nature portfolio

Peer Review File

Cellular reprogramming during anti-PD-1 and chemotherapy treatment in early-stage primary hormone receptor-positive breast cancer

Corresponding Author: Dr Eliezer Van Allen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors present the results of some genomic data analyses from a phase II neoadjuvant breast cancer clinical trial investigating PD-L1 expression in patients with HR-positive disease after the combination of neoadjuvant chemotherapy and immunotherapy (pembrolizumab). Leveraging tumor biopsies obtained at various occasions, including pre-treatment, beginning of combination treatments (week 3), in the middle of neoadjuvant treatments (week 7), after the completion of combination treatment, and at surgery, the authors performed single-nuclei multiome sequencing and attempted to characterize the assembled cells or nuclei. Several gene expression metaprograms (MPs) were identified through analyzing expression data from the malignant cells, CD8+ T cells, and macrophages. How these MPs were associated with treatment response (RCB 0/1 vs 2/3) was explored. Some of the findings were related to previous observations or validated via in TCGA HR+ breast cancer patients. Although the snRNA-seq analyses from a limited number of 40 tumor biopsy samples were pretty broad and provided some interesting biologic interpretation around those MPs, there are a couple of critical issues that need to be addressed.

A. The description on how the authors ended up with those 20 patients is misleading. The pilot study (NCT02999477) was designed to enroll 50 patients but the accrual was probably stopped after enrolling 32 patients. The authors need to clearly outline how they ended up with the 40 tumor biopsies from 20 patients. Because the numbers of included tumor biopsies are 12, 9, 8, 4, and 7 at those five occasions, respectively, whether or not those samples were representative of the study population at the corresponding occasions is important. Otherwise, the difference in the gene expressions between favorable responders and unfavorable responders, or between two occasions may very well be due to other factors rather than the response status or different occasions, respectively.

B. From the snRNA-seq data of 40 tumor biopsies from 20 patients, the authors detected about 250K nuclei. For each sequencing batch, “2,000 highly variable genes were identified and used for the principal component analysis (PCA). The top 50 principal components (PCs) were then used to compute the nearest neighbors distance matrix,…”. What’s the size of each batch and how were the batches set up? Presumably the selected top 50 PCs were very likely different by batches, how was that factored in the cluster analysis? When the authors started working on the selection of MPs, how was the gene expression matrix constructed in light of the batch factor and the above-mentioned procedure?

(Remarks on code availability)
1) Page 3. “specific mechanisms between tumor and microenvironmental cells that may result in therapeutic response to ICI plus chemotherapy in breast cancer are incompletely understood.” Suggest using “not well understood” rather than “incompletely”.
2) Single-nucleus transcriptome and chromatin accessibility profiling before and during monotherapy and combination therapy.
• The rationale for selecting RCB 0/1 instead of pCR (= RCB 0) is not adequate.
• “No cluster bias was evident when assessing pretreatment, monotherapy, and on-combination therapy biopsies within or between treatment arms (Fig. 1D; Supplementary Fig. 1B)” Not sure what this statement was based on. Did it mean no obvious patterns were observed in those plots?

3) Tumor-intrinsic molecular programs associated with combination therapeutic
Response.
• Supplemental Figure 2C. The lower panel includes all malignant cells from those two patients. Can the sample be annotated? Or separation between samples from the same patient is not great?
• Figure 2A and Supplemental Figure 2D. Was there any reason for why patient p27 was not included?
• The legend of Figure 2A, “Programs are clustered into nine metaprograms (MPs)”. Did you mean “Genes are clustered …”? Were the top 50 genes selected because they had the highest variation?
• Figure 2C. Whether the baseline samples were representative of the favorable responses and the unfavorable responses affects the validity of the comparisons.

4) CD8+ T cell states responsive to combination therapy
• All the samples were clustered into three groups based on the three MPs. Can the clustering of different samples from the same patient be discussed so that we can make sense out of the characterization?
• The comparisons described in Figures 3D-3F seem to be exploratory by nature. The Distributions largely overlapped each other even though there was statistically significant differences. How can such information be used to guide clinical practice? This question can be raised for the analyses/figures on macrophage cells.

Reviewer #2

(Remarks to the Author)
The authors of the manuscript entitled “Cellular reprogramming during anti-PD-1 and chemotherapy treatment in early-stage primary hormone receptor-positive breast cancer” have conducted a single cell-based analysis from longitudinal tissue samples obtained from patients enrolled in a prospective study of neoadjuvant nab-paclitaxel with pembrolizumab in patients with stage II-III HR+/HER2- breast cancer. By using single-nucleus RNA and ATAC sequencing, the authors identified distinct cellular states and gene expression metaprograms which are associated with differential therapy responses. This study provides insights on (chemo)immunotherapy resistance/response mechanisms in patients with early HR+/HER2- breast cancer -a subtype considered to display lower levels immune infiltration and immunogenicity- and several interesting TME-related implications for the biology of the disease.

Major comments:
1. The authors focus on the longitudinal samples single-cell analysis of. However, presentation of the primary and key secondary results of the trial would contribute to a more complete and overall presentation of the study, rationale and implications of the translational sub-studies
2. Although the study lies on single cell-based approaches for the different analyses and TME interactions, validation of these findings in situ (using a spatial biology method) would further enhance the results of the study.

Minor comments:
1. The authors focus mostly on two metaprograms, namely the MP7 and MP11. It would be interesting to provide more results on the other MPs, i.e. interferon, cell cycle.
2. The authors could consider presenting also longitudinal changes of the cancer cell states/MP, since they present longitudinal results for T-cells and macrophages
3. Why did the authors choose to focus on macrophages and not in other cells playing a role in immunotherapy response (i.e. B-cells)?
4. The authors could consider using bulk RNA-seq data from the same samples, in order to correlate with the single-cell data.
5. The authors could clarify how many patients were included in the validation cohort (eribulin + pembrolizumab) with bulk RNA-seq data.
6. Lines 235-241: It seems that the last sentence in line 237 ”…while a decrease in the cytotoxic effector gene signature..” is contradictory to what it is states in the concluding sentence in line 239 “…followed by the increase of the cytotoxic effector gene signature in CD8+ T cells”
7. The authors could consider discuss their results, in relation to other window-of-opportunity trials of immunotherapy in early breast cancer
8. The authors should correct the references 3 and 4 so that they correspond to the correct journal

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The manuscript entitled “Cellular reprogramming during anti-PD-1 and chemotherapy treatment in early-stage primary hormone receptor-positive breast cancer” submitted by Fu et al, describes the impact of immune checkpoint blockade (ICB) in combination with chemotherapy on transcriptional programs of tumor cells and immune cells in early hormone receptor-positive breast cancer. The team performed single-nuclei RNA-seq and for some samples paired single-nuclei ATAC-seq on 20 patients with multiple sampling before, during, and after treatment (although not for all patients consistently) to investigate the impact of chemo-ICB combination on tumor and immune cells with the goal to identify mechanisms of ICB response/non-

response. While study design and patient selection promise insightful findings into the effect of ICB/chemo combination treatment on tumor and immune cells, the study has limitations that should be addressed.

The patients are divided into two groups: patients that first receive chemo and then ICB-chemo combination and patients that receive ICB and then combination. How does ICB alone impact immune and tumor cell abundances and phenotypes compared to a combination with chemo or pre-chemotherapy? The study only compares responders to non-responders without taking the different treatment arms into account. It would be interesting to see differences or similarities between the two different treatment regimes. Additionally, how do cell abundances and phenotypes shift over the course of the treatment when comparing the different sample time points?
It would be great if the authors would explain RECIST response. It is unclear from the text or the methods of how patients are selected to be analyzed by snRNA or multiome profiling, WES or bulk RNAseq. It is also not clear how the sampling of patients was chosen since there are 5 different time points for sampling but these are not applied to all patients. As indicated in Figure 1B some patients are represented only with one timepoint. What are the inclusion/exclusion criteria?
Overall, the study often uses visualization and statistical tests that assume single-cell-level information but are indeed comparing sample-level information (favorable or unfavorable responses are sample-level information and therefore samples need to be compared and not individual cells of all samples combined). If the pool of cells is used for the test, then the test does not have the same meaning because cells from the same sample are correlated, so a summary stats for each sample should be computed, and then a test should be run between the groups. A few examples are Fig2C (assuming the boxplot is showing the expression in individual cells and not in cells normalized by sample), Figure2D, E,F, Fig 3 C D E F, etc,
The study finds that patients with tumor cell states of high EMT-III show better responses compared to patients with tumor cells with high ER-signaling, who show worse responses. Especially, enrichment of EMT-III in favorable response patients is minimal and might not exist anymore if calculated on the sample level, instead on the single-cell level as indicated above.
Cell numbers for all samples have to be reported. Fig 3C shows the activity of IL15 in different T cells, however cell numbers seem very small. This is the only graph, where cell numbers are depicted, which raises the question of cell numbers for the other analyses.
No stromal cells are described and analyzed in the study, although stromal cells compromise the majority of some samples. I would suggest to describe the stromal part as well.
How are the findings presented in this study are different/similar to other studies? In particular, the interaction studies, how do these findings relate to other similar studies? What effects are generalizable, which are BC specific, and which are ER+ BC specific? How does chemotherapy/ICB combi impact cell phenotypes and abundances compared to ICB alone?
Do the authors have tissue blocks collected from the patients to validate key findings?
ATACseq data are only used in a small part of the analysis in regard to different tumor cell states. For the immune cells, these data are not used at all.
Overall, the computational approach seems convoluted. I am not familiar with the cell state annotation that I assume is used for e.g. fig 2D, it would be great if the authors could explain their workflow in greater detail.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
In the revision, the authors provided considerable detail on sample selection and bioinformatic procedures leading to their findings. However, the more deeply one examines these steps, the more concerns arise about the robustness and reproducibility of the resulting observations. After standard quality control, the authors retained 22,125 nuclei with high-quality data for both snRNA-seq and snATAC-seq. In their analysis of tumor-intrinsic molecular programs, they identified malignant cells and derived multiple molecular programs per sample. Each of these programs consists of a set of genes with associated importance weights.
Based on hierarchical clustering of selected molecular programs (Figure 2A), the authors identified a few "metaprograms," each comprising a collection of molecular programs, represented again by top genes and possibly weights. However, the interpretation of each metaprogram rests heavily on the specific genes included, and it is unclear how consistent or robust this clustering is across subsamples or perturbations. Furthermore, the method used to calculate per-cell signature scores for each metaprogram (Figure 2C) is not described. There are multiple plausible approaches (e.g., weighted or unweighted averages), but the omission of this detail makes the interpretation and reproducibility of these scores uncertain.
Overall, the procedure used to derive the metaprograms appears to involve a number of ad hoc decisions and fine-tuning steps, some of which may be sound, while others seem less well justified. Given this complexity and subjectivity, it is unlikely that an independent group applying the same pipeline would recover the same metaprograms. While some readers may focus primarily on the biological associations reported (e.g., with treatment response), those associations depend fundamentally on the upstream definitions of the metaprograms.
To test the stability and robustness of their method, I encourage the authors to randomly split the malignant cells from each sample into two halves (1:1) and re-derive the metaprograms using the same procedure on each subset. A comparison of the resulting metaprograms, up to the point of Figures 2A–B, would provide useful evidence for whether the identified structures are biologically meaningful or overly sensitive to data composition and procedural choices.

