Accessing HuBMAP Data in HuBMAP Workspaces

Phil Blood
Scientific Director
PI, HuBMAP Infrastructure and Engagement
Pittsburgh Supercomputing Center
Carnegie Mellon University

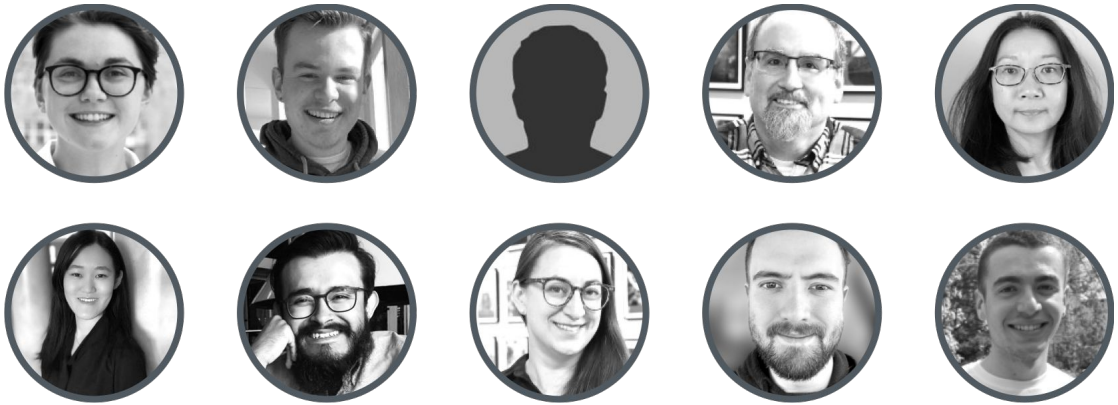
Overview

1. HuBMAP Workspaces
2. Workspaces Tutorial

Workspaces

Team

- John Conroy
- Juan Muerto
- Tiffany Liaw
- Thomas Smits
- Austen Money
- Mark Keller
- Nick Akhmetov
- Lisa Choy
- Gesina Phillips
- Chris Csonka



