FACEBASE WORKSHOP:
FAIR DATA FOR CRANIOFACIAL RESEARCH

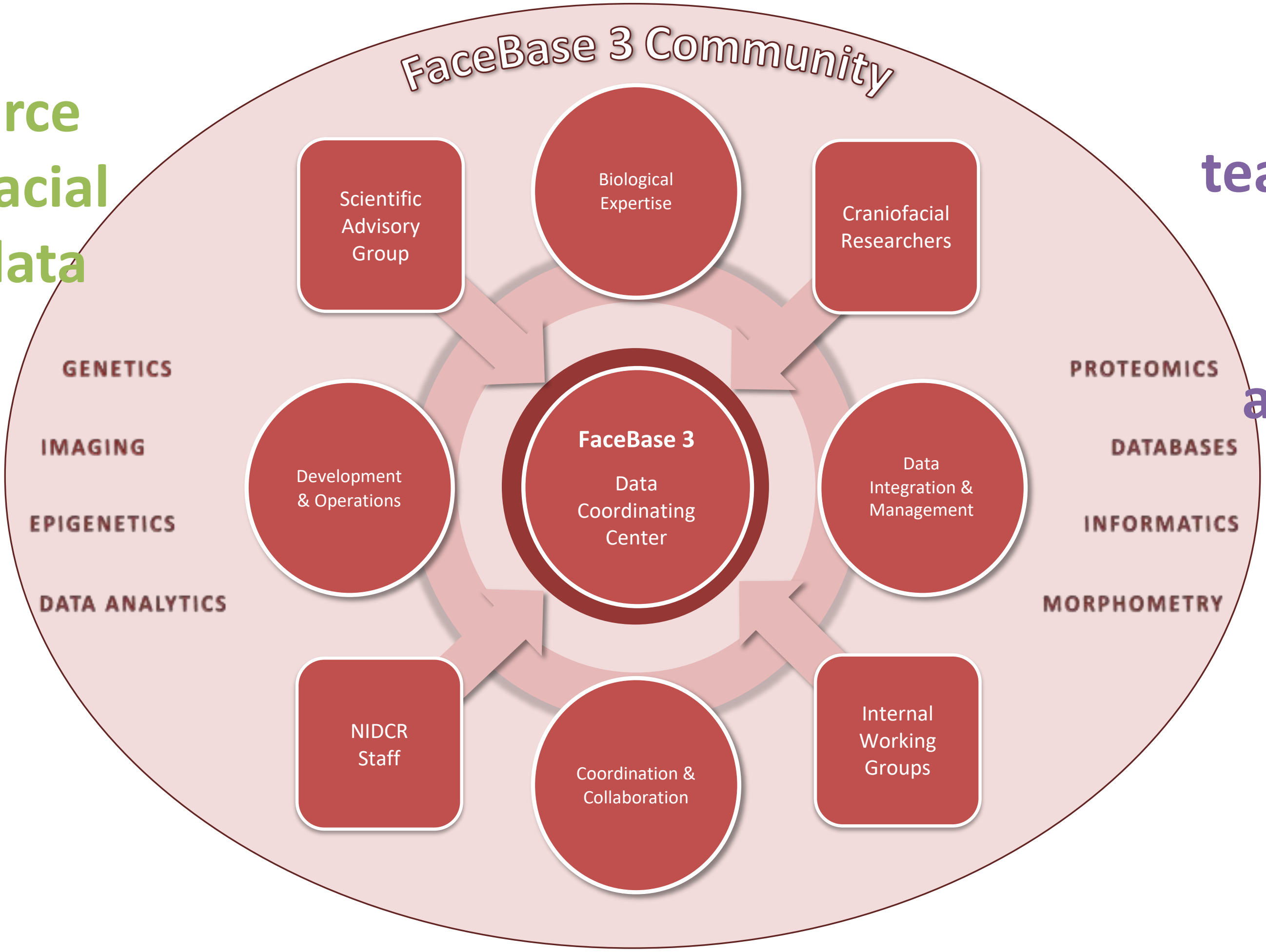
Robert Schuler

43rd Annual SCGDB Meeting

October 20, 2020

An Open Platform for Craniofacial and Dental Research

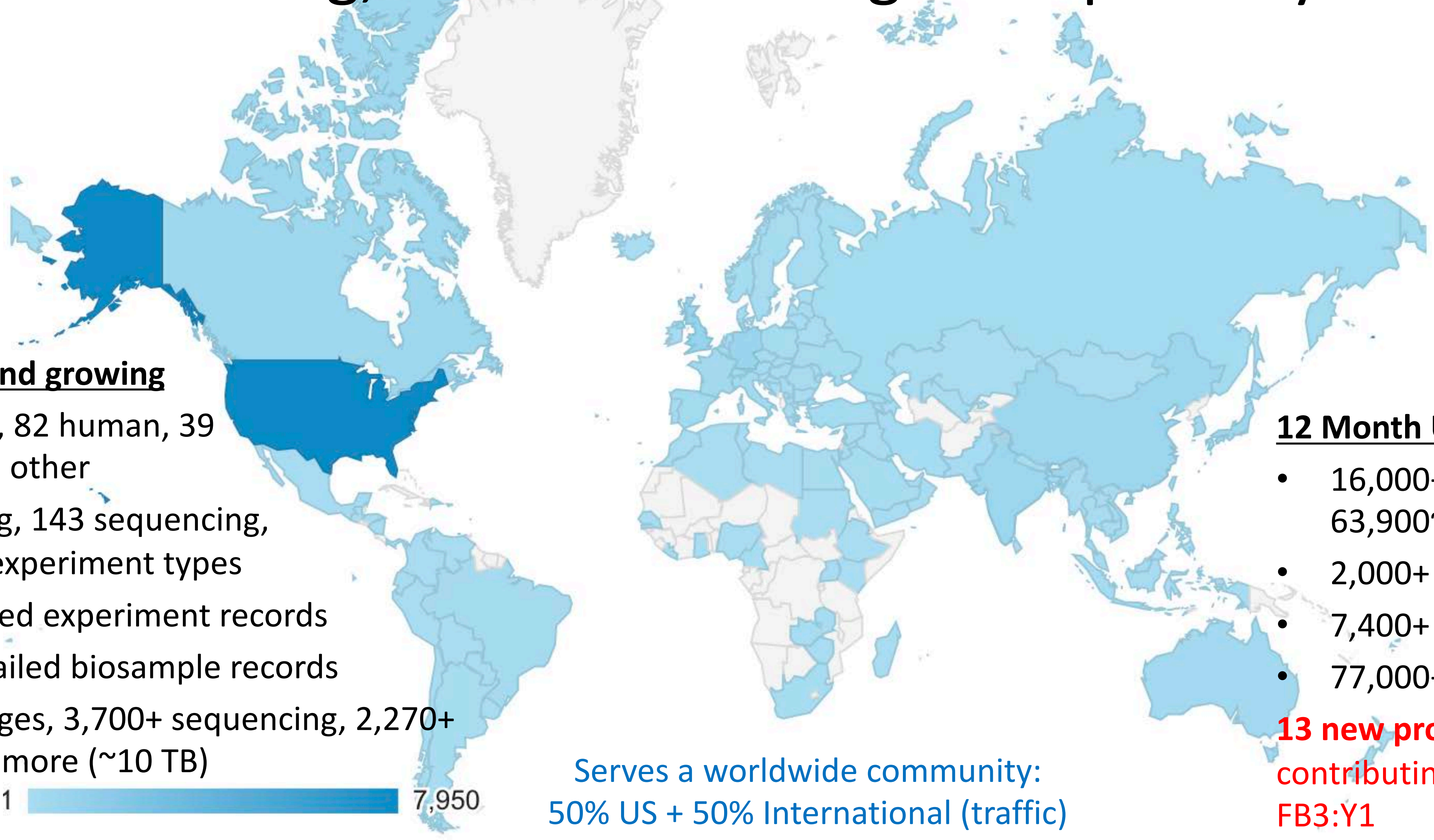
Comprehensive resource on dental and craniofacial research -- fostering data sharing and curated resources.



Comprised of a core team of craniofacial and computer science expertise w/ advisory and experts groups for oversight.

Transitioned from a "hub and spoke" consortium to a community-wide effort in 2019.

Growing, Multi-faceted Digital Repository



942 Datasets and growing

- 820 mouse, 82 human, 39 zebrafish, 3 other
- 634 imaging, 143 sequencing, and other experiment types
- 800+ detailed experiment records
- 3,500~ detailed biosample records
- 5,000+ images, 3,700+ sequencing, 2,270+ tracks, and more (~10 TB)

Serves a worldwide community:
50% US + 50% International (traffic)

12 Month Usage

- 16,000+ visitors, 63,900~ views
- 2,000+ downloads
- 7,400+ image views
- 77,000+ track reads

13 new projects contributing data from FB3:Y1

Mission to Improve the Stewardship of Research Data

Science is dependent on generating and analyzing data, but…

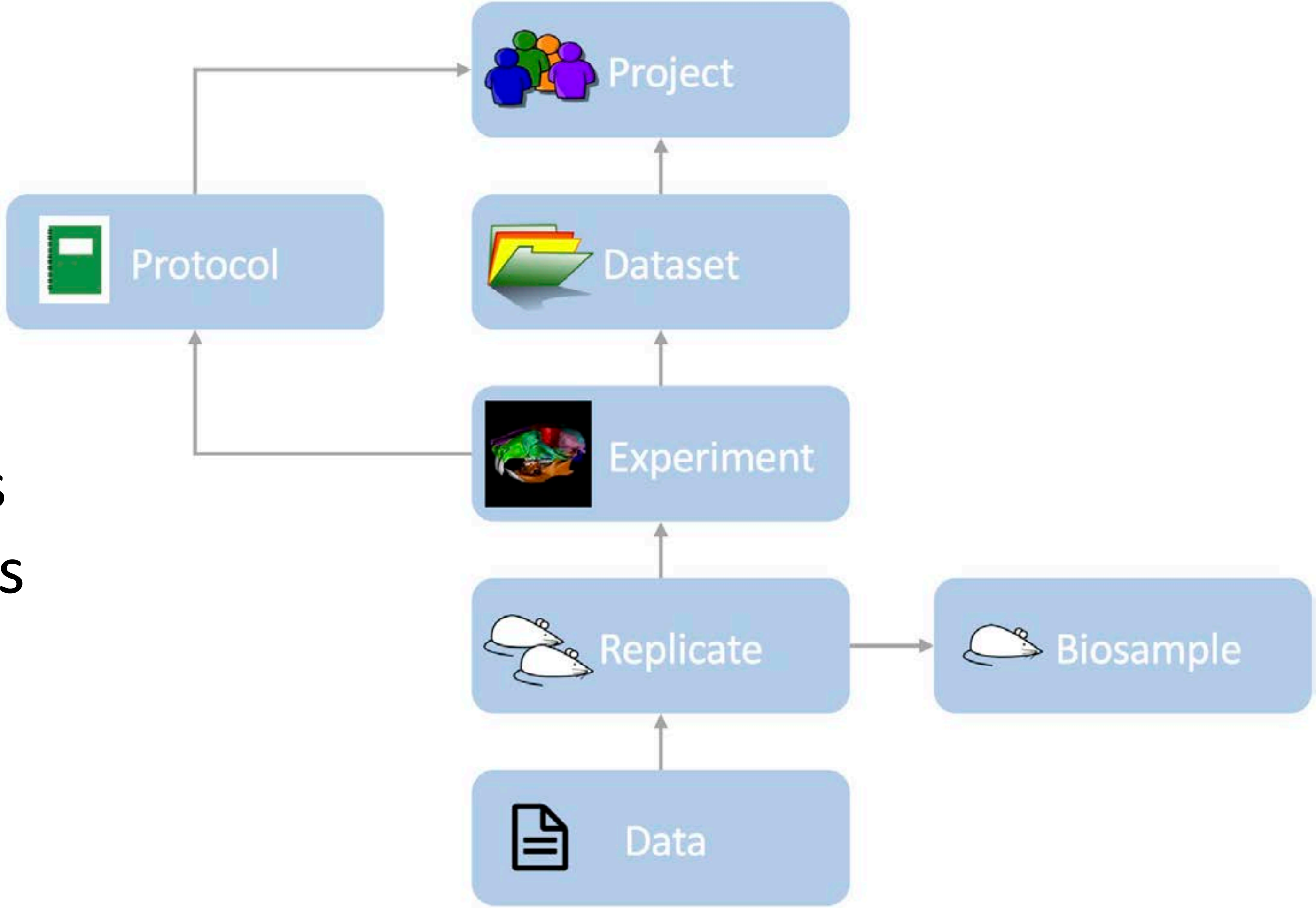
- 80% of time spent on accessing, cleaning, integrating data
- Scarcity of data sharing
- 10% reproducibility of data
- Recent high-profile retractions in COVID-19 research, for example.

Findable ***A***ccessible ***I***nteroperable ***R***eusable Data

- **Organization** of data files to understand what and how they were produced
- **Terminology** from community standards with broad acceptance
- **Description** of the experimental methods and biological materials
- **Protocols** that are enumerated or referenced for precise details

Organization of Data on FaceBase

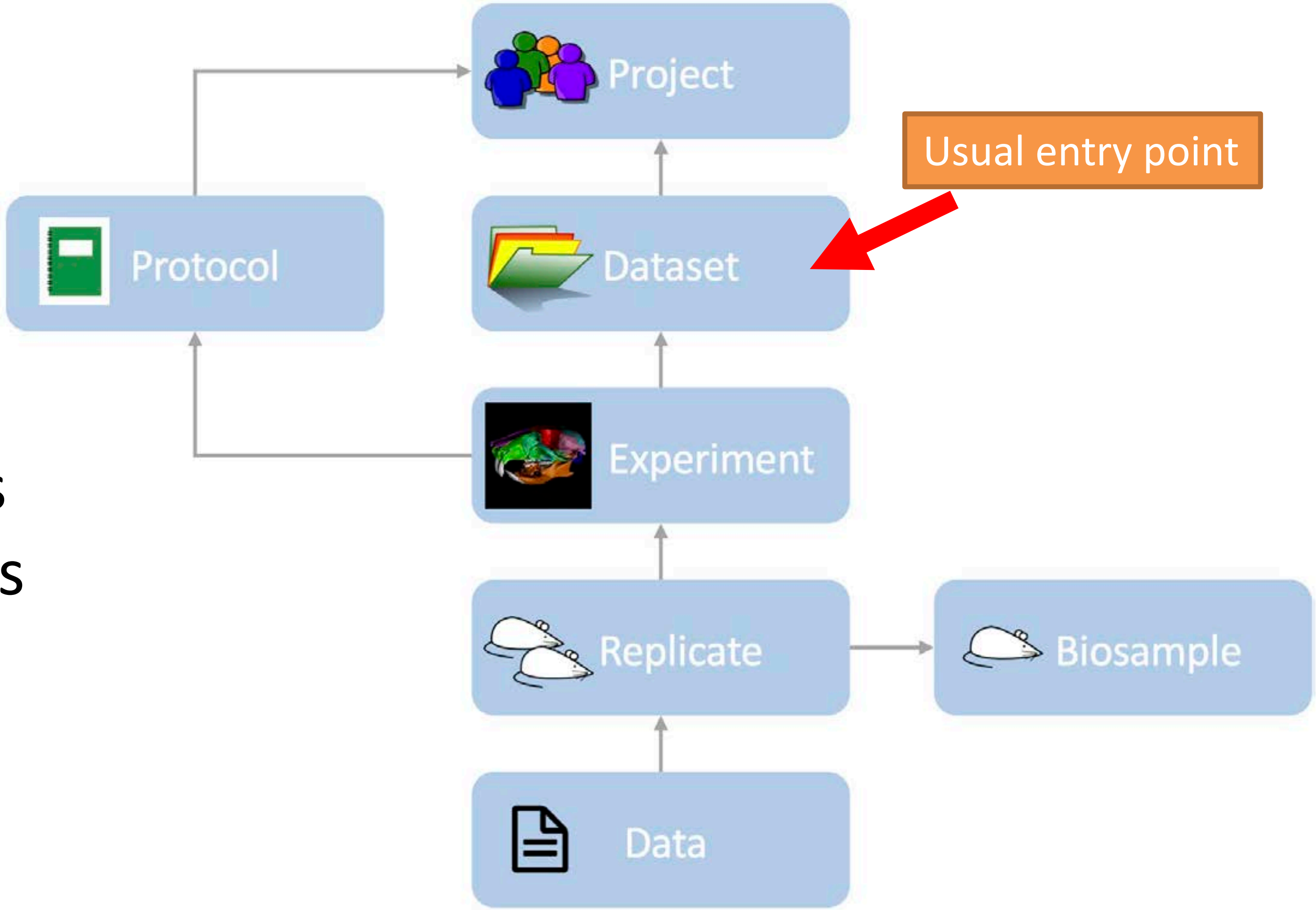
- **Projects** contribute **Datasets**
- **Datasets** organize **Experiments**, **Biosamples**, and **Data**
- **Protocols** give additional details on the materials and procedures conducted by Experiments



FaceBase data model (high-level view)

Organization of Data on FaceBase

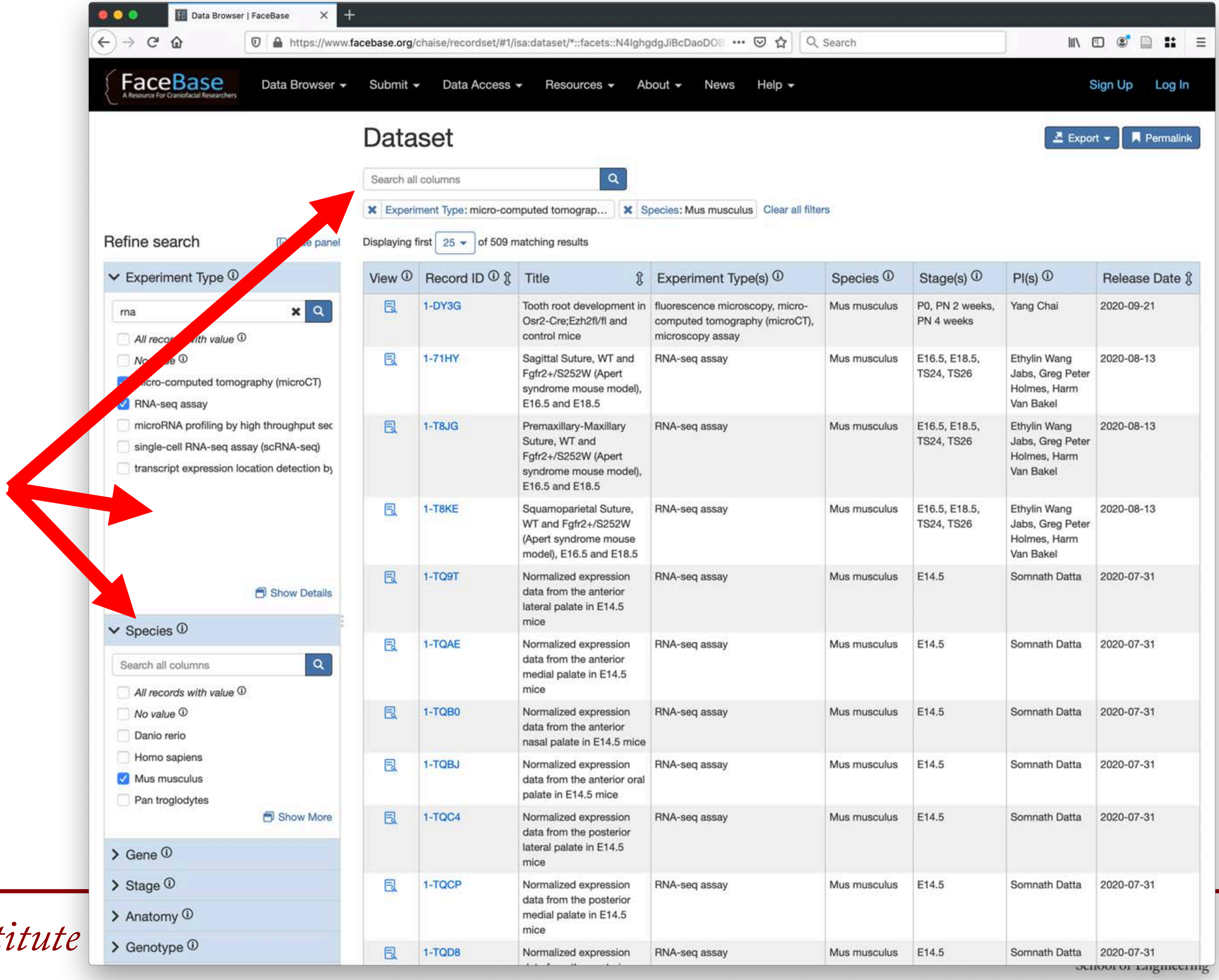
- **Projects** contribute **Datasets**
- **Datasets** organize **Experiments**, **Biosamples**, and **Data**
- **Protocols** give additional details on the materials and procedures conducted by Experiments
- *Search generally begins at the Dataset*



FaceBase data model (high-level view)

Search and Refine

- Refinements by experiment type, species, gene, age, anatomy, etc.



