



FDL 2021 ASTRONAUT HEALTH

SPACE MEDIC:CAUSAL INFERENCE FOR OUT-OF-DISTRIBUTION GENERALIZATION



FDL | FRONTIER DEVELOPMENT LAB

Invariant Risk Minimization for Cross-Organism Inference: Substituting Mouse Data for Human Data in Human Risk Factor Discovery

Odhran O'Donoghue, Giuseppe Ughi, Linus Scheibenreif, Paul Duckworth, Kia Khezeli, Nicholas Chia, Adrienne Hoarfrost, Samuel Budd, Patrick Foley, John Kalantari, Graham Mackintosh, Frank Soboczenski, Lauren M. Sanders

THE CHALLENGE

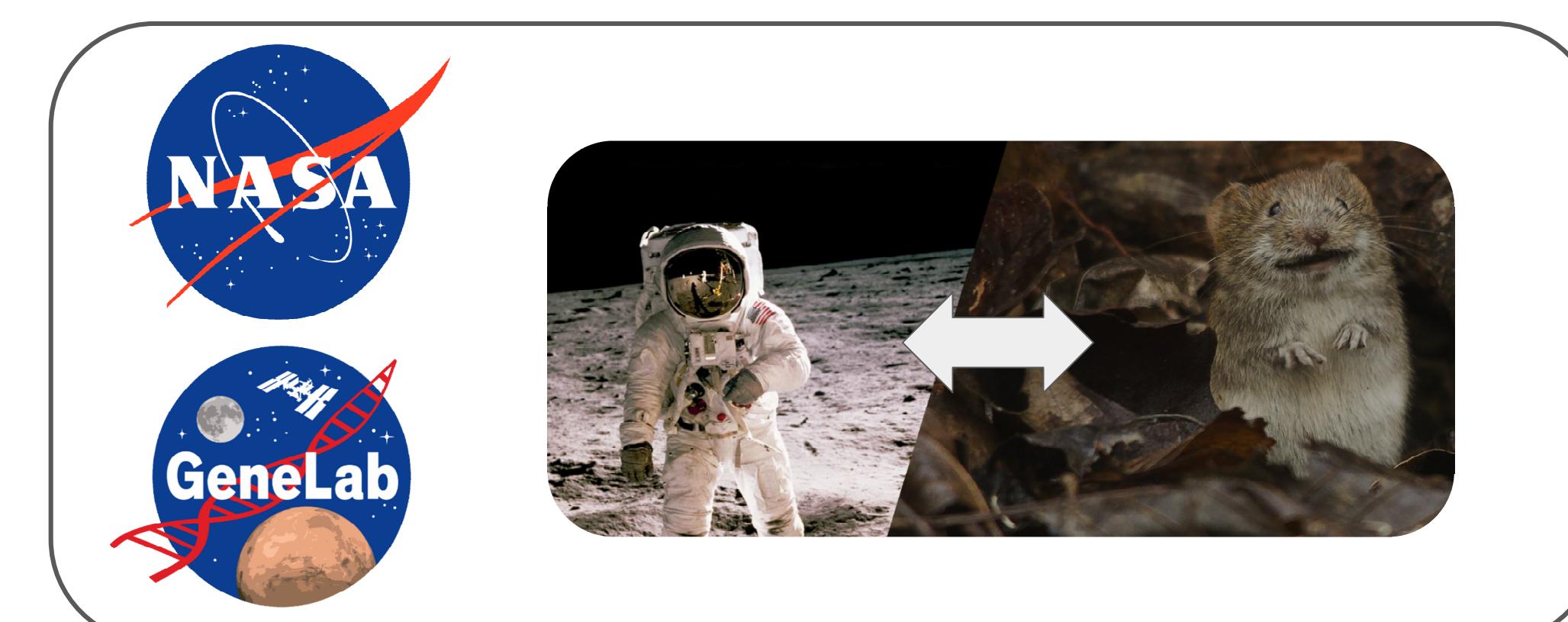
Can we supplement human biomedical data with mouse model data to achieve **higher statistical power** and characterize **causal factors** in human **radiation exposure**?

