

Improving conventional prognosticators in diffuse large B cell lymphoma using marker ratios

Kim-Anh LÊ CAO^{a,1}, Colm KEANE^a, Erica HAN^a, Dipti TALAULIKAR^{c,d} and Maher GANDHI^{a,b}

^a *The University of Queensland Diamantina Institute*

Translational Research Institute, Brisbane, Queensland, Australia

^b *Princess Alexandra Hospital, Brisbane, Queensland, Australia*

^c *Canberra Hospital, Canberra, Australian Central Territory*

^d *Australian National University Medical School, Australian Central Territory.*

Abstract. Risk stratification for diffuse large B-cell lymphoma (DLBCL) is required as patients with disease may not be cured despite initial R-CHOP treatment. We investigated gene ratio tests to predict survival outcome of DLBCL patients based on the relationship between immune-effector and inhibitory (checkpoint) genes from nanoString™ nCounter in 158 paraffin-embedded DLBCL tissues. We assessed the predictive value of several possible gene ratios using a tree-based survival statistical model, and investigated the predictive value of those gene ratios in an independent R-CHOP treated cohorts of 233 patients. We showed that an immune ratio composed of CD4*CD8*:(CD163/CD68)*PD-L1 was able to stratify overall survival better than single or combination of immune markers, distinguishing groups with disparate 4-year survivals (92% versus 47%). The immune ratio was independent of and added to the revised international prognostic index (R-IPI) and cell-of-origin (COO) and has potential implications for selection of patients for checkpoint-blockade within clinical trials.

Keywords. Survival analysis, gene expression, tumor associated macrophages.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is an aggressive B-cell lymphoma that has ~30% mortality despite R-CHOP chemotherapy treatment. Prognosticators such as the cell-of-origin (COO) and international prognostic index (IPI) and/or revised-international prognostic index (R-IPI) [1] were established to identify poor outcome patients. However, these established prognosticators are not sufficient to accurately predict outcome [2]. There is therefore a pressing need to develop tools that add to the predictive power of conventional prognosticators to more accurately predict response to initial therapy.

Some factors are known to have a great influence on the behavior of multiple tumor types. The role of infiltrating immune cells within tumor micro-environment (TME) in particular, are capable of influencing tumor progression. Other involved immunological cell types are macrophages, which are known to have a dual role in

¹ Corresponding Author.

cancer depending on their phenotype. Specifically, tumor associated macrophages (TAM) possess anti-tumoral capacities (M1) and pro-tumoral capacities (M2). This explains why the percentage of M2 macrophages of the total macrophage count (i.e. the CD163/CD68 ratio) and the M1/M2 ratio have been found to be good prognosticators to predict survival and metastatic ability of various types of tumors, such as melanoma, non-small cell lung carcinoma and angioimmunoblastic T-cell lymphoma (AITL). For example [3] and [4] reported a significantly better survival for patients based on their CD163/CD68 ratio in glioma and AITL tumors respectively. The question on how to define the optimal cutoff on such ratio to segregate good and poor outcomes remains however unanswered. A naïve median cut-off stratifying patients into two subgroups of equal sizes does not correspond to the actual survival rates for a given disease. Therefore, while cancer researchers have investigated gene expression ratios to predict good prognosis after cancer therapy, the challenge in need to be addressed is the identification of 1/ a robust gene expression ratio combined to the determination of 2/ an optimal cutoff value in the ratio to best segregate outcomes.

In our study, we quantified gene expression in 158 paraffin-embedded tissues from R-CHOP treated DLBCL patients from 2 Australian centers using the Nanostring™ digital multiplex platform. We investigated ratios of immune-effector genes (TNF α , CD137, CD3, CD4, CD8, CD56) with immune-checkpoints (PD1, PDL1, PDL2, LAG3, TIM3, CCL3, CD163, CD68, M2) with the rationale that a ratio of the antagonistic forces of immune-effectors to immune-checkpoints might best reflect net anti-intratumoral immunity, with low ratios indicating reduced immunity. We used tree-based survival statistical models as an innovative analytical tool to test all combinations of gene ratios and to determine optimal cut-offs to stratify overall survival independently from R-IPI and COO. The results were further tested in an external international R-CHOP treated DLBCL cohort drawn from a publically available database performed on a gene expression Affymetrix platform.

