Peer Review File

Article information: https://dx.doi.org/10.21037/tlcr-23-334

Reviewer A

This is a very interesting study indicating that the organic cation transporter (OCT)-3, encoded by the gene SCL22A3, is associated with poor outcome in human lung squamous cell carcinoma. The data indicate that SCL22A3 may be a surrogate marker, as it is regulated through methylation patterns affecting a larger gene set involved in the inflammatory response of the tumor microenvironment. I have only one major comment:

Comment 1: *The authors should provide staging-adjusted Kaplan-Meier curves comparing low and high expression of SCL22A3, either for TNM stages pT1,2,3, and 4 respectively, or better by UICC stages. This data could be added to Figure 3, where only data for the entire cohort stratified to therapy modalities are shown.*
This would provide better evidence for the solidity of the conclusion and show more clearly how balanced high-and low expression cohorts are and how much data have been censored.

Reply 1: Thank you for reviewing our manuscript. In agreement with your comment, we have included Figures 3K, 3L, 3M, and 3N to illustrate survival differences between *SLC22A3*-low and *SLC22A3*-high cohorts stratified by AJCC stage I, II, III, and IV, respectively.

Changes in the text: In lines 235-237, we added: " Staging-stratified Kaplan-Meier curves based on the AJCC staging system showed that *SLC22A3*-high was associated with worse outcomes in stage II and IV patients (Figure 3 K-N). "

Comment 2: *I also suggest that a native speaker reviews the manuscript with respect to language.*

Reply 2: We have taken your feedback seriously. After careful review, we made further

refinement to enhance the overall quality of the manuscript.

Changes in the text: Several changes were made to improve the manuscript's language.

Reviewer B:

This manuscript by Nguyen et al. reports on the expression on SLC22A3 as a potential prognostic factor for human lung squamous cell carcinoma (LSCC). Utilizing the databased for LSCC in the TCGA database, the authors report that high SLC22A3 expression is associated with worse prognosis and enhanced immune-related pathways. Although SLC22A3 has been reported in other tumor types, this paper is the first to report the correlation of SLC22A3 as a prognostic factor for LSCC. The overall study is descriptive but is still important for the field of LSCC biology.

Some weaknesses to address to improve the study:

Comment 1: *Novelty of the paper would be strengthened if the authors can validate the findings (or at least some of the findings) in a separate cohort of LSCC (Chinese LSCC database).*

Reply 1: Thank you for reviewing our manuscript. We tried but were unable to access the Chinese databases. Therefore, we utilized data from the GEO database.

For prognosis, validation was carried out in GSE74777 (n = 107), GSE37745 (n = 66), which have relatively large sample size compared to other datasets available on GEO.

For immunogenicity validation, we employed HOT score to classify LSCC into "hot" and "cold" tumor group and compared *SLC22A3* expression between the two groups. GSE162520 (n = 45) and GSE161537 (n = 17), which were initially used in the study developing the HOT score, and TCGA-LUSC cohort were used in this analysis.

The details can be found in the Methods and Results of our revised manuscript.

Changes in the text:

We added new lines to the manuscript for validation:

- We added a new section in Methods:

"**Validation analyses**

We downloaded GSE74777, GSE37745, GSE162520, and GSE161537 from Gene Expression Omnibus (GEO) database and extracted gene expression and clinical data of LSCC cases (n = 107, 66, 45, and 17, respectively) to validate the prognosis and immunogenicity significance of *SLC22A3* on LSCC.

We conducted Kaplan-Meier OS analyses on GSE74777 and GSE37745 separately and on the combined cohort of these datasets, which were generated using the value-binning technique. Value-binning is a commonly used technique in data analysis that discretizes data into a predefined number (B) of bins, thereby facilitating cross-platform data management [23]. We applied a simple value-binning method to categorize *SLC22A3* gene expression values into B bins, discretizing expression values from 1 to B. We selected B = 107 (corresponding to the sample number of the larger dataset) because each value can be assigned to a gene expression value in both datasets. This process can be deemed non-parametric scaling. Our goal was to remove the "magnitude" effect, which is easily confounded by cross-platform differences, and normalize the expression value, which creates the same range of value in both datasets for further non-parametric analyses. This technique was also performed by another study [24]. We used the maximally selected rank statistics (MSRS) technique to determine the survival cutpoint of the *SLC22A3* expression in validation data, which is also a non-parametric analysis, to find the most significant prognostic cut-off in each dataset. The cut-off bins of GSE37745 and GSE74777 were 40 and 95, respectively. This result indicated the differences in clinicopathological characteristics of the two examined cohorts. Therefore, we used the corresponding cut-off of each dataset to divide the sample into *SLC22A3*-high and *SLC22A3*-low samples.

