haematologica
Journal of the Ferrata Storti Foundation

A targeted gene signature stratifying mediastinal gray zone lymphoma into classical HL-like or PMBL-like subtypes

by Grazia Gargano, Maria Carmela Vegliante, Flavia Esposito, Susanna A. Pappagallo, Elena Sabattini, Claudio Agostinelli, Stefano A. Pileri, Valentina Tabanelli, Maurilio Ponzoni, Luisa Lorenzi, Fabio Facchetti, Arianna Di Napoli, Marco Lucioni, Marco Paulli, Lorenzo Leoncini, Stefano Lazzi, Stefano Ascani, Giuseppina Opinto, Gian Maria Zaccaria, Giacomo Volpe, Paolo Mondelli, Antonella Bucci, Laura Selicato, Antonio Negri, Giacomo Loseto, Felice Clemente, Anna Scattone, Alfredo F. Zito, Luca Nassi, Nicoletta Del Buono, Attilio Guarini, and Sabino Ciavarella

Received: Feb 22, 2024.
Accepted: July 8, 2024.

Citation: Grazia Gargano, Maria Carmela Vegliante, Flavia Esposito, Susanna A. Pappagallo, Elena Sabattini, Claudio Agostinelli, Stefano A. Pileri, Valentina Tabanelli, Maurilio Ponzoni, Luisa Lorenzi, Fabio Facchetti, Arianna Di Napoli, Marco Lucioni, Marco Paulli, Lorenzo Leoncini, Stefano Lazzi, Stefano Ascani, Giuseppina Opinto, Gian Maria Zaccaria, Giacomo Volpe, Paolo Mondelli, Antonella Bucci, Laura Selicato, Antonio Negri, Giacomo Loseto, Felice Clemente, Anna Scattone, Alfredo F. Zito, Luca Nassi, Nicoletta Del Buono, Attilio Guarini, and Sabino Ciavarella. A targeted gene signature stratifying mediastinal gray zone lymphoma into classical HL-like or PMBL-like subtypes.
Haematologica. 2024 July 18. doi: 10.3324/haematol.2024.285266 [Epub ahead of print]

A targeted gene signature stratifying mediastinal gray zone lymphoma into classical HL-like or PMBL-like subtypes

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], Sabino Ciavarella[1, #].

[*] These authors equally contributed to this work

Corresponding author

1. Hematology and Cell Therapy Unit, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari, Italy;
2. Department of Mathematics, University of Bari Aldo Moro, Bari, Italy;
3. Haematopathology Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Bologna, Italy;
4. Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, Bologna, Italy;
5. Division of Hematopathology, European Institute of Oncology IRCCS, Milan, Italy;
6. Ateneo Vita-Salute San Raffaele Milan and Pathology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy;
7. Pathology Unit, Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy;

8. Department of Clinical and Molecular Medicine, Sant'Andrea University Hospital, Sapienza University of Rome, Rome, Italy;
9. Department of Molecular Medicine, University of Pavia/Foundation IRCCS Policlinico San Matteo, Pavia, Italy;
10. Section of Pathology, Department of Medical Biotechnology, University of Siena, Siena, Italy;
11. Pathology Unit, Azienda Ospedaliera Santa Maria di Terni, University of Perugia, Terni, Italy;
12. Clinical Pathology and Microbiology Unit, Bonomo Hospital, Andria (BT), Italy
13. Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari, Bari, Italy;
14. Department of Pathology, IRCCS Istituto Tumori 'Giovanni Paolo II', Bari, Italy;
15. Department of Hematology, Careggi Hospital and University of Florence, Florence, Italy.

Correspondence: Dr. Sabino Ciavarella, MD, PhD
Hematology and Cell Therapy Unit,
IRCCS-Istituto Tumori 'Giovanni Paolo II', Bari, Italy.
Tel: +39 080 5555446; E-mail: s.ciavarella@oncologico.bari.it

Data sharing statement

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

Word count

Main text 1587 words; 2 figures; supplementary files: one PDF file for all supplementary data.

Acknowledgements

The authors 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).

Author 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 performed RNA extraction and digital expression analysis. AS, AFZ, GL, FC, LN and AG carried out samples collection and 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.

Disclosures. All the other authors have no conflicts of interest to disclose.

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 morphologic and immunophenotypic traits and a molecular straddling between PMBL and CHL[3,4]. However, MGZL diagnosis largely relies on morphological/immunophenotypic criteria[5] and unstandardized connotation as CHL- or PMBL-like entities may affect therapeutic choice and patient outcome. Retrospective studies revealed common diagnosis reclassification and heterogeneous treatments associated with high relapse rate 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 their transcriptomic proximity to either CHL or PMBL. The study, conducted in line with the Declaration of Helsinki and 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.

