A targeted gene signature stratifying mediastinal gray zone lymphoma into classical Hodgkin lymphoma-like or primary mediastinal B-cell lymphoma-like subtypes

Mediastinal gray zone lymphoma (MGZL), a B-cell lymphoma with overlapping features between primary mediastinal B-cell lymphoma (PMBL) and classical Hodgkin lymphoma (CHL), is a unique entity and a diagnostic challenge.[1,2] MGZL typically exhibits discordant morphological and immunophenotypic traits and a molecular profile straddling between PMBL and CHL.[3,4] However, the diagnosis of MGZL largely relies on morphological/immunophenotypic criteria[5] and unstandardized connotations such as CHL- or PMBL-like entities may affect therapeutic choices and patients' outcome. Retrospective studies revealed common reclassification of diagnosis and heterogeneous treatments associated with high relapse rates even following intensified chemotherapy,[6,7] emphasizing the need for new tools to improve the pathobiological stratification of MGZL. Here, we report the development of a signature - comprising both tumor- and tumor microenvironment (TME)-related genes - that enables MGZL categorization based on transcriptomic proximity to either CHL or PMBL. The study, conducted in line with the Declaration of Helsinki and with formal ethical approval (Comitato Etico Regionale per la Sperimentazione Clinica della Toscana with protocol number bioGZL-2020, Rif. CEAVC Em. 2022-263, Study number 18236_oss, 21/06/2022), was designed as reported in Figure 1.

With the speculative idea of a molecular allocation of MGZL between CHL and PMBL,[3,4] we first sought to identify a unique set of transcripts capable of discerning these two entities, considering potential unbalanced contributions to gene expression offered by tumor and TME cells. A discovery cohort comprising 84 CHL and 51 PMBL, further subdivided into training (50 CHL and 31 PMBL) and testing (34 CHL and 20 PMBL) sets, was generated by pooling Affymetrix-HG-133plus2 raw data from three different gene expression profile (GEP) datasets of fresh-frozen biopsy tissues (GSE17920, GSE11318, GSE87371),[8-10] and processed as two distinct expression matrices (*data not shown*). A combination of CIBERSORTx deconvolution (*http://cibersortx.stanford.edu*)[11] and a non-negative matrix factorization (NMF)-based approach[12] was used to identify gene sets that more accurately discriminate CHL from PMBL in an unsupervised fashion. CIBERSORTx was applied to create a customized signature matrix, including GEP of both tumor (N=2) and TME cytotypes (N=22) (Figure 2), used to derive two purified GEP matrices from the bulk transcriptome of the training cohort. Each purified matrix was independently decomposed by the NMF algorithm.[12] The method was run 100 times varying the rank value in the interval [2,7]. We adopted a cophenetic correlation coefficient (CCC) and consensus matrices to choose the optimal value of rank *r* for the factorization process. We choose as optimal rank the first value of rank *r* at which the CCC trend starts decreasing, and the one associated with clear block diagonal patterns in the consensus plots (*r*=2) (*Online Supplementary Figure S1A, B*, upper panels). NMF enabled an unsupervised selection of genes related to tumor (N=47 for CHL and N=653 for PMBL) and the TME (N=1,594 for CHL and N=637 for PMBL) (*Online Supplementary Figure S1C, D*). We consistently observed a considerable abundance of TME genes in CHL, whereas tumor-derived transcripts prevailed in PMBL. Unsupervised clustering highlighted the capacity of these genes, merged in a unique panel of 2,913 transcripts (*data not shown*), to fully and reproducibly separate CHL from PMCL (*Online Supplementary Figure S1C, D*). By further filter-based feature selection (ReliefF[13] and Laplacian score[14]), we selected a final panel of 168 genes, which retained this discriminative ability both in the training and testing cohorts (*Online Supplementary Figure 2A, B*). Interestingly, genes associated with T-cell receptor signaling (e.g., *CD28* and *CD3G*), inflammation (e.g., *PRDX2*) and STAT5A targets (e.g., *PRTFDC1*) stood out as related to CHL, whereas genes involved in the cell cycle (e.g., *BTRC, MCM6, SPC25* and *RNF8*), chromatin-modifying enzymes (e.g., *HCFC1*), and those involved in the regulation of TP53 activity (e.g., *TP63*) emerged as associated with PMBL, suggesting a peculiar enrichment of known pathways for each disease. A 168-code set (including 15 housekeeping genes) was then customized to digitally profile (NanoString nCounter Flex Analysis System, NanoString Technologies) a real-life independent collection of CHL, PMBL, and MGZL cases whose RNA was directly extracted from formalin-fixed paraffin-embedded specimens (MagMAX FFPE RNA/DNA Ultra Kit, ThermoFisher Scientific). Under the auspices of the Italian Lymphoma Foundation, a multicenter collection of 39 cases originally diagnosed as gray zone lymphoma underwent central pathology review by a panel of expert hematopathologists (ES, CA, SAP, ML, MP, LL, FF, AD, ML, MP, LL, SL, SA, AS and AZ) to assess whether these fulfilled the current World Health Organization classification[1] and International Consensus Classification[2] by hematoxylin/eosin and immunohistochemical staining, and were also classified according to Sarkozy *et al.*[5] *In situ* hybridization for Epstein-Barr virus was performed and only negative cases with mediastinal involvement were considered.

