



COLLEGE of AMERICAN
PATHOLOGISTS

ARCHIVES

of Pathology & Laboratory Medicine

NEW ARTICLE

This article was posted on the *Archives* Web site as a New Article. New Articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the *Archives*. New Articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The DOI for this manuscript is doi: 10.5858/arpa.2024-0366-OA

The print version of this manuscript will replace the New Article version at the above DOI once it is available.

The Impact of Scoring Method on Accuracy and Reproducibility of Hans Cell-of-Origin Prediction in Excisional Biopsies of Diffuse Large B-Cell Lymphoma, Not Otherwise Specified

Oleksandr Yanko, MD; Andrew G. Lytle, MD, PhD; Pedro Farinha, MD, PhD; Merrill Boyle, BS; Graham W. Slack, MD; David W. Scott, MBChB, PhD; Jeffrey W. Craig, MD, PhD

• **Context.**—Aided by tissue microarray (TMA) technology, several RNA-correlated immunohistochemistry-based algorithms have been developed for cell-of-origin (COO) prediction in diffuse large B-cell lymphoma, not otherwise specified (DLBCL-NOS). However, there is currently no empirical evidence to guide the optimal application of these algorithms to whole tissue sections (WTSs).

Objective.—To assess the impact of various scoring methods on the accuracy and reproducibility of the popular Hans algorithm.

Design.—We compared 3 different WTS-based scoring methods, designated as *global*, *selective*, and *hotspot* scoring, to a matched TMA evaluation and gold standard RNA analysis (Lymph2Cx; germinal center B cell $n = 64$; activated B cell/unclassified $n = 68$) using a representative series of 132 excisional biopsies of de novo DLBCL-NOS. Positivity scores (10% increments) were submitted by 3 expert lymphoma pathologists, with 30% or more defining positivity.

Results.—Sixty-eight of the 132 cases of DLBCL-NOS (52%) exhibited variation in Hans immunohistochemistry marker phenotype as a consequence of scoring method and/or interscorer discordance. Although this led to changes in Hans COO assignment in 27 of 132 cases (20%), none of the WTS-based scoring methods were statistically inferior to one another in terms of raw accuracy. Hotspot scoring yielded the lowest proportion of borderline scores (20%–40% range) for BCL6 transcription repressor (BCL6) and IRF4 transcription factor (MUM1) but negatively impacted the balance between sensitivity and specificity for these markers. Selective scoring was associated with significantly worse interscorer concordance compared to TMA evaluation, which it was designed to replicate.

Conclusions.—Overall, our data favor the use of global scoring for its noninferior accuracy, solid interscorer concordance, nonnegative influence on individual Hans markers, and current widespread use.

(Arch Pathol Lab Med. doi: 10.5858/arpa.2024-0366-OA)

Diffuse large B-cell lymphoma, not otherwise specified (DLBCL-NOS) remains the World Health Organization Classification of Haematolymphoid Tumours, 5th edition

Accepted for publication December 18, 2024.

Supplemental digital content is available for this article. See text for hyperlink.

From the Department of Pathology (Yanko, Craig), and the Hematology Laboratory (Craig), University of Virginia Health System, Charlottesville; and the BC Cancer Centre for Lymphoid Cancer, Vancouver, British Columbia, Canada (Lytle, Farinha, Boyle, Slack, Scott).

The study was supported by a Program Project Grant from the Terry Fox Research Institute (grant No. 1023) and funding from the BC Cancer Foundation.

Craig reports consultancy/expert testimony for Bayer and honoraria from BeiGene. Scott reports consultancy/honoraria from Abbvie, AstraZeneca, GenMab, Roche, and Veracyte; research funding from Roche/Genentech; and named inventor on patents describing the use of gene expression to subtype aggressive B-cell lymphoma, including one licensed to nanoString Technologies. The other authors have no relevant financial interest in the products or companies described in this article.

Corresponding author: Jeffrey W. Craig, MD, PhD, Department of Pathology, Hematology Laboratory, University of Virginia Health System, 1215 Lee St, PO Box 800904, Charlottesville, VA 22908-0904 (email: jwc3q@uvahealth.org).

Arch Pathol Lab Med

(WHO-HAEM5) and International Consensus Classification (ICC-2022) diagnostic category encompassing the vast majority (80%–85%) of morphologically defined DLBCLs that lack MYC proto-oncogene, bHLH transcription factor (MYC) and BCL2 apoptosis regulator (BCL2) and/or BCL6 transcription repressor (BCL6) rearrangements and do not conform to any of the more specifically defined lymphomas of large B cells.^{1–3} Roughly 20 years ago, microarray-based gene expression profiling (GEP) studies identified 2 major groups of DLBCL based on presumed cell of origin (COO): germinal center B-cell-like (GCB) subtype and activated B-cell-like (ABC) subtype, with approximately 10% to 20% of tumors remaining unclassified.^{4,5} The distinction between GCB and ABC subtypes was subsequently shown to be prognostically significant in patients treated with rituximab chemotherapy, with early studies showing 5-year overall survival rates of 80% versus 50%, respectively.⁶ Based on the divergent oncogenic signaling pathways active in these subtypes, numerous clinical trials have since used COO to guide enrollment and therapy selection, albeit to limited success.⁷

