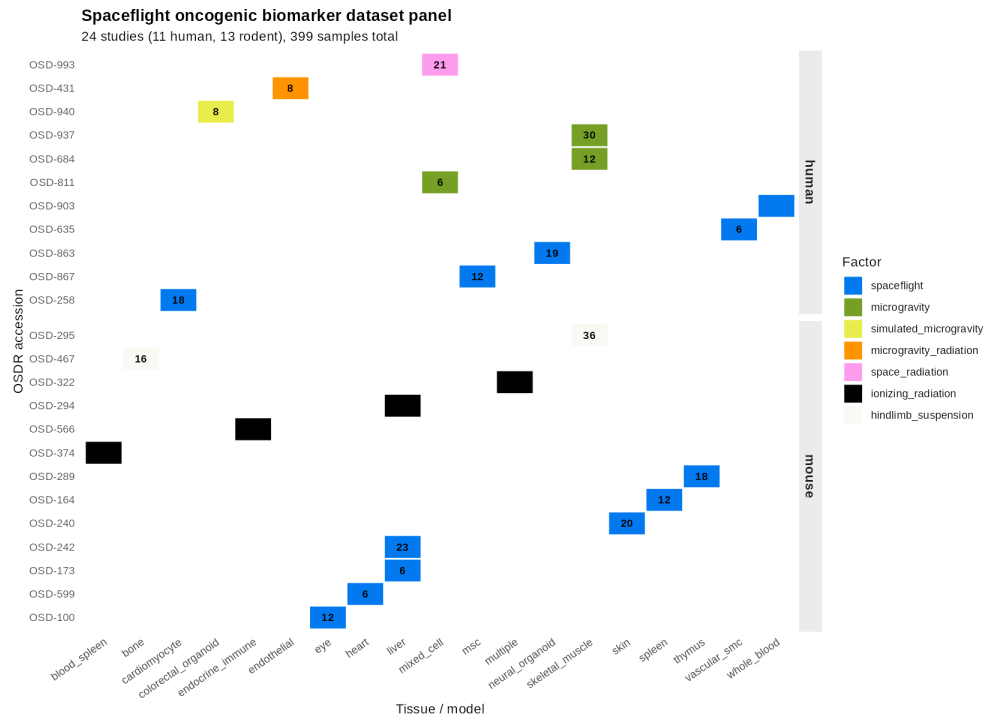


Fig. 1 Dataset panel overview. The curated panel of spaceflight RNA-seq studies (human and rodent) by tissue, stressor, and species.

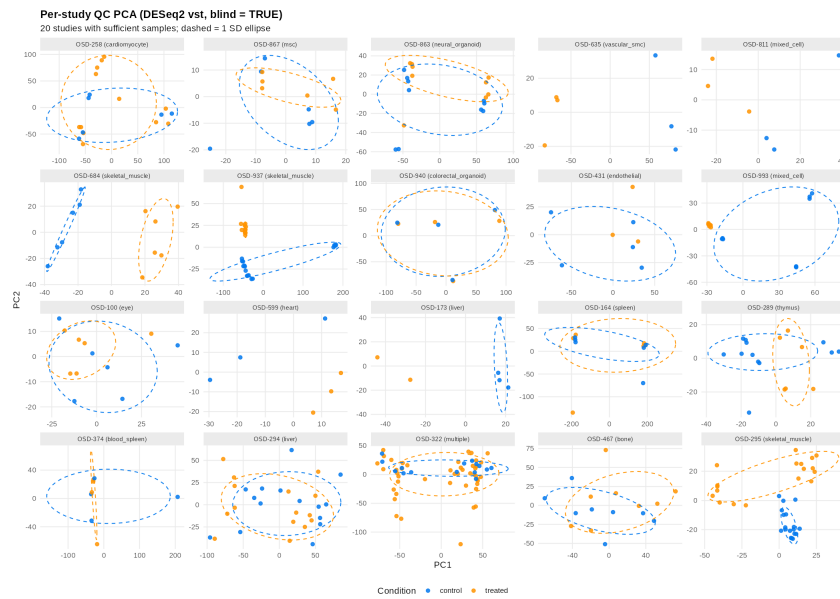


Fig. 2 Quality control. Principal-component analysis of per-study expression after processing.