1. Data and methods

1.1. Nanostring™ digital multiplex gene expression study

The study was approved by responsible Ethics Committees at participating sites. The initial tissue cohort comprised 252 patients with histologically confirmed DLBCL. All patients received R-CHOP, and were otherwise selected solely on the basis of Formalin Fixed Paraffin Embedded (FFPE) tissue availability. Only de-novo cases of DLBCL were included, and grade IIIB follicular lymphoma, transformed follicular lymphoma, HIV-positive and post-transplant patients were excluded, resulting in 158 samples with survival data available. The data were normalised against positive and negative controls in the assay. In addition, four housekeeping genes were employed to normalise the data to account for potential differences in RNA quality. Approximately 10% of samples did not pass QC and were discarded from the analysis. The data including the expression of 48 genes were further log₂ transformed.

1.2. Affymetrix gene expression for validation

For validation of the generated model's findings we used a publicly available gene expression data set using Affymetrix HG-U133 Plus 2.0 GeneChips platform [5] on

fresh-frozen samples. The normalised data include the expression levels of 54,676 transcripts for 233 patients treated with R-CHOP (GSE10846). Probes mapping to our genes of interest were identified through the jetset score [6]. .

1.3. Ratios of markers

Our study specifically investigates ratios of the antagonistic forces of immune-effectors to immune-checkpoints to reflect net anti-intratumoral immunity. We therefore considered the immune effectors genes CD4, CD8, CD56 and CD137 as part of ‘numerators’ in the ratios, and immune-checkpoints CD163, CD68, M2 (CD163/CD68), PD1 and PDL1 as part of ‘denominators’. We tested each gene marker individually (10 genes), simple ratios of one numerator gene divided by one denominator gene (12 ratios), and finally all possible combinations of these genes (44 ratios).

1.4. Tree-based survival models

A survival analysis based on a tree-structured statistical model [7] was applied to stratify patients into two survival subgroups (termed ‘high or low-risk’). Tree-structured regression models perform recursive binary partitioning to model the association between response and covariates. In our study, the response is the survival time of each patient, and a single covariate, the expression levels of a marker ratio from the Nanostring data is considered in the regression model. Therefore, for each ratio a survival tree with a single node split was modelled. The main aim of the approach is to determine the optimal cutoff value on the marker ratio (i.e. the ‘*best split*’) so that patients are stratified into two survival subgroups with the best possible segregation (see example in Fig 1A). The best split is optimally determined based on the entropy criterion, a criterion that is commonly used in tree-based partitioning approaches, as fully described in [8].

In summary, for each marker ratio from the Nanostring data, the patients were stratified according to the following rule: if the ratio value is less than or equal to the tree cutoff, then the patient is assigned to the ‘high-risk’ group; if the ratio value is strictly greater than the tree cutoff, then the patient is assigned to the ‘low-risk’ group. To assess the prognostic value of each marker ratio given the determined cutoff, a log rank test was applied to test the difference between the survival curves from both groups. A total of 66 survival trees were fitted on single gene markers (10), simple marker ratios (12) and all combinations of marker ratios (44) expression levels. The p-values from the log-rank tests along with the cutoff value and the proportion of 4-year survival were recorded.

We used the Affymetrix cohort to evaluate the ability of the survival trees to stratify patients into poor and good- risk groups. Marker ratios were calculated as in the Nanostring data. Using the rules described above, we stratified the Affymetrix patients into two subgroups and compared those subgroups with their actual survivals using a log rank test. P-values were adjusted using the Bonferroni multiple correction procedure. We used the R statistical software [9] and the R packages `party` and `partykit` for the tree-structure survival models [10].

2. Results

2.1. Comparison between single markers and ratios

As highlighted by previous studies, we found that the M2 ratio was prognostic ($P=0.0056$) whereas CD163 and CD68 alone were not ($P=1$).

The tree-structured survival models identified a number of prognostic immune molecule combinations. Within the top ten combinations (based on the tree-based survival curve tests), CD4, CD8, CD137, CD56 as numerators, and M2, PD-1 and PD-L1 as denominators were all present, with all except CD56 occurring frequently (\geq seven times). The highest ranked combination with respect to significantly different survival curves between two identified subgroups was the product of the immune-effectors CD4 and CD8, in a ratio with the products of M2 and PD-L1 (the CD4*CD8/M2*PD-L1 immune-ratio). This was selected for further analysis below. The optimal cut-off identified with the tree-structured model was of -0.279 . Patients with an immune-ratio ≤ -0.279 were stratified into the poor-risk group, and those with an immune-ratio > -0.279 were stratified into the good-risk group (Fig 1A). Using Kaplan–Meier and the log-rank test, the CD4*CD8/M2*PD-L1 ratio separated patients into disparate high and low-ratio 4 year survivals that differed by 45% ($P<0.0001$). The proportions of patients in high and low-ratio groups were 59% and 41%.

