THE EMBO JOURNAL

Single-cell Dissection of Cervical Cancer Reveals Key Subsets of the Tumor Immune Microenvironment

Guangxu Cao, Jiali Yue, Yetian Ruan, Ya Han, Yong Zhi, Jianqiao Lu, Min Liu, Xinxin Xu, Jin Wang, Quan Gu, Xuejun Wen, Jinli Gao, Jiuhong Kang, Qingfeng Zhang, Chenfei Wang, and Fang Li
DOI: 10.15252/embj.2022110757

Corresponding author(s): Fang Li (fang_li@tongji.edu.cn) , Chenfei Wang (08chenfeiwang@tongji.edu.cn)

Review Timeline:

Submission Date:	24th Jan 22
Editorial Decision:	16th Mar 22
Appeal Received:	17th Oct 22
Editorial Decision:	24th Jan 23
Editorial Decision - Follow up:	5th Feb 23
Revision Received:	5th Apr 23
Accepted:	19th May 23

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.)

Dear Dr Li,

Thank you again for your interest and sharing your manuscript (EMBOJ-2022-110757) for consideration by the EMBO Journal. Thank you also for your patience with our feedback which got delayed due to protracted referee input. We have sent your manuscript to three reviewers with expertise in tutor stream biology and cancer single cell analyses for evaluation and now received reports from all of them, which I copy below. I am afraid that in light of their comments we decided that we cannot offer publication in The EMBO Journal.

As you will see, the referees appreciate the potential interest of your resource analysis for the field. However, they also raise major concerns with the analysis that I am afraid preclude publication here. At the core of the matter, they state that the work remains too descriptive, and the claims on clinical relevance of identified subpopulations and their interactions not sufficiently supported by data. They are also critical towards the relatively small number of patients analysed. These shortcomings substantially decrease the experts' enthusiasm for the work.

We have discussed the referees' reports and cross-comments in the team. Given the negative opinions from good experts, and considering the journal's single round of major revisions, I am afraid we have concluded that we cannot offer to publish your study in The EMBO Journal.

Please note that we would per se be able to reassess a substantially reworked and amended version of the manuscript which would include functional validation of the findings presented along the lines of the referee comments but considering the major revision of the study apparently required, as well as the unclear outcome of such experiments, the result of such a re-evaluation would be entirely open at this point.

Thank you in any case for the opportunity to consider this manuscript. I regret we cannot be more positive on this occasion but hope nevertheless that you will the referees' comments useful.

Kind regards,

Daniel Klimmeck

Daniel Klimmeck, PhD
Senior Editor
The EMBO Journal

****

Referee #1:

In this manuscript, Cao et al. performed single cell RNA and TCR sequencing in six paired CC tumors and adjacent normal tissues to uncover the TME of CC. They investigated the heterogeneity of seven major cell lineages and the cellular characterization, interactions, and dynamic development of immune components. Notably, the authors unearthed tumor-specific GCB associated with TLS, and revealed the potential collaboration between T cells, myeloid cells and fibroblasts with B cells in induction and maintaining of TLS. While of potential interest, conclusions regarding the cell-cell interaction and functional role of several subsets are overstated. Several points should be addressed to strengthen the conclusions, as outlined below:

1. The study didn't provide functional information on the distinct subpopulations of each lineage. There was no information on whether one subset is indeed rate limiting for cancer progression or restraining. The analyses in the current study are comprehensive, but remain descriptive and conjectural with respect to functional contribution.
2. The authors mentioned in abstract that "Our study uncovers potential immune-resistance mechanisms in the TME". It was not

clearly stated in the results section. Though the results showed exhausted T cells and suppressive phenotype of NK cells were enriched in tumor sites, germinal center responses were also found in the tumor area which was associated with a better anti-tumor response. The authors should further discuss this issue.
3. For the sake of completeness, the manuscript would benefit from showing a comparison of scRNA-seq/immunofluorescence/immunohistochemistry data of tumour to normal tissue, especially immune cell components, and the spatial relationship and abundance of different subpopulations in supporting TLS formation.
4. Most of the supplementary figures were showed as Fig EV, while some were written as S1F in Line 119, S2E in Line 242. Beside, Fig EV2D in line 239 should be Fig EV2E, S2E should be Fig EV2F, and Fig EV2F in line251 should be Fig EV2G.

Referee #2:

In this manuscript, Cao, Yue, Ruan, et al. perform single cell RNA sequencing (scRNA-seq) paired with TCR sequencing of cervical cancer (CC) and matched adjacent normal samples. Pseudotime analysis revealed that exhausted T cells were likely established from clonal expansion of cytotoxic T cells and HPV-specific immunity is enriched in the tumor. Tumor-specific germinal center B cells associated with tertiary lymphoid structures were identified, and the presence of such cells was associated with improved outcomes and response to immunotherapy in other cancer types. The presence of antigen presenting CAFs was verified and their potential role in T cell infiltration and TLS formation was proposed. While scRNA-seq of CC has been reported previously (Gu et al. Front. Oncol 2021 and Li et al. Mol. Ther. Nucleic Acids 2021), this manuscript is the first to do in depth analysis of tumor microenvironment cell subsets and TCR repertoires. Generally, this reviewer notes that the data analyses were not clearly presented and as such appeared not to fully support the conclusions. For example, many of the settings used to define the cell clusters/subclusters were only superficially presented and none of the key subclusters was clearly shown in terms of their signature gene expression. Although each of the figures characterized one cell populations such as epithelial/cancer cells, T/NK cells, B cells, myeloid cells, and fibroblasts, the manuscript does not show functional validation for the proposed functions of these cell populations; albeit this may be outside the scope of the work.

Several points should be clarified, as outlined below.

1. In Figure EV1E, expression of cell type markers is presented; however, some cell types are missing (NK cells and DCs) and the markers distinguishing fibroblasts from perivascular cells (COL12A1 and MCAM) are not included. This figure should be revised to include expression of the missing markers. Gene signatures for the cancer cell population such as EP0-VIM (associated with better survival in TCGA dataset) and EP2-POSTN (associated with worse survival in TCGA dataset) should be shown. The violin plot, dot plot, and UMAP plot of these signature gene expression of these subclusters were not clearly presented. Given that such EP genes such as VIM and POSTN are highly expressed by other cell populations such as fibroblasts, the utilization of such gene clusters in TCGA bulk RNA-seq data can be problematic unless thoroughly explained.
2. Supplementary Figure 1C showed the UMAP plot of six samples (P1-P6). However, it seems that P6 sample contributed to a very high total cell numbers for all cell types, while the other samples seem to have much less cell numbers. Split UMAP view of this six samples (P1-P6) and cell number/composition table for six samples will clarify. If the other 5 samples have such low total cell yield, this entire study would be based on only one sample (P6). Please clarify.
3. The authors analyze abundance of cell subsets with respect to CTLs and state that B cell and PC subsets correlate with CTLs, but the correlations are relatively weak in some instances (i.e., R=0.199). As a result, the manuscript should be revised to make this point clearer and soften some of the conclusions derived from this correlation analysis.
4. Figure 2 showed the NK associated genes (FCGR3A and KLRC1) are associated with better survival in TCGA cohort. However, the expression of FCGR3A and KLRC1 are not strongly correlated with better survival in the total TCGA cervical cancer cohort using common cutoff threshold (median expression). This needs to be clarified/discussed.
5. The authors claimed that this study unearthed tumor-specific germinal center B cell (GCB) associated with tertiary lymphatic structures (TLS). High GCB proportion in CC patients is predictive for improved clinical outcomes and enhanced immunotherapy response, suggesting the pivotal role of B-cell mediated hormonal responses in anti-tumor immunity. However, the authors only provided one multiplex IHC image (Figure 4A) to support such interaction. Overall this claim is not robustly supported by the data presented here.
6. Can the author clarify why the fibroblast subclusters (such as F1_ACTA2) only express myofibroblast marker genes and lack COL1A1 and DCN (as shown in Figure 6C).
7. For the subclustering of epithelial/cancer cells, T/NK cells, B cells, myeloid cells, and fibroblasts, due to the lack of split UMAP plot of six samples (P1-P6) for these cell types, it is unknown whether the authors' subclustering strategy is consistently observed in these various samples (P1-P6).
8. Analysis of interactions between cell subsets revealed potential signaling between Tfh and B cells (Fig. 4B-C). The authors may want to validate by immunostaining that these cell types are within proximity to each other within the tumor microenvironment to signal to each other. A similar analysis could be performed for fibroblasts and lymphocytes to confirm their proximity to interact as proposed in Figure 6F.
9. The authors correlate the expression of factors expressed by CAFs (CCL4 and CH3L1) and overall survival in a TCGA cohort. The expression of these factors in other cell types in the tumor microenvironment should be included to demonstrate that CAFs are the major producers of CCL4 and CH3L1, otherwise their bulk expression does not entirely inform on whether CAFs play a role in tumor progression.

10. In some places in the figures, it is somewhat unclear whether tumor or tumor+normal samples are being analyzed, i.e., Fig. 2A, 2C, 2D, 2H-I. This should be clarified in the main text and/or figure legends.
11. The authors stated that this study uncovers potential immune-resistance mechanisms in the TME, but such mechanisms were only predicted based on scRNA-seq data and some staining results. . The immune-resistance related genes in immune cell clusters were not functionally tested. The authors may wish to revise the text for clarity.

Referee #3:

The manuscript by Cao, Yue et al. analyses in depth the multiple players of the tumour microenvironment in cervical cancer. The authors confirm several results, including known roles of T cell populations, myeloid cells and CAFs in cervical cancer, and delve more deeply into the less known role of B cells. A number of gene sets derived from the single-cell data are used for scoring ~350 cervical tumour samples from TCGA and investigating the prognostic value of the presence of multiple cell populations for patient survival, which is interesting. The data and analyses appear to be of good quality. The manuscript itself is has some weaknesses specified below, which at times make it hard to understand and assess the work. One obvious follow-up question of this work is how do the different prognostic measures combine and interact with one another? For individual tumours, do the different predictors of overall survival agree with one another (e.g. does a tumour with a high POSTN signature also have a high F6 CAF signature, what does it mean for survival if these metrics disagree)? As a suggestion for improving the clinical interest of the manuscript, could the authors use statistical modelling to put all of their predictors of survival into the same model?

Major concerns

Lack of details in the methods
The methods section is cursory and does not offer enough detail for the work to be assessed and reproduced. For example:
1) How is the significance of cell abundance changes determined? The methods section lumps together a bunch of tests in line 812 but does not say which method was used where. None of the methods listed in this section are appropriate for measuring statistically significant differences in proportional data.
2) The manuscript uses a large number of "signatures", but does not adequately explain what these are and how they were obtained. In the survival analyses, how were the gene sets defined for each specific sub-cluster? An explanation and a supplementary table listing the gene sets used for scoring should be provided.

Validation of the cellular proportions
Several of the figures in the paper refer to cellular proportions, however scRNA-seq protocols are not unbiased and will sample some cell-types more than others. Moreover, a change in proportions does not demonstrate cell infiltration (e.g. line 118).

Overstatement of some results
The authors find that a subset of malignant epithelial cells express POSTN, a protein associated with metastasis. Because of this known association of POSTN with metastasis and because of an association between the strength of the POSTN signature with survival data, the authors rightfully suggest that this cell population might be an invasive (line 153). Later on in the manuscript, this reasonable claim transforms into an assertion of fact (line 488) that "we identified the POSPN+ (sic) cells as an early invasive subtype". Another example is the section title "Tumor-specific iCAFs predict dismal outcomes in CC" (line 400). Of all the iCAFs, only one subset (F6) has an hazard ratio confidence interval that does not overlap 1 (no increased hazard). The F6 subset hazard ratio's lowest confidence boundary is rather close to 1, so making such a strong statement about dismal outcomes does not seem appropriate. Other instances of overstatement happen throughout the text (e.g. line 186).

Significance in Figure 4D
Is there a typo in the legend of this figure? If -log10(pvalue) = 0 then pvalue = 1, which is not significant.

Minor concerns

Contradictory information in the definition of the cell-types
The work described relies on the accurate identification of cell-types, however the information in the main text and in the methods sections as to which markers were used for cell-type assignment is confusing. For example:
1) In line 112 it is stated that the marker used for T-cells is CD8A, which would miss a large number of T-cell subsets, whereas in line 657 of the methods it is stated that CD3D is also used as a marker of T-cells.
2) For fibroblasts: the main text (line 114) says COL12A was used as a marker, whereas the methods (line 659) say it was COL1A1.
3) For the annotation of further cell-types the authors state in the methods that they were annotated by "the average expression of the respective gene sets", which does not mean anything if the gene sets are not given.
4) In line 425: if fibroblasts were defined through the expression of a collagen and F5_CCL4 does not express any collagen genes, then why was F5_CCL4 classified as a fibroblast cluster in the first place?
Given how crucial the definition of cell types is for this paper, and in general for the analysis of single-cell RNA-seq, it would be good to match the main text with the methods. Providing a comprehensive supplementary table listing the markers used in the

definition of cell-types would be very useful for other researchers.

Possible biased cell-type sampling
The cell proportions of 30% T/NK cells and < 10% epithelial cells in normal cervical samples appear biased. Given the morphology of this tissue I would expect to see a much higher proportion of epithelial cells represented. Was the tissue sufficiently digested? This in itself may not be an issue if the tumour samples are equally biased, which I assume they would be since the same protocol was used. In any case, if there is such a bias this should be acknowledged. In general, I would have liked to see the cell proportions claims (e.g. the reduction of stromal cells from 45% to 5% in SF1F) validated with an alternative assay.

Cryptic sentences and use of undefined acronyms
The manuscript should be carefully reviewed to remove cryptic sentences. E.g.:
- Line 786
- Line 93
- Line 749
- CC in the title
- TLS in line 322 (it's been defined in the abstract but that is not enough)
- GC in line 332
- OS in line 436

Suggestions for improvement

Joint modelling of the multiple predictors (see general comment at the beginning of this review).

The discussion is confusing, going back and forth between subjects, it would benefit from being cleaned up.

The manuscript would greatly benefit from being edited by a native speaker as many verbs are in the wrong tense, articles are missing and the use of singulars and plurals is often wrong.

Additional cross-comments referee #3:

I agree with both referee 1 (general comment and point 1) and referee 2 (point 11) that the functional conclusions in this manuscript are overstated. Nevertheless, the correlation between the presence of the immune populations and prognosis could have diagnostic value, even if the mechanism is unknown. Short of performing functional assays, I would advise the authors to tone down functional/mechanistic claims.

Point 2 from referee 1 gets at the same issue that I refer to when asking about how different predictors of overall survival agree with one another.

The number of patients is indeed small, any claims made solely from the scRNA-seq data need to take that into account. The issue is somewhat compensated by the addition of the TCGA data to bear on the prognostic claims. Referee 2 point 2 needs urgent clarification.

Additional cross-comments referee #1:

I agree with the other reviewers that some of the conclusions in this MS are overstated. My main concern is about the"tumor ecosystem" that the ligand-receptor pairs analysis illustrated cell-cell interaction would better be validated by immunostaining (Referee 1& Referee 2, point 5, 8) of clinical tumor samples to show the spatial relationship and abundance relevance of these cell types. Secondly, it's not clear how the immune cell composition or signature genes predicts the response to immunotherapy (Referee 1& Referee 3, general comment & Referee 2, point 11). Third, the issue raised by Referee 2 point 2 needs urgent clarification.

manuscript. For more details of this service, and to transfer your manuscript please click on Link Not Available. **

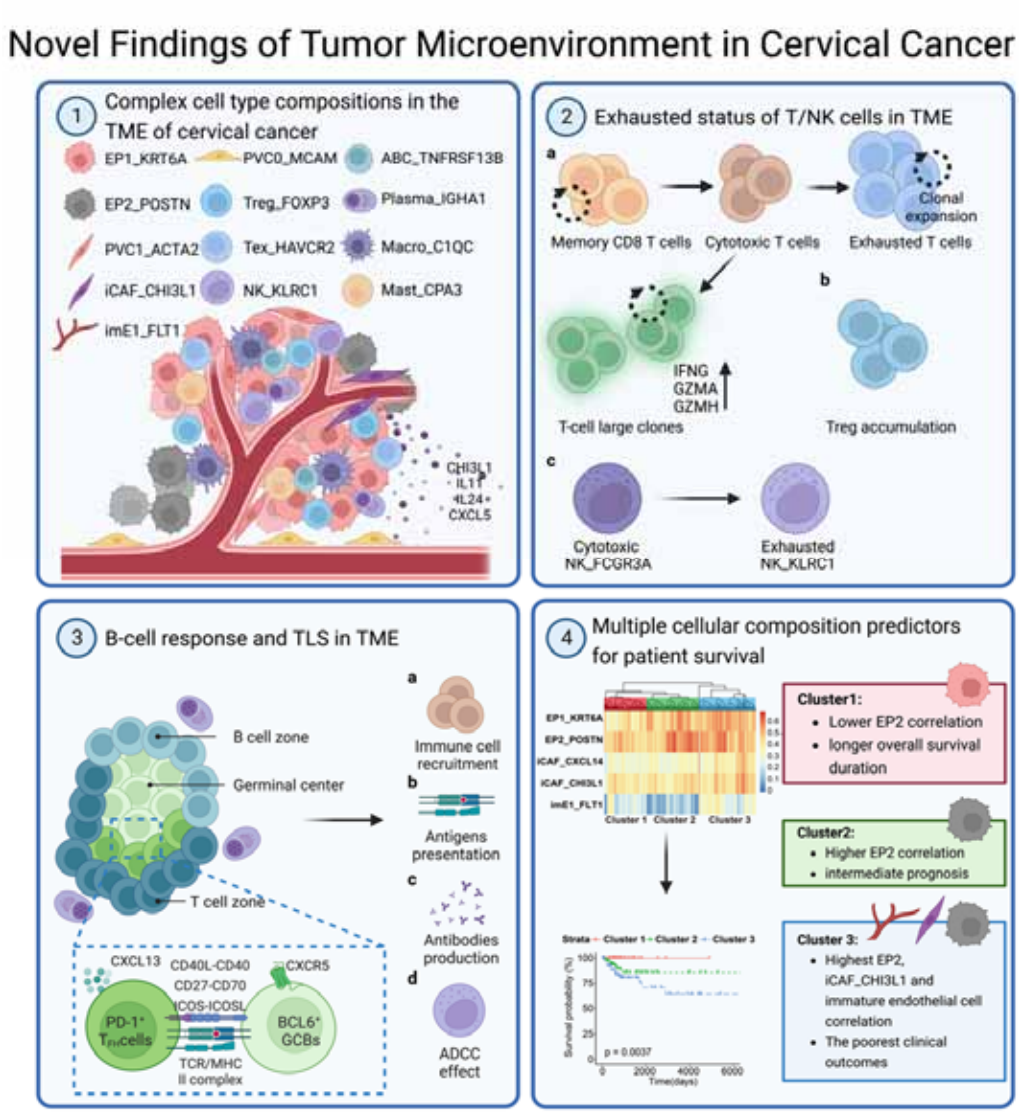
Please do not share this URL as it will give anyone who clicks it access to your account.

