Predicting relapse in medulloblastoma patients by integrating evidence from clinical and genomic features

Tamayo et al

Data Supplement

Table S1. Clinical characteristics of patients in the training and test datasets.

Table S2. Disease-subtype expression signatures used to predict disease-subtype in the test set.

Figure S1. Kaplan-Meier survival plots and ROC curves for the Pomeroy et al. 2002 signature in the training and test sets.

Figure S2. Set of 36 expression signatures used to predict disease-subtype in test samples. Six expression signatures were chosen from the top 10 according to the au-ROC scores for each disease-subtype {c1,…, c6}.

Figure S3. Heat map of the top 20 disease-independent expression signatures that discriminate relapse vs. no-relapse according to the au-ROC in the training dataset. We selected c-Myc as the representative disease- independent expression signatures to be used in the model.

Figure S4. Heat map of the top 20 (bottom 20) disease-dependent pathways that discriminate relapse vs. no-relapse according to the au-ROC for each disease-subtype {c,…, c} in the training dataset. The boxes show the ones we selected to be representative of disease-independent expression signatures to be used in the model.

Figure S5. Kaplan-Meier plots and log-rank test p-values of the associated rank test for the current clinical schema and Models A-D in: A) the training set, and B) the test set. The 95% confidence intervals are shown in blue.

Figure S6. Relative utility curves for Models A-D in the training and test sets. The plots show the relative utility, i.e., the fraction of the utility of perfect prediction that is achieved at the optimal cut point for a risk prediction model, vs. risk threshold. We note that the curves for Models C and D overlap for both the train and test sets.

Figure S7. Full Bayesian nomogram showing all the arms corresponding to disease-subtype dependent expression signatures conditional to the disease-subtype. A specific instance of this nomogram populated with the features of one patient is shown on Fig 3.

Figure S8. Flow-chart of the entire model methodology showing the steps that have to be performed for the analysis of training and test samples. The orange and pink squares represent experimental procedures, and the blue squares represent computational steps. The green squares represent datasets or computational models.

TABLE S1

Attribute	Training Dataset	Test Dataset
Number of Samples	96	78
Sex		
Female	38.5%	30.8%
Male	61.5%	69.2%
Relapse		
No	57.3%	56.4%
Yes	42.7%	43.6%
Status		
Alive	67.7%	66.7%
Deceased	30.2%	33.3%
Unknown	2.1%	-
Average Age (months)	107.4 ± 69.9	105.6 ± 88.3
Histology		
Classic	67.7%	71.8%
Desmoplastic	13.5%	17.9%
LGA (Large Cell Anaplastic)	16.7%	-
Unknown	2.10%	10.3%
Metastasis		
M0	75.0%	66.7%
M1	3.12%	8.97%
M2	2.08%	2.56%
M3	9.38%	12.8%
M4	-	2.18%
M+	2.08%	2.56%
Unknown	8.33%	-
Risk		
Standard	60.4%	65.3%
High	30.2%	28.2%
Unknown	9.40%	6.5%

