Title Page

Multicenter flow cytometry proficiency testing of canine blood and lymph node samples

Running header title: Flow cytometry proficiency testing

K. Meichner, T. Stokol, J. Tarigo, A. Avery, M. J. Burkhard, S. Comazzi, J. Fogle, D. Marks Stowe, B. Rütgen, D. Seelig, A. Wasserkrug-Naor, W. Vernau, D. Bienzle

From the Department of Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA (Meichner, Tarigo); Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY (Stokol); Department of Microbiology, Immunology & Pathology, University of Colorado, Fort Collins, CO (Avery); Department of Veterinary Biosciences, The Ohio State University, OH (Burkhard); Department of Veterinary Medicine, University of Milan (Comazzi); Department of Population Health and Pathobiology (Fogle, Marks Stowe), North Carolina State University, Raleigh, NC; University of Vienna, Vienna, Austria (Rütgen); Department of Veterinary Clinical Sciences, University of Minnesota (Seelig); Department of Diagnostic Medicine/Pathobiology, Kansas State University, Manhattan, KS (Wasserkrug Naor[a]); Department of Pathology, Microbiology & Immunology, University of California, Davis, CA (Vernau); Department of Pathobiology, University of Guelph, Guelph, ON (Bienzle)

[a] Current address: Novartis Institutes for BioMedical Research, East Hanover, NJ

Correspondence:
Dorothee Bienzle; Department of Pathobiology, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada
dbienzle@uoguelph.ca

Abstract

Background – Flow cytometry (FC) is used increasingly in veterinary medicine for further characterization of hematolymphoid cells. Guidelines for optimizing assay performance and interpretation of results are limited, and concordance of results across laboratories is unknown.

Objectives – To determine inter-investigator agreement on interpretation of FC results from split samples analyzed in different laboratories using various protocols, cytometers and software; and on interpretation of archived FC standard (FCS) data files contributed by different investigators.

Methods – Multicenter observational cross-sectional study. Anticoagulated blood or lymph node aspirate samples from 9 client-owned dogs were aliquoted and shipped to participating laboratories. Samples were analyzed with individual laboratory-developed protocols. In addition, FCS files from a set of separate samples from 11 client-owned dogs were analyzed by participating investigators. A study non-participant tabulated results and interpretations. Agreement of interpretations was assessed with Fleiss' kappa statistic.

Results – Prolonged transit times affected sample quality for some laboratories. Overall agreement among investigators regarding FC sample interpretation was strong ($\kappa = .86 \pm .19$, $P < .001$), and for specific categories ranged from moderate to perfect. Agreement on category of lymphoproliferation or other leukocyte sample from analysis of FCS files was weak ($\kappa = .58 \pm .05$, $P < .001$).

Conclusions – Lymphoproliferations were readily identified by FC but identification of categories of hematolymphoid neoplasia in fresh samples or archived files was variable. There is a need for a more standardized approach to maximize the enormous potential of FC in veterinary medicine.

Keywords: Assay performance; Dog; External laboratory quality assessment; Immunophenotyping

Introduction

Flow cytometry (FC) is a laser-based analytic technique whereby multiple concurrent light scatter and light emitting properties of cells are measured.[1] Assessment of hematolymphoid neoplasms in humans typically incorporates flow cytometric immunophenotyping of leukocytes with panels of fluorochrome-labeled antibodies, in addition to morphologic, cytogenetic and molecular evaluation.[2] In animals, FC is a commonly used research tool, but clinical applications for characterization of hematolymphoid neoplasms have only evolved in recent years.[1,3,4] Cytogenetic and molecular assays other than analysis of clonality of antigen receptor genes are rarely used for diagnostic purposes. Flow cytometry is a complex analytic technique with many potential variables introduced by sample collection, preparation, analysis, and interpretation, which can profoundly affect results.[1,5] Furthermore, most instruments used in veterinary medicine are not validated for diagnostic purposes, and voluntary or mandatory quality assurance (QA) or quality control (QC) programs are uncommon. However, with increasing knowledge regarding the prognosis of different immunophenotypes of hematolymphoid neoplasms in animals,[3,4,6-10] results of FC have the potential to profoundly impact patient management.

In human medicine, consensus documents to guide all analytical aspects of clinical FC have been in place for several decades, and instruments and reagents are designated specifically for clinical use with limited adjustability and variability.[11-16] Furthermore, clinical laboratories for human samples are subject to national or regional QA/QC programs.[17,18] Laboratories abide by such guidelines to fulfill legal and accreditation requirements, and to provide optimal patient care.[11-13,15-17,19-22] Proficiency testing (PT) is one component of QA. In the United States, a common PT program administered by the College of American Pathologists (CAP) consists of 2 to 3 shipments of 2 to 3 samples (blood, bone marrow or organ aspirates) per year sent to

participating laboratories for analysis and comparison of results.[18] Samples in individual laboratories are analyzed by FC in the same manner as other patient specimens, and results are reported back to the CAP. Deviation from expected results requires correction of assay performance to ensure accurate patient results and to meet requirements for laboratory accreditation.

