FaceBase Management and Coordination Hub

Jeff Murray
University of Iowa

Mary L. Marazita
University of Pittsburgh



YEAR FOUR HIGHLIGHTS

Statistics

- 430 Registered users (127 registered in the past 12 months)
- 6317 visits (FEB2012 FEB 2013)
- Average of viewing 5 pages per visit
- 102 different academic and/or research institutes
- 98 Countries



Data Status

	Investigators	File Type(s)	Number of Datasets	Total:	
Data Received & Not Posted	Hallgrimsson, Benedikt	.obj	1		
	22 A 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1 C 1		3	6	
			1		
	Sanchez, Pedro	microCT	1		
Data Awaiting Approval	Beaty, Terri	.map, .ped	1	28	
	Potter, Steve	Gene expression microarray & RNA-Seq	20		
	Sanchez, Pedro	microCT	7		
	Batzel, Peter	.fq	3		
	Chai, Yang	microarray	6		
Total Datasets Posted	David Clouthier	miRNA	1		
	Hallgrimsson, Benedikt	.aim, .txt, .hx	37		
	Pelikan, Richard	.cel	10		
	Postlethwait, John	miRNA .xls	2	264	
	Potter, Steve	microarray .cel		264	
	Ruffins, Seth	.nii, .xls, .jpg	41		
	Sanchez, Pedro	microCT	99		
	Shapiro, Linda	.model, .dcm	2		
	Visel, Axel	.xls, .pdf; .pdf, .avi, .tgz, .txt	51		
	Weinberg, Seth	3DMD	1		



Data Quality

- Data curation process and procedures
- Data quality verification check
 - Data Curator verified 100% of the submitted data set
 - Corrected the tagged data
 - Contacted Spoke PIs
- Monthly data status report reviewed with NIDCR



Hub processes/procedures

- New virtual development environments for developers to program and test
- Revamped bug tracking and testing
- Seven major releases throughout year 4

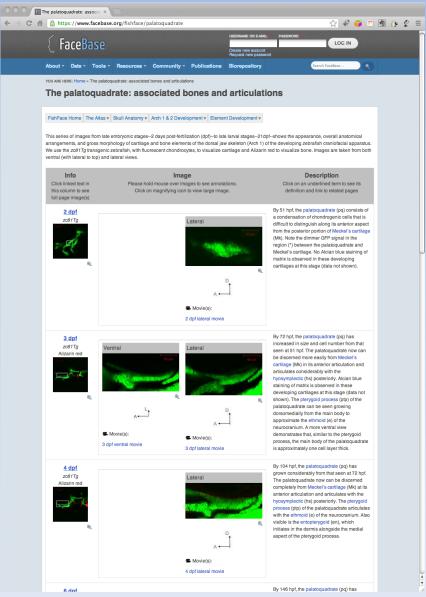
Year 4 Highlights: Spoke Specific

Update to 3D Normative Database
Generalized 3D module
Applicable to any 3D facial
image data
FishFace enhancements



Fishface enhancements

https://www.facebase.org/fishface/home





Year 4 Highlights: Spoke Specific (continued)

Linking Genome Browser to datasets

Enhancer datasets update

Jackson Labs

Cre driver project

Status and strain updates

Biorepository page enhancements



JAX Cre Driver Project

YOU ARE HERE: Home » JAX Cre Driver Project

JAX Cre Driver Project

View Edit Node export

The overall goal of the FaceBase Cre driver program is to generate a set of novel mouse tool strains to facilitate genetic analysis of the developing midface and palate. The current set of strains under development target structures and cell populations that are not effectively covered by the existing repertoire of cre strains available to the public. Current progress towards the release of these strains is indicated in the table, with links provided to a description of the strain, characterization data and All strains will be distributed by the JAX Repository. If you have any comments, questions or suggestions contact Steve Murray.