Dataset

Search all columns

✖ Experiment Type: micro-computed tomograp... ✖ Species: Mus musculus Clear all filters

Refine search

Experiment Type

rna

☐ *All records with value*
☐ *No value*
☑ micro-computed tomography (microCT)
☑ RNA-seq assay
☐ microRNA profiling by high throughput sec
☐ single-cell RNA-seq assay (scRNA-seq)
☐ transcript expression location detection by

Show Details

Species

Search all columns

☐ *All records with value*
☐ *No value*
☐ Danio rerio
☐ Homo sapiens
☑ Mus musculus
☐ Pan troglodytes

Show More

Gene

Stage

Anatomy

Genotype

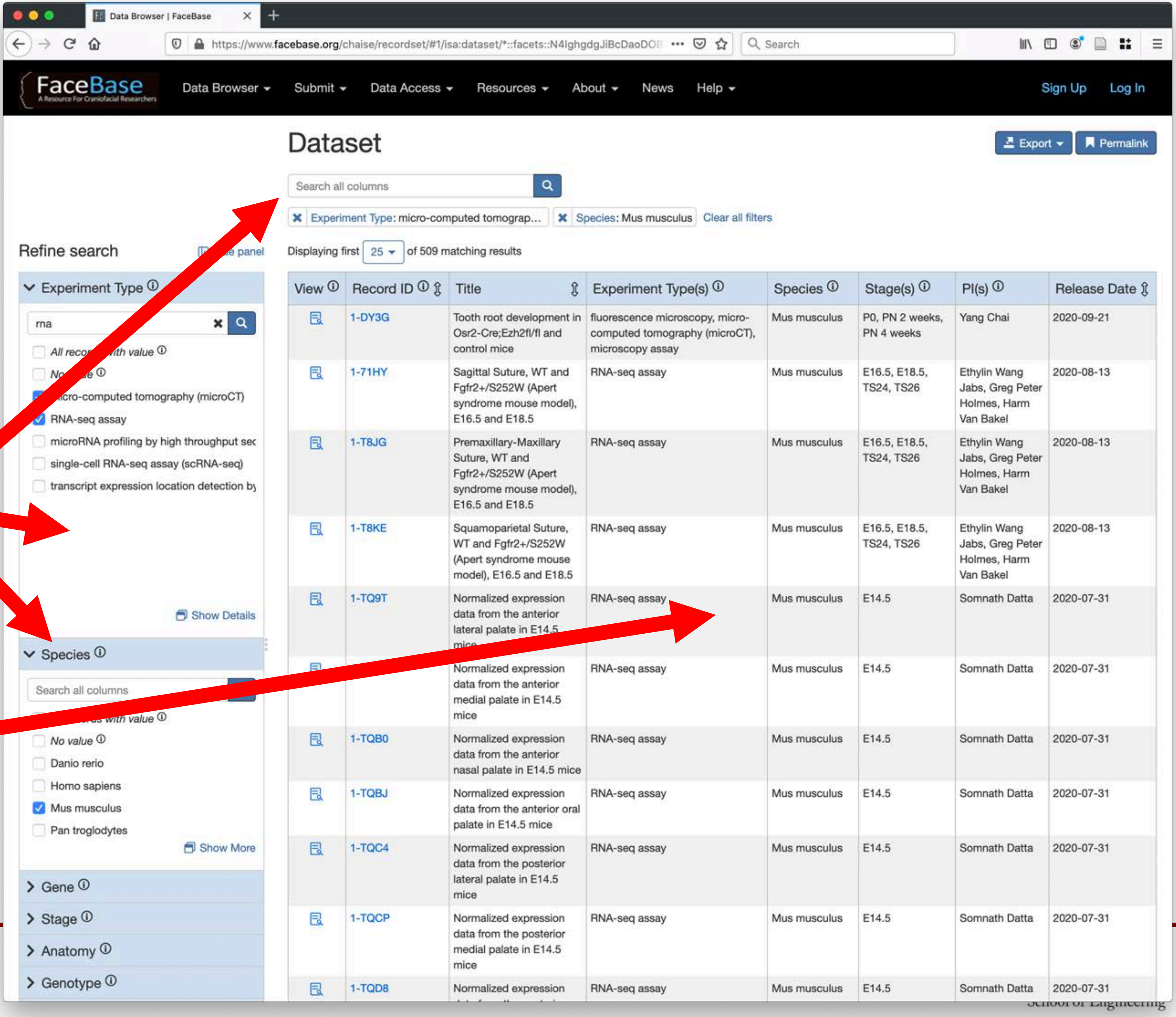
Displaying first 25 of 509 matching results

View	Record ID	Title	Experiment Type(s)	Species	Stage(s)	PI(s)	Release Date
	1-DY3G	Tooth root development in Osr2-Cre;Ezh2fl/fl and control mice	fluorescence microscopy, micro-computed tomography (microCT), microscopy assay	Mus musculus	P0, PN 2 weeks, PN 4 weeks	Yang Chai	2020-09-21
	1-71HY	Sagittal Suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5	RNA-seq assay	Mus musculus	E16.5, E18.5, TS24, TS26	Ethylin Wang Jabs, Greg Peter Holmes, Harm Van Bakel	2020-08-13
	1-T8JG	Premaxillary-Maxillary Suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5	RNA-seq assay	Mus musculus	E16.5, E18.5, TS24, TS26	Ethylin Wang Jabs, Greg Peter Holmes, Harm Van Bakel	2020-08-13
	1-T8KE	Squamoparietal Suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5	RNA-seq assay	Mus musculus	E16.5, E18.5, TS24, TS26	Ethylin Wang Jabs, Greg Peter Holmes, Harm Van Bakel	2020-08-13
	1-TQ9T	Normalized expression data from the anterior lateral palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQAE	Normalized expression data from the anterior medial palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQB0	Normalized expression data from the anterior nasal palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQBJ	Normalized expression data from the anterior oral palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQC4	Normalized expression data from the posterior lateral palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQCP	Normalized expression data from the posterior medial palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQD8	Normalized expression	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31

Search and Refine

- Refinements by experiment type, species, gene, age, anatomy, etc.

- Summary view with key properties listed.



Dataset

Search all columns

✖ Experiment Type: micro-computed tomograp... ✖ Species: Mus musculus Clear all filters

Refine search

Experiment Type

rna

☐ All reco with value
☐ No e
☑ icro-computed tomography (microCT)
☑ RNA-seq assay
☐ microRNA profiling by high throughput sec
☐ single-cell RNA-seq assay (scRNA-seq)
☐ transcript expression location detection by

Show Details

Species

Search all columns

☐ with value
☐ No value
☐ Danio rerio
☐ Homo sapiens
☑ Mus musculus
☐ Pan troglodytes

Show More

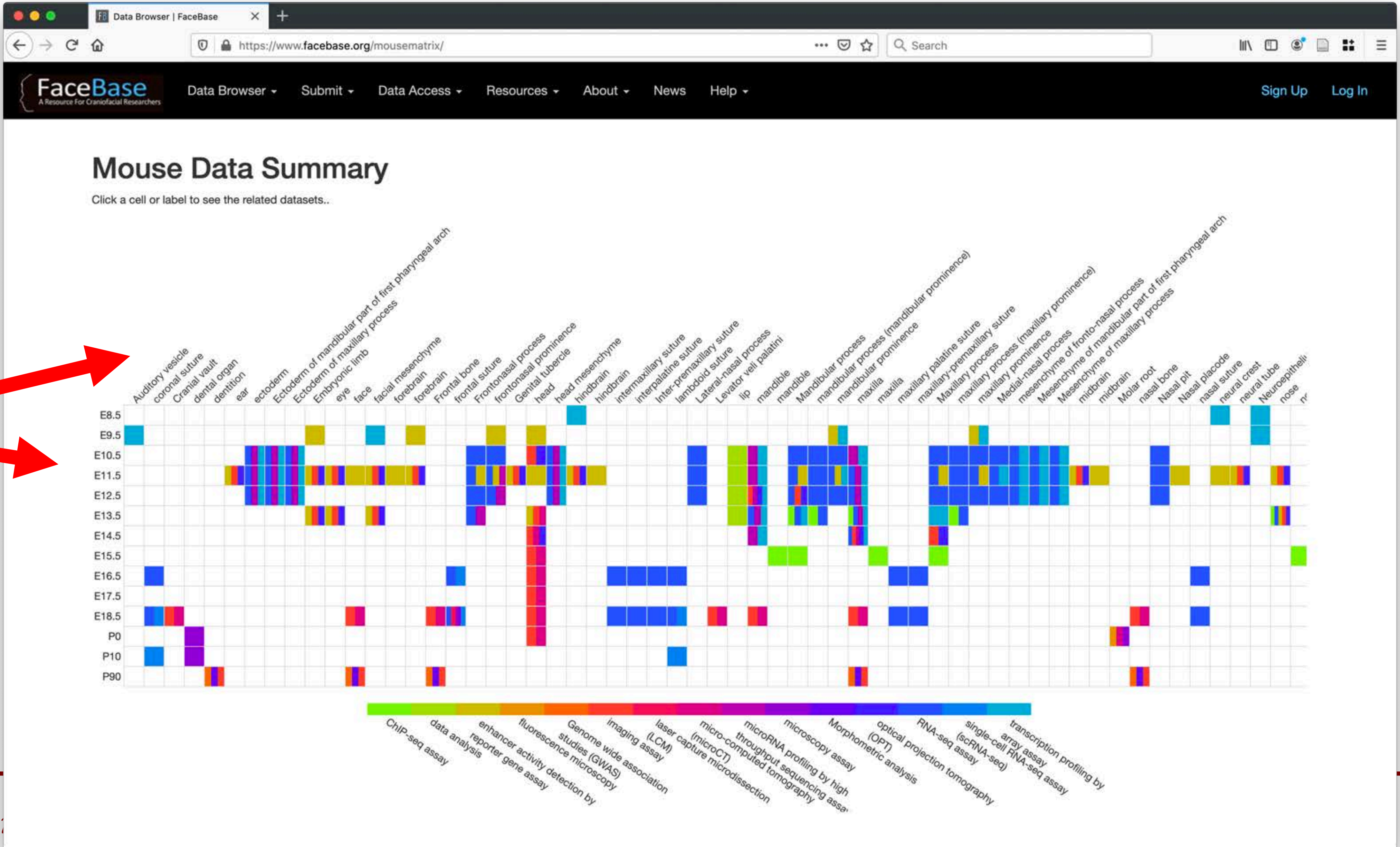
Gene
Stage
Anatomy
Genotype

Displaying first 25 of 509 matching results

View	Record ID	Title	Experiment Type(s)	Species	Stage(s)	PI(s)	Release Date
	1-DY3G	Tooth root development in Osr2-Cre;Ezh2fl/fl and control mice	fluorescence microscopy, micro-computed tomography (microCT), microscopy assay	Mus musculus	P0, PN 2 weeks, PN 4 weeks	Yang Chai	2020-09-21
	1-71HY	Sagittal Suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5	RNA-seq assay	Mus musculus	E16.5, E18.5, TS24, TS26	Ethylin Wang Jabs, Greg Peter Holmes, Harm Van Bakel	2020-08-13
	1-T8JG	Premaxillary-Maxillary Suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5	RNA-seq assay	Mus musculus	E16.5, E18.5, TS24, TS26	Ethylin Wang Jabs, Greg Peter Holmes, Harm Van Bakel	2020-08-13
	1-T8KE	Squamoparietal Suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5	RNA-seq assay	Mus musculus	E16.5, E18.5, TS24, TS26	Ethylin Wang Jabs, Greg Peter Holmes, Harm Van Bakel	2020-08-13
	1-TQ9T	Normalized expression data from the anterior lateral palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
		Normalized expression data from the anterior medial palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQB0	Normalized expression data from the anterior nasal palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQBJ	Normalized expression data from the anterior oral palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQC4	Normalized expression data from the posterior lateral palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQCP	Normalized expression data from the posterior medial palate in E14.5 mice	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31
	1-TQD8	Normalized expression	RNA-seq assay	Mus musculus	E14.5	Somnath Datta	2020-07-31

Summary of Mouse Datasets

Headings and Cells link directly to the Search interface

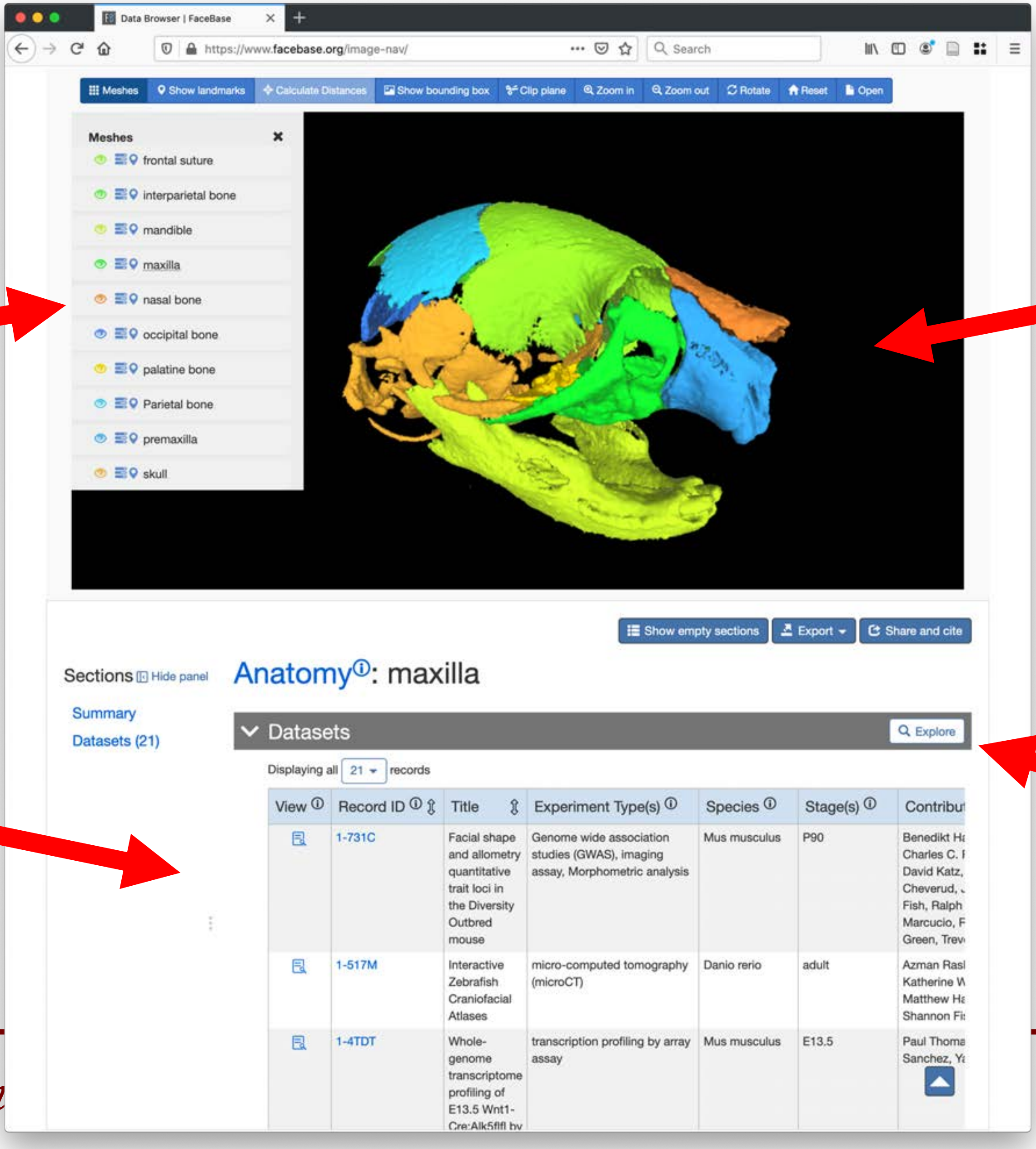


Mouse Data Summary

Click a cell or label to see the related datasets..

