nature portfolio

Peer Review File

Multimodal blood based proteomic profiling reveals insights into mechanisms of immunotherapy resistance

Corresponding Author: Dr Arnav Mehta

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)
In this manuscript, Wright and colleagues aim to create a blood-based atlas that pinpoints plasma proteins distinguishing responders from non-responders to anti-PD-1 ± CTLA-4 therapy, traces those proteins to matched tumours or immune-cell sources, and flags markers that may predict immune-related toxicity. They profiled 2,943 proteins with Olink in 465 plasma samples collected at baseline and week 6 from 250 patients with metastatic melanoma. Matched PBMCs were analysed by CyTOF, and selected tumour samples underwent bulk and/or single-cell RNA sequencing. Linear mixed-effects models, non-negative matrix factorisation (NMF) and ligand–receptor analyses then integrated NPX, CyTOF, scRNA-seq and spatial data, linking temporal plasma shifts to cellular circuits and clinical outcomes.

Clear circulating biomarkers of early progression and toxicity are still lacking in routine cancer-immunotherapy practice, so this study is timely. The large, well-annotated sample set offers considerable potential. The manuscript is comprehensive and data-rich; however, many sections are overly dense and loosely connected, which disrupts the narrative flow. Redundant discussion of well-known or only marginally relevant concepts, together with an unclear conclusion, further obscures the possible clinical application of the results. The points below may help improve readability and strengthen the paper's impact. If the text were more concise and included additional validation, it could suggest useful tools for clinical practice.

Major observations
Validation – The study's impact would increase if at least some of the proposed biomarkers were validated in an independent cohort or dataset.

Figures – Several figures appear slightly blurry; higher-resolution images would improve clarity.

Introduction – Emphasise that only a few biomarkers (PD-L1, high TMB, microsatellite instability) are currently approved for clinical use.

Abbreviations – Ensure that each abbreviation is defined at first mention (e.g., immune-checkpoint blockade, tumour micro-environment).

Page 5, paragraph 3 – Note that on-treatment decreases in serum IL-6 were also associated with better overall survival in pancreatic cancer (DOI 10.1200/JCO.21.02511).

Page 5, paragraphs 3–4 – Several findings—such as the link between effector-memory $CD8^+$ T cells and response, or Tregs and non-response—confirm established concepts; highlight that this dataset recapitulates these insights at the serum-protein level.

Page 6, paragraph 2 – The sentence "Cluster 34… associated with immune-related diseases such as sepsis, influenza and gout" adds little insight; consider focusing on the most relevant findings.

Page 6, paragraph 4 – Instead of "aggregate expression of transcripts from scRNA-seq," refer to a gene signature or score corresponding to the previously identified proteins.

Page 6, paragraph 6 – Phrases such as “CXCL10 was expressed highly only in CXCL10$^+$ macrophages” are redundant; streamline for brevity.

Reviewer #2

(Remarks to the Author)
The study presents a comprehensive plasma proteomic analysis of metastatic melanoma patients undergoing immune checkpoint blockade (ICB) therapy. Using Olink technology, the authors identify proteins in circulation that are associated with treatment response, non-response, and the time course of therapy. A major strength of the work is the longitudinal sampling of serum, allowing for dynamic assessment of immune activity. The finding that both responders and non-responders exhibit evidence of immune activation is important and raises meaningful questions about the modulatory mechanisms that determine clinical outcome. The combination of proteomic profiling with matched immune cell proportions in peripheral blood is another strong aspect of the study.

While the dataset is rich and the overall approach is promising, some methodological and interpretational concerns limit the strength of the conclusions. In particular, the integration of single-cell RNA-seq data from pre-treatment tumor samples with post-treatment plasma proteomics appears speculative and not fully justified. The authors correlate mRNA expression patterns from tumor-resident immune cells with circulating protein levels after ICB therapy, but without temporal or spatial concordance. No protein measurements from tumor tissue are provided to support this link. As such, these findings should be framed as hypothesis-generating rather than conclusive evidence of cellular origin.
Additionally, the ligand-receptor analysis based on scRNA-seq data involves multiple assumptions, including threshold-based inferences of signaling interactions, and does not account for post-transcriptional regulation or secretion. This approach may help generate potential interaction maps, but should not be overinterpreted in the absence of functional validation. It is particularly confusing that only responders are shown with clear sender-receiver relationships, whereas non-responder-related proteins appear to lack corresponding source cell types in the visualization. This asymmetry is not sufficiently addressed.
Moreover, the methodology used for the ligand-receptor analysis is relatively manual and includes several steps that require further clarification. The authors identify cytokines, chemokines, and their receptors among the proteins derived from the linear mixed model, and then use the CellPhoneDB database to define known ligand-receptor pairs. They compute the average expression of each ligand and receptor across major cell types in the scRNA-seq data and assess interactions by calculating the minimum of the average ligand expression in one cell type and receptor expression in the other. Interactions with a normalized expression above 0.1 are considered present. For multicomponent receptors, the lowest expressed component defines the expression value. While the approach is biologically plausible, the choice of the 0.1 threshold and the handling of receptor complexes are not sufficiently justified, and the analysis may be sensitive to these parameter choices.

The work is timely and relevant, particularly in the context of ongoing efforts to understand mechanisms of ICB resistance and response. The large-scale, longitudinal serum profiling is a strength, and the findings regarding shared immune activation in both responders and non-responders are conceptually important. However, certain extrapolations—especially those relying on tissue-plasma integration—are speculative and should be tempered accordingly. The manuscript could benefit from citing related recent work, e.g. Yoo et al. ("Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data"), which shares the goal of response stratification using peripheral data.

The methodology for the Olink proteomics appears sound, and the authors perform linear mixed-effects modeling with appropriate inclusion of patient ID as a random effect. However, the introduction of melanoma subtype as a covariate reveals a major limitation: the strong imbalance of cutaneous versus non-cutaneous subtypes across response groups confounds many of the initial associations. After adjusting for subtype, most response- and non-response-associated proteins lose significance, suggesting that many findings are at least partially driven by subtype differences rather than true response biology.
The use of the Deng et al. plasma proteome atlas (53,026 adults) is introduced but insufficiently explained. The authors identify a similarity between their protein cluster and cluster 34 from the atlas, which is associated with influenza, sepsis, and gout. It remains unclear whether the authors are claiming a biological similarity (i.e., inflammatory mechanisms) or simply a shared protein expression pattern. This needs clarification.
The definition and rationale of the thresholding strategy for ligand-receptor analysis is not described in sufficient detail. The term "thresholding" is used repeatedly without explaining what thresholds were applied, how they were defined, or why they are biologically meaningful.

The use of linear mixed models is appropriate for handling repeated measurements and patient-specific effects. The model design, with time, response, and interaction terms, is well-chosen. However, the confounding introduced by melanoma subtype is a key concern. The authors acknowledge this but appear to understate its impact. The claim that subtype had "limited influence" on timepoint-associated proteins is only partially reassuring, given how many response-related findings were lost after including subtype as a covariate.

The authors should clearly state the limitations of correlating pre-treatment tumor transcriptomics with post-treatment plasma proteomics.

The ligand-receptor thresholding strategy should be better explained, including the cutoffs used and rationale for inferring signaling.

The subtype imbalance should be acknowledged more directly as a major limitation, and response associations should be interpreted cautiously in this light.

Figures need clearer legends, especially regarding abbreviations and what is being shown (e.g., sender-receiver relationships, response group definitions).

Consider including tumor proteomics or matched tissue data (even retrospectively) to better support the tissue origin hypotheses.

The use of the Deng et al. reference atlas should be contextualized more clearly—what does similarity to cluster 34 actually imply?

The manuscript would benefit from editing for clarity, particularly in the figure legends and transitions between sections. The first results section titled "Plasma proteomic profiling performed on metastatic melanoma cohort undergoing ICB" is misleading—it does not present proteomic data but describes cohort characteristics and survival analyses. The narrative jumps quickly between Olink, tissue, and single-cell data, sometimes without clear rationale or structure.

Overall, references are appropriate but additional recent work (e.g. Yoo et al.) should be cited where relevant.

Reviewer #3

(Remarks to the Author)
The authors present high dimensional proteomic profiling of plasma from a large cohort of patients with metastatic melanoma treated with systemic therapies. With validation of candidate targets within single cell and bulk transcriptomics. The study has a strong clinical rationale, statistical approach and unique longitudinal sampling provide robust analysis. Overall, a very strong manuscript and there for minor suggested added below:
1) Can the authors provide unsupervised clustering of the patient's proteomic profiles (PCA or UMAP like) to explore before/after batch effect, patient grouping based on response and treatment timepoint.
2) It's stated the multi-omic data is available in supplementary tables. Can the authors confirm the data availability of the remaining omics, COsMX, RNAseq etc..
3) Given the role of melanoma subtype on the results of the study, did the authors have access to tumour mutation burden or broader melanoma molecular subtyping for comparison?
4) Did the proteomic profiles associate or cluster based on site of disease, liver lung brain specific profiles in the plasma?
5)Given that some samples are taken on-treatment. Can the authors confirm any role competitive binding or masking of PD-1/CTLA-4 might have on the onlink results for these antigens?

