Peer Review File

Article information: https://dx.doi.org/10.21037/tlcr-24-242

Reviewer A

In this study, the associations of circulating 91 protein profiles with clinical factors such as smoking status, disease stage, and survival outcomes were explored. The strengths of this study are the relatively large sample size, prospectively collected survival data, and strict statistical methodologies applied. However, there are issues that should be addressed as described below.

Comments:

1. This study enrolled patients with stage I-IIIA NSCLC. However, no description was provided regarding treatment. Please provide information regarding the type of surgery, applied (neo)adjuvant chemotherapy. These factors might have influence on survival outcomes. This information should also be presented in Abstract as well.

Reply 1: While we agree that treatment does influence the survival of patients with NSCLC, due to the study design and timing of blood withdrawal, we believe it could not have any impact on the observed estimates for the association between circulating proteins and survival. Please kindly note that the blood samples were collected from the enrolled patients before receiving any treatment. Also, all patients had undergone surgical resection of their tumor.

To address the reviewer's concern, for this revision, we performed a sensitivity analysis by further adjusting the models for chemotherapy and radiation therapy, that showed very similar estimates for the association between circulating proteins and survival.

To clarify these points, we made the following changes to the manuscript and tables:

- We added the number and percentages of participants who received chemotherapy and radiotherapy to **Table 1**.
- We slightly modified the first paragraph of the "Methods - Study population, recruitment, sample collection, and follow-up" section to emphasize the timing of blood collection as follows: "**Upon recruitment, blood samples were collected from the participants before receiving any treatment for their disease, and were then stored at a temperature of -70°C.**" (see Page 5 and 6, Line 28-2)
- We added two sentences to the third paragraph of the "Methods – Statistical analysis" section to emphasize further adjustment for treatment did not change the results as follows "**Further adjustment for treatment (chemotherapy and radiation therapy) did not affect the observed estimates. Therefore, we proceeded with the more parsimonious model for subsequent analyses**." (see Page 8, Line 3-5)
- We added the treatment information to the first paragraph of the "Results - Descriptive statistics" section as follows "**All patients underwent surgical resection for their tumor, while 26% further received chemotherapy, and 17% received radiation therapy.**" (see Page 9, Line 2-3)

2. Method: what type of blood samples were used for analysis, serum or plasma?

Reply 2: we confirm that plasma samples were used for the analysis. We revised the whole manuscript (i.e. the abstract, introduction, methods, and results) to clarify that the samples used in this study are "plasma samples".

3. Method: the definition of never smoker is not provided.

Reply 3: We modified the first paragraph of the "Methods - Study population, recruitment, sample collection, and follow-up" section to define never smokers as follows: "**Ever smokers were defined as participants who reported to have smoked at least 100 cigarettes through their lifetime, otherwise, the participants were considered as never smokers.**" (see Page 6, Line 2-3)

4. Method (Proteomic assay): the authors used the Olink proteomic platform. The reviewer considers that this technology is not commonly known, please provide brief introduction of this assay.

Reply 4: we added a paragraph to the "Methods – Proteomics assays" to introduce the Olink proteomic platform as follows "**We used the Olink proteomics platform (https://www.olink.com/) in Uppsala, Sweden to measure the relative concentrations of circulating proteins. The platform is based on proximity extension assays (PEA) that are highly sensitive, avoid cross-reactivity, and have high reproducibility(19). It provides high-throughput semi-quantitative concentration measurements of annotated proteins by quantitative PCR (qPCR), and has been applied widely for proteomic measurement in various studies(20).**" (see Page 6, Line 14-18)

5. Method (Proteomic assay): please provide more detailed information how the authors conducted normalization of concentrations of each protein.

