A single cell transcriptional atlas of early synovial joint development

Qin Bian, Yu-Hao Cheng, Jordan P Wilson, Emily Y. Su, Dong Won Kim, Hong Wang, Sooyeon Yoo, Seth Blackshaw and Patrick Cahan

DOI: 10.1242/dev.185777

Editor: Benoit Bruneau

Review timeline

Original submission:	18 October 2019
Editorial decision:	14 January 2020
First revision received:	9 June 2020
Accepted:	9 June 2020

Original submission

First decision letter

MS ID#: DEVELOP/2019/185777

MS TITLE: A single cell transcriptional atlas of early synovial joint development

AUTHORS: Qin Bian, Yu-Hao Cheng, Jordan P Wilson, Dong Won Kim, Hong Wang, Seth Blackshaw, and Patrick Cahan

I have now received all the referees reports on the above manuscript, and have reached a decision. The referees' comments are appended below, or you can access them online: please go to BenchPress and click on the 'Manuscripts with Decisions' queue in the Author Area.

The overall evaluation is positive and we would like to publish a revised manuscript in Development, provided that the referees' comments can be satisfactorily addressed. Please attend to all of the reviewers' comments in your revised manuscript and detail them in your point-by-point response. If you do not agree with any of their criticisms or suggestions explain clearly why this is so.

Reviewer 1

Advance summary and potential significance to field

In this manuscript, the authors employ single cell RNA sequencing of four stages of mouse knee joint development to infer pathway networks that promote the differentiation of different cell lineages which comprise synovial joints. The authors confirm some of the clustering analyses by testing expression in vivo by performing one cell differentiation experiment in vitro, and present the bulk of the data as a resource.
There are a high number of predictions concerning lineage progression and regulation made throughout the results section which are potentially useful for other investigators, and I readily accept that a predominantly descriptive study such as this can be valuable. However, I have three most substantive concerns that preclude my recommendation for publication of this study in its current form: factors that might bias the cell population of interest are not sufficiently clear, the proportion of inferences to confirmatory experiments is too high, and a substantial number of predictions are based on an analysis tool that is not published or sufficiently described.

Comments for the author

Is it known whether changes in the spatial distribution of Gdf5 +ve descendents from the initially periarticular positions to specific intraarticular tissues between E12.5-15.5 is due to cell movement, gain of gene expression in previously Gdf5 –ve cells, or death of formerly Gdf5 +ve cells? That information would be useful to better understand which subpopulation of cells is represented in the cluster-specific and RNA velocity analyses.
Even under live culture conditions, there would be concern that limbs at the stages employed here degrade ex utero. What happens to RNA during 4.5 h collagenase digestion, and does that process happen evenly to all RNA species in all cell populations?
It would be helpful to briefly define doublets and criteria for excluding low-quality libraries. Please consider the preceding point. In other words, how can one know whether the most degradation-vulnerable cells or transcripts were systematically excluded?
Is there a list of “canonical cell cycle genes” that were analysed? How did G2/M and S phase cells “confound downstream analysis”? Are all cells in the developing knee cycling at a similar rate such that analysis is not biased by selecting for cell cycle stage? Is “cell trimming” a standard practice? The visually apparent S phase cluster, in particular, seems to include a substantial number of G1 phase cells. Were those also excluded? How did removal of those cells, and the “cell trimming process” in general, alter the outcome?
Since Epoch is an analysis tool that is being prepared for submission and is only briefly described in this manuscript, it is hard to understand the relevance and validity of Epoch data.