Among the 39 cases with an original diagnosis of gray zone

lymphoma, 28 were confirmed as MGZL and only 24 passed the quality check for the final study phase. Their histopathological features were defined as closer to CHL (CHL-like), PMBL (PMBL-like) or intermediate (*Online Supplementary Figure S2D*), also according to the morpho-phenotypic subgroups of Sarkozy *et al.*[5] (*Online Supplementary Table S1*). Eight cases (Sarkozy group 0 [N=2], group 1 [N=5] and group 2 [N=1]) showed CHL-like morphology with Hodgkin and Reed-Sternberg cells within an inflammatory background and a variable degree of fibrosis, associated with

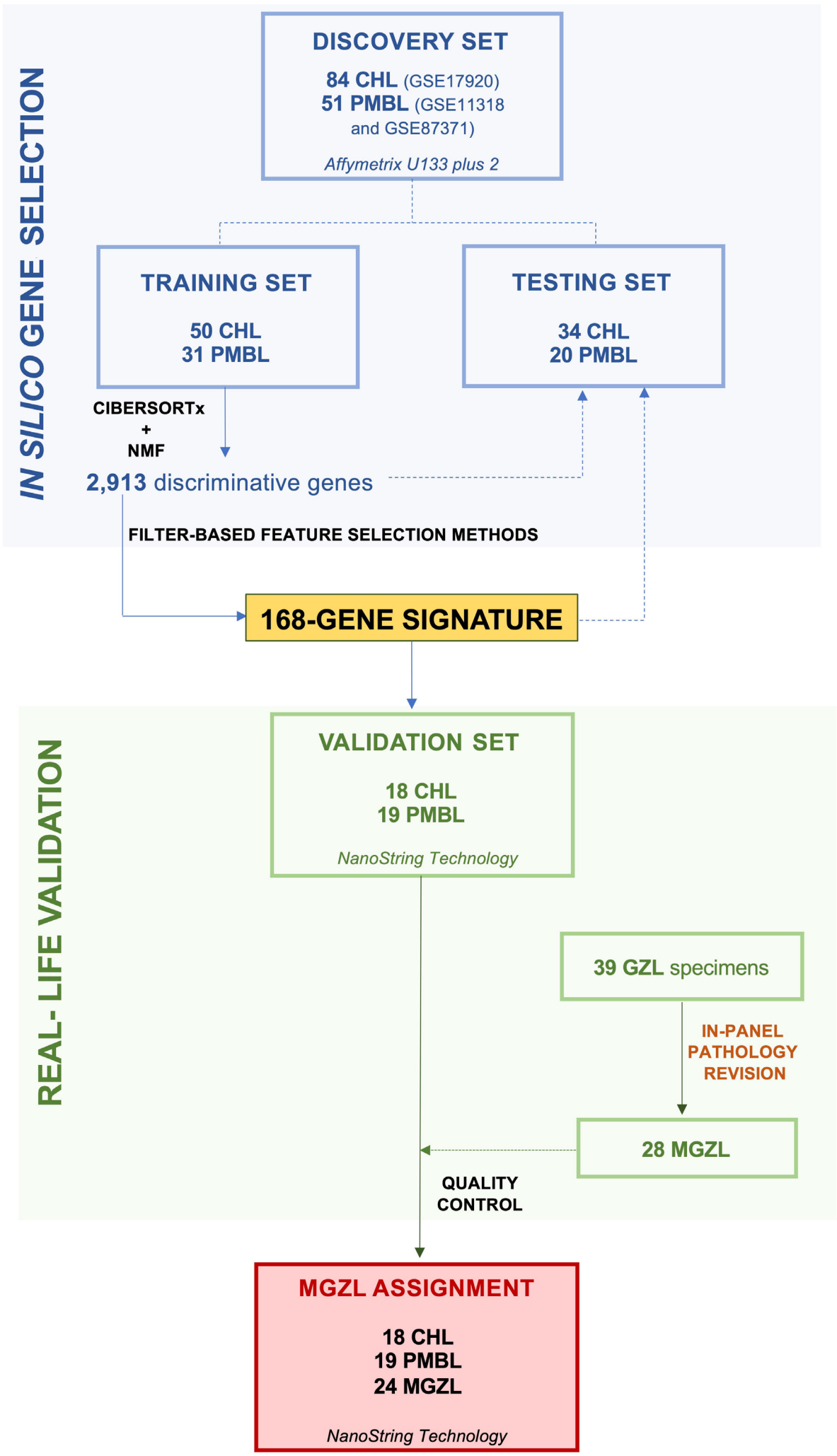

Figure 1. Schematic overview of the study design. The methodological workflow included three different phases and independent cohorts of patients. The gene selection phase related to the training/testing set was conducted *in-silico* (light blue), while the validation (light green) and MGZL assignment phases (red) were completed on a real-life set of cases (24 out of 28 centrally revised MGZL passed the quality control for the final study phase). CHL: classical Hodgkin lymphoma; PMBL: primary mediastinal B-cell lymphoma; GZL: gray zone lymphoma; MGZL: mediastinal gray zone lymphoma.

sheets of monomorphic mononuclear medium/large cells. Ten cases (Sarkozy group 1/group 2 [N=2], group 2 [N=6] and group 2/group 3 [N=2]) showed PMBL-like morphology with a predominance of medium/large tumor cells mixed with a variable number of Reed-Sternberg-like cells, and low inflammatory/fibrosis background. These cases were CD30-positive with variable expression of CD15. Two cases were negative for CD20 and CD79A with weak to moderate expression of PAX5. Other cases displayed incomplete expression of B-cell markers, a moderate inflammatory background, and variable expression CD30. The remaining six cases (Sarkozy group 1/group 2 [N=5] and group 2/group 3 [N=1]) displayed intermediate features, with sheets of mononuclear cells expressing of CD30 extensively, and a predominant full B-cell phenotype. These cases, on a pathological ground, were very challenging to categorize due their intermediate morpho-phenotype, even according to the Sarkozy classification.