Reply 5: The normalized protein expression (NPX) is Olink's relative protein quantification unit on a log2 scale. We have added more details on the normalization by citing the previous published paper (PMID: 34715355, added as reference 21) which introduced the detailed procedures and gave an overall description of normalization. Specifically, we added the following information to the "Methods – Proteomics assays" section: "**Protein measurements were expressed as normalized protein expression (NPX) values for each protein per individual. NPX is Olink's relative protein quantification unit on a log2 scale, and the detailed normalization procedure was described before(21). In brief, it's comprised of three main steps: normalization to the extension control, log2 transformation, and level adjustment using the plate control**." (see Page 6, Line 20-24)

6. Method (Proteomic assay): the list of 92 proteins should be shown at least in a Supplementary file.

Reply 6: we added the 92 proteins including the Uniprot ID as Supplementary Table 1. We further modified the "Methods – Proteomics assays" section as follows: "**We measured circulating levels of 92 proteins using the "immuno-oncology" panel based on their biological function (Supplementary Table 1), of which 44 are involved in the inflammatory response.**" (see Page 6, Line 18-20)

7. Method (Proteomic assay): it is not sure whether the handling of missing data as mean values are sound. Additionally, please provide the reasons why values were missed in <5% of participants using the multiplex assay?

Reply 7: To clarify these issues, we modified the "Methods – Proteomics assays" section of the manuscript as follows: "**We excluded one protein (IL33) which had no variance across the patients. Of the remaining 91 proteins, only 5 had missing values. These include IL4 which was missing for 47 participants (4.8%), and four other proteins (IL1-alpha, IL2, IL12RB1, and IL13) which were only missing for one random participant. Because of the randomness of the missing values that were due to failed datapoints (assay / chip failure) and the small number of individuals with missing data, we replaced the missing values for the five proteins with the mean values of the study population.**" (see Page 6-7, Line 26-3)

8. Method (Proteomic assay): please explain why "14 samples failed during the protein measurement"?

Reply 8: To clarify this issue, we modified the last two sentences of the "Methods – Proteomics assays" section of the manuscript as follows: "**Furthermore, of the selected 1000 samples, 14 were excluded due to contamination (n=11), and protein measurement failure (n=3). Consequently, 91 proteins from a**

total of 986 patients were included in the current analysis (Figure 1)." We also added these details to Figure 1 (Flowchart of the participants enrolment). (see Page 7, Line 3-5)

9. Method (Proteomic assay): I wonder how the authors handled the values beyond upper limit of detection?

Reply 9: Measured proteins data from the Olink platform has an S-curve (sigmoid) relationship with the true protein concentration in a sample. The upper limit of quantification (ULOQ) was calculated by using a standard curve with 2-fold serial dilutions of known protein concentrations. Data above ULOQ was not reported due to the high risk of misinterpretation from the high dose Hook effect. In our data, we didn't observe any individual with protein levels above ULOQ. Therefore, we added the below sentence in the methods as "**We didn't observe protein levels above upper limit of detection**". (see Page 6, Line 26)

10. Method (Proteomic assay): were standards used to measure absolute concentrations?

Reply 10: As explained in our response to the comments number 4 and 5, the protein levels were measured as relative concentrations. We added a paragraph to the "Methods – Proteomics assays" to introduce the Olink proteomic platform as follows "**We used the Olink proteomics platform (https://www.olink.com/) in Uppsala, Sweden to measure the relative concentrations of circulating proteins. The platform is based on proximity extension assays (PEA) that are highly sensitive, avoid cross-reactivity, and have high reproducibility(19). It provides high-throughput semi-quantitative concentration measurements of annotated proteins by quantitative PCR (qPCR), and has been applied widely for proteomic measurement in various studies(20).**" (see Page 6, Line 14-18)

11. Results (Figure 1): as a difference in overall survival according to smoking status was not the main objective of this study, survival curves shown in Figure 1 appears not to be appropriate as it is. I recommend to move this Figure into Supplementary material. Instead, a flowchart of patient selection is required as Figure 1.

Reply 11: as suggested by the reviewer, we moved the original Figure 1 into the supplement (Supplementary Figure 1) and **generated a new Figure 1** that illustrates the **flowchart** of the study.

