SHORT PAPER

NEOPLASTIC DISEASE

Short title: Accuracy of Cytological Examination of Feline Lymphoma

Running head: M Gambini *et al*

Cytology of Feline Nodal Lymphoma: Low Interobserver Agreement and Variable Accuracy in Immunophenotype Prediction

Matteo Gambini*,†, Valeria Martini*,†, Serena Bernardi*,†, Mario Caniatti*,†, Maria Elena Gelain‡, Paola Roccabianca*,† and Stefano Comazzi*,†

**Department of Veterinary Medicine and †Veterinary Teaching Hospital, University of Milan, Milan and ‡Department of Comparative Biomedicine and Food Science, University of Padua, Padua, Italy*

Correspondence to: Valeria Martini (e-mail: valeria.martini@unimi.it)

Summary

Nodal lymphomas are less common in cats than in dogs and, consequently, no specific studies have been published. Cytology is the first step in the diagnosis of nodal lymphoma but is highly subjective. Morphological features have been introduced for the cytological classification of canine lymphomas but not for cats. Therefore, the aim of this study was to evaluate interobserver agreement on various cytological features of feline nodal lymphomas and to investigate the accuracy in predicting B or T immunophenotypes. Four veterinary cytologists examined 25 feline

nodal and mediastinal lymphoma cytological samples by adapting the criteria used for the evaluation of canine lymphomas and setting histopathology and immunohistochemistry as the gold standard. High interobserver variability was found in the evaluation of most features except for the presence or absence of cytoplasmic vacuoles, which were more common in B cell lymphomas. Cytology training centre was the major factor influencing the extent of agreement among evaluators. Diagnostic accuracy in predicting lymphoma immunophenotype varied from 35% to 75% and did not appear to be correlated with the experience of the evaluators. We conclude that cytological criteria, commonly used to describe canine lymphomas, are not adaptable to the counterpart feline neoplasms. Cytology-based immunophenotyping of feline lymphomas from different laboratories, and different cytologists within the same laboratory, differ substantially and should not be considered reliable. Specific cytological criteria are needed to describe feline lymphoma.

Keywords: accuracy; cytology; feline nodal lymphoma; interobserver agreement

Nodal lymphoma is common in dogs and is most often diagnosed by cytology (Zandvliet, 2016). Conversely, cytological diagnosis of lymphoma is challenging in cats (Blackwood, 2013) and this may be particularly true for nodal lymphomas, likely because specific diagnostic criteria have been poorly described due to the low prevalence of this disease presentation (Gabor *et al*, 1998; Moore, 2013).

In dogs, cytology is considered to be a reliable technique for diagnosing lymphoma, given the remarkable prevalence of high-grade cases (Fournel-Fleury *et al*, 1997; Ponce *et al*, 2010). However, the classification of these tumours, based on their cytological features, is characterized by variable interobserver agreement, ranging from fair to almost perfect, depending on the classification system applied (Teske and van Heerde, 1996). Therefore, further laboratory analyses, such as immunohistochemistry (IHC) or flow cytometry, are generally required to confirm the

diagnosis and determine immunophenotype and lymphoma subtype (Burkhard and Bienzle, 2015). However, some specific cytological features have been described in the dog as potentially suggestive of T or B cell origin including cell size, cytoplasmic colour and granules, nuclear shape, chromatin pattern and number, size and distribution of nucleoli (Fournel-Fleury *et al*, 1997; Ponce *et al*, 2010). To the best of our knowledge, similar specific cytological criteria have not been applied to the classification of feline nodal lymphomas and no data are available on interobserver variability in the assessment of cytological features of feline nodal lymphoma. Therefore, the aim of this study was to evaluate the diagnostic performance of cytology in predicting the phenotype of feline nodal and mediastinal lymphomas. Precision was assessed by calculating interobserver agreement on various cytological features that might be useful in predicting immunophenotype (ie, B or T cell), whereas accuracy was calculated for each observer by using the results of IHC as a gold standard.