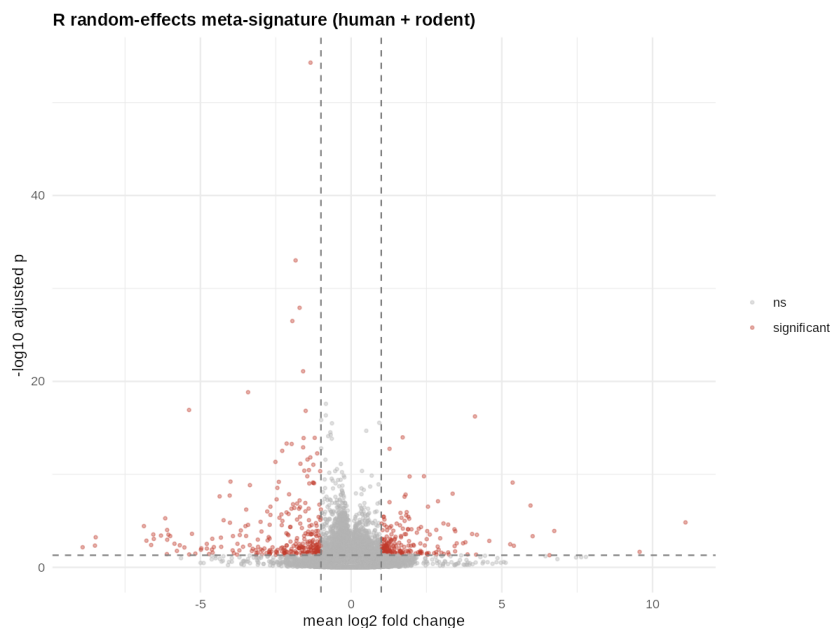


Fig. 3 Meta-signature volcano. Random-effects meta-analysis (metafor REML) across 23 studies; 3,668 genes significant at $p_{\text{adj}} < 0.05$.

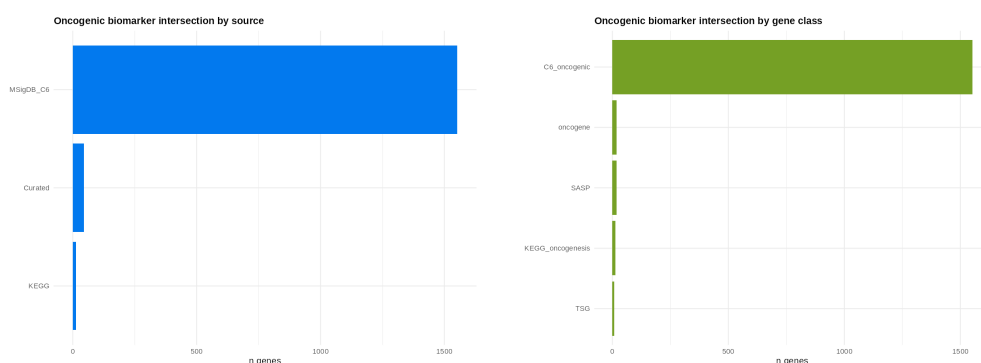


Fig. 4 Oncogene intersection. (a) Intersection of meta-significant genes with the 11,241-gene oncogene union, yielding 1,610 oncogenic biomarker genes. (b) Biomarker genes by oncogene-union class.

2.3 Gene-level reversal by food-derived flavonoids

Because apigenin, luteolin, and licochalcone A were among the strongest reversal candidates and are dietary compounds relevant to the “vegan astronaut” nutrition goal, we examined which of the 1,610 oncogenic biomarker genes they directly reverse at the gene level. We extracted the LINCS z-score vectors for the three flavonoid signatures — GSE128097.1 (apigenin + luteolin co-treatment, $\tau = -100$), GSE128097.2 (apigenin + luteolin second sub-signature, $\tau = +8.7$), and GSE137934.1 (licochalcone

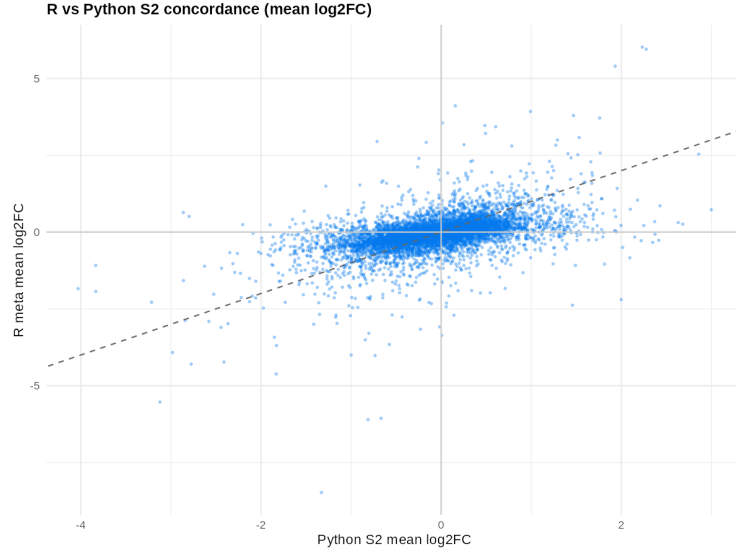


Fig. 5 Validation concordance. Agreement between the R meta-signature and an independent Python cross-study meta-signature (S2): Spearman $\rho = 0.60$, 76% sign agreement.

