

Gene expression dynamics of human and mouse craniofacial development at the single-cell level

Corresponding Author: Dr Justin Cotney

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript "Gene expression patterns of the developing human face at single cell resolution reveal cell type contributions to normal facial variation and disease risk" by Khouri-Farah et al. describes a new resource providing high-resolution single-cell transcriptome data for multiple stages of human embryonic development. Importantly, these stages cover critical morphogenetic processes relevant to human facial variation and craniofacial birth defects.

There is currently a sparsity of high-quality data from actual human tissues covering these processes. Considering differences in craniofacial development and regulatory sequences between human and mouse (a model for which various datasets are available), the data provided in this study will be a rich resource for craniofacial researchers interested in studying general developmental processes, as well as the impact of human genetic variation on these processes in the context of birth defects. The authors provide an easily usable interface for researchers to explore the gene expression profiles in specific cell-type clusters, which will facilitate the utilization of their data by others.

While the data provided through this study are without doubt valuable and deserve to be adequately described in a publication, the manuscript itself has significant room for improvement. The text is much longer and detailed than needed and many of the figures are convoluted and contain unnecessary detail. Also, the rationale for some of the analyses performed in this study is not clearly described.

Overall, the generation of human single-cell RNA sequencing data across multiple timepoints is an important addition to the craniofacial field, but the current manuscript would require major revisions to make this valuable resource easily accessible and provide the advance and insight that would make it a strong candidate for Nature Communications.

Major points

1. All figures need major revisions so that they can be easily interpreted. All plots, text, legends, axis labels need to be readable. The authors should think carefully about what is the key takeaway from each figure and which plots are absolutely required to make that point. Keeping a consistent color theme throughout would also improve the accessibility of the figures.
2. The main text is far too long and needs to be at least 50% reduced. The readability of this manuscript also needs improvement. There are many different analyses presented but the rationale for their inclusion and key conclusions for the reader are not always clear. I would recommend adding a Supplementary Information section where more details can be added to improve the structure of the main text and make it easier to follow.
3. The human single-cell data from this study was produced using 10X Multiome but only the RNA-seq component of the data is presented here. The ATAC component of the data is not needed for the purposes of this manuscript (which focuses on transcriptomes) but it would be good to mention for the reader why it is not reported. Was it of insufficient quality to provide useful results? Will it be described in a complementary manuscript? Some other reason?
4. It's surprising that Sox10 does not come up in any of the human CNCC subtype clusters, considering that it is a key marker in the mouse CNCC and given its known role in cranial neural cell development and differentiation. Why do the authors think that is? Given the overlap in markers with neural cells, is there any concern that human nuclei of CNCC were

accidentally filtered out when removing neuronal cell types? E.g., TUBB3 is expressed in premigratory cranial neural crest cells in birds and maintained in neural crest derived tissues/cell types. See: doi: 10.1016/j.gep.2019.119067.

Minor comments

1. The PDF files are very large (causing poor display/printing issues on some systems) because dot plots are pasted in vector format with large numbers of individual datapoints. It would be helpful to rasterize plots so that the Figures can render more quickly and make it easier to review.
2. The sentence in the introduction about Neanderthal-introgressed sequences seems a little out of place. Perhaps move this up to the section about normal facial variation and give some clearer rationale for why this analysis was done/included.
3. Typo on 105 - doublets.
4. Line 113: Which clusters were problematic? It's not easy to tell from S1b as the colors merge together.
5. Line 117: Please specify that nuclei are filtered specifically based on TUBB3 and MAP2 expression and not the other genes shown in S1.
6. Figure 2B: Perhaps only show top 5 gene markers per cluster so that the y-axis is more easily read.
7. Figure 2C - it's not possible to read the terms in the smaller font. Simplify to top 3-5 enriched terms per cluster that can all be shown and read.
8. The Venn Diagram in FigureS4B may not be necessary. It's not easy to interpret and the upset plot clearly shows the intersections.
9. Line 136-138: This statement is somewhat unclear as at this point in the manuscript no enrichments for craniofacial biology and developmental abnormalities has been shown. I think the conclusion of this paragraph can be kept as simply "Overall, these results suggest that our single-nucleus expressions data closely resembles the bulk gene expression data."
10. Reword sentence at 154-155. It's not necessary to state "some of the timepoints". I suggest the authors rephrase so that it's clear that at CS12 all samples were derived from female embryos and at CS13 all samples were derived from male embryos, but that this did not introduce any significant bias.
11. Typo Line 155 - main cluster -> main clusters.
12. Figure S5A - y-axis text is overlapping and not readable.
13. Line 258 sub-cluster cnl2 is not described - what is this cluster?
14. Line 271: clusters cnl1, cln2 and cnl3 need descriptions.
15. Figure 3C: Do the different colors represent the Carnegie stages in this plot? Can this plot be simplified to highlight the genes referred to in the text? Or perhaps a heatmap would be easier to interpret.
16. Figure 3D: Challenging to read the y-axis.
17. Figure 3E: I suggest changing the color scale for these plots so that cream or a light color represents lowly expressed genes and a darker color (e.g., red/blue) represents highly expressed genes
18. Line 300: Muscle cells clusters have been observed in the single-cell data from <https://doi.org/10.1038/s41467-024-46396-4> and <https://doi.org/10.1038/s41467-023-41349-9>
19. Line 301: It seems unlikely that sampling differences and/or timepoints would be the reason these cells are not observed but I would imagine it's due to the lack of specificity in resolved clusters. E.g., Sox10, one the main markers used by the authors for specifying CNCC cells, is also a specific marker for neural-crest derived Schwann cells. Is it also possible that CNCCs had not been resolved in preexisting single-cell datasets because they were instead labelled broadly as neural cell types?
20. What is the rationale for selecting E10.5-E12.5 in mouse for resolving muscle cells?
21. Figure 4 needs to be simplified. It has too many panels and it is difficult to understand what the key takeaway is meant to be. What are the absolute key plots to show? And can the rest be put in supplements? For example, Figure 4K and Figure 4N are not readable.
22. Figure 4G: Please specify what the subtype labels represent.

23. When comparing single-cell data between species the authors could consider joint-embedding with SAMap (<https://doi.org/10.7554/eLife.66747>) to identify homologous cell types.

24. For the cross-species cell type comparisons, what gene list was used as background for the enrichment analyses? Is it all human/mouse genes or genes that have 1-to-1 orthologs between mouse and human?

25. Figure 7: It is difficult to understand the main takeaway from these plots. The heatmap legend is not labelled and the values are not easily interpreted. Perhaps paring these heatmaps down to only those that are significant for the main figure could help. The x-axes are not readable so it's impossible to know which sub-cell types have specific variant enrichments. In Figure 7C, do the different colors represent the different human stages?

26. Line 621: What is this referring to? The HARs or overall for all gene lists?

27. Line 713: From Supplementary Table 7, Sox10 does not appear in the CNCC subtypes for human data.

(Remarks on code availability)

I did not review the code in detail or attempt to install/run the code.

Reviewer #2

(Remarks to the Author)

The manuscript by Khouri-Farah et al describes a novel dataset derived from single-nucleus RNA-seq analysis of human embryonic face and head development across weeks 4-8 after conception. They define several broad classes of cell types, including epithelial and mesenchymal, and a population defined as neural crest, and delve deeper into the cell identities by sub-clustering. Further, they perform a similar analysis using approximately equivalent stages of mouse embryogenesis and compare and contrast the data obtained between the two species. The authors next used a published human spatial transcriptomics dataset as a means to validate several of their findings concerning their human assignments. The human single-nucleus RNA-seq data are also used in conjunction with several public databases to make predictions concerning disease associations, human facial shape variation, connections between human and Neanderthal facial morphology, and clefting genetics. Overall, the dataset is a valuable additional resource for people interested in human embryonic head development that adds to other bulk, single-cell, and spatial datasets focused on the entire embryo. However, there are several significant issues with the manuscript, particularly with the analysis and interpretation of the datasets, that considerably reduce its impact and utility.

Major issues

1. The manuscript is too long, and there are many sections that are redundant or repetitive. The journal's instructions for authors state that ideally the main text should be 5,000 words. This manuscript is well over 10,000 words, with the discussion alone about 3,000 words. Much of the data from the results section is reiterated in the discussion and there are paragraphs dealing with disease correlation analyses of red blood cells and immune cells that are not pertinent for an analysis focused on the craniofacial region. For example, the paragraph beginning on line 191 concerning ontologies associated with red blood cells including malaria or hemolytic anemia or the paragraph beginning on line 203 concerning immune cell related diseases. The length of the manuscript and its multiple individual correlations with other datasets means that the focus of the paper is lost, and the key points are difficult to extract from the study. Overall, the manuscript would be more impactful if it was considerably shortened and focused on key findings concerning the craniofacial region.

2. A major flaw in the manuscript is a lack of clarity concerning the regions under analysis in the sn-RNA-seq study. For human, only the following statements are provided: Results "we profiled the entire craniofacial prominence" and Methods "Primary human craniofacial tissues were obtained". For mouse: Results "we generated single-nucleus gene expression from mouse craniofacial tissue. These samples reflected the major morphological landmarks of the human tissue profiled allowing a more direct comparison of cell types". Methods "primary mouse craniofacial tissues ... were pooled". Unfortunately, these extremely limited descriptions do not provide sufficient information to understand or review the manuscript with any degree of certainty. One illustration of how this lack of detail affects the manuscript is that – as illustrated in Figure 6 - the craniofacial human ectoderm samples contain tissues related to the optic and otic placodes as well as to the neck (thyroid). These findings indicate a much greater region of the head and neck were surveyed than just the craniofacial prominences shown in Figure 1 and it is imperative to know the morphological boundaries that were used for the dissections.

As another example, on line 116, describing clusters expressing TUBB3 and MAP2: "We reasoned these clusters were derived from developing brain tissue not directly part of the craniofacial structures". In removing these "brain clusters" their sample size decreased from 86,359 nuclei to 42,131, a drop of over half. These data indicate that their dissections of craniofacial tissue were not precise but included a majority of cells from other regions of the head. On the other hand, throwing away cells with expression of these two markers likely means they are losing cells from critical craniofacial structures, particularly the trigeminal ganglia. The authors indicate that the mouse data were obtained and processed in a similar manner, but it would be valuable to know if it required excluding Tubb3 and Map2 cells, and if so what percentage of mouse cells were removed from the analysis to help determine equivalence of dissections.

These considerations are especially important when the authors discuss the idea that there are multiple marker genes that were only expressed in a single species (Results: Section titled "Species-specific differences in marker gene expression

during human and mouse craniofacial development"). Examination of the data provided in Table 20, for example, states that genes including ALX3, DLX2, DLX3, DLX5, and DLX6 are only detected as expressed in the human facial mesenchyme and not in the mouse "Only called in a single species". This is certainly not supported by multiple other studies, raising concerns about the scope of the dissections, the data processing, and attention to detail in data analysis.

3. The question of anatomy and the regions dissected also has a direct bearing on the parts of the paper discussing the identification and analysis of clusters 1 and 6, called as mesenchyme and cranial neural crest cells, respectively. The mesenchyme of the head is derived from two major tissue-types, the mesoderm and the neural crest. Fate mapping studies have indicated that the majority of the mesenchyme in the craniofacial prominences is of neural crest origin, whereas mesodermally derived mesenchyme occupies areas of the head that are above and behind the developing eye. There is also a core of mesoderm tissue in the arches, including arch 1, that will form muscles of the head and neck. Therefore, it would be expected that a significant portion of the mesenchyme in cluster 1 would have a neural crest origin. However, the ratios of neural crest versus mesoderm derived mesenchyme will vary depending on the amount of tissue included from more posterior regions of the head. This consideration again makes it imperative to understand the boundaries of the dissections performed.

Even if a large amount of mesodermally-derived mesenchyme was present in cluster 1, there should still be a significant contribution of neural crest to this cluster. In this respect, when the authors subcluster the mesenchyme, they identify maxilla, mandibular and nasal prominence cells indicating the presence of neural crest derivatives. Therefore, cluster 6 should not be addressed as the only cranial neural crest cell population in their datasets. In fact, examination of cluster 6 suggests it may represent a few specialized neural crest derivatives – particularly Schwann cells as well as a few melanocytes. For example, as shown in Figure 3C, PLP1 - a prominent component of myelin - is expressed in all subclusters of cluster 6. Similarly, Mpz (myelin protein z) is broadly expressed in most subclusters. Together, these observations suggest that the mesenchyme has been mis-classified as not containing neural crest, while cluster 6 has been annotated as the major cranial neural crest population, whereas it represents highly specific populations of neural crest derivatives.

4. The authors define many subclusters in the mesenchyme, ectoderm, and neural crest but do not adequately describe how they were assigned to specific regions or structures. For example, for the ectoderm, TBX18 was assigned as a fusion zone marker, but it is not clear why, nor is it clear what fusion zone they are discussing anatomically. These issues of mapping and assignment are compounded by an absence of appropriate verification analyses. Here the authors rely on published data of spatial transcriptomics of two sagittal sections through a CS13 human embryo to overlay their expression data which are then presented in several figures throughout the manuscript. There are two problems with this approach. The first is that they are relying on just two sections and so it is not possible to determine the 3-dimensional expression patterns of any gene or cluster, meaning that considerable information is lost regarding assignment and anatomy. The second issue is that the resolution of spatial mapping was a 55µm spot, whereas an approximately 2µm spot would be required to accurately map a transcript to a single cell layer. Therefore, the study fails to link subcluster identity with specific anatomy at a meaningful resolution and in the absence of any other verification studies this significantly reduces the power of the dataset.

Minor issues

1. Line 462-463. "with LNP1 being more posterior and LNP1 being more anterior". One of those should not be LNP1.

2. Line 625. "We found no enrichment in any cell types...". This line is duplicated in the next sentence.

3. The authors refer to Li et al (reference 12) throughout the manuscript concerning mouse craniofacial data. However, this reference describes genome analysis of human brain data.

4. I have a number of suggestions for the images of the human sections. First, as the manuscript is focused on the head and face, it would be more useful to show only these structures rather than the whole embryo to reflect the focus of the paper. Second, orientation should be indicated on the panels: left/right (i.e. for Fig 4C), anterior/posterior, rostral/caudal. Third, it would be useful to have one or more annotated images describing the pertinent structures present in the images.

5. The figure legends for the Supplemental figures are missing. There was also no explanation of the information provided in the Supplemental tables.

6. Multiple instances: "this data" should be "these data".

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The paper by Khouri-Farah et al., "Gene expression patterns of the developing human face at single cell resolution reveal cell type contributions to normal facial variation and disease risk" reports a single-nucleus RNA sequencing (snRNA-seq) atlas of human craniofacial development from craniofacial tissues of 24 embryos that comprise six key time points spanning 4–8 post-conception weeks. The Authors describe seven major cell types, including muscle progenitors and cranial neural

crest cells (CNCCs), as well as multiple subtypes of ectoderm and mesenchyme. Comparative analyses reveal substantial conservation of major cell types, alongside human biased differences in gene expression programs. The Authors also claim that CNCCs exhibit the lowest marker gene conservation, underscoring their evolutionary plasticity. Spatial transcriptomics is also employed to further localize cell populations, which should provide a detailed view of their anatomical context. Lastly, the Authors attempt to link these developmental processes to genetic variation, identifying cell type-specific enrichments for common variants associated with facial morphology and rare variants linked to orofacial clefts.

Undoubtedly, this atlas has the potential to offer critical insights into the cellular and genetic mechanisms shaping the human face, highlighting conserved and distinctly human aspects of craniofacial biology and linking them to the developmental origins of craniofacial disorders and the genetic basis of facial variation. This work could certainly provide an important resource for exploring craniofacial biology, with implications for developmental genetics, evolutionary biology, and congenital anomalies. Therefore, the study could be of interest and could fill a poorly explored aspect of human craniofacial biology by reporting studies conducted on human embryos as well as mouse embryos. However, there are multiple serious concerns regarding the current manuscript that need to be fully addressed. I list below some of the most salient issues that need to be resolved in order to increase the rigor and accuracy of the described findings, as well as supporting many of the conclusions.