Dear Editor,

Thank you for the suggestions to our manuscript entitled "Single-cell sequencing links tumor ecosystem subsets to progression and anti-tumor immunity in CC" (EMBOJ-2022-110757). Our manuscript has been revised in accordance with the referees' comments and further analyses and experiments have been added. A revised manuscript has already been prepared for resubmission (EMBOJ-2022-112798).

We appreciate the time and efforts from you and the reviewers in improving our manuscript. We have performed additional experiments and computational analyses based on the suggestions of each reviewer and integrated the new findings into the revised manuscript. The major changes were as follows.

- We confirmed the infiltration of immune cells in the cervical cancer regions compared to peritumor normal tissues using multiplex immunohistochemistry (mIHC).
- We also validated the spatial proximal relationships between $CD4^+PD\text{-}1^+CXCL13^+$ T_{FH} cells and GCB in the TLS region, as well as the $CHI3L1^+$ iCAFs close to malignant cells via IHC staining. These analyses confirmed their potential interactions in the cervical cancer TME.
- We have visualized and documented the markers used for cell type annotation in the revised manuscript and Table EV8. Besides, we have also provided detailed gene signatures for each cell type subsets in Table EV9, which is used for survival and ICB response analyses.
- Using unsupervised hierarchical clustering, we developed a joint model for outcome prediction in TCGA cervical cancer samples and identified a cluster of cervical cancer patients that have high EP2_POSTN, iCAF_CHI3L1, and imE1_FLT1 signatures showed the worst prognosis.
- We have town down several over statements based on weak correlations or evidence, and clearly documented the methods for survival and ICB analyses as the reviewer suggested. We also polished the manuscript by a native speaker.
- We have summarized our novel findings on cervical cancer TME in the following figures.



As far as we know, our study is the first paper to comprehensively delineate the TME of cervical cancer from both the immune and stromal view. A recent scRNA-seq study has been published regarding cervical cancer (Chunbo Li, et al. Frontiers in Immunology, 2022). However, this work focuses specifically on the alterations of immune cell fractions during the progression of cervical cancer. Our results regarding T cells are consistent with this report, besides, we also revealed the lineage relationships of T and B cells in cervical cancer based on clonal analyses of TCR and BCR. A crucial role for GCB cells in the formation of TLS and anti-tumor immunity has also been highlighted in our research. Finally, our study also revealed the roles of cancer-associated fibroblasts (iCAFs) and endothelial cells in modulating tumor immunity and establishes a multiplex subset stratification model to group patients

into differential clusters with contrasting prognoses. In summary, our studies provided valuable resources for unraveling the TME of cervical cancer.

We hope that the new findings along with substantial revisions have made this work suitable for publication in the EMBO Journal. A point-to-point response to reviewer comments follows, with each reviewer's comments in black and our response in blue.

Referee #1

In this manuscript, Cao et al. performed single cell RNA and TCR sequencing in six paired CC tumors and adjacent normal tissues to uncover the TME of CC. They investigated the heterogeneity of seven major cell lineages and the cellular characterization, interactions, and dynamic development of immune components. Notably, the authors unearthed tumor-specific GCB associated with TLS, and revealed the potential collaboration between T cells, myeloid cells and fibroblasts with B cells in induction and maintaining of TLS. While of potential interest, conclusions regarding the cell-cell interaction and functional role of several subsets are overstated. Several points should be addressed to strengthen the conclusions, as outlined below:

1. The study didn't provide functional information on the distinct subpopulations of each lineage. There was no information on whether one subset is indeed rate limiting for cancer progression or restraining. The analyses in the current study are comprehensive, but remain descriptive and conjectural with respect to functional contribution.

We thank the reviewer for this helpful suggestion. In the previous manuscript, we characterized the potential functions of subpopulations from each lineage based on the transcriptomic profiles and GO functional analyses (Fig. EV1H for malignant cells, Fig. EV6A for fibroblasts, Fig. EV6E for endothelial cells). We have added the GO functional analyses for T, B, and myeloid subsets in the revised manuscript to better understand their potential functions in eliminating or protecting malignant cells (Revised Fig. EV2G for T-cells, Fig. 3C for B-cells and Fig. 5E for myeloid cells). Besides, we have also validated several findings using multiplex immunohistochemistry (mIHC). We first verified the alteration of immune cell

proportions in the tumor area of cervical cancer (CC) compared to the normal area. In addition, we confirmed the presence of $CD4^+CXCL13^+PD\text{-}1^+$ T_{FH} cells in the TLS of CC and $CHI3L1^+$ cancer-associated fibroblasts (CAFs) within tumor areas. We have described the experiments in detail in the following response as well as the revised manuscript.

2. The authors mentioned in abstract that "Our study uncovers potential immune-resistance mechanisms in the TME". It was not clearly stated in the results section. Though the results showed exhausted T cells and suppressive phenotype of NK cells were enriched in tumor sites, germinal center responses were also found in the tumor area which was associated with a better anti-tumor response. The authors should further discuss this issue.

Our study suggested that the dysfunction of several immune and stromal cell subtypes in the cervical cancer TME might prevent tumor cells from being eliminated and offers clues for potential immune-resistance mechanism. First, our data indicatedwo z that exhausted T cells (Tex_HAVCR2) and suppressive NK cells (NK_KLRC1) are more abundant in tumor tissues (Fig. 2B), suggesting the cytotoxic T lymphocytes (CTL) that infiltrated in the TME might be impaired by the suppressive environment. We further verified our findings in cervical cancer cohorts GSE63514 (n=128) and GSE9750 (n=57) (den Boon *et al*, 2015). Both exhausted T-cells and suppressive NK cells were consistently accumulated from normal and pre-cancerous (CIN2-3) to cancer samples (Fig. R1A-B). The trajectory and TCR analyses also revealed a transition from cytotoxic T cells to exhausted T cells (Fig. 2D-G). In the revised manuscript, we performed additional analyses and noticed that poliovirus receptor (PVR)-like protein signaling are the main co-inhibitory interactions between T/NK cells (CD96, and ITGIT) and malignant cells (PVR and NECTIN1). We also observed other interactions including LGALS9:HAVCR2, CD47:SIRPG, and PDCD1:CD274. All these interactions suggest the potential mechanism of T/NK cell exhaustion induced by malignant cells (Fig. R1C).

We further investigate the function of regulatory T cells (Tregs), which are consensus factors attenuating the cytotoxic function of CTLs in a variety of cancers (Togashi *et al*, 2019; Wherry & Kurachi, 2015), as well as other subsets promoting CTLs into a dysfunctional status. Tregs were identified by canonical markers IL2RA (CD25), and

FOXP3 (Fig. R1D). T-cell exhaustion may be attributed to Tregs, which are highly enriched in tumor samples (Fig. R1E) and express the T-cell suppressive genes CTLA4 and TIGIT (Fig. R1D). Furthermore, we quired which subpopulations within TME might induce T-cell dysfunction by TIDE (Jiang *et al*, 2018). The dysfunctional score calculated by TIDE indicates to which degree the subset can attenuate the effect of CTL on survival outcome. Interestingly, an immature endothelial subset imE1_FLT1 (Fig 6H) showed an elevated dysfunctional score, consistent with the known role of the aberrant tumor vasculature in promoting immune suppression (Huang *et al*, 2018). In addition, CAF and perivascular cells (PVC) subsets (PVC0_MACM, PVC1_ACTA2, eCAF_DCN, and iCAF_CHI3L1) also exhibit adverse correlations with CTL (Fig 6D). Overall, in this revision, we demonstrated that co-inhibitory interactions with cancer cells, the regulatory function of Tregs, T-cell dysfunction and exclusion induced by stromal subset may be involved in immune-resistance mechanisms of cervical cancer.

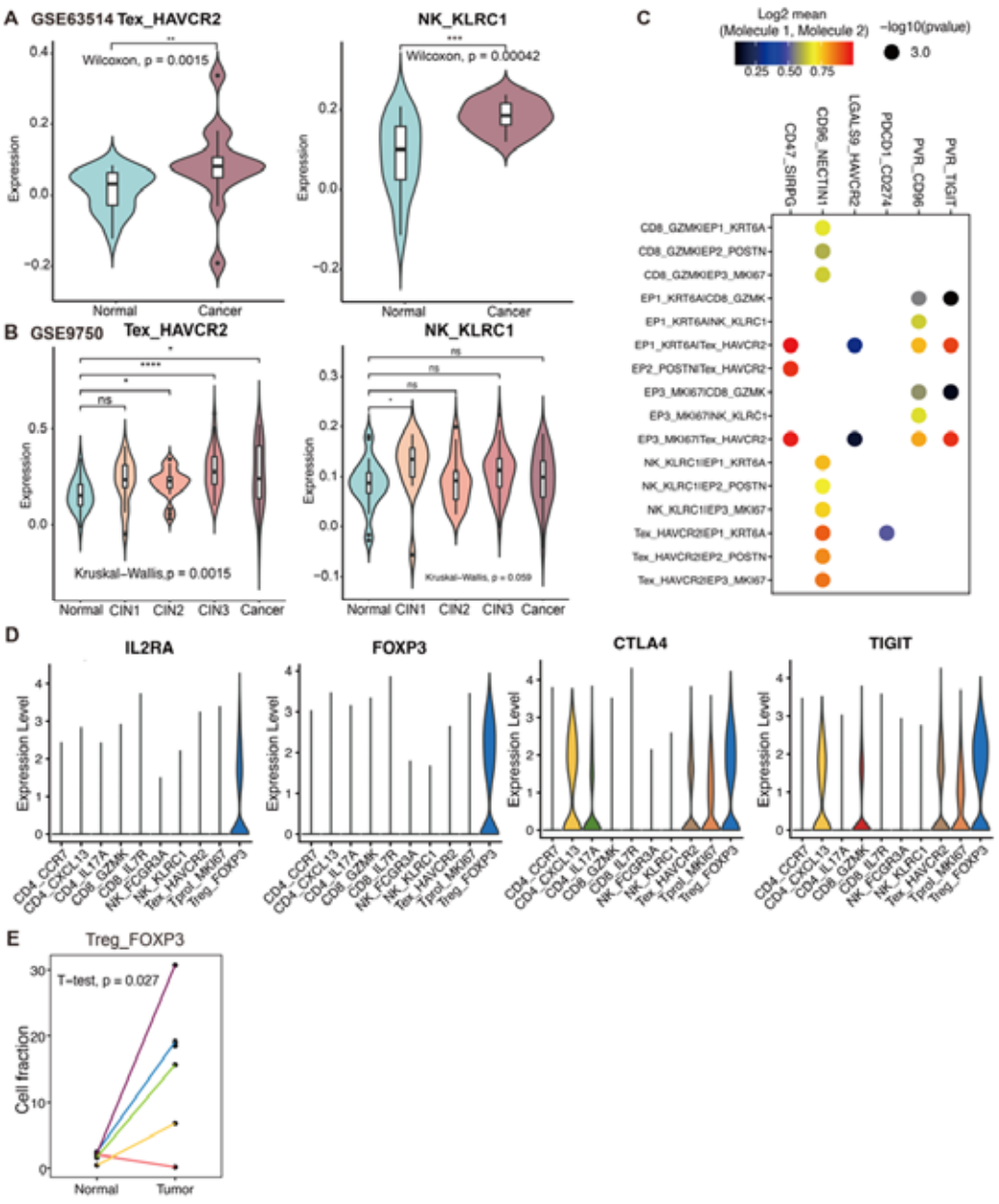


Figure R1 The potential immune-resistance mechanisms in the cervical cancer TME.

A-B Violin plot showing the expression of T and NK gene signatures in cervical cancer and normal tissues in GSE9750 (A) and GSE63514 (B).

C Bubble plots showing the interactions between epithelial cells and T/NK cell populations using CellPhoneDB.

D Violin plot showing the expression of Tregs related genes in T cells.

E Boxplots showing the cell-type fractions of T/NK cells for matched tumor and normal samples. Each line represents one patient. P values are from student's t-test and $P < 0.05$ was considered as statistical significance.

3. For the sake of completeness, the manuscript would benefit from showing a comparison of scRNA-seq/immunofluorescence/immunohistochemistry data of tumour to normal tissue, especially immune cell components, and the spatial relationship and abundance of different subpopulations in supporting TLS formation.

We detected the tumor-infiltrating immune markers (CD3: T-cells, CD20: B-cells, CD56: NK cells, and CD68: Macrophages) of representative tumor and adjacent normal tissue (patient 1) using mIHC (Fig. R2A). As expected, the cellular composition is distinct between the cervical tumor area and its peritumor stroma (Fig. R2B, revised Figure 1E). The number of stromal cells (35.01%) is roughly equivalent to the number of epithelial cells (43.76%) in the tumor area (Fig. R2C-D, revised Fig 1E), while stromal cells (77.01%) are dominated in normal tissues (Fig. R2E-F, revised Fig EV1F). Furthermore, $CD3^+$ T cells and $CD68^+$ macrophages show more infiltration and wider dispersion in the tumor area, consistent with the commonly T-cell inflamed TME of cervical cancer (O'Donnell *et al*, 2019) (Fig. R2C-D, revised Fig EV1F), whereas $CD56^+$ NK cells and $CD20^+$ B cells are seldomly observed in adjacent normal tissue (Fig. R2E-F, revised Fig EV1F). These results are in general consistent with the cell type compositions inferred from our scRNA-seq data (Fig. EV1E). We also investigated the presence of $CD4^+CXCL13^+PD\text{-}1^+$ Th cells, a T-cell subpopulation supporting TLS formation and maturation (details in Refree#2 Q9), as well as $VIM^+CHI3L1^+$ iCAFs (details in Refree#2 Q9). The representative images of CC and adjacent normal tissue have been updated in the revised manuscript.

