© 2020. Published by The Company of Biologists Ltd | Development (2020) 147, dev191213. doi:10.1242/dev.191213

TECHNIQUES AND RESOURCES

RESEARCH ARTICLE

FaceBase 3: analytical tools and FAIR resources for craniofacial and dental research

Bridget D. Samuels[1], Robert Aho[2], James F. Brinkley[3], Alejandro Bugacov[4], Eleanor Feingold[5], Shannon Fisher[6], Ana S. Gonzalez-Reiche[7], Joseph G. Hacia[8], Benedikt Hallgrimsson[9], Karissa Hansen[2], Matthew P. Harris[10], Thach-Vu Ho[1], Greg Holmes[7], Joan E. Hooper[11], Ethylin Wang Jabs[7], Kenneth L. Jones[12], Carl Kesselman[4], Ophir D. Klein[13], Elizabeth J. Leslie[14], Hong Li[15], Eric C. Liao[16], Hannah Long[17], Na Lu[7], Richard L. Maas[18], Mary L. Marazita[5,19,20], Jaaved Mohammed[17], Sara Prescott[17], Robert Schuler[4], Licia Selleri[2], Richard A. Spritz[21], Tomek Swigut[17], Harm van Bakel[7], Axel Visel[22,23,24], Ian Welsh[2], Cristina Williams[4], Trevor J. Williams[15], Joanna Wysocka[17], Yuan Yuan[1] and Yang Chai[1,*]

ABSTRACT

The FaceBase Consortium was established by the National Institute of Dental and Craniofacial Research in 2009 as a 'big data' resource for the craniofacial research community. Over the past decade, researchers have deposited hundreds of annotated and curated datasets on both normal and disordered craniofacial development in FaceBase, all freely available to the research community on the FaceBase Hub website. The Hub has developed numerous visualization and analysis tools designed to promote integration of multidisciplinary data while remaining dedicated to the FAIR principles of data management (findability, accessibility, interoperability and reusability) and providing a faceted search infrastructure for locating desired data efficiently. Summaries of the datasets generated by the FaceBase projects from 2014 to 2019 are provided here. FaceBase 3 now welcomes contributions of data on craniofacial and dental development in humans, model organisms and cell lines. Collectively, the FaceBase Consortium, along with other NIH-supported data resources, provide a continuously growing, dynamic and current resource for the scientific community while improving data reproducibility and fulfilling data sharing requirements.

KEY WORDS: FaceBase Consortium, Craniofacial development, Data resource, Human, Mouse, Zebrafish

Introduction

Over the past decade, the biomedical fields have witnessed tremendous growth and technological advancement, driven in part by the exponential growth of 'big data' assets and the computational resources necessary to unlock their potential. The FaceBase Consortium, funded by the National Institute of Dental and Craniofacial Research (NIDCR) of the National Institutes of Health (NIH), was established in 2009 with the goal of enabling the craniofacial research community to share in this data revolution. FaceBase seeks to provide a comprehensive, trustworthy data repository that integrates innovative analysis and visualization tools with educational resources on craniofacial development. All aspects of FaceBase have been designed with the FAIR (findability, accessibility, interoperability and reusability) data principles in mind (Wilkinson et al., 2016). FaceBase promotes multidisciplinary collaboration and research in craniofacial development, molecular genetics and genomics by providing a platform for researchers to analyze, integrate and annotate datasets before and after they are published. The curated content available through FaceBase empowers the research community to leverage the tremendous resources developed by laboratories worldwide to accelerate their own hypothesis-driven basic, translational and clinical research. Now, FaceBase is opening its doors to researchers who wish to make their datasets accessible through the Hub and take advantage of the toolkit it provides for analyzing, visualizing and integrating numerous data types.

Throughout its first (2009-2014) and second (2014-2019) iterations, FaceBase 1 and FaceBase 2, the Consortium operated as a 'Hub and Spoke' model, with 10-11 Spoke Projects

[1]Center for Craniofacial Molecular Biology, Herman Ostrow School of Dentistry, University of Southern California, Los Angeles, CA 90033, USA. [2]Program in Craniofacial Biology, Departments of Orofacial Sciences and of Anatomy, Institute of Human Genetics, University of California San Francisco, San Francisco, CA 94143, USA. [3]Structural Informatics Group, Department of Biological Structure, University of Washington, Seattle, WA 98195, USA. [4]Information Sciences Institute, Viterbi School of Engineering, University of Southern California, Marina del Rey, CA 90292, USA. [5]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15219, USA. [6]Department of Pharmacology and Experimental Therapeutics, Boston University School of Medicine, Boston, MA 02118, USA. [7]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [8]Department of Biochemistry and Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA. [9]Department of Cell Biology and Anatomy, Alberta Children's Hospital Research Institute, and McCaig Bone and Joint Institute, University of Calgary, Alberta, Canada. [10]Department of Orthopedic Research, Boston Children's Hospital and Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [11]Department of Cell and Developmental Biology, School of Medicine, University of Colorado, Aurora, CO 80045, USA. [12]Department of Biochemistry and Molecular Genetics, School of Medicine, University of Colorado, Aurora, CO 80045, USA. [13]Program in Craniofacial Biology, Departments of Orofacial Sciences and Pediatrics, Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94143, USA. [14]Department of Human Genetics, Emory University, Atlanta, GA 30322, USA. [15]Department of Craniofacial Biology, School of Dental Medicine, University of Colorado, Aurora, CO 80045, USA. [16]Massachusetts General Hospital, Plastic and Reconstructive Surgery, Boston, MA 02114, USA. [17]Departments of Chemical and Systems Biology and of Developmental Biology, Howard Hughes Medical Institute, School of Medicine, Stanford University, Stanford, CA 94305, USA. [18]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [19]Center for Craniofacial and Dental Genetics, Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA. [20]Clinical and Translational Science, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA. [21]Human Medical Genetics and Genomics Program, School of Medicine, University of Colorado, Aurora, CO 80045, USA. [22]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [23]U.S. Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [24]School of Natural Sciences, University of California Merced, Merced, CA 95343, USA.

*Author for correspondence (ychai@usc.edu)

B.D.S., 0000-0003-2712-1008; B.H., 0000-0002-7192-9103; K.H., 0000-0003-0352-5980; H.v.B., 0000-0002-1376-6916; Y.C., 0000-0003-2477-7247

Handling Editor: James Briscoe
Received 2 April 2020; Accepted 13 August 2020

Development (2020) 147, dev191213. doi:10.1242/dev.191213

independently selected through a peer review process to generate and share data during each of these 5-year periods. These Spokes received NIDCR support to generate datasets made available online through the Hub portal (facebase.org). The FaceBase 1 Spoke Projects, which focused on midface development in humans and animal models, have been described previously (Hochheiser et al., 2011). FaceBase 2 expanded its scope to craniofacial development more broadly (Brinkley et al., 2013). NIDCR has also supported secondary analyses of FaceBase datasets through the R03 mechanism (PAR-13-178 and PAR-16-362).

To date, FaceBase includes over 880 datasets on human, mouse, zebrafish and chimpanzee prenatal and postnatal development, including both typically and atypically developing individuals, which are available to the scientific community. These datasets represent a wide range of experiment types, including ATAC-seq, ChIP-seq, bulk and single-cell RNA-seq, two- and three-dimensional imaging, genome-wide association studies (GWAS), and accompanying metadata, as described in the sections to follow. Many of these datasets are interactive and enable users to perform their own custom analyses, thanks to the innovative web browser-based tools integrated into the Hub. The Hub's infrastructure and tools serve to integrate genomic and phenotypic data from multiple species. Moreover, FaceBase provides an ideal platform for collaborations through pre-publication access control, data curation tools, emphasis on reproducibility and integration across datasets. Approximately 200 publications to date refer to FaceBase datasets and other Hub resources.

FaceBase is now moving beyond the Hub and Spoke model. As of Autumn 2019, the Hub welcomes contributions of data relevant to the craniofacial community from all researchers. The Hub team is strongly committed to providing the training and resources necessary to enable researchers to upload and curate their own data in a user-friendly, efficient and scalable manner. A list of priorities for data recruitment over the next year have been identified, including expansion to include (among others) data on dental and salivary gland development, *Xenopus* and chick models, single-cell RNA sequencing, and characterization of cell lines pertinent to orofacial tissues (see https://www.facebase.org/submit/data-priorities/). In the sections that follow, we review the datasets deposited by the FaceBase 2 Spoke Projects, describe data analysis and visualization tools available through the Hub, and outline our vision for the future development and expansion of FaceBase as a dynamic nexus of craniofacial research.

FaceBase 2 spoke projects

We first provide brief overviews of the FaceBase 2 Spoke Projects that focused on generating data on craniofacial development in animal models, then on those that focused on human development or both humans and animal models.

Anatomical atlas and transgenic tools for late skull formation in the zebrafish (https://doi.org/10.25550/1WW2)

The zebrafish is an important model organism for studying human disease processes, including craniofacial abnormalities. This has been driven in part by the development of sophisticated tools for genome manipulations, largely based on CRISPR technology and rivaling those established in the mouse. Those advances, however, built on a foundation of traditional forward genetics, and on the expansive collected descriptions of normal and abnormal development in mutants. Together, these genetic tools have led to the development of a large collection of mutants and transgenic lines relevant to craniofacial development. Another advantage of zebrafish is its availability for observations and experimental manipulations throughout development. This includes the critical window of time of neurocranium formation for which the mouse is largely inaccessible for live observations. Nevertheless, leveraging these tools has been limited, largely from lack of detailed information on zebrafish cranial structure and form.

Shannon Fisher and Matthew Harris directed their Spoke Project with the aim of filling an important gap in our knowledge of zebrafish development, providing information about skull development in the time between late larval stages and the adult, providing reliable quantitative data on growth and morphogenesis of the craniofacial skeleton. This Spoke focused on acquiring two types of data: (1) confocal images of fluorescent transgene expression in live juvenile fish during the development of the cranium and (2) high-resolution microCT analysis of adult skulls for analysis of form and dysmorphology in defined genetic backgrounds. They analyzed both healthy control and mutant fish, focusing on genes relevant to human craniofacial disorders. All the raw data are available on the FaceBase Hub, as well as analysis of the anatomy of developing and adult fish, and annotated models of the adult zebrafish skull (Fig. 1). These datasets have been optimized to enable visualization of thumbnail images and interactive 3D views in any modern web browser.

One benefit of this work stems from the detailed description of transgene expression patterns during skull development. These include a reporter line for chondrocytes (Kanther et al., 2019), several for osteoblasts (Kanther et al., 2019) and most recently for osteoclasts (Caetano-Lopes et al., 2020). These lines are available for distribution through the community and provide a platform for investigating cell and tissue dynamics during development. The confocal and microCT data are also important illustrations of normal development and allow consistent comparisons across labs. Importantly, the developmental data are presented within a framework based on anatomical landmarks rather than chronological age, allowing standardization across different environments and rearing conditions (Parichy et al., 2009). In gathering the data, the team developed tools and approaches that have broader applicability in the community. These include low-magnification confocal imaging to document dynamic skull morphogenesis in live fish and the development of sensitizing stains to allow microCT imaging at earlier stages (Charles et al., 2017). There are also preliminary descriptions of mutants involving genes implicated in human diseases, including osteogenesis imperfecta and craniosynostosis (Gistelinck et al., 2018; Henke et al., 2017; Kague et al., 2016).