Single-nucleus RNA Sequencing (snRNA-seq) Data Generation and Processing (Human Samples)

- It would be important to show a cartoon schematizing the regions of the human embryonic heads that were dissected. The Authors state that they processed the “developing craniofacial region” (quote) of human embryos from 4 to 8 weeks post conception. Lacking more detailed information, the reader is left guessing which parts of the developing craniofacial region were processed.
- How did the Authors generate the initial clustering of the major cell populations? Which Seurat version did they use? Which other packages were employed? All relevant information is lacking.
- More details are required in the Methods sections describing how the single-cell data were embedded and clustered, e.g. how many highly variable genes, PC dimensions, and nearest neighbors used to generate the UMAPS appearing in the main figures. Was any batch correction used? Were distinct embeddings calculated for the subset of neural crest clusters?
- It would also be important to provide separate clusters for each gestational time analyzed. This would highlight differences that are to be expected in cluster transcriptomes across the different gestational days, which are currently not clear.
- Why do the Authors use only TUBB3 and MAP2 as markers for neuronal cells that they want to remove from the query set? While primarily known as a neuronal marker, TUBB3 expression is also observed in non-neuronal cells and tissues, including endothelial cells, fibroblasts, spermatogenic cells, and stem cells, suggesting a broader role beyond just neurons. The choice of neural markers appears to be ill-informed and extremely reductive.
- Regrettably, there appears to be lack of rigor in tissue collection. It is of concern that the Authors lose half of their cells through the first filtering step. According to their statements, they started from 86,359 individual nuclei, but ended up with 42,131 remaining nuclei for their analyses. They concluded: “We reasoned these clusters were derived from developing brain tissue not directly part of the craniofacial structures. We therefore excluded these nuclei from downstream analyses resulting in a total of 42,131 remaining nuclei with an average of 4095 nuclei from each sample”. In summary, either the contamination of facial tissues by brain tissue was massive, or the Authors also removed cells that are not neural cells and thus could be integral part of head developing structures.

- The Authors state: “Early samples had consistently higher mitochondrial reads (Fig. S2d), potentially reflecting their higher dependence on mitochondrial output or an artifact related to lower cell numbers in the processing of each sample.” This assertion is superficial and must be revisited. It is well known that high mitochondrial signatures are a strong indicator of apoptosis. The Authors have not unequivocally ruled out the presence of cell death in their samples.

Single-nucleus RNA Sequencing (snRNA-seq) Data Generation and Processing (Mouse Samples)

- Similar concerns are present regarding the isolation of mouse tissues. There is a lack of adequate information regarding the collection of the mouse tissues. Were mouse tissues also fresh frozen + thawed (like human tissues) or were they processed directly following tissue collection? Differences in tissue processing could heavily influence the snRNA-seq results obtained. There are serious concerns that the “species-specific” signatures the Author report might be due at least in part to differences in processing of the collected tissues, and they may not necessarily underpin species-specific identities.

The authors state that: “Samples were mechanically broken into single-cell suspensions” and that “Craniofacial prominences were collected in a very similar fashion to human samples”. What do the Authors mean by “similar”? Were neural tissue also removed in mouse craniofacial tissue dissections as it was done for human dissections? Or was there absence of neural contamination? All of these details, which have critical bearing on the reported findings of species-specific differences, are not provided.

- The Authors compare their datasets to those reported by the group of Trevor Williams (Li, H., Jones, K.L., Hooper, J.E. & Williams, T. The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. Development 146, 2019). But the latter datasets were generated only from murine midfaces and only at E11.5. This substantial difference in the datasets should at least be mentioned.

Neural Crest Cell Cluster (Cluster 6)

- What is cluster 6? It is not at all clear whether this is a biologically defined group of cells or a somewhat “mixed bag of cells” cluster. The Authors seem not to be completely sure themselves: “We then turned to the not yet concretely identified cluster 6 [...]. We did not detect any significant pathway enrichments for this particular cluster (201). “The wide variety of biological functions and specific disease enrichments all suggested that this cluster might be enriched for neural crest cells. (208-210)”.

These descriptions are rather confusing.

- The Authors state that CNCC markers are significantly higher in cluster 6.

“When we examined a module of genes from regulatory networks recently identified in cultured human and chimpanzee cranial neural crest cells (CNCCs)69, we found significantly higher expression in cluster 6 (Fig. 2D).”

Which test or package or statistics did they use to reach this conclusion? Again, no relevant information is provided.

- Do all developmental stages contribute equally to the population in cluster 6? This information is lacking and is very important.

- The Authors state: “SOX9 and COL20A1 were more specifically expressed across the (CNCC) clusters, but again in non-overlapping patterns.” (261-262).

This sentence is very confusing. What do the Authors mean?

- There appears to be lack of a definite proof that Cluster 6 is truly a Neural Crest Cell cluster.

Spatial Transcriptomics

- There appears to be huge differences in between the two replicates shown for spatial transcriptomics (Fig. 3 E,F,G). For certain markers and modules the two replicates shown in the figure appear to be completely different from one another regarding the respective gene expression spatial features. Clearly

these human samples are precious and hard to collect, making it difficult to generate multiple biological replicates.. However, it is very hard to draw firm conclusions from these spatial validations.

Enrichment of Disease Genes and Variants Linked to Facial Variation

These analyses are hard to follow and very cryptic. Disease associations are correlative. Certain statements and conclusions do not appear to be based on rigorous supporting data. One example is provided below. The Authors state: “Interestingly, while CNCCs certainly give rise to many of the downstream cell types and tissues, we found relatively few shape associations for CNCC subtypes” (531-532)

This statement is somewhat concerning, given that a rich body of literature directly links differences of facial shapes to variants in CNCC genes. The following review article comprehensively summarizes most of the relevant literature showing that variants in CNCC genes and enhancers underpin differences in face variation: Naqvi S, Hoskens H, Wilke F, Weinberg SM, Shaffer JR, Walsh S, Shriver MD, Wysocka J, Claes P. Decoding the Human Face: Progress and Challenges in Understanding the Genetics of Craniofacial Morphology. *Annu Rev Genomics Hum Genet.* 2022 Aug 31;23:383-412. Interestingly, the Authors cite this review article in their “Introduction”.

Minor points

- Inconsistency in size of figure labelling (A,B,C etc.), legends in multiple figures (partially) overlap, image quality in multiple figures is poor.
- Lettering in some figures is exceedingly small and/or warped and thus very hard to read, even when zooming in onto the image or cartoon.
- The Authors need to revise the nomenclature for genes. Mouse genes should not be capitalized.
- Some of the references cited are not correct. For example, Ref 12 (line 146) is not correct.

Style and Grammar (only a few issues of many are reported below):

- 105: “after filtering for doubles and quality of per nucleus data“. Confusing sentence.
- 106/107: “Samples obtained from human embryos are more difficult to control what is collected”. Sentence needs rewriting.
- 256: “TFAP2A was expression was observed in all clusters”. Sentence needs rewriting.
- 625-627: “We found no enrichment in any cell types for genes affected by de novo synonymous variants. We found no enrichment in any cell types for genes affected by de novo synonymous variants.” The sentence is repeated twice.
- 758: space and period missing: “embryo25By“
- 815 f. Clip -> lip: “Failure of this region to fuse in humans has been suggested to cause cleft lip that could also involve the nostril region and primary palate“
- 833: dominant or major: “Mesenchyme is by far the dominant major cell type“
- 993: described: “as describe for human“
- 1019: remove word duplication: “mouse genes were reduced and converted to those to those with one“
- 1058: space and period missing: “MAGMA89Frontal“

Lastly, this reviewer suggests careful editing of the text – in its present form, the manuscript appears to have been rushed and, regrettably, it is overall sloppy.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

The authors present a comprehensive single-nucleus RNA sequencing (snRNA-seq) atlas of human craniofacial development, derived from 24 embryonic samples spanning six key time points during the embryonic period. Additionally, they integrate spatial transcriptomics data to further localize cell populations. Our findings shed light on the developmental origins of craniofacial disorders, the genetic basis of facial variation, and the evolutionary legacy of ancient hominins. Overall, this work provides a valuable new atlas for craniofacial biology, supported by comprehensive experiments. However, there are still areas that require improvement. Specifically, some paragraphs should be refined or removed to enhance readability. At present, several sections in the Results are overly redundant, making it difficult to focus on the key findings. I encourage the authors to revise the “Results” section for improved clarity and readability.

My comments are as follows:

1. Beyond providing a more extensive atlas, have the authors identified any conclusions that Yankee, T.N. et al. (Ref 13) did not?
2. The authors state, “Closer inspection of the genes strongly expressed in these clusters revealed many canonical neuronal genes, such as TUBB3 and MAP2.” How did they quantify “strong” expression? Furthermore, they excluded nuclei derived from developing brain tissue. How can they ensure that all unrelated clusters from craniofacial structures were successfully removed?
3. Line 172: Why did the authors choose 100 marker genes per cluster? In contrast, in line 156, they selected only 10 marker genes to identify unknown clusters. What is the rationale behind these differing marker gene selection criteria?
4. Lines 172–216: The authors should refine these paragraphs to highlight the main findings rather than listing excessive details.
5. Line 973: It is unclear why the authors selected multiple resolution values ranging from 0 to 1 to identify clusters.
6. Line 1034: The authors should clarify how they preprocessed the spatial transcriptomics data quantitatively and in detail. Did they filter out low-quality spots?
7. Figures 2, 3, 4, 5, and 6 share a common issue: they contain excessive information, making them overwhelming. The authors should consider moving less critical subfigures to the supplementary materials. The main figures should emphasize distinct and important messages for readers.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #6

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have addressed my previous comments and present a substantially improved manuscript. I am supportive of publication. However, I encourage the authors to further tighten the text. The manuscript still appears to exceed the recommended word limit for this format by more than 50% (approximately 7,600 words for a 5,000-word limit). Condensing it to around 5,000 words would improve its focus, readability, and overall accessibility.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The revised manuscript by Khouri-Farah et al describing a novel dataset derived from single-nucleus RNA-seq analysis of human embryonic face and head development across weeks 4-8 after conception is significantly improved from the previous version. However, several major concerns still remain that should be rectified.

1. Although it has been shortened, the manuscript is still very long at about 7500 words through the Discussion section. The main value of the manuscript is that it describes the overall dataset that can be used as a resource for the community. However, the author's attempt to cover multiple topics – the human data, the mouse data, the human versus mouse data, human disease associations, human facial shape variation, and Neanderthal anatomy – with the data they have obtained. This makes the analysis very broad and it reads as one list after another. Moreover, verification of the data can only be superficial given the limited access to human embryo tissue samples. Therefore, these limits in resolution and in mapping of human data onto human anatomy reduces the utility of using such datasets to study shape variation in a robust manner. The manuscript would be improved by moving some of these sections on anatomy to a different manuscript to shorten and focus the current version.

2. The text of the abstract, introduction and results is still misleading in regards to the tissues used for the analysis, as for example, when the author's state that they have "constructed an atlas of human craniofacial development by collecting craniofacial tissues from 24 individual human embryos" (Section Main, line 61). Although they provide caveats at the very end of the discussion, it would be more accurate to state throughout the manuscript that they have "collected head and neck tissue containing the craniofacial complex and used bioinformatic methods to remove presumptive neural tissue from the analysis." This would provide readers with a more accurate picture of the data used throughout the manuscript.

3. The mesenchyme, cluster 0, is the largest cluster they identify in the UMAP analysis. In my previous review I had highlighted that the mesenchyme of the head is derived from two major tissue-types, the mesoderm and the neural crest. Therefore, it would be expected that a very large fraction of cluster 0 actually represents neural crest derivatives. For example, one of the markers they use for cluster 0 is ALX1. Expression of this gene is known to be strongly associated with the craniofacial neural crest in human, mouse, and zebrafish (e.g. Pini et al. EMBO Mol Med 2020 "ALX1-related frontonasal dysplasia results from defective neural crest cell development and migration" and Iyannar et al Front Cell Dev Bio 2022. "Alx1 Deficient Mice Recapitulate Craniofacial Phenotype and Reveal Developmental Basis of ALX1-Related Frontonasal Dysplasia").

My interpretation of the author's presentation is that only clusters 5 and 7 are neural crest. Perhaps the author's postulate that clusters 5 and 7 are migratory neural crest cells and can be viewed differently than other neural crest derivatives. However, as cluster 5 is associated with cell types differentiating towards Schwann cells and melanocytes, they would appear to be no different than neural crest cells in cluster 0 that are differentiating towards skeletal elements. Therefore, it is imperative that the authors directly discuss the contribution of neural crest to cluster 0. I would suggest that altering the following statement on line 100 would be one of the simplest ways to accomplish this - although there are others:

"ALX1 was most strongly expressed in cluster 0 (mesenchyme),"

changed to

"ALX1 was most strongly expressed in cluster 0 (mesenchyme), a tissue -type which has a dual origin from the mesoderm and cranial neural crest during facial development"

I would then also recommend changing:

"We inferred RNA velocity and transition states to identify trajectories in the data across the newly identified main cell types. We found that the trajectories emanate from the SOX2+/SOX10- population, confirming that these cells likely represent an early progenitor population (Fig. 2c)."

To:

"We inferred RNA velocity and transition states to identify trajectories in the data across the newly identified main cell types.

We found that the trajectories emanate from the SOX2+/SOX10- population, confirming that these cells likely represent an early progenitor population for other neural crest populations in clusters 0 and 5 (Fig. 2c)."

Defining the contribution of the neural crest to the craniofacial mesenchyme also has implications for how the data relating to critical tissues involved in craniofacial disease and shape variation are considered later in the results section.

4. Spatial transcriptomics on two human embryo slide sections is still utilized for verification analysis in several main figures as well as in multiple Supplementary figures. As stated in the previous review the resolution of spatial mapping in the current study was a 55µm spot, whereas an approximately 2µm spot would be required to accurately map a transcript to a single cell layer. Given this issue, it is very difficult to interpret the data shown. I would recommend that arrows are used highlight and label the regions we should be focusing on, or else many of the panels should be dropped if they do not contribute to the analysis so that a reader can focus on the salient points of the study.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The paper by Nagham Khouri-Farah and Alexandra Manche "Gene expression patterns of the developing human face at single-cell resolution reveal cell-type contributions to normal facial variation and disease risk" has been substantially improved after revision and is now stronger, shorter, and clearer.

The following issues still need to be addressed:

1) The authors state that: "Samples were mechanically broken into single-cell suspensions" and that "Craniofacial prominences were collected in a very similar fashion to human samples".

What do the Authors mean by "similar"? Was the neural tissue removed when conducting the mouse craniofacial tissue dissections, as it was done for human dissections? Was there absence of neural contamination? All of these details, which have critical bearing on the reported findings of species-specific differences, have not yet been clearly provided.

The Authors state that they have included additional information about the collection and processing of mouse craniofacial tissue in the Methods Section of the revised version of the manuscript:

"less significant contamination of neuronal cell types was observed in mouse data, which was identified and filtered as described for human."

The point has not been fully addressed; it would help to clearly explain how the dissections were carried out. In the Methods Section it should be precisely specified which mouse craniofacial prominences were dissected and analyzed for the transcriptomic analysis. The text states "craniofacial prominences". Was the whole mouse embryonic head/face homogenized, or which facial prominences were dissected?

2) The Authors compare their datasets to those reported by the group of Trevor Williams (Li, H., Jones, K.L., Hooper, J.E. & Williams, T. The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. Development 146, 2019).

But the latter datasets were generated only from murine midfacial prominences and only at E11.5.

This substantial difference in the datasets should at least be mentioned.

The Authors state: "We have now included this additional detail in the Results section of the revised version of the manuscript."

