

Feedback and Tutorial Session



Mikko Nieminen

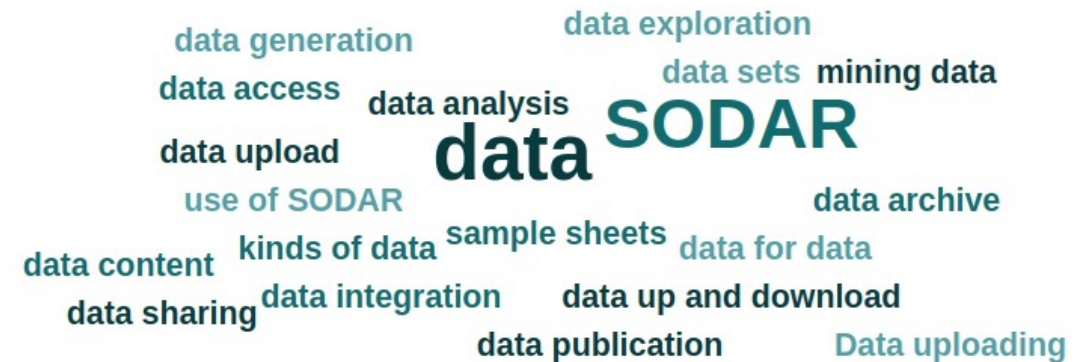
Senior Software Engineer, Core Unit Bioinformatics
SODAR User Meeting, 2024-10-18

Contents

1. Introduction
2. IRODS File Uploads and Access
3. Wet Lab Collaboration Example
4. Usability Issues
5. Open Discussion

Introduction

- We gathered feedback in the registration form
 - Topics you would like to see covered
 - Biggest challenges in using SODAR
- **Thank you for your responses!**
- Some topics already tackled in previous sessions
- We will address some of the most commonly mentioned topics
 - Open discussion and feedback at the end of the session



iRODS Data Uploads and Access (1/3)

- **Data Upload Tutorial**

- How to set up your data with .MD5 sum files
- How to create a landing zone
- Uploading your files to the landing zone with iRODS iCommands
- Validating your data and moving it to the sample repository

- Let's try to do this live..











The screenshot shows the SODAR Beta web interface. At the top, there is a search bar and navigation links for 'Manual' and 'Help'. Below the search bar, a breadcrumb trail reads 'Home / Test Category / Mikko Nieminen / Test Project'. A yellow notification banner at the top states: 'Landing zone "20220322_172043" creation initiated: see the zone list for the creation status. Collections will be created.' The main content area is titled 'Test Project' and 'Landing Zones'. A table titled 'Your Zones' lists three zones with their status and details.

Zone	Status Info	Status	Links
20211004_140255	Failed to move landing zone files: BatchValidateChecksumsTask failed: Checksums do not match for "test3_MODIFIED.txt" in resource "ceph-1" (File: df80abc3ee5a7eaa073aa2ce419a6b5; iRODS: None) 1 file (105 B)	FAILED	[Icons]
20220201_132730_deploy_test	Available with write access for user 0 files (0 B)	ACTIVE	[Icons]
20220322_172043	Creating landing zone in iRODS 0 files (0 B)	CREATING	[Icons]





iRODS Data Uploads and Access (2/3)





- **Data Access Tutorial**


- How to browse files in sample sheets
- How to browse files in browser via WebDAV
- How to open files in browser via WebDAV
- **TBD:** Demonstrate IGV integration?
- How to download data using iCommands

Sample	Links
Name	Study
HG00119-N1	 
HG00121-N1	 
HG00126-N1	 
HG00102-N1	 
HG00107-N1	 

Assay Shortcuts

Results and Reports    





Misc Files    



This is a directory listing of the CUBI iRODS server.

