Peer Review File

Article information: http://dx.doi.org/10.21037/tcr-20-2525

Reviewer A: This manuscript, written by Dr. Chunhui Zhou et al., with the title of " Identification of key mutations in central nervous DLBCL by comprehensive analysis between sequencing and TCGA database" describes the mutational profile of 12 cases of DLBCL from the central nervous system. Their primary data is compared with data from the TCGA database, which is publicly available. The manuscript is in general well written, it is well organized and have clear figures and tables.

Comment 1: The authors should make an additional effort to improve the description of the materials and methods. Please confirm that all the reagents that have been used have been properly described. Since the mutational profiling is the most important, please also expand the description of the methods, including the criteria for mutation and the variant classification.

For example, in some antibodies from CST there is discrepancy between the antibody clone and the catalog number.

CD20 (E7B7T) XP® Rabbit mAb #48750

CD20 (L26) Mouse mAb #74332

CD79A (D1X5C) XP® Rabbit mAb #13333

CD3ε (D7A6E™) XP® Rabbit mAb #85061

BCL6 (E5I8I) Rabbit mAb #89369

TNFRSF8/CD30 (E4L4I) XP® Rabbit mAb #54535

Syndecan 1 (D4Y7H) Rabbit mAb #12922

Reply 1: Thanks for the reviewer's suggestions, in the revised manuscript, we added and corrected related information of reagents (see Page 5, line 83-93), which were marked red.

Comment 2: Lines 132-133, "immunohistochemistry revealed that 12 cases were positive for CD10, Bcl-6, and MUM1 staining". Please confirm this statement. From the text I understand that all the cases have an immunophenotype positive for the 3 markers. If it is CD10 positive, the DLBCL will have a cell-of-origin subtype of germinal center B-cell-like (Hans' algorithm). In the table 1 it is written that case 4 is GCB and that 7 and 11 is ABC (activated). This is contradictory. If the IHC has been performed and the data is available the table could be completed better.

Reply 2: Thanks for reviewer's question. In the original manuscript, we didn't make the analysis of CD10, Bcl-6, and MUM1 staining clear. In the revised manuscript, we added the description of DLBCL typing in the 'Materials and Methods' section (see Page 5, line 90-93) and changed the statement in the results section (see Page 8, line 148-150).

Comment 3: Please compare your results with the mutational profile of the same subtype from other groups.

Reply 3: Thanks for the reviewers' important comments. In fact, before this study, we have retrieved the research concerning central nervous DLBCL. Finally 5 most relevant studies 5 were enrolled, including: Pouzoulet et al [1]. Klanova et al [2], Schmitz et al [3], Chapuy et al [4], Wright et al [5]. Among them, Pouzoulet collected 6 cases of PCNSL, and the CD79A, CD79B, PIM1, and Card11 were the most significant mutation. The study of Klanova identified the CDKN2A, MYD88, and CD79B as the most prominent mutation in PCNS. Our study identified MYD88, PIM1, CD79B, and BTG1 as an important mutation gene by sequencing combining the analysis of TCGA database. Together above results, there are many similarities between our and other studies, while some different mutations also identified. All these findings provide important view for comprehensive understanding the relationship between PCNSL and genetic mutation. These descriptions were added in the section of discussion (see Page 12-13, line 242-251).

Comment 4: Please provide more discussion about the most frequently mutated genes, their biological function and the possible pathological function in this tumor.

Reply 4: Thanks for reviewer's important comments. In the revised manuscript, we added some description of the high-frequency mutation genes in the discussion section, and marked it in red (see Page 12, line 227-232).

Reviewer B: The authors of this manuscript present an interesting characterization of a very rare subset of DLBCL by characterizing mutations in a small cohort of 12 CNS DLBCL cases. While the results presented here may serve as a useful validation of previously published work, the fact that the number of

CNS DLBCL cases is not large enough to draw strong conclusions, and that the mutated genes appear to be largely in-line with previously reported findings for CNS-DLBCL significantly hamper my enthusiasm for this manuscript.

Major comments

Comment 1: The authors should discuss Klanova M, et al. Blood 2019. The results from Klanova et al. are similar to those presented here, with the same number of CNS DLBCL cases and some useful characterization of the recurrently mutated genes.

Reply 1: Thank for the reviewer's professional and helpful comments. According to this comment, we carefully read and analyzed Klanova M's research. As you can see, DLBCL in the central system is indeed a relatively rare case. Therefore, even in the GOYA study, the author only screened out 12 cases for the mutational profile analysis of CNS DLBCL. It is worth noting that the Klanova M' study found that MYD88, CDKN2A, and CD79B are highly mutated genes, which are very similar to our results, which indicating that the Mutational profile analysis of CNS DLBCL based on the Chinese population have many similarities with foreign researches. Meanwhile some differences in results (such as PIM1 and BTG1) also reflect the feature between different studies. Therefore, our results provide an important supplement for previous studies concerning mutant genes of CNS DLBCL. Above discussions have been put in the revised manuscript and marked in red (see Page 12-13, line 242-251). Thanks for the important suggestions of reviewer again.

