

FaceBase: Data Resources for Craniofacial and Dental Development, Molecular Genetics, and Genomics

A. Bugacov¹, J. Feng², T. Guo², T-V. Ho², V. Nguyen², R. Schuler¹, C. Williams¹, P. Sedghizadeh², C. Kesselman¹, Y. Chai²

¹USC Information Sci. Inst., Marina del Rey, CA, ²Herman Ostrow Sch. of Dentistry of USC, Los Angeles, CA

INTRODUCTION

FaceBase (facebase.org), established in 2009, provides a freely available repository of data for the scientific community on dental, oral and craniofacial (DOC) development and diseases as well as in other anatomically or biologically relevant regions. FaceBase is a trusted source of research and educational resources across the translational spectrum on humans and model organisms.

FaceBase is supported by the National Institute of Dental and Craniofacial Research (NIDCR), National Institute on Deafness and Other Communication Disorders (NIDCD) and Office of Data Science Strategy (ODSS).

FaceBase is an NIH-supported resource for the dental, oral, and craniofacial (DOC) research community

- Boasts 1100+ datasets and growing – human, mouse, zebrafish, chick, and chimpanzee (other models are welcome!)
- Promotes multidisciplinary collaboration and research in development, molecular genetics, and genomics
- Integrates genomic and phenotype data from multiple species
- Includes research on the developmental biology and genetics of a variety of craniofacial structures and other areas of the head.
- Expanded focus to include clinical and public health data to embrace the entire translational spectrum
- Provides high-quality datasets to members of the research community for use in their own projects

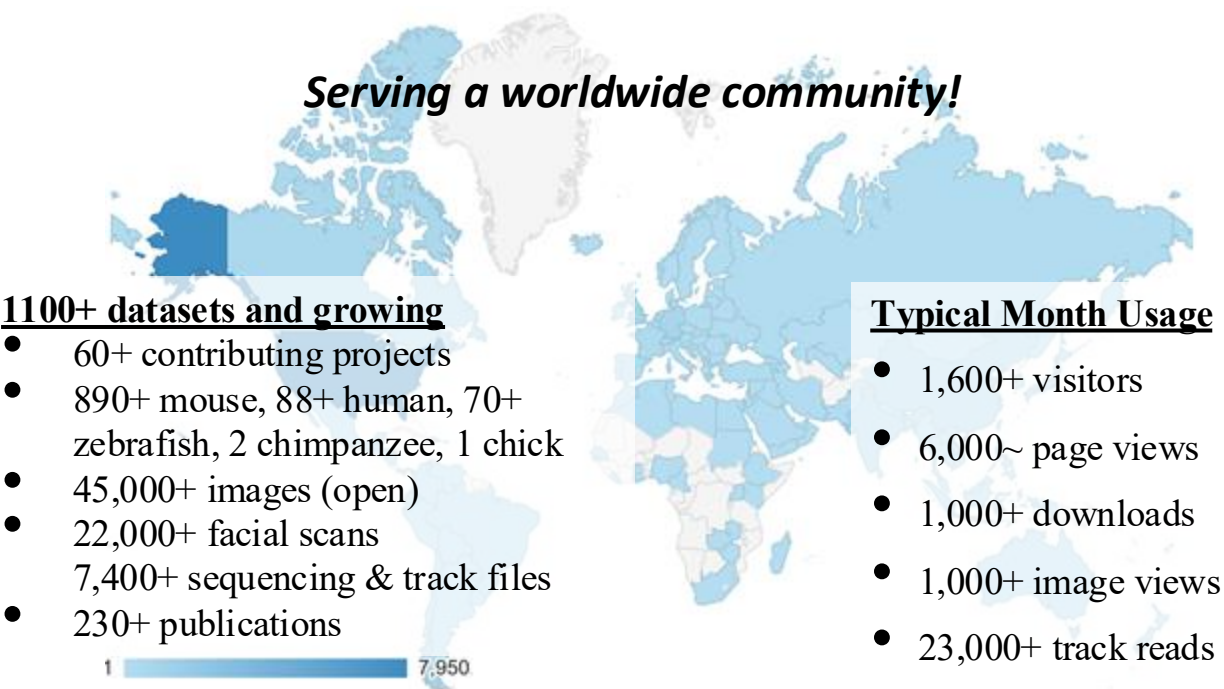


Figure 1. A growing, multi-faceted digital repository. FaceBase serves a worldwide community.

