FaceBase 3: Analytical Tools and FAIR Resources for Craniofacial and Dental Research

Bridget D. Samuels, Robert Aho, James F. Brinkley, Alejandro Bugacov, Eleanor Feingold, Shannon Fisher, Ana S. Gonzalez-Reiche, Joseph G. Hacia, Benedikt Hallgrimsson, Karissa Hansen, Matthew P. Harris, Thach-Vu Ho, Greg Holmes, Joan E. Hooper, Ethylin Wang Jabs, Kenneth L. Jones, Carl Kesselman, Ophir D. Klein, Elizabeth J. Leslie, Hong Li, Eric C. Liao, Hannah Long, Na Lu, Richard L. Maas, Mary L. Marazita, Jaaved Mohammed, Sara Prescott, Robert Schuler, Licia Selleri, Richard A. Spritz, Tomek Swigut, Harm van Bakel, Axel Visel, Ian Welsh, Cristina Williams, Trevor J. Williams, Joanna Wysocka, Yuan Yuan and Yang Chai
DOI: 10.1242/dev.191213

Editor: James Briscoe

Review timeline
Original submission: 2 April 2020
Editorial decision: 12 May 2020
First revision received: 17 July 2020
Accepted: 13 August 2020

Original submission

First decision letter

MS ID#: DEVELOP/2020/191213

MS TITLE: FaceBase 3: A FAIR Resource for Craniofacial and Dental Research

AUTHORS: Bridget D Samuels, Robert Aho, James F Brinkley, Alejandro Bugacov, Eleanor Feingold, Shannon Fisher, Ana S Gonzalez-Reiche, Joseph G Hacia, Benedikt Hallgrimsson, Karissa Hansen, Matthew P Harris, Harry Hochheiser, Thach-Vu Ho, Greg Holmes, Joan E Hooper, Ethylin Wang Jabs, Kenneth L Jones, Carl Kesselman, Ophir D Klein, Elizabeth J Leslie, Hong Li, Eric C Liao, Hannah Long, Na Lu, Richard L Maas, Mary L Marazita, Jaaved Mohammed, Sara Prescott, Robert Schuler, Licia Selleri, Richard A Spritz, Tomek Swigut, Harm van Bakel, Axel Visel, Ian Welsh, Cristina Williams, Trevor J Williams, Joanna Wysocka, Yuan Yuan, and Yang Chai

I have now received all the referees' reports on the above manuscript, and have reached a decision. I apologise that this has taken longer than usual, but COVID19 is affecting how quickly some of our referees can respond. The referees' comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

As you will see, the referees express considerable interest in your work and recognise the value of the resources for the community. However they have some significant criticisms and suggestions. Addressing these will require revisions of your manuscript before we can consider publication, but I believe will improve the paper. The referees have a series of questions and suggestions that need clarification and help readers navigate the resources you describe. If you are able to revise the manuscript along the lines suggested, which may involve further experiments, I will be happy receive a revised version of the manuscript. Your revised paper will be re-reviewed by one or more of the original referees, and acceptance of your manuscript will depend on your addressing satisfactorily the reviewers' major concerns. Please also note that Development will normally permit only one round of major revision.

We are aware that you may currently be unable to access the lab to undertake experimental revisions. If it would be helpful, we encourage you to contact us to discuss your revision in greater detail. Please send us a point-by-point response indicating where you are able to address concerns raised (either experimentally or by changes to the text) and where you will not be able to do so within the normal timeframe of a revision. We will then provide further guidance. Please also note that we are happy to extend revision timeframes as necessary.

Please attend to all of the reviewers' comments and ensure that you clearly highlight all changes made in the revised manuscript. Please avoid using 'Tracked changes' in Word files as these are lost in PDF conversion. I should be grateful if you would also provide a point-by-point response detailing how you have dealt with the points raised by the reviewers in the 'Response to Reviewers' box. If you do not agree with any of their criticisms or suggestions please explain clearly why this is so.

Reviewer 1

Advance summary and potential significance to field

This Techniques and Resources Article presents an overview of datasets generated from FaceBase 2 projects over the last five years, describes data analysis and visualization tools available through the FaceBase Hub website and touches on future plans to expand the Hub. The information presented here, together with the FaceBase Hub itself, will be of major value to members of the developmental biology community interested in both normal and disordered craniofacial development. The article does a commendable job of explaining how this resource conforms to the FAIR data principles, thereby ensuring that the datasets are accessible reproducible and reusable.

Comments for the author

I would encourage minor revision of the Article. Additional details should be added to the text for the purpose of clarity and to better showcase the available data.