Additional JAX resources

- · Other FaceBase tool strains at JAX
- All Cre strains at JAX
- JAX Cre strain characterization
- Other FaceBase Repository mouse models

BAC Transgenics							
Project	Driver (gene/enhancer)	Allele Type	Construct Complete?	Injection?	Founders?	Characterization?	Available?
dNp63CreERT2_2ACerulean	Trp63	BAC Tg	Yes	Yes	Yes	In Progress	In Progress
dNp63Cre_2ACerulean	Trp63	BAC Tg	Yes	Yes	Yes	Yes	Yes
Founder 3428			Yes	Yes	Yes	Yes	Yes
Founder 3430			Yes	Yes	Yes	Yes	Yes

Biorepository



YOU ARE HERE: Home » Biorepository

Biorepository

The FaceBase Biorepository is currently collecting biological samples from individuals with craniofacial disorders or defects along with their family members.

Information for Prospective Cases: Clefts of the lip and/or palate can be caused by a wide range of genetic, environmental and other factors. The FaceBase Biorepository will serve as a common source of both biological samples and information that can be made available to investigators trying to determine the underlying cause of these common birth defects. Genetic studies, in particular, will benefit from both family history information and having samples from affected individuals as well as their family members. DNA is the information containing molecules found in all the cells of our body and can be easily obtained from material such as blood or saliva samples. As part of the FaceBase Biorepository, we are requesting families to submit biological samples from specific family members as well as information from other family members that might be affected with either the same condition or a similar condition. The medical and family history information that is collected includes other relevant information such as exposure to possible environmental causes during pregnancy.

The biorepository is managed by Nichole Nidey, a research study coordinator, and Jeff Murray, a pediatric clinical geneticist and researcher. They are available to speak with family members regarding questions they may have, including providing information about the biorepository and making arrangements for the collection of samples for those who wish to participate. All participation is voluntary. Your name or other personally identifiable information (name, address, etc) will be removed before information is placed in the biorepository. Summary data to show how the database itself has been used overall as well as updates on whether specific findings might have been made using this database will be available on the FaceBase website at www.facebase.org. A newsletter containing this information will also be given to families and referring clinicians so that they may discuss the specifics with the families if there appears to be information that might be relevant in a particular case. Families will also need to sign a consent form that has been approved by the Institutional Review Board at the University of Iowa. Also, any submitted samples or data can also be removed from the database at any time should the family no longer wish to participate.

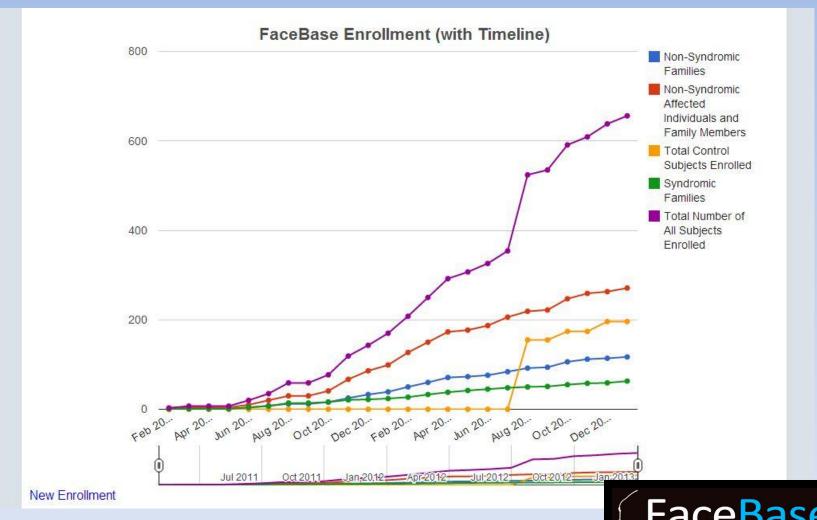
Investigation is interested in requesting DNA samples or for more information, please contact cleftresearch@ulowa.edu, Nichole Nidey, nichole-play@ulowa.edu or (319) 353-4365, or Jeff Murray, , ff-murray@ulowa.edu.

