

Scoring of MYC protein expression in diffuse large B-cell lymphomas: concordance rate among hematopathologists

Amer Z Mahmoud¹, Tracy I George¹, David R Czuchlewski¹, Qian-Yun Zhang¹, Carla S Wilson¹, Cordelia E Sever², Alexei G Bakhirev², Dahua Zhang², Nichole L Steidler¹, Kaaren K Reichard³, Huining Kang⁴, Kathryn Foucar¹ and Mohammad A Vasef¹

¹Department of Pathology, University of New Mexico, Albuquerque, NM, USA; ²Presbyterian Hospital, Albuquerque, NM, USA; ³Mayo Clinic, Rochester, MN, USA and ⁴Department of Internal Medicine, University of New Mexico, Albuquerque, NM, USA

Recent studies have shown that immunohistochemical evaluation of MYC protein expression in diffuse large B-cell lymphoma is a useful prognostic tool with high concordance rate among pathologists. Concordance in these studies was assessed among few pathologists from one institution by scoring tissue microarrays. In daily practice, MYC evaluation is performed on entire tumor sections by a diverse group of pathologists. In our study, nine hematopathologists from two institutions scored whole-tissue sections of two sets of cases. The training set included 13 cases of diffuse large B-cell lymphoma and 4 cases of Burkitt lymphoma. The validation set included 18 cases of diffuse large B-cell lymphoma and 1 case of Burkitt lymphoma. MYC positivity was defined as $\geq 40\%$ of tumor cells demonstrating nuclear staining similar to prior studies. The mean score for each case was used to determine MYC status with discrepant cases defined as having any score causing a different MYC status designation. Discrepant cases from the training set were characterized by staining heterogeneity, extensive necrosis or crush artifact and had mean scores within 15 percentage points of 40%. Cases from the validation set that demonstrated any of these features were scored twice on two different days. Overall concordance was moderate (Kappa score: 0.68, P -value < 0.001) with no significant change between the two sets (Kappa scores: 0.69 vs 0.67). Thirty-nine percent of cases were discrepant. The findings indicate that a significant number of diffuse large B-cell lymphomas are inherently difficult to score due to staining heterogeneity. The effect of heterogeneity can be under-represented when concordance is measured among few pathologists scoring tissue microarrays. Careful scoring strategy in our study failed to improve concordance. In the absence of specific instructions on how to deal with heterogeneity, caution is advised when evaluating MYC expression in diffuse large B-cell lymphoma.

Modern Pathology (2015) 28, 545–551; doi:10.1038/modpathol.2014.140; published online 28 November 2014

The MYC oncogene is involved in many types of human cancer. The discovery of a consistent balanced chromosomal translocation involving the MYC gene in Burkitt lymphoma was the first evidence to characterize MYC as a human oncogene.¹ Subsequently, MYC gene alterations have been discovered in B-cell neoplasms other than Burkitt lymphoma.² Among those neoplasms, diffuse large B-cell lymphoma is the most widely studied. The presence of MYC rearrangements in

patients with diffuse large B-cell lymphoma treated with Rituximab, Cyclophosphamide, Doxorubicin, Vincristine and Prednisone (R-CHOP) has been shown to be associated with poor prognosis.^{3,4} In particular, the so-called ‘double-hit’ lymphomas that are characterized by MYC rearrangement and a concurrent rearrangement of other B-cell lymphoma-associated genes such as BCL2 or BCL6 are associated with poor response to therapy, aggressive clinical course and dismal prognosis.^{5–7} These lymphomas are classified as ‘B-cell lymphoma, unclassifiable with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma’ in the current 2008 WHO classification of hematopoietic neoplasms.⁸

Evaluation for MYC alterations in diffuse large B-cell lymphoma is typically performed by fluorescence *in situ* hybridization (FISH). FISH is capable

Correspondence: Dr MA Vasef, MD, Department of Pathology, TriCore Reference Laboratories, University of New Mexico, 1001 Woodward Pl, NE, Albuquerque, NM 87102, USA.
E-mail: mvasef@salud.unm.edu

Received 15 May 2014; revised 10 August 2014; accepted 11 August 2014; published online 28 November 2014

of detecting *MYC* gene alterations that result from translocation or amplification of the gene. These genetic alterations result in *MYC* protein overexpression, which is ultimately responsible for the oncogenic effect.^{1,2} *MYC* protein overexpression has also been found to occur as a consequence of other genomic events not detected by FISH.⁹ Thus FISH might miss a subset of diffuse large B-cell lymphoma cases that demonstrate *MYC* protein overexpression.

