nature portfolio

Peer Review File

Single-cell sequencing reveals immune features of treatment response to neoadjuvant immunochemotherapy in esophageal squamous cell carcinoma

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
Zhenlin Yang, He Tian, and Bozhao Li et al. conducted a single-cell analysis on patients with ESCC who underwent neoadjuvant immunotherapy. The authors identified that mast cell abundance, B cells, and tertiary lymphoid structures were related to clinical response. They also showed a CXCL13 effect on tumor suppression when combined with ICI. In addition, lineage transition between T cell phenotypes by immunotherapy was analyzed. Finally, by analyzing epithelial cells, several gene expression was correlated to clinical response. Although they showed many data, this reviewer found that their analysis primarily showed only correlation but not underlying mechanisms and that the manuscript lacked clarity.

Major points
1) Neoadjuvant combinational therapy did modify the composition of epithelial and non-epithelial cells (Fig.2a, 3a, 4a, 6a, 7a, etc.). However, all comparisons between responders and non-responders were performed using combined data from pre and post-treatment samples. The effect of therapy was not uniform in each sample, undermining the validity of the analysis.

2) Pre-treatment samples were obtained through endoscopic biopsy, and post-treatment samples were from surgical specimens. Because biopsy samples were derived from only the superficial layer of the esophageal wall, the cell composition of biopsy samples may largely differ from surgical specimens. Were there no differences between data from endoscopic biopsy and surgical specimens?

3) Among 10 ESCC patients undergoing single-cell analysis, there were only two non-responders. This small number of non-responders compromises the validity of their results.

4) Follicular B cells and memory B cells were enriched in responders. Validation analysis in another cohort is strongly recommended to support their claim.

5) In the analysis of T cells, dysfunctional state markers were stronger in non-responders. However, CXCL13, which is also such a marker, was correlated with clinical response. How is this contradictory observation explained?

6) In the mouse experiments, they utilized a cell line sensitive to PD-1 antibody therapy. CXCL13 is already known to recruit immune cells to cancer tissue, as they described in their discussion. Can CXCL13 administration enhance PD-1 efficacy in other cell lines which are insensitive to PD-1 antibody alone?

7) In lines 498-501, does the lineage transition pattern discriminate responders from non-responders?

Reviewer #2:
Remarks to the Author:
This manuscript of Yang et al entitled “Single-cell analyses reveal the cellular and molecular dynamics of esophageal squamous cell carcinoma during neoadjuvant immunotherapy” described an analyzed of the TME of 10 ESCCs undergoing neoadjuvant chemo and PD-1/PD-L1 targeting therapy using scRNA sequencting technologies. Thereby they observed that that pretreatment mast cell abundance, post treatment TLS presence and CXCL13 expression was associated with efficacy of chemo+anti-PD-1/PD-L1 treatment. Mast cell and CXCL13 expression was also studies in a larger cohort of 70 patients undergoing the same treatment and CXCl13 and anti-PD-1 treatment combination were studied in a mouse models.

Although neoadjuvant immunotherapy is not standard of care for ESCCs, the strong point of this study is that the dynamics of the immune microenvironment could be analyzed due to the neoadjuvant setting in which pretreatment tumor biopsies could be analyzed in parallel to the surgical resection specimen and that results could be validated in a separate cohort. Furthermore, scRNAseq and TCR seq are very powerful tools to analyze the phenotypical features of the immune microenvironment. However, it needs to be stressed that for only from 7 pre- and 8 post treatment samples scRNA seq worked and that only from 5 samples, pre and post treatment tissues actually matched.
Also, within scRNA seq studies, all spatial relation of tumor and immune cells are lost in this setting which has shown to be very important.

In general I think this study helps to understand response to chemotherapy and immunotherapy combinations in ESCC but the way the data are described in this draft of the manuscript makes it difficult to appreciate the main findings.

Major concerns:
- Throughout the manuscript no differences are made for pre and post treatment samples while pre and post treatment samples should not be done combined in the same analysis. In general, there should only be one data point per patients and I have the impression that pre- and post treatment samples from the same patient are in the same response group (for instance in Figure 1 and 2). I clearly recommend to annotate in the text and in the figures whether the analyses are from pre or post treatment samples.
- While the authors state that they will analyze the dynamics of the immune infiltrate during treatment, there is only matched data from 5 patients. I recommend to to add more matched patients to the analyses, especially from nonresponders as this is the group where we need to identify mechanisms of chemo-immunotherapy resistance.
- Throughout the manuscript the authors discuss resistance to immunotherapy while patients are also treated with chemotherapy as well. This should be changed.

Other concerns per section:

Introduction:
- Please correct that the effects of neoadjuvant chemoradiotherapie are limited in ESCCs as the CROSS trial (Hagen et al, NEJM, 2012) showed that nCRT results in a complete pathological response in ~44% of ESCC.
- Please rephrase the wording “the mysteries of clinical efficacy” to make it more scientific.
Methods:
- Please indicate in which setting these patients received neoadjuvant chemoimmunotherapy. Is this standard of care of were these patients part of a trial? Please indicate which anti-PD-1 or anti-PD-l1 agents and which chemotherapy regime was used. How uniform was it?
- Please describe why it was chosen to analyze sample post neoadjuvant treatment while there was actually a complete pathological response….which area of the resection specimen were taken when no tumor was visable anymore.
- Please describe the process of tumor biopsy to ensure that biopsies were taken from the tumor area. Is anything known about how many tumor biopsies must be taken to ensure tumor presence?
- Were the IHC data from 45 patients the same patients from whom scRNA data were available?

Results
- Please describe in more detail for which patients there were matching pre- and posttreatment data. What kind of response did these patients have? How uniform was the treatment they received?
- Please indicate in Figure 1A from which samples you actually have scRNA seq data which is 7/10 pretreatment cases and not the 10 as shown in Figure 1A. To which response groups do these belong?
- Figure 1d: were pre- and post treatment data combined? Or is this only pretreatment?
- Figure 2: please indicate on every place whether the results are from the pre- or posttreatment sample. Pre and post treatment samples should be analyzed separately: I get the impression that pre-

and post samples are combined and that the pre and post treatment sample from the same patient are in the same R and NR group. Please explain.
- Figure 2e: please explain how the mast cell scoring was done: was this done by a pathologist? Also, please explain the level of response in this group as in the scRNAseq cohort an pCR, mCR and non-MPR scoring system was used.
- I don not agree with the statement that post treatment TLS presence indicates that TLSs might mediate a favorable response to immunotherapy in ESCC as this can only be done when you know something about pretreatment presence. Please revise this part of the manuscript.
- In general: please describe in more detail what was found about the matched tumor samples. What happens within the patient? How do immune features change during treatment and how does this correlate with response?
- Please explain how CXCL13, as a marker of dysfunctional CD8 T cells, can be a marker of a good response to immunotherapy? Cant it be that patients with high CXCL13 levels just had higher CD8 levels and are just more inflamed tumors?

Discussion:
- Please change immunotherapy for chemo-immunotherapy combinations throughout the discussion.
- Please add pre-treatment to line 530 to indicate in which clinical setting the mast cells were measured.
- Line 552: the authors assume that pretreatment TLS presence is association with response to immunotherapy while

Reviewer #3:
Remarks to the Author:
The paper "Single-cell analyses reveal the cellular and molecular dynamics of esophageal squamous cell carcinoma during neoadjuvant immunotherapy" by Yang et al reports a detailed analysis of single cell data of ESCC patients receiving neoadjuvant immunotherapy, aiming at the identification of the cellular populations that mostly influence the response to therapy.
The paper is very straightforward, high quality data and analysis are presented in a clear way.
However, I found some criticalities that undermine the conclusions.
First of all, some samples (p1 pre, p1 post, p2 pre and p8 post) show an extremely high percentage of epithelial cells. Given that all the analysis are performed on compositions, I wonder how those samples may impact results, also given that one of such samples is in the NR class which, in turn, is the smallest class of the entire dataset. Removing epithelial cells (but also muscle and neurons) will greatly shift the percentages.
One major claim of the paper is that some populations (Mast, B and CD8 cells) are enriched in responders when compared to non responders and that such changes may be at the basis of the differential response to therapy.
Nevertheless, analysis of compositional datasets should not be performed comparing percentages of a single populations across conditions, without any regard of the whole composition. For example, the lower amount of mast cells (Figure S2d) may be due to the lower amount overall of blood cells in NR, given that P8 is characterized by ~90% epithelial cells and the remaining NR samples come only from a single patient (P7pre and post).
I tried to retrieve the original cell counts for each population (given the reported percentages), removed all non-blood cells and performed compositional data analysis using scCODA (doi: 10.1038/s41467-021-27150-6). Using the default Bayesian model I could not find any credible change between R and NR:

	times_credible	pct_credible	is_credible
Bfoc	0	0.0	False
Blood_Endothelial_cells	0	0.0	False
Bmem	1	0.059	False