By applying our 168-gene signature on the NanoString platform to a real-life set of *CHL* (N=18) and PMBL (N=19), we first confirmed its capability to fully distinguish the two diseases (*Online Supplementary Figure 2C*). Then, the signature was used to stratify 24 MGZL based on their transcriptional proximity to either CHL or PMBL. As illustrated in *Online Supplementary Figure S3*, seven out of eight cases (pathologically annotated as CHL-like and in accordance with the Sarkozy subgroup) consistently clustered within the CHL subgroup, whereas only one showed a discordant allocation. Conversely, out of the ten PMBL-like MGZL, three fell concordantly within PMBL, whereas seven cases located within the CHL cluster. The three MGZL that consistently clustered into the PMBL subgroup shared strong PAX5 and moderate CD30 staining, with a more complete B-cell phenotype. In the remaining seven MGZL, PAX5 staining was weak or moderate, CD30 at strong intensity and other B-cell markers variably positive. The remaining six cases - with intermediate morphology/Sarkozy assignment – mostly clustered within CHL, while one was within the PMBL cluster. The transcriptional clustering resulted in a valuable separation of cases, mostly falling into the CHL subgroup, especially those with intermediate morphology and Sarkozy group 1/group 2, although a subset of cases unanimously classified as PMBL-like also clustered in this group. In line with previous reports,[4] 19 out of 24 MGZL in

