Peer Review File

Article Information: https://dx.doi.org/10.21037/tlcr-21-750

Reviewer A

This manuscript showed no causality between CRP levels and the occurrence of LC using Mendelian randomization based on a large genomic database including multiethnic population. The results were not differed in lung cancer subtype (adenocarcinoma and squamous cell carcinoma). The authors emphasized the two important things: First, the sample size for the analysis of the causal relationship between CRP levels and LC risk is large. This is the first study to demonstrate no causality between CRP and LC risk using Mendelian randomization with a large sample. Second, the authors selected the relevant SNPs data using a comprehensive and proved method using a worldwide gene database (GWAS) as an IV set.
I think the manuscript is well written and the results are interesting for many physicians and researchers. This is potentially acceptable to the Journal of Translational Lung Cancer Research, although this study includes some minor revision points to be improved.
First, this manuscript includes some ambiguous points. The assumptions for the Mendelian randomization study, which was pointed out in the method (P5 L173-176). It is difficult to interpret "(ii) the IVs affect LC merely via their effect on elevated CRP levels". In my interpretation, this means the IVs only have the effect on CRP levels but have no direct influence on the occurrence of LC (Please refer to the following figure of (iii)). I think it is difficult to ignore the effect of a genetic variants on the LC occurrence. If you have some data that demonstrate these SNPs have no direct effect on the LC occurrence, please add.

Second, the results that CRP does not increase the lung cancer risk contradicts that of the previous biological experiment. Many studies reveal CRP promotes carcinogenesis supported by the results of the biological experiments. The authors should add the biological explanation that supports the result that this study demonstrated.

Major comments
1. Method (P5 L173-176)
Please clarify and revise the unclear and ambiguous sentence concerning the assumption of the Mendelian randomization study.
It is difficult to interpret "(ii) the IVs affect LC merely via their effect on elevated CRP levels". In my interpretation, this means the IVs only have the effect on CRP levels but have no direct influence on the occurrence of LC (Please refer to the following figure of (iii)).
a) Please improve the ambiguous description.
b) I think it is difficult to ignore the effect of a genetic variants on the LC occurrence. If you have some data that demonstrate these SNPs have no direct effect on the LC occurrence, please add.

Reply to a): Thank you for your comments. According to the three basic assumptions of the Mendelian Randomization (MR) study, a unidirectional flow could not be violated for instrumental variables (IVs), that is, to affect the outcome (i.e., lung cancer) via exposure (i.e., C-reactive protein (CRP) concentrations). Otherwise, the study would violate the second assumption (i.e., the IVs are independent of the outcome given the exposure) (1), which constitute reverse causality and interfere the causal association revealed by the MR study.

Change in the text:

Page 9, line 232: "(ii) the IVs affect LC merely via their effect on elevated CRP concentrations (i.e., the IVs are independent of the outcome given the exposure)."

Reply to b): First of all, all CRP-associated Single-nucleotide polymorphisms (SNPs) used in our study have a common feature, as they all reached the significance threshold ($P < 5*10^{-8}$). Hence, they were significantly related to the regulation of circulating CRP concentrations. Second, after retrieving the SNP database of NCBI, there was no evidence to support that the CRP-related SNPs we used as instrumental variables have a significant correlation with lung cancer. In conclusion, we can elucidate that the effect of a genetic variants on the lung cancer occurrence was negligible in our study.

2. Discussion(P7L272-293)

The authors should add the biological explanation that supports the result which this study demonstrated.

Although the authors mentioned the two reasons(L282-289), I think this is not sufficient.

As I examined, CRP does not always act as a cancer promoter but sometimes has an anti-cancer effect. Sasaki et al. reported that the rate of lymph node metastasis in mice with subcutaneous squamous cell carcinoma was 2.3-times lower in mice injected with CRP (Sasaki T, Motoyama S, Sato Y, Yoshino K, Matsumoto G, Minamiya Y, Saito H, Murata K, and Ogawa J: C-reactive protein inhibits lymphangiogenesis and resultant lymph node metastasis of squamous cell carcinoma in mice. Surgery 154(5): 1087-1092, 2013). In addition, Shinohara et al. reported the better survival of elevated CRP in the acute postoperative period (Shinohara S, Sugaya M, Onitsuka T, et al. Prognostic Impact of Postoperative C-reactive Protein for Non-small Cell Lung Cancer Following Lobectomy. Anticancer Res 2018;38:3193-8).