Minor
In the abstract, please justify the term “functional analysis“ or change it.
Please explain the rationale for not performing single cell analysis of an unbiased population of all knee joint cells regardless of Gdf5 expression history.
Since mice were bred during the day (in the dark), midnight was taken as E0.5, and E12.5-15.5 stages were harvested at 3 pm, it seems those fetuses should more correctly be labelled as ~ E13.0 -16.0.
It may be simpler to note that the knee was isolated by transfemoral and transtibial division.
In Supp. Fig. 1, the labels Dissociation and Dissections may be reversed. The immune and neural crest-derived cells are not labelled. This figure would benefit from a proper legend.
Were cells from all four stages mixed and analysed together, or were they analysed separately and clustered together to derive Supp. Fig. 1C timpoint?
Fig. 2A legend: five or three superclusters?
The speculation that SC1 may include dedifferentiated chondrocytes at E12.5 based on proportionately low expression Col2a1 is not intuitive.
Would it not be useful to show separate Leiden clustering for each time point in order to better appreciate the evolution of Gdf5 +ve cell types?
Please explain how clustering of multiple time points affects RNA velocity analysis.

First revision

Author response to reviewers' comments

Response to Reviewers

Reviewer 1

In this manuscript, the authors employ single cell RNA sequencing of four stages of mouse knee joint development to infer pathway networks that promote the differentiation of different cell lineages which comprise synovial joints. The authors confirm some of the clustering analyses by testing expression in vivo, by performing one cell differentiation experiment in vitro, and present the bulk of the data as a resource. There are a high number of predictions concerning lineage progression and regulation made throughout the results section which are potentially useful for other investigators, and I readily

accept that a predominantly descriptive study such as this can be valuable. However, I have three most substantive concerns that preclude my recommendation for publication of this study in its current form:

1) factors that might bias the cell population of interest are not sufficiently clear,

This is an excellent point that we did not thoroughly address in the original manuscript. There are several factors that might bias the cell population of interest. As the reviewer points out below, certain sub-populations might be more susceptible to loss during isolation and thus be under-represented in the scRNAseq data. For example, some populations might be prone to cell death during enzymatic tissue dissociation. Moreover, some populations may tend to be more proliferative at the stages we sampled, and thus would have been excluded because of our cell cycle 'cell trimming' step. In the revised manuscript, we have described potential factors that might bias the population that we analyzed by scRNAseq in the Discussion on pg 14-15. Additionally, we re-analyzed the scRNAseq data after including cells that we had previously excluded due to stage of cell cycle. Because we included an additional 1,000 cells, this re-analysis resulted in cascade of changes that affected some of the sub-clusters. Namely, each super- cluster and major sub-cluster now includes at least one sub-cluster defined by a proliferation signature.

2) the proportion of inferences to confirmatory experiments is too high

We agree and have taken two steps to address this imbalance. First, we have performed additional experiments to test the scRNAseq clustering results. These experiments included ISH of Dkk2 and Egfl6 (Fig 6B), which allowed us to distinguish the location of Supercluster 3 (fibrous cells) from a small sub-cluster in SC2 with a similar profile. We have also used the RNAscope Hiplex (multiplex ISH) method to localize and assign putative identity to the following clusters: SC2_B2 as nascent intra-articular ligament with Htra1/Mfap4 (Fig 5F), SC2_B4 as synovial cavity lining cells and SC2_B5 as menisci cells with Emp1/Cd44 (Fig 5G), SC2_B2 as cells of the outer tendon with Scx/Tmnd (Fig 6F), and SC3_B3 as fibrochondrocytes of the synovium, enthesis, and perichondrium with Col8a2/Dlx5 (Fig 6G). We describe these results on pages 11-13 12.

Secondly, we have removed the results depending on Epoch, as this tool is still unpublished (see our response to the next issue below). This has substantially reduced the number of predictions reported here and has resulted in greater parity between hypothesis and experimental support.

3) and a substantial number of predictions are based on an analysis tool that is not published or sufficiently described.

We agree with the reviewer. The tool that we used in the original manuscript has still not been published. Therefore, we have removed sections of the manuscript derived from Epoch analysis. To identify putative regulators of trajectories, we used a previously published pseudotime method that tests for genes that are dynamically expressed. Also, we have limited the application of this method to SC1_B2 and SC1_B4.

