

Predicting the response to immunotherapy from gene expression data in HER2-negative breast cancer

Corresponding Author: Professor Caterina La Porta

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 the manuscript, "Predicting the response to immunotherapy from gene expression data in HER2-negative breast cancer" by La Porta et al., the authors utilize data from two existing cohorts (I-SPY2 and NCT03366844) to evaluate the ability of the ARIADNE algorithm and cytokine score to predict response to immunotherapy in patients with HER2- breast cancer. The I-SPY2 cohort included patients with both TNBC and ER positive, HER2- BC treated with immunotherapy in addition to chemotherapy compared to chemotherapy alone, whereas the NCT cohort included patients with TNBC treated with pembrolizumab and radiotherapy. The authors conclude that a low ARIADNE score and high cytokine score is associated with higher rates of pCR in patients who receive immunotherapy plus chemotherapy but not chemotherapy alone in patients with HER2- BC but this was not statistically significant in the subgroup of patients with TNBC alone. Additionally, single cell RNAseq similarly demonstrated that a low ARIADNE score and a high cytokine score in CAF but low cytokine score in cancer cells correlated with pCR. Finally, the authors demonstrated that patients responding to immunotherapy had a much larger fraction of exhausted T-cells than non-responders and the fraction increased during the course of the therapy.

This is a very interesting study evaluating an important clinical question that is, identifying markers of response to immunotherapy. This will be of significant interest to those in the breast cancer field given the widespread acceptance of immunotherapy as first line therapy for patients with early-stage TNBC. However, given that the authors utilized two different datasets that not only included different cohorts of patients but also different neoadjuvant therapies, it is difficult to keep track throughout the manuscript if it is appropriate to compare results between the groups. It would be helpful if the authors could be very clear which patient cohort was utilized for each analysis. Additionally, when the authors refer to patients with HER2- BC, it is not clear whether this is meant to be inclusive of TNBC or to contrast with TNBC and only indicate patients with hormone receptor positive, HER2- BC. This needs to be clarified throughout.

For specific comments and corrections of typos and grammatical errors (verb tense, etc.), please refer to the tracked document, including corrections in the figure legends.

Reviewer #2

(Remarks to the Author)

The paper describes how gene expression-based scores ARIADNE can predict responses to immunotherapy in HER2-breast cancer, and explores how this relates to cytokine scores, PD-L1 pathway and other immunological genomic scores.

It is difficult to assess the significance of the presented results, because some key information is either not included or unclear.

Major comments:

1. How were the high and low thresholds ((0,75) and (0.87,1)) determined for ARIADNE? Could the authors also explain why all TNBC iSPY samples (n=364) were included and not just the ones reported in this study (immunotherapy and chemotherapy arms)? Why were the thresholds established only on TNBC and not based on all HER2- samples included in this study?
2. Could the authors please report the number of samples per group for each analysis? For example, if (0.87,1) quartile

range is representative for the included samples, then would it only be around 9 patients (of 69) in the high-risk group for pembro arm?

3. Given the statements in the discussion about the potential combination of ARIADNE and cytokine scores for prediction of immunotherapy responses, the authors should explore the relationship between ARIADNE and cytokine scores more thoroughly. What is the correlation of the two features? What does the multivariate logistic regression analysis show?

4. It would also be helpful to show the correlation between each cytokine expression, and not just an aggregate average expression score. Likewise, for the PD-L1 pathway, could the authors show correlations for gene expression for high and low groups separately (including significance), for example as heatmaps, and not just differences in correlation?

5. For the PD-L1 pathway, a number of the included genes are also part of the PI3K/AKT and MAPK pathways, important for cancer cell proliferation. Could there be an independent explanation for differences in gene expression correlation, not related to the regulation of PD-L1 expression?

6. Figure 4e – it would help to visualise this as boxplots as well (in addition to boxplots Fig 4b-d which show individual cell scores rather than aggregate by patient).

