

# Single-cell transcriptomic landscape of cardiac neural crest cell derivatives during development

Wen Chen, Xuanyu Liu, Wenke Li, Huayan Shen, Ziyi Zeng, Kunlun Yin, James Priest, and Zhou Zhou

**DOI: 10.15252/embr.202152389**

*Corresponding author(s): Zhou Zhou (zhouzhoufuwai@126.com) , Xuanyu Liu (swissox2020@sina.com)*

---

## Review Timeline:

|                     |             |
|---------------------|-------------|
| Submission Date:    | 4th Jan 21  |
| Editorial Decision: | 11th Mar 21 |
| Revision Received:  | 30th Jun 21 |
| Editorial Decision: | 18th Aug 21 |
| Revision Received:  | 19th Aug 21 |
| Accepted:           | 27th Aug 21 |

---

*Editor: Martina Rembold*

## Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. Depending on transfer agreements, referee reports obtained elsewhere may or may not be included in this compilation. Referee reports are anonymous unless the Referee chooses to sign their reports.)

Dear Prof. Zhou

Thank you for the submission of your research manuscript to our journal. I apologize for the delay in handling your manuscript, but we have only recently received the last referee report and we have also discussed these further with the referees and within the team.

As you will see, the referees acknowledge that your dataset is potentially interesting, but they also point out a number of technical concerns that need to be addressed. The use of a single Cre line needs to be at least critically discussed. Further data based on another Cre line could be added but this is not a prerequisite for publication here. All concerns regarding cell numbers, quality control of scRNA-seq data and potential over-clustering need to be addressed. A histological validation of marker expression should be provided.

Given these constructive comments, we would like to invite you to revise your manuscript for potential publication in our Resource section, with the understanding that the referee concerns (as detailed above and in their reports) must be fully addressed and their suggestions taken on board. Please address all referee concerns in a complete point-by-point response. Acceptance of the manuscript will depend on a positive outcome of a second round of review. It is EMBO reports policy to allow a single round of revision only and acceptance or rejection of the manuscript will therefore depend on the completeness of your responses included in the next, final version of the manuscript.

We invite you to submit your manuscript within three months of a request for revision. This would be June 11th in your case. However, we are aware of the fact that many laboratories are not fully functional due to COVID-19 related shutdowns and we have therefore extended the revision time for all research manuscripts under our scooping protection to allow for the extra time required to address essential experimental issues. Please contact us to discuss the time needed and the revisions further.

\*\*\*\*\*

IMPORTANT NOTE: we perform an initial quality control of all revised manuscripts before re-review. Your manuscript will FAIL this control and the handling will be DELAYED if the following APPLIES:

- 1) A data availability section is missing.
- 2) Your manuscript contains error bars based on  $n=2$ . Please use scatter blots showing the individual datapoints in these cases. The use of statistical tests needs to be justified.

When submitting your revised manuscript, please carefully review the instructions that follow below. Failure to include requested items will delay the evaluation of your revision.

\*\*\*\*\*

When submitting your revised manuscript, we will require:

- 1) a .docx formatted version of the manuscript text (including legends for main figures, EV figures and tables). Please make sure that the changes are highlighted to be clearly visible.
- 2) individual production quality figure files as .eps, .tif, .jpg (one file per figure). Please download our Figure Preparation Guidelines (figure preparation pdf) from our Author Guidelines pages

<https://www.embopress.org/page/journal/14693178/authorguide> for more info on how to prepare your figures.

3) a .docx formatted letter INCLUDING the reviewers' reports and your detailed point-by-point responses to their comments. As part of the EMBO Press transparent editorial process, the point-by-point response is part of the Review Process File (RPF), which will be published alongside your paper.

4) a complete author checklist, which you can download from our author guidelines (). Please insert information in the checklist that is also reflected in the manuscript. The completed author checklist will also be part of the RPF.

5) Please note that all corresponding authors are required to supply an ORCID ID for their name upon submission of a revised manuscript (). Please find instructions on how to link your ORCID ID to your account in our manuscript tracking system in our Author guidelines

()

6) We replaced Supplementary Information with Expanded View (EV) Figures and Tables that are collapsible/expandable online. A maximum of 5 EV Figures can be typeset. EV Figures should be cited as 'Figure EV1, Figure EV2' etc... in the text and their respective legends should be included in the main text after the legends of regular figures.

- For the figures that you do NOT wish to display as Expanded View figures, they should be bundled together with their legends in a single PDF file called \*Appendix\*, which should start with a short Table of Content. Appendix figures should be referred to in the main text as: "Appendix Figure S1, Appendix Figure S2" etc. See detailed instructions regarding expanded view here:

- Additional Tables/Datasets should be labeled and referred to as Table EV1, Dataset EV1, etc. Legends have to be provided in a separate tab in case of .xls files. Alternatively, the legend can be supplied as a separate text file (README) and zipped together with the Table/Dataset file.

7) Data Availability section: please add a link that resolves to a page where the data can be accessed, following the template below (see also <

<https://www.embopress.org/page/journal/14693178/authorguide#dataavailability>>).

# Data availability

The datasets produced in this study are available in the following databases:

- [data type]: [name of the resource] [accession number/identifier/doi] ([URL or identifiers.org/DATABASE:ACCESSION])

\*\*\* Note - All links should resolve to a page where the data can be accessed. \*\*\*

8) We would also encourage you to include the source data for figure panels that show essential data. Numerical data should be provided as individual .xls or .csv files (including a tab describing the data). For blots or microscopy, uncropped images should be submitted (using a zip archive if multiple images need to be supplied for one panel). Additional information on source data and instruction on how to label the files are available .

9) Our journal encourages inclusion of \*data citations in the reference list\* to directly cite datasets that were re-used and obtained from public databases. Data citations in the article text are distinct from normal bibliographical citations and should directly link to the database records from which the data can be accessed. In the main text, data citations are formatted as follows: "Data ref: Smith et al, 2001" or "Data ref: NCBI Sequence Read Archive PRJNA342805, 2017". In the Reference list, data citations must be labeled with "[DATASET]". A data reference must provide the database name, accession number/identifiers and a resolvable link to the landing page from which the data can be accessed at the end of the reference. Further instructions are available at .

10) Regarding data quantification

The following points must be specified in each figure legend:

- the name of the statistical test used to generate error bars and P values,
- the number (n) of independent experiments (please specify technical or biological replicates) underlying each data point,
- the nature of the bars and error bars (s.d., s.e.m.)

Discussion of statistical methodology can be reported in the materials and methods section, but figure legends should contain a basic description of n, P and the test applied.

- Please also include scale bars in all microscopy images.

11) As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. This File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

You are able to opt out of this by letting the editorial office know ([emboreports@embo.org](mailto:emboreports@embo.org)). If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

We would also welcome the submission of cover suggestions, or motifs to be used by our Graphics Illustrator in designing a cover.

I look forward to seeing a revised version of your manuscript when it is ready. Please let me know if you have questions or comments regarding the revision.

Yours sincerely

Martina Rembold, PhD  
Senior Editor  
EMBO reports

\*\*\*\*\*

Referee #1:

This manuscript by Liu et. al describes the use of scRNAseq to define the transcriptome of neural crest cells migrating into the heart, with analysis conducted on 8 developmental stages from E10.5-14.5, 17.5, P1 and P7. A total of 34,131 CNCC were retrieved by FACS using a Wnt1Cre and ROSA-dtTomato reporter, and single cell RNA sequencing was conducted using the 10X Genomics

platform. A key finding of this study is the contribution of CNCC derived mural cells to pericytes and microvascular smooth muscle cells in the heart, an open question that was not fully addressed in previous studies. However, they did not find any evidence for contribution of CNCC to the cardiac conduction system, another outstanding question in CNCC biology. Using pseudotime and RNA velocity computational approaches, they provided evidence indicating the CNCC derivatives had differentiated prior to arriving at the heart, and constructed differentiation pathway using TF regulatory dynamics. Overall, this is an interesting paper that has significant merit and does provide some new insights into CNCC biology, but there are some questions of overriding importance that should be carefully addressed. Some additional fine points requiring clarification are also summarized.

1. The authors isolated cells using 10 embryos for the earlier stages and 7 (14.5), 5 (17.5), 3 (P1 and P3) embryos with older stages to generate the RNASeq data comprised of 34,131 CNCC cells. This is clearly not proportional to the rate at which the heart is growing in the embryo, and thus this would suggest much of the data will likely comprised of cells from P1 and P3. While in Figures 4A/B, the authors provided the proportion of each cluster in each stage (4A) and the proportion of cells from each stage in each cluster (4B), there was no figure showing the proportion of the 34,000 cells that were analyzed at each developmental stage. If only 10 embryos are used at E10.5 to 12.5, and 7 at 13.5, and 5 at 14.5, compared to 3 at E17.5, P1 and P7, it seems likely most of the 34,000 cells will be comprised of cells from P1 and P7. This data should be shown, ie. the proportion of cells from each stage included in the 34,000 cell dataset. Thus if the cardiac conduction system represents only a very small fraction of cells in the heart as would be expected, then this might suggest the current analysis would not have the ability to detect cells of the cardiac conduction system. This needs to be carefully addressed, since this is one of the two important outstanding questions that the authors had hoped to address with their analysis.

