Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-22-1000

Reviewer A

Comment 1: The rationale for the study has not been well established, and the main purpose is not clear in the introduction section. The manuscript would be strengthened by a more explicit purpose that indicates why this research is important. In addition, the authors should explain better the association between lung adenocarcinoma and histone modifications and what is the difference in this study compared with the previous one in that they collected the data. Also, the authors should write the mean of the LUAD abbreviation the first time it appears in the text.

Reply 1: In response to your question, we rearranged the introduction and introduced the relationship between lung adenocarcinoma and histone modification in detail (see Page 4, line 93). Compared with previous studies, we focusing on the acetylation modification regulators of histones and included multiple datasets of TCGA and geo databases for analysis in a larger sample size.
In addition, we write the mean of the LUAD abbreviation the first time it appears in the text. (see Page 3, line 68)

Changes in the text:
Various diseases, particularly cancer, involve epigenetic changes. Histone is the main protein component of chromatin and plays a central role in gene regulation, acting as a spool winding DNA. Histones undergo various modifications, including phosphorylation, acetylation, glycosylation, methylation, ubiquitination, and citrullination, which affect gene transcription. The importance of histone acetylation has been confirmed in cancer; it regulates gene transcription and cellular processes such as immune response, apoptosis, autophagy, cell cycle arrest, DNA damage repair, and metabolism. Abnormal histone acetylation is a hallmark of many cancers, particularly NSCLC.

Therapeutic agents targeting the epigenetic regulatory family of HDACs have demonstrated clinical success in treating some hematological malignancies. Preclinical models of melanoma and LUAD have demonstrated that HDAC inhibitors can upregulate the expression of PD-L1 and T-cell chemokines, thereby enhancing the sensitivity of immune responses to anti-PD-1/PD-L1 treatment and improving clinical

outcomes. Recent studies have reported that histone acetylation is closely related to tumor microenvironment (TME); specifically, histone acetylation regulates the cytolytic activity of $CD8^+$ T cells, polarization of macrophages, and immunosuppressive function of regulatory T cells.

Comment 2: In material and methods, it is mentioned survival R package was used. The authors should add more details about the pipeline used to better perform the prognostic model construction. The data retrieved from GEO was the raw data? What type of analysis was performed? The batch effect was performed to combine multiple datasets?

Reply 2: Thank you for your comments. In this study, we analyzed 36 histone acetylation regulatory genes, including the writers, the erasers and the readers. First, we downloaded mRNA expression profiles and corresponding clinical data of LUAD patients from the TCGA databases. Next, we analyzed the expression and mutation of these histone acetylation regulatory genes. After lasso cox regression analysis, 5 genes were screened to establish the prognosis model using the cox regression method, and they were used to build the prognostic model and calculate the risk score. The datasets we downloaded from the GEO datasets are all public original data. In order to be able to validate the prognostic model we constructed in more samples, we combined three datasets from the geo database. As there may be large technical differences between different batches of samples in the three datasets, which constitutes the batch effect and affects further analysis. The purpose of batch correction is to eliminate batch effect.

Comment 3. Material and methods should be revised for clarity; how was the process detailed to identify the 36 histone acetylation regulated genes? I suggest using some public repository (GitHub) to make available the pipelines for better clarification of the process involved in the analysis.

Reply 3: Thank you for your comments and valuable opinions. This is an oversight of our company. In our study, we searched the literature related to histone acetylation modification, collated and analyzed 36 recognized histone acetylation genes to determine different histone acetylation modification patterns. Histone acetylation is currently recognized to be regulated by two enzyme families: histone acetyltransferases (HATS) and histone deacetylases (HDACs). The main function of hats is to relax the concentrated chromatin, promote the activation of gene transcription, and provide binding sites for reading protein BRDs. In contrast to hats, HDACs remove acetyl groups from highly acetylated histones and inhibit gene transcription. HDACs can change the transcription of oncogenes and tumor suppressor genes by removing acetyl groups, reversing chromatin acetylation. Bromodomains (BRDs) are considered to be the first histone binding module. These proteins can specifically recognize acetylated lysine residues on histone tails, and can also bind acetylated lysine residues of non-histones. BRD containing proteins can be broadly divided into three categories

functionally: chromatin modifiers, such as acetylated writer P300, chromatin remodellers, such as SMARCA22 and SMARCA4, and chromatin readers, such as the BRT family of bromodomain proteins. The BET family includes BRD2, BRD3, Brd4, and BRDT, which regulate the transcription of oncogenes such as MYC and nut fusion oncoproteins. In this regard, we have supplemented this content in the material method. (see Page 5, line 122)

Changes in the text:
The literature related to histone acetylation modification was retrieved, and 36 acknowledged histone acetylation genes were curated and analyzed to identify distinct histone acetylation modification patterns.

