Tutorial & Demo: Submitting Data

Alejandro Bugacov

FaceBase Bootcamp (Nov. 5th 2025)

New Project Email Example:

Dear Dr. Bugacov,

We created your new FaceBase Project “FaceBase Demo” (https://doi.org/10.25550/1-RC7M)

Before you begin working on your datasets, you need to be added to the Globus group that controls access to your new Project. To be added to this group please follow these steps:

1) Click the URL: https://app.globus.org/groups/5eebf313-c0b2-11ea-adc2-0ed6a5b292e1/about

2) Choose your method of signing in to Globus:

   a) You can use your institution's credentials by searching your institution's name in the search field.

   b) If your institution is not available, click where it says "Then use Globus ID to sign in." to choose your own username/password and follow the instructions.

3) Request group membership: Once you are signed in, you should see the group page for your project. Click the "Join" button (towards the right side of the page) to request membership.

FaceBase Focus on Streamlined Data Sharing

- New evolved data model to streamline the dataset curation and data upload
- Unified platform that supports different strategies for different type of dataset scenarios
  - From the small and simple dataset to the large and complex to controlled-access clinical (human) data
- Fully web-based self-curated mechanism with strict privacy control and sophisticated UIs (e.g., batch editing and cloning)
- Robust tools and web APIs for batch data uploads and dataset curation
- Data processing pipelines and visualizations

FaceBase supports different strategies for various dataset scenarios

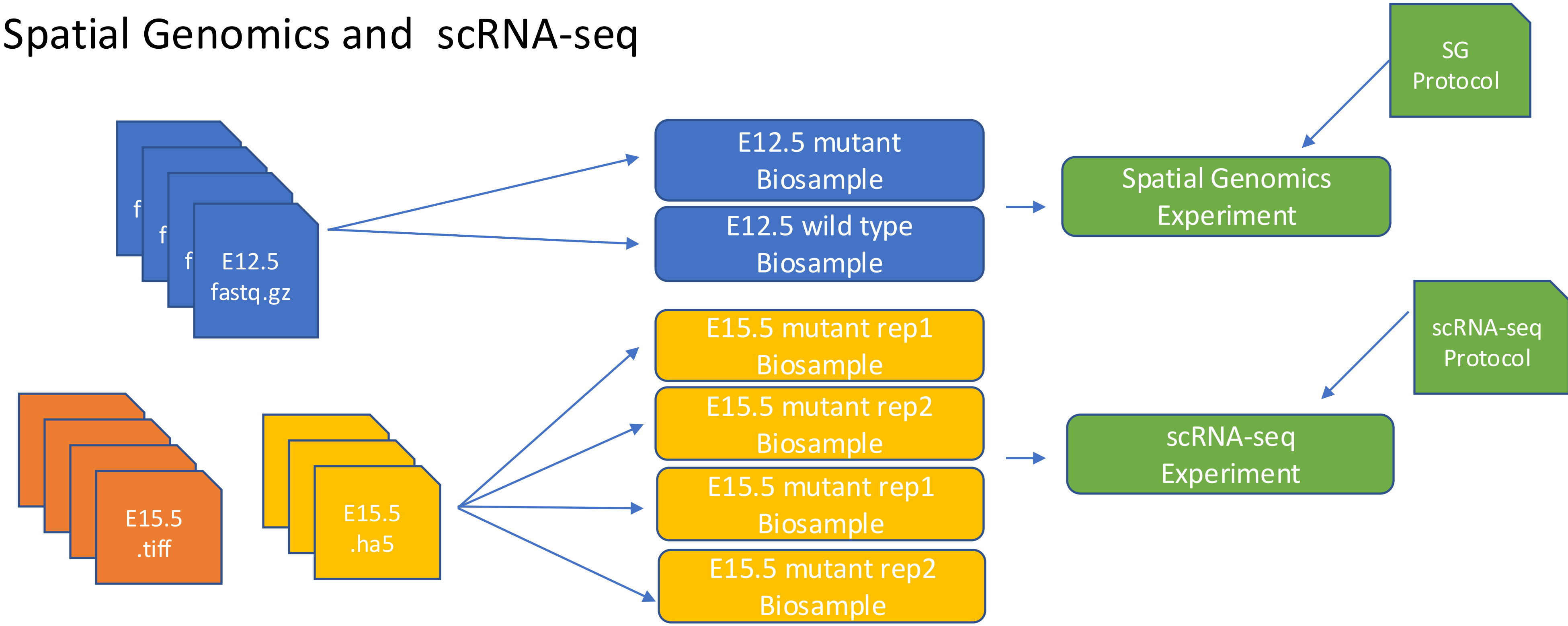
- Simple scenario:
  - About 10 or so Files and Biosamples and 1 or 2 Experiment types
  - Fully browser-based self-curation and data upload
- Large dataset scenario:
  - ~100 of files and Biosamples and ~10s of Experiment types
  - Hybrid strategy: Browser-based and DERIVA Upload Tool combination
    - Drop all files you want to share to a folder in your desktop or server
    - Batch upload all files at once using our robust DERIVA upload tool
    - Using the browser, associate uploaded files to corresponding Biosample taking advantage of UI batch editing support
- Clinical (human) data scenario:
  - *Continuous* sharing of data from clinical source
  - Combine browser-base, DERIVA Tools and web-service API to programmatically upload data and its metadata