TABLE S2

Feature	Subtype	Description	Database
Expression Signatures Used to Predict Disease-Subtype			
TRANSLATION_FACTORS	c_1	Translation factor genes (52) GenMAPP	MSigDB v2.5/C2
BRCA_BRCA1_POS	c_1	Genes whose expression is consistently positively correlated with brca1 germline status (107)	MSigDB v2.5/C2
FLUMAZENILPATHWAY	c_1	Flumazenil pathway (9 genes)	MSigDB v2.5/C2
AGUIRRE_PANCREAS_CHR8	c_1	Genes (61) on chr with copy-number-driven expression in pancreatic adenocarcinoma.	MSigDB v2.5/C2
MENSSEN_MYC_UP	c_1	Genes (34) up-regulated by MYC in HUVEC (umbilical vein endothelial cell)	MSigDB v2.5/C2
PENG_RAPAMYCIN_DN	c_1	Genes (229) downregulated in response to rapamycin starvation	MSigDB v2.5/C2
DSRNA_UP/DN	c_2	Upregulated (38) and downregulated (15) by dsRNA (polyI:C) in IFN-null GRE cells	MSigDB v2.5/C2
HINATA_NFKB_DN	c_2	Genes (25) downregulated by NF-kappa B	MSigDB v2.5/C2
P53_DN.v2	c_2	Genes upregulated (150) and downregulated (150) in siRNA knockdown of p53 (GEO dataset GDS1852).	OPAM.v3
IL22BPPATHWAY	c_2	IL-22 (13) is produced by T cells and induces the acute phase inflammatory response in hepatocytes.	MSigDB v2.5/C2
ABBUD_LIF_GH3_DN	c_2	Genes (7) that decreased after LIF treatment of GH3 cells	MSigDB v2.5/C2
NGUYEN_KERATO	c_2	Genes concomitantly upregulated (29) and downregulated (86) by activated Notch1 in mouse and human primary keratinocytes	MSigDB v2.5/C2
POMEROY_DESMOPLASIC_VS_CLASSIC_MD_UP	c_3	Genes (47) expressed in desmoplastic medulloblastomas.	MSigDB v2.5/C2
HSA04340_HEDGEHOG_SIGNALING_PATHWAY	c_3	Genes (57) involved in Hedgehog signaling pathway	MSigDB v2.5/C2
HINATA_NFKB	c_3	Genes upregulated (111) and downregulated (21) by NF-kappa B	MSigDB v2.5/C2
RADAEVA_IFNA_DN	c_3	Genes (10) down-regulated by interferon-alpha in primary hepatocyte	MSigDB v2.5/C2
RORIE_ES_PNET_UP	c_3	Genes (28) showing the greatest increase in expression in NBa Ews/Fli-1 infectants	MSigDB v2.5/C2
SARCOMAS_LIPOSARCOMA	c_3	Genes upregulated (10) and downregulated (11) in liposarcomas, versus other soft-tissue tumors.	MSigDB v2.5/C2