Some detailed comments and questions are in the following:
1) Supplemental Figure 2K. Comparing the signature scores for EMT-III between baseline and W3D1 on Pembro, the centers of those two distributions are very close and the p-value is questionable. Same goes with the last comparison in Supp. Figure 2L.
2) Page 8, lines 168-169. “Longitudinal analysis revealed that EMT-III-related MP11 levels significantly decreased during combination therapy, independent of response (Supplementary Fig. 2J-L).” The supplemental Figure 2L does not support that statement.
3) Page 8, line 189. Suspect that the authors meant Supp. Figure 2M instead of 2K.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have successfully addressed my comments and provided also additional analyses and results (based on the reviewers' suggestions) which overall improved the manuscript.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors significantly improved their manuscript during the revision process. They sufficiently addressed my concerns. They added new analysis and now are describing the used methods and workflows more transparently to improve interpretation of the results. The dataset and its analysis add value to the current understanding of the effect of ICB and chemo in breast cancer on tumor cell phenotypes, TME and resistance.

(Remarks on code availability)
I checked if the code is available, I didn't run it. Codes (python notbooks) are available for every figure and seem sufficient.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I appreciate the effort from authors to conduct extensive analyses to address previous concerns satisfactorily. Those additional analytical results will lead to more confidence in their interpretation.

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (also reviewed linked manuscript) (Remarks to the Author):

The authors present the results of some genomic data analyses from a phase II neoadjuvant breast cancer clinical trial investigating PD-L1 expression in patients with HR-positive disease after the combination of neoadjuvant chemotherapy and immunotherapy (pembrolizumab). Leveraging tumor biopsies obtained at various occasions, including pre-treatment, beginning of combination treatments (week 3), in the middle of neoadjuvant treatments (week 7), after the completion of combination treatment, and at surgery, the authors performed single-nuclei multiome sequencing and attempted to characterize the assembled cells or nuclei. Several gene expression metaprograms (MPs) were identified through analyzing expression data from the malignant cells, CD8+ T cells, and macrophages. How these MPs were associated with treatment response (RCB 0/1 vs 2/3) was explored. Some of the findings were related to previous observations or validated via TCGA HR+ breast cancer patients. Although the snRNA-seq analyses from a limited number of 40 tumor biopsy samples were pretty broad and provided some interesting biologic interpretation around those MPs, there are a couple of critical issues that need to be addressed.

A. The description on how the authors ended up with those 20 patients is misleading. The pilot study (NCT02999477) was designed to enroll 50 patients, but the accrual was probably stopped after enrolling 32 patients. The authors need to clearly outline how they ended up with the 40 tumor biopsies from 20 patients. Because the numbers of included tumor biopsies are 12, 9, 8, 4, and 7 at those five occasions, respectively, whether or not those samples were representative of the study population at the corresponding occasions is important. Otherwise, the difference in the gene expressions between favorable responders and unfavorable responders, or between two occasions may very well be due to other factors rather than the response status or different occasions, respectively.

Response: We thank the reviewer very much for their thoughtful review of our manuscript and helpful feedback for how we can improve the manuscript. We appreciate your concern about the representativeness of the tumor biopsies and regret our lack of clarity on this point. In our revised manuscript, we included a detailed explanation in the Methods section (please see the paragraph below), explaining how we arrived at the 40 tumor biopsies from 20 patients and discussed the representativeness of these samples in the context of the study population on each occasion. Briefly, we analyzed the dropout rate due to library construction failure in both favorable and unfavorable responders, finding comparable rates (favorable responders:

50%; unfavorable responders: 40%). Baseline patient and disease characteristics for these 12 patients are presented in Supplementary Table 4. The median age was 42 years, with 66.7% having clinical stage II breast cancer and 75% of tumors exhibiting pure ductal histology. Additionally, 16.7% of tumors were classified as HR low-positive. These characteristics align with the 29 patients included in the efficacy analysis outlined in Ada et al., 2024 (Supplementary Table 4). A comparison of baseline biopsy characteristics with other timepoints revealed no significant differences (Supplementary Table 4).

As we now state in the Methods section on page 21 of the highlighted revised manuscript:

Patient population and sample selection:
Tumor biopsies were collected from patients enrolled in the randomized and open-label clinical trial NCT02999477, which enrolled 32 patients. All participants provided written informed consent before undergoing any study-related procedures. This study was approved by the institutional review board of Dana-Farber/Harvard Cancer Center and was conducted in accordance with the principles of the Declaration of Helsinki. Among 32 patients, two patients withdrew before receiving study therapy, and one was found ineligible due to HER-2 positivity after a single dose of pembrolizumab. Of the remaining 29 patients eligible for efficacy analysis, only 20 were ultimately evaluated. This discrepancy arose because the single-cell sequencing study was initiated prior to full trial enrollment (with six trial patients enrolled after the single-cell sequencing efforts began, who were then not included in the single cell cohort), one patient failed to achieve successful single cell profiling at all timepoints, and the absence of baseline biopsies for two patients. Briefly, baseline biopsies were available for 21 patients for single-cell sequencing; however, due to a high rate of library construction failure, only 12 patients had successful library construction and single-cell profiling. Baseline patient and disease characteristics for these 12 patients are presented in Supplementary Table 4. The median age was 42 years, with 66.7% having clinical stage II breast cancer and 75% of tumors exhibiting pure ductal histology. Additionally, 16.7% of tumors were classified as HR low-positive. These characteristics align with the 29 patients included in the efficacy analysis outlined in Ada et al., 2024 (co-submitted). For W3D1 biopsies, 14 patient samples were available, with successful library construction and single-cell profiling achieved for 9 patients. Similarly, for W7D1 biopsies, 12 patient samples were available, of which 8 underwent successful library construction and single-cell profiling. Among pre-surgery biopsies, 5 samples were available, with 4 successfully processed. For surgery biopsies, 10 samples were available, with successful library construction achieved for 7. Overall, despite variability in sample availability and processing success,

the patient characteristics of those with successful single-cell profiling remained consistent with those in the efficacy analysis (Supplementary Table 4).

B. From the snRNA-seq data of 40 tumor biopsies from 20 patients, the authors detected about 250K nuclei. For each sequencing batch, “2,000 highly variable genes were identified and used for the principal component analysis (PCA). The top 50 principal components (PCs) were then used to compute the nearest neighbors distance matrix,…”. What’s the size of each batch and how were the batches set up? Presumably the selected top 50 PCs were very likely different by batches, how was that factored in the cluster analysis? When the authors started working on the selection of MPs, how was the gene expression matrix constructed in light of the batch factor and the above-mentioned procedure?

Response: Thank you for raising these important points. In response, we have updated the manuscript to provide greater clarity regarding the analysis strategy and associated results. Like prior cancer studies that employ single cell analysis, this dataset encompasses seven sequencing batches, with full details available in Supplementary Table 1A. To ensure rigor, clustering and cell compartment annotation were conducted independently for each batch. To visualize the overall clustering structure across the entire dataset (visualized in Figure 1C), we concatenated data from all batches and applied UMAP to the combined dataset while preserving the cell labels assigned during the batch-specific analyses. This approach, which is consistent with established methodologic workflows for analyzing single cell data [1–3], allowed us to maintain resolution of biological variation while harmonizing the data.

When working with the MPs, we did not rely on the top 50 PCs. Instead, we applied non-negative matrix factorization (NMF) to each sample, testing ranks 4–10 to generate 49 distinct NMF programs. To mitigate batch-driven artifacts and prioritize biologically relevant signals, we employed a strategy adapted from a hallmark prior study that pioneered this approach [4] to identify programs shared across patients. The workflow proceeded as follows:

- Gene Ranking and Program Definition: For each NMF program, genes were ranked by their weight, and programs were characterized by their top 50 genes.

- Overlap Analysis: We assessed gene overlaps both within individual samples (to exclude redundant programs) and across samples (to identify shared biology). Programs were deemed robust if they exhibited <35 overlapping genes within the

same sample (reducing redundancy) and >10 overlapping genes between samples (highlighting consistency).

- Metaprogram Clustering: Finally, robust programs were aggregated into metaprograms using a custom clustering framework adapted from Gavish et al. (2023).