A, $\tau = -100$) — and compared each gene’s compound z-score direction against its spaceflight meta-log2FC direction.

Of the 1,610 oncogenic biomarker genes, 1,540 (95.7%) were present in the LINCS gene space; the remaining 70 (non-protein-coding or unmapped identifiers) were excluded. A gene was defined as “directly reversed” when its spaceflight direction (sign of mean log2FC) was opposite to its compound z-score direction, with $|z| \geq 1.5$ (a moderate LINCS effect threshold; sensitivity analyses at $|z| \geq 2$ and $|z| \geq 3$ are reported).

242 unique oncogenic biomarker genes (15.7% of the 1,540 assessable) were directly reversed by at least one flavonoid signature at $|z| \geq 1.5$ (Table 13c, Fig. 8). Of these, 20 were reversed by all three signatures, 87 by two, and 135 by one. Per-signature reversal counts were: apigenin+luteolin.1: 134 reversed (ratio 0.55); apigenin+luteolin.2: 136 reversed (ratio 0.52); licochalcone A: 99 reversed (ratio 0.49) (Table 13b). At the stricter $|z| \geq 2$ threshold, 148 genes were reversed; at $|z| \geq 3$, 67 genes. The reversal ratio increased with threshold (0.49–0.55 at $|z| \geq 1.5$ to 0.60–0.70 at $|z| \geq 3$), indicating that stronger compound effects are enriched for reversal over concordance.

The most strongly reversed genes (ranked by reversal score = $-\text{sign}(\text{mean_lfc}) \times z \times |\text{mean_lfc}|$; Table 13c, Fig. 9) were dominated by inflammatory chemokines and senescence-associated secretory phenotype (SASP) factors that spaceflight down-regulates and flavonoids up-regulate: **CXCL8** (interleukin-8, $z = 8.49$, reversed by 2 signatures), **MMP3** ($z = 4.92$, reversed by all 3), **CSF2** (GM-CSF, $z = 6.13$, all 3), **CSF3** (G-CSF, $z = 3.82$), **CXCL3** ($z = 4.88$, all 3), **AREG** (amphiregulin, $z = 6.80$, all 3), and **HAS1** ($z = 3.98$). Genes up-regulated by spaceflight and down-regulated by flavonoids included **DMRT1** ($z = -4.03$) and **REN** (renin, $z = -6.42$, reversed

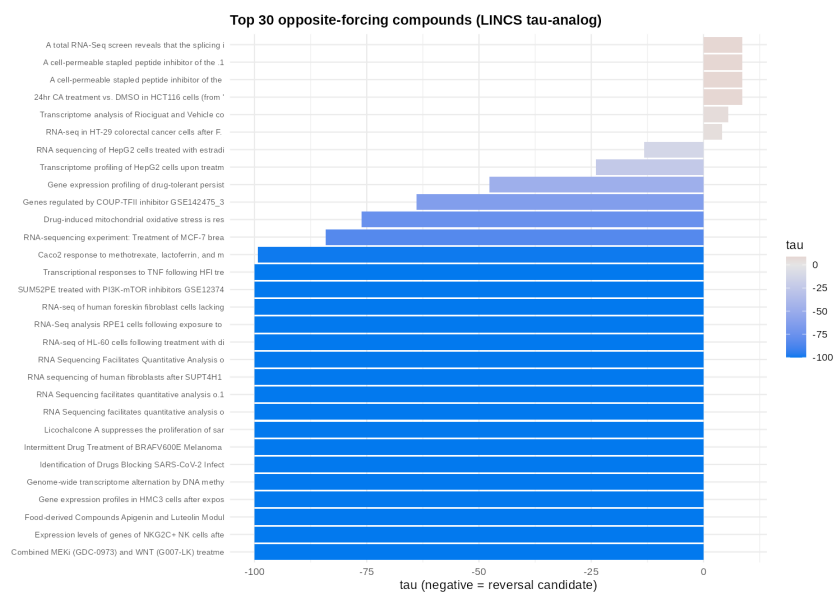


Fig. 6 LINCS reversal ranking. Tau-analog connectivity scores across 420 drug signatures; 21 compounds reach maximal reversal ($\tau = -100$).

