**Metabolic Heterogeneity on Baseline 18FDG-PET/CT Scan is a Predictor of Outcome in Primary Mediastinal B-Cell Lymphoma.**

Luca Ceriani1, Lisa Milan1, Maurizio Martelli2, Andrés J.M. Ferreri3, Luciano Cascione4, Pier Luigi Zinzani5, Alice Di Rocco2, Annarita Conconi6, Anastasios Stathis7, Franco Cavalli4, Monica Bellei8, Kelly Cozens9, Elena Porro4, Luca Giovanella1, Peter W. Johnson10\* and Emanuele Zucca4,7,11\*.

*1 Nuclear Medicine and PET/CT Centre, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland*

*2 Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy*

*3 Department of Onco-Hematology, Unit of Lymphoid Malignancies, IRCCS San Raffaele Scientific Institute, Milan, Italy*

*4 Institute of Oncology Research*, *Bellinzona, Switzerland*

*5 Institute of Hematology “Seràgnoli”, University of Bologna, Bologna Italy*

*6 Division of Hematology, Ospedale degli Infermi, Biella, Italy*

*7 Division of Medical Oncology, Oncology Institute of Southern Switzerland*, *Bellinzona, Switzerland*

*8 Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy*

*9 Clinical Trials Unit, University of Southampton, Southampton, United Kingdom*

*10 Cancer Research UK Centre, University of Southampton, Southampton, United Kingdom.*

*11 Medical Oncology, University of Bern, Bern, Switzerland.*

Presented in part at the 14th International Conference on Malignant Lymphoma, Lugano (Switzerland), 14 – 17 June, 2017

\*PWJ and EZ are co-senior authors

*Running title:* Metabolic heterogeneity predicts PMBCL outcome

*Correspondence:* Emanuele Zucca, IELSG Operating Group, Ospedale San Giovanni, CH-6500 Bellinzona, Switzerland. E-mail: ielsg@eoc.ch. Telephone: (+41) 91 811 9040. Fax: (+41) 91 811 9182.

*Text word count, 2861; abstract word count, 227; figure count, 4; table count, 3; reference count, 48; supplemental file, 1.*

**Key Points**

* MH on 18FDG-PET/CT may be a prognostic tool for PMBCL.
* High TLG combined with high MH at presentation identifies patients at high risk of progression following conventional therapy**.**

**Abstract**

An important unmet need in the management of primary mediastinal B-cell lymphoma (PMBCL) is to identify the patients for whom first line therapy will fail, in order to intervene before the lymphoma becomes refractory. High heterogeneity of intratumoral 18F-fluorodeoxyglucose (18FDG) uptake distribution on positron emission tomography/computed tomography (PET/CT) scans has been suggested as a possible marker of chemo-resistance in solid tumors. In the present study, we investigated the prognostic value of metabolic heterogeneity (MH) in 103 PMBCL patients prospectively enrolled in the International Extranodal Lymphoma Study Group (IELSG) 26 study, aimed at clarifying the role of PET in this lymphoma subtype. MH was estimated using the area under curve of cumulative SUV-volume histogram (AUC-CSH) method. Progression-free survival (PFS) at 5 years was 94% versus 73% in low and high MH groups, respectively (p=0.0001). In a Cox model of progression-free survival including dichotomized MH, metabolic tumor volume (MTV), total lesion glycolysis (TLG) and tumor bulk (mediastinal mass >10 cm), only TLG (p<0.001) and MH (p<0.001) retained statistical significance. Using these two features to construct a simple prognostic model resulted in early and accurate (PPV=89% and NPV≥90%) identification of patients at high risk of progression, at a point that would allow the use of risk-adapted treatments. This may provide an important opportunity for the design of future trials, aimed at helping the minority of patients who harbor chemorefractory PMBCL.

**Introduction**

Tumor heterogeneity describes the observation that cancer cells can show variable phenotype profiles, including cell morphology, gene expression, metabolism, motility, proliferation, and metastatic potential; this complex phenomenon also reflects genome instability and epigenetic variation, and has been associated with differences in outcome across several cancer types1-3. Since malignant tumors comprise a heterogeneous mixture of functionally distinct cells that may differ widely in their responses to therapy, it is possible that high heterogeneity may correspond to acquired and innate resistance, leading to incomplete responses and treatment failure4.

Positron emission tomography/computed tomography (PET/CT) with 18-Fluorodeoxyglucose (18FDG) is emerging as a means to characterize the metabolic patterns of intratumoral heterogeneity5,6. The intratumoral distribution of 18FDG uptake reflects the glucose metabolism of both tumor and microenvironment as well as other processes including necrosis, apoptosis, proliferation and angiogenesis5,7,8. Increased metabolic heterogeneity (MH) has been correlated with treatment failure and poor prognosis in several solid tumors and in sarcoma6,9-14, but very few data are available concerning MH in lymphomas15,16.

