Predictive modeling analysis of genomic regions associated with tumor phenotype in canine non-Hodgkin lymphoma

Predictive modeling analysis using multiple regression analysis was performed for the two genomic regions whose copy number status was associated most significantly with tumor phenotype in canine non-Hodgkins lymphoma (cNHL). For both regions, a contingency table was generated to indicate the copy number status at that locus for all known B-cell lymphoma (BCL) cases (kBCL) and all known T-cell lymphoma (TCL) cases (kTCL). Regression analysis was then performed to model the ability of each locus, in isolation, to classify tumor phenotype accurately, where pBCL indicates 'predicted BCL' and pTCL indicates 'predicted TCL'.

Summary measures of the predictive performance for both models are shown. The results show a decision matrix indicating the number of cases that were classified as pBCL or pTCL using the regression model, compared to the known phenotype status (kBCL or kTCL). Also shown are measures of classification accuracy and sensitivity and specificity, which, respectively, summarize true positive and true negative classification accuracies for each contrast. Multivariate analysis was then performed to determine the predictive ability of both models when used in combination.

Model A

Clone ID	P-value	Dog chromosome location	Human chromosome location	Association
CH82-325C12	2.90E-15	chr11:44256300-44428212	chr9:21971498-22175227	⬆ loss in T

Contingency table:

phenotype	LOSS	NORMAL	GAIN	Total
BCL	0	109	4	113
TCL	20	16	1	37
Total	20	125	5	150

Regression analysis:

Source	SS	df	MS		
Model	9.9612766	1	9.9612766	Number of obs	150
Residual	17.9120567	148	.12102741	F(1, 148)	82.31
Total	27.8733333	149	.187069351	Prob > F	0.0000
				R-squared	0.3574
				Adj R-squared	**0.3530**
				Root MSE	.34789

phenotypei~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ch82325c12~2	.6510638	.0717642	9.07	0.000	.509249	.7928787
_cons	.8184397	.0292976	27.94	0.000	.760544	.8763354

Decision matrix:

Classified	BCL	TCL	Total
not LOSS	113	17	130
LOSS	0	20	20
Total	113	37	150

Summary of predictive performance

pTCL if predicted Pr(kTCL) >= .5

Sensitivity	**Pr(pTCL|kTCL)**	**54.1%**
Specificity	**Pr(pBCL|kBCL)**	**100%**
Positive predictive value	Pr(kTCL|pTCL)	100%
Negative predictive value	Pr(kBCL|pBCL)	86.9%
False positive rate for kTCL	Pr(pBCL|kTCL)	46%
False negative rate for kBCL	Pr(pTCL|kBCL)	0%
False positive rate for pBCL	Pr(kTCL|pBCL)	13.1%
False negative rate for pTCL	Pr(kBCL|pTCL)	0%
Overall accuracy		**88.7%**

Outcome:

Model A indicates that the copy number status of the genomic region encompassed by clone CH82-325C12 (CFA 11; 44.3Mb) explains 35.3% of the variation in tumor phenotype within the cohort (adjusted R-squared value = 0.3530). **This model predicts TCL with 54.1% accuracy (sensitivity), and predicts BCL with 100% accuracy (specificity), with an overall accuracy of 88.7%.**

Model B

Clone ID	P-value	Dog chromosome location	Human chromosome location	Association
CH82-518M17	2.59E-14	chr26:30333217-30501573	chr22:23040362-23163490	⬆ loss in B

Contingency table:

```
phenotype  |       LOSS     NORMAL     GAIN |     Total
-----------+---------------------------------+----------
       BCL |         74         32          7 |       113
       TCL |          0         32          5 |        37
-----------+---------------------------------+----------
     Total |         74         64         12 |       150
```

Regression analysis:

```
Source |       SS       df       MS              Number of obs =     150
-------------+------------------------------           F(  1,   148) =   47.95
       Model |  6.82121319     1  6.82121319           Prob > F      =  0.0000
    Residual |  21.0521201   148  .142244055           R-squared     =  0.2447
-------------+------------------------------           Adj R-squared =  0.2396
       Total |  27.8733333   149  .187069351           Root MSE      =  .37715

------------------------------------------------------------------------------
phenotypei~r |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  ch82518m17 |  -.3361307   .0485394    -6.92   0.000    -.4320506   -.2402109
       _cons |   .6143993   .0367534    16.72   0.000       .54177    .6870286
```

