

Purpose

Obtain metadata from labs, data set from CESC Centers

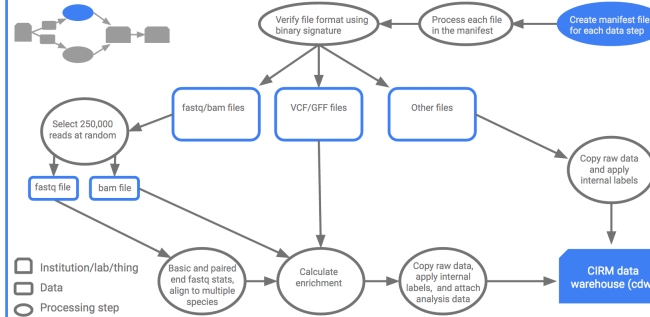
Unify data and metadata for internal consistency, perform validation and analysis QA on sequence data/files

Make submissions available to researchers through cdw

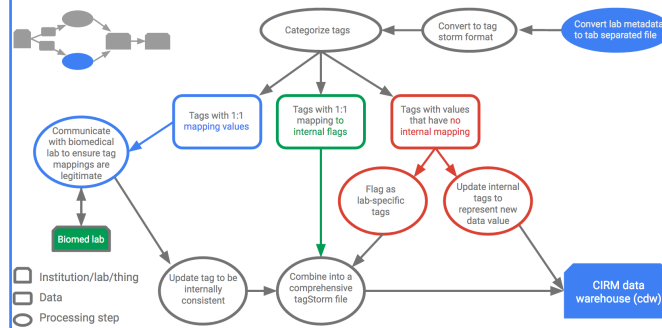
- Submission analysis for individual labs
- Download data from one or more labs (restricted access prior to publication)
- Some data types can be visualized on the genome browser

Data flow: Internal

Data validation: how we process sequence data and other files



Data unification: how we process metadata

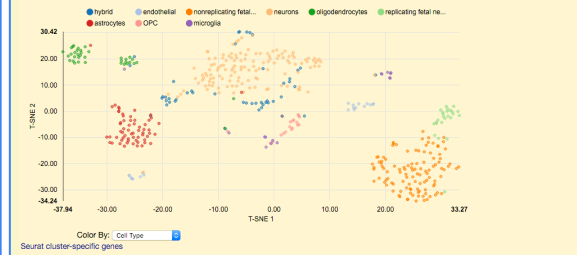


cdw: CIRM Data Warehouse

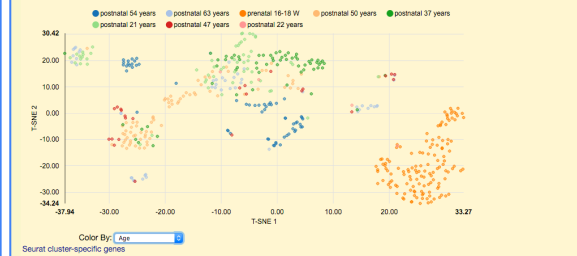
The screenshot shows the CIRM Data Warehouse (cdw) interface. At the top, there are links for 'Help', 'Test', and 'Login'. Below that is the 'Browser V0.49' section, which contains a text description: 'The CIRM Stem Cell Hub contains 1015 GB of data in 5,469 files from 8 labs. You have access to 3,999 files. Try using the browse menu on files or tracks. The query link allows simple SQL-like queries of the metadata.' Below the text are three pie charts showing the proportion of files in each category: 'lab', 'format', and 'assay'. The 'lab' chart shows 'salk' as the largest category. The 'format' chart shows 'bam' as the largest category. The 'assay' chart shows 'single-cell RNA-seq' as the largest category. At the bottom, there is a table with columns for 'id', 'lab', 'format', 'assay', and 'file'. The table contains a list of data entries.

Individual submission analysis

Seurat T-SNE: gene list trimming and non-linear dimensional reduction



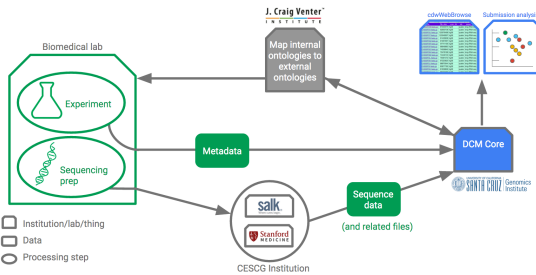
Seurat T-SNE: gene list trimming and non-linear dimensional reduction



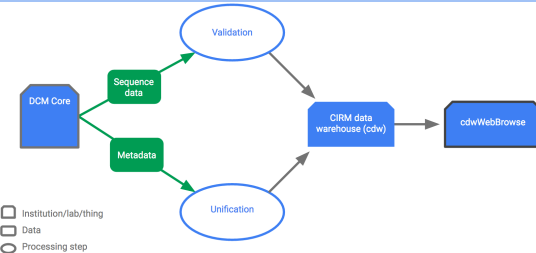
Darmanis S, Sloan SA, Zhang Y, Engle M, Caneda C, Shuer LM, Hayden Gephart MG, Barres BA, Quake SR. (2015). A survey of human brain transcriptome diversity at the single cell level. Proceedings of the National Academy of Sciences. 112(23):7285-90.

Data flow

External data flow: how labs fit in



Internal data flow: how data moves to the data warehouse



Funding and Acknowledgements

We would like to thank the California Institute for Regenerative Medicine for providing grant GC1R-06673-C. DCM staff involved: Josh Stuart, Jim Kent, Chris Eisenhart and Clay Fischer.