Explore By Anatomical Region of Interest

Linked to standard terminology



Meshes

- frontal suture
- interparietal bone
- mandible
- maxilla
- nasal bone
- occipital bone
- palatine bone
- Parietal bone
- premaxilla
- skull

Sections Hide panel

Summary

Datasets (21)

Anatomy: maxilla

Datasets

Displaying all 21 records

View	Record ID	Title	Experiment Type(s)	Species	Stage(s)	Contribu
	1-731C	Facial shape and allometry quantitative trait loci in the Diversity Outbred mouse	Genome wide association studies (GWAS), imaging assay, Morphometric analysis	Mus musculus	P90	Benedikt Ha Charles C. David Katz, Cheverud, Fish, Ralph Marcucio, Green, Trev
	1-517M	Interactive Zebrafish Craniofacial Atlases	micro-computed tomography (microCT)	Danio rerio	adult	Azman Rasl Katherine W Matthew Ha Shannon Fi
	1-4TDT	Whole-genome transcriptome profiling of E13.5 Wnt1-Cre:Alk5fl/fl by	transcription profiling by array assay	Mus musculus	E13.5	Paul Thoma Sanchez, Ya

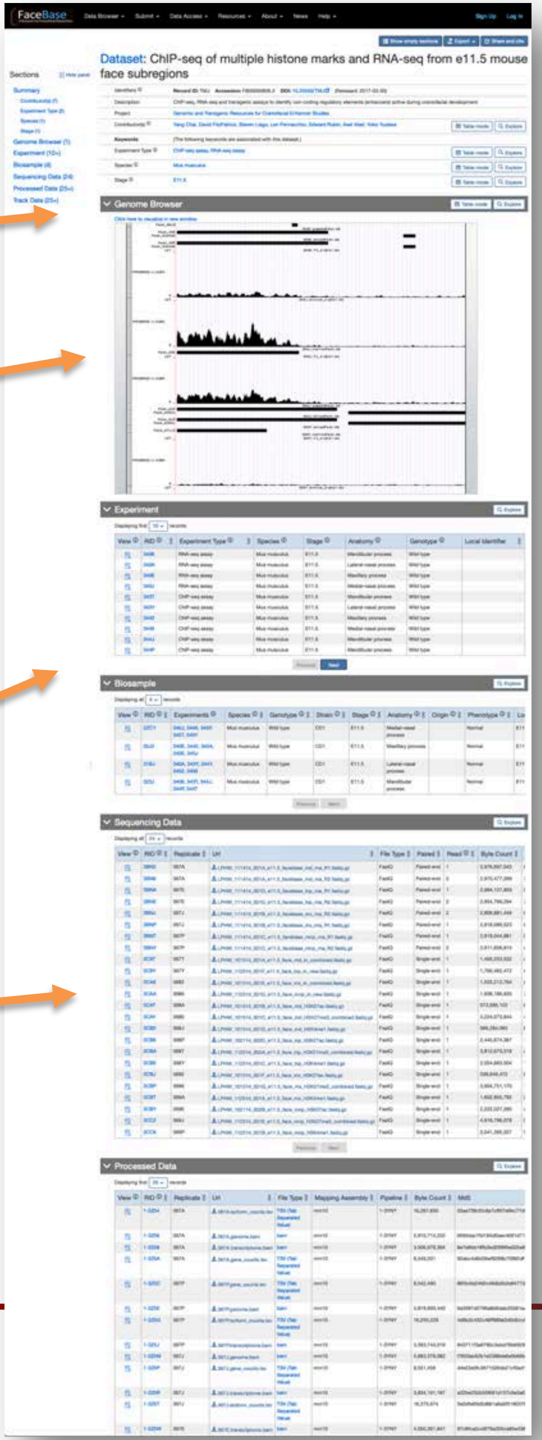
Surface mesh viewer with each structure coded to the terminology

Links above drive search results of related datasets

“Explore” to go deeper into the search and refine actions

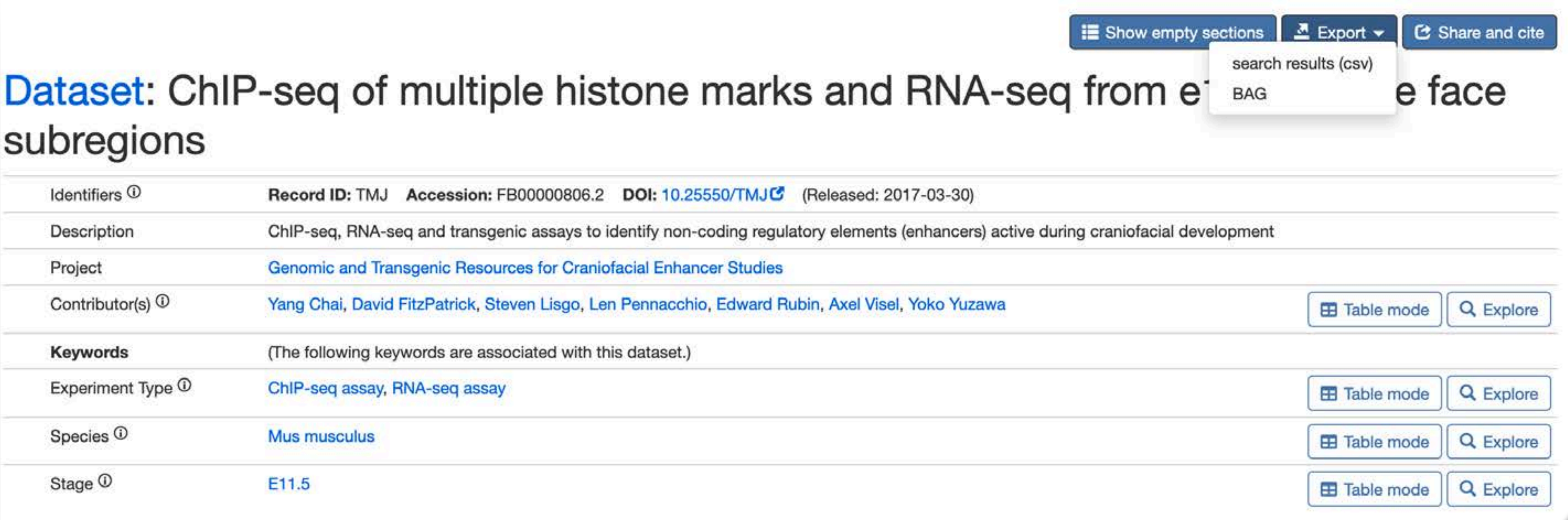
Overview of Dataset Display

- Summary
- Visualization
  - Genome Browser
  - Surface Models
  - Orthoslice Views
  - Images
- Experiments & Biosamples
- Data Files
  - Sequencing, Processed, Imaging, Array, Supplementary, other…



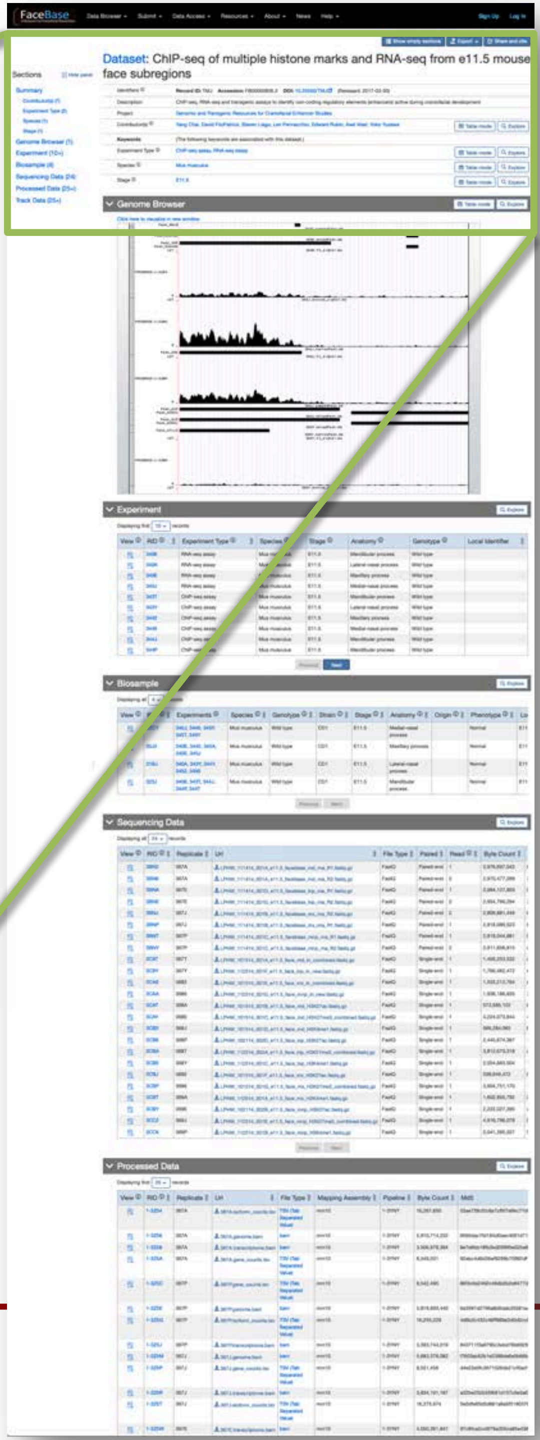
Dataset: ChIP-seq of multiple histone marks and RNA-seq from e11.5 mouse face subregions

Summary of the Dataset Contents



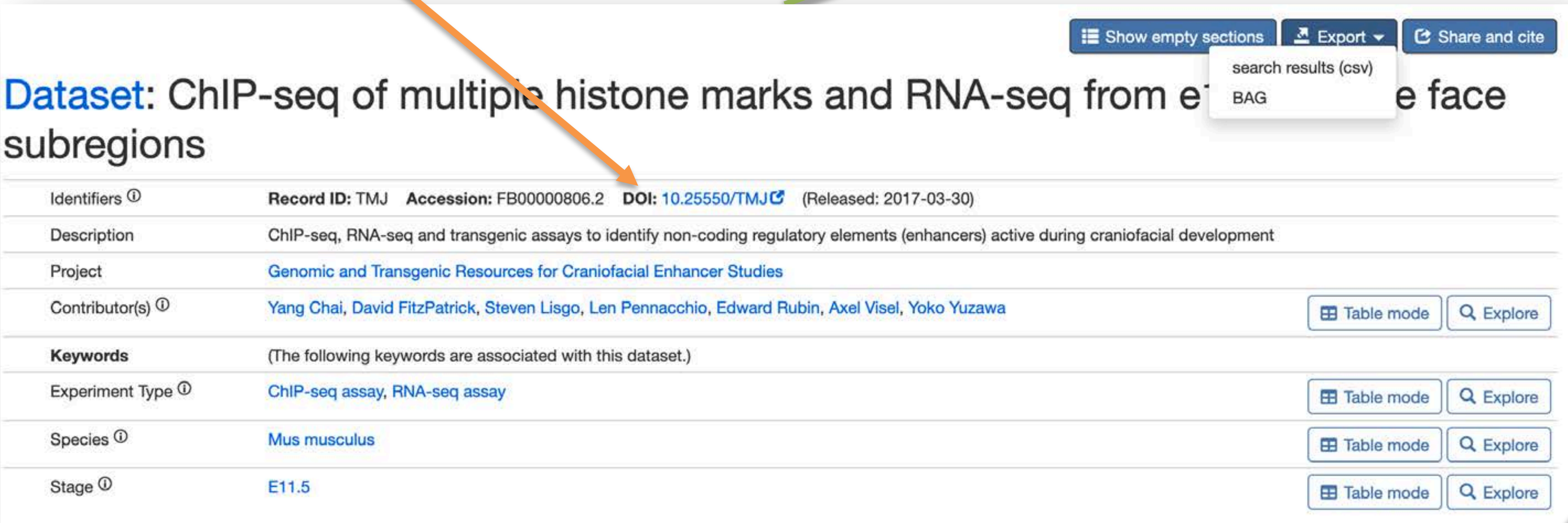
Dataset: ChIP-seq of multiple histone marks and RNA-seq from e1 e face subregions

Identifiers ⓘ	**Record ID:** TMJ **Accession:** FB00000806.2 **DOI:** 10.25550/TMJ (Released: 2017-03-30)
Description	ChIP-seq, RNA-seq and transgenic assays to identify non-coding regulatory elements (enhancers) active during craniofacial development
Project	Genomic and Transgenic Resources for Craniofacial Enhancer Studies
Contributor(s) ⓘ	Yang Chai, David FitzPatrick, Steven Lisgo, Len Pennacchio, Edward Rubin, Axel Visel, Yoko Yuzawa
Keywords	(The following keywords are associated with this dataset.)
Experiment Type ⓘ	ChIP-seq assay, RNA-seq assay
Species ⓘ	Mus musculus
Stage ⓘ	E11.5



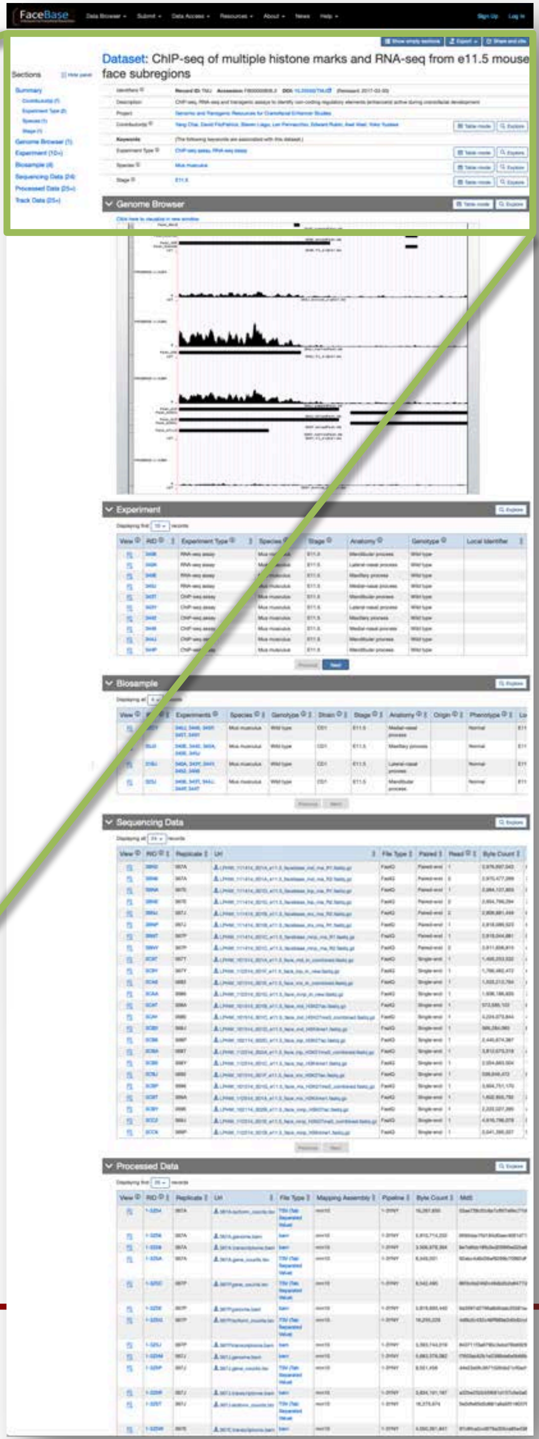
Summary of the Dataset Contents

Persistent Identifiers (e.g., DOI)



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e … e face subregions

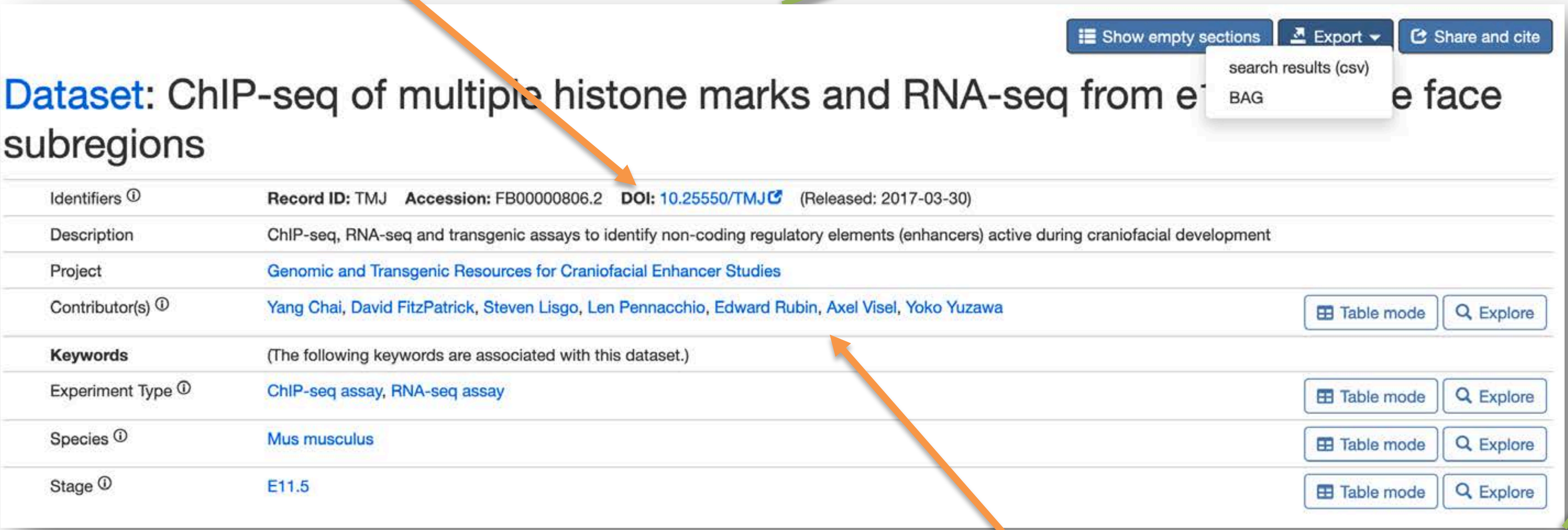
Identifiers ⓘ	**Record ID:** TMJ **Accession:** FB00000806.2 **DOI:** 10.25550/TMJ (Released: 2017-03-30)
Description	ChIP-seq, RNA-seq and transgenic assays to identify non-coding regulatory elements (enhancers) active during craniofacial development
Project	Genomic and Transgenic Resources for Craniofacial Enhancer Studies
Contributor(s) ⓘ	Yang Chai, David FitzPatrick, Steven Lisgo, Len Pennacchio, Edward Rubin, Axel Visel, Yoko Yuzawa
Keywords	(The following keywords are associated with this dataset.)
Experiment Type ⓘ	ChIP-seq assay, RNA-seq assay
Species ⓘ	Mus musculus
Stage ⓘ	E11.5



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e11.5 mouse face subregions

Summary of the Dataset Contents

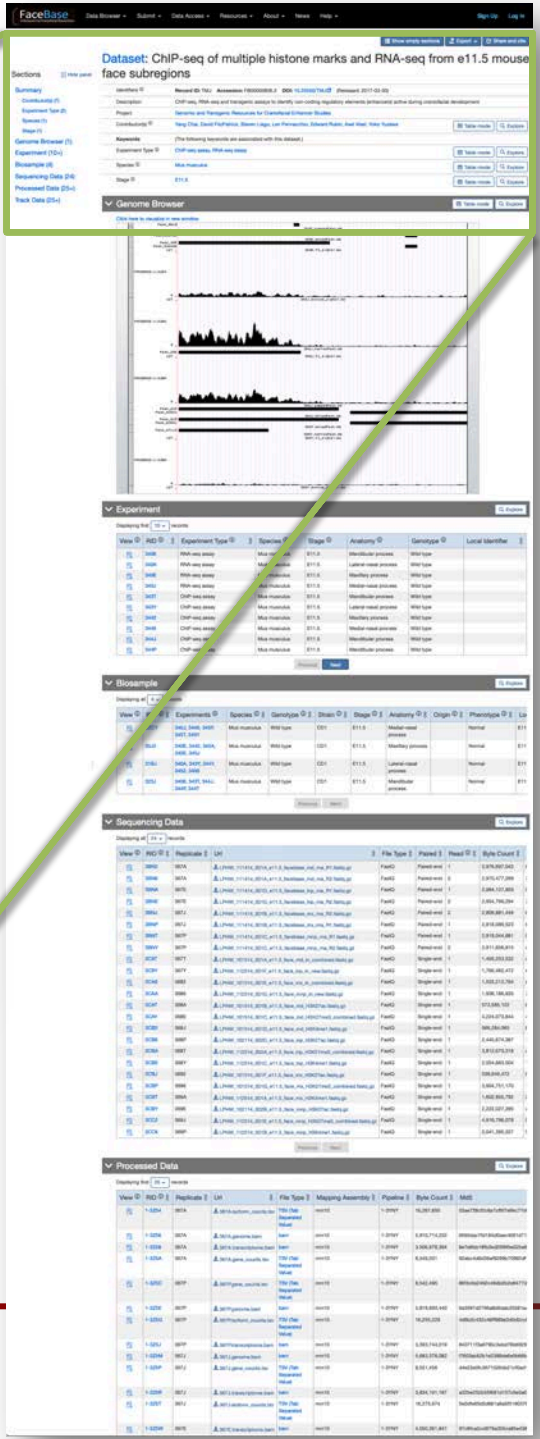
Persistent Identifiers (e.g., DOI)