This approach emphasizes cross-patient biological reproducibility while minimizing technical biases, aligning with established methodologies for integrative multi-sample analysis.

Reviewer #1 (Remarks on code availability):

1) Page 3. “specific mechanisms between tumor and microenvironmental cells that may result in therapeutic response to ICI plus chemotherapy in breast cancer are incompletely understood.” Suggest using “not well understood” rather than “incompletely”.

Response: Thank you for this suggestion. We have modified this language in the revised manuscript.

2) Single-nucleus transcriptome and chromatin accessibility profiling before and during monotherapy and combination therapy.
• The rationale for selecting RCB 0/1 instead of pCR (= RCB 0) is not adequate.

Response: We thank the reviewer for the opportunity to expand on this further. Our decision to use RCB 0-1 vs RCB 2-3 in genomic analyses was based on multiple considerations. Notably, while there are clear differences in outcomes between RCB-0/pCR and RCB-I in HER2+ and HR-/HER2- breast cancer, in HR+/HER2- breast cancer, EFS is similar between these two categories5,6. Additionally, while RCB is prognostic in HR+ breast cancer, with RCB 0-1 patients experiencing fewer long-events than those with RCB 2-3, RCB is not validated as a predictive biomarker in HR+ breast cancer5 and therefore is not incorporated into adjuvant therapy recommendations (unlike HER2+ breast cancer). Finally, there is the practical consideration that pCR is uncommon in HR+ breast cancer7. This approach is also in accordance with the clinical trial results presented in the companion study (Waks et al).

• “No cluster bias was evident when assessing pretreatment, monotherapy, and on-combination therapy biopsies within or between treatment arms (Fig. 1D; Supplementary Fig. 1B)” Not sure what this statement was based on. Did it mean no obvious patterns were observed in those plots?

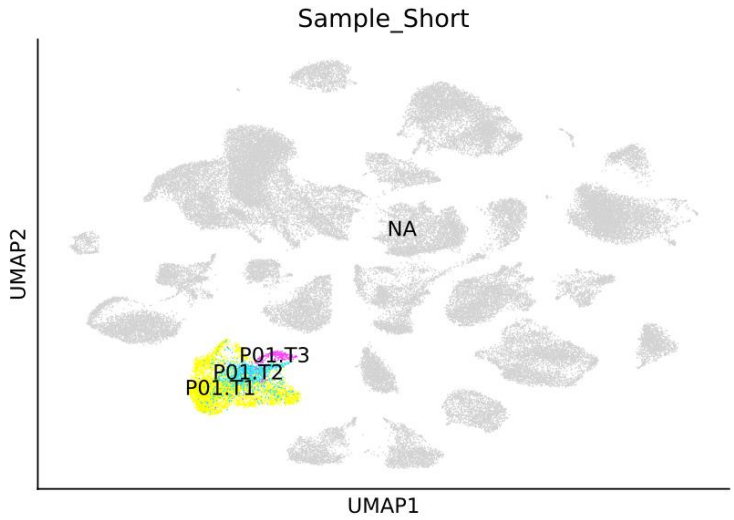
Response: The conclusion was drawn from our observation that the plots did not exhibit any pattern showing any cell type (e.g., stromal, epithelial, or immune populations) unique to the treatment timepoint, library technique or patient. Consequently, we have revised the manuscript on page 6 of the revised highlighted manuscript for improved clarity:

None of these three major cell classes were unique to any specific patient, library construction method, treatment, or treatment timepoint. (Fig. 1D; Supplementary Fig. 1B).

3) Tumor-intrinsic molecular programs associated with combination therapeutic Response.

• Supplemental Figure 2C. The lower panel includes all malignant cells from those two patients. Can the sample be annotated? Or separation between samples from the same patient is not great?

Response: We included the UMAP labeled by sample ID for each patient (Reviewer Attachment 1; see examples below). Consistent with prior cancer single cell studies, we hypothesized that inter-patient tumor heterogeneity is much greater than intra-tumor heterogeneity. While most samples from the same patient cluster together, a few patients show distinct clusters.



• Figure 2A and Supplemental Figure 2D. Was there any reason for why patient p27 was not included?

Response: Thank you for your detailed review. We did not include p27 in this section because only eight malignant cells were detected in the samples from this patient. Our methodology of filtering out samples with less than a specific number of cells has been noted on page 25 of the revised highlighted manuscript:

Within each cell type, we filter out samples with number of cells less than 100 and adopted the methodologies outlined in previous studies[14,15].

• The legend of Figure 2A, “Programs are clustered into nine metaprograms (MPs)”. Did you mean “Genes are clustered …”? Were the top 50 genes selected because they had the highest variation?

Response: Thank you for your question. The reference is robust NMF programs, not genes. We clustered these programs into metaprograms based on their shared genes for further grouping. The top 50 genes of MPs are ranked by their weights in the corresponding robust NMF program. We have revised the Methods to clarify this point.

• Figure 2C. Whether the baseline samples were representative of the favorable responses, and the unfavorable responses affects the validity of the comparisons.
Response: Thank you for your question. To confirm that our baseline samples accurately represent the trial cohort, we compared the dropout rates from library construction failures between favorable and unfavorable responders and evaluated the clinical characteristics of baseline biopsies with single-cell data against those of the overall cohort—an analysis that revealed no significant differences, which we have added to our revised study (new Supplementary Table 4).

4) CD8+ T cell states responsive to combination therapy
• All the samples were clustered into three groups based on the three MPs. Can the clustering of different samples from the same patient be discussed so that we can make sense out of the characterization?

Response: Thank you for your comments. We calculated the proportion of three CD8T cell states relative to the overall CD8T cell population and investigated the population changes during treatments. Consistent with the findings in Figure 3, we observed an

increase in cytotoxic cells during treatment for all patients, while a decrease in exhausted T cells was noted only in favorable responders (New Supplementary Fig. 3G).

• The comparisons described in Figures 3D-3F seem to be exploratory by nature. The Distributions largely overlapped each other even though there were statistically significant differences. How can such information be used to guide clinical practice? This question can be raised for the analyses/figures on macrophage cells.

Response: Thank you for your observation regarding the comparisons in Figures 3D-3F. While the distributions show overlap, the statistically significant differences provide potentially meaningful insights into CD8+ T cell states. Specifically, Figure 3D highlights that the dynamic changes in exhausted CD8+ T cells from non-responders may differ between chemotherapy and anti-PD1 treatment. Meanwhile, Figures 3E and 3F illustrate differences between the dynamics observed during combination therapy and those after its completion, shedding light on changes over shorter versus longer timeframes. Collectively, these findings suggest that variations in CD8+ T cell profiles could serve as potential biomarkers to identify patients who are more likely to respond to treatment. Additionally, the observed post-treatment dynamics may reflect underlying mechanisms of selective therapeutic response. However, we recognize that the overlapping distributions limit the immediate clinical applicability of these findings. To bridge this gap, further studies involving larger cohorts and more detailed analyses are needed to validate and extend these observations. Future efforts should focus on refining these insights to establish robust thresholds or composite biomarkers that better differentiate responders from non-responders. This approach could ultimately enhance patient stratification and guide more tailored therapeutic interventions. In line with this thoughtful point raised by the reviewer, we have modified our Discussion to further reflect such considerations.

Reviewer #2 - (also reviewed linked manuscript) (Remarks to the Author):

The authors of the manuscript entitled “Cellular reprogramming during anti-PD-1 and chemotherapy treatment in early-stage primary hormone receptor-positive breast cancer” have conducted a single cell-based analysis from longitudinal tissue samples obtained from patients enrolled in a prospective study of neoadjuvant nab-paclitaxel with pembrolizumab in patients with stage II-III HR+/HER2- breast cancer. By using single-nucleus RNA and ATAC sequencing, the authors identified distinct cellular states and gene expression metaprograms which are associated with differential therapy responses. This study provides insights on (chemo)immunotherapy resistance/response mechanisms in patients with early HR+/HER2- breast cancer -a subtype considered to display lower levels immune infiltration and immunogenicity- and several interesting TME-related implications for the biology of the disease.

Response: We thank the reviewer for their kind words and thoughtful review.

Major comments:
1. The authors focus on the longitudinal samples single-cell analysis of. However, presentation of the primary and key secondary results of the trial would contribute to a more complete and overall presentation of the study, rationale and implications of the translational sub-studies

Response: Thank you for your thoughtful comments. The companion manuscript (Waks et al., 2024) provides a detailed account of the primary and secondary outcomes of the trial. While we reference those findings here for context, we have intentionally not included them in the manuscript to avoid redundancy:

