

New year, new gEAR: Better performance, new features, and more data

Joshua D Orvis¹, Brian Gottfried¹, Dustin J Olley¹, Jayaram Kancherla⁶, Beatrice Milton⁴, Yang Song¹, Hector Corrada-Bravo⁶, Amiel Dror², Michael Kelly³, Anup A. Mahurkar¹, Ronna Hertzano^{1,3,4} Affiliations: ¹Institute for Genome Sciences, University of Maryland School of Medicine; ²Department of Otolaryngology, Galilee Medical Center, Nahariya, Israel; ³National Institutes of Health, NIDCD, Bethesda, MD; ⁴Departments of Otorhinolaryngology Head & Neck Surgery and ⁵Anatomy and Neurobiology, University of Maryland School of Medicine; ⁶University of Maryland Institute for Advanced Computer Studies

The gEAR portal (gene Expression Analysis Resource) is an online tool for multi-omic and multi-species data visualization, sharing, and analysis. Originally designed for auditory and vestibular researches, the gEAR portal has now been expanded for general use. The gEAR is unique in its ability to allow users to upload, view and analyze their own data in the context of previously published datasets, as well as confidentially share their data with collaborators prior to publication. It is also unique in combining not only multiple species but multiple data types including bulk RNA-seq, sorted cell RNA-seq, single cell RNA-seq (scRNA-seq) and epigenomics in a one page, user-friendly, browseable format. Individual expression datasets can be displayed in a variety of ways alongside each other, including interactive bar, line or violin plots, and colorized anatomical SVGs.

Most recently, scRNA-seq has matured as a commonly used technique for measuring gene expression across tissues. To provide researchers access to scRNA-seq data regardless of their programming knowledge, we have integrated a scRNA-seq workbench into the gEAR. The gEAR scRNA-seq workbench provides access to both the raw data of scRNA-seq datasets, as well as to saved expert analyses where cell types have already been assigned – giving researchers rapid insight into gene expression of their cell type of interest.

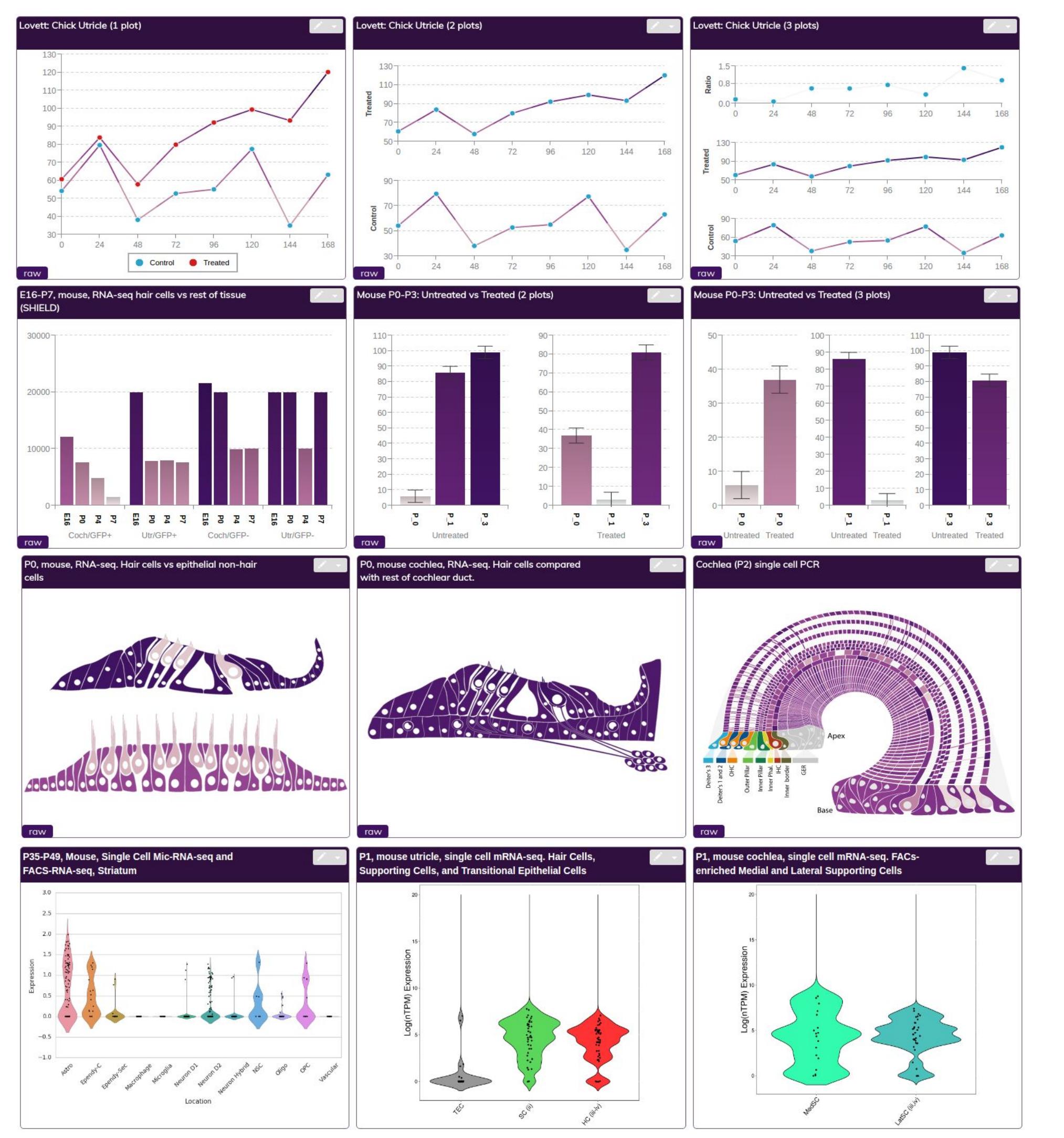
Recent integration with EpiViz allows evidence tracks based on genomic coordinates like epigenetic data (methylation sites, ChIP-seq, ATAC-seq, etc.)