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e e face subregions

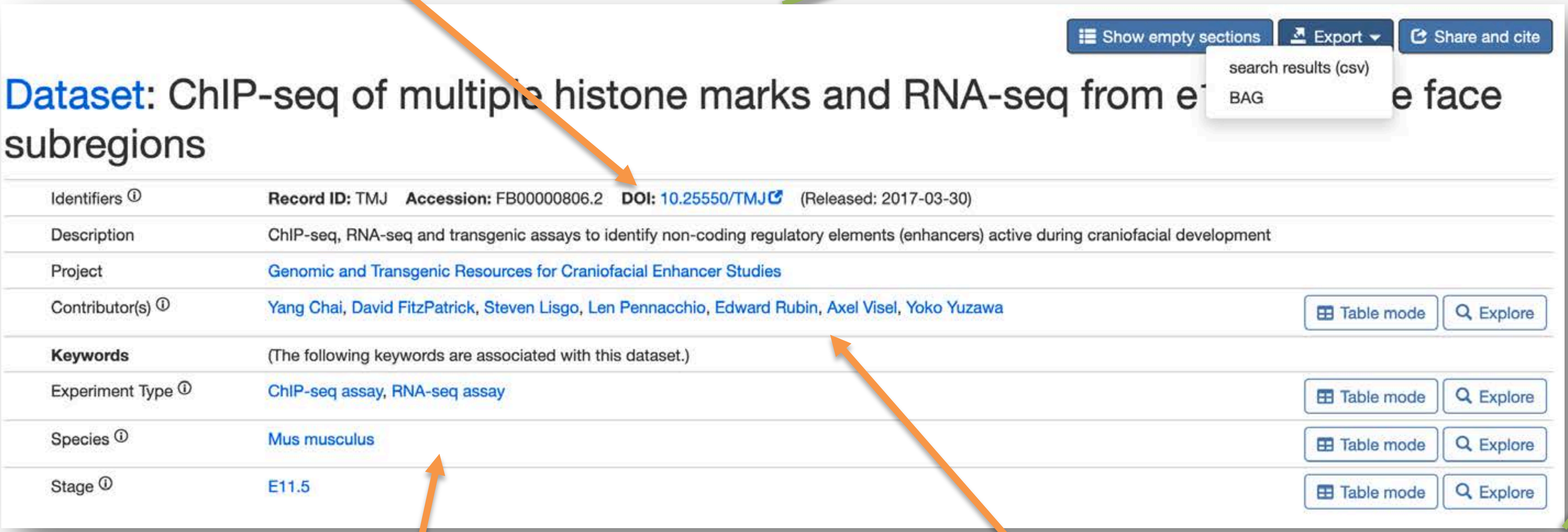
Identifiers ⓘ	**Record ID:** TMJ **Accession:** FB00000806.2 **DOI:** 10.25550/TMJ (Released: 2017-03-30)
Description	ChIP-seq, RNA-seq and transgenic assays to identify non-coding regulatory elements (enhancers) active during craniofacial development
Project	Genomic and Transgenic Resources for Craniofacial Enhancer Studies
Contributor(s) ⓘ	Yang Chai, David FitzPatrick, Steven Lisgo, Len Pennacchio, Edward Rubin, Axel Visel, Yoko Yuzawa
Keywords	(The following keywords are associated with this dataset.)
Experiment Type ⓘ	ChIP-seq assay, RNA-seq assay
Species ⓘ	Mus musculus
Stage ⓘ	E11.5

Attribution of contributors



Summary of the Dataset Contents

Persistent Identifiers (e.g., DOI)

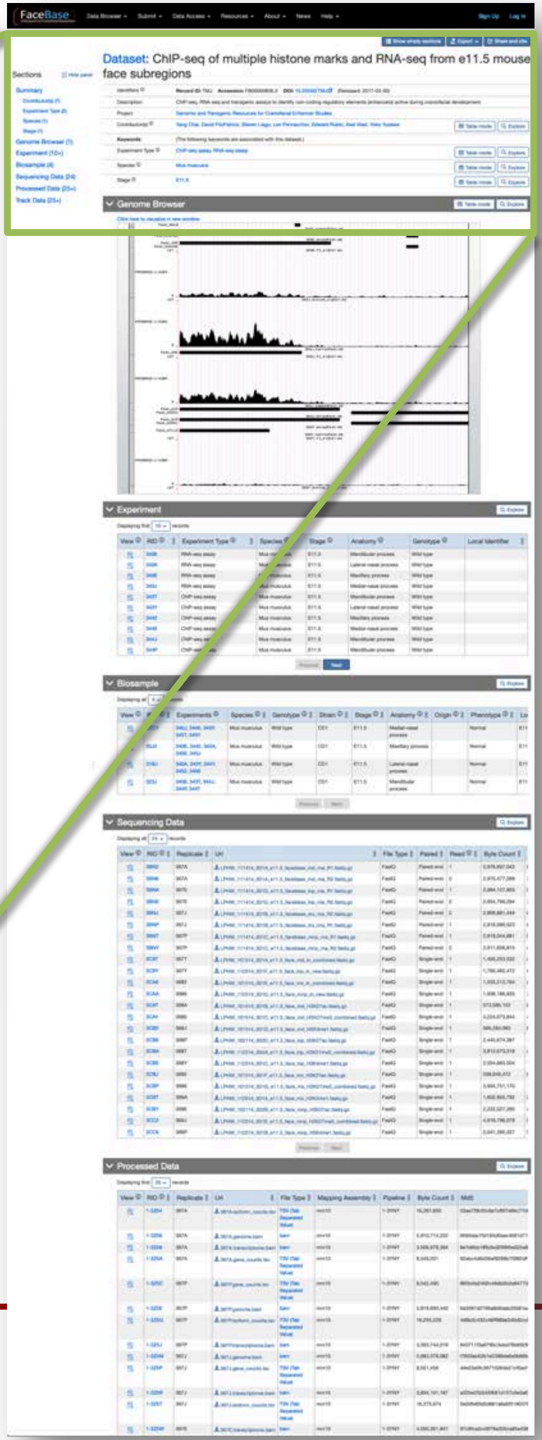


Dataset: ChIP-seq of multiple histone marks and RNA-seq from e e face subregions

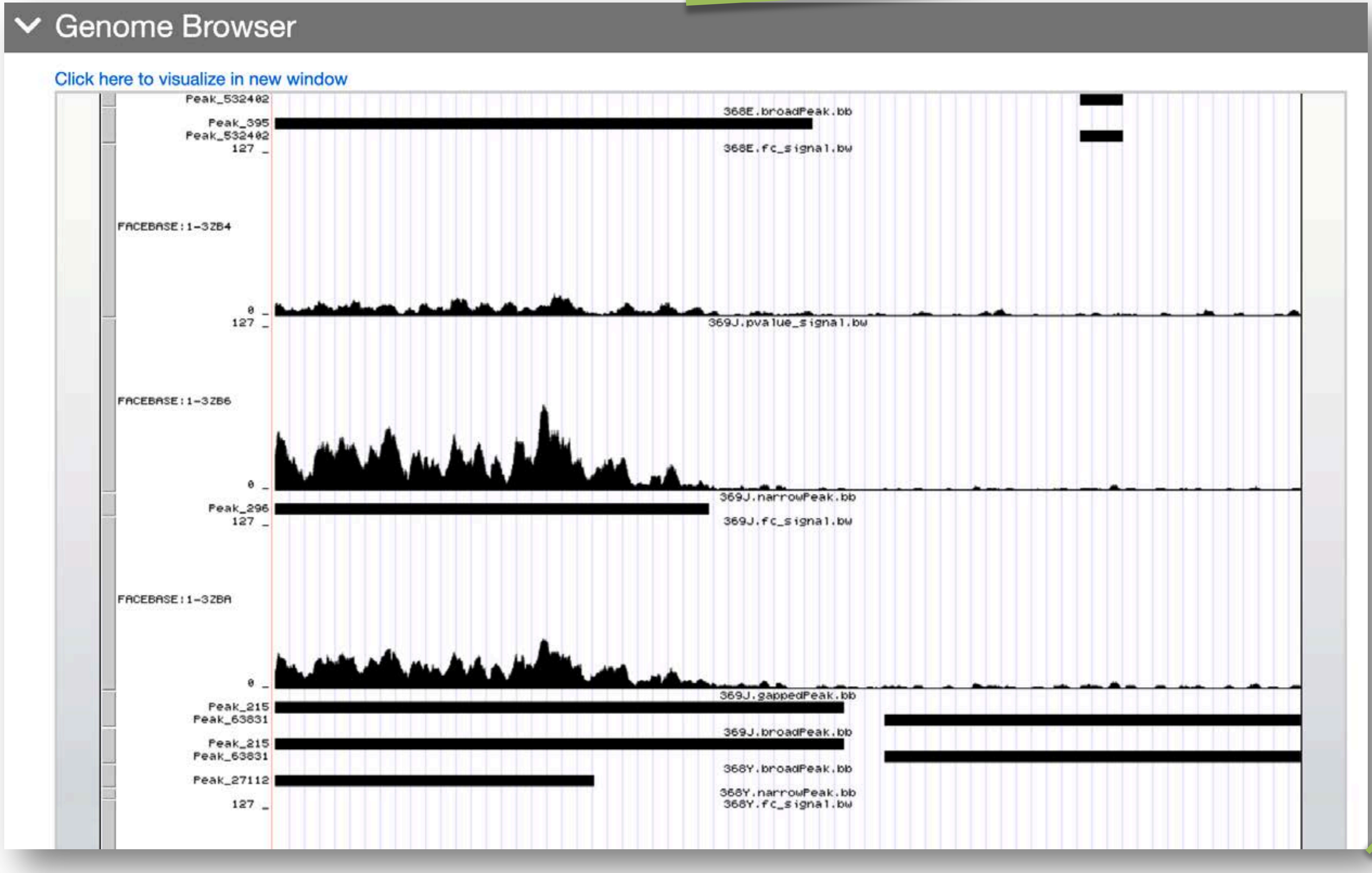
Identifiers ⓘ	**Record ID:** TMJ **Accession:** FB00000806.2 **DOI:** 10.25550/TMJ (Released: 2017-03-30)
Description	ChIP-seq, RNA-seq and transgenic assays to identify non-coding regulatory elements (enhancers) active during craniofacial development
Project	Genomic and Transgenic Resources for Craniofacial Enhancer Studies
Contributor(s) ⓘ	Yang Chai, David FitzPatrick, Steven Lisgo, Len Pennacchio, Edward Rubin, Axel Visel, Yoko Yuzawa
Keywords	(The following keywords are associated with this dataset.)
Experiment Type ⓘ	ChIP-seq assay, RNA-seq assay
Species ⓘ	Mus musculus
Stage ⓘ	E11.5

Keywords from standardized terminology

Attribution of contributors

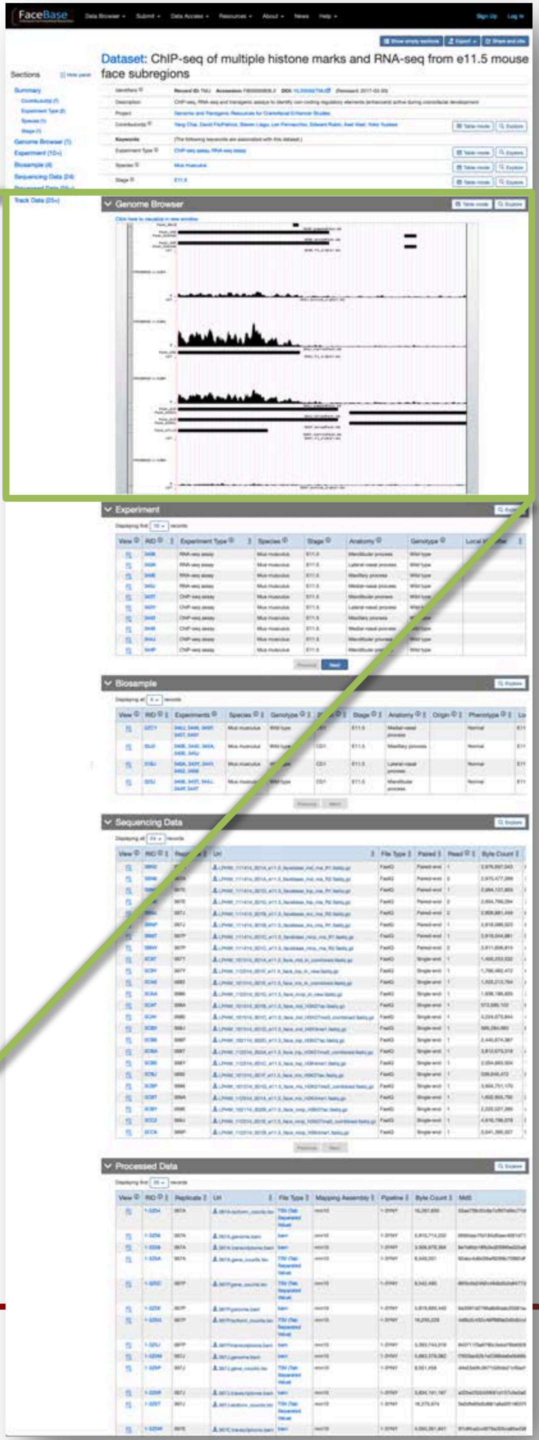


Embedded Visualization



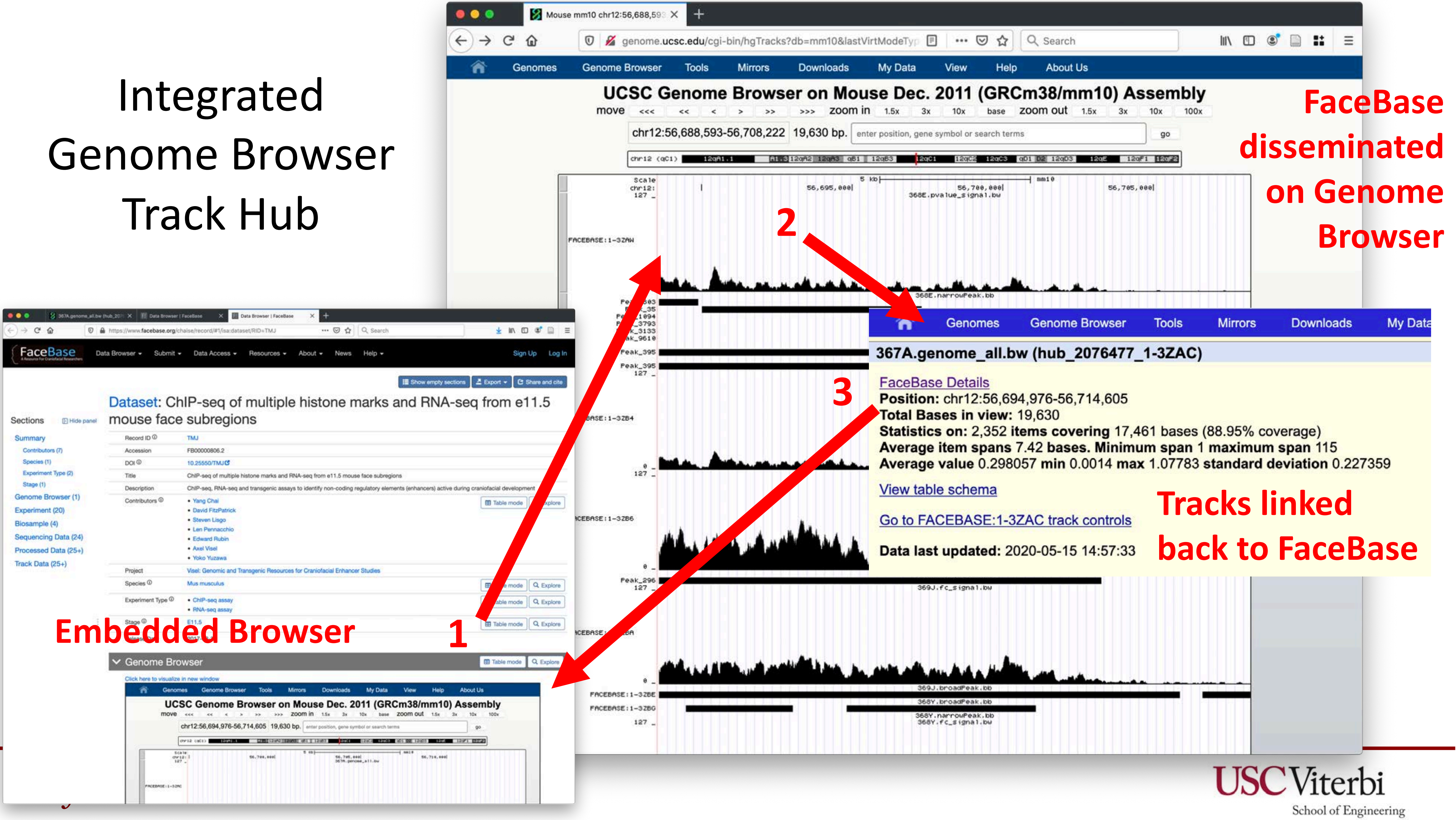
Genome Browser

Click here to visualize in new window



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e11.5 mouse face subregions

Integrated Genome Browser Track Hub



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e11.5 mouse face subregions

Record ID	TMJ
Accession	FB00000806.2
DOI	10.25550/TMJ
Title	ChIP-seq of multiple histone marks and RNA-seq from e11.5 mouse face subregions
Description	ChIP-seq, RNA-seq and transgenic assays to identify non-coding regulatory elements (enhancers) active during craniofacial development
Contributors	Yang Chai David FitzPatrick Steven Lisgo Len Pennacchio Edward Rubin Axel Visel Yoko Yuzawa
Project	Visel: Genomic and Transgenic Resources for Craniofacial Enhancer Studies
Species	Mus musculus
Experiment Type	ChIP-seq assay RNA-seq assay
Stage	E11.5

367A.genome_all.bw (hub_2076477_1-3ZAC)

FaceBase Details

Position: chr12:56,694,976-56,714,605
Total Bases in view: 19,630
Statistics on: 2,352 **items covering** 17,461 bases (88.95% coverage)
Average item spans 7.42 **bases. Minimum span** 1 **maximum span** 115
Average value 0.298057 **min** 0.0014 **max** 1.07783 **standard deviation** 0.227359