It does not appear that this information was included or cited in the revised version of the document. It should be made clear where this information was added within the Results section, as this reviewer still cannot find it.

3) The revised manuscript still needs additional clean-up. For instance, capital letters are used to label the figure panels, but then lowercase letters are used in the text to refer to the figure panels. Also, format should be consistent throughout the text. For instance, in the Figure Legends either the Authors add a dot or do not add a dot after the parenthesis when they list the figure panels. To give an example: some figure legends appear like this:

A). B). C) D). E)

If the Authors cannot spend the time, or do not care to fix these details, they should ask for the help of a capable editor. Otherwise, the paper will still give the overall impression of sloppiness.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

The authors have resolved all of my questions. I recommend accepting this paper.

(Remarks on code availability)

Reviewer #5

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Reviewer #6

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

(Remarks on code availability)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The second revision of the manuscript by Khouri-Farah et al has made a few modest changes compared to the previous version. Most of the other comments from the three remaining reviewers who had recommendations for alterations have instead only been considered in the rebuttal letter. Because of this, I believe that a review of the new manuscript is not warranted. Instead, I would suggest 3 things moving forward.

1.

The paper is too long. I find it hard to read and pull out the really important details.

This is also a point raised by reviewer 1 again:

“The manuscript still appears to exceed the recommended word limit for this format by more than 50% (approximately 7,600 words for a 5,000- word limit). Condensing it to around 5,000 words would improve its focus, readability, and overall accessibility.”

Response:

“We will work with the journal to further shorten, if necessary, but not all reviewers agree on this point.”

My reading of this comment is that they are not interested in the valid recommendations of the reviewers and want the paper to be accepted without further work. Asking the journal to adjust the length by a third – which would require alterations to the scientific content – does not seem appropriate and this responsibility should instead fall upon the authors themselves.

2.

I still have no clear idea of what human and mouse tissues were used for the analyses.

This is also a major concern for reviewer #3:

“What do the Authors mean by “similar”? Was the neural tissue removed when conducting the mouse craniofacial tissue dissections, as it was done for human dissections? Was there absence of neural contamination? All of these details, which have critical bearing on the on the reported findings of species-specific differences, have not yet been clearly provided. The Authors state that they have included additional information about the collection and processing of mouse craniofacial tissue in the Methods Section of the revised version of the manuscript: “less significant contamination of neuronal cell types was observed in mouse data, which was identified and filtered as described for human.” The point has not been fully addressed;

it would help to clearly explain how the dissections were carried out. In the Methods Section it should be precisely specified which mouse craniofacial prominences were dissected and analyzed for the transcriptomic analysis. The text states "craniofacial prominences". Was the whole mouse embryonic head/face homogenized, or which facial prominences were dissected?"

I am also unclear on this point. In the rebuttal, they state:

Authors rebuttal:

"This level of description of tissue from both human and mouse has been used by multiple groups outside of our own, including several published in this journal recently (e.g. Rajderkar et al 2024)."

Examination of the Rajderkar et al paper in Nature Communications shows that it was accompanied by a schematic of what was dissected in a Supplemental figure, which is helpful in knowing what they at least tried to obtain.

The authors of the submitted manuscript also state that the neck isn't present at the stages they are taking. But this answer avoids having to address why thyroid tissue is found in their samples if they are only taking the facial prominences.

Authors rebuttal:

"Thus, including more specific claims about what tissue was obtained, which we also cannot state with certainty, would only further add to confusion."

Clearly, two of the reviewers are already totally confused what has been used for the dissections. In this instance, my understanding of the responsibility of authors and reviewers is that the manuscript should provide sufficient detail that the studies can be replicated by an independent party. This is certainly not the case here.

My suggestion is that a schematic is provided to indicate the regions that they were theoretically dissected from both human and mouse samples. If they are uncomfortable providing such schematic then it indicates that the variation in the tissue obtained makes many the claims and conclusions they include in the manuscript less solid.

3.

Neural crest cells.

We probably agree here, but there is still some confusion and I think the authors need to be very specific with the language that is used. My original comment to the authors was:

"Defining the contribution of the neural crest to the craniofacial mesenchyme also has implications for how the data relating to critical tissues involved in craniofacial disease and shape variation are considered later in the results section."

Authors rebuttal:

"Thus, we are confident that the mesenchyme subclusters might be derived from neural crest but have transitioned to a much more differentiated state."

Authors discussion:

"Among these, the endothelium, mesenchyme, and erythrocyte clusters were the most functionally shared between human and mouse, with the CNCC cluster exhibiting the least conservation. This could reflect substantial functional differences in CNCCs across species indicating that this cell type may be particularly labile across evolution allowing innovation of craniofacial shape as others have proposed."

Specifically, it is really important that the definition of CNCCs is stated very clearly when considering things such as shape change and cleft databases. I would argue that the current authors are only using a limited subset of CNCCs for their analyses in this regard. Many other people in the field would use migratory CNCCs and their derivatives in the mesenchyme as the CNCC population.

So, my recommendation is that the authors are very specific on what they are using in these latter parts of the paper: a subset of neural crest that excludes those that are differentiating into mesenchymal derivatives (the latter will be a very significant percentage).

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The Authors have tried to improve the manuscript to the best of their abilities.

This Reviewer understands that working with human embryos is challenging and that certain issues (e.g. use of perfectly reproducible dissections of intact tissues) cannot be completely resolved.

This Reviewer also counts on the Editors of Nature Communications to fix some of the typos, awkward sentences, and formatting issues that are still present in the revised text and figures - and to polish the manuscript.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

REVIEWER COMMENTS

We thank the reviewers for their thoughtful commentary and feedback on our manuscript and have diligently considered these suggestions to improve our work. Specifically, we have considerably reduced the length of the text to refine the scope of the manuscript, improve readability, and strengthen the rationale. We have also included more detailed explanations of the anatomy and region dissected in both the human and mouse samples that were used for snRNA-seq data collection. Furthermore, we have reanalyzed the snRNA-seq data using previously presented as well as newly suggested methods. To maximize the number of nuclei for downstream analysis, we performed filtering on raw filtered barcodes from the single nucleus multiome experiment, resulting in an increase of over 44,000 nuclei (130,427 in new analysis vs. 86,359 in previous analysis). We also removed neuronal and radial glia populations using a new systematic method of generating module scores, resulting in the resolution of roughly 54,000 additional nuclei (95,925 in new analysis vs. 42,131 in previous analysis). When subclustering the putative CNCC's, we labeled plausible SOX10+ Schwann using a similar module scoring methodology as described for neuronal and radial glia cell types. We have also streamlined the comparisons with mouse craniofacial samples and focused on conservation of cell subtype annotations on a large atlas of mouse spatial transcriptomic data. Based on these new analyses, we have updated all main and supplemental figures to reflect these changes in addition to simplifying main figures to emphasize critical points and improve interpretability for the reader. We believe these substantial additions have improved the manuscript and addressed the reviewers' concerns. We have also addressed each of the reviewers' specific comments below (in blue). We also provide a version of the manuscript that highlights all tracked changes in this revision compared to the original submission. Lastly, we have substantially improved our online application that accompanies the paper to include the new clustering as well as spatial transcriptomics data. We hope the reviewers and editor find these revisions to be more than sufficient for publication.

Reviewer #1 (Remarks to the Author):

The manuscript "Gene expression patterns of the developing human face at single cell resolution reveal cell type contributions to normal facial variation and disease risk" by Khouri-Farah et al. describes a new resource providing high-resolution single-cell transcriptome data for multiple stages of human embryonic development. Importantly, these stages cover critical morphogenetic processes relevant to human facial variation and craniofacial birth defects.

There is currently a sparsity of high-quality data from actual human tissues covering these processes. Considering differences in craniofacial development and regulatory sequences between human and mouse (a model for which various datasets are available), the data provided in this study will be a rich resource for craniofacial researchers interested in studying general developmental processes, as well as the impact of human genetic variation on these processes in the context of birth defects.

The authors provide an easily usable interface for researchers to explore the gene expression profiles in specific cell-type clusters, which will facilitate the utilization of their data by others.

While the data provided through this study are without doubt valuable and deserve to be adequately described in a publication, the manuscript itself has significant room for improvement. The text is much longer and detailed than needed and many of the figures are convoluted and contain unnecessary detail. Also, the rationale for some of the analyses performed in this study is not clearly described.

Overall, the generation of human single-cell RNA sequencing data across multiple timepoints is an important addition to the craniofacial field, but the current manuscript would require major revisions to make this valuable resource easily accessible and provide the advance and insight that would make it a strong candidate for Nature Communications.

We appreciate the reviewer recognizing the value of the datasets to the craniofacial community. We understand the reviewers concerns and have made significant changes to the manuscript to reduce the length and increase the clarity of the text as well as add additional explanation to the rationale of the analyses performed. We have also included a paragraph in the discussion explicitly laying out the limitations of the human snRNA-seq data and how we have addressed these caveats. We believe these major changes have significantly strengthened the manuscript and hope this addresses all the weakness and suggestions indicated below.

Major points

1. All figures need major revisions so that they can be easily interpreted. All plots, text, legends, axis labels need to be readable. The authors should think carefully about what is the key takeaway from each figure and which plots are absolutely required to make that point. Keeping a consistent color theme throughout would also improve the accessibility of the figures.

We have made extensive updates to narrow the focus of the figures and to enhance their conciseness, clarity, and accessibility. This includes making the plots, text, legends, and axis labels readable as well as removing extraneous plots that do not add to the key takeaway of the figure. We have condensed the mouse figure substantially and removed the human-specific analyses.

2. The main text is far too long and needs to be at least 50% reduced. The readability of this manuscript also needs improvement. There are many different analyses presented but the rationale for their inclusion and key conclusions for the reader are not always clear. I would recommend adding a Supplementary Information section where more details can be added to improve the structure of the main text and make it easier to follow.

We have significantly reduced the length of the main text. We have taken the reviewer's suggestion and included a Supplemental Information section to add additional methods and results that are not essential for the main text.

3. The human single-cell data from this study was produced using 10X Multiome but only the RNA-seq component of the data is presented here. The ATAC component of the data is not needed for the purposes of this manuscript (which focuses on transcriptomes) but it would be good to mention for the reader why it is not reported. Was it of insufficient quality to provide useful results? Will it be described in a complementary manuscript? Some other reason?

We understand the reviewer's concerns. The ATAC data was of sufficient quality but will be discussed in a complementary manuscript that is focused on gene regulatory networks. As the paper was previously written we felt including chromatin accessibility would be too much information for a single paper. Indeed, all reviewers wanted a shorter manuscript, and we could not provide robust analysis of ATAC data in this condensed format. We have a brief statement in the methods mentioning these details (lines 644-647).

4. It's surprising that Sox10 does not come up in any of the human CNCC subtype clusters, considering that it is a key marker in the mouse CNCC and given its known role in cranial neural cell development and differentiation. Why do the authors think that is? Given the overlap in markers with neural cells, is there any concern that human nuclei of CNCC were accidentally filtered out when removing neuronal cell types? E.g., TUBB3 is expressed in premigratory cranial neural crest cells in birds and maintained in neural crest derived tissues/cell types. See: doi: 10.1016/j.jep.2019.119067.

We apologize that this was unclear in the original version of the manuscript, but the main CNCC cluster we originally described is indeed labeled by *SOX10*. Thus, it being a marker gene for the population would preclude it as a marker from subclusters derived from that cluster. Indeed, we showed in the previous manuscript and the new analysis here that *SOX10* is a marker gene for the *SOX2-/SOX10+* putative CNCC population (lines 116-117) as we have shown in Fig. 2a. However, *SOX10* again is not a marker gene for any of the 19 putative CNCC subclusters but is generally expressed in all of them (Figure 3c). Instead, many other genes are markers of the subclusters and are co-expressed with *SOX10*.

The reviewer raises an excellent point that we could be removing cells that are biologically important with the arbitrary choice of TUBB3. To address this concern, we have adopted a new method of systemically removing neuronal and radial glia cells that is not solely dependent on *TUBB3* and *MAP2* expression. We used the neuronal and radial glia marker genes from Eze et al. 2021 (doi:10.1038/s41593-020-00794-1) to generate module scores and removed the clusters with scores significantly above the median of all clusters. There was no concern that putative CNCC's were filtered out as this population was neither enriched for the neuronal nor the radial glia module score. Additionally, the

putative CNCC population is marked by significant *SOX10* expression, and the neuronal and radial glia populations were not *SOX10*⁺. We detail this approach in the Supplemental Note to keep the results section more compact.

Minor comments

1. The PDF files are very large (causing poor display/printing issues on some systems) because dot plots are pasted in vector format with large numbers of individual datapoints. It would be helpful to rasterize plots so that the Figures can render more quickly and make it easier to review.

The figures have been rasterized in the updated PDF files, making them easier to view.

2. The sentence in the introduction about Neanderthal-introgressed sequences seems a little out of place. Perhaps move this up to the section about normal facial variation and give some clearer rationale for why this analysis was done/included.

We have decided to remove this sentence from the introduction and have included additional detail in the results section to give clearer rationale for the Neanderthal-introgressed sequence analyses.

3. Typo on 105 - doublets.

We have corrected this typo.

4. Line 113: Which clusters were problematic? It's not easy to tell from S1b as the colors merge together.

We have changed this figure in the updated version of the manuscript to color the bar plot by stage instead of by sample, making it more readable (Fig. S1b).

5. Line 117: Please specify that nuclei are filtered specifically based on TUBB3 and MAP2 expression and not the other genes shown in S1.

We have changed our method of removing neuronal and radial glia clusters as described above.

6. Figure 2B: Perhaps only show top 5 gene markers per cluster so that the y-axis is more easily read.

We have updated Fig. 2b to include only the top 5 marker genes for each cluster.

7. Figure 2C - it's not possible to read the terms in the smaller font. Simplify to top 3-5 enriched terms per cluster that can all be shown and read.

We have updated this figure (now Fig. 2d) to include only the top 5 functional enrichment terms, making each term visible.

8. The Venn Diagram in FigureS4B may not be necessary. It's not easy to interpret and the upset plot clearly shows the intersections.

We have decided to keep this figure as an additional visual depiction of the results.

9. Line 136-138: This statement is somewhat unclear as at this point in the manuscript no enrichments for craniofacial biology and developmental abnormalities has been shown. I think the conclusion of this paragraph can be kept as simply "Overall, these results suggest that our single-nucleus expressions data closely resembles the bulk gene expression data."

We agree with the reviewer and have removed this sentence from the main text since we already make the point that single-nucleus expressions data closely resembles the bulk gene expression data. Specifically in lines 90-94, we state:

"Principal component and differential expression analyses of these pseudo-bulk profiles revealed very similar results to those previously obtained from bulk tissues¹³. Specifically, samples were ordered by developmental stage along the first principal component (Fig. 1b, Fig. S4a) and the greatest number of differentially expressed genes were observed between the earliest and latest comparable timepoints (CS13 vs CS17) (Fig. S4b,c)."

10. Reword sentence at 154-155. It's not necessary to state "some of the timepoints". I suggest the authors rephrase so that it's clear that at CS12 all samples were derived from female embryos and at CS13 all samples were derived from male embryos, but that this did not introduce any significant bias.

We have now included a paragraph in the Discussion section explaining this limitation. Specifically, in lines 559-564, we state:

"First our two earliest timepoints, CS12 and CS13, are made up of only female or male samples, respectively. While we did not observe any sex related bias in any of our main cell type clusters, some subtypes were made up of only one sex. Thus, some of the differences we observed in overall conservation of main cell types, particularly for CNCCs, could be influenced by this bias. However, most of the subtypes had a clear counterpart in mouse subclustering that were equally mixed for male and female samples."

11. Typo Line 155 - main cluster -> main clusters.

We have updated the text, changing "main cluster" to "main clusters".

12. Figure S5A - y-axis text is overlapping and not readable.

We have updated the figure (now Fig. S6), which is now easily readable.

