The Stem Cell Commons: An HSCI Resource for Stem Cell Genomics Data and Reproducible Analyses

Shannan Ho Sui[*,1,2], Sudeshna Das[*,3,4], Nils Gehlenborg[*,4], M. Emily Merrill[3], Ilya Sytchev[1], Richard Park[4], Psalm Haseley[4], John Hutchinson[1], Stéphane Corlosquet[3], Terah Hansen[1], Oliver Hofmann[1,2], Peter Park[4], and Winston Hide[1,2]

[*]Authors contributed equally, [1]Biostatistics, Harvard School of Public Health, [2]Harvard Stem Cell Institute,[3]Massachusetts General Hospital, [4]Harvard Medical School, Boston, MA, USA,

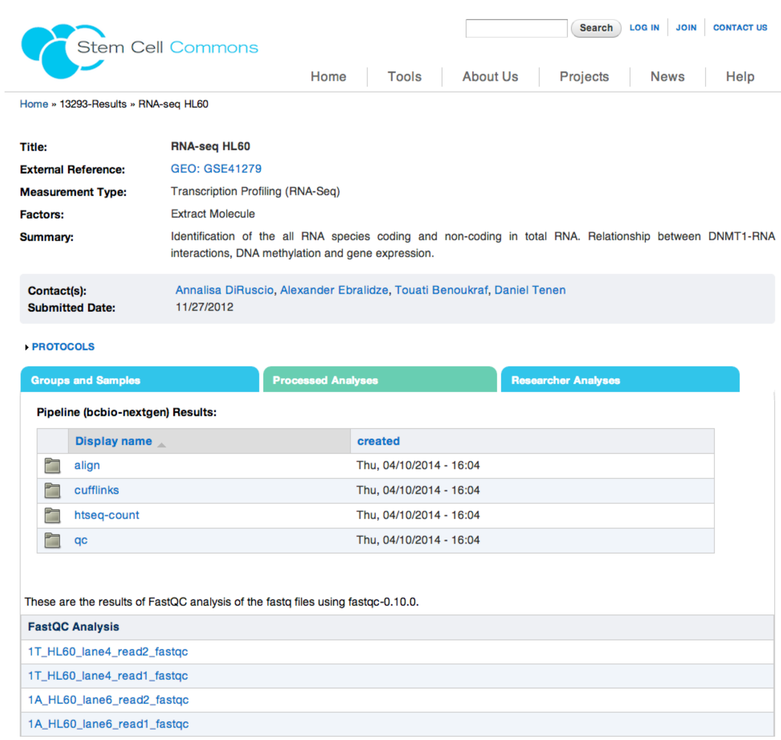
Abstract

The Stem Cell Commons (stemcellcommons.org) has been developed by the HSCI to create a community for stem cell bioinformatics. This open source environment for **sharing, processing and analyzing stem cell data** brings together genomics data sets with tools for curation, searching, dissemination and analysis. Standardizing the experimental metadata and analysis approaches provides us with opportunities to perform data integration and comparison of data sets at the HSCI.

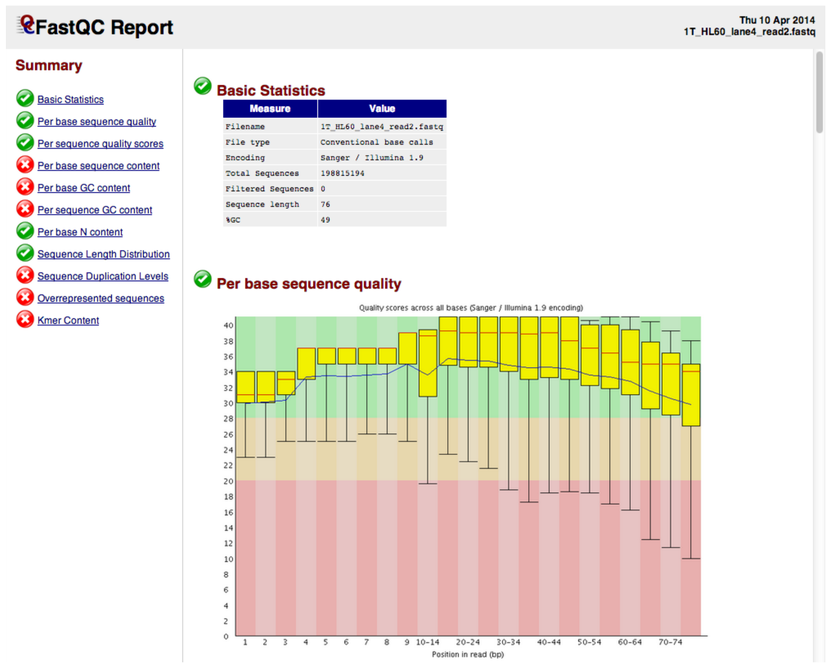
- Data curation and support services
- Query detailed metadata
- Export to standard formats (e.g. GEO)
- Run standardized pipelines
- Secure accounts for private data
- Share data and analyses
- Submit data
- Download raw and processed data