We validated the immunogenicity impact of *SLC22A3* on LSCC by utilizing the Hot Oral Tumor (HOT) score described below. Because the unit of read count was inconsistent between TCGA and validation data, it is impossible to utilize the cut-off of FPKM $\geq$ 5 to determine high expression in the validation data. Therefore, we used the MSRS technique to determine the survival cutpoint of the *SLC22A3* expression in validation data. Based on the HOT score, samples in the TCGA-LUSC cohort were

defined as "hot" or "cold" tumors. *SLC22A3* gene expression was compared between the "hot" and "cold" tumor groups in TCGA-LUSC as well as in validation datasets: GSE162520 and GSE161537 (these two datasets were used in the original study on developing HOT score [25]). Additionally, we conducted OS Kaplan-Meier analyses of *SLC22A3*-low and *SLC22A3*-high LSCC samples in the GSE162520 dataset. We excluded GSE161537 in this survival analysis since the patients of this cohort were treated with immunotherapy, which may confound the result.

Hot Oral Tumor (HOT) score

The HOT score was calculated by using GSVA of the 27-gene list in the original study [25], including *CCL19, CCR2, CCR4, CCR5, CD27, CD40LG, CD8A, CXCL10, CXCL11, CXCL13, CXCL9, CXCR3, CXCR6, FASLG, FGL2, GZMA, GZMH, IDO1, IFNG, IRF8, LAG3, LYZ, MS4A1, PDCD1, TBX21, TLR7*, and *TLR8*. LSCC tumors with -1<ES<0 or 0<ES<1 were classified as "cold" or "hot" tumors, respectively. The calculation procedure was performed as in the previous study [25]."

- In Results, we added a new section:

"***Validation on prognostic and immunogenic effects of SLC22A3 gene expression in LSCC***

When validating the prognostic significance of *SLC22A3* in LSCC patients from GSE37745 and GSE74777 cohorts separately, we observed a tendency towards poorer prognoses in *SLC22A3*-high patients compared to *SLC22A3*-low patients, albeit without statistical significance (p-value = 0.078 and 0.07, respectively) (**Supplementary Figure 2A and 2B**). These results may be influenced by the relatively limited sample sizes within these cohorts. Therefore, we employed the value-binning technique to merge these two datasets (please refer to the Methods for more details). The combined cohort's Kaplan-Meier OS curves showed that *SLC22A3*-high expression was linked to poorer prognoses ($p < 0.001$) **(Figure 6A)**.

Our study also found that *SLC22A3*-high tumors were positively correlated with immune-related pathways and an abundance of infiltrating immune cells in the tumor

microenvironment. To validate this finding, we utilized the HOT score [25], an interesting metric developed by J-P. Foy et al. specifically designed to identify tumors with high immunological activity, suggesting potential benefits from immunotherapies. We found that *SLC22A3* expression was consistently higher in "hot" tumors compared to "cold" tumors in both TCGA and validation datasets **(Figure 6B and 6C**). Despite not being statistically significant (p = 0.22), the results indicated a trend towards poorer prognosis in *SLC22A3*-high patients compared to *SLC22A3*-low patients **(Supplementary Figure 2C)**."

- We added Figure 6 and Supplementary Figure 2.

Comment 2: *Careful consideration should be given to state that SLC22A3 is a "prognostic factor" as there is no validation to the analysis.*

Reply 2: We appreciate your concern. We have validated the prognostic significance of *SLC22A3* on LSCC.

Changes in the text: No changes were made to the text.

