

The use of CART algorithms to combine serum acute phase protein levels as a diagnostic aid in canine lymphoma.

Background. Identification of some of the biomarkers used in the original Lymphoma Blood test confirmed a role for acute phase proteins (APPs) in the detection of lymphoma in dogs. However, it has long been recognised that measurement of signal alone does not give sufficient test performance due to poor specificity. However, measuring multiple APPs and combining the results in an analytical algorithm could dramatically improve performance.

Objectives. To determine the efficacy of algorithms using Classification and Regression Tree (CART) analysis which combine the measurement of two APPs, Haptoglobin (HAPT) and C-Reactive Protein (CRP), for the detection of lymphoma in dogs.

Dogs. A sample test set of 194 dogs comprising 64 diagnosed with lymphoma, 51 healthy and 79 other conditions which commonly present with symptoms similar to lymphoma.

Methods. Serum samples were collected from dogs undergoing differential diagnosis for lymphoma. Lymphoma positive samples were confirmed by either FNA or excisional biopsy. Non-lymphoma dogs were confirmed to be free of the disease at a minimum of six months after providing the serum sample.

Results. By combining serum HAPT and CRP levels using a specifically designed algorithm, it was possible to differentiate between dogs with lymphoma and other diseases with great precision giving specificity of 93% and sensitivity of 85%. By comparing lymphoma and healthy dogs only, the test showed 100% specificity.

Conclusions. Combining canine serum APP levels using CART analysis enables differentiation between dogs diagnosed with lymphoma and other commonly encountered conditions in differential diagnosis with high levels of diagnostic significance.

Introduction.

During previous mass spectrometry studies on canine lymphoma, a number of protein peaks were shown to be constantly different between serum from dogs with lymphoma and control samples from dogs with other diseases. In all, 13 peaks were shown to be significantly differentially expressed between the populations at the $p < 0.05$ level. These peaks ranged from 5.3 to 74.4 kDa in size.

On further analysis, it was observed that two peaks showed particularly pronounced significant differences between the groups of dogs. These showed molecular weights of 35.9 and 74.4 kDa. It was therefore decided to investigate these proteins further in order to characterise and identify them.

The two protein peaks were purified using a combination of chromatographic and electrophoretic processes and the subsequently purified bands were then analysed by liquid chromatography mass

spectrometry (LC MS) to determine the amino acid sequence of each protein. The sequences were then compared to protein sequences in the National Centre For Biotechnology Information database using the MASCOT search engine to determine the protein identity.

This approach definitively identified the 35.9 kDa peak to be HAPT. Further work is still on going to complete the analysis of the 74 kDa peak.

HAPT has been known as an acute phase protein for some time, but has yet to find a strong application in veterinary medicine. Mischle *et al*¹ have shown significantly elevated serum levels of HAPT in dogs with lymphoma and other lymphatic neoplasia. However, it was concluded that the HAPT by itself was not specific to lymphoma and therefore not a good diagnostic marker for this disease.

From the start, the PetScreen approach has been to work with multiple markers in order to improve test performance. Classification and Regression Tree (CART) analysis was used to develop algorithms which combine values from different markers to dramatically improve performance above that available from single biomarkers alone. A number of factors pointed us to investigate the potential of developing novel algorithms using acute phase proteins in combination:

- C-Reactive Protein (CRP) is now routinely used during the diagnostic work up of Non-Hodgkin's Lymphoma in humans.
- Several reports have shown the CRP levels are elevated in dogs with lymphoma, but like HAPT, its stand-alone performance was not adequate for diagnostic purposes.
- The FDA has recently approved a multi-marker test which uses an algorithm to combine known unspecific biomarkers to dramatically improve specificity for human ovarian cancer detection (the Ova-1 test)².
- Very sensitive assays for both HAPT and CRP in dogs already exist.

Consequently, we began testing both HAPT and CRP levels in canine serum taken from large numbers of dogs diagnosed with lymphoma and other conditions.

