Extraction of High Content Toxicity Biomarkers Data from Scientific Literature

Oleg Stroganov[1], Bing Zhou[1], Tyler Myers[1], James Petts[1], Scott Auerbach[2], Kelly Shipkowski[2], Warren Casey[2]

[1] *Rancho BioSciences, PO Box 7208, Rancho Santa Fe, CA 92067*

[2] *Division of Translational Toxicology, NIEHS, RTP, NC 27709*

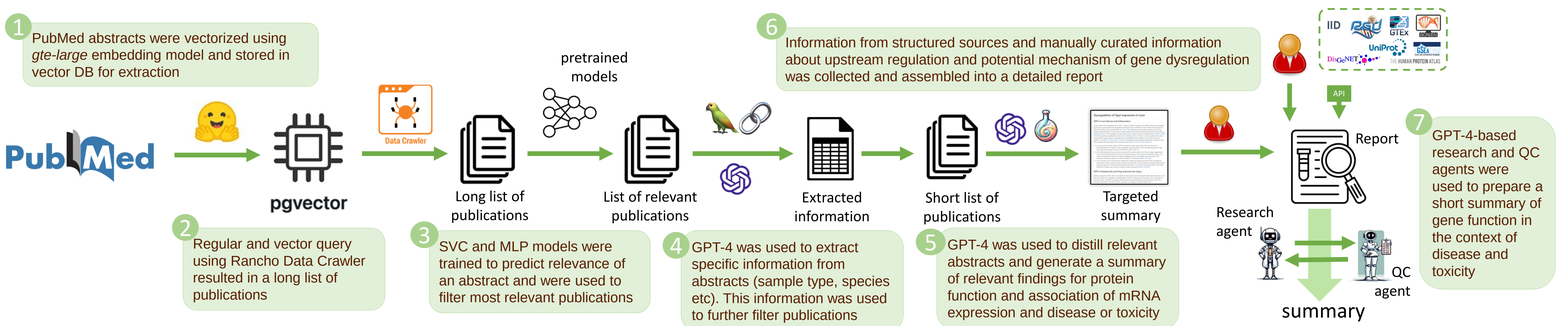
Abstract
Final ID 5016

Background

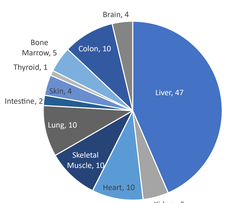
The Division of Translational Toxicology (DTT) at the National Institute of Environmental Health Sciences (NIEHS), NIH, focuses on environmental factors impacting human health, emphasizing toxicity biomarker identification in various organs.

Traditional methodologies, though effective, fall short in providing a comprehensive characterization of toxicological effects via transcriptional changes. This work aims to generate robust summaries of these biomarkers, facilitating their utilization in in vivo tox screens. The summaries will be incorporated into study reports, justifying the use of identified genes for characterizing adversity at the transcriptome level.

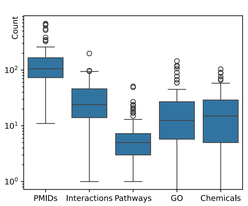
Methods



Results



Number of genes by organs

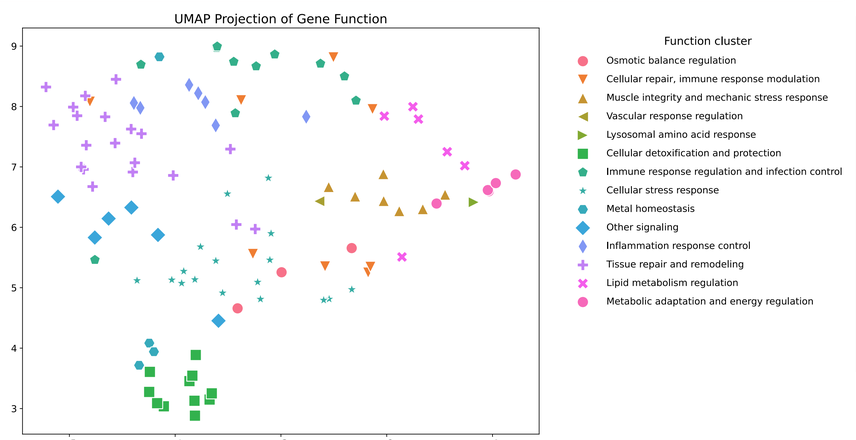


Dataset metrics

GO annotation associated with semantic cluster "Immune response regulation and infection control"	FDR
GO:0006954: inflammatory response	0.00013
GO:0006952: defense response	0.0013
GO:0002376: immune system process	0.0181

Developed methodology enabled focused annotation of >100 genes. Each report has information on association with toxicity and disease on transcriptional level, upstream regulation, description of function and mechanism as well as information on protein-protein interactions, chemicals eliciting response, pathways and other signatures and other data.

Short 2-paragraph summaries of the role of gene in disease and pathology were futher distilled to a single sentence descriptions. Clustering of these functional descriptions in latent embedding space revealed several clusters of genes that have similar role in different tissues, including detoxification, immune modulation and ECM remodeling.



Conclusions

This approach significantly advances the field of toxicogenomics by providing a comprehensive, empirically derived dataset of transcriptional biomarkers. These biomarkers are critical in assessing chemical agent impacts on human health.

The methodology's efficacy in extracting and summarizing large-scale literature data presents a paradigm shift in toxicological research, offering a scalable and precise tool for biomarker discovery and characterization.

NIH National Institute of Environmental Health Sciences
Your Environment. Your Health.

Rancho BioSciences