Comment 3: *It is unclear why FPKM of 5 is chosen as a cut off for High vs Low. If the sample has an FPKM of 6 or 4, are those considered high/low respectively? How different would the results be reflected if the top 25 quartiles vs the bottom 25 quartiles for SLC22A3 expression would have been chosen?*

Reply 3: Thank you for your question. Choosing the appropriate FPKM cut-off is an empirical issue. The "high expression" cut-offs of >1, >3, and >5 FPKM are usually chosen by previous studies [PMID: 26911985, 34714959, 33492573], while others have used higher cut-offs like >50 FPKM or even >100 FPKM [PMID: 33472597, 26973288]. The choice depends on the distribution of read counts for the gene of interest, in our case, *SLC22A3* (please refer to Figure 2A). With the cut-off of 5 FPKM, 17.8% of patients were classified as *SLC22A3*-high, which suited our analyses. Thus, we chose the 5 FPKM cut-off for our study.

Changes in the text: No changes were made to the text.

Reviewer C

The article entitled "High expression of SLC22A3 that encodes organic cation transporter-3 is associated with poor prognosis in human lung squamous cell carcinoma" shows an investigation into the role of the SLC22A3 gene in the prognosis of squamous lung cancer, and therefore its value as a prognostic biomarker in these tumours. Undoubtedly, it addresses a current issue on which there is literature described in other solid tumours, but not in non-small-cell lung cancer. The authors' article is well written and easy to understand and is not excessively long. Furthermore, the results obtained are interesting and allow us to infer hypotheses that should be proven in the future with independent studies.

Overall, the article is well structured without major flaws and the language does not require modification. The figures and tables are as requested by the article, and I don't think they need to be modified either. The title, abstract and keywords do need changes, which are expressed below. The references used are what the article needs and do not need changes either.

In my point of view, I think the article needs the following changes:

Major changes

Comment 1: *Introduction and discussion: I think it would be necessary to indicate and explain why the SLC22A3 gene has been chosen for study. Furthermore, the main advances in NSCLC are being made in adenocarcinomas. Why have squamous cell carcinomas been chosen?*

Reply 1: Thank you for reviewing our manuscript. We sincerely appreciate your insightful comments and suggestions.

The primary reason behind our choice to investigate on *SLC22A3* gene is its potential to influence prognosis and treatment responsiveness in certain carcinomas, as suggested by previous studies (PMID: 23934321, 34080124, 33574675, 30561001, 29088791, 31732877, and 19190342). However, its impact on lung cancer, particularly LSCC, remains unexplored.

We acknowledge that recent advancements in NSCLC treatment benefit LUAD more

than LSCC. Therefore, we believe that understanding the molecular aberrations underpinning LSCC tumors is essential for improving therapy success.

These considerations guided our study, and we mentioned in the introduction section.

Changes in the text: No changes were made to the text.

Comment 2: *Results: Have PD-L1 subgroup analyses been performed? Please specify.*

Reply 2: We have conducted a differentially expressed gene analysis (DEGA) for *CD274*, which encodes PD-L1, and found no significant difference between the *SLC22A3*-high and *SLC22A3*-low groups. We apologize for the oversight in not including this data in the initial version of the manuscript.

Changes in the text: We have added the DEGA data for *CD274* to Table 3. The main text remains unchanged.

Comment 3:

-Discussion: it would be important to specify how the results obtained could be extrapolated to the clinic, especially in the field of immunotherapy.

Reply 3: Thank you for your suggestions. We have included the potential implications of *SLC22A3* in clinical, particular the immunotherapy strategies.

Changes in the text: In lines 383-388, we added: "These findings suggest that high expression of *SLC22A3* that encodes organic cation transporter-3 in LSCC is associated with poor prognosis and immunogenicity of the tumor. Understanding the function implications of *SLC22A3* in LSCC and how it interacts with the immune system may help improve LSCC patient stratification for optimizing immunotherapy treatment, thereby potentially improving outcomes for LSCC patients."

Minor changes

Comment 4: *Title: this is the conclusion of the article; it should be modified so that it is a title of the subject being addressed.*

Reply 4: Thank you for your suggestion. We re-titled our paper as follow.

Changes in the text:

Title:

"SLC22A3 that encodes organic cation transporter-3 is associated with prognosis and immunogenicity of human lung squamous cell carcinoma."