2. The differentiation trajectory constructed in Figure 7 is potentially an important contribution of this study. The question that arises however is that since essentially all but two clusters (c.20/c.10) have contribution from the E10.5 embryos based on Figure 4B, do we then conclude that all the cells in the differentiation trajectory of Figure 7 are already present at E10.5, meaning that not only are neuronal differentiation already initiated prior to arrival in the heart, but that some of the different vascular smooth muscle and mesenchymal cells have already undergone differentiation prior to their arrival in the heart at E10.5? This can also be appreciated examining Figure 4A, as nearly all if not all of the different lineages appear to be present at E10.5.

3. Regarding the cycling score shown in the ridge plot in Figure 5, while overall panel 5A seems to make sense, drilling down into Figure 5B raises questions. For example, it is puzzling that cluster C18 which comprises mesenchymal cells that will give rise to the OFT cushion mesenchyme has the highest cycling score, while C5 with mesenchymal cells that according to the differentiation trajectory is the progenitor for all the vascular/mesenchyme lineages ranked fifth overall. How are we to interpret this cycling score, since each of these clusters are comprised of cells from all the different stages but in different proportions?

Additional points:

1. In the 139-141 line, authors said they did not find any CCS or myocardial cell clusters. In addition to the bigger question raised above, also important to address is quality control of the scRNASeq data, whether this could have led to lost/removal of rare clusters? Did the authors try different filtering conditions to examine for rare clusters that might represent small CCS or myocardial cell clusters?

2. In line 119-120 with text: "In addition, tdTomato-positive cells were found to be embedded in the walls of ventricles (Figure 1G)". Were these Wnt1-positive cells cardiomyocyte - was this checked with cardiomyocyte markers?

3. Figure. 3F, H - need legend to explain what the color gradation refer to. Should also indicate cell stages in Figure 3F and H with the pseudotime trajectory.

4. In the 250-251 line, authors stated: "Consistent with that Crabp1 is the top marker of CNCC-derived mesenchymal cells at E9.25 (Soysa et al., 2019), c5 showed high expression of Crabp1 and Crabp2 (Figure S7)". However, we did not find such information in the Soysa et al., 2019 paper.

5. Figure S1 legend indicated that the image provided has Pdgfra marking mesenchymal cells. However, the images showed no Pdgfra staining, only tdTomato and Myh11.

5. In the 268 line, path for "VSMCs of the coronary vasculature (c15->c2) said to be in Figure S7 was actually not replicated by the analysis shown in Figure S7.

Referee #2:

Liu and collaborators elaborate a very comprehensive single-cell analysis of cardiac neural crest population during embryonic development.

Using a neural crest cell specific marker and dissecting hearts, they select neural crest cells for single cell suspension and sequencing via 10X technology. Stages analyzed ranged from E10.5 to P7.

With this data, the authors are able to quantitatively describe the fates of neural crest cells within the heart. They report that they give rise to a mesenchymal lineage and vascular smooth muscle cells, but also melanocytes, mural cells and neurons. They also propose a transition from CNCC-derived pericytes to VSMC as well as the overall fate-decision stages of CNCC from the moment they enter the heart.

Using probabilistic methods, they are able to describe developmental paths followed by CNCCs, revealing the degree of plasticity within fates; which is progressively reduced with development (concomitant with loss of proliferative markers).

The authors provide a vast array of information obtained from single cell data and shed light over fate-decisions and behavior of CNCCs in a very detailed manner. Incorporation of an interactive online tool for data visualization is highly useful. Nonetheless, the results are very much descriptive in nature and therefore the tree of fate decisions proposed is not supported by experimental confirmation and remains a model to test. While the data produced will be an important results for specialists, they do not change our view of the overall contribution of the neural crest to heart development. Additional experiments would enhance the data and strengthen the manuscript.

- Analysis is performed on Wnt1-Cre population within the heart. Wnt-1 is a bona fide marker for neural crest cells and vastly used, but concerns arise when lineage tracing tools are used to define a population; both in terms of specificity and reliability. Do authors have tools to demonstrate the absence of undesired Wnt1-Cre recombination outside the neural crest lineage?

- The authors isolate cells from stages ranging from E10.5 to P7. The number of cells analyzed for several of the datasets seem insufficient. For example, authors can only identify 3 neurons cells at E10.5, for example, pooling 10 different hearts), to improve the reproducibility of the data.
- One of the major points of the study is the transition from CNCC-derived pericytes to mVSMC. Whereas data and further velocity and monocle assays point in this direction, authors analyze data only from P7 hearts to determine a transdifferentiation of the pericyte lineage. It would be very informative to see the subclustering behavior of cluster 19 across developmental stages with histological validation for expression of markers, further suggesting pericyte transdifferentiation. Moreover, a quantitative value of the amount of pericytes from CNCC that actually transdifferentiate as well as the relevance of the contribution should be assessed quantitatively.
- The authors are able to describe heterogeneity within the VSMC population arising from early stages up until P7. I find that this has not been sufficiently explored in terms of relative amount, contribution, and interaction with VSMC arising from other lineages (ie. Epicardium). Co-staining with specific markers for each subpopulation should provide more information, not only at P7 but also across development.
- The same could be said about the Tcf21 population mentioned in figure 6. How do these cells contribute in relation with the total amount of Tcf21 found in the developing and postnatal heart? Is the fate of Tcf21 cells arising from different origins different?
- Figure 7 should include the differentiation dynamics and trajectory of lineages not represented, such as neurons and melanocytes, to provide a whole, comprehensive view of CNCC population within the heart.
- Minor: Line 148: heterozygous is probably meant to have been heterogeneous.

Referee #3:

Overall, the manuscript is well written and outlines the journey a migrating CNCC takes to becoming terminally differentiated.

Liu and colleagues describe single cell transcriptional analyses of ~30,000 post-migratory cardiac neural crest cells (CNCCs) from Wnt-1-Cr/Rosa26tdTomato reporter mice. CNCCs were identified tissue dissection and FACS sorting for tdTomato + cells at six embryonic (E10.5-E17.5) and two post-natal (P1, P7) timepoints. The authors are congratulated for tackling these demanding experiments. Analyses lead to the description of CNCC clusters that contribute to - mural cells comprised pericytes and vascular smooth muscle cells (VASM; 50% of cells), mesenchymal (42% of cells), neural and Schwann cells and melanocytes. Further analyses led to 9 subclusters of VSMC and 8 subclusters of mesenchymal cells. Overall the manuscript has high quality data and the text is clearly written. Additional information would further strengthen the value of this work to the developmental community.

Major points:

1. Discuss the rationale for using a single Cre and point out limitations of this in the discussion. Wnt1-Cre is an established Cre line for CNCC cells, but there are many others (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6099459/>). Given the differences in these Cre lines,

how would this affect your current understanding of CNCC migration and differentiation?

2. More information on quality metrics is needed in methods or online sections.

- a. How were tissues from P1 and P7 selected for dissociation? Specifically, were full thickness of which vascular beds, which part of cardiac septa, etc were harvested for cell isolation.
- b. Provide numbers of mice and FACS-sorted cells obtained for each time point. How was viability of cells assessed?
- c. Provide the cutoff values or supplemental umap plots with mitochondrial counts, UMI, S-phase score, scrublet and G2M phase score. This is especially important as the manuscript discusses intermediate or transitional cell states between pericytes and mvSMCs - and these data would eliminate the possibility of doublets.
- d. Was only one replicate per sample obtained? This hinders analyses of batch effects. Please provide a PCA/UMAP plot from before/after CCA correction.

3. The text would be easier to follow if cluster numbers are provided along with marker genes. For example, Lines 130:133- "The VSMC (marked by the mature VSMC marker Myh11 and the immature VSMC marker Cxcl12) (Liu et al., 2019; Sinha et al., 2014) and mesenchymal (marked by Pdgfra and Lum) (Camp et al., 2015; Farahani and Xaymardan, 2015) lineages constituted the two largest lineages of the CNCC derivatives (accounting for 50.4% and 42.1% of the derivatives, respectively). Insert cluster numbers here and throughout the text. Additionally, an additional UMAP colored by lineage would be helpful.

4. Is the main object over-clustered at too high of a resolution? Only 30K cells were used to define 17 subcluster the comprise VASM and mesenchymal cells.

- a. Many of the cluster marker genes look indistinguishable between clusters (c1=vSMC and c3-vSMC look very similar to each other). Present markers that clearly distinguish these clusters from one another.
- b. The top markers from cluster 1 are also present in cluster 3 at high levels and vice versa.
- c. Overclustering this object may result in artificial cell populations that may skew your results. This is especially relevant for Figure 7, where many of the transcription factors in the figure have a log2FC that is negligible.