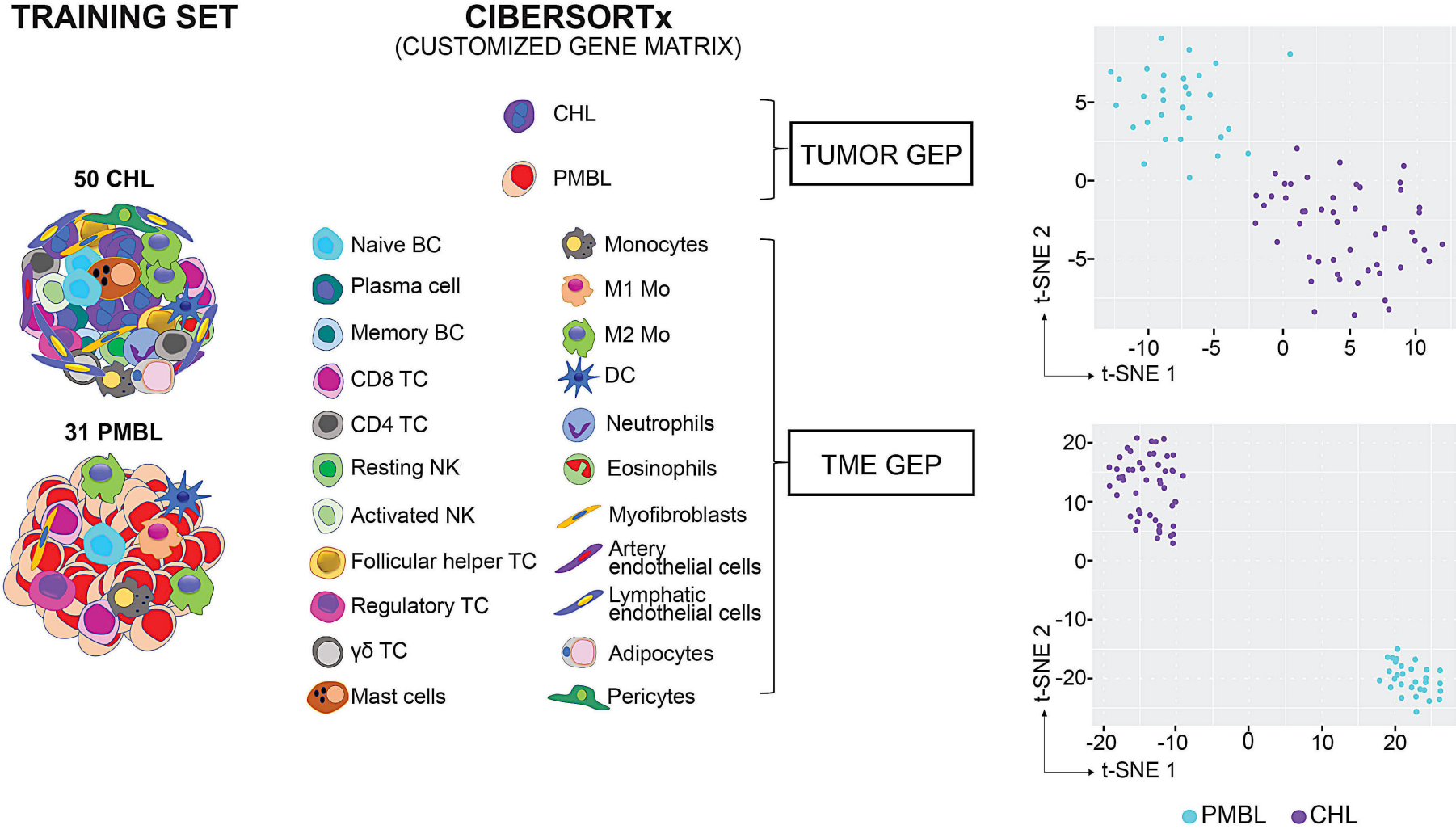

Figure 2. Identification of a molecular signature distinguishing classical Hodgkin lymphoma and primary mediastinal B-cell lymphoma. Schematic overview of the CIBERSORTx analysis. We applied CIBERSORTx for digital purification of a bulk mixture matrix (50 classical Hodgkin lymphomas and 31 primary mediastinal B-cell lymphomas as the training set) using a customized signature matrix of 24-cell types, composed of two tumor cytotypes and 22 tumor microenvironment (TME) cells. Tumor and TME compartments were run as merged classes to purify the gene expression profile for both components. The t-distributed stochastic neighbor embedding visualization of the two purified tumor and TME gene expression profiles are depicted on the right side. CHL: classical Hodgkin lymphoma; PMBL: primary mediastinal B-cell lymphoma; BC: B cells; TC: T cells; NK: natural killer cells; Mo: macrophages; DC: dendritic cells; GEP: gene expression profiles; t-SNE: t-distributed stochastic neighbor embedding.

our study appeared transcriptionally closer to CHL (*Online Supplementary Table S1*). The availability of clinical information for a subset of 14 MGZL prompted us to speculate on the hypothetical relationship between therapeutic outcome and pathological/molecular stratification of cases (*Online Supplementary Table S2*). A complete response was recorded for eight cases - all treated with R-CHOP (rituximab plus cyclophosphamide, doxorubicin, vincristine, prednisone)-like regimens - of which three were pathologically and molecularly assigned to the CHL cluster, three were PMBL-like