7. The link between the exhausted T-cell analysis and the rest of the paper is not obvious. At a minimum, it would be good to investigate how their fraction relates to patients with low and high-risk ARIADNE scores.

8. Please mention in the abstract and discussion that ARIADNE identifies epithelial-mesenchymal phenotypes.

Minor comments:

1. First sentence in the second last introduction paragraph is missing a comma after chemotherapy.

2. Please add the relevant statistics for Figure 4f-h to support the results “The ARIADNE score obtained from cancer cells is decreasing in time for patients not responding to immunotherapy while it remains roughly constant for responders (Fig. 4f). The cytokine score decreases for CAFs (Fig. 4g) and increases for cancer cells (Fig. 4h). after immunotherapy (T1).”

3. Fig S2 – was multiple testing correction performed?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors present interesting data about ARIADNE and a cytokine signature that is associated with pCR to immunotherapy in patients with HER- tumors, including TNBC. They have addressed my request for clarifications in their revised manuscript. However, I would ask that they review their manuscript thoroughly for typos and to utilize the correct verb tense when reporting Methods, Results, and Statistical Analyses. Most should be past tense, but they utilize present tense verbs throughout. Also, minor annoyances like listing the abbreviation for pembrolizumab later in the manuscript and not at first mention but then not using the abbreviated term anywhere in the manuscript should be corrected.

Overall, based on the content, I think this manuscript is appropriate for publication in Communications Medicine and will add value to the field.

Reviewer #2

(Remarks to the Author)

The authors have sufficiently addressed all of my requests.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We thank the reviewers for their constructive suggestions. We have now revised the manuscript accordingly and you can find following a point-by point response to their issues.

Reviewer #1

Reviewer: In the manuscript, “Predicting the response to immunotherapy from gene expression data in HER2-negative breast cancer” by La Porta et al., the authors utilize data from two existing cohorts (I-SPY2 and NCT03366844) to evaluate the ability of the ARIADNE algorithm and cytokine score to predict response to immunotherapy in patients with HER2- breast cancer. The I-SPY2 cohort included patients with both TNBC and ER positive, HER2- BC treated with immunotherapy in addition to chemotherapy compared to chemotherapy alone, whereas the NCT cohort included patients with TNBC treated with pembrolizumab and radiotherapy. The authors conclude that a low ARIADNE score and high cytokine score is associated with higher rates of pCR in patients who receive immunotherapy plus chemotherapy but not chemotherapy alone in patients with HER2- BC but this was not statistically significant in the subgroup of patients with TNBC alone. Additionally, single cell RNAseq similarly demonstrated that a low ARIADNE score and a high cytokine score in CAF but low cytokine score in cancer cells correlated with pCR. Finally, the authors demonstrated that patients responding to immunotherapy had a much larger fraction of exhausted T-cells than non-responders and the fraction increased during the course of the therapy. This is a very interesting study evaluating an important clinical question that is, identifying markers of response to immunotherapy. This will be of significant interest to those in the breast cancer field given the widespread acceptance of immunotherapy as first line therapy for patients with early-stage TNBC.

Authors: We thank the referee for her/his positive feedback. We really appreciate that she/he understands the goal of our paper and the clinical relevance.

Reviewer: However, given that the authors utilized two different datasets that not only included different cohorts of patients but also different neoadjuvant therapies, it is difficult to keep track throughout the manuscript if it is appropriate to compare results between the groups. It would be helpful if the authors could be very clear which patient cohort was utilized for each analysis.

Authors: We have now cited the two datasets as DATA1 or DATA2 in the methods section and in results and captions of figures, to be clearer on which cohort was used for each analysis, as suggested by the referee (see new version of methods and results sections and captions of all the figures).

Reviewer: Additionally, when the authors refer to patients with HER2- BC, it is not clear whether this is meant to be inclusive of TNBC or to contrast with TNBC and only indicate patients with hormone receptor positive, HER2- BC. This needs to be clarified throughout.