5. Are there any novel marker genes for subclusters? Most markers in the manuscript were largely previously established marker genes, which takes away from the novelty of your study. Novel marker discovery should be a distinct advantage of single cell RNA-seq.

6. Figure 4 needs supplemental information. Quantitative data is needed that informs the relative contribution of each cluster to its lineage and to the main object. Additionally, please provide a similar breakdown of each cluster by age.

7. Figure 6F: The text notes that the mesenchymal cluster c9 showed high expression of Penk. It looks as if only a subset of these cells (refer to as Penk9high) express this marker.

- a. Could this marker delineate two particular cellular states within the cluster?
- b. Alternatively, could the marker be only expressed in cluster 9 cells at a particular age (e.g. P7)? Please expand upon this point.
- c. Are there additional markers that segregate with Penk in this cluster?

Minor points:

9. The order of the figures might be better organized by starting with an overall analysis of the proportion of cell types in the main object, followed by individual analyses of clusters

10. Figure 1D-H, Figure 6 . The scale bar text is too small to read. Please increase the font size.
11. Line 118: "Compared with the aortic and pulmonary valves (Figure 1E), the CNCCs made a much smaller contribution to the atrioventricular valves (Figure 1F)". Cite evidence that CNCCs contribute to AV valves.
12. Figures 2-4: Please provide a scale for your UMAPs. Without an axis label it is difficult to tell how close these cell populations are to one another.
13. Figure S3: Please provide quantitation for the overlap between your staining over multiple slides.
14. Figure 5C: Please label the color scheme at the top of the figure.
15. Figure 6D: Please order the violin plot so that the clusters are in numerical order or grouped by lineage (clearly marked with bars)
16. Line 203-5: The intended meaning of two sentences are unclear. First states that CNCC lose proliferative markers and reach a "quiescent state with the progression of development." Next sentence defines clusters as highly proliferative. Do clusters specified in sentence 2 represent earlier developmental state? Or do they progress through development but retain proliferative capacity? Is proliferative capacity evident at P17?
17. Line 209-11: The intended meaning of two sentences seem to contradict each other.  
" We observe gradual loss of neural crest signatures with the development of CNCC derivatives  
Notably the CNCC-derived cell lineages exhibited differences in neural crest molecular signatures."
18. Line 228: Please reword this sentence: "So, c11 represents..." do not begin sentences with "So".

### Point-to-point Responses to Comments

#### Referee #1:

This manuscript by Liu et. al describes the use of scRNAseq to define the transcriptome of neural crest cells migrating into the heart, with analysis conducted on 8 developmental stages from E10.5-14.5, 17.5, P1 and P7. A total of 34,131 CNCC were retrieved by FACS using a Wnt1Cre and ROSA-dtTomato reporter, and single cell RNA sequencing was conducted using the 10X Genomics platform. A key finding of this study is the contribution of CNCC derived mural cells to pericytes and microvascular smooth muscle cells in the heart, an open question that was not fully addressed in previous studies. However, they did not find any evidence for contribution of CNCC to the cardiac conduction system, another outstanding question in CNCC biology. Using pseudotime and RNA velocity computational approaches, they provided evidence indicating the CNCC derivatives had differentiated prior to arriving at the heart, and constructed differentiation pathway using TF regulatory dynamics. Overall, this is an interesting paper that has significant merit and does provide some new insights into CNCC biology, but there are some questions of overriding importance that should be carefully addressed. Some additional fine points requiring clarification are also summarized.

1. The authors isolated cells using 10 embryos for the earlier stages and 7 (14.5), 5 (17.5), 3 (P1 and P3) embryos with older stages to generate the RNASeq data comprised of 34,131 CNCC cells. This is clearly not proportional to the rate at which the heart is growing in the embryo, and thus this would suggest much of the data will likely comprised of cells from P1 and P3. While in Figures 4A/B, the authors provided the proportion of each cluster in each stage (4A) and the proportion of cells from each stage in each cluster (4B), there was no figure showing the proportion of the 34,000 cells that were analyzed at each developmental stage. If only 10 embryos are used at E10.5 to 12.5, and 7 at 13.5, and 5 at 14.5, compared to 3 at E17.5, P1 and P7, it seems likely most of the 34,000 cells will be comprised of cells from P1 and P7. This data should be shown, ie. the proportion of cells from each stage included in the 34,000 cell dataset. Thus if the cardiac conduction system represents only a very small fraction of cells in the heart as would be expected, then this might suggest the current analysis would not have the ability to detect cells of the cardiac conduction system. This needs to be carefully addressed, since this is one of the two important outstanding questions that the authors had hoped to address with their analysis.

#### Response:

Thanks for your reviewing and comments! Given the relatively small tissue volume of the hearts in the earlier stages, we pooled cells of more embryos (10 embryos) to meet the cell number requirement of cell capture. However, the captured cell number from each sample (stage) is dependent on the number of cells loaded on the machine, rather than the number of pooled embryos. Actually, the cells from P1 and P7 (It is P7, and no P3 cells were considered in this study) only account for 21.7% of the total number of cells (34,131). In this revision, we added “Number of Cells Passing QC” for each sample in Table EV1. Consistent with the conclusions of many

previous reports based on staining, our data do not support the direct contribution of CNCC to the cardiac conduction system (CCS). We do not think that this resulted from the small number of “CNCC-derived CCS cells”, since we successfully captured rare cell population such as CNCC-derived pericytes, and CNCC-derived neurons at E10.5 (only 6 cells). In addition, although the cell number of each stage is not very large, we analyzed the combined data from all the stages, and each stage contains nearly all the cell states due to developmental asynchrony. This made the rare cellular states detectable by our dataset. Consistent with the single-cell data, our staining results do not support the direct contribution of CNCC to CCS cells, which are specialized cardiomyocytes (Figure EV2).

2. The differentiation trajectory constructed in Figure 7 is potentially an important contribution of this study. The question that arises however is that since essentially all but two clusters (c.20/c.10) have contribution from the E10.5 embryos based on Figure 4B, do we then conclude that all the cells in the differentiation trajectory of Figure 7 are already present at E10.5, meaning that not only are neuronal differentiation already initiated prior to arrival in the heart, but that some of the different vascular smooth muscle and mesenchymal cells have already undergone differentiation prior to their arrival in the heart at E10.5? This can also be appreciated examining Figure 4A, as nearly all if not all of the different lineages appear to be present at E10.5.

#### **Response:**

In Figure 7, we reconstructed the differentiation path and uncovered the relative position of each cluster along the path. In Figure 4A and 4B, we observed that cells of many clusters were present at the earliest stage E10.5, which indicates that many (not all) CNCC-derivatives had already committed or differentiated to a specific lineage including vascular smooth muscle and mesenchymal cells when they arrived at the heart. We think that this could be explained by developmental asynchrony which means that cells in differentiation do not progress in perfect asynchrony. Therefore, even in the earliest stage sample, it included cells of later clusters along the differentiation path. However, the undifferentiated cells still accounted for a large proportion of cells at the early stages. For example, the mesenchymal cluster c5, which was inferred to represent the earliest state of the CNCC-derived mesenchymal cells after migrating into the heart (Figure 7), comprised cells mainly from the early stages like E10.5 (Figure 4B).

3. Regarding the cycling score shown in the ridge plot in Figure 5, while overall panel 5A seems to make sense, drilling down into Figure 5B raises questions. For example, it is puzzling that cluster C18 which comprises mesenchymal cells that will give rise to the OFT cushion mesenchyme has the highest cycling score, while C5 with mesenchymal cells that according to the differentiation trajectory is the progenitor for all the vascular/mesenchyme lineages ranked fifth overall. How are we to interpret this cycling score, since each of these clusters are comprised of cells from all the different stages but in different proportions?

#### **Response:**

The cycling score reflects proliferative activity of each cell. In Figure 5B, although c18 ranked the first instead of c5 according to the density distribution of the cycling score, we think that it still makes sense, because c18, c12, c8 and c5, the top four clusters, all represent early states of CNCC

derivatives based on the developmental paths shown in Figure 7, and the differences of the cycling score distributions of these four clusters are actually not obvious. Biologically, it may be explained by that c18 will give rise to OFT cushion mesenchyme, which contains a large number of cells, and thus needs high proliferative activity. Each of the clusters are comprised of cells from all the stages due to developmental asynchrony. We interpreted the cycling score by comparing the score distribution of all cells instead of individual cells.

Additional points:

1. In the 139-141 line, authors said they did not find any CCS or myocardial cell clusters. In addition to the bigger question raised above, also important to address is quality control of the scRNASeq data, whether this could have led to lost/removal of rare clusters? Did the authors try different filtering conditions to examine for rare clusters that might represent small CCS or myocardial cell clusters?