The IELSG26 study was designed to evaluate the role of PET/CT in the treatment of primary mediastinal large B-cell lymphoma (PMBCL), a distinct clinicopathological and molecular subtype of diffuse large B-cell lymphoma arising from the B-cells in the thymus. It is characterized by a rapidly progressive anterior mediastinal bulky mass, often with local invasion and compressive syndromes17. PMBCL commonly affects young adults, and is highly curable with aggressive immunochemotherapy regimens with or without radiotherapy, with 5-year survival rates over 90% in recent studies18,19. However, it is crucial to maximize cure rates with initial therapy since salvage treatment for the few patients failed by front-line therapy has extremely poor results20,21. If it were possible to identify this important minority early, it would allow the development of risk-stratified approaches, but the International Prognostic Index (IPI) has limited utility in PMBCL22 because most patients present when the disease is localized, leaving an unmet need for reliable prognostic markers.

The IELSG26 study demonstrated that visual assessment of posttreatment PET/CT scans using a 5-point scale (Deauville score) can identify the patients who will be likely cured19,23 and that baseline quantitative 18FDG-PET/CT parameters, the maximum standardized uptake value (SUVmax), the metabolic tumor volume (MTV) and the total lesion glycolysis (TLG) can be powerful predictors of PMBCL outcomes24,25.

The IELSG26 also showed that PMBCL, generally limited to a single bulky lesion, has distinctive metabolic characteristics that may make MH more reproducible and more easily studied than in other lymphomas. Consequently, we characterized the MH patterns in the IELSG-26 cohort of PMBCL and examined the prognostic value of MH alone or combined with other functional PET parameters. Here we report the results of this assessment.

**Methods**

*Patient population*

Metabolic Heterogeneity was estimated on baseline 18FDG-PET/CT in 103 of 125 patients with histologically proven PMBCL prospectively enrolled in the International Extranodal Lymphoma Study Group (IELSG) 26 study (ClinicalTrial.gov identifier: NCT00944567). All were treated according to local policy with rituximab and doxorubicin-based immunochemotherapy regimens comprising either R-CHOP (rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone) or R-CHOP-like, R-VACOP-B (rituximab plus etoposide, leucovorin, doxorubicin, cyclophosphamide, vincristine, prednisone, and bleomycin) or R-MACOP-B (rituximab plus methotrexate, leucovorin, doxorubicin, cyclophosphamide, vincristine, prednisone, and bleomycin) regimens. Infusional chemotherapy regimens were not used. Ninety three patients also had consolidation mediastinal involved field radiotherapy. Treatment details have been reported previously.19,23

The study was conducted in accordance with the precepts of the Helsinki declaration and received approval from the local research ethical committee of each participating center. All patients gave their written consent.

*18FDG-PET/CT imaging procedures*

PET/CT studies were performed following standard technical procedures for the acquisition and elaboration of PET imaging, according to guidelines at the time when the study was planned26.

Baseline PET scans were performed within 14 days before commencing immune-chemotherapy. In cases where urgent treatment was required and the PET scan could not be performed before therapy started, the baseline scan was omitted after discussion with the clinical coordinators; these cases were excluded from this analysis.

PET/CT imaging was performed on full-ring integrated PET/CT systems. Each center was required to follow an active quality control and quality assessment protocol26. PET and CT images were acquired in the same session. CT scans obtained with a low dose protocol were used for attenuation-correction of the PET images. All patients were fasted for at least 6 hours before the injection of 250 - 370 MBq (4.5 MBq/Kg) of 18FDG. Blood glucose measured before injection of the radiotracer was <160 mg/dl in all patients. PET data were acquired in two or three dimensional mode from the mid-thigh towards the base of the skull after a standardized uptake time of 60 minutes (± 5 minutes). The PET acquisition time was at least 3 minutes per bed position. Images were reconstructed with validated and commercially available iterative algorithms according to the local protocols and standardized uptake values (SUV) were automatically calculated. For each examination the PET/CT image data were sent to the core lab where central review was performed19,24.

*18FDG-PET/CT parameters*

The 18FDG-PET/CT images obtained at baseline for initial staging were analyzed following a standard protocol on a dedicated workstation (Siemens SyngoMMWP Workstation VE36A; Siemens, Erlangen, Germany). Dedicated software (Syngo TrueD) automatically estimated the average and maximum SUV (SUVmean and SUVmax) and metabolic tumor volume (MTV) of the entire mediastinal lesion using an isocontour fixed threshold method based on 25% of the SUVmax, as previously described24. The total lesion glycolysis (TLG), which represents the sum of the SUV of the different voxels included in the segmented volume, was calculated as the product of SUVmean and MTV27.

MH was assessed using the cumulative SUV volume histograms (CSH) method13, as applied in previous studies in solid tumors14-16,28-30. In CSH the percentage of tumor volume (derived from PET-based tumor delineation methods31) with SUV above a certain threshold is plotted against that threshold value, which is varied from 0 to 100% of SUVmax (in this study from 25% to 100% according to the fixed threshold used for the MTV segmentation, therefore excluding intratumoral areas with absent or very low FDG uptake, mostly due to tumor necrosis). The area under the curve of CSH (AUC-CSH) is a quantitative index of tracer uptake heterogeneity where lower values correspond to increased heterogeneity32. MH value distribution in the study population was also evaluated by estimating the coefficient of variation (COV) of the intratumoral FDG uptake, which was calculated as the standard deviation of the SUV divided by the SUVmean of the segmented mediastinal lesion5,16.