Answer: HER2- BC also includes TNBC. This information was already provided in the results section at page 5. However, following the suggestion of the referee, we have now reported the same information in the methods section too (see page 3 of the new version of the manuscript).

Reviewer: For specific comments and corrections of typos and grammatical errors (verb tense, etc.), please refer to the tracked document, including corrections in the figure legends.

Authors: We changed the manuscript accordingly, responding to all the queries.

Reviewer: The paper describes how gene expression-based scores ARIADNE can predict responses to immunotherapy in HER2- breast cancer, and explores how this relates to cytokine scores, PD-L1 pathway and other immunological genomic scores. It is difficult to assess the significance of the presented results, because some key information is either not included or unclear.

Authors: We answered to all the queries made by the referee, and we thank him/her for her/his comments. Overall, the paper improved considerably after the changes.

Major comments:

Reviewer: How were the high and low thresholds ((0,75) and (0.87,1)) determined for ARIADNE?

Authors: First of all, we apologize for a typo that is now corrected in the revised version: The groups are defined as: low [0,75] and high (0.75, 1]. In the revised manuscript, we have now described better how we determined the thresholds (see page 4; *“We then defined ARIADNE groups by sorting and splitting the dataset into two risk groups (low and high) based on quantile discretization, with intervals [0, 0.75] and (0.75, 1], respectively. In other words, out of $n = 364$ TNBC samples, 75% (263) are assigned to low and 25% (91) are assigned to high. This process defines a score threshold that separates the two risk groups. We then computed the ARIADNE score [20] and assign risk group also for the other HER2- samples (not TNBC) present in the dataset.”*).

Reviewer: Could the authors also explain why all TNBC iSPY samples ($n=364$) were included and not just the ones reported in this study (immunotherapy and chemotherapy arms)? Why were the thresholds established only on TNBC and not based on all HER2-samples included in this study?

Authors: We included all the TNBC iSPY samples, because ARIADNE was originally defined for TNBC. Notice that the therapy is not relevant to define the threshold because ARIADNE is applied to biopsies before treatment and pCR information is not used to define the score. We mention this point in the text at page 4: *“Since the ARIADNE was originally defined with respect to TNBC gene expression data [20], we considered all the TNBC samples in the dataset ($n = 364$) and computed the ARIADNE landscape and scores as explained in [20].”*

Reviewer: Could the authors please report the number of samples per group for each analysis? For example, if (0.87,1) quartile range is representative for the included samples, then would it only be around 9 patients (of 69) in the high-risk group for pembro arm?

Authors: We have now reported the number of samples in each group in the revised manuscript at page 4. Notice that the quartile score boundaries were defined by considering all the 364 TNBC samples. Hence, the high-risk group in the HER2- pembro arm is composed by 35 patients, which is more than 25%. Now we have explained better the process at page 4: *“.....out of $n = 364$ TNBC samples, 75% (263) are assigned to low and 25% (91) are assigned to high. This process defines a score threshold that separates the two risk groups. We then computed the ARIADNE score [20] and assign risk group also for the other HER2- samples (not TNBC) present in the dataset. Out of the $n = 69$ HER2- samples receiving pembrolizumab (29 TNBC and 40 hormone-positive), 35 were classified as high and 34 as low.”*)

Reviewer: Given the statements in the discussion about the potential combination of ARIADNE and cytokine scores for prediction of immunotherapy responses, the authors should explore the relationship between ARIADNE and cytokine scores more thoroughly. What is the correlation of the two features? What does the multivariate logistic regression analysis show?

Authors: The correlation between the two features was already reported in figure 2c. We now perform multivariate logistic regression, as suggested by the referee. The result is reported in Table S1 and shows that ARIADNE and cytokine scores are both predictive of pCR in an independent way. We confirmed this by computing the log-likelihood ratio between a logistic regression with the ARIADNE score alone and when including also the cytokine score. The ratio is larger than one and statistically significant, indicating that the including the cytokine score provides improved predictions.