Select the data you would like to view pelow:

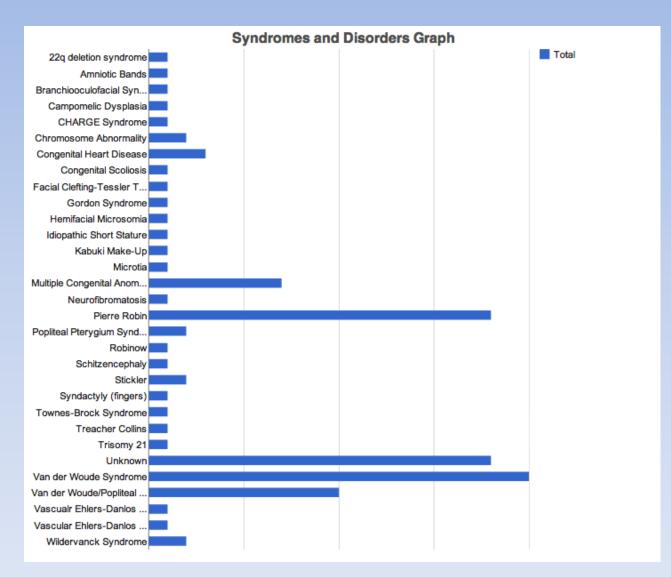
Syndromes and Disorders Enrollment Categories New Enrollment

w descriptions of the data categories here

Biorepository Enrollment



Syndromes and Disorders Graph





Year 4 Highlights: Viewer/Data Curation Improvements

OPT Viewer

Dataset versioning

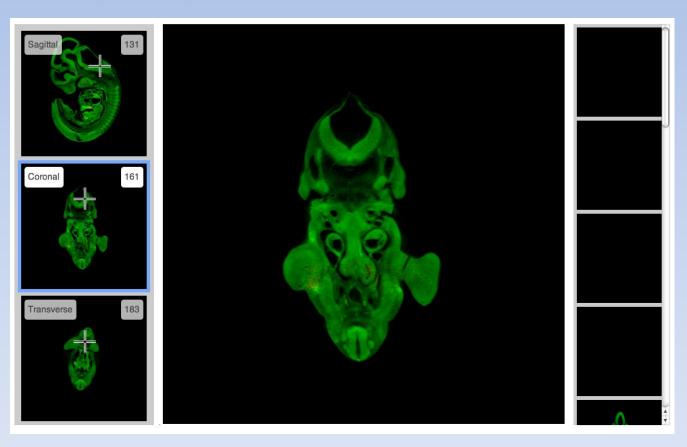
Email notification



Interactive OPT Viewer

Satya, et al. CMU

Active for all OPT datasets



Year 4 Highlights: Data Integration

- Gene research pages and updates
 - Requested and outlined at last annual meeting
- Timeline search