Bn 1 0.059 False
CD4_Treg 0 0.0 False
CD8_Teff 0 0.0 False
CD8_Tem 0 0.0 False
CD8_Tex 0 0.0 False
Dendritic_cells 0 0.0 False
GCB 1 0.059 False
Lymphatic_Endothelial_cells 0 0.0 False
Macrophages 0 0.0 False
Mast_cells 0 0.0 False
Other_CD4_T 0 0.0 False
Other_CD8_T 1 0.059 False
Plasma_B 0 0.0 False
Tn 0 0.0 False

I had the same conclusions using additive logratio (ALR) transform of the data paired with a t-test. I think authors must properly perform compositional data analysis on these data, also including the pre/post covariate (I find strange to weight samples before and after treatment in in the same way) before drawing any conclusion. Any analysis will suffer from a unbalanced representation of R/NR, it would be beneficial to increase the number of NR samples (although I understand it may be difficult). Lastly, I would point that even if the analytical procedure was correct, it is not clear which statistical test is performed throughout the text: it is written in the methods that t-test or Wilcoxon is used, but this is not reported alongside each comparison. This is important as different tests give different results and t-test may not be the appropriate test when comparing percentages. Wilcoxon may be in fact more appropriate (and more powerful, as percentages are possibly not normally distributed) but results in not significant tests, for example using the Bmem percentages in table S2:

```
> R <- c(0.983606557, 0.617723925, 0.858737656, 0.924855491, 3.646833013, 1.144640999,
6.115107914, 0.021463833, 16.15798923, 18.21862348, 8.230655495, 10.86172345)
> NR <- c(1.179245283, 0.127145582, 0.30647986)
> wilcox.test(R, NR)

Wilcoxon rank sum exact test

data: R and NR
W = 28, p-value = 0.1802
alternative hypothesis: true location shift is not equal to 0
```

Dear reviewers,

Thank you very much for giving us an opportunity to revise our manuscript. We thank the editor and reviewers very much for the constructive comments and suggestions on our manuscript entitled "Single-cell sequencing reveals immune features of treatment response to neoadjuvant immunochemotherapy in esophageal squamous cell carcinoma" (Manuscript Number: NCOMMS-23-13382A-Z). To address the reviewers' concern about insufficient sample size in the first draft, we collected another 8 samples from esophageal squamous cell carcinoma (ESCC) patients receiving neoadjuvant immunochemotherapy (nICT) and performed scRNA-seq paired with scTCR-seq analysis, expanding the sample size of single-cell data up to 23 samples. After revision, single-cell data of enrolled samples was classified by distinct treatment timepoints and efficacy groups, and then performed with a comprehensive comparison analysis. We next found several important immune features of non-responders, revealed the underlying mechanism of therapeutic resistance to nICT and uncovered novel promising targets for enhancing anti-PD-1 efficacy. Importantly, we also retrospectively extracted 108 matched pre- and post-treatment formalin-fixed paraffin-embedded samples from ESCC patients receiving nICT treatment from our center as an independent validation cohort, and conducted RNA-seq and spatial immunofluorescence analysis. These important findings from single-cell analysis were further validated by a larger validation cohort. In summary, we have studied reviewers' comments carefully and have revised our manuscript accordingly. Attached please find the revised version and relevant document, which we would like to submit for your kind consideration.

Responds to the reviewer's comments:

Reviewer #1:

Reviewer #1, expertise in ESCC genomics (Remarks to the Author):

The authors conducted a single-cell analysis on patients with ESCC who underwent neoadjuvant immunotherapy. The authors identified that mast cell abundance, B cells, and tertiary lymphoid structures were related to clinical response. They also showed a CXCL13 effect on tumor suppression when combined with ICI. In addition, lineage transition between T cell phenotypes by immunotherapy was analyzed. Finally, by analyzing epithelial cells, several gene expression was correlated to clinical response. Although they showed many data, this reviewer found that their analysis primarily showed only correlation but not underlying mechanisms and that the manuscript lacked clarity.

Major points

1) Neoadjuvant combinational therapy did modify the composition of epithelial and non-epithelial cells (Fig.2a, 3a, 4a, 6a, 7a, etc.). However, all comparisons between responders and non-responders were performed using combined data from pre- and

post-treatment samples. The effect of therapy was not uniform in each sample, undermining the validity of the analysis.

Author response: Thank you for your valuable comments. Neoadjuvant immunochemotherapy greatly changed tumor microenvironment (TME) of ESCC. Pre- and post-treatment samples showed prominent heterogeneity in TME contents. Yet as pointed out, analysis based on combined data from pre- and post-treatment samples overlooks the dynamic changes in the TME during immunochemotherapy. In response, we collected another 8 samples from non-responders and expanded the sample size of scRNA-seq analysis to 23. The scRNA-seq data were classified according to treatment timepoints and performed subsequent analysis using separated data from pre- and post-treatment samples. This refined approach led to the discovery of several novel underlying mechanisms of therapeutic resistance, including high tolerance induced by $CD4^+$ Tregs, terminal exhaustion phenotype of $CD8^+$ Tex, and the enrichment of $LRRC15^+$ fibroblasts and SPP1+ macrophages in post-treatment specimens, all closely associated with therapeutic resistance to immunochemotherapy.

2) Pre-treatment samples were obtained through endoscopic biopsy, and post-treatment samples were from surgical specimens. Because biopsy samples were derived from only the superficial layer of the esophageal wall, the cell composition of biopsy samples may largely differ from surgical specimens. Were there no differences between data from endoscopic biopsy and surgical specimens?

Author response: Thanks so much for the reviewer's question. In this study, the pre-treatment samples were collected from locally advanced ESCC tumors that were directly macroscopic under endoscopy. The sampling was performed directly at the site of visible tumor at a depth of at least submucosa layer, avoiding necrotic area. As shown in Figure 1g of the revised manuscript, pre-treatment biopsy samples were marked by an obviously higher cell proportion of epithelial cells, which was consistent with the characteristic of endoscopic biopsy consisting of a large amount of ESCC tumor cells from the superficial layer. In contrast, post-treatment surgery samples were observed with higher cell proportion of endothelial cells and fibroblasts, in line with our knowledge that malignant epithelial cells decreased due to drugs treatment, while fibroblasts infiltrate the regression area. In addition, surgical resection samples contained more mesenchymal components, with higher proportions of endothelial cells and fibroblasts. Of note, no significant difference was observed in the fraction of $CD45^+$ cells between pre- and post-treatment samples. Collectively, both endoscopic biopsy and surgical resection specimens showed a diversity of cell types and heterogeneity of cell components. The differences in cell proportions of epithelial cells, endothelial cells and fibroblasts, in consistent with our expectation, could also be explained by our current understanding of solid tumors. Given the differences between endoscopic biopsy and surgical specimens, we next classified scRNA-seq data into pre- and post-treatment groups, and performed a comprehensive comparison analysis to uncover the TME dynamics in ESCC during immunochemotherapy.

3) Among 10 ESCC patients undergoing single-cell analysis, there were only two non-responders. This small number of non-responders compromises the validity of their results.

Author response: The reviewer's question and suggestion are very valuable to our study. As displayed in the study overview (Figure 1a), we have collected another 8 samples from ESCC patients receiving neoadjuvant immunochemotherapy and performed scRNA-seq paired with scTCR-seq analysis. On the basis of the single-cell profile of 23 samples, we further reported several interesting findings. Moreover, we also retrospectively extracted 54 pairs of matched pre- and post-treatment formalin-fixed paraffin-embedded (FFPE) samples from ESCC patients receiving neoadjuvant immunochemotherapy from our center as an independent external validation cohort, and conducted RNA-seq analysis. The important findings revealed by single-cell analysis were further validated by transcriptomic analysis of FFPE samples from the validation cohort.

4) Follicular B cells and memory B cells were enriched in responders. Validation analysis in another cohort is strongly recommended to support their claim.