Within 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 contribution to gene expression offered by tumor and TME cells. A discovery cohort comprising 84 CHL and 51 PMBL, further subdivided in a training (50 CHL and 31 PMBL) and testing (34 CHL and 20 PMBL) sets, was generated by pooling Affymetrix-HG133plus2 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 nonnegative 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 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* which CCC trend starts decreasing, and the one associated with clear block diagonal patterns in the consensus plots (r = 2, Supplementary Figure S1 A-B, upper panels). NMF enabled an unsupervised selection of genes related to tumor (n= 47 for CHL and n= 653 for PMBL) and TME (n=1,594 for CHL and n= 637 for PMBL), respectively (Supplementary Figure S1 C-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 (Supplementary Figure S1 C-D). By further filter-based feature selection (Relief[13] and Laplacian Score[14]), we selected a final panel of 168 genes, which retained this discriminative ability both in training and testing cohorts (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 cell cycle (e.g., *BTRC, MCM6, SPC25, 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 customized 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 FFPE 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, AZ) to fulfill the current WHO classification and ICC criteria[1,2] by hematoxylin/eosin (H&E) and immunohistochemical (IHC) stainings, and also classified according to Sarkozy et al.[5] *In situ* hybridization for Epstein-Barr virus (EBV) was performed and only EBV-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 (Supplementary Figure S2 D), also according to the morpho-phenotypic subgroups by Sarkozy et al.[5] (Supplementary Table S1). Eight cases (Sarkozy's group 0 [n = 2], group 1 [n = 5] and group 2 [n = 1]) showed CHL-like morphology with HRS cells within an inflammatory background and variable degree of fibrosis, associated with sheets of monomorphic mononuclear medium/large cells. Ten cases (Sarkozy's 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 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, moderate inflammatory background, and variable expression CD30. The remaining six cases (Sarkozy's group 1/group 2 [n = 5] and group 2/group 3 [n=1]) displayed intermediate features, with sheets of mononuclear cells expressing CD30 extensively, and a predominant full B-cell phenotype. These cases, on pathological ground, were at very challenging categorization due their intermediate morpho-phenotype, even according to Sarkozy's 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 (Supplementary Figure 2C). Thus, the signature was used to stratify 24 MGZL based on their transcriptional proximity to either CHL or PMBL. As reported in Supplementary Figure S3, seven out of eight cases (pathologically annotated as CHL-like and in accordance with the Sarkozy's 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 placed 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 was weak or moderate, CD30 at strong intensity and other B-cell markers variably positive. The remaining six cases - at intermediate morphology/Sarkozy's assignment – mostly clustered within CHL, while one 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's group 1/group 2, although also a subset of cases unanimously classified as PMBL-like clustered in this group. In line with previous reports[4], 19 out of 24 MGZL in our study appeared transcriptionally closer to CHL (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 (Supplementary Table S2). A complete response was recorded for eight cases - all treated by R-CHOP-like regimens - of which three pathologically and molecularly assigned to the CHL cluster, three PMBL-like including only one case in molecular concordance, and two cases at intermediate morphology but fallen within the CHL and PMBL cluster, respectively. 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). Likewise, one case within the PMBL cluster but characterized pathologically as CHL-like was treated by a CHL-oriented therapy (ABVD) displaying refractoriness. The other three PMBL-like MGZL - but molecularly allocated into the CHL cluster - exhibited unfavorable outcomes after induction by R-CHOP-like protocols. A partial response was also recorded for a PMBL-like case (both morphologically and molecularly) treated by 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 may prompt the development of new frontline combinations of drugs, which to date remains heterogeneous and solely guided by the morpho-phenotypic depiction of the disease.

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 FFPE samples - that fully separates CHL from PMBL, and molecularly assigns MGZL to either entity. Such an approach might help in overcoming the histopathological challenge of MGZL categorization, despite efforts to apply Sarkozy's classification to describe their 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, prompt its translation into the clinical setting to optimize the treatment of these rare cases.

References

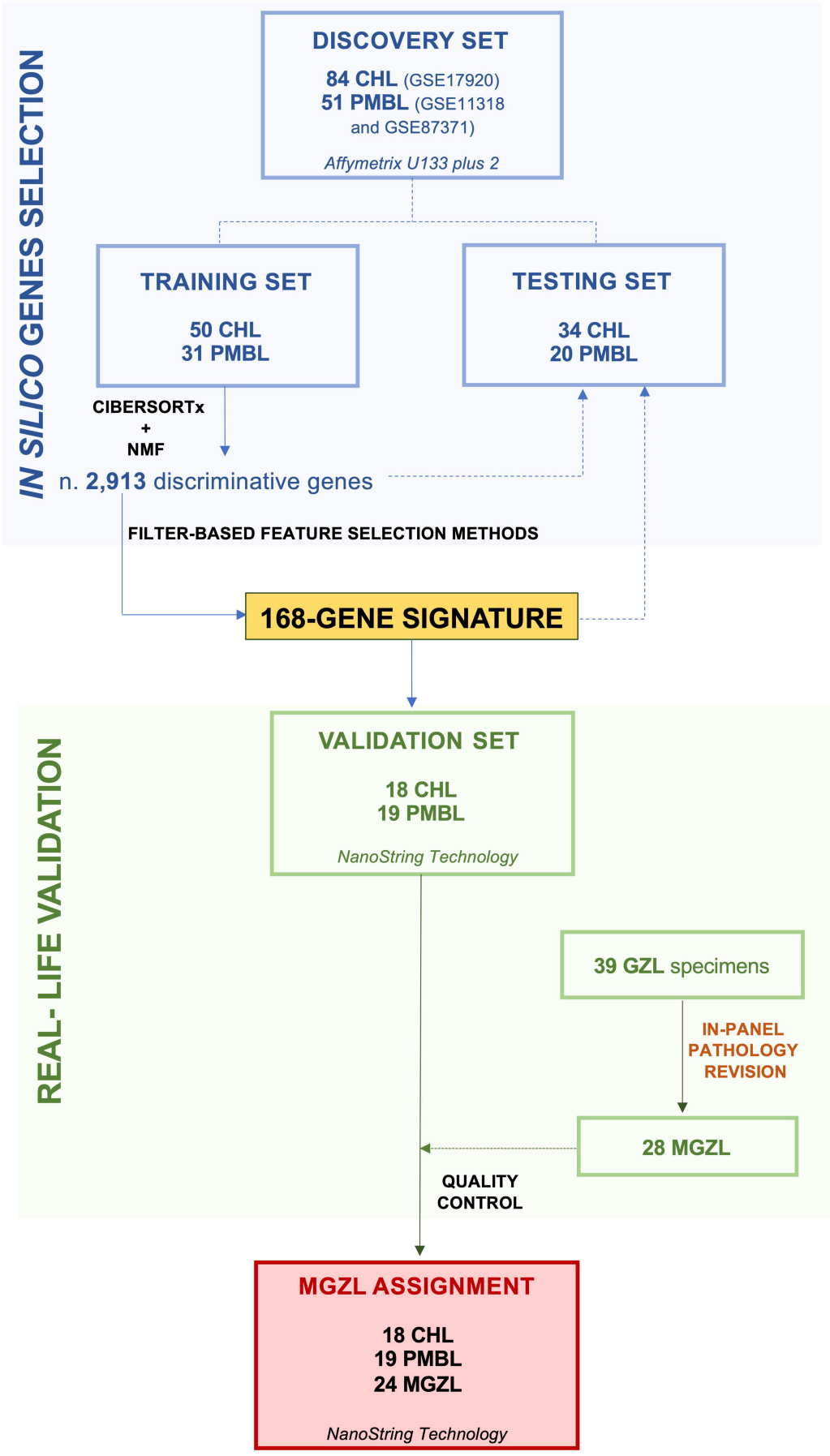
1. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. Leukemia. 2022;36(7):1720-1748.
2. Campo E, Jaffe ES, Cook JR, et al. The International Consensus Classification of Mature Lymphoid Neoplasms: 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.