Comment 5: *Abstract (methods): indicate the centre where the work was carried out.*

Reply 5: In this study, we utilized database from The Cancer Genome Atlas Program (TCGA) project, which is publicly available. Therefore, we did not mention the centre where the data was generated. To make it more transparent, we would like to emphasize the publicly available characteristics of the TCGA database in the main text.

Changes in the text: In line 32, we added "a publicly available database".

Comment 6: *Abstract (results): indicate more exact data than those expressed.*

Reply 6: Thank you for your advise. We saw that the abstract was not in good arrangement. Thus, we have modified the result of the abstract to indicate more data and also rewritten the method of abstract to be more concise.

Changes in the text: We have modified method and result of the abstract:

"***Methods:***

We analyzed gene expression, DNA methylation, and clinicopathological data from the TCGA-LUSC (n = 501), a publicly available database exclusively consisting of LSCC patients. Using a 5 FPKM cut-off, we divided LSCC patients into two groups: patients with tumors possessing high and low *SLC22A3* expression (*SLC22A3*-high and *SLC22A3*-low, respectively). Prognostic significance was determined through Cox analyses and Kaplan-Meier curves for overall survival (OS) and disease-free survival (DFS). Differential methylation position, differentially gene expression, and pathway analyses were performed. Validation was carried out in GSE74777 (n = 107), GSE37745 (n = 66), GSE162520 (n = 45) and GSE161537 (n = 17).

Results:

SLC22A3-high LSCC patients had lower OS and DFS rates than *SLC22A3*-low LSCC patients. The different expression levels of *SLC22A3* in LSCC were correlated

with the methylation status of the *SLC22A3* gene. Pathway analysis indicated that *SLC22A3* expression levels were positively correlated with immune-related pathways such as inflammatory response and abundance of infiltrating immune cells in the tumor microenvironment. Notably, in the *SLC22A3*-high group, many genes encoding immunological checkpoint inhibitory molecules were upregulated. In addition, *SLC22A3* expression positively correlated with the Hot Oral Tumor (HOT) score, indicating high tumor immunogenicity."

Comment 7: *Keywords: add "biomarker".*

Reply 7: Thank you for your suggestion. In this revision, after successfully validated our findings, we agreed that "biomarker" should be a keyword.

Changes in the text: We added "biomarker" into keywords.

Comment 8: *Introduction, line 64: indicate that small-cell lung cancer represents about 15% and non-small-cell lung cancer about 85%.*

Reply 8: We have indicated your suggestion in our revised manuscript.

Changes in the text: In lines 58-61, we added: " Lung cancer has two main types: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), which account for 85% and 15% of all cases, respectively (www.cancer.org, accessed on July 24, 2023). "

Comment 9: *Introduction, 68-69: update the data on survival, they are old.*

Reply 9: We have updated the latest data on survival from Surveillance, Epidemiology, and End Results (SEER) program.

Changes in the text: In lines 64-65, we changed: "The 5-year relative survival rate of patients with LSCC increased over time but remained poor, reaching only 24.2% in 2020 [3]."

Comment 10: *Genes should be indicated in italics*.

Reply 10: Thank you for your careful review. We have now corrected the missing gene

italic indications.

Changes in the text: In line 69, we corrected: […] such as *EGFR*, *KRAS* mutations, and *ALK* rearrangements […]

Comment 11: *Methods: indicate confirmation of approval by your centre's ethics committee.*

Reply 11: As mentioned, we utilized publicly available databases that do not require approval confirmation.

Changes in the text: No changes were made to the text.

Comment 12: *Results: the titles of the results sections are the conclusions of the results. They need to be adjusted to what is done at each point.*

Reply 12: In agreement with your comment, we have changed the titles of the result section.

Our flow of result section in previous manuscript is:

1. A total of 17.8% LSCCs possess high expression (FPKM ≥ 5) of the *SLC22A3* gene.
2. *SLC22A3*-high is associated with poor prognosis in TGCA-LUSC cohort.
3. *SLC22A3* expression patterns in LSCC are associated with cancer pathways, cancer-immune interaction, and cancer progression landscapes.
4. *SLC22A3*-high LSCC is associated with high activity of the immune-related pathways.
5. Many genes encoding immune checkpoint inhibitory molecules were enriched in *SLC22A3*-high LSCC.