Decision matrix:

```
Classified   |       BCL           TCL  |      Total
-----------+---------------------------+----------
not LOSS     |       106            32  |        138
    LOSS     |         7             5  |         12
-----------+---------------------------+----------
   Total     |       113            37  |        150
```

Summary of predictive performance

```
pBCL if predicted Pr(kBCL) >= .5
--------------------------------------------------
Sensitivity                     Pr(pBCL|kBCL)   93.8%
Specificity                     Pr(pTCL|kTCL)   13.5%
Positive predictive value       Pr(kBCL|pBCL)   76.8%
Negative predictive value       Pr(kTCL|pTCL)   41.7%
--------------------------------------------------
False positive rate for kTCL    Pr(pBCL|kTCL)   86.5%
False negative rate for kBCL    Pr(pTCL|kBCL)    6.2%
False positive rate for pBCL    Pr(kTCL|pBCL)   23.2%
False negative rate for pTCL    Pr(kBCL|pTCL)   58.3%
--------------------------------------------------
Overall accuracy                                74%
```

Outcome:

Model B indicates that the copy number status of the genomic region encompassed by clone CH82-518M17 (CFA 26; 30.4Mb) explains 24% of the variation in tumor phenotype within the cohort (adjusted R-squared value = 0.2396). **This model predicts BCL with 93.8% accuracy (sensitivity), and predicts TCL with 13.5% accuracy (specificity), with an overall accuracy of 74%.**

Multivariate analysis of models A and B

Clone ID	P-value	Dog chromosome location	Human chromosome location	Association
CH82-325C12	2.90E-15	chr11:44256300-44428212	chr9:21971498-22175227	⬆ loss in T
CH82-518M17	2.59E-14	chr26:30333217-30501573	chr22:23040362-23163490	⬆ loss in B

Regression analysis:

```
Logistic regression                               Number of obs     =        150
                                                  LR chi2(2)        =      74.52
                                                  Prob > chi2       =     0.0000
Log likelihood = -46.535779                       Pseudo R-squared  =     0.4447

------------------------------------------------------------------------------
phenotypei~r | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ch82325c12~2 |   64.15195   68.36178     3.90   0.000     7.946029    517.9282
  ch82518m17 |   .1853011   .0810088    -3.86   0.000     .0786602    .4365167
------------------------------------------------------------------------------
```

Decision matrix:

```
Logistic model for phenotypeindicator

-------- True --------
Classified |      BCL            TCL  |      Total
-----------+--------------------------+-----------
not LOSS   |       106             15  |        121
    LOSS   |         7             22  |         29
-----------+--------------------------+-----------
   Total   |       113             37  |        150
```

Summary of predictive performance:

```
pBCL if predicted Pr(kBCL) >= .5
--------------------------------------------------
Sensitivity                     Pr(pBCL|kBCL)   93.8%
Specificity                     Pr(pTCL|kTCL)   59.5%
Positive predictive value       Pr(kBCL|pBCL)   87.6%
Negative predictive value       Pr(kTCL|pTCL)   75.9%
--------------------------------------------------
False positive rate for kTCL    Pr(pBCL|kTCL)   40.5%
False negative rate for kBCL    Pr(pTCL|kBCL)    6.2%
False positive rate for pBCL    Pr(kTCL|pBCL)   12.4%
False negative rate for pTCL    Pr(kBCL|pTCL)   24.1%
--------------------------------------------------
Correctly classified                            85.3%
--------------------------------------------------
```

Outcome:

The results of the multivariate analysis indicate that, when evaluated in combination, the copy number status of the genomic regions encompassed by clones CH82-325C12 (CFA 11; 44.3Mb) and CH82-518M17 (CFA 26; 30.4Mb) explains 44.5% of the variation in tumor phenotype within the cohort (pseudo R-squared value = 0.4447). **This combined model predicts BCL with 93.8% accuracy and predicts TCL with 59.5% accuracy, with an overall accuracy of 85.3%.**