2.2. Immune-effectors: checkpoint ratios add to conventional prognosticators.

Both R-IPI and COO separated patients into prognostic groupings ($P=0.002$ and $P=0.018$ and 4-year survivals of 83% versus 61%, and 82% versus 60% respectively). We showed that our immune-ratio into low and high-ratio groupings sub-stratified R-IPI and COO (not shown). For R-IPI, 37% of patients were in the best survival groupings (R-IPI and high-ratio immune score), with $\sim 20\%$ in each of the other categories. Similarly for COO, 43% of patients were in the best survival grouping (GCB and high-immune score), with $\sim 20\%$ in each of the other categories. Combining the immune-ratio with R-IPI and COO created two clearly separate groupings ($P<0.001$): a poor-risk group with R-IPI-low-ratio and/or non-GCB-low-ratio (4-year OS=40%) versus a good-risk group (4-year OS=89%). Thirty percent of patients had a poor-risk combined R-IPI-COO-immune-score. Importantly, in a Cox multivariate regression analysis the significance remained, with the immune-ratio independent of COO and R-IPI (immune ratio $P<0.0001$, COO $P=0.048$ and R-IPI $P=0.014$).

2.3. External validation of immune-ratios using the Affymetrix platform on frozen tissue.

The immune-ratio was then applied to the independent cohort of R-CHOP-like treated patients in which gene expression was quantified on fresh-frozen samples using an Affymetrix platform. The immune-ratio segregated patients into groupings with significantly different survivals. In particular, the ratio CD4*CD8/M2*PD-L1 led to high-ratio 76% versus low-ratio 63% ($P=0.0065$, Fig 1B). 60% of patients had a high-ratio, almost identical to that seen for the Australian-tissue cohort. With a multivariate Cox regression analysis the ratio was independent of COO and R-IPI (immune-ratio $P<0.017$, COO $P=0.011$ and R-IPI $P<0.0001$).

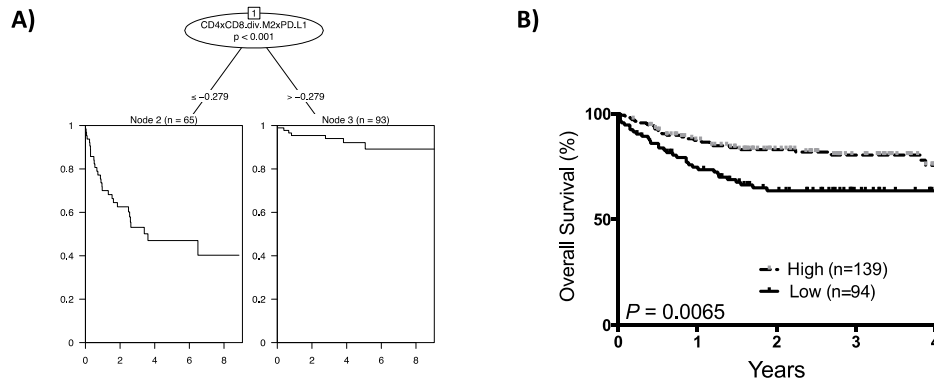


Figure 1. A) Nanostring cohort: Tree-structured survival model representation for the immune ratio $CD4*CD8/M2*PD-L1$ with optimal cut-off stratifying patients into poor and good risk groups, as displayed by the Kaplan Meier survival curves. Log rank test p-value is indicated in the split node. **B)** Affymetrix validation cohort: patients were stratified into the poor and good risk groups based on the same immune-ratio and cut-off determined in A (-0.279).

3. Discussion and future work

In this study, we have addressed some technical barriers that impede the translation of prospected clinical tests from discovery to clinical implementation.

Firstly, the study needed to address the problem of platform variability (digital count Nanostring and gene expression Affymetrix platforms), which is a systematic source of variation that is commonly observed when attempting to integrate or combine data generated from multiple platforms. Some batch-correction statistical approaches have been proposed but so far they have only been developed for microarray platforms [11, 12]. To our knowledge, it is the first time that ratio of several markers have been used to help reduce variability amongst platforms that largely differ in scale but also in the techniques that are used. It is also possible that the use of such ratio could enhance reproducibility of prognosticator for survival between different types of specimen (FFPE in our Nanostring cohort vs. fresh frozen tissues in the validation cohort). We acknowledge that our immune ratio would need to be further evaluated on an external cohort with FFPE tissues.