Neither consensus recommendations nor quality programs for veterinary clinical FC analysis have been established. For that reason, a FC interest group was formed at the joint annual meeting of the American College of Veterinary Pathologists (ACVP) and the American Society for Veterinary Clinical Pathology (ASVCP) in Atlanta, Georgia, in 2014. The group has subsequently met annually, and includes representatives from academic and commercial laboratories from North America and Europe performing or planning to establish diagnostic veterinary FC. As a first step towards establishing consensus recommendations, a PT program was initiated to compare immunophenotyping results between laboratories. Results of the PT initiative, and recommendations for minimum standards in veterinary FC, are presented here.

Material and Methods

Patient samples

Samples for FC were obtained between February 1, 2015 and July 31, 2017. All samples were of canine origin and had been submitted for diagnostic testing for suspected hematopoietic neoplasia to Cornell, North Carolina State, Georgia or Guelph University. Samples for PT were left over after diagnostic testing; therefore, ethics committee approval was not obtained but owners provided written consent for testing. Lymph node (LN) aspirates were placed into FC buffer (1x phosphate-buffered saline supplemented with 1% heat-inactivated fetal bovine serum, 1% 0.5 M potassium EDTA [K-EDTA], and 1% sodium azide), and peripheral blood (PB)

samples were placed into K-EDTA tubes. Samples were aliquoted and shipped by courier on ice overnight Monday through Wednesday within 24 hours after sample acquisition. Patient signalment and numerical CBC results were provided, but neither cytologic, histopathologic, nor additional clinical findings were provided before the FC analysis was completed.

Analysis of fresh samples by FC

Samples were selected for inclusion according to availability of adequate specimen volume, and ability to be shipped and analyzed during regular working hours. Immunophenotyping was performed using individual laboratory developed test (LDT) protocols, as previously reported by several laboratories.[6,7,10,23,24] In general, samples were aliquoted into FC polypropylene tubes. Red blood cells were lysed according to individual LDT protocols, which included ammonium-chloride-potassium (ACK) buffer or water lysis of blood, bone marrow and LN specimens. Then, antibodies were added as per LDT protocol (Table 1). Samples were analyzed using laboratory-specific FC instruments. Specimens with insufficient cells or poor viability (<50%) were excluded from analysis. Viability assessment was according to individual LDT protocols and included Trypan blue staining before FC cell preparation and/or incorporating a viability dye such as 7-aminoactinomycin D (7-AAD) or propidium iodide (PI) into the FC staining protocol.

Analysis of archived FC files

Participants also analyzed a set of archived FC standard (FCS) data files from dogs with hematopoietic neoplasia (hereafter referred to as '*files'*) distinct from those submitted as fresh samples (*'samples'*). The files were generated with instrument-specific acquisition software by each laboratory using LDT protocols and saved in generic FCS 3.0 format for opening with

different analytic software. Each file was provided with information on the combination of antibodies and fluorochromes, event number, nature of control samples, and the instrument used for acquisition.

Reporting of results

For fresh samples, each participant generated a written report including the percentage of viable cells, percentage of cells in the various leukocyte subsets as defined by a common set of antibodies (Table 2), and interpretations as routinely generated by each participant. An individual who did not participate in data acquisition or analysis compiled all results into a spreadsheet. Individual laboratory names were not recorded but rather a number was randomly assigned to each laboratory for each sample to allow blinded interpretation of results.

Results of fresh samples were grouped into the following categories: '*$CD4^+$ T cell lymphoproliferation*' (expanded population of $CD4^+/CD3^+$ or $CD4^+/CD5^+$ cells), '*$CD8^+$ T cell lymphoproliferation*' (expanded population of $CD8^+/CD3^+$ or $CD8^+/CD5^+$ cells), '*$CD4^-CD8^-$ T cell lymphoproliferation*' (expanded population of $CD3^+/CD4^-/CD8^-$ and/or $CD5^+/CD4^-/CD8^-$ cells) or '*B cell lymphoproliferation'* (expanded population of $CD21^+$ cells). An expanded cell population with particular light scatter and immunophenotypic features was defined by individual LDT protocols. Other categories were '*mixed*' (mixed population of cells with variable immunophenotypes inconsistent with neoplasia and therefore supportive of a reactive process) and '*other'* (samples for which the constellation of antigen detection was equivocal for a neoplastic or reactive process, or samples that did not meet criteria of the above categories). Finding a predominance of cells with an aberrant immunophenotype, such as lack of antigens usually expressed on leukocytes, or an atypical constellation of antigens, was also considered

supportive of neoplasia. Percentage of cells positive for an antigen was determined from investigator-set gates, which in turn was based on forward-light scatter (FSC) and side-light scatter (SCC) characteristics of cells and/or detection of a common leukocyte antigen.

Archived files were interpreted in a similar manner as described above. Each investigator was provided with a set of FCS files for interpretation. A category of '*$CD4^+CD8^+$ T cell lymphoproliferation'* was added for samples with an expanded population of $CD4^+/CD8^+/CD3^+$ or $CD4^+/CD8^+/CD5^+$ cells. '*Other'* could be an interpretation of 'equivocal' regarding neoplastic or reactive immunophenotype, or samples that did not meet criteria of the other categories.