THE OPPORTUNITY

CRISP v1.0 allows to investigate the causal structure present in datasets where all the data is stored in the same location [1].

Translation between *large* amounts of animal model data and *small* amounts of human data is needed to assess applicability of animal findings to human patients.

CRISP v2.0 includes new models, including **Nonlinear IRM** [2] and CausalNex [3] for out-of-distribution domain generalization.



THE SOLUTION

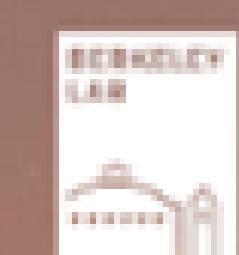
We leverage the Invariant Risk Minimization (IRM) method within **CRISP v2.0**.

- Identifying invariance across different data-generating environments improves **domain generalization for causal inference** [2,4,5]
- Environments:** subsets of data that do not share the same underlying populations, but for which the causal relationships to a target variable are assumed invariant
- We demonstrate that **environment-based mechanisms** are effective for identifying invariant relationships present in cross-organism datasets

$$\min_{\Phi: \mathcal{X} \rightarrow \mathcal{H}} \sum_{e \in \mathcal{E}_{tr}} R^e(w_0 \circ \Phi) + \lambda \|\nabla_{w|w=w_0} R^e(w \circ \Phi)\|^2 \quad \text{Equation (1)}$$

Experimental Design: IRM model classifies irradiated samples from non-irradiated controls, giving ranked list of features (genes) according to coefficients from model Φ in Equation (1), ranked by consistently predicting target variable across environments.

Google Cloud

USGS
science for a changing world

Microsoft

NVIDIA

LOCKHEED MARTIN

intel.

LSA LUXEMBOURG SPACE AGENCY

SK

MIT Portugal

planet.