including only one case in molecular concordance, and two cases had intermediate morphology, one of which fell within the CHL cluster and the other within the PMBL cluster. Intriguingly, six MGZL showing poorer outcomes were mostly characterized by a discrepancy between transcriptomic clustering and treatment. In particular, one case - pathologically annotated as CHL-like and concordantly assigned to the CHL cluster - reached a partial response after a typical PMBL-oriented protocol (da-EPOCH-R: dose adjusted etoposide, prednisone, vincristine, cyclophosphamide, doxorubicin, and rituximab). Likewise, one case within the PMBL cluster but characterized pathologically as CHL-like was treated with a CHL-oriented therapy (ABVD: adriamycin, bleomycin, vinblastine, dacarbazine) and displayed refractoriness. The other three PMBL-like MGZL - but molecularly allocated into the CHL cluster - exhibited unfavorable outcomes after induction with R-CHOP-like protocols. A partial response was also recorded for a PMBL-like case (both morphologically and molecularly) treated with da-EPOCH-R. Although obtained by an unprecedented approach, our results strengthen the idea that the majority of MGZL are biologically closer to CHL,[4] reinforcing the value of previous observations[15,16] on the efficacy of anti-CD30/anti-PD1 drugs in MGZL, although as salvage strategy after PMBL-oriented frontline treatments. More accurate MGZL categorization, which to date remains heterogeneous and guided solely by the morpho-phenotypic depiction of the disease, may prompt the development of new frontline combinations of drugs.