To facilitate use of the zebrafish model, the team created an annotated atlas of the complex anatomy of the zebrafish skull. It is challenging to draw parallels between the 74 separate ossifications of the adult zebrafish skull and the 22 bones of the mammalian skull, hindering the full application and appreciation of zebrafish as a model for human disease. To illustrate the developmental and anatomic analogies, they generated interactive 3D models, including one designed to facilitate comparison with the mouse (Ho et al., 2015). This atlas will be particularly valuable in evaluating potential zebrafish models for human craniofacial abnormalities. Moreover, it lays the groundwork for similar data collections on other fish species, including genetic model systems, such as medaka, cavefish and stickleback. As fishes comprise over half of all vertebrates, comparing phenotypes and differential responses to genetic and environmental perturbation across fish species can yield insight into dynamics of skeletal development (Witten et al., 2017). Acanthamorph fishes, e.g. medaka, are

Development (2020) 147, dev191213. doi:10.1242/dev.191213

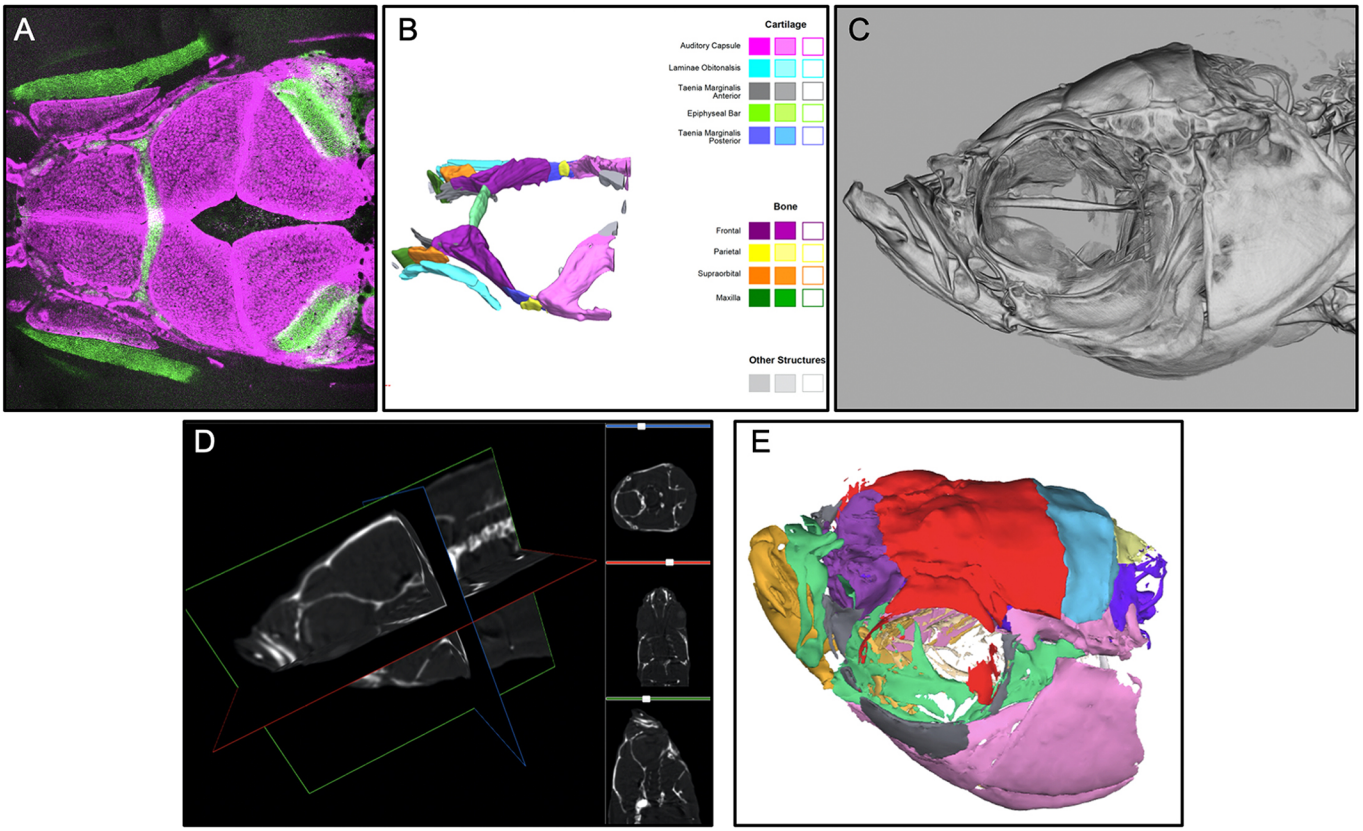

Fig. 1. Images and models of zebrafish craniofacial anatomy. (A) Confocal stack of the skull of a wild-type zebrafish at 11.83 mm standard length (41 days post fertilization); osteoblasts and chondrocytes are marked by expression of mCherry and eGFP, respectively. (B) 3D PDF model based on similar confocal data, showing the labeled buttons to hide or reveal individual elements. (C) High-resolution microCT of an adult zebrafish skull. (D) Orthotopic slices of the same data from the FaceBase online viewer. (E) 3-D model based on the microCT data. Different colors indicate distinct bones.

anosteocytic, or lack osteocytes (Davesne et al., 2019). Interestingly, these fishes can remodel bone and respond to strain similarly to mammals, a trait previously attributed to osteocytes (Ofer et al., 2019a,b). Additionally, zebrafish do not have oral teeth, whereas other fishes such as medaka, cichlids and sticklebacks retain them. All of these systems show replacement teeth and provide exceptional models for understanding tooth regeneration and repair (Fraser et al., 2009; Tucker and Fraser, 2014; Hulsey et al., 2016; Witten et al., 2017). Given that all teleost fishes share an ancestral whole-genome duplication, the differential retention and sub-functionalization of gene pairs provides unique windows to understand how the skull is formed and how it may vary (Harris et al., 2014; Caetano-Lopes, et al., 2020).

Transcriptome atlases of the craniofacial sutures (https://doi.org/10.25550/1WW8)

Normal human craniofacial development requires the integrated growth of the 22 bones of the human skull. These bones meet along their edges at sutures, which are major sites of bone growth. Sutures consist of osteogenic fronts (OFs), where preosteoblasts proliferate and differentiate to bone, and intervening suture mesenchyme (SM). Mutations in numerous genes, affecting many signaling pathways and biological processes, perturb suture development and result in a range of craniofacial dysostoses, including many forms of craniosynostosis in which sutures fuse prematurely. Sutures differ widely in their physical structure, cell lineage, mechanical environment and susceptibility to craniosynostosis (Heuzé et al., 2014; Richtsmeier and Flaherty, 2013).

Thorough knowledge of the transcriptional profiles of sutures is required to conduct hypothesis-driven research about their role in craniofacial development and dysgenesis. The project led by Greg Holmes, Harm van Bakel and Ethylin Wang Jabs provides murine RNA-seq datasets derived by laser capture microdissection from 11 craniofacial sutures at multiple embryonic ages to address this need (Fig. S1). In addition to providing RNA-seq datasets of wild-type mice, they include Apert and Saethre-Chotzen craniosynostosis syndrome models to allow study of premature suture ossification. The team also employed single-cell RNA-seq (scRNA-seq) analysis to identify heterogeneous cell types. The four major calvarial sutures (coronal, lambdoid, frontal and sagittal) were assayed via scRNA-seq in wild-type mice at E18.5 and postnatal days (P)10 and P28. These complement and extend the bulk RNA-seq atlases to postnatal ages at which stem cell populations have been identified in suture mesenchyme (Holmes et al., 2020a; Zhao and Chai, 2015). Collectively, these datasets provide a rich gene expression reference for gene discovery projects of interest to the wider scientific community, exemplified by other FaceBase projects. These include genes implicated in craniofacial defects uncovered in human genomics data from GWAS surveys or identified clinically, as well as genes identified in the craniofacial development of other species such as zebrafish. The expression of genes identified in such projects can be mapped to OFs or SM in the bulk RNA-seq datasets or to individual suture cell types within the scRNA-seq datasets. In addition, these datasets can be analyzed for differential gene expression and network analyses among various sutures, subregions, developmental stages, and wild-type and mutant

Development (2020) 147, dev191213. doi:10.1242/dev.191213

genotypes to identify novel biological, cellular, and molecular processes and pathways involved in normal skull development and craniosynostosis (Holmes et al., 2020b).

Genomic and transgenic resources for craniofacial enhancer studies (https://doi.org/10.25550/1WW4)

Genetic studies have shown that distant-acting regulatory sequences (enhancers) embedded in the vast non-coding region of the human genome play important roles in craniofacial development and susceptibility to craniofacial birth defects. However, the genomic locations and *in vivo* functions of most craniofacial enhancers remain unknown. During FaceBase 1, Axel Visel's Spoke Project generated the first sets of annotation and functional data for distal enhancers controlling craniofacial development (Attanasio et al., 2013). In FaceBase 2, they aimed to characterize the gene regulatory landscape of craniofacial development more comprehensively. To map predicted enhancers, they used ChIP-seq for a panel of histone modifications that are informative for the chromatin states of noncoding genomic regions. They also obtained ATAC-seq data for subsets of the samples investigated to map accessible chromatin. They applied these methods to all subregions of the developing mouse face at three stages of embryonic development, as well as to human embryonic face tissue to identify human-specific craniofacial enhancers not functionally conserved in mice.

These studies enable the targeted interrogation of genetic loci of interest for the presence of candidate enhancers, which may include the regulatory landscapes of genes known to be involved in craniofacial development, as well as non-coding risk intervals for craniofacial birth defects identified in genome-wide association and whole-genome sequencing studies. To enable in-depth studies of individual candidate sequences, the team used a transgenic mouse *in vivo* reporter system to determine the activity of individual enhancer sequences during crucial stages of embryonic development. Importantly, this system can also be used to evaluate the impact of specific human sequence variants (e.g. those associated with orofacial clefts) within known craniofacial enhancers. To enable the analysis of *in vivo* enhancer reporter activity patterns in three-dimensional space, the project also performed optical projection tomography (OPT) analysis of transgenic reporter embryos (Fig. S2).