[Index of /sodarZone/projects/a1/a12164ff-9753-4d32-a520-7f499f862910/sample_data/study_34dcf3de-e410-4e53-b717-38f6bf6407f0/assay_a435a0b1-9ef0-4fc7-b69f-faff954a260e/HG00119-N1-DNA1-WES1/raw_data/2019-09-18/ on sodarZone](#)

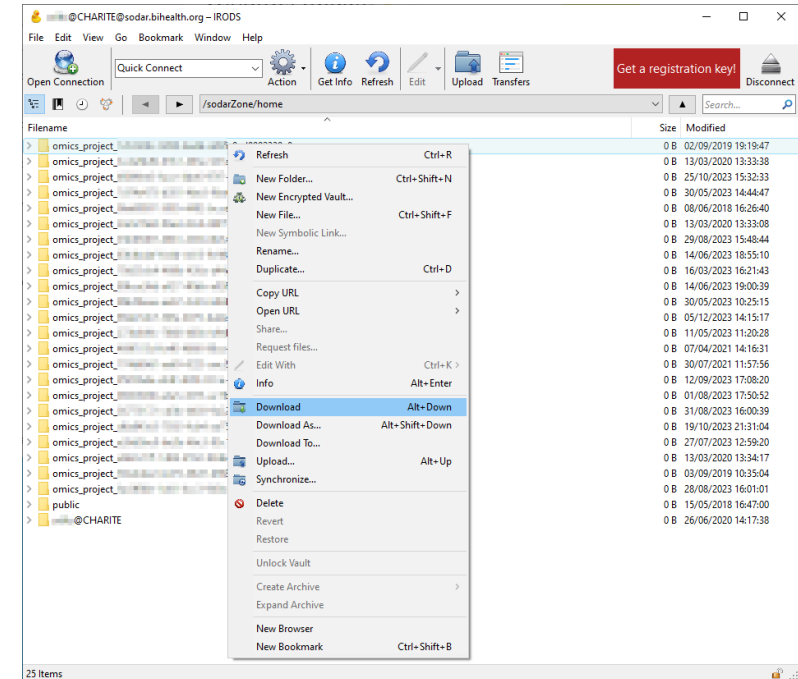
Parent collection

Name	Size	Owner	Last modified
 SRR099967_1.fastq.gz	7.4G	holtgrem@CHARITE	2019-09-23 08:27
 SRR099967_1.fastq.gz.md5	55	holtgrem@CHARITE	2019-09-23 08:27
 SRR099967_2.fastq.gz	7.5G	holtgrem@CHARITE	2019-09-23 08:27
 SRR099967_2.fastq.gz.md5	55	holtgrem@CHARITE	2019-09-23 08:27

iRODS Data Uploads and Access (3/3)

- **Other Data Upload and Access Methods**

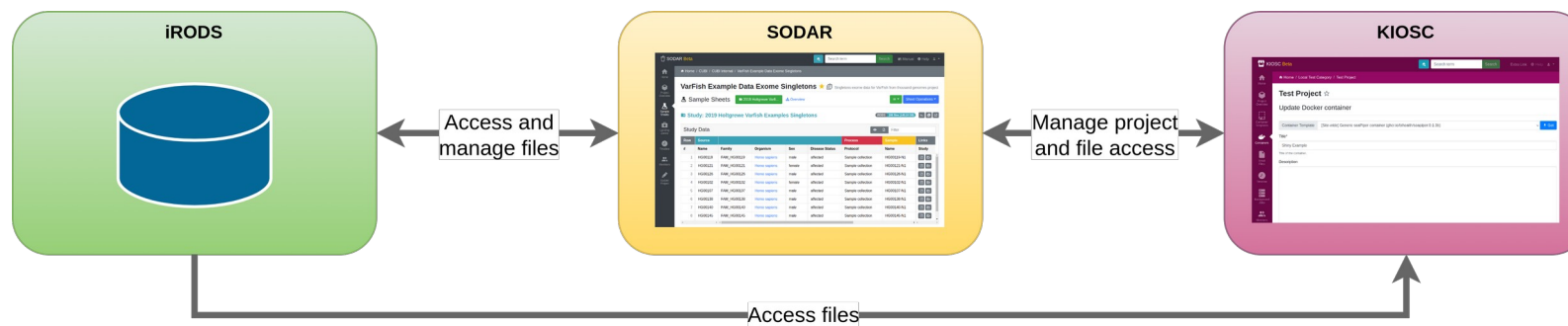
- **CUBI Toolkit** aka CUBI-TK
 - More generic command line access
- **WebDAV** network drive mounting
 - Works on some systems
- **Cyberduck**
 - Graphical interface
 - Requires some setup, see SODAR manual for documentation
- **iRODS Python client**
 - Powerful for writing Python tools but not user friendly
- **iBridges** wrapper on iRODS Python client
 - Recommended over base client



