Genes & Diseases (2023) 10, 37–40

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

RAPID COMMUNICATION

Simplifying genetic classifiers by six mutated genes in diffuse large B-cell lymphoma

Although diffuse large B-cell lymphoma (DLBCL) is considered as a curable disease after standard immunochemotherapy, approximately 30% of patients would succumb to short survival. To date, clinical presentations and gene expressions have been used to identify these high-risk patients. Recently, genetic alterations are used to model their proposed classifiers. For example, four, five and seven genetic subtypes were respectively found by Schmitz,[1] Chapuy[2] and George[3] colleagues as a method for tailored treatment. However, these classifiers were calculated using the complex tools such as the LymphGen classifier and consensus clustering algorithm.[2,3] Besides, mutated genes were not enough to estimate their genetic subtypes by their proposed tools, which need additional data such as copy number alteration and chromosome translocations. What's more, there is still not a consistent classifier to validate predefined subtypes. Thus, it is difficult to apply these genetic tools in clinical practice. In order to translate into clinical practice easily, a smaller gene-set using the easy-to-implement method will be practical instead of measuring genome-wide sequencing. In this context, we searched for a small number of genes using targeted next-generation sequencing, and established its predictive ability for DLBCL patients.

This study was approved by the Institutional Review Board (No. IIT20210082B-R2) and registered at chictr.org.cn (ChiCTR2100046974). The study design is seen in Figure 1A. The median age was 58 years (interquartile range, 50–66 years), and 46 (45.5%) patients older than 60 years were included. Clinical characteristics are shown in Table S1. The patients with a higher level of IPI, Non-GCB, or DEL had shorter progression-free survival (PFS) and overall survival (OS) (Fig. S1–S3). A total of 446 genes (Fig. 1B, Table S2) were measured in 101 patients. Additionally, 1210 somatic mutations, including non-sense, missense, splice site, and frame shift mutations were found along with a median of 11 (range: 1–42) mutations per sample (Fig. S4). The most recurrently mutated genes were *PIM1* (35.6%), *KMT2D* (26.7%), *CD79B* (22.8%), *MYD88* (22.8%), *KMT2C* (21.8%) (Fig. 1C and Table S3). Sixty-five genes with the mutated frequency more than 5% were enriched in 72 KEGG pathways such as B cell receptor signaling pathway, NF-κB signaling pathway, Notch signaling pathway and so on (Table S4).

We also detected 13 driver genes: *EP300, MYD88, CD79B, BCL10, IKZF3, CCND3, MEF2B, BCL2, DTX1, PIM1, IRF4, SYK* and *CDKN2A* (Table S5). Notably, *DTX1* coexisted with *PIM1* mutations, and excluded with *BCL2* mutations (Fig. S5). Additionally, most of them were frequently predominant in the patients with cell-of-origin (COO), double expresser lymphoma (DEL), rearrangements of *BCL2, MYC,* and *BCL6* (Fig. S6–S10), respectively. These results implied genetic sequencing can further improve classifications of DLBCLs. Nevertheless, there is no consensus classification in clinical practice now. The underlying reason of the discrepancy might be the differences in sequencing strategies, sequencing panel, biopsy material, statistical methods, sample size and studied population. When we compared significantly enriched genes across studies, some subtypes shared the same founder genes. For example, *BCL2* fusion and *EZH* mutations had contribution respectively from EZB subtype of Schmitz[1] and George W. Wright et al,[3] C3 cluster of Chapuy et al,[2] and BCL2 subtype of Stuart E. Lacy et al.[4] Based on these overlapping genes across different classifiers, Lucia Pedrosa et al created the "2-step" method to construct five genetic subtypes ($ST2^{2-S}$, $N1^{2-S}$, MCD^{2-S}, $BN2^{2-S}$, EZB^{2-S}; Table S6).[5] Owing to *BTG1*, *PIM2* and *UBE2A* mutations were not included in our sequencing panel, we modified this classifier by replacing *BTG1* with *BTG2*, *PIM2* with *IRF4*, *UBE2A* with *ETS1*, respectively; and added *MYC* translocations into EZB^{2-S} subtype and named it as $EZB^{2-S\ MYC+}$. Although genes, even of very similar

Peer review under responsibility of Chongqing Medical University.