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
All comments from the initial review have been addressed satisfactorily. I have one minor remaining point for clarification.

As results from plasma proteomics can be influenced by pre-analytical variability, it would be helpful if the Methods section could include a bit more detail on blood collection procedures (e.g., type of venipuncture device, needle gauge, use of a butterfly needle, tourniquet technique, patient position, and collection tubes) and sample processing (e.g., time from collection to processing and transport method). In addition, clarification regarding whether plasma or serum was used, and whether all samples were collected and processed in a consistent manner, would further strengthen the manuscript.

Reviewer #2

(Remarks to the Author)
Comments for Authors
This revised manuscript presents a large and carefully curated longitudinal plasma proteomics dataset from patients with metastatic melanoma treated with immune checkpoint blockade. The cohort size, clinical annotation, and use of linear mixed-effects models for longitudinal analysis remain major strengths. The additional analyses and clarifications included in this revision address several points raised in the initial review; however, several of the core interpretational concerns—particularly regarding subtype confounding and plasma–tissue integration—have been only partially addressed and remain

insufficiently resolved.

Noteworthy results and potential significance
The identification of robust treatment-associated plasma proteomic changes, including immune activation signatures observed in both responders and non-responders, is a notable and potentially valuable contribution. The longitudinal design and integration with peripheral immune profiling make this dataset a useful resource for the field. With appropriate framing, the study could help inform future efforts to develop blood-based biomarkers for immunotherapy monitoring.

Support for conclusions and remaining concerns
Despite improvements, several conclusions remain insufficiently supported by the presented evidence and require further clarification or reframing:
1) Melanoma subtype as a major confounder of response-associated signals.
The revised analyses confirm that melanoma subtype is the only clinical variable significantly associated with response, and that most response- and non-response-associated proteins lose significance after adjusting for subtype or restricting analyses to cutaneous melanoma. While these results are now more transparently presented, the manuscript still tends to frame certain findings as response-associated rather than largely subtype-driven. This distinction should be emphasized more consistently throughout the Results, Discussion, and figure legends to avoid overinterpretation.
2) Interpretation of plasma–tissue integration using scRNA-seq
The integration of pre-treatment tumor scRNA-seq data with post-treatment plasma proteomics remains conceptually problematic as currently framed. While the authors now acknowledge limitations, the manuscript still implies a closer correspondence between plasma protein abundance and the tumor microenvironment than is supported by the data. Tumor scRNA-seq provides transcript-level information from a distinct spatial and temporal compartment and cannot establish the cellular origin, secretion, or plasma contribution of measured proteins. These analyses should be framed consistently as hypothesis-generating contextualization rather than evidence of plasma protein origin or tumor-derived signaling.
3) Correlation of plasma protein modules with immune cell populations
The correlation analyses between plasma protein modules and PBMC immune cell proportions require clearer explanation and stronger discussion of confounding. Both measurements derive from peripheral blood and may reflect shared systemic drivers (e.g. tumor burden or generalized inflammation). In addition, CyTOF data represent relative cell proportions rather than absolute cell numbers, and wording should be adjusted accordingly.
4) Toxicity-related interpretations
The focus on pancreatitis as an immune-related adverse event is insufficiently justified. It is unclear whether other irAEs were observed in the cohort and, if so, why they were not analyzed. Moreover, interpretations suggesting tissue or organ origin of circulating proteins in the context of toxicity are speculative in the absence of orthogonal validation and should be clearly labeled as such. A summary of all observed irAEs (type, grade, frequency) should be provided, at least in the Supplementary Information.

Methodology, clarity, and presentation
The methodological framework is generally sound and described in sufficient detail for reproduction. However, several aspects would benefit from improved clarity:
• The definition and calculation of composite or aggregate protein scores should be explicitly described.
• Key heatmaps and clustering figures lack sufficient annotation of protein identities, limiting interpretability.
• Some sections remain overly long and descriptive, and further streamlining—particularly of validation and integration sections—would improve readability and focus.

Overall assessment
In summary, the dataset and analytical effort are strong, and the manuscript has improved in response to prior feedback. However, further revision is needed to ensure that conclusions are proportionate to the evidence, particularly with respect to response associations, plasma–tissue integration, and toxicity-related interpretations. Addressing these issues primarily through clearer framing, more cautious language, and improved transparency would substantially strengthen the manuscript.

Reviewer #3

(Remarks to the Author)
Thank you for performing the additional analysis requested. This has satisfied my comments.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed my last comment in a very satisfactory manner.

Reviewer #2

(Remarks to the Author)
The manuscript has improved following revision, and the authors have addressed the main concerns raised in the previous review round.

I have only a few remaining minor points:

1. Statistical terminology (“correlation”)

The term “correlation” is used in contexts where group comparison tests (e.g., Wilcoxon rank-sum or Fisher’s exact test, as in Table 1) are applied. This is not statistically precise and may be misleading. I recommend revising this wording accordingly (e.g., “differences between groups” or “association assessed by group comparison”).

2. Figure readability

Figure 1C: The x-axis labeling appears truncated (e.g., “200” is cut off).
Figure 2E: Axis labels/values are difficult to read due to overlap.
Figure 4B Similar readability issues as in Figure 2E.

These should be adjusted to improve clarity.

With these minor revisions, the manuscript will be suitable for publication.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Responses to reviewer comments

Reviewer comments in black | Responses in blue

Reviewer #1

In this manuscript, Wright and colleagues aim to create a blood-based atlas that pinpoints plasma proteins distinguishing responders from non-responders to anti-PD-1 ± CTLA-4 therapy, traces those proteins to matched tumours or immune-cell sources, and flags markers that may predict immune-related toxicity. They profiled 2,943 proteins with Olink in 465 plasma samples collected at baseline and week 6 from 250 patients with metastatic melanoma. Matched PBMCs were analysed by CyTOF, and selected tumour samples underwent bulk and/or single-cell RNA sequencing. Linear mixed-effects models, non-negative matrix factorisation (NMF) and ligand–receptor analyses then integrated NPX, CyTOF, scRNA-seq and spatial data, linking temporal plasma shifts to cellular circuits and clinical outcomes.

Clear circulating biomarkers of early progression and toxicity are still lacking in routine cancer-immunotherapy practice, so this study is timely. The large, well-annotated sample set offers considerable potential. The manuscript is comprehensive and data-rich; however, many sections are overly dense and loosely connected, which disrupts the narrative flow. Redundant discussion of well-known or only marginally relevant concepts, together with an unclear conclusion, further obscures the possible clinical application of the results. The points below may help improve readability and strengthen the paper's impact. If the text were more concise and included additional validation, it could suggest useful tools for clinical practice.

We appreciate the thoughtful response from the reviewer and expand on the specific concerns as they are mentioned below. Additionally, we focused on refining the messaging and clarity of the findings as suggested.

Major observations

Validation – The study's impact would increase if at least some of the proposed biomarkers were validated in an independent cohort or dataset.

We appreciate the comment from the reviewer, and agree with the importance of validating the markers in external, independent cohorts. We are limited in datasets available for external validation, which is in part why this dataset is so unique. We have spoken to several collaborators and colleagues who unfortunately do not have available representative cohorts for comparison at this time. Further samples obtained from our biobank in our mind would not suffice for an appropriate orthogonal dataset for validation. We have therefore focused our efforts on validating our findings, where able, with publicly available datasets. We note, however, that we are limited in that these other datasets almost always have fewer patients and fewer proteins measured, and the cohorts are not as carefully curated as the one in our paper. Our cohort represents almost a decade's worth of active biobanking of melanoma patients at MGH with careful clinical annotation.