Please add the biological explanation that supports the result.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 15, line 395 to 404: "CRP may not always act as a cancer promoter but sometimes play a role in treating cancer. Sasaki et al. observed that CRP can inhibit lymph node metastasis and lymphatic angiogenesis of squamous cell carcinoma through injecting CRP into mice model (2). As for the prognostic effects of CRP, Shinohara and colleagues found high serum circulating CRP levels on postoperative days were associated with enhanced five-year overall survival (OS) as well as recurrence-free survival (RFS) in NSCLC patients (3). However, Okada et al. reported the high serum CRP levels during perioperative period were a poor

prognostic factor for OS and RFS in NSCLC patients (4). Consequently, the impact of CRP levels on the incidence and prognosis of LC was nuanced and complexed, and further studies with larger sample size and better design are in an urgent need."

Minor comments

1. Material and methods(L154-170)

Please clarify the methodology to calculate the genome-wide significance threshold. 查

I think there are several methodologies to calculate it, including the Bonferroni correction, Sidak correction, False Discovery Rate (FDR), permutation test, and Bayesian approaches.

If you refer something in the method, please add the reference. In addition, please declare what software you use for the calculation.

Reply: Thank you for your comments. The genome-wide significant threshold ($P = 5 \times 10^{-8}$) was obtained through Bonferroni correction to control the family-wise error rate (FWER) given the linkage disequilibrium (LD) structure of the genome. The threshold was widely accepted for identifying an association between a common genetic variant and a trait of interest (5). The genome-wide significance for each single nucleotide polymorphism was originally provided by each genome-wide association studies provided by the MR-Base platform (http://app.mrbase.org/) (6), and therefore no additional calculation was needed in our study.

Change in the text:

Page 8, line 205-208: "Through Bonferroni correction to control the family-wise error rate (FWER), the P = 5 × 10−8 threshold was widely accepted for association identification between a common genetic variant and a trait of interest, given the linkage disequilibrium (LD) structure of the genome (33). Using the MR-Base platform, SNPs associated with elevated CRP concentrations were initially selected from Neale Lab, Pan-UK Biobank (UKB) team, RIKEN Center for Integrative Medical Sciences, and the European Bioinformatics Institute (EBI) database of complete GWAS summary data at the genome-wide significance threshold (P < 5×10-8)."

2. Bioinformatics and biological statistics

In the manuscript, the study does not include bioinformaticians and biological statisticians. I think the analysis need the help of these specialists. Please clarify the contribution of the bioinformatic analysis in the study.

Reply: Thank you for your comments. Our main authors Xiangrong Wu and Haoxin Peng undertook the bioinformatic analysis in this study. Their capabilities are widely recognized in the field of Mendelian Randomization (MR) research, and both had published several MR studies as the lead authors (7-9).

Regarding the contribution of the bioinformatic analysis in the study, MR is an analytical method that uses genetic variants as instrumental variables to estimate the causal effect of an exposure on an outcome (10). MR studies can provide reliable evidence on the effect of modifiable risk factors for disease or ill health and can overcome some limitations of traditional observational epidemiology, such as unmeasured confounding (11). In settings for which the instrumental variable assumptions are well justified, the findings could help optimize clinical

trials or drug development and inform clinical or public health decision making (12).

Change in the text:

Page 6, line 141: "Using genetic variations as instrumental variables (IVs), Mendelian randomization (MR) analysis is a novel epidemiology method to estimate the causation between an exposure and an outcome, with less impressionability to reverse causation and unmeasured confounders. In settings for which the IV assumptions are well justified, the findings could help optimize drug development or clinical trials and inform clinical or public health decision making (24)."

Reviewer B

The manuscript describes a study that has assessed the bidirectional MR analyses between lung cancer risk and genetically predisposed CRP concentrations in a large dataset that allowed for investigations in different histological subtypes and ethnic ancestries. Genetic summary data from ILCCO was included in the analyses. The null associations reported in the results are of interest to the readers interested in CRP and its association to lung cancer risk.

The study appears largely well performed and I have few comments. The novelty in this observational study lies in the large number of participants.

Specific comments:

First highlight: 'CRP levels' - CRP blood concentrations is more specific?