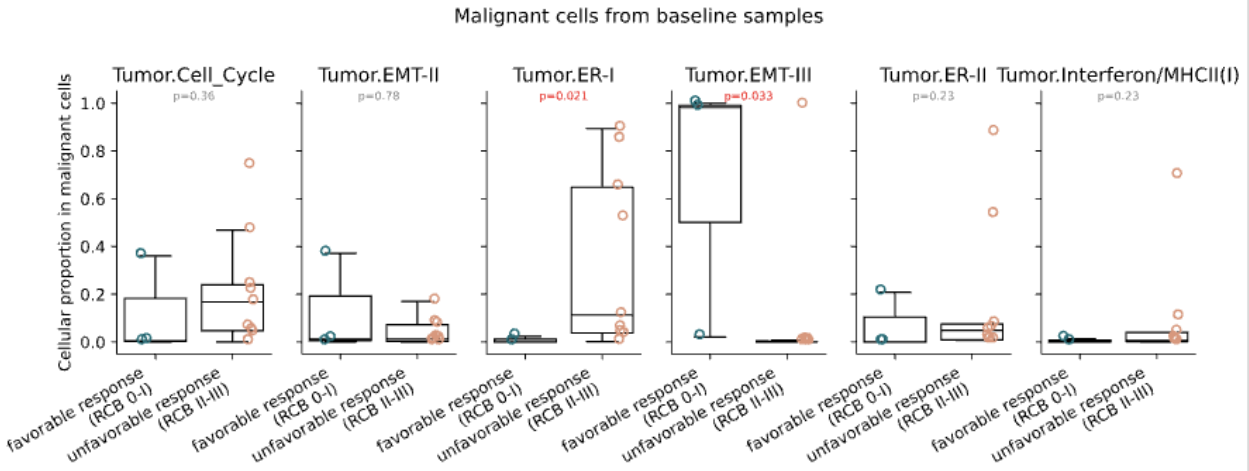
> ***Results:*** *Of 29 patients, 72% were node-positive. Residual cancer burden (RCB) 0-1 rate was 28% (inclusive of patients receiving additional neoadjuvant adriamycin/cyclophosphamide). No significant change in PD-L1 expression occurred following nab-paclitaxel or pembrolizumab run-in. Higher baseline PD-L1 expression and inflammatory gene signatures were associated with favorable response (RCB 0-1); higher expression of estrogen response genes, with unfavorable response (RCB 2-3).*
>
> ***Conclusions:*** *In this work, PD-L1 expression did not change significantly following single agent ICI or chemotherapy.*

2. Although the study lies on single cell-based approaches for the different analyses and TME interactions, validation of these findings in situ (using a spatial biology method) would further enhance the results of the study.

Response: Thank you for your comment. We agree that spatial biology approaches could further inform the interaction analyses; unfortunately, tissue blocks from patients on trial were exhausted (between the studies described in this manuscript and the companion clinical trial effort). Our analyses rely on the available data and may require further validation through additional studies with appropriate tissue collection. Given these issues, we presented hypothesized interactions in this study that had potential biological relevance and were related to clinical outcomes in this unique trial cohort, and we have acknowledged this further in our revised Discussion.

Minor comments:
1. The authors focus mostly on two metaprograms, namely the MP7 and MP11. It would be interesting to provide more results on the other MPs, i.e. interferon, cell cycle.

Response: Since we did not observe notable differences between favorable responders and unfavorable responders in other MPs, we only included MP7 and MP11 in our manuscript. However, to provide additional clarity, we have included the results for the other MPs as new Supplementary Fig. 2I in the revised study.



Supplementary Fig. 2: Detailed Analysis of tumor cells
I: Baseline comparison of the relative abundance of tumor cell states in the tumor population between favorable responders (R) and unfavorable responders (NR), with p-values calculated using the Mann–Whitney–Wilcoxon test. Each dot represents an individual sample.

2. The authors could consider presenting also longitudinal changes of the cancer cell states/MP, since they present longitudinal results for T-cells and macrophages

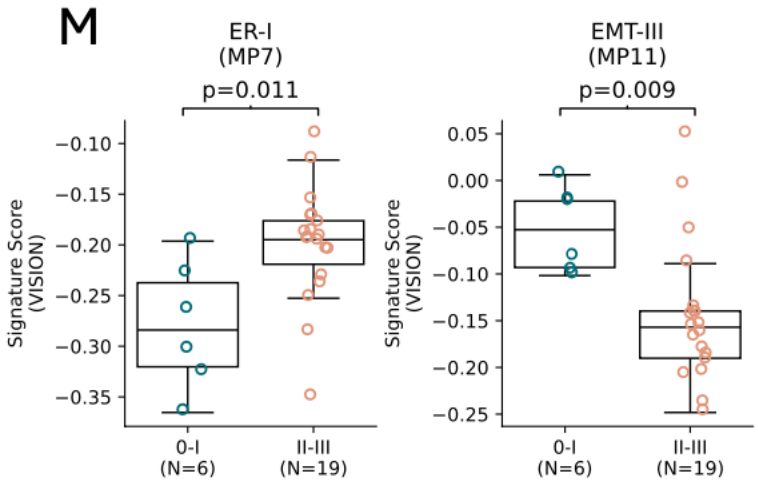
Response: Thank you for your valuable suggestions; we have included the results in the revised manuscript. Briefly, EMT-III-related MP11 was elevated by both monotherapy regimens but reduced by combination therapy. In contrast, ER-I-related MP7 increased preferentially in unfavorable responders to combination therapy while decreasing in favorable responders. This further suggests that ER-I-related MP7 is linked to resistance to combination therapy.

3. Why did the authors choose to focus on macrophages and not in other cells playing a role in immunotherapy response (i.e. B-cells)?

Response: We should have clarified this in the manuscript. Due to the insufficient number of other immune cells, such as B cells (N=2,954), with one-third of the samples having fewer than 10 detected B cells, we chose to focus on the immune cell subtype with sufficient representation for additional analysis (macrophages; N=13,375). We appreciate this key point and have revised the manuscript accordingly to acknowledge this future direction.

4. The authors could consider using bulk RNA-seq data from the same samples, in order to correlate with the single-cell data.

Response: Thank you for your great suggestions. We have included the result in the revised manuscript as new supplementary figure 2M.



Bulk RNA-seq from the same trial cohort

M: Normalized ER-I-related MP7 state and EMT-III-related MP11 gene signatures scores in bulk RNA-Seq data from baseline breast biopsies of early-stage HR+ breast cancer patients. Samples are grouped by the therapy response. Unadjusted two-sided Mann–Whitney–Wilcoxon p-values are provided for comparisons between patients with favorable response and those unfavorable response.

5. The authors could clarify how many patients were included in the validation cohort (eribulin + pembrolizumab) with bulk RNA-seq data.

Response: Thank you for the suggestion. The validation cohort consists of 13 patients, and we have included this information on page 8 of the revised highlighted manuscript.

6. Lines 235-241: It seems that the last sentence in line 237 "…while a decrease in the cytotoxic effector gene signature.." is contradictory to what it is states in the concluding sentence in line 239 "…followed by the increase of the cytotoxic effector gene signature in CD8+ T cells"

Response: Thank you for pointing that out. We have revised the manuscript to resolve the contradiction between lines 237 and 239 for clarity and consistency. It was a typo in line 237; it should indicate an increase in the cytotoxic effect gene signature (Fig. 3F; Supplementary Fig. 3G).

7. The authors could consider discuss their results, in relation to other window-of-opportunity trials of immunotherapy in early breast cancer

Response: Thank you for the suggestion. We've added sentences in the Discussion section describing our results in relation to other window-of-opportunity trials of immunotherapy in early breast cancer in the revised manuscript.

8. The authors should correct the references 3 and 4 so that they correspond to the correct journal

Response: Thank you for the note. We have addressed this issue in the revised manuscript.

Reviewer #3 (Remarks to the Author):

The manuscript entitled “Cellular reprogramming during anti-PD-1 and chemotherapy treatment in early-stage primary hormone receptor-positive breast cancer” submitted by Fu et al, describes the impact of immune checkpoint blockade (ICB) in combination with chemotherapy on transcriptional programs of tumor cells and immune cells in early hormone receptor-positive breast cancer. The team performed single-nuclei RNA-seq and for some samples paired single-nuclei ATAC-seq on 20 patients with multiple sampling before, during, and after treatment (although not for all patients consistently) to investigate the impact of chemo-ICB combination on tumor and immune cells with the goal to identify mechanisms of ICB response/non-response. While study design and patient selection promise insightful findings into the effect of ICB/chemo combination treatment on tumor and immune cells, the study has limitations that should be addressed.

The patients are divided into two groups: patients that first receive chemo and then ICB-chemo combination and patients that receive ICB and then combination. How does ICB alone impact immune and tumor cell abundances and phenotypes compared to a combination with chemo or pre-chemotherapy? The study only compares responders to non-responders without taking the different treatment arms into account. It would be interesting to see differences or similarities between the two different treatment regimes. Additionally, how do cell abundances and phenotypes shift over the course of the treatment when comparing the different sample time points?

Response: Thank you for your careful attention to our manuscript and for emphasizing these points. In the revised manuscript, we have incorporated findings from our analysis of tumor MPs across various timepoints, which also includes a comparative examination of differences between two monotherapy regimens. Unfortunately, between the limited sample size and comparable clinical outcomes observed between the two treatment arms, analyses comparing the treatment arms are not feasible. These outcomes are further explored in the companion clinical trial study co-submitted with this study (Waks et al 2024).

It would be great if the authors would explain the RECIST response. It is unclear from the text or the methods of how patients are selected to be analyzed by snRNA or multiome profiling, WES or bulk RNAseq. It is also not clear how the sampling of patients was chosen since there are 5 different time points for sampling, but these are not applied to all patients. As indicated in Figure 1B some patients are represented only with one timepoint. What are the inclusion/exclusion criteria?

Response: We attempted to profile all collected samples; however, due to a high library failure rate (in the setting of often very small and difficult to handle trial biopsy specimens), only half of the samples had successful sequencing profiles. In our revised manuscript, we included a detailed explanation in the Methods section (please see the paragraph below), explaining how we arrived at the 40 tumor biopsies from 20 patients and discussed the representativeness of these samples in the context of the study population on each occasion. Briefly, We analyzed the dropout rate due to library construction failure in both favorable and unfavorable responders, finding comparable rates (favorable responders: 50%; unfavorable responders: 40%). Baseline patient and disease characteristics for these 12 patients are presented in Supplementary Table 4. The median age was 42 years, with 66.7% having clinical stage II breast cancer and 75% of tumors exhibiting pure ductal histology. Additionally, 16.7% of tumors were classified as HR low-positive. These characteristics align with the 29 patients included in the efficacy analysis outlined in Ada et al., 2024 (Supplementary Table 4). A comparison of baseline biopsy characteristics with other timepoints revealed no significant differences (Supplementary Table 4).