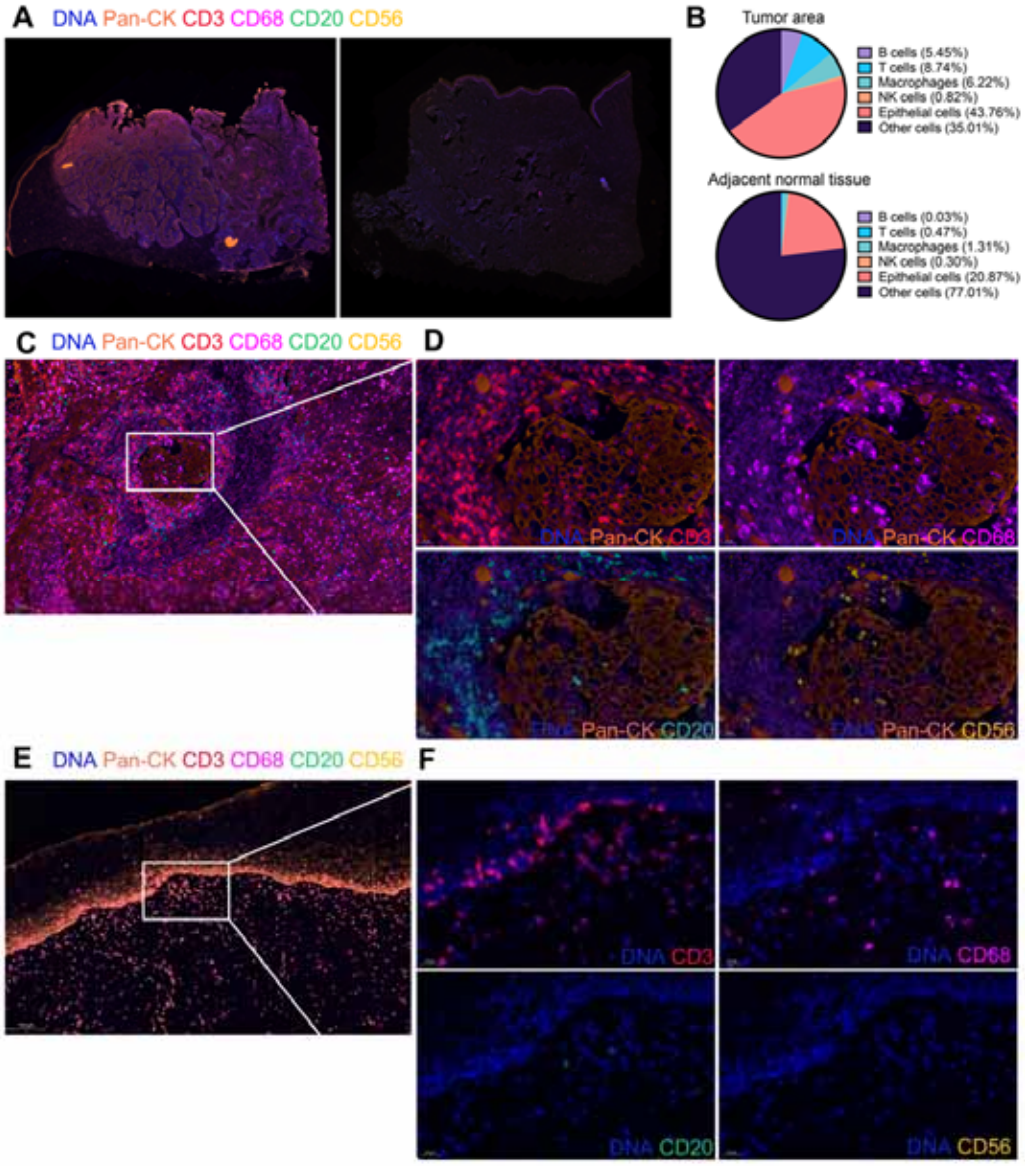


Figure R2 Tumor microenvironment of cervical cancer revealed by mIHC staining.

A MIHC stating showing the presence of CD3, CD20, CD68, CD56 and pan-CK. Magnification: 5x.

B Pie chart displaying the percentages of each kind of cell counted in (A) to describe the immune cell composition of the cervical cancer tissue and adjacent normal tissue.

C Merged image showing the colocalizations of DAPI, CD3, CD20, CD68, CD56 and pan-CK in cervical cancer tissue. Magnification: 100x.

D Cervical cancer tissue showing the location of each of CD3, CD20, CD68, CD56. Magnification: 400x.

E Merged image showing the colocalizations of DAPI, CD3, CD20, CD68, CD56 and pan-CK in adjacent normal tissue. Magnification: 100x.

F Adjacent normal tissue showing the location of each of CD3, CD20, CD68, CD56. Magnification: 400x.

4. Most of the supplementary figures were showed as Fig EV, while some were written as S1F in Line 119, S2E in Line 242. Besides, Fig EV2D in line 239 should be Fig EV2E, S2E should be Fig EV2F, and Fig EV2F in line251 should be Fig EV2G.

All supplementary figures are now referred to as "Fig. EV" in the revised manuscript.

Additional cross-comments referee #1:

I agree with the other reviewers that some of the conclusions in this MS are overstated. My main concern is about the "tumor ecosystem" that the ligand-receptor pairs analysis illustrated cell-cell interaction would better be validated by immunostaining (Referee 1& Referee 2, point 5, 8) of clinical tumor samples to show the spatial relationship and abundance relevance of these cell types. Secondly, it's not clear how the immune cell composition or signature genes predicts the response to immunotherapy (Referee 1& Referee 3, general comment & Referee 2, point 11). Third, the issue raised by Referee 2 point 2 needs urgent clarification.

We thank this and other reviewers' helpful suggestions to our study. According to the suggestions from referee #1 and referee #2, we confirmed the spatial relationship and abundance relevance of T_{FH} cells and CHI3L1^{+} CAFs via mIHC. Besides, these immunostaining slides corroborate the cell-type variation of tumor and normal samples. For prediction of response to immune checkpoint blockade (ICB) therapy, we compared the ssGSEA score of signatures from subsets of different lineages (Revised Table EV9) between responders and non-responders and calculated the statistical significance by two-sided Wilcoxon test. Additionally, we have confirmed that the cell numbers are comparable among patients, with a slightly higher number of cells in patient 6 (referee#2 point 2), the original bias is caused by the order of drawing points in the UMAP. Lastly, we have toned down the conclusions of potential immune-resistance mechanisms.

Referee #2:

In this manuscript, Cao, Yue, Ruan, et al. perform single cell RNA sequencing (scRNA-seq) paired with TCR sequencing of cervical cancer (CC) and matched adjacent normal samples. Pseudotime analysis revealed that exhausted T cells were likely established from clonal expansion of cytotoxic T cells and HPV-specific immunity is enriched in the tumor. Tumor-specific germinal center B cells associated with tertiary lymphoid structures were identified, and the presence of such cells was associated with improved outcomes and response to immunotherapy in other cancer types. The presence of antigen presenting CAFs was verified and their potential role in T cell infiltration and TLS formation was proposed. While scRNA-seq of CC has been reported previously (Gu et al. Front. Oncol 2021 and Li et al. Mol. Ther. Nucleic Acids 2021), this manuscript is the first to do in depth analysis of tumor microenvironment cell subsets and TCR repertoires. Generally, this reviewer notes that the data analyses were not clearly presented and as such appeared not to fully support the conclusions. For example, many of the settings used to define the cell clusters/subclusters were only superficially presented and none of the key subclusters was clearly shown in terms of their signature gene expression. Although each of the figures characterized one cell populations such as epithelial/cancer cells, T/NK cells, B cells, myeloid cells, and fibroblasts, the manuscript does not show functional validation for the proposed functions of these cell populations; albeit this may be outside the scope of the work.

1. In Figure EV1E, expression of cell type markers is presented; however, some cell types are missing (NK cells and DCs) and the markers distinguishing fibroblasts from perivascular cells (COL12A1 and MCAM) are not included. This figure should be revised to include expression of the missing markers. Gene signatures for the cancer cell population such as EP0-VIM (associated with better survival in TCGA dataset) and EP2-POSTN (associated with worse survival in TCGA dataset) should be shown. The violin plot, dot plot, and UMAP plot of these signature gene expression of these subclusters were not clearly presented. Given that such EP genes such as VIM and POSTN are highly expressed by other cell populations such as fibroblasts, the utilization of such gene clusters in TCGA bulk RNA-seq data can be problematic unless thoroughly explained.

We thank the reviewer for their helpful suggestions. We have added more markers in the original Figure EV1E to distinguish different cell types. We used EPCAM (epithelial), PECAM1 (endothelial), COL1A1 and COL12A1 (fibroblasts), ACTA2 and MCAM (perivascular cells), CD3D (T-cells), FCGR3A and KLRB1 (NK cells), CD79A (B-cells), CD38 (plasma), CD14 and FCN1 (mono/macrophages), HLA-DRA and FCGR3A (DCs), KIT (mast cells) to represent different cell type lineages, which showed a clear separation in our data (Fig. R3A, revised Figure 1C).

Epithelium are classified into malignant epithelium (EP1_KRT6A, EP2_POSTN, EP3_MKI67, a subset of EP0_MUC5B) and normal epithelium (EP4_EPCAM and majority of EP0_MUC5B). We renamed the previous EP0_VIM into EP0_MUC5B since MUC5B displayed more subtype-specific expression in this cluster. In the revised manuscript, besides the heatmap for visualizing subtype-specific markers (Fig. 1H), we also demonstrated the subtype-specific expression of 5 epithelium signatures using violin plots (Fig. R3B, revised Figure EV1G). E2_POSTN highly expressed EMT signatures which contribute to cancer progression by endowing cancer cells with the capacity to invade and metastasize (Lamouille *et al*, 2014) (Fig. 1H-I). In addition, these EMT signature genes did not show a comparable expression in fibroblast subsets (Fig. R3C). We performed survival analyses with EP2_POSTN signatures after eliminating the overlap genes with fibroblasts. The results showed that patients with the EP2_POSTN signature had poorer outcomes, indicating that the signature was specific to malignant cells (Fig. R3D). In summary, we have updated the example markers for separating the major cell type lineages, and demonstrated that the gene signatures for different epithelium subsets were highly specific to each subset and will not affluence the survival analysis.

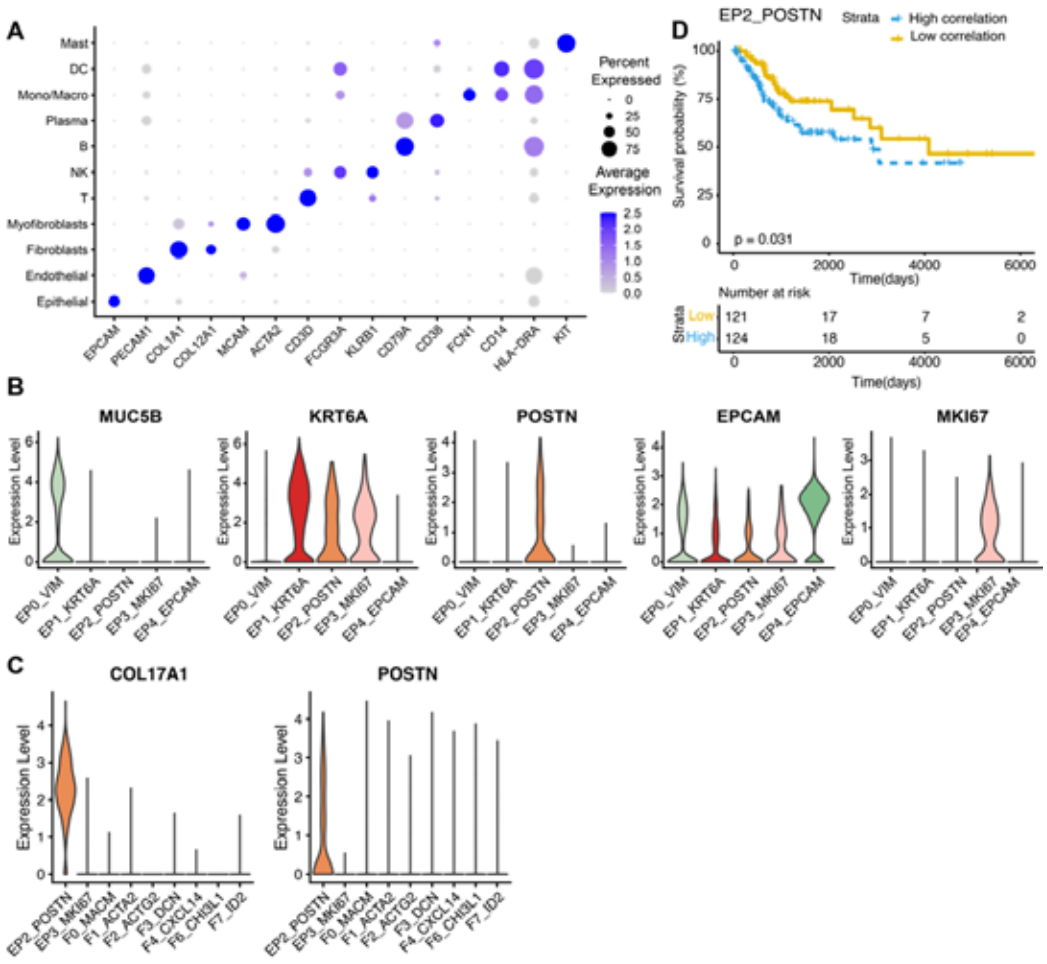


Figure R3 Marker genes distributions in major cell type lineage and epithelium subsets

A Dot plot showing the lineage-specific marker genes of T/NK, B, Plasma, Myeloid, Mast, Fibroblast, Endothelial, and Epithelium. The shade of color denotes the average gene expression level, the dot size denotes the percentage of gene expression in the corresponding lineage.

B Violin plot showing the epithelial marker genes in epithelial cell populations.

C Violin plot showing the expression of EP2 signature genes in EP2 and fibroblasts.

D Comparison of Overall survival (OS) rates for the high-correlation and low-correlation groups, stratified using the EP2 signatures with removing overlap genes of fibroblasts. P values are calculated using the log-rank test.

2. Supplementary Figure 1C showed the UMAP plot of six samples (P1-P6). However, it seems that P6 sample contributed to a very high total cell numbers for all

cell types, while the other samples seem to have much less cell numbers. Split UMAP view of these six samples (P1-P6) and cell number/composition table for six samples will clarify. If the other 5 samples have such low total cell yield, this entire study would be based on only one sample (P6). Please clarify.

The original UMAP in Fig. EV1C is caused by the drawing order of different patients, for which the last drawn P6 sample will cover the previous samples. We have generated the separated UMP for each patient, which showed that the cell number is comparable among patients, with a slightly higher number of cells in patient 6 (Fig R4). We have updated Fig. 1D and EV1C to avoid this bias and also attached the number of cells for each patient in the revised Table EV2.

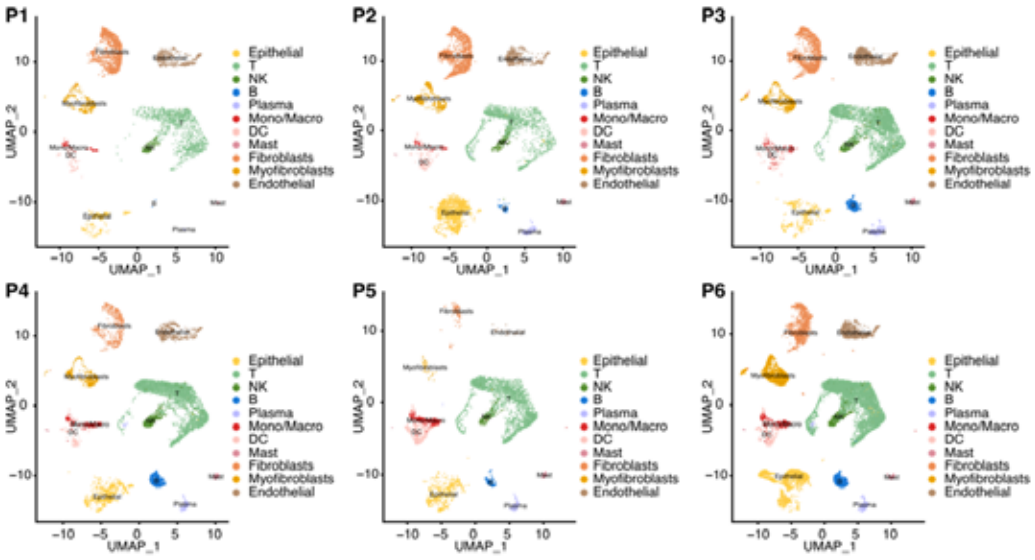


Figure R4 UMAP plot of 6 cervical cancer patients.

Cells were annotated based on known lineage-specific marker genes as T/NK, B, Plasma, Myeloid, Mast, Fibroblast, Endothelial, Epithelial cells (denoted by colors).

3. The authors analyze abundance of cell subsets with respect to CTLs and state that B cell and PC subsets correlate with CTLs, but the correlations are relatively weak in some instances (i.e., R=0.199). As a result, the manuscript should be revised to make this point clearer and soften some of the conclusions derived from this correlation analysis.

We agree with the reviewer that the conclusion based on the correlation analyses is overstated. We have modified the description as,

“Among the B-cell and PC subsets, GCB (B3_NEIL1) showed the highest positive correlation with CTL (R = 0.451), indicating a potential interaction between GCB and T-cells that could promote T-cell infiltration in the TME.” Other descriptions based on correlations have been revised to town down the statement.

4. Figure 2 showed the NK associated genes (FCGR3A and KLRC1) are associated with better survival in TCGA cohort. However, the expression of FCGR3A and KLRC1 are not strongly correlated with better survival in the total TCGA cervical cancer cohort using common cutoff threshold (median expression). This needs to be clarified/discussed.

The survival analyses for each cell type subsets were performed based on correlation with gene signature ratios (Fig. R5). Briefly, for each cell type subset such as NK_FCGR3A, we generated the averaged gene expression for all of the genes for that subset, then for each gene, we divided the gene expression by the corresponding averaged gene expression of all T and NK cells, which will generate a list of ratios for all genes. These ratios represent the specificity of each gene to the NK_FCGR3A cell subset. Then, for each patient in the TCGA cohorts, we calculated the Spearman correlation using the expression of all genes from that patient with ratios of all genes from NK_FCGR3A. Patients will be classified into higher correlation group and lower correlation group, for which higher correlation represents the patient might have higher infiltration of NK_FCGR3A, and lower correlation vice versa.