Major Comments:

1. The FaceBase 2 Spoke projects discussed should be presented in a more structured order, covering for example, projects relating to humans, model organisms (mouse then zebrafish) and finally cell lines.
Further, for clarity, it will be important to state early in the description of each project the species structure(s) and timepoints(s) analyzed.
2. When mutant samples are analyzed, for example, the mutant mouse models in project "Integrated research of functional genomics and craniofacial morphogenesis" and the mutant zebrafish in project
"Anatomical atlas and transgenic tools for late skull formation in the zebrafish", the number of models, if not the names of the models, should be stated as well as whether or not the genes involved function in specific pathways or cellular processes.
3. For the project "Ontology of Craniofacial Development and Malformation (OCDM)", the authors should touch on whether or not the terms used match the phenotype terms (and trees) available in other commonly-used databases such as Mouse Genome Informatics.
4. For the project "RNA dynamics in the developing mouse face", when discussing how this dataset could be mined, point ii focuses on upstream regulation of these features, when the effects of these events may be just as or more interesting and should be addressed. These might include differential isoform usage with unique functions across tissues and/or timepoints, differential isoform expression and transcript stability.
5. Two of the spoke projects, "Genomic and transgenic resources for craniofacial enhancer studies" and
"Rapid identification and validation of human craniofacial development genes" do not have an associated figure. For the former, OPT visualization of a transgenic reporter embryo would be helpful.
6. Within the "Genomic and transgenic resources for craniofacial enhancer studies" project a unified analytical workflow is discussed. The authors should clarify whether or not future data

submitters apply this workflow to their own data or whether the FaceBase Hub provides this service.
7. For the “Human genomics analysis interface” project, the authors should state the distribution of data across ages and ethnicities.
8. For the “Rapid identification and validation of human craniofacial development genes” project, the authors should discuss whether their analyses resulted in the discovery of novel genes/variants or genes/variants that were previously known to underlie craniofacial dysmorphologies. Further, the authors should state how often the animal models used recapitulated the human phenotype.

Minor Comments:

1. Figure 1 would benefit from addition of a separate panel to demonstrate “digital dissection” of the skull, as in Figure 2D for zebrafish.
2. Figure S2 A labels the “FNP”, which no longer exists at this timepoint as the development of the nasal pits has generated the MNP and LNP. The authors should specifically point out that the tissue taken included both of these structures. Further, in Figure S2 C, it may be more interesting to show a novel splicing event instead of the well-studied ectodermal vs. mesenchymal splicing of Fgfr2.
3. Within the project “RNA dynamics in the developing mouse face”, the authors should also reference the recent paper demonstrating a role for splicing factor Rbfox2 in craniofacial development: DOI 10.7554/eLife.45418.
4. The gene names in the Legend of Figure S3 should be italicized.
5. Given its repeated use in the description of the previous project, the definition of enhancers in the first line of project “Genomic and transgenic resources for craniofacial enhancer studies” should be included earlier in the Article.
6. Figures 2B and 2D should be enlarged.
7. The authors should state whether the reporter lines for chondrocytes, osteoblasts and osteoclasts described in “Anatomical atlas and transgenic tools for late skull formation in the zebrafish” were previously published or generated for this project. If the former, references should be provided.
8. The UMAP in the upper right of Figure S4 should have the different cell populations labeled as in Figure S2 D.
9. The legend for Figure 6 should indicate that data for additional anatomical sites are available but not shown in the figure.

Reviewer 2

Advance summary and potential significance to field

Samuels et al. present a manuscript outlining recent accomplishments, technical organization and features, as well as future goals of FaceBase - a consortium aimed at integrating ‘big data’ across basic, translational and clinical craniofacial research.
FaceBase not only serves as a repository for annotated and curated data sets relevant to craniofacial biology, but provides essential visualization and analysis pipelines for such data. Spoke projects described here address tissue- and species- specific expression patterns and transcriptional control of facial development, as well as multispecies morphological WT and mutant data sets that have generated new predictions for molecular mechanisms underpinning craniofacial development and dysmorphology. Further, phenotyping pipelines are providing novel diagnostic tools. While many of the achievements of FaceBase 2 have generated multidisciplinary biological data sets, another accomplishment is the generation of an Ontology for integrating the data available across the platform.
In developing infrastructure for robust data analysis and cross-experiment comparisons, FaceBase is now poised as a key platform for cross-discipline collaboration. FaceBase 3 aims to build responsibly upon the achievements of this report, to position itself as a cornerstone resource for craniofacial and developmental biology.

Comments for the author

This manuscript is appropriate for publication in Development as it presents an extremely rich resource for craniofacial, developmental and clinical biologists. As FaceBase 2, reported here, has used spoke projects to broaden the relevance of data available to a broader community than just craniofacial biology. However, the following comments must first be addressed.

1. Craniofacial development offers a platform to understand key principles in developmental, molecular and cell biology. The abstract and introduction do not convey the broad interest of FaceBase to the community, e.g. those interested in epithelia or/vs mesenchyme, generally interested in ossification, chondrogenesis signaling, stem cells etc. Adding some comment on this might extend the readership of this article.

2. The introduction fails to demonstrate the diversity of data sets or the tools generally available through FaceBase. Although this is explored later in the manuscript it might benefit the more broad readership of Development to emphasize this briefly earlier in the text.

3. In the introduction, it is written that FaceBase is expanding its scope to be broader but no example is used to demonstrate this point. While not essential adding to this would aid in managing expectations for the rest of the manuscript.

4. Is it 10-11 spokes funded per year or per version of Facebase?

5. The motivation for OCDM is not well emphasized. Building an ontology for this network is essential for data across FaceBase to be integrated. Further, this ontology adds to the utility of Facebase to the braoder community. Commenting on this idea would improve impact of this Spoke.

6. It is suggested that OCDM offers an understanding of how signaling is altered in craniofacial malformation. It would improve the broader impact of the paper to also indicate that this might be useful for understanding signaling complexity in general. Further, the tissue-specificity of signal function is also an important feature. While this is not essential, it is important to demonstrate that this resource is useful for researchers outside of the craniofacial community.