We have updated the manuscript with validation analysis as well as an additional supplemental figure and supplemental tables. Additional analyses in response to this suggestion are reflected in Table S12, S13 and S14 and Figure S8, and below is the revised text which has been included as well as Figure S8:

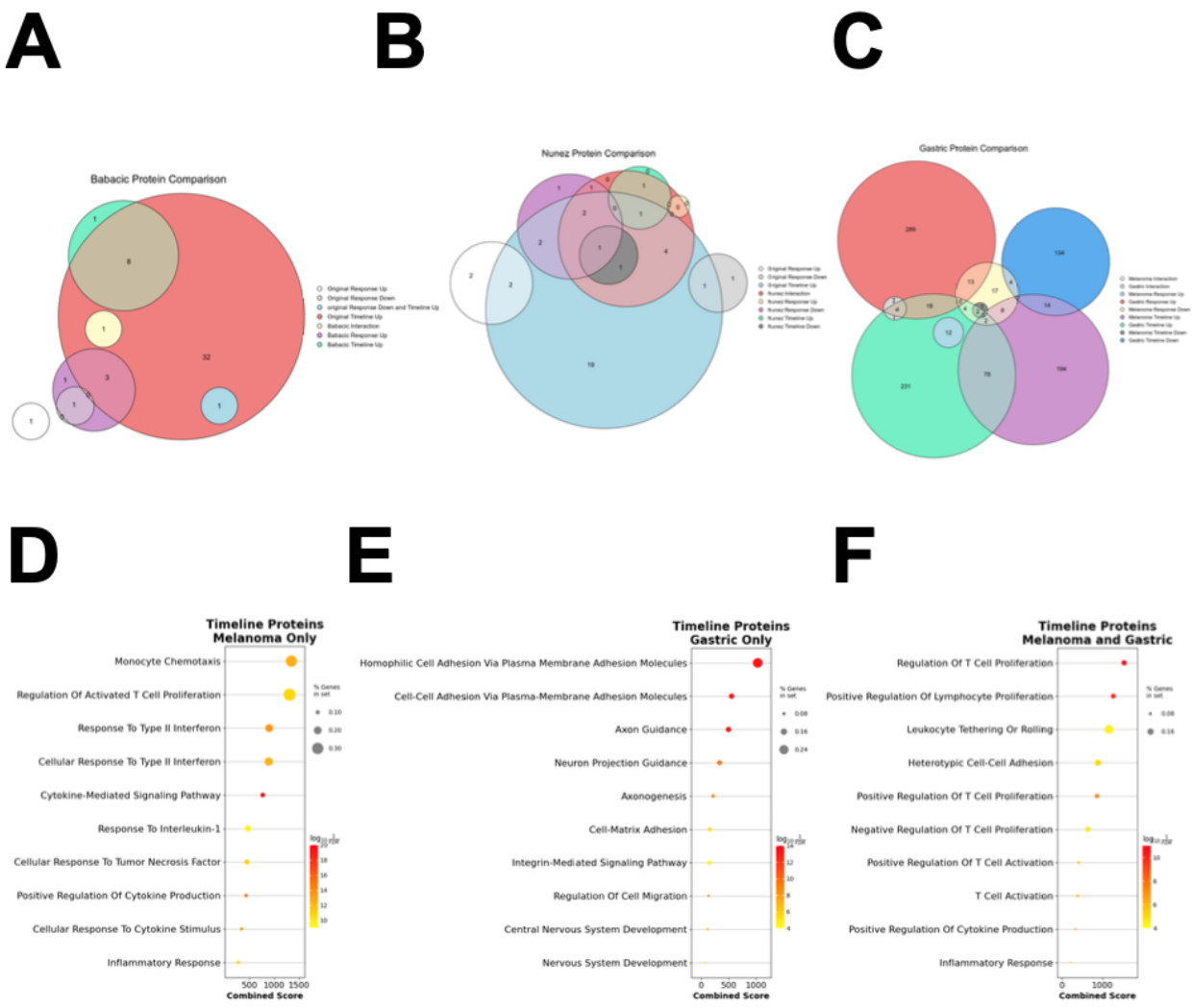
"To validate our treatment- and response-associated plasma proteomics signatures, we re-analyzed orthogonal data sets using our linear mixed effects model methodology. We found publicly available data sets with similar

structures involving plasma sampling of melanoma immunotherapy cohorts pre- and on-treatment with Babačić et al[1] and Nuñez et al[2]. We also utilized an internal plasma proteomics data set from a cohort of gastric cancer patients undergoing sequential chemo-immunotherapy (samples obtained from 46 patients pre-treatment, after one cycle of chemotherapy and after 7 cycles of combination chemo-immunotherapy). We acknowledge that studying gastric cancer patients may not be a direct comparison; however, we nevertheless felt that it was an opportunity to investigate shared trends in anti-PD1 response. While the Babačić et al. and Nuñez et al. data sets included a panel of around 90 proteins, the gastric data set included approximately 3000 proteins, which is in the range of what our melanoma immunotherapy data set had. More specifically, the Babacic data set included 15 patients (14 3 weeks on- and 13 pre-treatment samples) undergoing ICIs and the Nuñez data set included 41 patients with some combination of pre-, d7-14, 1 month pre-tox, and Tox time points.

There are 47 significant proteins from the original analysis that are present in the Babačić et al. data (new Table S12, Figure S8A). There were 15 proteins in the Babačić et al. data with significant effects. 1 with an interaction effect (with a negative coefficient), 5 with response effects (all with positive response effects), and 9 with timeline effects (all with positive timeline coefficients). IL6 was more abundant in non-responders in our data and more abundant in responders in Babačić et al. CD244, CRTAM, CXCL11, CXCL9, GZMH, CXCL10, CXCL13 and KLRD1 were more abundant post-ICB in both data sets. Babačić et al. called SD responders in their analysis, and we decided to change this to non-responders in our validation analysis to match the approach with our data. On-treatment samples were also collected at 3 weeks while ours were collected at 6 weeks, which may help explain limited overlap between significant effects, especially treatment-associated proteins.

There are 38 significant proteins from the original analysis that are present in the Nuñez et al. data (new Table S13, Figure S8B). There were 16 proteins in the Nuñez et al. data with significant effects. 13 with interaction effects (2 with negative coefficient, 11 with positive coefficient), 8 with response effects (7 negative coefficients, 1 positive coefficient), and 6 with timeline effects (3 with negative coefficient, 3 with positive coefficient). 2 of the significant Nuñez proteins were also significant in the original analysis in the same direction (KITLG was higher in responders in both cases and CXCL9 had positive timeline effects in both cases). However, there were some proteins with positive interaction effects in the Nuñez data that had positive timeline effects in our data (CXCL9, CXCL11, CXCL10, MMP10, TNF, IFNG, CCL20, IL10RA, IL17A, CDCP1, IL10RA). There were four more significant proteins from Nuñez et al. that overlapped with our significant proteins but that didn't have the same type of effect. The structure of the data was different because there were 4 sampling time points centered around detection of Tox (pre-ICI, d7-d14, 1 month pre-Tox, and Tox), which may help explain limited overlap between our results and theirs.

80 proteins that were significant for timeline effects and more abundant on-treatment were shared between the gastric and melanoma data sets (new Table S14, Figure S8C). GO term enrichment for those 80 overlapping proteins showed that they are enriched in inflammatory processes (Fig S8d), which indicates that the upregulation of these proteins may be a common feature of systemic response in circulation to therapeutics in cancer (including immunotherapy). **There was limited overlap between proteins with other effects, which may indicate that global host response to immunotherapy may be shared across cancers, response- and non-response-associated proteins may differ from disease to disease.** CXCL16 and PAEP were both more abundant pre-treatment in both data sets. The melanoma-only timeline proteins were most enriched in other immune-related processes, especially myeloid and cytokine signaling (Figure S8E), while the gastric-only timeline proteins are most enriched in a lot of adhesion/ECM-related processes (Figure S8F). This may imply that immunotherapy elicits specifically T cell activation/proliferation related changes in circulation, while melanoma may enrich for additional immune processes and gastric may enrich for EMT-related processes."



Figures – Several figures appear slightly blurry; higher-resolution images would improve clarity.

Thank you for bringing this to our attention. We have corrected this and submitted higher resolution figures in the revised manuscript.

Introduction – Emphasise that only a few biomarkers (PD-L1, high TMB, microsatellite instability) are currently approved for clinical use.

We appreciate the comment and have taken care to mention the biomarkers approved for clinical use in the Introduction section of the revised manuscript:

“Nevertheless, only a handful of these biomarkers, such as PD-L1 abundance, TMB and microsatellite status have been approved for clinical use (Reschke et al., *J. Immunotherapy Cancer* 2021)”

Abbreviations – Ensure that each abbreviation is defined at first mention (e.g., immune-checkpoint blockade, tumour micro-environment).

We again appreciate this comment to ensure clarity in the text and have reviewed the manuscript to ensure we have now defined all abbreviations at first mention.

Page 5, paragraph 3 – Note that on-treatment decreases in serum IL-6 were also associated with better overall survival in pancreatic cancer (DOI 10.1200/JCO.21.02511).

Thank you for bringing this study to our attention. We included this citation in referencing past studies that identify a link between IL-6 abundance and patient prognosis.

Page 5, paragraphs 3–4 – Several findings—such as the link between effector-memory $CD8^+$ T cells and response, or Tregs and non-response—confirm established concepts; highlight that this dataset recapitulates these insights at the serum-protein level.