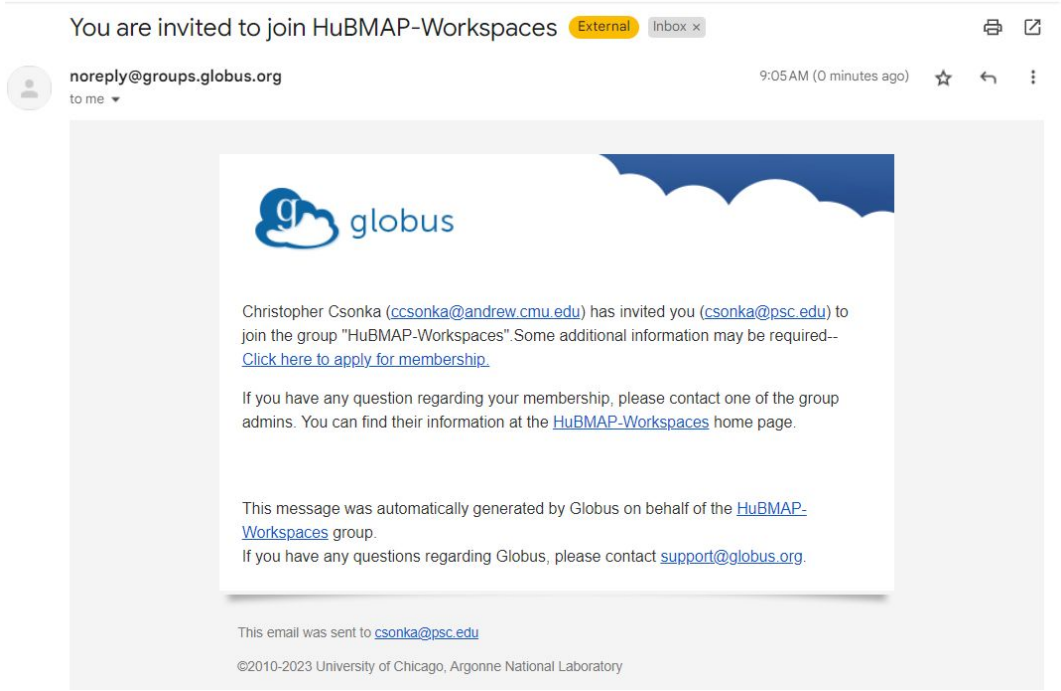
HuBMAP Data Portal Workspaces

- Joint effort by Harvard and UPitt/PSC
  - *Harvard:* user interfaces, portal integration, workspace templates (front end)
  - *UPitt/PSC:* computational infrastructure (back end)
- Target audiences
  - *Current:* computational analysts in HuBMAP and elsewhere
  - *Future:* students and trainees aiming to learn about spatial and single-cell data analysis approaches
- Support analysis of HuBMAP and user-provided data directly in the data portal
  - No need to download HuBMAP data but users can upload their own data (size currently limited)
  - Compute resources provided by PSC on hardware funded by HuBMAP
- Uses popular JupyterLab workspace environment and supports Python and R
  - Analysis templates demonstrate workspace capabilities and facilitate quick startup for common analyses

Getting Access to Workspaces

https://portal.hubmapconsortium.org/workspaces

- **HuBMAP Consortium Members:** You already have access. Just log in at the URL above.
- **All Others:** Accept the invitation to HuBMAP Globus you received in email by following the steps outlined in that message. Then log in at the URL above.



Managing Shared Resources

- Intended for lighter-weight, interactive analyses
- Take steps to use available resources efficiently
  - Use **minimum data** needed for development
  - **Stop workspaces** when not in use (you will be able to restart and resume where you left off)
- Be a good workshop citizen!

- Workspace resource **maximum** limits [please use the minimum needed]
- **CPUs:** 16 cores per Workspace (128 total available)
- **RAM:** 128 GB per Workspace (2-4 TB total available)
- **Session length:** 12 hours

portal.hubmapconsortium.org/workspaces

HuBMAP Data Resources

My Workspaces

Getting Started

Get a tutorial of how to explore workspaces to analyze HuBMAP data.

Navigate To The Workspace Tutorial

What are workspaces?

Workspaces enable lightweight exploration of public HuBMAP data and user-provided data using Python and R in a Jupyter Lab environment hosted by HuBMAP at no cost to community members.

How do I use workspaces?

Explore more about this platform through our workspace tutorials to optimize your experience with workspaces. To begin a new workspace, find datasets on our search page and launch a workspace from them.

What do workspaces currently support?

Workspaces launch with Python support by default, with the option to add support for R upon launch. Please note that workspaces with added R support may experience longer load times.

Questions/Suggestions

Please be aware that certain limitations currently exist on this platform due to its simplified exploration design. If you have any questions or suggestions about workspaces, contact us through the HuBMAP Help Desk.

17 Workspaces

☐	HBM976.ZTXZ.422 Workspace Status: No Jobs | ID: 220 | Created 2023-03-24	Launch Workspace
☐	HBM979.KNQM.454 Workspace Status: No Jobs | ID: 278 | Created 2023-07-21	Launch Workspace
☐	Kidney Slide-Seq Demo Status: No Jobs | ID: 234 | Created 2023-03-27	Launch Workspace
☐	Kidney snR Status: No Jobs | ID: 237 | Created 2023-04-06	Launch Workspace

portal.hubmapconsortium.org/workspaces

Launch New Workspace

Workspaces are currently Jupyter Notebooks that allows you to perform operations on HuBMAP data.

Three steps are shown for launching a workspace, but the only required field to launch a workspace is “Step 2: Configure Workspace”. “Step 1: Edit Datasets Selection” is only required if there are issues with any of the datasets selected, which will be indicated by an error banner.