Dataset creation in FaceBase

- Step 1: Conceptualization
  - Think about how many data **Files** you want to share, how many **Biosamples** (or **Participants/Visits** if clinical data) these files correspond to and what **Experiments** were performed on them
- Step 2: Dataset Curation and data Upload
  - Create Dataset
  - Create Experiments, Biosamples
    - **Metadata** and **Protocols**
  - Upload data Files
  - Associate Files to corresponding Biosamples and Experiments
- Step 3: Dataset Review
  - Decide if ready for public release or embargo

Demo Dataset

- How many files?
  - 10 data files
- How many Biosamples?
  - 6 Biosamples
    - 4 E15.5 soft palate: Acan -/- mutant and wild type
    - 2 E12.5 hard palate: Acan -/- mutant and wild type, 2 replicates
- How many Experiments?
  - 2 Experiments: Spatial Genomics and scRNA-seq



E12.5
fastq.gz

E15.5
.tiff

E15.5
.ha5

E12.5 mutant
Biosample

E12.5 wild type
Biosample

E15.5 mutant rep1
Biosample

E15.5 mutant rep2
Biosample

E15.5 mutant rep1
Biosample

E15.5 mutant rep2
Biosample

SG
Protocol

Spatial Genomics
Experiment

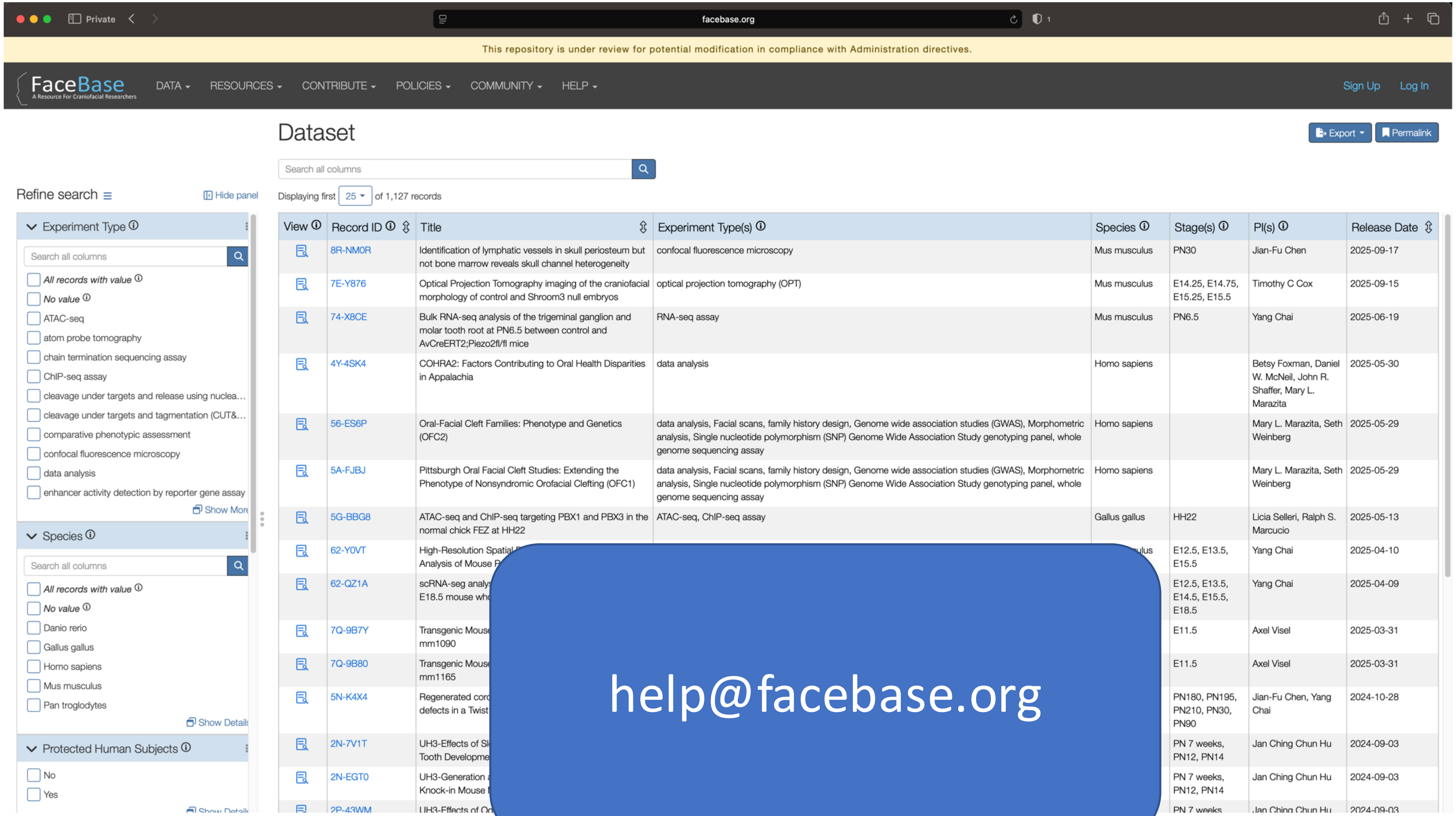
scRNA-seq
Protocol

scRNA-seq
Experiment

Data processing pipelines and visualization

Name	Data Types	Action	Type/Frequency