7. The strength of ‘Integrated research of functional genomics and craniofacial morphogenesis’ is that many different kinds of data sets are integrated to address how signaling is used to mediate morphogenesis in complex craniofacial structures. The first two paragraphs do not demonstrate this point, in part because the connection between the first and second paragraph of this section is not very clear. It would benefit understanding of the manuscript if this were further clarified.

8. This reviewer understands the final comment in the final paragraph of this section (see 7) to indicate how reproducible and integrated different datasets are in FaceBase. Is this correct? If so, this should be stated explicitly to make very clear the strengths of this spoke and may comprise a stand-alone point.

9. Further to comment 8, it is simply stated that ‘FaceBase provides an ideal platform for collaborations’ but why this is placed here is not entirely clear. If this point is to be made, it would be strengthened by linking to future directives discussed later; if relevant, how FaceBase helps mediate collaboration. Additionally adding specific motivation for focusing on mandibular data for this spoke would not only improve the narrative of this section but could be used to demonstrate how these data sets bring together different disciplines. It should, however, be made clear that the microCT data includes also the entire skull which will be useful for other studies.

10. Fig.1 Legend. It might be necessary to make clear that the different colors indicate distinct bones.

11. Fig. 1 What do the blue dots represent in this image? This should be included in the reference or they should be removed.

12. While it might be assumed in the section titled ‘Genomic and transgenic resources for craniofacial enhancer studies’ that OPT data can be visualized using FaceBase tools, however, this is not explicitly stated. It might benefit the reader to be reminded of such functionality.

13. While it is stated that the Zebrafish spoke paves the way for similar work in other fish, the reasons why this might be of interest to the broad community is not clearly articulated. Such impact could include understanding how bone adapts to distinct stress for example provided by different feeding schemes. Although not essential, the import of this spoke might be emphasized by including a brief discussion of such implications.

14. Fig. 2. Requires scale bars to be added to images where possible (in particular a, c and E) to emphasize the scales represented by the rich Zebrafish data set.

15. Fig. 2E. The figure might make more impact if the mouse comparison was included. In this way, the figure could also demonstrate how this data set is integrating into corresponding mammalian resources.

16. The following statement should be updated with relevant references: Sutures differ widely in their physical structure, cell lineage, mechanical environment, and susceptibility to craniosynostosis.

17. It is important to remind the reader that the models used in the multi-suture bulk transcriptomics are craniosynostotic and therefore models not only for craniofacial dysgenesis but premature suture ossification Further, though not essential the reader may understand the importance of this spoke if it is mentioned that sutures also harbor a skeletal stem cell population that is poorly understood.

18. Fig. S4. It would make it simpler for the reader to understand this image if the arrows from the diagram to the data set were labelled with seq type e.g. scSeq and BulkSeq, respectively. While a key for these cell clusters in the UMAP plot is not absolutely essential it would improve the figure by highlighting cell types of interest to the broader community.

19. Fig3B is not discussed in the text, is this included to demonstrate one parameter that can be plotted or is it intended to add information? The relevance/importance of this panel should be indicated.

20. Fig. 4. Some descriptions of the figure might help understanding, e.g. Does DB mean database? Do the coin purses represent exchange or that these exchanges require additional financial support from researchers? Do the arrows indicate physical information flow or work flow perhaps? In either case, why is the arrow to web data browser from core services not bidirectional?

21. Fig. 4 Can the expression data be viewed in the visualization tools - in some cases 3D? If so, this is an advantage could be further emphasized either in the text or added to the figure.

22. Fig. 4. Does the resource hub refer not only to the original hub but also the spokes? Indicating this may make the scheme more clear if comment 8 is addressed.

23. It is not quite clear what fig. 5 adds beyond what is noted in fig. 1. If kept, it would be interesting for the reader to know which part of the skull has been selected for viewing in fig. 5/what mesh type was used.

24. Why are teeth and salivary glands priority areas, vs other regions? While it is modestly touched upon in the following paragraph, this should be commented on in the text earlier.

25. Fig. 6. The anatomy labels are clipped on the right of the image.

26. Fig. S5B. What do the different color/shapes mean in this plot?

27. One question that could be addressed briefly, is how instructions for features such as automated analyses organized in FaceBase. This could be included or simply referenced.

28. The abstract fails to convey the breadth of craniofacial development or data types that FaceBase2 utilizes. This must be rectified as the general readership should be incited to know more about the service the FaceBase currently supplies.

29. Further to comment 28, the title also does not convey the technological and resource advancements of Facebase 2.

Reviewer 3

Advance summary and potential significance to field

This manuscript by the FaceBase Consortium summarises the datasets generated by the Phase 2 FaceBase projects from 2014-2019. While Phase 1 projects previously described focused on the midface development, FaceBase 2 covers craniofacial development more broadly. One of the most considerable assets and achievements of the Consortium is the creation of the FaceBase Data Hub which includes more than 850 datasets in free access to the scientific community, featuring new web browser-based tools and enabling users-led custom analyses.

Comments for the author

Facebase 2 Spoke Projects briefly introduced and presented are diverse, highly innovative and at the cutting-edge, covering different molecular, cellular, epigenetic, gene regulatory and disease-related aspects of craniofacial biology, including the development of data interactive tools, 3D craniofacial morphometry tools and establishment of ontology for use by the craniofacial filed.

