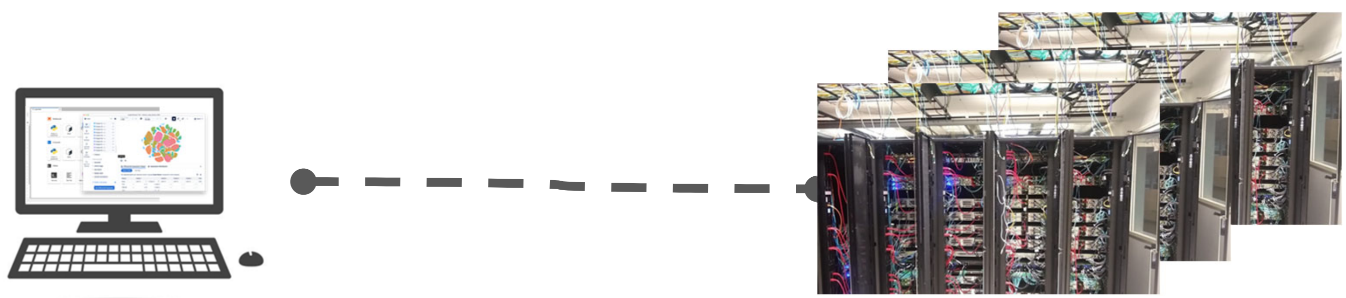
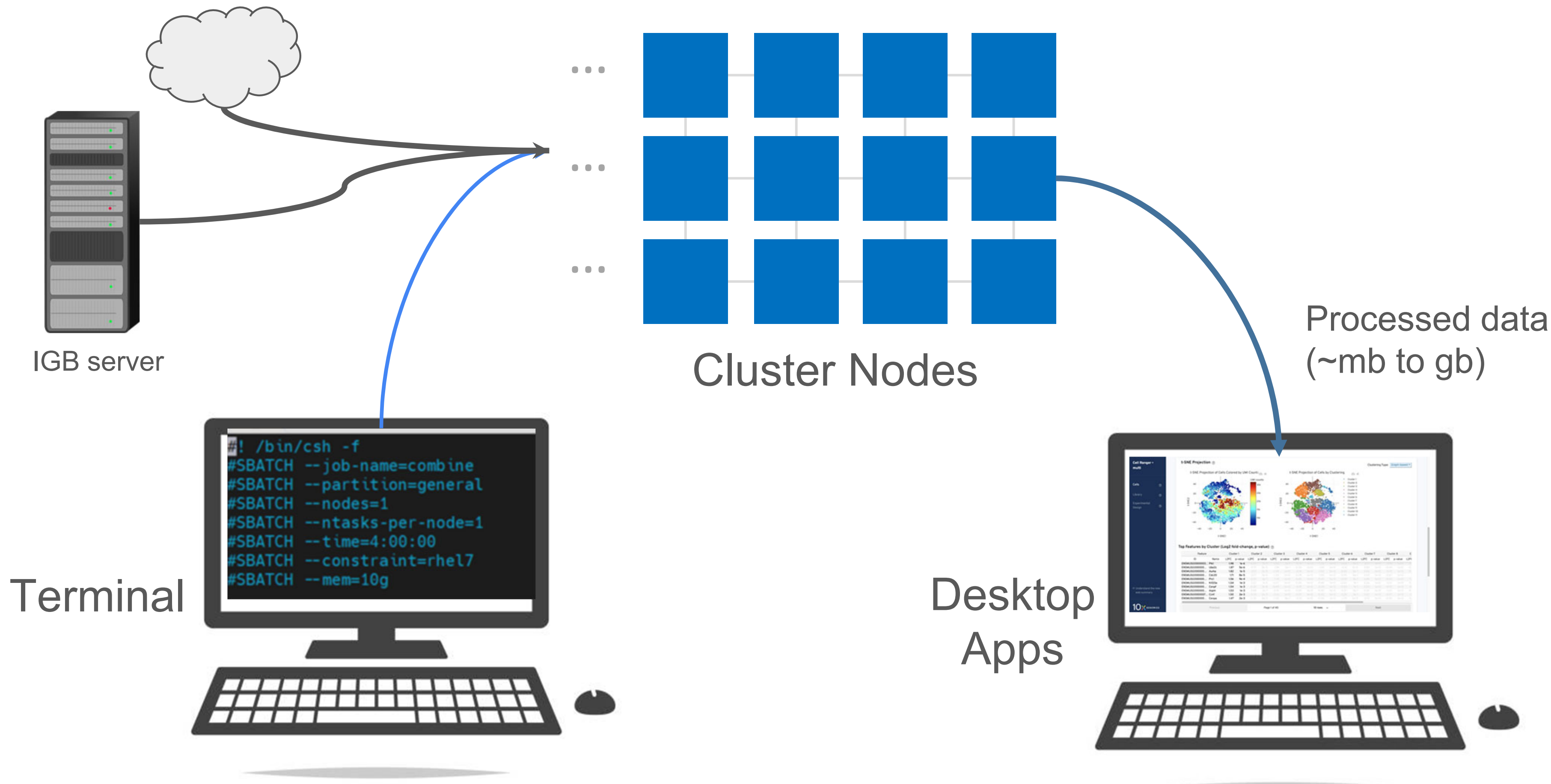
Data Wrangling from your computer to HPC3