The data were expressed as median and interquartile range (IQR). The PET-associated functional parameters were dichotomized using receiver-operating characteristic (ROC) analysis to identify the optimal cut-off point33. Progression-free survival (PFS) were defined according to the revised National Cancer Institute criteria34 and estimated using the Kaplan-Meier or life-table method as appropriate35. Follow-up was calculated as the median time to censoring using a reverse Kaplan-Meier analysis36. Differences between survival curves were analyzed by using the log-rank test37. For multivariable analysis, Cox regression with a stepwise backward selection process was performed on dichotomized variables to estimate the hazard ratio (HR) and its confidence interval (CI).38 The exact 95% CIs were calculated for incidence percentages. Negative predictive values (NPVs) and positive predictive values (PPVs) were calculated according to standard definitions39. Mann-Whitney U test was used to test differences between variables in two groups of patients40. Association between two variables was investigated through regression and correlation analysis (Pearson’s coefficient). P values of .05 or less (two-sided test) were considered to indicate statistical significance. Statistical analysis was conducted using the STATA 11 software package (Stata Corp 2009. Stata: Release 11 Statistical Software. College Station, TX, USA).

**Results**

Detailed clinical features and outcomes of the patients enrolled in the IELSG26 study have been published previously19,23,24. Baseline PET/CT scans were available in 103 out of 125 patients, having been omitted in 20 patients because of the urgency of treatment and excluded in 2 cases (after central control) for technical reasons. All were treated with rituximab and doxorubicin-based immunochemotherapy regimens, comprising R-CHOP either every 14 or 21 days or at intensified doses (n=16 ), R-VACOP-B or R-MACOP-B regimen (n=87). There was no difference in outcomes between the different chemotherapy regimens. Consolidation radiotherapy was considered standard policy in nearly all centers and was given to 93 patients. At a median follow-up of 62 months (IQR, 56-71 months), 12 patients had local progression during initial immunochemotherapy or within 3 months of its completion and 1 relapsed (also in the mediastinum) 17 months after the treatment start. Six had died, with an estimated 5-year PFS rate of 87% (95% CI, 79%-92%)24.

The main clinical characteristics of the 103 patients included in the present analysis and the description of the baseline functional PET parameters are summarized in Table 1.

AUC-CSH in the whole population ranged between 0.32 and 0.65 (median, 0.49; IQR, 0.43 – 0.54). Figure 1 provides an example of patients showing different degrees of MH. The AUC-CSH and COV methodologies provided a similar MH value distribution (see supplemental file).

The values of MH, irrespective of the methodology used, did not show a significant relationship with other baseline quantitative PET-derived parameters, namely SUVmax, MTV and TLG (see supplemental file), the only exception being a weak (r=-0.2) association with MTV when MH was measured using the AUC-CSH, which was not confirmed in multivariable analysis (data not shown).

The ROC analysis of the AUC-CSH values identified an optimal cut-off point of 0.45 to discriminate (p =0.011) patients experiencing disease progression or relapse with sensitivity of 69% (95% CI, 39-91%) and specificity of 72% (95% CI, 62-81%). Patients with progression or relapse had significantly lower AUC-CSH (corresponding to higher MH) than those remaining in continuous remission (AUC-CSH, 0.43; IQR, 0.37-0.47 vs 0.49; IQR, 0.44-0.54; Mann-Whitney U test p = 0.012) [Figure 2].

In univariate analysis of dichotomized parameters, PFS at 5 years was 94% (95% CI, 85-98%) for patients with low MH and 73% (95% CI, 55-85%) for those with high MH (log-rank test, p=0.0001) [Figure 3] . MH showed no significant correlation to other baseline prognostic factors including individual IPI factors, the presence of bulky disease or B-symptoms (data not shown). Patients with high MH had also a significantly worse overall survival (results shown in the supplemental file).

In stepwise Cox models (see supplemental file) including dichotomized MH, TLG, SUVmax MTV, bulky disease (>10 cm), IPI or aaIPI, as well as age as a continuous variable, only elevated MH (HR, 12.8; 95% CI, 3.3-49.9; p<0.001 and elevated TLG (HR, 46.5; 95% CI, 5.8-373.8; p<0.001) remained independently associated with significantly shorter PFS.

A prognostic model for PFS based on the combination of MH and TLG at baseline was then built and tested. It identified patients with significantly different outcomes [Table 2] . In comparison with either MH or TLG alone, this model resulted in a much higher PPV without a detrimental effect on the NPV [Table 3] and confirmed a significantly poorer outcome in patients with both high TLG and high MH (HR, 31.2; 95% CI, 9.9-97.8; p<0.001), while no treatment failure was seen in the group of patients with low TLG and low MH at baseline [Figure 4]. These risk groups showed no association with conventional prognostic indices such as IPI/aaIPI or with specific clinical features in univariate analysis. Among the 9 patients with high TLG and high MH, only 2 had stage III disease and unfavourable IPI or aaIPI.