Third highlight and throughout the text: the word level is used most often, please be consistent and use concentrations. Also, make sure to mention that you mean blood concentrations when appropriate.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

We have replaced "levels" as "concentrations" or "blood concentrations" throughout the manuscript.

Page 3, lines 97-98: Please provide a citation.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 5, line 115: "Particularly, lung cancer (LC) is currently the second most common cancer and the major cause of cancer-related death globally (1)"

Page 3, line 101: Prevention is mentioned. As smoking is the main risk factor, lung cancer prevention is largely dependent on limiting smoking habits. Instead, CRP as a marker of smoking or marker to indicate screening would be more appropriate to mention for me.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your

suggestions.

Change in the text:

Page 6, line 135-139: “However, given that cigarette smoking is by far the leading risk factor for LC, CRP is widely seen as a predictor of LC in smokers and former smokers (22), while cigarette smoking itself has been reported to increase circulating levels of CRP directly (23). Hence, conventional observational studies are prone to bias by potential confounding factors or reverse causation”

Page 3, lines 107-110: In my opinion the information mentioned here should include a mention of the marker as a systemic or local marker.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 6, line 126-127: “Simultaneously, circulating C-reactive protein (CRP) was considered as a systemic marker of chronic inflammation”

Page 3, lines 113-114: In line with my previous comment, CRP concentrations may also indicate a systemic response to ongoing disease or comorbidity.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 6, line 126-127: “Simultaneously, circulating C-reactive protein (CRP) was considered as a systemic marker of chronic inflammation”

Page 3, line 122: Mention and define SNPs where ‘genetic variants’ are mentioned?

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 6-7, line 146-166: “Single nucleotide polymorphism (SNP) is a single base-pair difference in the DNA sequence of individuals within a species, which are the most common type of genetic variation in humans (25).”

Page 4, line 124: What is meant by ‘heritability of CRP’? Concentrations of the protein. Please be more specific.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 7, line 169-170: “it has been estimated that the heritability of CRP blood concentrations was estimated from 25% to 40%”

E.g. Page 4, line 155 and page 5, lines 182+184: Please be consistent in defining abbreviations only upon first mention throughout the manuscript.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 9-10, line 240-251: "Five well-established MR analytical methods were applied, including inverse-variance weighted (IVW), weighted median, MR-Egger, weighted mode, and simple mode. IVW was applied to the combination of multiple IVs as sole estimation of genetic variants by weighted score, which has the most statistical power among the five methods (37)."

Page 5, lines 176-177: The sentence seems to be missing one word to make sense? Also, please present arguments or citations to provide the reader certainty that the assumption was indeed met.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 9, line 234-236: "As demonstrated by the previous studies (33) and the screening procedure of SNPs described in the previous paragraph, the first assumption was met."

Page 5, line 186: There seem to be a word missing in 'the rest methods'.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 10, line 251-252: "The rest of the methods were performed for sensitivity analysis and to indirectly test the second assumption."

Page 6, line 213: 'Regarding LC on CRP levels' – please rephrase to mention the association that I assume you mean.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 11, line 280: "Regarding the impact of LC upon CRP concentrations"

Page 6, line 219: What is meant by ' among all studies'? Please be more specific.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 11, line 286:" The results of sensitivity analyses among all study populations are shown in Supplementary table 2."

Page 6, lines 224-225: 'the collectivity effect of CRP on LC' – please rephrase to mention the association that I assume you mean. And is collectivity effect what you mean?
After full consideration, we think "overall effects" is more appropriate than "collectivity effect". Besides, we had revised the sentence.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 11, line 290-291: "Leave-one-out studies of the LC overall and subgroup analyses showed no evidence that a single SNP had an impact upon the overall effect of CRP variants on LC risks"

Page 6, line 223: Please rephrase to use another word than 'following'

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 12, line 305: "Additionally, no causality was found in subgroup analyses concerning pathologic types as well."

Page 6, lines 230-233: There are no further mention of the different subgroup analyses. Maybe mention that variation and sample size varied across the different groups?

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 12, line 300-305: "No bidirectional causation between genetically regulated elevated concentrations of CRP and LC risk was observed in our study among East Asian (P = 0.690, nSNP = 6, n = 75,391), Hispanic/Latin American (P = 0.980, nSNP = 13, n = 15,912), European (P = 0.397, nSNP = 174, n = 204,402), African American/Afro-Caribbean (P = 0.551, nSNP = 4, n = 6203), and South Asian populations (P = 0.984, nSNP = 7, n = 8397)."