The ICC-2022 supports the continued inclusion of COO designation in DLBCL-NOS, and the WHO-HAEM5 considers

COO to fall within its list of desirable diagnostic criteria for this entity. Given the time and resource requirements and technical challenges imposed by microarray-based GEP technologies, both classification systems accept the use of various GEP-correlated, immunohistochemistry (IHC)-based algorithms for COO determination in routine clinical practice (eg, Hans, Choi, Tally, Visco-Young).^{8–11} Among these, the Hans algorithm, which evaluates the expression of 3 cellular proteins (neprilysin [CD10], BCL6 transcription repressor [BCL6], and IRF4 transcription factor [MUM1], with $\geq 30\%$ tumor cell staining defining a positive result for each marker) to classify DLBCL-NOS into GCB or non-GCB subtypes (with the latter encompassing most ABC and GEP unclassified tumors), was the first reported and has achieved widespread clinical use because of its simplicity, reagent availability, and reasonable performance at predicting outcome.^{8,12–16} The utility of the Hans algorithm remains controversial, however, because of imperfect concordance with GEP (typically on the order of 70%–90%), the inability to recognize the GEP unclassified group, and failure to maintain prognostic significance in some studies.^{17–22} These shortcomings reflect the relative simplicity of Hans algorithm compared with GEP, as well as the potentially tremendous impact of technical variations across laboratories and poor interscorer concordance.²³

Suggestions to improve the accuracy and reproducibility of Hans and other IHC-based algorithms include standardization of reagents and laboratory techniques, centralized review of all DLBCL-NOS cases, and the development of guidelines for optimal IHC scoring by pathologists.^{17,23–25} As there are currently no official instructions or evidence-based literature to guide the clinical implementation of IHC-based COO algorithms, the manner of application has remained at the discretion of individual pathologists. When faced with excisional biopsy specimens, many pathologists choose to produce scores that are based on the complete assessment of whole tissue sections (WTSs) (herein referred to as the *global* assessment method), a strategy that seemingly avoids observer bias and also helps to account for any inherent tumor heterogeneity. However, the concept of global scoring is not addressed by the Hans study or any of the other IHC-based COO algorithms, all of which were instead developed using small-volume sampling via tissue microarray (TMA) technology, wherein each DLBCL specimen was ultimately represented by one or more 0.6- to 1.5-mm-diameter TMA core fragment(s) corresponding to an area of high tumor cell density as determined by CD20 B-lymphocyte antigen (CD20) staining or hematoxylin–eosin (H&E) analysis.^{8–11}

By assessing only a small portion of the available tumor, TMA scoring likely underrepresents the degree of staining heterogeneity present in the corresponding WTS and may also avoid some of the obstacles that regularly confound the evaluation of excisional specimens (eg, variable tumor cell fraction, zonal necrosis, tissue fixation/preservation gradients, cautery artifact, drying artifact, and crush artifact) (Supplemental Figure 1; see the [supplemental digital content](#) containing 5 figures and 2 tables). Further, by restricting the amount of tissue evaluated, TMAs may also promote greater concordance between scorers. For instance, several of the IHC-based COO algorithm studies reported excellent interscorer agreement, describing variations “by a single 10% increment”^{9,10} or “at 2%–6% ranges”¹¹ for individual IHC markers. Minimizing discordance is critical in this context, as Choi et al⁹ further found that introduction of as little as 10%

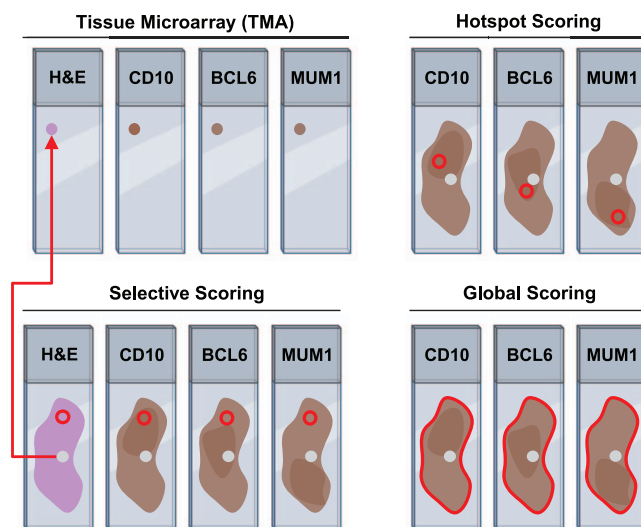


Figure 1. Four different scoring methods were compared in this study. In a conventional tissue microarray, cores of formalin-fixed, paraffin-embedded tumor are geometrically arrayed and all immunohistochemical marker scores are derived from evaluation of these limited core samples. In selective scoring, scores for each respective marker are based on the evaluation of a single “best” high-power field chosen by each independent scorer through the scorer’s evaluation of the hematoxylin–eosin slide; this area should be well preserved, show high tumor cell density, and be free of all potential artifacts. In hotspot scoring, scores for each respective marker are based on the single high-power field with the highest percentage of tumor cell staining. In global scoring, scores for each respective marker are based on the aggregate assessment of all tumor cells present. Created with BioRender.com. Abbreviations: BCL6, BCL6 transcription repressor; CD10, neprilysin; H&E, hematoxylin–eosin; MUM1, IRF4 transcription factor.

variability into antibody scoring reduced concordance with their original COO algorithm result to approximately 90%.

