Peer Review File
Article information: https://dx.doi.org/10.21037/jtd-22-829

Reviewer A

In this study, the authors looked at 46 laboratory variables that are routinely tested before surgery in lung cancer, to see if any of these variables are associated with the presence of KRAS mutation in non-small cell lung cancer (NSCLC). KRAS mutation detection is indeed important in NSCLC for guiding treatment. It is an interesting study to see if any of these correlate with the presence of KRAS mutations. Although I am not sure of the clinical applicability of the prediction model (because if the tumor is already obtained by surgery, it is easy to perform sequencing to test for KRAS), the study may reveal interesting biological insights into KRAS-mutant NSCLC. However, many factors can confound the results and hence need to be addressed carefully.

1. Some of the parameters used in prediction, such as high-density lipoprotein cholesterol may be confounded by common comorbidities in the age group of the patients, such as high cholesterol, heart diseases, or diabetes. Similarly, drugs such as cholesterol-lowering drugs, for example, can affect many of the 46 parameters. The authors need to assess if this affects the validity of the results.

Reply 1: Thanks for your question. We must acknowledge that comorbidities and medication used in patients could have an impact on the results of the study. We therefore included a sample size large enough to counteract the bias caused by individual differences as much as possible. In addition, the data were collected from the last preoperative blood sample of patients with an average age of 60 years who had been evaluated by clinicians as suitable for surgery. It could be proved that the majority of patients did not have serious complications that affected the findings. As you said, patients might be taking medication to reduce indicators such as cholesterol, resulting in some measures that fail to show differences in the study. In Reply to this issue, we will carry out more research to explore the biological phenomena related to the type of gene mutation. However, it also illustrated that those results that were statistically significant may have been more different when patients were not taking the drugs.
Changes in the text: Line 352-354: We included a sample size large enough to counteract the bias caused by individual differences as much as possible.

2. I understood that the blood results were derived from samples taken right before surgery. Perhaps the authors will also like to clarify if any of these patients received any treatment prior to surgery.

Reply 2: Thanks for your valuable revision. The data in the training cohort came from primary surgery patients who had not undergone anti-tumor therapy. We have revised and

supplemented in the manuscript.
Changes in the text: Line 122-123: All samples were obtained from the last preoperative blood test of primary surgery patients who had not undergone anti-tumor therapy.

3. “5 gene mutations” (line 163) actually refers to only EGFR, KRAS and ALK (3 mutations?) because “EGFR19del” and “L858R” are both mutations on EGFR. If all the KRAS mutations are analyzed as one group, why did the authors separately analyze EGFR19del and L858R?

Reply 3: Thanks for your question. EGFR 19del and 21 exon L858R mutations were previously considered to be a class of disease, but it has been confirmed that there are large differences between them in the wake of more and more studies data. First, they had distinct molecular structures. Their kinase domain spatial configuration was located in different regions, which might affect the phosphorylation, leading to discrepancy on subsequent functions. Secondly, the clinical characteristics of them were various. Patients with 19del were more dependent on the EGFR pathway and were significantly more sensitive to the drug than those with the 21 exon L858R mutation. In addition, the L858R mutation was associated with further mutations genes that might lead to bypass activation. Therefore, we believed that 19del and L858R mutations might have diverse biological characteristics, and it was more appropriate to study them separately.
Changes in the text: Line 285-296: EGFR 19del and 21 exon L858R mutations were previously considered to be a class of disease, but it had been confirmed that there were large differences between them in the wake of more and more studies data. First, they had distinct molecular structures. Their kinase domain spatial configuration was located in different regions, which might affect the phosphorylation, leading to discrepancy on subsequent functions. Secondly, the clinical characteristics of them were various. Patients with 19del were more dependent on the EGFR pathway and were significantly more sensitive to the drug than those with the 21 exon L858R mutation. In addition, the L858R mutation was associated with further mutations genes that might lead to bypass activation. Therefore, we believed that 19del and L858R mutations might have diverse biological characteristics. In this study, EGFR 19del and EGFR L858R were considered as independent genotypes as study factors.

4. The authors mentioned that “…17 showed significant inter-group differences..” (line 213), however, 9 variables were eventually used to construct the formula. Because not all readers are experts in statistical methods, it will help if the authors explain in simple terms why this was so.

Reply 4：18 haematological parameters showed significant inter-group differences in univariate analysis, among the 46 laboratory variables. Since the 46 laboratory variables of a sample were derived from the same patient, there were strong multicollinearity relationship. The influence of multicollinearity should be taken into account when building the model, therefore we applied LASSO analysis to filter the variables to minimize the potential collinearity and the confusion caused by overfitting. In the end, 9 meaningful variables were obtained.

5. The authors mentioned that the method could be applied to colorectal cancer. There is a paper published in 2020 by Cao Y et al (The value of haematological parameters and serum tumour markers for predicting KRAS mutations in 784 Chinese colorectal cancer patients: a retrospective analysis). Given that the themes are similar, perhaps the authors will like to cite and comment on their results in view of the results they have obtained.

