Peer Review File

Article information: https://dx.doi.org/10.21037/tlcr-23-14

Reviewer A

This a descriptive and exploratory study, where the authors evaluated lysosome genes as predictive and prognostic marker in lung adenocarcinoma using bioinformatics and public data base as basic tools. The study demonstrates deep knowledge on the part of the authors in the construction of mathematical models using a large cohort of samples and R statistical analysis. However, there are major points in the study design that should be addressed by the authors.

Comment 1: A Table containing clinic-pathologic characteristics of the patients should be include with information of age, gender, tobacco history, histotypes of adenocarcinoma, tumors stage, mutation status (p53, EGFR, ALK, BRAF, RET and ROS).
Reply 1: We thank the reviewer for this suggestion. Based on your suggestion, we have added Table S1 and Table S2 as the clinical information for the sample in the manuscript.
Changes in the text: see Page 6-7, line 135-140; Table S1; Table S2.

Comment 2: Were all adenocarcinomas invasive lesions? Please, confirm if there was AIS and MIA.
Reply 2: We thank the reviewer for this suggestion. The samples used in this study were all TCGA samples, and review of the dataset clinical information did not reveal AIS and MIA information in LUAD clinical practice.
Changes in the text: No

Comment 3: Besides transcriptome, protein expression determined by immunohistochemistry or immunofluorescence should be performed for practical implications.

Reply 3: We thank the reviewer for this suggestion. we increase the validation of Western blotting and human protein atlas (HPA) database according to review comments.
Changes in the text: see Page 3, line 55-57; Page 4, line 68-70; Page 13-14, line 278-300; Page 21, line 456-463; Figure 10; Figure S3.

Comment 4: The validation of the study should include cases of the authors Institution or culture cells or an experimental model and complemented by RT-PCR.
Reply 4: We thank the reviewer for this suggestion. we increase the validation of RT-PCR according to review comments.

Changes in the text: see Page 3, line 55-57; Page 4, line 67-68; Page 12-13, line 255-277; Page 21, line 454-456; Figure 10.

Reviewer B

Dear Authors, an interesting study on LUAD and prognostic lysosome-related genes; in my opinion, the paper is fair work. However, it needs various improvements prior to further processing. Please answer or consider the following:
Comment 1: Title: I think the "lysosome/lysosomal" should be connected with words like "associated" or "related" when referring to "genes" with the use of hyphen; it is present in some locations, but in some not. Please standardize it throughout the paper, including the title.
Reply 1: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 1, line 3; Page 6, line 116.

Comment 2: Line 39: you can avoid the repetition of "lung cancer" in single sentence via changing "of all lung cancer cases" to "of all cases of this tumor". This is of course optional.
Reply 2: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 2, line 38.

Comment 3: Line 42-43: The first sentence should be "RNA-Seq data of The Cancer Genome Atlas Lung Adenocarcinoma cohort (TCGA-LUAD) were downloaded via Genomic Data Commons Data Portal". This would improve English and clarify that TCGA data are deposited in GDC Portal.
Reply 3: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 2, line 41-42.

Comment 4: Line 46: Maybe add "previously" before "published"? This is optional.
Reply 4: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 2, line 44.

Comment 5: Line 48: Add comma before "respectively".
Reply 5: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 2-3, line 44-47.

Comment 6: Line 54-55: Last sentence in Methods could start with "Lastly" instead

of “Moreover”, and then “lastly” can be deleted from the middle part of sentence.
Reply 6: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 3, line 52-54.

Comment 7: Line 57: Obvious typo “gens” (genes).
Reply 7: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 3, line 58.

Comment 8: Line 62-63: Two sentences start with “Moreover”. Maybe choose one to change to “Furthermore”?
Reply 8: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 3, line 61-65.

Comment 9: Line 70: Consider changing “were selected for the prognosis model, and it is valuable for the prognosis of LUAD patients.” to “were selected for the model that is of prognostic significance for LUAD patients”.
Reply 9: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 4, line 71-72.

