THE EMBO JOURNAL

Regeneration of the human segmentation clock in somitoids *in vitro*

Yue Qin, Xingnan Huang, Zepo Cai, Baomei Cai, Jiangping He, Yuxiang Yao, Chunhua Zhou, Junqi Kuang, Yihang Yang, Huan Chen, Yating Chen, Sihua Ou, Lijun Chen, Fang Wu, Ning Guo, Yapei Yuan, Xiangyu Zhang, Wei Pang, Ziyu Feng, Shengyong Yu, Jing Liu, Shangtao Cao, Duanqing Pei
DOI: 10.15252/embj.2022110928

Corresponding author: Duanqing Pei (peiduanqing@westlake.edu.cn)

Review Timeline:

Submission Date:	14th Feb 22
Editorial Decision:	16th Mar 22
Revision Received:	5th Jul 22
Editorial Decision:	18th Aug 22
Revision Received:	2nd Sep 22
Accepted:	16th Sep 22

Editor: Daniel Klimmeck

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.)

1st Editorial Decision **16th Mar 2022**

Dear Duanqing,

Thank you again for the submission of your manuscript (EMBOJ-2022-110928) to The EMBO Journal. Your study has been sent to three reviewers with developmental expertise, however one reviewer got much delayed and, has not sent us his-her report so far. We have received feedback from the other two referees, which I enclose below, and decided to - in the interest of the timeliness of the data - proceed with our decision based on these reports.

As you will see, the referees acknowledge the potential interest and novelty of your methods advances and results, although they also express a number of issues that will have to be conclusively addressed before they can be supportive of publication of your manuscript in The EMBO Journal. In more detail, the experts state that consideration of additional markers for acquired pluripotent and differentiated states during the cell fate conversion is required to consolidate your claims (ref#1, pt2. 1,2; ref#2 pts.2,3). Further, reviewer #1 requests integration of published single-cell profiling datasets in order to better link the identities acquired in the current protocol to existing knowledge (ref#1, pt.4). In addition, the reviewers raise a number of points related to additional controls required, overall data discussion and literature references that would need to be conclusively addressed to achieve the level of robustness and clarity needed for The EMBO Journal.

I judge the comments of the referees to be generally reasonable and given their overall interest, we are in principle happy to invite you to revise your manuscript experimentally to address the referees' comments, pending there are no major concerns by referee #3 on the technical robustness of the work.

I will share the report of this expert as soon as we receive it.

As you know, we generally allow three months as standard revision time. As a matter of policy, competing manuscripts published during this period will not negatively impact on our assessment of the conceptual advance presented by your study. However, we request that you contact the editor as soon as possible upon publication of any related work, to discuss how to proceed. Should you foresee a problem in meeting this three-month deadline, please let us know in advance and we may be able to grant an extension.

When submitting your revised manuscript, please carefully review the instructions below.

Thank you for the opportunity to consider your work for publication.

I look forward to your revision.

Best regards,

Daniel

Daniel Klimmeck, PhD
Senior Editor
The EMBO Journal

Referee #1:

In this manuscript, Qin et al. present a system to directly reprogram human urine epithelial cells to presomitic mesoderm, which expresses markers of mesoderm differentiation and somitogenesis and shows oscillatory expression of cyclic genes. In addition, they describe a small molecule inhibitor cocktail that enables the long-term culture of the generated progenitor cells. This addresses several challenges in the study of human somitogenesis and is potentially very exciting for the field. However, there are several points that should be worked out further or clarified to substantiate those main findings.

Major points:

- Episomal vectors are used for reprogramming of urine epithelial cells. These episomal vectors induce expression of Oct4 (Pou5f1), Sox2, SV40LT and Klf4 (pEP4E02SET2K) and the miR-302-367 cluster (pCEP4-miR-302-367), a combination which has been used successfully to reprogram somatic cells to iPSCs.
These episomal vectors replicate and are lost only slowly during proliferation (5 % per cell division). In addition, it has been shown that episomal vectors can still be integrated into the genome. This raises several questions for the current study:
1. Do the cells go through a transient phase of "pluripotency" or multipotency? By analyzing expression of further pluripotency markers at further timepoints this can be tested.
2. What do the cells express that differentiate to PSM-like cells? In other words: How does induction of these different factors influence the (trans-)differentiation process? Co-staining of differentiated cells for these induced pluripotency markers and PSM markers can for instance be used to test this.
3. Do the cells that keep on proliferating in a progenitor-like state for various passages still contain the episomal DNA? By staining for the induced factors and by qPCR this can be tested.

4. The identity of the different cell types at the consecutive stages of reprogramming and differentiation to PSM does not become entirely clear. Even though RNAseq and qPCR of selected markers is performed, it remains difficult until the end to grasp what these cells exactly are. It would be useful to compare the obtained scRNAseq data to available scRNAseq data from human gastruloids (Morris et al. 2020), in vitro differentiated PSM and/ or mouse embryos at different stages (e.g. van den Brink et al. 2020). This would, by the way, also allow a better characterization of the cell identity of cells in the "failed branch(es)" of cell trajectories in Figure S1G (what are these cells?). Furthermore, expression of further pluripotency would be useful.

5. Various combinations of small molecule inhibitors are presented throughout the manuscript, e.g. during the reprogramming stage or when maintaining the UiPSM cells in a progenitor stages over several passages. It is said at several places that an optimized medium composition is used. However, the reasoning for using these particular combinations does not become clear and re-writing these sections and clarifying this would greatly improve the readability and understandability of the paper. For instance, I would expect to see a reference to a previous publication in which this or a similar combination has been used. Alternatively, I would expect to either read an explanation based on known literature why a certain drug was used or the description of some sort of inhibitor screen (as they do for the optimization of their gastruloid/ "somitoid" protocol).

6. A small molecule inhibitor cocktail is presented that is used to maintain UiPSM cells long-term in culture. If this is solely dependent on these small molecules (and not the reprogramming factors, see above), then it should be possible to use these molecules to maintain primary or in vitro differentiated progenitor cells in culture in a similar way. If this is possible, this would be an exciting finding for the field.

7. For the quantification of Hes7 and Mesp2 oscillations and the GFP control, I would expect to see the real raw or smoothened/detrended signal of several samples (Figure 3E and Figure S3G). In addition, the movies corresponding to Figure 3E, Figure S3G and Figure 4A should be added to the paper.

Minor points:

- In the introduction, it says that PSM is derived from the "CLE of the tailbud". This is not complete. In early stages of somitogenesis PSM is indeed generated from the CLE. At later stages, once the tailbud has formed, the NMPs reside in a region termed the chordoneural hinge (CNH).

- In the introduction, van den Brink et al. is cited twice. However, this reference is not really fitting at these points and other papers might be more useful.

- In the optimization of the gastruloid protocol, it is not clear how this optimization is done. Is elongation the only readout or are also the present cell types taken into account?

Further point:
The language is sometimes not perfect with typos and words missing. It would be useful to proofread this entirely again.

Referee #2:

General summary and opinion about the principle significance of the study, its questions and findings

The manuscript describes a novel method for the generation of human PSM cells using somatic cells (epithelial cells harvested from urine). This research is timely as it is the first model system for PSM of its kind and comes at a time where somitogenesis research is being transformed by the alternative approach of the generation of new human model systems using stem cells. The authors demonstrate a moderate efficiency in the generation of PSM cells which they use to generate somite like structures. The manuscript would benefit from some additional data and further clarification.

Specific major concerns essential to be addressed to support the conclusions
1. The authors use T expression as an indicator of the percentage of cells that achieve PSM status and present a representative example (Fig 1E) showing 44% T positive cells. It is not clear what this data represents as Fig 1L shows data for the three donors that are used in this study with 56, 34 and 17% T positive. These percentages do not suggest an efficient differentiation protocol and therefore the protocol would benefit from further improvement. Moreover, the number of donors used is very small (three) and analysing the reprogramming efficiency for a larger number of donors would improve the impact of this manuscript. From Fig 2 onwards the authors use optimised induction medium (DM). Does differentiation using this medium result in higher percentages of T positive cells? Data demonstrating the efficiency using DM is lacking.
2. The data in Fig S2A shows variable expression levels of PSM markers, on several occasions this is clearly reduced in the higher passage numbers. This at the very least suggest that these cell lines are not consistent in their expression patterns when being passaged and it possibly means that the PSM identity might be reduced.
3. Sox2 expression is used as a somitogenesis marker as well as a pluripotency marker. Therefore, the manuscript would be more convincing if an alternative somitogenesis marker would be chosen to focus on throughout the manuscript. Fig 1H shows that Tbx6 and Sox2 expression is complementary rather than co-expressing. The coexpression data in Fig 1J would benefit from an explanation as the top and bottom panel look rather different but still result in the same conclusion.
4. The description of Fig 3E states that Hes7 Pepper oscillating fluorescence is gradually decreasing but this is not obvious from the Figure. How were these Figures generated? There seems to be an unnatural uniformity to these oscillations, making them look like a model rather than actual data. Similarly, the graph in Fig SG is remarkably straight and level.
5. The data in Fig 4A doesn't convincingly support the statement that Hes7-GFP expression regresses posteriorly.