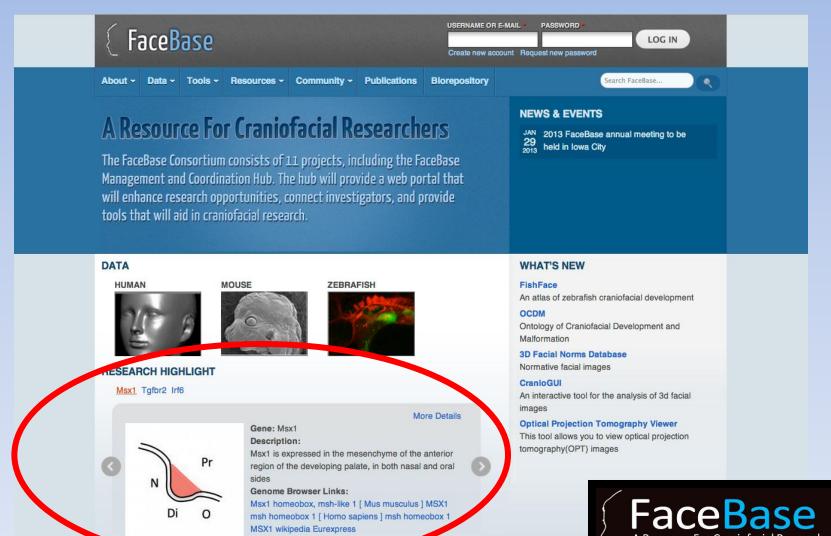


Research Highlight Pages

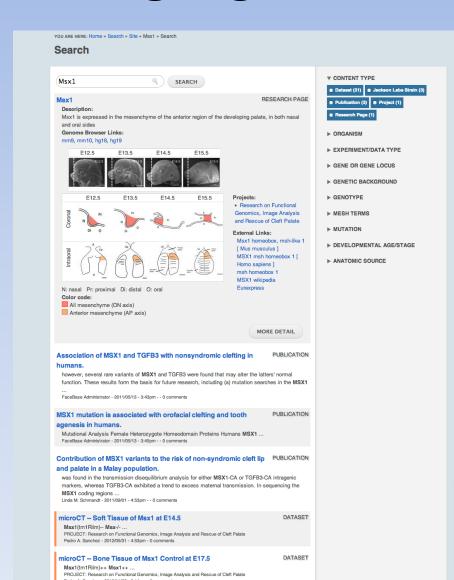
- Curated summaries of specific genes
 - Images, links, etc.
- Summaries and full pages
- Link to related datasets
- Many thanks to Yang's group
 - Production of images and content
 - -Curation



Research Highlight Summary

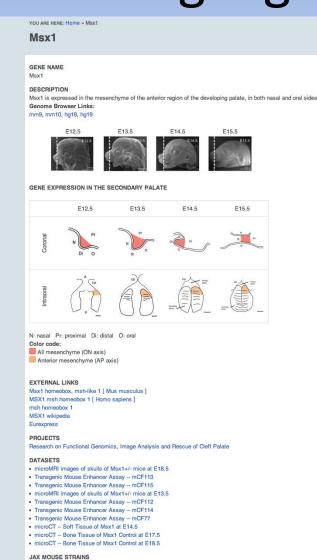


Research Highlight: Search Results





Research Highlight Details



B6.129S1-Osr2 /J

. STOCK Shh /J

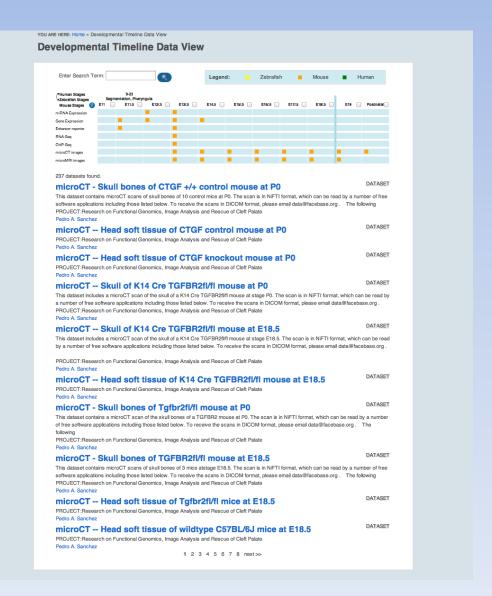
STOCK Tg(Wnt1-cre)11Rth Tg(Wnt1-GAL4)11Rth/J



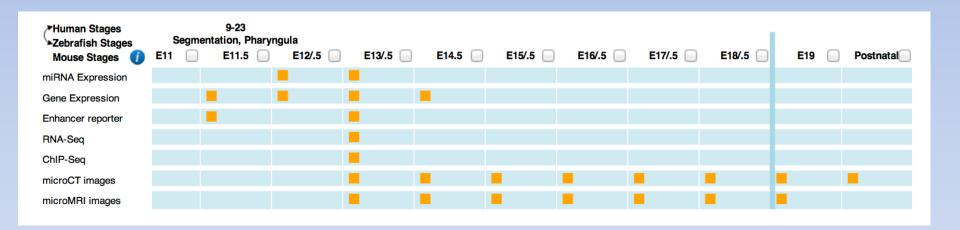
Timeline View

https://www.facebase.org/timeline

- Graphical view of datasets by developmental stage
- Approximate alignment of time points
- Navigate and filter results by time point