**Discussion**

Although treatment outcomes of PMBCL are generally favorable, there remains an important minority in whom the initial treatment fails, usually as a result of rapidly emerging chemorefractory disease, for which salvage treatments are often unsuccessful20,21. The present study was aimed at the early identification of these patients. In a retrospective Italian series of 138 consecutive patients treated with CHOP or MACOP-B/VACOP-B (with radiotherapy in 75% of the cases but without rituximab) all the patients with stable disease or progression during initial chemotherapy and all those relapsing after initial remission died of lymphoma, irrespective of first-line and salvage treatment type21. Another retrospective study from the Princess Margaret Hospital in Toronto, showed significantly inferior overall response to salvage chemotherapy (25% vs 48%) and overall survival rates after progression (15% vs 34% at 2 years) in 37 patients with relapsing or refractory PMBCL compared with a control group of 143 relapsing patients with other diffuse large B cell lymphomas20. Some studies have reported favorable outcomes in small groups of patients with chemo-sensitive relapse treated with autologous transplantation41, and chemosensitivity before high-dose therapy emerged as the strongest predictor of survival in these patients20,41. However, this can only be assessed when relapse has already occurred, and overall the prognosis following recurrence is very poor.

Malignant tumor cells are heterogeneous in various respects; the factors contributing to cancer MH have mainly been studied in solid tumors, and reflect cell metabolism, proliferation, blood flow, and hypoxia5. There is increasing evidence that quantifiable imaging parameters can be used in vivo to provide valid and reproducible estimates of cancer heterogeneity42. Increased metabolic heterogeneity (MH) on PET scans has been correlated with worse response to treatment in several solid tumors and may be an indicator of resistance to therapy 6,9-14. We have previously shown the prognostic impact of functional (quantitative) PET parameters such as SUVmax, MTV, and TLG in PMLBL at diagnosis24. Given that 18FDG uptake is not homogeneous within the large mediastinal lesions that characterize PMBCL, we investigated the hypothesis that the degree of MH in PMBCL may have prognostic value, and examined the relationship between baseline MH and other functional PET/CT parameters.

There are no data on reproducibility of visual estimation of MH; several computational methods have been proposed for its calculation and none has been clearly shown to be superior. We utilized the AUC-CSH, which has been widely described in the literature, mainly in solid tumors but also in lymphomas14-16,28-30. This methodology is straightforward to apply and has been shown to be highly reproducible43. Moreover, since AUC-CSH is independent of tumor volume32, the impact of segmentation can be controlled, making AUC-CSH particularly suitable for the study of PMBCL, where the presence of a bulky mass at presentation requires estimation of MTV using the isocontour fixed threshold method, based on 25% of the SUVmax44.

The calculation of AUC-CSH using the segmented MTV, thereby excluding necrotic areas with absent or very low FDG uptake, resulted in the most accurate estimation of metabolic distribution. Moreover, we compared our results estimating MH with an alternative method based on the SUVmax coefficient of variation5,16. The correlation between the two methodologies was good and both generated superimposable results when the impact of MH on treatment outcome was analyzed45.

Our results illustrate that the heterogeneity of 18FDG uptake does not hinge on the main quantitative PET parameters, but may be an additional biomarker in PMBCL treated with aggressive immunochemotherapy and radiotherapy.

A major challenge in the treatment of PMBCL is the need to identify, at diagnosis, the few patients for whom initial therapy will fail and who then risk an extremely poor outcome. In our previous studies we showed that the prognostic utility of functional PET parameters has a limitation in their low positive predictive value (PPV)24. This led us to study whether the effectiveness of TLG for risk stratification could be improved by combining TLG with other clinical and imaging parameters. We showed that the combination of baseline TLG and post-treatment Deauville Score results in a better PPV, however this approach identifies poor-risk patients only after the standard immunochemotherapy is completed and does not allow a risk-tailored strategy from the outset, such as treatment intensification25.

Although the restricted sample size and low event rate precluded testing in a validation cohort, the present study indicates that the combination of TLG and MH may offer a tool for early identification of patients with a high risk of initial treatment failure, which can be tested in other studies. In comparison with either MH or TLG alone, the proposed model showed much greater ability to specify patients with very unfavorable outcomes, with a markedly improved PPV. Further studies are warranted to ascertain whether this subset of poor-risk patients carries specific biological features that may engender the observed chemoresistance.

Few previous studies, all in solid tumors, have addressed the prognostic utility of combining intratumoral MH with other functional PET parameters46-48. Our observations are in keeping with the results of a study on oropharyngeal squamous cell carcinoma where MH was combined with TLG to build a risk score46.

In the present study, PMBCL patients with both low MH and low TLG had a very favorable outcome and might be suitable for treatment de-escalation. However, it must be noted that almost all patients had mediastinal irradiation, precluding any conclusion on the need for radiotherapy in this group.

On the other hand, and more importantly, the combination of high MH and high TLG accurately identifies at diagnosis the few patients with very poor prognosis and may, therefore, represent a powerful tool to allow patients to be selected for more intensive treatment. If confirmed in other cohorts this could inform the design of future clinical trials and may represent an important chance to reduce the number of patients with refractory PMBCL.

**Acknowledgements**

The IELSG26 study has been supported by grants ICP OCS-01709-04-2005 and ICP OCS-02062-03-2007 from OncoSuisse and was endorsed by the Italian Lymphoma Foundation and by Cancer Research United Kingdom (CRUKE/06/035). The participation of the UK centers was coordinated by the Southampton Clinical Trials Unit. A list of centers and investigators participating to the IELSG26 study has been previously reported19,24. The authors thank Rita Gianascio Gianocca for her editorial assistance and are also grateful for the precious contributions of Marina Cesaretti and Carol J. Tyas to study organization and data management. Finally we thank Dr. Georg Stüssi for critical reading of the manuscript.