MATERIALS & METHODS

FaceBase is a highly scalable, Cloud-based (STRIDES) data sharing and analysis hub. Cloud storage provides extremely robust, cost-effective, virtually unlimited capacity and improves accessibility while ensuring long-term sustainability. FaceBase embraces TRUST (Transparency, Responsibility, User-focus, Sustainability, Technology) principles throughout to ensure the integrity of FAIR (Findable, Accessible, Interoperable, Reusable) data resources. FaceBase supports the full translational spectrum of research through enhanced data curation procedures for clinical and public health data while strengthening basic research data handling. FaceBase provides resources to help the community meet the new NIH Data Management & Sharing (DMS) requirements.



OBJECTIVES

FaceBase places a high priority on good data management and sharing practices and supports research and community-building through a Cloud-based, TRUST-worthy repository of high-quality FAIR data resources. FaceBase welcomes new contributions across the full spectrum of translational research.

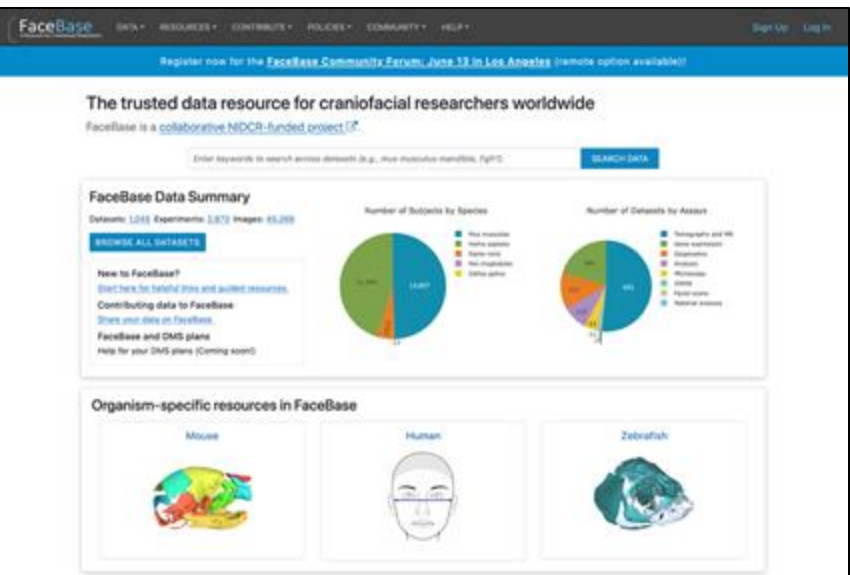


Figure 2. FaceBase Site. Cloud-based data services with a powerful online interface for searching, retrieving, and visualizing a broad and diverse collection of data resources.

FaceBase provides:

- Cloud-based (STRIDES) platform for data integration, analysis, and visualization
- High-quality datasets ready for Artificial Intelligence and Machine Learning (AI/ML) analyses
- Online visualization of single-cell and bulk omics, 2D and 3D imaging, surface meshes, and more
- Atlases of craniofacial development in multiple species, including humans
- Specialized resources such as the 3D Facial Norms database, Human Genome Analysis Interface (HGAI), and more
- Standardized terminologies and ontologies to accurately and precisely label and describe data
- User-friendly tools enabling data contributors to upload and annotate their data for re-use
- Proper data citations including Digital Object Identifiers (DOI) for all datasets

RESULTS

FaceBase includes 1,100+ datasets from human, zebrafish, mouse, chimpanzee, and chick, and seeks data on Xenopus and other model organisms. Data span micro-CT imaging, microscopy, single-cell and bulk RNA-seq, ChIP-seq, enhancer activity detection, exome sequencing, morphometric analysis, GWAS, and others. Data are findable via faceted search interfaces, accessible on the Web, interoperable using standardized vocabulary (e.g., UBERON, NCBI Gene, and MGI), and reusable with clear descriptions and detailed protocol documentation. FaceBase hosts the EnamelBase resource on morphological and functional data on tooth development, developed by NIDCR UH3-supported amelogenesis consortium. FaceBase's imaging and facial scan collections are ideally suited for AI/ML.