The gEAR portal is publicly available and can be accessed through umgear.org

Data Visualization Types

Currently data can be visualized in line and bar graphs, scalable vector graphics (SVG), and violin plots. Each visualization type is rendered on-the-fly to produce near instantaneous results.

Line and bar graphs can be split into 2 or 3 subplots. Each can include standard deviation bars. Titles can also be added for each subplot.



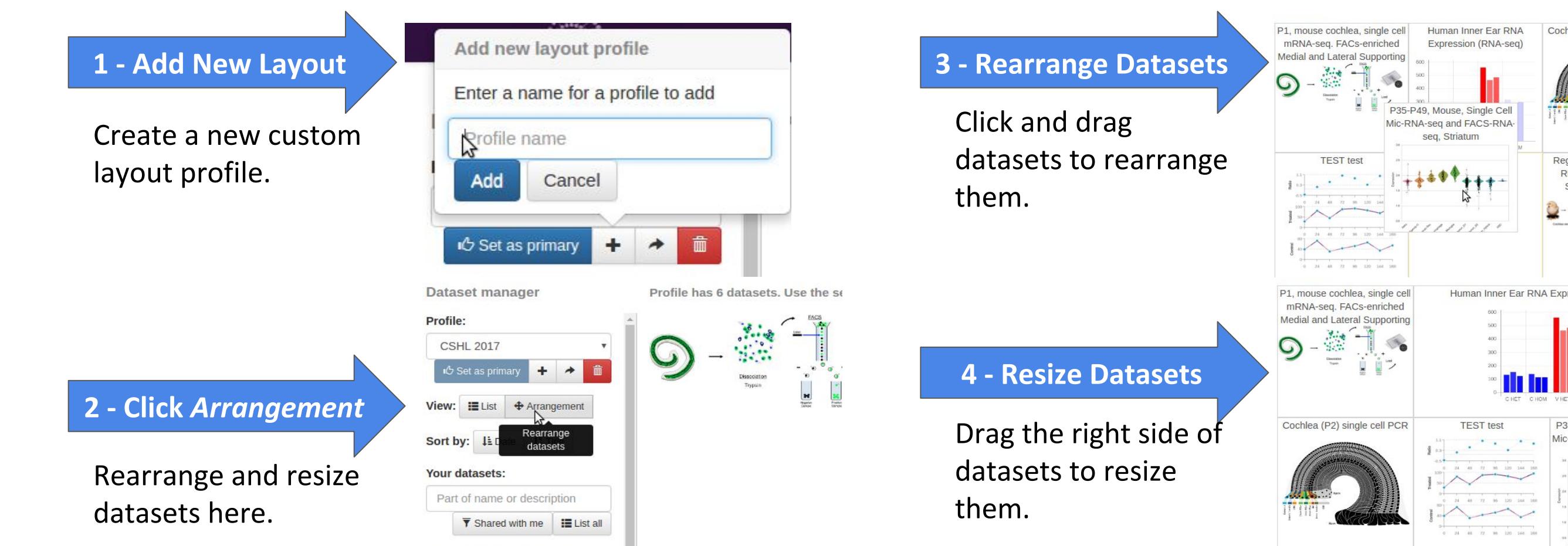
Data columns from the expression file are assigned to <path> elements to link expression values to regions of the SVG.