We retrospectively investigated the database and archives of the Veterinary Teaching Hospital of the University of Milan and the Department of Comparative Biomedicine and Food Science of the University of Padua from January 2010 to January 2019. The inclusion criteria were a definitive diagnosis of nodal or mediastinal lymphoma, and the availability of at least one good-quality cytological smear from a lymph node (LN) or mediastinal mass for review and one formalin-fixed paraffin-embedded (FFPE) tissue block of the corresponding lesion. Mediastinal masses were also included in the study, because lymphomas in this site may also arise from mediastinal or sternal LNs (Fabrizio *et al*, 2014). From each FFPE tissue block, at least five sections were cut. One section was stained with haematoxylin and eosin (HE) and four were used for IHC utilizing primary antibodies for CD20 directed against mature B cells (epitope-specific rabbit antibody; Thermo Fisher Scientific, Cheshire, UK; 1:800), CD79 directed against all stages of B cells (monoclonal mouse anti-human, clone HM57; Dako, Atlanta, Georgia, USA; 1:100), CD3 directed against T cells (mouse monoclonal, clone F7.2.38, Dako; 1: 100) and CD5 directed against T cells (monoclonal mouse anti-human, clone SP19; Abcam, Cambridge, UK; prediluted

and ready to use). IHC was performed with an automatic immunostainer (Ventana Benchmark XT; Roche Diagnostics, Monza, Italy). All reagents were dispensed automatically except for the primary antibodies, which were manually dispensed.

The diagnosis of lymphoma was confirmed in all cases by a European College of Veterinary Pathologists (ECVP) board-certified pathologist (PR) based on routine histopathology and IHC. The latter was further used to immunophenotype lymphomas as B cell or T cell types.

All cytological specimens were stained by the May–Grünwald Giemsa technique and blindly evaluated by four cytologists with different experience: an ECVCP board-certified clinical pathologist (SC, evaluator 1), an ECVP board-certified anatomical pathologist (MC, evaluator 3) and their respective PhD students (SB, evaluator 2, and MG, evaluator 4). All evaluators were aware of the final diagnosis of nodal lymphoma, but not of the subtype nor the immunophenotype of the neoplasms. Before the beginning of the study, the cytologists conferred to standardize the description and the corresponding categorization of the morphological features that had to be evaluated for each cytological specimen. The morphological features evaluated were based on those used to evaluate canine lymphomas with the addition of other criteria, including cell homogeneity, the presence of vacuoles or perinuclear halos and presence of accessory non-neoplastic cells (Table 1, Fig. 1). In general, deeply bluish cytoplasm, a perinuclear halo and a round nucleus with visible nucleoli, were considered suggestive of B cell phenotype, whereas slightly basophilic cytoplasm and the presence of cytoplasmic granules, an indented, convoluted or irregular nucleus without nucleoli and higher numbers of plasma cells and eosinophils were considered suggestive of T cell phenotype. Despite the lack of a supposed link with phenotype, the other morphological features were considered in the study because they are commonly included in cytological reports in the laboratory practice. Evaluators were free to decide on phenotype, based on the prevalence of these criteria, current literature reports and their personal experience.

Overall, interobserver agreement on the morphological features was calculated using free-marginal Fleiss' kappa, using an online calculator (http://justusrandolph.net/kappa/). The

coefficients were interpreted according to Landis and Koch (Landis and Koch, 1977) as follows: ≤0, no agreement; >0.00 and ≤0.20, low agreement; ≥0.21 and ≤0.40, fair agreement; ≤0.41 and ≥0.60, moderate agreement; ≥0.61 and ≤0.80, substantial agreement; ≥0.81 and ≤1.00, almost perfect agreement. Considering that the different level of expertise of the operators may have influenced our results, we assessed the agreement between the two board-certified operators, between the two PhD students and between each board-certified evaluator and the respective PhD student. Cohen's kappa was calculated accordingly using an online calculator (https://www.graphpad.com/quickcalcs/kappa1/) and interpreted according to Landis and Koch (Landis and Koch, 1977). Finally, the level of accuracy of each evaluator in predicting immunophenotype was investigated. The accuracy in correctly diagnosing lymphoma immunophenotype was calculated as the number of correctly identified cases divided by the total number of cases and expressed as a percentage.

Overall, 36 cases fulfilled the initial criteria for inclusion in the study. However, five of these cases were excluded because at least one examiner considered the quality of the cytological specimen to be suboptimal, while six cases were excluded because the immunophenotype could not be determined on the basis of IHC results due to poor fixation and conservation of the FFPE samples or because of re-diagnosis as lymphoid hyperplasia. Thus, 25 samples were finally enrolled in the study, including 13 (52%) B cell and 12 (48%) T cell lymphomas.