**Authorship Contributions**

L.C., L.M. and E.Z. designed the study, performed research and analyzed the data. L.C., M.M., L.C., P.W.M.J. and E.Z. wrote the paper. All authors contributed to patient data collection or management, critically reviewed and approved the manuscript and shared final responsibility for the decision to submit.

**Disclosure of Conflicts of Interest**

The authors declare no competing financial interests.

**References**

1. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*. 2010;1805(1):105-117.

2. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. *Science*. 2013;339(6127):1546-1558.

3. Gupta RG, Somer RA. Intratumor Heterogeneity: Novel Approaches for Resolving Genomic Architecture and Clonal Evolution. *Mol Cancer Res*. 2017;15(9):1127-1137.

4. Pribluda A, de la Cruz CC, Jackson EL. Intratumoral Heterogeneity: From Diversity Comes Resistance. *Clin Cancer Res*. 2015;21(13):2916-2923.

5. Chung HH, Kang SY, Ha S, et al. Prognostic value of preoperative intratumoral FDG uptake heterogeneity in early stage uterine cervical cancer. *J Gynecol Oncol*. 2016;27(2):e15.

6. Tixier F, Le Rest CC, Hatt M, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med*. 2011;52(3):369-378.

7. Basu S, Kwee TC, Gatenby R, Saboury B, Torigian DA, Alavi A. Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. *Eur J Nucl Med Mol Imaging*. 2011;38(6):987-991.

8. Pugachev A, Ruan S, Carlin S, et al. Dependence of FDG uptake on tumor microenvironment. *Int J Radiat Oncol Biol Phys*. 2005;62(2):545-553.

9. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging*. 2013;40(1):133-140.

10. Davnall F, Yip CS, Ljungqvist G, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging*. 2012;3(6):573-589.

11. Lee M, Lee H, Cheon GJ, et al. Prognostic value of preoperative intratumoral FDG uptake heterogeneity in patients with epithelial ovarian cancer. *Eur Radiol*. 2017;27(1):16-23.

12. Eary JF, O'Sullivan F, O'Sullivan J, Conrad EU. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. *J Nucl Med*. 2008;49(12):1973-1979.

13. El Naqa I, Grigsby P, Apte A, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognit*. 2009;42(6):1162-1171.

14. Tixier F, Vriens D, Cheze-Le Rest C, et al. Comparison of Tumor Uptake Heterogeneity Characterization Between Static and Parametric 18F-FDG PET Images in Non-Small Cell Lung Cancer. *J Nucl Med*. 2016;57(7):1033-1039.

15. Hanaoka K, Hosono M, Tatsumi Y, et al. Heterogeneity of intratumoral (111)In-ibritumomab tiuxetan and (18)F-FDG distribution in association with therapeutic response in radioimmunotherapy for B-cell non-Hodgkin's lymphoma. *EJNMMI Res*. 2015;5:10.

16. Watabe T, Tatsumi M, Watabe H, et al. Intratumoral heterogeneity of F-18 FDG uptake differentiates between gastrointestinal stromal tumors and abdominal malignant lymphomas on PET/CT. *Ann Nucl Med*. 2012;26(3):222-227.

17. Johnson PW, Davies AJ. Primary mediastinal B-cell lymphoma. *Hematology Am Soc Hematol Educ Program*. 2008:349-358.

18. Dunleavy K, Pittaluga S, Maeda LS, et al. Dose-Adjusted EPOCH-Rituximab Therapy in Primary Mediastinal B-Cell Lymphoma. *New England Journal of Medicine*. 2013;368(15):1408-1416.

19. Martelli M, Ceriani L, Zucca E, et al. [18F]fluorodeoxyglucose positron emission tomography predicts survival after chemoimmunotherapy for primary mediastinal large B-cell lymphoma: results of the International Extranodal Lymphoma Study Group IELSG-26 Study. *J Clin Oncol*. 2014;32(17):1769-1775.

20. Kuruvilla J, Pintilie M, Tsang R, Nagy T, Keating A, Crump M. Salvage chemotherapy and autologous stem cell transplantation are inferior for relapsed or refractory primary mediastinal large B-cell lymphoma compared with diffuse large B-cell lymphoma. *Leuk Lymphoma*. 2008;49(7):1329-1336.

21. Todeschini G, Secchi S, Morra E, et al. Primary mediastinal large B-cell lymphoma (PMLBCL): long-term results from a retrospective multicentre Italian experience in 138 patients treated with CHOP or MACOP-B/VACOP-B. *Br J Cancer*. 2004;90(2):372-376.

22. Abou-Elella AA, Weisenburger DD, Vose JM, et al. Primary mediastinal large B-cell lymphoma: a clinicopathologic study of 43 patients from the Nebraska Lymphoma Study Group. *J Clin Oncol*. 1999;17(3):784-790.

23. Ceriani L, Martelli M, Gospodarowicz MK, et al. Positron Emission Tomography/Computed Tomography Assessment After Immunochemotherapy and Irradiation Using the Lugano Classification Criteria in the IELSG-26 Study of Primary Mediastinal B-Cell Lymphoma. *Int J Radiat Oncol Biol Phys*. 2017;97(1):42-49.

24. Ceriani L, Martelli M, Zinzani PL, et al. Utility of baseline 18FDG-PET/CT functional parameters in defining prognosis of primary mediastinal (thymic) large B-cell lymphoma. *Blood*. 2015;126(8):950-956.