The Resource Hub presented by the Consortium enables access to the data produced and offers a comprehensive reference for the craniofacial community that can be browsed using either a rich web application or desktop client to mine and visualize the data.

The manuscript presents a clear overview of the second phase of this highly successful Consortium, not only clearly introducing the projects involved, but also highlighting the usability of the databases and interactive tools created, of high interest to the developmental biologists.

First revision

Author response to reviewers' comments

We are delighted to submit our revised manuscript with ID# DEVELOP/2020/191213 now titled “FaceBase3: Analytical Tools and FAIR Resources for Craniofacial and Dental Research” to *Development*.

We greatly appreciate your feedback as well as the comments we received from the reviewers and their interest in our work. Our point-by-point responses to the reviewers’ comments appear below in blue. We are confident that all concerns have been adequately addressed. We have also made slight changes throughout the manuscript to reduce the word count to 7,699.

As a result of these revisions, our manuscript is much improved. Per the journal guidelines, we have submitted the supplementary information as a single, clean PDF file. The changes we have made to the supplementary figures and legends are described in our point-by-point responses. If it would be helpful to have a copy of the supplementary information with the changes highlighted as we have done for the manuscript, we would be happy to provide it.

Additionally, Dr. Harry Hochheiser has requested to be removed as an author because he felt his contribution did not rise to the level of authorship. We have therefore honored this request. Dr.

Hochheiser has confirmed this in writing to Editorial Administrator Laeticia Beck.

Reviewer 1 Advance Summary and Potential Significance to Field:
This Techniques and Resources Article presents an overview of datasets generated from FaceBase 2 projects over the last five years, describes data analysis and visualization tools available through the FaceBase Hub website and touches on future plans to expand the Hub. The information presented here, together with the FaceBase Hub itself, will be of major value to members of the developmental biology community interested in both normal and disordered craniofacial development. The article does a commendable job of explaining how this resource conforms to the FAIR data principles, thereby ensuring that the datasets are accessible, reproducible and reusable.

Reviewer 1 Comments for the Author:
I would encourage minor revision of the Article. Additional details should be added to the text for the purpose of clarity and to better showcase the available data.

We thank the reviewer for the positive comments and helpful feedback.

Major Comments:

1. The FaceBase 2 Spoke projects discussed should be presented in a more structured order, covering, for example, projects relating to humans, model organisms (mouse then zebrafish) and finally cell lines. Further, for clarity, it will be important to state early in the description of each project the species, structure(s) and timepoints(s) analyzed.

We appreciate this suggestion and have re-organized the descriptions of the Spokes accordingly (first those using animal models grouped by model, then those focusing on humans, and finally those that developed new computational/analytical tools), and we have verified that each one includes these important details.

2. When mutant samples are analyzed, for example, the mutant mouse models in project "Integrated research of functional genomics and craniofacial morphogenesis" and the mutant zebrafish in project "Anatomical atlas and transgenic tools for late skull formation in the zebrafish", the number of models, if not the names of the models, should be stated as well as whether or not the genes involved function in specific pathways or cellular processes.

The FaceBase Hub provides a continually growing list of the models that are covered; in the descriptions of each Spoke, we provide an overview and have added mention of disorders modeled by some of the mutants analyzed. We feel that the importance of the description here is to show the type and availability of the datasets, rather than a comprehensive listing, which is not possible given the dynamic nature of the data repository.

3. For the project "Ontology of Craniofacial Development and Malformation (OCDM)", the authors should touch on whether or not the terms used match the phenotype terms (and trees) available in other commonly-used databases such as Mouse Genome Informatics.

We added a brief discussion of this issue to the second paragraph of this section: *"Terms and relations in existing ontologies are utilized wherever possible, but the OCDM adds rich detail not present in these ontologies."*

4. For the project "RNA dynamics in the developing mouse face", when discussing how this dataset could be mined, point ii focuses on upstream regulation of these features, when the effects of these events may be just as or more interesting and should be addressed. These might include differential isoform usage with unique functions across tissues and/or timepoints, differential isoform expression and transcript stability.

We appreciate this helpful suggestion. We did not want to highlight isoform usage in too much detail as the short reads available from RNAseq do not allow definitive identification of full-length transcript isoforms. Nevertheless, we concur with the reviewer and have added a note that *"determining how changes in splicing and/or promoter usage might impact the functionality of*

related transcripts and protein isoforms."

5. Two of the spoke projects, "Genomic and transgenic resources for craniofacial enhancer studies" and "Rapid identification and validation of human craniofacial development genes" do not have an associated figure. For the former, OPT visualization of a transgenic reporter embryo would be helpful.

We thank the reviewer for this suggestion. We have included visualizations of transgenic reporter embryos in the new Figure S2. Unfortunately, due to the current crisis which has closed laboratories and been very difficult for clinician-scientists, we were unable to obtain a figure from the "Rapid identification and validation of human craniofacial development genes."

6. Within the "Genomic and transgenic resources for craniofacial enhancer studies" project a unified analytical workflow is discussed. The authors should clarify whether or not future data submitters apply this workflow to their own data or whether the FaceBase Hub provides this service.

This is a service provided by the Hub that can also be performed on researchers' own computers. We have clarified this point in the text.

7. For the "Human genomics analysis interface" project, the authors should state the distribution of data across ages and ethnicities.