12. Results (page 4, line 36): "most were male (67%)" is better to be "67% of patients were male", something like that. Similarly, "one third of the measured proteins" appears not to be appropriate (page 4, line 47).

Reply 12: as suggested by the reviewer, we have revised the first paragraph of the "Results - Descriptive statistics" section of the manuscript as follows: "**For the 986 patients, the mean age at diagnosis was 64 years (SD 8.8 years), and 67% of patients were male.**" (see Page 9, Line 1) Also we revised the first paragraph of the "Results - Cross-sectional analysis on protein levels by smoking status and TNM stage" as follows "**After adjustment for potential confounders, protein levels were nominally different (p<0.05) for 32% (29/91) of the measured proteins between never and ever smokers (Figure 2A, Supplementary Table 2)**" (see Page 9, Line 10-11)

13. Results: Figure 4A-E are not indicated in the main text.

Reply 13: We added the citation in this revision to the second paragraph of the "Results - Protein levels and lung cancer survival" as follows: "**Before adjusting for stage (models 1 and 2), MUC16 and TRAIL were associated with survival in ever smokers (P ENT-corrected < 0.05), and MUC16, HGF, and IL8 were associated with survival among never smokers, showed ENT-significant association with survival (P ENT-corrected < 0.05) (Figure 4A, 4B, 4D, 4E).**" (see Page 10, Line 13-16)

14. Results: the subsection of "Protein levels and overall survival by histology". As this is not the objective of this study, I consider this subsection is unnecessary.

Reply 14: While we appreciate the reviewer's feedback, we believe these results are still important and could be informative and valuable for the design and interpretation of future research. Considering the reviewer's comment, we suggest keeping these results in the supplementary, so the interested researchers can access them.

15. Results: the authors described that the levels of IL-8 and HGF were higher in ever smokers than never smokers. However, both were significantly prognostic in the never smoker cohort. Please explain this unexpected finding. Also, this reviewer wants to see the distribution of concentrations of both proteins according to the smoking status. Actually, no exact concentrations and units are provided in this paper.

Reply 15: To clarify this issue, we would like to kindly bring the following two points to the attention of the reviewer:

1) The proteins levels measured by Olink were relative concentrations, which were expressed as normalized protein expression (NPX) values, and it is Olink's relative protein quantification unit on a log2 scale (**please kindly see our responses to comments 4, 5 and 10**). We further rescaled each protein to have a mean of 0 and a standard deviation (SD) of one. Therefore, the unit of proteins in our paper is illustrated as per standard deviation increase. To avoid this confusion, we revised the "Results - Protein levels and lung cancer survival" section of the manuscript and **replaced all "HR"** in this section **with "HR per SD increase".**

2) The distribution of proteins levels across smoking status is presented in **Figure 2A** and **Supplementary Table 2**. Please kindly note that these results are obtained from a cross-sectional analysis assessing the distribution of these proteins across never and ever smoking patients, regardless of their survival. The fact that we found IL-8 and HGF at diagnosis to be higher in ever smokers compared to never smokers, but to be associated with survival only among never smokers (but not ever smokers), clearly shows the heterogeneity of the protein levels across smoking status. The results supported our novel conclusion that circulating proteins in NSCLC patients vary by smoking history, and failure to account for smoking status "as an important confounder" in the studies analyzing protein levels and cancer survival, may result in observing biased estimates. Accordingly, as we showed in this work, only after stratifying the analysis by smoking status (to account for confounding by smoking history) we could observe the significant associations between these two proteins and survival among never smoking patients.

16. The authors randomly selected patients with a smoking history. However, I wonder why the authors did not adjust for sex and histology when they selected patients, because differences in the proportions of these factors should have been expected when stratified based on a smoking history. These factors might affect the results. This limitation should be addressed in the limitation section.