The calculation process using gene signature ratios has two advantages. First, using signatures rather than a single gene have better specificity, since many genes, such as KLRC1 also express in other T-cell subsets rather than NK. Second, since the gene expression level varied greatly, using the relative gene signature ratio will reduce the bias caused by highly expressed genes. We have clarified that the survival was performed using gene signature ratios in the revised result and method section.

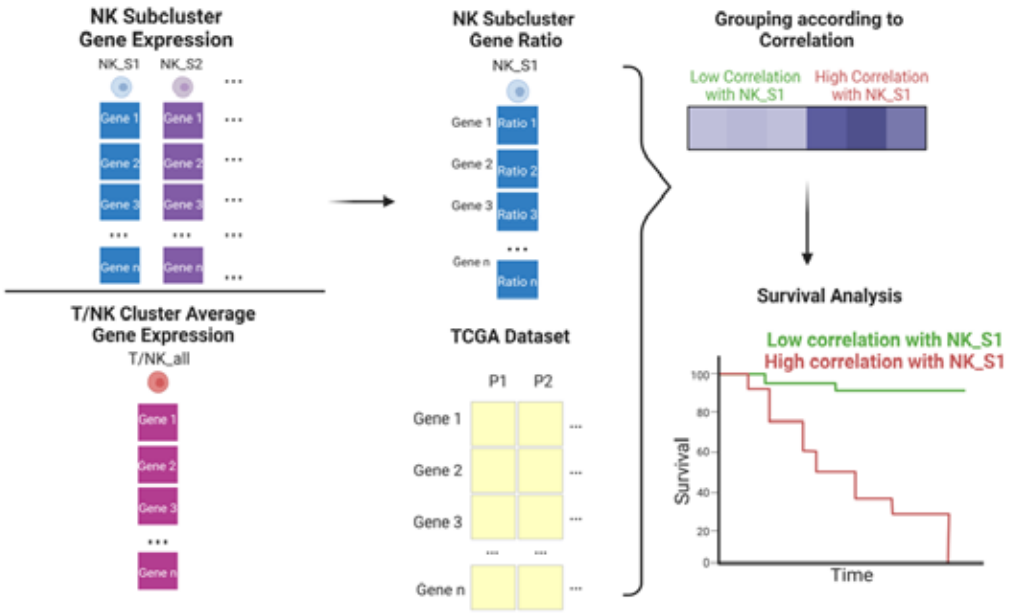


Figure R5 Schematic diagram of survival analysis

5. The authors claimed that this study unearthed tumor-specific germinal center B cell (GCB) associated with tertiary lymphatic structures (TLS). High GCB proportion in CC patients is predictive for improved clinical outcomes and enhanced immunotherapy response, suggesting the pivotal role of B-cell mediated hormonal responses in anti-tumor immunity. However, the authors only provided one multiplex IHC image (Figure 4A) to support such interaction. Overall this claim is not robustly supported by the data presented here.

We understand the concern of this reviewer. In the revised manuscript, we further confirmed the presence of GCB in TLSs using additional samples (Fig. R6A, revised Figure 4C, n=3). We quantified the cell numbers in the mIHC slides, and observed that both the B-cells, PCs, and GCBs showed exclusive high cellular density in the TLS area rather than the tumor or stromal area (Fig. R6A). Although B-cells and PCs are also enriched in the TLS, our analyses suggested the GCBs are the B-cell subsets that have the strongest interaction with $CXLC13^{+}T_{FH}$ and also have the most significant clinical outcome and ICB prediction power. A recent study by Meylan et al. confirmed that TLS could drive the *in-situ* maturation of B cells and anti-tumor antibody production and antibody-dependent cell-mediated cytotoxicity (ADCC) in renal cell cancer(Meylan *et al*, 2022). We further investigate whether CC patients

with higher GCB signatures have elevated ADCC levels (Fig. R6B, revised Fig 3J). As expected, patients with high CGB signatures showed higher expression of plasma marker (CD38), hallmarks of ADCC (FCGR1A and FCGR3A), and NK cells related-genes (KLRC1, PRF, and NKG7). Taken together, our analyses suggest GCBs have a pivotal role in predicting improved clinical outcomes of CC patients and ICB response, possibly due to their function in the TLSs through enhancing the ADCC effect. The anti-tumor functions of TLS in cervical cancer should be further investigated and we have added these analyses in the revised discussion section.

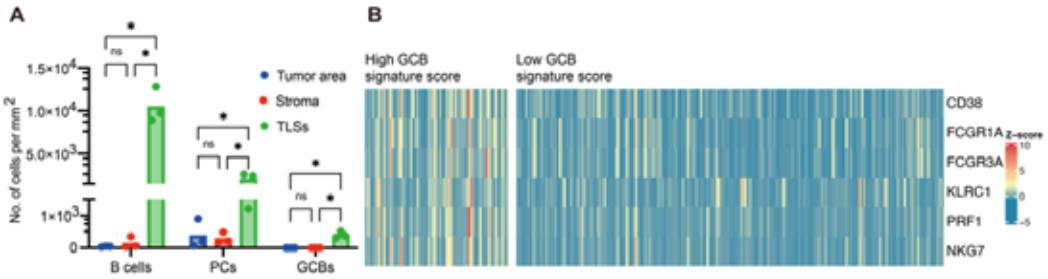


Figure R6 GCB subsets correlated with elevated ADCC effects.

A Quantification of B cells, PCs and GCBs in TLSs, stroma and pan-CK+ tumor area of three cervical cancer patients. *P < 0.05, two-way repeated measures analysis of variance (ANOVA) with Tukey's multiple comparison test.

B Expression of the genes that comprise the plasma cell and ADCC signatures ordered by GCB signature score.

6. Can the author clarify why the fibroblast subclusters (such as F1_ACTA2) only express myofibroblast marker genes and lack COL1A1 and DCN (as shown in Figure 6C).

We are sorry for the potential misleading in naming the fibroblast clusters. The original F0_MCAM and F1_ACTA2 are actually two pericyte clusters with high expression of perivascular cell markers MCAM and RGS5 (Fig. R7, PVC0_MACM, PVC1_ACTA2), but they still expressed low levels of COL1A1 but seldom express DCN. F2_ACTG2 is a myofibroblast cluster that also expresses low levels of COL1A1 and DCN, it has been renamed to mCAF_ACTG2. F3_DCN has been renamed as eCAF_DCN, as it only shows high ECM signature genes. iCAF_CXCL14 and iCAF_CHI3L1 are iCAFs that highly express immune-related genes, and F7_ID2

shows high expression of genes related to angiogenesis and is named angiCAF_ID2. All of the pericyte and CAF names have been updated throughout the manuscript.

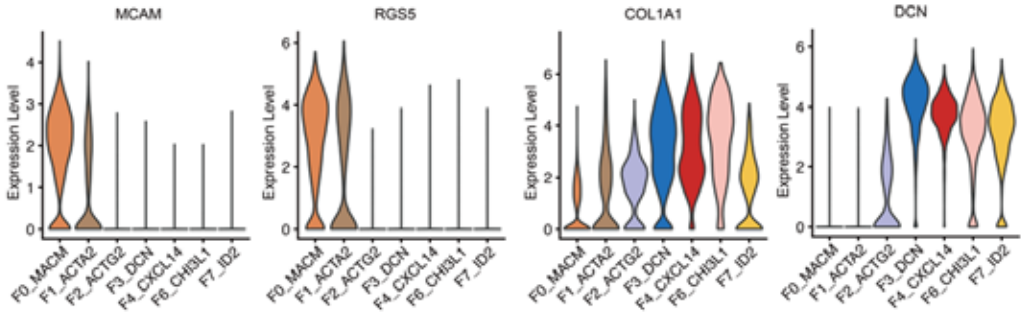


Figure R7 Violin plot showing the expression of pericyte and myofibroblast markers in fibroblast subsets.

7. For the subclustering of epithelial/cancer cells, T/NK cells, B cells, myeloid cells, and fibroblasts, due to the lack of split UMAP plot of six samples (P1-P6) for these cell types, it is unknown whether the authors' subclustering strategy is consistently observed in these various samples (P1-P6).

We have summarized the proportions of cell type subsets for each patient and also for both tumor and normal samples. The normal EP4_EPCAM cluster was highly enriched in P6, possibly due to the surgical variations of different patients, and does not affect our major conclusions. GCB (B3_NEIL1) subset was only observed in P4 and P6, suggesting a larger variation of GCBs and also potential related TLSs in different patients. Most other cell type subsets are in general observed in every patient with notable variations between tumor and normal samples (Fig. R8).

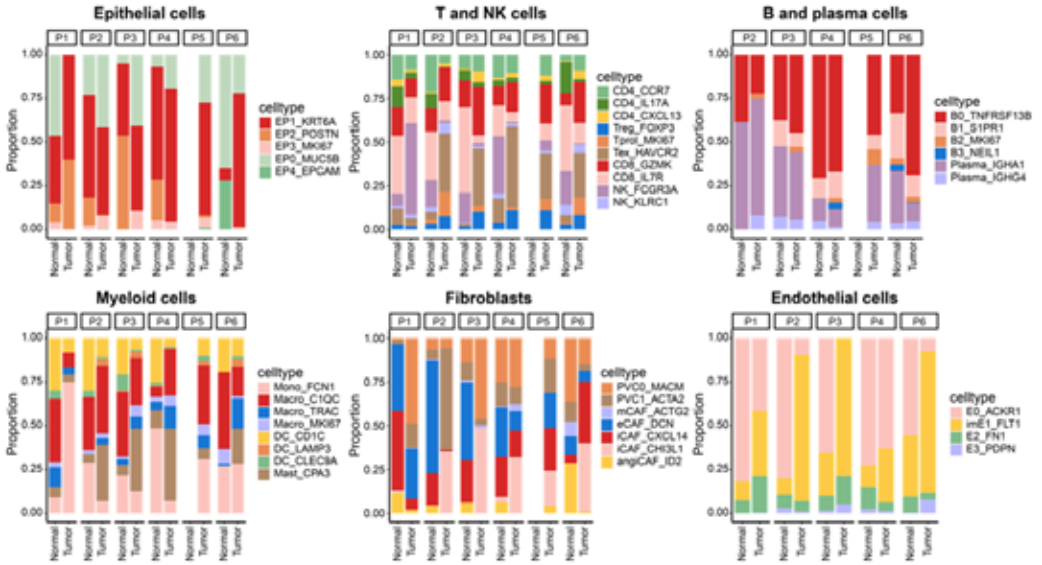


Figure R8 Stacked bar plot showing the subclusters of epithelial cells, T/NK cells, B cells, myeloid cells, fibroblasts and endothelial cells distribution across patients and categories of tissues.

8. Analysis of interactions between cell subsets revealed potential signaling between Tfh and B cells (Fig. 4B-C). The authors may want to validate by immunostaining that these cell types are within proximity to each other within the tumor microenvironment to signal to each other. A similar analysis could be performed for fibroblasts and lymphocytes to confirm their proximity to interact as proposed in Figure 6F.

To validate the potential interactions between T_{FH} and B cells, we performed mIHC of another panel (CD4, CD20, Ki67, PD-1, and CXCL13) and found that $CD4^{+}CXCL13^{+}PD\text{-}1^{+}$ T_{FH} cells are indeed close to B cells (Fig. R9A, revised Figure 4E), and the average distance is 29.81 um (Fig. R9C). Moreover, the $CXCL13^{+}T_{FH}$ cells colocalize with TLS, especially in immature TLS and germinal centers (Fig. R9B, revised Fig 4E), suggesting its potential role in TLS maturation and T-B interaction.

For the fibroblasts, we have double-checked the clustering result of fibroblasts, and confirmed that the F5_CCL4 fibroblast is a potential doublet-cluster that has high expression of T-cell-related genes but very low expression of collagen genes (see Fig.R12 in response to reviewer#3 point 5). All analyses that related to F5_CCL4 were removed in the revised manuscript. Instead, we analyzed the spatial location of F6_CHI3L1 (iCAF_CHI3L1 in the revised manuscript), a fibroblast subset that negatively correlated with the CTL infiltration (Fig 6D) and showed the worst prognosis in the TCGA CESC patients (Fig. 6I-J). Interestingly, the F6_CHI3L1 fibroblast seems to be closed around tumor areas, which indicates its potential immune exclusive functions by physically blocking the interaction of tumor and immune cells (Fig. R9D, revised Figure EV6C). We have updated these analyses in the revised manuscript.

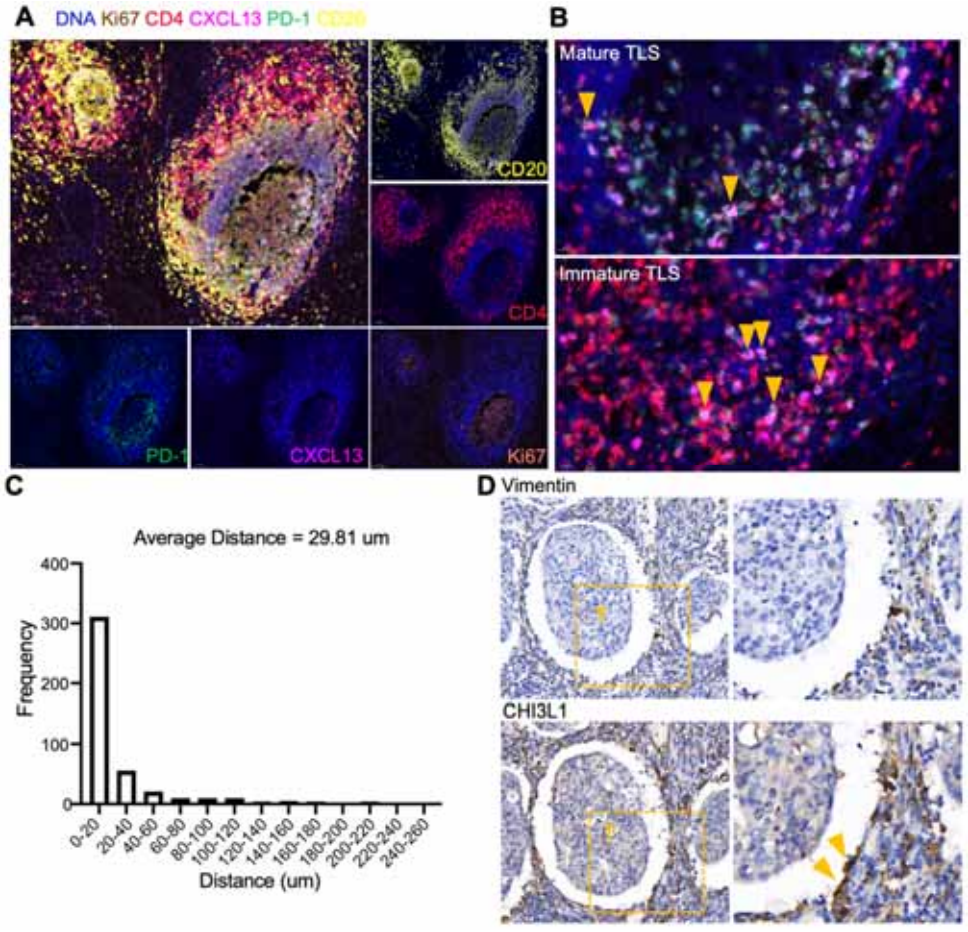


Figure R9 Immunostaining confirms the spatial relevance of specific cell types.

A-B MIHC staining of a representative tumor section showing the co-expression of CD20, CD4, Ki67, CXCL13 and PD-1 (A). The PD-1^{+}CXCL13^{+}CD4^{+} T_{FH} cells are abundant in germinal centers or immature TLSs (B).

C Histograms of distances between PD-1^{+}CXCL13^{+}CD4^{+} T_{FH} cells and CD20^{+} B cells.

D IHC staining of representative tumor section showing the expression of CHI3L1 in CAFs close to cancer cells.

9. The authors correlate the expression of factors expressed by CAFs (CCL4 and CH3L1) and overall survival in a TCGA cohort. The expression of these factors in other cell types in the tumor microenvironment should be included to demonstrate that CAFs are the major producers of CCL4 and CH3L1, otherwise their bulk expression does not entirely inform on whether CAFs play a role in tumor progression.

Thank you for the suggestion! As mentioned in point 8, we have removed the F5_CCL4 subcluster as it is a potential doublet cluster (see Fig.R12 in response to reviewer#3 minor point). For F6_CHI3L1, we confirmed that its maker genes (CHI3L1, IL11, and IL24) are mainly expressed by F6_CHI3L1 but not in other cell subsets (Fig. R10).