Wet Lab Collaboration Example (1/6)

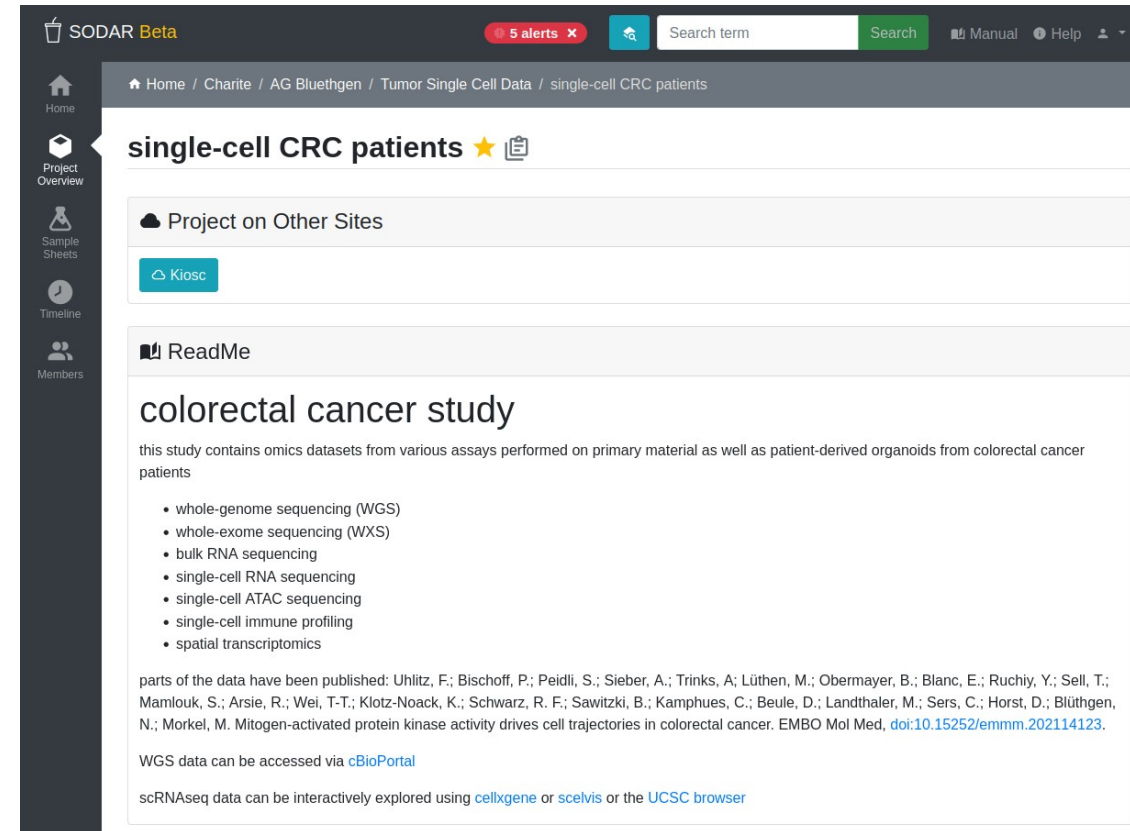
- **Enabling Collaboration with External Tools**

- Provide data exploration and visualization tools for wet lab scientists
- Project files and access control in SODAR
- External systems access SODAR data using iRODS access tickets
- Both in-house CUBI tools and third party software can be used
- **Kiosc** is commonly used to run data exploration tools with SODAR data



Wet Lab Collaboration Example (2/6)