GPCRS_CLASS_C_METABOTROPIC_GLUTAMATE_PHEROMONE	c_4	Genes (14) associated with G-protein coupled receptors related to metabotropic glutamate receptors.	MSigDB v2.5/C2
DFOSB_BRAIN_2WKS_UP	c_4	Genes (39) Up-regulated in nucleus accumbens of mice after 2 weeks of induction of transgenic deltaFosB	MSigDB v2.5/C2
PARKINPATHWAY	c_4	Genes (12) associated with Parkinson's disease	MSigDB v2.5/C2
HSA05020_PARKINSONS_DISEASE	c_4	Genes (17) involved in Parkinson's disease	MSigDB v2.5/C2
MITRPATHWAY	c_4	Genes (9) associated with the MyoD/MEF2 transcription factors	MSigDB v2.5/C2
HSA04080_NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	c_4	Genes (254) involved in neuroactive ligand-receptor interaction	MSigDB v2.5/C2
PHOTO_DN.v1	c_5	Photoreceptor upregulated (150) and downregulated (150) genes (Cepko et al. 2007[1])	OPAM.v3
CRX_DN.v1	c_5	CRX -/- upregulated (150) and downregulated (150) genes (Cepko et al. 2007[1])	OPAM.v3
CtIP_DN.v1	c_5	CtIP -/- upregulated (150) and downregulated (150) genes (GEO dataset GDS2189)	OPAM.v3
NRL_DN.v1	c_5	NRL -/- upregulated (150) and downregulated (150) genes (Cepko et al. 2007[1])	OPAM.v3
WNT_UP.v1	c_5	Genes upregulated (200) and downregulated (200) by wnt (GEO dataset GSE1899).	OPAM.v3
CHREBPPATHWAY	c_5	Genes (20) associated with chREBP, a transcription factor inhibited by cAMP	MSigDB v2.5/C2
NEUROTRANSMITTERSPATHWAY	c_6	Genes (6) associated with biosynthesis of neurotransmitters	MSigDB v2.5/C2
BCAT_UP.v1_UP	c_6	Genes (200) upregulated by beta-catenin (GEO dataset GDS748)	OPAM.v3
ST_WNT_BETA_CATENIN_PATHWAY	c_6	Genes (34) associated with Wnt/Beta-catenin signaling.	MSigDB v2.5/C2
LEF1_UP.v1_UP	c_6	Genes (200) upregulated by Lef1 (GEO dataset GSE3229)	OPAM.v3
TGFBETA_LATE_UP	c_6	Genes (33) upregulated by TGF-beta treatment of skin fibroblasts only at 1-4 hrs (clusters 4-6)	MSigDB v2.5/C2
UVB_NHEK3_C6	c_6	Genes (30) regulated by UV-B light in normal human epidermal keratinocytes, cluster 6	MSigDB v2.5/C2