Comment 4: In figure 1, the authors conclude that the "differential expression of acetylation regulatory genes in tumor and normal tissues and their high mutation plays an important role in the occurrence and progression of LUAD." How was this conclusion made? In the 214 samples with mutations, was the same mutation in all patients? Describe better the mutations and their impact on the patients.

Reply 4: Thank you for your comments. This is a mistake. Through our study, we found that up to 38.15% of the samples in the TCGA-LUAD cohort had mutations in histone acetylation regulatory genes, and the mutation type was mainly missense mutation, followed by nonsense mutation. Among the 36 histone acetylation regulation genes, the mutation rate of smarca4 accounted for 7%, and the mutation rate of HDAC9, BPTF, BAZ2B, and CREBBP was also higher than 3%. The high mutation rate indicates the instability of histone acetylation regulatory genes in the t TCGA-LUAD cohort, and suggests the potential relationship between histone acetylation modification and tumors. The specific internal relationship needs further experimental research. Therefore, we have modified and supplemented this content in the results. (see Page7, line 176)

Changes in the text:
In the TCGA-LUAD cohort, 38.15% of the samples had mutations in histone acetylation regulatory genes, and the mutation type was mainly missense mutation, followed by nonsense mutation. Among the 36 histone acetylation regulation genes, the mutation rate of smarca4 accounted for 7%, and the mutation rate of HDAC9, BPTF, BAZ2B, and CREBBP was also higher than 3%. The high mutation rate indicates the instability of histone acetylation regulatory genes in the TCGA-LUAD cohort, and suggests the potential relationship between histone acetylation modification and tumors.

Comment 5: Make figure 1 with a better resolution. Figure 7a is clear the differences, and compared figure 1A is hard to understand the differences in each gene and the statistical differences at the top of the figure.

Reply 5: Thank you for your comments. In response to your questions, we redrawn Figure 1 to highlight the expression differences of various genes and the statistical

differences at the top of the figure.

Comment 6: On page 10, line 273 authors mention “combined two cohorts,” but there are 3 different GSE databases. What databases were used for the analysis?

Reply 6: Thanks for your comments. I am very sorry for the mistake. We did merge three cohorts of GEO datebase. We have made corresponding modifications in the manuscript according to your requirements. (see Page11, line 290)

Changes in the text:
We combined three cohorts (GSE30219, GSE72094, and GSE50081) from GEO database to verify the above results.

Comment 7: The authors need to improve the English grammar across the text.

Reply 7: Thank you for your valuable opinions. This is also an aspect that we need to improve. In order to increase the readability of the article, we invited native English professionals to revise and polish the article.

Reviewer B

Comment 1: Introduction: When referencing cancer statistics, please make sure to include references. The reader cannot assess whether these numbers are local or global without proper references.

Reply 1: Thank you for your comments. This is an oversight of us. We are sure that the cancer statistics we quoted are from a source. At present, we have supplemented the reference materials. (see Page 3, line 66)

Changes in the text:
At present, the incidence and mortality rates of lung cancer ranks first among malignant tumors in China, and non-small-cell-lung-cancer (NSCLC) is the most common pathological type of lung cancer, accounting for 85% of all lung cancer(1). Lung adenocarcinoma (LUAD) is the most common subtype of NSCLC, accounting for 40%(2).

Comment 2: On line 86, you describe the methods of the study. The introduction sectin should be reserved for the background information, so these paragraphs should be moved to the methods section.

Reply 2: Thank you very much for your valuable comments on our writing. In response to your comments on the structure of the article, we have revised many parts of the

article. Not only adjusted the problematic paragraphs you pointed out, but also revised the citation part, supplemented a large amount of background information to explain the relationship between lung adenocarcinoma and histone modification, and indicated the importance of this study. (see Page 4, line 94)

Changes in the text:

Therapeutic agents targeting the epigenetic regulatory family of HDACs have demonstrated clinical success in treating some hematological malignancies. Preclinical models of melanoma and LUAD have demonstrated that HDAC inhibitors can upregulate the expression of PD-L1 and T-cell chemokines, thereby enhancing the sensitivity of immune responses to anti-PD-1/PD-L1 treatment and improving clinical outcomes. Recent studies have reported that histone acetylation is closely related to tumor microenvironment (TME); specifically, histone acetylation regulates the cytolytic activity of $CD8^+$ T cells, polarization of macrophages, and immunosuppressive function of regulatory T cells.

Because of technical limitations, current relevant research is mostly limited to one or two histone acetylation regulators and cell types; however, antitumor effect is characterized by the interaction of many tumor suppressors in a highly coordinated manner. In this regard, bioinformatic analysis is helpful. We developed histone acetylation modification scores to predict prognosis, clinical characteristics, and immune status in patients with LUAD. Our results suggested that histone acetylation modification score is a powerful prognostic marker that can accurately predict prognosis and immunotherapy response.