The datasets generated by this Spoke have already demonstrated their utility. Uslu et al. used transgenic reporter data generated by this project for a more detailed in-depth exploration of the noncoding major orofacial clefting interval at the *Myc* locus (Uslu et al., 2014). Prescott et al. used a collection of craniofacial enhancers identified and characterized by the Visel Spoke in a study of regulatory divergence of neural crest enhancers between chimpanzees and humans (Prescott et al., 2015). Shaffer et al. used enhancer data generated by this Spoke to identify a significant association between cleft palate and a branchial arch enhancer at the *FOXP1* locus (Shaffer et al., 2019). Finally, Carlson et al. used data from this Spoke to examine the regulatory basis of phenotypic modifiers of non-syndromic cleft lip with or without cleft palate (Carlson et al., 2017).

Importantly, during FaceBase 2, the Visel team developed unified processing workflows for RNA-seq and ChIP-seq data that applies standardized ENCODE pipelines to datasets generated by different researchers, enabling comparative analysis across data from different labs. This analytical workflow has been applied to FaceBase 2 datasets from multiple Spokes and will be available to future FaceBase submitters as a service provided by the Hub.

Integrated research of functional genomics and craniofacial morphogenesis (https://doi.org/10.25550/1WWE)

The principal goal of developmental biology is to understand how tissues are induced and patterned to generate different organs with the correct temporal and spatial specificity. Multiple molecules have been identified as crucial regulators of craniofacial morphogenesis; however, the challenge remains to determine how various signaling centers coordinate to build the complicated structures that make our face. It is necessary to integrate multiple types of data to reveal the signaling networks that guide craniofacial morphogenesis. Using this type of multifaceted approach, the Spoke led by Yang Chai established correlations between gene expression, cell lineage analysis and morphogenesis of mandible and maxilla, which will lead to new discoveries of molecular regulatory mechanisms of craniofacial development. Expanding on their work on palatogenesis in FaceBase 1, the Chai FaceBase 2 Spoke focused on jaw morphogenesis because deformities of the mandible and maxilla are relatively common and can affect the development of other facial structures; to take one example, maxillary hypoplasia is often associated with cleft palate and has been described in more than 60 syndromes (Hennekam et al., 2010; Jin et al., 2012). Despite their importance, the mechanisms that regulate facial bone development have not been well characterized.

The Chai Spoke generated comprehensive datasets of gene expression and dynamic imaging analyses during mandible, maxilla and palate development. Available on the Hub are global and specific gene expression profiling studies of mandible and maxilla development using microarray, RNA-seq and *in situ* hybridization analyses. The datasets include both healthy controls and mutant models with altered TGFβ signaling (e.g. *Wnt1-Cre;Tgfbr2*$^{fl/fl}$ and *Wnt1-Cre;Alk5*$^{fl/fl}$) at E10.5, E11.5, E12.5, E13.5 and E14.5. In parallel, this Spoke generated micro-computed tomography (microCT) images, highlighting the spatiotemporal morphogenesis of the mandible and maxilla in these mutant models and controls at E14.5, E16.5, E18.5 and newborn stages; these images may be of broad interest for studies on craniofacial development and malformations as they include the hard and soft tissues of the entire head.

These datasets have promoted the generation of novel hypotheses and collaboration with other Spoke Projects to investigate the role of TGFβ signaling in regulating craniofacial development and how modulation of defined signaling pathways may be beneficial for the prevention of craniofacial malformations (Iwata et al., 2012; Oka et al., 2007; Pelikan et al., 2013; Sugii et al., 2017). To verify the reproducibility of the results, gene expression microarray and RNA-seq data from the Chai Spoke were cross-validated against datasets from the Williams Spoke. Users can view specific gene expression patterns and correlate them with cellular contributions, such as cranial neural crest cells, myogenic cells and other cell types, during mandible and maxilla development (Chai et al., 2000; Chai and Maxson, 2006). The 3D microCT imaging datasets allow users to rotate the skull through 360 degrees and perform digital dissections (Fig. 2). Users can also isolate the mandible or maxilla and perform deep phenotyping using well-defined anatomical landmarks, which were developed in concert with the Ontology of Craniofacial Development and Malformation Spoke (described in a later section).

RNA dynamics in the developing mouse face (https://doi.org/10.25550/1WW6)

Embryonic development frequently requires the precise coordinated interaction of cell types of different origin that have distinct gene expression signatures. Face formation is no exception, and although

Development (2020) 147, dev191213. doi:10.1242/dev.191213

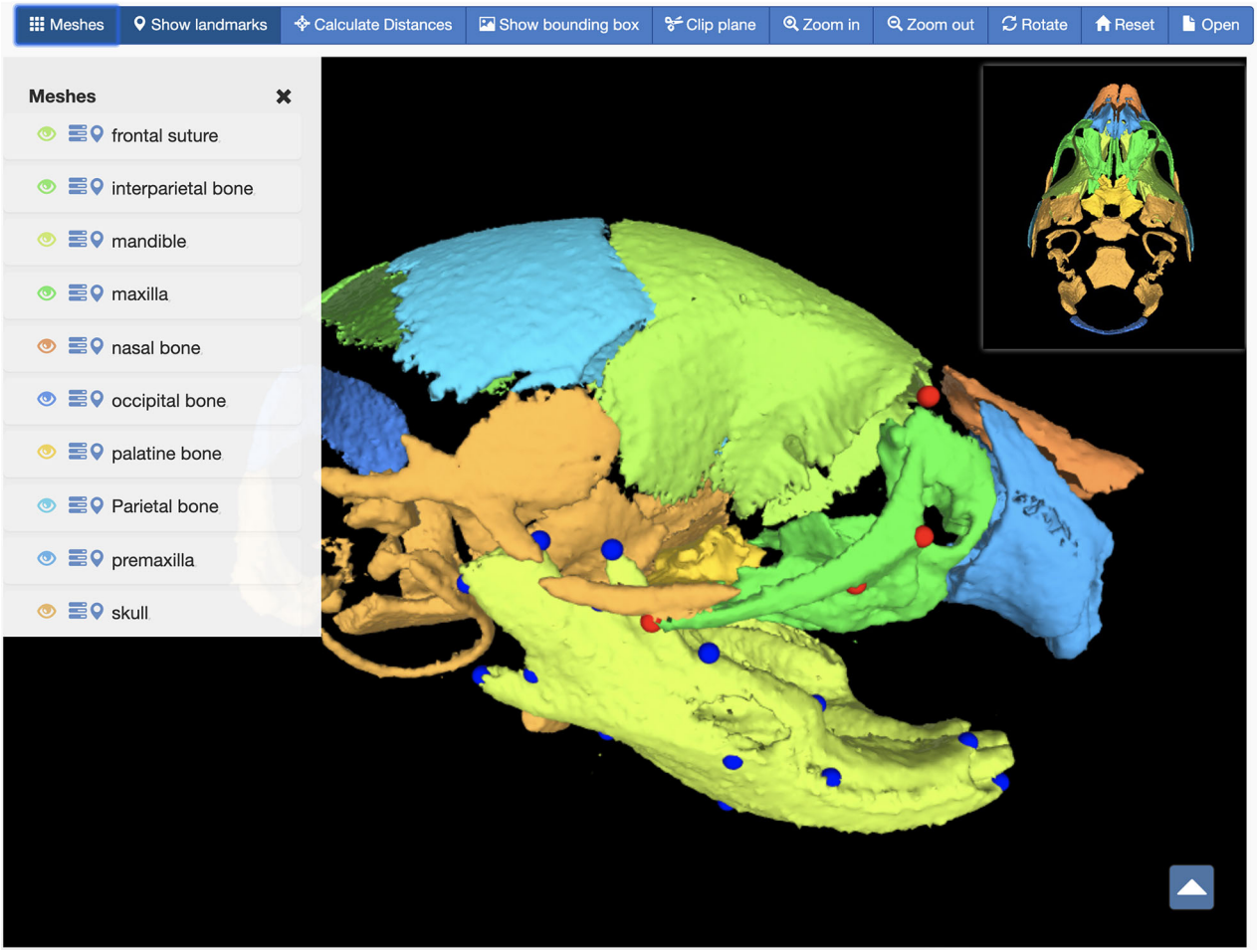

Fig. 2. Interactive 3D model of an E18.5 typically developing mouse skull based on microCT data. Different colors indicate distinct bones. Blue dots indicate anatomical landmarks of the mandible; red dots indicate anatomical landmarks of the maxilla. Descriptions of landmarks are provided in the 'Show landmarks' menu of the FaceBase 3D mesh viewer. Inset shows digital dissection of the same model, performed using the 'Rotate' and 'Clip plane' functions. FaceBase Record ID 3V4A.

most of the mammalian face is derived from the neural crest, correct growth, patterning and morphogenesis relies on reciprocal signaling with adjacent tissues including the neural tube, endoderm and ectoderm. The Spoke led by Joan Hooper, Kenneth Jones, and Trevor Williams developed innovative methods using microdissection and enzymatic digestion to isolate the ectodermal and mesenchymal components of the developing facial prominences for analysis (Li and Williams, 2013). Their focus was on E10.5-E12.5, the period most relevant to understanding the gene networks that operate during normal facial fusion, but which are disrupted in clefting of the lip and primary palate, one of the most frequent human birth defects (Dixon et al., 2011). Furthermore, by isolating each prominence separately, information could be obtained concerning the unique expression profiles present in different regions of the developing face (Fig. S3). The team also processed previous microarray gene expression data from these stages of mouse facial development (Feng et al., 2009; Hooper et al., 2017) to produce an indexed list of every gene detected so that individual expression profiles in the ectoderm and mesenchyme of each facial prominence can be readily visualized (Leach et al., 2017).

One major part of the project used RNA-seq analysis to study expression across 20 triplicate samples representing different prominences, layers and ages (Hooper et al., 2020). In addition, a custom mouse microarray designed in concert with Affymetrix was used to assess miRNAs, rRNAs, tRNAs, snRNAs and snoRNAs present in the developing face, and revealed differences in spatiotemporal expression. All these RNA-seq, gene expression microarray and expression profile datasets are available via FaceBase. The RNA-seq datasets have sufficient depth for robust analysis of differential splicing, promoter and poly A site use (see Fig. S3C). Differences in splicing across age stages and between the ectoderm and mesenchyme were particularly prevalent, correlating with the importance of differential splicing effectors, such as *Esrp1*, *Esrp2* and *Rbfox2*, that display tissue-specific expression and cause major defects in face formation when mutated (Bebee et al., 2015; Cibi et al., 2019; Lee et al., 2020; Warzecha et al., 2009). Finally, with the advent of scRNA-seq, it became possible to address crucial ectodermal and mesenchymal cell populations, as well as gene expression programs, that are associated with the fusing lambdoid junction at E11.5, the time point when fusion of the lateral and medial nasal prominences, together with the maxillary prominences, forms the upper lip and primary palate. This study revealed distinct gene expression programs associated with the regions of fusing epithelial seams, both within the ectoderm and mesenchyme, as well as changes in the distribution of periderm at the sites of fusion (Li et al., 2019). Fig. S3D shows the expression levels of the four genes involved in orofacial clefting noted in Fig. S3B – *Irf6*, *Rspo2*, *Sumo1* and *Bmp4* – as feature plots overlaid

Development (2020) 147, dev191213. doi:10.1242/dev.191213

on the t-distributed stochastic neighbor embedding (tSNE) plot of cell populations associated with the fusing epithelial seam, reiterating their varied distributions.