https://doi.org/10.1016/j.gendis.2022.03.014

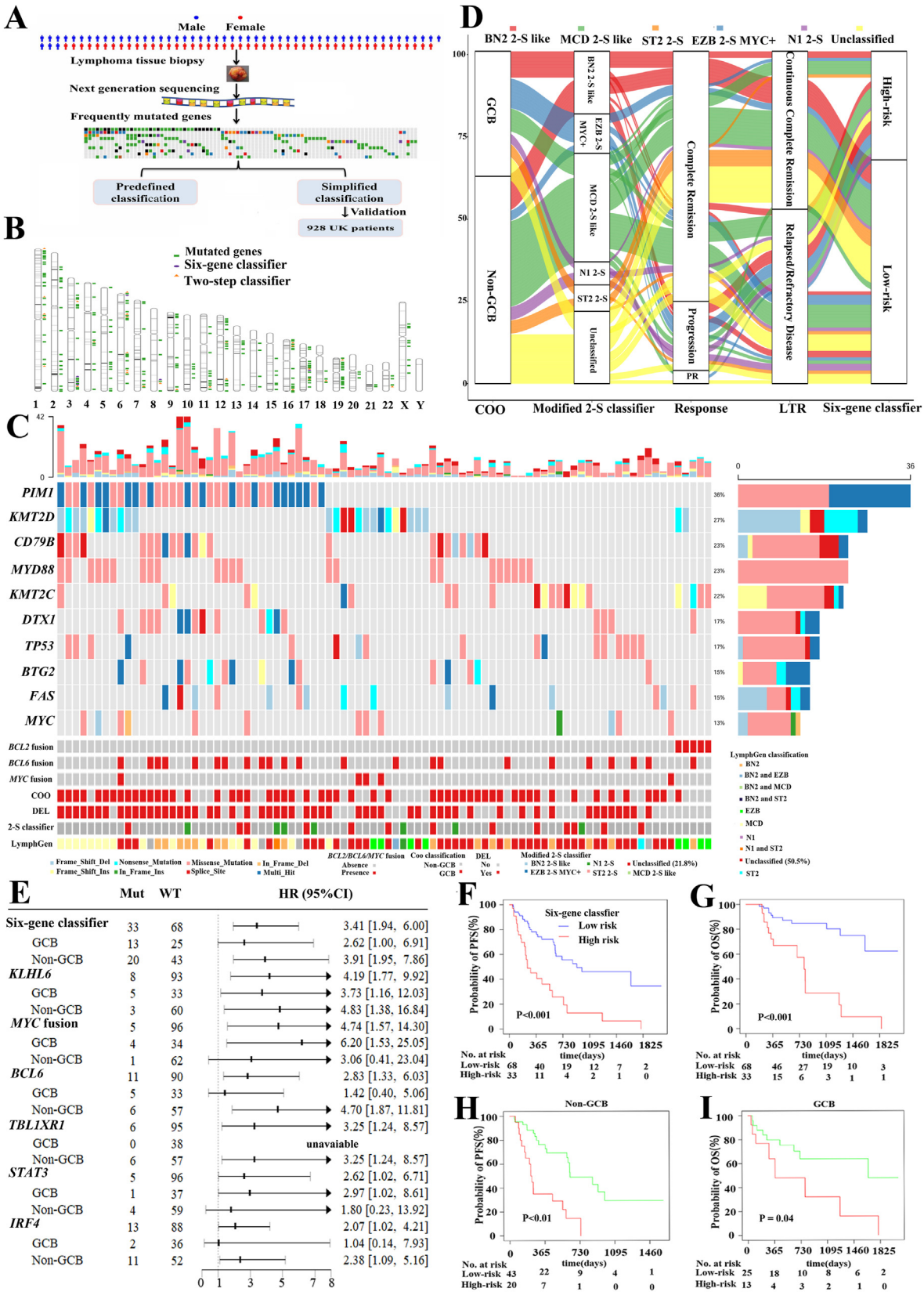

Figure 1 Six-gene classifier with prognostic significance was identified in DLBCL. We set up a simplified classifier using the training and validation study design **(A).** A total of 101 DLBCL patients were enrolled to perform targeted sequencing. The simplified classifier was studied in our patients and then validated in 928 UK patients. The founder genes were analyzed in this study **(B).** A total of 446 predefined cancer-related genes (black lines in chromosomes) were used to search for genetic mutations, and 222