1. Edit Datasets Selection (Optional)

To remove a dataset, select the dataset and press the delete button. If all datasets are removed, an empty workspace will be launched.

To add more datasets to a workspace, you must navigate to the dataset search page, select datasets of interests and follow steps to launch a workspace from the search page. As a reminder, once you navigate away from this page, all selected datasets will be lost so take note of any HuBMAP IDs of interest, or copy IDs to your clipboard by selecting datasets in the table below and pressing the copy button. You can also save datasets to the “My Lists” feature.

2. Configure Workspace (Required)

Name

Like “Spleen-Related Data” or “ATAC-seq Visualizations”

A workspace name is required. Please enter a workspace name.

All workspaces are launched with Python support, with the option to add support for R. Workspaces with added R support may experience longer load times.

Select Environment

- (•) Jupyter Lab (Python)
- ( ) Jupyter Lab (Python+R)

3. Select Templates (Optional)

Templates can be selected for your workspace for potential workflows revolving around data access, visualization, QA or other purposes. Multiple templates

Cancel Launch Workspace

Kidney snR

Launch Workspace

portal.hubmapconsortium.org/workspaces

Launch New Workspace

Workspaces are currently Jupyter Notebooks that allows you to perform operations on HuBMAP data.

Three steps are shown for launching a workspace, but the only required field to launch a workspace is “Step 2: Configure Workspace”. “Step 1: Edit Datasets Selection” is only required if there are issues with any of the datasets selected, which will be indicated by an error banner.

Filter workspace templates by tags

Recommended Tags

visualization | api

1 Templates Selected

Deselect All | Select All

Blank notebook

Notebook only inserting UUIDs of added datasets

Analyse neighborhoods of slide-seq spatial data with squidpy

Use squidpy to analyse and visualize neighborhoods of clusters of slide-seq data.

slide-seq | visualization | anndata

Cluster SPRM output of CODEX data and visualize with Vitessce

This template shows how to cluster CODEX data and load a new clustering into Vitessce

codex | visualization | vitessce

Compress Anndata

This notebook shows how to manipulate anndata objects in h5ad files.

anndata

JupyterLab HuBMAP API Tutorial

This notebook shows users how to utilize the HuBMAP APIs to perform analysis.

sdk | api

JupyterLab Vitessce Visualization

This notebook allows the user to visualize a specific dataset using the vitessce framework.