These datasets could be mined for studies including: (1) how individual genes identified by human clinical studies or from model system genetic analyses are expressed during mouse facial development to develop hypotheses concerning functional relevance or molecular mechanisms of action; (2) correlating splicing or poly A addition differences with the expression of RNA binding proteins, differential promoter use with transcription factor expression, and 5′ and 3′ UTR isoform differences with potential miRNA binding to identify the genetic programs and regulatory interactions that underlie facial morphogenesis; (3) determining how changes in splicing and/or promoter usage might impact the functionality of related transcripts and protein isoforms; (4) developing a systems-level analysis of gene interactions during facial development, including signaling interactions and transcription responses that occur within and between adjacent tissues; (5) investigating early stages in the development of distinct expression programs within the olfactory epithelium as it separates from the surface ectoderm; and (6) using the control datasets as a baseline to understand how cell populations and associated gene expression are altered in mouse models of facial dysmorphology.

Epigenetic landscapes and regulatory divergence of human craniofacial traits (https://doi.org/10.25550/1WWG)

Cranial neural crest cells (CNCCs) play major roles during development in establishing craniofacial morphology and determining species-specific variation. To understand distinctive human facial features, it is crucial to study human CNCCs and their derivatives in addition to neural crest from model organisms. Furthermore, although many genes and pathways involved in CNCC formation and differentiation are conserved across species, the non-coding sequences involved in gene regulation are often species specific.

The Spoke team led by Joanna Wysocka and Licia Selleri established and validated human pluripotent stem cell differentiation models that recapitulate induction, migration and differentiation of CNCCs *in vitro*, and facilitate modeling of human neurocristopathies (Bajpai et al., 2010; Bowen et al., 2019; Calo et al., 2018; Rada-Iglesias et al., 2012). This represents a major advance in our understanding of human neural crest formation, which occurs at 3 to 6 weeks of gestation and is largely inaccessible to molecular studies. The team extended their model to chimpanzee CNCCs, enabling the identification of molecular features that distinguish human cells from those of our closest living relatives (Prescott et al., 2015). They characterized epigenetic landscapes and transcriptomes of human and chimpanzee CNCCs, and provided genome-wide annotations of candidate regulatory elements, both those that are conserved and those that functionally diverged more recently in the human lineage. Specifically, the Spoke contributed ChIP-seq datasets from human and chimpanzee *in vitro*-derived CNCCs using antibodies against specific histone modifications and transcriptional co-activators, ATAC-seq datasets to map chromatin accessibility, and RNA-seq analyses of transcriptomes. In addition, they used transgenic mice to characterize spatiotemporal activity of select candidate enhancers in the context of the developing embryo. They prioritized candidate enhancers within loci associated with human craniofacial disorders, and those that showed strong changes in regulatory activity between humans and chimps. When tested *in vivo*, the majority of enhancers with *in vitro* signatures of species-specific bias showed robustly reproducible differences in spatial reporter activity (Fig. S4).

Datasets generated by this Spoke are a rich resource for studying chromatin-level regulation of key craniofacial genes, understanding non-coding regulatory regions involved in human craniofacial development and disease, and characterizing enhancers that may drive phenotypic divergence of the human craniofacial complex. They complement epigenomic and transcriptomic studies generated by the Visel Spoke. Furthermore, the datasets aid interpretation of GWAS of normal-range and disease-associated facial variation. As proof-of-principle, the team conducted comparative epigenomic analysis of ∼100 different human cell types (representing distinct embryonic, adult and *in vitro* derived cell types), which revealed significant enrichment of active chromatin marks at GWAS-identified regions associated with facial shape in the *in vitro*-derived CNCCs (Claes et al., 2018 and unpublished data). Furthermore, they found that candidate regulatory regions in the vicinity of the craniofacial GWAS-led SNPs were significantly enriched for predicted CNCC enhancers (Claes et al., 2018 and unpublished data). These observations suggest a developmental origin of the facial variation captured in GWAS studies of adults and further validate this Spoke's FaceBase datasets as a resource for the functional follow-up analysis of the candidate non-coding variants.

Rapid identification and validation of human craniofacial development genes (https://doi.org/10.25550/1WWA)

The Spoke led by Richard Maas and Eric Liao applied next-generation sequencing technologies and high-throughput validation techniques to enable rapid identification of candidate genes responsible for craniofacial disorders. The dysmorphoses analyzed included a broad range of disorders, including cleft lip and palate, oblique facial clefts, hemifacial microsomia and less commonly seen anomalies for which the genetic basis is not yet fully elucidated.

The group contributed 25 datasets, each of which include data from the proband and family members; these are available with the permission of the FaceBase Data Access Committee. Whole-exome sequencing was typically performed, with whole-genome sequencing as a follow-up in more difficult cases. Once a candidate gene variant was identified, the group then attempted to phenocopy using a murine or zebrafish model. Zebrafish models were generated for 14 of the analyzed cases. At least eight new craniofacial disease-causing genes were identified, and for at least six other genes, the range of associated phenotypes was expanded.

Resources developed by other FaceBase projects provide animal model data, including information on gene expression and regulatory elements that complement these datasets, improving the functional annotation of identified and validated genes. Data produced by this Spoke will be useful for human geneticists, who can probe for candidate genes or phenotypes of interest; other researchers may find the validating animal models useful for elucidating the molecular mechanisms underlying the phenotypes.

Developing 3D craniofacial morphometry data and tools to transform dysmorphology (https://doi.org/10.25550/1WWC)

The goal of this project, led by Richard Spritz, Benedikt Hallgrimsson and Ophir Klein, was to develop a foundation for application of craniofacial 3D morphometrics in clinical practice, to enable dysmorphologists to replace clinical gestalt with specific quantitative measures and tools. Specifically, the aims were to: (1) build a 3D morphometric scan 'library' of craniofacial dysmorphic syndromes across age groups and ethnicities; (2) characterize the aberrant facial shapes of specific human dysmorphic syndromes using 3D morphometrics for derivation of objective quantitative measures; and (3) develop a prototype diagnostic tool to accurately

Development (2020) 147, dev191213. doi:10.1242/dev.191213

distinguish among craniofacial dysmorphic syndromes. The team collected and analyzed 3D images from 3327 individuals with 396 different syndromes (see sample size distribution in Fig. 3A), as well as 727 of their unaffected clinically unaffected relatives and 3003 unaffected, unrelated individuals. The age distributions of the syndromic subjects and their relatives are shown in Fig. 3B. The team developed and tested various parametric and machine-learning approaches to automated syndrome diagnosis (Bannister et al., 2017, 2020), and applied these methods to compare their utility to automated diagnosis of syndromes with craniofacial dysmorphology. The best results came from the machine-learning approach, achieving balanced accuracy of 78.1% and sensitivity of 56.9% for syndrome diagnosis (Hallgrimsson et al., 2020). These studies demonstrated that facial deep phenotyping by quantitative facial 3D imaging has strong potential to be useful in clinical diagnosis.

Through application to the FaceBase Data Access Committee, users can access 3D facial images from more than 5300 individuals with over 500 different syndromes with facial dysmorphism, as well as over 800 of their unaffected relatives. These data will be useful for any researcher interested in a broad and deep library of 3D images of individuals with craniofacial syndromes. In addition to performing comprehensive analyses using the entire dataset, researchers will be able to perform in-depth analyses of individual conditions or to quantitatively compare craniofacial findings in a focused group of specific syndromes.

Ontology of craniofacial development and malformation (OCDM) (https://doi.org/10.25550/1WX2)

The goal of this Spoke Project, led by James F. Brinkley, was to create an ontology for use by FaceBase and other craniofacial communities (Brinkley et al., 2013). The ontology consists of (1) a set of standardized terms for data annotation and retrieval by keyword search, and (2) a set of relations among these terms for representation of knowledge and for 'intelligent' queries that follow these relations to integrate data annotated with different but related terms. Such well-defined terms and relations are essential for integrating highly diverse and distributed data, not only within FaceBase, but also in the larger craniofacial community.

The OCDM consists of a set of sub-ontologies organized by the three species representing most FaceBase data (human, mouse and zebrafish). Terms and relations in existing ontologies are used wherever possible, but the OCDM adds rich detail not present in these ontologies. Within each species are sub-ontologies describing normal adult and developmental anatomy, and sub-ontologies describing malformations. Sub-ontologies across species describe mappings between normal structures and between malformations, creating a large and detailed semantic network, a small proportion of which is shown in Fig. S5. Each component is available as a Web Ontology Language (OWL) file, where OWL is the standard representation for the semantic web. In addition, the team developed software tools for creating and maintaining the OCDM, and for making it available for queries by other applications. These include the OCDM browser, a web-based tool for exploring OCDM content. More details are available on the Structural Informatics OCDM project page (Structural Informatics Group, 2020).

The rich detail in the OCDM makes it a computable representation of developmental pathways and pathological variants that lead to craniofacial malformations. Such pathways are becoming increasingly difficult for even subject matter experts to comprehend, with the result that many computable signaling and