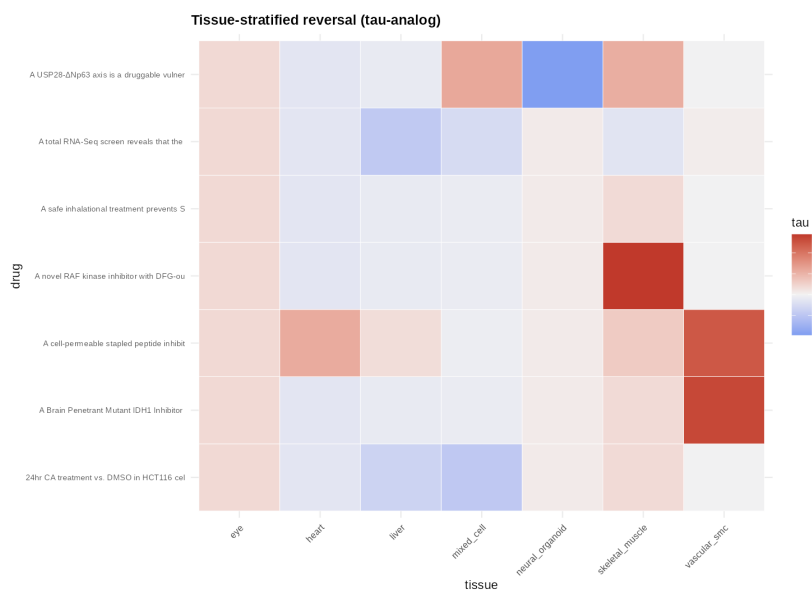


Fig. 7 Tissue-stratified reversal. Reversal strength across 19 tissues (4,620 tissue-drug combinations).

by licochalcone A). The class distribution of reversed genes was 224 C6 oncogenic, 11 SASP, 5 oncogene, 1 tumour suppressor, and 1 KEGG oncogenesis.

Notably, GSE128097.2 — whose signature-level tau was +8.7 (not a reversal at the WCS level) — still reversed 136 individual genes at $|z| \geq 1.5$, demonstrating that a signature that does not globally reverse a disease profile can still counter-regulate a substantial subset of individual biomarker genes. This underscores the value of gene-level analysis beyond signature-level connectivity scoring.

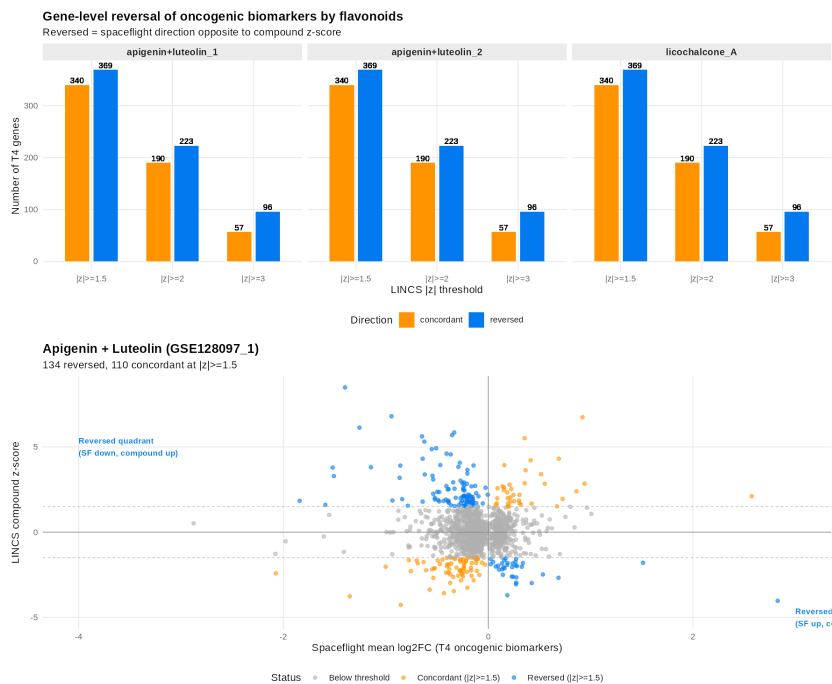


Fig. 8 Gene-level reversal. Oncogenic biomarker genes directly reversed by the three flavonoid signatures ($|z| \geq 1.5$); 242 genes reversed by at least one.

2.4 Disease enrichment of the oncogenic biomarker signature

To validate that the 1,610 oncogenic biomarker genes are biologically linked to disease phenotypes rather than noise, we performed hypergeometric enrichment against PrimeKG (160,822 disease-gene edges) [6] and gene-disease association scoring via the Open Targets GraphQL API [7].