MYC protein expression has been evaluated in formalin-fixed paraffin-embedded tissue by immunohistochemistry in multiple studies.^{9–13} These studies have shown that cases of diffuse large B-cell lymphoma with concurrent overexpression of *MYC* and *BCL2* proteins have a dismal prognosis similar to those with double-hit lymphomas. The rate of so-called double-hit lymphoma-like cases as determined by immunohistochemistry is larger than the one detected by FISH.^{9,10} Evaluation of *MYC* expression in these studies is performed by estimating the percentage of *MYC* protein expressing tumor cells. In the majority of studies, a cutoff of $\geq 40\%$ is used to define *MYC* protein overexpression.

Although high rates of inter-observer concordance have been reported in prior studies, scoring is performed on scant material in tissue microarrays.^{9–13} Tissue microarrays sample only a small portion of the tumor and therefore may underrepresent the heterogeneity of staining among tumor cells encountered in daily practice. The heterogeneity of *MYC* protein staining is not adequately addressed in these studies. Additionally, the concordance rate is assessed among only two^{10,11,13} or three⁹ pathologists that might underestimate inter-observer variability among practicing pathologists. We hypothesized that these concordance rates might not be reproducible in daily practice. In this study, we examined the concordance rate in *MYC* scoring among nine hematopathologists from two institutions when evaluating entire tumor sections and investigated whether scoring a tissue microarray-sized field instead of entire section will improve the concordance. We also identified some features that characterize discrepant cases and evaluated whether careful scoring of these cases can improve concordance. The impact of using an image analysis program was also assessed.

Materials and methods

Case Selection

Following institutional review board approval, two sets of high-grade B-cell lymphomas were selected. The training set contained 13 cases of diffuse large B-cell lymphoma and 4 cases of Burkitt lymphoma diagnosed between 2003 and 2011 at the University of New Mexico and Presbyterian Hospital in Albuquerque, NM, USA. The validation set

included 18 cases of diffuse large B-cell lymphoma and 1 case of Burkitt lymphoma diagnosed between 2013 and 2014 in the Department of Pathology at the University of New Mexico. The training set is used to identify potential factors leading to discrepant scoring while the validation set is used to evaluate whether careful scoring of cases characterized by these factors can improve concordance rate among hematopathologists. The cases in both sets were selected to represent various sites, including nodal (neck, mediastinal, axillary, pelvic, para-aortic and inguinal) and extra-nodal (tonsil, brain, thyroid, stomach, small bowel, spleen, uterine cervix, bone marrow and spine) and different specimen types, including resection (thyroid and spleen), excisional biopsy, needle core biopsy and bone marrow biopsy. Cases of Burkitt lymphoma were expected to have a very high *MYC* expression and served as a quality control.

Immunohistochemistry

Paraffin immunohistochemistry was performed using a monoclonal *MYC* antibody (clone Y69; Epitomics, Burlingame, CA, USA) at 1:50 dilution and with 24-min incubation. Briefly, four-micron thick recuts of representative paraffin-embedded tissue blocks were baked for at least 30 min in an oven at 60 °C. Deparaffinization, antigen retrieval (CC1 Ventana, pH 8), blockage of endogenous peroxidase activity, antibody dispense and incubation steps were all performed on automated Ultra Benchmark Instrument (Ventana, Tuscan, AZ, USA). Next, the slides were removed from the Ultra instrument after completion of the run, dipped 10–15 times in Dawn water to remove the oil, rinsed in tap water, dehydrated using a graded series of reagent alcohols, dipped in xylene and coverslipped for microscopic review.

Whole-Slide Digitalization

Cases were de-identified, and slides were scanned using the Aperio whole-slide digitalizer (scanscope CS system, Leica Biosystems, Buffalo Grove, IL, USA) at $\times 20$ magnification. A password-protected account was created, and the pathologists were provided access to this account. Slides were reviewed through the Aperio ImageScope software (v11.2.0.780). One-millimeter fields were marked on the digital slides using the circle annotation tool, and the diameter of the field was confirmed using the measure tool. The fields were selected based on having the highest amount of tumor and the least amount of necrosis and/or crush artifact. Image analysis was performed on the discrepant cases to generate an automated score using the Aperio immunohistochemistry nuclear algorithm (v9.1.19.1569). The parameters used in the algorithm were: threshold type, edge threshold method;

segmentation type, cytoplasmic rejection; lower threshold, 0; upper threshold, 230; and nuclear threshold, 220. The most critical parameter was the nuclear threshold, and it was selected by comparing various thresholds used to score a specific field to consensus score among pathologists.