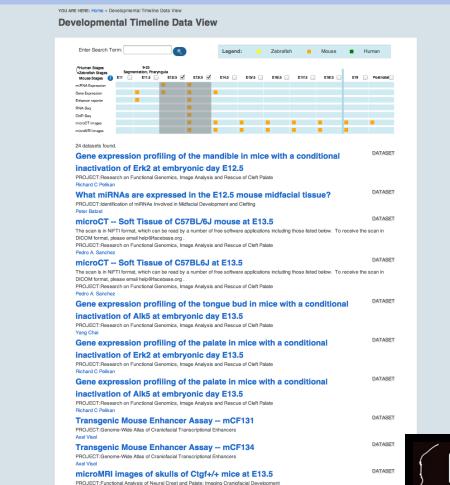


Timeline view - closeup





Timeline View – filter by developmental stage



Seth W Ruffins



Year 4 Highlights: General Site Enhancements

Annual meeting pages

Terms of use and data use certification

Statistics page

Cleft gallery

Working group pages

Administration views

Drupal updates

Site menu organization

Home page redesign

Data status log



YEAR FIVE PLANS

Year 5 Goals

- Final data uploads from spokes (by summer)
 - Continue data QC initiatives
 - Incorporate OCDM concepts
- Data integration
 - Finish year 4 projects
 - New modules (afternoon session)
- Develop analysis tools
 - Human data
 - Spoke specific
- Substantially increase the cases and data available through the Biorepository
- General site enhancement



Development Goals

- Data integration based on both metadata and analytic results
- Drill-down tools for extraction of individual files from within datasets
- Integration of OCDM for data annotation
- Deployment of fully functional IIP/Woolz
- Project-specific data displays



DATA

Data goals

- Upload remaining spoke data
- Continue to work with the PIs on improving data presentation
- Continue to update data process and supporting procedures: the ones focused on quality include:
 - Implement scientific quality assurance validation of 5% of the submitted data sets
 - Implement the standardized quality control verification procedure
- Incorporate non-FaceBase generated data on www.facebase.org



Data Quality Plans

- Data entry, approval, versioning, and deprecating procedures posted
- Quality control data curator verification of 100% of the submitted / approved
- Quality assurance scientific data validation of 5% of the approved data
- Continue to define and follow documented process and procedures
- Contact users for user experience improvements



DATA INTEGRATION

Towards a vision of meaningful integration

- Visualize the database network.
- Stimulate thoughts about scientific questions.
- "Gap analysis":
 - Find fruitful questions & hypotheses.
 - What gaps in the database network should be filled, to maximize the # of important questions made answerable?



Comparisons/Integration

Goal: focus on inter-spoke integration

Comparison type	Mapping type	Resource (examples)	
Time to time	Anatomic site	Imaging Expression Data	
Location to location	Adjacency	Ontology-dependent	
Species to species	Anatomic site	Ontology-dependent	
	Genetic homology	HomoloGene	
	Developmental stage	Carnegie Stage	
Platform to platform	Semantic identifier map	DAVID, Enfin,	



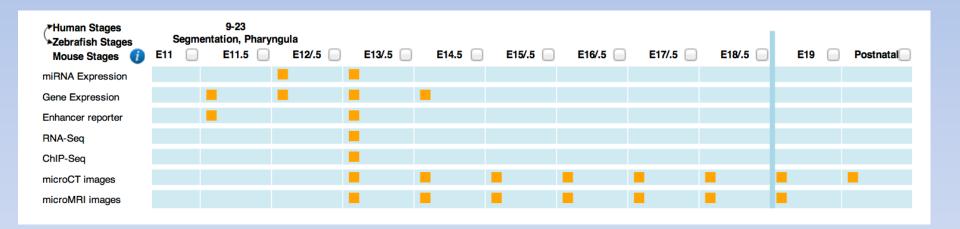
Data Integration

- CoGENE data incorporated into timeline view
- Microarray data Presentation
- Customized dataset across projects
- Gene WIKI