We have now cited previous studies that circulating effector-like T cells have been noted in circulation at higher frequencies in responding patients on immunotherapy treatment based on the T cell-related proteins we noticed that were more abundant in the on-treatment samples, though it should be noted that the increase in abundance in these proteins was observed generally in on-treatment samples in our results (both responders and non-responders):

“Increases in circulating effector T cell abundance have been noted in previous studies as a feature of immunotherapy response (Tooley et al., *Cell Rep. Med.* 2024 and Lao et al., *Immunobiology* 2025)”

Page 6, paragraph 2 – The sentence “Cluster 34… associated with immune-related diseases such as sepsis, influenza and gout” adds little insight; consider focusing on the most relevant findings.

We appreciate the comment. We agree that the content of this paragraph doesn’t strongly relate to the results or help contextualize them. As such, we have decided to remove the paragraph discussing the relationship between Deng et al.’s results and ours. We believe, as per the opening comments from the reviewer, removing this section improves the readability and clarity of the overall manuscript.

Page 6, paragraph 4 – Instead of “aggregate expression of transcripts from scRNA-seq,” refer to a gene signature or score corresponding to the previously identified proteins.

We have edited this statement to read “the composite score of genes corresponding to the post-ICB proteins”.

Page 6, paragraph 6 – Phrases such as “CXCL10 was expressed highly only in $CXCL10^+$ macrophages” are redundant; streamline for brevity.

We removed mention of CXCL10 expression in CXCL10+ macrophages due to redundancy.

Reviewer #2

The study presents a comprehensive plasma proteomic analysis of metastatic melanoma patients undergoing immune checkpoint blockade (ICB) therapy. Using Olink technology, the authors identify proteins in circulation that are associated with treatment response, non-response, and the time course of therapy. A major strength of the work is the longitudinal sampling of serum, allowing for dynamic assessment of immune activity. The finding that both responders and non-responders exhibit evidence of immune activation is important and raises meaningful questions about the modulatory mechanisms that determine clinical outcome. The combination of proteomic profiling with matched immune cell proportions in peripheral blood is another strong aspect of the study.

We appreciate the reviewer's comments and will expand on specific concerns below.

While the dataset is rich and the overall approach is promising, some methodological and interpretational concerns limit the strength of the conclusions. In particular, the integration of single-cell RNA-seq data from pre-treatment tumor samples with post-treatment plasma proteomics appears speculative and not fully justified. The authors correlate mRNA expression patterns from tumor-resident immune cells with circulating protein levels after ICB therapy, but without temporal or spatial concordance. No protein measurements from tumor tissue are provided to support this link. As such, these findings should be framed as hypothesis-generating rather than conclusive evidence of cellular origin.

We appreciate the comments from the reviewer. Despite this being one of the largest and most well-curated cohorts of plasma samples from patients on anti-PD1 therapy, the practicality of when it would be appropriate to collect patient biopsies (due to safety to the patient, their willingness to undergo biopsy, or overall health) limits our ability to collect time-matched biopsies with plasma samples. In addition, due to the limitations in tissue available and requirements to use some of the tissue for clinical purposes, we were unable to collect RNA-sequencing, proteomics and spatial data concurrently. Proteomics data from tissue, in particular, requires significant tissue mass to generate lysate sufficient for most proteomic approaches, and we were therefore particularly limited in our ability to generate this data.

We agree that these are all pertinent limitations to the results we present in our study and that because we could not obtain time point-matched samples with spatial resolution or with bulk proteomics, our results are primarily hypothesis-generating. We have added further comments in the Conclusion section and include mention that our findings need to be validated in the future with orthogonal datasets.

Additionally, the ligand-receptor analysis based on scRNA-seq data involves multiple assumptions, including threshold-based inferences of signaling interactions, and does not account for post-transcriptional regulation or secretion. This approach may help generate potential interaction maps, but should not be overinterpreted in the absence of functional validation. It is particularly confusing that only responders are shown with clear sender-receiver relationships, whereas non-responder-related proteins appear to lack corresponding source cell types in the visualization. This asymmetry is not sufficiently addressed.

We agree that there are limitations in the ligand-receptor interaction approach used, and have provided additional comments in the body of the text to clarify these results are hypothesis-generating and that the sender-receiver relationships are meant to support the hypotheses we introduced previously, but that future functional validation will be required to verify them. We highlighted interactions involving genes corresponding to proteins more abundant in non-responders because CellPhoneDB did not identify any interactions involving genes

corresponding to proteins more abundant in responders, and we have clarified this detail in the text. We had made specific changes in the text as below:

“We acknowledge that both the thresholding analysis (Fig. 3) and the CellPhoneDB analysis (Fig. 4) rely on multiple assumptions and do not account for post-translational modifications of RNA or secretion of proteins. The approach of thresholding expression of cytokines/chemokines and their receptors (Fig. 3) is arbitrary and as such we performed a similar analysis with lower or higher thresholds (Fig. S14). Making the threshold more promiscuous adds more intuitive interactions, some of which are seen in Fig. 3, and making it more stringent prunes interactions in a way that limit interpretation. For this reason, it is important to acknowledge that this approach has limitations and that the identification of these interactions through an alternative approach such as CellPhoneDB (Fig. 4) helps support these. Nevertheless, the results from both approaches are hypothesis-generating and meant to support biological models proposed by proteomics and scRNA-seq results presented earlier, and further functional experiments will be required to validate them.”

Moreover, the methodology used for the ligand-receptor analysis is relatively manual and includes several steps that require further clarification. The authors identify cytokines, chemokines, and their receptors among the proteins derived from the linear mixed model, and then use the CellPhoneDB database to define known ligand-receptor pairs. They compute the average expression of each ligand and receptor across major cell types in the scRNA-seq data and assess interactions by calculating the minimum of the average ligand expression in one cell type and receptor expression in the other. Interactions with a normalized expression above 0.1 are considered present. For multicomponent receptors, the lowest expressed component defines the expression value. While the approach is biologically plausible, the choice of the 0.1 threshold and the handling of receptor complexes are not sufficiently justified, and the analysis may be sensitive to these parameter choices.

We acknowledge that the choice of expression threshold for a ligand-receptor relationship to be considered is based on intuition from previous analyses, and that regardless of threshold functional experiments are required to verify the presence of these interactions in the melanoma tumor microenvironment. To address the reviewers’ concerns, we have also included a supplemental figure (new Figure S14) showing how the inferred interactions change as the threshold is modulated. Making the threshold more promiscuous adds more intuitive interactions that were also seen in Figure 3, while making the threshold more stringent prunes some interactions that might impede interpretation. As such, we will mention the alternatively thresholded figures and include acknowledgement that this measurement is hypothesis generating and can’t objectively identify sender-receiver relationships without further validation. We have included Figure S14 as a reference for readers to review and interpret based on their own scientific judgement.