Comment 10: Line 81: “of lung cancer” can be “of this tumor” to avoid repetition.
Reply 10: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 5, line 98.

Comment 11: Line 83: If you explain “EGFR” abbreviation, I think you should do the same for “KRAS” that is just after the first one. The entire sentence should be improved; one should delete “activating mutations” or change the final part “are the most frequently mutated genes”. If you want to refer to EGFR and KRAS genes, you cannot write that “activating mutations are the most frequently mutated genes” since this may confuse the Reader.
Reply 11: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 5, line 99-101.

Comment 12: Line 85: “is less than 5 years, and that of LUAD patients has markedly improved” could be “is less than 5 years but has markedly improved”.
Reply 12: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.

Changes in the text: see Page 5, line 102.

Comment 13: Line 94: Capitalize the first letter of “lysosomes” at the beginning of sentence.
Reply 13: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 5, line 111.

Comment 14: Line 95: Change the entire line from “role in many non-tumor and tumor diseases. For example, in non-tumor diseases,” to “role in many tumor and non-tumor diseases. For the latter,”
Reply 14: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 5, line 111.

Comment 15: Line 100: “will inhibit”? Maybe just “inhibits”? I noticed few areas where you use future tense but I think it is unnecessary.
Reply 15: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 6, line 117.

Comment 16: Line 104: change “The” to “the” after semicolon, or change semicolon to full stop. Please also work on similar examples in lines 107 and 108. Decide where sentences should end or if some are too short and could be combined with a semicolon.
Reply 16: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 6, line 120-126.

Comment 17: Line 111: “relationship between the model and immunity, mutation” could be “relationship between the model and immunity or mutational status”.
Reply 17: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 6, line 127-128.

Comment 18: Line 117: “if” should be “of”, whereas “TCGA database” should be “GDC Portal”.
Reply 18: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 6, line 133-134.

Comment 19: Line 120: I understand that only LUAD samples were limited from 526

to 513? All normal (59) were included for further steps?
Reply 19: We thank the reviewer for this suggestion. The TCGA-LUAD dataset in this study included 526 LUAD samples and 59 normal samples, of which 513 LUAD samples with survival and clinical information were used for the screening of prognostic genes and the construction of models. It should be noted that LUAD samples were used in building the model and normal samples were not required.
Changes in the text: No.

Comment 20: Line 121: Please rewrite the sentence with GSE dataset and GEO database. I suggest to avoid many short inclusions in sentences; some details can be put in the brackets next to website link for example GSE identifier.
Reply 20: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 7, line 137-139.

Comment 21: Line 124-125: Change “144 lysosome-related genes (LRGs) were selected according to Vairo et al” to “The list of 144 lysosome-related genes (LRGs) was acquired from Vairo et al study”.
Reply 21: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 7, line 140-141.

Comment 22: Line 129: Obvious typo “differentilly” (differentially).
Reply 22: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 7, line 145.

Comment 23:Line 130-131: Why did you choose such criteria for DEGs? Typically, one should use Log2FC > 0.58 (or better 0.6), and p < 0.01. Was it because the standard threshold did not yield satisfactory amount of significant results?
Reply 23: We thank the reviewer for this suggestion. Here differential analysis p measures the significance of a gene, and Log2FC measures the fold difference between two groups of genes. When performing differential analysis, the appropriate screening threshold can be selected according to the actual situation of the dataset. Here we select Pvalue < 0.05 and | Log2FC | > 0.5 to obtain the optimal intersection of differential genes and subsequent lysosome-related genes at this threshold. If a stricter threshold is used (| Log2FC | > 1), it will lead to less differential lysosome-related genes obtained by the intersection, and then less prognostic lysosomal genes cannot better reflect the role of differential lysosome-related genes in the target disease. If the threshold is lowered, the number of genes obtained will be higher and the reliability will be reduced. An analysis to adjust the threshold according to the actual situation is presented in Reference [PMID: 34225739] (1), which refers to

(Pvalue < 0.05 and | Log2FC | > 0.5) [PMID: 29547407, PMID: 28058013] (2,3).
Changes in the text: No.