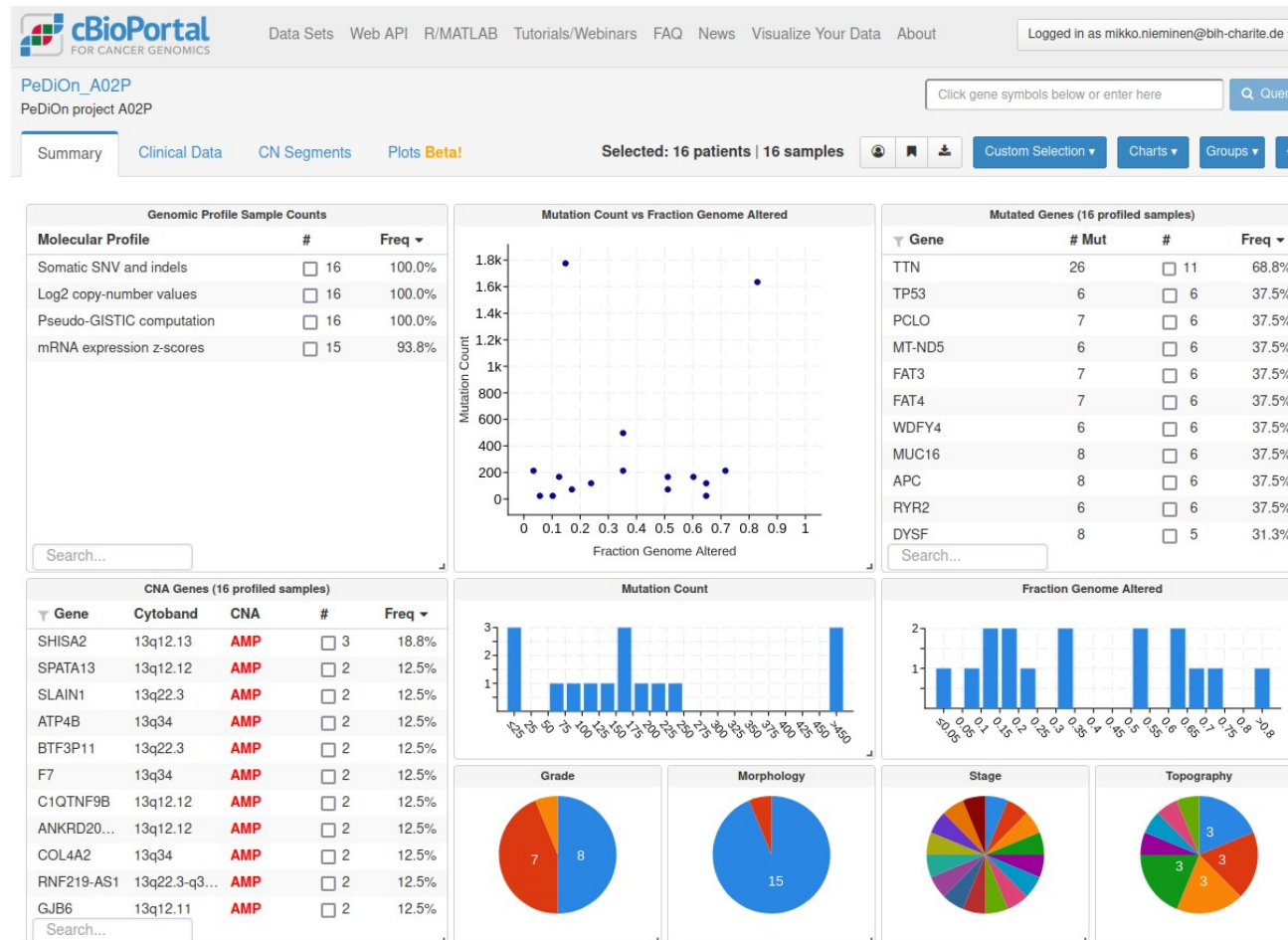
- **Example Project: Single-Cell Colorectal Cancer Study**
 - Links to other services using data in project readme
 - WGS data in **cBioPortal**
 - scRNAseq data can be interactively explored using multiple tools
 - **Cellxgene**
 - **Scelvis**
 - **UCSC Genome Browser** is used here for scATAC