Scoring

Six hematopathologists from the University of New Mexico and three hematopathologists from the Presbyterian Hospital scored each set of cases at two different time intervals. Scoring was performed on the digital slides ensuring that the exact same section was scored by all the pathologists. Prior to scoring of both sets, relevant publications were discussed in a journal club, and pathologists were asked to assign their scores based on their understanding of the literature.^{9,10} For the training set, pathologists were instructed to avoid areas of necrosis, assign each case a specific score (ie, no range was permitted) and provide their comments on cases they perceived to be difficult to score. Pathologists estimated the percentage of positive tumor cells and reported their scores in increments of 5%.

For the validation set, pathologists were instructed to avoid areas of necrosis and to assign each case a specific score. Additionally, if pathologists identified any of the factors that could explain the reasons for their discrepant score in training set or if their score was within 15 percentage points of the 40%, they were instructed to score that case twice, on two different days, and provide a mean of the two scores.

Scoring of tissue microarray-sized fields in discrepant cases was performed by eight hematopathologists. These fields were clearly marked on the digital slides ensuring that the exact same field was scored by all the pathologists.

Definition of MYC Positivity and Discrepancy

Cases with a mean score of $\geq 40\%$ MYC nuclear expression in tumor cells were defined as being MYC positive. A discrepant case was defined as any case having ≥ 1 discrepant scores. A discrepant score was defined as any score that resulted in a different MYC status designation (ie, from negative to positive and *vice versa*) than that of the mean score of the case.

Statistical Analysis

Each case was given a total of nine individual scores. The mean, s.d. and range of these scores were calculated for each case. Concordance rate was evaluated by Kappa score. A pairwise comparison between every pair of pathologists was first per-

formed to calculate the Cohen's Kappa score and the corresponding *P*-value.¹⁴ The Fleiss' Kappa score was then calculated as an index of concordance among all pathologists.¹⁵ The analyses were performed using the R software (<http://www.R-project.org/>) with package irr (R package version 0.84, <http://CRAN.R-project.org/package=irr>).

Results

Training Set

There was moderate concordance among hematopathologists for scoring MYC expression in diffuse large B-cell lymphoma cases in the training set with a Fleiss Kappa score of 0.69 ($P < 0.001$) and a Fleiss Kappa of 0.71 ($P < 0.001$) for all cases (Table 1). Seven out of the 17 (41%) were discrepant, including three cases of diffuse large B-cell lymphoma and one case of Burkitt lymphoma. Among the discrepant diffuse large B-cell lymphoma cases, the number of discrepant scores were 3 out of 9 (33%) in one case, 2 out of 9 (22%) in two cases and 1 out of 9 (11%) in three cases. The discrepant Burkitt lymphoma case had 4 out of 9 (44%) discrepant scores (Table 2). Of note, the reviewing hematopathologists commented on all the discrepant cases except for case 8 (Table 2). The factors that contributed to discrepant results were identified as: geographic variation of MYC staining, variation in intensity of MYC stain, necrosis and crush artifact (Figures 1–4). Only one non-discrepant case was commented on by pathologists (Table 2). All the discrepant cases had a mean score within 15 percentage points of the 40% threshold.

Validation Set

The concordance rate for scoring the validation set was similar to that of the training set with a Fleiss Kappa score of 0.69 ($P < 0.001$) for all cases and a score of 0.67 ($P < 0.0001$) for diffuse large B-cell lymphoma cases (Table 1). Six out of the 19 cases (32%) were discrepant, all of which were diffuse large B-cell lymphoma. Among the discrepant cases,

Table 1 Fleiss's Kappa score calculated for different combinations of cases

	Kappa score ^a
All cases ($n = 36$)	0.70
All DLBCL cases ($n = 31$)	0.68
All cases from training set ($n = 17$)	0.71
All cases from validation set ($n = 19$)	0.69
DLBCL cases from training set ($n = 13$)	0.69
DLBCL cases from validation set ($n = 18$)	0.67
All discrepant cases—entire sections ($n = 13$)	0.17
All discrepant cases—1-mm fields ($n = 13$)	0.66

^aAll statistical analyses showed *P*-values < 0.001 .