Violin or tSNE plots generally represent single-cell datasets.

The latest update includes the ability to display user-generated analysis components on the front page. After using the Single-Cell Workbench (see poster center) to generate an analysis, the output of some individual steps can be used on the front page to interrogate individual genes. The first of these are tSNE/UMAP plots with expression levels of the selected gene colored.

Customized Dataset Layouts

Custom layouts let you decide what datasets to search, and arrange your layout profile how you want it. The order and width of the datasets can be changed in a custom layout.

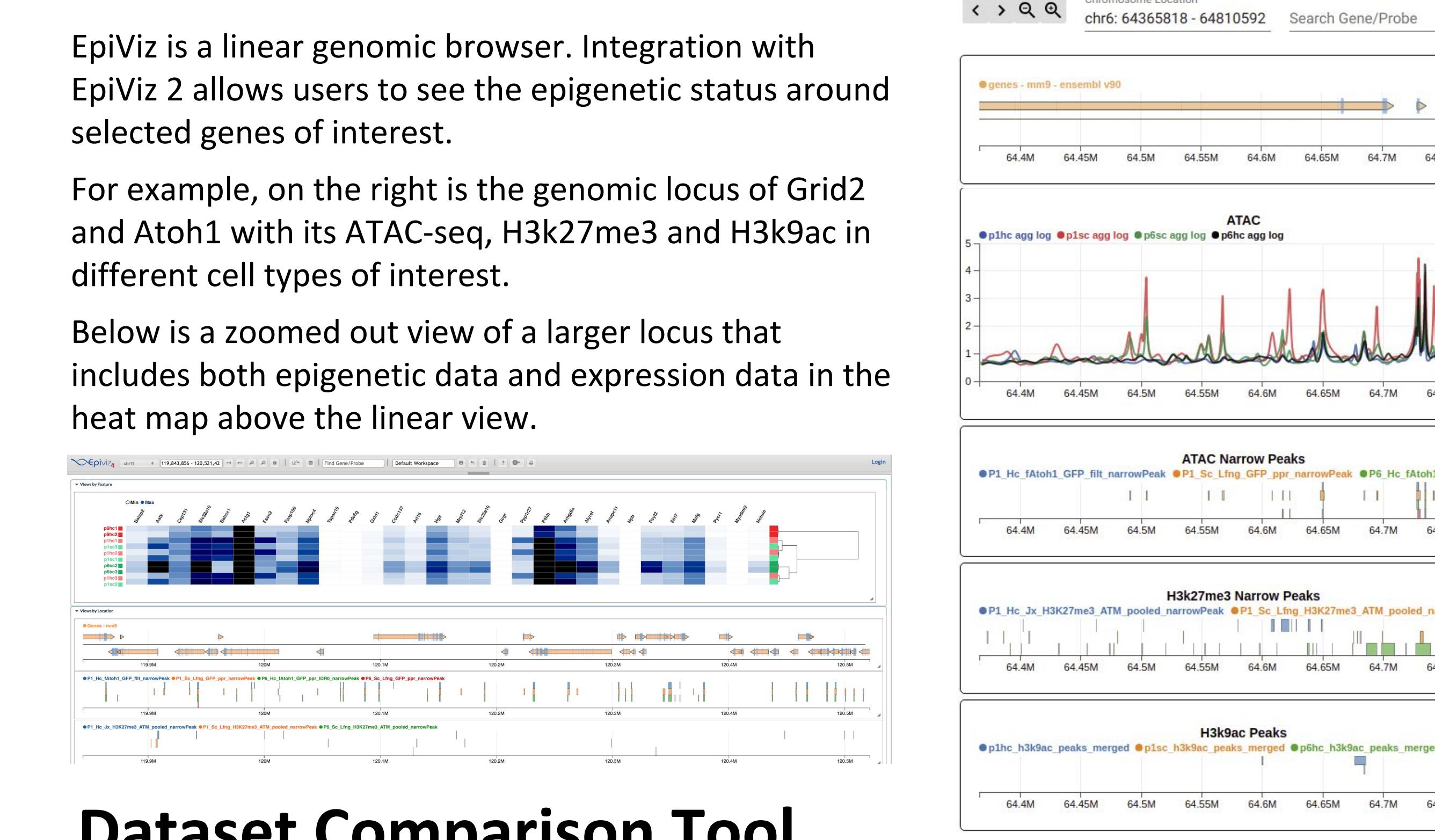


Integration with EpiViz 2

EpiViz is a linear genomic browser. Integration with EpiViz 2 allows users to see the epigenetic status around selected genes of interest.

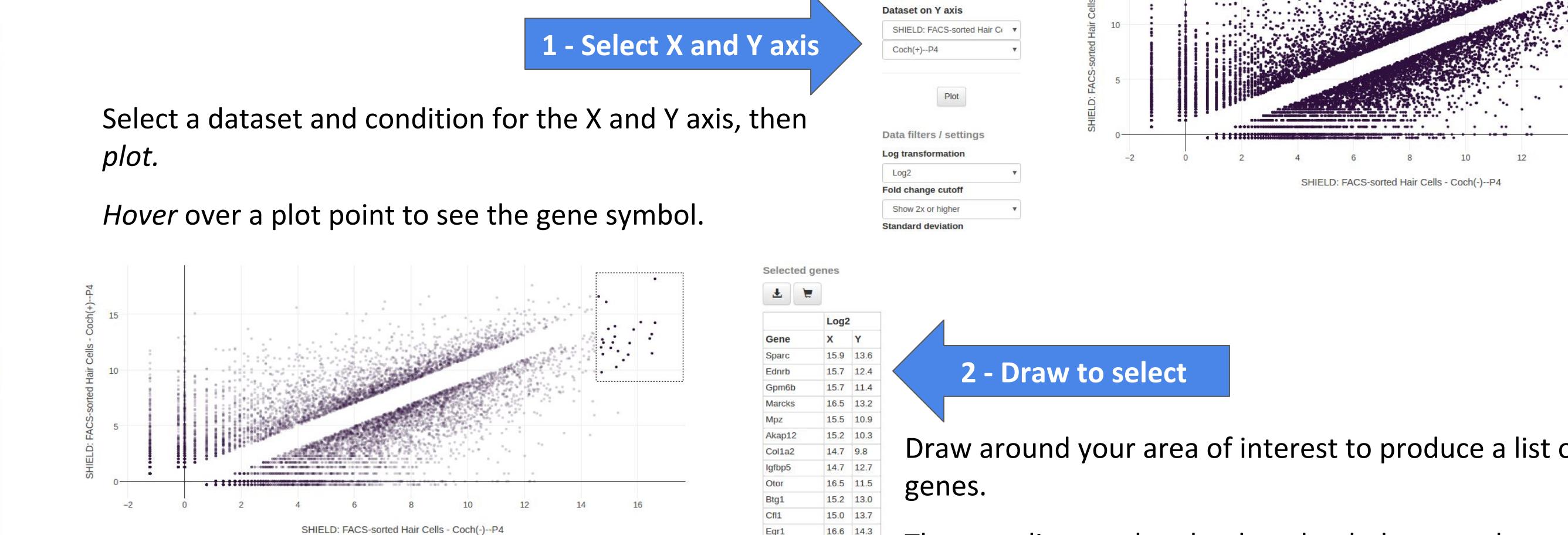
For example, on the right is the genomic locus of Grid2 and Atoh1 with its ATAC-seq, H3k27me3 and H3k9ac in different cell types of interest.

Below is a zoomed out view of a larger locus that includes both epigenetic data and expression data in the heat map above the linear view.