As we now state in the Methods section on pages 20-21 of the highlighted revised manuscript:

Patient population and sample selection:
Tumor biopsies were collected from patients enrolled in the randomized and open-label clinical trial NCT02999477, which enrolled 32 patients. All participants provided written informed consent before undergoing any study-related procedures. This study was approved by the institutional review board of Dana-Farber/Harvard Cancer Center and was conducted in accordance with the principles of the Declaration of Helsinki. Among 32 patients, two patients withdrew before receiving study therapy, and one was found ineligible due to HER-2 positivity after a single dose of pembrolizumab. Of the remaining 29 patients eligible for efficacy analysis, only 20 were ultimately evaluated. This discrepancy arose because the single-cell sequencing study was initiated prior to full trial enrollment (with six trial patients enrolled after the single-cell sequencing efforts began, who were then not included in the single cell cohort), one patient failed to achieve successful single cell profiling at all timepoints, and the absence of baseline biopsies for two patients. Briefly, baseline biopsies were available for 21 patients for single-cell sequencing; however, due to a high rate of library construction failure, only 12 patients had successful library construction and single-cell profiling. Baseline patient and disease characteristics for these 12 patients are presented in Supplementary Table 4. The median age was 42 years, with 66.7% having clinical stage II breast cancer and 75% of tumors exhibiting pure ductal histology. Additionally, 16.7% of tumors were classified as HR low-positive. These characteristics align with the 29 patients included

in the efficacy analysis outlined in Ada et al., 2024 (co-submitted). For W3D1 biopsies, 14 patient samples were available, with successful library construction and single-cell profiling achieved for 9 patients. Similarly, for W7D1 biopsies, 12 patient samples were available, of which 8 underwent successful library construction and single-cell profiling. Among pre-surgery biopsies, 5 samples were available, with 4 successfully processed. For surgery biopsies, 10 samples were available, with successful library construction achieved for 7. Overall, despite variability in sample availability and processing success, the patient characteristics of those with successful single-cell profiling remained consistent with those in the efficacy analysis (Supplementary Table 4).

Overall, the study often uses visualization and statistical tests that assume single-cell-level information but are indeed comparing sample-level information (favorable or unfavorable responses are sample-level information and therefore samples need to be compared, and not individual cells of all samples combined). If the pool of cells is used for the test, then the test does not have the same meaning because cells from the same sample are correlated, so a summary stats for each sample should be computed, and then a test should be run between the groups. A few examples are Fig2C (assuming the boxplot is showing the expression in individual cells and not in cells normalized by sample), Figure2D, E,F, Fig 3 C D E F, etc,

Response: Thank you for your insightful feedback. I fully agree that cells from the same sample are correlated. To account for this, we conducted the tests using a linear mixed-effects model. Briefly, the gene signature scores were compared between two groups of interests to identify signatures with significant differences. The cell type-specific signatures changes over treatment time were modeled using linear mixed-effects regression model:
$y=\beta_0+ \beta_1 t+Zd+\epsilon$, where y is the observed signature score for a signature across all cells in the given cell type, β_0 is the intercept, β_1 is the fixed effect of treatment time t on signature level, Z is a binary design matrix indicating if the cells are from the same patient or not, d is a vector of random effect for patient, which is normally distributed with mean zero and represents the deviation from the overall mean of the mean signature scores for each patient, and ϵ is random errors. To identify signatures that significantly differentiate favorable responders from unfavorable responders, we applied the similar formulate, where β_1 is the fixed effect of therapy response t on signature level.
We have included this clarification in the revised manuscript (Methods: Differential gene signature analysis) on page 27 of the revised highlighted manuscript.

The study finds that patients with tumor cell states of high EMT-III show better responses compared to patients with tumor cells with high ER-signaling, who show worse responses. Especially, enrichment of EMT-III in favorable response patients is minimal and might not exist anymore if calculated on the sample level, instead on the single-cell level as indicated above.

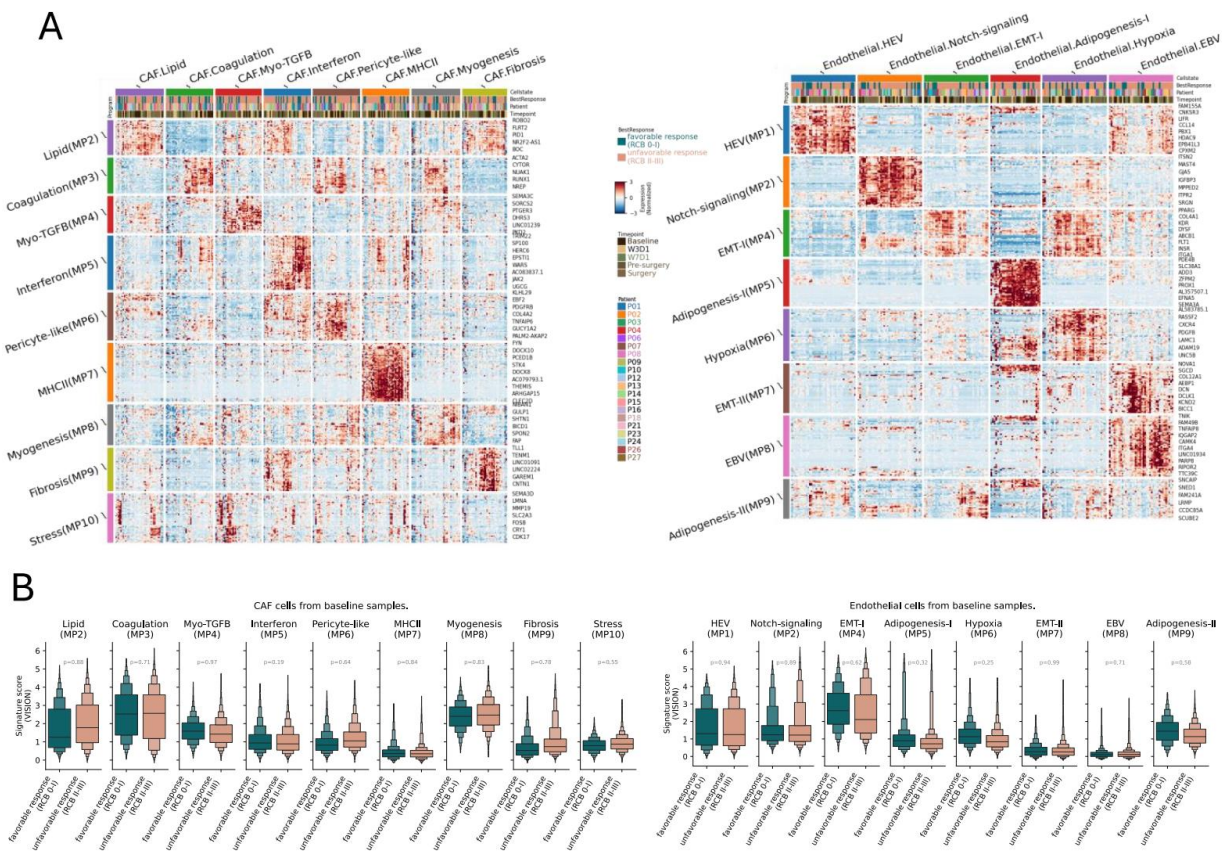
Response: Thank you for your comment. As mentioned above, we used a linear mixed-effects model to address this issue. Additionally, as shown in Figure 2D, the comparisons were made at the sample level. We observe an enrichment of EMT-III in favorable responders. We agree that the effect size may not be such that it is sufficient for determining selective response in totality and denote the need for further validation considerations in future efforts. (see Methods: Differential gene signature analysis on page 27 of the revised highlighted manuscript).

Cell numbers for all samples have to be reported. Fig 3C shows the activity of IL15 in different T cells, however cell numbers seem very small. This is the only graph, where cell numbers are depicted, which raises the question of cell numbers for the other analyses.

Response: Thank you for your comments; we apologize for the lack of clarity on our part. The dot in Figure 3C represents samples, not cells. Within the CD8 T cell population, we identified 9,453 cytotoxic CD8 T cells, 5,392 exhausted CD8 T cells, and 2,385 naive CD8 T cells. To avoid confusion, we have revised the Figure 3C legend and included the cell counts for the various cell states in the updated manuscript Supplementary Table 1.

No stromal cells are described and analyzed in the study, although stromal cells compromise the majority of some samples. I would suggest describing the stromal part as well.

Response: Thanks for your comments. We agree that the stroma may be biologically relevant in this context, so following this suggestion we conducted a similar analysis on stromal cells. Unfortunately, this investigation did not yield significant findings. Nevertheless, we are including the results here, in case they can inform future investigations by the research community:



Supplementary Fig. X: Detailed Analysis of stromal cells

A: Heatmap showing the average expression of CAF-related genes across CAF cell states (left) and Endothelial-related genes across Endothelial cell states. Rows list gene names, with cell state, patient, and timepoint information indicated at the top.

B: Boxplot illustrating the distribution of metaprogram (MP) gene signatures in pretreatment CAFs (left) and Endothelial cells (right), comparing favorable and unfavorable responders. Unadjusted p-values from the linear mixed model, with patient ID as a random variable, are displayed.

To delineate how monotherapies and combination therapy reshape stromal cell states, we analyzed metaprograms (Supplementary Fig.X1 A) in cancer-associated fibroblasts (CAFs) and endothelial cells across treatment regimens. No differential metaprograms distinguished favorable versus unfavorable responders in baseline samples (Supplementary Fig.X1 B)

How are the findings presented in this study different/similar to other studies? In particular, the interaction studies, how do these findings relate to other similar studies?