Comment 24: Line 133-136: This entire part is just one sentence, it needs to be rewritten or even split in two. Change "visualized to" should be "visualized using". Moreover, was "ggplot" the real "ggplot" package or the second edition i.e. "ggplot2" (typically the latter one is used nowadays)? Moreover, the "heatmap" was package on its own, or the function within ggplot? Lastly, correct obvious typo (dufferential).
Reply 24: We thank the reviewer for this suggestion. First, we divided this part into two sentences according to your suggestion. Spelling errors have also been modified. Wayne diagram is obtained by the ggplot2 package (version 3.3.2, Fig. https://link.springer.com/book/10.1007/978-3-319-24277-4), the heat map was drawn by the Heatmap package (version 4.1.0, Fig. https://link.springer.com/content/pdf/10.1007/978-3-540-70928-2_29.pdf pdf = inline link).
Changes in the text: see Page 7, line 148-152.

Comment 25: Line 139-141: Change "The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene ontology (GO) enrichment analyses which includes cellular components, (CC), molecular functions, (MF), and biological process (BP) were utilized for" to "The Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene ontology (GO; containing cellular components [CC], molecular functions [MF], and biological processes [BP]) were employed for".
Reply 25: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 7, line 155-159.

Comment 26: Line 145: remove capitalized letters in the middle of sentences. This applies to mentioned location, as well as few others in the paper. Please correct accordingly.
Reply 26: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 8, line 159-163.

Comment 27: Line 152: Change "with setting" to "with the following settings".
Reply 27: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 8, line 169.

Comment 28: Line 154: Change "optimize" to "optimized".
Reply 28: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 8, line 171.

Comment 29: Line 156: Change “by the” to “from”. Moreover, what is “the model genes expression levels”? Do you refer to expression of genes included in the risk model that you created?
Reply 29: We thank the reviewer for this suggestion. Risk scores were obtained based on the expression of genes in samples and the coefficients of genes obtained by LASSO and multivariate Cox analysis. Thus, expression levels of pattern gene genes in the manuscript refer to expression levels of genes included in the risk model.
Changes in the text: see Page 8, line 174.

Comment 30: Line 159: By “ground on” you mean “based on”? Please also check line 214.
Reply 30: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 8, line 176.

Comment 31: Line 161-162: The package for ROC curves is “survival ROC” or “survivalROC”? From what I saw in few sources, it should be the latter?
Reply 31: We thank the reviewer for this suggestion. According to your suggestion we have revised the manuscript and should be survivalROC.
Changes in the text: see Page 8, line 180.

Comment 32: Line 168: You can mention that T, N, and M are all part of staging but can also be considered separately. Moreover, explain to the audience what T/N/M stands for. In the same sentence, you can delete the first “of patients”.

Reply 32: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 9, line 186-191.

Comment 33: Line 179: If you mention about R-related functions, one could also add brackets after the function’s name, e.g., “cph()”. In the same sentence, what is the name of the R package that contains cph() function? Is it “rms”?
Reply 33: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion. The cph function is the rms package (version 6.1.0, https://CRAN.R-project.org/package=rms).
Changes in the text: see Page 9, line 199.

Comment 34: Line 185-186: Version 7.4 refers to MSigDB, not GO and KEGG collections/gene sets. Please rewrite to something like: “Based on the C5:GO(BP+CC+MF) and C2:CP:KEGG gene sets downloaded from Molecular Signatures Database (MSigDB) v7.4 (http://www.gsea-msigdb.org)”. In the same

paragraph, explain “NES” abbreviation (normalized enrichment score).
Reply 34: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 10, line 206-209.

Comment 35: Line 194-195: change “which would be presented as stromal scores, immune scores, and ESTIMATE composite scores” to something like “these scores (stromal, immune, combined) can also be analyzed separately”.
Reply 35: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 10, line 217-218.