Comment 3a: TCGA stores RNA-seq data whereas the GEO repositories were the Affymetrix U133 array. These do not directly match. How did you ensure transferability between these two transcriptomic assessment methods?

Reply 3a: Thank you for your comments. In view of the inconsistency between the data platforms of your databases, although the two sequencing platforms are not consistent, theoretically speaking, the level of gene expression should be consistent. No matter which transcriptome evaluation method is used, the expression of a gene will not change due to the change of transcriptome evaluation method. We only consider the expression of multiple genes when constructing the prognosis model. Using this prognostic model, each sample can be scored and classified (high and low risk). At

present, there are many methods for transcriptome evaluation. We also hope that the constructed prognostic model can be better popularized in clinical application, and it is not necessary to use a certain sequencing method.

Comment 3b: As I understand it, you are actually not investigating histone acetylation although the title and the abstract suggests this, but rather transcriptomic changes in effector genes. How did you choose genes of interest, which did you choose and why?

Reply 3b: Thank you for your comments. We agree with you that the transcriptomic changes of histone acetylation are by no means the same as the actual changes, but subject to the current technical conditions, bioinformatics mostly focuses on studying the transcriptomic changes of effector genes, and changes in gene expression levels can be used as a method to predict the corresponding protein levels or functions. To determine the actual histone acetylation level, expensive histone acetylation sequencing is required. Currently, there is no available public database for acetylation modification. At the same time, it has to be mentioned that the prediction model of studying multiple genes at the same time greatly improves the accuracy of the prediction results compared with the change of the expression level of a single gene. Considering the expression difference between tumor and normal tissues and the difference between the expression of this gene and the prognosis of patients, the five core genes were screened out after dimension reduction by Lasso, and the scores of each gene were calculated. Among them, kat2b is considered to be the most important one of these 36 genes, which can be further studied in the future.

Comment 3c: If I understand your methods correctly, you included 36 gene expression levels in a cox regression model including some 497 patients. Whether you also include clinical variables (TNM stages etc) is not described, but is very important. Also, including such as vast amount of covariates drastically increase your risk of overfitting, thus bringing the validity of the results of the regression model into question.

Reply 3c: In our study, we did not use all 36 genes in constructing Cox regression model. Considering the expression difference between tumor and normal tissues and the difference between the gene expression and patient prognosis, we screened the five core genes after dimension reduction by Lasso, and calculated the score of each gene. Among them, kat2b is considered to be the most important one of these 36 genes, which can be further studied later. In fact, in constructing the Cox regression model, we only used the five core genes screened and did not include the clinical variables (TNM stage, etc.), so there is not much risk of overfitting in theory when the five genes are used as covariates. The 1 -, 3 -, and 5-year survival rates of patients can be better predicted by combining TNM stage, age, sex and other factors.

Comment 3d: Although not described in the methods section, you seem to be analyzing copy number variations in the results section. This is, at best, confusing. Also, I suddenly find the genes of interest listed here, although no explanation for how you

arrived at this set is provided.

Reply 3d: Thank you for your comments. This is an oversight. we obtained genomic mutation data (including somatic mutation and copy number variation) of TCGA-LUAD from the UCSC Xena database. (see Page 5, line 118). As for the histone acetylation regulation genes mentioned by you, in our study, we searched the literature related to histone acetylation modification and sorted and analyzed 36 recognized histone acetylation regulation genes to determine different histone acetylation modification modes. Histone acetylation is currently recognized to be regulated by two enzyme families: histone acetyltransferases (HATS) and histone deacetylases (HDACs). The main function of hats is to relax the concentrated chromatin, promote the activation of gene transcription, and provide binding sites for reading protein brds. In contrast to hats, HDACs remove acetyl groups from highly acetylated histones and inhibit gene transcription. HDACs can change the transcription of oncogenes and tumor suppressor genes by removing acetyl groups, reversing chromatin acetylation. Bromodomains (BRDs) are considered to be the first histone binding module. These proteins can specifically recognize acetylated lysine residues on histone tails, and can also bind acetylated lysine residues of non-histones. BRD containing proteins can be broadly divided into three categories functionally: chromatin modifiers, such as acetylated writer P300, chromatin remodellers, such as SMARCA2 and SMARCA4, and chromatin readers, such as the bet family of bromodomain proteins. The bet family includes BRD2, BRD3, Brd4, and BRDT, which regulate the transcription of oncogenes such as MYC and nut fusion oncoproteins. In this regard, we have supplemented this content in the material method.

Changes in the text:
Genomic mutation data, including somatic mutation and copy number variation (CNV) of TCGA-LUAD were obtained from the UCSC Xena database.

Comment Minor issues1: Please explain the abbreviation LUAD prior to using it in the text (both abstract and main manuscript)

Reply Minor issues1: Thanks for your comments. I am very sorry for the mistake. We have made corresponding modifications in the manuscript according to your requirements. We write the mean of the LUAD abbreviation the first time it appears in the text.