We discuss these results at page 7: *“We next investigated if the cytokine score has a predictive value for the response to immunotherapy that is independent from the ARIADNE score. To facilitate the comparison between the two scores, we performed a multivariate logistic regression considering standardized versions of the ARIADNE and cytokine scores, as described in the methods section. We found that both scores are significantly associated with pCR with similar effects: ARIADNE score has an OR of 0.5 (0.3-0.9 95% CI, $p < 0.05$) and the cytokine score an OR 2.0 of (CI 1.1-3.7 95% CI, $p < 0.05$) for each standardized score point (Table S1). As a further test, we computed the log-likelihood ratio L_r between a logistic regression with the ARIADNE score alone and when including also the cytokine score ($L_r = 6$, $p < 0.05$), indicating that the cytokine score improves the predicting abilities of ARIADNE.”*

Reviewer: It would also be helpful to show the correlation between each cytokine expression, and not just an aggregate average expression score. Likewise, for the PD-L1 pathway, could the authors show correlations for gene expression for high and low groups

separately (including significance), for example as heatmaps, and not just differences in correlation?

Authors: We added a supplementary figure with the association between the individual cytokines with respect to the ARIADNE groups as requested (Figure S3). No significant association was found. However, considering the complex network between cytokines and other immune mediators, the analysis of individual cytokines might be misleading. In a recent paper published by our group (Garrone O. ESMO Open, 2025) supported by prior literature (Kartikasari aer et al: Front. Oncol. 2021), we have demonstrated the value of cumulative cytokine profiling for a more accurate capture of the immune landscape and its functional implications.

We now commented in the revised manuscript this supplementary figure on page 6: *“The expression value of individual cytokines has no association with the ARIADNE class (Fig. S3). Considering the complex network between cytokines and other immune mediators, the analysis of individual cytokines might be misleading [19].”*.

Reviewer: For the PD-L1 pathway, a number of the included genes are also part of the PI3K/AKT and MAPK pathways, important for cancer cell proliferation. Could there be an independent explanation for differences in gene expression correlation, not related to the regulation of PD-L1 expression?

Authors: We mentioned this point in the discussion section on page 8 : *“Notice, however, that a number of the genes involved in the PD-L1/PD-1 pathway are also part of the PI3K/AKT and MAPK pathways, important for cancer cell proliferation, so an independent explanation for differences in gene expression correlation could also be possible”*.).

Reviewer: Figure 4e – it would help to visualise this as boxplots as well (in addition to boxplots Fig 4b-d which show individual cell scores rather than aggregate by patient).

Authors: We report the boxplots in Fig. S5 of the revised manuscript.

Reviewer: The link between the exhausted T-cell analysis and the rest of the paper is not obvious. At a minimum, it would be good to investigate how their fraction relates to patients with low and high-risk ARIADNE scores.

Authors: We checked the possible association between the ARIADNE groups and the fractions of exhausted CD8+ and regulatory CD4+ T-cells, but the results were not statistically significant. The results are reported in Fig. S6 and mentioned in page 8.

Reviewer: Please mention in the abstract and discussion that ARIADNE identifies epithelial-mesenchymal phenotypes.

Authors: Done (see abstract and page 8 of the revised manuscript)

Minor comments:

Reviewer: First sentence in the second last introduction paragraph is missing a comma after chemotherapy.

Authors: Done

Reviewer: Please add the relevant statistics for Figure 4f-h to support the results “The ARIADNE score obtained from cancer cells is decreasing in time for patients not responding to immunotherapy while it remains roughly constant for responders (Fig. 4f). The cytokine score decreases for CAFs (Fig. 4g) and increases for cancer cells (Fig. 4h). after immunotherapy (T1).”

Authors: The differences are all statistically significant ($p < 10^{-4}$), but the effect sizes are small. We report this information in the caption of the figure.

Reviewer: Fig S2 – was multiple testing correction performed?

Authors: Notice that there is only two significant different genes out of 28 genes, the correction with multiple comparison leads to no significant difference. We now mention this point in the caption.