Response: Previous studies have identified distinct tumor subtypes in primary breast cancer[8] and the cell origin of luminal and basal tumor cells[9]. Another study investigated changes in the tumor microenvironment, specifically focusing on its relationship to T cell expansion [2]. However, it remains unclear how intrinsic tumor features and interactions with immune cells influence therapy response during chemotherapy, anti-PD-1 inhibitors, or combination therapy. Therefore, this study aims to explore how initial treatments (chemotherapy versus anti-PD-1 inhibitors) affect the tumor microenvironment, and how these effects vary between responders and non-responders during and after combination therapy in early-stage HR+ breast cancer patients. Our findings reveal distinct differences in the dynamic tumor-immune interactions between these groups.

Notably, previous studies have identified CTLA-4 as a marker of exhausted CD8+ T cells, which plays a crucial role in regulating immune responses by suppressing T cell activation and effector functions. In our analysis, we observed differential ligands expressed on malignant cells interacting with CTLA-4 on CD8+ T cells following combination therapy. Specifically, responders exhibited CD86/CTLA-4 interactions, while non-responders showed FAM200A/CTLA-4 and WBP1/CTLA-4 interactions. Additionally, prior research has demonstrated that macrophages expressing NRP1 and NRP2 are associated with tumor progression. Our study further identified VEGFA, a key ligand expressed on malignant cells, as a potential driver of these interactions. VEGFA likely engages with NRP1/NRP2-expressing macrophages, contributing to immune evasion and tumor survival by inhibiting tumor-killing mechanisms. We have revised the Discussion to clarify this point based on this helpful question.

What effects are generalizable, which are BC specific, and which are ER+ BC specific?

Response: Thank you for pointing this out. Previous studies have revealed a positive correlation between EMT and PD-L1 expression in breast cancer patients [10]. Additionally, high immune infiltration in lung adenocarcinoma has been strongly linked to the induction of EMT, which is associated with increased expression of inhibitory immune checkpoints, such as CTLA-4 and PD-L1 [11]. A pan-cancer study[12] further demonstrated that the crosstalk between immune evasion and EMT correlates with responses to immune checkpoint blockade across several cancer types. Collectively, these findings suggest that tumor specific EMT may serve as a biomarker for immunotherapy response not only in breast cancer, but also more broadly across other cancers. We have included a Discussion on this on page 18 of the revised highlighted manuscript.

How does chemotherapy/ICB combi impact cell phenotypes and abundances compared to ICB alone?

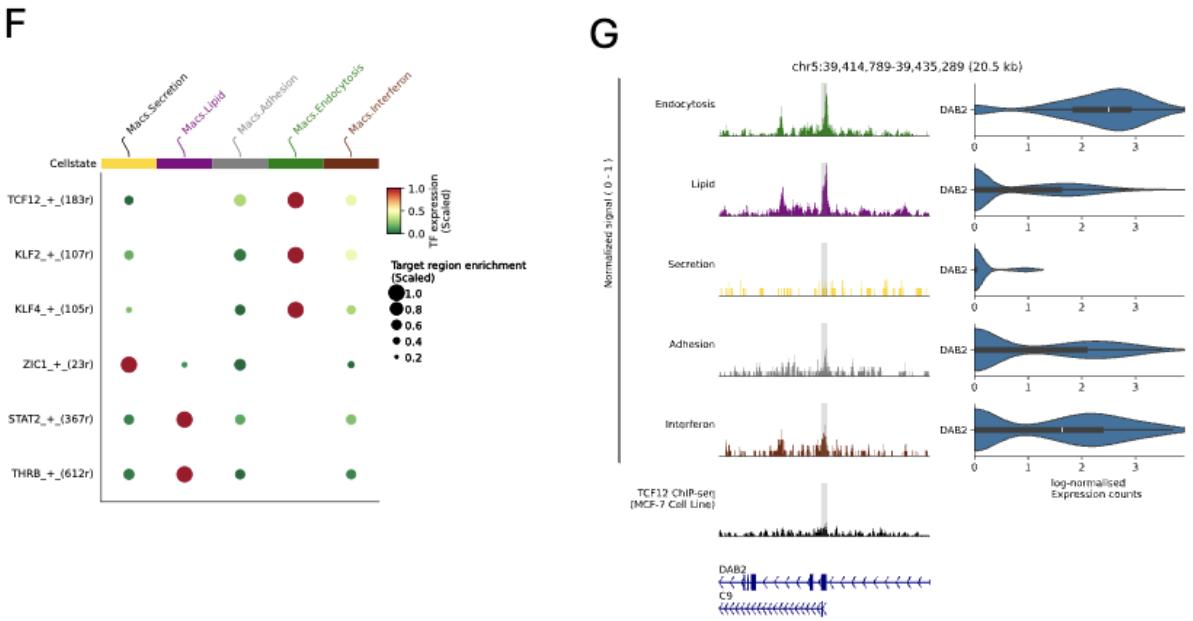
Response: In Figures 3 and 4, we examined the effects of chemotherapy alone versus ICB alone on the tumor microenvironment. It is important to note that these differences were observed after a short period of exposure, as patients transitioned to combination therapy three weeks following their initial run-in treatment. Our findings revealed that exhausted CD8+ T cells were preferentially enriched in non-responders following ICB treatment, but not chemotherapy. Additionally, combination therapy further enhanced the exhausted cell state in CD8+ T cells from non-responders. We also observed a decrease in the endocytosis-related MP1 gene signature in macrophages after either type of monotherapy. Interestingly, this decrease was only observed in responders following combination therapy, suggesting a potential link between this macrophage phenotype and treatment response.

Do the authors have tissue blocks collected from the patients to validate key findings?

Response: Thank you for your comment. Unfortunately, we did not have access to such samples for validation purposes, as the tissue was used for this study and the associated efforts from the companion clinical trial study co-submitted with this paper. Our analyses rely on the available data and may require further validation through additional studies with appropriate tissue collection.

ATACseq data are only used in a small part of the analysis in regard to different tumor cell states. For the immune cells, these data are not used at all.

Response: We appreciate the reviewer's insightful observation. During multi-modal integration of high-quality cells, we observed a substantial reduction in cell counts, with only malignant cells and macrophages retaining sufficient numbers (>1,000 cells; Supplementary Table 1). Given the predominance of malignant cells in our dataset, we prioritized ATAC-seq analysis for this population in our original submission to ensure robust biological interpretation. However, to maximize the utility of our data, we extended our analysis to macrophages, where we identified an enriched transcriptional regulatory network (TCF12-DAB2) linked to endocytic activity—a finding consistent with their functional role. The results have been provided in the revised manuscript.



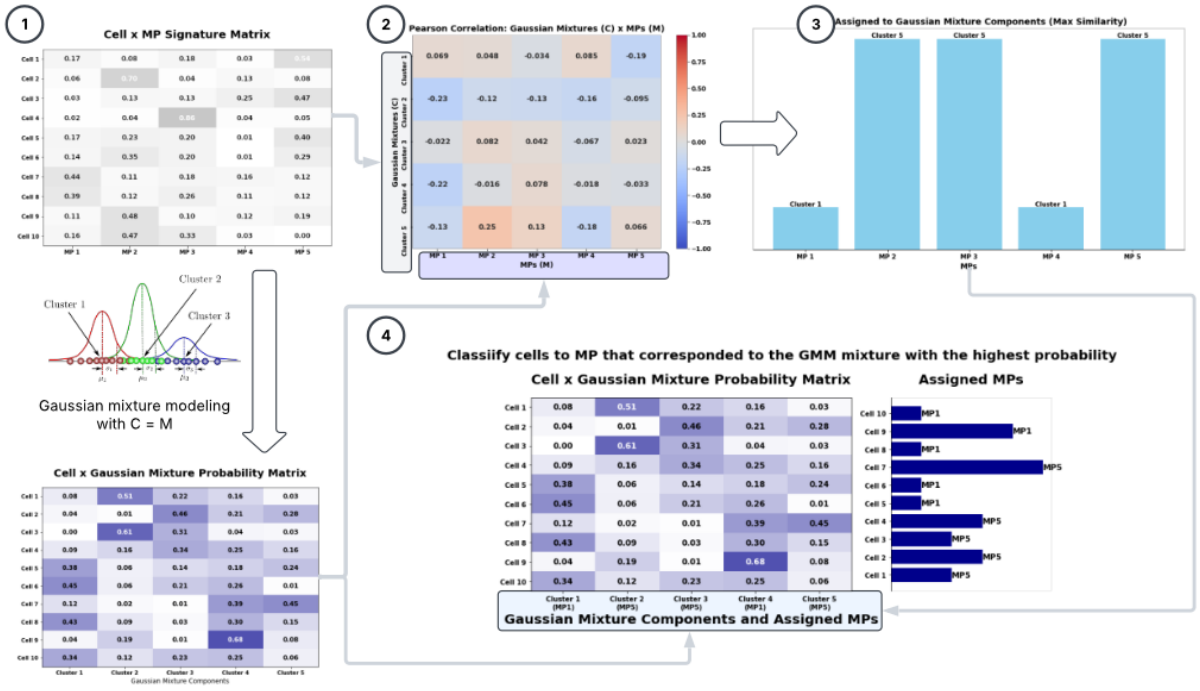
Supplementary Fig. 4: Detailed Analysis of Macrophages
F: Dotplot showing scaled transcription factor expression (color) and scaled target region enrichment score (dot size). Rows are gene regulation networks (GRNs), named after the transcription factor and its putative regulatory region. Columns represent four tumor cell states.
G: Chromatin accessibility profiles for the DAB2 gene region are shown across five distinct macrophage states. On the right, violin plots illustrate the expression levels of DAB2 across these states. Predicted TCF12 target sites are marked with colored ticks and semi-transparent boxes, and the lower left panel presents TCF12 ChIP-seq data for this region from the MCF-9 cell line[41-43].