**Response:**

Just as the conclusions of many previous studies, our current dataset does not support any myocardial or CCS (specialized myocardial cells) clusters. Although different quality filtering thresholds would filter out different number of cells, normally they would lead to the filtering of some relatively low quality cells, rather than the removal of a whole cluster. Given that the rare cell population such as CNCC-derived pericytes and CNCC-derived neurons at E10.5 could be retained after quality filtering, we do not think that CNCC-derived myocardial cells could get lost due to filtering. In this revision, we performed immunofluorescence staining to examine whether the tdTomato-positive cells embedded in the walls of ventricles could be co-stained by the cardiomyocyte marker cTnT (Figure EV2). Consistent with the single-cell data, we could not find any co-stained cells. Nevertheless, we don't deny the presence of direct or indirect contribution of CNCC to myocardial or CCS cells, which may be dependent on driver lines, recombination systems and other factors. Given the large number of myocardial cells in the heart, we hold the view that the contribution of CNCC to cardiomyocytes, if exists, would be very limited and negligible.

2. In line 119-120 with text: "In addition, tdTomato-positive cells were found to be embedded in the walls of ventricles (Figure 1G)". Were these Wnt1-positive cells cardiomyocyte - was this checked with cardiomyocyte markers?

**Response:**

In this revision, we performed immunofluorescence staining to check whether the tdTomato-positive cells embedded in the walls of ventricles could be co-stained with the cardiomyocyte marker, i.e., cTnT. As shown in Figure EV2, the tdTomato-positive cells embedded in the walls of ventricles were not co-stained by cTnT. They just represent CNCC-derived cells, for example, VSMCs, pericytes or mesenchymal cells, rather than cardiomyocytes.

3. Figure. 3F, H - need legend to explain what the color gradation refer to. Should also indicate cell stages in Figure 3F and H with the pseudotime trajectory.

**Response:**

As suggested, we added "Scale expression" in the legends of Figure 3F and 3H to explain the color gradation. We also added a density curve above the figures to indicate the cell subpopulations with the pseudotime.

4. In the 250-251 line, authors stated: "Consistent with that *Crabp1* is the top marker of CNCC-derived mesenchymal cells at E9.25 (Soysa et al., 2019), c5 showed high expression of *Crabp1* and *Crabp2* (Figure S7)". However, we did not find such information in the Soysa et al., 2019 paper.

**Response:**

Please see Figure 1d and 1e of Soysa et al. (2019)'s paper. For convenience, we attached the figure below. As indicated in the legend, the NC cluster refers to neural crest-derived mesenchyme, which is mainly from E9.25 (1d). The top marker of NC is *Crabp1* (1e).

[Figure 1d, e from de Soysa TY *et al.* (2019), Nature 572: 120-124; not shown]

5. Figure S1 legend indicated that the image provided has *Pdgfra* marking mesenchymal cells. However, the images showed no *Pdgfra* staining, only tdTomato and Myh11.

**Response:**

Please see the lower right panel of Appendix Figure S1. *Pdgfra*-expressing cells were stained in blue.

5. In the 268 line, path for "VSMCs of the coronary vasculature (c15->c2) said to be in Figure S7 was actually not replicated by the analysis shown in Figure S7.

**Response:**

We think that it may be partially affected by the limitation of visualization in a 2D UMAP space.

Nevertheless, in this revision, we removed “VSMCs of the coronary vasculature (c15->c2)” from the text.

## **Referee #2:**

Liu and collaborators elaborate a very comprehensive single-cell analysis of cardiac neural crest population during embryonic development.

Using a neural crest cell specific marker and dissecting hearts, they select neural crest cells for single cell suspension and sequencing via 10X technology. Stages analyzed ranged from E10.5 to P7.

With this data, the authors are able to quantitatively describe the fates of neural crest cells within the heart. They report that they give rise to a mesenchymal lineage and vascular smooth muscle cells, but also melanocytes, mural cells and neurons. They also propose a transition from CNCC-derived pericytes to VSMC as well as the overall fate-decision stages of CNCC from the moment they enter the heart.

Using probabilistic methods, they are able to describe developmental paths followed by CNCCs, revealing the degree of plasticity within fates; which is progressively reduced with development (concomitant with loss of proliferative markers).

The authors provide a vast array of information obtained from single cell data and shed light over fate-decisions and behavior of CNCCs in a very detailed manner. Incorporation of an interactive online tool for data visualization is highly useful. Nonetheless, the results are very much descriptive in nature and therefore the tree of fate decisions proposed is not supported by experimental confirmation and remains a model to test. While the data produced will be an important result for specialists, they do not change our view of the overall contribution of the neural crest to heart development. Additional experiments would enhance the data and strengthen the manuscript.

- Analysis is performed on Wnt1-Cre population within the heart. Wnt-1 is a bona fide marker for neural crest cells and vastly used, but concerns arise when lineage tracing tools are used to define a population; both in terms of specificity and reliability. Do authors have tools to demonstrate the absence of undesired Wnt1-Cre recombination outside the neural crest lineage?

## **Response:**

Thank you very much for your reviewing and comments! We agree that lineage tracing models sometimes may cause undesired recombination. Given that *Wnt1-Cre* mice have been widely used in CNCC studies since the first CNCC research using *Wnt1-Cre* by Jiang et al. (2000), our study provides single-cell resolution dataset based on this classic model. Currently, our lab has no tools or additional tracing models to demonstrate the absence of undesired recombination. However, we think that it is a meaning topic for further investigation but it is out of the scope of the current study. In this revision, we added related discussion (please see Line 360-363).

- The authors isolate cells from stages ranging from E10.5 to P7. The number of cells analyzed for

several of the datasets seem insufficient. For example, authors can only identify 3 neurons cells at E10.5, for example, pooling 10 different hearts), to improve the reproducibility of the data.

**Response:**

For some CNCC-derived lineages, for example, neurons, the number of cells is rare at early stages. While previously studies failed to identify CNCC-derived neurons at E10.5 (the earliest stage previously reported is E11.5) (Hildreth et al., 2008; Poelmann et al., 2004), our single-cell dataset captured CNCC-derived neurons at E10.5 (6 cells). On one hand, it suggests that CNCC-derived neurons are indeed rare at the early developmental stages. On the other hand, it reflects that the number of cells in our dataset is sufficient to identify such rare cell population. In addition, although the cell number of each stage is not very large, we analyzed the combined data from all the stages, and each stage contains nearly all the cell states due to developmental asynchrony. This made the rare cellular states detectable by our dataset.

- One of the major points of the study is the transition from CNCC-derived pericytes to mVSMC. Whereas data and further velocity and monocle assays point in this direction, authors analyze data only from P7 hearts to determine a transdifferentiation of the pericyte lineage. It would be very informative to see the subclustering behavior of cluster 19 across developmental stages with histological validation for expression of markers, further suggesting pericyte transdifferentiation. Moreover, a quantitative value of the amount of pericytes from CNCC that actually transdifferentiate as well as the relevance of the contribution should be assessed quantitatively.

**Response:**

In this revision, we performed immunofluorescence staining of the hearts with the pericyte marker Kcnj8 and tdTomato across all the studied stages (Appendix Figure S5). As shown in the figure, CNCC-derived pericytes were present at all the studied stages. However, we think that it is difficult to perform accurate quantitative analysis of the amount of pericytes that transdifferentiated based on staining. The reasons are as follows: 1) there is no markers clearly distinguishing the transdifferentiating pericytes from others; 2) the staining results only reflect the cell states of a 2D section, whereas the distribution of CNCC-derived pericytes is not spatially homogenous in a 3D heart.

- The authors are able to describe heterogeneity within the VSMC population arising from early stages up until P7. I find that this has not been sufficiently explored in terms of relative amount, contribution, and interaction with VSMC arising from other lineages (ie. Epicardium). Co-staining with specific markers for each subpopulation should provide more information, not only at P7 but also across development.

**Response:**

To explore the relationship between CNCC-derived VSMCs and VSMCs derived from other lineages, in this revision, we performed immunofluorescence staining of the hearts with Cxcl12 (immature VSMC marker), Myh11 (mature VSMC marker) and td-Tomato (CNCC derivative marker) across all the studied stages. As shown in Appendix Figure S3, CNCC-derived VSMCs

mainly contribute to the structures close to the arterial pole of the heart such as OFT (early stages), the base of the great arteries (later stages), and coronary arteries. For OFT and the base of the great arteries, CNCC-derived VSMCs mainly populated the inner media, while VSMCs of other embryonic origins such as SHF-derived VSMCs populated the outer media. Although we added a figure (Figure EV3) to clearly show the subpopulation-specific markers, we found that it is difficult to co-staining with the markers for each subpopulation, since many of these subpopulations represent a continuum of cellular states and they are hard to be clearly distinguished from one another by staining with a few markers. Instead, we focused on some subpopulations of interest, for example, we added staining results of the *Tcf21*<sup>high</sup> *tdTomato*<sup>+</sup> cells across all the studied stages (Appendix Figure S8).

- The same could be said about the *Tcf21* population mentioned in figure 6. How do these cells contribute in relation with the total amount of *Tcf21* found in the developing and postnatal heart? Is the fate of *Tcf21* cells arising from different origins different?