The screenshot displays the SODAR Beta web interface. The top navigation bar includes the SODAR Beta logo, a search bar with the text 'Search term', and a 'Search' button. A notification for '5 alerts' is visible. The breadcrumb trail shows the path: Home / Charité / AG Bluethgen / Tumor Single Cell Data / single-cell CRC patients. The main content area is titled 'single-cell CRC patients' and features a 'Project on Other Sites' section with a 'Kiosk' button. Below this is a 'ReadMe' section for the 'colorectal cancer study'. The ReadMe text states: 'this study contains omics datasets from various assays performed on primary material as well as patient-derived organoids from colorectal cancer patients'. A bulleted list of assays includes: whole-genome sequencing (WGS), whole-exome sequencing (WXS), bulk RNA sequencing, single-cell RNA sequencing, single-cell ATAC sequencing, single-cell immune profiling, and spatial transcriptomics. A citation is provided: 'parts of the data have been published: Uhlitz, F.; Bischoff, P.; Peidli, S.; Sieber, A.; Trinks, A.; Lüthen, M.; Obermayer, B.; Blanc, E.; Ruchiy, Y.; Sell, T.; Mamlouk, S.; Arsie, R.; Wei, T-T.; Klotz-Noack, K.; Schwarz, R. F.; Sawitzki, B.; Kamphues, C.; Beule, D.; Landthaler, M.; Sers, C.; Horst, D.; Blüthgen, N.; Morkel, M. Mitogen-activated protein kinase activity drives cell trajectories in colorectal cancer. EMBO Mol Med, doi:10.15252/emmm.202114123.' At the bottom, it notes that WGS data can be accessed via cBioPortal and scRNAseq data can be interactively explored using cellxgene or scelvis or the UCSC browser.

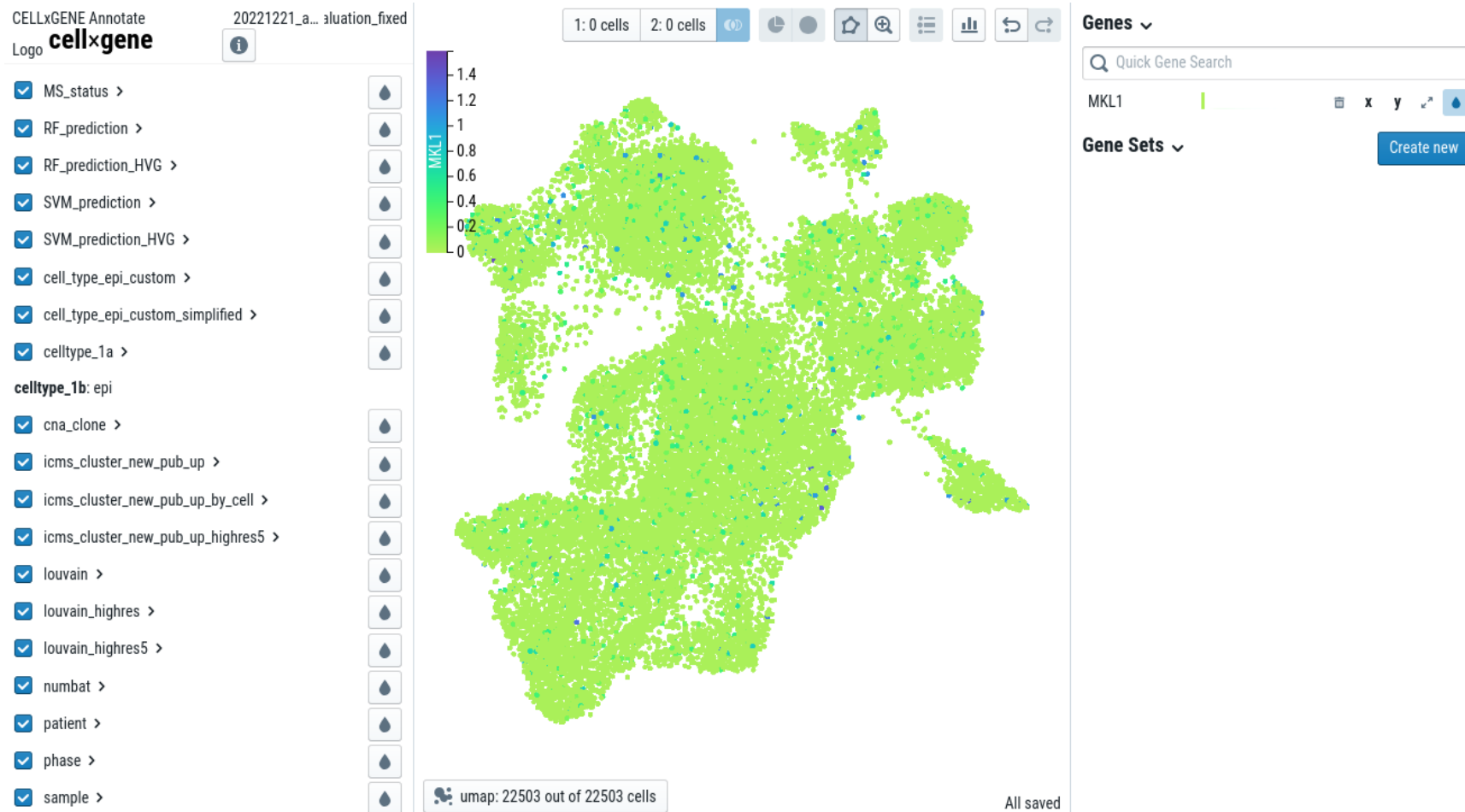
Wet Lab Collaboration Example (3/6)