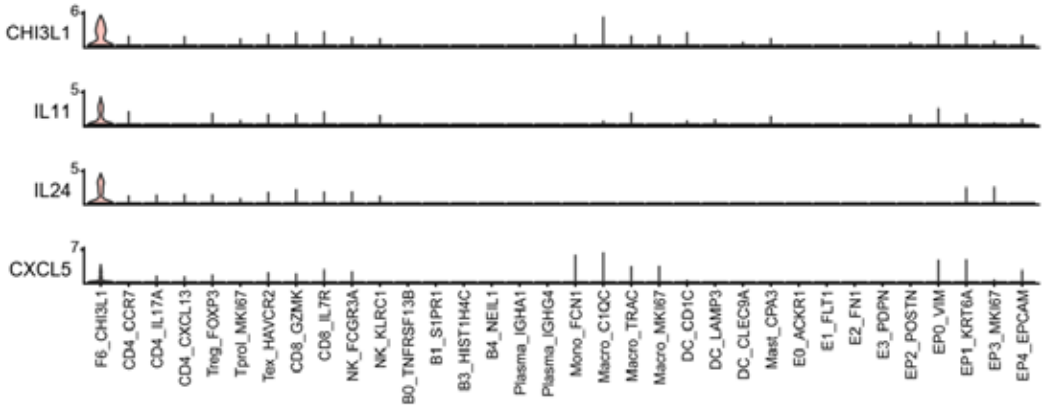


Figure R10 Violin plot showing the expression of F6_CHI3L1 signature genes in F6 and other cell lineages.

10. In some places in the figures, it is somewhat unclear whether tumor or tumor + normal samples are being analyzed, i.e., Fig. 2A, 2C, 2D, 2H-I. This should be clarified in the main text and/or figure legends.

All the analyses were performed based on integrated tumor and normal samples. We have modified the text as well as the corresponding figure legends in the revised manuscript.

11. The authors stated that this study uncovers potential immune-resistance mechanisms in the TME, but such mechanisms were only predicted based on scRNA-seq data and some staining results. The immune-resistance related genes in immune cell clusters were not functionally tested. The authors may wish to revise the text for clarity.

In the revised manuscript, we have toned down the statement of potential immune-resistance mechanisms. Instead, we have highlighted the complex and heterogenous immune and stromal landscapes reflected by our data and emphasized more on the combination of important cell subsets for predicting disease outcomes. Future functional experiments will further validate the immune-resistance mechanisms such as $CHI3L1^+$ fibroblast, $CPA3^+$ mast cells, and $FLT1^+$ endothelial cells.

Referee #3

The manuscript by Cao, Yue et al. analyses in depth the multiple players of the tumour microenvironment in cervical cancer. The authors confirm several results, including known roles of T cell populations, myeloid cells and CAFs in cervical cancer, and delve more deeply into the less known role of B cells. A number of gene sets derived from the single-cell data are used for scoring ~350 cervical tumour samples from TCGA and investigating the prognostic value of the presence of multiple cell populations for patient survival, which is interesting. The data and analyses appear to be of good quality. The manuscript itself is has some weaknesses specified below, which at times make it hard to understand and assess the work. One obvious follow-up question of this work is how do the different prognostic measures combine and interact with one another? For individual tumours, do the different predictors of overall survival agree with one another (e.g. does a tumour with a high POSTN signature also have a high F6 CAF signature, what does it mean for survival if these metrics disagree)? As a suggestion for improving the clinical interest of the manuscript, could the authors use statistical modelling to put all of their predictors of survival into the same model?

Major concerns

1. Lack of details in the methods

The methods section is cursory and does not offer enough detail for the work to be assessed and reproduced. For example:

How is the significance of cell abundance changes determined? The methods section lumps together a bunch of tests in line 812 but does not say which method was used where. None of the methods listed in this section are appropriate for measuring statistically significant differences in proportional data.

We thank the reviewer for their helpful suggestion. The cell type subset proportions were calculated for each normal and tumor sample of each patient, then the significance of the difference was evaluated using the two-sided student's t-test. * represents P-value < 0.05. Similar calculations were also performed for TCR clone size, TCR diversity, and clonality comparison between normal and tumor samples, CTL dysfunction scores between different cell subsets, as well as the cell subsets signatures for responders and non-responders in the ICB cohorts, the significance were evaluated using one-sided student's t-test. * Represents P-value < 0.05. The differentially expressed genes between different cell subsets were determined using Seurat with the non-parametric Wilcoxon rank sum test with a P-value threshold of 0.05 and an FDR threshold of 0.25. For the EMT score calculation, statistical significance was determined using a Kruskal–Wallis test with a P-value threshold of $2.2*10^{-16}$. For survival analyses and CTL dysfunction analyses, hazard ratios were calculated using the Cox regression, and p values were calculated using the log-rank test. All statistical analyses were performed in R (version 4.0.5). We have added the description of the statistical analyses in the revised method section.

The manuscript uses a large number of "signatures", but does not adequately explain what these are and how they were obtained. In the survival analyses, how were the gene sets defined for each specific sub-cluster? An explanation and a supplementary table listing the gene sets used for scoring should be provided.

The differentially expressed genes for different cell subsets were determined using Seurat with the non-parametric Wilcoxon rank sum test with a P-value threshold of 0.05 and an FDR threshold of 0.25. We further selected the top 100-200 genes as

the gene signature for each cell type subset depending on the total number of DE genes in each cell type subset. A table of gene signatures (Revised Table EV9) has been attached to the revised manuscript.

2. Validation of the cellular proportions

Several of the figures in the paper refer to cellular proportions, however scRNA-seq protocols are not unbiased and will sample some cell-types more than others. Moreover, a change in proportions does not demonstrate cell infiltration (e.g. line 118).

We agree with the reviewer that the cellular proportions from the scRNA-seq data might be biased by different scRNA-seq protocols, for example, the 10X Genomics protocol is based on immune cells. However, all the samples in our study were generated using the same 10X Genomics platform, and the bias should be similar for both the tumor and normal samples and thus are comparable. To further demonstrate the infiltration of immune cells in the CC tumor region, we also performed mIHC analyses of immune cells and stromal cells in both tumor and normal regions (see Fig. R2, point 3 to reviewer#1), and confirmed the infiltration of T, B, and macrophages in the tumor region. We have tone-downed the statement of other descriptions related to cell infiltration without functional or mIHC validations.

3. Overstatement of some results

The authors find that a subset of malignant epithelial cells express POSTN, a protein associated with metastasis. Because of this known association of POSTN with metastasis and because of an association between the strength of the POSTN signature with survival data, the authors rightfully suggest that this cell population might be an invasive (line 153). Later on in the manuscript, this reasonable claim transforms into an assertion of fact (line 488) that "we identified the POSPN+ (sic) cells as an early invasive subtype". Another example is the section title "Tumor-specific iCAFs predict dismal outcomes in CC" (line 400). Of all the iCAFs, only one subset (F6) has an hazard ratio confidence interval that does not overlap 1 (no increased hazard). The F6 subset hazard ratio's lowest confidence boundary is rather close to 1, so making such a strong statement about dismal outcomes does

not seem appropriate. Other instances of overstatement happen throughout the text (e.g. line 186).

We have revised the overstatement of several conclusions in the manuscript as follows.

Line 488: “We identified the POSPN+ (sic) cells as an early invasive subtype.” To “We revealed the intra- and inter-tumoral heterogeneity of cancer cells and identified POSPN^{+} cells as a potential early invasive subtypes.”

line 400: “Tumor-specific iCAFs predict dismal outcomes in CC” To “Tumor-specific iCAF and endothelial subset are potentially associated with tumor progression and immunosuppression”

Line 186: “Collectively, our data revealed the exhausted status of cytotoxic T cells and NK cells and an imbalanced CD4 response in the CC tumor area.” To “In summary, our findings suggest that CD8^{+} T cells exhibit a cytotoxic-exhausted phenotype and a clear local expansion trajectory, the infiltrated Tregs and PVR-like protein signaling may contribute to the immune-resistance microenvironment of CC.”

4. Significance in Figure 4D

Is there a typo in the legend of this figure? If -log10(pvalue) = 0 then pvalue = 1, which is not significant.

Yes, there is a typo in the legend and we have replaced Figure 4D with Fig. R11 in this revision.

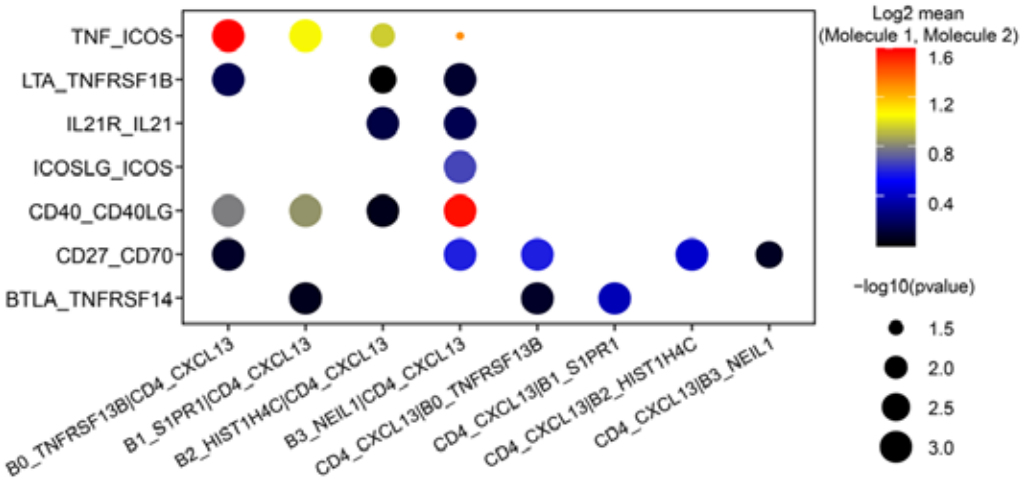


Figure R11 Bubble plots showing the interactions between TFH (CD4_CXCL13) and B cell populations using CellPhoneDB.

Minor concerns

5. Contradictory information in the definition of the cell-types

The work described relies on the accurate identification of cell-types, however the information in the main text and in the methods sections as to which markers were used for cell-type assignment is confusing. For example:

In line 112 it is stated that the marker used for T-cells is CD8A, which would miss a large number of T-cell subsets, whereas in line 657 of the methods it is stated that CD3D is also used as a marker of T-cells.

We thank the reviewer for pointing this out. In the revised manuscript, we have clearly stated that the CD3D was used to determine T-cells for the major lineage UMAP (see Fig. R3 in point 1 to reviewer#2). Besides, we have added an expanded view table that includes the cell type marker genes for each cell type subset (Revised Table EV8).

Exchange cell markers

For fibroblasts: the main text (line 114) says COL12A was used as a marker, whereas the methods (line 659) say it was COL1A1.

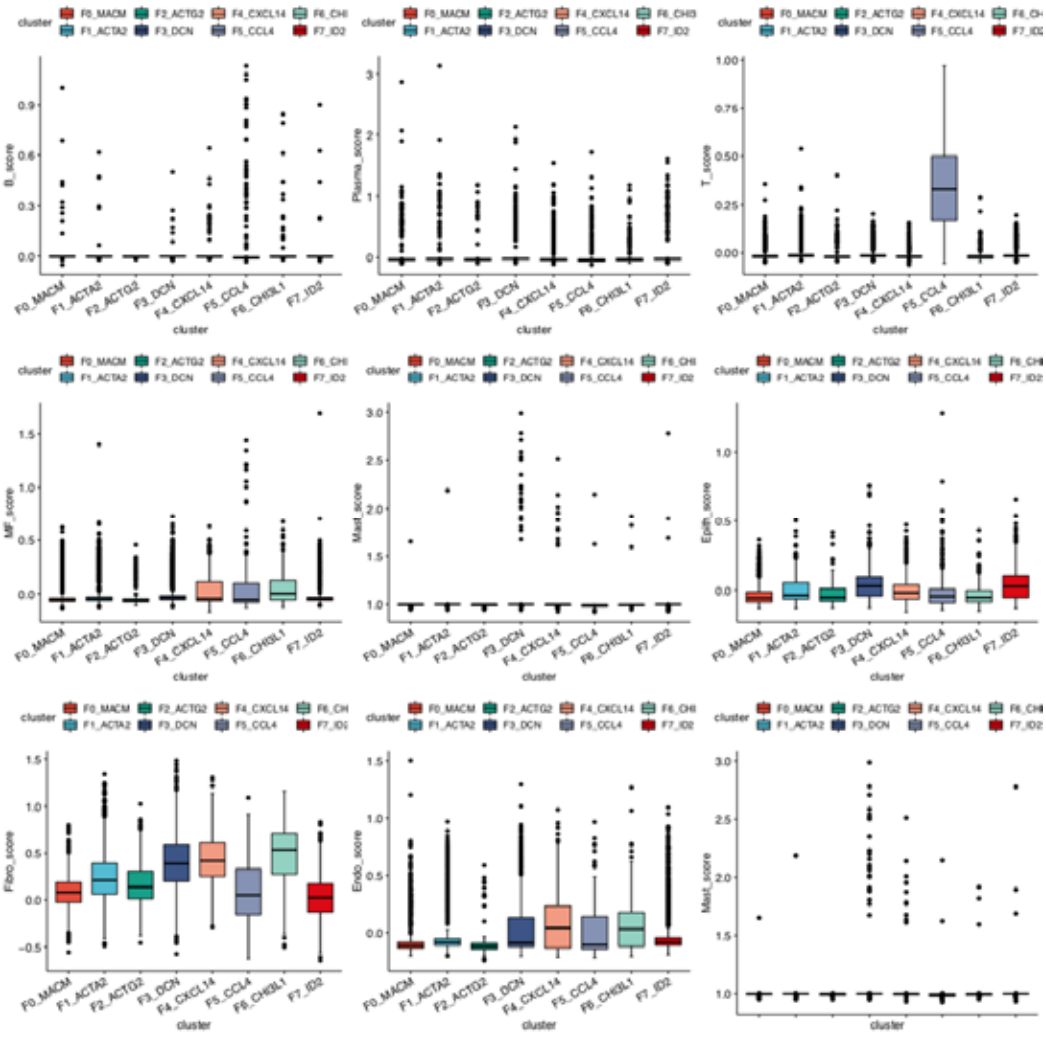
Both of COL12A and COL1A1 were used as markers for fibroblast in the revised manuscript (see Fig. R3 in point 1 to reviewer#2).

For the annotation of further cell-types the authors state in the methods that they were annotated by "the average expression of the respective gene sets", which does not mean anything if the gene sets are not given.

We have added an expanded view table that includes the cell type marker genes for each cell type subset (Revised Table EV8).

In line 425: if fibroblasts were defined through the expression of a collagen and F5_CCL4 does not express any collagen genes, then why was F5_CCL4 classified as a fibroblast cluster in the first place?

F5_CCL4 expresses genes such as VIM, PDGFRA, and PDGERB, and in the original manuscript, we defined it as a subset of fibroblast. However, we further



checked the expression of cell type subset signatures and found that F5_CCL5 express both T-cell and fibroblast genes (Fig. R12). Besides, as described by the reviewer, the F5_CCL4 subset has a very low-level expression of collagen genes. These evidences collectively suggest that the F5_CCL4 subset might be potential doublets of T and Fibroblast cells, we thus removed this cluster in the revised manuscript and clearly state the reasons in the revised method section.

Figure R12 Box plots showing the signature scores of fibroblast cell subtypes.

Given how crucial the definition of cell types is for this paper, and in general for the analysis of single-cell RNA-seq, it would be good to match the main text with the

methods. Providing a comprehensive supplementary table listing the markers used in the definition of cell-types would be very useful for other researchers.

We thank the reviewer for this helpful suggestion, the cell type markers are listed in Table EV8 of the revised manuscript.

6. Possible biased cell-type sampling

The cell proportions of 30% T/NK cells and < 10% epithelial cells in normal cervical samples appear biased. Given the morphology of this tissue I would expect to see a much higher proportion of epithelial cells represented. Was the tissue sufficiently digested? This in itself may not be an issue if the tumour samples are equally biased, which I assume they would be since the same protocol was used. In any case, if there is such a bias this should be acknowledged. In general, I would have liked to see the cell proportions claims (e.g. the reduction of stromal cells from 45% to 5% in SF1F) validated with an alternative assay.

We agree with the reviewer that there is bias in cell type sampling for the 10X Genomics protocol. The 10X Genomics scRNA-seq protocol is based on microfluidic droplets, and has a bias toward immune cells in cellular estimation as they are relatively smaller, in a free state thus are easier to capture than epithelial cells, we have clearly discussed this in the revised method section.

However, the disparity in cell-type ratios between tumors and normal tissues is mainly due to the original variation in cell composition. Alzamil, Lama et al provided a schematic(Alzamil *et al*, 2021), which showed the morphology and the approximate proportions of the human cervix (Fig. R13A). We also referred to a typical HE-staining slide (case ID: TCGA-C5-A1MP) from the TCGA database (Fig. R13B), and this slide shows the distinct cellular composition of the tumor and adjacent normal tissue. In normal cervix tissues, a thin layer of epithelium covers the surface of the cervix. The cervix epithelium is supported by an abundant cervical stroma containing fibrous tissues and blood vessels (Fig. R13C). However, in the tumor area, irregular infiltrated cancer nests are the majority of cellular composition, and immune cells and stromal cells can only be observed in the area outside the cancer nests (Fig. R13D). Furthermore, the mIHC staining of adjacent normal cervical tissues also demonstrated a 40% proportion of stromal cells (Fig. R2B). Additionally, a previous

meta-analysis showed that the number of infiltrating T-cells in normal cervical tissue was similar to the number in cancerous tissue(Litwin *et al*, 2021). In conclusion, our cell-type sampling is related to the differences in cell composition between tumor and normal cervical tissues, and is generally in agreement with previous reports.