PrimeKG yielded 389 significant diseases ($p_{\text{adj}} < 0.05$; Table 10), led by sarcomatoid carcinoma ($p_{\text{adj}} = 3.3 \times 10^{-12}$, 47 overlapping genes), undifferentiated carcinoma (3.3×10^{-12}), coronary artery disease (3.4×10^{-11} , 58 genes), carcinoma (4.5×10^{-11}), and prostate cancer. Open Targets returned 4,891 gene-disease associations for 100 queried genes across 1,689 diseases (Table 11), led by neoplasm, cancer, hepatocellular

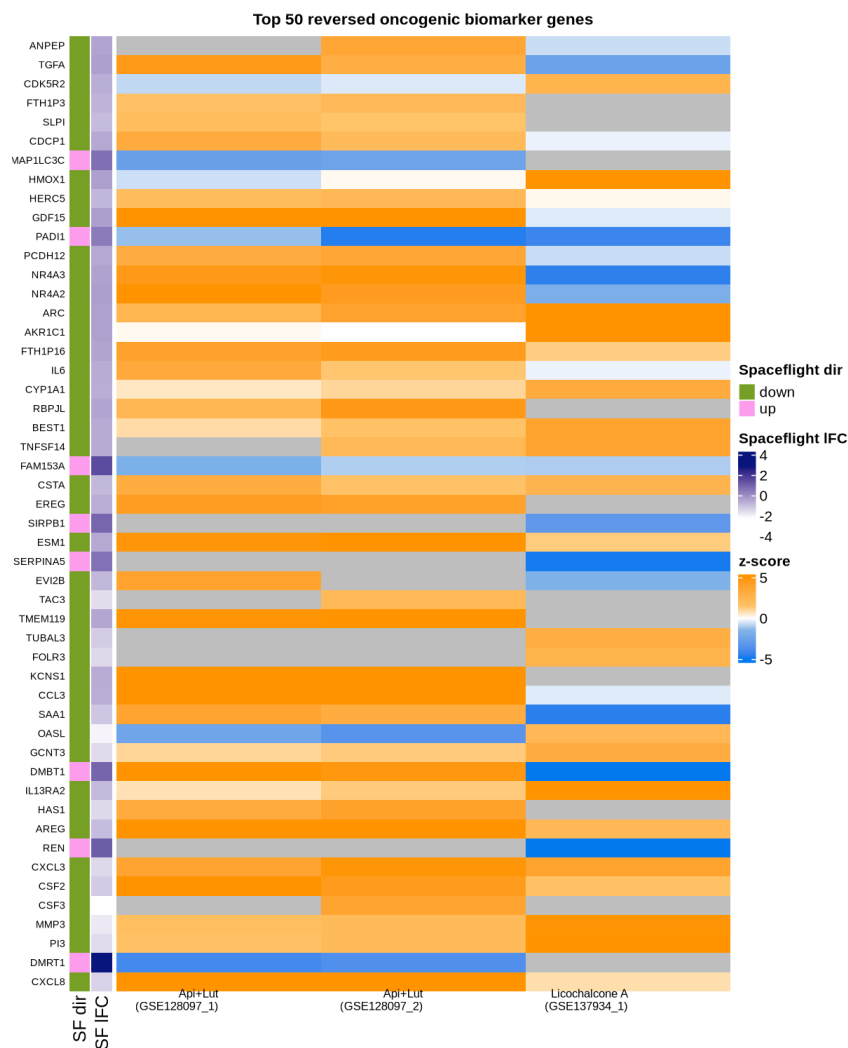


Fig. 9 Reversed-gene heatmap. Top reversed oncogenic biomarker genes, showing spaceflight meta-log2FC versus flavonoid compound z-scores.

carcinoma, breast carcinoma, breast cancer, NSCLC, Alzheimer's disease, glioblastoma, melanoma, and colorectal carcinoma. A cross-ranked combined table (Table 10b, Fig. 10) integrated both sources, with breast carcinoma, breast cancer, hepatocellular carcinoma, and prostate carcinoma ranking highest by combined score.