Author response: Thank you for highlighting this issue. As mentioned above, in order to improve the validity, we expanded sample size for single-cell analysis to 23 specimens and developed an external validation cohort of 108 samples. As shown in Supplementary Figure 4, in both $CD45^+$ cells and B cells, fraction of memory B cells was relatively higher in responders. Additionally, in the validation cohort, the infiltrating level of memory B cells in post-treatment samples from responders significantly increased compared to pre-treatment tumors, higher than that of post-treatment samples from non-responders. Yet, no significant difference was found in the infiltrating level of follicular B cells. According to the reviewer's recommendation, the revised manuscript features an independent validation cohort, furthering supporting our claim.

5) In the analysis of T cells, dysfunctional state markers were stronger in non-responders. However, CXCL13, which is also such a marker, was correlated with clinical response. How is this contradictory observation explained?

Author response: We sincerely thank the reviewer for the valuable feedback. In the revised manuscript, we changed to analyze pre- and post-treatment ESCC samples separately instead of using combined data. This refined approach provided additional evidence supporting the critical role of CXCL13 during neoadjuvant immunochemotherapy in ESCC. As shown in Figure 2, post-treatment non-responders were characterized by highly increased infiltration of $CD8^+$ Tex-CXCL13 cells with a more remarkable terminal exhaustion phenotype compared to responders. While in pre-treatment samples, tumors with higher infiltration score of $CD8^+$ Tex-CXCL13 showed improved therapeutic response to immunochemotherapy. Multiplex immunohistochemistry staining was performed and further validated the favorable role of baseline enrichment of $CXCL13^+CD8^+$ T cells in enhancing immunotherapy efficacy. $CD8^+$ Tex-CXCL13 cells in pre-treatment tumors were thought to be tumor-reactive progenitor $CD8^+$ Tex with high reprogrammability to respond to immune checkpoint blockade. In contrast, $CD8^+$ Tex-CXCL13 cells from

post-treatment non-responders were mainly terminal $CD8^+$ Tex, showing limited response to PD-1/PD-L1 inhibitors. To conclude, this study revealed the heterogeneity of $CD8^+$ Tex-CXCL13 phenotype among distinct treatment timepoints and efficacy groups. Enrichment of $CD8^+$ Tex-CXCL13 cells served as a predictor for better therapeutic response in pre-treatment tumors but also a marker for therapeutic resistance in post-treatment samples. We have comprehensively discussed these findings in the discussion part of the revised manuscript. Hopefully, our explanation could help understand the contradictory observation about the role of CXCL13 expression in $CD8^+$ Tex-CXCL13 cells.

6) In the mouse experiments, they utilized a cell line sensitive to PD-1 antibody therapy. CXCL13 is already known to recruit immune cells to cancer tissue, as they described in their discussion. Can CXCL13 administration enhance PD-1 efficacy in other cell lines which are insensitive to PD-1 antibody alone?

Author response: We sincerely thank the reviewer for the question. We agree that the mouse experiment using cell lines insensitive to PD-1 antibody would be more powerful to validate the favorable role of CXCL13 administration. In the present study, the mouse ESCC cell line used for animal experiment was mEC25, which was kindly provided by Prof. Li Fu. Currently, mEC25 is thought to be the first reliable syngeneic mouse ESCC cell line, proved to largely recapitulate the clinical features, molecular genetics, biological heterogeneity, and immune microenvironment of human ESCC. This makes it a valuable tool for exploiting the mechanism of anti-PD-1 therapy. Nevertheless, we have not yet found another reliable mouse ESCC cell line which is insensitive to PD-1 antibody alone and also suitable for designing novel anti-PD-1 based therapies with enhanced anti-tumor efficiency. We are keeping in touch with Prof. Li Fu and making efforts for developing a suitable mouse ESCC cell line with primary resistance to PD-1 antibody. Once the stable cell line is established, we will further validate our findings in the current study.

7) In lines 498-501, does the lineage transition pattern discriminate responders from non-responders?

Author response: We have considered the reviewer's question carefully. In Figure 4i of the revised manuscript, we present the analysis of lineage transition pattern based on T cell receptor (TCR) clonotypes in pre-treatment samples and post-treatment samples from both responders and non-responders. We observed a significant decrease in lineage transition between $CD8^+$ Teff cells with $CD8^+$ Tem or $CD8^+$ Tex cells in post-treatment samples from non-responders compared to responders. Equally importantly, non-responders also revealed lower TCR clonal expansion in $CD8^+$ Teff cells, indicating that the inadequately recognized tumor antigen in $CD8^+$ Teff cells may be a potential mechanism contributing to resistance to neoadjuvant immunochemotherapy in ESCC.

Reviewer #2, expertise in ESCC immunotherapy (Remarks to the Author):

This manuscript of Yang et al entitled "Single-cell analyses reveal the cellular and molecular dynamics of esophageal squamous cell carcinoma during neoadjuvant immunotherapy" described an analyzed of the TME of 10 ESCCs undergoing

neoadjuvant chemo and PD-1/PD-L1 targeting therapy using scRNA sequencting technologies. Thereby they observed that that pretreatment mast cell abundance, post treatment TLS presence and CXCL13 expression was associated with efficacy of chemo+anti-PD-1/PD-L1 treatment. Mast cell and CXCL13 expression was also studies in a larger cohort of 70 patients undergoing the same treatment and CXCl13 and anti-PD-1 treatment combination were studied in a mouse models.

Although neoadjuvant immunotherapy is not standard of care for ESCCs, the strong point of this study is that the dynamics of the immune microenvironment could be analyzed due to the neoadjuvant setting in which pretreatment tumor biopsies could be analyzed in parallel to the surgical resection specimen and that results could be validated in a separate cohort. Furthermore, scRNAseq and TCR seq are very powerful tools to analyze the phenotypical features of the immune microenvironment. However, it needs to be stressed that for only from 7 pre- and 8 post treatment samples scRNA seq worked and that only from 5 samples, pre and post treatment tissues actually matched. Also, within scRNA seq studies, all spatial relation of tumor and immune cells are lost in this setting which has shown to be very important. In general, I think this study helps to understand response to chemotherapy and immunotherapy combinations in ESCC but the way the data are described in this draft of the manuscript makes it difficult to appreciate the main findings.

Major concerns:

- Throughout the manuscript no differences are made for pre and post treatment samples while pre and post treatment samples should not be done combined in the same analysis. In general, there should only be one data point per patients and I have the impression that pre- and post-treatment samples from the same patient are in the same response group (for instance in Figure 1 and 2). I clearly recommend to annotate in the text and in the figures whether the analyses are from pre or post treatment samples.

Author response: We highly acknowledge the reviewer's recommendation. Neoadjuvant immunochemotherapy has greatly remodeled the tumor microenvironment (TME) of locally advanced ESCC. The TME landscape was highly heterogenous in pre- and post-treatment samples. Recognizing the importance of preserving information on TME dynamics during immunochemotherapy, we followed the reviewer advice. We separated single-cell profiles of pre-treatment biopsy specimens from post-treatment surgery samples, and conducted comparison analysis between pre- and post-treatment samples, as well as responders and non-responders. This allowed us to identify new mechanisms of therapeutic resistance. And as advised, we detailed the annotation in the revised text and figures, specifying the timepoint and efficacy group of samples for each analysis.

- While the authors state that they will analyze the dynamics of the immune infiltrate during treatment, there is only matched data from 5 patients. I recommend to add more matched patients to the analyses, especially from non-responders as this is the group where we need to identify mechanisms of chemo-immunotherapy resistance.

Author response: We have carefully considered the comments and made corresponding revisions accordingly. An additional eight ESCC patients showing

therapeutic resistance to neoadjuvant immunochemotherapy were recruited. We then performed scRNA-seq paired with scTCR-seq analysis. Through the single-cell profile of 23 specimens, we uncovered several new and important immune features of immunochemotherapy resistance. In addition, we retrospectively collected matched pre- and post-treatment formalin-fixed paraffin-embedded (FFPE) samples from 54 ESCC patients receiving neoadjuvant immunochemotherapy from National Cancer Center as an independent validation cohort, and performed transcriptomic sequencing. Our findings about the underlying mechanisms of therapeutic resistance to immunochemotherapy derived from scRNA-seq analysis were subsequently validated by transcriptome profile of the validation cohort.

- Throughout the manuscript the authors discuss resistance to immunotherapy while patients are also treated with chemotherapy as well. This should be changed.

Author response: Thank you for your kind suggestion. All ESCC patients enrolled in current study received neoadjuvant immunochemotherapy followed by surgical resection. We have carefully replaced "immunochemotherapy" to "immunotherapy" in the manuscript, to avoid confusion.

Other concerns per section:

Introduction:

- Please correct that the effects of neoadjuvant chemoradiotherapies are limited in ESCCs as the CROSS trial (Hagen et al, NEJM, 2012) showed that nCRT results in a complete pathological response in ~44% of ESCC.