Overall, the computational approach seems convoluted. I am not familiar with the cell state annotation that I assume is used for e.g. fig 2D, it would be great if the authors could explain their workflow in greater detail.

Response: Thank you for your feedback. We appreciate your suggestion to provide more clarity on the computational approach and regret our lack of clarity on this point.

We've expanded the explanation of the cell state annotation process and provide a more detailed description of our workflow in the revised manuscript (Methods: Cell state annotation; Supplementary Fig.6).

Supplementary Fig. 6: workflow of cell state annotation in each cell type.



Reference

1. Klughammer J, Abravanel DL, Segerstolpe Å, Blosser TR, Goltsev Y, *et al.* A multi-modal single-cell and spatial expression map of metastatic breast cancer biopsies across clinicopathological features. *Nat. Med.* **2024**; 30: 3236–3249.
2. Bassez A, Vos H, Van Dyck L, Floris G, Arijs I, *et al.* A single-cell map of intratumoral changes during anti-PD1 treatment of patients with breast cancer. *Nat. Med.* **2021**; 27: 820–832.
3. Wolf FA, Angerer P, & Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **2018**; 19: 15.
4. Gavish A, Tyler M, Greenwald AC, Hoefflin R, Simkin D, *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **2023**; 618: 598–606.
5. Yau C, Osdoit M, van der Noordaa M, Shad S, Wei J, *et al.* Residual cancer burden after neoadjuvant chemotherapy and long-term survival outcomes in breast cancer: a multicentre pooled analysis of 5161 patients. *Lancet Oncol.* **2022**; 23: 149–160.
6. Symmans WF, Yau C, Chen Y-Y, Balassanian R, Klein ME, *et al.* Assessment of residual cancer burden and event-free survival in neoadjuvant treatment for high-risk breast cancer: An analysis of data from the I-SPY2 randomized clinical trial. *JAMA Oncol.* **2021**; 7: 1654–1663.
7. Litton JK, Regan MM, Pusztai L, Rugo HS, Tolaney SM, *et al.* Standardized Definitions for Efficacy End Points in neoadjuvant breast cancer clinical trials: NeoSTEEP. *J. Clin. Oncol.* **2023**; 41: 4433–4442.
8. Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat. Genet.* **2021**; 53: 1334–1347.

9. Iglesia MD, Jayasinghe RG, Chen S, Terekhanova NV, Herndon JM, *et al.* Differential chromatin accessibility and transcriptional dynamics define breast cancer subtypes and their lineages. *Nat. Cancer* **2024**; 5: 1713–1736.
10. Alsuliman A, Colak D, Al-Harazi O, Fitwi H, Tulbah A, *et al.* Bidirectional crosstalk between PD-L1 expression and epithelial to mesenchymal transition: significance in claudin-low breast cancer cells. *Mol. Cancer* **2015**; 14: 149.
11. Lou Y, Diao L, Cuentas ERP, Denning WL, Chen L, *et al.* Epithelial–mesenchymal transition is associated with a distinct tumor microenvironment including elevation of inflammatory signals and multiple immune checkpoints in lung adenocarcinoma. *Clin. Cancer Res.* **2016**; 22: 3630–3642.
12. Wang G, Xu D, Zhang Z, Li X, Shi J, *et al.* The pan-cancer landscape of crosstalk between epithelial-mesenchymal transition and immune evasion relevant to prognosis and immunotherapy response. *NPJ Precis. Oncol.* **2021**; 5: 56.

Responses to the Reviewers

To the reviewers:

Thank you for your helpful feedback on our manuscript. We feel that your suggestions have greatly enhanced our paper, and we have incorporated them as outlined below. All references to page and line numbers refer to the revised tracked edits version of the manuscript.

We welcome any further recommendations you may have.

Sincerely,

The authors

Reviewer #1 (Remarks to the Author):

In the revision, the authors provided considerable detail on sample selection and bioinformatic procedures leading to their findings. However, the more deeply one examines these steps, the more concerns arise about the robustness and reproducibility of the resulting observations. After standard quality control, the authors retained 22,125 nuclei with high-quality data for both snRNA-seq and snATAC-seq. In their analysis of tumor-intrinsic molecular programs, they identified malignant cells and derived multiple molecular programs per sample. Each of these programs consists of a set of genes with associated importance weights. Based on hierarchical clustering of selected molecular programs (Figure 2A), the authors identified a few "metaprograms," each comprising a collection of molecular programs, represented again by top genes and possibly weights. However, the interpretation of each metaprogram rests heavily on the specific genes included, and it is unclear how consistent or robust this clustering is across subsamples or perturbations. Furthermore, the method used to calculate per-cell signature scores for each metaprogram (Figure 2C) is not described. There are multiple plausible approaches (e.g., weighted or unweighted averages), but the omission of this detail makes the interpretation and reproducibility of these scores uncertain.

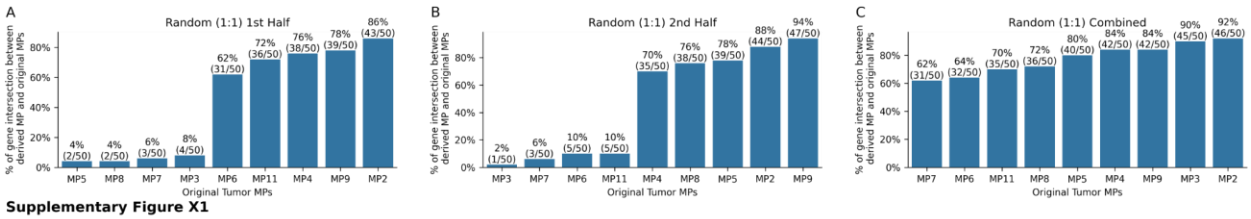
Response: Thank you for your insightful comments. We apologize for any lack of clarity on our part as it relates to concerns about robustness and reproducibility of the analysis. First, we would like to clarify that our analysis did not involve deriving sample-specific metaprograms. As illustrated in Supplementary Figure 2C, where cells are clustered by patient IDs, inter-tumor heterogeneity notably exceeds intra-tumor variance. Therefore, we adopted a previously published and widely utilized method for metaprogram inference from the tumor heterogeneity landscape study [1], which is specifically designed to identify molecular programs/metaprograms recurrent across multiple samples. Furthermore, we would like to clarify that the method used for signature calculation is described in the Methods section (Cell state annotation), where we

applied VISION [2] to quantify cell-type-specific metaprograms (MPs). Thank you again for your valuable feedback and the opportunity to clarify these points.

Overall, the procedure used to derive the metaprograms appears to involve a number of ad hoc decisions and fine-tuning steps, some of which may be sound, while others seem less well justified. Given this complexity and subjectivity, it is unlikely that an independent group applying the same pipeline would recover the same metaprograms. While some readers may focus primarily on the biological associations reported (e.g., with treatment response), those associations depend fundamentally on the upstream definitions of the metaprograms.
To test the stability and robustness of their method, I encourage the authors to randomly split the malignant cells from each sample into two halves (1:1) and re-derive the metaprograms using the same procedure on each subset. A comparison of the resulting metaprograms, up to the point of Figures 2A–B, would provide useful evidence for whether the identified structures are biologically meaningful or overly sensitive to data composition and procedural choices.

Response: Thank you for your insightful suggestions to examine this point in more detail. We have subsequently performed the comparison as recommended. Given that MPs reflect common molecular features shared by subpopulations across multiple samples, we anticipated that each randomly selected half would retain approximately 50% of the original MPs. Consistent with this expectation, each half recovered around 50% (4/9) of the original MPs, with an overlap exceeding 70% (35/50) (see Supplementary Figure X1 A-B, below, which is not included in the paper, but we are including here for the reviewers to see).

To evaluate the robustness of our method, we combined robust recurrent heterogeneous programs (RPHs) derived from the two random halves and conducted clustering on the concatenated RPHs. This approach successfully recaptured all original MPs, with a minimum overlap of genes greater than 60% (30/50) (see Supplementary Figure X1 C, below).



Supplementary Figure X1

A : Bar plot illustrating the interaction of genes between original tumor MPs and MPs derived from a randomly selected half of the tumor cells per sample (1st Half). The x-axis represents the original MPs, while the y-axis indicates the percentage of interaction.

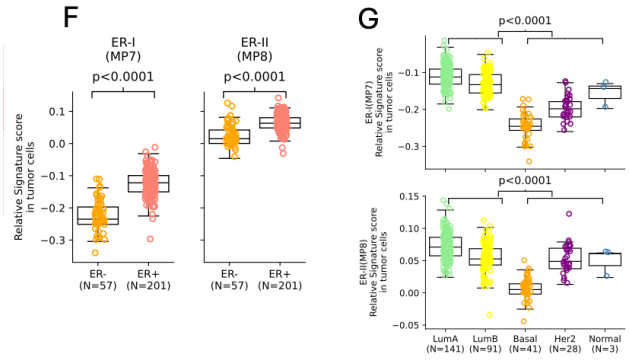
B : Bar plot illustrating the interaction of genes between original tumor MPs and MPs derived from a randomly selected half of the tumor cells per sample (2nd Half). The x-axis represents the original MPs, while the y-axis indicates the percentage of interaction.

C : Bar plot illustrating the interaction of genes between original tumor MPs and MPs derived from re-clustering RPHs from two randomly selected halves. The x-axis represents the original MPs, while the y-axis indicates the percentage of interaction.