- Exploring Data in cBioPortal



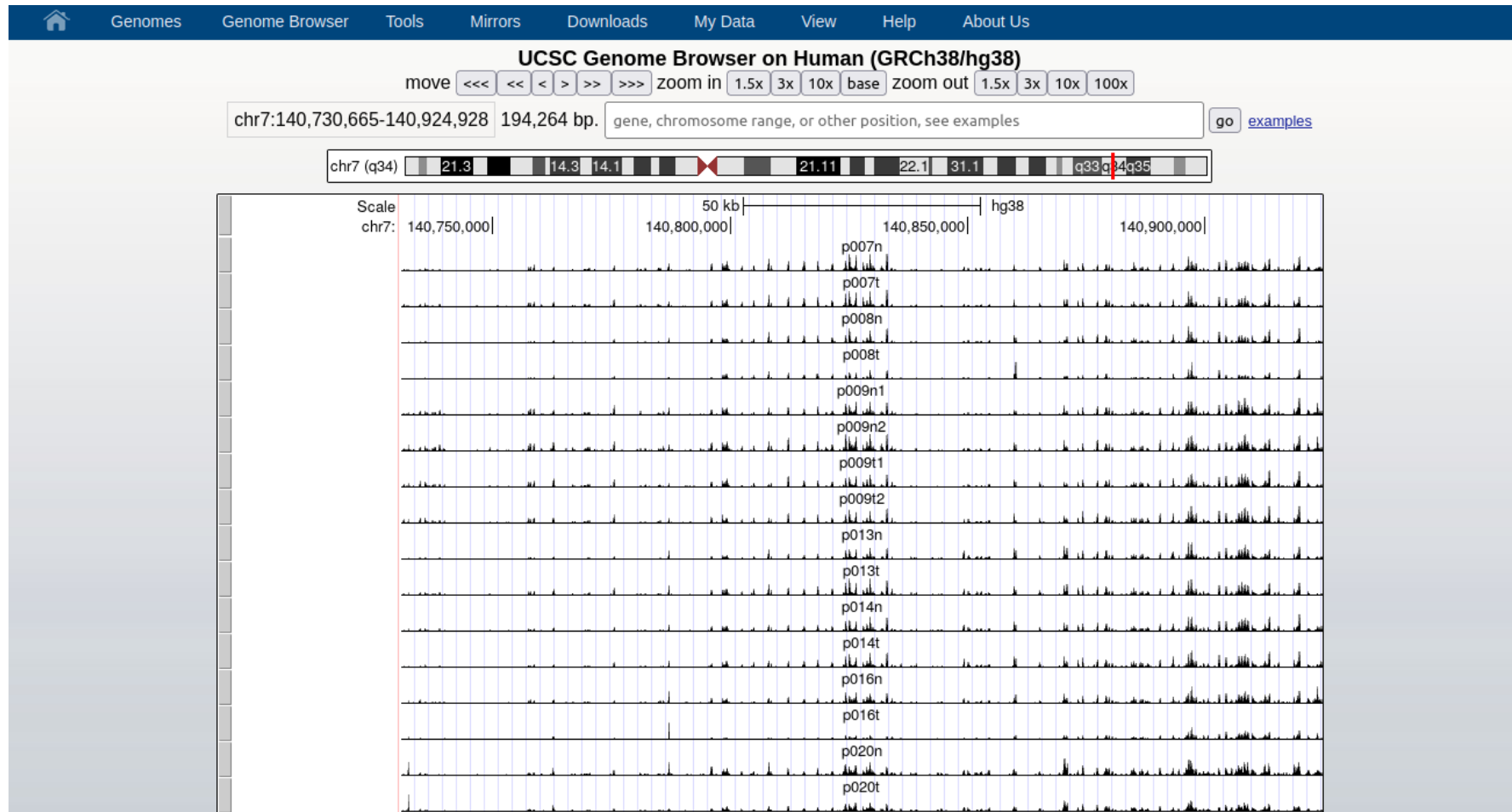
Wet Lab Collaboration Example (4/6)