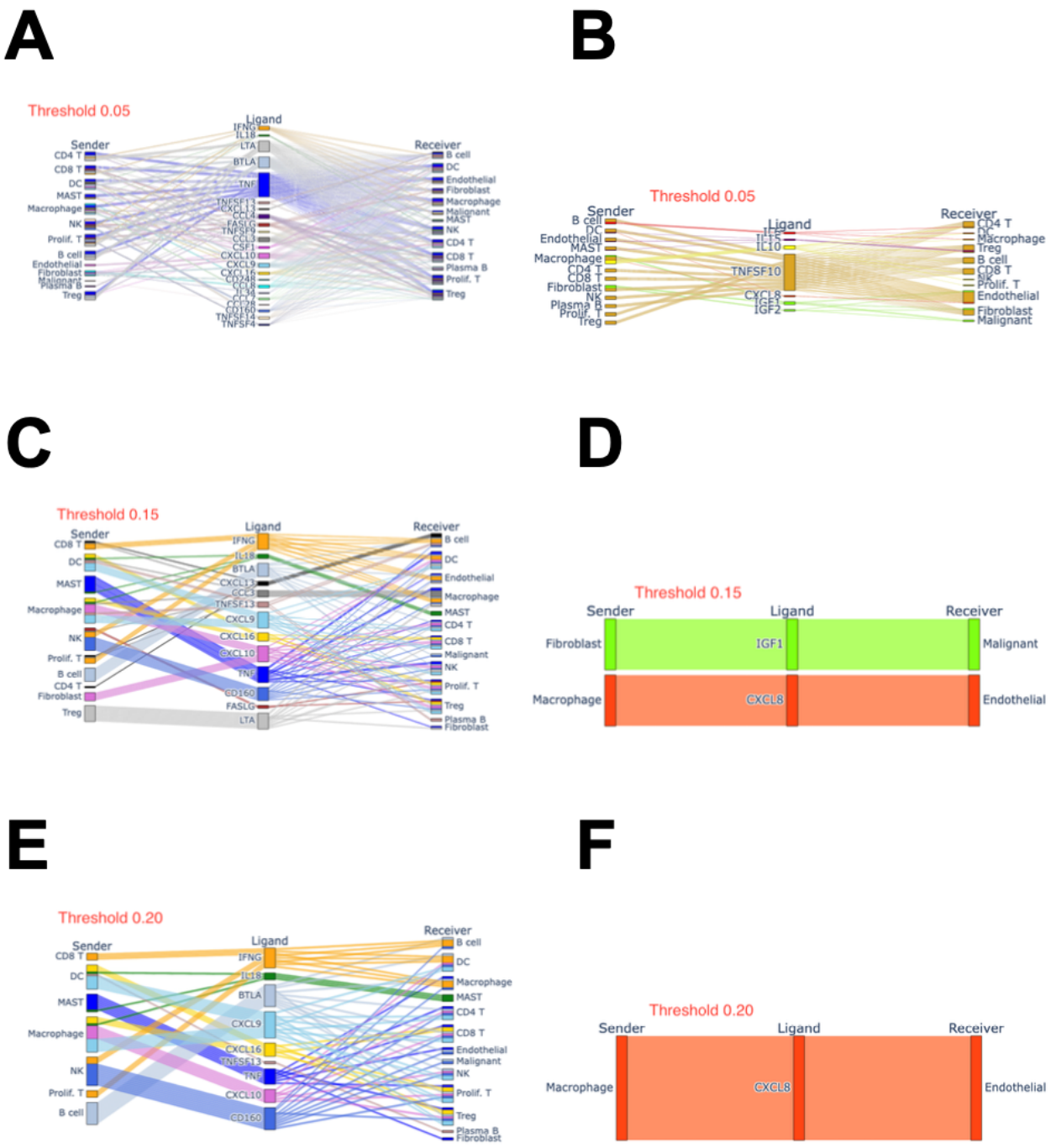

A
Threshold 0.05
Sender
Ligand
Receiver
B
Threshold 0.05
Sender
Ligand
Receiver
C
Threshold 0.15
Sender
Ligand
Receiver
D
Threshold 0.15
Sender
Ligand
Receiver
Fibroblast
IGF1
Malignant
Macrophage
CXCL8
Endothelial
E
Threshold 0.20
Sender
Ligand
Receiver
F
Threshold 0.20
Sender
Ligand
Receiver
Macrophage
CXCL8
Endothelial

Figure S14

In addition, we have included the following revised text in the manuscript:

“We acknowledge that both the thresholding analysis (Fig. 3) and the CellPhoneDB analysis (Fig. 4) rely on multiple assumptions and do not account for post-translational modifications of RNA or secretion of proteins. The approach of thresholding expression of cytokines/chemokines and their receptors (Fig. 3) is arbitrary and as such we performed a similar analysis with lower or higher thresholds (Fig. S14). Making the threshold more promiscuous adds more intuitive interactions, some of which are seen in Fig. 3, and making it more stringent prunes interactions in a way that limit interpretation. For this reason, it is important to acknowledge that this approach has limitations and that the identification of these interactions through an alternative approach such as CellPhoneDB (Fig. 4) helps support these. Nevertheless, the results from both approaches are hypothesis-generating and meant to support biological models proposed by proteomics and scRNA-seq results presented earlier, and further functional experiments will be required to validate them.”

The work is timely and relevant, particularly in the context of ongoing efforts to understand mechanisms of ICB resistance and response. The large-scale, longitudinal serum profiling is a strength, and the findings regarding shared immune activation in both responders and non-responders are conceptually important. However, certain extrapolations—especially those relying on tissue-plasma integration—are speculative and should be tempered accordingly. The manuscript could benefit from citing related recent work, e.g. Yoo et al. ("Prediction of checkpoint inhibitor immunotherapy efficacy for cancer using routine blood tests and clinical data"), which shares the goal of response stratification using peripheral data.

We agree that moving between proteomics data in the plasma and transcriptomic data in melanoma tissue is challenging and have added mention of this limitation in the Discussion section. Thank you for bringing the study from Yoo et al. to our attention. We have made mention of this in the Introduction section.

The methodology for the Olink proteomics appears sound, and the authors perform linear mixed-effects modeling with appropriate inclusion of patient ID as a random effect. However, the introduction of melanoma subtype as a covariate reveals a major limitation: the strong imbalance of cutaneous versus non-cutaneous subtypes across response groups confounds many of the initial associations. After adjusting for subtype, most response- and non-response-associated proteins lose significance, suggesting that many findings are at least partially driven by subtype differences rather than true response biology.

We acknowledge these subtype differences. We have included analyses using our lineage mixed regression models including subtype as a covariate in the manuscript (Fig. S4 to S6). The results of these are reflected in the below quoted text. We note, however, that the subtype differences are reflective of the natural history of responders and non-responders collected in our biobank. In the current treatment paradigm, all histologies are treated the same even though as a field we know the responses are not the same. Our analyses therefore also offer some insight into the distinct biology of different subtypes and provide foundation for future work in separating out histology specific features. Several groups are working on this at the level of tumor tissue, and may leverage our proteomic dataset as a resource.

“We refit the linear mixed regression models using subtype as an additional covariate and noted that most time point-associated proteins remained the same suggesting that this variable did not account for the time point-associated proteomic changes. For response-associated proteins, 8 of the 27 remained significant (CDH17, ITGB6, BMP4, CPA2, MYOC, ITGAV, LEFTY2, CD1C, HMCN2), while for non-response-associated proteins, only 2 of 58 remained significant (STC1, MIA) suggesting that there may be more distinct mechanisms of

nonresponse across histologies. The 137 proteins that were associated with subtypes included IGF1R, GDF15, TYRO3, BCAN and PODXL2 (Fig. S4, Table S7-8)."

The use of the Deng et al. plasma proteome atlas (53,026 adults) is introduced but insufficiently explained. The authors identify a similarity between their protein cluster and cluster 34 from the atlas, which is associated with influenza, sepsis, and gout. It remains unclear whether the authors are claiming a biological similarity (i.e., inflammatory mechanisms) or simply a shared protein expression pattern. This needs clarification.

We agree that the relationship between Deng et al.'s results and ours isn't very clear and have decided to remove the paragraph discussing the cross-over between our results and Deng et al.'s. We believe removing this section improves the readability and clarity of the overall manuscript.

The definition and rationale of the thresholding strategy for ligand-receptor analysis is not described in sufficient detail. The term "thresholding" is used repeatedly without explaining what thresholds were applied, how they were defined, or why they are biologically meaningful.

We have provided an explanation of the thresholding approach in the Methods section explaining the process for thresholding the expression of interactions. We acknowledge that choosing a threshold for expression to infer interactions is based largely on intuition from previous datasets and as such have produced supplemental figures (Figure S14) showing the inferred interactions at different expression thresholds to give a more complete understanding of how this choice changes the results and interpretations.

The use of linear mixed models is appropriate for handling repeated measurements and patient-specific effects. The model design, with time, response, and interaction terms, is well-chosen. However, the confounding introduced by melanoma subtype is a key concern. The authors acknowledge this but appear to understate its impact. The claim that subtype had "limited influence" on timepoint-associated proteins is only partially reassuring, given how many response-related findings were lost after including subtype as a covariate.

We appreciate the comment from the reviewer and acknowledge this limitation. As in our previous comment, we have included analyses including subtype as a covariate, and have additionally included a statement in our discussion to highlight this limitation. We note again that the subtype differences are reflective of the natural history of responders and non-responders collected in our biobank. In the current treatment paradigm, all histologies are treated the same even though as a field we know the responses are not the same. Our analyses therefore also offer some insight into the distinct biology of different subtypes and provide foundation for future work in separating out histology specific features.

As is addressed below, we found that fitting linear mixed effects models on cutaneous-only samples results in slopes for response that are correlated with the original models (in which all subtypes are included), so it appears as though the cutaneous-only models are still identifying similar relationships between response and protein abundance, but that they are just not reaching the level of significance after multiple hypothesis correction.

"We also refit the linear mixed regression models with only cutaneous samples, and interestingly no proteins had significant response effects (**Table S9**). We do note, however, that the proteins which had significant response effects in the original models have highly correlated slopes with the same proteins in the cutaneous-only models (the cutaneous-only model response slopes tended to be lower in magnitude) (**Fig. S5a**), indicating that there is likely some association between the abundances of these proteins and response that fails to reach the level of significance. When comparing the skewness of these protein distributions in responder samples in the overall

setting to the cutaneous-only setting, we note that these are similar across response proteins (**Fig. S5b**). However, we noticed that the skewness in the overall setting was higher than the skewness in the cutaneous-only (**Fig. S5c**). In other words, an arm of the protein distribution of non-responder samples that differentiates it from responder samples is lost when non-cutaneous samples are removed from the models."