View table schema

Go to FACEBASE:1-3ZAC track controls

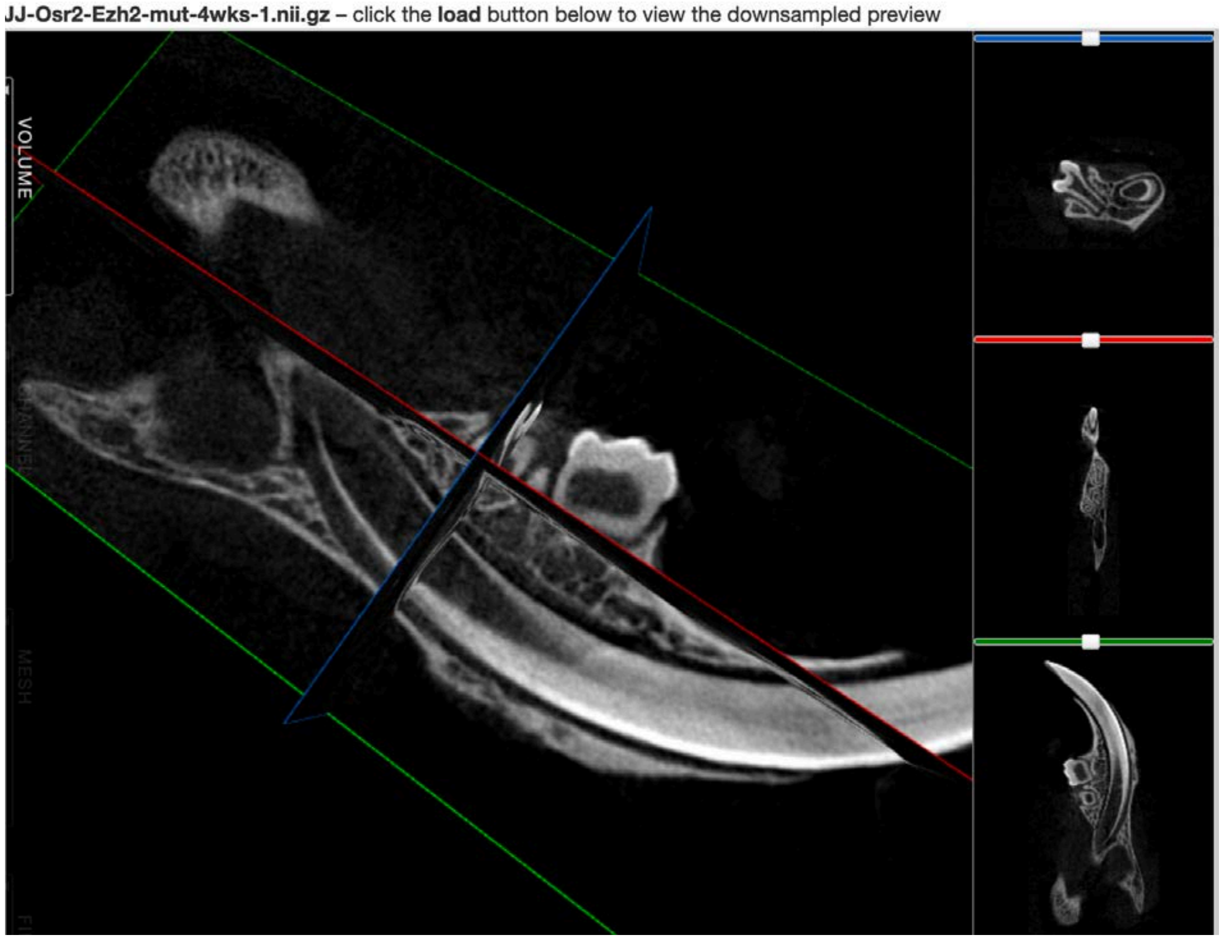
Data last updated: 2020-05-15 14:57:33

FaceBase disseminated on Genome Browser

Embedded Browser

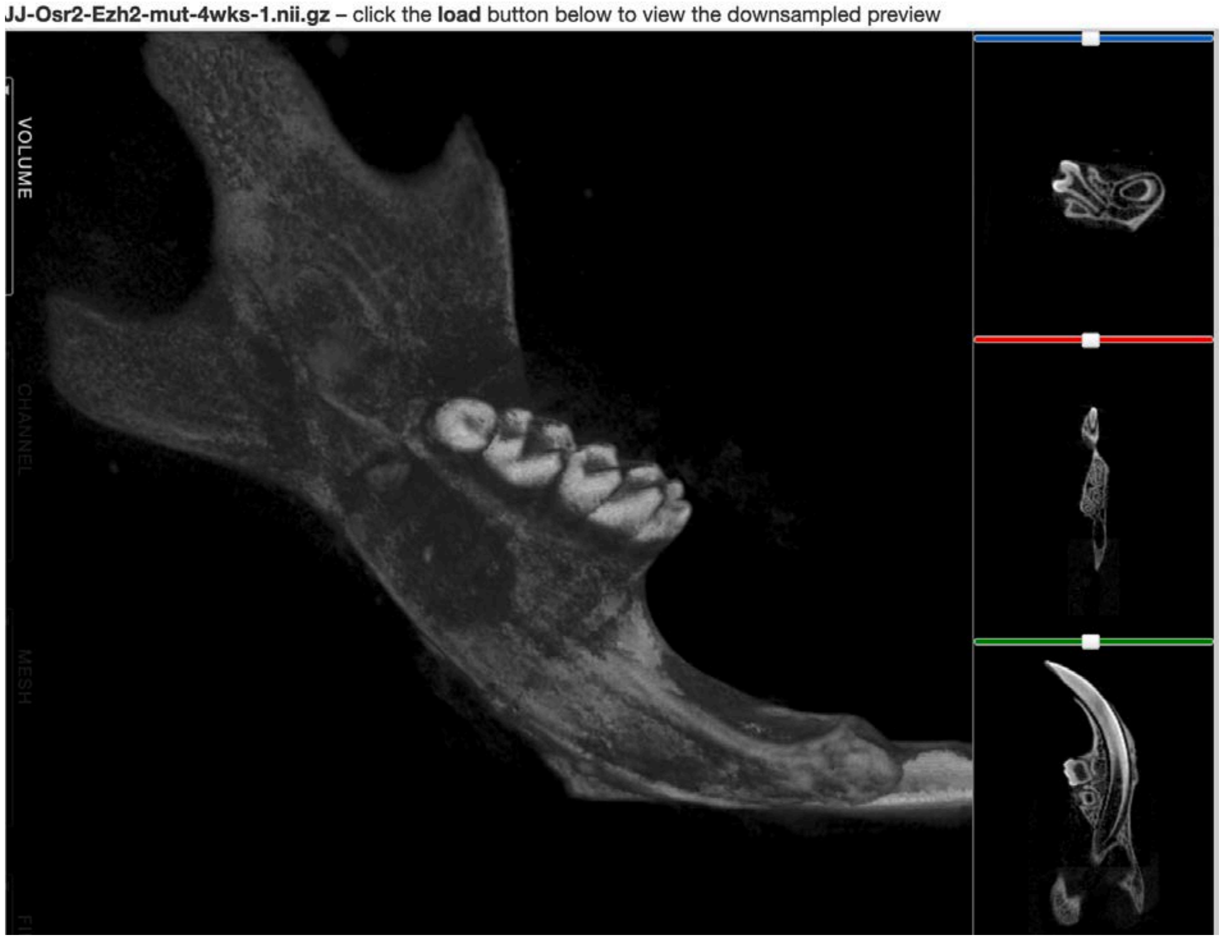
Tracks linked back to FaceBase

Orthogonal Slice Viewer with Volume Rendering



JJ-Osr2-Ezh2-mut-4wks-1.nii.gz – click the **load** button below to view the downsampled preview

Orthoslice View



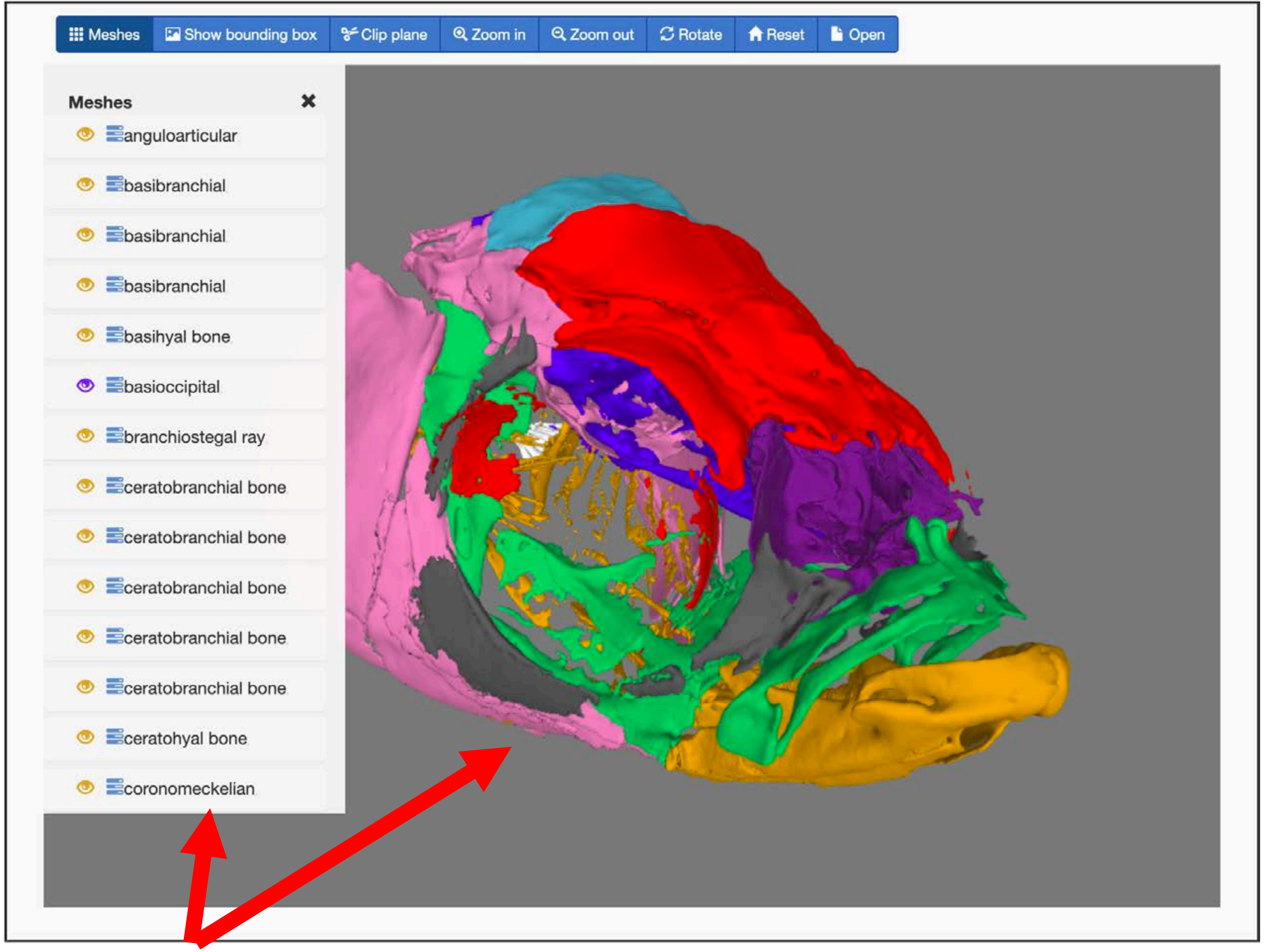
JJ-Osr2-Ezh2-mut-4wks-1.nii.gz – click the **load** button below to view the downsampled preview

3D Volume View

Web-based Viewer

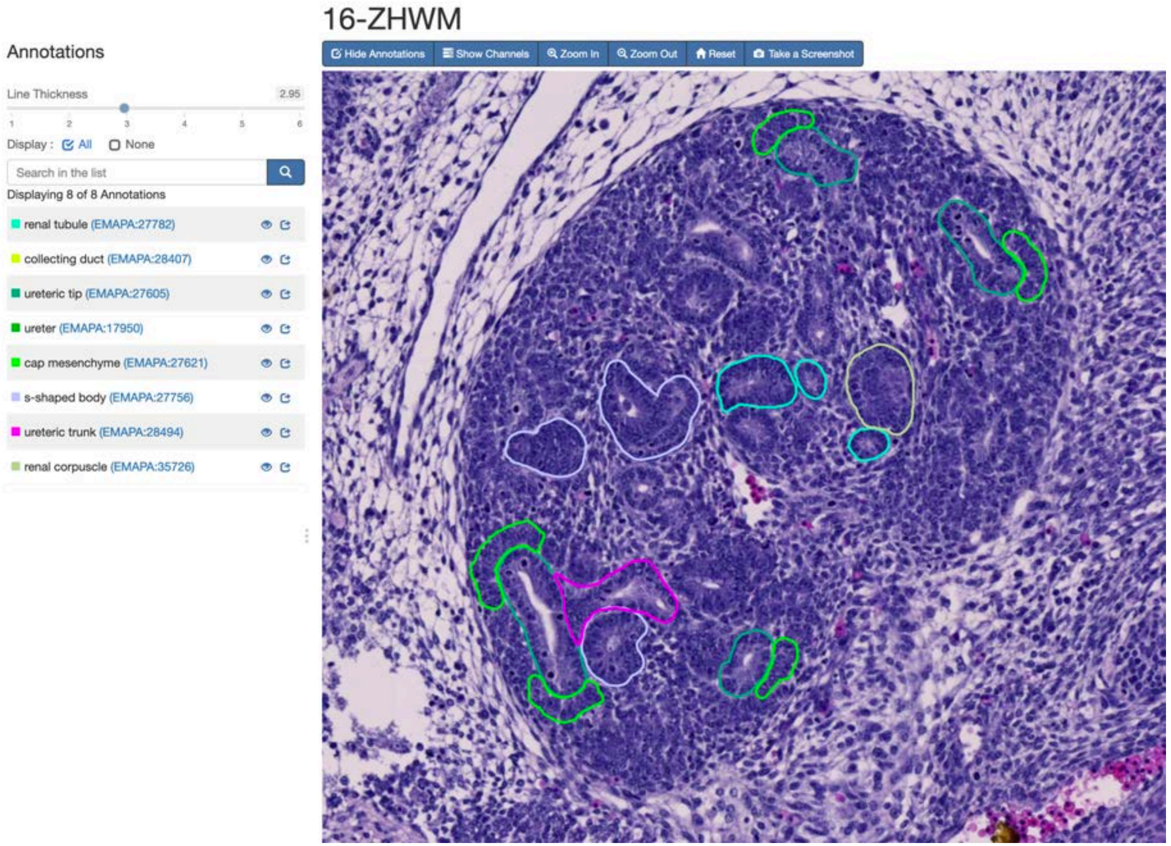
Highly Detailed Surface Models

- Surface models may contain any number of anatomical structures
- Linked to underlying (meta)data
- Linked to standard vocabulary (ZFA, MA, UBERON,…)

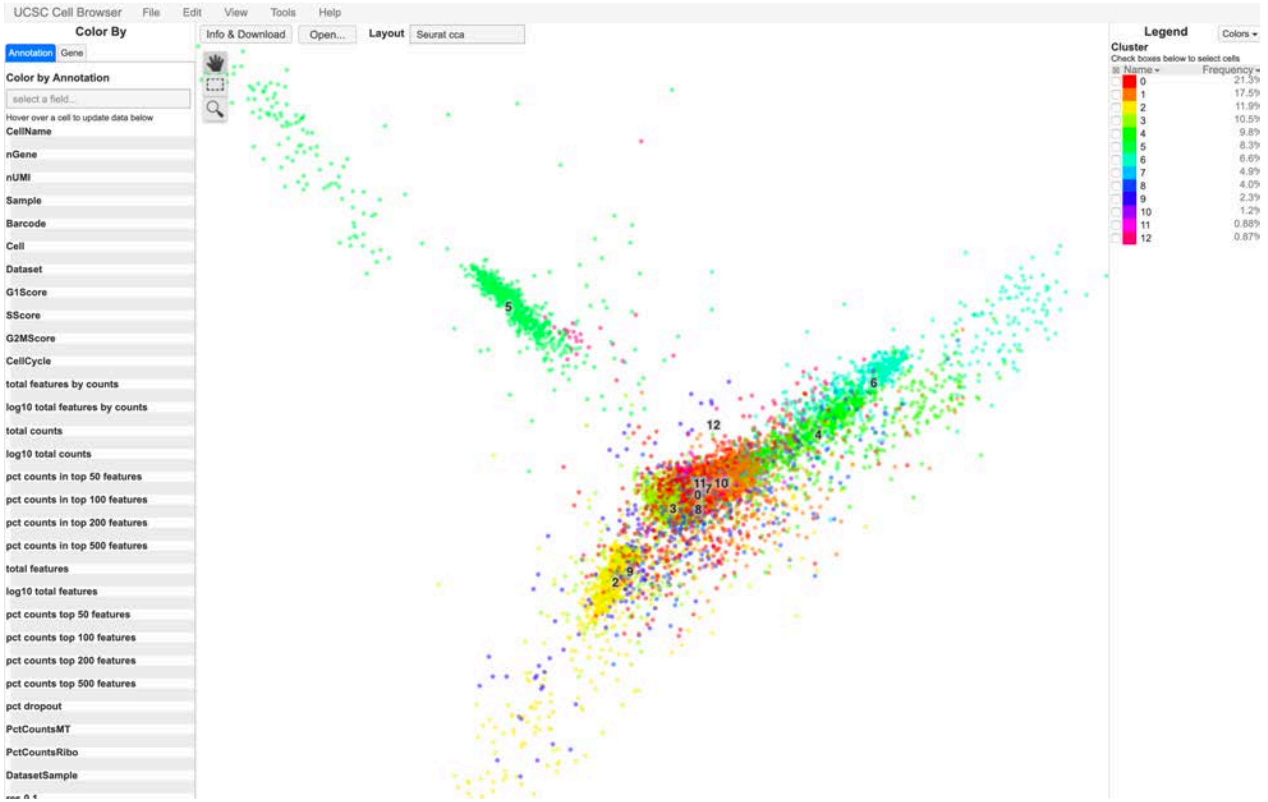


60+ anatomical structures linked to standard terminology in Zebrafish atlas (pictured)