Data Repository

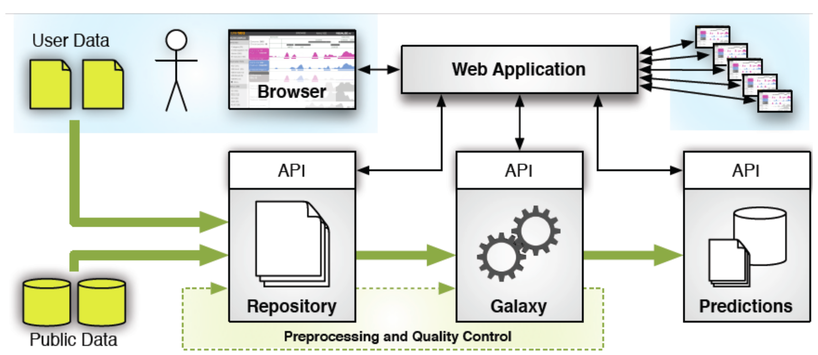
The Stem Cell Commons database, built on the eXframe system, contains data from **20** HSCI laboratories and **202** datasets made up of over **3500** assays. These include microarray and NGS data spanning multiple cell types and tissues from **human**, **mouse**, **rat** and **zebrafish**. We have both **raw data** and **results of standardized pipelines**, Custom analysis results are made available to researchers as well.



Quality control reports for NGS data



Refinery Analysis/Visualization Framework

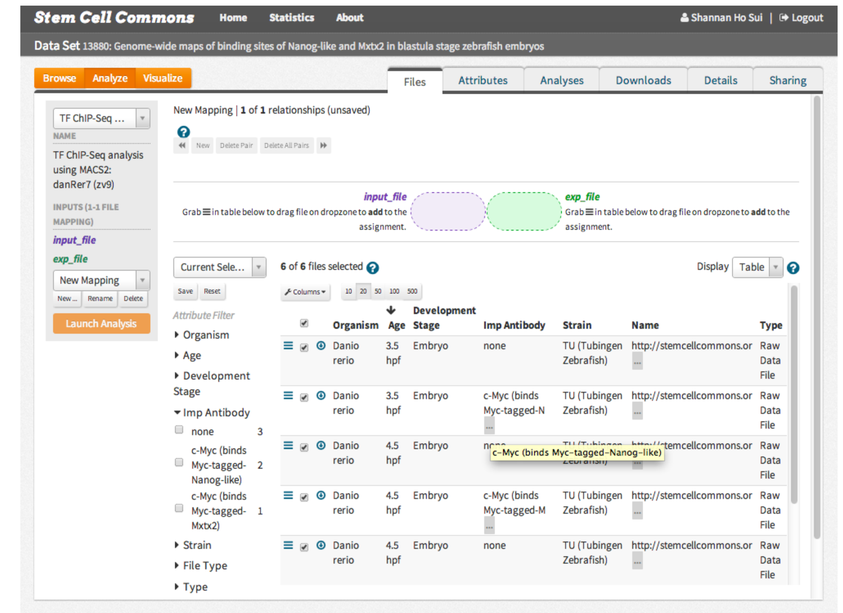


We have linked the Stem Cell Commons data repository with the Refinery analysis and visualization platform (refinery-platform.org), a flexible environment for managing data and running diverse analysis workflows. We **invite new users to try our new ChIP-seq pipeline** and provide feedback on the user interface and other workflows to be added. Refinery will ultimately serve as a platform for development of novel visual exploration tools.

Acknowledgements

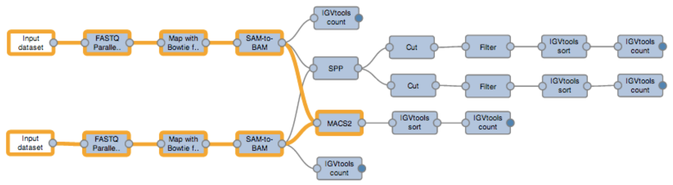
This project was supported by funding from the HSCI and an NIH/NHGRI K99 award to NG. Special thanks to Yi Zhou and Song Yang (Zon Lab) and Annalisa DiRuscio (Tenen Lab) for curation and testing, and to HSCI researchers for submitting their data.

User friendly interface to assign inputs to workflows

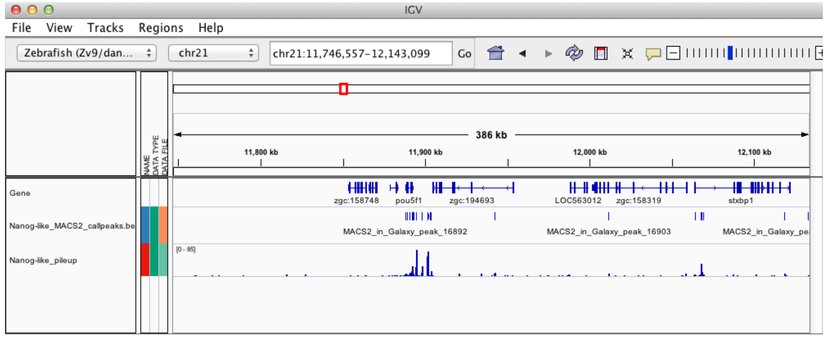


Workflow visualization for data provenance

Collaboration with Marc Streit & Stefan Luger, Johannes Kepler University Linz, Austria



View results in the IGV Genome Browser



Contact Us

We **support new data submissions** to the Stem Cell Commons and invite users to help us test our ChIP-seq analysis pipeline. Email us at hsci_bioinformatics@harvard.edu to add your data and/or get an account for Refinery. We welcome questions, suggestions and comments!