Author response: Thank you for your thorough review and valuable feedback. For patients with resectable locally advanced ESCC, neoadjuvant chemoradiotherapy (nCRT) followed by esophagectomy is currently the standard-of-care treatment that improves survival compared with surgery alone. After nCRT, pathological complete response (pCR) rates were reported as 48.6% in the CROSS trial (37 cases) and 43.2% in the NEOCRTEC5010 trial (451 cases). We have corrected the statement about the efficacy of nCRT for ESCC in the first paragraph of the Introduction section.

- Please rephrase the wording "the mysteries of clinical efficacy" to make it more scientific.

Author response: Thank you for this suggestion. In response, we rephrased the wording "the mysteries of clinical efficacy" in the revised manuscript. The summary sentence now reads: "Collectively, our study elucidates the immune features of distinct response to nICT treatment and provides novel insights for potentiating anti-PD-1 efficacy in ESCC", which was more scientific and concise.

Methods:

- Please indicate in which setting these patients received neoadjuvant chemoimmunotherapy. Is this standard of care? were these patients part of a trial?

Author response: Thank you for your careful review and kind reminder. We have revised the introduction to include a detailed description about the clinical setting of these patients receiving nICT treatment. Currently, nCRT followed by surgical resection is the standard of care for patients with resectable locally advanced ESCC. Nevertheless, the limitations of nCRT, including increased postoperative mortality

risk and a high rate of distant recurrence, underscore the need for more effective and safer treatments. Accumulating evidences have proved that nICT treatment shows superior clinical benefits and is shifting the treatment paradigm across a variety of solid tumors, including operable locally advanced ESCC, which is also the clinical setting of our study. In our study, 9 samples from the single-cell cohort (23 patients) and 19 cases from the external validation cohort (96 patients) were part of clinical trial (NCT04460066), a phase II study published in 2023 by our team. Importantly, there is no overlap between these trial samples. And the remaining part was collected from real-world cohort. Detailed information was provided in Supplementary Data 1 and Data 5.

Please indicate which anti-PD-1 or anti-PD-L1 agents and which chemotherapy regime was used. How uniform was it?

Author response: Thank you for your question. The anti-PD-(L)1 agents in our study included Pembrolizumab, Camrelizumab, Tislelizumab, Sintilimab, and Socazolimab. These agents have demonstrated prominent clinical benefits in multiple solid tumors. As for chemotherapy, the regime used in patients in this study was Platinum-based chemotherapy, which was also widely applied in the neoadjuvant and adjuvant therapy of locally advanced ESCC and first-line therapy of metastatic ESCC. Detailed information of the neoadjuvant agents for each ESCC patients was described in Supplementary Data 1.

- Please describe why it was chosen to analyze sample post neoadjuvant treatment while there was actually a complete pathological response. Which area of the resection specimen were taken when no tumor was visable anymore?

Author response: We appreciate your detailed feedback. Those post-treatment ESCC samples with complete pathological response to nICT treatment also provide crucial immune feature information that were critical for enhanced therapeutic efficacy. Notably, responders' post-treatment samples exhibited significantly higher TCR clonal expansion of $CD8^+$ effector T cells, suggesting the importance of sufficient tumor antigen recognition in these cells. In our study, all ESCC patients in single-cell sequencing cohort received esophagoscopy before and after nICT treatment to precisely measure and localize primary tumor site. After surgical resection, Lugol's iodine was applied to classify regressed primary tumor bed and deeply stained para-tumor epithelial mucosa. Dissection and sampling were performed at the located primary tumor bed site, followed by rapid freezing pathology examination to confirm the tissue's pathological nature. Only samples confirmed to be primary tumor site underwent single-cell sequencing.

- Please describe the process of tumor biopsy to ensure that biopsies were taken from the tumor area. Is anything known about how many tumor biopsies must be taken to ensure tumor presence?

Author response: We sincerely appreciate your insightful question. In our study, we collected the pre-treatment biopsy samples through esophagoscopy at baseline. All pre-treatment samples enrolled in scRNA-seq were collected from locally advanced ESCC tumors that were clearly visable under esophagoscopy and could be obtained through biopsy. Importantly, the biopsy samples for diagnosis, collected at the same

site, underwent hematoxylin and eosin (H & E) staining, and were confirmed as tumors by experienced pathologists. Of note, results from CopyKAT analysis verified the presence of malignant epithelial cells in all biopsy samples. We have revised the methods section with a more detailed description of tumor biopsy process.

- Were the IHC data from 45 patients the same patients from whom scRNA data were available?

Author response: Thank you for the detailed review. As shown in Supplementary Data 5, after revision, the external validation cohort now comprises 96 patients, whose pre-treatment biopsy samples underwent H & E staining and post-treatment samples multiplex immunohistochemistry (IHC) staining. The 18 patients enrolled in single-cell analysis have no overlap with those in the external validation cohort.

Results:

- Please describe in more detail for which patients there were matching pre- and post-treatment data. What kind of response did these patients have? How uniform was the treatment they received?

Author response: Thank you for your question. In the revised manuscript, we have enhanced the details regarding the clinical setting for the 5 patients with matched pre- and post-treatment samples in single-cell sequencing. Supplementary Data 1 provides a breakdown: 2 cases achieved complete pathological response (pCR), 2 cases had major pathological response (MPR), and 1 case did not major pathological response (non-MPR). All these patients received PD-(L)1 blockade in combination with Platinum-based chemotherapy. For detailed clinical information of ESCC patients with matched pre- and post-treatment scRNA-seq data, please refer to Supplementary Data 1.

- Please indicate in Figure 1A from which samples you actually have scRNA seq data which is 7/10 pretreatment cases and not the 10 as shown in Figure 1A. To which response groups do these belong?

Author response: Thank you for your suggestion. After revision, scRNA-seq was performed on a total of 23 samples from 18 ESCC patients. This set included 7 pre-treatment esophagoscopy biopsy samples and 16 post-treatment surgical resection samples. The 7 patients with baseline biopsy samples included 3 cases with MPR, 3 cases with pCR, and 1 case with non-MPR, among which 5 cases had matched pre- and post-treatment specimens. Detailed information is provided in revised Figure 1a and Supplementary Data 1.

- Figure 1d: were pre- and post-treatment data combined? Or is this only pretreatment?

Author response: Figure 1d of the primary version showed the cell proportion of each patient based on the combination of pre- and post-treatment data. In our revised manuscript, the cell proportion of major cell types was displayed in pre- and post-treatment samples (Figure 1g), respectively.

- Figure 2: please indicate on every place whether the results are from the pre- or posttreatment sample. Pre and post treatment samples should be analyzed separately: I get the impression that pre- and post- samples are combined and that the pre and post

treatment sample from the same patient are in the same R and NR group. Please explain.

Author response: We appreciate your thorough review. TME of locally advanced ESCC was greatly remodeled during nICT treatment. The immune landscape displayed heterogeneity between pre- and post-treatment samples. Recognizing the importance of preserving information of TME dynamics during immunochemotherapy, we followed the reviewer's recommendation by differentiating single-cell profiles of pre-treatment biopsy specimens from post-treatment surgery samples. For better clarity and conciseness, we provided a detailed annotation in the revised text and figures about the timepoint and efficacy group of samples that each analysis based on.

- Figure 2e: please explain how the mast cell scoring was done: was this done by a pathologist? Also, please explain the level of response in this group as in the scRNAseq cohort an pCR, MPR and non-MPR scoring system was used.

Author response: Thank you for your valuable comment. In the initial manuscript, mast cell scoring was conducted via IHC staining, involving the assessment of average mast cell number across six randomly selected high-power fields (×400), as counted by two experienced pathologists. Following an expansion of the single-cell analysis to 23 specimens, we observed that the difference in mast cell proportions between responders and non-responders was no longer significant. Hence, the results of mast cells have been removed from the revised manuscript.

In evaluating the pathologic response of tumor specimens from ESCC patients receiving nICT treatment, we adhered to pan-tumor immune-related pathologic response criteria. We conducted H & E staining on post-treatment surgical resection specimens at primary tumor site. Then we calculated percent residual viable tumor (RVT) by the formula: RVT surface area/total tumor bed surface area * 100. The total tumor bed surface area included RVT, tumor-associated stroma, necrosis, and the regression bed where viable tumor had been present before treatment but was no longer. The %RVT for each sample was evaluated by two experienced pathologists blinded to the study design. Post-treatment specimens with %RVT = 0%, 0% < %RVT ≤ 10% and %RVT > 10% were defined as pCR, MPR and non-MPR, respectively. Responders were identified as patients with pCR or MPR following nICT, while non-responders as those with non-MPR.