We have additionally added the below text in the discussion to acknowledge the limitation of the subtype imbalance: "Our cohort is representative of the natural subtype distribution of melanoma responders and non-responders, and as such, melanoma subtypes are not evenly distributed between the two groups. However, in the current treatment paradigm, all histologies are treated the same even though as a field we know the responses are not the same. This therefore represents an opportunity to understand underlying biological differences between the subtypes."

The authors should clearly state the limitations of correlating pre-treatment tumor transcriptomics with post-treatment plasma proteomics.

In the Discussion section, we have mentioned this limitation and that our results are meant for hypothesis-generation, requiring further validation.

We have included the following text in the Discussion section in reference to this:

"Making comparisons between plasma and melanoma tissue and between proteomics and scRNA-seq data is also challenging as post-transcriptional regulation and protein translation may limit the translation of signals from one data modality to another. As such, we stress that our plasma-to-tissue comparisons are speculative and require further validation. Additionally, time point-matched tumor scRNA-seq samples were not available, which further limits our ability to validate signals from the proteomics data set in the scRNA-seq data, especially genes corresponding to proteins that are more abundant on treatment. Because of the limitations we encountered with sample collection, we emphasize that the results we present are primarily hypothesis-generating."

The ligand-receptor thresholding strategy should be better explained, including the cutoffs used and rationale for inferring signaling.

As mentioned above, we acknowledge that choosing a threshold to cut off ligand-receptor expression to infer an interaction is largely based on intuition and not systematic, and as such we have produced analogous figures (Figure S14) using different expression thresholds to demonstrate how these results change as we modulate the expression threshold.

The subtype imbalance should be acknowledged more directly as a major limitation, and response associations should be interpreted cautiously in this light.

We have added the below text in the discussion to acknowledge the limitation of the subtype imbalance: "Our cohort is representative of the natural subtype distribution of melanoma responders and non-responders, and as such, melanoma subtypes are not evenly distributed between the two groups. However, in the current treatment paradigm, all histologies are treated the same even though as a field we know the responses are not the same. This therefore represents an opportunity to understand underlying biological differences between the subtypes."

Figures need clearer legends, especially regarding abbreviations and what is being shown (e.g., sender-receiver relationships, response group definitions).

We have reviewed and attempted to clarify the figure legends in ambiguous cases, and have taken special care to better define abbreviations and plots related to sender-receiver relationships. We hope that this aids with ease of understanding of the manuscript.

Consider including tumor proteomics or matched tissue data (even retrospectively) to better support the tissue origin hypotheses.

Unfortunately we do not currently have tumor proteomics data available to us. Matched proteomics will be difficult to generate because we have limited tissue in the biobank and it is stored as FFPE, and we used all the tissue available to us for scRNA-seq. This is a fundamental limitation of any well curated patient cohort. As discussed previously, the practicality of when it would be appropriate to collect patient biopsies (due to safety to the patient, their willingness to undergo biopsy, or overall health) limits our ability to collect time-matched biopsies with plasma samples and sufficient material for proteomics studies, which require more tumor material than does RNA-seq.

The use of the Deng et al. reference atlas should be contextualized more clearly—what does similarity to cluster 34 actually imply?

As mentioned above, we have chosen to remove the paragraph discussing Deng et al.'s results. We believe removing this section improves the readability and clarity of the overall manuscript.

The manuscript would benefit from editing for clarity, particularly in the figure legends and transitions between sections. The first results section titled "Plasma proteomic profiling performed on metastatic melanoma cohort undergoing ICB" is misleading—it does not present proteomic data but describes cohort characteristics and survival analyses. The narrative jumps quickly between Olink, tissue, and single-cell data, sometimes without clear rationale or structure.

We have changed the heading of the first results subsection to better reflect the results we are presenting. We have made an effort to include more rationale at the beginning of each subsection in the results section to aid with transition from the previous section. As mentioned above, we have also made an effort to edit the figure legends to improve their readability.

Overall, references are appropriate but additional recent work (e.g. Yoo et al.) should be cited where relevant.

Thank you for bringing the Yoo et al. paper to us. We have cited this study in the Introduction section:

"Additionally, taking biological measurements from patient blood has shown promise in predicting immunotherapy response across different cancers, including melanoma (Yoo et al., *Nature Medicine* 2025)"

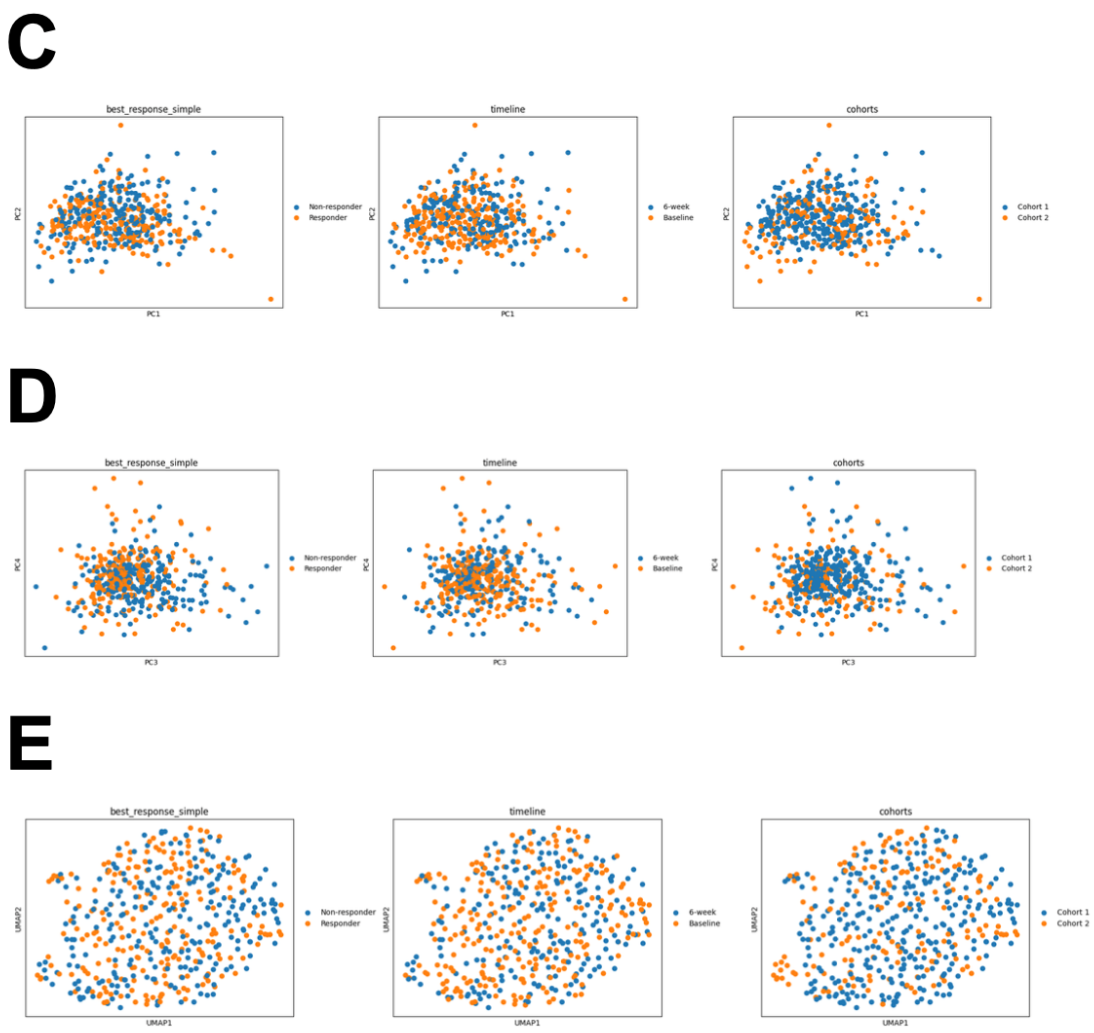
Reviewer #3

The authors present high dimensional proteomic profiling of plasma from a large cohort of patients with metastatic melanoma treated with systemic therapies. With validation of candidate targets within single cell and bulk transcriptomics. The study has a strong clinical rationale, statistical approach and unique longitudinal sampling provide robust analysis. Overall, a very strong manuscript and there for minor suggested added below:

We thank the reviewer for their response and will address specific concerns below.

1) Can the authors provide unsupervised clustering of the patient's proteomic profiles (PCA or UMAP like) to explore before/after batch effect, patient grouping based on response and treatment timepoint.

We provide PCA and UMAPs below labeled by different metadata variables below after batch correction (also in Figure S2). The raw data prior to batch correction was unfortunately not provided to us by the Olink team; however, it was noted by Olink that batch correction is always required for runs on separate days. In this case, our samples were run in distinct batches by Olink across the different protein panels, and we included overlapping samples in addition to the standard plate normalization in order to perform bridge normalization as described in the methods. Figure S2 is shown below:



2) It's stated the multi-omic data is available in supplementary tables. Can the authors confirm the data availability of the remaining omics, COsMX, RNAseq etc.