Table 2 Results of training set scoring

Case	Diagnosis	Mean (s.d.)	Range	DS	Comments	MYC gene status
1	DLBCL	4.6 (2.8)	0–10	0	None	One extra copy ^{a, b}
2	DLBCL	8.3 (5.6)	0–20	0	None	Not tested
3	DLBCL	13.7 (4.4)	8–20	0	None	Not tested
4	DLBCL	17.8 (10.3)	5–35	0	Intensity variation	Normal pattern ^b
5	DLBCL	27.2 (9.4)	20–50	1	Geographic variation	Not tested
6	DLBCL	27.8 (7.1)	20–40	1	Necrosis	Not tested
7	DLBCL	29.4 (15.3)	5–50	2	Crush artifact	Normal pattern ^b
8	DLBCL	35.6 (6.8)	30–50	3	None	Normal pattern ^a
9	BL	42.8 (17)	20–65	4	Variation ^c , necrosis	Not tested
10	DLBCL	43.3 (10)	30–60	2	Intensity variation	Normal pattern ^a
11	DLBCL	47.8 (10.6)	30–60	1	Variation ^c , necrosis	Not tested
12	DLBCL	49.4 (12.9)	40–70	0	None	Not tested
13	DLBCL	57.2 (6.7)	50–70	0	None	Normal pattern ^b
14	BL	82.8 (12.8)	60–90	0	None	Rearranged ^b
15	BL	88.3 (8.7)	70–95	0	None	Rearranged ^b
16	DLBCL	88.9 (4.9)	85–95	0	None	Not tested
17	BL	91.7 (2.5)	90–95	0	None	Rearranged ^b

Abbreviations: BL, Burkitt lymphoma; DLBCL, diffuse large B-cell lymphoma; DS, discrepant scores.

Mean, s.d. and range of scores are provided for each case. Cases are arranged in ascending order according to their mean score. Horizontal black line indicates the 40% threshold. Discrepant cases are highlighted.

^aMYC FISH break apart probe.

^bMYC/IGH FISH dual fusion probe.

^cIndicates both geographic and intensity variation of MYC staining.

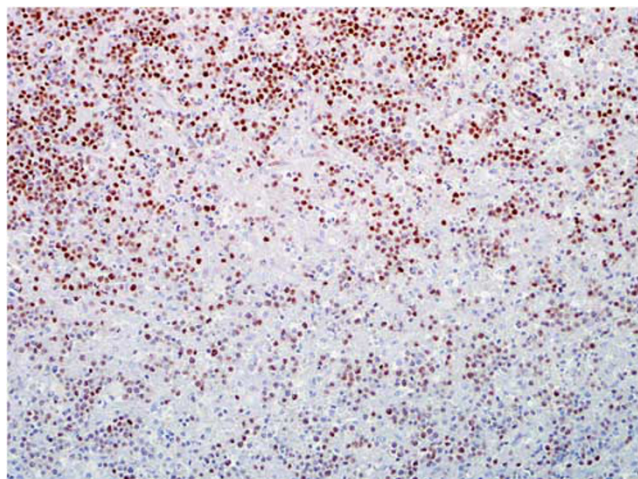


Figure 1 Geographic variation of MYC staining. Variability in MYC expression among different areas of tumor; immunoperoxidase stain, original magnification, $\times 200$.

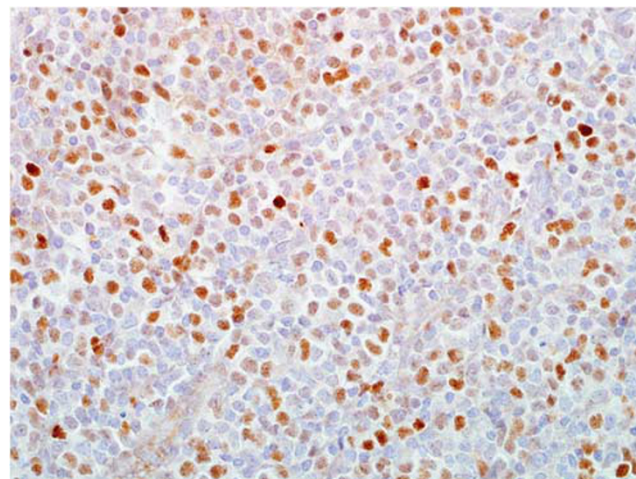


Figure 2 Intensity variation of staining. Variability in intensity of MYC expression among tumor cells; immunoperoxidase stain; original magnification $\times 200$.

the number of discrepant scores were 4 out of 9 (44%) in two cases, 3 out of 9 (33%) in two cases and 1 out of 9 (11%) in two cases (Table 3).