- I do not agree with the statement that post treatment TLS presence indicates that TLSs might mediate a favorable response to immunotherapy in ESCC as this can only be done when you know something about pretreatment presence. Please revise this part of the manuscript.

Author response: We thank the reviewer for pointing out this imprecise statement in our article. Tertiary lymphoid structures (TLSs) were reported to be closely associated with a favorable prognosis and improved immunotherapy efficacy in multiple cancers. Researchers have confirmed the correlation between TLSs and prolonged survival in treatment naïve ESCC, while the relevance of TLSs to ESCC immunotherapy efficacy remained unsettled. In the present study, we validated a significantly higher TLS abundance in post-treatment surgical resection specimens of responders and a positive

association between baseline CXCL13+CD8+ T cells infiltrating level and post-treatment TLS abundance. We have provided more evidences for the potential role of TLSs during ESCC immunotherapy and revised the content about TLSs in the manuscript.

- In general: please describe in more detail what was found about the matched tumor samples. What happens within the patient? How do immune features change during treatment and how does this correlate with response?

Author response: Thank you for your question. In addition to the single-cell profile of 5 pairs of matched pre-treatment biopsy samples and post-treatment surgery samples, we also retrospectively extracted 54 pairs of matched pre- and post-treatment FFPE samples from ESCC patients receiving nICT treatment following surgical resection from our hospital as an independent external validation cohort, and conducted RNA-seq analysis. Through the combination of scRNA-seq and RNA-seq, we further reported several interesting findings. In non-responders, post-treatment samples were marked by significantly increased infiltration of $CD8^+$ Tex-CXCL13 and $CD4^+$ Treg-TNFRSF4 cells compared with matched baseline biopsy samples. While in responders, a contrary tendency was observed in the distribution of $CD4^+$ Treg-TNFRSF4 cells, and no prominent difference in the infiltration level of $CD8^+$ Tex-CXCL13 cells was found between pre- and post-treatment samples. These findings indicated that the immune landscape of tumors with resistance to immunochemotherapy was characterized by terminal exhaustion and high tolerance, which might be the potential mechanism of therapeutic resistance and promising target to enhance efficacy. Moreover, post-treatment samples from responders revealed largely decreased infiltrating abundance of tCAF-MMP11 and Macro-SPP1 cells versus matched baseline biopsy samples, different from non-responders. Interestingly, both tCAF-MMP11 and Macro-SPP1 cells were found to positively correlate with the cell proportion of $CD4^+$ Treg-TNFRSF4 cells, and potentially recruit and activate $CD4^+$ Treg-TNFRSF4 cells through ligand-receptor mediated cell-cell communication network. These findings have been thoroughly described and discussed in the revised manuscript.

- Please explain how CXCL13, as a marker of dysfunctional CD8 T cells, can be a marker of a good response to immunotherapy? Can it be that patients with high CXCL13 levels just had higher CD8 levels and are just more inflamed tumors?

Author response: We appreciate your professional review. When progressing along with the (pre)dysfunctional axis driven by antigen recognition in TME, $CD8^+$ T cells acquire the capacity to express and secrete CXCL13. $CXCL13^+CD8^+$ T cells from baseline treatment-naïve tumors are considered as tumor-reactive $CD8^+$ Tex with high reprogrammability to respond to immune checkpoint blockade. Importantly, CXCL13 is also a well-established B cell attractant. CXCL13-CXCR5 chemokine axis is essential for organizing B cell follicles in TLSs. CXCL13-dependent TLS formation has been reported as critical for effective immunotherapy in solid tumors. Consistently, in current study, pre-treatment tumors with a higher infiltration score of $CD8^+$ Tex-CXCL13 showed improved therapeutic response to immunochemotherapy. We further validated the positive association between pre-treatment infiltrating level

of $CXCL13^+CD8^+$ T cells and post-treatment TLS abundance, and revealed CXCL13 to potentiate anti-PD-1 efficacy in vivo. Taken together, the enrichment of $CD8^+$ Tex-CXCL13 cells in baseline tumors was a predictor for favorable therapeutic response to immunochemotherapy. We have thoroughly described and discussed these findings in the revised manuscript, in hope for helping understand the role of $CD8^+$ Tex-CXCL13 cells during nICT treatment for ESCC.

Discussion:

- Please change immunotherapy for chemo-immunotherapy combinations throughout the discussion.

Author response: Thank you for the valuable feedback. All ESCC patients enrolled in this study underwent neoadjuvant immunochemotherapy followed by surgical resection. As advised, we have replaced the wording 'immunotherapy' with 'immunochemotherapy' throughout the revised manuscript.

- Please add pre-treatment to line 530 to indicate in which clinical setting the mast cells were measured.

Author response: Thanks for your thoughtful reminder. We have carefully revised the manuscript, and provided an accurate clinical setting for each analysis to avoid confusion for readers.

- Line 552: the authors assume that pretreatment TLS presence is association with response to immunotherapy while

Author response: Thank you for your valuable questions and suggestions, which have improved the depth and clarity of our findings. Studies have reported the favorable role of TLS in antitumor immune response and the underlying predictive value of TLS for immunotherapy responses in several kinds of solid tumors. Yet in lack of enough pre-treatment tumor samples for exploring the role of TLS in immunotherapy retrospectively or prospectively, there is currently no study validating the hypothesis that baseline TLS presence enhances immunotherapy efficacy. In current study, we found a significantly higher infiltrating level of $CD8^+$ Tex-CXCL13 cells in pre-treatment tumors of responders, which was also positively associated with TLS abundance in post-treatment samples. Despite baseline biopsy samples could hardly reflect the TLS abundance before neoadjuvant therapy. We highlighted the favorable role of enrichment of $CD8^+$ Tex-CXCL13 cells in pre-treatment samples and provided more evidences for the underlying mechanism that CXCL13 enhancing immunotherapy efficacy probably through promoting TLS formation. And we have revised our discussion on TLS accordingly.

Reviewer #3, expertise in scRNAseq/scTCRseq (Remarks to the Author):

The paper "Single-cell analyses reveal the cellular and molecular dynamics of esophageal squamous cell carcinoma during neoadjuvant immunotherapy" by Yang et al reports a detailed analysis of single cell data of ESCC patients receiving neoadjuvant immunotherapy, aiming at the identification of the cellular populations that mostly influence the response to therapy.

The paper is very straightforward, high-quality data and analysis are presented in a clear way. However, I found some criticalities that undermine the conclusions.

First of all, some samples (p1 pre, p1 post, p2 pre and p8 post) show an extremely high percentage of epithelial cells. Given that all the analysis are performed on compositions, I wonder how those samples may impact results, also given that one of such samples is in the NR class which, in turn, is the smallest class of the entire dataset. Removing epithelial cells (but also muscle and neurons) will greatly shift the percentages.

Author response: Thank you for your valuable question. The comparison analysis of scRNA-seq data in the first draft used the number of all cells as the denominator to calculate various cell components. But this approach led to perturbations, primarily driven by the diverse and sample-dependent nature of the major cell components present at primary tumor site itself and the sampling process. In the revised manuscript, the analysis of immune cells was only based on CD45+ cells and was independent of epithelial cells, which limited the influence of heterogeneity in epithelial cell proportions. Additionally, as shown in the study overview of Figure 1a, we have enrolled another 8 samples from non-responders (NR) and performed scRNA-seq paired with scTCR-seq analysis. We also retrospectively extracted 54 pairs of matched pre- and post-treatment FFPE samples from ESCC patients receiving neoadjuvant immunochemotherapy from our center as an independent validation cohort, and performed RNA-seq analysis. The important findings revealed by single-cell analysis were further validated by transcriptomic analysis of a much larger validation cohort.

One major claim of the paper is that some populations (Mast, B and CD8 cells) are enriched in responders when compared to non-responders and that such changes may be at the basis of the differential response to therapy.

Nevertheless, analysis of compositional datasets should not be performed comparing percentages of a single populations across conditions, without any regard of the whole composition. For example, the lower amount of mast cells (Figure S2d) may be due to the lower amount overall of blood cells in NR, given that P8 is characterized by ~90% epithelial cells and the remaining NR samples come only from a single patient (P7pre and post).