13. Line 258 sub-cluster cnl2 is not described - what is this cluster?

The subclustering of the putative CNCC population has changed in the updated version. The function of the neural crest-like (ncl) populations have been described in lines 176-181 – “The ncl1 cluster was enriched for several categories shared with early CNCCs and Schwann cells suggesting this was an early neural crest type with possible downstream fate commitment towards a Schwann cell phenotype. The ncl2 cluster was enriched for processes overlapping with the mesenchymal committed CNCC’s including hair follicle development and skin development (Fig. S9, Supplemental Table 3). The ncl3 cluster was enriched for synaptic signaling, while the ncl4 was enriched for gas transport processes.”

14. Line 271: clusters cnl1, cln2 and cnl3 need descriptions.

We have addressed this in comment #13.

15. Figure 3C: Do the different colors represent the Carnegie stages in this plot? Can this plot be simplified to highlight the genes referred to in the text? Or perhaps a heatmap would be easier to interpret.

The genes shown in Fig. 3c correspond to known CNCC gene markers and the colors differ according to the gene that is plotted. We have decided to keep the stacked violin plot to show the variation in gene expression for each of the putative CNCC subcluster.

16. Figure 3D: Challenging to read the y-axis.

We have updated Fig. 3d to make the y-axis more readable.

17. Figure 3E: I suggest changing the color scale for these plots so that cream or a light color represents lowly expressed genes and a darker color (e.g., red/blue) represents highly expressed genes

We thank the reviewer for their suggestion. We have changed the color scales for all Dot Plots showing transcription factor marker gene expression such that a light color represents lowly expressed gene and a darker color represents highly expressed genes. However, we believe that the blue-yellow-red color scale for the spatial transcriptomics plot strikes a better visual difference between high and low expression, thereby highlighting the regions with higher expression.

18. Line 300: Muscle cells clusters have been observed in the single-cell data

from <https://doi.org/10.1038/s41467-024-46396-4> and <https://doi.org/10.1038/s41467-023-41349-9>

We apologize about not mentioning these papers in the original manuscript. We are aware of these papers and have cited them in the revised version of the manuscript. However, the description of major cell types has been reduced and instead we focus on the *SOX2*-/ *SOX10*+ putative cranial neural crest cells. We do not make any statements about the presence or lack thereof in these previous mouse papers in the revised manuscript.

19. Line 301: It seems unlikely that sampling differences and/or timepoints would be the reason these cells are not observed but I would imagine it's due to the lack of specificity in resolved clusters. E.g., *Sox10*, one the main markers used by the authors for specifying CNCC cells, is also a specific marker for neural-crest derived Schwann cells. Is it also possible that CNCCs had not been resolved in preexisting single-cell datasets because they were instead labelled broadly as neural cell types?

We agree with the reviewer that the pre-existing single-cell datasets may have labeled these cell clusters more broadly as neural cell types. However, there has not been in depth analysis performed on these clusters as we have conducted in the revised version of the manuscript. More specifically, we have re-annotated the putative CNCC subtypes in the following ways as stated in lines 150-164 of the revised version of the manuscript:

“We analyzed the genes enriched in each of these subclusters to label based on functionality. For instance, CNCC cluster 16 was enriched for melanocyte-specific gene markers *MITF* and *PMEL* (Fig S8d), while two other clusters (clusters 10 and 14) were significantly enriched for the G2M stage of the cell cycle and labeled as cycling (Fig S8e). Since *SOX10* is also a known marker for neural crest derived Schwann cells⁶⁹, we generated a Schwann cell module score using marker genes from human fetal development²⁴. We identified five clusters that were enriched for the Schwann cell module score (Fig. S8f) and were biased towards the later time points profiled (Fig S8c). To identify additional subclusters we generated additional module scores from markers of E9.5 mouse neural crest⁷⁰ labelled by *Wnt1*^{Cre}/*R26R*^{Tomato}, which consist of various functionally specific, fate committed CNCC subtypes. Using this methodology, we identified one cluster of mesenchymal committed CNCCs and one cluster of sensory committed CNCCs. We also identified four more punctate clusters, which we annotated as neural crest like (ncl). The resulting subclusters were not enriched for any of the previously computed module scores and were labeled according to stage-specific bias, i.e., early (early.CNCC) vs. intermediate (int.CNCC) (Fig. 3B, Fig. S8c).”

20. What is the rationale for selecting E10.5-E12.5 in mouse for resolving muscle cells?

We are not certain exactly what the reviewer is referring to. We use all data from E10.5 to E15.5 in one day increments to resolve all cell types. These stages

encompass all of the same morphological stages of the human embryos we profiled (CS12 to CS20).

21. Figure 4 needs to be simplified. It has too many panels and it is difficult to understand what the key takeaway is meant to be. What are the absolute key plots to show? And can the rest be put in supplements? For example, Figure 4K and Figure 4N are not readable.

We agree with the reviewer and have simplified Fig. 4 to display the key findings.

22. Figure 4G: Please specify what the subtype labels represent.

We refer the reviewer to the text above in which we describe our new method of annotating the putative CNCC subtypes.

23. When comparing single-cell data between species the authors could consider joint-embedding with SAMap (<https://doi.org/10.7554/eLife.66747>) to identify homologous cell types.

We thank the reviewer for this suggestion and have now incorporated the SAMap cell annotation method into our scheme of cell subtype annotation for mesenchyme, ectoderm, and CNCC populations.

24. For the cross-species cell type comparisons, what gene list was used as background for the enrichment analyses? Is it all human/mouse genes or genes that have 1-to-1 orthologs between mouse and human?

The background gene list is all 1-to-1 orthologs between human and mouse that are expressed in the dataset. This is described in the methods section on lines 725-726.

25. Figure 7: It is difficult to understand the main takeaway from these plots. The heatmap legend is not labelled and the values are not easily interpreted. Perhaps parring these heatmaps down to only those that are significant for the main figure could help. The x-axes are not readable so it's impossible to know which sub-cell types have specific variant enrichments. In Figure 7C, do the different colors represent the different human stages?

We have simplified Fig. 7 and have now properly labeled all legends and axes, making it easier to interpret. We have also reduced Fig 7a to include only the Xiong et al. 2019 craniofacial landmark GWAS data. The Bonfante et al profile landmark GWAS data has been moved to Supplemental Figure 31.

26. Line 621: What is this referring to? The HARs or overall for all gene lists?

This is referring to the gene list labeled “Random” in Fig. 8a, which we have clarified in the text on lines 847-848: “As an additional negative control, we generated random gene lists based on the mean number of all the lists tested here.”.

27. Line 713: From Supplementary Table 7, Sox10 does not appear in the CNCC subtypes for human data.

We understand the reviewer’s concern. However, *SOX10* is a marker gene for the CNCC main cluster but not a marker gene for any of the CNCC subclusters as it is expressed generally in all of them.

Reviewer #1 (Remarks on code availability):

I did not review the code in detail or attempt to install/run the code.

Reviewer #2 (Remarks to the Author):

The manuscript by Khouri-Farah et al describes a novel dataset derived from single-nucleus RNA-seq analysis of human embryonic face and head development across weeks 4-8 after conception. They define several broad classes of cell types, including epithelial and mesenchymal, and a population defined as neural crest, and delve deeper into the cell identities by sub-clustering. Further, they perform a similar analysis using approximately equivalent stages of mouse embryogenesis and compare and contrast the data obtained between the two species. The authors next used a published human spatial transcriptomics dataset as a means to validate several of their findings concerning their human assignments. The human single-nucleus RNA-seq data are also used in conjunction with several public databases to make predictions concerning disease associations, human facial shape variation, connections between human and Neanderthal facial morphology, and clefting genetics. Overall, the dataset is a valuable additional resource for people interested in human embryonic head development that adds to other bulk, single-cell, and spatial datasets focused on the entire embryo. However, there are several significant issues with the manuscript, particularly with the analysis and interpretation of the datasets, that considerably reduce its impact and utility.

We thank the reviewer for recognizing the importance of the data generated and the extensive analysis we originally performed. We have made significant changes to the manuscript based on the reviewer’s concerns. We hope this addresses all the weakness and suggestions indicated below.

Major issues

1. The manuscript is too long, and there are many sections that are redundant or repetitive. The journal’s instructions for authors state that ideally the main text should be

5,000 words. This manuscript is well over 10,000 words, with the discussion alone about 3,000 words. Much of the data from the results section is reiterated in the discussion and there are paragraphs dealing with disease correlation analyses of red blood cells and immune cells that are not pertinent for an analysis focused on the craniofacial region. For example, the paragraph beginning on line 191 concerning ontologies associated with red blood cells including malaria or hemolytic anemia or the paragraph beginning on line 203 concerning immune cell related diseases. The length of the manuscript and its multiple individual correlations with other datasets means that the focus of the paper is lost, and the key points are difficult to extract from the study. Overall, the manuscript would be more impactful if it was considerably shortened and focused on key findings concerning the craniofacial region.

We have made extensive updates to shorten the length and narrow the focus of main text. This includes significantly reducing the length of the results by removing unnecessary results related to erythrocytes and immune cells and removing parts of the discussion that reiterate text in the results.

2. A major flaw in the manuscript is a lack of clarity concerning the regions under analysis in the sn-RNA-seq study. For human, only the following statements are provided: Results “we profiled the entire craniofacial prominence” and Methods “Primary human craniofacial tissues were obtained”. For mouse: Results “we generated single-nucleus gene expression from mouse craniofacial tissue. These samples reflected the major morphological landmarks of the human tissue profiled allowing a more direct comparison of cell types”. Methods “primary mouse craniofacial tissues ... were pooled”. Unfortunately, these extremely limited descriptions do not provide sufficient information to understand or review the manuscript with any degree of certainty. One illustration of how this lack of detail affects the manuscript is that – as illustrated in Figure 6 - the craniofacial human ectoderm samples contain tissues related to the optic and otic placodes as well as to the neck (thyroid). These findings indicate a much greater region of the head and neck were surveyed than just the craniofacial prominences shown in Figure 1 and it is imperative to know the morphological boundaries that were used for the dissections.

We appreciate the reviewer’s concern about the precision of the dissection. The fact of the matter is that these samples are not dissected as one would do for an intact mouse embryo. We receive samples that have already been separated from the rest of the embryo and do not have all the additional anatomical landmarks. As we are sure the reviewer can appreciate that human embryonic samples are rare and extremely precious. HDBR does not provide whole embryos due to the method of termination, and projects must specify specific anatomical structures of tissues to be collected. We do our best to remove regions that are clearly non-facial but ultimately this is impossible. We thus must use computational methods to remove these extraneous tissues/cell types. While we can remove many of the contributions of brain tissue, other ectodermal types are not so easily extracted. Our analysis readily identifies them based on marker genes and ontological enrichments. Thus, we take advantage of this finding and utilize these additional

types as excellent controls for many of our analyses. The mouse tissues are much more precisely dissected and do not contain many of these additional subtypes.

As another example, on line 116, describing clusters expressing *TUBB3* and *MAP2*: “We reasoned these clusters were derived from developing brain tissue not directly part of the craniofacial structures”. In removing these “brain clusters” their sample size decreased from 86,359 nuclei to 42,131, a drop of over half. These data indicate that their dissections of craniofacial tissue were not precise but included a majority of cells from other regions of the head. On the other hand, throwing away cells with expression of these two markers likely means they are losing cells from critical craniofacial structures, particularly the trigeminal ganglia. The authors indicate that the mouse data were obtained and processed in a similar manner, but it would be valuable to know if it required excluding *Tubb3* and *Map2* cells, and if so what percentage of mouse cells were removed from the analysis to help determine equivalence of dissections.

In the revised manuscript, we have adopted a new method of systemically removing neuronal and radial glia cells that is not solely dependent on *TUBB3* and *MAP2* expression. We used the neuronal and radial glia marker genes from Eze et al. 2021 (doi:10.1038/s41593-020-00794-1) to generate module scores and removed the clusters with scores above the median (Methods and Supplemental Note). We obtained 130,427 nuclei after filtering for doublets and nuclei quality, which was then reduced to 95,892 nuclei after removing neuronal and radial glia clusters.

Additionally, we have now included a description of the limitations of the human scRNA-seq data in lines 564-571:

“Second, the nature of the human tissue we received precluded precise dissection, as we can perform in mouse guided by anatomical landmarks of the entire embryo. We attempted to identify and remove extraneous cell types from the human samples using module score-based approaches. However, these could have resulted in over or under filtering of cell types. Indeed, we were able to identify subclusters that clearly reflected thyroid and pituitary gland progenitors. However, inclusion of these subclusters as well as immune and erythrocyte clusters served as excellent negative controls demonstrating the specificity of our analyses.”

These considerations are especially important when the authors discuss the idea that there are multiple marker genes that were only expressed in a single species (Results: Section titled “Species-specific differences in marker gene expression during human and mouse craniofacial development”). Examination of the data provided in Table 20, for example, states that genes including *ALX3*, *DLX2*, *DLX3*, *DLX5*, and *DLX6* are only detected as expressed in the human facial mesenchyme and not in the mouse “Only called in a single species”. This is certainly not supported by multiple other studies,

raising concerns about the scope of the dissections, the data processing, and attention to detail in data analysis.

We apologize for being imprecise with our language in the original manuscript. We did not want to imply that these genes are not expressed in mouse, simply they are not identified as marker genes for the same cluster. They are expressed robustly in mouse, but not in a cluster specific way that passes the marker gene thresholds we set. To address this more thoroughly we perform differential gene expression and the SAMap procedure to identify the same cell types. In an attempt to shorten the length and narrow the focus of the revised manuscript, we have decided to remove the results section where we had previously discussed the species-specific differences in marker gene expression during human and mouse craniofacial development.

3. The question of anatomy and the regions dissected also has a direct bearing on the parts of the paper discussing the identification and analysis of clusters 1 and 6, called as mesenchyme and cranial neural crest cells, respectively. The mesenchyme of the head is derived from two major tissue-types, the mesoderm and the neural crest. Fate mapping studies have indicated that the majority of the mesenchyme in the craniofacial prominences is of neural crest origin, whereas mesodermal derived mesenchyme occupies areas of the head that are above and behind the developing eye. There is also a core of mesoderm tissue in the arches, including arch 1, that will form muscles of the head and neck. Therefore, it would be expected that a significant portion of the mesenchyme in cluster 1 would have a neural crest origin. However, the ratios of neural crest versus mesoderm derived mesenchyme will vary depending on the amount of tissue included from more posterior regions of the head. This consideration again makes it imperative to understand the boundaries of the dissections performed.

Even if a large amount of mesodermally-derived mesenchyme was present in cluster 1, there should still be a significant contribution of neural crest to this cluster. In this respect, when the authors subcluster the mesenchyme, they identify maxilla, mandibular and nasal prominence cells indicating the presence of neural crest derivatives. Therefore, cluster 6 should not be addressed as the only cranial neural crest cell population in their datasets. In fact, examination of cluster 6 suggests it may represent a few specialized neural crest derivatives – particularly Schwann cells as well as a few melanocytes. For example, as shown in Figure 3C, PLP1 - a prominent component of myelin - is expressed in all subclusters of cluster 6. Similarly, Mpz (myelin protein z) is broadly expressed in most subclusters. Together, these observations suggest that the mesenchyme has been mis-classified as not containing neural crest, while cluster 6 has been annotated as the major cranial neural crest population, whereas it represents highly specific populations of neural crest derivatives.

We thank the reviewer for their comment and agree that cluster 5 (previously cluster 6 in the initial version of the manuscript) may not be the only cranial neural crest-derived population. Furthermore, cluster 0 (previously cluster 1 in the initial version of the manuscript) may still contain a significant amount of

neural crest derivatives. To account for this point, we have now labeled the previous CNCC cluster as “SOX2-/SOX10+”, as this population not only contains a highly specific population of CNCC’s but also contains plausible Schwann cells. We have decided to keep the name of cluster 0 as “mesenchyme” despite some nuclei in this population potentially being derived from neural crest as it contains more terminally differentiated, less transient mesenchymal cells. We have identified this using RNA velocity and transition state analyses, which show emanation of trajectories from the SOX2+ early progenitor population (Fig. 2c). We will also point out that again previous efforts that have relied on digestion of mouse tissues have generally not been able to identify a robust neural crest population. This could suggest that these are particularly sensitive cells that are easily lost in preparation of samples. Indeed, a rich body of literature points to neurocristopathies being due to this very sensitive cell type. This is beyond the scope of the paper.