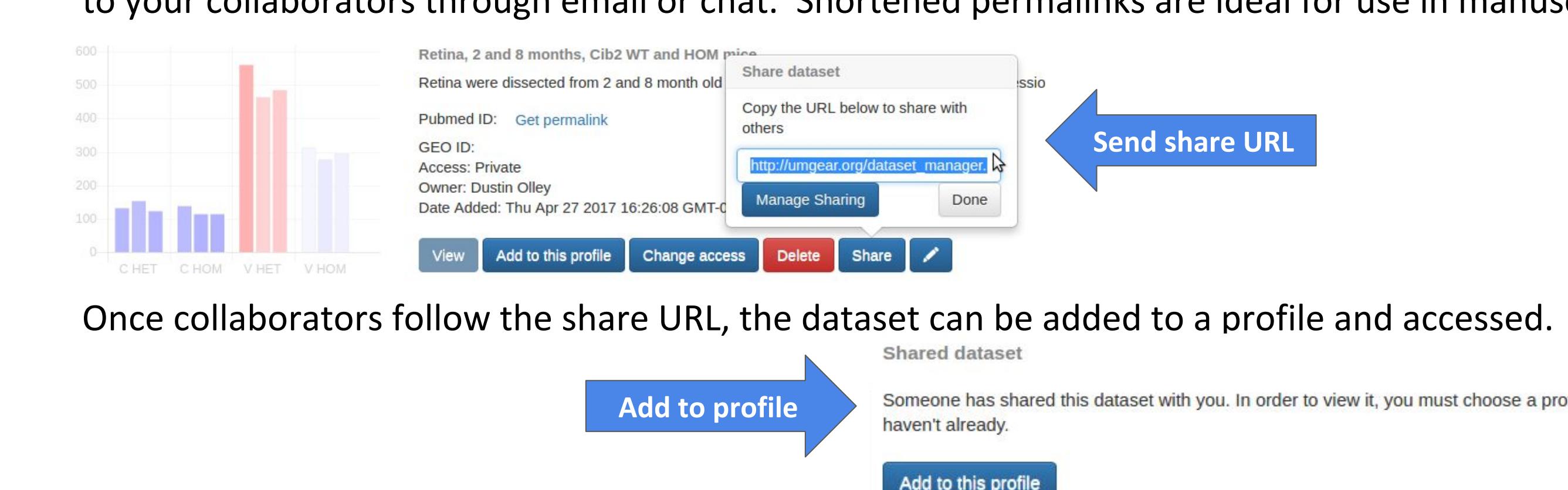
Dataset Comparison Tool

Compare conditions within a dataset or between datasets.



Collaborate Via Sharing

Sharing datasets enhances collaboration. You can share your dataset by sending a dataset unique URL to your collaborators through email or chat. Shortened permalinks are ideal for use in manuscripts.



The Single Cell Analysis Workbench

The Single Cell Analysis Workbench is a tool under active development. This tool currently implements much of the Seurat pipeline in a graphical user-interface and in an implementation that allows faster data processing and visualization than standard R-based pipelines.

The analysis workbench is accessible from the dropdown menu

After choosing the dataset (1) and obtaining the 'dataset structure' (2) work through the steps of the analysis using the 'toggle buttons'

A quality control step follows (3). This filters cells based on their mitochondrial content.

To perform the downstream principal component and tSNE analysis first identify a group of **highly variable genes** (4). For efficiency, the workbench limits users to a maximum of 2,000 genes.

The PCA (5) outlines the number of principle components identified and allows users to examine the expression of genes in the first and second PCs. Coming soon – Heat Maps for the genes contributing to each PC.

Generate a tSNE plot (6) by selecting the number of PCs to incorporate. You may check expression of genes of interest by typing their symbols in the search box.

Following the Louvain clustering (7) **marker genes** can be identified (8) and downloaded as a table to then use as input to search other datasets for validation. For example – population '6' (top gene Otor) is clearly mesenchymal cells.

Finally, you can compare genes/clusters (10) by producing a bidirectional comparison of cluster pairs to visualize the genes which best differentiate between them.

After you find marker genes (8) click cells in the marker gene table to select them for visualization (9), generating both stacked violin and dot plots to visualize expression in each cluster.

Your Data in Context of Others

Compare and query your data along beside other data. No other resource allows you to dynamically compare your data across datasets and organisms on-the-fly.



Customizable dataset views

Multiple views of a single dataset can now be created using the Dataset Curator by dataset owners and users, allowing full customization of what and how data are shown. You can also toggle between available views on the front page on the fly.

The curator allows you to create as many of these views as you like, including drag/drop property ordering, X/Y axis selection, color gradients, and more.