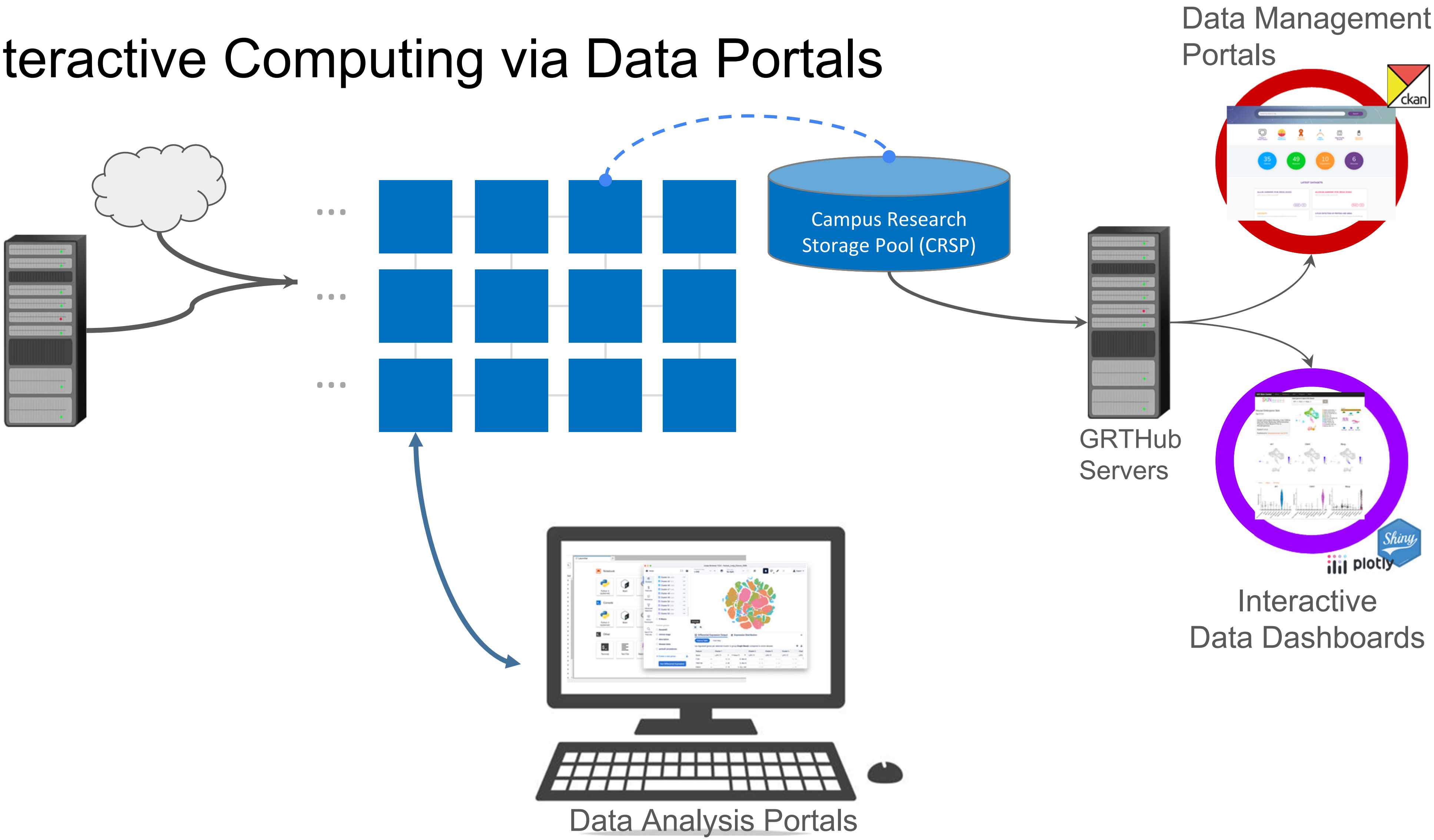
Ivan Chang



Traditional HPC Workflow

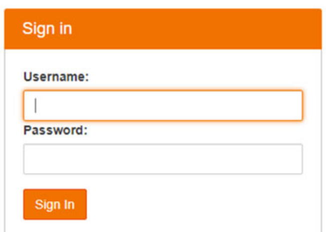


Interactive Computing via Data Portals

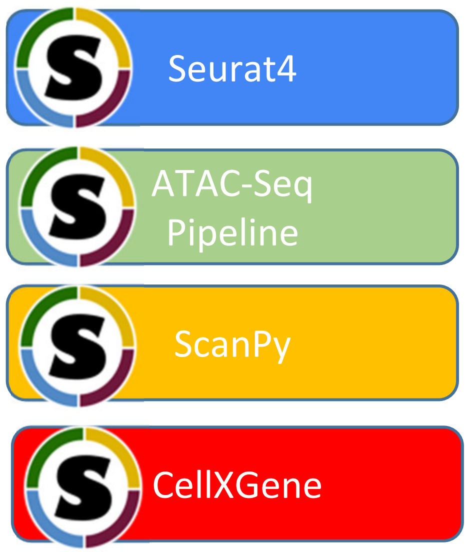


Jupyterhub-based Data Analysis Portals

User Authentication

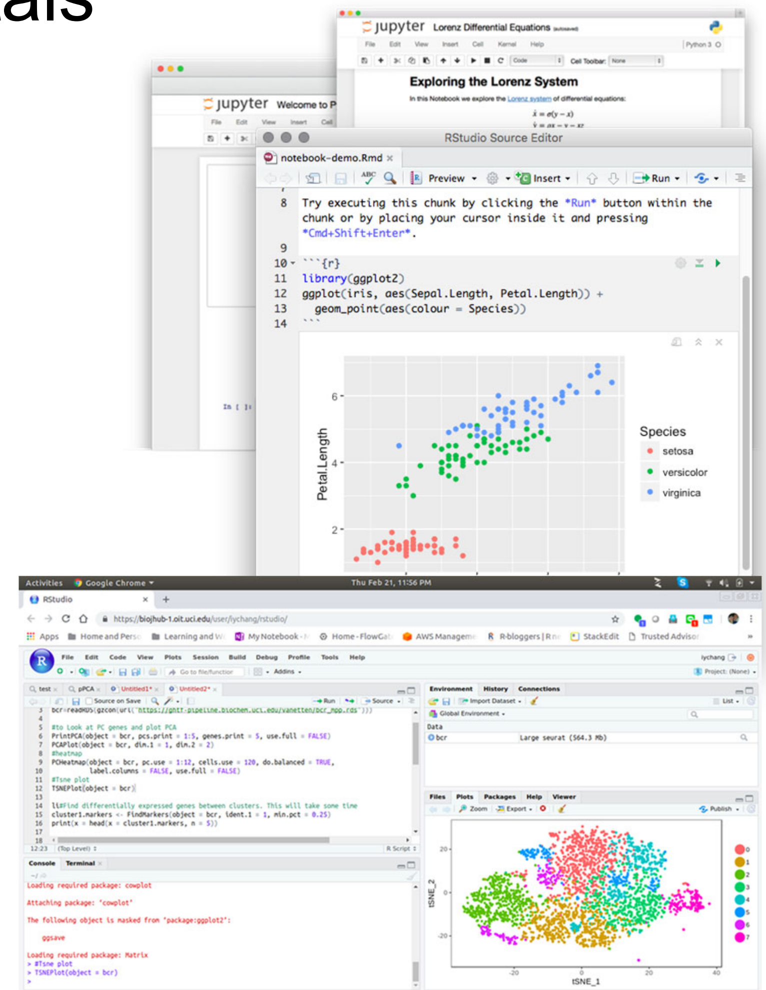


Container Selection



. . .

Computational Notebooks



Two-types of Jupyterhub Data Analysis Portals