**Response:**

In this revision, to explore the relationship of CNCC-derived *Tcf21*<sup>high</sup> cells and *Tcf21*-expressing cells derived from other sources. we added staining results of the *Tcf21*<sup>high</sup> *tdTomato*<sup>+</sup> cells across all the studied stages (Appendix Figure S8). As shown in Appendix Figure S8 and Figure 6E, the *Tcf21*<sup>high</sup> *tdTomato*<sup>+</sup> cells were mainly localized in the proximal regions of the OFT where the aortic and pulmonary valves were formed at the early stages E10.5-E12.5, and they were mainly localized in the valve annulus region at the late stages E13.5-P7. In contrast, the *Tcf21*<sup>high</sup> *tdTomato*<sup>-</sup> cells mainly populated the epicardium. So we think that generally the fate of *Tcf21*-expressing cells of different origins is different.

- Figure 7 should include the differentiation dynamics and trajectory of lineages not represented, such as neurons and melanocytes, to provide a whole, comprehensive view of CNCC population within the heart.

**Response:**

Given that the cell fate bifurcations of lineages including neuron, glia and mesenchyme are reported to have occurred during early migration of neural crest cells (Soldatov et al., 2019), here we focused on the mesenchymal and VSMC lineages which differentiate mainly after migrating into the heart. Using a probabilistic method, the relative positions along the developmental path were inferred for CNCC-derived mesenchymal and VSMC lineages, which are complex and have many subpopulations.

- Minor: Line 148: heterozygous is probably meant to have been heterogeneous.

**Response:**

Sorry! It is corrected.

### Referee #3:

Overall, the manuscript is well written and outlines the journey a migrating CNCC takes to becoming terminally differentiated.

Liu and colleagues describe single cell transcriptional analyses of ~30,000 post-migratory cardiac neural crest cells (CNCCs) from *Wnt1-Cr/Rosa26dtTomato* reporter mice. CNCCs were identified tissue dissection and FACS sorting for tdTomato + cells at six embryonic (E10.5-E17.5) and two post-natal (P1, P7) timepoints. The authors are congratulated for tackling these demanding experiments. Analyses lead to the description of CNCC clusters that contribute to - mural cells comprised pericytes and vascular smooth muscle cells (VASM; 50% of cells), mesenchymal (42% of cells), neural and Schwann cells and melanocytes. Further analyses led to 9 subclusters of VSCM and 8 subclusters of mesenchymal cells. Overall the manuscript has high quality data and the text is clearly written. Additional information would further strengthen the value of this work to the developmental community.

Major points:

1. Discuss the rationale for using a single Cre and point out limitations of this in the discussion. *Wnt1-Cre* is an established Cre line for CNCC cells, but there are many others (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6099459/>). Given the differences in these Cre lines, how would this affect your current understanding of CNCC migration and differentiation?

### Response:

Thank you very much for your reviewing and comments! Given that *Wnt1-Cre* mice have been widely used in CNCC studies since the first CNCC research using *Wnt1-Cre* by Jiang et al. (2000), our study provides single-cell resolution dataset and staining results based on this classic model, which facilitates comparing the findings with that of previous studies and the future reuse of the dataset by the community. We agree that phenotypic variances may exist among different tracing models. It is therefore meaningful to used additional driver lines (e.g., *P0-Cre* and *Pax3-Cre*) or a different recombination system such as *Wnt1-Flpe* to confirm the findings of this study in the future. However, it may be out of the scope of the current study. In this revision, we added related discussion to point out the potential limitation and cited the mentioned paper (Line 360-363).

2. More information on quality metrics is needed in methods or online sections.

a. How were tissues from P1 and P7 selected for dissociated? Specifically, were full thickness of which vascular beds, which part of cardiac septa, etc were harvested for cell isolation.

b. Provide numbers of mice and FACS-sorted cells obtained for each time point. How was viability of cells assessed?

c. Provide the cutoff values or supplemental umap plots with mitochondrial counts, UMI, S-phase score, scrublet and G2M phase score. This is especially important as the manuscript discusses intermediate or transitional cell states between pericytes and mvSMCs - and these data would

eliminate the possibility of doublets.

d. Was only one replicate per sample obtained? This hinders analyses of batch effects. Please provide a PCA/UMAP plot from before/after CCA correction.

**Response:**

a. As indicated in the Methods, the whole heart tissues (the base of the great arteries, atrial septum and ventricular septum were included) from P1 and P7 were harvested for dissociation, and the tissues were not selected. Compared with the embryonic heart tissues, measures were taken to avoid inadequate enzymatic digestion of the tissues especially from these two neonatal stages. The tissues were carefully minced into small pieces in dishes containing cold HBSS. Then, the tissues were shaken every ten minutes during the digestion process.

b. The number of mice for each stage was shown in Figure 1B. As suggested, we added the number of FACS-sorted cells in Table EV1. Please note that the number of FACS-sorted cells from E17.5, P1, and P7 does not represent all the tdTomato-positive cells in the heart tissues, as we only isolate enough number of cells for experiments.

Cell viability and concentrations were measured with a TC20 automated cell counter (Bio-Rad, USA). Cell viability was determined with 0.4% trypan blue staining solution. This information was added in the Methods.

c. As suggested, we added UMAP plots (Figure EV1) showing UMI count, gene count, S phase score, G2M phase score, proportion of the UMI count of mitochondrial genes and Scrublet doublet score. As shown in the Scrublet doublet score plot, the mural cells in cluster c19 did not have high doublet scores, thus excluding the possibility of doublets.

d. Yes, we only have one replicate per sample. We added UMAP plots before/after CCA integration (Figure EV1). Compared with the UMAP plot after CCA integration, the cells from different samples were aligned discretely in the UMAP space before CCA integration, reflecting the importance of CCA integration.

3. The text would be easier to follow if cluster numbers are provided along with marker genes. For example, Lines 130:133- "The VSMC (marked by the mature VSMC marker *Myh11* and the immature VSMC marker *Cxcl12*) (Liu et al., 2019; Sinha et al., 2014) and mesenchymal (marked by *Pdgfra* and *Lum*) (Camp et al., 2015; Farahani and Xaymardan, 2015) lineages constituted the two largest lineages of the CNCC derivatives (accounting for 50.4% and 42.1% of the derivatives, respectively). Insert cluster numbers here and throughout the text. Additionally, an additional UMAP colored by lineage would be helpful.

**Response:**

Thanks for your suggestion! We added cluster numbers along with marker genes throughout the text. For example, "The VSMC (c1, c2, c3, c4, c7, c8, c14, c15 and c16; marked by the mature VSMC marker *Myh11*". As suggested, we added a UMAP plot colored by lineage (the small panel in Figure 2B).

4. Is the main object over-clustered at too high of a resolution? Only 30K cells were used to define 17 subcluster the comprise VASM and mesenchymal cells.

a. Many of the cluster marker genes look indistinguishable between clusters (c1=vSMC and c3-vSMC look very similar to each other). Present markers that clearly distinguish these clusters from one another.

b. The top markers from cluster 1 are also present in cluster 3 at high levels and vice versa.

c. Overclustering this object may result in artificial cell populations that may skew your results. This is especially relevant for Figure 7, where many of the transcription factors in the figure have a log2FC that is negligible.

#### **Response:**

a. We think that the number of subclusters is highly dependent on the complexity and heterogeneity of captured cells, rather than the number of cells. Given the high complexity of CNCC-derived VSMC and mesenchymal lineages (which involved various tissue structures such as OFT cushion mesenchyme, valves, the great arteries, and the coronary vasculature as well as multiple differentiation stages spanning embryonic and neonatal stages), we think that it makes sense to have 17 subclusters. In this revision, we added a supplemental figure (Figure EV3) to show markers that distinguish the VSMC subclusters as well as the mesenchymal clusters.

b. The VSMC cluster c1 and c3 are closely related, and they are both VSMCs of the great arteries. Nevertheless, as shown in Figure EV3, they are distinguishable by the expression levels of markers such as *Bcam*, *Cebpb* and *Ppp1r15a*. Based on our analysis in Figure 7, c3 may represent a more mature state.

c. As shown in Figure EV3, these clusters are distinguishable from one another. Therefore, we do not think that overclustering exists in our case. Figure 7 shows key transcription factors, of which the expression changes normally have great impacts to the transcriptome. That also means that they do not necessarily have remarkable changes in expression to alter the transcriptomic program.

5. Are there any novel marker genes for subclusters? Most markers in the manuscript were largely previously established marker genes, which takes away from the novelty of your study. Novel marker discovery should be a distinct advantage of single cell RNA-seq.

#### **Response:**

We used previously established marker genes to assign cells cellular identity, which made our study linked with previous knowledge. But that does not mean that the novelty is affected. Many subpopulations identified in this study have not identified before. For example, the markers shown in Figure EV3 may represent novel knowledge about the signature of CNCC derivatives.

6. Figure 4 needs supplemental information. Quantitative data is needed that informs the relative contribution of each cluster to its lineage and to the main object. Additionally, please provide a similar breakdown of each cluster by age.