Comment 36: Line 199: Is “file” word intentional, or should it be deleted?
Reply 36: We thank the reviewer for this suggestion. We have deleted the “file” word
Changes in the text: see Page 10, line 221

Comment 37: Line 202: Not sure if the explanation of “ICIs” should be earlier in the main text? Please double-check.
Reply 37: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion. We did not comment in the text.
Changes in the text: No.

Comment 38: Line 211: “would be” -> “was”. Same in line 213
Reply 38: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 11, line 235; Page 11, line 236.

Comment 39: Line 230: You can rename subfigure D as C so it will be in appropriate order in the main text. In line 231, correct “gens”.
Reply 39: We thank the reviewer for this suggestion. Following your advice we have revised the manuscript and Figure 1.
Changes in the text: see Page 15, line 320-321; Figure 1.

Comment 40: Line 234: In my opinion, the bottom part of subfigure C (which I believe should be D?) could be removed as it does not add anything more when compared to Venn diagram and is illegible due to large differences in groups. Such removal can also help in enlarging the font in the entire figure, which should definitely be increased in size, similar in other figures.
Reply 40: We thank the reviewer for this suggestion. Following your advice we have revised the manuscript and Figure 1.
Changes in the text: see Page 15, line 323; Figure 1.

Comment 41: Line 238 and 240: is response to TNF mentioned twice? Is comma missing after the first one (before the word “embryonic”)?
Reply 41: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 15, line 327-329.

Comment 42: Line 240-242: BP/CC/MF are not process(es), better to call them “gene sets” of GO collection.
Reply 42: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 15, line 329-331.

Comment 43: Line 244-245: Remove capitalized letter in the middle of sentence, the names of revealed processes can be separate just by comma, there is no need to capitalize them in such situation. Alternatively, capitalize all of them (e.g. lines 238-239) and be consistent throughout the entire paper.
Reply 43: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 15, line 332-333.

Comment 44: Line 249: I think you should not double up periods even if “etc” is at the end of a sentence. Same in line 302.
Reply 44: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 16, line 338; Page 18, line 389.

Comment 45: Line 250: add “interactions” next to “predicted” and remove capitalized letters.
Reply 45: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 16, line 339.

Comment 46: Line 266: Change “worse survival situations” to “worse outcome”.
Reply 46: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 16, line 351.

Comment 47: Line 271: The paragraph ends with reference to Figure 3G-I, the next section refers to Figure 4. Therefore, I think you missed Figure 3J-L in the main text.
Reply 47: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 17, line 357.

Comment 48: Line 274: I think the “RiskScore(s)” (in some parts with a space mark) requires standardization in the main text. In my opinion, the sense will be remained if you just write it normally as “risk score(s)”.
Reply 48: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 17, line 360.

Comment 49: Line 275: Maybe add “nearly all” before “subgroups”, since some of them were not significant? Of course only a small part, but still.
Reply 49: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 17, line 361.

Comment 50: Line 282: “prognosis” -> “prognostic”. Line 286: add “respectively” after “were”.
Reply 50: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 17, line 368; Page 17, line 373.

Comment 51: Line 284: It appears that “N” was the most important within Staging, as investigated by two Cox analyses. “T” was also found significant in univariate analysis. On the other hand, “M” was not indicated in any of them. Maybe mention that in this part, or Discussion, that the importance of staging was most probably due to the “N” status/parameter (lymph nodes that have cancer)?
Reply 51: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 17, line 370-371.

Comment 52: Line 291: Add “that” after “found”.
Reply 52: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 18, line 378.

Comment 53: Line 309: Maybe “composite” -> “combined”? In line 310: “them” -> “that”. Line 311, remove colon after “such as”.
Reply 53: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 18, line 395.