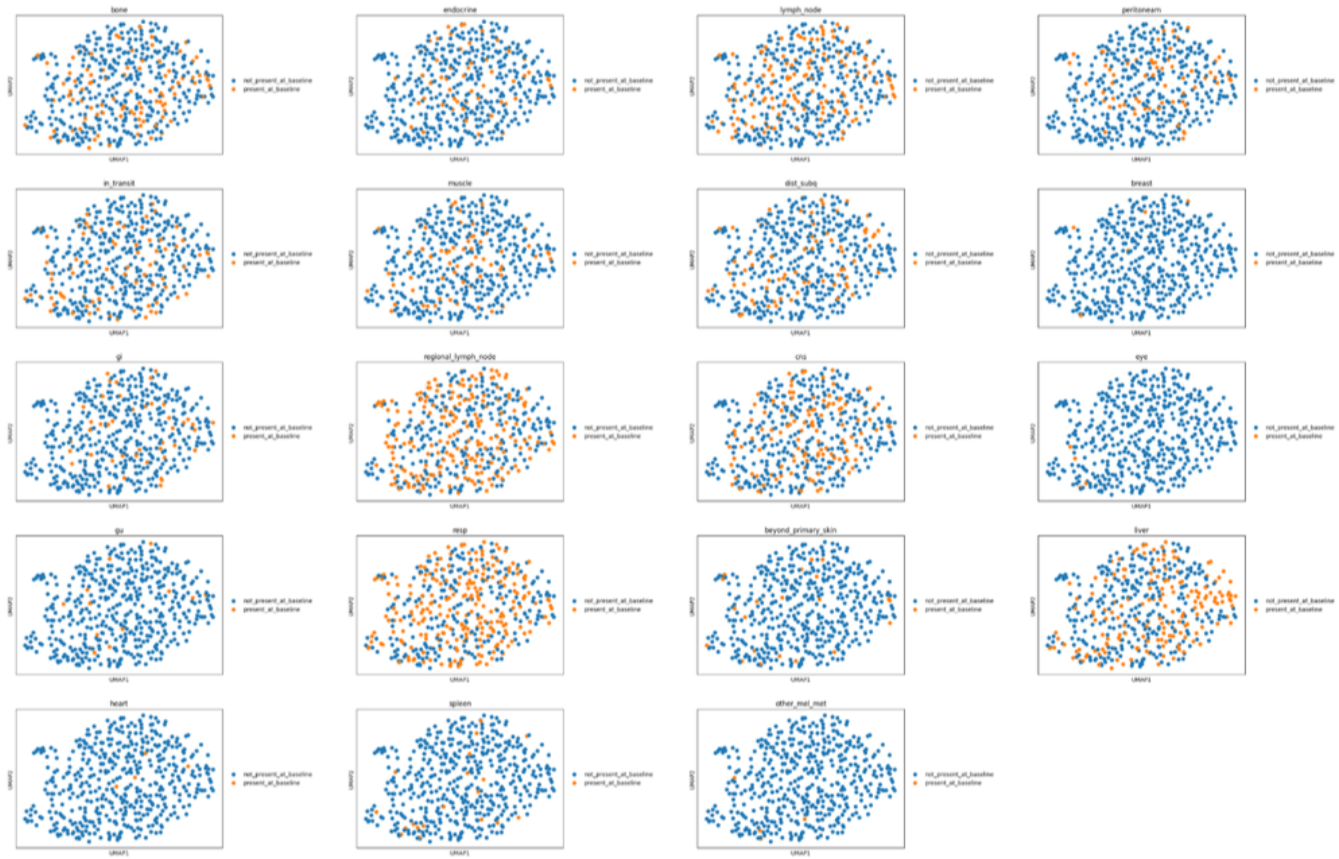
We have made an effort to make sure all data used in this manuscript will be available for further analysis and for reproducibility. Plasma proteomic and PBMC mass cytometry data are available in Table S3 and Table S4, respectively. The CosMx data is published and accessible online in Love et al. (https://www.science.org/doi/10.1126/sciadv.adm8206). Bulk RNA-seq and single-cell RNA-seq samples are available (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE295942).

3) Given the role of melanoma subtype on the results of the study, did the authors have access to tumour mutation burden or broader melanoma molecular subtyping for comparison?

Most patients did not have TMB testing, and we have included a supplemental table (Table S2) showing TMB for the patients who did. Given the small number of patients who did have TMB testing (n=8), we don't believe we're sufficiently powered to make comparisons on the basis of melanoma subtype of protein abundances. In addition, we have included BRAF mutational status (Fig. 1B) and additional information on mutational profiles where available (Table S5).

4) Did the proteomic profiles associate or cluster based on the site of disease, liver lung brain specific profiles in the plasma?

We did not notice any apparent clustering of proteomic data on the basis of metastatic site and have included a supplemental figure (Figure S3) showing the UMAP embedding of proteomic samples and colored by metastasis at a given site. Figure S3 is shown below. Table S5 was also included for sites of metastasis when available.



5) Given that some samples are taken on-treatment. Can the authors confirm any role competitive binding or masking of PD-1/CTLA-4 might have on the onlink results for these antigens?

We appreciate this important question from the reviewer. We have explored this extensively with the Olink team; the antibodies being used in the Olink assay are proprietary, but we have confirmed they bind to distinct epitopes compared to clinic anti-PD1 antibodies (for the PD1 assay) and clinical anti-CTLA4 antibodies (CTLA4 assay). The antibodies used in the assay itself bind different bins and the Olink team has shown, in-house, that they do not compete with clinical antibodies. As such, we are able to detect endogenous levels of PD1 and CTLA4 without masking.

Reviewer #4

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for taking the time to review the manuscript.

References

1. Babačić, H., Lehtiö, J., Pico de Coaña, Y., Pernemalm, M. & Eriksson, H. In-depth plasma proteomics reveals increase in circulating PD-1 during anti-PD-1 immunotherapy in patients with metastatic cutaneous melanoma. *J Immunother Cancer* **8**, (2020).
2. Nuñez, N. G. *et al.* Immune signatures predict development of autoimmune toxicity in patients with cancer treated with immune checkpoint inhibitors. *Med* **4**, 113–129.e7 (2023).

Responses to reviewer comments

Reviewer comments in black | Responses in blue

Reviewer #1:

All comments from the initial review have been addressed satisfactorily. I have one minor remaining point for clarification.

As results from plasma proteomics can be influenced by pre-analytical variability, it would be helpful if the Methods section could include a bit more detail on blood collection procedures (e.g., type of venipuncture device, needle gauge, use of a butterfly needle, tourniquet technique, patient position, and collection tubes) and sample processing (e.g., time from collection to processing and transport method). In addition, clarification regarding whether plasma or serum was used, and whether all samples were collected and processed in a consistent manner, would further strengthen the manuscript.

Thank you for bringing this to our attention, and we agree that this information would add helpful information to the methods. We added new information regarding the blood collection procedure to the Methods sections under the **Plasma, PBMC and tissue pre-processing** subsection so that this subsection now reads, in its entirety:

“All patients provided written informed consent to allow the collection of blood and tumor tissue for research (Dana Farber/Harvard Cancer Center protocol identifier 11-181). Samples were collected and processed at Massachusetts General Hospital. Blood was collected using a butterfly needle while the patient was seated. A tourniquet was applied 3-4 inches above the venipuncture side for a maximum of 1 minute. 10-30cc whole blood was collected in Sodium Citrate Mononuclear Cell Preparation (CPT) Tubes at room temperature. Plasma was isolated according to the BD manufacturer protocol within two hours of collection and stored at -80C. Plasma was transferred to Olink Proteomics for testing. All samples were collected and processed in a consistent manner.

Tissue was collected fresh in phosphate-buffered saline. For single-cell RNA-sequencing, fresh melanoma tumors were dissociated and scRNA-seq libraries were prepared using the 10x Genomics chromium controller and reagent kits as described below. Bulk RNA sequencing was performed on tissue preserved in FFPE blocks.

Sample/patient identification numbers were not known to anyone outside of the research group.”

Reviewer #2:

This revised manuscript presents a large and carefully curated longitudinal plasma proteomics dataset from patients with metastatic melanoma treated with immune checkpoint blockade. The cohort size, clinical annotation, and use of linear mixed-effects models for longitudinal analysis remain major strengths. The additional analyses and clarifications included in this revision address several points raised in the initial review; however, several of the core interpretational concerns—particularly regarding subtype confounding and plasma–tissue integration—have been only partially addressed and remain insufficiently resolved.

Noteworthy results and potential significance

The identification of robust treatment-associated plasma proteomic changes, including immune activation signatures observed in both responders and non-responders, is a notable and potentially valuable contribution. The longitudinal design and integration with peripheral immune profiling make this dataset a useful resource for the field. With appropriate framing, the study could help inform future efforts to develop blood-based biomarkers for immunotherapy monitoring.