25. Ceriani L, Martelli M, Conconi A, et al. Prognostic models for primary mediastinal (thymic) B-cell lymphoma derived from 18-FDG PET/CT quantitative parameters in the International Extranodal Lymphoma Study Group (IELSG) 26 study. *Br J Haematol*. 2017;178(4):588-591.

26. Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. *J Nucl Med*. 2006;47(5):885-895.

27. Larson SM, Erdi Y, Akhurst T, et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging*. 1999;2(3):159-171.

28. Yoon HJ, Kim Y, Kim BS. Intratumoral metabolic heterogeneity predicts invasive components in breast ductal carcinoma in situ. *Eur Radiol*. 2015;25(12):3648-3658.

29. Nakajo M, Nakajo M, Jinguji M, et al. The value of intratumoral heterogeneity of (18)F-FDG uptake to differentiate between primary benign and malignant musculoskeletal tumours on PET/CT. *Br J Radiol*. 2015;88(1055):20150552.

30. Kang SR, Song HC, Byun BH, et al. Intratumoral Metabolic Heterogeneity for Prediction of Disease Progression After Concurrent Chemoradiotherapy in Patients with Inoperable Stage III Non-Small-Cell Lung Cancer. *Nucl Med Mol Imaging*. 2014;48(1):16-25.

31. Nestle U, Kremp S, Schaefer-Schuler A, et al. Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. *J Nucl Med*. 2005;46(8):1342-1348.

32. van Velden FH, Cheebsumon P, Yaqub M, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur J Nucl Med Mol Imaging*. 2011;38(9):1636-1647.

33. Pepe MS. The statistical evaluation of medical tests for classification and prediction. New York: Oxford University Press; 2003.

34. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. *J Clin Oncol*. 2007;25(5):579-586.

35. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53:457-481.

36. Altman DG, De Stavola BL, Love SB, Stepniewska KA. Review of survival analyses published in cancer journals. *Br J Cancer*. 1995;72(2):511-518.

37. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data New York: John Wiley & Sons; 1980.

38. Cox DR. Regression models and life tables. *J R Stat Soc*. 1972;34:187-220.

39. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ*. 1994;309(6947):102.

40. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 1947;18:50-60.

41. Sehn LH, Antin JH, Shulman LN, et al. Primary diffuse large B-cell lymphoma of the mediastinum: outcome following high-dose chemotherapy and autologous hematopoietic cell transplantation. *Blood*. 1998;91(2):717-723.

42. Asselin MC, O'Connor JP, Boellaard R, Thacker NA, Jackson A. Quantifying heterogeneity in human tumours using MRI and PET. *Eur J Cancer*. 2012;48(4):447-455.

43. van Velden F, Nissen I, Jongsma F, et al. Test-retest variability of a cumulative SUV-volume histogram method for quantification of FDG uptake heterogeneity. *Journal of Nuclear Medicine*. 2012;53(supplement 1):2232.

44. Meignan M. Quantitative FDG-PET: a new biomarker in PMBCL. *Blood*. 2015;126(8):924-926.

45. Ceriani L, Milan L, Martelli M, et al. Metabolic heterogeneity of baseline 18-FDG PET-CT scan predicts outcome in primary mediastinal B-cell lymphoma. *Hematological Oncology*. 2017;35:60-61.

46. Cheng NM, Fang YH, Chang JT, et al. Textural features of pretreatment 18F-FDG PET/CT images: prognostic significance in patients with advanced T-stage oropharyngeal squamous cell carcinoma. *J Nucl Med*. 2013;54(10):1703-1709.

47. Chong GO, Lee WK, Jeong SY, et al. Prognostic value of intratumoral metabolic heterogeneity on F-18 fluorodeoxyglucose positron emission tomography/computed tomography in locally advanced cervical cancer patients treated with concurrent chemoradiotherapy. *Oncotarget*. 2017;8(52):90402-90412.

48. Mena E, Sheikhbahaei S, Taghipour M, et al. 18F-FDG PET/CT Metabolic Tumor Volume and Intratumoral Heterogeneity in Pancreatic Adenocarcinomas: Impact of Dual-Time Point and Segmentation Methods. *Clin Nucl Med*. 2017;42(1):e16-e21.

**Tables**

*Table 1.* Main clinical characteristics and functional PET parameters at presentation in the studied cohort (n= 103).

|  |  |  |
| --- | --- | --- |
| Characteristics | Median (IR) | No. of patients (%) |
| Age (years)  ≤ 60 years  ≤ 40 years |  | 97 (94)  75 (73) |
| Female sex |  | 63 (61) |
| ECOG Performance Status > 1 |  | 15 (15) |
| Mediastinal bulky disease >10 cm |  | 54 (52) |
| Ann Arbor stage I-II |  | 97 (94) |
| LDH >normal upper value |  | 77 (75) |
| Low and low-intermediate IPIrisk |  | 99 (96) |
| Low and low-intermediate aaIPIrisk |  | 82 (85) |
| Front-line treatment   R-CHOP or R-CHOP like regimen\*  R-VACOP-B or R-MACOP-B regimen\*  with RT |  | 16 (16) 87 (84) 93 (90) |
| SUVmax  Elevated (>22.2)# | 18.8 (15.5-23) | 30 (29) |
| MTV  Elevated (>703, PFS cut-point)# | 406 (267-641) | 20 (19) |
| TLG  Elevated (>5814, PFS cut-point)# | 4261 (2363-6398) | 33 (32) |

Abbreviation: IR, interquartile range; LDH, serum lactate dehydrogenase; IPI, international prognostic index; RT, consolidation mediastinal radiotherapy; SUVmax, maximum standard uptake value; MTV, metabolic tumor volume; TLG, total lesion glycolysis; PFS, progression-free survival.