4. The authors define many subclusters in the mesenchyme, ectoderm, and neural crest but do not adequately describe how they were assigned to specific regions or structures. For example, for the ectoderm, TBX18 was assigned as a fusion zone marker, but it is not clear why, nor is it clear what fusion zone they are discussing anatomically. These issues of mapping and assignment are compounded by an absence of appropriate verification analyses. Here the authors rely on published data of spatial transcriptomics of two sagittal sections through a CS13 human embryo to overlay their expression data which are then presented in several figures throughout the manuscript. There are two problems with this approach. The first is that they are relying on just two sections and so it is not possible to determine the 3-dimensional expression patterns of any gene or cluster, meaning that considerable information is lost regarding assignment and anatomy. The second issue is that the resolution of spatial mapping was a 55µm spot, whereas an approximately 2µm spot would be required to accurately map a transcript to a single cell layer. Therefore, the study fails to link subcluster identity with specific anatomy at a meaningful resolution and in the absence of any other verification studies this significantly reduces the power of the dataset.

While we agree with the reviewer that it is not optimal to use only two sections of a single human embryo, unfortunately that is all that currently exists in the literature. As we detailed in our response about dissection above, we do not receive intact human embryos and thus are currently unable to generate additional high quality spatial transcriptomics data. We therefore rely heavily on what is known in mouse and for which we can identify functionally conserved linkage in humans. Cell subtype annotations for mesenchyme and ectoderm are defined at a much higher resolution in mouse based on multiple recent single cell and spatial atlases. However, the SOX10+ putative CNCC and early progenitor subtypes are less clearly defined. Therefore, the mouse ectoderm and mesenchyme subtypes were defined first. Subtypes were assigned to specific regions and structures according to the following publicly available resources:

13. Yankee, T.N. et al. Integrative analysis of transcriptome dynamics during human craniofacial development identifies candidate disease genes. *Nat Commun* 14, 4623 (2023).
14. Rajderkar, S.S. et al. Dynamic enhancer landscapes in human craniofacial development. *Nat Commun* 15, 2030 (2024).
15. Sun, J. et al. Single-cell RNA-Seq reveals transcriptional regulatory networks directing the development of mouse maxillary prominence. *J Genet Genomics* 50, 676-687 (2023).
16. Han, X. et al. Runx2-Twist1 interaction coordinates cranial neural crest guidance of soft palate myogenesis. *Elife* 10(2021).

We lay this out in the Methods section on lines 733-737:

“Marker genes were identified for each cluster and functional enrichments were determined using clusterProfiler. Annotations for each cluster were manually assigned based on those originally described¹³⁻¹⁶. Mouse cell subtype assignments were further confirmed with ToppGene¹⁴⁰ using the toppFun function from the scToppR package¹⁴¹.”

Following the mouse ectoderm and mesenchyme subcluster annotation, we annotated the human mesenchymal and ectodermal subtypes. This was conducted in the following manner as described in the Methods section on lines 741-759:

“Following annotation of subtypes and for comparison with human data, an intermediate data set was created where mouse genes were reduced and converted to those to those with one-to-one fig. orthology with human genes using the orthogene R package (v. 1.10.0) as described above. The intermediate dataset was used to transfer mouse subtype annotations to human subtypes by first identifying shared features across clusters using FindTransferAnchors in Seurat with log normalization and canonical correlation analysis (cca) or reciprocal PCA projection (rpca). Predicted subtype labels were transferred to human subtypes using TransferData in Seurat and further confirmed with a confusion matrix. As an orthogonal approach to label transferring in Seurat, we used SAMap⁷² (v. 1.0.15), an algorithm implemented in python that uses self-assembling manifolds to map homologous cell types with shared expression programs across species. Briefly, we used the map_genes function within the SAMap package, which performs a BLAST search between species, with t1 and t2 both set to ‘nucl’, tr1 set to the gencode v32 human fasta file, tr2 set to the the gencode v10 mouse fasta file, n1 set to ‘hu’ for human, and n2 set to ‘mo’ for mouse. This generates a “maps” directory with the transcriptome mapping BLAST results. Next, we used the SAMAP, run, and get_mapping_scores functions within the SAMap package to calculate alignment scores between cell types and subtypes across species. To finalize cell subtype annotations, we used a combinatorial approach considering the three annotation methods (label transfer by Seurat using cca, label transfer by Seurat using rpca, and label mapping using SAMap). Final annotations were confirmed using contingency matrices, significant marker gene overlap between

human and mouse subtypes using the GeneOverlap R package¹⁵⁸, and functional enrichments using the compareCluster function in the clusterProfiler R package¹⁶².”

The *SOX10*+ putative CNCC and early progenitor clusters were annotated in the following ways as described in the Methods section on lines 761-773:

“The putative CNCC and early progenitor clusters were annotated by generating functional and/or regional module scores using marker genes from various publicly available databases and resources. To label the subclusters of the putative CNCC population, we generated a Schwann cell module score using the marker genes from Cao et al. 2020. Additionally, we used the gene markers from various functionally specific, fate committed CNCCs in Soldatov et al. 2019 and identified one cluster of mesenchymal committed CNCCs and one cluster of sensory committed CNCCs. The remaining putative CNCC subclusters were labeled according to Carnegie stage-specific biases. We randomly downsampled the number of cells to 200 cells per stage and quantified the percentage of cells in each cluster; those that were biased towards CS12 and CS13 were labeled as early (early.CNCC) and those that were biased towards CS13-C17 were labeled as intermediate (int.CNCC). When subclustering the early progenitor population, we generated module scores for neural progenitors using markers genes from the ScType database, neural stem cells using marker genes from Wang et al. 2024, mesenchymal progenitors using markers from Wang et al. 2024 and epithelial progenitors using markers from the human CellMarker2.0 database.”

After annotating the subtypes, marker genes for the human subtypes were identified using FindAllMarkers (logfc.threshold = 0.25, min.pct = 0.1, test.use = “wilcox”, min.cells.feature = 3, min.cells.group = 3, pseudocount.use = 1, and return.thresh = 0.01). This is how we identified specific genes to the various human subtypes. We then used the spatial transcriptomics slices to confirm the localization of the previously labeled subclusters and plot the subtype-specific marker genes and calculated module scores. Therefore, we were not intrinsically using the spatial transcriptomics data for subcluster annotation, but rather we used it for confirming our previously defined regional and/or structural annotations.

Minor issues

1. Line 462-463. “with LNP1 being more posterior and LNP1 being more anterior”. One of those should not be LNP1.

This has been removed in the revised version of the manuscript.

2. Line 625. “We found no enrichment in any cell types...”. This line is duplicated in the next sentence.

We apologize for the oversight and have removed this in the revised version of the manuscript.

3. The authors refer to Li et al (reference 12) throughout the manuscript concerning mouse craniofacial data. However, this reference describes genome analysis of human brain data.

We apologize for this and have updated the references accordingly.

4. I have a number of suggestions for the images of the human sections. First, as the manuscript is focused on the head and face, it would be more useful to show only these structures rather than the whole embryo to reflect the focus of the paper. Second, orientation should be indicated on the panels: left/right (i.e. for Fig 4C), anterior/posterior, rostral/caudal. Third, it would be useful to have one or more annotated images describing the pertinent structures present in the images.

We thank the reviewer for these suggestions and have better annotated the images of the human sections.

5. The figure legends for the Supplemental figures are missing. There was also no explanation of the information provided in the Supplemental tables.

We apologize for this oversight and have now included figure legends and supplemental table legends.

6. Multiple instances: “this data” should be “these data”.

We have made this text change throughout the revised version of the manuscript.

Reviewer #3 (Remarks to the Author):

The paper by Khouri-Farah et al., “Gene expression patterns of the developing human face at single cell resolution reveal cell type contributions to normal facial variation and disease risk” reports a single-nucleus RNA sequencing (snRNA-seq) atlas of human craniofacial development from craniofacial tissues of 24 embryos that comprise six key time points spanning 4–8 post-conception weeks. The Authors describe seven major cell types, including muscle progenitors and cranial neural crest cells (CNCCs), as well as multiple subtypes of ectoderm and mesenchyme. Comparative analyses reveal substantial conservation of major cell types, alongside human biased differences in gene expression programs. The Authors also claim that CNCCs exhibit the lowest marker gene conservation, underscoring their evolutionary plasticity. Spatial transcriptomics is also employed to further localize cell populations, which should provide a detailed view of their anatomical context. Lastly, the Authors attempt to link these developmental processes to genetic variation, identifying cell type-specific enrichments for common variants associated with facial morphology and rare variants

linked to orofacial clefts.

Undoubtedly, this atlas has the potential to offer critical insights into the cellular and genetic mechanisms shaping the human face, highlighting conserved and distinctly human aspects of craniofacial biology and linking them to the developmental origins of craniofacial disorders and the genetic basis of facial variation. This work could certainly provide an important resource for exploring craniofacial biology, with implications for developmental genetics, evolutionary biology, and congenital anomalies. Therefore, the study could be of interest and could fill a poorly explored aspect of human craniofacial biology by reporting studies conducted on human embryos as well as mouse embryos. However, there are multiple serious concerns regarding the current manuscript that need to be fully addressed. I list below some of the most salient issues that need to be resolved in order to increase the rigor and accuracy of the described findings, as well as supporting many of the conclusions.

We thank the reviewer recognizing the value of the datasets as an important resource for the craniofacial biology community. In the revised version of the manuscript and our response to the previous two reviewers we hope we have addressed all of the reviewer's major concerns to increase the rigor of the manuscript.

Single-nucleus RNA Sequencing (snRNA-seq) Data Generation and Processing (Human Samples)

- It would be important to show a cartoon schematizing the regions of the human embryonic heads that were dissected. The Authors state that they processed the “developing craniofacial region” (quote) of human embryos from 4 to 8 weeks post conception. Lacking more detailed information, the reader is left guessing which parts of the developing craniofacial region were processed.

We have included a schematic in Fig. 1a that encompass major milestones of human craniofacial development from 4 to 8 post conception weeks, which were collected and processed for snRNA-seq analysis. As we detail in our response to Reviewer 2 above, human embryos that we receive are not whole. We have done our best to provide detail about the tissues selected but the fact is that the human tissue is messy. Despite this, our systematic, unbiased analyses identify extraneous cell types and remove them when possible. When not possible, the ontological analyses identify them, and they are retained in the analyses to demonstrate specificity of genetic enrichments.

- How did the Authors generate the initial clustering of the major cell populations? Which Seurat version did they use? Which other packages were employed? All relevant information is lacking.

We have now included this information in the Methods section of the revised version of the manuscript in lines 657-690:

“Raw barcode matrices from each human samples generated by CellRanger ARC were utilized for downstream processing to maximize total nuclei. Raw barcode matrices for each sample were individually loaded with Read10X_h5 command in Seurat ¹⁵² (v. 5.2.1). Subsequently, the emptyDrops function within the DropletUtils R package (v. 1.24.0) was used to identify nuclei containing ambient RNA for each sample (emptyDrops FDR < 0.001). Individual sample Seurat objects were then merged into a single object. Percentage of mitochondrial reads were calculated for each cell and filtering was performed to only retain cells with less than ten percent mitochondrial derived. Further filtering was performed based on number of counts per cell ($500 < x < 25000$) and number of genes detected per cell ($500 < x < 7000$), resulting in a total of 130,427 nuclei. Filtered data were normalized with default values and cell cycle scores were calculated using Seurat. Data was scaled based on S and G2M score regression and dimensionality reduction with principal component analysis (PCA) were performed using respective commands in Seurat. The top 2000 variable features were identified, and sample-level batch correction was performed using the harmony R package¹⁵³ (v. 1.2.3). Nearest neighbors based on harmony corrected embeddings were calculated with up to 30 dimensions and clusters were identified with multiple resolutions from 0.1 to 1 in Seurat. Resolution parameters were selected for clustering based on generated elbow plots and identification of the elbow point to determine the optimal number of clusters. We then performed Uniform Manifold Approximation and Projection (UMAP)³⁶ dimensionality reduction using harmony corrected embeddings in Seurat (dimensions = 30, minimum distance = 0.3). Resulting clusters were inspected for expression of multiple craniofacial markers from Li et al 2019¹² and marker genes were identified for each cluster. Module scores were generated using the AddModuleScore function in Seurat for neuronal markers and radial glia markers from Eze et al 2021¹⁵⁴. Cells from clusters identified with high neuronal or radial glia module were removed (See Supplementary Note), resulting in 95,892 nuclei. The process of normalization, harmonization, and clustering was repeated with remaining cells from all samples. We observed that early samples had consistently higher mitochondrial reads (Fig. S2d), however, this was not due to increased apoptosis as tested by generating an apoptosis module score using gene markers from the Hallmark Apoptosis gene set within the human MSigDB Core Collection (Fig. S2d). The sex of each embryo was also identified using known sex determination genes for males (*UTY*, *DDX3Y*, *KDM5D*, and *NLG4Y*) and females (*XIST* and *TSIX*) (Fig. S2e), and each of the main cell type clusters contained roughly the same number of male and female samples (Fig. S2f). Marker genes for major cell types were identified using FindAllMarkers (logfc.threshold = 0.25, min.pct = 0.1, test.use = “wilcox”, min.cells.feature = 3, min.cells.group = 3, pseudocount.use = 1, and return.thresh = 0.01). The top 100 marker genes for each cluster ($p < 0.05$, ranked by log2fold change versus all other clusters) were then analyzed for gene and disease ontology enrichments using compareCluster in clusterProfiler R package (v. 4.12.6). Cranial neural crest genes were compiled based on markers identified by regulatory network construction in human cultured CNCC and craniofacial tissue data⁶⁵. Major cell type labels were applied to each cluster.”

- More details are required in the Methods sections describing how the single-cell data were embedded and clustered, e.g. how many highly variable genes, PC dimensions, and nearest neighbors used to generate the UMAPS appearing in the main figures. Was any batch correction used? Were distinct embeddings calculated for the subset of neural crest clusters?

We describe our method of clustering and dimensionality reduction for the main cell types above in lines 659 to 693 and in the Supplemental Note. For each major cell type, normalization, scaling with regressed cell cycle impacts, harmonization, and subclusters were identified using the same procedure. We provide full code on our GitHub that can be readily inspected (https://github.com/cotneylab/craniofacial_snrna).

- It would also be important to provide separate clusters for each gestational time analyzed. This would highlight differences that are to be expected in cluster transcriptomes across the different gestational days, which are currently not clear.

We have now included UMAPs colored by stage as well as stacked boxplots highlighting the stage-specific bias for the main cell types (Fig. 1d-e), SOX10+ putative CNCC subtypes (Fig. 3b; Fig. S8b-c), mesenchymal subtypes (Fig. 5b; Fig. 23b), and ectodermal subtypes (Fig. 6b; Fig. 26b).

- Why do the Authors use only TUBB3 and MAP2 as markers for neuronal cells that they want to remove from the query set? While primarily known as a neuronal marker, TUBB3 expression is also observed in non-neuronal cells and tissues, including endothelial cells, fibroblasts, spermatogenic cells, and stem cells, suggesting a broader role beyond just neurons. The choice of neural markers appears to be ill-informed and extremely reductive.

We agree with the reviewer and have consequently changed our method of removing neuronal and radial glia cell clusters. We have extensively described this new systematic method of cell removal above and in the Methods section of the revised version of the manuscript on lines 675-678 and in the Supplemental Note.