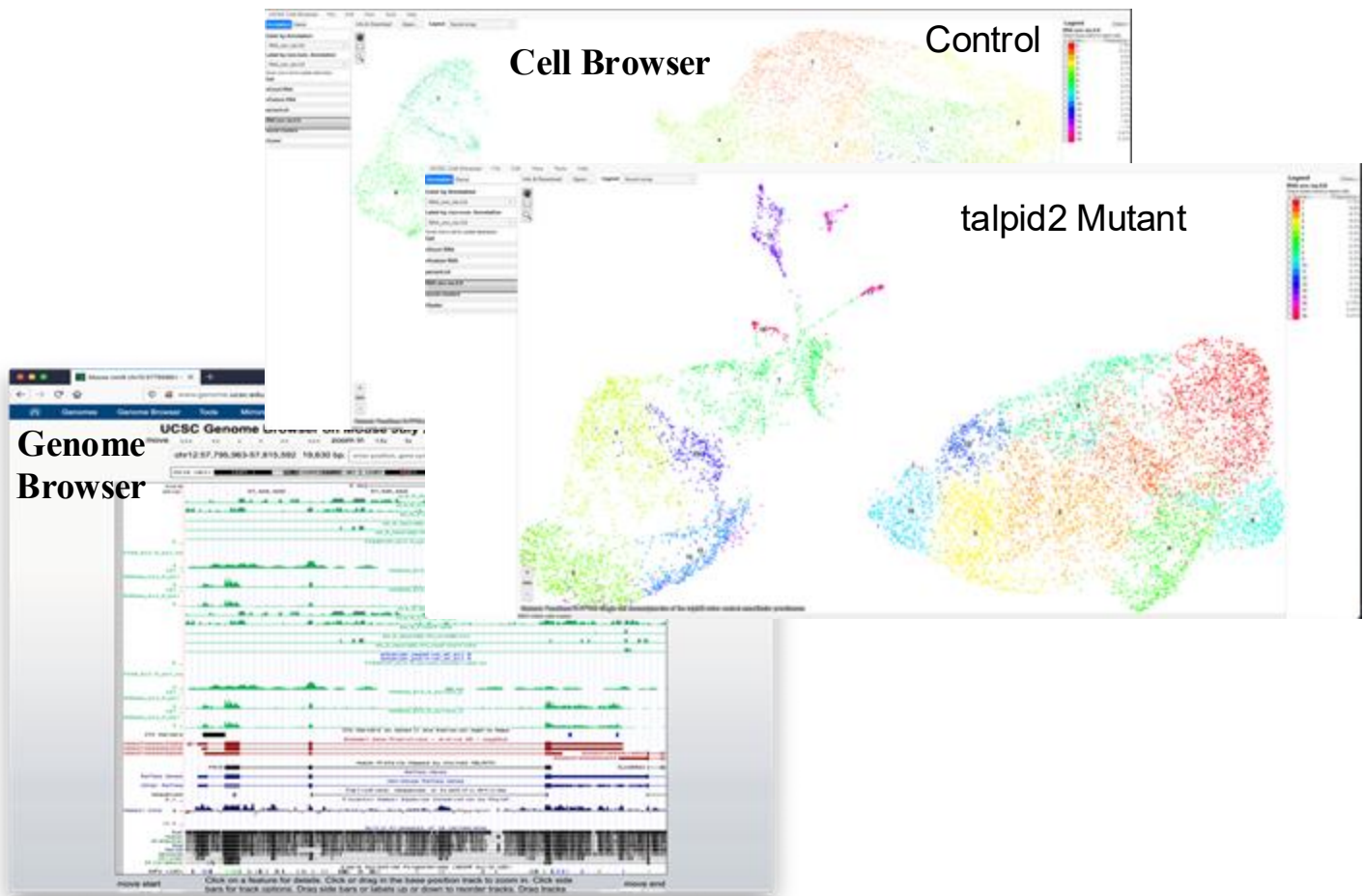


Figure 4. FaceBase genome (left) and single cell visualization (right). FaceBase operates a public, registered "track hub" integrated with the UCSC Genome Browser for visualization of bulk RNA-seq and ChIP-seq data. In addition, FaceBase generates visualizations of its single-cell RNA-seq data using the UCSC Cell Browser.