\*Details on the immunochemotherapy regimens and radiotherapy plans have been previously reported19.  
# Dichotomized using receiver operating characteristics (ROC) analysis to identify the optimal cut point, as previously reported24

*Table 2.* Progression-free survival according to risk groups defined by the model built on baseline TLG and MH

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Risk group | Parameter | N | PD/relapse | Median PFS | 5-yr PFS | 95% CI |
| *Low risk* | |  |  |  |  |  |
|  | Low TLG + Low MH | 45 | 0 | n.r. | 100% | \_\_ |
| *Intermediate risk* | | *49* | *5* | *n.r.* | *90%* | *77-96%* |
|  | Low TLG + High MH | 25 | 1 | n.r. | 96% | 75-99% |
|  | High TLG + Low MH | 24 | 4 | n.r. | 83% | 61-93% |
| *High risk* | |  |  |  |  |  |
|  | High TLG + High MH | 9 | 8 | 4.3 months | 11% | 1-39% |

Abbreviation: PD, disease progression; PFS, progression-free survival; CI, confidence interval; MH, metabolic heterogeneity; TLG, total lesion glycolysis, n.r., not reached.

*Table 3.* Comparison of the prognostic power of the model combining baseline TLG and MH versus either TLG alone or MH alone.

|  |  |  |
| --- | --- | --- |
| Parameter | PPV (N) | NPV (N) |
| High MH | 26% (9/34) | 94% (65/69) |
| High TLG | 36% (12/33) | 99% (69/70) |
| High TLG + High MH | 89% (8/9) | 95% (89/94)  - Low TLG + Low MH, 100% (45/45)  - High TLG or High MH, 90 % (44/49) |

Abbreviation: NPV, negative predictive value; PPV, positive predictive value; MH, metabolic heterogeneity;   
TLG, total lesion glycolysis.

**Figure Legends**

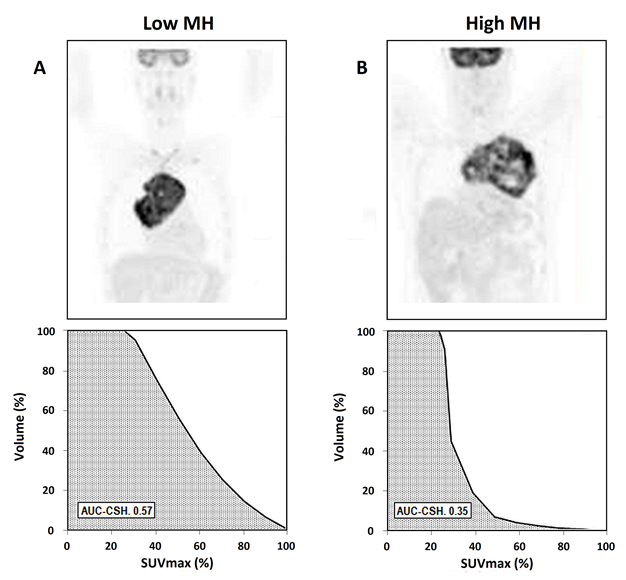
*Figure 1.* Examples of the various degrees of metabolic heterogeneity (MH) in primary mediastinal B-cell lymphoma. Panel A shows the PET imaging and the corresponding AUC-CSH histogram in a patients with low heterogeneity (AUC-CSH, 0.57) and panel B the PET imaging and the corresponding AUC-CSH histogram in a patient with high heterogeneity (AUC-CSH, 0.35). Lower AUC-CSH values correspond to higher metabolic heterogeneity .

*Figure 2.* Metabolic heterogeneity in the patients with continuous complete remission (CCR, white box) and in those with lymphoma relapse or progression (PD, gray box). Lower AUC-CSH values correspond to higher metabolic heterogeneity.

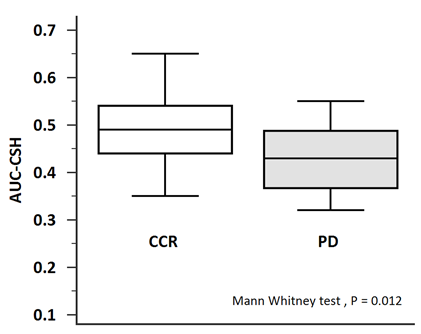
*Figure 3.* Kaplan-Meier estimate of progression-free survival according to the metabolic heterogeneity (MH) of the mediastinal lymphomatous mass in the baseline PET scans. Lower AUC-CSH values correspond to higher metabolic heterogeneity, shadows indicate 95% confidence intervals.