Comment 54: Line 325: If your analysis presented that in the top 10 mutated, the last one is KRAS, but there is no EGFR – maybe the mention in lines 83-84 should be

provided with more details, or modified?
Reply 54: We thank the reviewer for this suggestion. In the introduction, epidermal growth factor receptor (EGFR) and KRAS activating 83 mutations are the most common mutated genes in LUAD, which are the findings of the references, while in our findings, KRAS is only present in the top 10 mutated genes, but not EGFR, which may be caused by different databases and analytical methods, and is also an innovation of our study, that is, we both agree with some results in the references and supplement them, which can provide new theoretical support for subsequent studies to further explore genetic mutations in LUAD.
Changes in the text: No.

Comment 55: Line 334-335: Change the last sentence in the paragraph to "TP53 and TTN were the genes with highest mutation frequency both in the high-risk and low-risk group."
Reply 55: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 19, line 416-417.

Comment 56: Line 337: Remove "the" between "higher" and "TMB".
Reply 56: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 19, line 418.

Comment 57: Line 338: Consider changing "correlation […] was significant" to "correlation […] was weak-to-moderate yet significant".
Reply 57: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 19, line 419.

Comment 58: Line 343: Change "Among 56 types of chemotherapeutic 12 chemotherapeutic drugs" to "Among 56 chemotherapeutics, 12 drugs".
Reply 58: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 20, line 425.

Comment 59: Line 352: Add "is" before "decreased".
Reply 59: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 22, line 469.

Comment 60: Line 361: do not capitalize "Normal" in the middle of sentence.

Reply 60: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 22, line 478.

Comment 61: Line 382: Change “researches” to “research” or “experiments” or “studies”.
Reply 61: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 23, line 501.

Comment 62: Line 383: Change “P53” to “TP53” (if gene) or “p53” (if protein). Remember to italicize gene symbols throughout the paper.
Reply 62: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 24, line 514.

Comment 63: Line 386-387: Change “cause immune escape and promote tumor progression” to “leading to immune escape and tumor progression”.
Reply 63: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 24, line 517.

Comment 64: Line 389: Change “load” to “burden”. Line 403: what is that excessive space mark? Line 404: remove period in “Masts.cells”, use “Mast” instead of “Masts”.
Reply 64: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 24, line 519; Page 25, line 532; Page 25, line 533.

Comment 65: Line 407: “ii” ◊ “II”.
Reply 65: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.
Changes in the text: see Page 25, line 536.

Comment 66: General comment: In many locations, the full stop is not followed by space mark; instead, the first capitalized letter is put. Please correct throughout the paper.

Reply 66: We thank the reviewer for this suggestion. We have revised the manuscript according to your suggestion.

Comment 67: General comment: what I also found lacking in this paper (which can

greatly increase its value) is the analysis focused on two transcription factors (TFs) that you have in your prognostic model, i.e. TFAP2A and GATA2. Moreover, these TFs were found to be opposite (the first as risk factor, the second as protective factor). This could be leveraged to some sort of analysis that can reveal what biological functions may be exclusively regulated by each of them, rendering specific observation on the clinical level. Could you perform analysis with proposed workflow: (a) acquire targets separately for TFAP2A and GATA2; I suggest using GTRD [https://gtrd.biouml.org]. (b) check what part of these targets are exclusive for TFAP2A and exclusive for GATA2 (Venn diagram might be helpful). (c) perform functional analysis (GO/KEGG) to check what these targets regulate; network (GeneMania? STRING?) might also reveal some interactions between these targets on various levels. (d) infer about potential effect that can have impact on the clinic, maybe some processes will be well-known for LUAD context? Such inclusion would immensely increase the value of study.