In conclusion, aiming at laying the basis for a more accurate stratification of MGZL, we diversified our approach from previous studies[3,4] by exploiting computational tools to obtain a restricted gene panel - easily measurable on RNA from formalin-fixed paraffin-embedded samples - which fully separates CHL from PMBL, and molecularly assigns MGZL to one or other of the entities. Such an approach might help in overcoming the histopathological challenge of MGZL categorization, despite efforts to apply the Sarkozy classification to describe proximity to CHL or PMBL. Validation of our signature on larger, independent sets of MGZL would be critical to decipher the underlying relationship between molecular and phenotypic traits, to build a combined histopathological/transcriptomic model of MGZL stratification and, ultimately, to prompt its translation into the clinical setting in order to optimize the treatment of these rare cases.

Authors

Grazia Gargano,[1,2*] Maria Carmela Vegliante,[1*] Flavia Esposito,[2] Susanna A. Pappagallo,[1] Elena Sabattini,[3] Claudio Agostinelli,[3,4] Stefano A. Pileri,[5] Valentina Tabanelli,[5] Maurilio Ponzoni,[6] Luisa Lorenzi,[7] Fabio Facchetti,[7] Arianna Di Napoli,[8] Marco Lucioni,[9] Marco Paulli,[9] Lorenzo Leoncini,[10] Stefano Lazzi,[10] Stefano Ascani,[11] Giuseppina Opinto,[12] Gian Maria Zaccaria,[13] Giacomo Volpe,[1] Paolo Mondelli,[1] Antonella Bucci,[1] Laura Selicato,[2] Antonio Negri,[1] Giacomo Loseto,[1] Felice Clemente,[1] Anna Scattone,[14] Alfredo F. Zito,[14] Luca Nassi,[15] Nicoletta Del Buono,[2] Attilio Guarini[1] and Sabino Ciavarella[1]

[1]Hematology and Cell Therapy Unit, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari; [2]Department of Mathematics, Aldo Moro University of Bari, Bari; [3]Hematopathology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna; [4]Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna; [5]Division of Hematopathology, European Institute of Oncology IRCCS, Milan; [6]Ateneo Vita-Salute San Raffaele Milan and Pathology Unit, IRCCS San Raffaele Scientific Institute, Milan; [7]Pathology Unit, Department of Molecular and Translational Medicine, University of Brescia, Brescia; [8]Department of Clinical and Molecular Medicine, Sant'Andrea University Hospital, Sapienza University of Rome, Rome; [9]Department of Molecular Medicine, University of Pavia/Foundation IRCCS Policlinico San Matteo, Pavia; [10]Section of Pathology, Department of Medical Biotechnology, University of Siena, Siena; [11]Pathology Unit, Azienda Ospedaliera Santa Maria di Terni, University of Perugia, Terni; [12]Clinical Pathology and Microbiology Unit, Bonomo Hospital, Andria; [13]Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, Bari; [14]Department of Pathology, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari and [15]Department of Hematology, Careggi Hospital and University of Florence, Florence, Italy

**GG and MCV contributed equally as first authors.*

Correspondence:
S. CIAVARELLA - s.ciavarella@oncologico.bari.it

https://doi.org/10.3324/haematol.2024.285266

Received: February 22, 2024.
Accepted: July 8, 2024.
Early view: July 18, 2024.

Disclosures

No conflicts of interest to disclose.