*Figure 4.*Kaplan-Meier estimate of progression-free survival according to a prognostic score based on the combination of MH and TLG at baseline, with shadows indicating 95% confidence intervals. Low risk patients are defined by the presence of both low MH and low TLG at presentation. The intermediate risk group comprises the patients with either low MH and high TLG or high MH and low TLG. High risk patients are defined by the presence of both high MH and high TLG. The trend test for the overall model was statistically significant, as well as the comparison (log-rank test) of the individual curves (low risk vs. intermediate risk, P=0.0286; low risk vs. high risk, P<0.0001; intermediate risk vs. high risk, P<0.0001).

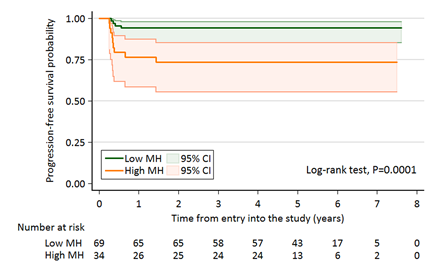
**Figures**

****

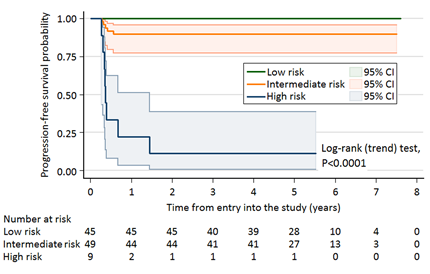
*Figure 1.*



*Figure 2.*

**

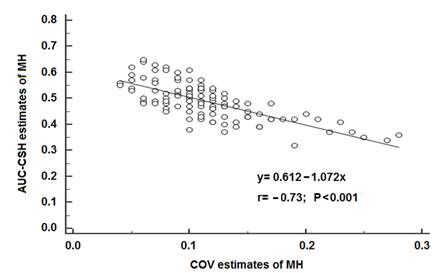
*Figure 3.*



*Figure 4.*

**SUPPLEMENTAL MATERIAL**

**a. Correlation between COV and AUC-CSH**



***Figure S1.*** Correlation between COV and AUC-CSH for the estimation of metabolic heterogeneity

**b. Lack of correlation between MH and other functional PET parameters**

.

****

***Figure S2.*** The values of MH, irrespectively of the methodology used (either COV or AUC-CSH) did not show a significant relationship with other baseline quantitative PET-derived parameters (namely, SUVmax, MTV and TLG) with the only exception of a weak (r=-0.2) association with MTV when MH was measured using the AUC-CSH, which was not confirmed at multivariate analysis (data not shown).

**c. Overall survival (OS) analysis**

***Table S1****.* ROC analysis of the AUC-CSH values for the identification of their optimal cut point

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Endpoint | ROC curve AUC (95%CI) | P-value | Cutoff | Sensitivity (95%CI) | Specificity (95%CI) |
| PFS | 0.72 (0.62-0.80) | 0.011 | 0.45 | 69% (39-91%) | 72% (62-81%) |
| OS | 0.65 (0.55-0.74) | 0.290 | 0.36 | 33 (4-78%) | 98% (93-99%) |

Abbreviation: ROC, receiver operating characteristics; CSH, cumulative SUV(standard uptake value )-volume histograms; CI, confidence interval; OS, overall survival; PFS, progression-free survival; AUC, area under the curve.



***Figure S3.*** Kaplan-Meier estimate of overall survival according to the metabolic heterogeneity (MH) of the mediastinal lymphomatous mass in the baseline PET scans. The apparent conundrum of overall survival rate worse than the progression-free one in the patients with high MH is related to the presence (determined by means of the receiver operating characteristic curve) of different cut-off values for progression (AUC-CSH, ≤ 0.45) and for death (AUC-CSH, ≤ 0.36). This is also the reason for the different number of patients at risk.

**d. Multivariable analysis of the impact of metabolic heterogeneity on progression-free survival  
(controlling for the other functional PET parameters, IPI and age)**

Stepwise backward Cox regression (significance level for removal from the model P≥0.05)

Model included dichotomized baseline AUC-CSH, TLG, MTV, SUVmax, Bulk>10cm, IPI, aaIPI as well as age as a continuous variable.

Begin with full model:

p = 0.8062 > 0.05, removing Bulk>10cm

p = 0.2371 > 0.05, removing SUVmax

p = 0.3389 > 0.05, removing age

p = 0.2303 > 0.05, removing aaIPI

p = 0.0931 > 0.05, removing MTV

p = 0.0851 > 0.05, removing IPI

Cox regression - no ties

Number of subjects = 97

Number of observations = 97

Number of failures = 11

Time at risk = 458.205493

Likelihood ratio chi2(2) = 34.33

Log likelihood = -32.567757 Probability > chi2 < 0.0001

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***VARIABLE*** | ***Hazard Ratio*** | ***Standard Error*** | ***P-value*** | ***95% Confidence Intervals*** |
| AUC-CSH | 12.83 | 8.89 | <0.001 | 3.30 - 49.93 |
| TLG | 46.52 | 49.47 | <0.001 | 5.79 - 373.85 |

The inclusion of aaIPI in the model above, with the removal of patients older than 60 reduced the number of subjects (from 103 to 97) and of failures (from 13 to 11). A model including only IPI (data not shown) and comprising the entire cohort (N=103) resulted in nearly superimposable results (HR, 10.11 for AUC-CSH and 50.10 for TLG respectively), confirming that metabolic heterogeneity (AUC-CSH) and TLG are independently associated with outcome.