Minor concerns that should be addressed
1. For qPCR data much more info needs to be provided regarding quality control (e.g. MIQE). The expression levels are normalised against GAPDH only. Can the authors provide evidence that GAPDH expression is consistent throughout the differentiation protocol? The qPCR data would benefit from the use of more than one housekeeping gene for normalisation.
2. The description of the generation of the CRISPR cell lines is rather minimal. There is a need for more details, especially including information on the quality control experiments performed to establish that the new cell lines are exactly what they are meant to be and no additional changes have occurred. More information is needed regarding the quality control of the RNA seq samples. Additional information is required on the histological analysis method. There is insufficient information on quantification, sample numbers and statistics throughout the paper.
3. Towards the end of the introduction there is a statement that suggests that human gastruloids are the only species that contain three germlayers. This sentence would benefit from reformulating.
4. The details of the media used in Fig1 and S1 are not described in the methods section.
5. Fig 2I-J show two clusters (M5+6) that are consistent with muscle development. These clusters are rather small in comparison to the other clusters. Further explanation of what this means would make it easier to interpret the data.
6. The discussion states limitations regarding hESC and hiPS for studying PSM function. This is followed by an example of reporter studies demonstrated in the manuscript suggesting that this is an advantage of the described system. However, these type of experiments have been described in hiPS derived PSM cells too and these cells can be differentiated much more efficiently into PSM.
7. The discussion also states that the UiPSM system may serve as a chronological model for ageing and rejuvenation research. This statement is too strong and is premature based on the data provided.
8. The colour schemes of several figures don't allow of proper analysis of the data as the colours used are too similar. Moreover, some of the colours are so pale they don't print well and in some figures the colours green and red are used which aren't very accessible for colour blind readers. Some figures contain extremely small text.
9. Is there a reason why the growth curve for UC is so much shorter than for UiPSM (Fig 2C)?
10. In Fig S1 it would enhance the data if the three controls (DE, NSC and hESC) were included in all three graphs as they would function as either positive or negative controls.
11. Fig S2A there are no error bars or an indication of the number of biological repeats.
12. Fig S3C-E would benefit from a larger number of examples for each condition. Additional info in the legends of Fig S3F regarding top, middle and bottom panels is required.

Additional non-essential suggestions for improving the study
1. The manuscript would benefit from corrections of spelling, grammar and typos.

2. The results section that describes Fig 2E-J would benefit from an introducing sentence explaining the objective of these experiments.
3. In Fig 1J the top and bottom figures have been swapped (based on the legend).
4. References to Brachury/T/Tbxt are not consistent. In the main text T is mostly used while in some of the figures it is referred to as Tbxt.

9) We replaced Supplementary Information with Expanded View (EV) Figures and Tables that are collapsible/expandable online (see examples in https://www.embopress.org/doi/10.15252/embj.201695874). 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: https://www.embopress.org/page/journal/14602075/authorguide#expandedview >.

- Additional Tables/Datasets should be labelled 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.

10) When assembling figures, please refer to our figure preparation guideline in order to ensure proper formatting and readability in print as well as on screen: http://bit.ly/EMBOPressFigurePreparationGuideline

Please remember: Digital image enhancement is acceptable practice, as long as it accurately represents the original data and conforms to community standards. If a figure has been subjected to significant electronic manipulation, this must be noted in the figure legend or in the 'Materials and Methods' section. The editors reserve the right to request original versions of figures and the original images that were used to assemble the figure.

11) For data quantification: please specify the name of the statistical test used to generate error bars and P values, the number (n) of independent experiments (specify technical or biological replicates) underlying each data point and the test used to calculate p-values in each figure legend. The figure legends should contain a basic description of n, P and the test applied. Graphs must include a description of the bars and the error bars (s.d., s.e.m.).

Further information is available in our Guide to Authors: https://www.embopress.org/page/journal/14602075/authorguide

We realize that it is difficult to revise to a specific deadline. In the interest of protecting the conceptual advance provided by the work, we recommend a revision within 3 months (14th Jun 2022). Please discuss the revision progress ahead of this time with the editor if you require more time to complete the revisions. Use the link below to submit your revision:

https://emboj.msubmit.net/cgi-bin/main.plex

Referee #1:

In this manuscript, Qin et al. present a system to directly reprogram human urine epithelial cells to presomitic mesoderm, which expresses markers of mesoderm differentiation and somitogenesis and shows oscillatory expression of cyclic genes. In addition, they describe a small molecule inhibitor cocktail that enables the long-term culture of the generated progenitor cells. This addresses several challenges in the study of human somitogenesis and is potentially very exciting for the field. However, there are several points that should be worked out further or clarified to substantiate those main findings.

Major points:
- Episomal vectors are used for reprogramming of urine epithelial cells. These episomal vectors induce expression of Oct4 (Pou5f1), Sox2, SV40LT and Klf4 (pEP4E02SET2K) and the miR-302-367 cluster (pCEP4-miR-302-367), a combination which has been used successfully to reprogram somatic cells to iPSCs. These episomal vectors replicate and are lost only slowly during proliferation (5 % per cell division). In addition, it has been shown that episomal vectors can still be integrated into the genome. This raises several questions for the current study:
1. Do the cells go through a transient phase of "pluripotency" or multipotency? By analyzing expression of further pluripotency markers at further timepoints this can be tested.

Answer: See supplemental Fig2 on this issue.

The reviewer pointed out a very critical question and we have performed the experiments suggested by the reviewer. Based on the data generated in supplemental Fig 2C, we found no evidence that pluripotency or multipotency occur during reprogramming. In sum, we have not detected the pluripotency markers (endo*POU5F1, NANOG*) in such a nine-day reprogramming process via qPCR data analysis (Extended Data Fig.2A), or flow cytometry for NANOG in Extended Data Fig2C. Moreover, we extended the reprogramming induction time as suggested by the reviewer, and analyzed with qPCR more pluripotency markers, such as endo*POU5F1,* endo*KLF4,* endo*SOX2, NANOG*, *ESRRB*, *SALL4*, *LIN28A*, *DPPA3*, *DPPA5*, *DNMT3L*, *GATA3*, and show that only three markers, endo*SOX2, SALL4* and *LIN28A, are further induced* (Extended Data Fig.2D). However, all these three genes have been previously shown to be important for mesoderm differentiation (Tahara et al., 2019) (Takemoto et al., 2011) (Robinton et al., 2019). So, these results suggest that the canonical pluripotent network of TFs are not activated during this process.

2. What do the cells express that differentiate to PSM-like cells? In other words: How does induction of these different factors influence the (trans-)differentiation process? Co-staining of differentiated cells for these induced pluripotency markers and PSM markers can for instance be used to test this.

Answer: Again, we appreciate these questions and indeed paid special attentions to them during the course of our studies. We have detected endo*SOX2, SALL4* and LIN28A being further induced during the UiPSM reprogramming (Extended Data Fig.2D), but not other pluripotent markers such as *endoPOU5F1*. Based on this, we hypothesize that Sall4 may play a role in PSM generation. To test this, we overexpressed *SALL4, LIN28A* and additional *POU5F1* with lentivirus vector during the UiPSM induction process, show that *SALL4* could improve the PSM induction efficiency, while extra expression of *POU5F1* reduced, Lin28A has no impact on, the efficiency of UiPSM colonies formation (Fig.1A below). To further clarify the role of *POU5F1* and *SALL4* in the UiPSM colonies maintaining, we generated UiPSM subclones with knockout of total POU5F1 (endogenous expression and episomal vectors carried) and SALL4, we found that *SALL4* deletion significantly decreases UiPSM subclones, while total *POU5F1* deletion had little effect (Fig.1B,C below). These results suggest that *SALL4, as suggested by the reviewer, may play a critical role* in both induction processes and UiPSM maintenance, while

POU5F1 may even potentially suppress the induction process, and *POU5F1* is not required for the UiPSM self-renewal.

Figure for reviewers removed

In addition to SALL4 and Lin28A, we show that *T, MIXL1, TBX6* and *CDX2* are induced during the process (Fig. 1D, article). In the extra datasets we prepared for this rebuttal, Fig 2A (below) shows that CDX2 is co-expressed with endogenous SOX2 (See above for the fact that SOX2 is expressed in both pluripotent and PSM states). The pluripotent marker NANOG is not induced (Fig.2B), consistent with answers in 1. So, it is plausible that these mesoderm and PSM specific TFs are responsible for the differentiation into PSM-like cells. As for the cocktail, we show that CHIR99021 plays a decisive role during UiPSM induction (Fig.2C, D).

Figure for reviewers removed

3. Do the cells that keep on proliferating in a progenitor-like state for various passages still contain the episomal DNA? By staining for the induced factors and by qPCR this can be tested.

Answer: Please see Fig 3 below with additional data for this rebuttal.

We appreciate this question and have tested this as suggested. As shown in Fig 3 below, the episomal vectors are present in cells at various stages in minute quantity, presumably integrated into the genome as shown with qPCR data analysis of UiPSM colonies at various passages, including tg*POU5F1* , tg*KLF4,* tg*SOX2* and *SV40LT*, but did not detect endogenous expression of *POU5F1* and *KLF4,* only activated endogenous expression of *SOX2*. The expression patterns of *miR302-367* cluster showed variable integration (Fig.3A). We next used PCR to identify the integration of the components of episomal vector onto the host genome, suggesting integration of pEP4EO2SET2K and an individual specific integration of pCEP4-miR302-367 cluster (Fig.3B). EBAN1 as a key factor that determines the episomal vector's integrity is also expressed (Fig.3A,C).

Our lab had earlier successfully reprogrammed human derived urine cells into induced pluripotent stem cell (UiPSC) and induced neural progenitor cell (UiNPC) via transducing the same episomal vectors(Li et al., 2016; Wang et al., 2013), suggesting that these vectors work nicely in urine cells, making it easy to respond to the inductive chemical environment. The episomal vectors in our system mainly drove urine cells into a ‘plastic intermediate’ or ‘progenitor’ state, in a similar fashion as reported for HHFs (Kurian et al., 2013) (Zhu et al., 2014).

In the case of POU5F1 or OCT4, we have performed knockout experiments for both endogenous and ‘tg’ and show in Fig 1 (above) that it is not required for maintaining PSM state. So, it seems that these vectors may play minimal role in maintaining PSM.