Reply 67: We thank the reviewer for this suggestion. Following your suggestion we added the corresponding analysis and presented it in Figure S2. TFAP2A has been shown to critically promote LUAD progression, while GATA2 expression is significantly reduced in lung cancer. Therefore, we further investigated the relationship of these two prognostic biomarkers and LUAD. Firstly, 1338 genes targeting TFAP2A and 5240 genes targeting GATA2 were obtained. After taking the intersection, 514 common target genes of TFAP2 and GATA2 were acquired (Figure S2A, Table S3). Then, we mined the potential function of target genes. The results revealed that TFAP2A was associated with pathways related to cell replication, protein modifications, and metabolic processes, such as negative regulation of cell cycle process, cell cycle checkpoint signaling, carbon metabolism, TCA cycle and so on (Figure S2B). Moreover, GATA2 was associated with RNA splicing, ribonucleoprotein complex biogenesis, mitochondrial matrix and immune-related pathways, and p53 signaling pathway and so on (Figure S2C). The PPI network of TFAP2A, GATA2, and their interacting genes showed strong interactions between TFAP2A and TFAP2D, GATA2 and ZFPM1 (Figure S2D). Furthermore, PPI network demonstrated target genes of GATA2 were mainly correlated to positive regulation of transcription by RNA polymerase II, endothelial cell migration, and cell differentiation and so on. Finally, we researched the discrepancies of TFAP2A and GATA2 to different sub-types in different clinical factors. The TFAP2A gene was significantly different in sex, T1 vs. T2, while GATA2 gene was significantly different in age, sex, T1 vs. T3 (Figure S2E). In summary, TFAP2 and GATA2 were linked with the prognosis of LUAD patients.

Changes in the text: see Page 212, line 248-253; Page 20-21, line 430-451; Page 23-24, line 503-512; Figure S2; Table S3.

Comment 68: General comment: I also suggest to contact native speaker to improve the paper thoroughly; some parts would benefit from it.

Reply 68: We thank the reviewer for this suggestion. Based on your suggestion we have touched up the manuscript and uploaded the proof of touch-up.

Comment 69: General comment: Is “Ethical Statement” necessary in this paper? If I understood correctly, you used public data from GDC/TCGA, as well as from GEO? The GSE dataset is deposited by other researchers?
Reply 69: We thank the reviewer for this suggestion. The datasets used for the analysis of this study are from the public databases GEO and TCGA, gene expression databases created and maintained by NCBI, USA. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 26, line 574-575.

Comment 70: Data availability statement: Change the explanation of “GEO” (Gene Expression Omnibus). Generally, I suggest to write it as follows: “The datasets analyzed for this study can be found in the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA)”. (remove “of USA” and “database”).
Reply 70: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 26, line 574-575.

Comment 71: TRIPOD checklist: correct typo “Abstrac” (lacks “t”).
Reply 71: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.

Comment 72: Figures are generally okay, but the font size is too small. Please correct wherever possible. To be more specific, in Figure 1 legend change “volcanic map” (volcano plot), correct description of subfigure 1C (this is definitely not Cox analysis but rather Venn diagram), and rewrite description of 1E to “Expression of differential prognostic LRGs”. In Figure 2, the text “GeneMania report” and the below parameters, as well as the words “1 of 11” should be deleted. If possible, please change the color palette for GeneMania network (in my opinion, the default one on the server is illegible, some colors are too similar). Moreover, I think that the entire Figure 2 could fit well within single row, or subfigure B could switch sites with subfigure C. Figure 3: in subfigure A, is the second lamba value truncated? Subfigure B: you can add gene names to corresponding coefficients. Figure 5: subfigure D – correct typo (prediced -> predicted). Moreover, I think that the current subfigure D should be C, while the nomogram should be subfigure D (change symbols, do not move graphical objects). Figure 7: the font in subfigure A must definitely be increased. Subfigure E: some “ns” could be moved higher so to not overlap with outliers. To conform with other markings, you can change “ns No significance” to “ns p>0.05” at the end of figure’s description. By the way, in heatmaps – the legend presents z-scores? Figure 8: subfigure A – explain abbreviations in the top-middle