1. Corbo JC, Myers CA, Lawrence KA, et al: A typology of photoreceptor gene expression patterns in the mouse. Proc Natl Acad Sci U S A 104:12069-74, 2007

Figure S1

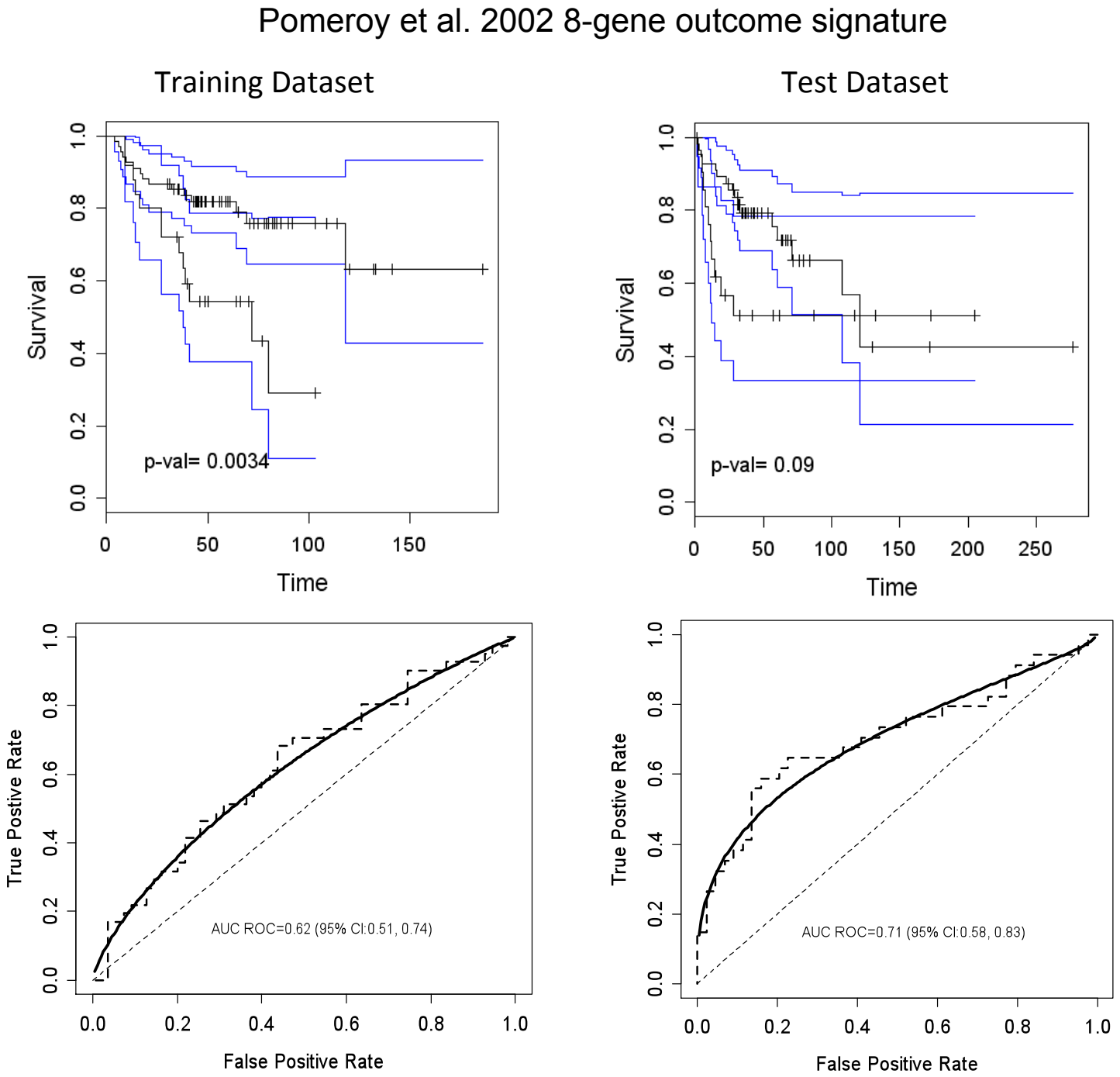

Pomeroy et al. 2002 8-gene outcome signature
Training Dataset
Test Dataset
Survival
Time
p-val= 0.0034
p-val= 0.09
True Postive Rate
False Positive Rate
AUC ROC=0.62 (95% CI:0.51, 0.74)
AUC ROC=0.71 (95% CI:0.58, 0.83)