Now, the result titles are as follows:

Changes in the text:

“A total of 17.8% LSCCs possess high expression (FPKM ≥ 5) of the *SLC22A3* gene.”

" Clinicopathological analysis showed worse prognoses in *SLC22A3*-high patients of TCGA-LUSC cohort."

"Dimension reduction analyses showed *SLC22A3* expression patterns were associated

with tumor pathway, immune, and progression landscapes".

"Pathway analyses showed vastly different oncogenic and TME signals in *SLC22A3*-high LSCCs".

"DEAs illustrated many significant DEGs between *SLC22A3*-low and *SLC22A3*-high tumors".

“Validation on prognostic and immunogenic effects of *SLC22A3* gene expression in LSCC”

In summary, I think the authors have done a good job that is worthy of publication in the journal. However, some changes would be necessary before consideration. If the authors, make them correctly I think it would be a promising article for publication.

Thank you for your positive comments on this study.

Reviewer D

Comment 1: *The description of the results is far-fetched, especially the results from the t-SNE dimension reduction. Looking at the t-SNE dimension reduction plots, I do not know on what basis the authors found differences between SLC22A3-low and SLC22A3-high groups. There are no differences there, patients from these two groups mix with each other, not creating any separate clusters (Fig 4A, 4D, 4G).*

Reply 1: Thank you for your comments. We understand your concerns about the interpretation of our results. Regarding the t-SNE dimension reduction plots, we determine the differences between *SLC22A3*-low and *SLC22A3*-high based on two approaches.

Firstly, we subjectively observed that *SLC22A3*-low distribution shows scattered patterns, while *SLC22A3*-high tends to accumulate in one region (lower region in Fig 4A and 4D, upper region in Fig 4G).

To further validate these observations, we conducted an in-depth analysis using scatter plots (Fig 4B, C, E, F, H, and I) to explore the linear relationship between each dimension of the t-SNE plot and the continuous normalized read counts of the *SLC22A3* gene. Notably, our Spearman's correlation analyses produced significant Rho values different from 0, along with p-values < 0.05, signifying the statistical significance of these relationships.

Changes in the text: No changes were made to the text.

Comment 2: *Similarly, the heatmaps (Fig 5) also do not show any significant and visible differences between groups.*

Reply 2: Regarding the heatmaps in Fig.5, we acknowledged that several pathways do not display significant differences between the two groups. However, we also noticed that many of the pathways exhibited highly significant differences, as illustrated by the p-value along with the heatmaps.

Changes in the text: No changes were made to the text.

Comment 3: *I also have doubts about the size of the compared groups - one of them has 412 patients and the other only 89.*

Reply 3: Compared to the *SLC22A3*-low group (n = 412), the *SLC22A3*-high group size (n = 89) is relatively small. However, the sample size of 89 patients is still considered large enough for the analyses. We believe that this difference does not affect the analyses.

Changes in the text: No changes were made to the text.

Comment 4:

The study also lacks the validation of analyzes on another cohort - there are certainly many of them in the GEO database.

Reply 4: As per your suggestion, we have validated our findings on GEO datasets.

For prognosis, validation was carried out in GSE74777 (n = 107), GSE37745 (n = 66), which have relatively large sample size compared to other datasets available on GEO.

For immunogenicity validation, we employed HOT score to classify LSCC into "hot" and "cold" tumor group and compared *SLC22A3* expression between the two groups. GSE162520 (n = 45) and GSE161537 (n = 17), which were initially used in the study developing the HOT score, and TCGA-LUSC cohort were used in this analysis.

The details can be found in the Methods and Results of our revised manuscript.

Changes in the text:

We added new lines to the manuscript for validation:

- We added a new section in Methods:

"**Validation analyses**

We downloaded GSE74777, GSE37745, GSE162520, and GSE161537 from Gene Expression Omnibus (GEO) database and extracted gene expression and clinical data of LSCC cases (n = 107, 66, 45, and 17, respectively) to validate the prognosis and immunogenicity significance of *SLC22A3* on LSCC.