**Response:**

As suggested, we added plots (Figure EV4A) to show the relative contribution of each cluster to its lineage (VSMC and mesenchymal lineages) and to all the captured cells (the main object). We also added plots to show the relative contribution of each cluster to its lineage by stage (Figure EV4B).

7. Figure 6F: The text notes that the mesenchymal cluster c9 showed high expression of *Penk*. It looks as if only a subset of these cells (refer to as *Penk*<sup>high</sup>) express this marker.

- a. Could this marker delineate two particular cellular states within the cluster?
- b. Alternatively, could the marker be only expressed in cluster 9 cells at a particular age (e.g. P7)? Please expand upon this point.
- c. Are there additional markers that segregate with *Penk* in this cluster?

**Response:**

a. Thanks for your comments! In this revision, we performed secondary clustering of c9. As shown in Figure EV5A, it could be clustered into 3 subclusters, sc0, sc1 and sc2. While sc1 and sc0 expressed high levels of *Penk*, the expression of *Penk* in sc2 is relatively low (Figure EV5C).

b. As shown in Figure EV5B and Figure 4B, *Penk* is relatively specifically expressed in embryonic stages when the aortopulmonary septum of the OFT was formed, i.e., E13.5-E14.5. The cells in the subcluster sc2 are mainly from early stages (E10.5-E12.5), which may represent progenitor states of the *Penk*<sup>high</sup> cells.

c. Yes. As shown in Figure EV5C and D, we found that the expression pattern of *Sfrp2* (Secreted Frizzled Related Protein 2) is most like that of *Penk*.

Minor points:

9. The order of the figures might be better organized by starting with an overall analysis of the proportion of cell types in the main object, followed by individual analyses of clusters

**Response:**

We generally organize the results as overall analysis followed individual clusters. However, we identified the presence of CNCC-derived pericytes when assigning the clusters lineage identity (Figure 2). And we think that it is better to put the results of pericytes-to-SMC transition (Figure 3) just after Figure 2, rather than at the end of manuscript.

10. Figure 1D-H, Figure 6. The scale bar text is too small to read. Please increase the font size.

**Response:**

We removed all the scale bar text on the figures. Instead, we added scale bar information in the legends.

11. Line 118: "Compared with the aortic and pulmonary valves (Figure 1E), the CNCCs made a much smaller contribution to the atrioventricular valves (Figure 1F)". Cite evidence that CNCCs contribute to AV valves.

**Response:**

We added the reference supporting the contribution of CNCCs to AV valves based on staining (please see Line 122).

12. Figures 2-4: Please provide a scale for your UMAPs. Without an axis label it is difficult to tell how close these cell populations are to one another.

**Response:**

Axis labels were added in Figure2-4.

13. Figure S3: Please provide quantitation for the overlap between your staining over multiple slides.

**Response:**

Thanks for your suggestion! Appendix Figure S4 (Figure S3 in the previous version) shows the results of smFISH validation of CNCC-derived melanocytes (marked by Dct) in aortic valves (Figure S4A) and mitral valves (Figure S4B). In this revision, we added more results to demonstrate the presence of CNCC-derived melanocytes in tricuspid valves (Figure S4C) and pulmonary valves (Figure S4D). We think that the results shown here are sufficient to prove the presence of CNCC-derived melanocytes in valves and the quantitation may not be necessary.

14. Figure 5C: Please label the color scheme at the top of the figure.

**Response:**

The color scheme was labelled in Figure 5C.

15. Figure 6D: Please order the violin plot so that the clusters are in numerical order or grouped by lineage (clearly marked with bars)

**Response:**

Now the clusters in the violin plots are in a numerical order.

16. Line 203-5: The intended meaning of two sentences are unclear. First states that CNCC lose proliferative markers and reach a "quiescent state with the progression of development." Next sentence defines clusters as highly proliferative. Do clusters specified in sentence 2 represent earlier developmental state? Or do they progress through development but retain proliferative capacity? Is proliferative capacity evident at P17?

**Response:**

Yes, these clusters (c18, c12, c8, c5 and c7) represent earlier developmental states as demonstrated

by the relative proportion analysis shown in Figure 4B (they are mainly from stages E10.5-E13.5), as well as the developmental path analysis shown in Figure 7. The proliferative capacity of CNCC derivatives at neonatal stages P1 and P7 is greatly decreased compared with the embryonic stages, which is shown by the Cycling Score in Figure 5A and G2M/S score in Figure EV1.

17. Line 209-11: The intended meaning of two sentences seem to contradict each other." We observe gradual loss of neural crest signatures with the development of CNCC derivatives (Figure 5D). Notably the CNCC-derived cell lineages exhibited differences in neural crest molecular signatures (Figure 5E)."

**Response:**

Figure 5D shows the gradual loss of neural crest signatures with the developmental stages. Figure 5E shows the differences in neural crest molecular signatures among lineages/subclusters. Thus, the two sentences do not contradict each other.

18. Line 228: Please reword this sentence: "So, c11 represents..." do not begin sentences with "So".

**Response:**

We reworded the sentence as "The cluster c11 thus represents neither contaminated epicardial cells nor epicardium-derived cells."

Dear Prof. Zhou

Thank you for the submission of your revised manuscript to EMBO reports. We have now received the full set of referee reports that is copied below.

As you will see, referee 2 still considers the study descriptive but also of value to scientists working in the cardiac neural crest field. Since your study will be published in our Resource section, since no major technical concerns remain and since referee 3 supports publication, we have decided to move forward. Both referees remain concerned about the use of a single Cre line and this limitation should be clearly stated and discussed in the manuscript. Please also provide the uMAP plots and address point 3 of referee 3.

From the editorial side, there are also a few things that we need before we can proceed with the official acceptance of your study.

- Please reduce the number of keywords to 5.
- Author contributions: Please change the abbreviations for Z Zhou to ZZh and for Z Zeng to ZZe.
- Please add all information on funding to the relevant fields in our online submission system. This section is currently incomplete if compared to the information given in the manuscript.
- Please add a callout to Fig 3C and callouts to the panels of Fig EV3, EV4, EV5, Appendix Fig S2, S3, S4, S5, S6, S8, and S9 in the text where appropriate.
- Author checklist: Please complete section D - 9 with information on the committee approving the animal experiments or indicate where in the manuscript this information is given (currently page 20). N/A does not apply here. Please also check compliance with the ARRIVE guidelines (D-10).
- Methods: Please double-check whether the term "current water" is correct (page 23, line 577).
- Figure legends:  
Please use the header "Data information" if the description relates to several panels in the figure, e.g., the definition of abbreviations used.  
Figure EV2: Please remove the numbers from the scale bars and define their size in the figure legend.
- Appendix:  
Please specify the size of scale bars in the figure legends instead of in the figure itself for Fig. S, S2, S3, S4, S5, S8  
Fig. S2: I suggest listing abbreviations and the color code for stainings separately. Currently these are sorted alphabetically and this mixture is confusing.  
Fig. S4: please define the meaning of the arrows. The arrows in panel C and D could be enlarged to make them more visible.  
May I suggest moving Figure S7 to its own page together with its legend on page 8?
- Please call Tables EV2, EV3, EV6 & EV7 'DatasetEV#'. The remaining EV tables need renumbering to Table EV1 - EV3 and the callouts in the text should be corrected.

- I made some minor changes to the abstract and also rewrote it in present tense. Please see my suggestion below my signature.

- Finally, EMBO reports papers are accompanied online by A) a short (1-2 sentences) summary of the findings and their significance, B) 2-3 bullet points highlighting key results and C) a synopsis image that is 550x200-600 pixels large (width x height) in .png format. You can either show a model or key data in the synopsis image. Please note that the size is rather small and that text needs to be readable at the final size. Please send us this information along with the revised manuscript.

We look forward to seeing a final version of your manuscript as soon as possible.

Yours sincerely,

Martina Rembold, PhD  
Senior Editor  
EMBO reports

\*\*\*\*\*

## Abstract

The migratory cardiac neural crest cells (CNCCs) contribute greatly to cardiovascular development. A thorough understanding of the cell lineages, developmental chronology and transcriptomic states of CNCC derivatives during normal development is essential for deciphering the pathogenesis of CNCC-associated congenital anomalies. Here, we perform single-cell transcriptomic sequencing of 34,131 CNCC-derived cells from mouse hearts covering eight developmental stages between E10.5 and P7. We report the presence of CNCC-derived mural cells that comprise pericytes and microvascular smooth muscle cells (mVSMCs). Furthermore, we identify the transition from the CNCC-derived pericytes to mVSMCs, and identify potential key regulators of this transition. In addition, our data support that many CNCC-derivatives had already committed or differentiated to a specific lineage when migrating into the heart. We explore the spatial distribution of some critical CNCC-derived subpopulations with single-molecule fluorescence in situ hybridization. Finally, we computationally reconstruct the differentiation path and regulatory dynamics of CNCC derivatives. Our study provides novel insights into the cell lineages, developmental chronology and regulatory dynamics of CNCC derivatives during development.