In order to assess the impact of scoring method on COO prediction in excisional DLBCL-NOS biopsies using the popular Hans algorithm, we compared global scoring of WTSs with 2 additional WTS scoring methods, which we have designated as *selective* and *hotspot* scoring, respectively, as well as to scoring of a matched TMA preparation (Figure 1). Our selective scoring method was designed to mimic TMA-based scoring, but without requiring the construction of a physical TMA. Hotspot scoring, as its name implies, targets the regions of tumor that exhibit the greatest degree of staining for each marker. A total of 132 DLBCL-NOS tumors were separately scored by 3 independent observers using each of the scoring methods outlined above and in Figure 1. The resulting Hans IHC marker scores and COO assignments were then used to assess algorithm accuracy compared with gold standard GEP using the Lymph2Cx assay and to assess the degree of interscorer concordance.^{26,27}

MATERIALS AND METHODS

Case Selection

From an established population-based registry cohort of patients aged 16 years or older treated with R-CHOP (rituximab plus cyclophosphamide, doxorubicin, vincristine and prednisone) with curative intent at the British Columbia Cancer Agency (Vancouver, Canada),^{27,28} we randomly selected 132 pretreatment formalin-fixed, paraffin-embedded excisional biopsies of de novo DLBCL-NOS that met the following criteria: (1) were well represented in TMA preparations, (2) had been successfully assigned a COO using the Lymph2Cx digital GEP assay, and (3) had sufficient material remaining in WTS for

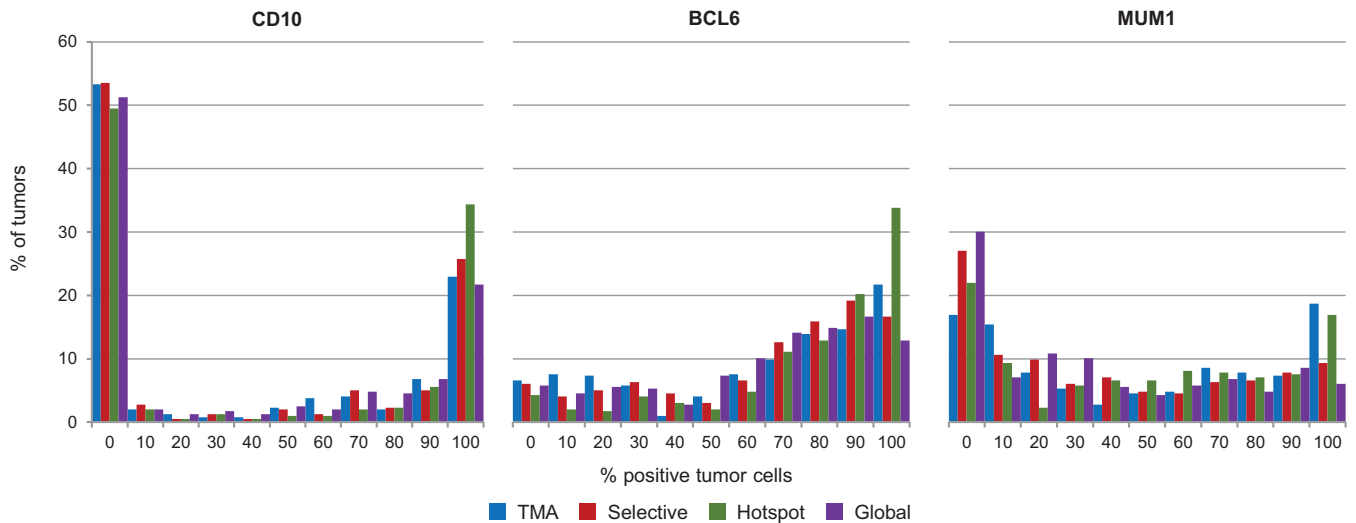


Figure 2. Histograms of CD10 (left), BCL6 (middle), and MUM1 (right) scores, across all scorers and stratified by scoring method, illustrate the typical profiles of these markers when applied to diffuse large B-cell lymphoma. Abbreviations: BCL6, BCL6 transcription repressor; CD10, nephrilysin; MUM1, IRF4 transcription factor; TMA, tissue microarray.

additional IHC testing. DLBCL-NOS diagnoses were confirmed through standardized review of original case materials, pathology reports, fluorescence in situ hybridization testing, and relevant clinical records, and confirmed to be consistent with current WHO-HAEM5 and ICC-2022 standards. This study received institutional review board approval and was performed in accordance with the Declaration of Helsinki.