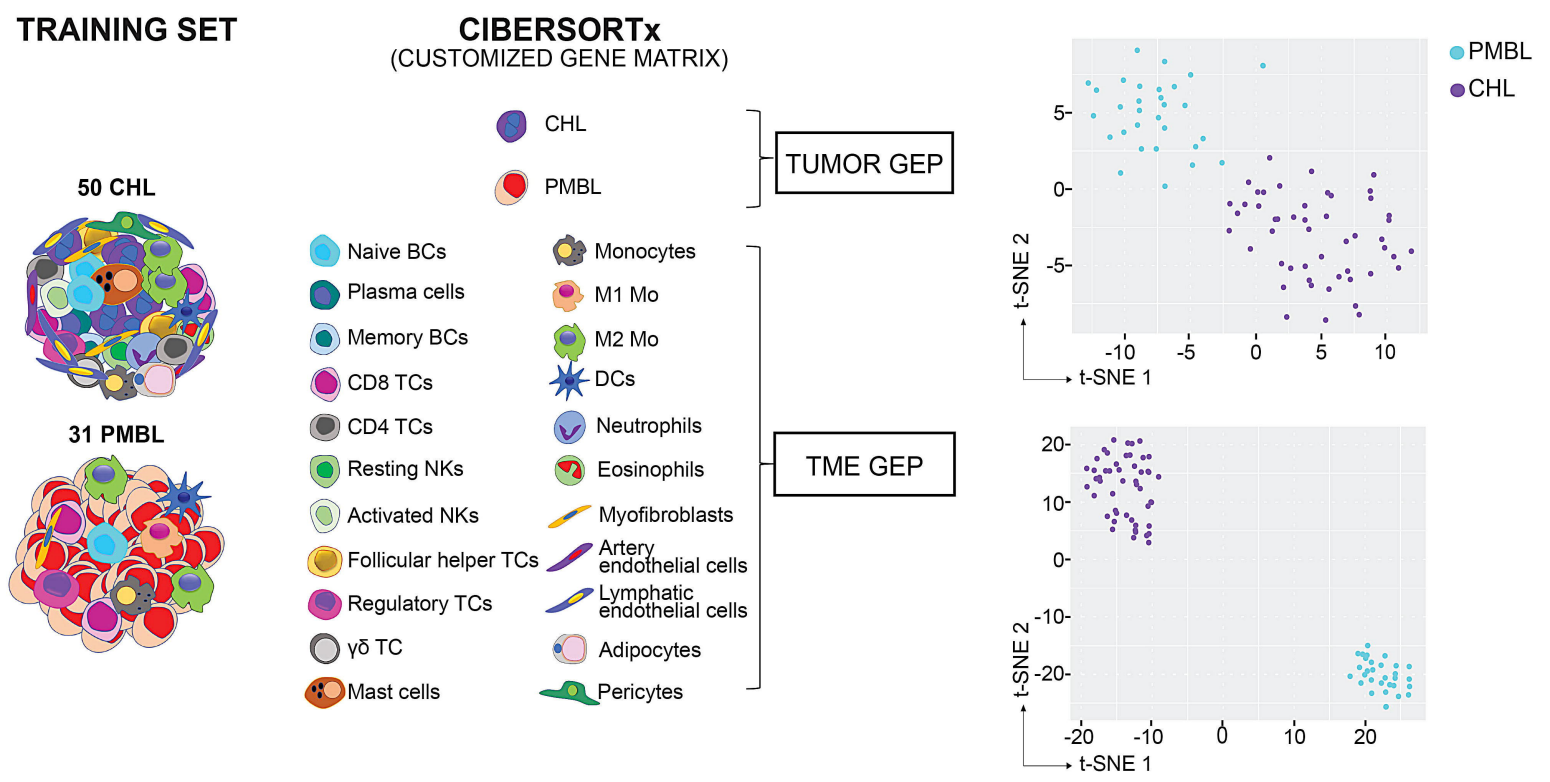
10. Dubois S, Viailly PJ, Bohers E, et al. Biological and clinical relevance of associated genomic alterations in MYD88 L265P and non-L265P-mutated diffuse large B-cell lymphoma: Analysis of 361 cases. Clin Cancer Res. 2017;23(9):2232-2244.
11. Newman AM, Steen CB, Liu CL, et al. Determining cell type 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”, Vol 18: 507-514; 2005.
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.

Figure Legends

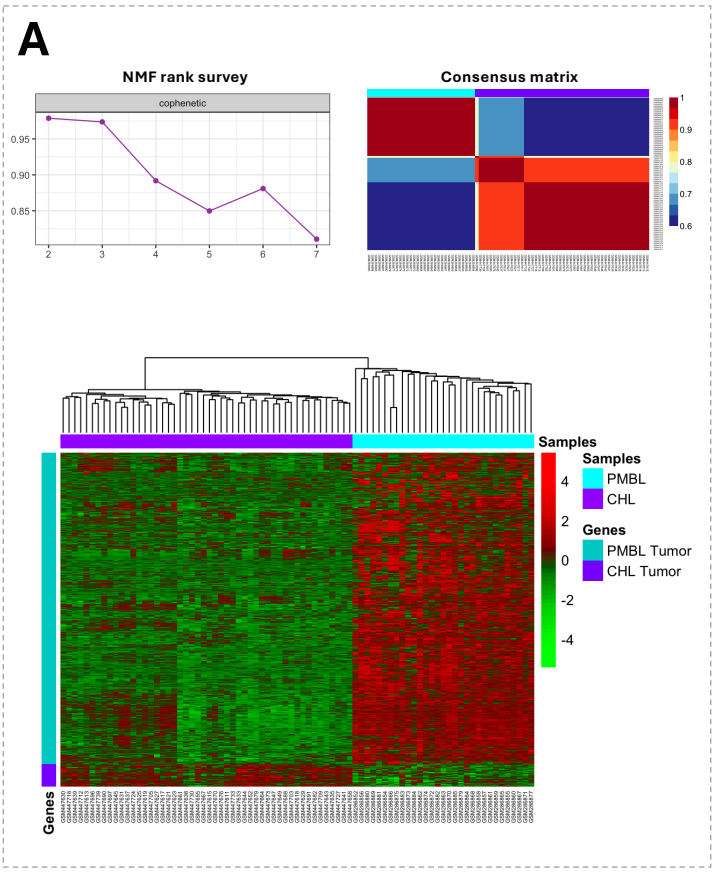
Figure 1. Schematic overview of the study design. The methodological workflow included three different phases and independent patient cohorts. The gene selection phase related to the training/testing set was conducted in-silico (light blue), while the validation (light green) and MGZL assignment phases (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.