Statistical analysis

Agreement among participants was assessed using Fleiss' kappa analysis in MS Excel 2013.[25] Agreement was determined for overall sample and file interpretations, and for identification of individual categories. Graphs were generated with GraphPad Prism (version 7). For Fleiss' kappa statistics, each sample needs to be evaluated by an equal (fixed) number of raters. Raters do not necessarily have to be the same for each sample. To meet this requirement, the number of raters (n = x) was restricted to the minimum number of participating investigators for each sample, meaning that each sample had to be evaluated at least x times. Therefore, cases with more than n = x raters (n = x + y), n = y raters were randomly excluded from the analysis using the Excel randomization function. For example, if overall the minimum number of investigators per sample was 7, each sample had to be evaluated at least 7 times, therefore, if a sample was analyzed by 8 investigators, one investigator had to be randomly excluded from the statistical analysis. Agreement was defined as '*no agreement*' for κ = -.10 – .2; '*minimal agreement*' for κ = .21 – .39; '*weak agreement*' for κ = .4 – .59; '*moderate agreement*' for κ = .6 – .79; '*strong*

agreement' for κ = .8 – .9; '*almost perfect agreement*' for κ = .91 – .99, and '*perfect agreement*' κ = 1.0.[26] Kappa values with standard errors were calculated, and $P \leq .05$ was considered statistically significant.

Results

Source of fresh FC samples

Nine fresh samples were analyzed, and 9 laboratories participated in the analysis of the fresh samples (Table 3). Not all laboratories received samples suitable for analysis or were able to analyze samples at particular time points; therefore, between 4 and 9 results were available for any particular fresh sample. Samples consisted of peripheral blood (PB) from 7 dogs, labeled as sample numbers: 1 – female spayed (FS) German shorthaired Pointer, 0.9 years, marked leukocytosis and systemic blastomycosis; 4 – FS mixed breed dog, 11 years, lymphocytosis; 5 – FS mixed breed dog, 10 years, no clinical abnormalities; 6 – FS mixed breed dog, 11 years, cytologically unclassifiable leukocytes; 7 – FS Golden retriever, 11 years, lymphocytosis; 8 – male neutered (MN) Golden retriever, 12 years, pancytopenia and unclassifiable leukocytes; 9 – FS Doberman Pinscher, 12 years, lymphocytosis and cytological diagnosis of lymphoid neoplasia. Two LN aspirates were derived from dogs with a cytological diagnosis of lymphoma: 2 – FS Weimaraner, 5 years; 3 – MN German Shepherd mixed breed dog, 3 years.

Pre-analytical aspects

Pre-analytical factors precluded sample assessment in some cases (Figure 1). For example, sample 1 (PB) could not be analyzed due to a transit delay of >72 hours resulting in hemolysis and poor (<50%) cell viability or tube breakage (3 and 5 laboratories, respectively). Sample 2

arrived with insufficient or poorly viable cells (<5%) in 2 and 1 laboratories, respectively. Sample 3 was not interpreted by 2 investigators because of poor cell viability (15% and <5%, respectively). No pre-analytical problems were encountered for the remaining 6 samples. Cytometers used were Accuri C6, LSR II, FACSCalibur, FACSCanto II, LSR Fortessa X-20 (all BD Biosciences, San Jose, CA) in 1, 3, 2, 1 and 1 laboratory, respectively, and Gallios (Beckman Coulter, Brea, CA) in 1 laboratory.

Interpretation of fresh samples

At least four interpretations per sample were available. The overall agreement between all participating investigators regarding the immunophenotype of the 9 fresh samples was strong ($\kappa = 0.86 \pm 0.19$; $P < .001$, Table 3 and Figure 1). Agreement was moderate for '*$CD4^+$ T cell lymphoproliferation*' ($\kappa = .64 \pm .41$; $P = .119$), perfect for '*$CD8^+$ T cell lymphoproliferation*' ($\kappa = 1 \pm .41$; $P < .014$), perfect for '*$CD4^-CD8^-$ T cell lymphoproliferation*' ($\kappa = 1 \pm .41$; $P < .014$), moderate for '*B cell lymphoproliferation*' ($\kappa = .75 \pm .20$; $P < .001$), perfect for '*mixed*' immunophenotype ($\kappa = 1 \pm .29$; $P < .001$) and moderate for '*other*' ($\kappa = .77 \pm .29$; $P = .008$). Within the '*other*' category, sample 2 was interpreted as inconclusive by one investigator but as a B cell lymphoproliferation by the other 3; sample 8 was interpreted as an undifferentiated leukemia by all investigators due to absence of antigen detection and abnormal cells in circulation.