IBM

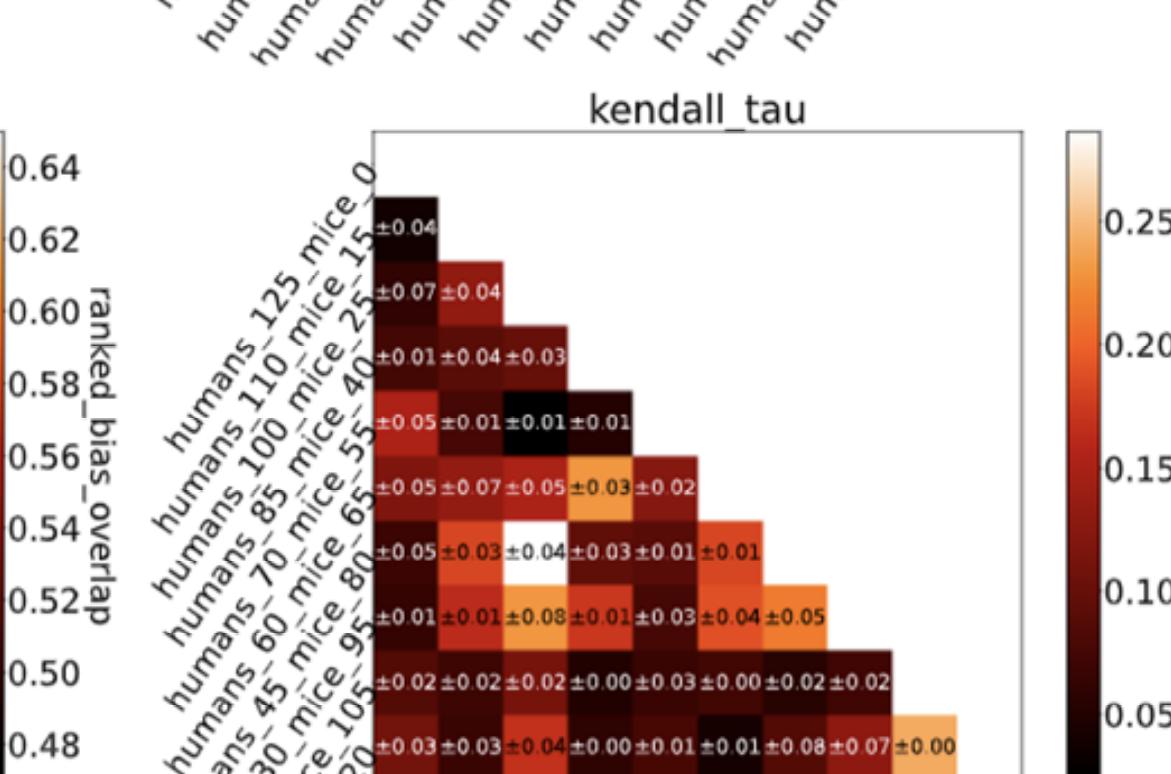
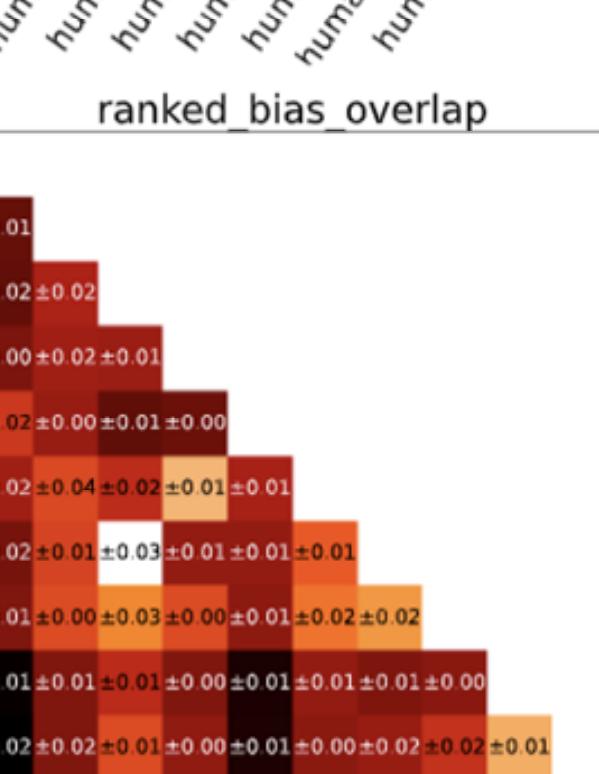