Tissue Microarray-Sized Fields on Discrepant Cases

The concordance rate for scoring preselected tissue microarray-sized fields in the discrepant cases was significantly much higher than in scoring entire sections with Fleiss Kappa scores of 0.66 and 0.17, respectively (P -values < 0.001). After preselecting

for tissue microarray-sized fields, the total number of discrepant cases decreased from 13 to 7 and the total number of discrepant scores decreased from 30 to 9 (Table 4).

Scoring of Discrepant Cases Using Image Analysis Program

Nine out of the 13 (69%) and 12 out of the 13 (92%) discrepant cases had an automated score with MYC designation concordant to that of manual scoring

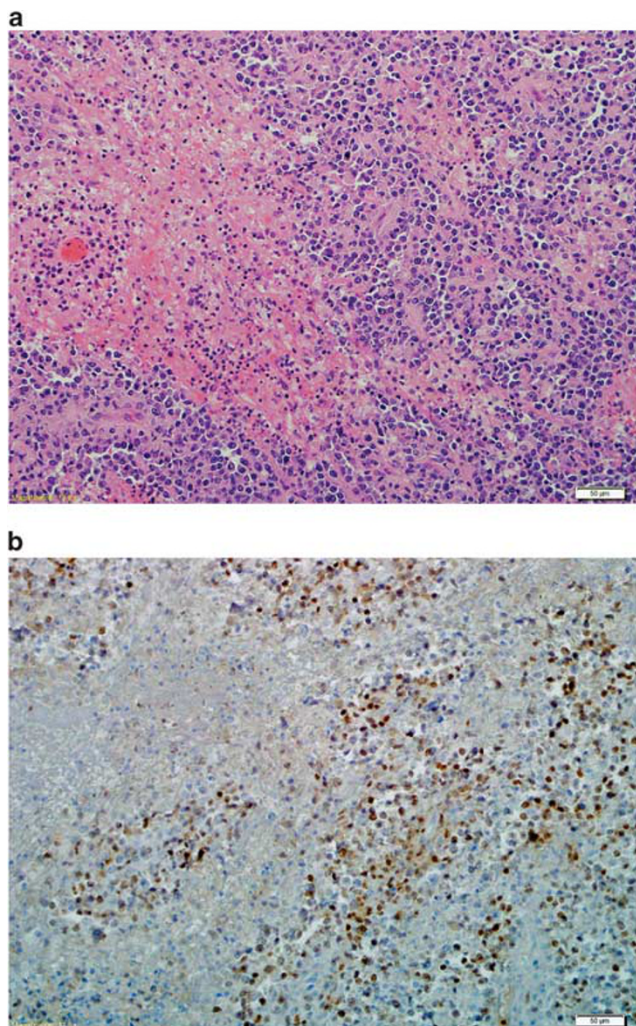


Figure 3 (a, b) Inconsistency in MYC staining in necrotic foci. Geographic necrosis evident on this hematoxylin and eosin-stained histological section of a diffuse large B-cell lymphoma case; (a) hematoxylin and eosin stain; (b) immunoperoxidase stain showing artifactual lack of MYC expression among necrotic portion of the tumor, original magnification $\times 200$.

when entire sections and 1-mm fields were scored, respectively (see Table 5).

Discussion

Evaluation of MYC protein overexpression by immunohistochemistry is becoming an important tool in prognostic stratification of patients with diffuse large B-cell lymphoma. Expression of MYC protein by $\geq 40\%$ of the neoplastic cells has been applied as a cutoff in most studies. These studies report high concordance in scoring of MYC expression in diffuse large B-cell lymphoma among pathologists. However, concordance has been assessed among few pathologists who performed scoring on tissue microarrays. An accurate scoring of MYC expression on tissue microarrays can be