Figure for reviewers removed

4. The identity of the different cell types at the consecutive stages of reprogramming and differentiation to PSM does not become entirely clear. Even though RNAseq and qPCR of selected markers is performed, it remains difficult until the end to grasp what these cells exactly are. It would be useful to compare the obtained scRNAseq data to available scRNAseq data from human gastruloids (Morris et al. 2020), in vitro differentiated PSM and/ or mouse embryos at different stages (e.g. van den Brink et al. 2020). This would, by the way, also allow a better characterization of the cell identity of cells in the "failed branch(es)" of cell trajectories in Figure S1G (what are these cells?). Furthermore, expression of further pluripotency would be useful.

Answer: See Fig1 and supplement Fig2 in the article on this issue.

Naomi Moris et al established the human gastruloids, they applied tomo-sequencing(tomo-seq), while no scRNAseq data was provided in the article(Moris et al., 2020). van den Brink et al actually collected scRNA-seq dataset from E8.5 mouse embryos(van den Brink et al., 2020), which was not applicable for analysis and comparison with our UiPSM data.

We ultimately referred to the scRNA-seq dataset from human iPSC differentiated PSM established by Margarete Diaz-Cuadros et al(Diaz-Cuadros et al., 2020). We then analyzed the published differentiation data and the whole UiPSM reprogramming data and found that the data of UiPSM reprogrammed on day9 could partially map onto the data of differentiated PSM (d1 NMP, d2 MPC, d3-4 aPSM), but no overlapping distribution with d0 iPSC (Fig.1E), which suggested Day9 UiPSM has both PSM features and the features of early differentiated cells, but no pluripotent cells. Furthermore, we analysed the data of Day9 UiPSM combining with d3-4 aPSM, UMAP projection and showed a significant overlap, and the data could be distributed into 8 clusters (Fig.1F,G). GO analysis further show that Day9 UiPSM and d3-4 aPSM are distributed in cluster1/2/4/5 cells related to the somitogenesis, somite development and pattern specification process, and highly expressed PSM specific genes, including *HES7, DLL3, CDX1, FGF8, NOTCH1, WNT5A, SALL4, SOX2, TBX6, T, MIXL1, LEF1, WNT5B, HOXB1, MSX1*; cluster4 mainly distributed d3-4 aPSM cells related to the stem cell development, which might be stem cells that are not fully differentiated into PSM, and also highly expressed PSM specific genes. These results suggest that there is significant similarity between Day9 UiPSM and d3-4 aPSM, while Day9 UiPSM still have a small group of cells failing to become PSM. Moreover, the unique metabolism-related characteristics may be due to the complexity of the UiPSM reprogramming process.

Please refer to supplement Fig2 in the article about 'failed branches'.

We reconstructed trajectory from all single-cell transcriptomic data throughout the whole UiPSM reprogramming process and the above published human iPSC differentiated PSM data (d1 NMP, d2 MPC and d3-4 aPSM) by Monocle2(Diaz-Cuadros et al., 2020). All of these cells could be divided into three states (P1, successful branch; P2, failed branch; P3, pre-branch), the only one failed branch could match the larger failed branch in the UiPSM reprogramming, mainly distributed cells of UiPSM reprogramming on day3 and day6, a little of d3-4 aPSM, a few of d2 MPC and Day9 UiPSM (Extended Data Fig. 2G-I). Further analysis indicated that cells in this branch not only expressed pre-branch specific genes, *PAX2, PAX8, SAA1,* associated with nephron tubule development, but also expressed successful branch specific genes, *CDX2, DLL3, HES7, TBX6, SOX2, SALL4, LEF1, MIXL1, T,* relevant to somite development (Extended Data Fig.2J,K), suggested cells this branch both contained characteristic features of uninduced urine cells and PSM. GO analysis showed that these branch cells have no corresponding tissue cell type, while mainly correlated with cell adhesion and DNA damage (Extended Data Fig.2L). Above all, the major cells in failed branch were undergoing a reprogramming process, this could be attributed to the timing of urine cells

in response to reprogramming induction. In conclusion, the failed branches are shared between reprogramming and differentiation based on the datasets from others.

5. Various combinations of small molecule inhibitors are presented throughout the manuscript, e.g. during the reprogramming stage or when maintaining the UiPSM cells in a progenitor stages over several passages. It is said at several places that an optimized medium composition is used. However, the reasoning for using these particular combinations does not become clear and re-writing these sections and clarifying this would greatly improve the readability and understandability of the paper. For instance, I would expect to see a reference to a previous publication in which this or a similar combination has been used. Alternatively, I would expect to either read an explanation based on known literature why a certain drug was used or the description of some sort of inhibitor screen (as they do for the optimization of their gastruloid/ "somitoid" protocol).

Answer: We have made the following additions and modifications in the document.
For the induction medium of UiPSM:

We apologize for not including those information regarding the inhibitors and signaling molecules. Here are the details as described below. Based on analysis in other systems, the presomitic mesoderm could be induced from hPSCs in vitro, by activating Wnt ß-catenin signaling pathway (Chu et al., 2019; Moris et al., 2020; Xi et al., 2017). Besides, the winged-helix transcription factor Foxh1, a regulator of Nodal signaling during the development process of anterior-posterior (A-P) patterning, could be stimulated by inhibited histone H3K79 methyltransferase DOT1L expression (Halstead and Wright, 2015; Wang et al., 2019). In addition, the epidermal growth factor (EGF) signaling is essential for self-renewal niche stem cells (Date and Sato, 2015). Taken together, the epithelial cells were electroporated with pEP4E02SET2K and pCEP4-miR-302-367, followed culturing with the chemical cocktails (WNT agonist CHIR99021, bFGF, Dot1L inhibitor EPZ5676 and EGF).

For the maintaining medium of UiPSM, we must use a different medium as the picked UiPSM colonies are not suitable for long-term culture in the above inductive medium. To this end, we show that removing EPZ5676 could improve maintenance. Moreover, Diaz-Cuadros et al treated human ES and iPSC with CHIR99021 supplemented TGF-ß signal inhibitor (LDN193189), could form PSM cells (Diaz-Cuadros et al., 2020). Therefore, after selecting a suitable TGF-ß signal inhibitor, we developed a defined medium (hereafter, DM medium) including CHIR99021, TGF-ß inhibitor A8301, bFGF and EGF for maintaining UiPSM.

For the aggregation and elongation medium of UiSomitoid:

We referred to the protocol of Naomi Moris et al established gastruloid(Moris et al., 2020). We seeded 100~1000 UiPSM cells in a well of low-adherence plates, found 300~600 UiPSM cells could aggregate into a sphere up to 150 μm in diameter with a high probability when treated with 3uM CHIR99021 (Extended Data Fig. 5A,C), while these spheres could not extend symmetrically. Recent studies have illustrated the dynamic of Wnt, Nodal and BMP signaling controlling fate patterning of human self-organized gastruloids(Chhabra et al., 2019; Moris et al., 2020). Therefore, we optimized conditions for somitoid elongation with inhibitors for BMP, WNT, Nodal signaling (Extended Data Fig. 5B), and show that CHIR99021 (3μM) and SB431542(5μM) are optimal for 300~600 seeding cells (Extended Data Fig. 5C-E).

6. A small molecule inhibitor cocktail is presented that is used to maintain UiPSM cells long-term in culture. If this is solely dependent on these small molecules (and not the reprogramming factors, see above), then it should be possible to use these molecules to maintain primary or in vitro differentiated progenitor cells in culture in a similar way. If this is possible, this would be an exciting finding for the field.

Answer: Indeed, the maintaining medium, DM, may be useful for future experiments to determine if it can be used to maintain PSM isolated by other means.

In preliminary results presented below, we show that DM cannot maintain the differentiated cells. Firstly, hES treated with a published medium (DMEM/F12 supplemented with 1% ITS and 3μM CHIR99021) for three days to differentiate into PSM, then changed into UiPSM maintaining medium (DM: Advanced DMEM/F12 supplemented with 3μM CHIR99021, 1μM A8301, 10ng/μl bFGF and 5ng/μl EGF), unfortunately, these hiPSM cells could not survive (Data not show). Secondly, we tried to use inductive medium like urine cells (IM: Advanced DMEM/F12 supplemented with 3μM CHIR99021, 5μM EPZ5676, 10ng/μl bFGF and 5ng/μl EGF) to induce hES for three days, and then, these HiPSM cells were digested and planted with Y27632 (5μM) into new plate in a 1:6 ratio. The results of immunofluorescence and qPCR data analysis indicated that HiPSM induced in IM highly expressed PSM specific genes (*T, MIXL1, TBX6, CDX2, DLL3*), while these genes gradually reduced in HiPSM colonies when proliferated five passages (Fig.4A-D). As shown throughout our paper, there are differences among the resulting cells generated by reprogramming and differentiation. We need to identify better conditions that can keep both in vitro.

Figure for reviewers removed

7. For the quantification of Hes7 and Mesp2 oscillations and the GFP control, I would expect to see the real raw or smoothened/detrended signal of several samples (Figure 3E and Figure S3G). In addition, the movies corresponding to Figure 3E, Figure S3G and Figure 4A should be added to the paper.

Answer: See Fig3 and supplement Fig6 in the article about HES7 and MESP2 oscillations.

We appreciate these comments and have now provided the info as requested. Specifically, the recorded picture signal is converted into a digital signal via ImageJ, the

obtained 8pepper fluorescence signal of the UiSomitoids were calculated via ImageJ and then recorded the mean fluorescence intensity for subsequent analysis. Detrend was performed first to remove trend and enhance data quality. When appropriate, the moving average was subtracted with window size of 10 units and then data was normalized between 0 and 1. To remove noise, the Sgolay filtering was applied. Finally, for smoothening data, we applied RLOESS. All operations were done in MATLAB. The plots were generated in R using ggplot2 package. We also applied scale function to make it much clearer.