TMA Construction, IHC, and COO Determination by GEP

Duplicate 0.6-mm cores from each formalin-fixed, paraffin-embedded excisional biopsy, corresponding to preselected areas with high tumor cell content, were used to construct a single TMA according to established protocols, as described previously.²⁷ IHC staining for the TMA and the WTS was performed on 4- μ m slides on a Dako Omnis platform (Agilent Technologies, Inc, Santa Clara, California) using the following antibodies: CD10 (clone 56C6, ready-to-use, GA648, Dako), BCL6 (clone PG-B6p, ready-to-use, GA625, Dako), and IRF4/MUM1 (clone MUM1p, ready-to-use, GA644, Dako). All IHC markers were scored independently by 3 expert hematopathologists (A.G.L., J.W.C., P.F.). Protein expression was recorded for each IHC marker as the percentage of tumor cells showing appreciable staining (defined as any definitive reactivity, above background, but without specific regard to intensity level) in 10% intervals. According to Hans algorithm, a threshold of 30% or greater was used to define positivity for all IHC markers. Gold standard COO assignments were provided by the Lymph2Cx 20-gene digital GEP assay performed on the NanoString platform (NanoString Technologies, Seattle, Washington), as described previously.^{26,27}

WTS-Based Scoring Methods

As illustrated in Figure 1, global scoring was defined as the scoring of each IHC marker across all evaluable tumor present on the WTS. Selective scoring was defined as the scoring of each IHC marker within the single “best” high-power field (using a standard $\times 40$ objective lens), representing an area chosen by each independent scorer through H&E evaluation to be of high tumor cell density and maximally free of necrosis, tissue fixation/preservation issues, and other obscuring factors. Hotspot scoring was defined as the scoring of each respective IHC marker within the single high-power field with the highest percentage of tumor cell staining. Results for individual Hans algorithm markers were then used to determine scoring method-specific COOs for comparison against one another, against matched TMA scoring, and against GEP-derived

COO assignments. Although 2 separate TMA tissue cores were available for most tumors, scorers were instructed to base their interpretations solely on the tissue core with the highest percentage of tumor cell staining, as previously described.^{8–10} In total, this work generated 4752 independent IHC marker scores along with an equal number of positive/negative calls and 1584 COO assignments.

Statistical Analysis

The performance of individual Hans IHC markers and COO predictions for each scoring method were compared with gold standard GEP-based COO to establish basic test characteristics (accuracy, sensitivity, specificity, positive predictive value [PPV], and negative predictive value [NPV]). Cohen κ scores were used to measure pairwise interscorer agreement,²⁹ and Fleiss κ scores were calculated to provide an index of concordance among all pathologists.³⁰ Analyses were performed using Microsoft Excel version 2018 in conjunction with the Real Statistics Resource Pack software (release 8.9.1; www.real-statistics.com). All reported *P* values are 2-sided, and those less than .05 were considered statistically significant.

RESULTS

Of the 132 DLBCL-NOS tumors included in this study, the GEP-based Lymph2Cx assay assigned 64 tumors (48.5%) to the GCB subtype, 51 (38.6%) to the ABC subtype, and 17 (12.9%) as unclassified, in keeping with established population-based data from North American DLBCL cohorts.³¹ The overall distributions of positive-staining tumor cells were as expected for CD10 (strongly biphasic), BCL6 (positively skewed), and MUM1 (weakly biphasic) across all scorers and scoring methods (Figure 2). Scores derived from the hotspot method were shifted upward compared with the other WTS-based scoring methods, by an average of roughly one 10% increment per marker (Supplemental Table 1). Accordingly, hotspot scoring resulted in a higher proportion of positive calls for each marker (ie, those $\geq 30\%$) compared with the other WTS-based scoring methods, significantly so for both BCL6 and MUM1 (pairwise Fisher exact tests, *P* < .01) (Supplemental Figure 2). These changes were accompanied by a significant decrease in the proportion of borderline scores (ie, those ranging from 20% to 40%) for hotspot scoring (101 of 1188 [8.5%])