The interface developed for this project gives access to data from a number of different studies, such that this information is not easily summarized. We have added the following sentence to the text: *"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/)."*

8. For the "Rapid identification and validation of human craniofacial development genes" project, the authors should discuss whether their analyses resulted in the discovery of novel genes/variants or genes/variants that were previously known to underlie craniofacial dysmorphologies. Further, the authors should state how often the animal models used recapitulated the human phenotype.

We have added the information that 14 zebrafish models were generated from analyzed cases and that at least 8 new craniofacial disease-causing genes were identified, in addition to expansion of the phenotypes associated with at least 6 others. Unfortunately, we regret that we were unable to obtain further information from the "Rapid identification and validation of human craniofacial development genes" Spoke project in this difficult time.

Minor Comments:

1. Figure 1 would benefit from addition of a separate panel to demonstrate "digital dissection" of the skull, as in Figure 2D for zebrafish.

We appreciate this suggestion. The surface mesh format shown in this figure (now Figure 2) does not allow visualization in the same manner as the zebrafish model, which shows a different image format (volume rather than mesh). However, the mesh viewer also has a "digital dissection" (clip plane) function which we now show in an inset.

2. Figure S2 A labels the "FNP", which no longer exists at this timepoint as the development of the nasal pits has generated the MNP and LNP. The authors should specifically point out that the tissue taken included both of these structures. Further, in Figure S2 C, it may be more interesting to show a novel splicing event instead of the well-studied ectodermal vs. mesenchymal splicing of Fgfr2.

Thank you for these suggestions. We have added the # sign to FNP in the figure referring a note in the legend: *"# Note that over the time course of the analysis each FNP becomes divided by the invagination of the nasal pit into a medial and lateral nasal process. The "FNP" sample always included these two processes combined."* In Figure S2C, our original intention was to indicate that our separations of ectoderm and mesenchyme yield samples in which there is little contamination. On reflection, we concur with the reviewer and have replaced the *Fgfr2* data with images derived

from *Tpm1*. In this latter example, we see both changes in differential splicing and promoter usage, not only across layers, but also between time points in the same layer.

We noticed that this point regarding the FNP also applies to Figure S4 as the lacZ reporter transgenic mouse assay is performed at E11.5 when the medial and lateral nasal processes (MNP/LNP) are present. We have updated the figure to highlight the lacZ signal detected in the MNP/LNP (as opposed to the FNP which no longer exists at this stage) and updated the figure legend to reflect this change.

3. Within the project "RNA dynamics in the developing mouse face", the authors should also reference the recent paper demonstrating a role for splicing factor Rbfox2 in craniofacial development: DOI 10.7554/eLife.45418.

We thank the reviewer for bringing this to our attention and added this reference plus an additional one concerning *Esrp1*.

4. The gene names in the Legend of Figure S3 should be italicized.

Thank you for this attention to detail. We have fixed the italicization of *FAM222A* and *CACNA1C*.

5. Given its repeated use in the description of the previous project, the definition of enhancers in the first line of project "Genomic and transgenic resources for craniofacial enhancer studies" should be included earlier in the Article.

We appreciate this helpful suggestion and have moved this definition to the first mention of enhancers in this section (which now precedes the other enhancer project, after the thematic reorganization of the sections on the Spokes).

6. Figures 2B and 2D should be enlarged.

Efforts to edit this figure (now Figure 1) have unfortunately been hampered by a lack of access to laboratories shut down by SARS-COV-2. We feel that the current figure is satisfactory for the intended purposes.

7. The authors should state whether the reporter lines for chondrocytes, osteoblasts and osteoclasts described in "Anatomical atlas and transgenic tools for late skull formation in the zebrafish" were previously published or generated for this project. If the former, references should be provided.

We have added the appropriate citations.

8. The UMAP in the upper right of Figure S4 should have the different cell populations labeled as in Figure S2 D.

Thank you for this suggestion. The different cell populations in the UMAP have now been labeled. This figure has been fully revised. Please see the response to Reviewer #2, point 18 below for details.

9. The legend for Figure 6 should indicate that data for additional anatomical sites are available but not shown in the figure.

We have replaced this figure (now Figure 5) with an expanded version that displays all currently available anatomical sites.

Reviewer 2 Advance Summary and Potential Significance to Field:
Samuels et al. present a manuscript outlining recent accomplishments, technical organization and features, as well as future goals of FaceBase - a consortium aimed at integrating 'big data' across basic, translational and clinical craniofacial research. FaceBase not only serves as a repository for annotated and curated data sets relevant to craniofacial biology, but provides essential visualization and analysis pipelines for such data. Spoke projects described here address tissue- and

species- specific expression patterns and transcriptional control of facial development, as well as multispecies morphological WT and mutant data sets that have generated new predictions for molecular mechanisms underpinning craniofacial development and dysmorphology. Further, phenotyping pipelines are providing novel diagnostic tools. While many of the achievements of FaceBase 2 have generated multidisciplinary biological data sets, another accomplishment is the generation of an Ontology for integrating the data available across the platform. In developing infrastructure for robust data analysis and cross-experiment comparisons, FaceBase is now poised as a key platform for cross-discipline collaboration. FaceBase 3 aims to build responsibly upon the achievements of this report, to position itself as a cornerstone resource for craniofacial and developmental biology.