In addition, the mean fluorescence intensity and corresponding time points were fitted as sine wave with nonzero baseline (Y=Amplitude*sin((2*pi*X/Wavelength) +PhaseShift) + Baseline) in prism 8.0. A phase change is defined as the time for the reporter cell line undergoing an oscillation, the oscillation transfer period. Moreover, we drew scatter plots of the calculated periods, Mean±s.d.

Based on this, we detected the oscillation signal value of three HES7-8pepper report cell lines and three MESP2-8pepper reporter cell lines. Besides, we also build a lenti-GFP UiPSM cell line as a negative control for the waken GFP expression (Extended Data Fig.6H). We show the latter is almost impossible to oscillate, or a few samples with weak oscillating signals also show disordered periods and minimal amplitude changes. Conversely, HES7-8pepper reporter and MESP2-8pepper reporter cell lines maintain relatively stable periodic variation (Fig.3E and Extended Data Fig.6E, F).

Minor points:

- In the introduction, it says that PSM is derived from the "CLE of the tailbud". This is not complete. In early stages of somitogenesis PSM is indeed generated from the CLE. At later stages, once the tailbud has formed, the NMPs reside in a region termed the chordoneural hinge (CNH).

Answer: we appreciate this and made changes as the follows:

"Somites emerge anteriorly from presomitic mesoderm or PSM which is derived from the caudal lateral epiblast (CLE) and adjacent node streak border (NSB) of E8.5 mouse embryos, then migrate from E10.5 chordoneural hinge (CNH) of tail bud(Cambray and Wilson, 2007; McGrew et al., 2008; Wilson et al., 2009)."

- In the introduction, van den Brink et al. is cited twice. However, this reference is not really fitting at these points and other papers might be more useful.

Answer: Thanks, and we have corrected this.

- In the optimization of the gastruloid protocol, it is not clear how this optimization is done. Is elongation the only readout or are also the present cell types taken into account?

Answer:

We appreciate this question. The well-known somitogenesis clock mechanism depends on time–space translation that distinguishes zones along the anterior–posterior (A–P) axis, providing precision in the posterior part of the axis(Durston, 2015). Thus, we needed to establish such an organization with A–P axis to explore whether UiPSM cells can mimic somitogenesis. Therefore, we decided to focus on the elongation of UiSomitoid to build A–P axis for optimization. Of course, we have also constantly monitored the expression of somitogenesis relevant genes when we optimize the extension medium via q-PCR and bulk RNA-seq (Data not shown). As the cells responded well, so we did not consider any other cell types.

Further point:

The language is sometimes not perfect with typos and words missing. It would be useful to proofread this entirely again.

Answer: we have made changes.

Referee #2:

General summary and opinion about the principle significance of the study, its questions and findings

The manuscript describes a novel method for the generation of human PSM cells using somatic cells (epithelial cells harvested from urine). This research is timely as it is the first model system for PSM of its kind and comes at a time where somitogenesis research is being transformed by the alternative approach of the generation of new human model systems using stem cells. The authors demonstrate a moderate efficiency in the generation of PSM cells which they use to generate somite like structures. The manuscript would benefit from some additional data and further clarification.

Specific major concerns essential to be addressed to support the conclusions
1. The authors use T expression as an indicator of the percentage of cells that achieve PSM status and present a representative example (Fig 1E) showing 44% T positive cells. It is not clear what this data represents as Fig 1L shows data for the three donors that are used in this study with 56, 34 and 17% T positive. These percentages do not suggest an efficient differentiation protocol and therefore the protocol would benefit from further improvement. Moreover, the number of donors used is very small (three) and analysing the reprogramming efficiency for a larger number of donors would improve the impact of this manuscript. From Fig 2 onwards the authors use optimised induction medium (DM). Does differentiation using this medium result in higher percentages of T positive cells? Data demonstrating the efficiency using DM is lacking.

Answer: Please also See supplement Fig1 in the article.

We appreciate these questions. We try to answer them in the following manner: First, we wish to stress that the percentages presented vary from experiments to experiments as well as donor to donor. But, we have been successful in generating PSM 100% despite the varying degree of efficiencies (we wish to stress that even a small percentage such as 17% is in fact 100% success as we can expand the resulting PSM and obtain self renewal PSM indistinguishable from other “more” efficient experiments. So, the efficiency may matter very little based on what we wish to achieve, i.e., obtaining PSM through reprogramming). Since this is not a method paper, we did not attempt to standardize the protocol which may require additional experiments or optimizations. We plan to do so in coming years so that this may become a robust platform for PSM and somitogenesis studies. Therefore, we decided to focus on the characterization of the PSM generated. Without demonstrating the function of PSM, it may not be productive to undergo those optimizations and standardizations. So, please forgive our biased focus on functional characterization. Nevertheless, we did apply this method to 18 urine cell samples from 10 healthy volunteers aged 20-40 years in our lab (Extended Data Fig.1H). We found that the induction efficiency varies from 15% to 60% due to individual differences regardless of gender. The representative example (Fig. 1C) shows 44% T positive cells reprogrammed from the Donor1 Urine cell (UC01), this UiPSM clone provided the main data in the article without additional description.

Secondly, In the early exploration of reprogramming induction system, the medium ‘DM’ could induce a few scattered UiPSM colonies and show an extremely low induction efficiency. **As show below**, CHIR99021 and EPZ5676 in induction medium played major roles in the inductive process, while, continuing adding EPZ5676 is not conducive to the growth and maintenance of UiPSM colonies (Fig.5A, B below). In contrast, we found that A8301 in ‘DM’ would inhibit the reprogramming process, but benefit to the UiPSM maintenance or self renewal.

Figure for reviewers removed

2. The data in Fig S2A shows variable expression levels of PSM markers, on several occasions this is clearly reduced in the higher passage numbers. This at the very least suggest that these cell lines are not consistent in their expression patterns when being passaged and it possibly means that the PSM identity might be reduced.

Answer: See supplement Fig4 in the article on this issue.

Actually, qPCR and immunofluorescence data analysis showed that the expression of PSM specific genes decreases after the 18th generation of the UiPSM colonies. We next compared the scRNA-seq data of the UiPSM clone at P10 with Day9 UiPSM and the published data (d3-4 aPSM), all these data could be clustered into 6 groups (Extended Data Fig.4A,B). Moreover, cells in cluster 0,2 both were distributed with Day9 UiPSM, UiPSM_P10 and d3-4 aPSM, highly expressed PSM specific genes, *TBX6*, *CDX2, HOXB1, DLL3, HES7, TBX6, T, MIXL1* and *WNT5A,* related to somitogenesis and somite development. Furthermore, UiPSM_P10 had a higher distribution in cluster 1,4 showed a higher proportion of cells expressing NMP specific genes, *SOX2* and *SALL4*, and a lower proportion of cells expressing PSM specific genes (Extended Data Fig.4C-F), suggested cells of UiPSM_P10 clone were still better enriched cells maintaining PSM characteristic features, increased the proportion of cells expressing NMP marker genes. In addition, we counted the proportion of cells co-expressing PSM specific genes in the sequenced cells of UiPSM_P10, SOX2 and TBX6, TBX6 and CDX2, both sharply decreased to 30%. T and MIXL1, T and SALL4 both slightly reduced to 40% as well. SOX2 and T could still maintain 50% (Extended Data Fig.4G).

In conclusion, UiPSM_P10 clone in the medium 'DM' did show some reduction of the PSM specific cells, but still maintained the PSM characteristic features. Nevertheless, the reviewer is correct that some clones do experience reduction of marker expressions. However, they remain competent in generating somitoids. It is possible, like early studies of ESCs, that they oscillate through cycles of naive/primed/other stages of fate under the culture conditions applied. As stated earlier, we are trying to further optimized the culture conditions such that more “perfect” PSM can be obtained and maintained.

3. Sox2 expression is used as a somitogenesis marker as well as a pluripotency marker. Therefore, the manuscript would be more convincing if an alternative somitogenesis marker would be chosen to focus on throughout the manuscript. Fig 1H shows that Tbx6 and Sox2 expression is complementary rather than co-expressing. The coexpression data in Fig 1J would benefit from an explanation as the top and bottom panel look rather different but still result in the same conclusion.

Answer: See supplement Fig1 in the article on this issue.

Figure1 E showed the scRNAseq data of UiPSM on day9 can cover the human iPSC differentiated d1 NMP, d2 MPC and d3-4 aPSM, suggested the UiPSM cells in our system as progenitor cells had all the characteristics of NMP and PSM, PSM progenitor cells. Although *SOX2* repressed mesodermal control genes *T* and *TBX6* to drive the NMP cells into neural lineages, *SOX2* as NMP markers competitively co-expressed with *T* or *TBX6* on the posterior tail bud, so we counted the frequency of cells co-expressed PSM specific genes, *SOX2* and *TBX6*, *SOX2* and *T*, *TBX6* and *CDX2*, *T* and *SALL4*, all could reach around 50%, *T* and *MIXL1* could also exceed 70% (Extended Data Fig.1E), supporting that UiPSM had the features of tail bud. We also counted the proportion of cells co-expressing PSM specific genes, *TBX6* and *CDX2*, *T* and *SALL4*, *T* and *MIXL1*, that also reach 50% (Extended Data Fig.1E), suggesting that the reprogramming efficiency from donor1 could achieve 50%.

4. The description of Fig 3E states that Hes7 Pepper oscillating fluorescence is gradually decreasing but this is not obvious from the Figure. How were these Figures generated? There seems to be an unnatural uniformity to these oscillations, making them look like a model rather than actual data. Similarly, the graph in Fig SG is remarkably straight and level.

Answer: See Fig3 and supplement Fig6 in the article about HES7 and MESP2 oscillations.

We appreciate these questions and have made the appropriate revisions in the text and figures.