RCIC provided computing resources:

Biojhub4 and Physicsjhub

- For all HPC3 users
- Curated analysis pipeline containers
- Access to workshop Jupyter notebooks and bioinformatics tutorials
- Uses free partition or user computing credits
- Limited to 6hrs continuous usage
- Limited to 16-cores and roughly 96gb ram

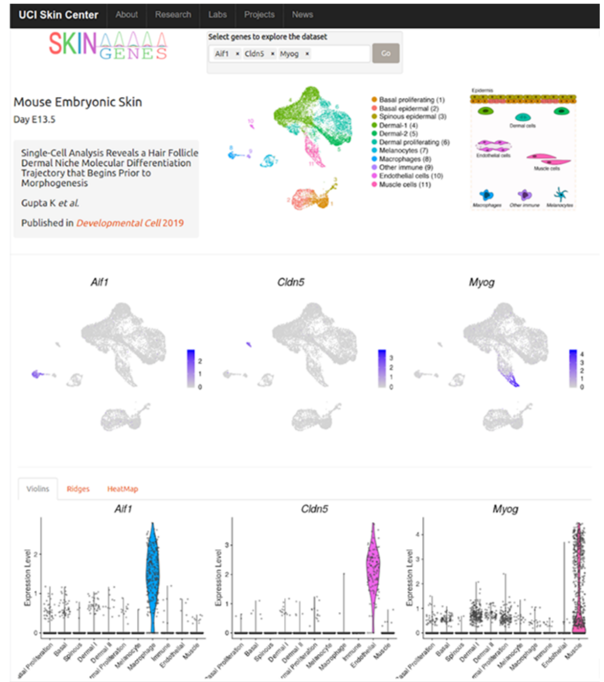
Research Center funded portal nodes