I tried to retrieve the original cell counts for each population (given the reported percentages), removed all non-blood cells and performed compositional data analysis using scCODA (doi: 10.1038/s41467-021-27150-6). Using the default Bayesian model I could not find any credible change between R and NR:

times_credible pct_credible is_credible
Bfoc 0 0.0 False
Blood_Endothelial_cells 0 0.0 False
Bmem 1 0.059 False
Bn 1 0.059 False
CD4_Treg 0 0.0 False
CD8_Teff 0 0.0 False
CD8_Tem 0 0.0 False
CD8_Tex 0 0.0 False
Dendritic_cells 0 0.0 False

GCB 1 0.059 False
Lymphatic_Endothelial_cells 0 0.0 False
Macrophages 0 0.0 False
Mast_cells 0 0.0 False
Other_CD4_T 0 0.0 False
Other_CD8_T 1 0.059 False
Plasma_B 0 0.0 False
Tn 0 0.0 False

Author response: We sincerely thank the reviewer for the valuable feedback that helped us improve the quality of our manuscript. As mentioned above, cell proportion of various immune cell clusters was calculated on the basis of CD45+ cells and corresponding immune cell types, instead of all cells as a whole. For example, when exploring the infiltrating level of CD4+ Treg-TNFRSF4 cells between samples of responders and non-responders, we first analyzed the fraction of CD4+ T cells in CD45+ cells, and then calculated the proportion of CD4+ Treg-TNFRSF4 cells in CD4+ T cells, which reduced the influence of the whole composition to an acceptable level. Moreover, given the great heterogeneity between pre- and post-treatment samples and the TME dynamics during immunotherapy, scRNA-seq data and analysis based on that of post-treatment samples were separated from pre-treatment samples. After expanding the number of samples performed with single-cell sequencing to 23 cases, we performed compositional data analysis by removing all non-immune cells. This revealed several cell populations as immune features for immunotherapy resistance. Importantly, we identified the top 30 DEGs for each cell cluster—ranked by log-transformed ratios of normalized gene expression relative to other clusters within corresponding major immune compartments (log2_avg)—as corresponding cell type specific gene signatures. Then, we calculated the infiltration level of distinct cell clusters by ssGSEA using RNA-seq data of the validation cohort to validate the former findings.

I had the same conclusions using additive logratio (ALR) transform of the data paired with a t-test. I think authors must properly perform compositional data analysis on these data, also including the pre/post covariate (I find strange to weight samples before and after treatment in in the same way) before drawing any conclusion. Any analysis will suffer from a unbalanced representation of R/NR, it would be beneficial to increase the number of NR samples (although I understand it may be difficult).

Lastly, I would point that even if the analytical procedure was correct, it is not clear which statistical test is performed throughout the text: it is written in the methods that t-test or Wilcoxon is used, but this is not reported alongside each comparison. This is important as different tests give different results and t-test may not be the appropriate test when comparing percentages. Wilcoxon may be in fact more appropriate (and more powerful, as percentages are possibly not normally distributed) but results in not significant tests, for example using the Bmem percentages in table S2:

> R <- c(0.983606557, 0.617723925, 0.858737656, 0.924855491, 3.646833013, 1.144640999, 6.115107914, 0.021463833, 16.15798923, 18.21862348, 8.230655495, 10.86172345)

> NR <- c(1.179245283, 0.127145582, 0.30647986)
> wilcox.test(R, NR)
Wilcoxon rank sum exact test
data: R and NR
W = 28, p-value = 0.1802
alternative hypothesis: true location shift is not equal to 0

Author response: We appreciate the reviewer's insightful comments, and have revised the manuscript accordingly. We collected another 8 samples from ESCC patients receiving neoadjuvant immunochemotherapy and performed scRNA-seq paired with scTCR-seq analysis. In response to the reviewer's suggestion that Wilcoxon test is more appropriate for the significance test in analysis of cell proportions using acRNA-seq data, we have corrected the significance test method in the revised manuscript and applied Wilcoxon test to perform compositional data analysis. In addition, we have carefully revised the manuscript, and provided an accurate annotation for the statistical test method using in each analysis.

Responds to the editor:
Thanks again to both reviewers and editor for the thoughtful and thorough review. We have carefully revised our manuscript. Hopefully we have addressed all of your concerns. The revised content in our manuscript were marked as light blue.
If you have any question about this paper, please don't hesitate to let us known.
Kind regards,
Shugeng Gao
Department of Thoracic Surgery
National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences
Tel: (+86) 010-87787155
E-mail: gaoshugeng@cicams.ac.cn

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
In this revision, Zhenlin Yang, He Tian, Xiaowei Chen, and Bozhao Li et. al. performed single-cell data analysis at distinct treatment timepoints and efficacy groups. In addition, to compensate only one paired samples from non-responders, they collected an additional FFPE matched pre- and post-treatment samples to confirm their claims. Some of this reviewer's concerns are resolved; however, the manuscript lack clarity and statistical tests are not correctly performed.

1) Because the data distributions of terminally exhaustion score of CD8_Tex_STMN1 in Post-R and Post-NR groups appear bimodal (Figure 2e), a t-test may not be appropriate for the statistical test. A non-parametric test like the Mann-Whitney U test is recommended in this case. Similar issues were seen in Figures 2g, 2i, 2l, 5g, 5h, 5l, 5n, and 6i.

2) In line 76, is "nICT" correct, or should it be "nCRT"?

3) In line 127, is "visable" correct, or should it be "visible"?

4) In line 374, there is a incorrect expression "bar charts show applied to showed the percentage of …".

5) According to Fig. 1i, there are 5 clusters in B cells, 9 clusters in macrophages, and 10 clusters in T cells. However, the number is discordant with lines 376-379. Could the aurhors clarify this discrepancy?

6) In lines 451-453, the authors stated an interaction between CD8+ Tex-CXCL13 and B cells expressing CXCR5. However, according to Supplementary Fig. 4j, this reviewer cannot ascertain which fraction of B cells expresses CXCR5.

7) In Fig. 3d, the effect of CXCL13 on anti-PD-1 therapy was evaluated. However, the authors compared the combinational therapy with vehicle and monotherapy groups. This is not relevant. This reviewer recommends comparing the combinational therapy with anti-PD-1 monotherapy to confirm the additional effect of CXCL13 on anti-PD-1 therapy.

8) In lines 536-539, the authors stated the significance of the transition from Texp to terminal Tex for improved efficacy of immunotherapy. However, according to Figure 4m, the proportion of Texp was still higher than non-responders at post-treatment timing. There is a possibility that not increase of terminal Tex but sustained fraction of Texp is crucial for the improved therapy response.

9) In the Data Availability section, the authors have deposited data in GSA for Human. This reviewer visited the site and searched HRA003591(https://ngdc.cncb.ac.cn/search/specific?db=hra&q=HRA003591). Then, a link was shown to access the data (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA003591). However, when trying to access the link, the server returned "ERR_EMPTY_RESPONSE". Additionary, there seems no description on restriction of data usage. Please clarify this.

10) The word counts of the abstract and main text are over the format for articles of Nature Communications.

Reviewer #2:
Remarks to the Author:

I kindly thank the authors for the updated version of the manuscript. The authors answered all questions. The extra patient series and analyses made a clear improvement of the manscript and I think the data are highly valuable for a larger audience of translational scientist in the IO field. I have no further comments.

Reviewer #3:
Remarks to the Author:
The revised paper has changed a lot during the revision. I surely appreciate the extension of the dataset to other individuals and the inclusion of retrospective validation on FFPE samples. I'm a bit confused (not necessarily in a negative way) by the large amount of text that has been changed, possibly shifting the scope of the paper.
In their answer, the authors say that they solved the problem of compositional data analysis focusing on immune compartment only.
Disregarding some cell types doesn't change the nature of the compositional problem. Moreover, a considerable part of the paper is concerning the changes in CAF. I still would have applied a proper CODA framework to these data, especially given the high variability (Supp fig 2C). This applies to FFPE data as well.

Some minor concerns:
- It is not clear if GO enrichment has been performed with or without background. If no background has been used, some of the GO terms more related to the cells of interest could be overscored. The analysis of immune cells is a typical case in which a proper background (e.g. the set of genes actually analyzed in the experiment and/or the set of genes normally expressed in the immune compartment)
- The size of the FFPE set of samples is not clear. Depending on the section it is either 96 pairs (line 135), 96 specimens (line 291) or 54 paired samples (line 252, 289)
- I'm not convinced by the need to include some analysis, such as RNA velocity and also CNV analysis. As none of them is further discussed, they don't add much value to the paper.
- I think line 366 ("using dimension reduction, unsupervised clustering, and cell cluster annotation") can safely be removed from the text, as long as it is in the methods.
- line 127: correct "visable" to "visible"

Response to the reviewer's comments:
Reviewer #1 (Remarks to the Author):

In this revision, Zhenlin Yang, He Tian, Xiaowei Chen, and Bozhao Li et. al. performed single-cell data analysis at distinct treatment timepoints and efficacy groups. In addition, to compensate only one paired sample from non-responders, they collected an additional FFPE matched pre- and post-treatment samples to confirm their claims. Some of this reviewer's concerns are resolved; however, the manuscript lack clarity and statistical tests are not correctly performed.