Specifically, the recorded picture signal is converted into a digital signal via ImageJ. The obtained 8pepper fluorescence signals of the UiSomitoids were calculated via ImageJ and then recorded the mean fluorescence intensity for subsequent analysis. Detrend was performed first to remove trend and enhance data quality. When appropriate, the moving average was subtracted with window size of 10 units and then data was normalized between 0 and 1. To remove noise, the Sgolay filtering was applied. Finally, for smoothening data, we applied RLOESS. All operations were done in MATLAB. The plots were generated in R using ggplot2 package. We also applied scale function to make it much clearer.

In addition, the mean fluorescence intensity and corresponding time points fitted sine wave with nonzero baseline (Y=Amplitude*sin((2*pi*X/Wavelength) +PhaseShift) + Baseline) in prism 8.0. A phase change is defined as the time for the reporter cell line undergoing an oscillation, the oscillation transfer period. Moreover, we drew scatter plots of the calculated periods, Mean±s.d.

On this basis, we detect the oscillation signal value of three HES7-8pepper report cell lines and three MESP2-8pepper reporter cell lines. Besides, we also build a lenti-GFP UiPSM cell line as a negative control for the GFP expression (Extended data Fig.6H), we show the latter does not oscillate. Conversely, HES7-8pepper reporter and MESP2-8pepper reporter cell lines maintain relatively stable periodic variation (Fig.3E and Extended Data Fig.6E, F).

5. The data in Fig 4A doesn't convincingly support the statement that Hes7-GFP expression regresses posteriorly.

Answer: See Fig4 and supplement Fig7 in the article on this issue.

We have selected the relevant representative photos and used the model diagram to represent the description.

Minor concerns that should be addressed

1. For qPCR data much more info needs to be provided regarding quality control (e.g. MIQE). The expression levels are normalised against GAPDH only. Can the authors provide evidence that GAPDH expression is consistent throughout the differentiation protocol? The qPCR data would benefit from the use of more than one housekeeping gene for normalization.

Answer: As shown below.

The most commonly used housekeeping proteins are β-actin, β-tubulin, and glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) has long been recognized as an important enzyme for energy metabolism and the production of ATP and pyruvate through anaerobic glycolysis in the cytoplasm. We actually did not test whether our system was involved in the glycolysis metabolic pathway. But most importantly, the normalized gene counts of *GAPDH, CDX2, TBX6, SALL4, SOX2, DLL3* of bulk RNA-seq sequencing data from various samples in one experiment with same treatment, showed a consistent expression of GAPDH and hESC-specific expressions of *SALL4* and *SOX2* (Fig.6A). *GAPDH* also consistently expressed in the samples sequenced at the same time in the two experiments of UiPSM reprogramming, PSM specific genes, *CDX2, TBX6, SALL4, SOX2, DLL3,* also expressed

incrementally during UiPSM reprogramming (Fig.6B). In addition, the normalized gene counts of *GAPDH* in each slice of sample in UiSomitoid showed a constant expression via Geo-seq sequencing, and the expression pattern of PSM specific genes conformed with the differential expression patterns along the anterior and posterior axes of UiSomitoid (Fig4. C). Therefore, the normalized gene counts of GAPDH for all samples were quite constant in one experiment, and the count value varied from each experiment depending on the depth of sequencing

Moreover, for the qPCR experimental operation, we collected 2μg total RNA from cell samples dissociated with trizol to use for reverse transcription, and then controlled constant loading sample (7ng), the threshold cycle (Ct) of *GAPDH* of all samples in all our qPCR experiments ranged from 15 to 17 under strict control when fixed the baseline, these effective samples could be used for subsequent statistical analysis. These results suggested that *GAPDH appears to be expressed quite* constantly, thus, as an internal reference control in our system. Based on these arguments, we did not use another housekeeping genes. If the reviewer feels strongly about this, we could repeat this with fresh samples.

Figure for reviewers removed

2. The description of the generation of the CRISPR cell lines is rather minimal. There is a need for more details, especially including information on the quality control experiments performed to establish that the new cell lines are exactly what they are meant to be and no additional changes have occurred. More information is needed regarding the quality control of the RNA seq samples. Additional information is required on the histological

analysis method. There is insufficient information on quantification, sample numbers and statistics throughout the paper.

Answer: See supplement Fig6 in the article about CRISPR cell lines.

We apologize for these. We have provided those information in Extended Data Fig 6. Specifically, we have increased the number of the UiPSM colonies generated CRISPR cell lines including UiPSM-HES7-8pepper and UiPSM-MESP2-8pepper cell lines on display.

We have made some additions for sample numbers and statistics, as stated in legends. All the sequencing data in this text are to further support the ideas put forward from our experimental data. If repeating all the sequencing data, the trial period will also be longer, and the cost will be very high. In addition, experimental data and sequencing data can be mutually verified to support our points.

3. Towards the end of the introduction there is a statement that suggests that human gastruloids are the only species that contain three germlayers. This sentence would benefit from reformulating.

Answer: we have made some modifications. As shown below.

However, human gastruloids generated from pluripotent cells still possess endoderm development and express neuroectoderm cell associated genes (OTX2, SOX1 and SOX9).

4. The details of the media used in Fig1 and S1 are not described in the methods section.

Answer: we have made some additions on the induction medium in the document.

For the induction medium of UiPSM:

We apologize for not including those information regarding the inhibitors and signaling molecules. Here are the details as described below. Based on analysis in other systems, the presomitic mesoderm could be induced from hPSCs in vitro, by activating Wnt ß-catenin signaling pathway (Chu et al., 2019; Moris et al., 2020; Xi et al., 2017). Besides, the winged-helix transcription factor Foxh1, a regulator of Nodal signaling during the development process of anterior-posterior (A-P) patterning, could be stimulated by inhibited histone H3K79 methyltransferase DOT1L expression (Halstead and Wright, 2015; Wang et al., 2019). In addition, the epidermal growth factor (EGF) signaling is essential for self-renewal niche stem cells (Date and Sato, 2015). Taken together, the epithelial cells were electroporated with pEP4E02SET2K and pCEP4-miR-302-367, followed culturing with the chemical cocktails (WNT agonist CHIR99021, bFGF, Dot1L inhibitor EPZ5676 and EGF).

5. Fig 2I-J show two clusters (M5+6) that are consistent with muscle development. These clusters are rather small in comparison to the other clusters. Further explanation of what this means would make it easier to interpret the data.

Answer: As statement below.

M5+6 are small indeed, but they show more mature muscle or mesoderm features so we highlighted them. M1 is also clearly mesoderm, but of early features. We show that cluster M1 can be further divided into 7 subgroups in Fig S3. F and G, and even further divided.Actually, the GO analysis showed that cells in M1 related to Wnt signaling pathway and axis specification, suggesting cells in M1 may directly related to somite development. We wished to show the diversity of cells from PSM. The exact ratio and balance remains difficult for us to determine accurately. We should pay special attention in the future on this.

6. The discussion states limitations regarding hESC and hiPS for studying PSM function. This is followed by an example of reporter studies demonstrated in the manuscript

suggesting that this is an advantage of the described system. However, these types of experiments have been described in hiPS derived PSM cells too and these cells can be differentiated much more efficiently into PSM.

Answer: As shown below.

The reviewer is correct. What we mean is that the PSM from hiPSC or ESC also contain cells from other lineages such as *GATA4* and *GATA6*, although hES and hiPSC could differentiate much more efficiently into PSM (see below in Fig 7a, b)

Figure for reviewers removed

7. The discussion also states that the UiPSM system may serve as a chronological model for ageing and rejuvenation research. This statement is too strong and is premature based on the data provided.

Answer: We appreciate this advice, we have removed the statement in the article..

8. The colour schemes of several figures don't allow of proper analysis of the data as the colours used are too similar. Moreover, some of the colours are so pale they don't print well and in some figures the colours green and red are used which aren't very accessible for colour blind readers. Some figures contain extremely small text.

Answer: See supplement Fig.1E in the article on this issue.
we have made the required modifications. We replaced the color chart with a bar chart

9. Is there a reason why the growth curve for UC is so much shorter than for UiPSM (Fig 2C)?

Answer: As shown below.

Referring to the description of Shantaram Bharadwaj's (Bharadwaj et al., 2011), the proliferation ability of urine cell is poor. We isolated and collected epithelial-like cells from urine samples displaying a distinctive morphology and rapid proliferation within 3 passages when cultured in REGM medium, while the proliferation rate is much lower than UiPSM clone. Only a few urine cells could be passaged for more than 10 generations. Most of urine cells gradually undergo apoptosis when passaging more than 3 to 5 generations.

10. In Fig S1 it would enhance the data if the three controls (DE, NPC and hESC) were included in all three graphs as they would function as either positive or negative controls.

Answer: See supplement Fig.2B in the article on this issue.

Due to adjustment of part of the figures, the original Fig S1.A has been rearranged to supplement Fig.2A. The enhanced data(IF results) were arranged on Fig.2B.

The hESC(H9) was purchased from ATCC, we also detected the co-expression of DE marker, SOX17 and FOXA2 and NPC marker, Nestin and Tuj1 via immunofluorescence (Fig. 8). We selected DE, NPC and hESC as positive control for detecting the expression of endodermal and ectodermal specific genes, supported that there was no detected expression of these dermal genes during the whole UiPSM reprogramming process.

11. Fig S2A there are no error bars or an indication of the number of biological repeats.

Answer: See supplement Fig S3A in the article on this issue.
We have made some additions. The number of biological repeats reached 3.

12. Fig S3C-E would benefit from a larger number of examples for each condition. Additional info in the legends of Fig S3F regarding top, middle and bottom panels is required.

Answer: See supplement Fig S5C-E in the legend on this issue.
We did not make it clear about the experiment repeats in the early manuscript. We have added the illustration of experimental data in the legend.

Additional non-essential suggestions for improving the study
1. The manuscript would benefit from corrections of spelling, grammar and typos.