Reply 5: Thanks for your significative suggestion. We carefully read the study published in 2020 by Cao Y et al and added citations and comments on their results to the manuscript.
Changes in the text: Line 257-266: A similar study by Cao Y et al in 2020 showed that hematological parameters (WBC, MON, MLR, HCT, HGB, AVEMPV, MCH, MCHC) were significantly associated with KRAS gene mutations in colorectal cancer. Similar with our conclusions, the values of hematological parameters in patients with KRAS mutation were lower than in wild-type patients. The widespread distribution of KRAS gene might contribute to worse physical conditions in patients with KRAS mutations than wild-type patients through multiple pathways. However, metabolism-related parameters were not included in the study. Therefore, our study provided new evidence that KRAS gene mutations might affect metabolic processes in humans specifically.

6. In figure 1, there is a heat map but the columns are not labelled. To improve clarity, the authors can consider highlighting the 9 parameters that were eventually selected to construct the formula, on both the heat map and the graphs in figure 1B-1R

Reply 6: Thanks for your valuable revision. We are sorry to make this mistake. We have added the columns on the heat map. In addition, we highlighted the 9 parameters on the heat map and 9 parameters in figure 1B-1S as your suggestion.

7. In table 2, the nine parameters used in the formula are listed. Will the authors please add a column to indicate the unit of measurement so that readers can use the formula too?

Reply 7: Thanks for your suggestion. We added a column in table 2 to indicated the unit of measurement so that readers can use the formula.

8. "…showed worse levels" (line 279). Will the authors clarify what they mean by worse levels?

Reply 8: Thanks for your question. A worse level means that the value of the parameters deviates more from the normal value. For example, the median and overall level of UA in patients with KRAS mutation were higher than those with other gene types.
Changes in the text: Line 299-302: The values of the metabolic-related parameters (such as TG, HDL, UA, etc.) in patients with KRAS mutation showed more deviated from the normal values compared to the remaining four mutant lung cancer patients.

9. The authors discussed in detail about KRAS and its effect on cell metabolism. However, some of these parameters, such as basophil and eosinophil count may not be related. Perhaps

the authors can discuss briefly on this observation.

Reply 9: Thanks for your question. We added a brief description of the elevation of basophil and eosinophil count in patients with KRAS mutations in the manuscript.
Changes in the text: Line 302-305: Some of these parameters, such as basophil and eosinophil count might not be related to cell metabolism. We considered that the abnormality of these parameters was associated with immuno-inflammatory phenotypes and immunogenic enhancements in KRAS mutant lung cancer.

10. Will the authors provide a table of the parameters for each patient to derive the results? So that the readers can use the formulas if they are interested.

Reply 10: Thanks for your question. Here is the formula for calculating the probability of KRAS mutation: probability of KRAS mutation= (triglyceride x 0.491) + (partial arterial oxygen pressure x -0.034) + (uric acid x 0.007) + (basophil count x 13.8) + (fibrinogen x 0.352) + (standard bicarbonate x -0.159) + (high density lipoprotein cholesterol x -1.723) + (alpha-L-fucosidase x 0.070) + (eosinophil count x 1.816). Readers can make their own calculations if they are interested.

Reviewer B

KRAS-driven tumors have been reported to be associated with metabolic reprogramming. In the current study, the authors report on the association KRAS mutation status and blood parameters and have developed a model that can predict the presence of a KRAS mutation in lung cancer.

Major comments
The claim that model can aid in therapy selection and reduced costly molecular testing is overreaching. The sensitivity and specificity of the formula is too low, i.e. both <0.95, to safely not test tumors for actionable targets. All patients predicted to not have a KRAS mutation require sequencing to determine presence of actionable targets. In addition, when a KRAS mutation is predicted specifics of the KRAS mutation (i.e. p.(G12C) yes/no) that is important for KRAS-targeted therapy currently in second line remains unknown. Although the research nicely confirms that KRAS-driven lung cancers are biochemically distinct from non-KRAS-driven tumors, the authors provide no evidence for the clinical utility of this algorithm in predicting which patients require molecular testing. Only at the end of the paper, the authors state that their screening model predicts KRAS status in lung cancer patients, but more expensive laboratory test (incl. sequencing?) are required. Then what is the use of the KRAS assessment based on blood values, even in a financially challenging environment?

Reply: Thanks for your question. We acknowledge that there was considerable room for improvement in the model. We supposed that our model might provide some auxiliary value

for some patients with financial difficulties who refuse to undergo genetic testing. The model could provide evidence for the early application of immunotherapy, assuming clinicians need to treat patients empirically for whom genetic testing is not available. In addition, the model could also remind clinicians to monitor patients more closely on the effectiveness of treatment.