vitessce | visualization

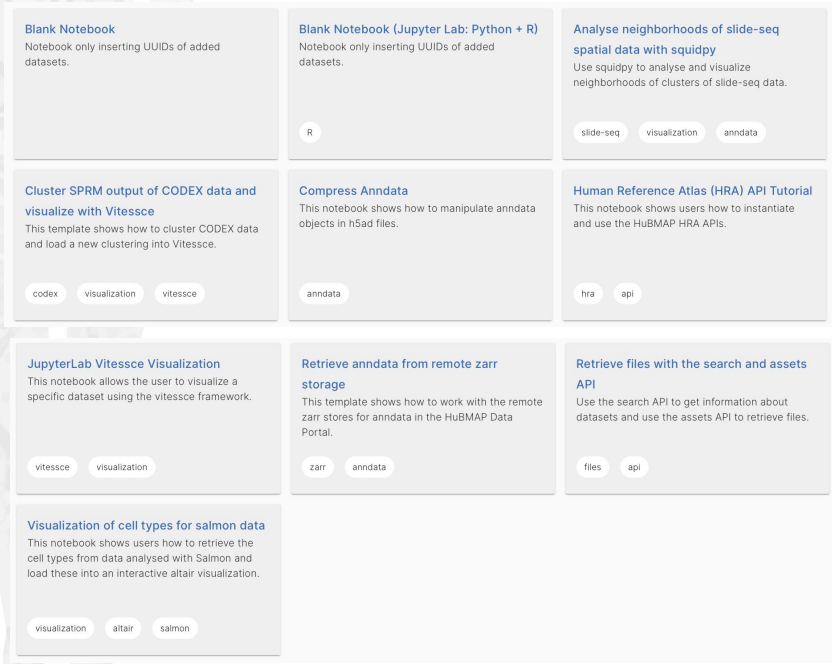
Cancel | Launch Workspace

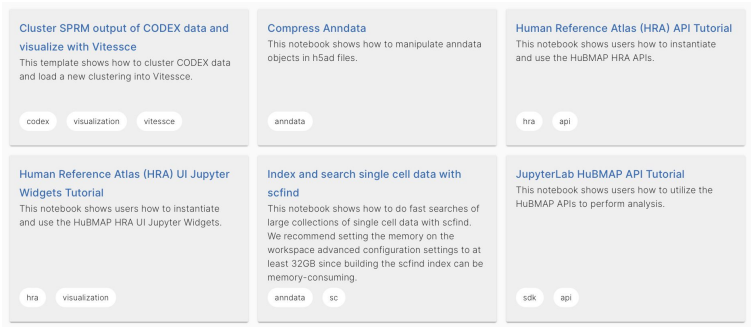
Kidney snR

Launch Workspace

HuBMAP Workspace Templates

Templates are pre-set notebooks that can be launched with different datasets.





You will learn how to contribute your own *Workspace Templates*.

It's great way to disseminate your algorithms and tools.

workspaces-pt.hubmapconsortium.org/passthrough/a001.pvt.hive.psc.ed

File Edit View Run Kernel Tabs Settings Help

Filter files by name

/

Name	Modified
datasets	1 minute ago
celltypes_salmon.ipynb	1 minute ago

celltypes_salmon.ipynb

Markdown

Python 3 (ipykernel)

Determining the composition of cell types in Salmon data

This notebook shows how to extract cell types from data analysed with salmon, and visualize them in several ways with Altair.

```
[ ]: # !pip install --upgrade pip
     # !pip install numpy pandas requests wheel matplotlib matplotlib-inline altair seaborn anndata
```

```
[ ]: import warnings
     import requests
     import json

     import os
     import os.path

     from csv import DictReader, excel_tab
     from io import StringIO

     import pandas as pd

     from matplotlib import pyplot as plt
     import seaborn as sns
     import altair as alt

     import anndata as ad
```

```
[ ]: # One of pandas's functions depends on 'convert_dtype' which generates a FutureWarning,
     # but as it's called many times, the notebook becomes crowded with this same warning.
     # This cell surpresses this warning.
     warnings.simplefilter(action='ignore', category=FutureWarning)
```

Linked datasets

The following datasets were symlinked to the workspace when this template was added:

```
[ ]: # linked datasets
     uuids = ['e81c9c9753998b2ca69e8bceadf2409d']

     # required file_types
     required_filestypes = ['secondary_analysis.h5ad']

     # search_api
     search_api = 'https://search.api.hubmapconsortium.org/v3/portal/search'
```

The following checks if the datasets are compatible with this template.

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 celltypes_salmon.ipynb 1

Workspaces Tutorial

portal.hubmapconsortium.org/workspaces

This Repository Is Under Review For Potential Modification In Compliance With Administration Directives.

HuBMAP Data Resources

Workspaces

Getting Started

Get a tutorial of how to explore workspaces to analyze HuBMAP data.

Navigate To The Workspace Tutorial

What are workspaces?

Workspaces provides a lightweight exploration platform tailored for researchers to easily access HuBMAP data and perform analyses directly within the portal. Effortlessly upload dataset files to a Jupyter notebook using provided templates to get started on analyzing HuBMAP data.

How do I use workspaces?

Explore more about this platform through our workspace tutorials to optimize your experience with workspaces. To begin a new workspace, find datasets on our search page and launch a workspace from them. To learn more about getting started, explore workspace templates to help you start analyzing HuBMAP data.