Upcoming Visualization Features

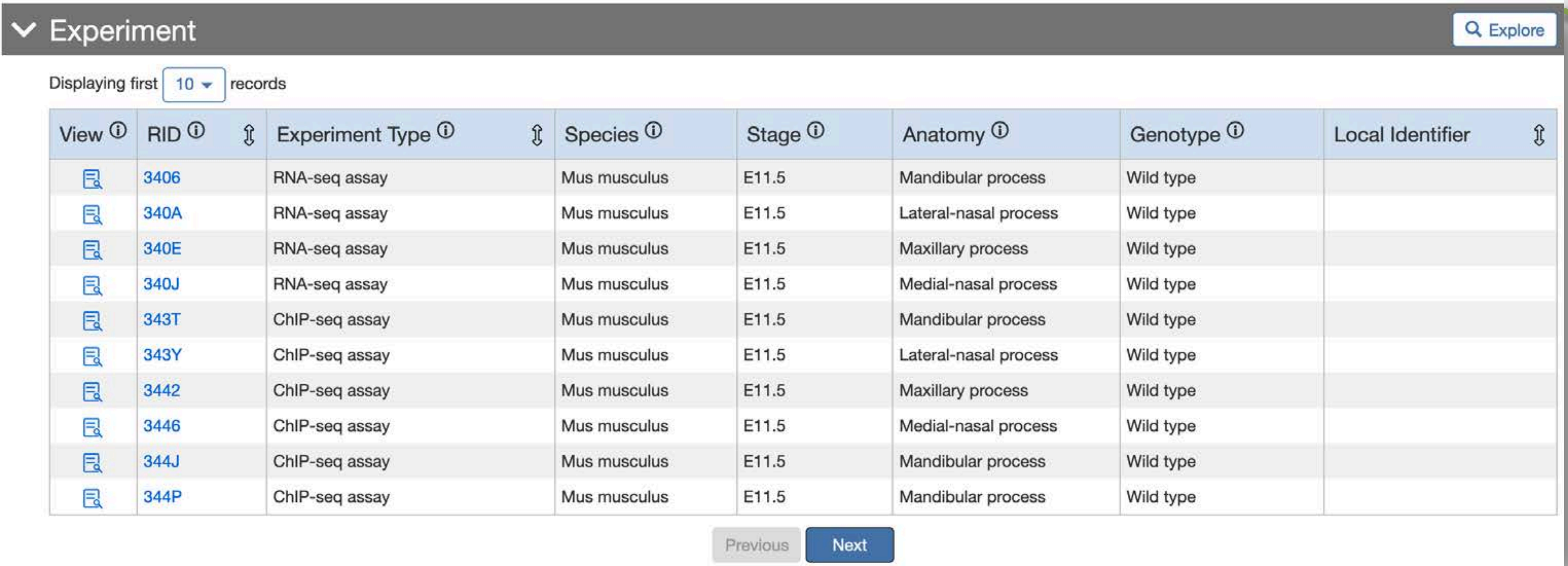


High-resolution Microscopy Images with Annotations



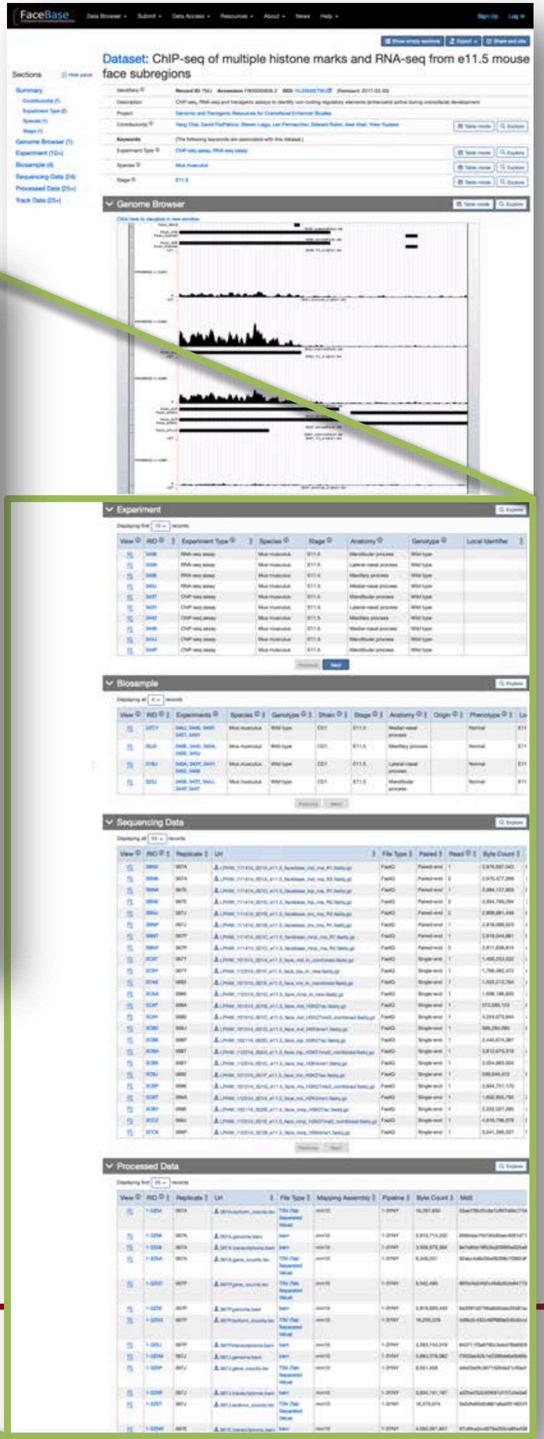
Single Cell Browser through Integration with UCSC

Details on Experiments, Biosamples, etc.



Displaying first 10 records

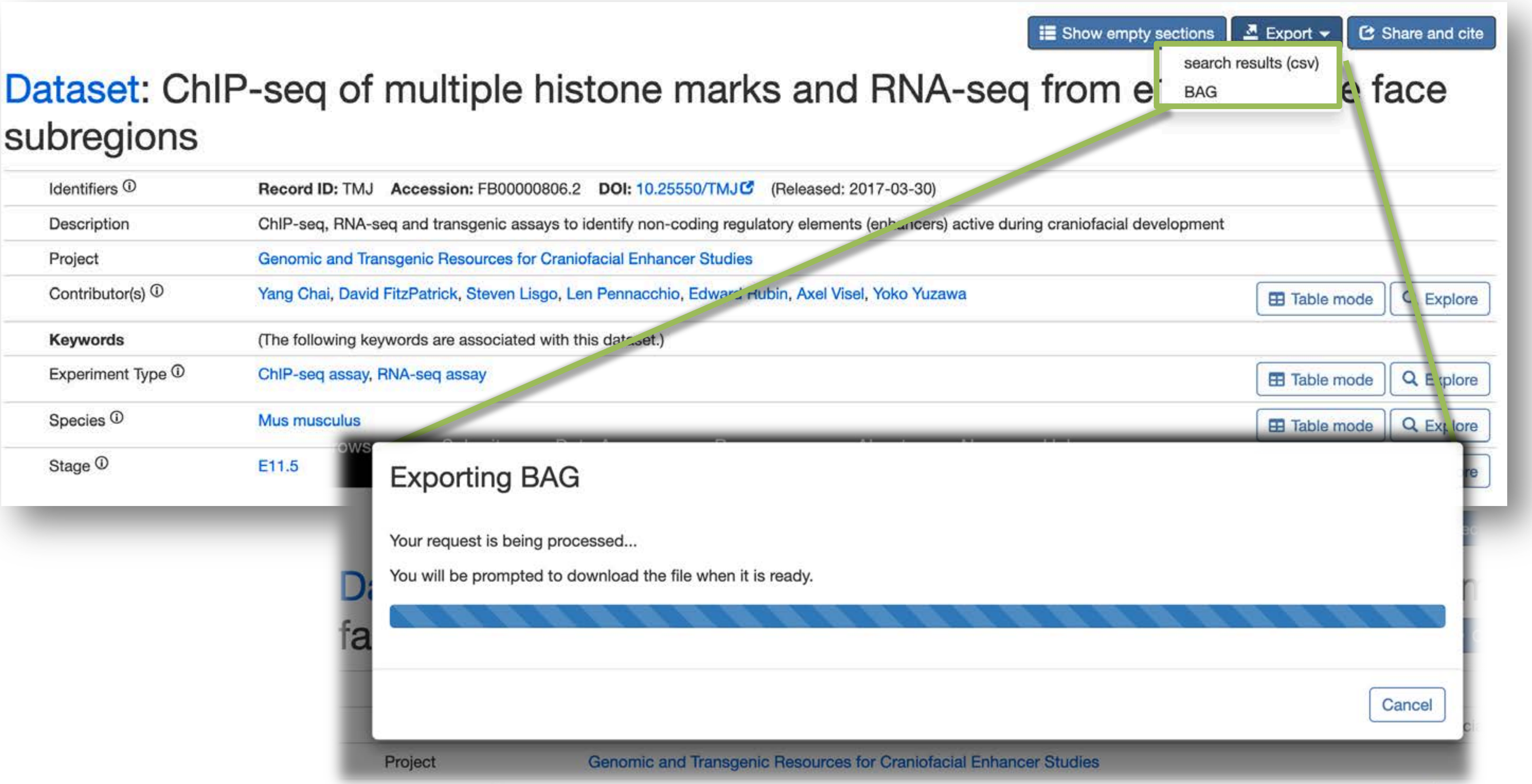
View	RID	Experiment Type	Species	Stage	Anatomy	Genotype	Local Identifier
	3406	RNA-seq assay	Mus musculus	E11.5	Mandibular process	Wild type	
	340A	RNA-seq assay	Mus musculus	E11.5	Lateral-nasal process	Wild type	
	340E	RNA-seq assay	Mus musculus	E11.5	Maxillary process	Wild type	
	340J	RNA-seq assay	Mus musculus	E11.5	Medial-nasal process	Wild type	
	343T	ChIP-seq assay	Mus musculus	E11.5	Mandibular process	Wild type	
	343Y	ChIP-seq assay	Mus musculus	E11.5	Lateral-nasal process	Wild type	
	3442	ChIP-seq assay	Mus musculus	E11.5	Maxillary process	Wild type	
	3446	ChIP-seq assay	Mus musculus	E11.5	Medial-nasal process	Wild type	
	344J	ChIP-seq assay	Mus musculus	E11.5	Mandibular process	Wild type	
	344P	ChIP-seq assay	Mus musculus	E11.5	Mandibular process	Wild type	



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e11.5 mouse face subregions

Further drill-down to related biosamples, files, and other related data.

Export (i.e., Download) Dataset



Dataset: ChIP-seq of multiple histone marks and RNA-seq from e e face subregions

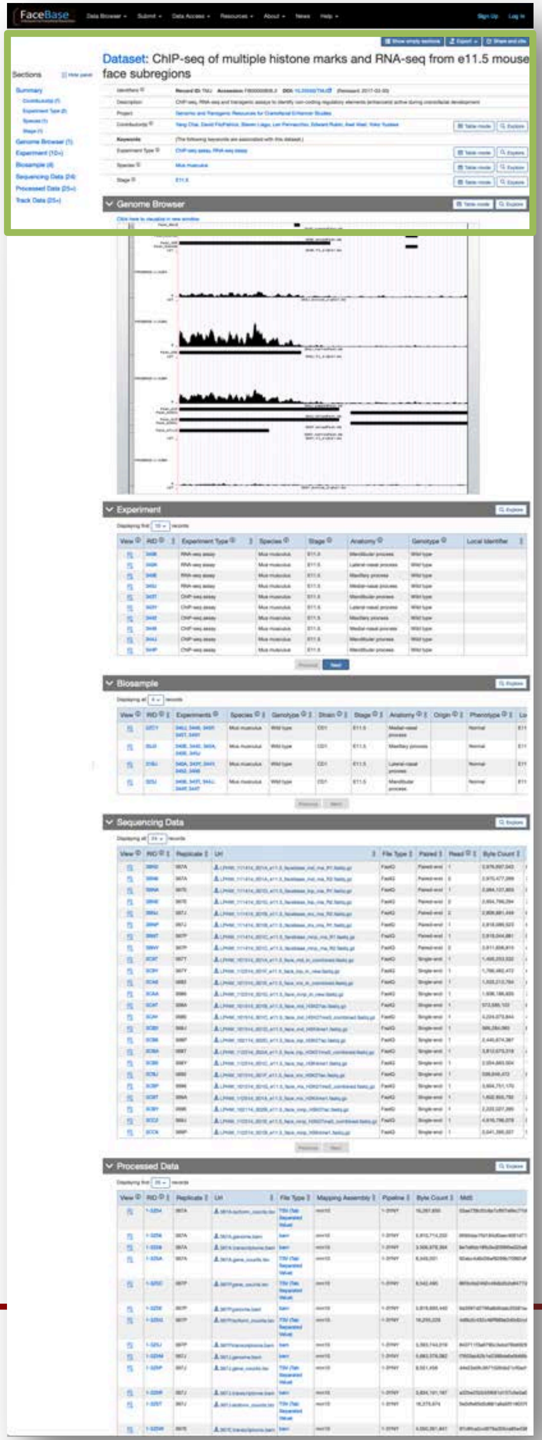
Identifiers	**Record ID:** TMJ **Accession:** FB00000806.2 **DOI:** 10.25550/TMJ (Released: 2017-03-30)
Description	ChIP-seq, RNA-seq and transgenic assays to identify non-coding regulatory elements (enhancers) active during craniofacial development
Project	Genomic and Transgenic Resources for Craniofacial Enhancer Studies
Contributor(s)	Yang Chai, David FitzPatrick, Steven Lisgo, Len Pennacchio, Edward Rubin, Axel Visel, Yoko Yuzawa
Keywords	(The following keywords are associated with this dataset.)
Experiment Type	ChIP-seq assay, RNA-seq assay
Species	Mus musculus
Stage	E11.5

Exporting BAG

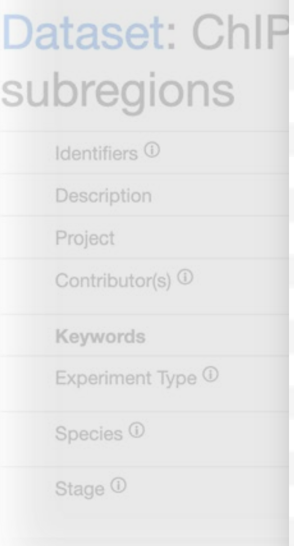
Your request is being processed...

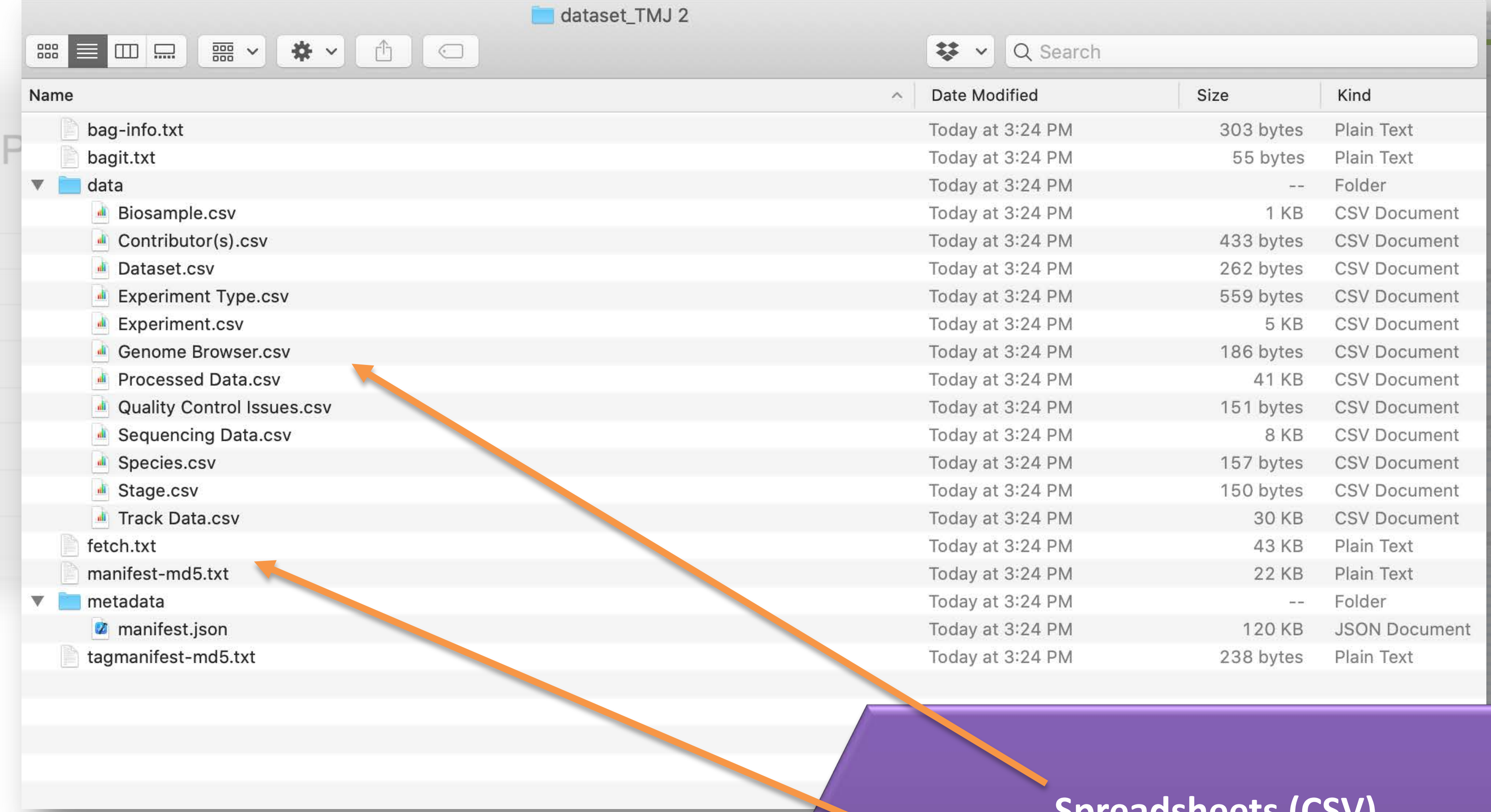
You will be prompted to download the file when it is ready.