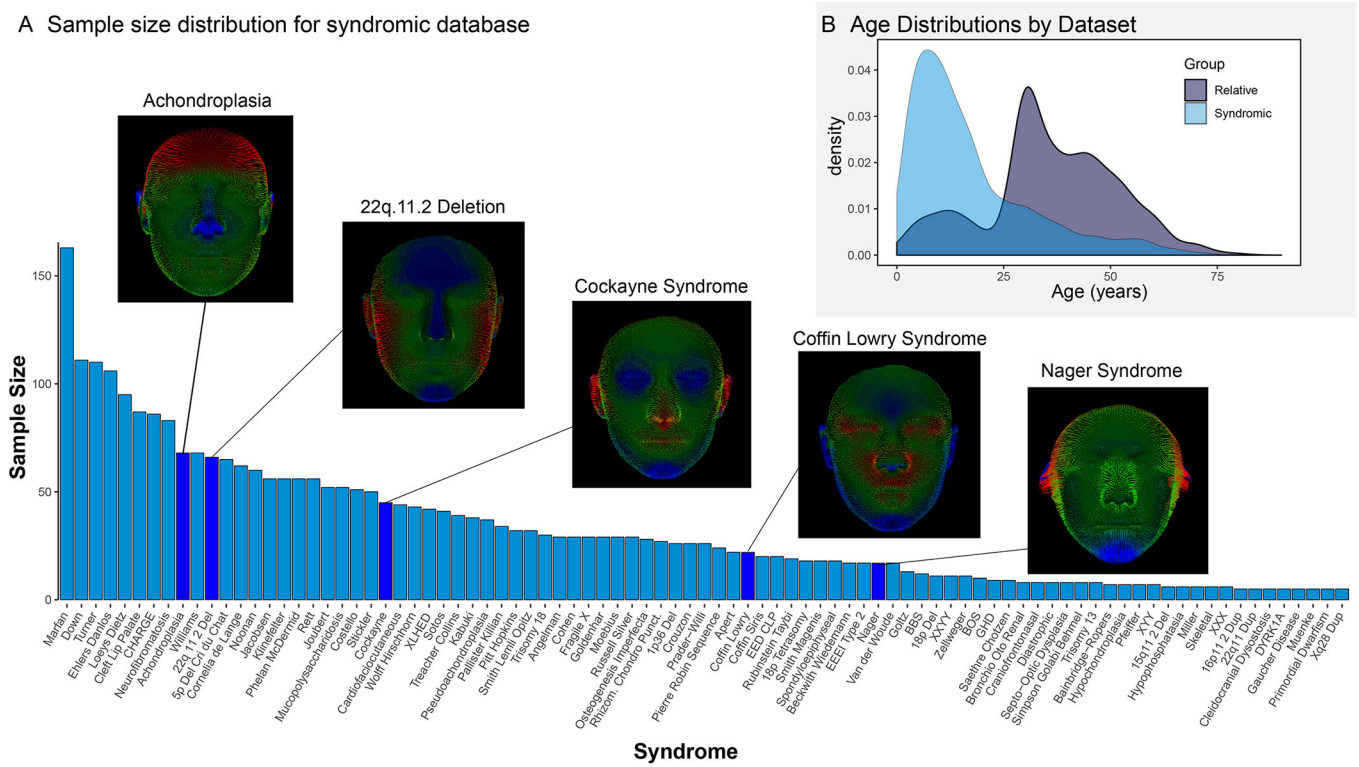

Fig. 3. 3D morphometric library of craniofacial dysmorphic syndromes. (A) Sample size distribution for the database of 3D facial images of subjects with genetic syndromes. The images show average facial shapes for select syndromes with a heatmap vector distribution overlay to highlight the regions of greatest difference. Blue indicates an area is smaller in syndromic individuals than in unaffected unrelated individuals; red indicates an area is larger. (B) Age distributions for syndromic subjects and their relatives in the database.

Development (2020) 147, dev191213. doi:10.1242/dev.191213

pathway databases, often represented as OWL ontologies, have been developed. When these efforts are complemented with highly detailed ontologies like the OCDM, the combined, queryable resources can greatly facilitate our understanding of craniofacial malformations and their relations to broader conditions.

Human genomics analysis interface (https://doi.org/10.25550/1WX0)

The Human Genomics Analysis Interface for FaceBase 2 (HGAI), directed by Mary Marazita, was developed to provide an easy way to visualize results from human craniofacial genomics projects without needing to access the individual data records. There are now several large human genomics databases relevant to craniofacial research, including multiple databases funded in part through FaceBase. HGAI identified nine appropriate craniofacial genetics projects, including studies of orofacial clefts, facial variation and dental disorders. The individual-level data for each project were analyzed in the aggregate and with respect to any appropriate subsets, creating 119 total results databases across the nine projects. These projects encompass a wide range of ages, ethnicities and phenotypes. More detailed information is available on the descriptive statistics tab for each project within the HGAI interface (http://facebase.org/hgai/). The results are available from the FaceBase Hub, visualized in the form of static Manhattan plots (Fig. S6A) and interactive LocusZoom plots (Fig. S6B).

HGAI is broadly useful for researchers who focus on animal models or who conduct human genetic studies. It offers an excellent opportunity to integrate genetic/genomic and expression data from animal models with human genomic data, enabling researchers who identify genes or pathways of interest in animal models to explore them further in human genomic datasets that we anticipate will continue to grow rapidly in the future.

The FaceBase Hub: a FAIR data resource for complex, diverse, evolving research data to advance craniofacial research

New discoveries in craniofacial development and dysmorphology are increasingly dependent on large, diverse and evolving collections of data generated through interdisciplinary research collaborations. The ability to share and locate datasets of interest, reuse them in an investigation and derive new results is crucial to these endeavors. Unfortunately, data are often poorly organized and annotated, such that it hinders reuse and reproducibility, in part because expert biocuration of data tends to be expensive and unsustainable. Furthermore, broad research communities have difficulty catalyzing a culture of data sharing in the absence of sufficient incentives. Last, research value that can be mined from data is limited due to the lack of access and interoperability necessary to analyze and visualize them. To address these issues, the FaceBase Consortium created an open, sustainable research community that transforms scholarly communication and facilitates a deeper understanding of craniofacial development. This is achieved through a number of key innovations in the areas of promoting data reuse, reproducibility, interoperability and interpretation through standards-based annotations and organization; creating a sustainable resource through automation and community-sourced data submissions;

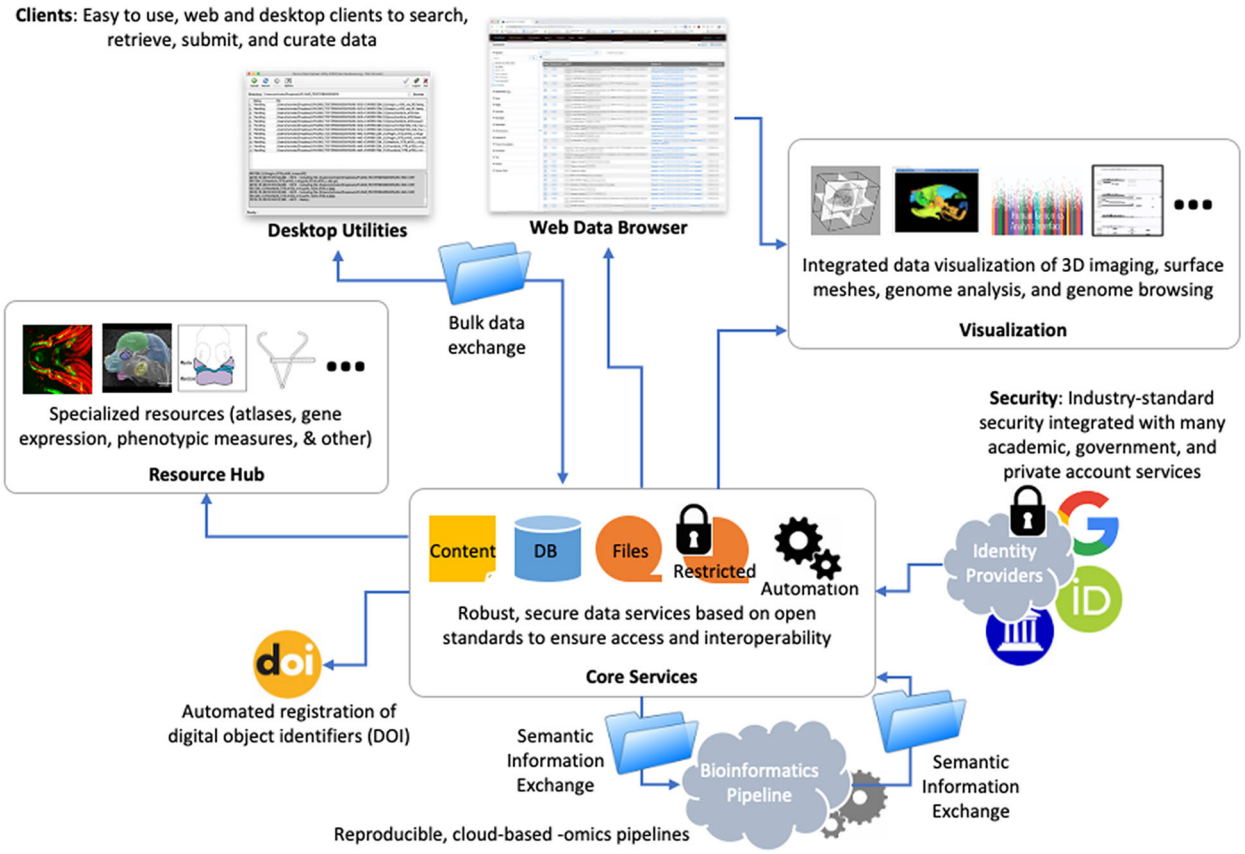

Fig. 4. Overview of the FaceBase platform and integrated services. The Hub's core data services drive the web-based data browser and search interface for accessing data (http://facebase.org/chaise/), visualization tools, analytical pipelines and the Resources Hub (https://www.facebase.org/resources/), which goes beyond the datasets accessible in the repository. Data submitters can use desktop utilities to upload data in bulk. Third-party identification providers including ORCID, Globus and Google are used to authenticate users. DOIs are minted for each dataset, facilitating accessibility and citation.

Development (2020) 147, dev191213. doi:10.1242/dev.191213

fostering data sharing through publication and citation facilitated by the generation of digital object identifiers (DOIs) for datasets; and facilitating research outcomes through integrated capabilities for data mining and visualization. In the rest of this section, we give an overview of the technology that powers the FaceBase platform, then highlight some of the key technical innovations that it enables.

Technical overview

The FaceBase platform (see Fig. 4) provides an integrated set of data, content, client, and visualization services and utilities, enabling secure sharing and collaboration over large data resources with community curated descriptive metadata. These core services ensure the accessibility and interoperability of data by providing a rich but flexible structure for describing and contextualizing raw and derived data from multiple research protocols involving different species, anatomical sites, phenotypes and more. Our online web service provides standard web (HTTP) data access (for unrestricted data) and allows for complex ad hoc queries over all metadata in FaceBase. The restricted (sensitive human subjects) data services are physically isolated from the rest of the core services and are organized in a two-layer structure behind increasingly restrictive firewalls that permit very limited access. Finally, the automation services perform scheduled and on-demand back-end processes such as the nightly registration of new datasets with a DOI provider. The core services are backed by an industry-grade, on-premises, secure storage service that provides a vast amount of capacity for future growth. The capabilities of the platform are extended with cloud-based services for running data analysis pipelines and for user account management (Chard et al., 2018), the latter of which integrates with many universities, government laboratories, ORCID and Google to allow users to login with their existing accounts.

Resource hub

FaceBase serves a dual role: in addition to being a data-sharing hub, it also facilitates the hosting of specialized information resources. The resource hub (http://facebase.org/resources/) describes community-contributed resources hosted by FaceBase as well as external resources. FaceBase acts as a registry for independently developed and operated resources, serving as a one-stop shop for comprehensive reference information for the craniofacial community.

Web and desktop clients

The FaceBase platform uses an adaptive data browser for common usage scenarios of searching, browsing, display and editing of data. The data browser is delivered as a rich web application that adapts to changes in the database schema, allowing the Hub to focus efforts on accurate and detailed modeling of data. We complement the web-based clients with native applications for Windows, MacOS and Linux that facilitate bulk transfer of data to and from the Hub's data services using a robust file archive format enriched with metadata and provenance (Chard et al., 2016). The clients are capable of checkpointing and restarting data transfers, a crucial feature when dealing with gigabytes of images or even terabytes of sequencing data.