Source of archived FCS files

Eight laboratories provided 11 archived FCS files for analysis (Table 4); 3 laboratories provided 2 cases each and 5 laboratories provided 1 case each. FCS files were generated from PB of 3

dogs: File 1 - FS mixed breed dog, 13 years; file 6 – MN Shih Tzu, 10 years; file 7 – FS mixed breed dog, 12 years. LN aspirates were from 6 dogs: File 2 – MN Golden retriever, 9 years; file 3 – FS Cavalier King Charles spaniel, 4 years; file 4 – FS Dogue de Bordeaux, 7 years; file 5 – FS German Shepherd, 12 years; file 8 – MN Boxer, 6 years; file 11 – FS mixed-breed, adult. Additional samples originated from an aspirate of a mediastinal mass in a FS Blue Heeler, 8 years (file 9), and PB, LN, and bone marrow from a male Jack Russell terrier, 9 years (file 10). All dogs had a cytologic diagnosis of lymphoid neoplasia.

Interpretation of archived FCS files

FlowJo, FACSDiva and Cell Quest (all BD Biosciences) software was used by 4, 2 and 1 investigator, respectively, and Kaluza (Beckman Coulter) was used by 1 investigator for analysis. Software used by some investigators was unsuitable for the FCS format generated by acquisition software of some cytometers, i.e. files 1, 2, and 4 were not analyzable by three, four and five of eight investigators, respectively (‘*uninterpretable’*; Table 4, Figure 2). Files that were categorized into ‘*other’* included file 4 (interpreted as presumptive unclassified leukemia), file 7 (interpreted as equivocal for a neoplastic or reactive process), file 8 (interpreted equivocal as B or T cell neoplasm), and file 9 (interpreted as thymoma) by one investigator each. Overall agreement between investigators was weak ($\kappa = .58 \pm 0.05$, $P < .001$). For individual categories, agreement was strong for ‘*B cell lymphoproliferation*’ ($\kappa = .85 \pm 0.11$; $P < .001$), moderate for ‘*CD4*$^{+}$’ ($\kappa = .76 \pm 0.08$; $P < .001$) and ‘*CD8*$^{+}$’ ($\kappa = .76 \pm .11$; $P < .001$) ‘*T cell lymphoproliferation*’, minimal for ‘*CD4*$^{+}$*CD8*$^{+}$*’ T cell lymphoproliferation*’ ($\kappa = .29 \pm .09$; $P = .002$) and ‘*uninterpretable’* ($\kappa = .37 \pm .11$; $P < .001$), with no agreement for the diagnosis of

‘$CD4^{-}CD8^{-}$*T cell lymphoproliferation*’ (κ = .03 ± .09; P = .77) and ‘*other’* (κ = -.05 ± .08; P = .54) (Table 4, Figure 2). No files were interpreted as ‘*mixed’* or reactive.

Discussion

In human medicine, FC immunophenotyping is a standardized clinical test using cytometers with limited adjustability and automatic gating algorithms subject to proficiency assessment.[27] In animals, FC has been applied as a research tool for decades but use as a diagnostic test is in the very early stages. At this point, there are no consistent protocols for cell preparation, antibody type, antibody amount, use of controls, data analysis or interpretation. For example, cell preparation can vary substantially between laboratories, ranging from samples prepared in tubes or 96-well plates, single-fluorochromes applied in two-step format or 2 to 8 directly conjugated antibodies applied concurrently. A unified format for reporting of FC results by the European canine lymphoma network has been proposed but a similar recommendation for pre-analytical or analytical FC aspects remains to be constructed.[28] If primary and secondary antibodies are used, both need to be titrated using appropriate target cells, and multiple concurrent antibodies need to be tested in combination for fluorochrome interference and spectral overlap. Concerning human clinical samples, many approaches have been described for this purpose, and discussions are ongoing regarding improved preparatory and analytical methods.[29,30] Various cytometers are used in veterinary diagnostic laboratories and they are often designed for adjustability to serve multiple cell types and species in research rather than for clinical purposes. Different cytometers also have variable acquisition software. Other challenges toward establishing FC as a standardized test in veterinary laboratory medicine are limited availability of validated antibodies directed to animal leukocyte antigens, and lack of reactivity of most of such antibodies with

antigens that have been formalin exposed. Hence, only fresh samples can be analyzed but that poses challenges with timely shipment. Finally, there are idiosyncrasies of animal leukocytes such as expression of CD4 on canine neutrophils and loss of CD45 on T zone lymphoma cells, which require specific expertise for interpretation.[31,32] This study was a first voluntary effort involving institutions that perform diagnostic veterinary FC for the purpose of 1) describing reagents and instruments being used; 2) assessing concordance of results from analysis of split samples; and 3) assessing concordance of interpretation of archived FCS files.