Project	Genomic and Transgenic Resources for Craniofacial Enhancer Studies



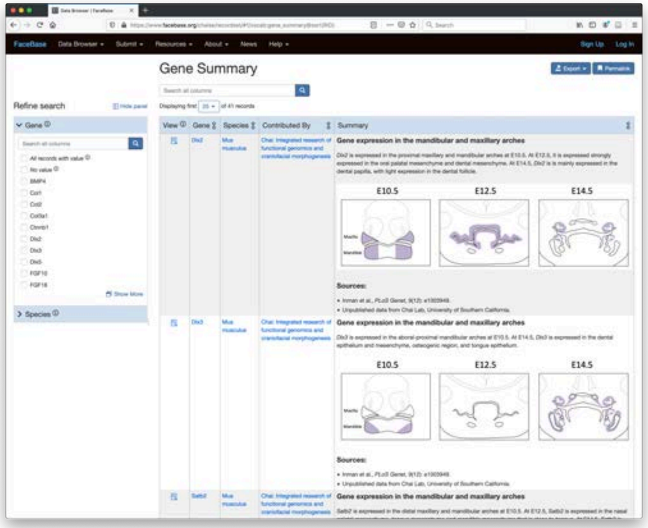
“Big Data Bag” Format for Large Datasets





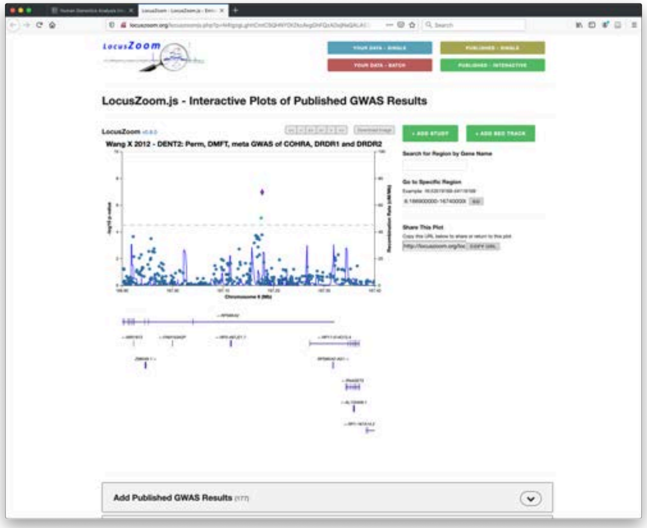
Name	Date Modified	Size	Kind
bag-info.txt	Today at 3:24 PM	303 bytes	Plain Text
bagit.txt	Today at 3:24 PM	55 bytes	Plain Text
▼ data	Today at 3:24 PM	--	Folder
Biosample.csv	Today at 3:24 PM	1 KB	CSV Document
Contributor(s).csv	Today at 3:24 PM	433 bytes	CSV Document
Dataset.csv	Today at 3:24 PM	262 bytes	CSV Document
Experiment Type.csv	Today at 3:24 PM	559 bytes	CSV Document
Experiment.csv	Today at 3:24 PM	5 KB	CSV Document
Genome Browser.csv	Today at 3:24 PM	186 bytes	CSV Document
Processed Data.csv	Today at 3:24 PM	41 KB	CSV Document
Quality Control Issues.csv	Today at 3:24 PM	151 bytes	CSV Document
Sequencing Data.csv	Today at 3:24 PM	8 KB	CSV Document
Species.csv	Today at 3:24 PM	157 bytes	CSV Document
Stage.csv	Today at 3:24 PM	150 bytes	CSV Document
Track Data.csv	Today at 3:24 PM	30 KB	CSV Document
fetch.txt	Today at 3:24 PM	43 KB	Plain Text
manifest-md5.txt	Today at 3:24 PM	22 KB	Plain Text
▼ metadata	Today at 3:24 PM	--	Folder
manifest.json	Today at 3:24 PM	120 KB	JSON Document
tagmanifest-md5.txt	Today at 3:24 PM	238 bytes	Plain Text

Spreadsheets (CSV)
File Manifest (links + checksums)
Client tool for reliable download



Gene Summary

Detailed gene summaries

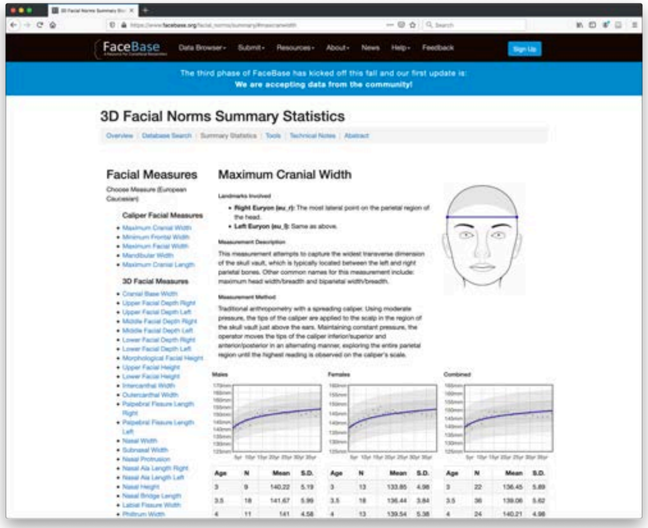


LocusZoom.js - Interactive Plots of Published GWAS Results

Wang X 2012 - DENT2: Perm, DMFT, meta GWAS of COHRA, DRDR1 and DRDR2

Add Published GWAS Results

Interactive plots of GWAS data



The third phase of FaceBase has kicked off this fall and our first update is:
We are accepting data from the community!

3D Facial Norms Summary Statistics

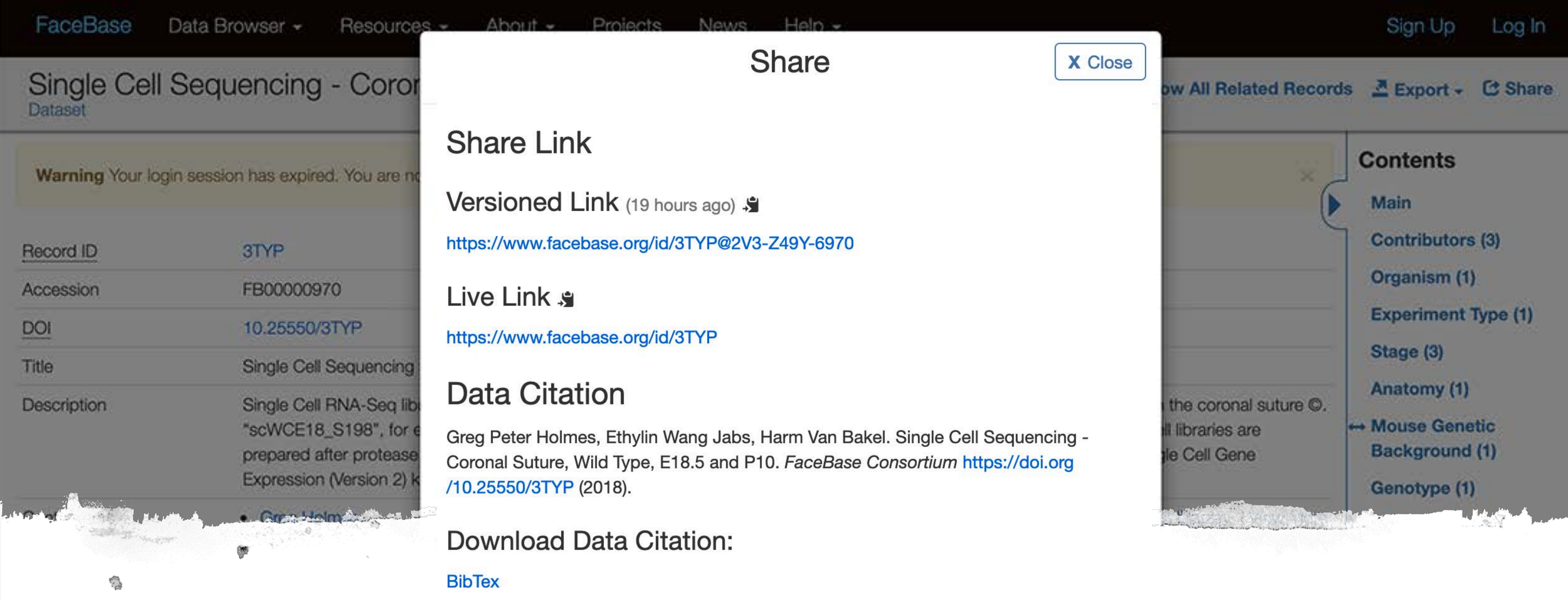
Facial Measures

Maximum Cranial Width

Facial norms measures

More at the Resources Hub…

https://www.facebase.org/resources/



Single Cell Sequencing - Coror

Dataset

Warning Your login session has expired. You are nc

Record ID 3TYP

Accession FB00000970

DOI 10.25550/3TYP

Title Single Cell Sequencing

Description Single Cell RNA-Seq lib "scWCE18_S198", for e prepared after protease Expression (Version 2) k … the coronal suture ©. … ll libraries are … le Cell Gene

Share

Share Link

Versioned Link (19 hours ago)

https://www.facebase.org/id/3TYP@2V3-Z49Y-6970

Live Link

https://www.facebase.org/id/3TYP

Data Citation

Greg Peter Holmes, Ethylin Wang Jabs, Harm Van Bakel. Single Cell Sequencing - Coronal Suture, Wild Type, E18.5 and P10. *FaceBase Consortium* https://doi.org/10.25550/3TYP (2018).

Download Data Citation:

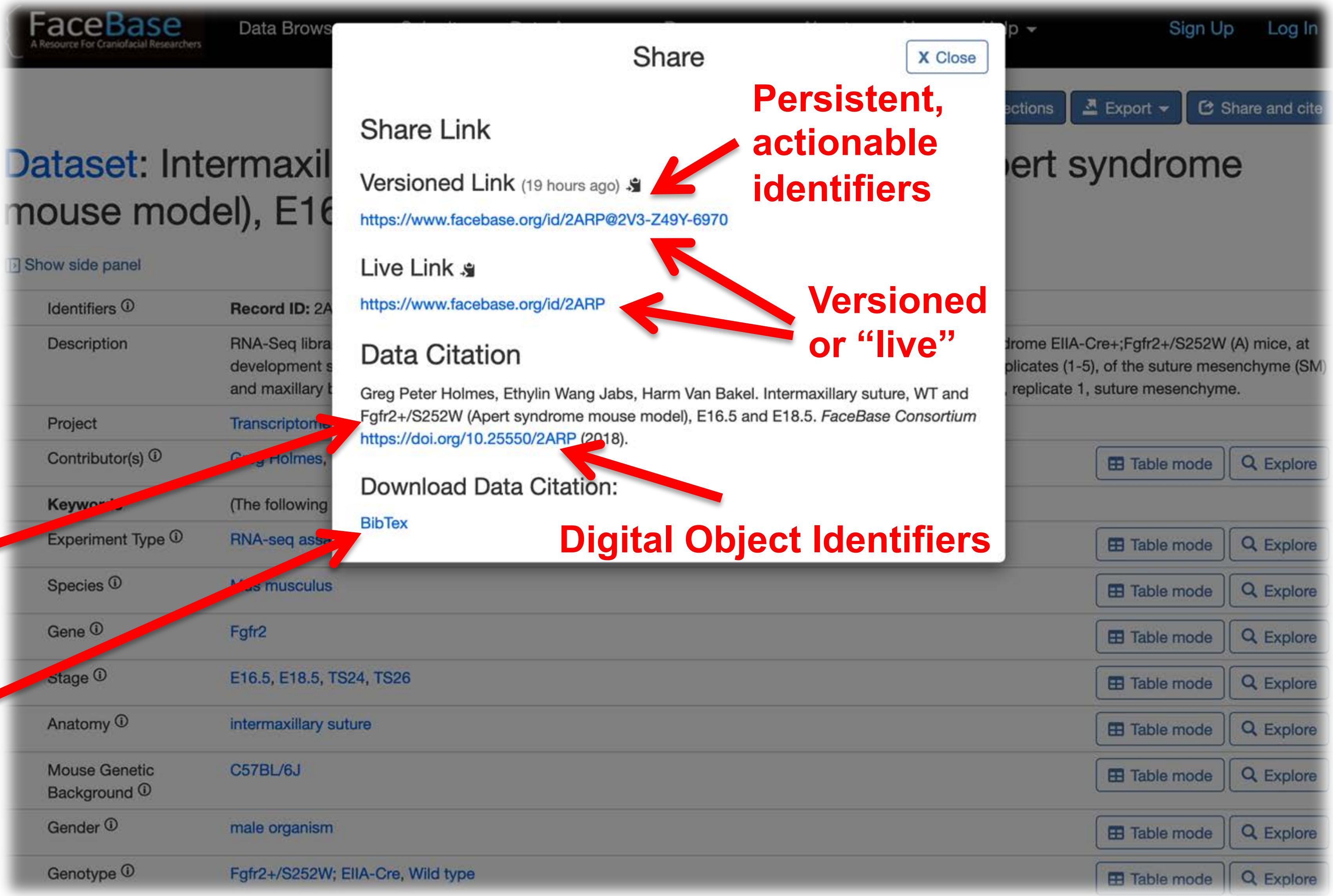
BibTex

Why Share Your Data?

- **Dedicated focus on craniofacial and dental**
- **Increase the visibility and impact of your research**
- **Cross-reference with publication**
- **Data are "published" like first-class academic works**
- **Satisfy data sharing requirements for grants and publications**

Data Citation

- **Persistent actionable identifiers** for *all database entities*
- Explore any **historical version** of *entire database*



Dataset: Intermaxil … ert syndrome mouse model), E16

Identifiers Record ID: 2A

Description RNA-Seq libra … development s … and maxillary … drome EIIA-Cre+;Fgfr2+/S252W (A) mice, at … plicates (1-5), of the suture mesenchyme (SM) … replicate 1, suture mesenchyme.

Project Transcriptome

Contributor(s) Greg Holmes,

Keywords (The following

Experiment Type RNA-seq assa

Species Mus musculus

Gene Fgfr2

Stage E16.5, E18.5, TS24, TS26

Anatomy intermaxillary suture

Mouse Genetic Background C57BL/6J

Gender male organism

Genotype Fgfr2+/S252W; EIIA-Cre, Wild type

Share

Share Link

Versioned Link (19 hours ago)

https://www.facebase.org/id/2ARP@2V3-Z49Y-6970

Live Link

https://www.facebase.org/id/2ARP

Data Citation

Greg Peter Holmes, Ethylin Wang Jabs, Harm Van Bakel. Intermaxillary suture, WT and Fgfr2+/S252W (Apert syndrome mouse model), E16.5 and E18.5. *FaceBase Consortium* https://doi.org/10.25550/2ARP (2018).

Download Data Citation:

BibTex

Persistent, actionable identifiers

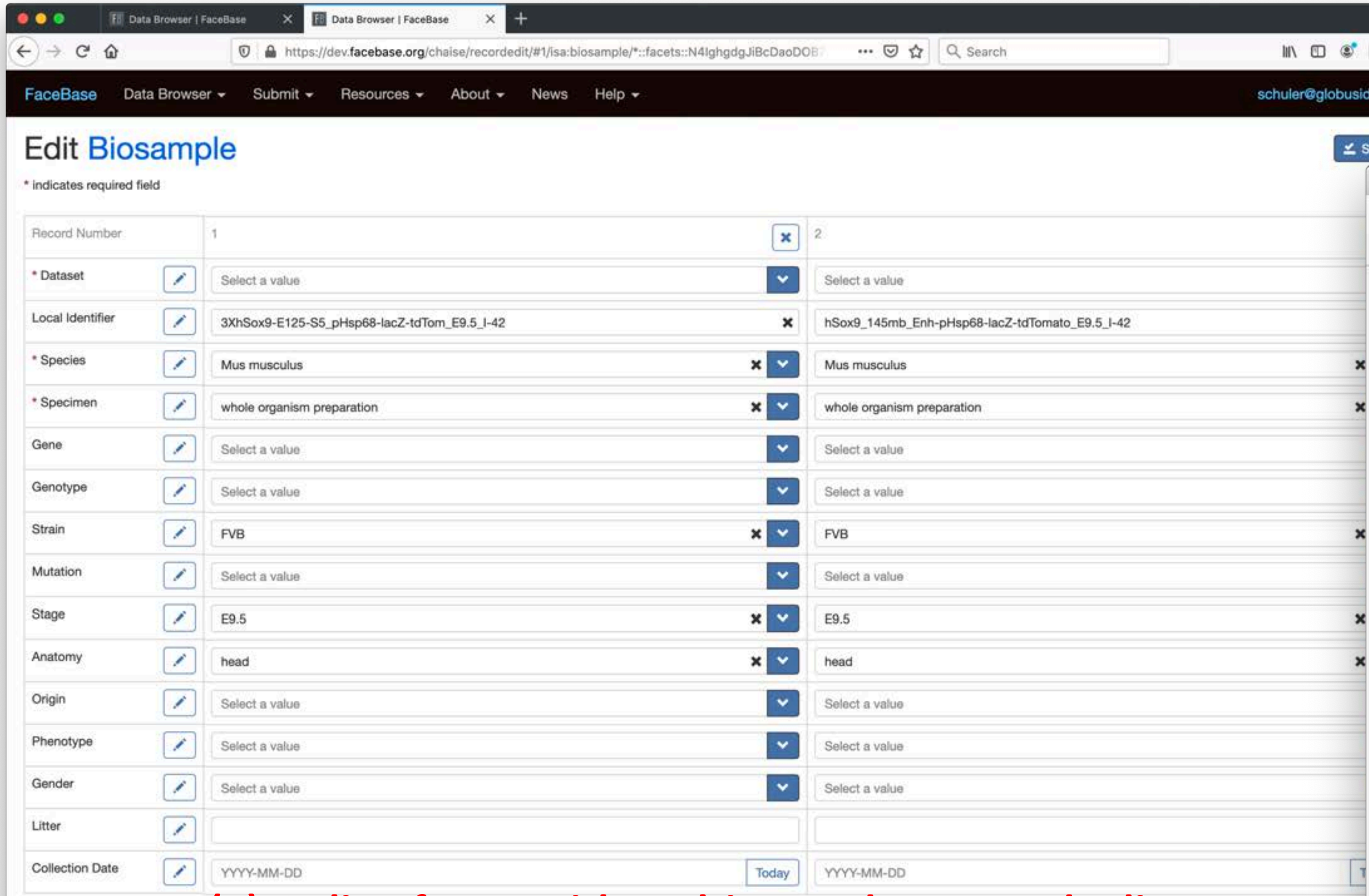
Versioned or "live"

Digital Object Identifiers

Follows leading publisher's recommendations

Import into reference managers

Empower Contributors to Submit Valuable Data



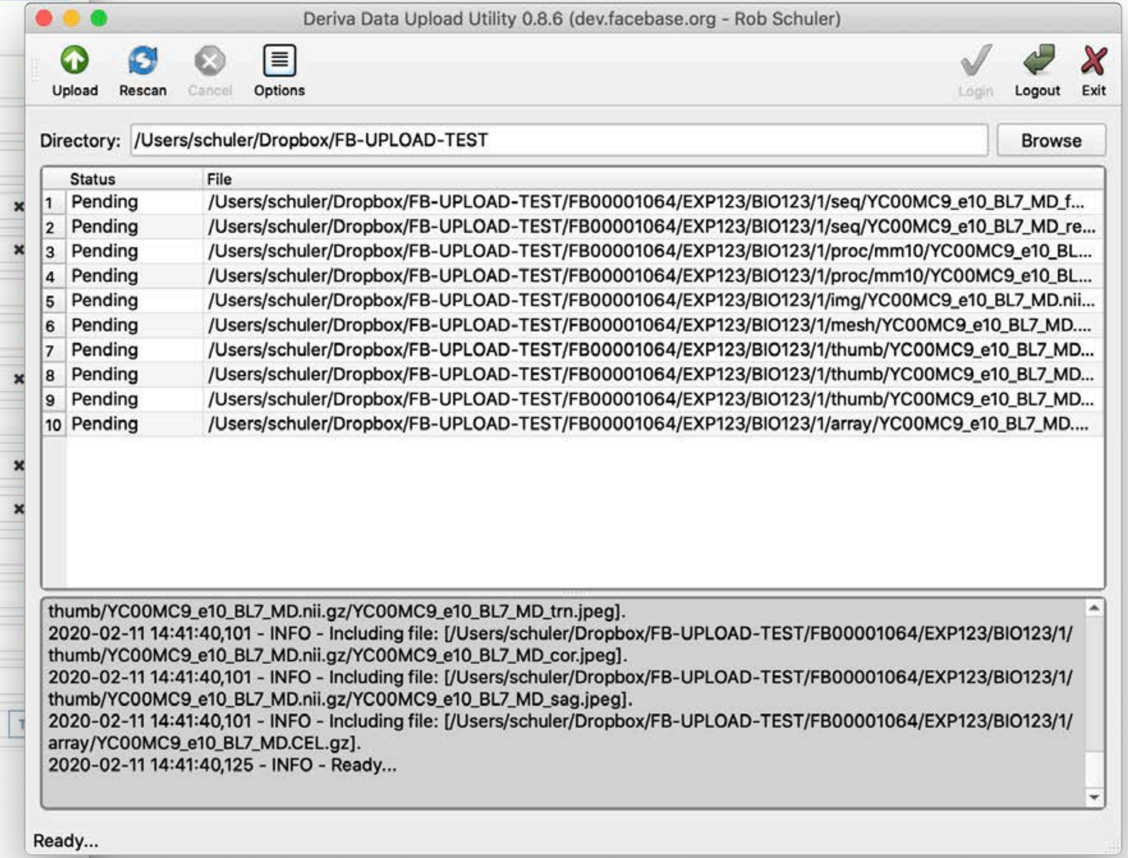
Edit Biosample

* indicates required field

Record Number	1	2
* Dataset	Select a value	Select a value
Local Identifier	3XhSox9-E125-S5_pHsp68-lacZ-tdTom_E9.5_I-42	hSox9_145mb_Enh-pHsp68-lacZ-tdTomato_E9.5_I-42
* Species	Mus musculus	Mus musculus
* Specimen	whole organism preparation	whole organism preparation
Gene	Select a value	Select a value
Genotype	Select a value	Select a value
Strain	FVB	FVB
Mutation	Select a value	Select a value
Stage	E9.5	E9.5
Anatomy	head	head
Origin	Select a value	Select a value
Phenotype	Select a value	Select a value
Gender	Select a value	Select a value
Litter		
Collection Date	YYYY-MM-DD Today	YYYY-MM-DD