The results of the analysis of the overall interobserver agreement on the morphological features of the 25 cases are listed in Table 2. The level of agreement was fair to moderate for nine of the 15 parameters evaluated, whereas it was almost perfect for the presence or absence of cytoplasmic granules and vacuoles, substantial for the number of eosinophils and low for the presence or absence of a perinuclear halo or chromatin pattern. No agreement was found among evaluators when asked to predict immunophenotype after morphological assessment.

High levels of agreement were detected among the four operators when evaluating the presence or absence of cytoplasmic granules and the number of eosinophils. However, these results

were likely affected by the low prevalence of samples with these characteristics. Indeed, granules were detected only in two (8%) samples (one by two evaluators and the other by a single evaluator). Similarly, eosinophils were detected in high numbers (≥3 in 5 high-power fields) in only two (8%) samples by three evaluators.

Almost perfect interobserver agreement was found also for the presence or absence of cytoplasmic vacuoles. Vacuoles were detected in nine (36%) samples (six by all four operators, two by three operators and one by two operators). Interestingly, B cell lymphomas were overrepresented in this subset of samples (seven samples; 78%), highlighting the presence of cytoplasmic vacuoles as a feature of potential value for predicting immunophenotype. Further studies on a larger scale are required to confirm this hypothesis.

The results of pairwise interobserver agreement between the evaluators are shown in Table 3. In general, the level of agreement was highly variable and usually slight or fair. Specifically, when considering the two board-certified evaluators, only slight to fair agreement was found for almost all the parameters evaluated (12 of 15; 80%), whereas agreement was almost perfect for the presence or absence of cytoplasmic vacuoles. Similar results were obtained when evaluating the agreement between the two PhD students and between evaluator 1 and their PhD student. Conversely, better agreement was obtained when comparing evaluator 3 and their PhD student, with moderate to substantial agreement on 10 of 15 features (66.7%), slight to fair agreement on three parameters (20%) and no agreement on the presence or absence of cytoplasmic granules. According to the latter finding, it is noteworthy that evaluators 1 and 3 did not report in any case the presence of granules or perinuclear halos, respectively. This likely affected the results of the comparison with other evaluators on evaluation of these features. In particular, the apparent discrepancy between Fleiss' kappa coefficient and Cohen's kappa coefficient in relation to the presence of granules might rely on the fact that the latter analysis was not performed for all the possible evaluator pairings. The low prevalence of granules might have additionally influenced results.

The low kappa values obtained between the board-certified evaluators and between the PhD students suggest that experience is not a leading factor influencing interobserver agreement. Conversely, teaching centre seems to be a major influencing factor. Indeed, higher kappa values were obtained between each board-certified evaluator and the respective PhD student, rather than between the two board-certified evaluators.

Regarding immunophenotype prediction for evaluators no. 1, 2, 3 and 4, the accuracy in correctly diagnosing a lymphoma as B or T was 36% (95% confidence interval [CI] 18.0–57.5%), 56% (95% CI 34.9–75.6%), 64% (95% CI 42.5–82.0%) and 76% (95% CI 54.9–90.6%), respectively. Thus, the accuracy of immunophenotype prediction seems to have been operator dependent, varying from less than 40% to more than 75%. This finding indicates that the morphological features generally used for tentative immunophenotype prediction in dogs do not apply to feline nodal lymphoma. Therefore, we highly recommend applying immunophenotyping techniques, such as IHC or flow cytometry, to define the immunophenotype of feline lymphomas. This is considered to be essential even for canine lymphomas (Burkhard and Bienzle, 2015), although interobserver agreement in dogs is higher than that observed in the current study (Teske and van Heerde, 1996).

The causes for the low agreement found among the observers in this study need further elucidation. One possible explanation for our results might be that samples from feline lymphomas are often composed of heterogeneous populations of cells. Therefore, the choice of different regions during the microscopic evaluation by each evaluator could have strongly biased the results. This finding further underlines that cytology alone should be used with caution in predicting feline lymphoma immunophenotype and that laboratory testing is essential for accurate immunophenotyping.

Interestingly, an unexpected finding was that PhD students had a higher accuracy in immunophenotype prediction than their respective tutors. Young cytologists may be more prone to adsorb knowledge from different schools, thereby compensating for their limited experience.

Furthermore, evaluators with longer experience might have been biased by their former evaluation habits, although all reached an initial consensus on the classification of each morphological feature, trying to mitigate as far as possible confounding factors related to their variable experience in haemato-oncology.