part. Moreover, what is the purpose of bottom-left and bottom-middle graphs? Subfigure F – font must definitely be increased. In description of Figure 8 – is “Catastrophe landscape waterfall map” a proper name? Or maybe use something like “mutational landscape”? Figure 9: if there is only subfigure “A”, I suggest to remove it completely and leave only figure’s title.
Reply 72: Thank you for comments.According to your suggestion, we have modified Figure 1 and its legend.According to your suggestion, we have modified Figure 2.According to your suggestion, we have modified Figure 5D, because Figure D is the nomogram calibration curve and is used to evaluate the accuracy of Figure C. Therefore, we believe that nomogram should be Figure C and the calibration curve is Figure D.According to your suggestion, we have modified Figure 7.TNP trinucleotide polymorphisms, SNP single nucleotide polymorphisms, ONP polymorphisms oligonucleotide polymorphisms, INS insertion insertion insertion insertion, DEL deletion deletion.This figure shows the overall situation of TCGA-LUAD mutation data, and statistically describes various mutation types, respectively. Mainly, we look at the other two figures reflecting the differences between high and low risk groups. The left lower figure shows the total number of mutations in each patient. The ordinate indicates the number of mutations, and the abscissa indicates the number of patients.The bottom and middle panels show boxplots for different mutation types, with colors consistent with the notes in Figure I.According to your suggestion, we have modified Figure 9.
Changes in the text: see Figure 1; Figure 2; Figure 5; Figure 7; Figure 8; Figure 8.

Reviewer C

Li et al. explore lysosomal related genes (LRGs) as biomarkers for Lung adenocarcinoma (LUAD). Gene expression data from LUAD and “normal” tissue was obtained from the TCGA database. Differential expression from normal tissue of 144 genes previously associated with lysosomal diseases was determined using limma R package and with the criteria p-value 0.5. Cox regression survival analysis was used to obtain prognostic set of genes. Prognostic LRGs were compared to differential expression genes in a Venn diagram to find overlap, they plotted the heatmap of those genes. KEGG and GO were used to look at biological function, GeneMania was used to look at protein-protein interactions. Internal validation was made for the LRGs, LASSO logistic regression was done, Significant LRGs were used for multivariate Cox survival analysis for the Risk Model Construction. All LUAD cases were computed, cases were separated into high and low risk groups and ROC curves made plotting survival intervals. Validation was done to determine effectiveness. Further analysis was made.

Comment 1: Major Points: Construct a workflow chart of the model construction. Explain why Lasso logistic regression is used.

Reply 1: We thank the reviewer for this suggestion. Following your advice we added the flowchart, see Figure S1. LASSO is a shrinkage estimation method, whose basic idea is to minimize the residual sum of squares under the constraint that the sum of absolute values of regression coefficients is less than a constant, which can produce some regression coefficients that are strictly equal to 0 and further obtain interpretable models. We use LASSO algorithm to screen the model genes in the training set, set parameters famil as binomial and type.measure as class, realize LASSO logistic regression, select strong correlation features, and screen out 8 signature genes when the cross-validation error is the lowest.
Changes in the text: see Figure S1.

Comment 2: Major Point: Prognostic gene expression signatures for LUAD have been made, they have been through peer review and are published (e.g. Liu et al. Hindawi/ Oxidative Medicine and Cellular Longevity Volume 2020 | Article ID 8847226 | https://doi.org/10.1155/2020/8847226). To be fair, Li et al. follow this kind of methodology with some deviations looking at lysosomal "related" genes previously identified in lysosomal disease (not just cancer). However, what is being called "differential expression" in these studies rests on normalization using tissue that are "blood derived", TCGA did not have normal adjacent tissue in most of their studies. When cBioPortal computed TCGA "differential expression" they took the LUAD cases that had two normal chromosomes and compared other cases to those, which wasn't perfect either. Another method would be to compute a z score for the entire LUAD cohort and find cases significantly up and down expressed. To this reviewer, taking the raw expression data from blood and finding genes with a p value under 0.05 in LUAD that are different by a log 2fold-change of 0.5 comes across as a major experimental design flaw. However, there are studies that have been peer reviewed that accept it. Authors should address the criticism.
Reply 2: We thank the reviewer for this suggestion. Liu et al 's study is mainly based on the TCGA-LUAD dataset to screen prognostic genes, while this study is mainly based on lysosome-related genes to screen prognostic genes for LUAD and construct prognostic models, which will provide a theoretical basis for studying the effect of lysosome-related genes on the development and progression of LUAD.In differential expression analysis, p value is used to measure the significance of genes, Log2FC measures the fold difference of genes between two groups, and appropriate screening threshold can be selected according to the actual situation of the dataset when performing differential analysis.| Log2FC | > 2 and p.Adj < 0.01 was used as a screening threshold to obtain differentially expressed genes by the edgeR package.In this study, we selected the differential genes obtained by pvalue < 0.05 and | Log2FC | > 0.5 at this threshold and subsequent lysosome-related genes to take the optimal intersection, and if a stricter threshold is used (| Log2FC | > 1), it will lead to fewer differential lysosome-related genes obtained by the intersection, and then fewer prognostic lysosomal genes are screened to better reflect the role of differential