Imaging pipeline (https://www.facebase.org/id/3KW6)	Image file in large number of imaging formats from single image in PNG, JPG or TIFF to multichannel 3D and 4D stacks	Pipeline creates thumbnails and visualization viewer with annotations. Contributor can manually tag anatomical terms from UBERON, text, regions, etc	Automatic/Daily
Genome Browser (https://doi.org/10.25550/1-7780)	bigWig and bigBed files	FaceBase embeds a UCSC Genome Browser visualization of all the uploaded tracks in the Dataset page	Automatic/Daily
Cellbrowser pipeline (https://www.facebase.org/id/V-J7QP , https://www.facebase.org/id/R-W170)	Seurat (v3 and v4) and Cell Ranger matrix files (matrix.mtx, features.tsv and barcodes.tsv)	FaceBase embeds a UCSC Cellbrowser visualization of each uploaded file at the Data File page (Under upgrade/revision)	Automatic/Triggered by file upload
3D Mesh viewer (https://www.facebase.org/id/1-517M)	3D meshes in OBJ format	FaceBase shows a 3D visualization of the created model. Multi mesh, can tag anatomical terms from UBERON	Manual model creation, Automatic visualization
3D Volumetric viewer (https://www.facebase.org/id/TY8)	3D files in NIfTI format	FaceBase shows a 3D volume rendering of the created model	Manual resolution reduction, Automatic visualization
DOI assignment	Each FaceBase Dataset and Project	A unique and persistent Data Object Identifier (DOI) is assigned to each Dataset and Project page	Automatic/Daily