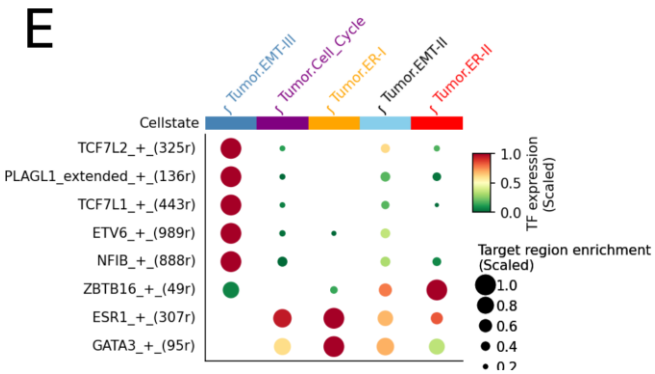
To further interpret the biological relevance of these MPs, we projected the signatures onto bulk RNA-Seq data from external studies, which yielded results consistent with our single-cell findings. For example, ER-related signatures were highly enriched in luminal and luminal B subtypes (Supplementary Figure 2G; also provided below), and ER+ patients from the TCGA cohort exhibited higher levels of ER-associated MP signatures compared to ER– patients (Supplementary Figure 2F; also provided below). Additionally, by integrating paired ATAC-Seq data, we identified canonical transcription factors that specifically regulate the MPs associated

with the expected phenotype (Main Figure 2E; also provided below). Taken together with the above results, biological consistency, and - most importantly - multiple forms of external replication, we propose these metaprograms as relevant in this context.

Supplementary Figure 2



Main Figure 2



Given the small cohort sizes in our discovery set that led to these definitions, we still recognize that further validation would additionally inform the significance of the metaprograms defined here as noted as a limitation in our Discussion.

Some detailed comments and questions are in the following:
1) Supplemental Figure 2K. Comparing the signature scores for EMT-III between baseline and W3D1 on Pembro, the centers of those two distributions are very close and the p-value is questionable. Same goes with the last comparison in Supp. Figure 2L.

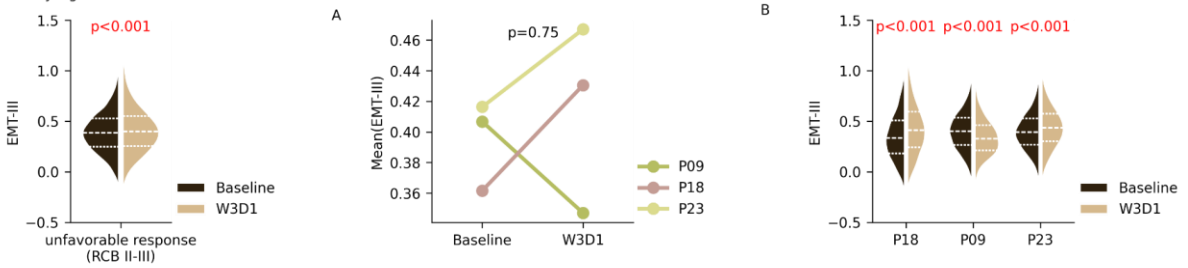
Response: Thank you for your thoughtful comments. We acknowledge that despite the small p-value denoted in this figure (Fig. 2K, 2L), the distributions between baseline and W3D1 are quite similar. However, when assessing average scores at the patient level, a difference becomes apparent in these paired biopsy analyses, as shown in Supplementary Figure X2A and X3A, below (corresponding to Supplementary Figures 2K and 2L, respectively), even though given cohort size that difference is not statistically significant. Furthermore, comparisons of cell-level distributions within individual patients reveal a clear separation between the two conditions, as shown in Supplementary Figure X2B and X3B (corresponding to 2K and 2L, respectively). While

comparing patient-level means is a reasonable approach, the limited patient sample size and our goal to fully leverage single-cell resolution led us to adopt an alternative statistical strategy that is reported herein. Specifically, to account for potential patient-specific confounding, we calculated the p-value using a linear mixed-effects model, as detailed in the Methods section (Differential gene signature analysis; also provided below). These linear mixed-effects model calculations result in the p values reported in Supplementary Figure 2K and 2L (and are noted in the legends). Even so, since patient P9 exhibited a significantly different trend in the opposite direction (Supplementary Figure X2B), and ultimately we are examining features from paired biopsies in small cohort subsets, we did not highlight the observed changes from baseline to W3D1 (Supplementary Figure 2K) in the main results.

We hope this clarifies the strategy and statistical testing plus results presented in this supplementary subpanel, and thank the reviewer for raising this issue. We also recognize the inherent limitations of the current dataset size, and as noted in the Discussion, additional studies with larger cohorts will be essential to further validate these findings.

> *Linear mixed-effects model details: $y=\beta_0+ \beta_1 t+Zd+\epsilon$, where y is the observed signature score for a signature across all cells in the given cell type, β_0 is the intercept, β_1 is the fixed effect of treatment time t on signature level, Z is a binary design matrix indicating if the cells are from the same patient or not, d is a vector of random effect for patient, which is normally distributed with mean zero and represents the deviation from the overall mean of the mean signature scores for each patient, and ϵ is random errors. To identify signatures that significantly differentiate favorable responders from unfavorable responders, we applied the similar formulate, where β_1 is the fixed effect of therapy response t on signature level.*

Supplementary Figure 2K

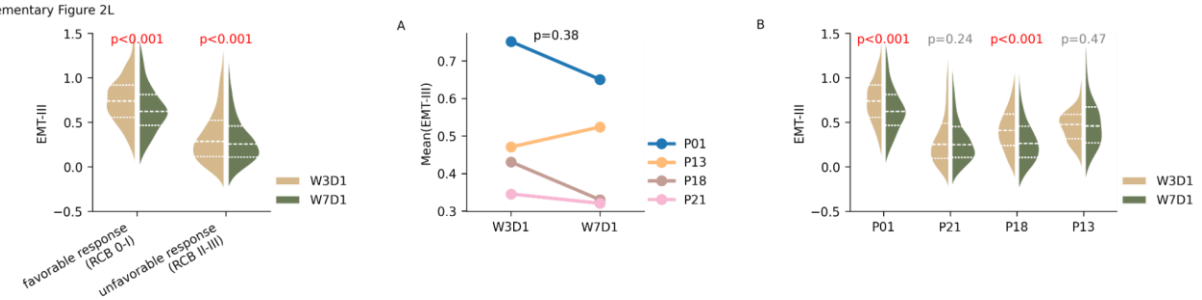


Supplementary Figure X2

A : Point plot depicting the average change in EMT-III signature from baseline to W3D1 for each patient.
Each point represents a sample and colored by patient id, and samples from the same patient are connected by a line.

B : Violin plot showing the distribution of EMT-III signature scores at baseline and W3D1, grouped by patient ID.
The x-axis represents patient IDs, the y-axis shows EMT-III signature scores, and violin plots are color-coded by timepoint.

Supplementary Figure 2L



Supplementary Figure X3

A : Point plot depicting the average change in EMT-III signature from W3D1 to W7D1 for each patient.
Each point represents a sample and colored by patient id, and samples from the same patient are connected by a line.

B : Violin plot showing the distribution of EMT-III signature scores at W3D1 and W7D1, grouped by patient ID.
The x-axis represents patient IDs, the y-axis shows EMT-III signature scores, and violin plots are color-coded by timepoint.

2) Page 8, lines 168-169. “Longitudinal analysis revealed that EMT-III-related MP11 levels significantly decreased during combination therapy, independent of response (Supplementary Fig. 2J-L).” The supplemental Figure 2L does not support that statement.

Response: Thank you for the careful review and we apologize for the issue noted here. We realized that Supplementary Fig. 2J–K illustrate changes from baseline to on-monotherapy (W3D1), whereas Supplementary Fig. 2L depicts changes from monotherapy to on-combination treatment (W7D1). As shown in Supplementary Fig. 2L, the EMT-III–related MP11 signature decreases from monotherapy to combination therapy, supporting our statement. Additionally, Supplementary Fig. X3 A and B show a clear decreasing trend at the patient level, based on average scores per patient. Therefore, on page 8, line 170 of the revised manuscript, we have replaced the reference to “Supplementary Fig. 2J” with “Supplementary Fig. 2L”.

3) Page 8, line 189. Suspect that the authors meant Supp. Figure 2M instead of 2K.

Response: Thanks for carefully reviewing our manuscript. We have corrected the typo in our revised manuscript (now on page 8, line 181 from “Supplementary Fig. 2K” to “Supplementary Fig. 2M”.

1. Gavish A, Tyler M, Greenwald AC, Hoefflin R, Simkin D, *et al.* Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **2023**; 618: 598–606.
2. DeTomaso D, Jones MG, Subramaniam M, Ashuach T, Ye CJ, *et al.* Functional interpretation of single cell similarity maps. *Nat. Commun.* **2019**; 10: 4376.

Reviewer #2 (Remarks to the Author):

The authors have successfully addressed my comments and provided also additional analyses and results (based on the reviewers' suggestions) which overall improved the manuscript.

Response: We are glad that we were able to address your comments and that you found our additional analyses and results helpful.

Reviewer #3 (Remarks to the Author):

The authors significantly improved their manuscript during the revision process. They sufficiently addressed my concerns. They added new analysis and now are describing the used methods and workflows more transparently to improve interpretation of the results. The dataset and its analysis add value to the current understanding of the effect of ICB and chemo in breast cancer on tumor cell phenotypes, TME and resistance.

Response: Thank you for your kind words regarding our new analysis and the dataset.

Reviewer #3 (Remarks on code availability):

I checked if the code is available, I didn't run it. Codes (python notbooks) are available for every figure and seem sufficient.

Response: We are glad that you found the code provided to be sufficient.

Responses to the Reviewers

Reviewer #1 (Remarks to the Author):

I appreciate the effort from authors to conduct extensive analyses to address previous concerns satisfactorily. Those additional analytical results will lead to more confidence in their interpretation.

Response: We are glad that we were able to address your comments and that you found our additional analyses and results helpful.