lysosome-related genes in the target disease, and if the threshold is lowered, the number of genes obtained will be higher and the reliability will be reduced.An analysis to adjust the threshold according to the actual situation is presented in Reference [PMID: 34225739], Reference [PMID: 29547407, PMID: 28058013] for this threshold (pvalue < 0.05 and | Log2FC | > 0.5) (1-3).

Comment 3:Minor (but important) sentence construction, grammar, and spelling issues throughout suggest the manuscript would benefit from an editorial service review. It is readable, but clearly a second language for the authors. Out of respect for the work, it would be worth the trouble.
Reply 3: We thank the reviewer for this suggestion. Based on your suggestion we have touched up the manuscript and uploaded the proof of touch-up.

Comment 4: Title: “Identification and prognosis of lysosomal associated genes in lung adenocarcinoma” Suggest instead: “Identification of lysosomal genes associated with prognosis in lung adenocarcinoma”.
Reply 4: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 1, line 1-2.

Comment 5: Abstract (Methods section)- instead of saying “limma” please identify as “R/Bioconductor software limma” in Abstract itself.
Reply 5: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 3, line 46-47.

Comment 6: Line 48, When authors write “...which were acquired by univariate Cox regression analysis and limma respectively” it means first they did univariate Cox Regression, and then they did something called “limma”. You can tell that cox regression was done using R/Bioconductor software (limma)- but it should be stated clearly.
Reply 6: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments. Differential prognostic LRGs were acquired by overlapping survival-related genes obtained via univariate Cox regression analysis and differentially expressed genes (DEGs) obtained via R package limma.
Changes in the text: see Page 2-3, line 44-47.

Comment 7: Line 52, “Then, rank-sum test, Tumor Immune Dysfunction and Exclusion (TIDE), and differential analysis were applied to immune checkpoint inhibitors (ICIs) to explore immunotherapy responses”. Probably more easily understood if you change to:
“Then, rank-sum test, Tumor Immune Dysfunction and Exclusion (TIDE), and

differential analysis were used to predict patient response to immune checkpoint inhibitors".
Reply 7: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 3, line 50-52.

Comment 8: Introduction -lines 88-90: "In the era of precision medicine, researchers have realized that identify of new therapeutic targets requires recognition key genes driving carcinogenesis". suggest "In the era of precision medicine, researchers have realized that the identification of new therapeutic targets requires recognition of key genes driving carcinogenesis".
Reply 8: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 5, line 105-106.

Comment 9: line 129: "...samples and 526 LUAD samples to screen out the differentilly...".
The word "differentially" is misspelled.
Reply 9: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 7, line 145.

Comment 10: line 135 "...and the heatmap of dufferential prognostic LRGs were plotted 136 by Heatmap (version 4.1.0) ". The word "differentially" is misspelled again, different spelling.
Reply 10: We thank the reviewer for this suggestion. We have added relevant content to the manuscript based on your comments.
Changes in the text: see Page 7, line 151.