Canine Samples.

All samples were obtained from first opinion practices in the UK and were subjected to rigorous follow up. For a dog to be classified as healthy it must be free of lymphoma for a minimum of six months following collection of the blood sample. Healthy samples were also obtained from the Pet Blood Bank where samples were taken from rigorously health screened dogs (generally large breeds), who were donating blood for transfusion. Lymphoma samples were all positive on either FNA or excisional biopsy and only pre-treatment samples were accepted into the study. Other diseases were diagnosed by

relevant procedures.

A sample training set 139 samples comprising dogs diagnosed with lymphoma, healthy dogs and those diagnosed with other diseases was initially used to develop the algorithms using CART analysis. The resultant algorithm was then tested using a further 194 samples also comprising lymphoma, healthy and other disease dogs as illustrated in tables 1 and 2.

Methods.

Assay of Acute Phase Proteins.

1 ml serum samples were collected in serum gel tubes and shipped to the PetScreen laboratory on ice. On arrival, the samples were aliquoted and stored at -20°C until tested.

The samples were tested in batches using commercially available assay kits for HAPT and CRP (Tridelta Development Ltd). All tests were performed in 96 well microplates and read in a standard 96 well microplate photometer.

Generation of Diagnostic Algorithms.

Ciphergen Biomarker Pattern Software (BPS) was used to generate a series of algorithms. BPS is a statistics package which automates the Classification and Regression Tree (CART) procedure for handling complex, multi-parameter data contained in a database. CART was first introduced by Leo Breiman³ *et al* from Berkeley and Stanford in 1984. CART presents results in the form of a "decision tree" which allows highly complex data to be expressed using easy to read diagrams based on "yes/no" questions. CART analysis is now widely used to represent complex data in diverse areas such as medical, marketing, environmental, banking and commercial applications. By providing a convenient means of adding extra parameters to an analysis, CART has the potential to increase the accuracy of the predictions.

BPS looks at the values at each node (i.e. is

the threshold for biomarker 1 greater than or less than X) and automatically varies them in order to produce a decision tree with the greatest analytical power. In the case of biomarker discovery, this is based on providing the software with a training data set which contains spectra from known healthy and diseased individuals. The resultant decision trees are then tested using a second data series to objectively validate each tree's performance. In some cases, it is possible to "prune" the trees to remove nodes which are actually inducing errors. In all cases of CART analysis, the larger the data sets employed, the more accurate the resultant predictions will be.

The algorithm takes the value of one parameter and determines if it is about or below a defined threshold. If it is above the Root Node threshold, the algorithm will show the sample to be positive. If below the threshold, the algorithm will assign data to Node 1, where it determines the value of the second parameter (marker) against a specified threshold. The results are then split again, either giving a negative answer or moving the data onto a third node, where further splitting of the sample is achieved. Through a process of iteration, the software uses the training set data to build decision trees in this manner, up to a point when optimal differentiation between the populations is achieved. At this point the resultant algorithm can then be tested against blinded samples contained in the sample test set.

Sample Training Set.

Initially, the HAPT and CRP results from 139 separate known serum samples comprising healthy, lymphoma and other conditions were fed into Biomarker Pattern Software along with the diagnosis for each sample. This sample cohort represented the "training" set from which the software (through a process of iteration) produces a range of algorithms with different performance characteristics. Eight individual algorithms were developed in this way, each with varying degrees of performance and complexity. The algorithm

which gave the best overall performance was then tested for performance using a second sample series (the Test Set) which were presented blind to the algorithm.

Sample Test Set.

A total of 194 samples were collected for the test set. The composition of samples is shown in table 1:

Condition of Dog	Number in Group
Lymphoma	64
Healthy	51
Other diseases	79

The mean age of all dogs in the test set was 7.9 years. 47.2% were female and 52.8% were male. 60.8% of all the dogs were neutered.