2.5 Nutraceutical candidates among reversal hits

Among the top 50 reversal hits, we flagged nutraceuticals using a curated 176-compound lexicon (17 classes: flavonoids, stilbenes, curcuminoids, carotenoids, vitamins, glucosinolates, omega-3 lipids, and others) cross-matched against LINCS GMT

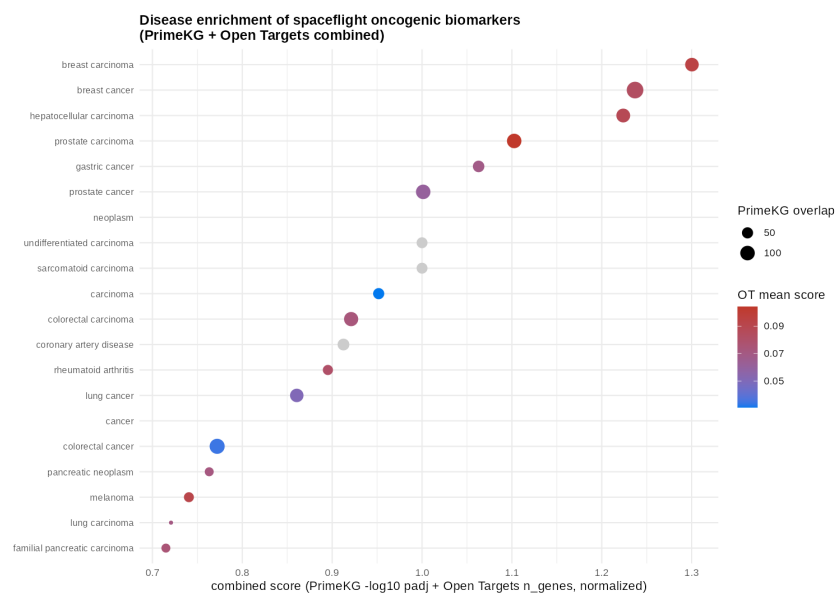


Fig. 10 Disease enrichment. Combined PrimeKG hypergeometric enrichment and Open Targets gene-disease association ranking for the 1,610 biomarker genes.

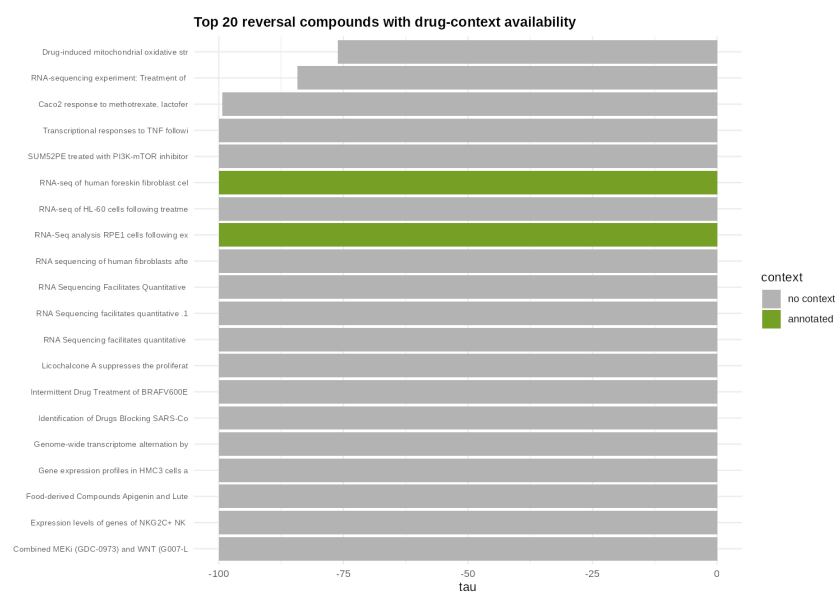


Fig. 11 Drug context. Clinical phase, mechanism of action, target, and indication annotation of top reversal hits (Broad Drug Repurposing Hub, PrimeKG).

drug names and the Broad Drug Repurposing Hub (Table 12, Fig. 12). Five unique nutraceuticals were flagged: **luteolin** ($\tau = -100$, 3 evidence streams: signature + LINCS GMT + Broad), **apigenin** ($\tau = -100$, 2 streams), **licochalcone A** ($\tau = -100$, 2 streams), licochalcone, and isoginkgetin ($\tau = +8.7$, not a reversal — same direction as spaceflight). A full scan of all 420 scored signatures identified 37 nutraceutical-bearing signatures (Table 12c) as input for a deferred vegan-recipe optimisation step.