The major limitation of the present study is the low number of cases included, which likely derived from the low prevalence of nodal lymphomas in cats and from the lack of a consistent diagnostic approach for feline lymphoma, based on cytology and histopathology. Additionally, intraobserver agreement was not evaluated, unlike in previous studies on canine lymphomas (Teske and van Heerde, 1996).

In conclusion, high interobserver variability affects the evaluation of morphological features of feline nodal lymphomas, thus preventing comparison of the results from different laboratories and even among different cytologists within the same laboratory. This variability may be even higher in a routine diagnostic setting, considering that in the current study the evaluators conferred before commencement of the study to standardize the morphological criteria used to describe the cells, and that this may have enhanced the level of agreement. Although we only included samples with a final diagnosis of lymphoma, the inclusion of non-lymphomatous lesions would likely have resulted in even lower agreement among operators.

Our results confirm the limitations of cytology in the immunophenotyping of feline lymphoma and that further tests are essential. Our observations should encourage veterinarians towards the discussion and creation of a shared definition and cytological classification of feline nodal lymphomas.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. The authors thank Mr Davide Trez for his technical support on immunohistochemistry.

Conflict of Interest Statement

The authors declare no conflict of interest with respect to the research, authorship or publication of this manuscript.

References

Blackwood L (2013) Cats with cancer: Where to start. *Journal of Feline Medicine and Surgery*, **15**, 366–377.

Burkhard MJ, Bienzle D (2015) Making sense of lymphoma diagnostics in small animal patients. *Clinics in Laboratory Medicine*, **35**, 591–607.

Fabrizio F, Calam AE, Dobson JM, Middleton SA, Murphy S *et al* (2014) Feline mediastinal lymphoma: a retrospective study of signalment, retroviral status, response to chemotherapy and prognostic indicators. *Journal of Feline Medicine and Surgery*, **16**, 637–644.

Fournel-Fleury C, Magnol JP, Bricaire P, Marchal T, Chabanne L *et al* (1997) Cytohistological and immunological classification of canine malignant lymphomas: comparison with human non-Hodgkin's lymphomas. *Journal of Comparative Pathology*, **117**, 35–59.

Gabor LJ, Malik R, Canfield PJ (1998) Clinical and anatomical features of lymphosarcoma in 118 cats. *Australian Veterinary Journal*, **76**, 725–732.

Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.

Moore A (2013) Extranodal lymphoma in the cat: prognostic factors and treatment options. *Journal of Feline Medicine and Surgery*, **15**, 379–390.

Ponce F, Marchal T, Magnol JP, Turinelli V, Ledieu D *et al* (2010) A morphological study of 608 cases of canine malignant lymphoma in France with a focus on comparative similarities between canine and human lymphoma morphology. *Veterinary Pathology*, **47**, 414–433.

Teske E, van Heerde P (1996) Diagnostic value and reproducibility of fine-needle aspiration cytology in canine malignant lymphoma. *Veterinary Quarterly*, **18**, 112–115.

Zandvliet M (2016) Canine lymphoma: a review. *Veterinary Quarterly*, **36**, 76–104.

Legends to Figures

Fig. 1. Cytological features of feline nodal and mediastinal lymphomas as determined by consensus among four examiners. (A) T cell lymphoma. Homogeneous cell population composed of small (black circle) to medium (white circle) neoplastic cells, characterized by round (black arrow) or indented (white arrow) nuclei with homogeneous (black arrowhead) or partially clumped (white arrowhead) chromatin. May–Grünwald Giemsa. Bar, 50 μm. (B) T cell lymphoma. Homogeneous cell population composed of small to medium neoplastic cells characterized by abundant slightly basophilic cytoplasm with intracytoplasmic magenta granules (black circle) and partially clumped (black arrowhead) or clumped (white arrowhead) chromatin. May–Grünwald Giemsa. Bar, 50 μm. (C) B cell lymphoma. Heterogeneous cell population including large neoplastic cells characterized by irregular (black arrow) or convoluted nuclei (white arrow) with prominent nucleoli (black arrowhead). May–Grünwald Giemsa. Bar, 33.5 μm. (D) B cell lymphoma. Heterogeneous cell population including medium to large neoplastic cells characterized by scant to moderately abundant, deeply basophilic cytoplasm (black arrow), perinuclear halo (black arrowhead) and intracytoplasmic clear vacuoles (white arrowhead). May–Grünwald Giemsa. Bar, 33.5 μm.