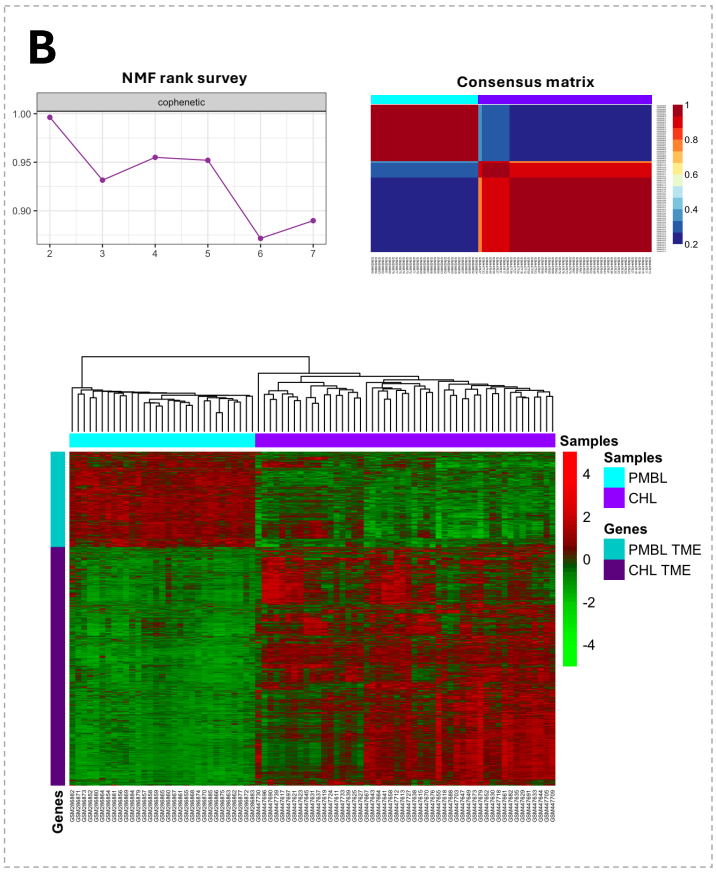
Figure 2. Identification of a molecular signature distinguishing CHL and PMBL. A. Schematic overview of CIBERSORTx analysis. We applied CIBERSORTx for digital purification of a bulk mixture matrix (50 CHL and 31 PMBL as training set) using a customized signature matrix of 24-cell type, composed of two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

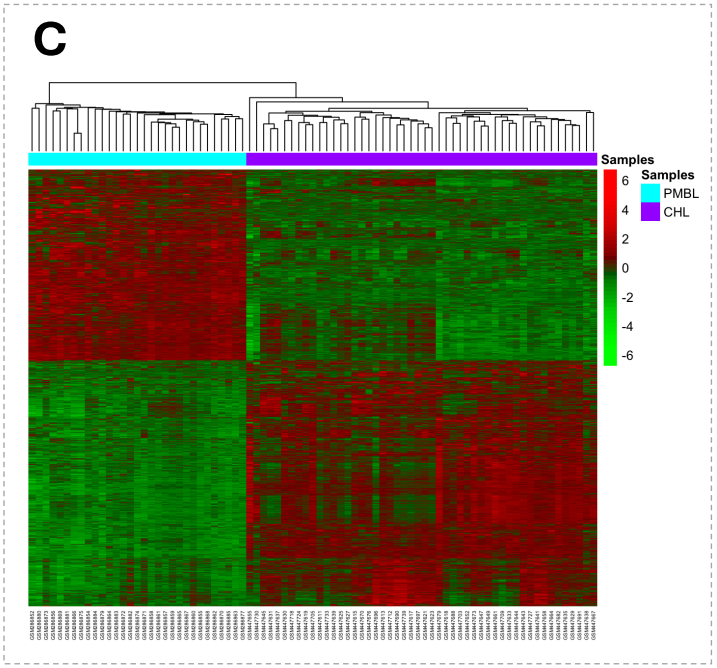

IN SILICO GENES SELECTION
DISCOVERY SET
84 CHL (GSE17920)
51 PMBL (GSE11318 and GSE87371)
Affymetrix U133 plus 2
TRAINING SET
50 CHL
31 PMBL
TESTING SET
34 CHL
20 PMBL
CIBERSORTx
+
NMF
n. 2,913 discriminative genes
FILTER-BASED FEATURE SELECTION METHODS
168-GENE SIGNATURE
REAL-LIFE VALIDATION
VALIDATION SET
18 CHL
19 PMBL
NanoString Technology
39 GZL specimens
IN-PANEL PATHOLOGY REVISION
28 MGZL
QUALITY CONTROL
MGZL ASSIGNMENT
18 CHL
19 PMBL
24 MGZL
NanoString Technology


TRAINING SET
CIBERSORTx
(CUSTOMIZED GENE MATRIX)
50 CHL
31 PMBL
CHL
PMBL
TUMOR GEP
Naive BCs
Plasma cells
Memory BCs
CD8 TCs
CD4 TCs
Resting NKs
Activated NKs
Follicular helper TCs
Regulatory TCs
γδ TC
Mast cells
Monocytes
M1 Mo
M2 Mo
DCs
Neutrophils
Eosinophils
Myofibroblasts
Artery endothelial cells
Lymphatic endothelial cells
Adipocytes
Pericytes
TME GEP
PMBL
CHL
t-SNE 2
t-SNE 1
5
0
-5
-10
10
20
-20