1) Because the data distributions of terminally exhaustion score of CD8_Tex_STMN1 in Post-R and Post-NR groups appear bimodal (Figure 2e), a t-test may not be appropriate for the statistical test. A non-parametric test like the Mann-Whitney U test is recommended in this case. Similar issues were seen in Figures 2g, 2i, 2l, 5g, 5h, 5l, 5n, and 6i.
Author response: Thank you for your valuable comments. According to the reviewer's recommendation, for the data with bimodal distribution, Mann-Whitney U test was applied for the statistical test. Those figures, including Figure 2e, 2g, 2i, 2l, S3e, S3f, 5g, 5h, and 6i, have been revised using the above non-parametric test.
2) In line 76, is "nICT" correct, or should it be "nCRT"?
Author response: Thank you for your kind reminder. The CROSS trial and the NEOCRTEC5010 trial report that more than 40% of ESCC patients achieve pathological complete response after neoadjuvant chemoradiotherapy (nCRT). We have corrected the clerical error.
3) In line 127, is "visable" correct, or should it be "visible"?
Author response: Thank you for your careful review and kind reminder. We have corrected this clerical error in the revised manuscript.
4) In line 374, there is an incorrect expression "bar charts show applied to showed the percentage of …".
Author response: Thank you for your thorough review and valuable feedback. We have rewritten the sentence as followed "bar charts were applied to show the percentage of each cell type across different samples, timepoints and efficacy groups".
5) According to Fig. 1i, there are 5 clusters in B cells, 9 clusters in macrophages, and 10 clusters in T cells. However, the number is discordant with lines 376-379. Could the authors clarify this discrepancy?
Author response: Thanks so much for the reviewer's valuable feedback. In current study, a total of 11 clusters in T cells, 6 clusters in B cells, 12 clusters in fibroblasts, 11 clusters in myeloid cells (including eight clusters of macrophages, two clusters of dendritic cells and one cluster of mast cells), 6 clusters in endothelial cells and 2 clusters in epithelial cells were identified. The Figure li mentioned above was a misuse of results from the intermediate analysis. We have redrawn a novel Figure li on the basis of the final results. Hopefully we have clarified the reason of this discrepancy and addressed your concern.

6) In lines 451-453, the authors stated an interaction between $CD8^+$ Tex-CXCL13 and B cells expressing CXCR5. However, according to Supplementary Fig. 4j, this reviewer cannot ascertain which fraction of B cells expresses CXCR5.

Author response: We highly acknowledge the reviewer's recommendation. We have studied the distribution of CXCR5 in six clusters of B cells. As shown in Supplementary Figure 4k, CXCR5 was highly expressed by Bmem-TNFRSF13B and Bfoc-TCL1A, which were also important components of tertiary lymphoid structures. Consistently, subsequent CellChat analysis also revealed an obviously high cell-cell interaction potential between $CD8^+$ Tex-CXCL13 and these two clusters of B cells.

7) In Fig. 3d, the effect of CXCL13 on anti-PD-1 therapy was evaluated. However, the authors compared the combinational therapy with vehicle and monotherapy groups. This is not relevant. This reviewer recommends comparing the combinational therapy with anti-PD-1 monotherapy to confirm the additional effect of CXCL13 on anti-PD-1 therapy.

Author response: We sincerely thank the reviewer for the valuable feedback. In the present study, mouse ESCC cell line mEC25 was applied to build the animal model. Seven days after tumor implantation, mice were divided into four groups at random and received distinct treatment strategies. Saline control, rmCXCL13, anti-murine PD-1, rmCXCL13 plus anti-murine PD-1 were injected intraperitoneally for four groups of mice, respectively. After eight injections, the tumor volume and weight of the combinational treatment group with rmCXCL13 plus anti-murine PD-1 were significantly lower than the monotherapy groups that received rmCXCL13 or anti-murine PD-1 alone (Figure 3c, 3d). As the reviewer recommended, we compared the therapeutic efficacy of combinational therapy with anti-PD-1 monotherapy and found a significantly enhanced anti-tumor efficacy of PD-1 blockade after the combination of rmCXCL13 (Figure 3c, 3d), further validating the favorable role of CXCL13 during immunotherapy.

8) In lines 536-539, the authors stated the significance of the transition from Texp to terminal Tex for improved efficacy of immunotherapy. However, according to Figure 4m, the proportion of Texp was still higher than non-responders at post-treatment timing. There is a possibility that not increase of terminal Tex but sustained fraction of Texp is crucial for the improved therapy response.

Author response: The reviewer's question and suggestion are very valuable to our study. We highly agreed with the reviewer's point of view. Previous studies highlighted that precursor tumor-reactive $CD8^+$ T cells were likely to be the major cell population responding to immunotherapy. Although RNA velocity analysis on Tex-relevant cells, which could also be considered as tumor-reactive $CD8^+$ T cells, revealed a potential transition route from Texp to terminal Tex. In responders, along with the sustained antigen presentation, there would be overactivation of Texp to some extent and thus resulting in a transition from Texp to terminal Tex, indicating that a light increase of terminal Tex fraction was properly the common response during anti-tumor process. Importantly, we also found that the proportion of Texp cells in non-responder samples was lower than responders at both pre- and post-treatment timepoints, indicating that the sustained fraction of Texp might be crucial for the improved therapy response to

immune checkpoint blockade. Herein, we revised the discussion about Texp and further highlighted the potential favorable role of the sustained high infiltration of Texp during immunotherapy.

9) In the Data Availability section, the authors have deposited data in GSA for Human. This reviewer visited the site and searched HRA003591(https://ngdc.cncb.ac.cn/search/specific?db=hra&q=HRA003591). Then, a link was shown to access the data (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA003591). However, when trying to access the link, the server returned "ERR_EMPTY_RESPONSE". Additionally, there seems no description on restriction of data usage. Please clarify this.

Author response: We have considered the reviewer's question carefully. The raw data of single-cell RNA sequencing and TCR sequencing for 23 samples enrolled in our study is now with controlled access for the public in dataset HRA003591. Researchers can get access to the data after passing the Data Access Committee. For the reviewers of our manuscript, we generate a shared URL (https://ngdc.cncb.ac.cn/gsa-human/s/lu3My4v0) that links to the data directly. Hopefully we have addressed your concern about the Data Availability section.

10) The word counts of the abstract and main text are over the format for articles of Nature Communications.

Author response: Thank you for highlighting this issue. As the reviewer recommended, we have revised the abstract and main text of the present manuscript to make it brief, and further reduced the word counts. We hope that the improvements will meet with approval.

Reviewer #2 (Remarks to the Author):

I kindly thank the authors for the updated version of the manuscript. The authors answered all questions. The extra patient series and analyses made a clear improvement of the manuscript and I think the data are highly valuable for a larger audience of translational scientist in the IO field. I have no further comments.

Author response: The reviewer thoroughly reviewed the first version of our manuscript and provided professional comments on the content of neoadjuvant immunotherapy from a clinical standpoint. After careful consideration of the reviewers' comments, we revised our manuscript accordingly. We thank the reviewer very much for the constructive comments and suggestions which help a lot to improve the quality of our manuscript.

Reviewer #3 (Remarks to the Author):

The revised paper has changed a lot during the revision. I surely appreciate the extension of the dataset to other individuals and the inclusion of retrospective validation on FFPE samples. I'm a bit confused (not necessarily in a negative way) by the large amount of text that has been changed, possibly shifting the scope of the paper.

In their answer, the authors say that they solved the problem of compositional data analysis focusing on immune compartment only. Disregarding some cell types doesn't

change the nature of the compositional problem. Moreover, a considerable part of the paper is concerning the changes in CAF. I still would have applied a proper CODA framework to these data, especially given the high variability (Supp fig 2C). This applies to FFPE data as well.