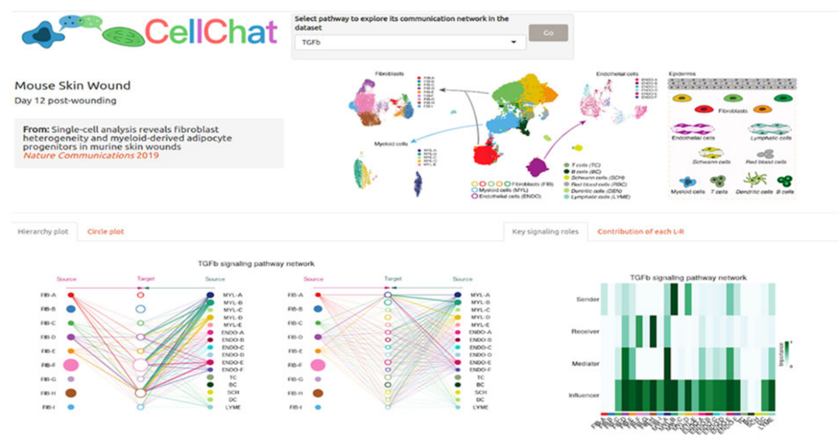
Bioportal-1 (U54/P30) and Bioportal-2 (P01 ALS)

- For center-wide research community
- Shared portal node computing resources (members self-regulated)
- Curated analysis pipeline containers
- Research specific pipeline containers
- **Access to bioinformatics desktop applications** such as Loupe browser, IGV Genome Browser, Cellpose GUI, CLC workbench, Segmentation software, etc.