Supplementary Data







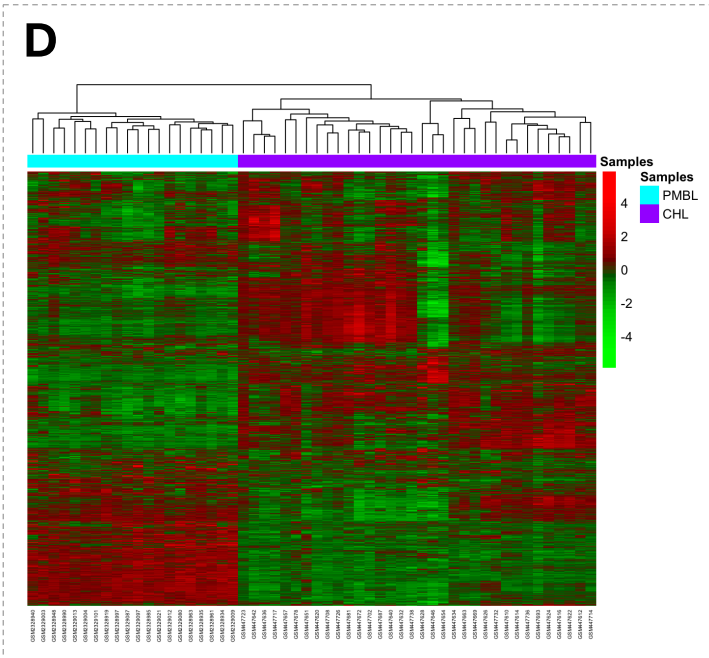
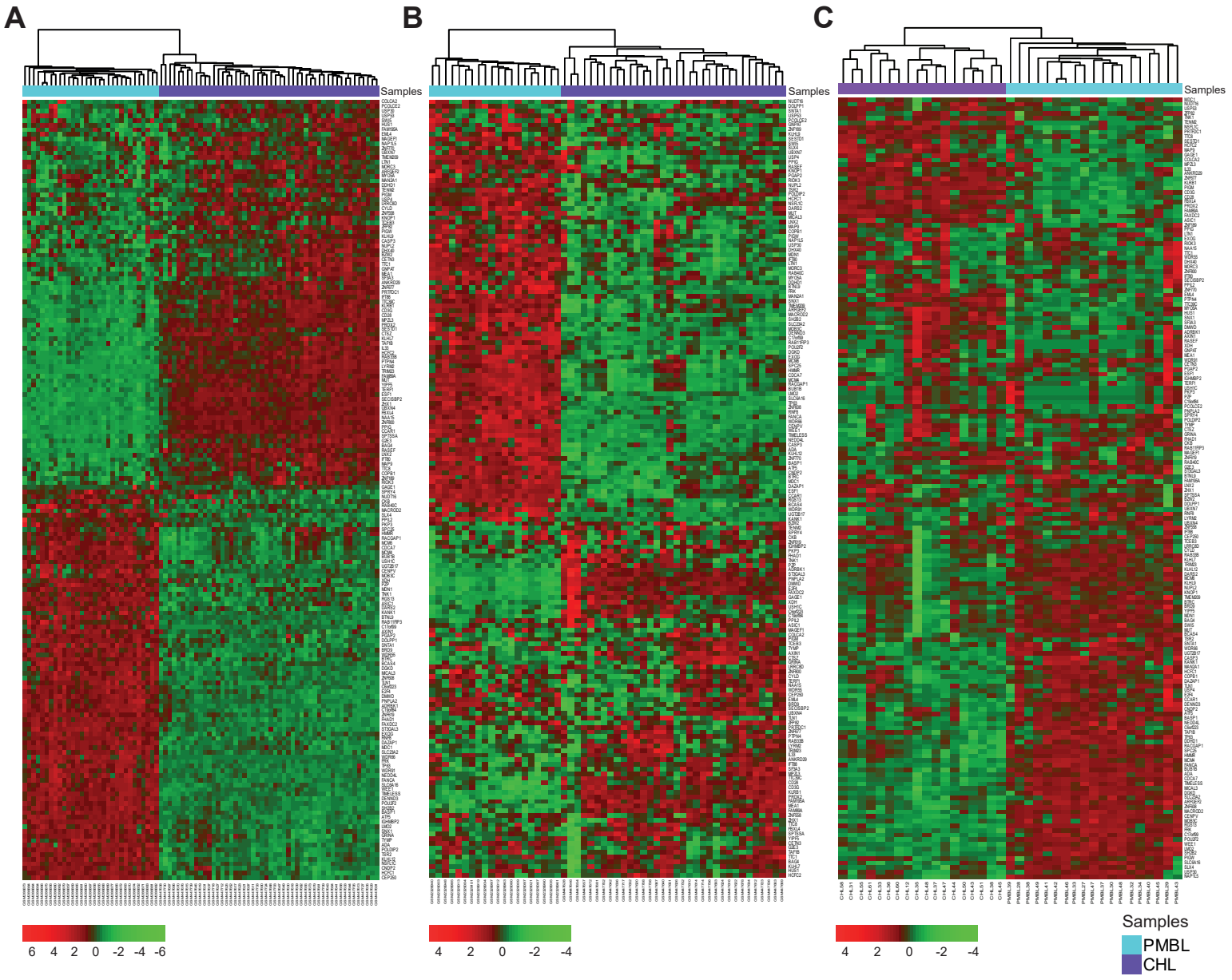


Figure S1. NMF application and performance of 2,913-gene signature in the training and testing sets.

A. Optimal rank determination by NMF for tumor GEP derived from CIBERSORTx: CCC against different rank *r* in the range [2,7] and consensus matrix for optimal rank *r* = 2 (upper panel). Heatmap representing the expressions of 700 tumor-related genes on the training cohort (CCC = 0.92, bottom panel). **B.** Optimal rank determination by NMF for TME GEP: CCC against different rank *r* in the range [2,7] and consensus matrix for optimal rank *r* = 2 (upper panel). Heatmap representing the expressions of 2,231 TME-associated genes on the training cohort (CCC = 0.88, bottom panel). NMF was performed by using the R package *NMF* (version 0.24.0). Hierarchical clustering analyses were performed using average linkage with Pearson correlation distance metric according to the highest CCC (*cluster* R-package, version 2.1.3); clustered heatmaps were drawn using the R package *pheatmap* (version 1.0.12). **C.** Heatmap showing clustering results of the overall 2,913 tumor/TME related genes distinctive of CHL and PMBL in the training set (CCC = 0.90). **D.** Heatmap depicting clustering results of 2,913-gene signature in the testing set (CCC = 0.90). NMF, nonnegative matrix factorization; GEP, gene expression profiles; CCC, cophenetic correlation coefficient; TME, tumor microenvironment. MGZL, mediastinal gray zone lymphoma; CHL, classical Hodgkin lymphoma; PMBL, primary mediastinal B-cell lymphoma.