Reviewer 2 Comments for the Author:
This manuscript is appropriate for publication in Development as it presents an extremely rich resource for craniofacial, developmental and clinical biologists. As FaceBase 2, reported here, has used spoke projects to broaden the relevance of data available to a broader community than just craniofacial biology. However, the following comments must first be addressed.

We thank the reviewer for the positive comments and helpful feedback.

1. Craniofacial development offers a platform to understand key principles in developmental, molecular and cell biology. The abstract and introduction do not convey the broad interest of FaceBase to the community, e.g. those interested in epithelia or/vs mesenchyme, generally interested in ossification, chondrogenesis, signaling, stem cells etc. Adding some comment on this might extend the readership of this article.

2. The introduction fails to demonstrate the diversity of data sets or the tools generally available through FaceBase. Although this is explored later in the manuscript it might benefit the more broad readership of Development to emphasize this briefly earlier in the text.

Thank you for these two very helpful suggestions, as well as #28 and #29 below, which speak to similar points. We have revised the title, abstract, and introduction accordingly. The title has been changed to "*FaceBase 3: Analytical Tools and FAIR Resources for Craniofacial and Dental Research.*" We added further information to the abstract and the introduction and revised the third paragraph of the introduction to read: *"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."*

3. In the introduction, it is written that FaceBase is expanding its scope to be broader but no example is used to demonstrate this point. While not essential, adding to this would aid in managing expectations for the rest of the manuscript.

We have developed a list of priorities for the coming year, which we have now highlighted in the introduction: *"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/)."*

4. Is it 10-11 spokes funded per year or per version of Facebase?

We have clarified this in the text to read: *"Throughout its first (2009-2014) and second (2014-2019) iterations, known as FaceBase 1 and FaceBase 2, the consortium operated as a 'Hub and Spoke' model, with 10-11 Spoke Projects independently selected through a peer review process to generate and deposit data during each of these five-year periods."*

5. The motivation for OCDM is not well emphasized. Building an ontology for this network is

essential for data across FaceBase to be integrated. Further, this ontology adds to the utility of Facebase to the braoder community. Commenting on this idea would improve impact of this Spoke.

6. It is suggested that OCDM offers an understanding of how signaling is altered in craniofacial malformation. It would improve the broader impact of the paper to also indicate that this might be useful for understanding signaling complexity in general. Further, the tissue-specificity of signal function is also an important feature. While this is not essential, it is important to demonstrate that this resource is useful for researchers outside of the craniofacial community.

We agree with these important points about the far-reaching utility of the OCDM and have added the following to the first paragraph of this section: *"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."* We have also expanded the end of the third paragraph: *"Such pathways are becoming increasingly difficult for humans to comprehend, with the result that many computable signaling and pathway databases, often represented as OWL ontologies, have been developed. When these g efforts are complemented with highly detailed and specific ontologies like the OCDM, the combined, queryable resources can greatly facilitate our understanding of craniofacial malformations and their relations to broader conditions."*

7. The strength of 'Integrated research of functional genomics and craniofacial morphogenesis' is that many different kinds of data sets are integrated to address how signaling is used to mediate morphogenesis in complex craniofacial structures. The first two paragraphs do not demonstrate this point, in part because the connection between the first and second paragraph of this section is not very clear. It would benefit understanding of the manuscript if this were further clarified.

Thank you for this helpful suggestion. We have thoroughly reworked the section on this Spoke to read more easily and clarify this connection.

8. This reviewer understands the final comment in the final paragraph of this section (see 7) to indicate how reproducible and integrated different datasets are in FaceBase. Is this correct? If so, this should be stated explicitly to make very clear the strengths of this spoke and may comprise a stand-alone point.

We believe this is a strength of FaceBase in general. We have moved the sentence about collaborations to which the reviewer is referring here ("*FaceBase provides an ideal platform for collaborations*") to the Introduction and revised it to read: *"Moreover, FaceBase provides an ideal platform for collaborations through pre-publication access control, data curation tools, emphasis on reproducibility, and integration across datasets."*

9. Further to comment 8, it is simply stated that 'FaceBase provides an ideal platform for collaborations' but why this is placed here is not entirely clear. If this point is to be made, it would be strengthened by linking to future directives discussed later; if relevant, how FaceBase helps mediate collaboration. Additionally, adding specific motivation for focusing on mandibular data for this spoke would not only improve the narrative of this section but could be used to demonstrate how these data sets bring together different disciplines. It should, however, be made clear that the microCT data includes also the entire skull which will be useful for other studies.

Please see the response to #8 above. In the thoroughly revised section about this Spoke, we have described the motivation for focusing on jaw development: *"the Chai FaceBase 2 Spoke focused on jaw morphogenesis because deformities of the mandible and maxilla are relatively common; to take one example, maxillary hypoplasia is often associated with cleft palate has been described in more than sixty syndromes (Hennekam et al., 2010; Jin et al., 2012). Despite their importance, the mechanisms that regulate facial bone development have not been well characterized."*

We have also added the statement that *"these microCT images may be of broad interest for studies on craniofacial development and malformations since they include the hard and soft tissues of the entire head."*

10. Fig.1 Legend. It might be necessary to make clear that the different colors indicate distinct bones. 11.Fig. 1 What do the blue dots represent in this image? This should be included

in the reference or they should be removed.

We appreciate this suggestion and clarified these points in the figure legend (now Figure 2). The blue and red dots indicate the anatomical landmarks that this spoke established for mandible and maxilla, which can be used for cross-model comparison.