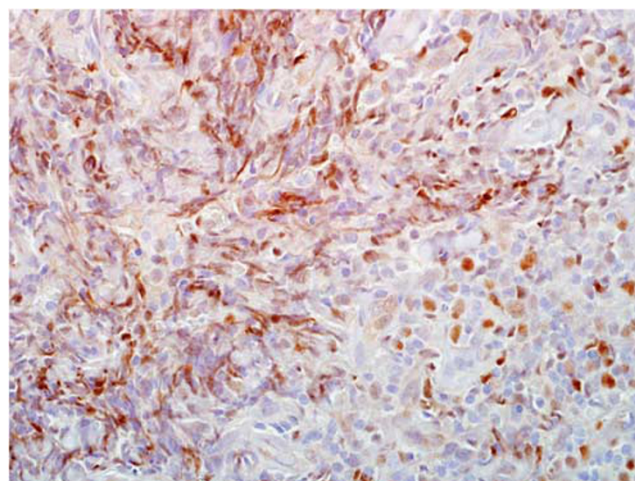


Figure 4 Crush artifact. The majority of tumor in this section is crushed precluding adequate evaluation of MYC expression; immunoperoxidase stain, original magnification $\times 200$.

Table 3 Results of validation set scoring

Case	Diagnosis	Mean (s.d.)	Range	Discrepant scores	MYC gene status
18	DLBCL	4.8 (2.5)	1–10	0	Normal ^a
19	DLBCL	9.4 (8.8)	0–25	0	Normal ^a
20	DLBCL	13.1 (8.4)	5–28	0	Not tested
21	DLBCL	18 (9.1)	5–30	0	Normal ^a
22	DLBCL	20.6 (8.7)	10–28	0	Not tested
23	DLBCL	21.3 (7.6)	10–35	0	Normal ^a
24	DLBCL	26.3 (11.1)	10–50	1	Normal ^a
25	DLBCL	30.2 (13.1)	10–45	3	Normal ^a
26	DLBCL	35.2 (18.4)	5–67	4	Normal ^a
27	DLBCL	40.2 (10.8)	25–58	4	Normal ^a
28	DLBCL	42.4 (11.3)	30–60	3	Normal ^a
29	DLBCL	45.8 (14.1)	20–60	1	Normal ^a
30	DLBCL	48.6 (9)	40–70	0	Normal ^a
31	DLBCL	54.8 (10.1)	40–70	0	One extra copy ^a
32	DLBCL	61 (10.5)	40–70	0	Not tested
33	DLBCL	66.3 (13.6)	53–90	0	Normal ^a
34	DLBCL	91.7 (5)	90–95	0	Not tested
35	DLBCL	96.6 (5)	85–100	0	Rearranged ^b
36	BL	97.1 (2.5)	95–100	0	Rearranged ^{c,b}

Abbreviations: BL, Burkitt lymphoma; DLBCL, diffuse large B-cell lymphoma.

Mean, s.d. and range of scores are provided for each case. Cases are arranged in ascending order according to their mean score. Horizontal black line indicates the 40% threshold. Discrepant cases are highlighted.

^aMYC break apart probe.

^bt (8;14) (q24.1;q32) as detected by chromosomal analysis.

^cMYC/IGH dual fusion probe.

problematic given the limited tissue present for evaluation, which might not be representative. In this study, we investigated the concordance rate among a larger number of hematopathologists who performed the MYC scoring on the entire biopsy sections of DLBCL cases. The study also identified

Table 4 Comparison of scoring entire sections *versus* 1-mm fields only in discrepant cases

Case	Diagnosis	Mean (s.d.) (ES)	Mean (s.d.) (1 mm)	DS (ES)	DS (1 mm)
5	DLBCL	27.2 (9.4)	24.4 (4.1)	1	0
6	DLBCL	27.8 (7.1)	29.4 (8.2)	1	1
7	DLBCL	29.4 (15.3)	33.8 (5.8)	2	2
8	DLBCL	35.6 (6.8)	48.8 (6.4)	3	1
9	BL	42.8 (17)	60.6 (17.2)	4	0
10	DLBCL	43.3 (10)	50 (11)	2	1
11	DLBCL	47.8 (10.6)	58.1 (5)	1	0
24	DLBCL	26.3 (11.1)	17.5 (7.5)	1	0
25	DLBCL	30.2 (13.1)	46.3 (11.3)	3	1
26	DLBCL	35.2 (18.4)	14.4 (6.2)	4	0
27	DLBCL	40.2 (10.8)	46.9 (8.8)	4	1
28	DLBCL	42.4 (11.3)	51.9 (16.2)	3	2
29	DLBCL	45.8 (14.1)	70 (7.6)	1	0

Abbreviations: BL, Burkitt lymphoma; DLBCL, diffuse large B-cell lymphoma; DS, discrepant scores; ES, entire section.