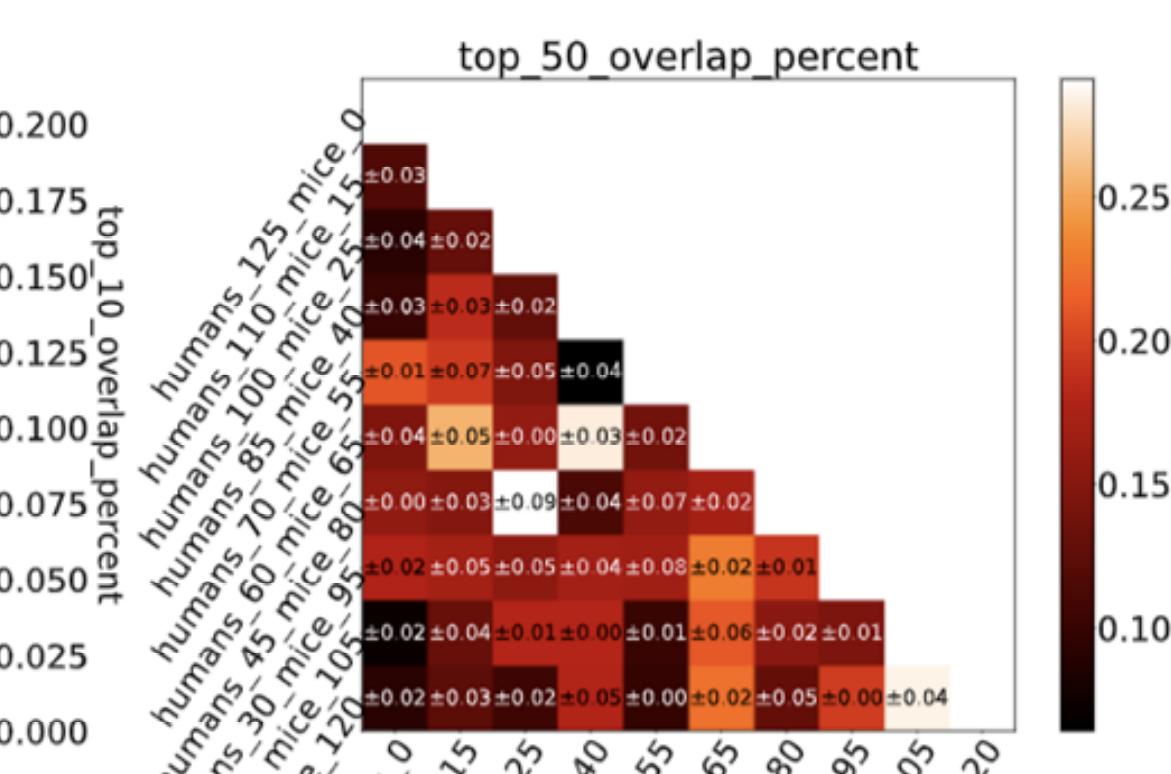
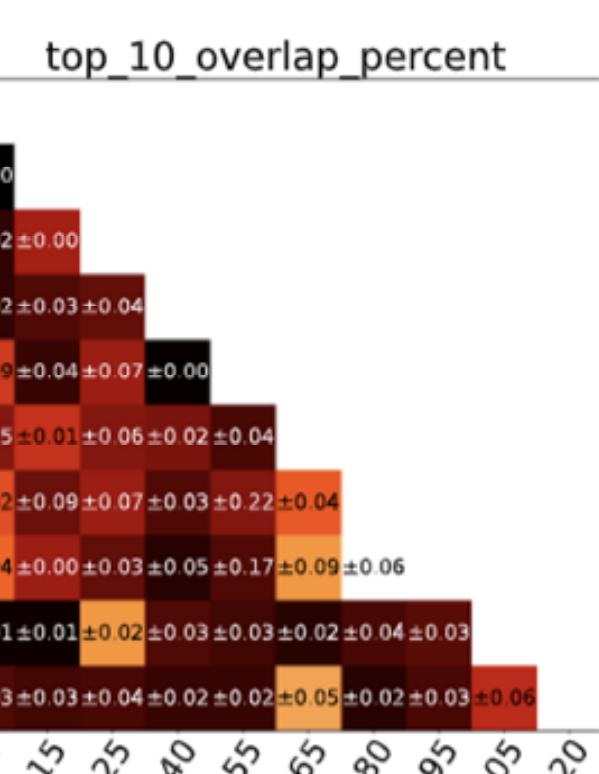
Institute for Simulation INTELLIGENCE