- Regrettably, there appears to be lack of rigor in tissue collection. It is of concern that the Authors lose half of their cells through the first filtering step. According to their statements, they started from 86,359 individual nuclei, but ended up with 42,131 remaining nuclei for their analyses. They concluded: “We reasoned these clusters were derived from developing brain tissue not directly part of the craniofacial structures. We therefore excluded these nuclei from downstream analyses resulting in a total of 42,131 remaining nuclei with an average of 4095 nuclei from each sample”. In summary, either the contamination of facial tissues by brain tissue was massive, or the Authors also removed cells that are not neural cells and thus could be integral part of head

developing structures.

On this issue we strongly rebut the reviewer that there is a lack of rigor in tissue collection. As detailed above, working with early human embryonic tissues obtained from ethical sources yields more complex samples. We have thus relied on increasing both replicate numbers at each timepoint and number of individual samples profiled. This is therefore one of the largest numbers of individual human embryos that have been sampled for any tissue to date. That being said, to make this clearer to the reader we have now included a section in the Discussion section of the revised manuscript that lays out the caveats and limitations of the human snRNA-seq data. One of the main limitations is that “the nature of the human tissue we received precluded precise dissection, as we can perform in mouse guided by anatomical landmarks of the entire embryo.” That said, we have adopted a new method of processing the data as described in the Methods section on lines 657-690, in which we resolve over 2x the number of nuclei (95,892 using new method vs. 42,131 nuclei using old method):

“Raw barcode matrices from each human samples generated by CellRanger ARC were utilized for downstream processing to maximize total nuclei. Raw barcode matrices for each sample were individually loaded with Read10X_h5 command in Seurat ¹⁵² (v. 5.2.1). Subsequently, the emptyDrops function within the DropletUtils R package (v. 1.24.0) was used to identify nuclei containing ambient RNA for each sample (emptyDrops FDR < 0.001). Individual sample Seurat objects were then merged into a single object. Percentage of mitochondrial reads were calculated for each cell and filtering was performed to only retain cells with less than ten percent mitochondrial derived. Further filtering was performed based on number of counts per cell ($500 < x < 25000$) and number of genes detected per cell ($500 < x < 7000$), resulting in a total of 130,427 nuclei. Filtered data were normalized with default values and cell cycle scores were calculated using Seurat. Data was scaled based on S and G2M score regression and dimensionality reduction with principal component analysis (PCA) were performed using respective commands in Seurat. The top 2000 variable features were identified, and sample-level batch correction was performed using the harmony R package¹⁵³ (v. 1.2.3). Nearest neighbors based on harmony corrected embeddings were calculated with up to 30 dimensions and clusters were identified with multiple resolutions from 0.1 to 1 in Seurat. Resolution parameters were selected for clustering based on generated elbow plots and identification of the elbow point to determine the optimal number of clusters. We then performed Uniform Manifold Approximation and Projection (UMAP)³⁶ dimensionality reduction using harmony corrected embeddings in Seurat (dimensions = 30, minimum distance = 0.3). Resulting clusters were inspected for expression of multiple craniofacial markers from Li et al 2019¹² and marker genes were identified for each cluster. Module scores were generated using the AddModuleScore function in Seurat for neuronal markers and radial glia markers from Eze et al 2021¹⁵⁴. Cells from clusters identified with high neuronal or radial glia module were removed (See Supplementary Note), resulting in 95,892 nuclei. The process of normalization, harmonization, and clustering was repeated with remaining cells from all samples. We observed that early samples had consistently higher mitochondrial

reads (Fig. S2d), however, this was not due to increased apoptosis as tested by generating an apoptosis module score using gene markers from the Hallmark Apoptosis gene set within the human MSigDB Core Collection (Fig. S2d). The sex of each embryo was also identified using known sex determination genes for males (*UTY*, *DDX3Y*, *KDM5D*, and *NLGN4Y*) and females (*XIST* and *TSIX*) (Fig. S2e), and each of the main cell type clusters contained roughly the same number of male and female samples (Fig. S2f). Marker genes for major cell types were identified using FindAllMarkers (logfc.threshold = 0.25, min.pct = 0.1, test.use = “wilcox”, min.cells.feature = 3, min.cells.group = 3, pseudocount.use = 1, and return.thresh = 0.01). The top 100 marker genes for each cluster (p < 0.05, ranked by log2fold change versus all other clusters) were then analyzed for gene and disease ontology enrichments using compareCluster in clusterProfiler R package (v. 4.12.6). Cranial neural crest genes were compiled based on markers identified by regulatory network construction in human cultured CNCC and craniofacial tissue data⁶⁵. Major cell type labels were applied to each cluster.”

- The Authors state: “Early samples had consistently higher mitochondrial reads (Fig. S2d), potentially reflecting their higher dependence on mitochondrial output or an artifact related to lower cell numbers in the processing of each sample.” This assertion is superficial and must be revisited. It is well known that high mitochondrial signatures are a strong indicator of apoptosis. The Authors have not unequivocally ruled out the presence of cell death in their samples.

We agree with the reviewer and understand their concern. In the revised version of the manuscript, we have ruled out that the early samples having higher mitochondrial reads is not due to apoptosis. More specifically, in lines 679-681, we state:

“We observed that early samples had consistently higher mitochondrial reads (Fig. S2d), however, this was not due to increased apoptosis as tested by generating an apoptosis module score using gene markers from the Hallmark Apoptosis gene set within the human MSigDB Core Collection (Fig. S2d).”

Single-nucleus RNA Sequencing (snRNA-seq) Data Generation and Processing (Mouse Samples)

- Similar concerns are present regarding the isolation of mouse tissues. There is a lack of adequate information regarding the collection of the mouse tissues. Were mouse tissues also fresh frozen + thawed (like human tissues) or were they processed directly following tissue collection? Differences in tissue processing could heavily influence the snRNA-seq results obtained. There are serious concerns that the “species-specific” signatures the Author report might be due at least in part to differences in processing of the collected tissues, and they may not necessarily underpin species-specific identities.

Mouse samples were dissected fresh from whole embryos and then flash frozen to mimic the condition of the human samples (lines 624-635). Additionally, we

have removed the text discussing species-specific signatures to remove extraneous detail and narrow the focus of the revised version of the manuscript.

The authors state that: “Samples were mechanically broken into single-cell suspensions” and that “Craniofacial prominences were collected in a very similar fashion to human samples”. What do the Authors mean by “similar”? Were neural tissue also removed in mouse craniofacial tissue dissections as it was done for human dissections? Or was there absence of neural contamination? All of these details, which have critical bearing on the reported findings of species-specific differences, are not provided.

We have included additional information about the collection and processing of mouse craniofacial tissue in the Methods section of the revised version of the manuscript in lines 624-635. Additionally, in lines 713-715, we state:

“less significant contamination of neuronal cell types was observed in mouse data, which was identified and filtered as described for human.”

- The Authors compare their datasets to those reported by the group of Trevor Williams (Li, H., Jones, K.L., Hooper, J.E. & Williams, T. The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. Development 146, 2019). But the latter datasets were generated only from murine midfaces and only at E11.5. This substantial difference in the datasets should at least be mentioned.

We have now included this additional detail in the Results section of the revised version of the manuscript.

Neural Crest Cell Cluster (Cluster 6)

- What is cluster 6? It is not at all clear whether this is a biologically defined group of cells or a somewhat “mixed bag of cells” cluster. The Authors seem not to be completely sure themselves: “We then turned to the not yet concretely identified cluster 6 [...]. We did not detect any significant pathway enrichments for this particular cluster (201). “The wide variety of biological functions and specific disease enrichments all suggested that this cluster might be enriched for neural crest cells. (208-210)”.

We now refer to this cluster as the SOX2-/SOX10+ putative CNCC cluster in the revised version of the manuscript. When we perform subclustering of this cluster, we find SOX10+ putative CNCC’s as well as Schwann cells, which we identified by module score-based analysis as described above.

These descriptions are rather confusing.

- The Authors state that CNCC markers are significantly higher in cluster 6. “When we examined a module of genes from regulatory networks recently identified in cultured human and chimpanzee cranial neural crest cells (CNCCs)⁶⁹, we found significantly higher expression in cluster 6 (Fig. 2D).”

To confirm that our initial labeling of this cluster as putative CNCC's, we generated this module score. We were able to confirm such annotation based on our finding that this cluster is significantly higher than all others using a generated CNCC module score of genes from regulatory networks identified in cultured human and chimpanzee CNCCs⁶².

Which test or package or statistics did they use to reach this conclusion? Again, no relevant information is provided.

We used the AddModuleScore function within the Seurat package. We have included this detail in the Methods section of the revised version of the manuscript.

- Do all developmental stages contribute equally to the population in cluster 6? This information is lacking and is very important.

We have now included UMAPs colored by stage and boxplots highlighting the stage-specific bias in Fig. 3b and Figs. S8b-c.

- The Authors state: "SOX9 and COL20A1 were more specifically expressed across the (CNCC) clusters, but again in non-overlapping patterns." (261-262). This sentence is very confusing. What do the Authors mean?

We have adjusted this sentence in the Results section (lines 165-166) to state

"Interestingly, SOX9 was specifically expressed in the mesenchymal committed CNCCs (Fig. 3c)."

- There appears to be lack of a definite proof that Cluster 6 is truly a Neural Crest Cell cluster.

In the revised version of the manuscript, we have confirmed the identity of this cluster. After subclustering, we were also able to distinguish the SOX10+ CNCC's from the SOX10+ Schwann cells.

Spatial Transcriptomics

- There appears to be huge differences in between the two replicates shown for spatial transcriptomics (Fig. 3 E,F,G). For certain markers and modules the two replicates shown in the figure appear to be completely different from one another regarding the respective gene expression spatial features. Clearly these human samples are precious and hard to collect, making it difficult to generate multiple biological replicates. However, it is very hard to draw firm conclusions from these spatial validations.

We understand the reviewer's concern. However, the spatial transcriptomics data comes from two non-identical serial sections, resulting in the variability shown. We describe in the updated Methods section in lines 784-786:

“Slice 1 is more laterally localized while Slice 2 is more medially localized. Thus, representative plots for ectoderm are shown in Slice 1 and for putative CNCC's and mesenchyme are shown in Slice 2.”

Therefore, the spatial transcriptomics figures shown for each of the main cell types should be more representative of true subtype-specific regional localization.

Enrichment of Disease Genes and Variants Linked to Facial Variation

These analyses are hard to follow and very cryptic. Disease associations are correlative. Certain statements and conclusions do not appear to be based on rigorous supporting data. One example is provided below. The Authors state:

“Interestingly, while CNCCs certainly give rise to many of the downstream cell types and tissues, we found relatively few shape associations for CNCC subtypes” (531-532)

This statement is somewhat concerning, given that a rich body of literature directly links differences of facial shapes to variants in CNCC genes. The following review article comprehensively summarizes most of the relevant literature showing that variants in CNCC genes and enhancers underpin differences in face variation: Naqvi S, Hoskens H, Wilke F, Weinberg SM, Shaffer JR, Walsh S, Shriver MD, Wysocka J, Claes P. Decoding the Human Face: Progress and Challenges in Understanding the Genetics of Craniofacial Morphology. Annu Rev Genomics Hum Genet. 2022 Aug 31;23:383-412. Interestingly, the Authors cite this review article in their “Introduction”.

While we understand the reviewer's point that these previous papers have purported to show enrichment of facial variants in CNCC genes, we have shown in both the Wilderman et al 2018 and Yankee et al 2023 papers that the culture system they employ is more terminally differentiated down a mesenchymal lineage present later in development. Our comparisons of gene expression between other methods of CNCC generation and primary tissues in Yankee et al 2023 supplemental Figure 6D clearly show this discrepancy. In our own analysis of facial variant enrichment in primary tissues in Wilderman et al 2018, we observed significant enrichments at the very end of the embryonic phase (~ 8 post conception weeks). This is after fusion of many of the facial structures and after neural crest have differentiated into many other cell types. Thus, our findings here agree with both our previous findings and those discussed in the review: neural crest-derived mesenchymal cell types are significant contributors to facial shape. Nonetheless, we have streamlined the discussion to stick to the most salient points.

Minor points

- Inconsistency in size of figure labelling (A,B,C etc.), legends in multiple figures (partially) overlap, image quality in multiple figures is poor.

In the revised version of the manuscript, we have corrected all of these issues including the size of figure labels, the overlap of legends in figures, and the image quality.

- Lettering in some figures is exceedingly small and/or warped and thus very hard to read, even when zooming in onto the image or cartoon.

We have corrected these errors in the revised version of the manuscript.

- The Authors need to revise the nomenclature for genes. Mouse genes should not be capitalized.

We apologize for this oversight and have checked and corrected the nomenclature of the human and mouse gene names.

- Some of the references cited are not correct. For example, Ref 12 (line 146) is not correct.

We apologize for this error and have corrected this in the revised version of the manuscript.

Style and Grammar (only a few issues of many are reported below):

- 105: “after filtering for doubles and quality of per nucleus data“. Confusing sentence.
- 106/107: “Samples obtained from human embryos are more difficult to control what is collected“. Sentence needs rewriting.
- 256: “TFAP2A was expression was observed in all clusters“. Sentence needs rewriting.
- 625-627: “We found no enrichment in any cell types for genes affected by de novo synonymous variants. We found no enrichment in any cell types for genes affected by de novo synonymous variants.”
The sentence is repeated twice.
- 758: space and period missing: “embryo25By“
- 815 f. Clip -> lip: “Failure of this region to fuse in humans has been suggested to cause cleft lip that could also involve the nostril region and primary palate”
- 833: dominant or major: “Mesenchyme is by far the dominant major cell type”
- 993: describeD: “as describe for human”
- 1019: remove word duplication: “mouse genes were reduced and converted to those to those with one”
- 1058: space and period missing: “MAGMA89Frontal”

We thank the reviewer for diligently reading through the manuscript and pointing out these errors. In the revised version of the manuscript, we have addressed all of these issues to improve style and grammar.

Lastly, this reviewer suggests careful editing of the text – in its present form, the manuscript appears to have been rushed and, regrettably, it is overall sloppy.

To improve the quality of the manuscript, we have since made significant changes to correct for grammar, spelling, and all other language issues. We hope this addresses all of the reviewer's concerns.

Reviewer #4 (Remarks to the Author):

The authors present a comprehensive single-nucleus RNA sequencing (snRNA-seq) atlas of human craniofacial development, derived from 24 embryonic samples spanning six key time points during the embryonic period. Additionally, they integrate spatial transcriptomics data to further localize cell populations. Our findings shed light on the developmental origins of craniofacial disorders, the genetic basis of facial variation, and the evolutionary legacy of ancient hominins. Overall, this work provides a valuable new atlas for craniofacial biology, supported by comprehensive experiments. However, there are still areas that require improvement. Specifically, some paragraphs should be refined or removed to enhance readability. At present, several sections in the Results are overly redundant, making it difficult to focus on the key findings. I encourage the authors to revise the "Results" section for improved clarity and readability.

We thank the reviewer for their acknowledgement of the value that the collected atlas will provide to the field of craniofacial biology. We understand the reviewer's concerns and have significantly refined the manuscript in its revised version to remove redundancy, enhance readability, and improve clarity. We hope these changes address all of the reviewer's comments.

My comments are as follows:

1. Beyond providing a more extensive atlas, have the authors identified any conclusions that Yankee, T.N. et al. (Ref 13) did not?

We performed substantial subclustering of the major cell types and thoroughly explore the newly identify CNCC population. Furthermore, we perform systematic enrichment analyses related to both common variants related to facial shape and disease risk as well rare de novo variation in specific subtypes of ectoderm. Only a single timepoint was addressed in the previous paper. The substantial analysis of bulk gene expression in that work precluded an expansive section on the single cell data that we have thus described here.

