

Supplementary Table 2 Cross-species comparison of mouse HCC and human diffuse large B-cell lymphoma

Predicted subclass	CCP		1NN		3NN		NC		SVM		LDA	
	AB	GCB	AB	GCB	AB	GCB	AB	GCB	AB	GCB	AB	GCB
human DLBCL												
AB-like (<i>n</i> = 21)	16	5	15	6	15	6	16	5	16	5	16	5
GCB-like (<i>n</i> = 21)	3	18	2	19	1	20	2	19	6	15	3	18
percent correctly classified *	81%		81%		83%		83%		74%		81%	
Mouse HCC												
DENA (<i>n</i> = 3)	3	0	1	2	2	1	3	0	3	0	3	0
<i>Myc</i> (<i>n</i> = 8)	5	3	2	6	3	5	6	2	5	3	5	3
<i>E2f1</i> (<i>n</i> = 10)	4	6	3	7	3	7	4	6	4	6	4	6
<i>Myc/E2f1</i> (<i>n</i> = 9)	4	5	5	4	3	6	5	4	6	3	4	5
<i>Myc/Tgfa</i> (<i>n</i> = 9)	5	4	5	4	3	6	5	4	5	4	6	3

*Percentage for correct prediction during leave-one-out cross validation

CCP, Compound Covariate Predictor; **1NN**, 1 Nearest Neighbor; **3NN**, 3 Nearest Neighbor; **NC**, Nearest Centroid; **SVM**, Support Vector Machines; **LDA**, Linear Discriminator Analysis.

Analysis procedure

Previous microarray-based gene expression profiling study demonstrated that diffuse large B-cell lymphoma (DLBCL) divided into two main subgroups; activated B-like (AB) and germinal centre B-like (GCB) subgroups (Reference 23 from main text). AB group showed higher proliferative gene expression signature and poorer prognosis than GCB. Out of 42 DLBCL examined, half of tumors (21 each) were assigned to AB and GCB group respectively based on their hierarchical clustering analysis. Gene expression data were downloaded from Stanford microarray database (<http://genome-www5.stanford.edu>) to compare gene expression pattern with mouse HCC. Before combining two independent data, orthologous genes were selected in both microarray platform and genes were filtered by applying same criteria used previously (genes with less than 30% of missing expression data and a expression ratios of 2 or more in at least 10% of samples were only selected), expression ratios were then standardized as described previously. Total 296 orthologous genes were used in prediction methods. Human DLBCL data were used for training prediction methods and mouse data were assigned to prediction set. Outcome of prediction is shown in table above.