We conducted Kaplan-Meier OS analyses on GSE74777 and GSE37745 separately and on the combined cohort of these datasets, which were generated using the value-

binning technique. Value-binning is a commonly used technique in data analysis that discretizes data into a predefined number (B) of bins, thereby facilitating cross-platform data management [23]. We applied a simple value-binning method to categorize *SLC22A3* gene expression values into B bins, discretizing expression values from 1 to B. We selected B = 107 (corresponding to the sample number of the larger dataset) because each value can be assigned to a gene expression value in both datasets. This process can be deemed non-parametric scaling. Our goal was to remove the "magnitude" effect, which is easily confounded by cross-platform differences, and normalize the expression value, which creates the same range of value in both datasets for further non-parametric analyses. This technique was also performed by another study [24]. We used the maximally selected rank statistics (MSRS) technique to determine the survival cutpoint of the *SLC22A3* expression in validation data, which is also a non-parametric analysis, to find the most significant prognostic cut-off in each dataset. The cut-off bins of GSE37745 and GSE74777 were 40 and 95, respectively. This result indicated the differences in clinicopathological characteristics of the two examined cohorts. Therefore, we used the corresponding cut-off of each dataset to divide the sample into *SLC22A3*-high and *SLC22A3*-low samples.

We validated the immunogenicity impact of *SLC22A3* on LSCC by utilizing the Hot Oral Tumor (HOT) score described below. Because the unit of read count was inconsistent between TCGA and validation data, it is impossible to utilize the cut-off of FPKM ≥ 5 to determine high expression in the validation data. Therefore, we used the MSRS technique to determine the survival cutpoint of the *SLC22A3* expression in validation data. Based on the HOT score, samples in the TCGA-LUSC cohort were defined as "hot" or "cold" tumors. *SLC22A3* gene expression was compared between the "hot" and "cold" tumor groups in TCGA-LUSC as well as in validation datasets: GSE162520 and GSE161537 (these two datasets were used in the original study on developing HOT score [25]). Additionally, we conducted OS Kaplan-Meier analyses of *SLC22A3*-low and *SLC22A3*-high LSCC samples in the GSE162520 dataset. We

excluded GSE161537 in this survival analysis since the patients of this cohort were treated with immunotherapy, which may confound the result.

Hot Oral Tumor (HOT) score

The HOT score was calculated by using GSVA of the 27-gene list in the original study [25], including *CCL19, CCR2, CCR4, CCR5, CD27, CD40LG, CD8A, CXCL10, CXCL11, CXCL13, CXCL9, CXCR3, CXCR6, FASLG, FGL2, GZMA, GZMH, IDO1, IFNG, IRF8, LAG3, LYZ, MS4A1, PDCD1, TBX21, TLR7*, and *TLR8*. LSCC tumors with -1<ES<0 or 0<ES<1 were classified as "cold" or "hot" tumors, respectively. The calculation procedure was performed as in the previous study [25].”

- In Results, we added a new section:

“***Validation on prognostic and immunogenic effects of SLC22A3 gene expression in LSCC***

When validating the prognostic significance of *SLC22A3* in LSCC patients from GSE37745 and GSE74777 cohorts separately, we observed a tendency towards poorer prognoses in *SLC22A3*-high patients compared to *SLC22A3*-low patients, albeit without statistical significance (p-value = 0.078 and 0.07, respectively) (**Supplementary Figure 2A and 2B**). These results may be influenced by the relatively limited sample sizes within these cohorts. Therefore, we employed the value-binning technique to merge these two datasets (please refer to the Methods for more details). The combined cohort's Kaplan-Meier OS curves showed that *SLC22A3*-high expression was linked to poorer prognoses ($p < 0.001$) **(Figure 6A)**.

Our study also found that *SLC22A3*-high tumors were positively correlated with immune-related pathways and an abundance of infiltrating immune cells in the tumor microenvironment. To validate this finding, we utilized the HOT score [25], an interesting metric developed by J-P. Foy et al. specifically designed to identify tumors with high immunological activity, suggesting potential benefits from immunotherapies. We found that *SLC22A3* expression was consistently higher in "hot" tumors compared to "cold" tumors in both TCGA and validation datasets **(Figure 6B and 6C**). Despite

not being statistically significant (p = 0.22), the results indicated a trend towards poorer prognosis in *SLC22A3*-high patients compared to *SLC22A3*-low patients **(Supplementary Figure 2C)**."

- We added Figure 6 and Supplementary Figure 2.