Author response: We have carefully considered the comments. According to the reviewer's recommendation, we applied scCODA to detect the compositional changes of various cell types in this study, which was a Bayesian model and could address these issues associated with high variability of samples, thus enabling the study of complex cell type effects in disease. As shown in Supplementary Figure 8f, we used bEndo_NEAT1 as the reference and detected the compositional changes of various cell types between post-treatment samples from responders and non-responders. Consistently, the results of scCODA also revealed a significantly higher infiltration level of $CD4^+$ Treg-TNFRSF4 and tCAF-MMP11 in post-treatment samples of non-responders compared to that of responders. Nevertheless, the infiltration level of various cells derived from the FFPE RNAseq data was calculated by ssGSEA, measured by enrichment score instead of compositional percentage, which might not be suitable for the usage of scCODA algorithm. The FFPE RNAseq data is composed of 54 pairs of pre- and post-treatment samples, which is currently the largest ESCC cohort with neoadjuvant immunotherapy enrolled in transcriptomic sequencing. Moreover, the Mann-Whitney U test was applied to detect the significance of differential infiltration score of various cells between two efficacy groups.

Some minor concerns:

- It is not clear if GO enrichment has been performed with or without background. If no background has been used, some of the GO terms more related to the cells of interest could be overscored. The analysis of immune cells is a typical case in which a proper background (e.g. the set of genes actually analyzed in the experiment and/or the set of genes normally expressed in the immune compartment)

Author response: Thank you for your valuable comments. We highly agreed with the reviewer's point of view. All GO enrichment analyses in the current study were performed with the background that tumors samples were derived from ESCC patients receiving neoadjuvant immunochemotherapy. Cell clusters that closely correlated with immunotherapy efficacy were selected for subsequent GO enrichment analyses. Under the background of distinct therapeutic response to immunochemotherapy, those biological processes enriched in above selected cell clusters will be more prominent in GO enrichment analyses. Detailed description of each GO enrichment analysis has been provided in the part of figure legends and results.

- The size of the FFPE set of samples is not clear. Depending on the section it is either 96 pairs (line 135), 96 specimens (line 291) or 54 paired samples (line 252, 289)

Author response: We sincerely appreciate your insightful question. In the present study, we retrospectively obtained 96 pairs of matched pre- and post-treatment FFPE tumor tissues from 96 ESCC patients who underwent neoadjuvant immunochemotherapy followed by surgery in our center. Among these samples, 96 pre-treatment FFPE biopsy

samples were conducted with immunofluorescence staining to detect the distribution of $CD8^+CXCL13^+$Tex, and 96 matched post-treatment FFPE surgical samples were performed with H & E staining to evaluate the pathological response rate and the abundance of tertiary lymphoid structure. For validation, based on the above retrospective extracted ESCC cohort, we screened out 54 patients with paired high-quality pre- and post-treatment FFPE sections (DV200 > 30%), and performed RNA sequencing. The RNA-seq data of paired pre- and post-treatment FFPE samples from these 54 patients was applied as an external independent validation cohort, named nICT-ESCC, to verify these findings derived from single-cell analysis. The detailed information of these matched pre- and post-treatment FFPE tumor tissues from 96 ESCC patients was displayed in Supplementary Data 5. Moreover, in the revised manuscript, we have described the size of samples enrolled in each analysis in detail to avoid confusion.

- I'm not convinced by the need to include some analysis, such as RNA velocity and also CNV analysis. As none of them is further discussed, they don't add much value to the paper.

Author response: The reviewer's question and suggestion are very valuable to our study. In this study, the results of RNA velocity and CNV analysis were used as additional validation for our main findings. The RNA velocity analysis for Tex-relevant cells revealed a potential transition route from Texp to terminal Tex, which further validate the findings in the TCR analysis. Through the CNV analysis, we divided epithelial cells into malignant and non-malignant groups, and revealed SPP1 to be highly expressed in malignant epithelial cells that mainly residue in post-treatment samples of non-responders, which again indicating SPP1 to be a potential therapeutic target to overcome immune suppression and enhance the response of ESCC tumors to immunotherapies. According to the reviewer's comments, we have systematically discussed the results of these analyses in the revised manuscript.

- I think line 366 ("using dimension reduction, unsupervised clustering, and cell cluster annotation") can safely be removed from the text, as long as it is in the methods.

Author response: Thanks so much for the reviewer's recommendation. The scRNA-seq raw data of 23 ESCC samples were processed using dimension reduction, unsupervised clustering, and cell cluster annotation. The process was described in detail in the part of methods. As the reviewer recommended, we have removed the description about data processing from the part of results.

- line 127: correct "visable" to "visible"

Author response: Thank you for the detailed review. We have corrected this clerical error in the revised manuscript.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The authors have addressed this reviewer’s remaining points and revised their manuscript accordingly. This study sounds well for readers of translational researchers.

Reviewer #3:
Remarks to the Author:
In this revised version, the authors addressed all my points. In particular, they applied a proper framework for compositional data analysis to evaluate the differences in cell proportions in R/NR patients. I have a further minor concern on this: the supplementary figure added to the manuscript is never mentioned nor discussed in the paper. Similarly, CODA is not described in the methods section. Compositional data analysis surely confirms some important changes previously highlighted (such as CD4+ Treg-TNFRSF4 and tCAF-MMP11) but suggests that some changes that have been commented (such as Macro-SPP1, Macro-EREG and Macro-STMN1 to name a few) may be due to some sampling effect.
Also, it would be appropriate to harmonize plots ordering R/NR consistently across the paper to ease readability (in S8F is NR/R, in other figures such as 5C is R/NR). More importantly, the y-axis scale apparently don't match (for example, comparing 5C and S8F for tCAF-MMP1 the boxplots vary in completely different ranges). Also, I wonder whether the y-axis label in figure S8F is correct: if the chart reports percentages (as stated, "Proportion (%)") 0.20 equals 0.2%.

Response to the reviewer's comments:

Reviewer #1 (Remarks to the Author):
The authors have addressed this reviewer's remaining points and revised their manuscript accordingly. This study sounds well for readers of translational researchers.
Author response: We sincerely thank the reviewer for the valuable feedback. The reviewer thoroughly reviewed the previous two versions of our manuscript and provided professional comments which help a lot to improve the quality of our manuscript.

Reviewer #3 (Remarks to the Author):
In this revised version, the authors addressed all my points. In particular, they applied a proper framework for compositional data analysis to evaluate the differences in cell proportions in R/NR patients. I have a further minor concern on this: the supplementary figure added to the manuscript is never mentioned nor discussed in the paper. Similarly, CODA is not described in the methods section. Compositional data analysis surely confirms some important changes previously highlighted (such as $CD4^+$ Treg-TNFRSF4 and tCAF-MMP11) but suggests that some changes that have been commented (such as Macro-SPP1, Macro-EREG and Macro-STMN1 to name a few) may be due to some sampling effect.
Author response: We highly acknowledge the reviewer's recommendation. According to the reviewer's comments, we have revised the manuscript, and mentioned and discussed each supplementary figure in the appropriate paragraph. In addition, a detailed description about scCODA used for compositional data analysis has been added in the methods section. As shown in the Supplementary Fig. 8f, $CD4^+$ Treg-TNFRSF4 and tCAF-MMP11 were found to remarkably increase in post-treatment samples from non-responders. While the infiltrating proportion of Macro-SPP1, Macro-EREG and Macro-STMN1 showed a relatively higher trend in post-treatment samples from non-responders but without statistical significance, indicating these differences to be potentially caused by some sampling effects. The results of scCODA analysis have also been fully discussed in the revised manuscript. Future single-cell analysis based on larger cohort of ESCC patients with neoadjuvant immunotherapy is needed to verify our findings in current study.
Also, it would be appropriate to harmonize plots ordering R/NR consistently across the paper to ease readability (in S8F is NR/R, in other figures such as 5C is R/NR). More importantly, the y-axis scale apparently doesn't match (for example, comparing 5C and S8F for tCAF-MMP1 the boxplots vary in completely different ranges). Also, I wonder whether the y-axis label in figure S8F is correct: if the chart reports percentages (as stated, "Proportion (%)") 0.20 equals 0.2%.
Author response: Thank you for your careful review and kind reminder. According to the reviewer's recommendation, we have revised the R/NR order in Figure S8F and checked all relevant figures to make sure the order of R/NR is consistent across the paper. Another concern of the reviewer is the seemingly mismatched y-axis scale in different figures. In Figure 5C, the y-axis represents the proportion of distinct CAF

clusters in all CAFs. While in Figure S8F, the y-axis represents the proportion of distinct cell clusters in all cells. The denominators used for calculating these two kinds of cell proportions are not the same. Herein, to avoid confusions in readers, we have provided a more detailed description for each figure used for showing cell proportion. Once again, thanks to the reviewer for the thoughtful and thorough review. We hope that the corrections will meet with approval.