

GEN3VA: Aggregation and Analysis of Gene Expression Signatures from Related Studies

Gregory W. Gundersen, Kathleen M. Jagodnik, Holly Woodland, Nicholas F. Fernandez, Kevin Sani, Anders B. Dohlman, Peter Man-Un Ung, Caroline D. Monteiro, Avner Schlessinger, Avi Ma'ayan

Department of Pharmacology and Systems Therapeutics, BD2K-LINCS Data Coordination and Integration Center (DCIC) - Icahn School of Medicine at Mount Sinai, Fluid Physics and Transport Processes Branch - NASA Glenn Research Center, Center for Space Medicine - Baylor College of Medicine and Daylesford, the Fairway, Weybridge, Surrey, KT13 0RZ, United Kingdom

Genome-wide gene expression profiling of mammalian cells is becoming a staple of many published biomedical and biological research studies. Such data is deposited into data repositories such as the Gene Expression Omnibus (GEO) for potential reuse. However, these databases currently do not provide simple strategies to systematically analyze collections of related studies. Here we present GENE Expression and Enrichment Vector Analyzer (GEN3VA), a web-based system that enables the integrative analysis of aggregated collections of tagged gene expression signatures identified and extracted from GEO. Each tagged collection of signatures is presented in a report that consists of heatmaps of the differentially expressed genes; principal component analysis of all signatures; enrichment analysis with several gene set libraries across all signatures, which we term *enrichment vector analysis*; and global mapping of small molecules that are predicted to reverse or mimic each signature in the aggregate. We demonstrate how GEN3VA can be used to identify common molecular mechanisms of aging by analyzing tagged signatures from 244 studies that compared young vs. old tissues in mammalian systems. In a second case study, we collected 86 signatures from treatment of human cells with dexamethasone, a glucocorticoid receptor (GR) agonist. Our analysis confirms consensus GR target genes and predicts potential drug mimickers. GEN3VA can be used to identify, aggregate, and analyze themed collections of gene expression signatures from diverse but related studies. Such integrative analyses can be used to address concerns about data reproducibility, confirm results across labs, and discover new collective knowledge by data reuse. GEN3VA is an open-source web-based system that is freely available at: <http://amp.pharm.mssm.edu/gen3va>.