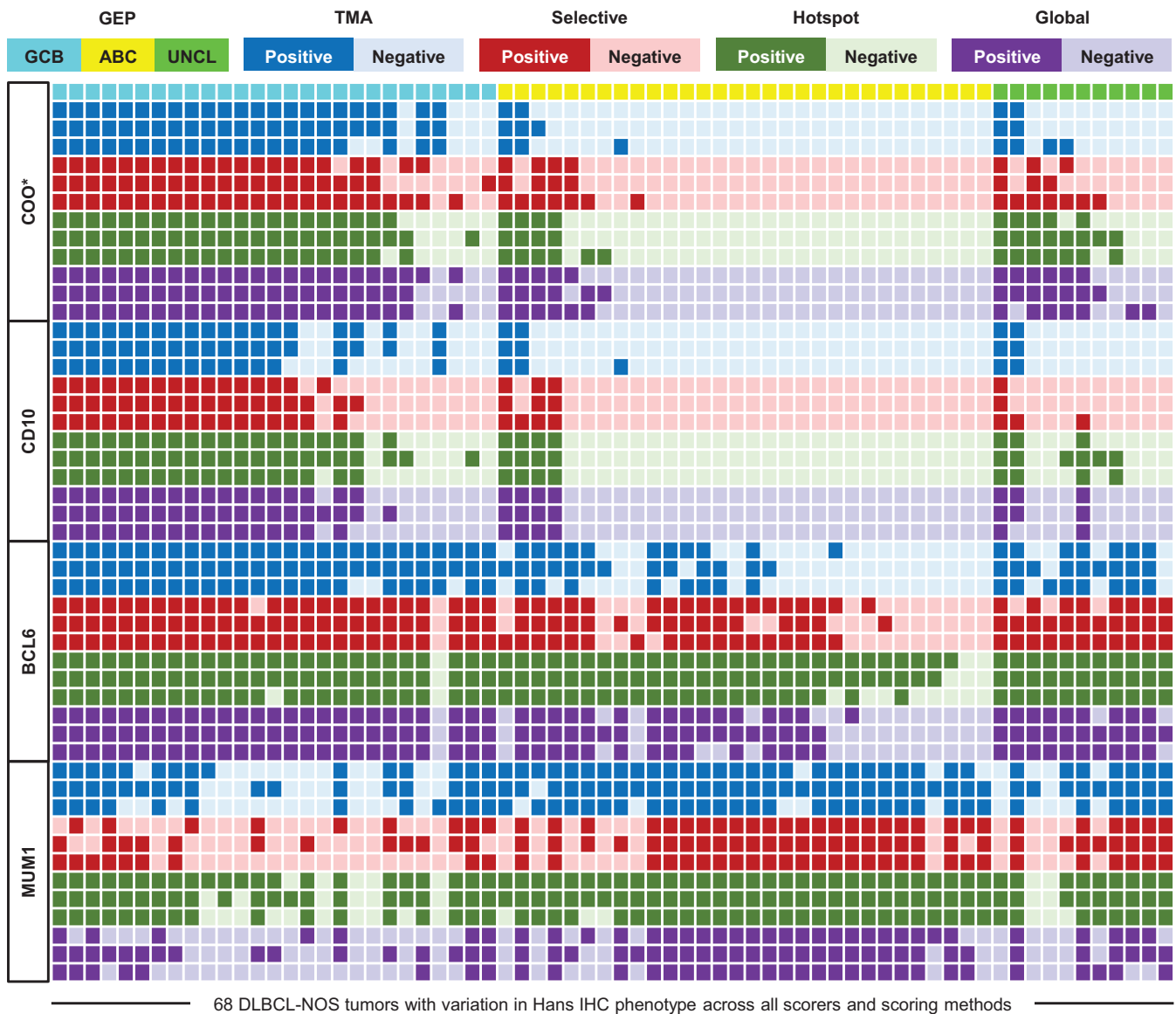


Figure 3. Slightly more than half of all diffuse large B-cell lymphoma, not otherwise specified (DLBCL-NOS) tumors evaluated in this study (68 of 132; 52%) showed inconsistent Hans immunohistochemistry marker phenotypes because of the influence of scoring method and/or interscorer discordance. Individual tumors are represented by columns. Data are derived from 3 independent scorers, resulting in triplicate entries across all scoring methods and markers. *For non-gene expression profiling (GEP) cell of origin (COO), dark colors/positive = germinal center B cell (GCB), light colors/negative = non-GCB. Abbreviations: ABC, activated B cell; IHC, immunohistochemistry; UNCL, unclassified.

across all markers) compared with selective scoring (163 of 1188; 13.7%) and global scoring (176 of 1188; 14.8%) (pairwise Fisher exact tests, $P < .001$). The distributions of raw IHC marker scores and positive/negative calls were highly similar for selective scoring and global scoring, and both were comparable to matched TMA scoring based on these parameters.

Sixty-eight of the 132 DLBCL-NOS tumors (52%) exhibited variation (ie, a change in the positive versus negative call for at least one marker) in Hans IHC phenotype as a function of scoring method and/or interscorer discordance (Figure 3). Specifically, 18 of 132 tumors (14%) showed variation in CD10, 36 of 132 (27%) in BCL6, and 43 of 132 (33%) in MUM1. These changes impacted the final Hans COO assignment in 27 of 132 tumors (20%). The 64 remaining tumors (48%) exhibited stable Hans COO and

IHC phenotypes across all scorers and scoring methods (Supplemental Figure 3). By gold standard GEP, a numerically higher proportion of unclassified tumors exhibited immunophenotypic variation (11 of 17; 65%) than GCB (27 of 64; 42%) or ABC (30 of 51; 59%), but this was not statistically significant (χ^2 test, $P = .11$). When restricting the analysis to WTS-based scoring methods only (ie, excluding TMA assessment), 59 of 132 tumors (45%) exhibited variation in Hans IHC marker phenotype, including 13 of 132 (10%) with variation in CD10, 24 of 132 (18%) with variation in BCL6, and 38 of 132 (29%) with variation in MUM1. Of the 132 DLBCL-NOS tumors evaluated, 20 (15%) were associated with inconsistent Hans COO assignments as a consequence of different WTS-based scoring methods.

Despite the widespread immunophenotypic instability described above, the overall proportion of tumors classified

Hans Cell-of-Origin (COO) Predictions Across 132 Diffuse Large B-Cell Lymphomas, Not Otherwise Specified, Along With Raw Accuracy and Reproducibility Metrics ^a				
Scoring Method	COO, No. (%)		Accuracy, % ^b	Interscorer Agreement, κ ^c
	GCB	Non-GCB		
TMA	197 (50)	199 (50)	87.1	0.939
Selective	188 (47)	208 (53)	83.8	0.858
Hotspot	201 (51)	195 (49)	82.1	0.919
Global	204 (52)	192 (48)	82.3	0.908

Abbreviations: GCB, germinal center B cell; TMA, tissue microarray.