Reply 16: We agree that the distribution of sex and histology differs across smoking status. However, due to the limitations in the number of the available sample size, we were not able to consider sex and histology during the sample selection. To account for this limitation, we have already adjusted all models for sex and histology, and all presented results have accounted for the differences in sex and histology by statistical adjustments. We added this limitation to the sixth paragraph of the "Discussion" section of the manuscript as follows: "**We did not consider sex and histology during the sample selection, which are differently distributed among never and ever smokers. However, we tried to address this limitation in the statistical analysis, by adjusting all models for sex and histology".** (see Page 13, Line 6-8)

17. The associations of protein profiles with other survival outcomes such as recurrence-free survival is warranted.

Reply 17: As suggested by the reviewer, we added a sensitivity analysis assessing the disease progression (recurrence, metastasis, and death whichever occurred first) as an alternative endpoint. Accordingly, we added a sentence to the second paragraph of the "Methods – Statistical analysis" section of the manuscript to illustrate this analysis as follows: "**We further performed a sensitivity analysis to assess the risk of disease progression (recurrence, metastasis, and death whichever occurred first) as an alternative endpoint.**" (see Page 7, Line 26-27)

We also **added a supplementary figure (Supplementary Figure 2)** to present the results of this analysis, and added a sentence to the first paragraph of the "Results - Protein levels and lung cancer survival" section of the manuscript as follows "**Similar results were obtained when we assessed disease progression as the outcome (Supplementary Figure 2), and when we stratified the analysis by tumor stage as IA-IIA and IIB-IIIA subgroups (Supplementary Table 4).**" (see Page 10, Line 5-7)

18. This study reported no statistical improvement in a prognostic model after integrating some proteins levels. In the analyses, the authors calculated hazard ratios per 1-SD increase. However, it might be possible to reach some significant results when the dichotomized status separated at an optimal cutoff point is used, as some proteins might harbor non-quantitative correlations with survival outcomes. Could you address this point?

Reply 18: We agree with the reviewer that the cutoff analysis is very important for the protein levels when their levels are measured as absolute concentrations. However, given that in our analysis protein levels were measured as relative concentrations, we think assigning a cutoff point would probably be misleading and could result in wrong interpretations.

Reviewer B

In this study, which I find quite interesting, the authors attempt to establish the prognostic value of several serum proteins in a relatively large cohort of NSCLC patients. I would like to make a few comments:

1. I find the analysis on never smokers quite intriguing. Do you happen to know the EGFR/ALK/other mutational status in these patients to correlate prognosis and tumor expression as well?

Reply 1: Unfortunately, we do not have tumor mutation information in this study.

2. I noticed a lack of data regarding neoadjuvant/adjuvant chemotherapy in these patients, which can significantly affect survival prognosis.

Reply 2: While we agree that treatment does influence the survival of patients with NSCLC, due to the study design and timing of blood withdrawal, we believe it could not have any impact on the observed estimates for the association between circulating proteins and survival. Please kindly note that the blood samples were collected from the enrolled patients before receiving any treatment. Also, all patients had undergone surgical resection of their tumor.

To address the reviewer's concern, for this revision, we performed a sensitivity analysis by further adjusting the models for chemotherapy and radiation therapy, that showed very similar estimates for the association between circulating proteins and survival.

To clarify these points, we made the following changes to the manuscript and tables:

- We added the number and percentages of participants who received chemotherapy and radiotherapy to **Table 1**.
- We slightly modified the first paragraph of the "Methods - Study population, recruitment, sample collection, and follow-up" section to emphasize the timing of blood collection as follows: "**Upon recruitment, blood samples were collected from the participants before receiving any treatment for their disease, and were then stored at a temperature of -70°C.**" (see Page 5 and 6, Line 28-2)
- We added two sentences to the third paragraph of the "Methods – Statistical analysis" section to emphasize further adjustment for treatment did not change the results as follows "**Further adjustment for treatment (chemotherapy and radiation therapy) did not affect the observed estimates. Therefore, we proceeded with the more parsimonious model for subsequent analyses**." (see Page 8, Line 3-5)
- We added the treatment information to the first paragraph of the "Results - Descriptive statistics" section as follows "**All patients underwent surgical resection for their tumor, while 26% further received chemotherapy, and 17% received radiation therapy.**" (see Page 9, Line 2-3)

3. Have you considered performing the analysis specifically in locally advanced stages (i.e. IIB-IIIA)? These tumors could be more indicative of "systemic" diseases rather than the very early ones.