2. The authors state, "Closer inspection of the genes strongly expressed in these clusters revealed many canonical neuronal genes, such as TUBB3 and MAP2." How did they quantify "strong" expression? Furthermore, they excluded nuclei derived from

developing brain tissue. How can they ensure that all unrelated clusters from craniofacial structures were successfully removed?

We understand the reviewer's concern and consequently have changed our method of removing neuronal and radial glia cell clusters. We have described this new systematic method of cell removal above, in the Methods section of the revised version of the manuscript on lines 675-678, and in the Supplemental Note. Briefly, we used a module-based method to quantitatively remove clusters enriched for neuronal and radial glia scores above the median. Prior to this removal, we obtained 130,427 nuclei after filtering for doublets and nuclei quality, which was then reduced to 95,892 nuclei after removing neuronal and radial glia clusters.

3. Line 172: Why did the authors choose 100 marker genes per cluster? In contrast, in line 156, they selected only 10 marker genes to identify unknown clusters. What is the rationale behind these differing marker gene selection criteria?

We understand the reviewer's concern and apologize for any confusion. We used the top 100 marker genes for each cluster for downstream functional analysis but used the top 5 marker genes for each cluster for visualization in the dot plot (Fig. 2b). This is not our only method of main cell type annotation as we have also characterized our clusters using the expression of known marker genes from Li et al 2019. Finally, we identified enrichment of gene ontologies and diseases relevant to our labeled cell types.

4. Lines 172–216: The authors should refine these paragraphs to highlight the main findings rather than listing excessive details.

We agree with the reviewer and have significantly refined these paragraphs to a single paragraph in the revised version of the manuscript, highlighting the main findings and removing extraneous detail.

5. Line 973: It is unclear why the authors selected multiple resolution values ranging from 0 to 1 to identify clusters.

We have selected the resolution values based on the optimal number of clusters as computed using an elbow plot. We have included these elbow plots with the elbow point (optimal number of clusters) quantitatively determined for each of the clustering analyses. We have detailed this in the revised Methods section on lines 670-672.

6. Line 1034: The authors should clarify how they preprocessed the spatial transcriptomics data quantitatively and in detail. Did they filter out low-quality spots?

We have detailed the spatial transcriptomics preprocessing steps in the “Processing of spatial transcriptomics” section of the Methods in the revised version of the manuscript on lines 783-812.

“Spatial transcriptomics data for two sagittal sections of a human CS13 embryo²⁵ were retrieved from <https://heoa.shinyapps.io/code/>. Slice 1 is more laterally localized while Slice 2 is more medially localized. Thus, representative plots for ectoderm are shown in Slice 1 and for putative CNCC’s and mesenchyme are shown in Slice 2.

Raw sequence count matrices were loaded using the Read10x command of Seurat¹⁵² and converted to HDF5 format. These counts were then combined with spot coordinates and section images using CreateSeuratObject. Data from both slices were merged and variable features were determined using Seurat. The percentage of mitochondrial reads was determined for each cell and was used to transform all data in the merged object using SCTransform¹⁶⁹ from Seurat. Data was clustered using UMAP and plotted which revealed a strong batch effect between the two spatial objects. Data was further normalized using Harmony (v. 1.2.3), projected using UMAP, and clustered with a resolution of 0.8. Marker genes for the 22 clusters were identified using FindAllMarkers in Seurat. The top 100 marker genes for each cluster ($p < 0.05$, ranked by log2fold change versus all other clusters) were then analyzed for gene and disease ontology enrichments using compareCluster in clusterProfiler R package (v. 4.12.6). Enrichments and spatial localization were compared to previous annotations by Xu et al 2023²⁵ and labelled accordingly. The top 100 marker genes from each of the subclusters identified in human craniofacial data were used to calculate module scores across the merged spatial object and plotted using SpatialFeaturePlot in Seurat.

E9.5-16.5 spatial transcriptomics datasets were retrieved from the MOSTA resource⁷¹ and loaded into Seurat as for human data above. For representative images in the supplement, we chose to use the mouse E11.5 spatial transcriptomics data as it is most morphologically similar to CS13 human embryos. For the E9.5-E16.5 combined stages object, we converted the mouse genes to those with one-to-one orthology with human genes using the orthogene R package (v. 1.10.0). Following, we used FindTransferAnchors in Seurat with log normalization and canonical correlation analysis (cca) to transfer the human main cell type and subtype annotations to the mouse spatial transcriptomics dataset using only the highly variable genes identified using FindVariableFeatures in Seurat. Predicted labels were then transferred to mouse spatial transcriptomics data using TransferData in Seurat. For each E11.5 section, module scores of top 100 marker genes for each main cluster or subcluster were calculated. Gene spatial feature plots for selected genes and module scores were then generated with Seurat.”

7. Figures 2, 3, 4, 5, and 6 share a common issue: they contain excessive information, making them overwhelming. The authors should consider moving less critical subfigures to the supplementary materials. The main figures should emphasize distinct and

important messages for readers.

We agree with the reviewer and have moved some of the less critical subfigures to the supplement. We hope that this narrows the focus of the figures to enhance their conciseness, clarity, and accessibility.

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We again thank the reviewers for their time and thoughtful comments in evaluating our revision. We appreciate they recognized our significant efforts to streamline the text and provide additional insightful analyses of this substantial craniofacial resource. Indeed 4 of the six reviewers had no additional comments or suggestions and recommended publication. However, there were still a few areas of contention related to source of tissue and identification of specific cell-types/subtypes based on single genes. On this first point we have not tried to oversell what tissues we have obtained and make claims about precision of tissue collection that we cannot defend. This is in line with many previous papers' description of the source of human and mouse craniofacial samples, including a recent paper in this journal (Rajderkar et al 2024). We instead leveraged unbiased, systematic filtering and identification of clusters to present to the reader what cell types we obtained. That being said we have added a simple statement to further describe what portions of the head we tried to exclude in the Methods section. Specifically we have added the statement "Primary human craniofacial prominences excluding the eyes and more anterior tissue from CS12, CS13, CS14, CS16, CS17 and CS20, each stage represented by a minimum of 3 replicates, were obtained from HDBR." Furthermore while we have made our best efforts to collect comparable samples from both human and mouse there is no way to make them identical. We again err on the side of caution and contend that samples are very similar but not identical.

On the second point, we apologize if there was confusion about the use of published marker genes described by Li et al 2019. Since this paper was one of the first to use single cell gene expression on craniofacial samples and has been frequently cited by other papers in the field, we used those markers in a qualitative way in our initial exploratory analysis of the craniofacial clusters. In the example of ALX1, our analysis found it had biased expression of this gene in the large cluster initially labelled as mesenchyme but was far from being a top marker (number 523 on our list). Instead our results highlight many other mesenchymal associated genes which we provided in Figure 2B and Supplemental Table 1 of the previous revision. To reduce the possibility of confusion we have altered Figure 2A heading to "Li et al. 2019 Cell-Type Markers" to highlight how they were identified and used.

Our full detailed response to each point raised by Reviewers 2 and 3 can be found in line below.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have addressed my previous comments and present a substantially improved manuscript. I am supportive of publication. However, I encourage the authors to further tighten the text. The manuscript still appears to exceed the recommended word limit for this format by more than 50% (approximately 7,600 words for a 5,000-

word limit). Condensing it to around 5,000 words would improve its focus, readability, and overall accessibility.

We thank the reviewer for their support in publishing our manuscript and recognizing the substantial changes and additional analyses we performed for the revision. We will work with the journal to further shorten, if necessary, but not all reviewers agree on this point.

Reviewer #2 (Remarks to the Author):

The revised manuscript by Khouri-Farah et al describing a novel dataset derived from single-nucleus RNA-seq analysis of human embryonic face and head development across weeks 4-8 after conception is significantly improved from the previous version. However, several major concerns still remain that should be rectified.

1. Although it has been shortened, the manuscript is still very long at about 7500 words through the Discussion section. The main value of the manuscript is that it describes the overall dataset that can be used as a resource for the community. However, the author's attempt to cover multiple topics – the human data, the mouse data, the human versus mouse data, human disease associations, human facial shape variation, and Neanderthal anatomy – with the data they have obtained. This makes the analysis very broad and it reads as one list after another. Moreover, verification of the data can only be superficial given the limited access to human embryo tissue samples. Therefore, these limits in resolution and in mapping of human data onto human anatomy reduces the utility of using such datasets to study shape variation in a robust manner. The manuscript would be improved by moving some of these sections on anatomy to a different manuscript to shorten and focus the current version.

We appreciate the reviewer's concerns; however, we believe that the analyses presented in the manuscript offer unique and critical insights into the regulatory and transcriptional programs occurring during craniofacial development. We extend beyond the superficial descriptions that have been offered for previous mouse single cell craniofacial papers and the human genome-wide association findings for craniofacial variation. This includes using the newly generated human and mouse single cell atlases of craniofacial tissue to determine cell type contributions to disease risk and facial variation. In doing so, we have identified pertinent cell subtypes enriched for *de novo* protein altering variants in orofacial clefting patients. Conversely many of the papers that have identified genetic associations with facial variation have either not performed any functional enrichments or used irrelevant adult cell type data. Unless something we have written is factually incorrect or not grounded in the data, which the reviewer raises no examples, we will elect to retain the bulk of the text in its revised form.

2. The text of the abstract, introduction and results is still misleading in regards to the tissues used for the analysis, as for example, when the author's state that they have "constructed an atlas of human craniofacial development by collecting craniofacial

tissues from 24 individual human embryos” (Section Main, line 61). Although they provide caveats at the very end of the discussion, it would be more accurate to state throughout the manuscript that they have “collected head and neck tissue containing the craniofacial complex and used bioinformatic methods to remove presumptive neural tissue from the analysis.” This would provide readers with a more accurate picture of the data used throughout the manuscript.

We understand the reviewer’s concerns; however, at the stages that we obtained tissues the neck does not yet exist. While the tissue dissection method was less precise than that of the mouse, as these samples are the result of elective terminations, tissue was collected from craniofacial prominences excluding the eyes and more anterior parts of the head. We have updated the methods section to include that specific detail (Lines 646 and 658 in this revision). This level of description of tissue from both human and mouse has been used by multiple groups outside of our own, including several published in this journal recently (e.g. Rajderkar et al 2024). Thus, including more specific claims about what tissue was obtained, which we also cannot state with certainty, would only further add to confusion.

Still, we reasoned this dissected tissue is likely to contain slight contamination from brain regions, so we utilized bioinformatic methods to remove presumptive neural tissue. The full details of our filtering steps and choice of genes for calculating the modules scores are found in the Supplemental Note in the first two pages of the Supplemental Information Document which we referenced on line 804 in the Methods section of the previous revision. Our mouse samples had one neuronal cluster that we removed using the same bioinformatic methods but no radial glia (Supplemental Figure 12). Following that bioinformatic filtering, we let our unbiased analyses reveal the cell types that are present in the sample and focused on those that can be clearly identified in both human and mouse. We added caveats we thought were reasonable, which the reviewer agrees, and cannot be fully articulated in a short abstract.

3. The mesenchyme, cluster 0, is the largest cluster they identify in the UMAP analysis. In my previous review I had highlighted that the mesenchyme of the head is derived from two major tissue-types, the mesoderm and the neural crest. Therefore, it would be expected that a very large fraction of cluster 0 actually represents neural crest derivatives. For example, one of the markers they use for cluster 0 is ALX1. Expression of this gene is known to be strongly associated with the craniofacial neural crest in human, mouse, and zebrafish (e.g. Pini et al. EMBO Mol Med 2020 “ALX1-related frontonasal dysplasia results from defective neural crest cell development and migration” and Iyyanar et al Front Cell Dev Bio 2022. “Alx1 Deficient Mice Recapitulate Craniofacial Phenotype and Reveal Developmental Basis of ALX1-Related Frontonasal Dysplasia”).

My interpretation of the author’s presentation is that only clusters 5 and 7 are neural crest. Perhaps the author’s postulate that clusters 5 and 7 are migratory neural crest

cells and can be viewed differently than other neural crest derivatives. However, as cluster 5 is associated with cell types differentiating towards Schwann cells and melanocytes, they would appear to be no different than neural crest cells in cluster 0 that are differentiating towards skeletal elements. Therefore, it is imperative that the authors directly discuss the contribution of neural crest to cluster 0. I would suggest that altering the following statement on line 100 would be one of the simplest ways to accomplish this - although there are others:

“ALX1 was most strongly expressed in cluster 0 (mesenchyme),”
changed to

“ALX1 was most strongly expressed in cluster 0 (mesenchyme), a tissue -type which has a dual origin from the mesoderm and cranial neural crest during facial development”

I would then also recommend changing:

“We inferred RNA velocity and transition states to identify trajectories in the data across the newly identified main cell types. We found that the trajectories emanate from the SOX2+/SOX10- population, confirming that these cells likely represent an early progenitor population (Fig. 2c).”

To:

“We inferred RNA velocity and transition states to identify trajectories in the data across the newly identified main cell types. We found that the trajectories emanate from the SOX2+/SOX10- population, confirming that these cells likely represent an early progenitor population for other neural crest populations in clusters 0 and 5 (Fig. 2c).”

Defining the contribution of the neural crest to the craniofacial mesenchyme also has implications for how the data relating to critical tissues involved in craniofacial disease and shape variation are considered later in the results section.

We apologize if the reviewer thought these genes were ascertained as formal marker genes. To begin our analysis, we used *ALX1* and several other cell type specific markers highlighted in the first single cell gene expression datasets available from mouse craniofacial tissues (Li et al 2019). They described *ALX1* as a pan-mesenchymal marker, and as we have shown likely lacked significant populations of neural crest. This qualitative, exploratory first step revealed several clusters in which none of these markers from Li et al 2019 showed expression. This finding motivated the downstream unbiased cluster identification we performed. Indeed, when we examine cluster 0 *ALX1* is biased in expression toward this cluster, however it is number 523 on our ranked list of makers for cluster 0. This result was previously included in Supplemental Table 1. It is thus not a very specific marker of mesenchyme and why we did not highlight it in subsequent figures or text related to main cell types. To reduce the confusion of how these genes were selected and presented in Figure 2 we have changed the heading of panel A to “Li et al. 2019 Cell-Type Markers”.

In agreement with the reviewer's assertion that *ALX1* is likely to play a role in CNCCs (and which we agree as we were among the authors of one of the cited papers), we found that *ALX1* is a marker for a subcluster of the CNCCs which we labeled mesenchymal committed CNCCs. We highlighted this finding in the first revision in Figure 3D, Supplemental Table 3 and listed in the main text on Lines 165 to 168 (Lines 167 to 170 in this revision). Thus, our findings are aligned with multiple papers in the field related to *ALX1* function and demonstrate expression in primary human craniofacial cell types for the first time. Moreover, these findings demonstrate that *ALX1*, and likely most genes, play multiple roles in development, regulating the specification and function of multiple cell states across different stages. Thus, it is difficult to ascribe one gene as definitive for a specific cell type and instead should be thought of in a network concept as we have attempted to do so throughout the manuscript.

We agree with the reviewer that cluster 0, defined as mesenchyme, is derived from multiple primordial cell types and have updated the specific text as suggested (Lines 100 to 101 in this revision). However, we do not agree that these are neural crest cells as the reviewer seems to indicate. To confirm our initial labeling of putative CNCCs vs. mesenchyme, we generated a violin plot of CNCC module scores across main clusters and found significant differences in scores as we showed in Figure 2E of the previous revision. When we extended this analysis to all cell subtypes, we find that all subtypes labeled as *SOX2*-/ *SOX10*+ putative CNCC have high CNCC module scores, while those from other main cell type clusters are considerably lower (Lines 180 to 183 this revision). We generated this CNCC module score previously, but did not include a plot of the results in our previous revision. We have included in this revision as panel J in Supplemental Figure 8. Thus, we are confident that the mesenchyme subclusters might be derived from neural crest but have transitioned to a much more differentiated state.