^a Excisions scored by 3 independent pathologists using 3 different whole tissue section–based scoring methods and a matched TMA.

^b Average across all 3 scorers, compared with gold standard gene expression profiling (Lymph2Cx).

^c Fleiss κ score encompassing all 3 scorers.

as GCB versus non-GCB by Hans algorithm did not differ significantly based on scoring method (Table). However, in head-to-head matchups, WTS-based scoring methods produced differing COO assignments from one another in nearly 6% of tumors on average, as well as differing from the corresponding TMA-derived COO assignments in roughly 8% to 9% of tumors on average (Supplemental Table 2). COO predictions from all scoring methods showed strong accuracy (>80%) with respect to gold standard GEP (Table), with TMA-derived COO predictions being slightly more accurate across all observers and the WTS-based scoring methods showing no significant differences in accuracy among themselves (Figure 4, A). Conventional test characteristics (sensitivity, specificity, PPV, NPV) indicated that the contribution of each marker to the final Hans COO prediction is also influenced by scoring method (Supplemental Figure 4). This was particularly true for hotspot scoring and the nuclear markers BCL6 and MUM1, for which these metrics became notably unbalanced. For example, whereas the NPV and sensitivity of

MUM1 for detecting non-GCB COO (GEP-defined) was significantly higher for hotspot scoring than for selective or global scoring, the corresponding specificity of MUM1 was significantly lower with hotspot scoring than for the other WTS-based scoring methods, with PPV also trending strongly downwards (Supplemental Figure 4). Fleiss κ , a measure of overall interscorer agreement, exceeded 0.9 for hotspot and global scoring, as well as for TMA scoring, indicating substantial concordance among pathologists (Table). Selective scoring yielded significantly lower average pairwise interscorer agreement (Cohen κ) for Hans COO assignment than for TMA-based scoring, which it was designed to mimic (Figure 4, B). Hotspot and global scoring produced similar pairwise Cohen κ scores for Hans COO assignment and were both noninferior to TMA scoring in this regard. In terms of individual IHC markers, average pairwise interscorer agreement was greatest for MUM1 when using hotspot scoring, whereas CD10 and BCL6 trended slightly better with global scoring (Supplemental Figure 5).

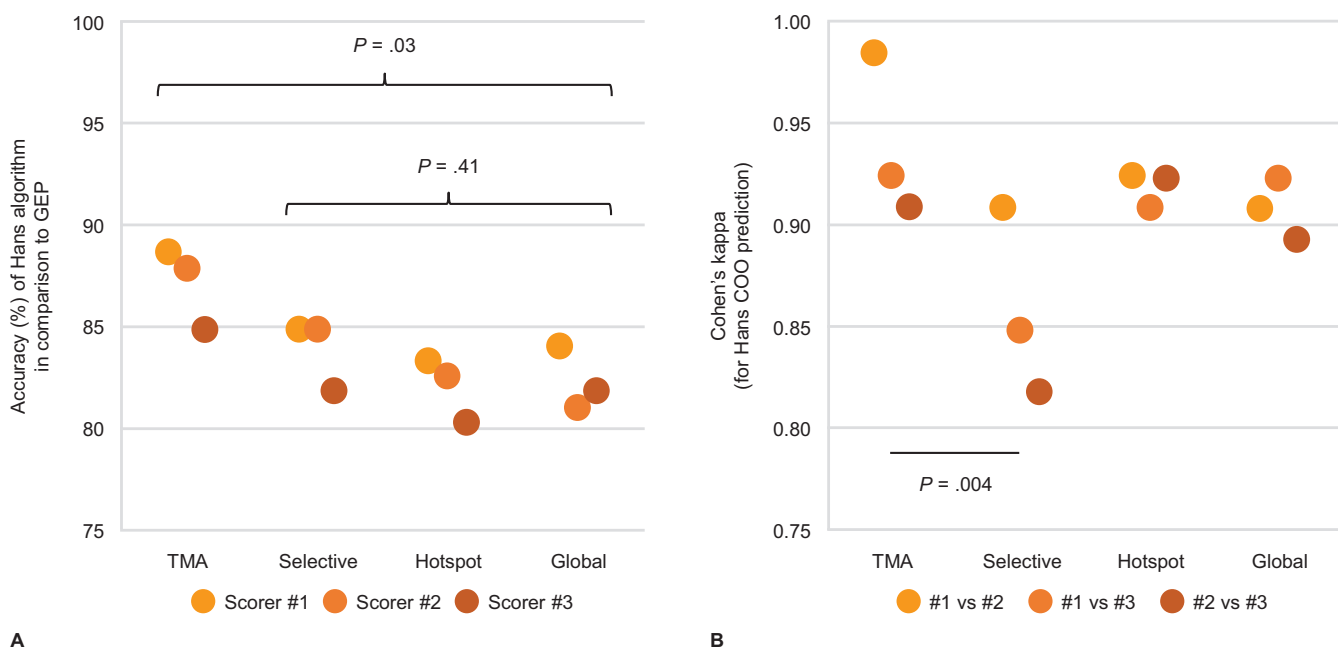


Figure 4. A, The accuracy of Hans algorithm cell-of-origin (COO) predictions was established in comparison with gold standard gene expression profiling (GEP). The statistical tests shown here represent 1-way/single-factor analysis of variance encompassing the indicated groups. B, Interscorer agreement for Hans COO assignment was assessed using pairwise Cohen κ scores. The paired 2-sample t test between tissue microarray (TMA) and selective scoring is significant; however, none of the remaining pairs reach statistical significance.