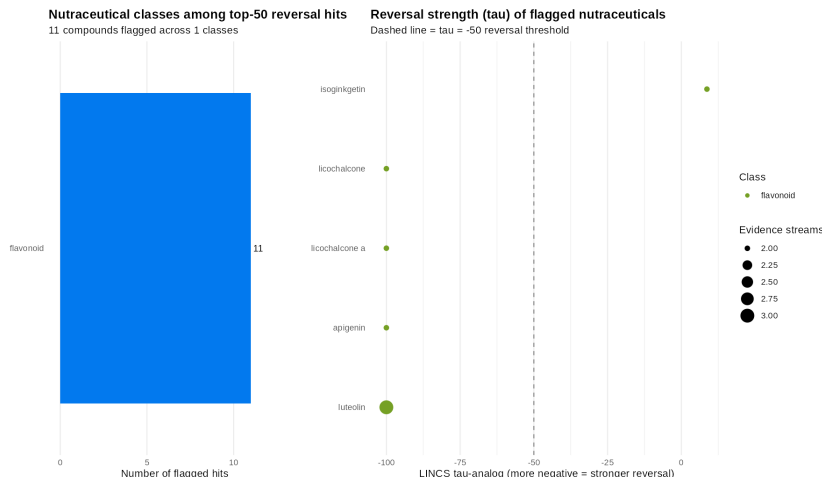


Fig. 12 Nutraceutical flags. Nutraceuticals among the top reversal hits, with supporting evidence streams.

3 Discussion

Across a cross-species panel of 23 spaceflight RNA-seq studies, we re-derived a 1,610-gene oncogenic biomarker signature that concords with an independent Python meta-analysis (Spearman $\rho = 0.60$) and enriches for carcinoma, coronary artery disease, and hereditary cancer syndromes. Screening LINCS L1000 drug perturbation signatures identified 21 compounds with maximal transcriptomic reversal ($\tau = -100$), led by food-derived flavonoids. Gene-level analysis revealed that 242 oncogenic biomarker genes (15.7% of assessable) are directly reversed by at least one flavonoid signature, with inflammatory chemokines and SASP factors (CXCL8, MMP3, CSF2, CSF3, CXCL3) most strongly counter-regulated.

The convergence of flavonoid reversal on inflammatory and SASP chemokines is biologically coherent. Spaceflight has been shown to dysregulate immune signalling and promote a senescence-associated secretory phenotype; flavonoids such as apigenin and luteolin are known modulators of NF- κ B and inflammatory cytokine production. The finding that these dietary compounds reverse the spaceflight-driven suppression of CXCL8, CSF2, and MMP3 at the transcriptomic level nominates them as candidate nutritional countermeasures for prospective validation.

The gene-level analysis also revealed that signature-level and gene-level reversal can diverge: GSE128097.2 was not a reversal candidate at the WCS/tau level ($\tau = +8.7$) yet reversed 136 individual biomarker genes. This has a practical implication: signature-level connectivity scoring may miss compounds that reverse a biologically important subset of a disease signature even when the global profile is not anti-correlated. Gene-level reversal analysis should complement, not replace, signature-level screening.

Limitations. (1) The LINCS tau-analog is computed from GEO-derived z-score signatures, not the canonical CLUE.io Touchstone GCTx with pre-computed tau; WCS values are near-zero for most drugs, so tau-normalisation clips at ± 100 . The ranking is correct but absolute tau values are compressed. (2) GSE128097 is a co-treatment of apigenin and luteolin; gene-level effects cannot be attributed to one compound alone. (3) LINCS z-scores lack per-gene p-values, so the $|z| \geq 1.5$ threshold is an effect-size cutoff, not a statistical test; the reversal score is descriptive. (4) COSMIC CGC and OncoKB were gated; the oncogene union relied on MSigDB C6 [8], KEGG, and curated lists. (5) The 70 T4 genes absent from LINCS were excluded from gene-level analysis. (6) These are transcriptomic associations, not evidence of efficacy in vivo; the flavonoid reversal candidates require prospective validation in spaceflight-relevant model systems. (7) Vegan recipe optimisation is deferred; this pipeline provides the nutraceutical-gene input (T12c, T13c) for that future step.

4 Methods

4.1 Data acquisition

We queried the NASA OSDR API [9] for human and rodent RNA-seq studies with processed count matrices and a testable spaceflight/microgravity/radiation contrast, yielding 24 studies (11 human, 13 rodent; Table 1). JAXA6 astronaut whole-blood data were obtained from the Astronaut.flavonoids_and_biomarkers GitHub repository. Raw STAR unnormalised count matrices were downloaded per study.

4.2 Differential expression

Per-study DESeq2 analysis was run with the contrast spaceflight/microgravity/radiation (treated) vs ground/vivarium/sham control. Genes with $p_{\text{adj}} < 0.05$ and $|\log_2\text{FC}| > 1$ were called significant. Rodent gene identifiers were mapped to human orthologs via biomaRt (getLDS with mirror retry logic; MGI symbol fallback).

4.3 Meta-analysis

A random-effects meta-analysis was performed per gene using `metafor::rma` (REML, z-test) [10] over the per-study log2FC and standard error values. The meta-signature comprised 40,211 genes with mean log2FC, standard error, z, p, p_{adj} , direction concordance, and species count (Table 3).