function (for example, *BTG2* is very biologically similar to *BTG1*), should not be used interchangeably for comparisons, and the reasons that we modified the "2-step" Classifier were as following. Firstly, genes with the frequent mutations were selected for replacement of those undetected genes in the same predefined subtypes; secondly, the addition of *MYC* translocation was not only because it was most enriched within the BCL2 cluster,[4] but also it contributed to the worse prognosis for the EZB subtype; thirdly, compared with the AIC cluster algorithms, the modified subtypes had the highly sensitive and specific classification for their respective genetic risk groups (Fig. S11). As a result, 19 (18.8%) patients in $BN2^{2-S\ like}$, 12 (11.9%) $EZB^{2-S\ MYC+}$, 33 (32.7%) $MCD^{2-S\ like}$, 7 (6.9%) $N1^{2-S}$, 8 (7.9%) $ST2^{2-S}$ and 22 (21.8%) unclassified subgroups were found (Fig. 1C). Consistent with previous reports,[1] $MCD^{2-S\ like}$ and $N1^{2-S}$ subtypes were dominated by non-GCB cases, and $EZB^{2-S\ MYC\ +}$ subtype mostly enriched in GCB cases (Fig. S12; Fig. 1D). The similar relationship was also observed in DEL subgroups (Fig. S13). Finally, we got five subtypes whose PFS and OS are significantly different. The survival of $ST2^{2-S}$, $MCD^{2-S\ like}$ and $BN2^{2-S\ like}$, $N1^{2-S}$ and $EZB^{2-S\ MYC\ +}$ decreased in turn (Fig. S14). This modified "2-step" classifier was independent of COO classification but not IPI (Table S7, 8). However, approximately 20%—50% of patients could not be assigned to the modified "2-step" and LymphGen classifiers based on our mutated data (Fig. 1C). Additionally, some of genetic subtypes were related to COO classifications, which will confound their prognostic significances. Therefore, we conducted stratified analyses by COO classifications. Here, we found out *MYC* translocation, *KLHL6*, *STAT3*, *BCL6*, *IRF4*, and *TBL1XR1* mutations with the prognostic significance (Table S9 and Fig. 1E). *MYC* fusion is known to be associated with poor prognosis of DLBCLs; *KLHL6* mutation favors DLBCL cell growth; *BLC6* alterations could promote the development of lymphoma; *STAT3*, *IRF4* and *TBL1XR1* mutations were detected in refractory/relapsed DLBCLs. Next, using the hazard ratio-weighted scoring, we incorporated prognostic information from these six genes to devise a prognostic scoring model. Specifically, one point was assigned to each patient if one of the 6 signatures was present (Table S9). The score of each patient was then calculated as the sum of aforementioned signatures present in each case. Subsequently, each patient was assigned a sum score of 0—3 points: 0 ($n = 68$), 1 ($n = 23$), 2 ($n = 5$), and 3 ($n = 5$). These 4 groups were divided into two risk-groups based on the similarity of the intergroup survival curves (Fig. S15). Finally, this genomic scoring system can distinguish low-(score 0) and high-risk (score:1—3) disease with significant differences in PFS and OS (Fig. 1F, G). Notably, this classifier can further stratify GCB and Non-GCB patients into two groups with distinct outcomes (Fig. 1H, I). In multivariate analysis, we found this novel genetic classifier was independently associated with OS [HR (95%CI), 3.655(1.603, 8.33), $p = 0.002$] and PFS [HR (95%CI), 2.482(1.396, 4.414), $p = 0.002$] after adjusting IPI, COO and DEL in our cohort (Table S10, 11). Notably, AUC values increased from 0.66 to 0.71 with lengthening follow-up times, implying this novel classifier had the moderate ability to predict long-term PFS (Fig. S16).

Furthermore, this simple classifier was validated in a contemporary UK "real-world" population-based patient cohort (Fig. S17, S18).[4] Even considering the underlying confounders like IPI and COO classification, the presence of 6-gene alterations was independently associated with poor OS [HR (95%CI), 1.31(1.093,1.571), $P = 0.003$] and PFS [HR (95%CI), 1.284(1.047, 1.574), $P = 0.016$] in the 928 DLBCL patients (Table S10, S11). Taken together, our 6-gene classifier and its prognostic scoring system are applicable to predict clinical outcomes in the large independent cohort of DLBCLs. In order to investigate essential pathways related to the high risk group, we reanalyzing the TCGA dataset with 37 DLBCL patients. We found 628 genes downregulated and 1352 genes upregulated in 16 patients as the high risk group (Fig. S19). Notably, the immune health related pathways such as ECM-receptor interaction, hematopoietic cell lineage, rheumatoid arthritis, systemic lupus erythematosus, tuberculosis and autophagy were downregulated in the high-risk group (Fig. S20), implying the anti-lymphoma immunity and autophagy should be defective in these patients. Therefore, these patients might benefit from bispecific T-cell engager therapy and/or chimeric antigen receptor (CAR) T cells to restore effective anti-tumor immunity. However, autophagy has a divergent role in lymphoma, and caution should be taken when the modulation of autophagy is considered as a therapeutic strategy.