DISCUSSION

The 3-marker Hans algorithm remains the most widely used strategy for assigning COO in DLBCL-NOS, effectively distinguishing GCB from non-GCB tumors in routine clinical practice. Roughly half of the 132 DLBCL-NOS tumors evaluated in this study were susceptible to variation in Hans IHC phenotype based on the implemented scoring method. These changes often translated to differing COO predictions, which occurred in 15% (20 of 132) of all tumors based on our 3 WTS-based scoring methods and 20% (27 of 132) of all tumors when TMA scoring was included (Figure 3). Despite these considerable differences, none of the WTS-based scoring methods were statistically inferior to the others in terms of accuracy against gold standard GEP-based COO (Figure 4, A). Nevertheless, each method was associated with its own set of advantages and disadvantages. For example, hotspot scoring achieved a significant reduction in the number of borderline scores (ie, those ranging from 20% to 40%), which may foster the perception of being easier to implement because of the lower frequency of challenging calls. On the other hand, hotspot scoring produced dramatic imbalances between sensitivity and specificity, as well as PPV and NPV, for both BCL6 and MUM1 (Supplemental Figure 4). Performance almost certainly would have suffered were it not for the built-in redundancy of this multi-marker panel and the intrinsically dichotomous profile of CD10 expression in DLBCL (Figure 2), which allowed this leading marker to lessen the impact of hotspot scoring on the overall algorithm.

Entering into this study, we hypothesized that selective scoring, which intentionally mimics the effect of TMA-based assessment by restricting the scoring of all IHC markers to a chosen area of high tumor cell burden and high quality based on H&E assessment (thereby mirroring the earliest phases of TMA construction) (Figure 1), would perhaps outperform global scoring because of the smaller area requiring evaluation and the avoidance of common histology and specimen artifacts that may introduce noise and uncertainty into semiquantitative measurements (Supplemental Figure 1). Not only was this not the case, but unlike hotspot and global scoring, selective scoring was associated with statistically inferior interscorer concordance compared with TMA scoring. In hindsight, this outcome might have been predicted, as most scorers will quickly agree as to which area(s) of tumor show the highest degree of staining (hotspot scoring) and both TMA and global scoring require scorers to literally assess the exact same tissue, albeit in dramatically differing quantities. Meanwhile, with selective scoring, each independent scorer is free to identify their own best area of tumor, and although several different areas may appear equally optimal by H&E microscopy, there is no guarantee that these different areas will be immunophenotypically equivalent.

Given the arguments above, our data favor the use of global scoring over the other WTS-based scoring options because of its noninferior accuracy, solid interscorer concordance, and nonnegative influence on the test characteristics of individual Hans markers. Global scoring also benefits from being naturally intuitive and is currently in widespread use, with most practicing pathologists already considering this to be standard of care. Future lymphoma research will likely continue to rely heavily on TMA technology because of the improved efficiency and cost reduction, both of which are critical in an era of big data and larger cohorts.

However, the application of TMA-derived knowledge to clinical care should always be met with a healthy degree of caution. The GEP-correlated, TMA-based algorithms for DLBCL COO, which were all products of TMA-based research, are a worthy example in this regard. This study did not assess the impact of scoring method on nonexcisional tissues such as core needle biopsies and cell blocks from fine-needle specimens. Additional research is needed to confirm that our findings translate to these comparatively limited samples, which have become increasingly common in daily practice. This study does not address several other important aspects of IHC interpretation, such as the impact of varying laboratory techniques and reagents, as well as the impact of scoring method on nonexpert scorers. Such issues were beyond the scope of the present research. We wish to note that our study design may be readily applied to other IHC markers and TMA-based algorithms, an analogous example of which would be the 4-marker IHC algorithm for recapitulating GEP-defined subtypes of peripheral T-cell lymphoma, not otherwise specified.³² Finally, although the longstanding emphasis on COO in DLBCL may eventually give way to newer genomics-based classifications, the latter may not be universally accessible, and alternative IHC-based solutions may once again come into play.

We would like to acknowledge the staff members at BC Cancer and its Centre for Lymphoid Cancer for the construction of the tissue microarray and the production of clinical-grade immunohistochemical stains. We are also thankful to Elizabeth Courville, MD, for constructive feedback during manuscript preparation.