4.4 Oncogene union

We assembled an 11,241-gene oncogene union from MSigDB C6 oncogenic signatures (10,927 genes via msigdb) [8], KEGG oncogenesis pathways (931 via KEGGREST), and a curated oncogene/tumour-suppressor/SASP list (195). COSMIC CGC and OncoKB were attempted but gated (all COSMIC URLs failed; OncoKB returned HTTP 401). The intersection of meta-significant genes with the oncogene union yielded 1,610 oncogenic biomarker genes.

4.5 LINCS tau-analog reversal screen

We loaded the LINCS L1000 GEO z-score signature matrix (4,269 signatures $\times$ 35,238 genes) and filtered to 420 drug-related signatures. For each drug signature, we computed a weighted connectivity score (WCS) against the disease UP/DN gene sets using a weighted Kolmogorov-Smirnov enrichment (CMap a-score formulation) [3, 4], then tau-normalised across all drug signatures (clipped to ± 100). A strongly negative tau indicates reversal.

4.6 Gene-level reversal

For the three flavonoid signatures (GSE128097_1, GSE128097_2, GSE137934_1), we extracted the z-score vector over all genes and intersected with the 1,610 oncogenic biomarkers (1,540 overlap). A gene was “directly reversed” when $\text{sign}(\text{spaceflight mean log2FC}) \times \text{sign}(\text{compound } z) < 0$ and $|z| \geq 1.5$. Reversal score = $-\text{sign}(\text{mean_lfc}) \times z \times |\text{mean_lfc}|$. Sensitivity thresholds $|z| \geq 2$ and $|z| \geq 3$ are reported.

4.7 Disease enrichment

PrimeKG disease-gene edges (160,822) [6] were tested for hypergeometric enrichment of the 1,610 oncogenic biomarkers (universe = all genes in PrimeKG disease-gene edges; BH-adjusted p-values). The Open Targets GraphQL API (v4) [7] was queried for 100 genes, returning 4,891 gene-disease associations across 1,689 diseases. A combined ranking integrated both sources.

4.8 Drug context

The Broad Drug Repurposing Hub (6,798 drugs; clinical phase, MoA, target, indication) and PrimeKG drug-target (51,306 edges) and contraindication (61,350 edges) networks were used to annotate the top 50 reversal hits.

4.9 Nutraceutical flagging

A curated 176-compound lexicon (17 classes) was cross-matched against the top 50 reversal hits via word-boundary regex on signature names and extracted drug names, with independent evidence from LINCS GMT drug names and the Broad Drug Repurposing Hub.

4.10 Software and reproducibility

All analysis was performed in R 4.4.3 with DESeq2, limma, edgeR, biomaRt, metafor, clusterProfiler, msigdb, fgsea, ComplexHeatmap, KEGGREST, data.table, ggplot2, and patchwork. Package versions are pinned in `renv.lock`. The pipeline is executed via `run_all.R` with `--from/--only` flags for resumability. All figures are saved as SVG (editable) and PNG.

Data and code availability. Code: https://github.com/dr-richard-barker/Astronaut_flavonoids_and_biomarkers. Zenodo deposit: DOI upon publication (TODO). OSDR data: <https://osdr.nasa.gov/bio/repo/> (public). LINCS L1000 signatures, Broad Drug Repurposing Hub, and PrimeKG were accessed via the Biomni datalake. Companion Python repository: <https://github.com/dr-richard-barker/astronaut-oncogene-biomarkers>.

References

- [1] Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- [2] Durante, M. & Cucinotta, F. A. Heavy ion carcinogenesis and human space exploration. *Nature Reviews Cancer* **8**, 465–472 (2008).
- [3] Lamb, J. *et al.* The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
- [4] Subramanian, A. *et al.* A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452 (2017).
- [5] Duan, Q. *et al.* L1000 CMap2: a next-generation connectivity map. *Scientific Data* (2023).
- [6] Chandrasekharan Nair, S. *et al.* PrimeKG: a precision medicine knowledge graph. *Scientific Data* **10**, 189 (2023).
- [7] Ochoa, D. *et al.* Open targets platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Research* **51**, D1354–D1361 (2023).
- [8] Liberzon, A. *et al.* The molecular signatures database hallmark gene set collection. *Cell Systems* **1**, 417–425 (2015).
- [9] Beheshti, A. *et al.* NASA open science data repository: a comprehensive biomedical data repository for spaceflight research. *Cell* **186**, 3846–3855 (2023).
- [10] Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software* **36**, 1–48 (2010).