- Exploring Data with Cellxgene



Wet Lab Collaboration Example (5/6)

- Exploring Data with UCSC Genome Browser



Wet Lab Collaboration Example (6/6)

- Project in Kiosc

The screenshot displays the KIOSC Beta web interface. The top navigation bar includes a search bar and links for Docs and Help. The breadcrumb trail shows the path: Home / Charite / AG Bluethgen / Tumor Single Cell Data / single-cell CRC patients. The left sidebar contains navigation icons for Home, Project Overview, Container Templates, Containers, Files, Timeline, Background Jobs, and Members.

Containers overview

Title	Repository:Tag	State
2022/12/12/ CRC patients scelvis	ghcr.io/bihealth/scelvis:0.8.9-0	running
2022/12/12 CRC patients cellxgene	quay.io/biocontainers/cellxgene:1.1.1--pyhdfd78af_0	running

Files Overview

Item	Size	Description	Owner	Updated
No files or links.				

Timeline Overview

Timestamp	Event	User	Description	Status
2024-10-17 14:48:14	remote_role_create	admin	add role "project guest" for mikkopen@CHARITE from site SODAR	OK
2024-10-17 13:08:16	container_task	obermayb@CHARITE	start container 2022/12/12 CRC patients cellxgene	OK
2024-10-17 10:26:12	container_task	obermayb@CHARITE	start container 2022/12/12/ CRC patients scelvis	OK
2024-10-17 10:23:18	container_task	obermayb@CHARITE	start container 2022/12/12 CRC patients cellxgene	OK
2024-10-17 10:20:53	container_task	obermayb@CHARITE	start container 2022/12/12 CRC patients cellxgene	OK

Usability Issues (1/4)

- One of the more major concerns raised in feedback
- We do our best to make things as accessible as possible
 - ..but we may not always succeed
- Feedback and suggestions from users are very much welcome
 - Especially regarding specific tasks or functions with which you may be struggling
- We'll discuss some commonly raised issues here

Usability Issues (2/4)

- **Difficulties in Using iRODS**

- iCommands may be challenging for people not used to the Unix shell
- To some degree this is out of our hands

- **Possible Solutions**

- Provide improved documentation and links to helpful iRODS resources
- Using and documenting Cyberduck, iBridges or other wrappers
- Possible simple drag-n-drop upload tools to be introduced in the future
 - Will not solve issues with mass uploads, checksum file generating or accessing uploaded files

Usability Issues (3/4)

- **Unclear Terminology**

- A lot of the terminology comes from 3rd party systems used or generally used terms
 - That said, we should do our best to reduce confusion
- First step would be to improve the **glossary section in the SODAR manual**
 - Feel free to suggest specific terms which need clarification
 - ..or better yet, submit a pull request if you're familiar with missing terms :)
- Other ideas?

Usability Issues (4/4)

- **General User Friendliness**

- SODAR may admittedly have a high bar to enter
- Hard to pinpoint issues without specific feedback
- Changing established user interfaces creates outcry from existing users
- **Possible solutions**
 - Improved documentation
 - More video tutorials for common tasks
 - Improvements to UI/software e.g. where necessary and feasible
 - Implementing „simple mode“ for non-expert users
- **How to help**
 - Provide specific feedback on difficulties
 - Can be done through helpdesk or issue trackers

Open Discussion

- Any feedback or points of discussion you want to raise?

Thank You

www.cubi.bihealth.org

BIH Berlin Institute
of Health
@Charité