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. Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. *Pathology*. 2018;50(1):74–87.
4. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503–511.
5. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(25):1937–1947.
6. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med*. 2008;359(22):2313–2323.
7. Barralough A, Hawkes E, Sehn LH, Smith SM. Diffuse large B-cell lymphoma. *Hematol Oncol*. 2024;42(6):e3202.
8. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275–282.
9. Choi WW, Weisenburger DD, Greiner TC, et al. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res*. 2009;15(17):5494–5502.
10. Meyer PN, Fu K, Greiner TC, et al. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J Clin Oncol*. 2011;29(2):200–207.
11. Visco C, Li Y, Xu-Monette ZY, et al. Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study. *Leukemia*. 2012;26(9):2103–2113.
12. Hwang HS, Park CS, Yoon DH, Suh C, Huh J. High concordance of gene expression profiling-correlated immunohistochemistry algorithms in diffuse large B-cell lymphoma, not otherwise specified. *Am J Surg Pathol*. 2014;38(8):1046–1057.
13. Culpin RE, Sieniawski M, Angus B, et al. Prognostic significance of immunohistochemistry-based markers and algorithms in immunochemotherapy-treated diffuse large B cell lymphoma patients. *Histopathology*. 2013;63(6):788–801.
14. Battle-Lopez A, Gonzalez de Villambrosia S, Francisco M, et al. Stratifying diffuse large B-cell lymphoma patients treated with chemoimmunotherapy: GCB/non-GCB by immunohistochemistry is still a robust and feasible marker. *Oncotarget*. 2016;7(14):18036–18049.

15. Fu K, Weisenburger DD, Choi WW, et al. Addition of rituximab to standard chemotherapy improves the survival of both the germinal center B-cell-like and non-germinal center B-cell-like subtypes of diffuse large B-cell lymphoma. *J Clin Oncol*. 2008;26(28):4587–4594.
16. Abdulla M, Hollander P, Pandzic T, et al. Cell-of-origin determined by both gene expression profiling and immunohistochemistry is the strongest predictor of survival in patients with diffuse large B-cell lymphoma. *Am J Hematol*. 2020;95(1):57–67.
17. Scott DW. Cell-of-origin in diffuse large B-cell lymphoma: are the assays ready for the clinic? *Am Soc Clin Oncol Educ Book*. 2015:e458–e466.
18. Coutinho R, Clear AJ, Owen A, et al. Poor concordance among nine immunohistochemistry classifiers of cell-of-origin for diffuse large B-cell lymphoma: implications for therapeutic strategies. *Clin Cancer Res*. 2013;19(24):6686–6695.
19. Gutierrez-Garcia G, Cardesa-Salzmann T, Climent F, et al. Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Blood*. 2011;117(18):4836–4843.
20. Read JA, Koff JL, Nastoupil LJ, Williams JN, Cohen JB, Flowers CR. Evaluating cell-of-origin subtype methods for predicting diffuse large B-cell lymphoma survival: a meta-analysis of gene expression profiling and immunohistochemistry algorithms. *Clin Lymphoma Myeloma Leuk*. 2014;14(6):460–467 e2.
21. Ott G, Ziepert M, Klapper W, et al. Immunoblastic morphology but not the immunohistochemical GCB/nonGCB classifier predicts outcome in diffuse large B-cell lymphoma in the RICOVER-60 trial of the DSHNHL. *Blood*. 2010;116(23):4916–4925.
22. Ott G. Aggressive B-cell lymphomas in the update of the 4th edition of the World Health Organization classification of haematopoietic and lymphatic tissues: refinements of the classification, new entities and genetic findings. *Br J Haematol*. 2017;178(6):871–887.
23. de Jong D, Rosenwald A, Chhanabhai M, et al. Immunohistochemical prognostic markers in diffuse large B-cell lymphoma: validation of tissue microarray as a prerequisite for broad clinical applications—a study from the Lunenburg Lymphoma Biomarker Consortium. *J Clin Oncol*. 2007;25(7):805–812.
24. Zu Y, Steinberg SM, Campo E, et al. Validation of tissue microarray immunohistochemistry staining and interpretation in diffuse large B-cell lymphoma. *Leuk Lymphoma*. 2005;46(5):693–701.
25. Tzankov A, Zlobec I, Went P, Robl H, Hoeller S, Dirnhofer S. Prognostic immunophenotypic biomarker studies in diffuse large B cell lymphoma with special emphasis on rational determination of cut-off scores. *Leuk Lymphoma*. 2010;51(2):199–212.
26. Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood*. 2014;123(8):1214–1217.
27. Scott DW, Mottok A, Ennishi D, et al. Prognostic significance of diffuse large B-cell lymphoma cell of origin determined by digital gene expression in formalin-fixed paraffin-embedded tissue biopsies. *J Clin Oncol*. 2015;33(26):2848–2856.
28. Ennishi D, Mottok A, Ben-Neriah S, et al. Genetic profiling of MYC and BCL2 in diffuse large B-cell lymphoma determines cell-of-origin-specific clinical impact. *Blood*. 2017;129(20):2760–2770.
29. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
30. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378–382.
31. Nowakowski GS, Chiappella A, Witzig TE, et al. Variable global distribution of cell-of-origin from the ROBUST phase III study in diffuse large B-cell lymphoma. *Haematologica*. 2020;105(2):e72–e75.
32. Amador C, Greiner TC, Heavican TB, et al. Reproducing the molecular subclassification of peripheral T-cell lymphoma-NOS by immunohistochemistry. *Blood*. 2019;134(24):2159–2170.