Data mining and visualization

FaceBase integrates data mining and visualization capabilities so that users can explore data in-depth while on the site. Visualization tools include a web-based 3D image viewer with orthoslice and 3D modes, along with thumbnail images; a surface mesh (also known as model) viewer that displays models that may be composed of several distinct mesh objects and supports user-supplied landmarks with anatomical annotations (Fig. 2); an integrated genome track hub (a formatted layout suitable for rendering in a number of genome browser visualization services); interactive analysis of GWAS data through LocusZoom (Fig. S7); and custom plotting interfaces such as the dynamically generated matrix of mouse datasets that provides an at-a-glance overview of currently available data (Fig. 5).

Technical innovations

The FaceBase Hub team has developed multiple key technical innovations in order to support data representation, efficient data curation, emerging bioinformatics and scholarly communication as highlighted below.

Uniform data representation

Craniofacial research involves human and model organisms, complex anatomical structures, and diverse experimental methods

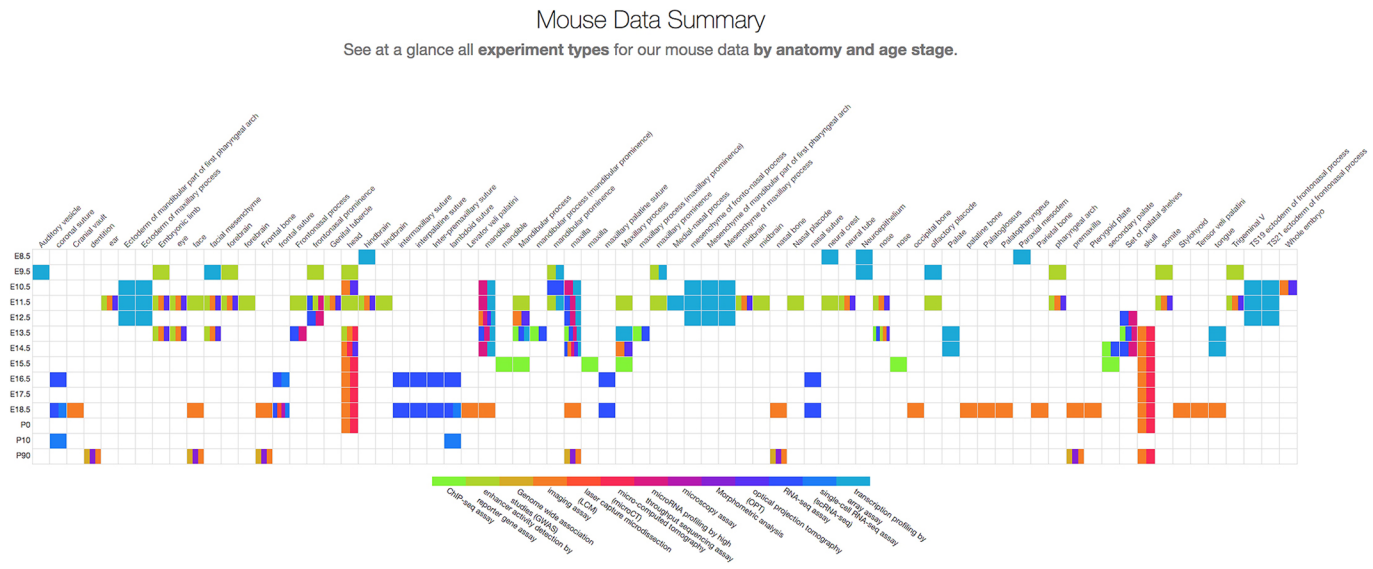

Fig. 5. Dynamically updated mouse data summary. Dynamically rendered matrix of available mouse datasets by age and anatomical source, color coded by experiment type.

Development (2020) 147, dev191213. doi:10.1242/dev.191213

and instruments. Researchers often need to explore data across species and experiment types. Doing so requires a data structure that can accommodate this heterogeneity. FaceBase has worked with consortium members to develop a structure that represents a level of detail necessary for users to find specific datasets of interest and reuse them for downstream analyses. Where possible, the allowable metadata terms are aligned with external, standard ontologies (e.g. OCDM, Uberon, Mammalian Anatomy, Mammalian Phenotype, Human Phenotype, and NCBI Taxonomy).

The key concepts represented in the uniform data structure (see Fig. S8) include: 'project', which represents a research project (e.g. a FaceBase Spoke Project, R01 investigation, etc.); 'person', a directory of persons (name, email, etc.) related to a project; and 'dataset', which represents a unit of data collected and submitted to FaceBase. Typically, a dataset represents a whole investigation or a self-contained study within a larger investigation or project. Within a dataset, an 'experiment' represents a particular assay in fine-grained detail. An experiment will generally be conducted on multiple biological replicates, which are represented as 'biosample' and 'replicate' entries.

The FaceBase model has been designed to facilitate machine interpretation and reuse of data without human language understanding. For example, crucial details, such as the relationship of sequencing data to associated bioreplicates are directly encoded in the structure. This property makes it possible to drive reproducible bioinformatics pipelines (described in a later section) and for third-party consumers of the data to know unambiguously how the data were generated or derived. At the same time, we focus on the minimal information necessary to support findability and reuse. Additional details may be entered at the discretion of the curator but are not required. This balance is crucial for sustainable data curation through community contributions.

Sustainable data curation

Generally, biocuration is both difficult and expensive to sustain. The volume of data is growing so fast that it is difficult for teams of curators to keep pace with demand. An innovative approach taken by FaceBase is to shift the responsibility for curation to the researchers themselves. To achieve this, we combined our simplified model, which strikes a balance between descriptive quality and curation effort, with online tools that are streamlined for bulk data entry, desktop clients for batch upload and automated linkage, and a lightweight process for review and curation. Future efforts include more automation in the initial 'triage' phase of data curation and quality control.

The Hub has tailored the data submission process to reduce manual effort. The process begins with registering a new project on FaceBase, organized around a single investigation on the scale of an R01 from the NIH. Such a project could conduct multiple experiments and submit them to FaceBase as individual or aggregated units (i.e. datasets). Project membership establishes user authorization to edit entries attributed to a given project. Visibility of data is limited during pre-release phases so that investigators can take advantage of FaceBase tools while the experiments are being conducted and coordinate the embargo of data while preparing publications. Online metadata entry permits batch editing to reduce redundant data entry. The data submission and curation processes are thoroughly documented on the FaceBase Data Curation Wiki (https://github.com/informatics-isi-edu/facebase-curation/wiki). When the dataset is curated and ready to release, the Hub conducts a review per FaceBase's data quality standards. Finally, after the approval of the submitter, the Hub releases the dataset for public viewing.

Reproducible bioinformatics

FaceBase's core services have been integrated with a cloud-based bioinformatics pipeline for processing sequence data. Most of the sequence data submitted during FaceBase 2 have been re-processed with a uniform pipeline to improve their comparability. FaceBase adopted the Big Data Bag (BDBag) format (Chard et al., 2016) to provide bulk data exchange that is semantically annotated with descriptive metadata and provenance. BDBag exchange facilitates robust reproducible data sharing and is leveraged in the bioinformatics pipelines, e.g. to bundle raw sequencing data and metadata for input to the pipeline and to capture derived data results from the output with metadata necessary to link the derived data back into the database and object store. To ensure broad data reusability, FaceBase adopted the uniform processing pipelines originally developed by the ENCODE DCC (ENCODE Project Consortium, 2012). Currently, the pipeline is implemented over a cloud-based service (DNAnexus) and can also be run by any researcher on their own hardware.

Transforming scholarly communication

To incentivize data-centric collaboration, we have prioritized building and deploying key capabilities to support the publication, citation and attribution of data. FaceBase recognizes that data sharing is a means to an end: the ultimate objective is to generate new knowledge. The traditional metrics of research outcomes are publications and their impact through citations. Through facilitating researchers to share FaceBase data, we can extend the research impact of our contributors by providing formal citation services (e.g. BibTeX format suitable for importing into citation managers such as EndNote, Mendeley or JabRef), cross-referencing FaceBase data with publications and other knowledge resources in the field, and socializing the craniofacial research community to the practice and importance of data citation, thus promoting data as a key contribution to science. The FaceBase platform has adopted best practices on research resource identifiers (Madduri et al., 2019), the BDBag semantic information exchange format with ability to describe data and its provenance (Chard et al., 2016), widely used vocabularies for clear description of data, and FAIR research principles (Wilkinson et al., 2016).

The future of FaceBase

FaceBase aims to transform scholarly communication in craniofacial research and drive new discovery through data-centric collaborations on a community-wide scale. The FaceBase 3 Steering Committee and Scientific Expert Panel regularly solicit feedback from the craniofacial community and convene to discuss data recruitment priorities and review the progress of our evolution as a knowledgebase. The data recruitment priorities are reviewed and approved annually by the Steering Committee, Scientific Expert Panel and NIDCR program staff. The current priorities are available on the Hub (https://www.facebase.org/submit/data-priorities/). Over the next year, in alignment with these priorities, we aim to bolster our strengths in data describing human, mouse and zebrafish development while targeting expansion to other significant model organisms, including chick and *Xenopus*, as well as characterization of commonly used cell lines. The FaceBase data model has been designed to encompass a broad range of anticipated future data and experiment types. For example, we expect that single-cell analyses (e.g. scRNA-seq) will be increasingly more widely used in the coming years and have the infrastructure to accommodate these datasets readily.

Human and murine dental and salivary gland development are also among our identified priority expansion areas for the coming

Development (2020) 147, dev191213. doi:10.1242/dev.191213

year. We look forward to making available a wide range of data on normally and abnormally developing teeth, including the characterization of the physical and chemical properties of mineralized tissues. We further expect that both transcriptome- and imaging-based data will be highly valuable to those researching dental development and disease. Improved understanding of the salivary gland can benefit individuals who suffer from xerostomia due to autoimmune conditions, such as Sjögren's syndrome, or as a side-effect of radiation therapy.

It is increasingly understood that dental and craniofacial health are closely linked to our general well-being, and that the oral cavity provides a unique diagnostic window for assessing overall health. Consequently, we see FaceBase as a crucial part of future atlases that will encompass the entire body, revealing the common mechanisms and signaling pathways that underlie diverse aspects of development and disease in various tissues. Towards this end, we aim to foster interdisciplinary collaborations, cross-link and integrate with complementary data repositories, and weave together efforts conducted with the support of different branches of the NIH.

The FaceBase community welcomes and greatly values new data contributors. More information can be found at the FaceBase Hub (facebase.org), where interest in contributing data can be indicated. Ultimately, the success of FaceBase depends on our scientific communities coming together to seize the opportunity to promote a culture of data sharing and collaboration. Collectively, we have the power to create an unrivaled resource that can in turn accelerate our research and transform care for the individuals who need it most.