Nine laboratories participated in analysis of fresh samples, but not all received suitable samples in a timely manner nor was sufficient sample available for all participating laboratories. Since fresh samples needed to be shipped across long distance and borders, cell viability was poor in several instances. Samples with viability <50% were excluded, but reduced cell integrity might still have contributed to non-specific antibody binding and therefore discrepant interpretations. Agreement between investigators on identification of lymphoproliferation was strong, but agreement on categories such as $CD4^+$ T cell and B cell tumors was moderate. Moderate agreement constitutes a quality problem for clinical laboratory tests in human medicine and is considered inadequate in healthcare research.[26] Moderate agreement would also be of concern for immunophenotyping canine lymphoid neoplasms. Response to therapy and survival vary among dogs with different T cell tumors, and also among dogs with B or T cell tumors.[6,7,10] Reliable identification of T versus B cell predominance is essential for basic immunophenotyping, and lack thereof precludes more detailed assessment of subcategories. Discrepancies in sample interpretation such as sample 3 being considered '$CD4^+$ T cell lymphoma' by three investigators, and 'B cell lymphoma' by a fourth investigator, might arise

from unfamiliarity with the constellation of antibodies being used, inappropriate instrument set-up, inappropriate fluorochrome compensation, lack of assessment of a corresponding blood or lymph node aspirate smear, differing gating strategies or limited experience with a highly complex analytic technique such as FC. Similar reasons might account for sample 2 being interpreted as 'B cell lymphoma' by three investigators, and as 'equivocal' by a fourth investigator. Such variability in interpretation might in part be addressed through consensus on use of reagents, methods and analytic approaches, and increased training. Achieving uniformity in cytometer use is cost-prohibitive and therefore unrealistic at this time.

In principle, analysis of archived FCS files should generate concordant results regardless of the type of analytical software used. However, variable agreement was observed, which is also of concern due to the potential impact on patient management. Disparate interpretations may reflect differences in gating strategies, experience with certain antibody-fluorochrome combinations, types of controls used, compensation approaches, and types of instruments and analytical software. For example, FCS files generated by some instruments have pre-set scales for light scatter and fluorescence, which require manual adjustment with some other analytic software to visualize all cell populations. Variable computational software contributing to differences in interpretation were also reported for analysis of human FC samples.[33] The recently introduced open source software *CytoML*, an R/Bioconductor package, is reported to facilitate cross-platform import, export and analysis of cytometry data, and may be useful for future studies in veterinary FC.[34]

There are several limitations of this study. Samples did not have a gold standard diagnosis but rather all results were considered independent and the goal was not to compare results of different investigators relative those of submitting investigator, but rather to assess overall agreement. A gold standard would not be easy to generate since most antibodies reactive with formalin-fixed tissues are different than those used for FC. Nevertheless, establishment of a gold standard diagnosis from a combination of morphologic combined with immunohistochemical or immunocytochemical assessment of concurrent patient samples, or Bayesian statistics taking all pertinent clinical and diagnostic information into account, should be considered in future studies. The majority of samples were PB, since obtaining sufficiently cellular samples for 9 laboratories was rarely feasible from LN aspirates. Leukocytes are better preserved in blood than in FC buffer used for LN or organ aspirates; therefore, PB is more suitable for possible time-delayed analysis. Sample tubes containing proprietary preservative have been tested for prolonging the analytical lifespan of canine lymphocytes, but decreased immunoreactivity and viability occurred after 3 days.[35] Thus, future studies will likely still have to rely on fresh samples.

Results of this study do not invalidate previous findings using FC as a diagnostic assay for immunophenotyping canine lymphoproliferative diseases. Concordance between FC and IHC for immunophenotyping canine T and B cell lymphomas was previously reported to be high (Thalheim 2013), and entities such as T zone lymphoma were reproducibly identified using FC by multiple investigators (6, 10). However, findings in the present study indicate need for improved concordance in the analysis of canine FC samples. This will require development of consensus standards for all analytical aspects of clinical FC. It would be highly desirable to arrive at a peer-reviewed Optimized Multicolor Immunofluorescence Panel (OMIP) for veterinary FC, as established for human leukocytes in general and for specific leukocyte

subpopulations.[36,37] A sample canine multicolor panel with a rationale is provided in Appendix 1 (Supplementary File). In the interim, FC should be recognized as a very powerful technique to be used in conjunction with morphologic cell and tissue assessment, clonality and immunohistochemical assays.

Acknowledgments

The authors thank Gail Babcock, Cornell University, for compilation and masking of results prior to distribution for analysis and Dr. Deborah Keys, University of Georgia, for statistical consulting.

References

1. Reggeti F, Bienzle D. Flow cytometry in veterinary oncology. *Vet Pathol*. 2011;48:223-235.
2. Harris NL, Campo E, Jaffe ES et al. Introduction to the WHO classification of tumours of haemtopoietic and lymphoid tissues. In: Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW, eds. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissue*. 4th ed. Lyon, France: International Agency for Research on Cancer (IARC), 2008:14-15.
3. Burkhard MJ, Bienzle D. Making sense of lymphoma diagnostics in small animal patients. *Vet Clin North Am Small Anim Pract*. 2013;43:1331-1347
4. Rout ED, Avery PR. Lymphoid Neoplasia: Correlations Between Morphology and Flow Cytometry. *Vet Clin North Am Small Anim Pract*. 2017;47:53-70.
5. Donnenberg VS, Donnenberg AD. Coping with artifact in the analysis of flow cytometric data. *Methods*. 2015;1;82:3-11.
6. Seelig DM, Avery P, Webb T, et al. Canine T-zone lymphoma: unique immunophenotypic features, outcome, and population characteristics. *J Vet Intern Med*. 2014;28:878-886.
7. Avery PR, Burton J, Bromberek JL, et al. Flow cytometric characterization and clinical outcome of CD4+ T-cell lymphoma in dogs: 67 cases. *J Vet Intern Med*. 2014;28:538-546.
8. Bromberek JL, Rout ED, Agnew MR, et al. Breed Distribution and Clinical Characteristics of B Cell Chronic Lymphocytic Leukemia in Dogs. *J Vet Intern Med*. 2016;30:215-222.