Reply 3: Thanks for this suggestion, we added a new analysis stratifying the patients into IA-IIA and IIB-IIIA and assessing the association between circulating proteins and overall survival. We presented the results in a **new supplementary table (Supplementary Table 4)**. In general, we found the association between most protein levels and overall mortality not to be significantly different between patients with stage IA-IIA vs. IIB-IIIA tumors. However, among the proteins which were nominally significant with overall mortality in the overall patients, only 2 proteins showed differences in the strength of their association with survival between earlier vs. later stage tumors.

In this revision, we added the following paragraph to the end of first paragraph of the "Results - Protein levels and lung cancer survival" section: "**Similar results were obtained when we assessed disease progression as the outcome (Supplementary Figure 2), and when we stratified the analysis by tumor stage as IA-IIA and IIB-IIIA subgroups (Supplementary Table 4). However, among the proteins with nominally significant association with overall mortality in overall patients (Figure 3C). ICOSLG showed stronger inverse association in patients with stage IIB-IIIA than patients with stage IA-IIA (HR per SD increase: 0.76 [95% CI: 0.66-0.87] vs. 0.96 [95% CI: 0.84-1.10]), and MUC16 showed stronger positive association in patients with stage IA-IIA than patients with stage IIB-IIIA (HR per SD increase: 1.31 [95% CI: 1.15-1.49] vs. 1.08 [95% CI: 0.95-1.22]) (Supplementary Table 4).**" (see Page 10, Line 5-12)

4. I would appreciate an explanation of why you believe IL-8 and HGF are associated with prognosis prediction in these patients. Have you hypothesized any biological basis for this association?

Reply 4: We would like to kindly emphasize that while IL-8 and HGF were associated with survival among never smoking patients, as we have mentioned in the abstract and the results, the integration of the protein markers into a statistical model that contains demographics and clinical factors did not improve the prediction of NSCLC survival among never smoking patients [C-index of 0.68 (clinical) vs. 0.72 (clinical+proteins) for never smokers, p=0.24].

To address this comment from the reviewer, we added the following sentences to the end of the fourth paragraph of the "Discussion" section: "**After correction for multiple testing, IL-8 and HGF remained significantly associated with overall mortality in NSCLC patients who never smoked. In previous studies, elevated IL-8 was linked to an unfavorable tumor microenvironment and was shown to potentially serve as a therapeutic target for NSCLC(38). Similarly, HGF (Hepatocyte Growth Factor) acts as a stromal cell-derived factor that strongly affects cancer cell invasiveness in the tumor microenvironment, that has also been targeted in anticancer drug discovery over the past decade(39).**" (see Page 12, Line 23-28)

We also **cited two new references** (number 38 and 39) for this paragraph.

5. I am also curious if the predictive value of these proteins and other soluble biomarkers such as CTCs or ctDNA has been explored. It might be worth mentioning given their importance in this context.

Reply 5: Unfortunately, other soluble biomarkers such as CTCs or ctDNA have not been measured in this study. We added the following sentences to the last paragraph of the "Discussion" section in the study "**While our study highlights the importance of considering smoking status and tumor stage in future analysis of circulating proteins in relation to cancer outcomes, well-powered future studies are needed to investigate a broader panel of the blood proteome and other biomarkers such as circulating tumor DNA (ctDNA) and Circulating Tumor Cells (CTCs), in relation to survival in NSCLC patients.**" (see Page 13, Line 17-20)