\*\*\*\*\*

## Referee #2:

The authors have answered several of my concerns, however the study remains descriptive and without major discoveries or genetic confirmation of the findings. Furthermore, the use of a single NC tracer is an important limitation of the study. I think this work will be useful as a resource for scientists working in the cardiac neural crest field.

## Referee #3:

The resubmission of the manuscript by Liu and colleagues provide additional information about experimental procedures, quality control, and clustering. These data have elevated the manuscript's quality and minimizes concerns regarding small clusters and that analyses of trajectories are driven by one particular cluster or sample. The authors' justification of the large number of clusters derived from a small number of cells by providing marker genes that isolate each individual cluster, provides potential new markers for individual cells and states. While these changes address the majority of the major concerns in this manuscript, a few additional textual changes would be appropriate.

1. In the discussion the author should specifically address that using a single Cre has limitations as data derived from one Cre line might subsets of migrating and differentiating neural crest cells.
2. The addition of information on quality metrics provide in this revision are helpful. However, more would be better. Please consider adding uMAP plots before and after CCA integration as well as scrublet score to provide clear information on doublets and batch effect.
3. The authors addressed the concern about over clustering by providing marker genes (Figure VE3). Specifying in the text if some of these genes are novel markers of cell states and information about their functions would be helpful.

### **Point-to-point Responses to Comments**

#### **Response to the Editor:**

Thank you for the submission of your revised manuscript to EMBO reports. We have now received the full set of referee reports that is copied below.

As you will see, referee 2 still considers the study descriptive but also of value to scientists working in the cardiac neural crest field. Since your study will be published in our Resource section, since no major technical concerns remain and since referee 3 supports publication, we have decided to move forward. Both referees remain concerned about the use of a single Cre line and this limitation should be clearly stated and discussed in the manuscript. Please also provide the uMAP plots and address point 3 of referee 3.

From the editorial side, there are also a few things that we need before we can proceed with the official acceptance of your study.

- Please reduce the number of keywords to 5.

#### **Response:**

Thank you very much for your suggestion! The number of keywords has been reduced to 5 (Line 27).

- Author contributions: Please change the abbreviations for Z Zhou to ZZh and for Z Zeng to ZZe.

#### **Response:**

They were corrected (Line 495-497).

- Please add all information on funding to the relevant fields in our online submission system. This section is currently incomplete if compared to the information given in the manuscript.

**Response:**

As suggested, we added all information on funding to the relevant fields in the online submission system.

- Please add a callout to Fig 3C and callouts to the panels of Fig EV3, EV4, EV5, Appendix Fig S2, S3, S4, S5, S6, S8, and S9 in the text where appropriate.

**Response:**

We added callouts for Fig 3C (Line 165), Fig EV3A and B (Line 138), Fig EV4A and B (Line 193-194), Fig EV5A-D (Line 246-247), Appendix Fig S2A and B (Line 114), Appendix Figure S3A-H (Line 117), Appendix Figure S4A, B, C and D (Line 147), Appendix Figure S5A-H (Line 158), Appendix Figure S6A and B (Line 222), Appendix Figure S8A-H (Line 239), as well as Appendix Figure S9A and B (Line 264).

- Author checklist: Please complete section D - 9 with information on the committee approving the animal experiments or indicate where in the manuscript this information is given (currently page 20). N/A does not apply here. Please also check compliance with the ARRIVE guidelines (D-10).

**Response:**

We added the committee approving the animal experiments in the D-9 of the Author Checklist. All our animal experiments complied with the ARRIVE guidelines. We added a statement in D-10 and the manuscript (Line 381-382).

- Methods: Please double-check whether the term "current water" is correct (page 23, line 577).

**Response:**

It is correct.

- Figure legends:

Please use the header "Data information" if the description relates to several panels in the figure, e.g., the definition of abbreviations used.

**Response:**

“Data information” was added in legends of related figures.

Figure EV2: Please remove the numbers from the scale bars and define their size in the figure legend.

**Response:**

For Figure EV2, the numbers were removed from the scale bars and the size was defined in the figure legend.

- Appendix:

Please specify the size of scale bars in the figure legends instead of in the figure itself for Fig. S1, S2, S3, S4, S5, S8

**Response:**

For Appendix Fig S1, S2, S3, S4, S5, and S8, the numbers were removed from the scale bars and the size was defined in the figure legend.

Fig. S2: I suggest listing abbreviations and the color code for stainings separately. Currently these are sorted alphabetically and this mixture is confusing.

**Response:**

As suggested, we listed the abbreviations and the color code for stainings separately for the legend of Fig. S2.

Fig. S4: please define the meaning of the arrows. The arrows in panel C and D could be enlarged to make them more visible.

**Response:**

The arrows were defined in the legend of Fig. S4. The arrows in panel C and D were enlarged.

May I suggest moving Figure S7 to its own page together with its legend on page 8?

**Response:**

As suggested, we moved Figure S7 to its own page together with its legend.

- Please call Tables EV2, EV3, EV6 & EV7 'DatasetEV#'. The remaining EV tables need renumbering to Table EV1 - EV3 and the callouts in the text should be corrected.

**Response:**

As suggested, we renamed the Tables to Dataset EV1 – EV4 and Table EV1 – EV3. The callouts in the text were corrected accordingly.

- I made some minor changes to the abstract and also rewrote it in present tense. Please see my suggestion below my signature.

**Abstract**

The migratory cardiac neural crest cells (CNCCs) contribute greatly to cardiovascular development. A thorough understanding of the cell lineages, developmental chronology and transcriptomic states of CNCC derivatives during normal development is essential for deciphering the pathogenesis of CNCC-associated congenital anomalies. Here, we perform single-cell transcriptomic sequencing of 34,131 CNCC-derived cells from mouse hearts covering eight developmental stages between E10.5 and P7. We report the presence of CNCC-derived mural cells that comprise pericytes and microvascular smooth muscle cells (mVSMCs). Furthermore, we identify the transition from the CNCC-derived pericytes to mVSMCs, and identify potential key regulators of this transition. In addition, our data support that many CNCC-derivatives had already committed or differentiated to a specific lineage when migrating into the heart. We explore the spatial distribution of some critical CNCC-derived subpopulations with single-molecule fluorescence in situ hybridization. Finally, we computationally reconstruct the differentiation path and regulatory dynamics of CNCC derivatives. Our study provides novel insights into the cell lineages, developmental chronology and regulatory dynamics of CNCC derivatives during development.

**Response:**

Thanks a lot for rewriting the abstract! We made changes according to your suggestion.

- Finally, EMBO reports papers are accompanied online by A) a short (1-2 sentences)

summary of the findings and their significance, B) 2-3 bullet points highlighting key results and C) a synopsis image that is 550x200-600 pixels large (width x height) in .png format. You can either show a model or key data in the synopsis image. Please note that the size is rather small and that text needs to be readable at the final size. Please send us this information along with the revised manuscript.

**Response:**

We submitted a short summary, bullet points, and a synopsis image along with the revised manuscript.

**Referee #2:**

The authors have answered several of my concerns, however the study remains descriptive and without major discoveries or genetic confirmation of the findings. Furthermore, the use of a single NC tracer is an important limitation of the study. I think this work will be useful as a resource for scientists working in the cardiac neural crest field.

**Response:**

Thank you very much for your comments! In this revision, we clearly stated the limitation of using one tracing model in the Discussion section (Line 365-368). We pasted the related paragraph below:

“Although the Wnt1-Cre line is well-established for neural crest lineage tracing, phenotypic variances may exist among different tracing models (Debbache et al, 2018). Thus, the current study may be limited by using only one tracing model. It is helpful to use additional driver lines (e.g., P0-Cre and Pax3-Cre) or a different recombination system such as Wnt1-Flpe to confirm the findings of this study in the future.”

**Referee #3:**

The resubmission of the manuscript by Liu and colleagues provide additional information about experimental procedures, quality control, and clustering. These data have elevated the manuscript's quality and minimizes concerns regarding small clusters and that analyses of trajectories are driven by one particular cluster or sample. The authors' justification of the large number of clusters derived from a small number

of cells by providing marker genes that isolate each individual cluster, provides potential new markers for individual cells and states.

While these changes address the majority of the major concerns in this manuscript, a few additional textual changes would be appropriate.

1. In the discussion the author should specifically address that using a single Cre has limitations as data derived from one Cre line might subsets of migrating and differentiating neural crest cells.

**Response:**

Thank you very much for your comments! In this revision, we clearly stated the limitation of using one tracing model in the Discussion section (Line 365-368). We pasted the related paragraph below:

“Although the Wnt1-Cre line is well-established for neural crest lineage tracing, phenotypic variances may exist among different tracing models (Debbache et al, 2018). Thus, the current study may be limited by using only one tracing model. It is helpful to use additional driver lines (e.g., P0-Cre and Pax3-Cre) or a different recombination system such as Wnt1-Flpe to confirm the findings of this study in the future.”