9. Rao S, Lana S, Eickhoff J, et al. Class II major histocompatibility complex expression and cell size independently predict survival in canine B-cell lymphoma. *J Vet Intern Med*. 2011;25:1097-1105.
10. Deravi N, Berke O, Woods JP, Bienzle D. Specific immunotypes of canine T cell lymphoma are associated with different outcomes. *Vet Immunol Immunopathol*. 2017;191:5-13.
11. Davis BH, Holden JT, Bene MC, et al. 2006 Bethesda International Consensus recommendations on the flow cytometric immunophenotypic analysis of hematolymphoid neoplasia: medical indications. *Cytometry B Clin Cytom*. 2007;72 Suppl 1:S5-13.
12. Stetler-Stevenson M, Davis B, Wood B, Braylan R. 2006 Bethesda International Consensus Conference on Flow Cytometric Immunophenotyping of Hematolymphoid Neoplasia. *Cytometry B Clin Cytom*. 2007;72 Suppl 1:S3.
13. D'Archangelo M. Flow cytometry: new guidelines to support its clinical application. *Cytometry B Clin Cytom*. 2007;72:209-210.
14. Borowitz M, Bauer KD, Duque RE et al. Clinical Applications of Flow Cytometry: Immunophenotyping of Leukemic Cells; Approved Guideline. *Clinical and Laboratory Standards Institute (CLSI) – H43-A.* 1998 Vol. 18, No. 8:1-80.
15. Stetler-Stevenson M, Ahmad E, Barnett D et al. Clinical Flow Cytometric Analysis of Neoplastic Hematolymphoid Cells; Approved Guideline – Second Edition. *Clinical and Laboratory Standards Institute (CLSI) – H43-A2*. 2007 Vol. 27, No. 11:1-81.

16. McCoy JP Jr, Overton WR. Quality control in flow cytometry for diagnostic pathology: II. A conspectus of reference ranges for lymphocyte immunophenotyping. *Cytometry*. 1994;18:129-139.

17. Owens MA, Vall HG, Hurley AA, Wormsley SB. Validation and quality control of immunophenotyping in clinical flow cytometry. *J Immunol Methods*. 2000;243:33-50.

18. American College of Pathologists 2019 Surveys and Pathology Education Programs (Proficiency Testing catalog). Available at: https://cap.objects.frb.io/documents/2019-surveys-catalog.pdf. Accessed May 23, 2019.

19. Davis BH, Holden JT, Bene MC, et al. 2006 Bethesda International Consensus recommendations on the flow cytometric immunophenotypic analysis of hematolymphoid neoplasia: medical indications. *Cytometry B Clin Cytom*. 2007;72 Suppl 1:S5-13.

20. Greig B, Oldaker T, Warzynski M, Wood B. 2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: recommendations for training and education to perform clinical flow cytometry. *Cytometry B Clin Cytom*. 2007;72 Suppl 1:S23-33.

21. Wood BL, Arroz M, Barnett D, et al. 2006 Bethesda International Consensus recommendations on the immunophenotypic analysis of hematolymphoid neoplasia by flow cytometry: optimal reagents and reporting for the flow cytometric diagnosis of hematopoietic neoplasia. *Cytometry B Clin Cytom*. 2007;72 Suppl 1:S14-22.

22. D'Angelo R, Weiss R, Wolfe D, et al. Facing the Inevitable: Being Prepared for Regulatory Requirements for Laboratory Developed Tests. *Am J Clin Pathol*. 2018;149:484-498.

404 23. Meichner K, Fogle JE, English L, Suter SE. Expression of Apoptosis-regulating Proteins
405 Bcl-2 and Bax in Lymph Node Aspirates from Dogs with Lymphoma. *J Vet Intern Med*.
406 2016;30:819-826.
407 24. Stokol T, Schaefer DM, Shuman M, Belcher N, Dong L. Alkaline phosphatase is a useful
408 cytochemical marker for the diagnosis of acute myelomonocytic and monocytic leukemia
409 in the dog. *Vet Clin Pathol*. 2015;44:79-93.
410 25. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological*
411 *Bulletin*. 1971;76:378-382.
412 26. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* (Zagreb).
413 2012;22:276-282.
414 27. Finak G, Langweiler M, Jaimes M, et al. Standardizing Flow Cytometry
415 Immunophenotyping Analysis from the Human ImmunoPhenotyping Consortium. *Sci*
416 *Rep*. 2016;10;6:20686.
417 28. Comazzi S, Avery PR, Garden OA, et al.; European Canine Lymphoma Network.
418 European canine lymphoma network consensus recommendations for reporting flow
419 cytometry in canine hematopoietic neoplasms. *Cytometry B Clin Cytom*. 2017;92:411-
420 419.
421 29. Roederer M. Spectral compensation for flow cytometry: visualization artifacts,
422 limitations, and caveats. *Cytometry*. 2001;45:194-205.
423 30. Ryherd M, Plassmeyer M, Alexander C, et al. Improved panels for clinical immune
424 phenotyping: Utilization of the violet laser. *Cytometry B Clin Cytom*. 2018;94:671-679.