Is it known whether changes in the spatial distribution of Gdf5 +ve descendants from the initially periarticular positions to specific intraarticular tissues between E12.5-15.5 is due to cell movement, gain of gene expression in previously Gdf5 –ve cells, or death of formerly Gdf5 +ve cells? That information would be useful to better understand which subpopulation of cells is represented in the cluster-specific and RNA velocity analyses.

This is an excellent point. First, a recent lineage tracing study indicated that early Gdf5 +ve descendants (labeled on E10.5 to E12.5) were seen at specific intra-articular tissues between E13.5 to E15.5[1]. But many Gdf5-expressing cells observed at intra-articular tissues did not have a Gdf5 expression history, indicating that some of Gdf5 +ve descendants could move, and some newly Gdf5 +ve cells gain gene expression in previously Gdf5 -ve cells. It is possible that some formerly Gdf5 +ve cells die during joint formation. Some studies have found little-to-no morphological or biochemical signs of cell death, in particular apoptosis, in the interzone at any embryonic stage in the rat knee joint[2]. We excluded dead cells before 10X processing. We believe cell death did not affect either subpopulation or RNA velocity analyses in our manuscript.

We have added a brief discussion of these points on Pg 14.

Even under live culture conditions, there would be concern that limbs at the stages employed here degrade ex utero. What happens to RNA during 4.5 h collagenase digestion, and does that process happen evenly to all RNA species in all cell populations?

There are two good points here. First, to what extent does 4.5 h collagenase digestion affect RNA? We have found that digestion overnight caused an increase in ECM degradation-related genes in adult mice knee joint (our unpublished data). In order to decrease the influence of collagenase D on cells, we optimized the original protocol by shortening overnight digestion at 0.5 mg/mL to 3hr-digestion at 1 mg/mL. It is still likely that the digestion will impact cells expression state. The second point is whether this effect is equal across all cells. It is certainly possible that it is not even across all cells. Unfortunately, we are unable to assess these potential influences with our single cell data. We acknowledge these issues on pg 15 and suggest that newer technologies such as Spatial Transcriptomics will enable the field to estimate and lessen their impact.

It would be helpful to briefly define doublets and criteria for excluding low-quality libraries. Please consider the preceding point. In other words, how can one know whether the most degradation-vulnerable cells or transcripts were systematically excluded?

We have added this information on pg 5-6. We defined doublets as the cells in the top 5% of total reads per capture based on the estimated doublet rate of the 10x platform. We labeled libraries as low-quality if they had fewer than 500 genes detected or if their total transcriptome consisted of greater than 5% of mitochondrially encoded genes. When we lowered these thresholds for including cells, we did not detect additional, biologically meaningful clusters and therefore we do not think that this routine QC step excluded interesting degradation-vulnerable cells.

Is there a list of “canonical cell cycle genes” that were analysed? How did G2/M and S phase cells “confound downstream analysis”? Are all cells in the developing knee cycling at a similar rate such that analysis is not biased by selecting for cell cycle stage? Is “cell trimming” a standard practice? The visually apparent S phase cluster, in particular, seems to include a substantial number of G1 phase cells. Were those also excluded? How did removal of those cells, and the “cell trimming process” in general, alter the outcome? Any paper to support it???

The list of cell cycle related genes was first reported in the original Seurat paper[3] and can be accessed here: https://satijalab.org/seurat/v3.0/cell_cycle_vignette.html

We agree with the reviewer that excluding these cells may have biased our overall results. Therefore, we went back and re-analyzed the complete dataset without filtering out these cells. We still identified many of the same or similar clusters. However, in each super-cluster, and each 'level 2' sub-cluster, we have populations that are defined primarily by expression of genes involved in active cell cycle.

Since Epoch is an analysis tool that is being prepared for submission and is only briefly described in this manuscript, it is hard to understand the relevance and validity of Epoch data.

Please see our response to major point 1 above.

Minor

In the abstract, please justify the term “functional analysis“ or change it.

We have changed “functional analysis of prospectively isolated populations” to “in vitro characterization of prospectively isolated populations”.