Answer: We have done further editing and corrections.

2. The results section that describes Fig 2E-J would benefit from an introducing sentence explaining the objective of these experiments.

Answer: We have made addition. See as below.
To further determine the differentiation potential and tumorigenicity of UiPSM with self-renewal ability in vivo.

3. In Fig 1J the top and bottom figures have been swapped (based on the legend).

Answer: We have revised the above-mentioned contents.

4. References to Brachury/T/Tbxt are not consistent. In the main text T is mostly used while in some of the figures it is referred to as Tbxt.

Answer: We have revised the above-mentioned contents.

References:

Bharadwaj, S., Liu, G., Shi, Y., Markert, C., Andersson, K.E., Atala, A., and Zhang, Y. (2011). Characterization of urine-derived stem cells obtained from upper urinary tract for use in cell-based urological tissue engineering. Tissue Eng Part A *17*, 2123-2132.

Cambray, N., and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. Development *134*, 2829-2840.

Chhabra, S., Liu, L., Goh, R., Kong, X., and Warmflash, A. (2019). Dissecting the dynamics of signaling events in the BMP, WNT, and NODAL cascade during self-organized fate patterning in human gastruloids. PLoS Biol *17*, e3000498.

Diaz-Cuadros, M., Wagner, D.E., Budjan, C., Hubaud, A., Tarazona, O.A., Donelly, S., Michaut, A., Al Tanoury, Z., Yoshioka-Kobayashi, K., Niino, Y., *et al.* (2020). In vitro

characterization of the human segmentation clock. Nature *580*, 113-118.

Durston, A.J. (2015). Time, space and the vertebrate body axis. Semin Cell Dev Biol *42*, 66-77.

Kurian, L., Sancho-Martinez, I., Nivet, E., Aguirre, A., Moon, K., Pendaries, C., Volle-Challier, C., Bono, F., Herbert, J.M., Pulecio, J.*, et al.* (2013). Conversion of human fibroblasts to angioblast-like progenitor cells. Nat Methods *10*, 77-83.

Li, D., Wang, L., Hou, J., Shen, Q., Chen, Q., Wang, X., Du, J., Cai, X., Shan, Y., Zhang, T.*, et al.* (2016). Optimized Approaches for Generation of Integration-free iPSCs from Human Urine-Derived Cells with Small Molecules and Autologous Feeder. Stem Cell Reports *6*, 717-728.

McGrew, M.J., Sherman, A., Lillico, S.G., Ellard, F.M., Radcliffe, P.A., Gilhooley, H.J., Mitrophanous, K.A., Cambray, N., Wilson, V., and Sang, H. (2008). Localised axial progenitor cell populations in the avian tail bud are not committed to a posterior Hox identity. Development *135*, 2289-2299.

Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schroder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature *582*, 410-415.

Robinton, D.A., Chal, J., Lummertz da Rocha, E., Han, A., Yermalovich, A.V., Oginuma, M., Schlaeger, T.M., Sousa, P., Rodriguez, A., Urbach, A.*, et al.* (2019). The Lin28/let-7 Pathway Regulates the Mammalian Caudal Body Axis Elongation Program. Developmental Cell *48*, 396-405.e393.

Tahara, N., Kawakami, H., Chen, K.Q., Anderson, A., Yamashita Peterson, M., Gong, W., Shah, P., Hayashi, S., Nishinakamura, R., Nakagawa, Y.*, et al.* (2019). Sall4 regulates neuromesodermal progenitors and their descendants during body elongation in mouse embryos. Development *146*.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D.M., Lovell-Badge, R., Papaioannou, V.E., and Kondoh, H. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. Nature *470*, 394-398.

van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivie, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A.*, et al.* (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature *582*, 405-409.

Wang, L., Wang, L., Huang, W., Su, H., Xue, Y., Su, Z., Liao, B., Wang, H., Bao, X., Qin, D.*, et al.* (2013). Generation of integration-free neural progenitor cells from cells in human urine. Nat Methods *10*, 84-89.

Wilson, V., Olivera-Martinez, I., and Storey, K.G. (2009). Stem cells, signals and vertebrate body axis extension. Development *136*, 1591-1604.

Zhu, S., Rezvani, M., Harbell, J., Mattis, A.N., Wolfe, A.R., Benet, L.Z., Willenbring, H., and Ding, S. (2014). Mouse liver repopulation with hepatocytes generated from human fibroblasts. Nature *508*, 93-97.

Dear Duanqing,

Thank you for submitting your revised manuscript (EMBOJ-2022-110928R) to The EMBO Journal, as well as for your patience with our feedback, which got protracted by delayed reviewer input. Your amended study was sent back to the two referees for their re-evaluation, and we have received their comments, which I enclose below. As you will see, the experts stated that the work has been substantially improved by the revisions and they are now broadly in favour of publication, pending minor revision.

Thus, we are pleased to inform you that your manuscript has been accepted in principle for publication in The EMBO Journal.

Please consider the remaining minor comments of the reviewers carefully and amend the text discussion and data presentation accordingly where appropriate. We concur that showing the rebuttal data in the manuscript supplemental part (ref#1, pt.1, ref#2, pt.3) is a reasonable request.

Also, we need you to take care of a number of minor issues related to formatting and data annotation as detailed below, which should be addressed at re-submission.

Please contact me at any time if you have additional questions related to below points.

As you might have noted on our web page, every paper at the EMBO Journal now includes a 'Synopsis', displayed on the html and freely accessible to all readers. The synopsis includes a 'model' figure as well as 2-5 one-short-sentence bullet points that summarize the article. I would appreciate if you could provide this figure and the bullet points.

Thank you for giving us the chance to consider your manuscript for The EMBO Journal. I look forward to your final revision.

Again, please contact me at any time if you need any help or have further questions.

with
Best regards,

Daniel

Daniel Klimmeck PhD
Senior Editor
The EMBO Journal

Formatting changes required for the revised version of the manuscript:

>> Please provide maximally five keywords for the manuscript.

>> Amend the 'Disclosure and Competing Interests Statement' by indicating EMBO Journal editorial advisory board membership of D.P. .

>> Remove the author contributions information from the manuscript text. Note that CRediT has replaced the traditional author contributions section as of now because it offers a systematic machine-readable author contributions format that allows for more effective research assessment. and use the free text boxes beneath each contributing author's name to add specific details on the author's contribution. More information is available in our guide to authors.
https://www.embopress.org/page/journal/14602075/authorguide

>> Funding information: make sure all funding information is consistently entered in our online system.

>> Enter author information and correct email contact for W. P., J. K., J. H., Z. C., B. C. and S. O. in our online system.

>> Provide figures as individual, high-resolution image files.

>> Add the animal ethics information to the main manuscript material & methods part.

>> Please rename the current 'Data quantification' section to 'Statistical analysis'.

>> Dataset EV legends: The three movies need renaming to Movie EV1-3 and legends zipped.

>> Data availability section: please add a hyperlink to the database entry and make sure privacy is released before online publication of your article.

>> Please indicate redisplay data Fig3A in the legend of Figure S5B.

>> Appendix file: merge table with figures into one appendix figure with a table of contents list on its first page. Correct the nomenclature to 'Appendix Figure S1, S2...'; adjust callouts in the main text accordingly. The table needs title/description added.

>> Please consider additional changes and comments from our production team as indicated by the .doc file enclosed and leave changes in track mode.

Use the link below to submit your revision:

Link Not Available

Referee #1:

In the revised version of their manuscript, Qin et al. address many of the points I myself and the other reviewer had raised.

However, there are still a few points that require clarification.

- One of the points I had raised was the contribution of the episomal DNA to the transdifferentiation process and the long-term maintenance of generated UiPSM cells. The authors now clarify that this DNA does indeed get integrated into the genome and is expressed. In the rebuttal letter, data is shown that indicates that knockout of some of these factors does not influence the differentiation process (rebuttal letter figure 1B). In contrast, when depleting Sall4, differentiation is indeed impaired (rebuttal letter figure 1C). This part is not discussed in detail in the manuscript, but would be important information, in my opinion. And what is the contribution of other reprogramming factors?
Also in light of their preliminary data that other cells differentiated towards the PSM state cannot be maintained in the small molecular inhibitor cocktail to maintain cells in the PSM state (shown in rebuttal letter figure 4), this indicates that this integrated DNA might have an inducing function besides reprogramming.
I therefore think that further details should be added to the results section and this should also be discussed in the discussion section.

- As suggested, the authors compare their scRNAseq data to published data. However, in contrast to comparing to a scRNAseq dataset containing diverse cell types, they decided to compare their results to a dataset of iPSCs specifically differentiated towards PSM cells (Diaz-Cuadros et al.). They find that "the data of UiPSM reprogrammed on day9 could partially map onto the data of differentiated PSM (d1 NMP, d2 MPC, d3-4 aPSM)". It would have been more unbiased and clear if their results had been compared to a dataset that contains various different cell types of a developing embryo. The definition of the cell type would have potentially been easier to make.

- Two movies were added to show oscillations of Hes7 and Mesp2. The Hes7 movies (I assume, I couldn't find a description of the movies) do not show oscillations clearly, in my opinion. It seems as if the cells drift out of focus continuously (and therefore the signal decreases) and then the focus is suddenly readjusted (and the signal is stronger again). To rule out that oscillations are observed because of the repeated re-adjustment of the focus, cells should be imaged together with another non-oscillating channel (e.g. Hoechst can be used to stain the cells). At least the brightfield movie should be shown to visualize that the focus does not change periodically.

Minor points:
- Two movies lack a timestamp and scale bar.

- The language could still be improved.