Figure S2

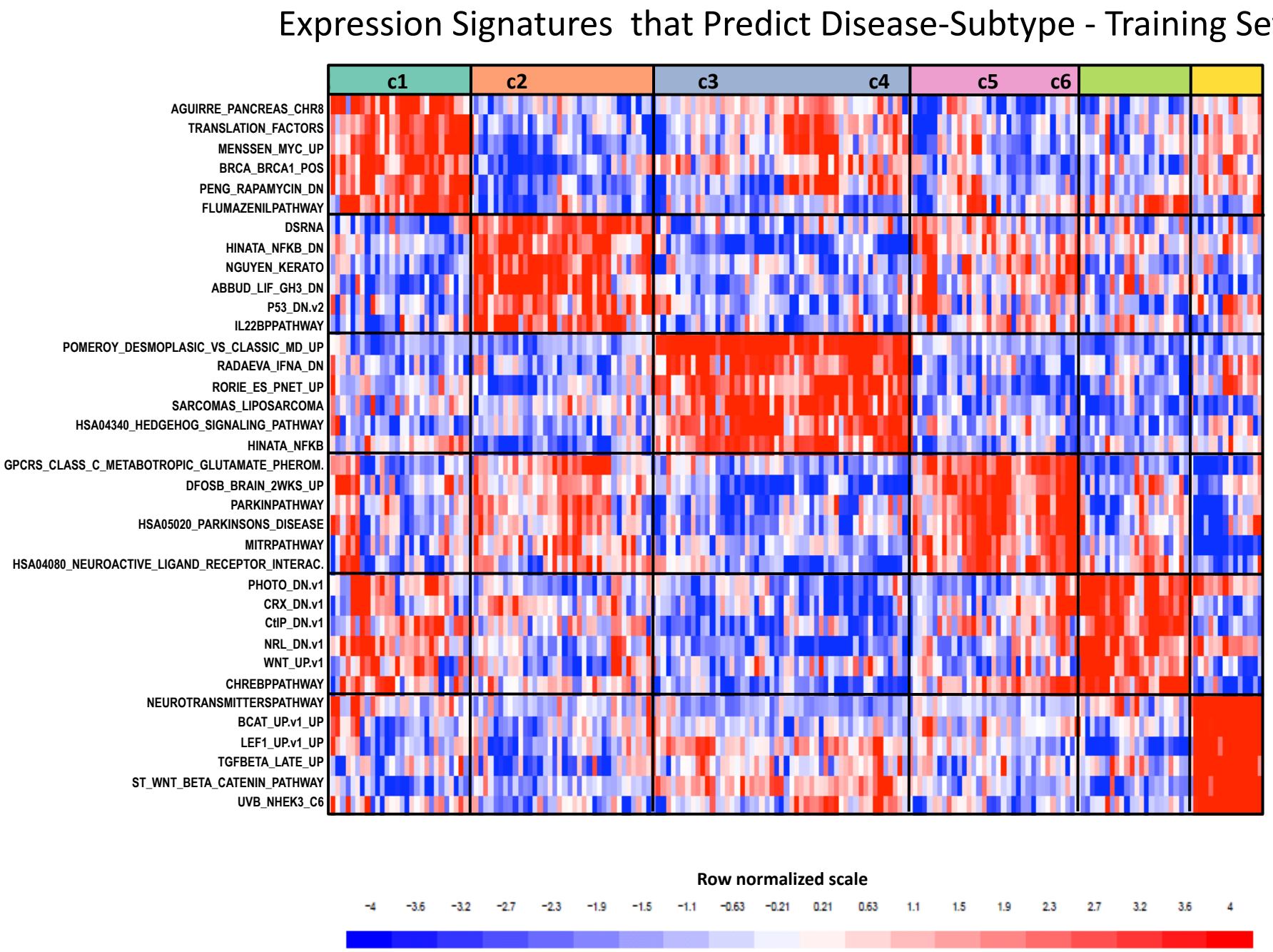

Expression Signatures that Predict Disease-Subtype - Training Set
c1
c2
c3
c4
c5
c6
AGUIRRE_PANCREAS_CHR8
TRANSLATION_FACTORS
MENSSEN_MYC_UP
BRCA_BRCA1_POS
PENG_RAPAMYCIN_DN
FLUMAZENILPATHWAY
DSRNA
HINATA_NFKB_DN
NGUYEN_KERATO
ABBUD_LIF_GH3_DN
P53_DN.v2
IL22BPPATHWAY
POMEROY_DESMOPLASIC_VS_CLASSIC_MD_UP
RADAEVA_IFNA_DN
RORIE_ES_PNET_UP
SARCOMAS_LIPOSARCOMA
HSA04340_HEDGEHOG_SIGNALING_PATHWAY
HINATA_NFKB
GPCRS_CLASS_C_METABOTROPIC_GLUTAMATE_PHEROM.
DFOSB_BRAIN_2WKS_UP
PARKINPATHWAY
HSA05020_PARKINSONS_DISEASE
MITRPATHWAY
HSA04080_NEUROACTIVE_LIGAND_RECEPTOR_INTERAC.
PHOTO_DN.v1
CRX_DN.v1
CtIP_DN.v1
NRL_DN.v1
WNT_UP.v1
CHREBPPATHWAY
NEUROTRANSMITTERSPATHWAY
BCAT_UP.v1_UP
LEF1_UP.v1_UP
TGFBETA_LATE_UP
ST_WNT_BETA_CATENIN_PATHWAY
UVB_NHEK3_C6
Row normalized scale
-4
-3.6
-3.2
-2.7
-2.3
-1.9
-1.5
-1.1
-0.63
-0.21
0.21
0.63
1.1
1.5
1.9
2.3
2.7
3.2
3.6
4