Comment 2: How did the authors process the sequencing data? Are silent mutations included in their analysis of recurrently mutated genes? Do they do anything to identify likely pathogenic mutations? They need to include significantly more detail in their methods to describe this information. And, if they have not made any effort to filter the mutations, they should.

Reply 2: Reviewer put forward an important suggestion. Due to our negligence, the sequencing method was not described in detail in the section of materials and methods. In the revised manuscript, we have supplemented the relevant methods in this section and marked them in red (see Page 6, line 113-120).

Comment 3: The authors' focus on recurrently mutated genes doesn't seem like the most relevant comparison. It would be much more valuable to directly compare the prevalence of mutations in CNS DLBCL with primary DLBCL. They start to do this analysis in section 4, but fall short by focusing on similarities, rather than differences between primary and CNS DLBCL cases. In fact, the analysis that is presented seems of little interest. Why do we care about the genes in common between the two, except as a validation that their sequencing method is reasonable? It's the differences that are of prime importance in understanding CNS DLBCL.

Reply 3: Thank for reviewer's comments. Indeed, from the view of scientific significance, it is more important to compare the mutation between CNS DLBCL and primary DLBCL. However, due to the difference in detection methods, enrolled populations, and process of collected samples between different studies, the identified results were also diverse. Therefore, the main purpose of this research is to identify some common mutant genes so that revealing the pathogenesis of CNS DLBCL. On the other hand, by the method of integrating sequencing results with TCGA database to find common and important mutations have also been used in many studies, such as Hu Y [6], Zhang R [7]. Anyhow for the significance of this study, we should indeed compare CNS DLBCL and primary DLBCL mutant genes. However, patient specimens were directly collected in central nervous system, and the primary lesions were not obtained, leading to a limitation to further detection. In future research, we wish a more complete research could be carried out.

Comment 4: To the point above, the mutational heterogeneity of DLBCL has been extremely well-studied (Intlekofer et al. Blood Cancer J 2018, Morin et al. Nature 2011, Pasqualucci et al. Nat Genet 2011, Bolen et al. Haematologica 2019), and a number of recent papers have described subsets of DLBCL based on their mutational profiles (Schmitz et al. NEJM 2018, Chapuy et al. Nat Med 2018, Wright et al. Cancer Cell 2020). Notably, none of these papers are referenced by the authors. If the goal of the TCGA analysis was to validate that the mutations they found were reasonable, they should consider starting there. The mutational subsets described by Schmitz, Chapuy, and Wright would also be extremely

helpful for further characterizing the differences between primary and CNS DLBCL. In fact, the authors should consider using the classifier from Wright et al to characterize the mutational subtypes of their 12 samples.

Reply 4: Thanks to the reviewers for this comment. In the discussion section of the revised manuscript, we compared and discussed the above related research with our study, so that further expand the scientific significance of this research (see Page 12-13, line 242-251).

Comment 5: By the same token, both the introduction and discussion feel largely incomplete. There's no discussion of the etiology of CNS DLBCL, nor of Cell of Origin or the different mutational subtypes referenced above, all of which would be highly relevant for these results.

Reply 5: Thanks to the reviewer's comments. In the introduction of revised manuscript, we added the relevant description and marked it in red (see Page 3, line 49-61).

Comment 6: Based on figure 2F -- are two of the CNS DLBCL cases completely unmutated? There is no discussion of this, and it seems likely that this is a result of sample failure.

Reply 6: Thanks for reviewer's question. For the 12 samples obtained, we were actually failed to detect the insertion-deletion. However, since other type of mutations was detected, we then think this reason was mainly due to the characteristics of the sample itself. Related descriptions have been put in the discussion section (see Page 11, line 216-220).

<u>Minor comments</u>

Comment 7: Intro, line 41 – you mention gene expression profiles, but the rest of the intro is on mutations. Did you mean "somatic mutation profiles"?

Reply 7: Yes, the "gene expression profile" that we mentioned means the "somatic mutation profiles".

Comment 8: The “sequencing experiment” methods section is confusingly written. Did the authors perform whole exome sequencing – i.e. encompassing the entire genome – or only perform targeted exome sequencing of ~50 genes?

Reply 8: Thanks for the reviewer's question. In fact, in the original manuscript, we have shown the sequencing method that “our panel was designed to encompass the whole exon regions (including parts of the intron regions)” (see Page 5, line 106-108).

Comment 9: Results, section 2 – “the pattern figure was performed.” Are the authors referring to mutational signature analysis? Or I guess the trinucleotide context of mutations. They don’t discuss this analysis in the methods section.

Reply 9: Thanks for reviewers’ comments. What we are referring to in this section is Single nucleotide variant (SNV) analysis, which is the analysis for variation caused by the substitution of a single nucleotide in the genome. In the original manuscript, we failed to make it clear in the material method. In the revised manuscript, we added description in detail (see Page 6, line 113-120).