Please explain the rationale for not performing single cell analysis of an unbiased

population of all knee joint cells regardless of Gdf5 expression history.

Our focus was on interzone and interzone derived cells, as they give rise to the major joint lineages. Many studies have provided data supporting the notion that most, if not all, interzone cells have a Gdf5 expression history, which is what motivated our strategy for selecting cells to analyze.

Since mice were bred during the day (in the dark), midnight was taken as E0.5, and E12.5-15.5 stages were harvested at 3 pm, it seems those fetuses should more correctly be labelled as ~ E13.0 -16.0.

This was typo which we have corrected from 3 pm to 3 am.

It may be simpler to note that the knee was isolated by transfemoral and transtibial division.

We agree and have changed this.

In Supp. Fig. 1, the labels Dissociation and Dissections may be reversed.

We agree and have changed this.

The immune and neural crest-derived cells are not labelled. This figure would benefit from a proper legend.

We have added detailed legend to Supplementary Fig 2.

Were cells from all four stages mixed and analyzed together, or were they analyzed separately and clustered together to derive Supp. Fig. 1C timepoint?

We sequenced four stages separately and analyzed together. Supplementary Fig1C was added to illustrate it.

Fig. 2A legend: five or three superclusters?

There are three superclusters.

The speculation that SC1 may include dedifferentiated chondrocytes at E12.5 based on proportionately low expression Col2a1 is not intuitive.

We have removed this statement.

Would it not be useful to show separate Leiden clustering for each time point in order to better appreciate the evolution of Gdf5 +ve cell types?

We have performed this analysis and now include it as Supp Fig 6.

Please explain how clustering of multiple time points affects RNA velocity analysis.

This question relates to the validity of applying RNA velocity to data from multiple timepoints. The answer to this question depends on how far in the future can RNA velocity (RV) extrapolate future state. La Manno et al 2018 explored this question by scaling RV-determined pseudotime to experimentally determined chromaffin differentiation[4]. They showed that individual cell velocity extrapolation times are around 2 hours, and that longer-scale phenomenon can be modeled with RV by linking single cell extrapolations (see La Manno Figure 3, 2F-G and Supp Note 2.6 and 2.7). That La Manno et al's application of RV to data spanning post-natal day 0 to day 5 (see La Manno et al figure 3) successfully recapitulated known developmental trajectories of the murine hippocampus supports the notion that RV can be applied across time scales longer than the individual cell extrapolation time.

We sampled GLE cells at 24 hour intervals. If we assume that at each timepoint, we sampled cells

from across the spectrum of differentiation, then it is likely that there will be overlap in differentiation stage in each adjacent time point and therefore we can safely apply RV. This assumption is supported for SC2 and SC3 by the fact that these clusters have a mixture of cells from e13-e15. However, SC1 is almost entirely made up of cells from e12.5. Therefore, we kept our RV analysis within superclusters, but we removed Figure 7A, which depicted RNA velocity analysis across all superclusters.

Bibliography

1. Shwartz, Y., Viukov, S., Krief, S. & Zelzer, E. Joint Development Involves a Continuous Influx of Gdf5-Positive Cells. *Cell Rep.* **15,** 2577–2587 (2016).
2. Ito, M. M. & Kida, M. Y. Morphological and biochemical re-evaluation of the process of cavitation in the rat knee joint: cellular and cell strata alterations in the interzone. *J. Anat.* **197 Pt 4,** 659–679 (2000).
3. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33,** 495–502 (2015).
4. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560,** 494–498 (2018).

Second decision letter

MS ID#: DEVELOP/2019/185777

MS TITLE: A single cell transcriptional atlas of early synovial joint development

AUTHORS: Qin Bian, Yu-Hao Cheng, Jordan P Wilson, Emily Y. Su, Dong Won Kim, Hong Wang, Sooyeon Yoo, Seth Blackshaw, and Patrick Cahan
ARTICLE TYPE: Techniques and Resources Article

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