**SPARC portal transcriptomic and genetic data visualization workflow in o²S²PARC—the GO project**

**Table of Contents**

* **About NIH SPARC**
* **FAIR Policy**
* **Our goal**
* **Transcriptomic data on SPARC Portal**
* **Why do we need visualization tools for gene expression and transcriptome data?**
* **What can we do?**
* **How to use the pipe?**
* **How does it work?**
* **How to utilize so the dataset -An example: multiple sclerosis (MS) case studies**
* **Contributors**
* **License**

**About NIH** **SPARC**

The SPARC Portal is an NIH-funded, open-access data management and sharing science platform, belonging to the Stimulating Peripheral Activity to Relieve Conditions (SPARC) program. The portal aims to advance clinical and academic understanding and interventions in the peripheral nervous system through data integration, knowledge, computational modeling, and spatial mapping. Please refer to the [SPARC website](https://sparc.science/about) for more information.

**FAIR Policy**

The SPARC consortium mandates a data sharing policy called the FAIR data sharing principle, which is “Findable, Accessible, Interoperable and Reusable”. This requires all the contributors and projects published on SPARC to be uniformly adapted to the platform data management and sharing agreement, which can be found [here](https://commonfund.nih.gov/sites/default/files/SPARC_material%20sharing%20policy%2026jan17_508.pdf).

**Our Goal**

At present, the platform has no data processing or visualization system for transcriptomic and genetic data analysis purposes. In fact, there are multiple transcriptomic datasets available on SPARC data browser, as well as the gene expression data from oSPARC template. These datasets would be very much useful for researchers and clinicians in uncovering the underlying Gene Ontology(GO) **(**molecular functions, cellular components, and biological processes) related to the peripheral nervous system and related diseases. **With Gene Ontology(GO) tools,** **we could uncover the myth of any interested pathophysiological proces****ses with gene expression and transcriptome data in SPARC portal.** This will help promote the discovery of potential targets for future interventions and treatments on peripheral nervous system and related diseases.

**Transcriptomic data on SPARC Portal**

Till Aug 8, 2022, there are 19 projects available on the SPARC Portal, containing the species of humans, pigs, mice, and rats. Involved anatomical structures include cervical ganglion neurons, celiac ganglion neurons, stellate ganglion neurons, right atrial ganglionic plexus (RAGP) neurons, enteric nervous system, nodose ganglia, sympathetic stellate ganglia, intrinsic cardiac nervous system, interscapular brown adipose tissue (iBAT)- related ganglia, and inguinal white fat (iWAT)-related ganglia. Analysis methods include RNA sequencing, real-time PCR (quantitative PCR or qPCR), small molecule FISH (RNAScope) probes, and gene ontology analysis. All datasets and metadata files are available for download.

**Why do we need visualization tools for gene expression and** **transcriptome data?**

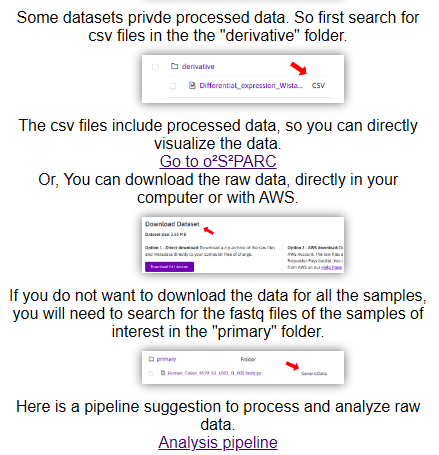
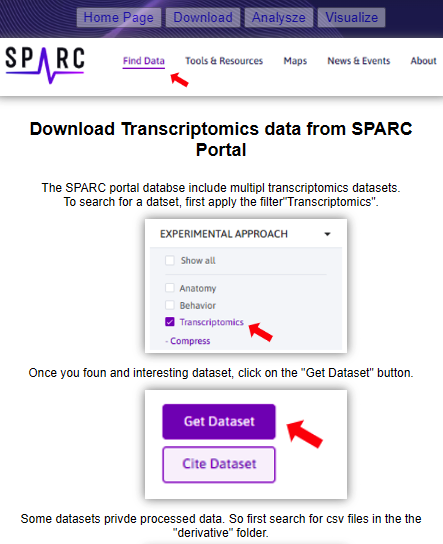
The present database provides a wide variety of species, organs, and datatypes differentiated transcriptome and gene expression information pools, which has a great potential for discovering new pathways and molecular. However, the separated and intricate information from transcriptomic and genetic data made it hard for the researchers and clinicians to get an impression of which pathways and molecular matter most in the disease and functions, which are the targets of the following interventions. Thus, we utilized visualization tools for gene expression and transcriptome data by clustering group genes or samples which share similar patterns of gene expression and transcriptomic profiles. This will expedite the discovery cycle in the **SPARC** program and attract more scientists and clinicians to contribute to the **SPARC**.