AFTERNOON SESSION



Timeline view - closeup





CoGENE dataset

- "Craniofacial and Oral Gene Expression Network"
 - Mike Lovett et al, web site formerly at Wash U
- Expression data generated with Affymetrix U95A and U95Av2 chips
- Data available on 17 different tissue types across several different embryological stages, as well as whole embryos from two stages - at least two replicates per tissue/stage
- Probe ids mapped to hg19 RefSeq genes to enable integration with other data sets

CoGENE Data – Human Craniofacial Development (M. Lovett)

Tissue/Timepoint	26 days (E10- 10.5)	4 weeks (E10.5)	5 weeks (E11.5)	6 weeks (E13)	8.5 weeks (E16)
1st pharangeal arch					
2nd pharangeal arch					
3rd (& 4th) pharangeal arch					
anterior rhombomere					
anterior tongue					
dental lamina					
frontal nasal prominence					
lateral nasal prominence					
lower lip					
mandible					
maxilla					
medial nasal prominence					
palatal shelf					
posterior rhombomere					
posterior tongue					
salivary gland					
upper lip					
whole embryo					



DATA SOFTWARE TOOLS

Spoke/Project Specific

- Gene mapping, sequencing
 - -GWAS
 - -PLINK, PSEQ
- MicroFace



Human Genomics Data

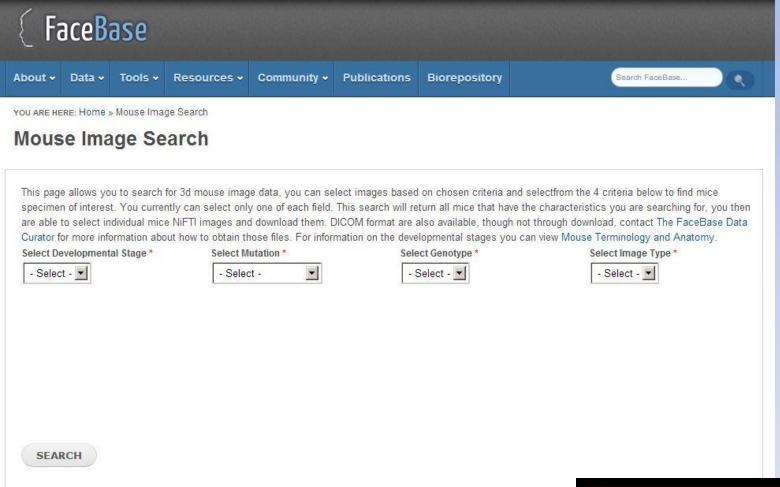
- Sequencing, GWAS
- dbGaP is the repository for individual-level data
- FaceBase will have a copy of the data (will not distribute data)
 - Develop analytical tools to do initial data mining analyses (simple stats, plus PLINK, PSEQ, etc) that can be implemented thorugh web site—only results returned
 - User can then chose to request individual level data from dbGaP

Viewers / Data Search Improvements

- Mouse image search module
- •IIP3D Woolz Viewer
 - -technical issues with file translation
- Side by side viewer comparison (Woolz and OPT)
- Image driven search and navigation
 - Annotated image for navigation