A
B
C
Samples
6 4 2 0 -2 -4 -6
4 2 0 -2 -4
4 2 0 -2 -4
Samples
PMBL
CHL

D

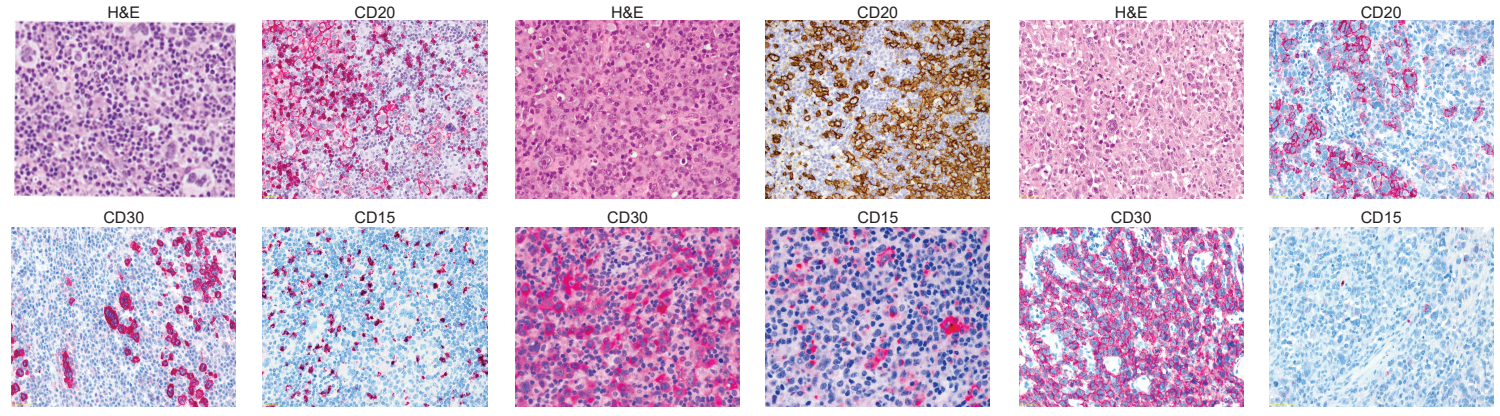

MGZL CHL-like
(Sarkozy's group 1)
MGZL PMBL-like
(Sarkozy's group 2)
MGZL intermediate
(Sarkozy's group 1/ group2)
H&E
CD20
CD30
CD15

Figure S2. Clustering analysis of the 168-gene signature in different cohorts and immunohistochemical staining of representative MGZL.

A-B-C. Heatmaps showing the unsupervised clustering of samples in the training set (*n* =81, CCC = 0.98), in the independent testing cohort (*n*= 54, CCC = 0.93), and in the real-life cohort (n =37, CCC = 0.87) based on the expression of the 168-gene signature. **D.** IHC images of H&E, CD20, CD30 and CD15 staining from three MGZL representatives of CHL-like, PMBL-like and intermediate morphology respectively, labeled also according to Sarkozy's classification. Magnification at 20X. CCC, cophenetic correlation coefficient; IHC, immunohistochemistry; H&E, hematoxylin and eosin; MGZL, mediastinal gray zone lymphoma; CHL, classical Hodgkin lymphoma; PMBL, primary mediastinal B-cell lymphoma.