Referee #2:

The manuscript describes a novel method for the generation of human PSM cells using somatic cells (epithelial cells harvested

from urine). This research is timely as it is the first model system for PSM of its kind and comes at a time where somitogenesis research is being transformed by the alternative approach of the generation of new human model systems using stem cells. The authors demonstrate a moderate efficiency in the generation of PSM cells which they use to generate somite like structures. The manuscript would benefit from some additional clarification.

1) The data in Fig S1E suggests that there are much bigger differences than similarities in expression patterns of differentiated UiPSM with the differentiation protocol used by Diaz-Cuadros et al. This suggests that the origin of the differentiated cells is more important than their differentiated state. However the text in the rebuttal (reviewer 1 - major point 4) suggests they are quite similar. Please explain and adjust the text.

2) Please provide more information regarding the quality control of the RNA seq samples as well as on the histological analysis method.

3) It would be worthwhile to include at least some of data used for the rebuttal in the supplementary section of the paper. The same applies for some of the explanations provided in the rebuttal. Including some of these would aid the reader.

Referee #1:

In the revised version of their manuscript, Qin et al. address many of the points I myself and the other reviewer had raised. However, there are still a few points that require clarification.

- One of the points I had raised was the contribution of the episomal DNA to the transdifferentiation process and the long-term maintenance of generated UiPSM cells. The authors now clarify that this DNA does indeed get integrated into the genome and is expressed. In the rebuttal letter, data is shown that indicates that knockout of some of these factors does not influence the differentiation process (rebuttal letter figure 1B). In contrast, when depleting Sall4, differentiation is indeed impaired (rebuttal letter figure 1C). This part is not discussed in detail in the manuscript, but would be important information, in my opinion. And what is the contribution of other reprogramming factors?
Also in light of their preliminary data that other cells differentiated towards the PSM state cannot be maintained in the small molecular inhibitor cocktail to maintain cells in the PSM state (shown in rebuttal letter figure 4), this indicates that this integrated DNA might have an inducing function besides reprogramming.
I therefore think that further details should be added to the results section and this should also be discussed in the discussion section.

Answer: Please see Appendix Fig S5 in the article.
we appreciate these questions and indeed paid special attentions to them during the course of our studies. Modifications has been added into the section of Results and Discussion.

For the results section: Notably, we have previously reported that iPSCs and NPCs colonies induced from urine cells have no detectable pEP4E02SET2K and pCEP4-miR-302-367 after 10 passages (Li et al., 2016; Wang et al., 2013). By contrast, no obvious loss of episomal vectors occurred as expected in the UiPSM colonies, until 18 passages. Indeed, EBAN1 as a key factor for episomal vectors is stably expressed (Malecka et al., 2019). We detected EBAN1 steadily expressed in UiPSM reprogramming process and UiPSM colonies, as well as episomal vector carrying factors *POU5F1, KLF4* and *SOX2* with exogenous expression at various UiPSM colonies. However, only *SOX2* was endogenously activated in induced PSM state. The miR302-367 cluster has variable integration (Appendix Fig S5A-C). Furthermore, urine cells failed to enter reprogramming process without the induction medium (Data not shown). We knocked out total POU5F1 and another activated factor (SALL4) in UiPSM reprogramming and show that POU5F1 is dispensable for maintaining PSM state, whereas SALL4 is requried for PSM maintenance (Appendix Fig S5D,E). These results suggest that these vectors, even being integrated at various sites, play minimal role in maintaining PSM.

For the discussion section: Another technical point awaits further clarification in the near future is the detection of residual reprogramming vectors. It would be highly desirable to generate integration free UiPSM. A few approaches may be attempted. First, the vectors may be further optimized such that integration is not possible. Secondly, mRNA based reprogramming may be utilized to avoid DNA based vectors. Lastly, Culture condition may be further optimized to lessen the length required for reprogramming, presumably minimizing integration.

- As suggested, the authors compare their scRNAseq data to published data. However, in contrast to comparing to a scRNAseq dataset containing diverse cell types, they decided to compare their results to a dataset of iPSCs specifically differentiated towards PSM cells (Diaz-Cuadros et al.). They find that "the data of UiPSM reprogrammed on day9 could partially map onto the data of differentiated PSM (d1 NMP, d2 MPC, d3-4 aPSM)".

It would have been more unbiased and clear if their results had been compared to a dataset that contains various different cell types of a developing embryo. The definition of the cell type would have potentially been easier to make.

Answer: Please see Fig 1 in the article.

Again, we appreciate these questions and indeed paid special attentions to them during the course of our studies. Unfortunately, we were unable to find the dataset of developing embryo during the period. However, we investigated the similarity by comparing data between Day9 and data of differentiated PSM (d1 NMP, d2 MPC, d3-4 aPSM). The results have been added into the figure. The results have revealed that Day9 UiPSM cells in cluster 0/1 almost completely covered the majority of d3-4 aPSM, showing expression of HOXB2, LEF1, DLL3, HES7, TBX6, WNT5A, FGF8 and CDX2, which were enriched for somitogenesis, somite development and pattern specification process. Those cells in cluster 3 highly resembled d1 NMP, showing expression of T, NODAL, SOX2, MIXL1 and SALL4. Besides, Day9 UiPSM cells in cluster 2 had a partial coincidence with d2 MPC and, related to Wnt signaling pathway (Fig. 1F-K). These results reveal our UiPSM possessed features of previously reported d1 NMP, d3-4 aPSM.

- Two movies were added to show oscillations of Hes7 and Mesp2. The Hes7 movies (I assume, I couldn't find a description of the movies) do not show oscillations clearly, in my opinion. It seems as if the cells drift out of focus continuously (and therefore the signal decreases) and then the focus is suddenly readjusted (and the signal is stronger again). To rule out that oscillations are observed because of the repeated re-adjustment of the focus, cells should be imaged together with another non-oscillating channel (e.g. Hoechst can be used to stain the cells). At least the brightfield movie should be shown to visualize that the focus does not change periodically.

Answer: We appreciate this question and indeed that was a concern. We have made the modifications. We added all brightfield movies to show oscillations clearly. The description of the movies showed in text like this: Indeed, we show that both HES7 and MESP2 RNAs oscillate ~300min in UiSomitoids (Fig 3E-G, Appendix Fig S7C-F, Movie EV1,2).

Minor points:
- Two movies lack a timestamp and scale bar.
Answer: Again, we appreciate this advice. We have made the supplementations.

- The language could still be improved.
Answer: Again, we appreciate this advice. We have made the modifications.

Referee #2:

The manuscript describes a novel method for the generation of human PSM cells using somatic cells (epithelial cells harvested from urine). This research is timely as it is the first model system for PSM of its kind and comes at a time where somitogenesis research is being transformed by the alternative approach of the generation of new human model systems using stem cells. The authors demonstrate a moderate efficiency in the generation of PSM cells which they use to generate somite like structures. The manuscript would benefit from some additional clarification.

1) The data in Fig S1E suggests that there are much bigger differences than similarities in expression patterns of differentiated UiPSM with the differentiation protocol used by Diaz-Cuadros et al. This suggests that the origin of the differentiated cells is more important than their differentiated state. However the text in the rebuttal (reviewer 1 - major point 4) suggests they are quite similar. Please explain and adjust the text.

Answer: Please see Fig 1 in the article.

Again, we appreciate this question and indeed that was a concern.

The text in the rebuttal (reviewer 1 - major point 4) described there was a similarity between Day9 UiPSM and d3-4 aPSM. It did not describe precisely here. Actually, the data Day9 UiPSM covered d1 NMP, d2 MPC d3-4 aPSM shown in Fig 1E. Via further analysis on this part of data, we found Day9 UiPSM not only shows the d3-4 aPSM feature, but also with d1 NMP feature and a small part of d2 MPC (Fig 1K-L). Therefore, there are differences between PSM differentiated from human iPSC and our UiPSM.

Hence, we adjusted these data in Fig 1F-K, and modified the former description.

2) Please provide more information regarding the quality control of the RNA seq samples as well as on the histological analysis method.

Answer: Please see Materials and Methods section in the article.

Again, we appreciate this question. The relevant details have been added into the section of Materials and Methods.

Quality control: All RNA-seq samples were assessed by FastQC before downstream analysis. Reads less than Q30 were considered as the low-quality reads and removed by Trimgalore. As for single-cell RNA-seq data, Q30 reads were more than 85%. Mapping rate were nearly 98%, where the average ratio of reads mapped onto exonic region were over 75%. More than 1200 genes were detected in each cell. As for bulk RNA-seq samples, mapping rates were no less than 80% in most samples. Furthermore, the number of detected genes ranged from 10000 to 15000. We do not display biological replicates in this project.

Histological analysis: The fresh graft tissues of UiPSM and UiPSC were isolated from mice after 2 months and a month. These tissues were fixed in 4% paraformaldehyde more than 48 hours and paraffin embedded. We transected each tissue into four segments of equal thickness. Then we continuously cut tissues into 8-μm sections along the cross section. 24 transverse sections corresponding to a 192-μm length in each segment were used for quantitative analysis. Histological analysis was done on these HES-stained sections.

3) It would be worthwhile to include at least some of data used for the rebuttal in the supplementary section of the paper. The same applies for some of the explanations provided in the rebuttal. Including some of these would aid the reader.

Answer: Please see Appendix Fig S5 in the article.

Again, we appreciate this advice and indeed paid special attentions to them during the course of our studies. We have supplemented the data for rebuttal in Appendix Fig S5 to help explain the function of episomal vectors carrying factors.

Dear Duanqing,

Thank you for submitting the revised version of your manuscript. I have now evaluated your amended manuscript and concluded that the remaining minor concerns have been sufficiently addressed.

Thus, I am pleased to inform you that your manuscript has been accepted for publication in the EMBO Journal.

Please note that it is EMBO Journal policy for the transcript of the editorial process (containing referee reports and your response letter) to be published as an online supplement to each paper. I would thus like to ask for your consent on keeping the additional referee figures included in this file.