Looking forward to working with you and help you share your data thru' FaceBase!!!



This repository is under review for potential modification in compliance with Administration directives.

Dataset

Refine search

Experiment Type

- ☐ *All records with value*
- ☐ *No value*
- ☐ ATAC-seq
- ☐ atom probe tomography
- ☐ chain termination sequencing assay
- ☐ ChIP-seq assay
- ☐ cleavage under targets and release using nuclea…
- ☐ cleavage under targets and tagmentation (CUT&…
- ☐ comparative phenotypic assessment
- ☐ confocal fluorescence microscopy
- ☐ data analysis
- ☐ enhancer activity detection by reporter gene assay

Species

- ☐ *All records with value*
- ☐ *No value*
- ☐ Danio rerio
- ☐ Gallus gallus
- ☐ Homo sapiens
- ☐ Mus musculus
- ☐ Pan troglodytes

Protected Human Subjects

- ☐ No
- ☐ Yes

Displaying first 25 of 1,127 records

View	Record ID	Title	Experiment Type(s)	Species	Stage(s)	PI(s)	Release Date
	8R-NM0R	Identification of lymphatic vessels in skull periosteum but not bone marrow reveals skull channel heterogeneity	confocal fluorescence microscopy	Mus musculus	PN30	Jian-Fu Chen	2025-09-17
	7E-Y876	Optical Projection Tomography imaging of the craniofacial morphology of control and Shroom3 null embryos	optical projection tomography (OPT)	Mus musculus	E14.25, E14.75, E15.25, E15.5	Timothy C Cox	2025-09-15
	74-X8CE	Bulk RNA-seq analysis of the trigeminal ganglion and molar tooth root at PN6.5 between control and AvCreERT2;Piezo2fl/fl mice	RNA-seq assay	Mus musculus	PN6.5	Yang Chai	2025-06-19
	4Y-4SK4	COHRA2: Factors Contributing to Oral Health Disparities in Appalachia	data analysis	Homo sapiens		Betsy Foxman, Daniel W. McNeil, John R. Shaffer, Mary L. Marazita	2025-05-30
	56-ES6P	Oral-Facial Cleft Families: Phenotype and Genetics (OFC2)	data analysis, Facial scans, family history design, Genome wide association studies (GWAS), Morphometric analysis, Single nucleotide polymorphism (SNP) Genome Wide Association Study genotyping panel, whole genome sequencing assay	Homo sapiens		Mary L. Marazita, Seth Weinberg	2025-05-29
	5A-FJBJ	Pittsburgh Oral Facial Cleft Studies: Extending the Phenotype of Nonsyndromic Orofacial Clefting (OFC1)	data analysis, Facial scans, family history design, Genome wide association studies (GWAS), Morphometric analysis, Single nucleotide polymorphism (SNP) Genome Wide Association Study genotyping panel, whole genome sequencing assay	Homo sapiens		Mary L. Marazita, Seth Weinberg	2025-05-29
	5G-BBG8	ATAC-seq and ChIP-seq targeting PBX1 and PBX3 in the normal chick FEZ at HH22	ATAC-seq, ChIP-seq assay	Gallus gallus	HH22	Licia Selleri, Ralph S. Marcucio	2025-05-13
	62-Y0VT	High-Resolution Spatial Analysis of Mouse		ulus	E12.5, E13.5, E15.5	Yang Chai	2025-04-10
	62-QZ1A	scRNA-seg analy E18.5 mouse wh			E12.5, E13.5, E14.5, E15.5, E18.5	Yang Chai	2025-04-09
	7Q-9B7Y	Transgenic Mouse mm1090			E11.5	Axel Visel	2025-03-31
	7Q-9B80	Transgenic Mouse mm1165			E11.5	Axel Visel	2025-03-31
	5N-K4X4	Regenerated cor defects in a Twist			PN180, PN195, PN210, PN30, PN90	Jian-Fu Chen, Yang Chai	2024-10-28
	2N-7V1T	UH3-Effects of S Tooth Developme			PN 7 weeks, PN12, PN14	Jan Ching Chun Hu	2024-09-03
	2N-EGT0	UH3-Generation Knock-in Mouse			PN 7 weeks, PN12, PN14	Jan Ching Chun Hu	2024-09-03
	2P-43WM	UH3-Effects of O			PN 7 weeks	Jan Ching Chun Hu	2024-09-03