Acknowledgements
The HGAI Spoke thanks the staff of the University of Pittsburgh Center for Craniofacial and Dental Genetics who contributed to the Marazita FaceBase 2 project, notably Nandita Mukhopadhyay (PhD), Myoung Keun Lee, Annette Werner, Tamra Mitchell, Lance Kennelty and James Gallagher. The Brinkley thanks Landon T. Detwiler, Jose L. V. Mejino, Michael L. Cunningham and Timothy C. Cox for their contributions.

Competing interests
The authors declare no competing or financial interests.

Author contributions
Conceptualization: C.K., Y.C.; Methodology: C.K., R.S.; Software: A.B., M.L.M., R.S., A.V., C.W.; Validation: A.V., T.J.W.; Formal analysis: R.A., J.F.B., E.F., S.F., A.S.G.-R., B.H., K.H., M.P.H., T.H., G.H., J.E.H., E.W.J., K.L.J., O.D.K., E.J.L., H. Li, E.C.L., H. Long, N.L., R.L.M., M.L.M., J.M., S.P., L.S., R.A.S., T.S., H.v.B., A.V., I.W., T.J.W., J.W., Y.Y.; Investigation: R.A., J.F.B., E.F., S.F., A.S.G.-R., B.H., K.H., M.P.H., T.H., G.H., J.E.H., E.W.J., K.L.J., O.D.K., E.J.L., H. Li, E.C.L., H. Long, N.L., R.L.M., M.L.M., J.M., S.P., L.S., R.A.S., T.S., H.v.B., A.V., I.W., T.J.W., J.W., Y.Y.; Data curation: B.D.S., A.B., J.G.H., B.H., T.H., M.L.M., R.S., T.S., A.V., C.W., Y.C.; Writing - original draft: B.D.S., R.A., J.F.B., E.F., S.F., A.S.G.-R., B.H., K.H., M.P.H., T.H., G.H., J.E.H., E.W.J., K.L.J., C.K., O.D.K., E.J.L., H. Li, E.C.L., H. Long, N.L., R.L.M., M.L.M., J.M., S.P., R.S., L.S., R.A.S., T.S., H.v.B, A.V., I.W., C.W., T.J.W., J.W., Y.Y., Y.C.; Writing - review & editing: B.D.S., J.F.B., S.F., J.G.H., B.H., M.P.H., T.H., G.H., J.E.H., E.W.J., K.L.J., C.K., O.D.K., E.C.L., R.L.M., M.L.M., R.S., L.S., R.A.S., H.v.B., A.V., C.W., T.J.W., J.W., Y.C.; Visualization: B.D.S., R.S., C.W.; Supervision: E.F., B.H., M.P.H., G.H., J.E.H., E.W.J., K.L.J., R.L.M., L.S., R.A.S., H.v.B., A.V., T.J.W., J.W., Y.C.; Project administration: B.D.S., S.F., C.K., C.W., Y.C.; Funding acquisition: J.F.B., S.F., B.H., M.P.H., G.H., J.E.H., E.W.J., K.L.J., C.K., O.D.K., E.C.L., R.L.M., M.L.M., L.S., R.A.S., H.v.B., A.V., T.J.W., J.W., Y.C.

Funding
The FaceBase 3 Hub is supported by the National Institute of Dental and Craniofacial Research (U01-DE028729 to C.K. and Y.C.). The FaceBase 2 Consortium projects were supported by the following awards from the National Institute of Dental and Craniofacial Research: U01-DE024449 (C.K.), U01-DE024434 (S.F. and M.P.H.), U01-DE024440 (R.S. and O.D.K.), U01-DE024430 (J.W. and L.S.), U01-DE024427 (A.V.), U01-DE024425 (M.L.M.), U01-DE024421 (Y.C.), U01-DE024417 (J.F.B.), U01-DE024443 (R.M.), U01-DE024429 (T.J.W., J.G.H. and K.L.J.) and U01-DE024448 (E.W.J., G.H. and H.v.B.). Research conducted at the E.O. Lawrence Berkeley National Laboratory was performed under Department of Energy Contract DE-AC02-05CH11231, University of California. Deposited in PMC for release after 12 months.

Data availability
The datasets referenced in this article are available through the FaceBase Hub (facebase.org)

Supplementary information
Supplementary information available online at
https://dev.biologists.org/lookup/doi/10.1242/dev.191213.supplemental

Peer review history
The peer review history is available online at
https://dev.biologists.org/lookup/doi/10.1242/dev.191213.reviewer-comments.pdf

References
Attanasio, C., Nord, A. S., Zhu, Y., Blow, M. J., Li, Z., Liberton, D. K., Morrison, H., Plajzer-Frick, I., Holt, A., Hosseini, R. et al. (2013). Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006. doi:10.1126/science.1241006

Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., Chang, C.-P., Zhao, Y., Swigut, T. and Wysocka, J. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature* **463**, 958-962. doi:10.1038/nature08733

Bannister, J. J., Larson, J. R., Hallgrimsson, B., Lim, P. H., Spritz, R. A., Klein, O. D., Bernier, F. P. J. and Forkert, N. D. (2017). Registration and landmarking of polygonal mesh facial scans using a point feature based iterative point correspondence algorithm. *Int. J. Comput. Assist. Radiol. Surg.* **12**, S216-S218.

Bannister, J. J., Crites, S. R., Aponte, J. D., Katz, D. C., Wilms, M., Klein, O. D., Bernier, F. P. J., Spritz, R. A., Hallgrímsson, B. and Forkert, N. D. (2020). Fully automatic landmarking of syndromic 3D facial surface scans using 2D images. *Sensors (Basel)* **20**, 3171. doi:10.3390/s20113171

Bebee, T. W., Park, J. W., Sheridan, K. I., Warzecha, C. C., Cieply, B. W., Rohacek, A. M., Xing, Y. and Carstens, R. P. (2015). The splicing regulators Esrp1 and Esrp2 direct an epithelial splicing program essential for mammalian development. *eLife* **4**, e08954. doi:10.7554/eLife.08954

Bowen, M. E., McClendon, J., Long, H. K., Sorayya, A., Van Nostrand, J. L., Wysocka, J. and Attardi, L. D. (2019). The spatiotemporal pattern and intensity of p53 activation dictates phenotypic diversity in p53-driven developmental syndromes. *Dev. Cell* **50**, 212-228.e216. doi:10.1016/j.devcel.2019.05.015

Brinkley, J. F., Borromeo, C., Clarkson, M., Cox, T. C., Cunningham, M. J., Detwiler, L. T., Heike, C. L., Hochheiser, H., Mejino, J. L. V., Travillian, R. S. et al. (2013). The ontology of craniofacial development and malformation for translational craniofacial research. *Am. J. Med. Genet. C Semin. Med. Genet.* **163C**, 232-245. doi:10.1002/ajmg.c.31377

Caetano-Lopes, J., Henke, K., Urso, K., Duryea, J., Charles, J. F., Warman, M. L. and Harris, M. P. (2020). Unique and non-redundant function of csf1r paralogues in regulation and evolution of post-embryonic development of the zebrafish. *Development* **147**, dev181834. doi:10.1242/dev.181834

Calo, E., Gu, B., Bowen, M. E., Aryan, F., Zalc, A., Liang, J., Flynn, R. A., Swigut, T., Chang, H. Y., Attardi, L. D. et al. (2018). Tissue-selective effects of nucleolar stress and rDNA damage in developmental disorders. *Nature* **554**, 112-117. doi:10.1038/nature25449

Carlson, J. C., Taub, M. A., Feingold, E., Beaty, T. H., Murray, J. C., Marazita, M. L. and Leslie, E. J. (2017). Identifying genetic sources of phenotypic heterogeneity in orofacial clefts by targeted sequencing. *Birth Defects Res.* **109**, 1030-1038. doi:10.1002/bdr2.23605

Chai, Y. and Maxson, R. E.Jr. (2006). Recent advances in craniofacial morphogenesis. *Dev. Dyn.* **235**, 2353-2375. doi:10.1002/dvdy.20833

Chai, Y., Jiang, X., Ito, Y., Bringas, P., Jr., Han, J., Rowitch, D. H., Soriano, P., McMahon, A. P. and Sucov, H. M. (2000). Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. *Development* **127**, 1671-1679.

Chard, K., D'Arcy, M., Heavner, B., Foster, I., Kesselman, C., Madduri, R., Rodriguez, A., Soliland-Reyes, S., Goble, C., Clark, K. et al. (2016). I'll take that to go: big data bags and minimal identifiers for exchange of large, complex datasets. In 2016 IEEE International Conference on Big Data, pp. 319-328. doi:10.1109/BigData.2016.7840618

Chard, K., Dart, E., Foster, I., Shifflett, D., Tuecke, S. and Williams, J. (2018). The Modern Research Data Portal: a design pattern for networked, data-invensive science. *Peerj Comput Sci.* doi:10.7717/peerj-cs.144

Charles, J. F., Sury, M., Tsang, K., Urso, K., Henke, K., Huang, Y., Russell, R., Duryea, J. and Harris, M. P. (2017). Utility of quantitative micro-computed tomographic analysis in zebrafish to define gene function during skeletogenesis. *Bone* **101**, 162-171. doi:10.1016/j.bone.2017.05.001

Cibi, D. M., Mia, M. M., Guna Shekeran, S., Yun, L. S., Sandireddy, R., Gupta, P., Hota, M., Sun, L., Ghosh, S. and Singh, M. K. (2019). Neural crest-specific deletion of *Rbfox2* in mice leads to craniofacial abnormalities including cleft palate. *eLife* **8**, e45418. doi:10.7554/eLife.45418

Claes, P., Roosenboom, J., White, J. D., Swigut, T., Sero, D., Li, J., Lee, M. K., Zaidi, A., Mattern, B. C., Liebowitz, C. et al. (2018). Genome-wide mapping of global-to-local genetic effects on human facial shape. *Nat. Genet.* **50**, 414-423. doi:10.1038/s41588-018-0057-4

Davesne, D., Meunier, F. J., Schmitt, A. D., Friedman, M., Otero, O. and Benson, R. B. J. (2019). The phylogenetic origin and evolution of acellular bone in teleost fishes: insights into osteocyte function in bone metabolism. *Biol. Rev. Camb. Philos. Soc.* **94**, 1338-1363. doi:10.1111/brv.12505

Dixon, M. J., Marazita, M. L., Beaty, T. H. and Murray, J. C. (2011). Cleft lip and palate: understanding genetic and environmental influences. *Nat. Rev. Genet.* **12**, 167-178. doi:10.1038/nrg2933

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74. doi:10.1038/nature11247

Feng, W., Leach, S. M., Tipney, H., Phang, T., Geraci, M., Spritz, R. A., Hunter, L. E. and Williams, T. (2009). Spatial and temporal analysis of gene expression during growth and fusion of the mouse facial prominences. *PLoS ONE* **4**, e8066. doi:10.1371/journal.pone.0008066