In conclusion, we set up a 6-gene panel as an effective tool to predict clinical outcome for DLBCL patients.

Author contributions

JHW, FY and JJ designed the research and/or analyzed the data. WWW, YS and JRY carried out the molecular genetic studies, LPM, WJY, and WZX provided clinical data. JHW, YF and JJ wrote the manuscript. All authors read and approved the final manuscript.

mutated genes (blue boxes along the chromosomes) were identified in this study. Six purple dots represent six genes (*MYC*, *KLHL6*, *STAT3*, *BCL6*, *IRF4*, and *TBL1XR1*) used to devise the six-gene classifier and 28 pink triangles represent genes used for the modified "2-step" classifier. Heat map illustrates top 10 mutated genes along with the predefined predictors **(C)**. Patients with the unclassified subtype account for 21.8% and 50.5% by the modified "2-step" classifier and LymphGen classification tool, respectively. The relationships among different genetic subtypes, COO classification, and treatment response were illustrated in **(D)** Cox stratified analysis of PFS by COO classification was used to identify six genes **(E)**. Because patients with *TBL1XR1* mutations did not exist in the GCB subgroup, the stratified analysis was not conducted and labeled as "unavailable". Survival curves of DLBCL patients for PFS **(F)** and OS **(G)** by the six-gene classifier. Survival curves of PFS and OS by the six-gene classifier in non-GCB **(H)** and GCB **(I)** DLBCL patients. LCR, Long term responses. RRD, relapse and refractory disease. COO, cell-of-origin.

Conflict of interests

Authors declare no conflict of interests.

Funding

This work was supported by Zhejiang Provincial Natural Science Foundation of China (No. LY19H080009, and LY20H080008); Pediatric Leukemia Diagnostic and Therapeutic Technology Research Center of Zhejiang Province (China) (No. JBZX-201904). This study is also supported by the major projects of Zhejiang Provincial Department of Science and Technology 491010+J3210NZJ (China) and the Research Project of Jinan Microecological Biomedicine Shandong Laboratory, Jinan, China.

Acknowledgements

We would like to thank the patients for donating specimens.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2022.03.014.

References

1. Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med*. 2018;378(15):1396—1407.
2. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24(5):679—690.
3. Wright GW, Huang DW, Phelan JD, et al. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer Cell*. 2020;37(4):551—568.
4. Lacy SE, Barrans SL, Beer PA, et al. Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report. *Blood*. 2020;135(20):1759—1771.
5. Pedrosa L, Fernández-Miranda I, Pérez-Callejo D, et al. Proposal and validation of a method to classify genetic subtypes of diffuse large B cell lymphoma. *Sci Rep*. 2021;11(1):1886.

Jinghan Wang[a,b,c,d,1], Fang Yu[e,1], Wenwen Wei[a,b,c,d,1], Jiansong Huang[b,c], Yang Shao[f], Junrong Yan[f], Liping Mao[a,b,c], Wenjuan Yu[a,b,c], Wanzhuo Xie[a,b,c], Jie Jin[a,b,c,d,*]

[a] *Department of Hematology, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang 310003, China*
[b] *Key Laboratory of Hematologic Malignancies, Diagnosis and Treatment, Hangzhou, Zhejiang 310003, China*
[c] *Zhejiang Provincial Clinical Research Center for Hematologic Diseases, Hangzhou, Zhejiang 310003, China*
[d] *Cancer Center, Zhejiang University, Hangzhou, Zhejiang 310003, China*
[e] *Department of Pathology, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang 310003, China*
[f] *Nanjing Geneseeq Technology Inc., Nanjing, Jiangsu 210008, China*

*Corresponding author. No.79, Qingchun Road, Hangzhou, Zhejiang 310003, China. Fax: +86 571 87236702.
E-mail address: jiej0503@zju.edu.cn (J. Jin)

9 November 2021
Available online 8 April 2022

[1] These authors are co-first authors.