Figure S3

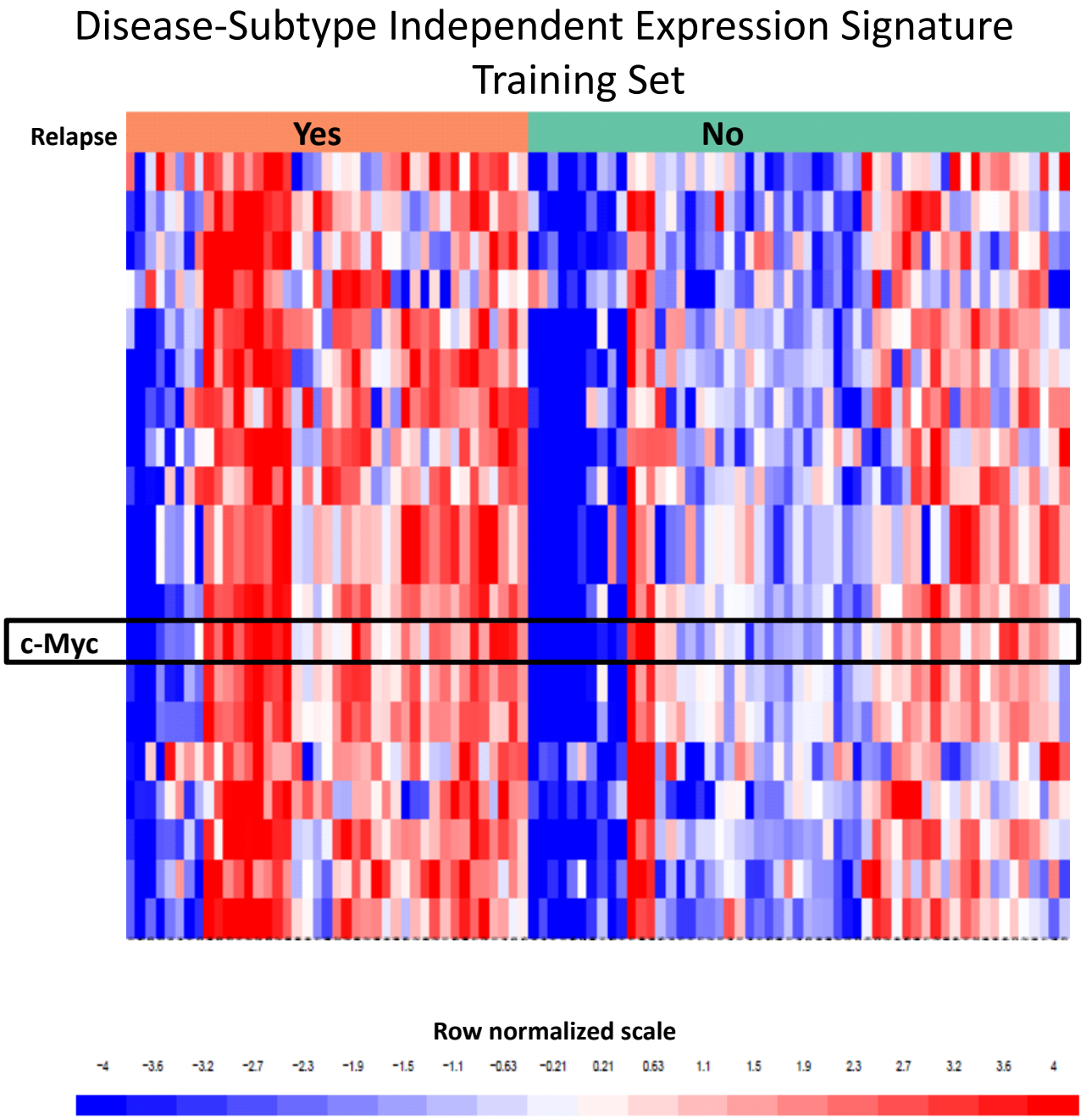

Disease-Subtype Independent Expression Signature
Training Set
Relapse
Yes
No
c-Myc
Row normalized scale
-4
-3.6
-3.2
-2.7
-2.3
-1.9
-1.5
-1.1
-0.63
-0.21
0.21
0.63
1.1
1.5
1.9
2.3
2.7
3.2
3.6
4