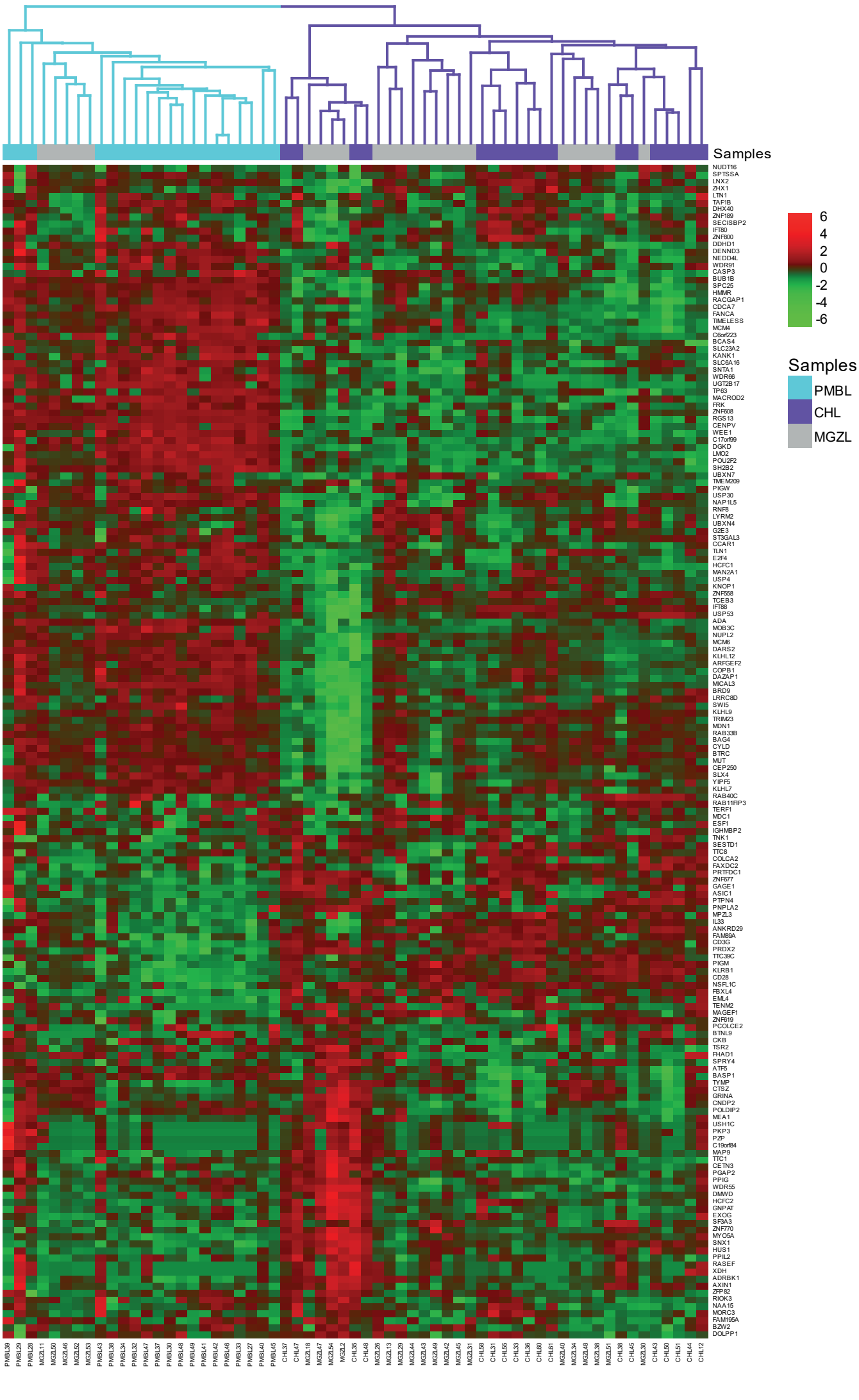

Samples
NUDT16
SPTSSA
LNX2
ZHX1
LTN1
TAF1B
DHX40
ZNF189
SECISBP2
IFT80
ZNF800
DDHD1
DENND3
NEDD4L
WDR91
CASP3
BUB1B
SPC25
HMMR
RACGAP1
CDCA7
FANCA
TIMELESS
MCM4
C6orf223
BCAS4
SLC23A2
KANK1
SLC6A16
SNTA1
WDR66
UGT2B17
TP63
MACROD2
FRK
ZNF808
RGS13
CENPV
WEE1
C17orf69
DGKD
LMO2
POU2F2
SH2B2
UBXN7
TMEM209
PIGW
USP30
NAP1L5
RNF8
LYRM2
UBXN4
G2E3
ST3GAL3
CCAR1
TLN1
E2F4
HCFC1
MAN2A1
USP4
KNOP1
ZNF568
TCEB3
IFT88
USP53
ADA
MOB3C
NUPL2
MCM6
DARS2
KLHL12
ARFGEF2
COPB1
DAZAP1
MICAL3
BRD9
LRRC8D
SWI5
KLHL9
TRIM23
MDN1
RAB33B
BAG4
CYLD
BTRC
MUT
CEP250
SLX4
YIPF5
KLHL7
RAB40C
RAB11FIP3
TERF1
MDC1
ESF1
IGHMBP2
TNK1
SESTD1
TTC8
COLCA2
FAXDC2
PRTFDC1
ZNF677
GAGE1
ASIC1
PTPN4
PNPLA2
MFZL3
IL33
ANKRD29
FAM89A
CD3G
PRDX2
TTC39C
PIGM
KLRB1
CD28
NSFL1C
FBXL4
EML4
TENM2
MAGEF1
ZNF619
PCOLCE2
BTNL9
CKB
TSR2
FHAD1
SPRY4
ATF5
BASP1
TYMP
CTSZ
GRINA
CNDP2
POLDIP2
MEA1
USH1C
PKP3
PZP
C19orf84
MAP9
TTC1
CETN3
PGAP2
PPIG
WDR55
DMWD
HCFC2
GNPAT
EXOG
SF3A3
ZNF770
MYO5A
SNX1
HUS1
PPIL2
RASEF
XDH
ADRBK1
AXIN1
ZFP62
RIOK3
NAA15
MORC3
FAM195A
BZW2
DOLPP1
6
4
2
0
-2
-4
-6
Samples
PMBL
CHL
MGZL
PMBL39
PMBL29
PMBL28
MGZL11
MGZL50
MGZL46
MGZL52
MGZL53
PMBL43
PMBL38
PMBL34
PMBL32
PMBL47
PMBL37
PMBL30
PMBL48
PMBL49
PMBL41
PMBL42
PMBL46
PMBL33
PMBL27
PMBL40
PMBL45
CHL37
CHL47
MGZL18
MGZL47
MGZL54
MGZL2
CHL35
CHL48
MGZL26
MGZL13
MGZL29
MGZL44
MGZL43
MGZL49
MGZL42
MGZL45
MGZL31
CHL58
CHL31
CHL55
CHL33
CHL36
CHL60
CHL61
MGZL40
MGZL34
MGZL48
MGZL38
MGZL51
CHL38
CHL45
MGZL30
CHL43
CHL50
CHL51
CHL44
CHL12

Figure S3. Transcriptional assignment of MGZL.

Heatmap showing the clustering analysis of 18 CHL, 19 PMBL, and 24 MGZL based on the 168-gene signature expression (CCC = 0.80). The two main clusters, respectively including all PMBL and all CHL samples, also incorporated the MGZL samples based on their transcriptional proximity to PMBL or CHL. CHL, classical Hodgkin lymphoma; PMBL, primary mediastinal B-cell lymphoma; MGZL, mediastinal gray zone lymphoma; CCC, cophenetic correlation coefficient.

Table S1. Morphologic and immunohistochemical features of MGZL cases and GEP cluster assignment.