One circular field with a diameter of 1 mm has been marked on each of the discrepant cases and scored by eight hematopathologists. The mean, s.d. and the number of discrepant scores for each case are provided and compared with those obtained by scoring entire sections.

Table 5 Comparison of manual *versus* automated scoring on entire sections and 1-mm fields of discrepant cases

Case	Diagnosis	Manual score (entire section)	Automated score (entire section)	Manual score (1 mm)	Automated score (1 mm)
5	DLBCL	27.2	35.4	24.4	33.8
6	DLBCL	27.8	20.5	29.4	36.5
7	DLBCL	29.4	40.9	33.8	46.1
8	DLBCL	35.6	33.3	48.8	50.7
9	BL	42.8	21.1	60.6	53.4
10	DLBCL	43.3	45.3	50	48.2
11	DLBCL	47.8	24.3	58.1	58.4
24	DLBCL	26.3	26.4	17.5	27
25	DLBCL	30.2	41.5	46.3	59.5
26	DLBCL	35.2	37	14.4	24.1
27	DLBCL	40.2	28.3	46.9	52.4
28	DLBCL	42.4	35.9	51.9	47.2
29	DLBCL	45.8	54.2	70	70.9

Abbreviations: BL, Burkitt lymphoma; DLBCL, diffuse large B-cell lymphoma.

Manual and automated scores on entire sections and 1-mm fields of discrepant cases are provided. Instances where automatic scoring resulted in a different MYC designation than manual scoring are highlighted.

features of discrepant cases and investigated whether careful scoring of such cases can improve concordance rate.

The overall concordance rate among the nine hematopathologists who participated in our study was lower than the one reported previously in the literature (Table 1). Twelve out of the 31 cases of diffuse large B-cell lymphoma (39%) from both training and validation set showed discrepant results in MYC scoring. Two of the discrepant cases

(cases 26 and 27) were particularly difficult to score, and the pathologists were almost divided on the MYC status in these two cases (Table 3). The most important feature noted in the discrepant cases was variation of MYC staining across the tumor, including variation in distribution of staining, intensity of the staining or both (Figures 1 and 2). Additional features that contributed to discrepant results included necrosis and crush artifact (Figures 3 and 4). Our findings indicate that MYC expression scoring on a representative biopsy section could be a significantly challenging task due to heterogeneity in distribution and intensity of MYC staining that may not be a factor in a tissue microarray. As there are no established criteria in the current literature that addresses the optimal approach to the issue of staining heterogeneity, significant inter-observer discrepancy on MYC scoring is expected when adequate tissue samples are evaluated.

The impact of a careful scoring strategy was evaluated in the validation set. Pathologists were instructed to score certain cases twice on two separate days and provide a mean of the two scores. Cases demonstrating staining heterogeneity, necrosis and crush artifact were all scored twice. Additionally, as all discrepant cases from the training set had a mean score within 15 percentage points of the 40% cutoff, any case in the validation set that was assigned a score between 25% and 55% was also scored twice. Despite this additional re-evaluation step, there was no significant difference in concordance between the two sets (Table 1).

The effect of preselecting tissue microarray-sized fields was evaluated among the discrepant cases. When pathologists scored 1-mm diameter circular fields instead of entire sections, the concordance was significantly much higher as indicated by kappa scores (see Table 1). The number of discrepant cases dropped by 46% (from 13 cases to 7 cases), and the total number of discrepant scores dropped by 70% (from 30 to 9 scores). This indicates that preselecting for tissue microarray-sized fields significantly improves concordance among pathologists but does not entirely eliminate discordant cases.

The potential effect of using an image analysis program to improve scoring of challenging cases was also investigated by using the Aperio immunohistochemistry nuclear algorithm. We applied this algorithm to score entire sections and tissue microarray-sized fields in discrepant cases and compared the automated scores with the mean score rendered by pathologist on each case. The automated and manual scores resulted in concordant MYC designation in 9 out of the 13 (69%) and 12 out of the 13 (92%) cases when entire sections and tissue microarray-sized fields were scored, respectively. This indicates that image analysis might be helpful in cases that are difficult to score. Preselection for tissue microarray-sized field still had a positive effect on concordance even when automated scoring was employed. One important caveat for using

image analysis is that its performance is significantly impacted by the parameters used in the algorithm. In our study, the most significant parameter is nuclear threshold. The specific value selected for this parameter will determine how intense the nuclear staining has to be for the software to call it positive.