Fraser, G. J., Hulsey, C. D., Bloomquist, R. F., Uyesugi, K., Manley, N. R. and Streelman, J. T. (2009). An ancient gene network is co-opted for teeth on old and new jaws. *PLoS Biol.* **7**, e31. doi:10.1371/journal.pbio.1000031

Gistelinck, C., Kwon, R. Y., Malfait, F., Symoens, S., Harris, M. P., Henke, K., Hawkins, M. B., Fisher, S., Sips, P., Guillemyn, B. et al. (2018). Zebrafish type I collagen mutants faithfully recapitulate human type I collagenopathies. *Proc. Natl. Acad. Sci. USA* **115**, E8037-E8046. doi:10.1073/pnas.1722200115

Hallgrimsson, B., Aponte, D. J., Katz, D. C., Bannister, J. J., Riccardi, S. L., Mahasuwan, N., McInnes, B. L., Ferrara, T. M., Lipman, D., Spitzmacher, J. et al. (2020). Automated syndrome diagnosis by three-dimensinoal facial photogrammetric imaging. *Genet. Med.* (In press). doi:10.1038/s41436-020-0845-y

Harris, M. P., Henke, K., Hawkins, M. B. and Witten, P. E. (2014). Fish is Fish: the use of experimental model species to reveal causes of skeletal diversity in evolution and disease. *J. Appl. Ichthyol.* **30**, 616-629. doi:10.1111/jai.12533

Henke, K., Daane, J. M., Hawkins, M. B., Dooley, C. M., Busch-Nentwich, E. M., Stemple, D. L. and Harris, M. P. (2017). Genetic screen for postembryonic development in the Zebrafish (Danio rerio): dominant mutations affecting adult form. *Genetics* **207**, 609-623. doi:10.1534/genetics.117.300187

Hennekam, R. C. M., Krantz, I. D. and Allanson, J. E. (2010). *Gorlin's Syndromes of the Head and Neck*, 5th edn. New York, USA: Oxford University Press.

Heuzé, Y., Holmes, G., Peter, I., Richtsmeier, J. T. and Jabs, E. W. (2014). Closing the gap: genetic and genomic continuum from syndromic to nonsyndromic craniosynostoses. *Curr. Genet. Med. Rep.* **2**, 135-145. doi:10.1007/s40142-014-0042-x

Ho, T.-V., Iwata, J., Ho, H. A., Grimes, W. C., Park, S., Sanchez-Lara, P. A. and Chai, Y. (2015). Integration of comprehensive 3D microCT and signaling analysis reveals differential regulatory mechanisms of craniofacial bone development. *Dev. Biol.* **400**, 180-190. doi:10.1016/j.ydbio.2015.02.010

Hochheiser, H., Aronow, B. J., Artinger, K., Beaty, T. H., Brinkley, J. F., Chai, Y., Clouthier, D., Cunningham, M. L., Dixon, M., Donahue, L. R. et al. (2011). The FaceBase Consortium: a comprehensive program to facilitate craniofacial research. *Dev. Biol.* **355**, 175-182. doi:10.1016/j.ydbio.2011.02.033

Holmes, G., Gonzalez-Reiche, A. S., Lu, N., van Bakel, H. and Jabs, E. W. (2020a). Skeletal stem cells in craniofacial bone. In *Encyclopedia of Bone Biology* (ed. M. Zaidi), pp. 111-149. Oxford: Academic Press.

Holmes, G., Gonzalez-Reiche, A. S., Lu, N., Zhou, X., Rivera, J., Kriti, D., Sebra, R., Williams, A. A., Donovan, M. J., Potter, S. S. et al. (2020b). Integrated transcriptome and network analysis reveals spatiotemporal dynamics of calvarial suturogenesis. *Cell Rep.* **32**, 107871. doi:10.1016/j.celrep.2020.107871

Hooper, J. E., Jones, K. L., Smith, F. J., Williams, T. and Li, H. (2020). An alternative splicing program for mouse craniofacial development. *Front. Physiol.* **11**, 1099. doi:10.3389/fphys.2020.01099

Hooper, J. E., Feng, W., Li, H., Leach, S. M., Phang, T., Siska, C., Jones, K. L., Spritz, R. A., Hunter, L. E. and Williams, T. (2017). Systems biology of facial development: contributions of ectoderm and mesenchyme. *Dev. Biol.* **426**, 97-114. doi:10.1016/j.ydbio.2017.03.025

Hulsey, C. D., Fraser, G. J. and Meyer, A. (2016). Biting into the genome to phenome map: developmental genetic modularity of cichlid fish dentitions. *Integr. Comp. Biol.* **56**, 373-388. doi:10.1093/icb/icw059

Iwata, J.-I., Hacia, J. G., Suzuki, A., Sanchez-Lara, P. A., Urata, M. and Chai, Y. (2012). Modulation of noncanonical TGF-β signaling prevents cleft palate in Tgfbr2 mutant mice. *J. Clin. Invest.* **122**, 873-885. doi:10.1172/JCI61498

Jin, Y.-R., Han, X. H., Taketo, M. M. and Yoon, J. K. (2012). Wnt9b-dependent FGF signaling is crucial for outgrowth of the nasal and maxillary processes during upper jaw and lip development. *Development* **139**, 1821-1830. doi:10.1242/dev.075796

Kague, E., Roy, P., Asselin, G., Hu, G., Simonet, J., Stanley, A., Albertson, C. and Fisher, S. (2016). Osterix/Sp7 limits cranial bone initiation sites and is required for formation of sutures. *Dev. Biol.* **413**, 160-172. doi:10.1016/j.ydbio.2016.03.011

Kanther, M., Scalici, A., Rashid, A., Miao, K., Van Deventer, E. and Fisher, S. (2019). Initiation and early growth of the skull vault in zebrafish. *Mech. Dev.* **160**, 103578. doi:10.1016/j.mod.2019.103578

Leach, S. M., Feng, W. and Williams, T. (2017). Gene expression profile data for mouse facial development. *Data Brief* **13**, 242-247. doi:10.1016/j.dib.2017.05.003

Lee, S. K., Sears, M. J., Zhang, Z., Li, H., Salhab, I., Krebs, P., Xing, Y., Nah, H.-D., Williams, T. and Carstens, R. P. (2020). Cleft lip and cleft palate in *Esrp1* knockout mice is associated with alterations in epithelial-mesenchymal crosstalk. *Development* **147**, dev187369. doi:10.1242/dev.187369

Li, H. and Williams, T. (2013). Separation of mouse embryonic facial ectoderm and mesenchyme. *J Vis Exp* **74**, 50248. doi:10.3791/50248

Li, H., Jones, K. L., Hooper, J. E. and Williams, T. (2019). The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. *Development* **146**, dev174888. doi:10.1242/dev.174888

Madduri, R., Chard, K., D'Arcy, M., Jung, S. C., Rodriguez, A., Sulakhe, D., Deutsch, E., Funk, C., Heavner, B., Richards, M. et al. (2019). Reproducible big data science: a case study in continuous FAIRness. *PLoS One* **14**, e0213013. doi:10.1371/journal.pone.0213013

Ofer, L., Dean, M. N., Zaslansky, P., Kult, S., Shwartz, Y., Zaretsky, J., Griess-Fishheimer, S., Monsonego-Ornan, E., Zelzer, E. and Shahar, R. (2019a). A novel nonosteocytic regulatory mechanism of bone modeling. *PLoS Biol.* **17**, e3000140. doi:10.1371/journal.pbio.3000140

Ofer, L., Dumont, M., Rack, A., Zaslansky, P. and Shahar, R. (2019b). New insights into the process of osteogenesis of anosteocytic bone. *Bone* **125**, 61-73. doi:10.1016/j.bone.2019.05.013

Oka, K., Oka, S., Sasaki, T., Ito, Y., Bringas, P., Jr., Nonaka, K. and Chai, Y. (2007). The role of TGF-β signaling in regulating chondrogenesis and osteogenesis during mandibular development. *Dev. Biol.* **303**, 391-404. doi:10.1016/j.ydbio.2006.11.025

Parichy, D. M., Elizondo, M. R., Mills, M. G., Gordon, T. N. and Engeszer, R. E. (2009). Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev. Dyn.* **238**, 2975-3015. doi:10.1002/dvdy.22113

Pelikan, R. C., Iwata, J., Suzuki, A., Chai, Y. and Hacia, J. G. (2013). Identification of candidate downstream targets of TGFβ signaling during palate development by genome-wide transcript profiling. *J. Cell. Biochem.* **114**, 796-807. doi:10.1002/jcb.24417

Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F. H., Swigut, T. and Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83. doi:10.1016/j.cell.2015.08.036

Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S. A., Swigut, T. and Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633-648. doi:10.1016/j.stem.2012.07.006

Richtsmeier, J. T. and Flaherty, K. (2013). Hand in glove: brain and skull in development and dysmorphogenesis. *Acta Neuropathol.* **125**, 469-489. doi:10.1007/s00401-013-1104-y

Shaffer, J. R., LeClair, J., Carlson, J. C., Feingold, E., Buxó, C. J., Christensen, K., Deleyiannis, F. W. B., Field, L. L., Hecht, J. T., Moreno, L. et al. (2019). Association of low-frequency genetic variants in regulatory regions with nonsyndromic orofacial clefts. *Am. J. Med. Genet. A* **179**, 467-474. doi:10.1002/ajmg.a.61002

Structural Informatics Group (2020). Ontology of Craniofacial Development and Malformation (OCDM). http://si.washington.edu/projects/ocdm.

Sugii, H., Grimaldi, A., Li, J., Parada, C., Vu-Ho, T., Feng, J., Jing, J., Yuan, Y., Guo, Y., Maeda, H. et al. (2017). The Dlx5-FGF10 signaling cascade controls cranial neural crest and myoblast interaction during oropharyngeal patterning and development. *Development* **144**, 4037-4045. doi:10.1242/dev.155176

Tucker, A. S. and Fraser, G. J. (2014). Evolution and developmental diversity of tooth regeneration. *Semin. Cell Dev. Biol.* **25-26**, 71-80. doi:10.1016/j.semcdb.2013.12.013

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C. and Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nat. Genet.* **46**, 753-758. doi:10.1038/ng.2971

Warzecha, C. C., Sato, T. K., Nabet, B., Hogenesch, J. B. and Carstens, R. P. (2009). ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Mol. Cell* **33**, 591-601. doi:10.1016/j.molcel.2009.01.025

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018. doi:10.1038/sdata.2016.18

Witten, P. E., Harris, M. P., Huysseune, A. and Winkler, C. (2017). Small teleost fish provide new insights into human skeletal diseases. *Methods Cell Biol.* **138**, 321-346. doi:10.1016/bs.mcb.2016.09.001

Zhao, H. and Chai, Y. (2015). Stem cells in teeth and craniofacial bones. *J. Dent. Res.* **94**, 1495-1501. doi:10.1177/0022034515603972