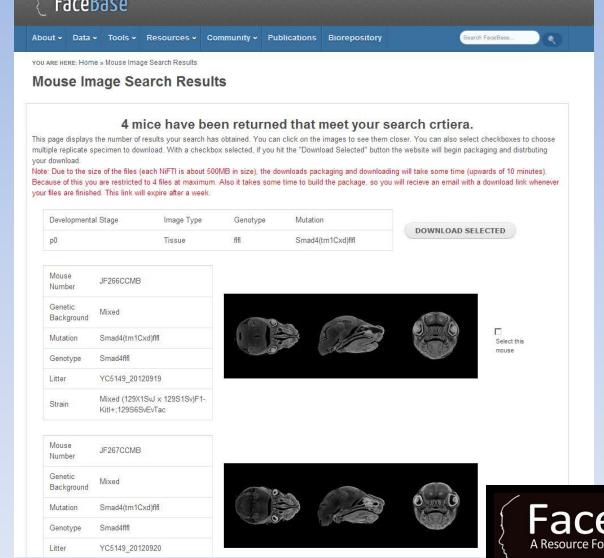


Mouse Image Module





Mouse image search results

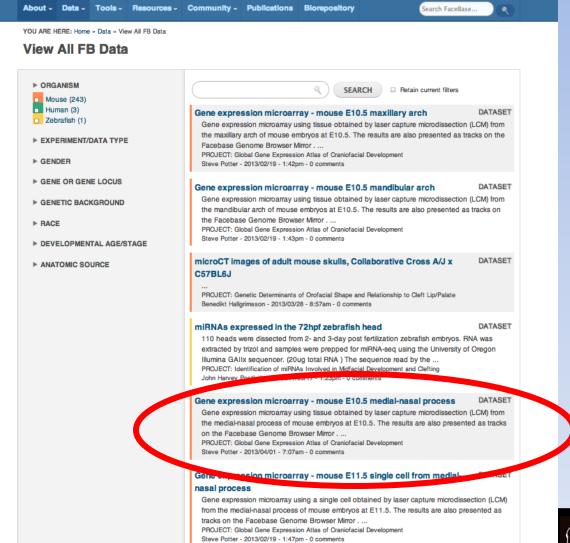


Side by side viewer



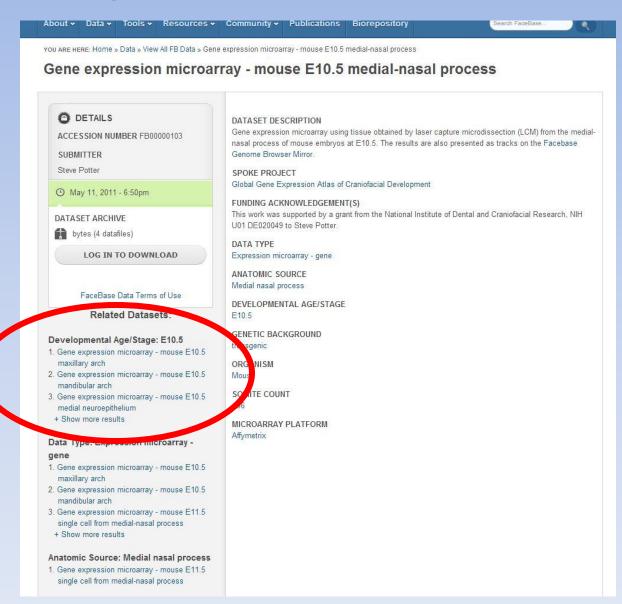


Data Search updates





Using metadata to select related datasets





BIOREPOSITORY

Available data Types

Biological Specimens

- DNA derived from-
 - Saliva
 - Cheek Swabs
 - Tissue
 - Cord Blood
 - Blood

Data Types

- 3D facial mesh data
- Ultrasound Video Data of Orbicularis Oris Muscle
- Dermatoglyphic Lip Print Image Data
- Intraoral Image data
- Video VPI data
- Laterality data
- Medical, Pregnancy,
 Demographic, Psychological
 Measures and Familial Health
 History Data

Team

Pittsburgh

- CCDG: Cristy Spino, Tom Maher, Lance Kennelty, Justin Stickle, Sean Schellinger, Annette Krag-Jensen
- DBMI: Harry Hochheiser, Roger Day, Chuck Borromeo, Shiyi
 Shen, Bill Shirey, Johnson Paul, John Milnes
- Other Pitt: Michael Barmada (Human Genetics)
- CMU: Mahadev Satyanarayanan

lowa

Nichole Nidey, Nancy Davin, Martine Dunnwald