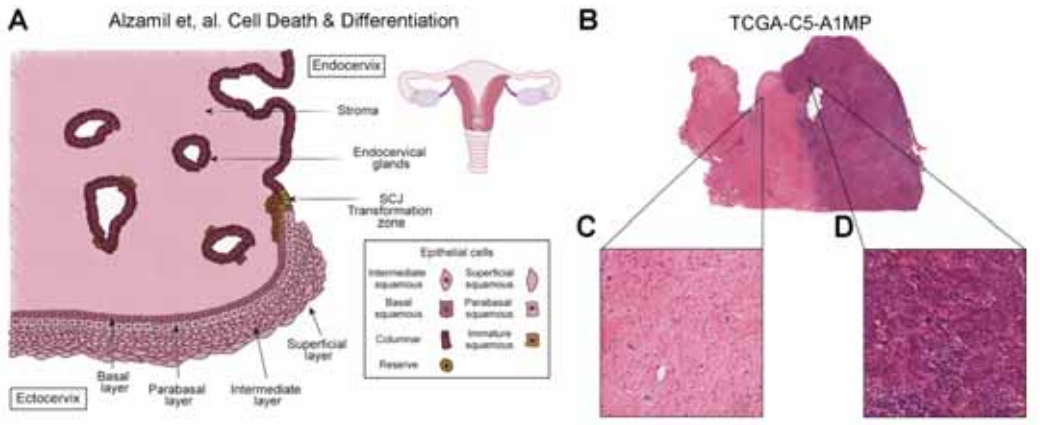


Figure R13 Cellular composition of normal cervix and cervical cancer.

A Schematic diagram of the anatomy of the human cervix. Figure is from Alzamil et, al. Cell Death & Differentiation. 2021 Fig. 4 (Alzamil *et al.*, 2021).

B-D Representative HE section contains adjacent normal cervical tissue (**C**) and tumor area (**D**).

7. Cryptic sentences and use of undefined acronyms

The manuscript should be carefully reviewed to remove cryptic sentences. E.g.:

- Line 786
- Line 93
- Line 749
- CC in the title
- TLS in line 322 (it's been defined in the abstract but that is not enough)
- GC in line 332
- OS in line 436

We have modified the original descriptions as follows.

Line 93: “Our data provide a transcriptomic profiling basis of the CC microenvironment and broadens the understanding of CC and other HPV-related cancers.” **To** “In this study, a single-cell transcriptomic profile of the CC

microenvironment was provided, contributing to a deeper understanding of the disease."

Line 786: "Tide set pivot genes as CD8A, CD8B, GZMA, GZMB, and PRF1 and calculate the average expression level of these genes to estimate the cytotoxic T-lymphocyte (CTL) level in bulk RNA-seq data." **To** "To estimate CTL levels, TIDE uses pivot genes such as CD8A, CD8B, GZMA, GZMB, and PRF1 and calculates the average expression of these genes; CESC patients from the TCGA cohort were grouped by CTL level, with the mean value as a cutoff point."

Line 749: "Clinical and survival analyses were conducted across one patient cohort receiving bulk RNA-sequencing with various treatments and endpoints-the TCGA CESC cohort (both cancer-specific survival and recurrence)." **To** "In the TCGA CESC cohort, patients receiving bulk RNA sequences with various treatments and endpoints were analyzed for clinical and survival outcomes (cancer-specific survival and recurrence)."

In addition, Cryptic abbreviations have been removed in the title, line 322, 332, and 436.

8. Suggestions for improvement

Joint modelling of the multiple predictors (see general comment at the beginning of this review).

We thank the referee's comments. Our previous study examined the prognostic value of gene signatures derived from a single subset (described in referee#2 Q4), and in this revision, we used unsupervised hierarchical clustering to learn if multiplex cell subsets could improve outcome prediction efficiency. As shown in Fig. R14A (Revised Figure 6I), a joint model incorporating classic and invasive tumor subsets, iCAFs, and immature endothelial subsets provided a promising prognosis for CC patients. The patients in cluster 3 showed high a correlation with EP2_POSTN, iCAF_CHI3L1, and imE1_FLT1, suggesting that they harbored invasion, immune suppression, and angiogenesis characteristics. In cluster 2, only EP2_POSTN

showed a significant correlation, whereas patients in cluster 1 had no significant correlation with these subsets. Based on this, patients in cluster 1 have a better overall survival rate than patients in clusters 2 and 3. There was an intermediate risk of death in cluster 2, while the poorest clinical outcomes were found in cluster 3 (Fig. R14B, revised Figure 6J). In contrast, when immune components were used as predictors, similar approaches failed to demonstrate significant results. Partially, this is due to the complicated status of immune cells in TME that could not be categorized into specific patterns. These results collectively suggest gene signatures derived from multiplex subsets of scRNA-seq can be used as outcome predictors in cervical cancer patients, and the stromal and malignant cell heterogeneities are highly associated with patients' survival.

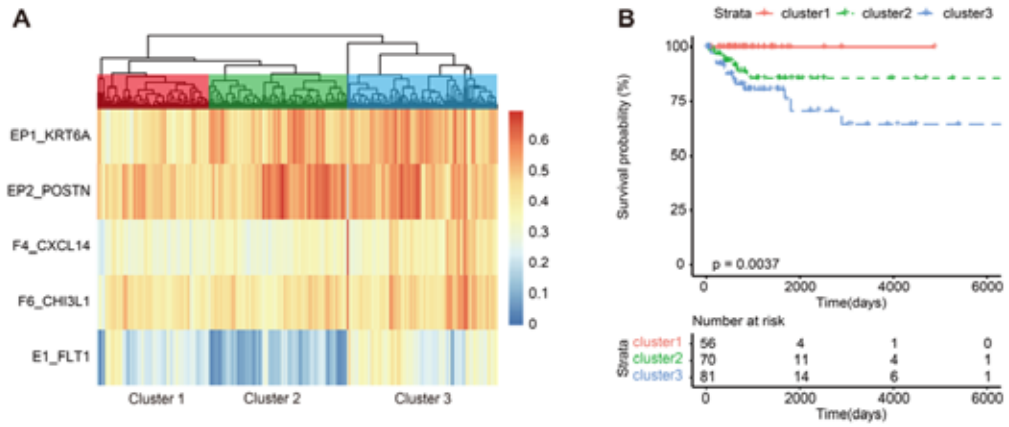


Figure R14 An joint model for prognostic prediction established by hierarchical clustering.

A Unsupervised hierarchical clustering for 255 patients from TCGA dataset based on the correlation to EP1, EP2, F4, F6 and E1. Three clusters were identified.

B Kaplan–Meier curves of overall survival for three clusters indicating the prognostic value of the joint model. P value was tested by the log-rank test.

The discussion is confusing, going back and forth between subjects, it would benefit from being cleaned up.

We have revised the discussion into the follower structures as follows.

We first summarized the findings of our study with a focus on T and B cells. Then, we expanded the analyses and discussed the dynamic status of T cells, their transition in cervical cancer, and the potential role of HPV-specific T cells in HPV-related

cancers. After that, we explored the anti-tumor mechanism of B-cells and TLS functions in our data as well as other cancer types, we also discussed the potential biomarker function of TLS in immunotherapy. Finally, we discussed using stromal cells as a predictor of CC patients' survival and the limitation of our study.

The manuscript would greatly benefit from being edited by a native speaker as many verbs are in the wrong tense, articles are missing and the use of singulars and plurals is often wrong.

We have performed language editing by a native speaker for the revised manuscript.

Additional cross-comments referee #3:

I agree with both referee 1 (general comment and point 1) and referee 2 (point 11) that the functional conclusions in this manuscript are overstated. Nevertheless, the correlation between the presence of the immune populations and prognosis could have diagnostic value, even if the mechanism is unknown. Short of performing functional assays, I would advise the authors to tone down functional/mechanistic claims.

Point 2 from referee 1 gets at the same issue that I refer to when asking about how different predictors of overall survival agree with one another.

The number of patients is indeed small, any claims made solely from the scRNA-seq data need to take that into account. The issue is somewhat compensated by the addition of the TCGA data to bear on the prognostic claims. Referee 2 point 2 needs urgent clarification.

We thank this and other reviewers' helpful suggestions to our study. The statement on functional conclusions has been toned down. Further, we evaluated whether a multi-cell population model could be developed to predict outcomes for cervical cancer patients. The model included five subsets of epithelial and stromal cells. Patients in cluster 1 showed a lower correlation to EP2_POSTN, iCAF_CHI3L1, and imE1_FLT1, and had the most favorable clinical outcomes, whereas patients in cluster 3 vice versa. In addition, we found the cell number to be comparable across patients, with a slightly higher number of cells in patient 6 (Referee#2 point 2).

Reference

Alzamil L, Nikolakopoulou K, Turco MY (2021) Organoid systems to study the human female reproductive tract and pregnancy. *Cell Death Differ* 28: 35-51

den Boon JA, Pyeon D, Wang SS, Horswill M, Schiffman M, Sherman M, Zuna RE, Wang Z, Hewitt SM, Pearson R *et al* (2015) Molecular transitions from papillomavirus infection to cervical precancer and cancer: Role of stromal estrogen receptor signaling. *Proc Natl Acad Sci U S A* 112: E3255-3264

Huang Y, Kim BYS, Chan CK, Hahn SM, Weissman IL, Jiang W (2018) Improving immune-vascular crosstalk for cancer immunotherapy. *Nat Rev Immunol* 18: 195-203

Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B *et al* (2018) Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 24: 1550-1558

Lamouille S, Xu J, Derynck R (2014) Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* 15: 178-196

Litwin TR, Irvin SR, Chornock RL, Sahasrabuddhe VV, Stanley M, Wentzensen N (2021) Infiltrating T-cell markers in cervical carcinogenesis: a systematic review and meta-analysis. *Br J Cancer* 124: 831-841

Meylan M, Petitprez F, Becht E, Bougouin A, Pupier G, Calvez A, Giglioli I, Verkarre V, Lacroix G, Verneau J *et al* (2022) Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* 55: 527-541 e525

O'Donnell JS, Teng MWL, Smyth MJ (2019) Cancer immunoediting and resistance to T cell-based immunotherapy. *Nat Rev Clin Oncol* 16: 151-167

Togashi Y, Shitara K, Nishikawa H (2019) Regulatory T cells in cancer immunosuppression - implications for anticancer therapy. *Nat Rev Clin Oncol* 16: 356-371

Wherry EJ, Kurachi M (2015) Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* 15: 486-499

Dear Dr Li,

Thank you for submitting your revised manuscript for consideration by the EMBO Journal. It has now been re-evaluated by three referees whose comments are shown below.

As you will see, the referees state that the work has been improved considerably by the revisional experiments. At the same time, they point to persistent important shortcomings of the analysis. In light of the overall positive recommendations, I would

like to invite you to submit a revised version of the manuscript. I need to stress though that addressing the remaining concerns of all three reviewers relating to additional experimentation required (ref#2, pt.4), adjustment of data display and methods annotation (ref#2, pts 1. 3; ref#3, pts.2,3), statistics applied (ref#3, pts. 1,4) and discussion of the findings (ref#3, pt2) need to be achieved to completeness to ensure further pursual of this study for publication at the EMBO Journal.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website: https://www.embopress.org/page/journal/14602075/authorguide#transparentprocess

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.

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Daniel Klimmeck

Daniel Klimmeck, PhD
Senior Editor
The EMBO Journal

Instructions for preparing your revised manuscript:

Please make sure you upload a letter of response to the referees' comments together with the revised manuscript.

Please also check that the title and abstract of the manuscript are brief, yet explicit, even to non-specialists.

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:
https://bit.ly/EMBOPressFigurePreparationGuideline
See also guidelines for figure legends: https://www.embopress.org/page/journal/14602075/authorguide#figureformat

At EMBO Press we ask authors to provide source data for the main manuscript figures. Our source data coordinator will contact you to discuss which figure panels we would need source data for and will also provide you with helpful tips on how to upload and organize the files.

IMPORTANT: When you send the revision we will require

- a point-by-point response to the referees' comments, with a detailed description of the changes made (as a word file).
- a word file of the manuscript text.
- individual production quality figure files (one file per figure)
- a complete author checklist, which you can download from our author guidelines (https://www.embopress.org/page/journal/14602075/authorguide).
- Expanded View files (replacing Supplementary Information)
Please see out instructions to authors
https://www.embopress.org/page/journal/14602075/authorguide#expandedview

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.

Further information is available in our Guide For 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 (24th Apr 2023). Please discuss the revision progress ahead of this time with the editor if you require more time to complete the revisions.

Referee #1:

In this revised MS, the authors had added correlation analysis between GCB/ joint model of tumor and stromal cells with clinical prognosis, also provided mIHC of clinical tumor samples to show the spatial relationship and abundance relevance of interacted immune cells. I thus suggest the acceptance of the manuscript.

Referee #2:

In this manuscript, Cao, Yue, Ruan, et al. perform scRNA-seq and scTCR-seq on cervical cancer (CC) tumor and adjacent normal tissues to elucidate the immune and stromal landscape. The authors identify an enrichment of immune cells in CC tumors, with cytotoxic large-clone T cells as potential effectors of the antitumor response, germinal center B cells as being associated with increased survival, and inflammatory fibroblasts as predictors of poor outcomes. This is a comprehensive dataset of CC samples and the revisions were substantial and added significant clarity. The mIF validation of scRNA-seq findings remains limited, as it does not account for inherent variability across samples in its current form. Several points should be addressed:

1. The authors confirm differences in immune cell subsets detected by scRNA seq with multiplexed IF, and the absolute percentages of immune cell subsets are discordant between the methods, a known limitation of scRNA-seq analysis. The methods section and legends are not clear on whether the samples used for mIF are the same as the samples for scRNA-seq, and the variation in immune cell abundance detected by mIF is not apparent based on the way the data is currently presented. Inclusion of such information is critical for interpreting the results and conclusions, e.g., that there is a decrease in stromal cells in tumor areas. The legend for Fig. EV1F should also be updated to reflect that this is quantification of mIF data and not scRNA-seq data.

2. TCRs that recognize HPV and CMV were analyzed, and both are detected in tumor samples, but few HPV-specific TCRs are detected. As a result, the conclusion that HPV-specific cellular immunity is elevated in the tumor area is not strongly supported by the data. The authors should provide potential explanations for this observation and discrepancies with previous studies.

3. GC gene signatures are identified and validated using mIF (Fig. EV4D). Standard deviations are not included in the data, and the plots should be updated to either include statistical variation or additional datapoints included to confirm that this phenotype occurs in multiple patients. Moreover, CellChat analysis identified myeloid cells as interacting with B cells and their spatial proximity to B cells in TLS should be confirmed.

4. cDC1s and cDC2s are detected in the myeloid cell cluster and can potentially give rise to mature DCs. The authors should confirm if plasmacytoid DCs are present and can also give rise to mDCs in order to clarify the conclusion that cDC1s and cDC2s are the origins of mDCs.

5. The colors in the legend in Figure 1D top panel do not appear to match the UMAP plots.

6. The DC_CLEC9A signature is referenced as being correlated with better outcomes; however, this correlation is not included in the figures (Figure 5H).

7. Figure 6: MACM should presumably be MCAM.

8. Line 223: Figure 2F should be referenced instead of Figure 2G.

9. Line 265: B lymphocyte functions in cancer have been studied, albeit are less characterized in CC. This statement should be updated accordingly.

10. Line 424: PVCs and CAFs having an identical source of MSCs is debatable and likely dependent on the subset of PVCs or CAFs. Both cell types express mesenchymal genes, but their origins are not completely understood, specifically in CC. This statement should be revised.

11. Line 446: the conclusion that fibroblast subsets alter CTL abundance through ECM deposition and remodeling should be softened, as such subsets of fibroblasts also express a number of secretory factors that are immunomodulatory.

12. The title is unclear: what do the authors mean by 'ecosystem'. Consider revising for clarity.

Referee #3:

The authors have addressed many of my concerns, thank you. I have a couple more comments that I would like to pass on.

1) The description of the methods has been improved, thank you. However, if I understood correctly the statistical analyses used appear to have some deficiencies.

1a) The method for comparing cell-type proportions as the t-test used is not appropriate for compositional data. See reference [1] below for an explanation and [2] for an example of a practical implementation (I do not particularly endorse the latter, alternative implementations are available which the authors may use if they please).

1b) The statistical test used in Figure 6H does not appear to model biological replicates adequately. From the description provided, the t-tests are performed across all single-cells pooled across replicates, this leads to pseudo-replication and artificial highly significant p-values (see for example 6H with four asterisks). The authors have 5 tumour-normal matched samples so these data can be readily modelled using mixed models with individual as random effect, see [3] for an example.

2) The table legends do not match the table files (e.g. table EV9 contains the antibody list, but the table legend says this should be in table EV10, etc).