Contributions

SC, MCV and GG conceived and planned the study. MCV, GG, FE, LS, NDB and GMZ defined and conceptualized the methods for data analysis. MCV, GG, FE, LS and NDB analyzed the data. MCV, GG, GV and SC prepared the figures and wrote the manuscript. GO, SAP, PM, AB

and AN extracted RNA and performed digital expression analysis. AS, AFZ, GL, FC, LN and AG collected samples and carried out clinical annotation. ES, CA, SAP, ML, MP, LL, FF, AD, ML, MP, LL, SL, SA, AS and AZ performed the pathological review. All authors critically reviewed the manuscript and approved the final draft for submission.

Acknowledgments

GG, FE, LS and NDB are members of the Gruppo Nazionale Calcolo Scientifico - Istituto Nazionale di Alta Matematica (GNCS-INdAM).

Funding

This work was supported by grants from the Italian Ministry of Health Ricerca Corrente 2024 deliberation N. 91/2024, AIRC 5x1000 (grant number 21198 to SAP), ERC Seeds Uniba project "Biomes Data Integration with Low-Rank Models" (CUP H93C23000720001 to FE), and "Finanziamento dell'Unione Europea – NextGenerationEU - missione 4, componente 2, investimento 1.1. - PRIN PNRR 2022 "Computational approaches for the integration of multi-omics data" CUP H53D230088700 (to NDB).

Data-sharing statement

The data generated in this study are available upon request from the corresponding author.

References

1. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. Leukemia. 2022;36(7):1720-1748.
2. Campo E, Jaffe ES, Cook JR, et al. The International Consensus Classification of mature lymphoid neoplasms: a report from the Clinical Advisory Committee. Blood. 2022;140(11):1229-1253.
3. Sarkozy C, Chong L, Takata K, et al. Gene expression profiling of gray zone lymphoma. Blood Adv. 2020;4(11):2523-2535.
4. Pittaluga S, Nicolae A, Wright GW, et al. Gene expression profiling of mediastinal gray zone lymphoma and its relationship to primary mediastinal B-cell lymphoma and classical Hodgkin lymphoma. Blood Cancer Discov. 2020;1(2):155-161.
5. Sarkozy C, Copie-Bergman C, Damotte D, et al. Gray-zone lymphoma between cHL and large B-cell lymphoma: a histopathologic series from the LYSA. Am J Surg Pathol. 2019;43(3):341-351.
6. Pilichowska M, Pittaluga S, Ferry JA, et al. Clinicopathologic consensus study of gray zone lymphoma with features intermediate between DLBCL and classical HL. Blood Adv. 2017;1(26):2600-2609.
7. Sarkozy C, Molina T, Ghesquières H, et al. Mediastinal gray zone lymphoma: clinico-pathological characteristics and outcomes of 99 patients from the Lymphoma Study Association. Haematologica. 2017;102(1):150-159.
8. Steidl C, Lee T, Shah SP, et al. Tumor-associated macrophages and survival in classic Hodgkin's lymphoma. N Engl J Med. 2010;362(10):875-885.
9. Lenz G, Wright GW, Emre NCT, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci U S A. 2008;105(36):13520-13525.
10. Dubois S, Viailly PJ, Bohers E, et al. Biological and clinical relevance of associated genomic alterations in MYD88 L265P and non-L265P-mutated diffuse large B-cell lymphoma: analysis of 361 cases. Clin Cancer Res. 2017;23(9):2232-2244.
11. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773-782.
12. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A. 2004;101(12):4164-4169.
13. Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF. Mach Learn. 2003;53(1-2):23-69.
14. He X, Cai D, Niyogi P. Laplacian score for feature selection. Advances in Neural Information Processing Systems. 2005;18:507-514.
15. Melani C, Major A, Schowinsky J, et al. PD-1 Blockade in mediastinal gray-zone lymphoma. N Engl J Med. 2017;377(1):89-91.
16. Santoro A, Moskowitz AJ, Ferrari S, et al. Nivolumab combined with brentuximab vedotin for relapsed/refractory mediastinal gray zone lymphoma. Blood. 2023;141(22): 2780-2783.