Secondly, in order to determine an optimal cutoff in our immune ratio, we used a tree-structure survival statistical model, which despite its sophistication is easy to visualise and understand. To our knowledge, this is the first time that such approach is used to stratify survival in patients. The advantage of the tree-based approach is that the stratification cutoff value is driven by survival data, instead of arbitrary cutoffs, or cutoffs driven by other information than survival as were previously proposed in the literature. For example in their glioma and angioimmunoblastic T-cell lymphoma studies, [3, 4] set up a 0.3 cutoff based on the $CD163/CD68$ ratio with no justification on such arbitrary choice. Other studies as those from [13] demonstrated the relevance of using gene expression ratios in lung and other thoracic malignancies. The cutoffs were determined based on a student t-test with pathological distinction (malignant

pleural mesothelioma vs adenocarcinoma) as outcome of interest. Such approach cannot be applied in our study as we do not have such information. For comparison we applied the naïve cutoff based on the median to segregate the patients cohort in two equal parts, bearing in mind that such partition does not reflect the mortality rate with R-CHOP chemotherapy treatment. In our Nanostring cohort, 21 out of the 66 marker ratios led to subgroups with significant differences between survival curves ($P \leq 0.05$). However, none of these markers were validated in the Affymetrix cohort. Using the survival trees, we obtained 54 marker ratios whose cutoffs led to significant differences between survival curves, amongst which four promising ratios were validated in the Affymetrix cohort. Such results could be improved by estimating a confidence interval or range around those cutoffs in order to fully assess the predictive value of our immune-ratio. These future directions will enable to translate expression profiling data into simple clinical tests.

Our ratio analysis solely included specific immune effectors and checkpoint gene markers as targeted by our Nanostring experiment, but our findings and statistical analyses could be generalised to a broader set of genes from high-throughput experiments and could also be applied to other marker ratios in other cancer studies.

References

1. Sehn, L.H., et al., *The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP*. *Blood*, 2007. **109**(5): p. 1857-1861.
2. Coiffier, B., *State-of-the-art therapeutics: diffuse large B-cell lymphoma*. *Journal of Clinical Oncology*, 2005. **23**(26): p. 6387-6393.
3. Komohara, Y., et al., *Possible involvement of the M2 anti-inflammatory macrophage phenotype in growth of human gliomas*. *The Journal of pathology*, 2008. **216**(1): p. 15-24.
4. Niino, D., et al., *Ratio of M2 macrophage expression is closely associated with poor prognosis for Angioimmunoblastic T-cell lymphoma (AITL)*. *Pathology international*, 2010. **60**(4): p. 278-283.
5. Lenz, G., et al., *Stromal gene signatures in large-B-cell lymphomas*. *New England Journal of Medicine*, 2008. **359**(22): p. 2313-2323.
6. Li, Q., et al., *Jetset: selecting the optimal microarray probe set to represent a gene*. *BMC bioinformatics*, 2011. **12**(1): p. 474.
7. Hothorn, T., K. Hornik, and A. Zeileis, *Unbiased recursive partitioning: A conditional inference framework*. *Journal of Computational and Graphical statistics*, 2006. **15**(3): p. 651-674.
8. Breiman, L., et al., *Classification and regression trees*. 1984: CRC press.
9. Team, R.C., *R: A language and environment for statistical computing [Internet]*. Vienna, Austria: *R Foundation for Statistical Computing; 2013*. Document freely available on the internet at: <http://www.r-project.org>, 2015.
10. Hothorn, T. and A. Zeileis, *partykit: A modular toolkit for recursive partytioning in R*. 2014, Working Papers in Economics and Statistics.
11. Johnson, W.E., C. Li, and A. Rabinovic, *Adjusting batch effects in microarray expression data using empirical Bayes methods*. *Biostatistics*, 2007. **8**(1): p. 118-127.
12. Le Cao, K.-A., et al., *YuGene: A simple approach to scale gene expression data derived from different platforms for integrated analyses*. *Genomics*, 2014. **103**(4): p. 239-51.
13. Gordon, G.J., et al., *Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma*. *Cancer research*, 2002. **62**(17): p. 4963-4967.