Our findings indicate that a significant number of diffuse large B-cell lymphoma cases are inherently difficult to score for MYC protein expression. Careful scoring of potentially difficult cases does not improve concordance in our study. Prior studies indicated that discrepant cases were resolved through group review at multi-headed microscope but did not provide any further details.^{9,10,13} As staining heterogeneity is the most significant factor causing discrepancy, specific instructions on how to address this problem are needed to improve concordance.

As expected, four cases of Burkitt lymphoma had very high MYC expression and perfect scoring consistency. Interestingly, one case of Burkitt lymphoma (case 9) was very difficult to score. This case was re-reviewed in its entirety by three hematopathologists. Based on morphological and immunohistochemical findings, two hematopathologists agreed with the original diagnosis of Burkitt lymphoma, and one hematopathologist thought it represented a B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma. FISH analysis did not show evidence of *MYC-IGH* rearrangement. The discrepant scoring in this case could be due to the fact that this case actually represents a grey zone lymphoma rather than Burkitt lymphoma although up to 10% of Burkitt lymphoma cases may lack a demonstrable *MYC* translocation by FISH.¹⁶ Other explanations include the presence of extensive necrosis or antigen decay due to the lengthy storage of paraffin block for 9 years in this case.

In summary, our findings in this study indicate that an accurate evaluation of MYC protein overexpression by immunohistochemistry is more challenging than previously described and may lead to discrepant MYC status designation among pathologists in a significant proportion of cases. Until specific instructions about how to deal with staining heterogeneity becomes available, pathologists are advised to exercise caution when interpreting MYC protein expression by immunohistochemistry, especially in cases with staining heterogeneity or scores close to 40%.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

- Dang CV. MYC on the path to cancer. *Cell* 2012;149:22–35.
- Ott G, Rosenwald A, Campo E. Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification. *Blood* 2013;122:3884–3891.
- Savage KJ, Johnson NA, Ben-Neriah S, *et al*. MYC gene rearrangements are associated with a poor prognosis in diffuse large B-cell lymphoma patients treated with R-CHOP chemotherapy. *Blood* 2009;114:3533–3537.
- Barrans S, Crouch S, Smith A, *et al*. Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J Clin Oncol* 2010;28:3360–3365.
- Aukema SM, Siebert R, Schuurin E, *et al*. Double-hit B-cell lymphomas. *Blood* 2011;117:2319.
- Johnson NA, Savage KJ, Ludkovski O, *et al*. Lymphomas with concurrent BCL2 and MYC translocations: the critical factors associated with survival. *Blood* 2009;114:2273–2279.
- Snuderl M, Kolman OK, Chen YB, *et al*. B-cell lymphomas with concurrent IGH-BCL2 and MYC rearrangements are aggressive neoplasms with clinical and pathologic features distinct from Burkitt lymphoma and diffuse large B-cell lymphoma. *Am J Surg Pathol* 2010;34:327–340.
- Jaffe ES, Stein H, Swerdlow SH, *et al*. B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and classical Hodgkin lymphoma. In: Swerdlow SH, Campo E, Lee Harris N, *et al*. (eds). WHO Classification of Tumors of Haematopoietic and Lymphoid Tissues, 4th edn. IARC Press: Lyon, France; 2008, pp 267–268.
- Johnson NA, Slack GW, Savage KJ, *et al*. Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J Clin Oncol* 2012;30:3452–3459.
- Green TM, Young KH, Visco C, *et al*. Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. *J Clin Oncol* 2012;30:3460–3467.
- Horn H, Ziepert M, Becher C, *et al*. MYC status in concert with BCL2 and BCL6 expression predicts outcome in diffuse large B-cell lymphoma. *Blood* 2013;121:2253–2263.
- Hu S, Xu-Monette ZY, Tzankov A, *et al*. MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: a report from The International DLBCL Rituximab-CHOP Consortium Program. *Blood* 2013;121:4021–4031.
- Perry AM, Alvarado-Bernal Y, Laurini JA, *et al*. MYC and BCL2 protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with rituximab. *Br J Haematol* 2014;165:382–391.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;20:37–46.
- Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971;76:378–382.
- Leoncini L, Raphael M, Stein H, *et al*. Burkitt lymphoma. In: Swerdlow SH, Campo E, Lee Harris N, *et al*. (eds). WHO Classification of Tumors of Haematopoietic and Lymphoid Tissues, 4th edn. IARC Press: Lyon, France; 2008, pp 80–81.