Since these were real world samples obtained from vets who suspected the dog had lymphoma, the other diseases represented those commonly encountered during the differential diagnosis of canine lymphoma. From our experience with the original Lymphoma blood test, the relative proportion of lymphoma (33%), healthy (26%) and other diseases (41%) is typical of the samples received by our laboratory.

The composition of the other diseases group is shown in table 2:

Condition of Dog	Number in Group
Lymphadenopathy including benign lymphatic hyperplasia	34
Other malignancies	11
Benign masses	5
Infection	3
General Malaise/weight loss	6
Hypercalcaemia	4
Hyperthyroidism	3
Inflammatory condition	4
Various other diseases	9

The HAPT and CRP levels were measured in all 194 test set samples. Both values for each sample were then fed into the diagnostic

algorithm which produced a positive or negative result for lymphoma.

Results.

After analysis, the results were compared to the known diagnosis of each sample and each result expressed as TN, TP, FN or FP. From these data, the sensitivity and specificity was determined.

The results for the total 194 samples in the test set are shown in table 3:

Outcome	Number in Group
TP	54
FN	10
TN	121
FP	9
Total	194
Specificity	93%
Sensitivity	85%
NPV	93%
PPV	86%

These results were then further broken down by disease in table 4:

Condition of Dog	TN	FP
Healthy	51	0
Lymphadenopathy including benign lymphatic hyperplasia	29	5
Other malignancies	11	0
Benign masses	3	2
Infection	3	0
General Malaise/weight loss	6	0
Hypercalcaemia	4	0
Hyperthyroidism	3	0
Inflammatory condition	3	1
Various other diseases	8	1
Total	121	9

Discussion.

Finding reliable and accurate means of describing test performance is highly complex. Most performance is described in terms of the

sensitivity and specificity of the test. However, these parameters can be very easily and dramatically affected by the composition of the test population. Unfortunately, levels of sensitivity and specificity are frequently quoted based upon comparisons between the specific disease population and a completely healthy cohort. This approach has the effect of over stating specificity when the test is subject to real world samples which come from a wide range of different diseases. Samples encountered during differential diagnosis can give false positive due results due to cross reactivity and can therefore have a negative effect on the quoted specificity. In veterinary medicine, where good research sample acquisition is widely recognised problem, the low sample numbers often reported can further bias accurate measurement of sensitivity and specificity.

In this study, we have based our results on real world samples, collected under conditions encountered during every day veterinary practice. We have achieved this by conducting extensive follow up on samples to determine the final diagnosis. From this work, we can see that the sample composition used in this study very closely reflects the type of samples routinely submitted during the process of canine lymphoma diagnosis.

With reference to table 4, it can be seen that if only healthy and lymphoma dogs are used in the analysis, this would give 100% specificity from 115 dogs. However, when we introduce the 79 other diseases commonly encountered during lymphoma diagnosis, we get a specificity of 93% from a total population of 194 dogs, with a sensitivity of 85%.

Of particular note with these data, is the very high number of lymphadenopathy samples that were tested. Retaining such a high specificity when 43% of the other diseases came from dogs with non-malignant lymphadenopathies points the way to using the test early in the diagnostic process when a dog presents with swollen lymph nodes. The very high specificity, and same day test

turnaround time, will enable vets to make urgent diagnostic decisions rapidly and confidently when facing a disease which develops so rapidly in dogs.

References.

1. Mischle, R *et al*, 2007. Changes in C-reactive protein and HAPT in dogs with lymphatic neoplasia, *The Veterinary Journal*, 174, 188-192
2. Miller, R. W. *et al*, 2011. Performance of the American College of Obstetricians and Gynecologists' Ovarian Tumor Referral Guidelines with a Multivariate Index Assay. *Obstetrics & Gynecology*: 117, Issue 6, 1298-1306
3. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Chapman & Hall (Wadsworth, Inc.): New York, 1984.