(1) Online forms with multi-record entry and edit

(2) Graphical desktop applications



Directory: /Users/schuler/Dropbox/FB-UPLOAD-TEST

	Status	File
1	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/seq/YC00MC9_e10_BL7_MD_f...
2	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/seq/YC00MC9_e10_BL7_MD_re...
3	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/proc/mm10/YC00MC9_e10_BL...
4	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/proc/mm10/YC00MC9_e10_BL...
5	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/img/YC00MC9_e10_BL7_MD.nii...
6	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/mesh/YC00MC9_e10_BL7_MD....
7	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/thumb/YC00MC9_e10_BL7_MD...
8	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/thumb/YC00MC9_e10_BL7_MD...
9	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/thumb/YC00MC9_e10_BL7_MD...
10	Pending	/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/array/YC00MC9_e10_BL7_MD....

```
thumb/YC00MC9_e10_BL7_MD.nii.gz/YC00MC9_e10_BL7_MD_trn.jpeg].
2020-02-11 14:41:40,101 - INFO - Including file: [/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/
thumb/YC00MC9_e10_BL7_MD.nii.gz/YC00MC9_e10_BL7_MD_cor.jpeg].
2020-02-11 14:41:40,101 - INFO - Including file: [/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/
thumb/YC00MC9_e10_BL7_MD.nii.gz/YC00MC9_e10_BL7_MD_sag.jpeg].
2020-02-11 14:41:40,101 - INFO - Including file: [/Users/schuler/Dropbox/FB-UPLOAD-TEST/FB00001064/EXP123/BIO123/1/
array/YC00MC9_e10_BL7_MD.CEL.gz].
2020-02-11 14:41:40,125 - INFO - Ready...
```

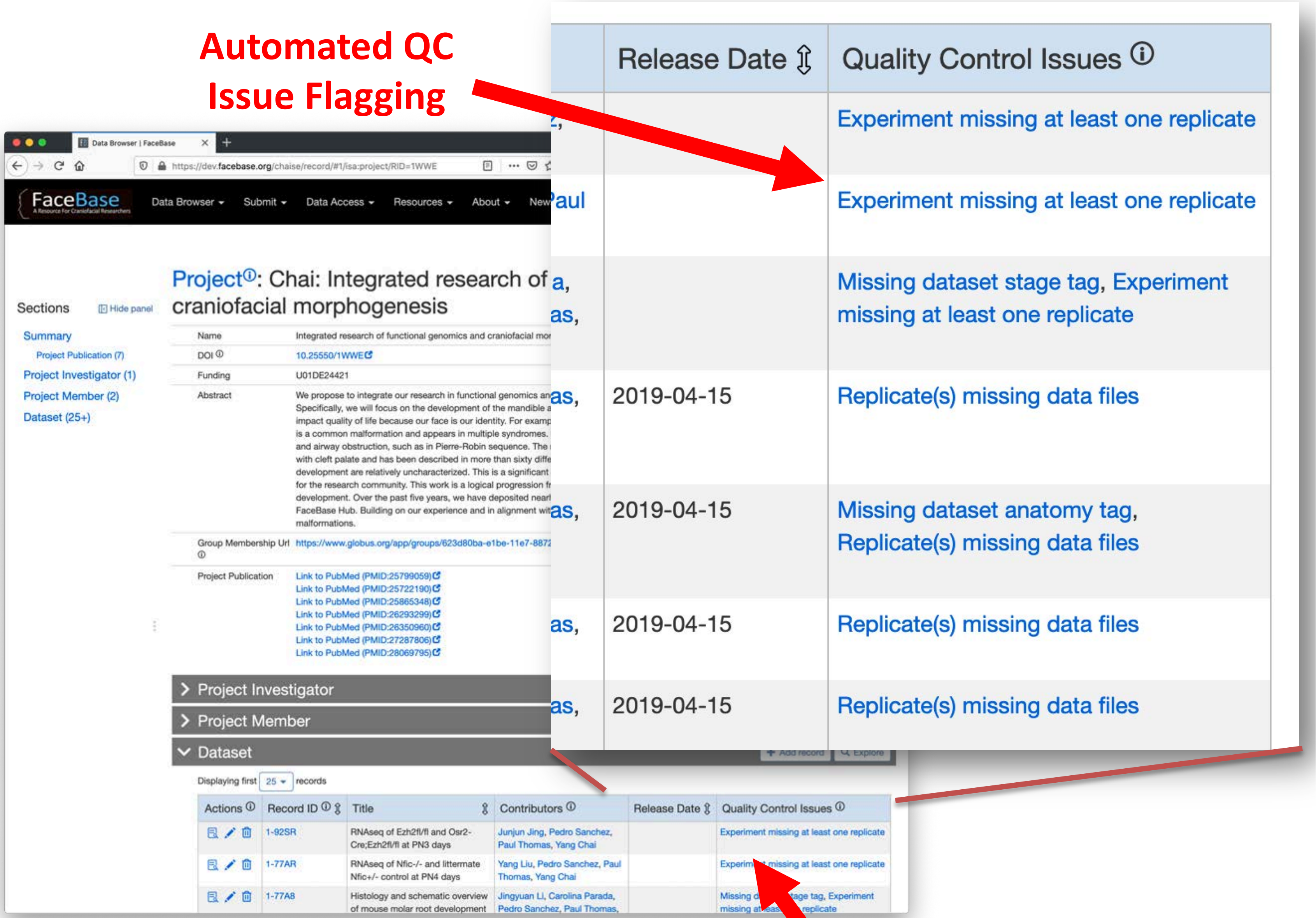
Ready...

(3) Command-line, Python and Web services interfaces for custom data submission workflows (not pictured)

Researchers have submitted their own datasets with 100s to 1000s of files, usually in a few days

Quality Control Dashboards

- Automated Quality Control Rules to check each dataset
- QC issues are displayed on project pages (visible only to contributors)



Automated QC Issue Flagging

Release Date	Quality Control Issues ⓘ
	Experiment missing at least one replicate
	Experiment missing at least one replicate
	Missing dataset stage tag, Experiment missing at least one replicate
2019-04-15	Replicate(s) missing data files
2019-04-15	Missing dataset anatomy tag, Replicate(s) missing data files
2019-04-15	Replicate(s) missing data files
2019-04-15	Replicate(s) missing data files

Projectⓘ: Chai: Integrated research of craniofacial morphogenesis

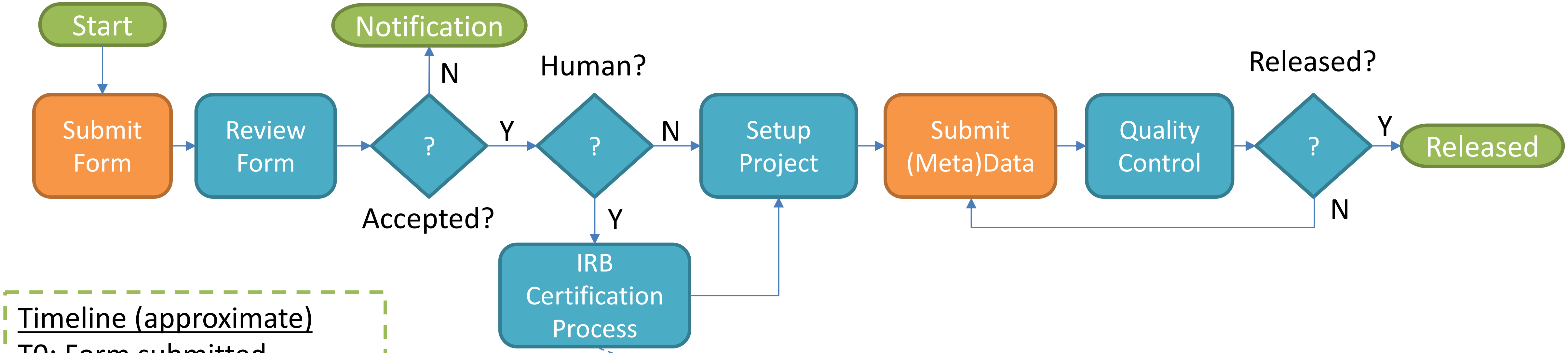
Name	Integrated research of functional genomics and craniofacial mor
DOI ⓘ	10.25550/1WWE
Funding	U01DE24421
Abstract	We propose to integrate our research in functional genomics an Specifically, we will focus on the development of the mandible a impact quality of life because our face is our identity. For examp is a common malformation and appears in multiple syndromes. and airway obstruction, such as in Pierre-Robin sequence. The with cleft palate and has been described in more than sixty diffe development are relatively uncharacterized. This is a significant for the research community. This work is a logical progression fr development. Over the past five years, we have deposited near FaceBase Hub. Building on our experience and in alignment wit malformations.
Group Membership Url ⓘ	https://www.globus.org/app/groups/623d80ba-e1be-11e7-8872
Project Publication	Link to PubMed (PMID:25799059) Link to PubMed (PMID:25722190) Link to PubMed (PMID:25865348) Link to PubMed (PMID:26293299) Link to PubMed (PMID:26350960) Link to PubMed (PMID:27287806) Link to PubMed (PMID:28069795)

Displaying first 25 records

Actions ⓘ	Record ID ⓘ	Title	Contributors ⓘ	Release Date	Quality Control Issues ⓘ
	1-92SR	RNAseq of Ezh2fl/fl and Osr2-Cre;Ezh2fl/fl at PN3 days	Junjun Jing, Pedro Sanchez, Paul Thomas, Yang Chai		Experiment missing at least one replicate
	1-77AR	RNAseq of Nfic-/- and littermate Nfic+/- control at PN4 days	Yang Liu, Pedro Sanchez, Paul Thomas, Yang Chai		Experiment missing at least one replicate
	1-77A8	Histology and schematic overview of mouse molar root development	Jingyuan Li, Carolina Parada, Pedro Sanchez, Paul Thomas,		Missing dataset stage tag, Experiment missing at least one replicate

Quality Control "Dashboard" Integrated into Project Pages

Streamlined Process for Data Submission



Start

Submit Form

Review Form

?

Accepted?

N

Notification

Y

Human?

?

N

Y

IRB Certification Process

Setup Project

Submit (Meta)Data

Quality Control

Released?

?

Y

Released

N

<u>Timeline (approximate)</u>

T0: Form submitted
T+2 weeks: Review decision
T+3 weeks: Project setup
T+5 weeks: Submit data*
T+6 weeks: QC review
* Based on user averages

<u>IRB Certification Process</u>

- Individual level data classified as human subjects
- Requires USC certification of your IRB decision
- Tracks are not considered restricted data
- Timeline Varies

🟧 User Activity
🟦 Hub Activity

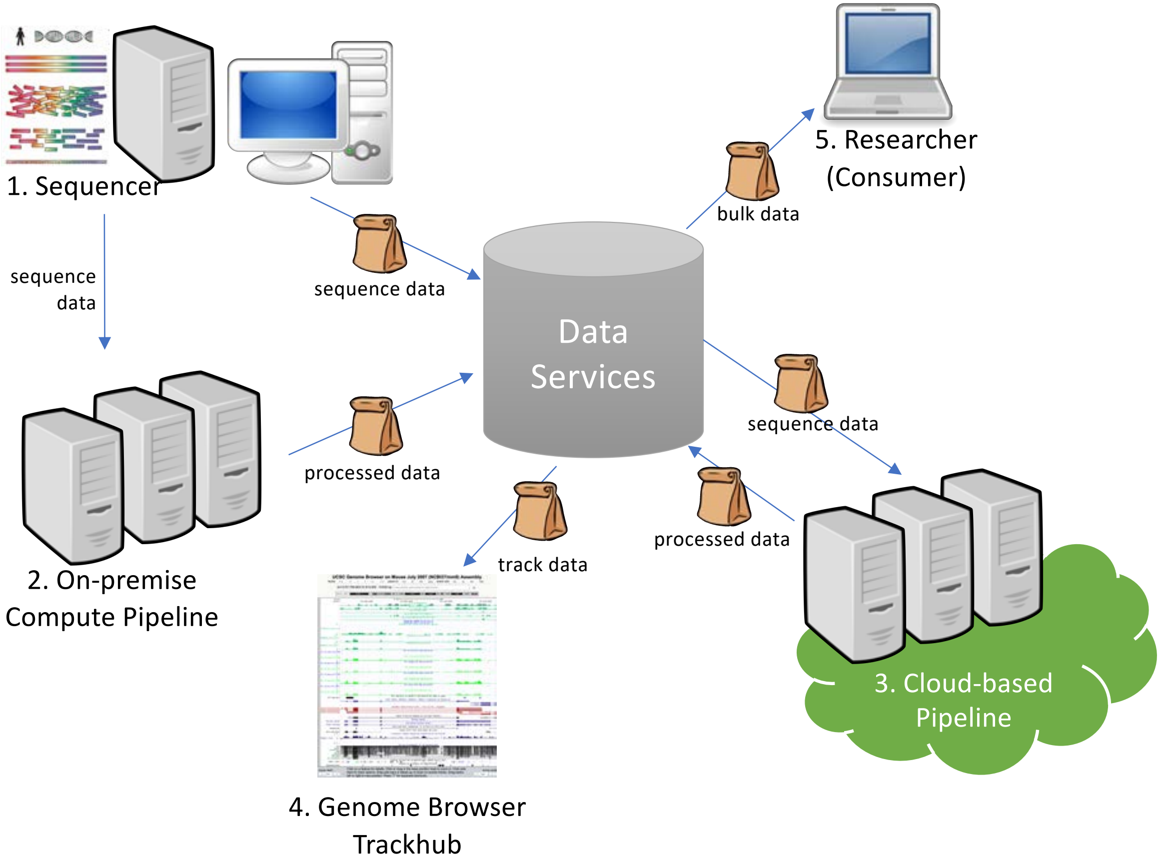
A Test of the Reproducibility of FaceBase Data

Evaluation: 3 labs, 13 datasets, on 192 RNA-Seq or ChIP-Seq experiments (1000+ data files)

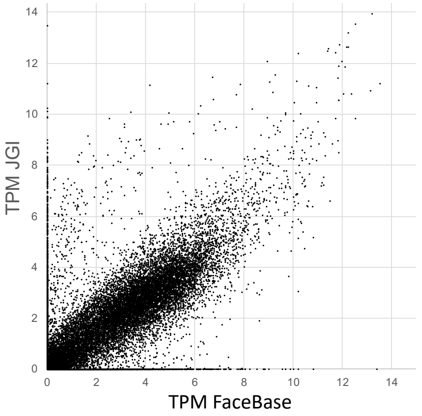
a) Uniform processing pipeline

b) Output strongly correlated with researcher's results

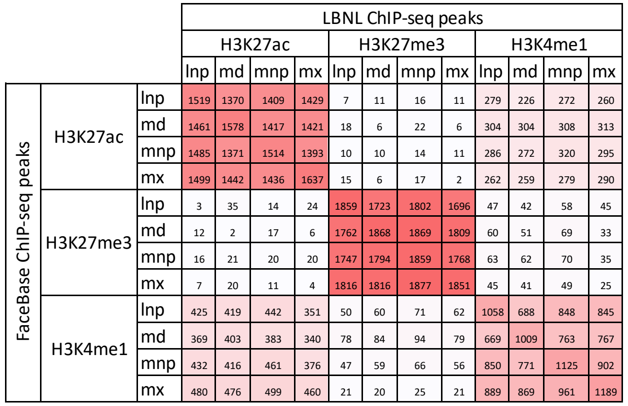
(a)



(b)



RNA-seq log2 counts



			LBNL ChIP-seq peaks											
			H3K27ac				H3K27me3				H3K4me1			
			lnp	md	mnp	mx	lnp	md	mnp	mx	lnp	md	mnp	mx
FaceBase ChIP-seq peaks	H3K27ac	lnp	1519	1370	1409	1429	7	11	16	11	279	226	272	260
		md	1461	1578	1417	1421	18	6	22	6	304	304	308	313
		mnp	1485	1371	1514	1393	10	10	14	11	286	272	320	295
		mx	1499	1442	1436	1637	15	6	17	2	262	259	279	290
	H3K27me3	lnp	3	35	14	24	1859	1723	1802	1696	47	42	58	45
		md	12	2	17	6	1762	1868	1869	1809	60	51	69	33
		mnp	16	21	20	20	1747	1794	1859	1768	63	62	70	35
		mx	7	20	11	4	1816	1816	1877	1851	45	41	49	25
	H3K4me1	lnp	425	419	442	351	50	60	71	62	1058	688	848	845
		md	369	403	383	340	78	84	94	79	669	1009	763	767
		mnp	432	416	461	376	47	59	66	56	850	771	1125	902
		mx	480	476	499	460	21	20	25	21	889	869	961	1189

ChIP-seq peaks (overlap top-2000)

Thank You

- **Co-PIs**: Carl Kesselman; Yang Chai
- **Core Team**: Rob Schuler (CS & Technical); Bridget Samuels (Biocuration); Alejandro Bugacov (Data science); Cris Williams (Communications); Joe Hacia (Bioinformatics); Thach Vu Ho (Data curation)

- **Website**: www.facebase.org
- **Sponsor**: NIH / NIDCR (U01DE028729)

USC
Viterbi
Information
Sciences Institute

Center for
Craniofacial
Molecular
Biology

Multi-Disciplined Partnership