OUTCOME

Acute gamma radiation blood draw (H) or whole blood (M)			Chronic heavy ion radiation lung cell culture (H) or lung tissue (M)		
	Human	Mouse		Human	Mouse
GLDS-152 [6]	50		GLDS-73 [12]	95	
GLDS-157 [7]	75		GLDS-148 [13]		41
GLDS-156 [9]		18	Total	95	41
GSE124612 [10]		248			
GSE62623 [11]		48			
Total	125	314			

1. Open source cross-organism datasets with matched gene-homologues (left):

We provide open-source mouse-human microarray gene expression datasets with matched gene homologues between organisms. Individual datasets sourced from NASA GeneLab data system (GLDS*) or Gene Expression Omnibus (GSE*).

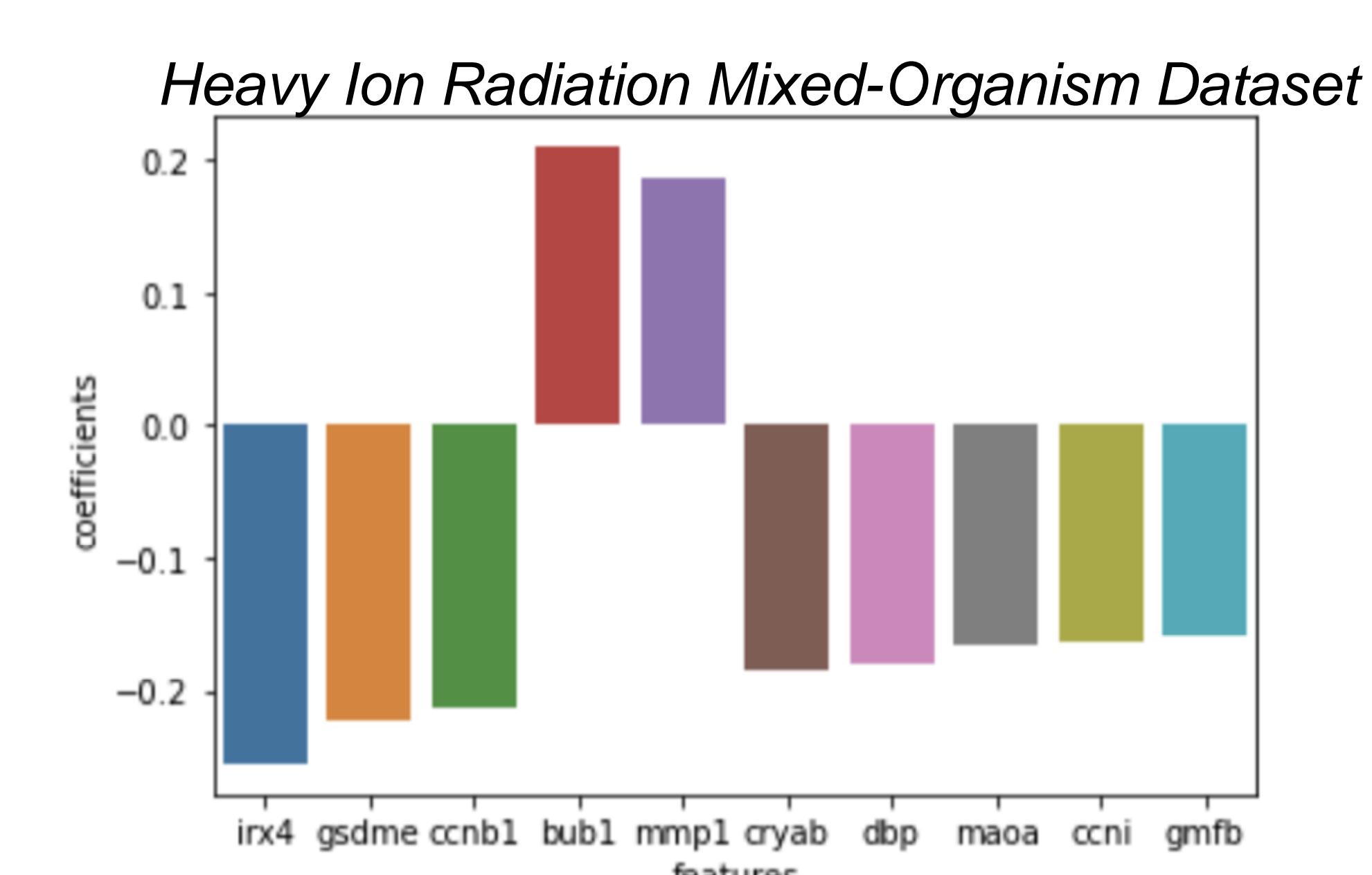
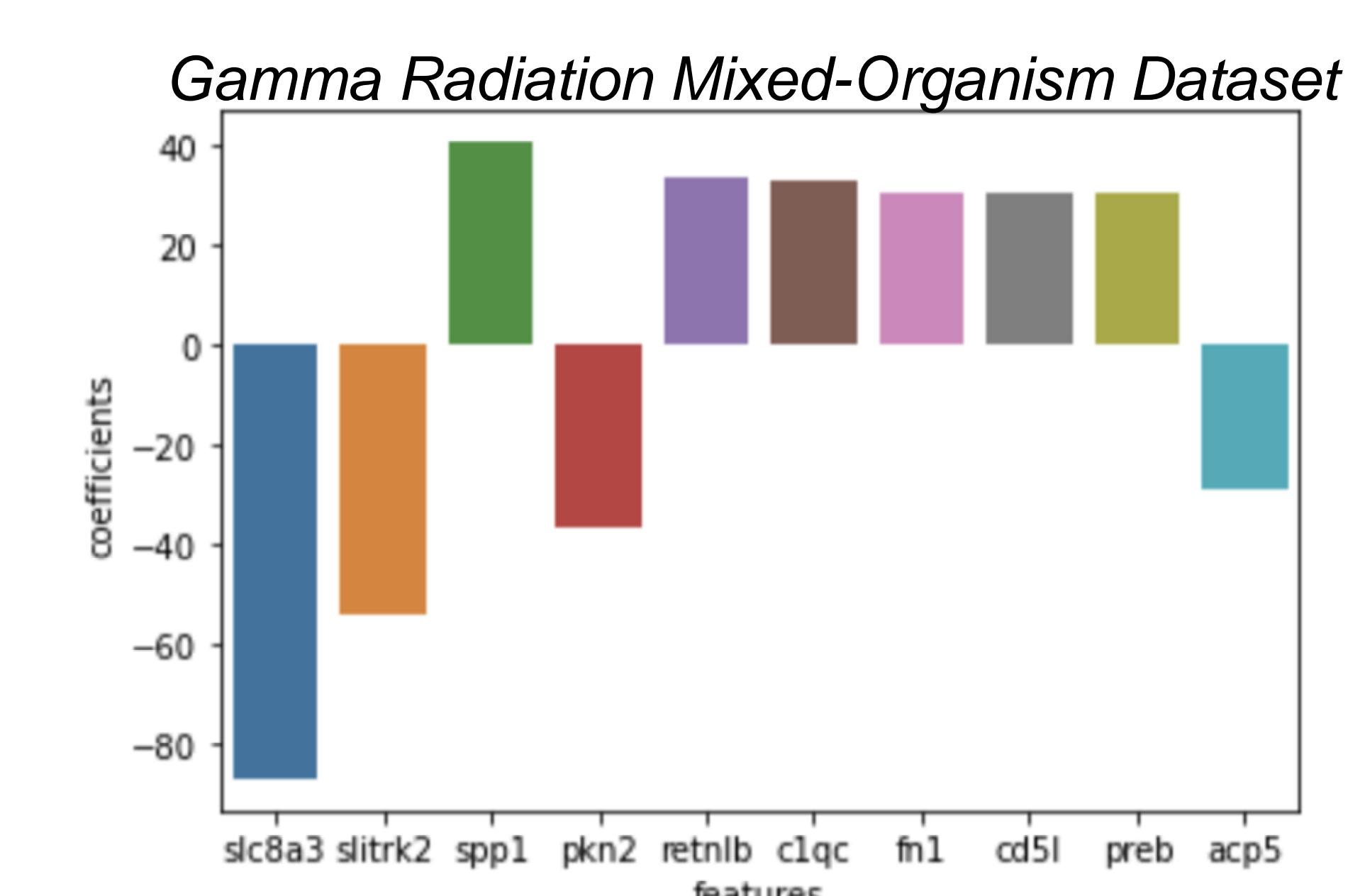


2. Coefficient Overlap (right):

We quantify the coefficient overlap with varying amounts of mouse samples in the cross-organism gamma radiation dataset, using 4 metrics.

- The Ranked-Bias overlap similarity is **consistently >0.5 correlation**.
- At least **10% of the top 50 features overlap** in most comparisons.

3. Biological relevance of top 10 highest ranked genes (below): IRM identifies genes **known to be relevant to cancer pathophysiology** in gamma radiation mixed dataset (*pkn2* [14], *retnlb* [15], *spp1* [16]) and in chronic heavy ion mixed dataset (*irx4* [17], *gsdme* [18], *ccnb1* [19], *bub1* [20], *mpp1* [21]).



NEXT STEPS

- Evaluate IRM domain generalization for causal inference with mixed organism datasets from different data types (omics, histology)
- Develop a comprehensive causal model of radiation exposure effects that translates from mouse models to human astronauts
- Evaluate biological results obtained through federated learning structure via Intel OpenFL (enabled in CRISPv2.0)

Scan for References