Interactive Data Dashboards

- Apps to allow general research public to visually explore published datasets.
- GRTHub provides the hardware and the hosting
- Turnkey R Shiny, plotly containers to quickly generate data dashboards for Seurat, SpatialLibd, Scanpy, etc.
- Consulting for designing and hosting customized dashboard sites





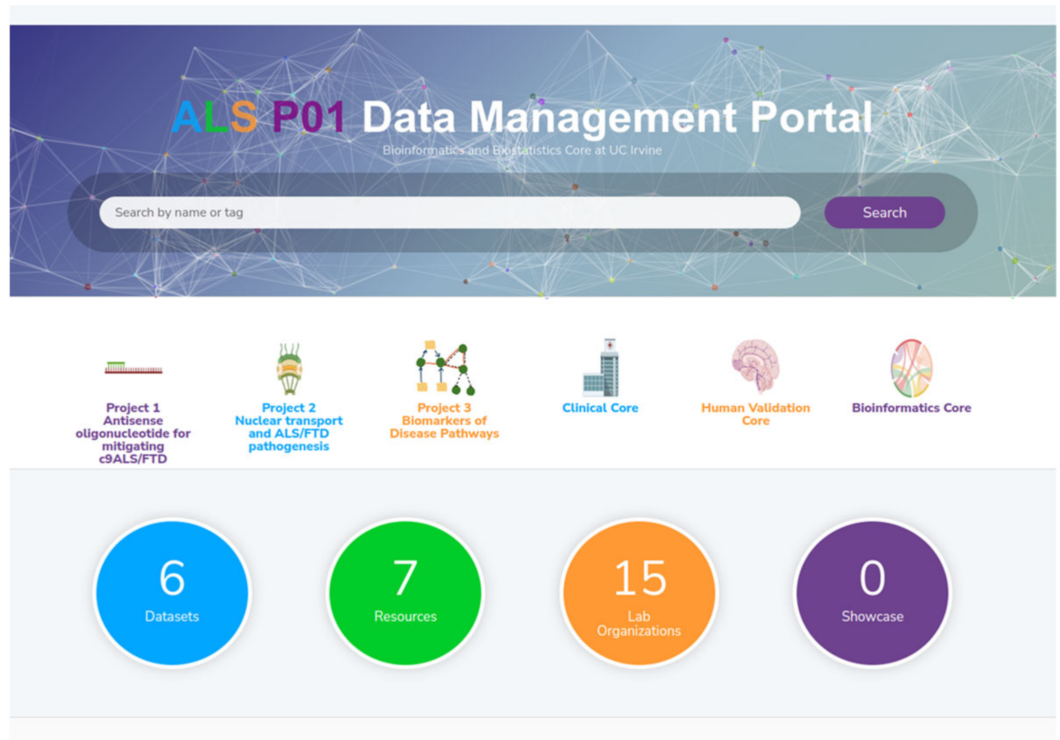
Shiny

plotly

Center-wide Data Management Portal

Support for research centers to create a public accessible and searchable data portal

- Hosted on GRTHub servers
- Customizable look and feel
- Data schema designed for the center's organization
- Control over the accessibility of dataset
- Interactive views of data tables
- Data retrieved from UCI CRSP via a mirrored local copy.



Coming soon…

- Integrating Cromwell/WDL pipeline workflow engine into Jupyterhubs
- More analysis pipelines
  - iScan Admixture
  - DeepVariant
  - Cellpose

Special thanks to the RCIC and GRTHub bioinformatics teams!

- **Philip Papadopoulos**
- **Francisco Lopez**
- **Nick Santucci**
- **Joulien Tatar**
- **Imam Toufique**
- **Nadya Williams**

- **Jenny Wu**
- **Christina Lin**