31. Moore PF, Rossitto PV, Danilenko DM, et al. Monoclonal antibodies specific for canine CD4 and CD8 define functional T-lymphocyte subsets and high-density expression of CD4 by canine neutrophils. *Tissue Antigens*. 1992;40:75-85.
32. Martini V, Cozzi M, Aricò A, et al. Loss of CD45 cell surface expression in canine T-zone lymphoma results from reduced gene expression. *Vet Immunol Immunopathol*. 2017;187:14-19.
33. Rawstron AC, Kreuzer KA, Soosapilla A, et al. Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: An European Research Initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) Harmonisation project. *Cytometry B Clin Cytom*. 2018;94:121-128.
34. Finak G, Jiang W, Gottardo R. CytoML for cross-platform cytometry data sharing. *Cytometry A*. 2018;93:1189-1196.
35. Cian F, Guzera M, Frost S, et al. Stability of immunophenotypic lymphoid markers in fixed canine peripheral blood for flow cytometric analysis. *Vet Clin Pathol*. 2014;43:101-108.
36. Moncunill G, Han H, Dobaño C, McElrath MJ, De Rosa SC. OMIP-024: pan-leukocyte immunophenotypic characterization of PBMC subsets in human samples. *Cytometry A*. 2014;85:995-998.
37. Donaldson MM, Kao SF, Foulds KE. OMIP-052: An 18-Color Panel for Measuring Th1, Th2, Th17, and Tfh Responses in Rhesus Macaques. *Cytometry A*. 2019;95:261-263.
38. Thalheim L, Williams LE, Borst LB, Fogle JE, Suter SE. Lymphoma immunophenotype of dogs determined by immunohistochemistry, flow cytometry, and polymerase chain reaction for antigen receptor rearrangements. *J Vet Intern Med*. 2013;27:1509-16.

448

449

Table 1. Antigens detected and antibodies used routinely for flow cytometric characterization of dog leukocytes by different laboratories.

			Laboratory[a]									
Antigen	Clone	Target species	1	2	3	4	5	6	7	8	9	10
CD1a[b]	CA9.AG5	Canine			x							
CD1a	CA13.9H11	Canine			x							
CD3	CA17.2A12	Canine	x	x	x	x	x	x	x	x	x	x
CD3	CD3-12	Human						x			x	
CD4	CA13.1E4	Canine			x							
CD4	YKIX302.9	Canine	x	x		x	x	x	x	x	x	x
CD5	YKIX322.3	Canine	x	x	x	x	x	x	x	x	x	x
CD8α	CA9.JD3	Canine			x							
CD8α	YCAT 55.9	Canine	x	x		x	x	x	x	x	x	x
CD8β	CA15.4G2	Canine			x							
CD11/18	YKIX490.6.4	Canine				x				x	x	
CD11a	CA11.4D3	Canine			x							
CD11a	HI111	Human									x	
CD11b	CA16.3E10	Canine	x		x							
CD11c	CA11.6A1	Canine	x		x							
CD11d	CA16.3D3	Canine			x							
CD11d	CA11.8H2	Canine	x		x							
CD14	TUK4	Human	x	x	x	x		x	x			x
CD18	CA1.4E9	Canine	x		x	x						
CD18	YFC118.3	Human		x								x
CD21	CA2.1D6	Canine		x	x	x		x		x	x	x
CD21	B-ly4	Human	x				x		x			
CD22	RFB4	Human	x	x		x						
CD25	P4A10	Canine	x	x								x
CD34	1H6	Canine	x	x	x	x	x	x	x	x	x	x
CD45	CA12.10C12	Canine			x							
CD45	YKIX716.13	Canine	x	x		x	x	x	x	x	x	x
CD45RA	CA4.1D3	Canine			x						x	
CD49d	CA4.5B3	Canine			x							
CD79a	HM57	Human								x	x	
CD79b	AT107-2	Murine						x				
CD90	CA1.4G8	Canine	x		x	x						
CD90	YKIX337.217	Canine										
MHC II[c]	CA2.1C12	Canine			x	x	x					
MHC II	YKIX334.2	Canine	x	x					x	x	x	x
TCR α/β[d]	CA15.8G7	Canine	x		x							
TCR γ/δ	CA20.6A3	Canine			x							
B5	Clone B5	Canine						x				

[a] 1, Cornell University; 2, Colorado State University; 3, University of California Davis; 4, University of Guelph; 5, Kansas State University; 6, North Carolina State University; 7, The Ohio State University; 8, University of Milan; 9, University of Vienna; 10, University of Georgia

[b] Allele-specific reactivity

[c] Major histocompatibility complex

[d] TCR, T-cell receptor

459 **Table 2**. Antigens detected in this study.
460

Antigen	Normal cell expression
CD3[a]	T-lymphocytes
CD4	Helper T-lymphocytes; neutrophils
CD5	Most T-lymphocytes
CD8	Cytotoxic T-lymphocytes
CD21	B-lymphocytes
CD45	Leukocytes
MHC II[b]	Lymphocytes, monocytes, macrophages, dendritic cells

461 [a] CD, Cluster of differentiation
462 [b] MHC, major histocompatibility complex
463

Table 3 – Categorization of fresh blood (PB) and lymph node (LN) samples from dogs by flow cytometric analysis. Nine investigators participated, and a minimum of 4 investigators interpreted each sample. Overall agreement between investigators regarding the type of hematolymphoid proliferation was strong ($\kappa = .86 \pm .19$; $P < .001$).