Page 6, lines 234-235: Please rephrase to mention lung cancer in this context.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 12, line 307-308: "The results are inconsistent with most previous prospective cohort studies (3,11,12,14,16-19) using measured plasma concentrations of circulating CRP to explore

the risk of LC."

Page 6, line 249: How are the results similar? Please be more specific? Do you mean in terms of strength of associations?

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 13, line 336: "Our study showed similar results in effect directions compared with previous MR studies"

Page 7, line 260: Please avoid the word 'proved'

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 13, line 346: "some of their chosen SNPs were also proposed to have nothing to do with elevated CRP concentrations subsequently"

Page 7, lines 280-282: This is the kind of information that could be mentioned also in the introduction in my opinion.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 5, line 119-120: "Growing studies have demonstrated that chronic inflammation, especially occurred within the respiratory system, might be a risk factor for LC"

Page 7, line 296: What is meant by 'proved'?

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 14, line 376: "It has been reported that soluble tumor necrosis factor receptor-2 (sTNFRII), serum amyloid A (SAA), and monokine induced by gamma interferon (CXCL9/MIG) were also related to LC risk"

Page 8, line 309: The assumptions are not really mentioned much and it is unclear to the reader why the assumptions could not be thoroughly tested. The sample size is large in this study but if the assumptions are not clearly addressed it does not strengthen the results in my opinion.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 16, line 423: “Besides, due to the methodology limitations currently, MR assumptions could not be thoroughly tested, and therefore potential violations against the assumptions may occur. To overcome this difficulty, instead of directivity evaluated in the second assumption, we implemented additional sensitivity analyses”

Page 8, line 311: Is the sentence missing one word? ‘Evaluated by/in the’?

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 16, line 424: “To overcome this difficulty, instead of directivity evaluated in the second assumption, we implemented additional sensitivity analyses”

Page 8, line 320: are elevated CRP concentrations and lung cancer two symptoms? Please rephrase.

Reply:

Thank you very much for your valuable comment. We have revised in accordance with your suggestions.

Change in the text:

Page 16, line 434: “Further investigation of the profound relationship between both phenotypes is required to be unveiled in pathologic and biochemistry aspects.”

Throughout: there are some minor misspellings throughout that should be checked.

1. Sekula P, Del Greco M F, Pattaro C, et al. Mendelian Randomization as an Approach to Assess Causality Using Observational Data. Journal of the American Society of Nephrology : JASN 2016;27:3253-65.
2. Sasaki T, Motoyama S, Sato Y, et al. C-reactive protein inhibits lymphangiogenesis and resultant lymph node metastasis of squamous cell carcinoma in mice. Surgery 2013;154:1087-92.
3. Shinohara S, Sugaya M, Onitsuka T, et al. Prognostic Impact of Postoperative C-reactive Protein for Non-small Cell Lung Cancer Following Lobectomy. Anticancer Res 2018;38:3193-8.
4. Okada S, Shimomura M, Tsunezuka H, et al. Prognostic Significance of Perioperative C-Reactive Protein in Resected Non-Small Cell Lung Cancer. Semin Thorac Cardiovasc Surg 2020;32:1046-55.
5. Chen Z, Boehnke M, Wen X, et al. Revisiting the genome-wide significance threshold for common variant GWAS. G3 (Bethesda) 2021;11.
6. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 2018;7.
7. Peng H, Wu X, Wen Y, et al. Association between systemic sclerosis and risk of lung cancer: results from a pool of cohort studies and Mendelian randomization analysis. Autoimmun Rev 2020;19:102633.
8. Wu X, Peng H, Wen Y, et al. Rheumatoid arthritis and risk of lung cancer: Meta-analysis and Mendelian randomization study. Semin Arthritis Rheum 2021;51:565-75.
9. Peng H, Li C, Wu X, et al. Association between systemic lupus erythematosus and lung cancer:

results from a pool of cohort studies and Mendelian randomization analysis. J Thorac Dis 2020;12:5299-302.
10. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37:658-65.
11. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32.
12. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ (Clinical research ed) 2018;362:k601.