4. Spatial transcriptomics on two human embryo slide sections is still utilized for verification analysis in several main figures as well as in multiple Supplementary figures. As stated in the previous review the resolution of spatial mapping in the current study was a 55µm spot, whereas an approximately 2µm spot would be required to accurately map a transcript to a single cell layer. Given this issue, it is very difficult to interpret the data shown. I would recommend that arrows are used highlight and label the regions we should be focusing on, or else many of the panels should be dropped if they do not contribute to the analysis so that a reader can focus on the salient points of the study.

We wholeheartedly agree that it would be desirable to have higher resolution spatial data from human. Unfortunately, none exist anywhere we can publicly access, and we do not have otherwise intact whole human embryos to generate such a resource. To bridge this gap in our previous revision, we demonstrated accurate projection of human subtypes onto much higher resolution mouse spatial transcriptomics data (Figure 4 G and H). Given the resolution of the

human data, we are hesitant to draw arrows or domains on the sections given many of the issues the reviewer raises. Instead, we leave these plots up to reader's interpretation since they qualitatively show biased expression of groups of genes across the embryo. Some are biased for the head/face but others are expressed in multiple regions reflecting our earlier comment on using any one gene, like *ALX1*, to try and define a single cell type, tissue, or organ.

Reviewer #3 (Remarks to the Author):

The paper by Nagham Khouri-Farah and Alexandra Manche "Gene expression patterns of the developing human face at single-cell resolution reveal cell-type contributions to normal facial variation and disease risk" has been substantially improved after revision and is now stronger, shorter, and clearer.

The following issues still need to be addressed:

1) The authors state that: "Samples were mechanically broken into single-cell suspensions" and that "Craniofacial prominences were collected in a very similar fashion to human samples".

What do the Authors mean by "similar"? Was the neural tissue removed when conducting the mouse craniofacial tissue dissections, as it was done for human dissections? Was there absence of neural contamination? All of these details, which have critical bearing on the reported findings of species-specific differences, have not yet been clearly provided.

The Authors state that they have included additional information about the collection and processing of mouse craniofacial tissue in the Methods Section of the revised version of the manuscript:

"less significant contamination of neuronal cell types was observed in mouse data, which was identified and filtered as described for human."

The point has not been fully addressed; it would help to clearly explain how the dissections were carried out. In the Methods Section it should be precisely specified which mouse craniofacial prominences were dissected and analyzed for the transcriptomic analysis. The text states "craniofacial prominences". Was the whole mouse embryonic head/face homogenized, or which facial prominences were dissected?

While we use the same mechanical process and downstream multiome generation on frozen tissue from both species, we started with tissue dissected from whole mouse embryos and tissue pieces from human. Additionally, given that we received the human samples frozen we cannot be absolutely certain that we have frozen our samples exactly the same way. As we responded to reviewer 2 above, our description of the tissue harvest is in line with many other papers in the field. Thus, we elect to stay with a similar description and not overstate what was performed in the study.

Due to more precise dissection of the mouse craniofacial prominences, we could control extraneous structures and tissues. We excluded eyes and more anterior portions of the head and brain from all mouse embryos and did not simply grind up the whole head. Thus, from the very beginning of our data generation we reasoned there was less contamination of the tissue from neuronal cells. However, any remaining neuronal cell clusters were removed using the same metrics as discussed for human. In the supplemental note of the revised version of the manuscript from our previous submission (Supplemental Note Lines 13-17), we explained our systematic method of removing neuronal and radial cell clusters. Briefly, we state that “We generated module scores for neurons and radial glia using their gene markers from Eze et al. 2021⁴ and labeled these cell types in our data when the module scores were above the median. We reasoned these clusters were derived from developing brain tissue and thus excluded these nuclei from downstream analyses.”

2) The Authors compare their datasets to those reported by the group of Trevor Williams (Li, H., Jones, K.L., Hooper, J.E. & Williams, T. The molecular anatomy of mammalian upper lip and primary palate fusion at single cell resolution. Development 146, 2019). But the latter datasets were generated only from murine midfacial prominences and only at E11.5.

This substantial difference in the datasets should at least be mentioned.

The Authors state: “We have now included this additional detail in the Results section of the revised version of the manuscript.”

It does not appear that this information was included or cited in the revised version of the document. It should be made clear where this information was added within the Results section, as this reviewer still cannot find it.

We apologize about the confusion around this point as both Reviewers 2 and 3 thought we did a more systematic analysis of the Li et al. 2019 data. As this reviewer astutely pointed out that these data were collected in a very different way and processed for single cell RNA-seq not multiome as done here. We agreed with this overall assertion and thus did not directly integrate or analyze the data from Li et al 2019 with our newly generated mouse and human scRNA-seq atlases. Instead, to help initially characterize the main cell type clusters in our human and mouse scRNA-seq datasets we qualitatively examined the expression of craniofacial cell type-specific marker genes reported by Li et al 2019 (*ALX1*, *EPCAM*, *CDH5*, and *FCERG1* as markers of mesenchyme, ectoderm, endothelium, and immune cells, respectively in Lines 98-100). In the human scRNA-seq data, we detailed the following in Lines 100-104: “*ALX1* was most strongly expressed in cluster 0 (mesenchyme), *EPCAM* in cluster 1 (ectoderm), *CDH5* in cluster 4 (endothelium), and *FCERG1* in cluster 6 (immune). We also examined markers of erythrocytes including *SPTA1*, *ALAS2*, *RHAG*, and *HEGMN*, which were strongly enriched in cluster 3 (Fig. S5). Clusters 2, 5 and 7 did not show signal for any of these genes.” The absence of signals for these genes in these additional clusters

warranted the more unbiased marker gene annotation we performed in downstream analysis to identify what these clusters were likely to represent.

As we detailed in our response above to Reviewer 2, we did not identify *ALX1* as a top pan-mesenchymal marker through unbiased analyses. To mitigate this confusion about “marker gene” status we have altered Figure 2A to make it more clear that these genes were markers from a previous paper (Li et al 2019).

3) The revised manuscript still needs additional clean-up. For instance, capital letters are used to label the figure panels, but then lowercase letters are used in the text to refer to the figure panels. Also, format should be consistent throughout the text. For instance, in the Figure Legends either the Authors add a dot or do not add a dot after the parenthesis when they list the figure panels. To give an example: some figure legends appear like this:

A). B). C) D). E)

If the Authors cannot spend the time, or do not care to fix these details, they should ask for the help of a capable editor. Otherwise, the paper will still give the overall impression of sloppiness.

We regret these minor typos and formatting errors and apologize if they detract from the reviewer’s ability to assess the manuscript. As we hope the reviewer can appreciate with multiple people working on a large, complex document with many track changes, comments, etc. merging into a flat document can result in such typographical errors. We are committed to working with the copyeditors upon acceptance to identify any remaining issues so that the final published product is more polished.

Reviewer #4 (Remarks to the Author):

The authors have resolved all of my questions. I recommend accepting this paper.

We thank the reviewer for their support in publishing our manuscript.

Reviewer #5 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Reviewer #6 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review

and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We are pleased that the reviewers' recognized our efforts to address their concerns. However as detailed below, Reviewer 2, has some additional issues that were not initially conveyed to the authors. We address the additional minor and/or subjective points in-line below in [blue](#).

REVIEWERS' COMMENTS

Reviewer #2 (Remarks to the Author):

Please see comments to editors. I do not believe a detailed review of the new version of the manuscript is warranted as so little has changed in response to the previous reviews.

Reviewer #3 (Remarks to the Author):

The Authors have tried to improve the manuscript to the best of their abilities.

This Reviewer understands that working with human embryos is challenging and that certain issues (e.g. use of perfectly reproducible dissections of intact tissues) cannot be completely resolved.

This Reviewer also counts on the Editors of Nature Communications to fix some of the typos, awkward sentences, and formatting issues that are still present in the revised text and figures - and to polish the manuscript.

Other comments from Reviewer #2 (not to Authors):

The second revision of the manuscript by Khouri-Farah et al has made a few modest changes compared to the previous version. Most of the other comments from the three remaining viewers who had recommendations for alterations have instead only been considered in the rebuttal letter. Because of this, I believe that a review of the new manuscript is not warranted. Instead, I would suggest 3 things moving forward.

1.

The paper is too long. I find it hard to read and pull out the really important details.

This is also a point raised by reviewer 1 again:

"The manuscript still appears to exceed the recommended word limit for this format by more than 50% (approximately 7,600 words for a 5,000- word limit). Condensing it to around 5,000 words would improve its focus, readability, and overall accessibility."

Response:

"We will work with the journal to further shorten, if necessary, but not all reviewers agree on this point."

My reading of this comment is that they are not interested in the valid recommendations of the reviewers and want the paper to be accepted without further work. Asking the journal to adjust the length by a third – which would require alterations to the scientific content – does not seem appropriate and this responsibility should instead fall upon the authors themselves.

[With all previous revisions a minority of reviewers complained about length. However, none could give specific sections, passages, or lines that should be removed. Nonetheless in our](#)

revision we removed a large section discussing the Neanderthal results. It is a minor point in the results and while we feel very interesting it could streamline the manuscript.

2.

I still have no clear idea of what human and mouse tissues were used for the analyses.

This is also a major concern for reviewer #3:

“What do the Authors mean by “similar”? Was the neural tissue removed when conducting the mouse craniofacial tissue dissections, as it was done for human dissections? Was there absence of neural contamination? All of these details, which have critical bearing on the on the reported findings of species-specific differences, have not yet been clearly provided. The Authors state that they have included additional information about the collection and processing of mouse craniofacial tissue in the Methods Section of the revised version of the manuscript: “less significant contamination of neuronal cell types was observed in mouse data, which was identified and filtered as described for human.” The point has not been fully addressed; it would help to clearly explain how the dissections were carried out. In the Methods Section it should be precisely specified which mouse craniofacial prominences were dissected and analyzed for the transcriptomic analysis. The text states “craniofacial prominences”. Was the whole mouse embryonic head/face homogenized, or which facial prominences were dissected?”

I am also unclear on this point. In the rebuttal, they state:

Authors rebuttal:

“This level of description of tissue from both human and mouse has been used by multiple groups outside of our own, including several published in this journal recently (e.g. Rajderkar et al 2024).”

Examination of the Rajderkar et al paper in Nature Communications shows that it was accompanied by a schematic of what was dissected in a Supplemental figure, which is helpful in knowing what they at least tried to obtain.

The authors of the submitted manuscript also state that the neck isn't present at the stages they are taking. But this answer avoids having to address why thyroid tissue is found in their samples if they are only taking the facial prominences.

Authors rebuttal:

“Thus, including more specific claims about what tissue was obtained, which we also cannot state with certainty, would only further add to confusion.”

Clearly, two of the reviewers are already totally confused what has been used for the dissections. In this instance, my understanding of the responsibility of authors and reviewers is that the manuscript should provide sufficient detail that the studies can be replicated by an independent party. This is certainly not the case here.

My suggestion is that a schematic is provided to indicate the regions that they were theoretically dissected from both human and mouse samples. If they are uncomfortable providing such schematic then it indicates that the variation in the tissue obtained makes many the claims and conclusions they include in the manuscript less solid.

We apologize for not making this more clear but the diagram in Figure 1a, which has been in every version of the paper handled by the reviewers, has always been intended to indicate the approximate regions that were sampled. We have added a dotted line around the estimated regions collected. We have also included additional text in the legend for Figure 1. Specifically, we write:

“a). Dashed regions indicate approximate anatomical regions of the developing craniofacial region with the exclusion of the eyes from 4 to 8 weeks post conception that were collected by

HDBR for processing in this study. The same tissue source and collection methods were previously described in Wilderman et al 2018²⁸ and Rajderkar et al 2024¹⁴. Individual Carnegie Stages (CS) and replicates at stage are indicated below images. Adapted from Selleri and Rijli 2023²¹. Orthologous regions were collected in this study from mouse samples ranging from E10.5 to E12.5.”

That being said, this reviewer seems to be intentionally obtuse. We obviously cannot collect the samples from the two different species in “exactly” the same way. This would require performing a medically induced termination in mice, collecting material from the dam in whatever condition it was excreted, flash freezing, and then thawing to perform final dissection. We have gone to great lengths to provide multiple replicates for all of our human and mouse samples and describe the differences between them without overstating what was performed.

3.

Neural crest cells.

We probably agree here, but there is still some confusion and I think the authors need to be very specific with the language that is used. My original comment to the authors was: “Defining the contribution of the neural crest to the craniofacial mesenchyme also has implications for how the data relating to critical tissues involved in craniofacial disease and shape variation are considered later in the results section.”

Authors rebuttal:

“Thus, we are confident that the mesenchyme subclusters might be derived from neural crest but have transitioned to a much more differentiated state.”

Authors discussion:

“Among these, the endothelium, mesenchyme, and erythrocyte clusters were the most functionally shared between human and mouse, with the CNCC cluster exhibiting the least conservation. This could reflect substantial functional differences in CNCCs across species indicating that this cell type may be particularly labile across evolution allowing innovation of craniofacial shape as others have proposed.”

Specifically, it is really important that the definition of CNCCs is stated very clearly when considering things such as shape change and cleft databases. I would argue that the current authors are only using a limited subset of CNCCs for their analyses in this regard. Many other people in the field would use migratory CNCCs and their derivatives in the mesenchyme as the CNCC population.

So, my recommendation is that the authors are very specific on what they are using in these latter parts of the paper: a subset of neural crest that excludes those that are differentiating into mesenchymal derivatives (the latter will be a very significant percentage).

Respectfully, this reviewer is cherry picking his own comments and taking our responses out of context. They simply highlighted the last sentence of an extensive reply we wrote in our original response. Moreover, the reviewer’s own comments are not self-consistent.

“Many other people in the field would use migratory CNCCs and their derivatives in the mesenchyme as the CNCC population.”

If cells are derived from CNCCs and have now become some other clearly defined cell type like cartilage or osteoblasts I don’t think anyone in the field would call them neural crest. While this

reviewer may think of these cell types in this way, this is not how the rest of the field does. This is exemplified by multiple craniofacial single cell papers that have been published, including now three in Nature Communications (Radjerkar 2024, Murillo-Rincon 2025, and this paper). These papers and multiple others clearly define a large mesenchymal cluster. All these papers agree that these are largely derived from neural crest but are more differentiated and clearly identified with these molecular techniques. What we show here for the first time, due to the large number of cells and stages we profiled, is a substantial cluster of cells that enriched for gene expression signatures of neural crest. We go into great detail to define specific CNCC subpopulations as being early, more plastic cells, those that are committed to specific lineages, and those that are more differentiated into cell types like melanocytes or Schwann cells.

However to attempt to address the reviewer's concern we have added the following text to the CNCC subclustering section:

"In summary while the SOX2-/SOX10+ cluster is clearly enriched for gene expression signatures of CNCCs, not all subclusters are CNCCs. Instead some subclusters represent more terminally differentiated cell states including Schwann cells and melanocytes. Moreover since the neural crest have been somewhat elusive to detect in previous single cell atlases it is possible other CNCC states are present in the developing embryo that we are unable to sample here."

We have also added the following to the discussion:

"The decreased conservation of CNCC gene expression programs could indicate this cell type may be particularly labile across evolution allowing innovation of craniofacial shape as others have proposed^{1,18-20,65,106}. Alternatively these differences could be reflective of our narrow definition of CNCCs or technical differences in the non-CNCC subclusters obtained in each species. Further analysis of differentially expressed genes from these subtype clusters in developing mouse embryos and human culture models could help to disentangle these possibilities."

The reviewer also seems to not grasp that we use all clusters for our analyses and show them in all figures. The reviewer suggests:

"So, my recommendation is that the authors are very specific on what they are using in these latter parts of the paper: a subset of neural crest that excludes those that are differentiating into mesenchymal derivatives (the latter will be a very significant percentage)."

We do not exclude any clusters in any analyses and plot them in all figures and present in all tables. What is abundantly clear in our final figure is that the genetic enrichments are biased toward marker genes of clusters with epithelial/ectodermal identity.