Sample	Source	Type of proliferation					
		T cell			B cell [d]		
		$CD4^{+}$ [a]	$CD8^{+}$ [b]	$CD4^{-}CD8^{-}$ [c]		Mixed [e]	Other [f]
1	PB	-	-	-	-	4	-
2	LN	-	-	-	3	-	1
3	LN	3	-	-	1	-	-
4	PB	-	-	-	4	-	-
5	PB	-	-	-	-	4	-
6	PB	-	4	-	-	-	-
7	PB	-	-	4	-	-	-
8	PB	-	-	-	-	-	4
9	PB	-	-	-	4	-	-
κ		.64	1	1	.75	1	.77
SE		.41	.41	.41	.20	.29	.29
P		.119	.014	.014	<.001	< .001	.008
95% CI		-.16-1.43	.19-1.80	.19-1.80	.35-1.15	.43-1.15	.20-1.33

CD, cluster of differentiation; CI, confidence interval; κ, kappa; SE, standard error
[a] Expanded population of $CD4^{+}/CD3^{+}$ or $CD4^{+}/CD5^{+}$ cells
[b] Expanded population of $CD8^{+}/CD3^{+}$ or $CD8^{+}/CD5^{+}$ cells
[c] Expanded population of $CD4^{-}/CD8^{-}/CD3^{+}$ or $CD4^{-}/CD8^{-}/CD5^{+}$ cells
[d] Expanded population of $CD21^{+}$ cells
[e] Population of cells with variable immunophenotypes inconsistent with neoplasia
[f] Constellation of antigen expression equivocal for a neoplastic or reactive process

Table 4 – Categorization of flow cytometry standard (FCS) files from blood (PB), lymph node (LN) and other tissue aspirate samples by 8 different investigators. Overall agreement regarding the type of proliferation of hematolymphoid cells was weak ($\kappa = .58 \pm .05$; P <.001). The category '*uninterpretable'* includes files that investigators could not analyze due to software incompatibilities.

File	Source	Type of proliferation							
		T cell				B cell	Mixed	Other	Uninterpretable
		$CD4^+$	$CD8^+$	$CD4^+CD8^+$ [a]	$CD4^-CD8^-$				
1	PB	-	-	5	-	-	-	-	3
2	LN	1	-	1	1	1	-	-	4
3	LN	-	1	-	-	7	-		
4	LN	-	-	-	2	-	-	1	5
5	LN	-	-	1	-	8	-		
6	PB	7	-	-	-	-	-		
7	PB	-	7	-	-	-	-	1	
8	LN	7	-	-	-	-	-	1	
9	Mass[b]	6	-	1	-	-	-	1	
10	Multiple[c]	-	7	-	1	-	-		
11	LN	8	-	-	-	-	-		
κ		.76	.76	.29	.03	.85	n/a	-.05	.37
SE		.08	.11	.09	.09	.11	n/a	.08	.11
P		< .001	< .001	.002	.77	< .001	n/a	.54	< .001
95% CI		.59-.93	.55-.97	.11-.48	-.16-.21	.63-1.06	n/a	-.20-.10	.15-.58

[a] Expanded population of $CD4^+/CD8^+/CD3^+$ or $CD4^+/CD8^+/CD5^+$ cells.
[b] Mediastinal mass
[c] Blood, lymph node, and bone marrow
For other abbreviations, see legend Table 3

Figure Captions

Figure 1. Graph plot depicting flow cytometric interpretation of individual freshly analyzed samples separated by sample type and immunophenotypic category. Samples of poor quality are shown as '*uninterpretable*' and were excluded from analysis. At least 4 investigators interpreted each sample, hence agreement was determined by random exclusion of any number of interpretations greater than four (see Table 3). Cells in sample 8 lacked expression of differentiating antigens and this sample was interpreted as undifferentiated leukemia ('*other'*). *Mixed*: Populations of cells with variable immunophenotypes inconsistent with neoplasia. *Other*: Constellation of antigen expression equivocal for a neoplastic or reactive process. *Uninterpretable*: Samples with insufficient cell number and/or viability for analysis.

Figure 2. Graph plot depicting interpretations of flow cytometry standard (FCS) files by immunophenotypic category. Eight investigators interpreted each file. *Other*: Constellation of antigen expression equivocal for a neoplastic or reactive process. *Uninterpretable*: Files that investigators could not analyze due to software incompatibilities.