**What can we do?**

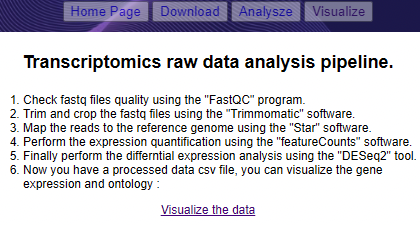
We developed a chrome extension that provides a guide for downloading and analyzing transcriptomic data available on the SPARC portal and a link to access the oSPARC template. Besides, we built gene expression visualization functions through the oSPARC template. The users can import processed csv files to see the gene expression graphs. The audience can visualize an independent dataset by generating a volcano plot, tables, and ontology graphs, or they can compare different datasets by generating Venn diagrams and tables.

**How to use the pipe?**

1. **Open the Chrome extension, and download transcriptomics data from the SPARC portal**
2. **Click on the download button, and get a guide on how to download your data.**



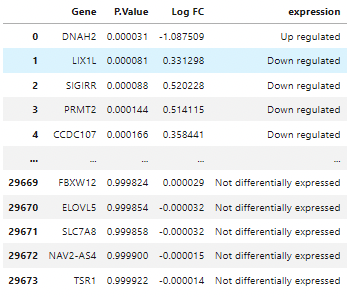
1. Click on the analysis button, and you’ll get a suggested pipeline to analyze the raw data.



1. Click on the visualize button, and it will provide a tutorial to use the [OSPARC](https://osparc.io/) study.

**How does it work?**

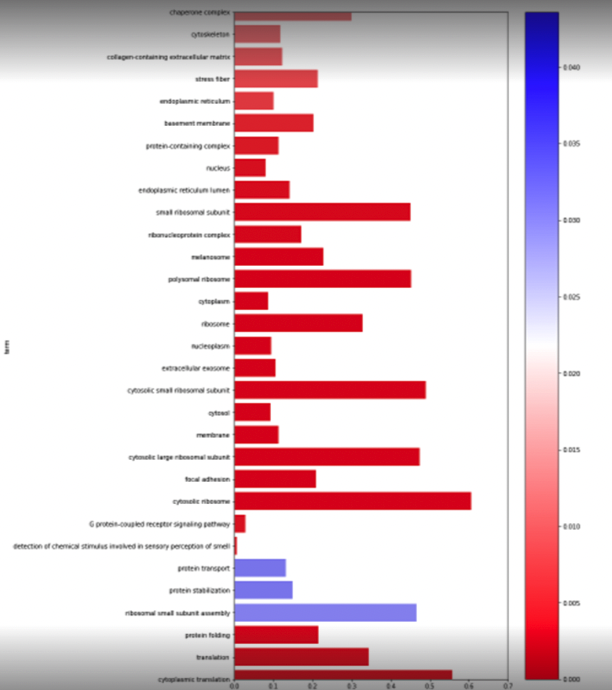
Here is what the OSPARC provides: you import your processed transcriptomics data, and the pipeline will identify the expression level of each gene, this will tell if the gene is upregulated, downregulated, or there is no significant difference.



Then, it will present the data in a volcano plot format as follows:



The top ontologies are represented in graphs:



By doing this, you will get six ontological graphs: three for upregulated genes and three for downregulated genes, one for each category: cellular components, biological processes, and molecular functions.

The Gene Ontology overview could be found [here](http://geneontology.org/docs/ontology-documentation/), and it consists of three components:

* The cellular component(CC) ontological graphs: Describes subcellular structures and macromolecular complexes. We often use CC to annotate cellular locations of gene products.
* The biological process(BP) ontological graphs: Describes the biological programs consisting of multiple molecular activities, such as DNA repair or signal transduction.
* The molecular function(MF) ontological graphs: Describes molecular-level activities performed by gene products, such as “catalysis” or “transport”.

**With these six graphs, we could uncover the myth of any interested pathophysiological processes with gene expression and transcriptome data using GO tools in SPARC portal.**

**How to utilize so the dataset -An example: multiple sclerosis (MS) case studies**

We use normal-appearing brain tissues from multiple sclerosis patients and healthy donors, and we want to compare the transcriptomics profiles between the two groups: MS vs healthy controls.

To determine the genes associated with the first stages of the diseases, and thanks to the pipeline results, we could determine the pathways that are upregulated and downregulated in MS compared with the controls.

Besides, we can also compare two independent datasets. For example, if we have the expression of a common gene seens in the two datasets, the MS dataset and another demyelinated dataset such as Guillain-Barre Syndrome (GBS), we can also analyze data from a demyelinating lesion perspective. As well as the controls.

This will allow us to know which genes expression leads to demyelination and which gene absence leads to the same thing.

**Contributors**

Hiba Benaribi (Leader, extension, and the oSPARC template)

Anmol Kiran (extension, and the oSPARC template)

David Nickerson (SysAdmin)

Hiba Benaribi, Mengyuan Ding (Writer - Documentation)

**License**

This project is fully Open Source and all generated graphs could be used in articles directly. All above is developed under the MIT License.