What do workspaces currently support?

Workspaces launch with Python support by default, with the option to add support for R upon launch. Please note that workspaces with added R support may experience longer load times.

Questions/Suggestions

Please be aware that certain limitations currently exist on this platform due to its simplified exploration design. If you have any questions or suggestions about workspaces, submit feedback at the HuBMAP #workspaces-feedback Slack channel, or contact us through the help desk.

Relevant Pages

Tutorials | Templates | Dataset Search Page

Workspace Invitations

Manage your workspace collaboration with notifications for invitations shared with you or sent by you. For received invitations, you can preview the workspace details before deciding to accept or decline the invitation.

Received | Sent

No received workspace invitations.

My Workspaces (5)

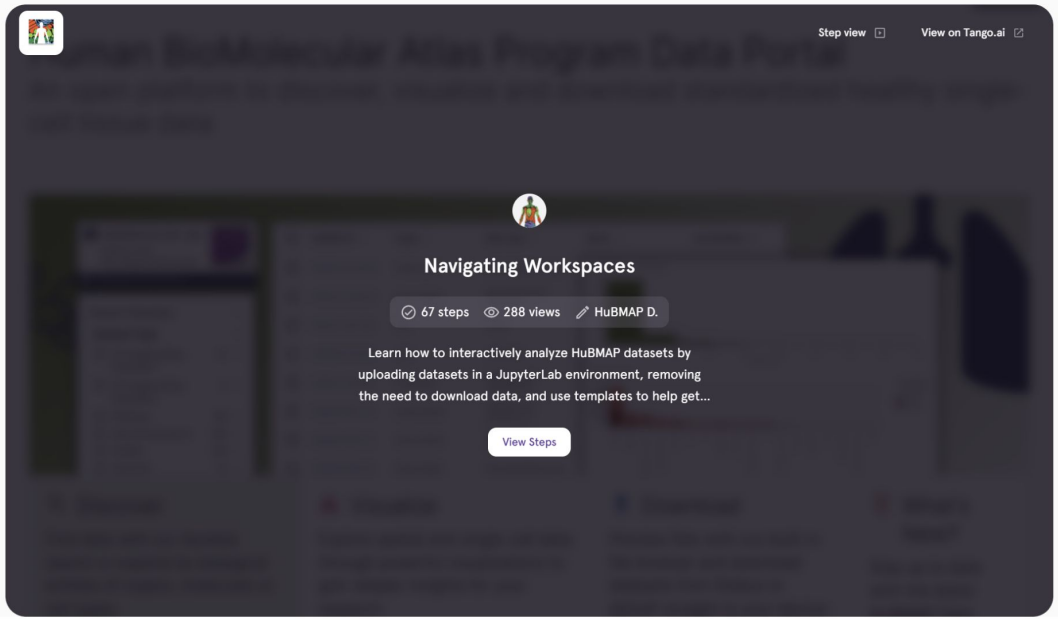
portal.hubmapconsortium.org/tutorials/workspaces

This Repository Is Under Review For Potential Modification In Compliance With Administration Directives.

HuBMAP Data Resources

Navigating Workspaces

Learn how to use workspaces to analyze HuBMAP data by initiating Jupyter notebooks and choosing from a variety of pre-established templates.



HuBMAP

About
Project Website
Documentation
Diversity
Submit Feedback

Software
GitHub
Services
APIs
Portal Usage Analytics

Policies
Overview
Data Sharing Policy
Citing HuBMAP

Funding
NIH Common Fund

Goals

Workshop Goals

Participants

- Learn about HuBMAP data and resources
- Share your methods and tools with the HuBMAP community through new *Workspace Templates*
- Influence the *Workspace* development agenda for the next few months

HuBMAP HIVE

- Onboard new *Workspace* users
- Assess *Workspace* infrastructure under real-world conditions
  - UI
  - Backend
- Prioritize feature development for future sprints