Case ID	Age	Sex	Mediastinal involvment	EBV	Tumor cells (%)	Cytoarchitecture and morphology	Sarkozy's groups	Reed-Sternberg cells (%)	Degree of inflammatory background	Degree of fibrosis	Necrosis	CD15 (%, intensity)	CD20 (%, intensity)	CD30 (%, intensity)	CD79A (%, intensity)	PAX5 (%, intensity)	OCT-2 (%, intensity)	BOB-1 (%, intensity)	CD45 (%, intensity)	GEP CLUSTER
MGZL34	71	M	yes	neg	30	CHL-like	group1	>10%	low	moderate	absent	25-50, moderate	75-100, strong	75-100, strong	75-100, moderate	75-100, moderate	75-100, strong	na	na	CHL-like
MGZL13	77	F	yes	neg	60	CHL-like	group2	<10%	low	moderate	present	50-75, moderate	negative	75-100, strong	negative	75-100, moderate	negative	negative	negative	CHL-like
MGZL31	63	F	yes	neg	60	CHL-like	group1	>10%	moderate	moderate	absent	rare, moderate	25-50, weak	75-100, strong	50-75, moderate	75-100, strong	negative	25-50, moderate	na	CHL-like
MGZL43	40	M	yes	neg	20	CHL-like	group0	>10%	high	moderate	absent	negative	75-100, strong	75-100, strong	75-100, strong	75-100, moderate	75-100, moderate	75-100, moderate	negative	CHL-like
MGZL48	30	M	yes	neg	40	CHL-like	group0	>10%	moderate	high, with thick bands	absent	75-100, moderate	75-100, strong	50-75, strong	75-100, strong	75-100, strong	75-100, moderate	75-100, moderate	negative	CHL-like
MGZL51	52	F	yes	neg	30	CHL-like	group1	>10%	high	high, with thick bands	absent	negative	75-100, strong	50-75, moderate	75-100, strong	75-100, strong	75-100, strong	50-75, weak	negative	CHL-like
MGZL45	46	M	yes	neg	55	CHL-like	group1	>10%	moderate	moderate	absent	75-100, strong	75-100, strong	75-100, moderate	50-75, moderate	75-100, moderate	25-50, weak	negative	negative	CHL-like
MGZL52	37	M	yes	neg	60	CHL-like	group1	>10%	moderate	absent	absent	negative	75-100, strong	75-100, moderate	75-100, moderate	75-100, moderate- strong	75-100, strong	75-100, moderate	negative	PMBL-like
MGZL47	45	M	yes	neg	20	Intermediate	group1	>10%	moderate	moderate	absent	negative	75-100, moderate	75-100, strong	50-75, strong	75-100, strong	75-100, moderate	25-50	25-50, strong	CHL-like
MGZL18	55	M	yes	neg	50	Intermediate	group1/group2	>10%	na	moderate	absent	25-50, strong	75-100, strong	75-100, moderate	75-100, moderate	75-100, moderate	75-100, strong	75-100, strong	na	CHL-like
MGZL2	29	F	yes	neg	60	Intermediate	group1/group2	<10%	low	high	absent	negative	50-75, strong	50-75, strong	75-100	50-75, strong	75-100	75-100	50-75	CHL-like
MGZL26	37	F	yes	neg	70	Intermediate	group1/group2	>10%	low	moderate	absent	rare, moderate	25-50, moderate	75-100, strong	50-75, moderate	75-100, strong	negative	na	negative	CHL-like
MGZL46	28	F	yes	neg	45	Intermediate	group1/group2	>10%	moderate	moderate	absent	negative	75-100, strong	70, variable strong	75-100, strong	75-100, strong	75-100, strong	75-100, strong	50-75	PMBL-like
MGZL30	52	F	yes	neg	85	Intermediate	group1/group2	>10%	moderate	moderate	absent	50-75, strong	75-100, strong	75-100, strong	75-100, strong	75-100, strong	75-100, moderate	75-100, strong	negative	CHL-like
MGZL38	25	M	yes	neg	50	PMBL-like	group2/group3	<10%	moderate	high, w/o thick bands	present	negative	negative	75-100, strong	negative	50-75, weak to moderate	50-75, weak	negative	75-100	CHL-like
MGZL40	42	M	yes	neg	na	PMBL-like	group1/group2	<10%	low/moderate	high, w/o thick bands	absent	50-75, strong	75-100, strong	75-100, strong	negative	75-100, weak	50-75, strong	negative	50-75, strong	CHL-like
MGZL49	51	M	yes	neg	80	PMBL-like	group2	<10%	moderate	high, with thick bands	absent	rare	rare, weak	75-100, strong	50-75, moderate	75-100, moderate	rare, strong	negative	negative	CHL-like
MGZL54	31	M	yes	neg	70	PMBL-like	group2	<10%	low	low	present	75-100, strong	75-100, strong	75-100, strong	negative	75-100, moderate	75-100	75-100	negative	CHL-like
MGZL11	41	M	yes	neg	80	PMBL-like	group2	>10%	low	low	absent	50-75, moderate	25-50, strong	50-75, moderate	75-100	25-50, strong	25-50, strong	na	25-50	PMBL-like
MGZL42	17	M	yes	neg	70	PMBL-like	group2/group3	>10%	low	moderate	present	25-50, moderate	25-50, weak	75-100, moderate	negative	75-100, weak	rare, weak	negative	negative	CHL-like
MGZL50	33	M	yes	neg	50	PMBL-like	group2	>10%	low	moderate	present	rare	75-100, strong	75-100, moderate	25-50	75-100, strong	75-100, strong	75-100, strong	negative	PMBL-like
MGZL29	69	M	yes	neg	90	PMBL-like	group2	>10%	absent	absent	absent	rare, moderate	negative	75-100, strong	negative	75-100, moderate	negative	50-75, moderate	negative	CHL-like
MGZL53	25	M	yes	neg	50	PMBL-like	group2	<10%	moderate	absent	absent	negative	75-100, strong	50-75, weak-moderate	50-75,weak	75-100, strong	25-50, moderate	75-100	75-100	PMBL-like
MGZL44	50	F	yes	neg	na	PMBL-like	group1/group2	<10%	moderate	high, with thick bands	absent	rare	75-100, strong	75-100, strong	50-75, moderate	75-100, moderate	75-100, strong	75-100, moderate	75-100, moderate	CHL-like

Table S2. Clinical characteristic and outcome of a subset of 14 MGZL patients

	CHL-cluster (N=10)	PMBL-cluster (N=4)
Morphology		
CHL-like	4 (40.0%)	1 (25.0%)
Intermediate	1 (10.0%)	1 (25.0%)
PMBL-like	5 (50.0%)	2 (50.0%)
Sarkozy's groups		
group0	2 (20.0%)	0 (0%)
group1	3 (30.0%)	1 (25.0%)
group1/group2	2 (20.0%)	1 (25.0%)
group2	2 (20.0%)	2 (50.0%)
group2/group3	1 (10.0%)	0 (0%)
Stage		
1	0 (0%)	1 (25.0%)
2	8 (80.0%)	1 (25.0%)
3	0 (0%)	1 (25.0%)
4	2 (20.0%)	1 (25.0%)
First-line therapy		
EPOCH	7 (70.0%)	2 (50.0%)
CHOP-like	2 (20.0%)	1 (25.0%)
MACOPB	1 (10.0%)	0 (0%)
ABVD	0 (0%)	1 (25.0%)
Therapy response rate		
CR	6 (60.0%)	2 (50.0%)
PD	2 (20.0%)	1 (25.0%)
PR	1 (10.0%)	1 (25.0%)
SD	1 (10.0%)	0 (0%)
PFS (months)		
Median [Min, Max]	34.3 [2.23, 106]	55.0 [2.93, 62.0]
Age		
Median [Min, Max]	43.5 [17.0, 52.0]	30.5 [25.0, 37.0]

CHL, classical Hodgkin lymphoma; PMBL, primary mediastinal B-cell lymphoma; CR, complete response; PD, partial disease; PR, partial response; SD, stable disease; PFS, progression-free survival.