12. While it might be assumed in the section titled 'Genomic and transgenic resources for craniofacial enhancer studies' that OPT data can be visualized using FaceBase tools, however, this is not explicitly stated. It might benefit the reader to be reminded of such functionality.

Thank you for this helpful suggestion. FaceBase indeed offers a 3D viewer through which OPT data can be visualized online, although improvements to the viewer will be necessary to show the data from this Spoke project in color so that the red/green signals can be properly appreciated; this work is planned for the coming year. Other Spoke projects including the zebrafish atlas already make extensive use of FaceBase's 2D and 3D image viewers, so we have added a statement to the description of the zebrafish atlas project noting that "*These datasets have been optimized to enable visualization of thumbnail images and interactive 3D views in any modern web browser.*"

13. While it is stated that the Zebrafish spoke paves the way for similar work in other fish, the reasons why this might be of interest to the broad community is not clearly articulated. Such impact could include understanding how bone adapts to distinct stress for example provided by different feeding schemes.
Although not essential, the import of this spoke might be emphasized by including a brief discussion of such implications.

We appreciate this important note about the cross-species utility of the zebrafish data and have highlighted it in the text.

14. Fig. 2. Requires scale bars to be added to images where possible (in particular, a, c and E) to emphasize the scales represented by the rich Zebrafish data set.

15. Fig. 2E. The figure might make more impact if the mouse comparison was included. In this way, the figure could also demonstrate how this data set is integrating into corresponding mammalian resources.

We are unfortunately unable to provide scale bars for B, D, and E or to edit this figure (now Figure 1) further due to lack of access to laboratories shut down by SARS-CoV-2. As these are representative figures and not intended to be used as a judgment of relative size between the samples, we feel that these are not essential for the intent and data within the figure.

16. The following statement should be updated with relevant references: Sutures differ widely in their physical structure, cell lineage, mechanical environment, and susceptibility to craniosynostosis.

We have placed the reference to Heuze et al. 2014 at the end of the paragraph and added an additional reference, Richtsmeier and Flaherty 2013, to encompass these statements.

17. It is important to remind the reader that the models used in the multi-suture bulk transcriptomics are craniosynostotic and therefore models not only for craniofacial dysgenesis but premature suture ossification Further, though not essential the reader may understand the importance of this spoke if it is mentioned that sutures also harbor a skeletal stem cell population that is poorly understood.

To make the presence of the sutures from craniosynostosis models more clear, and to address the issues of stem cells, we have revised this portion of the text and added two references for the statement about stem cells. This section now reads: *"In addition to RNA-seq datasets of wildtype (WT) mice, they include those for 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 WT 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).”

18. Fig. S4. It would make it simpler for the reader to understand this image if the arrows from the diagram to the data set were labelled with seq type e.g. scSeq and BulkSeq, respectively. While a key for these cell clusters in the UMAP plot is not absolutely essential it would improve the figure by highlighting cell types of interest to the broader community.

Thank you for these suggestions about this figure, which is now Figure S1. We have thoroughly revised the figure. Each element is now labeled A, B, C, and D. We have expanded the schematic of calvarial bones (A) to include calvarial and facial bones with all sutures labeled and indicated by an individual color. We have revised the generic suture schematic (B) to make it larger and with more readable font. As the reviewer requests we have labeled the bulk RNA-seq heatmap (C) and single-cell RNA-seq UMAP figures (D), although the arrows in the original figure have been removed. The relationships between the figure panels are clearly labeled with bulk and single-cell RNA-seq and the figure legend is revised accordingly. As the reviewer requests we have labeled cell clusters in the new UMAP plot in D.

19. Fig3B is not discussed in the text, is this included to demonstrate one parameter that can be plotted or is it intended to add information? The relevance/importance of this panel should be indicated.

We have revised the reference to Figure 3 to include both panels: *“Using data that were collected during the course of the project, the team analyzed 3D images from >3000 individuals with hundreds of different syndromes (see sample size distribution in **Figure 3A**), as well as thousands of unaffected related and unrelated individuals; the age distributions of syndromic subjects and their relatives are shown in **Figure 3B**.”*

20. Fig. 4. Some descriptions of the figure might help understanding, e.g. Does DB mean database? Do the coin purses represent exchange or that these exchanges require additional financial support from researchers? Do the arrows indicate physical information flow or work flow perhaps? In either case, why is the arrow to web data browser from core services not bidirectional?

We appreciate these suggestions for improving Figure 4 and have revised it accordingly. The coin purses were only intended to represent exchange of data, not funds. We have changed these to file folders to avoid this confusion and have expanded the caption to describe all the relationships represented by the arrows.

21. Fig. 4 Can the expression data be viewed in the visualization tools - in some cases 3D? If so, this is an advantage could be further emphasized either in the text or added to the figure.

The current interface for accessing expression data in FaceBase is not equipped with browser-based viewers. However, this is a feature (e.g. volcano plots, Manhattan plots) that the Hub development team has discussed and is considering implementing in the future.

22. Fig. 4. Does the resource hub refer not only to the original hub but also the spokes? Indicating this may make the scheme more clear if comment 8 is addressed.

The “Resources Hub” (https://www.facebase.org/resources/) is a particular portion of the FaceBase website that includes resources above and beyond the datasets produced by the spoke projects. We now note this in the legend and in the subsection entitled “Resource hub.”