3) I assume the table with signature genes is in EV8 (the title in the excel file suggests this is so). I could not find many of the signatures I was expecting in EV8, for example what were the genes used in 6I for the EP1, EP2_POSTN, F4 (is this column F4_CXCL14 in the table?), iCAF_CHI3L1, and imE1_FLT1 signatures?

4) Table EV4 does not contain the expected list of differentially expressed genes.

Minor comments:
line 424: "To decrypt stromal cells in the cervical cancer microenvironment" -> the formulation "to decrypt stromal cells" makes no sense.

References:
[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6084572/
[2] https://www.nature.com/articles/s41467-021-27150-6
[3] https://www.nature.com/articles/s41467-021-21038-1

1st Revision - Editorial Decision - Follow up **5th Feb 2023**

Dear Dr Li,

Further to below message I herewith send you a list of issues related to formatting and data annotation as detailed below, which should be addressed at re-submission.

Thank you in advance for your consideration.

Please let me know any time should there be additional questions related to this.

Kind regards,

Daniel Klimmeck

Daniel Klimmeck, PhD
Senior Editor
The EMBO Journal

****

Formatting changes required for the revised version of the manuscript:

>> Adjust the title of the 'Declaration of Interests' section to 'Disclosure and Competing Interests Statement'.

>> Add a completed Author Checklist to the manuscript.

>> Author Contributions: Please 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

>> EV Figures: please limit EV figures to maximally five. Turn additional figure material into an 'Appendix' .pdf file with a ToC on its first page. Rename additional figures to "Appendix Figure S1" etc and adjust callouts in the text.

>>Remove the 'data not shown' statement on p.14 or add the respective data to the manuscript.

>> Data availability section: Please rename the current 'Data and code availability' section to 'Data availability'; make sure to release data privacy for the GSA dataset and make codes unique to this project available.

>> Dataset EV Legends: Table EV4, EV5, EV8 and EV9 should be renamed Dataset EV1-EV4, and the remaining EV table numbers will need to be adjusted accordingly. Please double-check the numbering of the tables, as there are discrepancies (Table EV5 is mislabeled Table EV4 etc.). Please also update the table callouts in the manuscript text.

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

Referee #1:

In this revised MS, the authors had added correlation analysis between GCB/ joint model of tumor and stromal cells with clinical prognosis, also provided mIHC of clinical tumor samples to show the spatial relationship and abundance relevance of interacted immune cells. I thus suggest the acceptance of the manuscript.
We thank the referee's comments.

Referee #2:

In this manuscript, Cao, Yue, Ruan, et al. perform scRNA-seq and scTCR-seq on cervical cancer (CC) tumor and adjacent normal tissues to elucidate the immune and stromal landscape. The authors identify an enrichment of immune cells in CC tumors, with cytotoxic large-clone T cells as potential effectors of the antitumor response, germinal center B cells as being associated with increased survival, and inflammatory fibroblasts as predictors of poor outcomes. This is a comprehensive dataset of CC samples and the revisions were substantial and added significant clarity. The mIF validation of scRNA-seq findings remains limited, as it does not account for inherent variability across samples in its current form. Several points should be addressed:

1. The authors confirm differences in immune cell subsets detected by scRNA seq with multiplexed IF, and the absolute percentages of immune cell subsets are discordant between the methods, a known limitation of scRNA-seq analysis. The methods section and legends are not clear on whether the samples used for mIF are the same as the samples for scRNA-seq, and the variation in immune cell abundance detected by mIF is not apparent based on the way the data is currently presented. Inclusion of such information is critical for interpreting the results and conclusions, e.g., that there is a decrease in stromal cells in tumor areas. The legend for Fig. EV1F should also be updated to reflect that this is quantification of mIF data and not scRNA-seq data.

We thank the referee's helpful suggestions. In the previous manuscript, we selected representative tissue slides of patient 1 to show the variation of cell abundance between the tumor and adjacent normal tissue. In this revision, the remaining five patients (patients 2-6) whose samples were used for scRNA-seq previously were further stained with multiplex immunohistochemistry (mIHC, Figure R1A). Immune components tended to enrich in the tumor area (T) compared to adjacent normal tissue (N), and T cells ($CD3^+$) and macrophages ($CD68^+$) were significantly more prevalent in the tumor area (Figure R1B). The density of B cells ($CD20^+$) is variable among different patients, but still showed relatively less infiltration in normal tissue. These data were in line with our conclusion from scRNA-seq data and supported a higher immune cell abundance in cervical cancer.

We updated the figure and legends in Fig.EV1F to reflect that the quantification of cell density in the tumor and adjacent normal tissue were based on mIHC data.

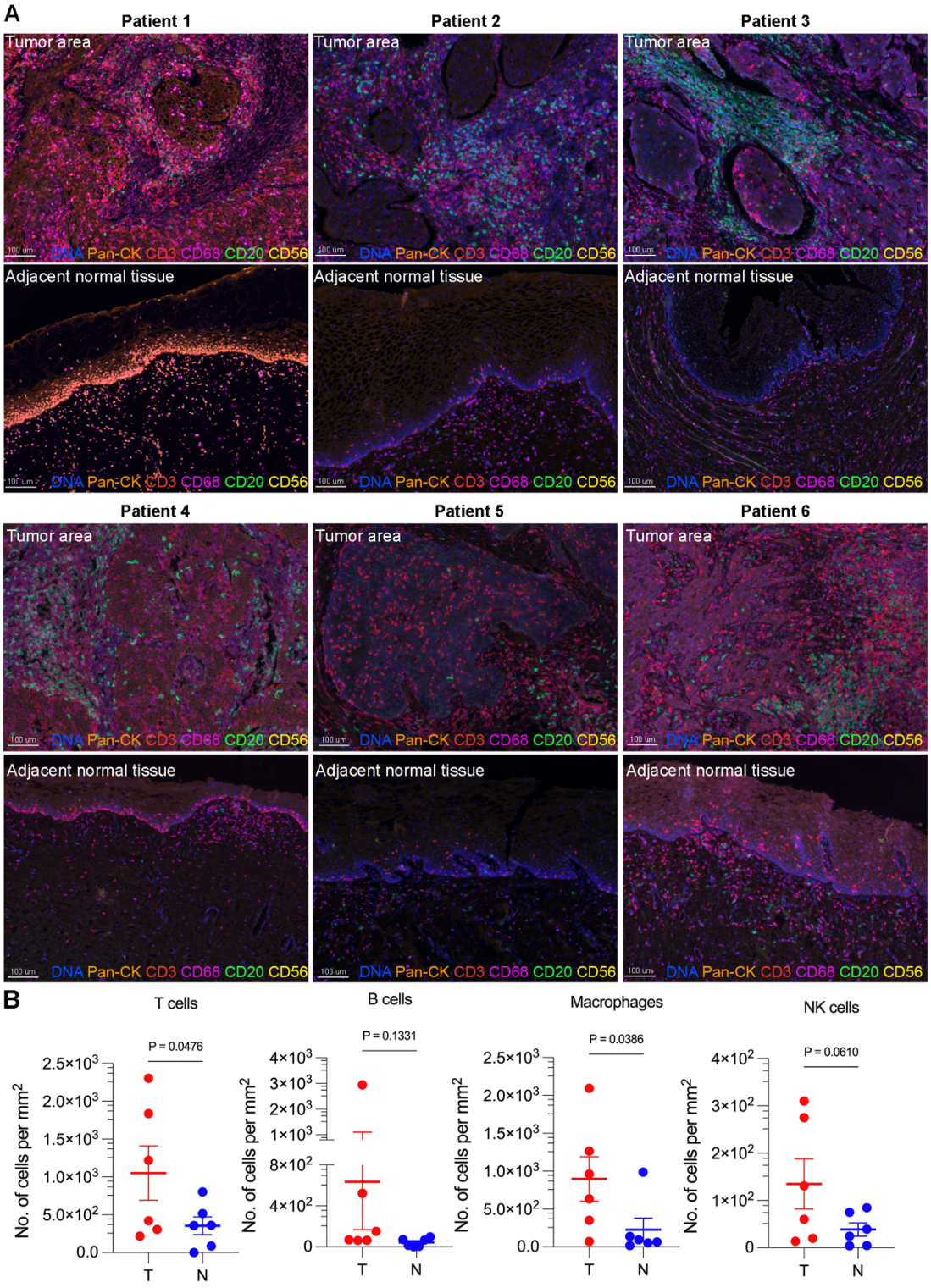


Figure R1 Quantification of immune cell subsets in tumor and adjacent normal regions using mIHC

A Representative mIHC of cervical cancer and adjacent normal tissue (n=6). Scale bars are 100 μm.

B Scatter plots showing the quantification of T cells ($CD3^+$), B cells ($CD20^+$), macrophages ($CD68^+$), and NK cells ($CD56^+$) in tumor area ($Pan\text{-}CK^+$) and adjacent

normal tissue (n=6). Represented Mean ± SEM. P value was measured by paired Student's t-test. Quantifications of immune cells were generated using HALO image analysis platfrom based on the mIHC images.

2. TCRs that recognize HPV and CMV were analyzed, and both are detected in tumor samples, but few HPV-specific TCRs are detected. As a result, the conclusion that HPV-specific cellular immunity is elevated in the tumor area is not strongly supported by the data. The authors should provide potential explanations for this observation and discrepancies with previous studies.

Our data indicated that the clonal size of HPV-specific T cells in the tumor is larger than the adjacent normal area, although the proportion of HPV-specific T cells is low. We identified HPV-specific T cells *in silico* by querying the VDJdb database (Shugay *et al*, 2018), which only provides 14 CDR3 sequences targeting the HPV E7 protein. However, the HPV-specific T cell responses from the tumor were not constrained to E7, they also recognized E1, E2, E4, E5, E6, and L1 proteins as dominant targets (Bhatt *et al*, 2020). Hence, only a small fraction of T cells targeting E7 could be identified as HPV-specific T cells, and this could probably explain why few HPV-specific TCRs were detected due to a lack of known HPV-specific TCRs. We have removed the conclusion from the abstract, and the potential explanations have been added in the discussion section as,

"We queried the TCRs in the VDJdb database to identify HPV-specific T cells, but the epitopes collected in this database were limited to E7 protein, which might lead to a severe underestimation of HPV-specific cells."

3. GC gene signatures are identified and validated using mIF (Fig. EV4D). Standard deviations are not included in the data, and the plots should be updated to either include statistical variation or additional datapoints included to confirm that this phenotype occurs in multiple patients. Moreover, CellChat analysis identified myeloid cells as interacting with B cells and their spatial proximity to B cells in TLS should be confirmed.

In Fig. EV4D, we summarized the distance between $CD4^+PD\text{-}1^+CXCL13^+$ Tfh cells and B cells in one mIHC slide to show their spatial proximity. Due to the numbers of Tfh cells being variable among patients, a single frequency distribution histogram cannot display each data point. In consideration of the referee's suggestion about statistical variation, we analyzed the frequency and density of Tfh cells in the TLS area, tumor area, and stroma. Tfh cells showed elevated frequency and density in TLSs, compared with parenchyma and stroma, indicating the predominant localization of Tfh in TLSs (Figure R2A). Additionally, we also verified the abundant presence of $CD14^+$ myeloid cells in TLSs (Figure R2B), supporting the cell-cell interactions between B cells and myeloid cells. However, $CD14^+$ myeloid cells were not specifically proximate to the TLS, but evenly distributed in the tumor and stroma area.

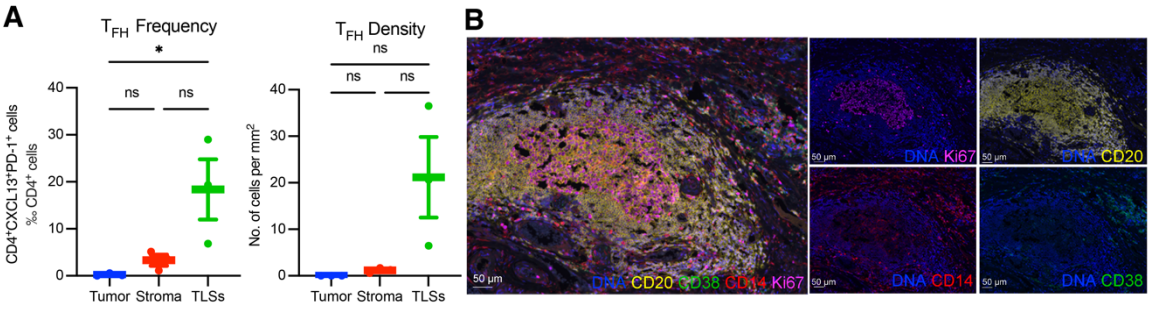


Figure R2 Spatial proximity of Tfhs and myeloid cells to TLSs

A Scatter plots showing the quantification of Tfh cells ($CD4^+PD\text{-}1^+CXCL13^+$) in TLSs, tumor and stroma (n=3). Represented Mean ± SEM. P value was measured by Holm-Sidak's multiple comparisons tests.

B Representative mIHC of $CD14^+$ cell localization in TLS. Scale bars are 50 μm.

4. cDC1s and cDC2s are detected in the myeloid cell cluster and can potentially give rise to mature DCs. The authors should confirm if plasmacytoid DCs are present and can also give rise to mDCs in order to clarify the conclusion that cDC1s and cDC2s are the origins of mDCs.

We thank the referee's suggestions. Plasmacytoid dendritic cells (pDCs) play an important role in the innate immune response to viral infections. We detected the expression of pDCs markers LILRA4, GZMB, and IL3RA. (Cheng *et al*, 2021), and a small fraction cDC2 cells (DC_CD1C) expressed LILRA4 and IL3RA (Figure R3A-B). However, due to the limited number of myeloid cells (n=3792) that were captured in our data, we cannot distinguish these cells from the cDC2 subset. A previous study also reported the presence of pDCs in CC, and suggested the potential role of pDCs in the innate immune response against HPV (Bontkes *et al*, 2005). Thus, we believe that pDCs are worthy of further exploration in future research.

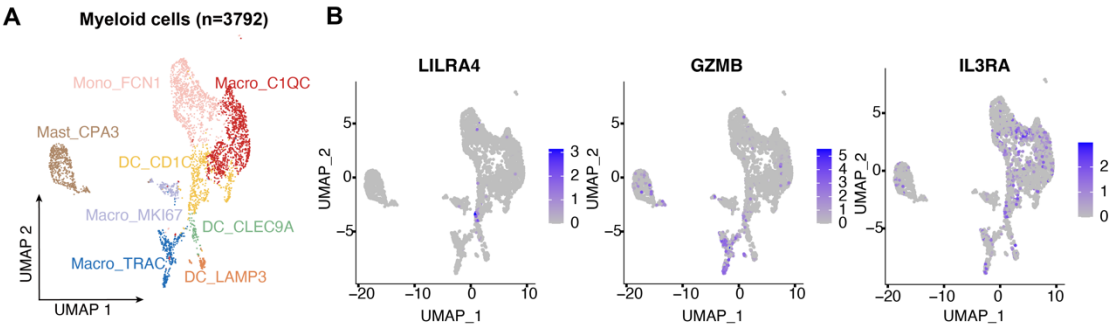


Figure R3 Expression of pDC markers in the CC scRNA-seq dataset

A UMAP plot showing the subtypes of 3,792 Myeloid cells. Cluster annotations are denoted and colored corresponding to cell types in the figure.

B UMAP plot showing the expression distribution of pDCs related genes in myeloid cells.

5. The colors in the legend in Figure 1D top panel do not appear to match the UMAP plots.

We have corrected the error in the figure legends (Figure R4).

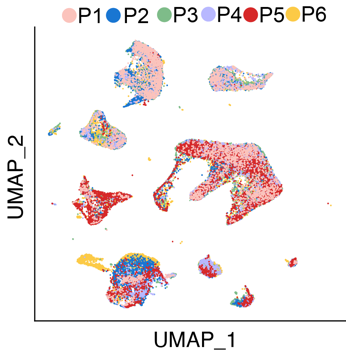


Figure R4 Updated figure showing the mixture of different patient samples in the C scRNA-seq dataset

UMAP plot of single cells profiled in the presenting work colored by patients.

6. The DC_CLEC9A signature is referenced as being correlated with better outcomes; however, this correlation is not included in the figures (Figure 5H).

The DC_CLEC9A signature is also related to better outcomes, and the survival graph has been added in Figure 5G.

7. Figure 6: MACM should presumably be MCAM.

We thank the careful review of this referee. This mistake has been corrected now.

8. Line 223: Figure 2F should be referenced instead of Figure 2G.

The sentence In line 223 has been modified as,

“Consistently, high-level Jaccard indices among CD8_GZMK, Tex_HAVCR2, and Tprol_MKI67, the hallmark of ongoing expansion (Li et al., 2019), indicate that these cells were mainly from local expansion (Fig 2F).”

9. Line 265: B lymphocyte functions in cancer have been studied, albeit are less characterized in CC. This statement should be updated accordingly.

We agree with the referee that the function of B cells has been studied before our study, and we modified the statement in line 265 as,

“Currently, the anti-tumor properties of B-lymphocytes have been increasingly recognized, yet are less characterized in CC."