Also, in case you might NOT want the transparent process file published at all, you will also need to inform us via email immediately. More information is available here:
https://www.embopress.org/page/journal/14602075/authorguide#transparentprocess

---

Please note that in order to be able to start the production process, our publisher will need and contact you shortly regarding the page charge authorisation and licence to publish forms.

Authors of accepted peer-reviewed original research articles may choose to pay a fee in order for their published article to be made freely accessible to all online immediately upon publication. The EMBO Open fee is fixed at $6,100 USD / £4,950 GBP / €5,500 EUR (+ VAT where applicable).

We offer two licenses for Open Access papers, CC-BY and CC-BY-NC-ND.
For more information on these licenses, please visit:
http://creativecommons.org/licenses/by-nc-nd/3.0/deed.en_US

Should you be planning a Press Release on your article, please get in contact with embojournal@wiley.com as early as possible, in order to coordinate publication and release dates.

On a different note, I would like to alert you that EMBO Press is currently developing a new format for a video-synopsis of work published with us, which essentially is a short, author-generated film explaining the core findings in hand drawings, and, as we believe, can be very useful to increase visibility of the work. This has proven to offer a nice opportunity for exposure i.p. for the first author(s) of the study. Please see the following link for representative examples and their integration into the article web page:
https://www.embopress.org/video_synopses
https://www.embopress.org/doi/full/10.15252/embj.2019103932

Please let me know, should you be interested to engage in commissioning a similar video synopsis for your work. According operation instructions are available and intuitive.

If you have any questions, please do not hesitate to call or email the Editorial Office.

Thank you for this contribution to The EMBO Journal and congratulations on a successful publication!

Please consider us again in the future for your most exciting work.

with
Best regards,

Daniel

Daniel Klimmeck, PhD
Senior Editor

The EMBO Journal
EMBO
Postfach 1022-40
Meyerhofstrasse 1
D-69117 Heidelberg
contact@embojournal.org
Submit at: http://emboj.msubmit.net

EMBO Press Author Checklist

Corresponding Author Name: Duanqing Pei
Journal Submitted to: The EMBO Journal
Manuscript Number: EMBOJ-2022-110928

USEFUL LINKS FOR COMPLETING THIS FORM
The EMBO Journal - Author Guidelines
EMBO Reports - Author Guidelines
Molecular Systems Biology - Author Guidelines
EMBO Molecular Medicine - Author Guidelines

Reporting Checklist for Life Science Articles (updated January 2022)
This checklist is adapted from Materials Design Analysis Reporting (MDAR) Checklist for Authors. MDAR establishes a minimum set of requirements in transparent reporting in the life sciences (see Statement of Task: 10.31222/osf.io/9sm4x). Please follow the journal's guidelines in preparing your manuscript.
Please note that a copy of this checklist will be published alongside your article.

Abridged guidelines for 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.
- ideally, figure panels should include only measurements that are directly comparable to each other and obtained with the same assay.
- plots 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. Any statistical test employed should be justified.
- Source Data should be included to report the data underlying figures according to the guidelines set out in the authorship 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 χ2 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.

Please complete ALL of the questions below.
Select "Not Applicable" only when the requested information is not relevant for your study.

Materials

Newly Created Materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
New materials and reagents need to be available; do any restrictions apply?	Yes	The information about constructing plasmid were provided in the section of 'CRISPR-Cas9 gene knockin and knockout' in Materials and Methods.

Antibodies	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
For **antibodies** provide the following information: - Commercial antibodies: RRID (if possible) or supplier name, catalogue number and or/clone number - Non-commercial: RRID or citation	Yes	All the information of antibodies we used in our study was provided in the section of 'Immunofluorescence staining in Materials and Methods'.

DNA and RNA sequences	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Short novel DNA or RNA including primers, probes: provide the sequences.	Yes	The primers we used mainly involved in qRT-PCR primers, we list all of primers in the supplementary table1.

Cell materials	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Cell lines: Provide species information, strain. Provide accession number in repository **OR** supplier name, catalog number, clone number, and/**OR** RRID.	Yes	The human ES(h9) cell line was supplied from WiCell.
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Yes	The information of primary culture for human derived urine cells was provided in the section of 'Generation of UiPSM' in the Materials and Methods.
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Not Applicable	All cell lines were confirmed to be mycoplasma free using the MycoAlertTM My coplasma Detection Kit (Lonza.LT07-318).

Experimental animals	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Laboratory animals or Model organisms: Provide species, strain, sex, age, genetic modification status. Provide accession number in repository **OR** supplier name, catalog number, clone number, **OR** RRID.	Yes	Relevant information was provided in the section of 'MITRG mice' in the Materials and Methods.
Animal observed in or captured from the field: Provide species, sex, and age where possible.	Yes	Relevant information were provided in the section of 'Teratoma Assay' in the Materials and Methods
Please detail **housing and husbandry conditions**.	Yes	The mice were uniformly raised by the Animal Center of GIBH, and were kept in the SPF laboratory.

Plants and microbes	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Plants: provide species and strain, ecotype and cultivar where relevant, unique accession number if available, and source (including location for collected wild specimens).	Not Applicable	
Microbes: provide species and strain, unique accession number if available, and source.	Not Applicable	

Human research participants	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If collected and within the bounds of privacy constraints report on age, sex and gender or ethnicity for all study participants.	Yes	Relevant information were provided on Figure.1H

Core facilities	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If your work benefited from core facilities, was their service mentioned in the acknowledgments section?	Yes	All the information of core facilities we benefited was displayed in the section of 'Acknowledgements' in our manuscript.

Design

Study protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
If study protocol has been **pre-registered, provide DOI in the manuscript**. For clinical trials, provide the trial registration number **OR** cite DOI.	Not Applicable	
Report the **clinical trial registration number** (at ClinicalTrials.gov or equivalent), where applicable.	Not Applicable	

Laboratory protocol	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Provide DOI OR other citation details if **external detailed step-by-step protocols** are available.	Yes	All the experiment protocols were described in detail in the section of 'Materials and Methods' in our manuscript.

Experimental study design and statistics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Include a statement about **sample size** estimate even if no statistical methods were used.	Yes	The sample size has been mentioned in 'Figure Legends.
Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. **randomization procedure**)? If yes, have they been described?	Yes	The major procedures have been described in the section of 'Oscillation assay and Comparison analysis of public scRNA-seq data' in Materials and Methods.
Include a statement about **blinding** even if no blinding was done.	Not Applicable	
Describe **inclusion/exclusion criteria** if samples or animals were excluded from the analysis. Were the criteria pre-established? If sample or data points were omitted from analysis, report if this was due to attrition or intentional exclusion and provide justification.	Not Applicable	
For every figure, are **statistical tests** justified as appropriate? Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it. Is there an estimate of variation within each group of data? Is the variance similar between the groups that are being statistically compared?	Yes	Relevant information about statistical tests has been provided in the section of 'Figures and corresponding Legends'.

Sample definition and in-laboratory replication	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
In the figure legends: state number of times the experiment was **replicated** in laboratory.	Yes	The number of replicate has been mentioned in 'Figure Legends'.
In the figure legends: define whether data describe **technical or biological replicates**.	Yes	Detail information was available in 'Figure Legends'.

Ethics

Ethics	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Studies involving **human participants**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval.	Yes	we applied Institutional Review Board Approval,number,GDL-IRB2020-004.Major ethical concerns including human derived urine cells from healthy volunteers.
Studies involving **human participants**: 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.	Yes	All volunteers provide urine samples with their informed consent. See the 'Ethics policy' section of Materials and Methods.
Studies involving **human participants:** For publication of **patient photos**, include a statement confirming that consent to publish was obtained.	Not Applicable	
Studies involving experimental **animals**: State details of **authority granting ethics approval** (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Yes	See the section of Teratoma Assay. The IACUC number, 2016008
Studies involving **specimen and field samples:** State if relevant **permits** obtained, provide details of authority approving study; if none were required, explain why.	Not Applicable	

Dual Use Research of Concern (DURC)	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Could your study fall under dual use research restrictions? Please check biosecurity documents and list of **select agents and toxins** (CDC): https://www.selectagents.gov/sat/list.htm	Not Applicable	
If you used a select agent, is the security level of the lab appropriate and reported in the manuscript?	Not Applicable	
If a study is subject to dual use research of concern regulations, is the name of the **authority granting approval and reference number** for the regulatory approval provided in the manuscript?	Not Applicable	

Reporting

The MDAR framework recommends adoption of discipline-specific guidelines, established and endorsed through community initiatives. Journals have their own policy about requiring specific guidelines and recommendations to complement MDAR.

Adherence to community standards	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
State if relevant guidelines or checklists (e.g., **ICMJE, MIBBI, ARRIVE, PRISMA**) have been followed or provided.	Not Applicable	
For **tumor marker prognostic studies,** we recommend that you follow the **REMARK** reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	Not Applicable	
For **phase II and III randomized controlled trials**, please refer to the **CONSORT** flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	Not Applicable	

Data Availability

Data availability	Information included in the manuscript?	In which section is the information available? (Reagents and Tools Table, Materials and Methods, Figures, Data Availability Section)
Have **primary datasets** been deposited according to the journal's guidelines (see 'Data Deposition' section) and the respective accession numbers provided in the Data Availability Section?	Yes	The primary datasets have been deposited in GEO platform. Relevant accession numbers were in the Data Availability Section.
Were **human clinical and genomic datasets** deposited in a public access-controlled repository in accordance to ethical obligations to the patients and to the applicable consent agreement?	Not Applicable	
Are **computational models** that are central and integral to a study available without restrictions in a machine-readable form? Were the relevant accession numbers or links provided?	Not Applicable	
If publicly available data were reused, provide the respective **data citations in the reference list**.	Yes	Data citations were in the reference list of manuscript.