YOUR DATASETS HERE

help@facebase.org

Citing and Referencing your FaceBase *DATA*

- A DOI is assign to your Dataset the moment you create it
  - Don’t need to wait for Dataset to be released to the public
- You can include the DOI in your manuscript right away
  - Contact us at help@facebase.org for anonymous journal reviews
- DOIs are assigned at each Dataset and Project level, are unique and persistent
- Share Link: Automatically generated for every page
  - Click “Share and Cite” (top-right corner). E.g., https://www.facebase.org/id/R-W170
- Citation: Automatically generated for every Dataset and Project
  - Click “Share and Cite” (top-right corner)
  - This is how FaceBase would like you to Cite your ***data***

Data Citation Example:

> Samantha Brugmann, Christian Bonatto Paese. Single cell transcriptomics of the talpid2 avian mutant mandibular prominence. *FaceBase Consortium* https://doi.org/10.25550/R-PPWJ (2022).

Exporting Data from FaceBase

- Individual File Download
  - Click on any (highlighted) file and the browser will prompt you to save the file to your computer
  - Do not need to be logged in
- Bulk Export:
  - CSV
    - Downloads a CSV file with all metadata from the result of a Search
    - Available both at the Results and Record page
    - Most useful to download results of a search in tabular form
  - BDBAG
    - Useful download *all data* from a dataset or a project
    - Very robust tool
      - Data validation, console showing progress and status, can be restarted from where it left off
    - Need to login with FaceBase User credentials
    - First Step: Download a BDBag (Big Data Bag) file to your computer
    - Second Step: Use the DERIVA Bag Tool to export the all data files and metadata to your computer
      - In the afternoon session we'll see how to install this tool
- Detail documentation: https://www.facebase.org/help/exporting/

Installing DERIVA Client Tools

- Suite of reliable tools for batch uploading and downloading data
- GUI and command line tool for Mac, Windows and Linux
- Detailed documentation:
  - https://github.com/informatics-isi-edu/facebase-curation/wiki/Deriva-Clients
  - Mac and Windows:
    - Download bundle from
      - https://github.com/informatics-isi-edu/deriva-client-bundle/releases
  - Linux:
    - Use pip install:
      - pip3 install --user deriva-client

Data Submission Resources

- Starting point
  - https://www.facebase.org/submit/submitting-data/
- FaceBase data curation Wiki
  - https://github.com/informatics-isi-edu/facebase-curation/wiki
  - Detailed step-by-step documentation
- Material for self-directed tutorial can be found at:
  - Email help to be added to the Demo group to do the self-directed tutorial
- Need a tutorial?
  - Contact the Hub (help@facebase.org) to setup a 1-hour walk-through conference call
- Need help while working on your datasets?
  - Email us at help@facebase.org !!!
  - FaceBase Monthly Office Hours: Last Wednesday of the Month