23. It is not quite clear what fig. 5 adds beyond what is noted in fig. 1. If kept, it would be interesting for the reader to know which part of the skull has been selected for viewing in fig. 5/what mesh type was used.

Upon further reflection on these figures, we decided to remove Figure 5 entirely as we came to agree with the reviewer that it was redundant.

24. Why are teeth and salivary glands priority areas, vs other regions? While it is modestly touched upon in the following paragraph, this should be commented on in the text earlier.

The data recruitment priorities were developed by the FaceBase Hub and external scientific advisors before final approval by NIDCR program staff, building on suggestions gleaned from various stakeholders. These priorities are revised on an annual basis. We have added a description of this process and a link to the data recruitment priorities document in the text.

25. Fig. 6. The anatomy labels are clipped on the right of the image.

Thank you for this attention to detail. We have replaced the figure (now Figure 5) with an expanded image in which all anatomical labels are visible.

26. Fig. S5B. What do the different color/shapes mean in this plot?

The colors indicate r^2 values, from < 0.2 (blue) to > 0.8 (red). We have added this information to the legend of this figure, now Figure S6.

27. One question that could be addressed briefly, is how instructions for features such as automated analyses organized in FaceBase. This could be included or simply referenced.

We have provided a link in the Sustainable Data Curation subsection to the data curation/submission wiki that includes thorough documentation. Because the procedures may change over time, it is best to refer readers to this documentation.

28. The abstract fails to convey the breadth of craniofacial development or data types that FaceBase2 utilizes. This must be rectified as the general readership should be incited to know more about the service the FaceBase currently supplies.

29. Further to comment 28, the title also does not convey the technological and resource advancements of Facebase 2.

Thank you for these very important suggestions, which we discuss in our response to comments #1 and
#2 from this reviewer.

Reviewer 3 Advance Summary and Potential Significance to Field:
This manuscript by the FaceBase Consortium summarises the datasets generated by the Phase 2 FaceBase projects from 2014-2019. While Phase 1 projects previously described focused on the midface development, FaceBase 2 covers craniofacial development more broadly. One of the most considerable assets and achievements of the Consortium is the creation of the FaceBase Data Hub, which includes more than 850 datasets in free access to the scientific community, featuring new web browser-based tools and enabling users-led custom analyses.

Reviewer 3 Comments for the Author:
Facebase 2 Spoke Projects briefly introduced and presented are diverse, highly innovative and at the cutting-edge, covering different molecular, cellular, epigenetic, gene regulatory and disease-related aspects of craniofacial biology, including the development of data interactive tools, 3D craniofacial morphometry tools and establishment of ontology for use by the craniofacial filed.

The Resource Hub presented by the Consortium enables access to the data produced and offers a comprehensive reference for the craniofacial community that can be browsed using either a rich web application or desktop client to mine and visualize the data.

The manuscript presents a clear overview of the second phase of this highly successful Consortium, not only clearly introducing the projects involved, but also highlighting the usability of the databases and interactive tools created, of high interest to the developmental biologists.

We very much appreciate the reviewer's positive comments about our manuscript.

Second decision letter

MS ID#: DEVELOP/2020/191213

MS TITLE: FaceBase 3: Analytical Tools and FAIR Resources for Craniofacial and Dental Research

AUTHORS: Bridget D Samuels, Robert Aho, James F Brinkley, Alejandro Bugacov, Eleanor Feingold, Shannon Fisher, Ana S Gonzalez-Reiche, Joseph G Hacia, Benedikt Hallgrimsson, Karissa Hansen, Matthew P Harris, Thach-Vu Ho, Greg Holmes, Joan E Hooper, Ethylin Wang Jabs, Kenneth L Jones, Carl Kesselman, Ophir D Klein, Elizabeth J Leslie, Hong Li, Eric C Liao, Hannah Long, Na Lu, Richard L Maas, Mary L Marazita, Jaaved Mohammed, Sara Prescott, Robert Schuler, Licia Selleri, Richard A Spritz, Tomek Swigut, Harm van Bakel, Axel Visel, Ian Welsh, Cristina Williams, Trevor J Williams, Joanna Wysocka, Yuan Yuan, and Yang Chai
ARTICLE TYPE: Techniques and Resources Article

I am happy to tell you that your manuscript has been accepted for publication in Development, pending our standard ethics checks.

Reviewer 1

Advance summary and potential significance to field

This Techniques and Resources Article presents an overview of datasets generated from FaceBase 2 projects over the last five years, describes data analysis and visualization tools available through the FaceBase Hub website and touches on future plans to expand the Hub. The information presented here, together with the FaceBase Hub itself, will be of major value to members of the developmental biology community interested in both normal and disordered craniofacial development. The article does a commendable job of explaining how this resource conforms to the FAIR data principles, thereby ensuring that the datasets are accessible reproducible and reusable.

Comments for the author

All previous requests for revision have been satisfactorily addressed.

Reviewer 2

Advance summary and potential significance to field

Samuels et al. presents the progress made in a key resource for the craniofacial and developmental biology community.

Comments for the author

The significant changes to the submitted article by Samuels et al are excellent and have much improved the impact. They present an outstanding resource for Development's readership.

Reviewer 3

Advance summary and potential significance to field

The manuscript has been revised appropriately to address the reviewer's concerns.

Comments for the author

N/A