We appreciate the positive comments from the reviewer on the value of the dataset.

Support for conclusions and remaining concerns

Despite improvements, several conclusions remain insufficiently supported by the presented evidence and require further clarification or reframing:

1) Melanoma subtype as a major confounder of response-associated signals. The revised analyses confirm that melanoma subtype is the only clinical variable significantly associated with response, and that most response- and non-response-associated proteins lose significance after adjusting for subtype or restricting analyses to cutaneous melanoma. While these results are now more transparently presented, the manuscript still tends to frame certain findings as response-associated rather than largely subtype-driven. This distinction should be emphasized more consistently throughout the Results, Discussion, and figure legends to avoid overinterpretation.

We appreciate the importance of making this distinction, and as such we have attempted to further clarify upfront when describing the cohort that differences in response may also be related to subtype so that readers can have this in mind as they continue to assess the results. We have included the following statement:

“We acknowledge that in downstream analysis of the proteomic data, when testing for a relationship for response, particular proteins may be indeed associated with a particular melanoma subtype as well. As such, it may be difficult to disentangle the contribution of immunotherapy response status and histological subtype may contribute to proteomic abundance, since these two covariates are inherently related.”

2) Interpretation of plasma–tissue integration using scRNA-seq. The integration of pre-treatment tumor scRNA-seq data with post-treatment plasma proteomics remains conceptually problematic as currently framed. While the authors now acknowledge limitations, the manuscript still implies a closer correspondence between plasma protein abundance and the tumor microenvironment than is supported by the data. Tumor scRNA-seq provides transcript-level information from a distinct spatial and temporal compartment and cannot establish the cellular origin, secretion, or plasma contribution of measured proteins. These analyses should be framed

consistently as hypothesis-generating contextualization rather than evidence of plasma protein origin or tumor-derived signaling.

We agree that it is important to temper the conclusions related to the analysis of the single-cell RNA-seq data in relation to the results of the proteomics analysis. As such, we have included a statement when introducing the single-cell RNA-seq data that makes the intention of this analysis clear upfront:

“We acknowledge that single-cell RNA-seq samples are not time point-matched with the plasma proteomics samples, and that this analysis is meant to generate hypotheses with regard to potential tumor-derived origins of circulating proteins. The correspondence for transcript-level and proteomic-level measurements is imperfect especially given that the sources of these two data modalities are from different compartments of the body, further emphasizing that the analysis of genes corresponding to proteins of interest in circulation is purely hypothesis-generating. Further, time point-conserved collection of tumor biopsy and plasma was not practical or safe for patients in this cohort.”

3) Correlation of plasma protein modules with immune cell populations. The correlation analyses between plasma protein modules and PBMC immune cell proportions require clearer explanation and stronger discussion of confounding. Both measurements derive from peripheral blood and may reflect shared systemic drivers (e.g. tumor burden or generalized inflammation). In addition, CyTOF data represent relative cell proportions rather than absolute cell numbers, and wording should be adjusted accordingly.

We thank the reviewer for noting this. To address this concern we have included the following statement when introducing this data:

“Importantly, the mass cytometry captures relative cell proportions for immune cells rather than absolute cell count.”

Further, when discussing the correlations of the immune cell proportions with the proteomic modules, we have included the following statement:

“As both of these data derive from peripheral blood, apparent correlations between immune cell proportions and proteins in circulation may be due system drivers of both. As such, it is important to acknowledge that relationships between immune cell proportions and proteins does not imply that one inherently causes the other but instead that these may be co-regulated changes reflective of systemic response to immunotherapy.” (Note that because we moved much of the validation work to the Supplemental Information file, this additional statement now appears in the Supplemental Information file rather than the main text).

4) Toxicity-related interpretations. The focus on pancreatitis as an immune-related adverse event is insufficiently justified. It is unclear whether other irAEs were observed in the cohort and, if so, why they were not analyzed. Moreover, interpretations suggesting tissue or organ origin of circulating proteins in the context of toxicity are speculative in the absence of orthogonal validation and should be clearly labeled as such. A summary of all observed irAEs (type, grade, frequency) should be provided, at least in the **Supplementary Information**.

We appreciate the comment and the importance of further analysis on the toxicities within the patient cohort. The remainder of the toxicity data from this cohort is part of another follow-on study that is unpublished at present. These data will be published in the future and we felt were beyond the scope of the current

manuscript which we have attempted to cut down further in this revision. We have included an additional supplemental table (**Table S17**), which includes annotations of which patient IDs in the cohort corresponded to those with pancreatitis as appeared in the analysis in **Figure 5**.

Methodology, clarity, and presentation

The methodological framework is generally sound and described in sufficient detail for reproduction. However, several aspects would benefit from improved clarity:

1) The definition and calculation of composite or aggregate protein scores should be explicitly described.

As mentioned in the Methods section, the score_genes function from the Scanpy package (v1.9.8) was used to derive aggregate protein scores. We have also updated the text to include reference to this function when relevant.

2) Key heatmaps and clustering figures lack sufficient annotation of protein identities, limiting interpretability.

We understand that some heatmaps such as Figure 2C and Figure 5E do not have all of the protein labeled. For Figure 2C, there were too many proteins to make it feasible to visualize on a single plot, and as such we included color bars at the top of each heatmap to indicate which effect category those proteins fall into, and we hope that referring to Table S6 will be sufficient for referencing which proteins appear in each group. For Figure 5E, only a subset of the proteins appearing in Figure 5D had a significant association with pancreatitis incidence and as such we chose to focus on these so that the text could be sufficiently large to read for the reader.

3) Some sections remain overly long and descriptive, and further streamlining—particularly of validation and integration sections—would improve readability and focus.

We agree that the description of the validation of our results is a bit lengthy. For this reason, we have moved the description of the results to a **Supplementary Information** file and included brief references to this file in the main text where appropriate. We hope this makes the main text more concise and improves readability

The references in the main text to the **Supplementary Information** file appear as follows:

“We refit linear mixed effects models using alternative approaches (using subtype as an additional covariate, using cutaneous-only melanoma samples and categorizing stable-disease patients as responders) to learn about the effects these factors have on the outcomes (Fig. S4-6, Table S7-10). These results are described in the **Supplementary Information** file.”

“We performed validation of our proteomics results using external orthogonal proteomics datasets (Fig. S8, Table S12-14) and the results are described in the **Supplementary Information** file. Additionally we leveraged the circulating immune proportions quantifications available to us to validate our results (Fig. S9, Table S15) and these results also appear in the **Supplementary Information** file.”

Overall assessment

In summary, the dataset and analytical effort are strong, and the manuscript has improved in response to prior feedback. However, further revision is needed to ensure that conclusions are proportionate to the evidence, particularly with respect to response associations, plasma–tissue integration, and toxicity-related interpretations. Addressing these issues primarily through clearer framing, more cautious language, and improved transparency would substantially strengthen the manuscript.

We thank the reviewer for their time and constructive feedback on the manuscript. We feel the revision is now a much stronger and clearer manuscript, and hope we have sufficiently addressed the reviewers comments.

Reviewer #3 (Remarks to the Author):

Thank you for performing the additional analysis requested. This has satisfied my comments.

We thank the reviewer for their time in reviewing this manuscript.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for their time in reviewing this manuscript.

Responses to reviewer comments

Reviewer comments in black | Responses in blue

Reviewer #1:

The authors have addressed my last comment in a very satisfactory manner.

We thank the reviewer for their time in reviewing this manuscript.

Reviewer #2:

The manuscript has improved following revision, and the authors have addressed the main concerns raised in the previous review round.

We thank the reviewer for their time in reviewing this manuscript.

I have only a few remaining minor points:

1. Statistical terminology (“correlation”)

The term “correlation” is used in contexts where group comparison tests (e.g., Wilcoxon rank-sum or Fisher’s exact test, as in Table 1) are applied. This is not statistically precise and may be misleading. I recommend revising this wording accordingly (e.g., “differences between groups” or “association assessed by group comparison”).

We agree that these cases can benefit from more precise language, and as such we have updated the language to reflect “association” on the basis of a given group comparison instead of “correlation” where appropriate.

2. Figure readability

Figure 1C: The x-axis labeling appears truncated (e.g., “200” is cut off).
Figure 2E: Axis labels/values are difficult to read due to overlap.
Figure 4B Similar readability issues as in Figure 2E.

Thank you for bringing these items to our attention. We have edited the axes and labels of Figures 1C, 2E and 4B as well as Figure 3C to improve readability.

These should be adjusted to improve clarity.

With these minor revisions, the manuscript will be suitable for publication.

Reviewer #4:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

We thank the reviewer for their time in reviewing this manuscript.