RESULTS (cont.)

FaceBase fosters a world-wide research community for data sharing that jumpstarts research. We encourage data submissions from the scientific community and accept a broad range of research data from basic to translational. Domain experts curate and integrate datasets to provide a high-quality, comprehensive resource for the broader community by making data FAIR – findable, accessible, interoperable, and reusable.

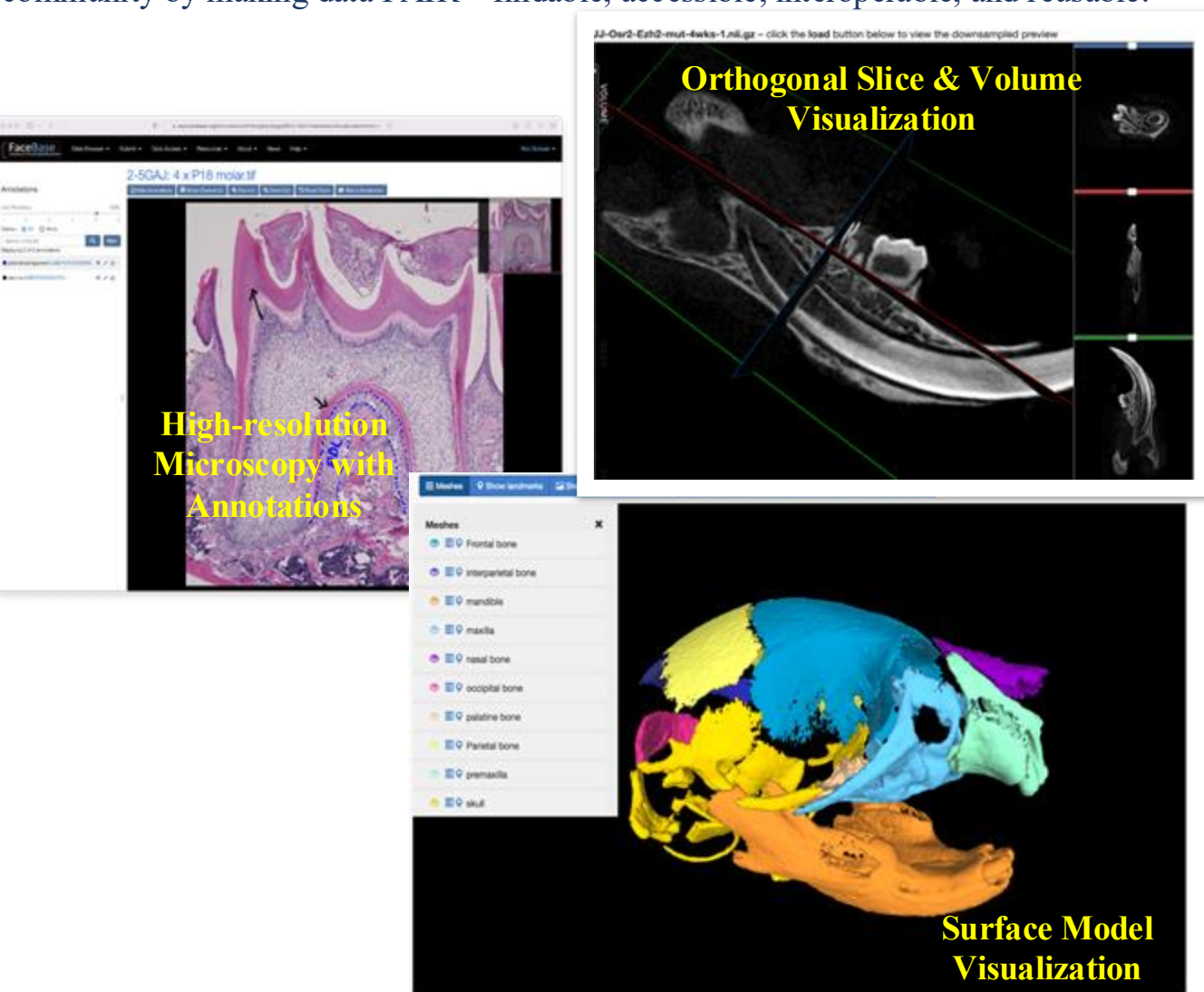
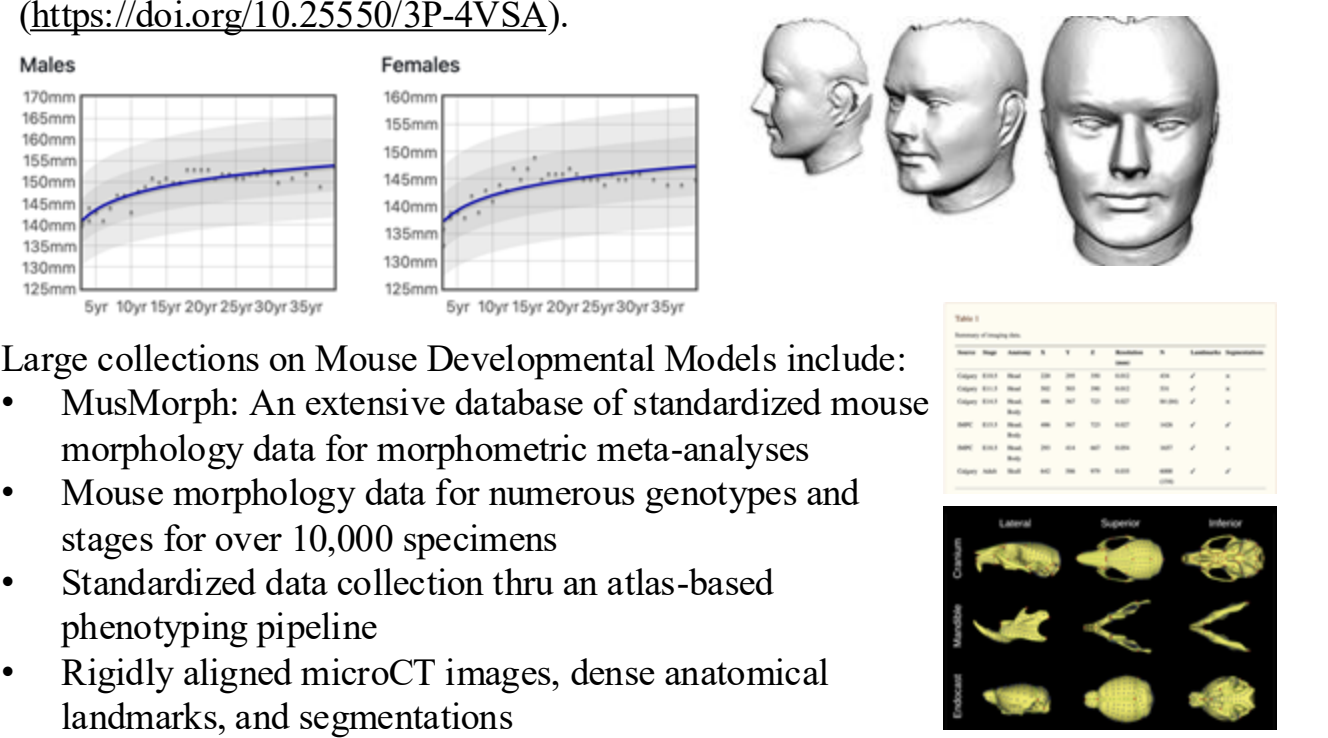