2. The addition of information on quality metrics provide in this revision are helpful. However, more would be better. Please consider adding uMAP plots before and after CCA integration as well as scrublet score to provide clear information on doublets and batch effect.

**Response:**

As suggested, we added UMAP plots in Figure EV1 showing Scrublet doublet score as well as before and after CCA integration.

3. The authors addressed the concern about over clustering by providing marker genes (Figure VE3). Specifying in the text if some of these genes are novel markers of cell states and information about their functions would be helpful.

**Response:**

Since the subclusters we identified are generally previously unrecognized cell states, many of the genes represent novel markers for these cell states that have not been reported. For example, the VSMC subpopulations c1 and c3 are distinguishable by the

expression levels of markers such as *Bcam*, *Cebpb*, and *Ppp1r15a* (Figure EV1). As suggested, we added related statements in the text (Line 136-140).

Prof. Zhou Zhou  
Fuwai Hospital, the Chinese Academy of Medical Sciences  
No. 167 Beilishi Street  
Xicheng District  
Beijing, Beijing 100037  
China

Dear Prof. Zhou,

I am very pleased to accept your manuscript for publication in the next available issue of EMBO reports. Thank you for your contribution to our journal.

At the end of this email I include important information about how to proceed. Please ensure that you take the time to read the information and complete and return the necessary forms to allow us to publish your manuscript as quickly as possible.

As part of the EMBO publication's Transparent Editorial Process, EMBO reports publishes online a Review Process File to accompany accepted manuscripts. As you are aware, this File will be published in conjunction with your paper and will include the referee reports, your point-by-point response and all pertinent correspondence relating to the manuscript.

If you do NOT want this File to be published, please inform the editorial office within 2 days, if you have not done so already, otherwise the File will be published by default [contact: [emboreports@embo.org](mailto:emboreports@embo.org)]. If you do opt out, the Review Process File link will point to the following statement: "No Review Process File is available with this article, as the authors have chosen not to make the review process public in this case."

Should you be planning a Press Release on your article, please get in contact with [emboreports@wiley.com](mailto:emboreports@wiley.com) as early as possible, in order to coordinate publication and release dates.

Thank you again for your contribution to EMBO reports and congratulations on a successful publication. Please consider us again in the future for your most exciting work.

Yours sincerely,

Martina Rembold, PhD  
Senior Editor  
EMBO reports

\*\*\*\*\*

#### THINGS TO DO NOW:

You will receive proofs by e-mail approximately 2-3 weeks after all relevant files have been sent to our Production Office; you should return your corrections within 2 days of receiving the proofs.

Please inform us if there is likely to be any difficulty in reaching you at the above address at that time. Failure to meet our deadlines may result in a delay of publication, or publication without your corrections.

All further communications concerning your paper should quote reference number EMBOR-2021-52389V3 and be addressed to [emboreports@wiley.com](mailto:emboreports@wiley.com).

Should you be planning a Press Release on your article, please get in contact with [emboreports@wiley.com](mailto:emboreports@wiley.com) as early as possible, in order to coordinate publication and release dates.

**YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND** ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Zhou Zhou

Journal Submitted to: EMBO Reports

Manuscript Number: EMBOR-2021-52389V1

### Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

#### A- Figures

##### 1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if  $n < 5$ , the individual data points from each experiment should be plotted and any statistical test employed should be justified.
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

##### 2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple  $\chi^2$  tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g.,  $P$  values =  $x$  but not  $P$  values  $< x$ ;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

#### B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

|                                                                                                                                                                                         |                                                                                                                                                                                                                                                                 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?                                                                                     | NA                                                                                                                                                                                                                                                              |
| 1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.                                                                       | Here is the statement in the main text: "The developmental stages we selected spanned from the very early time when CNCC derivatives arrived at the cardiac OFT during embryonic development (i.e., E10.5) (Lin et al, 2012) to neonatal stage P7 (Figure 1B)." |
| 2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?                                                      | NA                                                                                                                                                                                                                                                              |
| 3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.                | NA                                                                                                                                                                                                                                                              |
| For animal studies, include a statement about randomization even if no randomization was used.                                                                                          | Here is the statement in the main text: "Randomization and blinding are not applicable for this study".                                                                                                                                                         |
| 4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe. | NA                                                                                                                                                                                                                                                              |
| 4.b. For animal studies, include a statement about blinding even if no blinding was done                                                                                                | Here is the statement in the main text: "Randomization and blinding are not applicable for this study".                                                                                                                                                         |
| 5. For every figure, are statistical tests justified as appropriate?                                                                                                                    | Yes.                                                                                                                                                                                                                                                            |
| Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.                                                                      | NA                                                                                                                                                                                                                                                              |
| Is there an estimate of variation within each group of data?                                                                                                                            | Yes.                                                                                                                                                                                                                                                            |

<http://www.antibodypedia.com>

<http://1degreebio.org>

<http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repor>

<http://grants.nih.gov/grants/olaw/olaw.htm>

<http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm>

<http://ClinicalTrials.gov>

<http://www.consort-statement.org>

<http://www.consort-statement.org/checklists/view/32-consort/66-title>

<http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tum>

<http://datadryad.org>

<http://figshare.com>

<http://www.ncbi.nlm.nih.gov/gap>

<http://www.ebi.ac.uk/ega>

<http://biomodels.net/>

<http://biomodels.net/miriam/>

<http://jij.biochem.sun.ac.za>

[http://oba.od.nih.gov/biosecurity/biosecurity\\_documents.html](http://oba.od.nih.gov/biosecurity/biosecurity_documents.html)

<http://www.selectagents.gov/>

|                                                                                   |      |
|-----------------------------------------------------------------------------------|------|
| Is the variance similar between the groups that are being statistically compared? | Yes. |
|-----------------------------------------------------------------------------------|------|

### C- Reagents

|                                                                                                                                                                                                                                                                                                                                                                  |                          |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|
| 6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia ( <a href="#">see link list at top right</a> ), 1DegreeBio ( <a href="#">see link list at top right</a> ). | Catalog number provided. |
| 7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.                                                                                                                                                                                                            | NA                       |

\* for all hyperlinks, please see the table at the top right of the document

### D- Animal Models

|                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------|
| 8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.                                                                                                                                                                                                                                                            | They are stated in the manuscript.                                                                                                              |
| 9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.                                                                                                                                                                                                                                                                       | As indicated in the Materials and Methods of the manuscript, all experiments were approved by the ethics committee of Fuwai Hospital (2017-877) |
| 10. We recommend consulting the ARRIVE guidelines ( <a href="#">see link list at top right</a> ) (PLOS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH ( <a href="#">see link list at top right</a> ) and MRC ( <a href="#">see link list at top right</a> ) recommendations. Please confirm compliance. | As stated in the Materials and Methods of the manuscript, all animal experiments complied with the ARRIVE guidelines.                           |

### E- Human Subjects

|                                                                                                                                                                                                                                                                                                                                                        |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 11. Identify the committee(s) approving the study protocol.                                                                                                                                                                                                                                                                                            | NA |
| 12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.                                                                                                | NA |
| 13. For publication of patient photos, include a statement confirming that consent to publish was obtained.                                                                                                                                                                                                                                            | NA |
| 14. Report any restrictions on the availability (and/or on the use) of human data or samples.                                                                                                                                                                                                                                                          | NA |
| 15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.                                                                                                                                                                                                                                             | NA |
| 16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram ( <a href="#">see link list at top right</a> ) and submit the CONSORT checklist ( <a href="#">see link list at top right</a> ) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list. | NA |
| 17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines ( <a href="#">see link list at top right</a> ). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.                                                                                          | NA |

### F- Data Accessibility

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|
| 18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'.                                                                                                                                                                                                                                                                                                                                                                                                              | Data have been deposited in SRA. |
| Data deposition in a public repository is mandatory for:<br>a. Protein, DNA and RNA sequences<br>b. Macromolecular structures<br>c. Crystallographic data for small molecules<br>d. Functional genomics data<br>e. Proteomics and molecular interactions                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                  |
| 19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad ( <a href="#">see link list at top right</a> ) or Figshare ( <a href="#">see link list at top right</a> ).                                                                                                                                                                                                                                              | NA                               |
| 20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP ( <a href="#">see link list at top right</a> ) or EGA ( <a href="#">see link list at top right</a> ).                                                                                                                                                                                                                                       | NA                               |
| 21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines ( <a href="#">see link list at top right</a> ) and deposit their model in a public database such as Biocompare ( <a href="#">see link list at top right</a> ) or JWS Online ( <a href="#">see link list at top right</a> ). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information. | NA                               |

### G- Dual use research of concern

|                                                                                                                                                                                                                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 22. Could your study fall under dual use research restrictions? Please check biosecurity documents ( <a href="#">see link list at top right</a> ) and list of select agents and toxins (APHIS/CDC ( <a href="#">see link list at top right</a> )). According to our biosecurity guidelines, provide a statement only if it could. | NO |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|