10. Line 424: PVCs and CAFs having an identical source of MSCs is debatable and likely dependent on the subset of PVCs or CAFs. Both cell types express mesenchymal genes, but their origins are not completely understood, specifically in CC. This statement should be revised.

We thank the referee’s comments, and the sentence in line 424 has been modified as,

“We further conducted analyses of perivascular cells (PVCs) and cancer-associated fibroblasts (CAFs, Fig 6A-B).”

11. Line 446: the conclusion that fibroblast subsets alter CTL abundance through ECM deposition and remodeling should be softened, as such subsets of fibroblasts also express a number of secretory factors that are immunomodulatory.
The conclusion has been revised as,
“CTL correlation analysis demonstrated that PVC0_MCAM, PVC1_ACTA2, eCAF_DCN, and iCAF_CHI3L1 were inversely correlated with CTL (Fig 6D), partly as a result of ECM remodeling and immunomodulatory factors secreted by these subsets (Di Modugno et al, 2019; Sahai et al., 2020).”

12. The title is unclear: what do the authors mean by 'ecosystem'. Consider revising for clarity.
Our previous title tried to summarize that our manuscript revealed the baseline cellular landscape of the cervical cancer microenvironment and emphasized subsets contributing to tumor progression or anti-tumor immunity. We use the word ‘ecosystem’ here to describe the complex compositions and interactions of heterogeneous cancer cells, immune cells, and stroma cells in cervical cancer, which is like an ecosystem that depends on each other to maintain a steady state. We agreed with the referee that the previous title might be confusing. We have changed to a more straightforward title as
“Single-cell dissection of cervical cancer reveals key subsets of the tumor immune microenvironment”

Referee #3:

The authors have addressed many of my concerns, thank you. I have a couple more comments that I would like to pass on.

1) The description of the methods has been improved, thank you. However, if I understood correctly the statistical analyses used appear to have some deficiencies.

1a) The method for comparing cell-type proportions as the t-test used is not appropriate for compositional data. See reference [1] below for an explanation and [2] for an example of a practical implementation (I do not particularly endorse the latter, alternative implementations are available which the authors may use if they please).
We agree with this reviewer that cell-type proportions from scRNA-seq are compositional and the t-test is not appropriate. We used two recommended approaches including scCODA (Buttner *et al*, 2021) and ALDEx2 (Fernandes *et al*, 2014) recommended by this reviewer, and re-analyzed the changes of cell-type proportions between tumor and adjacent normal tissue from scRNA-seq data. For scCODA, we defined cell-types with significant changes at an FDR level of 0.2 (red line in Figure R5). For ALDEx2, the p-value between normal and tumor tissue is determined by Welch's t-test (stars in Figure R5).

The result of scCODA and ALDEx2 were in general consistent. Fibroblasts were abundant in adjacent normal tissues, and T-cells were enriched in the tumor region (Figure R5A), although it might not reach a statistical significance due to a small sample size, the trends were consistent with the immunostaining results (Figure R1). For the T-cell sub-lineage, Tregs (Treg_FOXP3) and exhausted T cells (Tex_HAVCR2) were significantly enriched in the tumor region and validated both by scCODA and ALDEx2 (Figure R5B). Other sub-lineage analyses were also consistent with our previous analyses (Figure R5C-E). We have updated these analyses in the revised manuscript, and kept both the results of scCODA and ALDEx to ensure the reliability of the results.

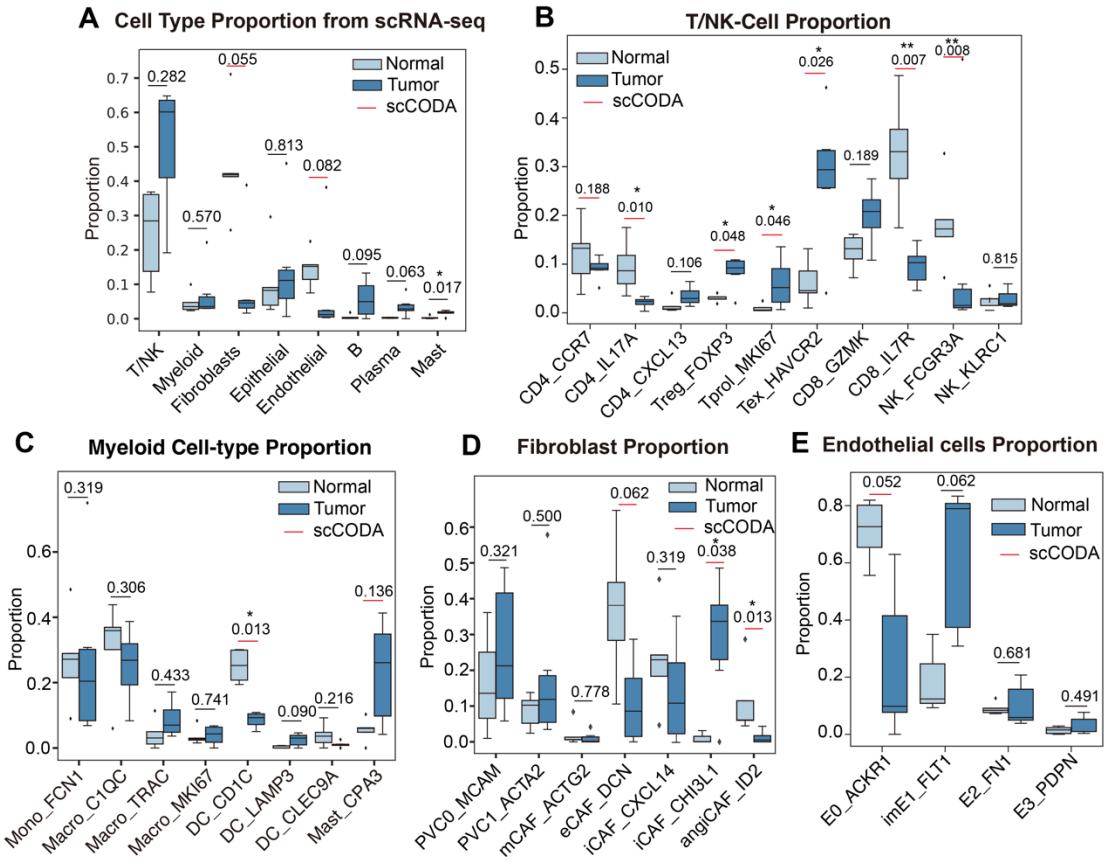


Figure R5 Statistical analyses of cell-type proportion difference between tumor and the matched normal samples

Boxplots showing the cell-type proportions of lineages (A), T/NK cells (B), myeloid cells (C), fibroblasts (D) and endothelial cells (E) for matched tumor and normal samples (n=5). Credible and significant results of scCODA are depicted as red bars. Stars indicate the significance calculated by ALDEx2 model (* $P < 0.05$, ** $P < 0.01$).

1b) The statistical test used in Figure 6H does not appear to model biological replicates adequately. From the description provided, the t-tests are performed across all single-cells pooled across replicates, this leads to pseudo-replication and artificial highly significant p-values (see for example 6H with four asterisks). The authors have 5 tumour-normal matched samples so these data can be readily modelled using mixed models with individual as random effect, see [3] for an example.

We apologize for the misleading description of the T-cell dysfunctional score for the endothelial marker genes. We used the TIDE algorithm (Jiang *et al*, 2018) to calculate the T-cell dysfunction score of each gene in the TCGA CESC cohort. The TIDE algorithm assumes that the high infiltration of cytotoxic T-cells should be correlated with better patient prognosis. Then it applied an interaction test based on Cox-PH regression to identify those genes which can perturb the beneficial effect of T-cell infiltration on patient prognosis. The higher T-dysfunction score represents the genes might be related to the dysfunction status of T-cells, thus affecting its association with survival.

For the top 50 marker genes of each endothelial subset, we calculated their T-dysfunction score in the TCGA CESC cohort. In the revised manuscript, we reanalyzed the statistical significance of T-dysfunction scores with the Kruskal-Wallis test, and the differences between imE1 (E1_FLT1) with another subset were calculated by Wilcox rank-sum test with Benjamini-Hochberg correction.

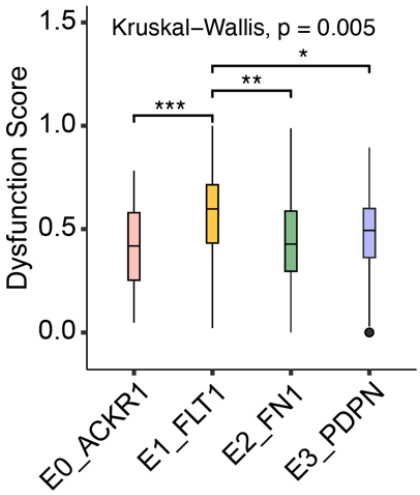


Figure R6 Difference of T-cell dysfunction score of the top 50 marker genes between endothelial populations
Boxplot showing the T-cell dysfunction score of marker genes (n=50) in endothelial cell populations; **P < 0.01, ***P < 0.001 by Wilcox rank-sum test with Benjamini-Hochberg correction.

2) The table legends do not match the table files (e.g. table EV9 contains the antibody list, but the table legend says this should be in table EV10, etc).
We apologize for the error here, and all table legends have been updated in the revised manuscript.

3) I assume the table with signature genes is in EV8 (the title in the excel file suggests this is so). I could not find many of the signatures I was expecting in EV8, for example what were the genes used in 6I for the EP1, EP2_POSTN, F4 (is this column F4_CXCL14 in the table?), iCAF_CHI3L1, and imE1_FLT1 signatures?
In the last revision, we changed the names of several stromal subsets to highlight their

characteristics, but they were not updated in the table EV8. All subset names have now been corrected, and EP1's signature has also been updated. In this revision, Dataset EV4 provides signature genes for survival analyses and ICB response prediction.

4) Table EV4 does not contain the expected list of differentially expressed genes.
In this revision, we have updated all the supplementary tables and datasets. Dataset EV1 provide lists of differentially expressed genes for all the subsets.

Minor comments:
Line 424: "To decrypt stromal cells in the cervical cancer microenvironment" -> the formulation "to decrypt stromal cells" makes no sense.
We thank the referee's comments, and the sentence in line 424 has been modified as, "We further conducted analysis of perivascular cells (PVCs) and cancer-associated fibroblasts (CAFs) (Figs 6A-B)."

References:
[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6084572/
[2] https://www.nature.com/articles/s41467-021-27150-6
[3] https://www.nature.com/articles/s41467-021-21038-1

Formatting changes required for the revised version of the manuscript:

Adjust the title of the 'Declaration of Interests' section to 'Disclosure and Competing Interests Statement'.
We have revised the title to 'Disclosure and Competing Interests Statement'.

Add a completed Author Checklist to the manuscript.
The Author Checklist for this manuscript has been attached in this revision.

Author Contributions: Please 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
The author contribution information has been removed from the main text.

EV Figures: please limit EV figures to maximally five. Turn additional figure material into an 'Appendix' .pdf file with a ToC on its first page. Rename additional figures to "Appendix Figure S1" etc and adjust callouts in the text.
EV Figures have been rearranged to meet the requirement, and an 'Appendix.pdf' file

with a context has been added into the submission.

Remove the 'data not shown' statement on p.14 or add the respective data to the manuscript.
The ‘data not shown’ statement has been removed.

Data availability section: Please rename the current 'Data and code availability' section to 'Data availability'; make sure to release data privacy for the GSA dataset and make codes unique to this project available.
We have renamed data availability section to ‘Data availability’, and have release the data privacy.

Dataset EV Legends: Table EV4, EV5, EV8 and EV9 should be renamed Dataset EV1-EV4, and the remaining EV table numbers will need to be adjusted accordingly. Please double-check the numbering of the tables, as there are discrepancies (Table EV5 is mislabeled Table EV4 etc.). Please also update the table callouts in the manuscript text.
The table legends and numbers have updated as requirement.

Consider additional changes and comments from our production team as indicated by the .doc file enclosed and leave changes in track mode.
We thank the careful examination of you and your colleagues, we have modified the manuscript as your suggestions.

To Dr. Killet Vivian:

Fig. 1I
Please define the number and the nature, i.e. biological or technical, of the replicates.
We thank for your detailed review and suggestions.
The Kruskal-Wallis test is a non-parametric method for testing whether samples originate from the same distribution. To test whether the EMT score of malignant cell subpopulations differed significantly, we applied Kruskal-Wallis test as previously reported (Ji *et al*, 2020). In this situation, epithelial cells should be considered as the replicates, and it is generally unspecified. We then conducted Benjamini-Hochberg correction to reduce probability of false positive.

Fig. 2J
Please define the number and the nature, i.e. biological or technical, of the replicates.
We tested whether the size of clones specific to HPV or CMV differed by Wilcox test. The sample size is determined by the number of T-cell clones that are specific to HPV or CMV. The number of replicates is the number of patients (n=6) and have been labelled in the figure legend.

Fig. 3H

Please define the number and the nature, i.e. biological or technical, of the replicates.
The number of patients which are classified as responders or non-responders are noted in the figure. No replicates were used in the analyses.

Fig. 6H
Please define the number and the nature, i.e. biological or technical, of the replicates.
The number of analyzed genes were labeled in Figure 6H.

In addition, we have modified the figure legends in accordance with your suggestions.

Reference:

Bhatt KH, Neller MA, Srihari S, Crooks P, Lekieffre L, Aftab BT, Liu H, Smith C, Kenny L, Porceddu S *et al* (2020) Profiling HPV-16-specific T cell responses reveals broad antigen reactivities in oropharyngeal cancer patients. *J Exp Med* 217

Bontkes HJ, Ruizendaal JJ, Kramer D, Meijer CJ, Hooijberg E (2005) Plasmacytoid dendritic cells are present in cervical carcinoma and become activated by human papillomavirus type 16 virus-like particles. *Gynecol Oncol* 96: 897-901

Buttner M, Ostner J, Muller CL, Theis FJ, Schubert B (2021) scCODA is a Bayesian model for compositional single-cell data analysis. *Nat Commun* 12: 6876

Cheng S, Li Z, Gao R, Xing B, Gao Y, Yang Y, Qin S, Zhang L, Ouyang H, Du P *et al* (2021) A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell* 184: 792-809 e723

Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB (2014) Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2: 15

Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, Guo MG, George BM, Mollbrink A, Bergenstrahle J *et al* (2020) Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Cell* 182: 497-514 e422

Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B *et al* (2018) Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* 24: 1550-1558

Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, Komech EA, Sycheva AL, Koneva AE, Egorov ES *et al* (2018) VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res* 46: D419-D427

Dear Dr Li,

Thank you for submitting the revised version of your manuscript. We have now evaluated your amended manuscript and concluded that the remaining minor concerns by the experts and on the formatting side 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 again 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.

Kind regards,

Daniel Klimmeck

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

Corresponding Author Name: Fang Li	
Journal Submitted to: The EMBO Journal	
Manuscript Number: EMBOJ-2022-110757R-Q	

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

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
- 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

2. Captions

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

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements.
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:

  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

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?	Not Applicable	

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	Details for all antibodies used in this study are included in Reagents tools table. This table includes supplier names and catalog numbers.

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.	Not Applicable	

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.	Not Applicable	
Primary cultures: Provide species, strain, sex of origin, genetic modification status.	Not Applicable	
Report if the cell lines were recently **authenticated** (e.g., by STR profiling) and tested for mycoplasma contamination.	Not Applicable	

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.	Not Applicable	
Animal observed in or captured from the field: Provide species, sex, and age where possible**.**	Not Applicable	
Please detail **housing and husbandry conditions**.	Not Applicable	

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	Table EV1 contains the age, gender, and ethnicity of all participants.

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?	Not Applicable	

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.	Not Applicable	

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 was not predetermined or formally justified for statistic power for each expreiment, according to the recognized standards in this field. The statement could be found in Method section.
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?	Not Applicable	
Include a statement about **blinding** even if no blinding was done.	**Yes**	Blinding of the investigator was not considered in this study. The statement could be found in Method section.
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.	**Yes**	The inclusion/exclusion criteria for sample collection and single cell data analysis were described in Methods.
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**	The statistical tests were justified as appropriate and were described in Methods.

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.	Not Applicable	
In the figure legends: define whether data describe **technical or biological replicates**.	**Yes**	The information is 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**	This study was supervised and assessed by the Ethics Committee Board of Tongji University-affiliated East Hospital (No. 2019-133). The statement can be found in the Methods section.
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**	Written consent was obtained from all patients prior to collection of tissue following pre-approved protocols, and the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report. The statement can be found in the Methods section.
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 approva**l (IRB or equivalent committee(s), provide reference number for approval. Include a statement of compliance with ethical regulations.	Not Applicable	
Studies involving **specimen and field samples:** State if relevant **permits** obtained, provide details of authority approving study; if none were required, explain why.	**Yes**	Patient tissues used in this work were collected under protocols detialed described in Methods section.

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**	Single-cell RNA sequencing dataset is available in Genome Sequence Archive (GSA) repository with the accession number HRA001742. The information can be found in the Data and code 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**	The information can be found in the Data and code availability section.