Figure 5. FaceBase visually-guided search and online imaging visualization. FaceBase provides an orthogonal slice viewer with volume visualization feature (top right); a high-resolution microscopy image viewer and annotation editor (left); and an interactive, fully customizable 3D surface model visualization (bottom right) that enables users to examine microCT data including anatomical landmarks.

EXAMPLE DATASETS

Large and diverse collections of high quality Human Facial Scans, Genomic and Morphometric data include: comprehensive studies to identify genetic determinants of orofacial shape, 3D facial surface scans and genotypic data; Facial Norms Database: Study of about 3000 healthy Caucasian subjects (<https://doi.org/10.25550/VWP>); and Tanzania Dataset: Study of 3600 normal subjects from Tanzania (<https://doi.org/10.25550/TX4>); Dental score data on over 500 patients (<https://doi.org/10.25550/3P-4VSA>).



EXAMPLE DATASETS (cont.)

Sampling of human health translational studies include:

Oral Health Research in Appalachia (<https://doi.org/10.25550/2Q-M86G>)

- Studies the covariation between genetic, behavioral, family, and community factors associated with oro-dental disease and tooth loss in rural and urban northern Appalachian families
- Cohort from about 800 households
- Panels of SNPs from DNA samples also available

Whole Exome and Whole Genome Sequencing of Craniofacial Microsomia Patients and Their Families (<https://doi.org/10.25550/49-7STP>)

- Sequencing data for a total of 591 individuals
- Controlled access data include an indexed BAM file for each participant and sequencing method and a joint VCF with genotype calls across all samples
- Public data include a sites-only VCF, a PED file indicating family relationships, and a bigWig file indicating coverage for each BAM

Genetics, Genomics and Deep Phenotyping of Orofacial Clefts

- (<https://doi.org/10.25550/51-YGX8>)
- Identifies etiologic factors, primarily genetic loci and variants, involved in orofacial cleft birth defects, including isolated cleft lip (CL), isolated cleft palate (CP), and CL with or without CP (CL/P)
 - Includes OFC probands at several sites in the United States and multiple international sites (Europe, Central/South America, Asia, and Africa).
 - Includes data from over 24,000 human participants
 - With 3D and 2D images, videos, and extensive survey data

CONCLUSIONS

FaceBase is home to a growing collection of data from the community across the full translational spectrum as we collectively build a comprehensive, trustworthy data repository and educational resource. FaceBase promotes multidisciplinary collaboration and research in dental, oral, and craniofacial (as well as other anatomically or biologically relevant) development and diseases, molecular genetics and genomics.

REFERENCES

- Schuler, R.E. et al. (2022) FaceBase: A Community-Driven Hub for Data-Intensive Research. J. Dent Res. 101(11): 1289-1298. PMID: 35912790.
- Samuels, B.D. et al. (2020) FaceBase 3: Analytical tools and FAIR resources for craniofacial and dental research. Development 147: dev19121.
- Brinkley, J.F. et al. (2016) The FaceBase Consortium: A comprehensive resource for craniofacial researchers. Development 143: 2677-2688.
- Hochheiser, H. et al. (2011). The FaceBase Consortium: A comprehensive program to facilitate craniofacial research. Dev Biol 355(2): 175-182.

ACKNOWLEDGEMENTS

FaceBase is supported by U24DE034163 from the NIDCR, NIDCD and ODSS.

Scan to visit [FaceBase.org](https://facebase.org)

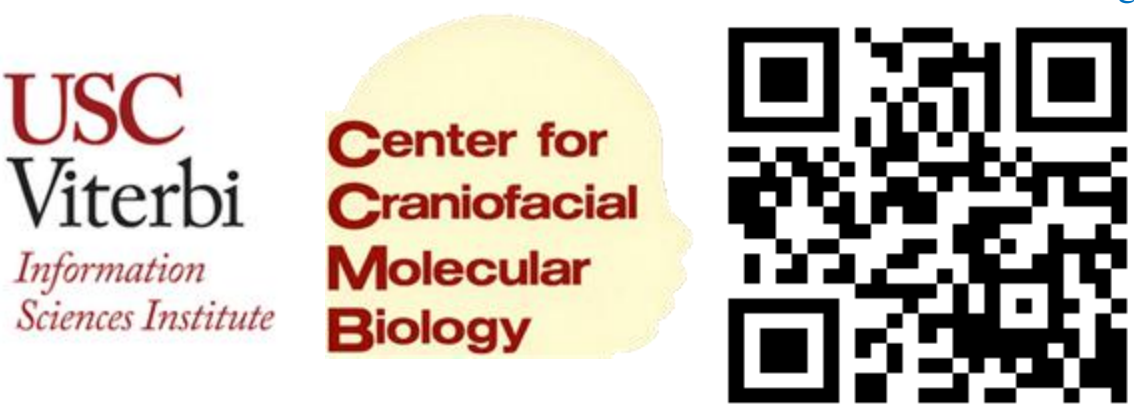


Figure 3. Data submission and publication pipeline. User-friendly online forms and graphical desktop applications, automated data processing and quality control. Release data with your manuscript publication.