Figure S4

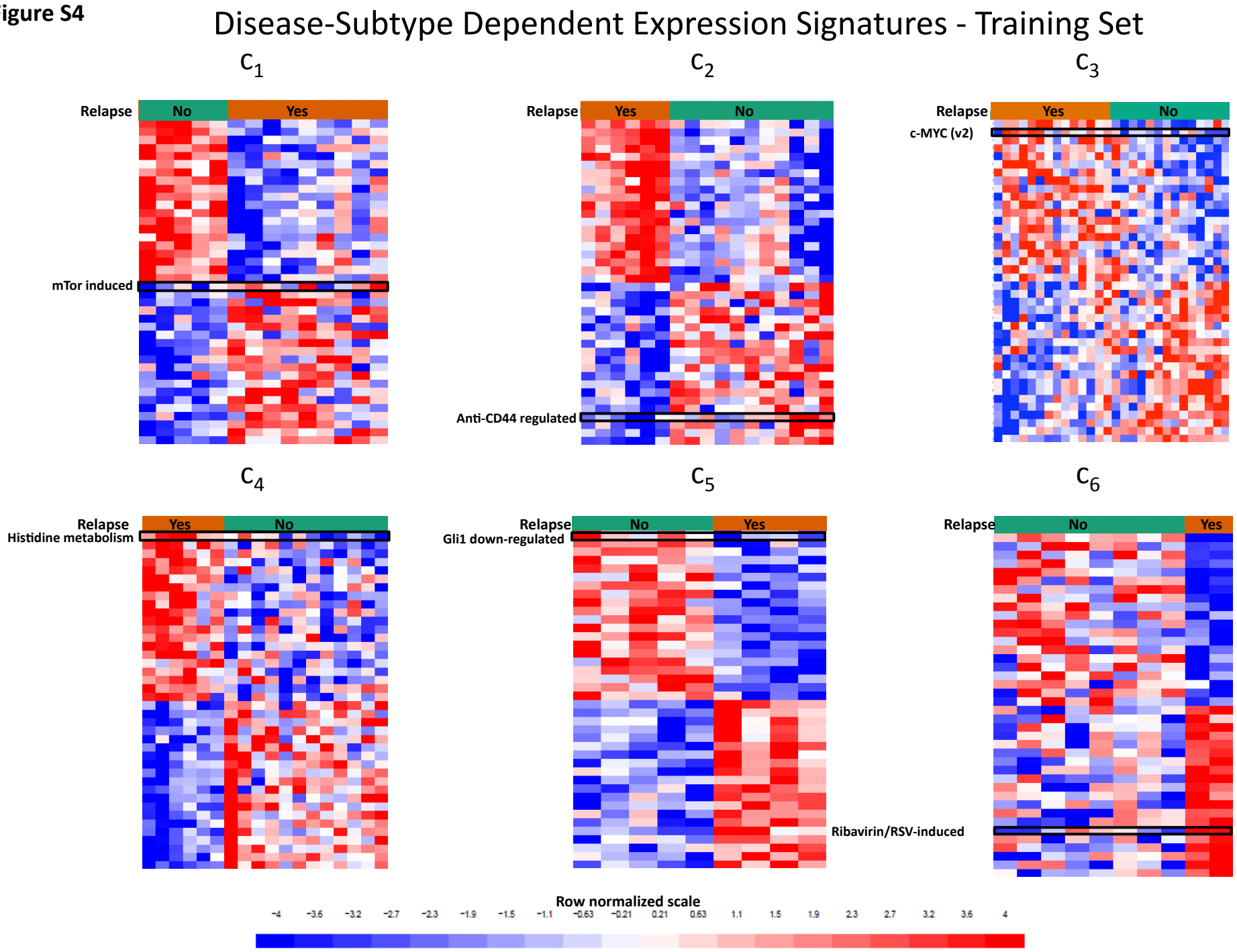

Disease-Subtype Dependent Expression Signatures - Training Set
C1
Relapse
No
Yes
mTor induced
C2
Relapse
Yes
No
Anti-CD44 regulated
C3
Relapse
Yes
No
c-MYC (v2)
C4
Relapse
Yes
No
Histidine metabolism
C5
Relapse
No
Yes
Gli1 down-regulated
C6
Relapse
No
Yes
Ribavirin/RSV-induced
Row normalized scale
-4
-3.6
-3.2
-2.7
-2.3
-1.9
-1.5
-1.1
-0.63
-0.21
0.21
0.63
1.1
1.5
1.9
2.3
2.7
3.2
3.6
4

Figure S5-A

Training Set

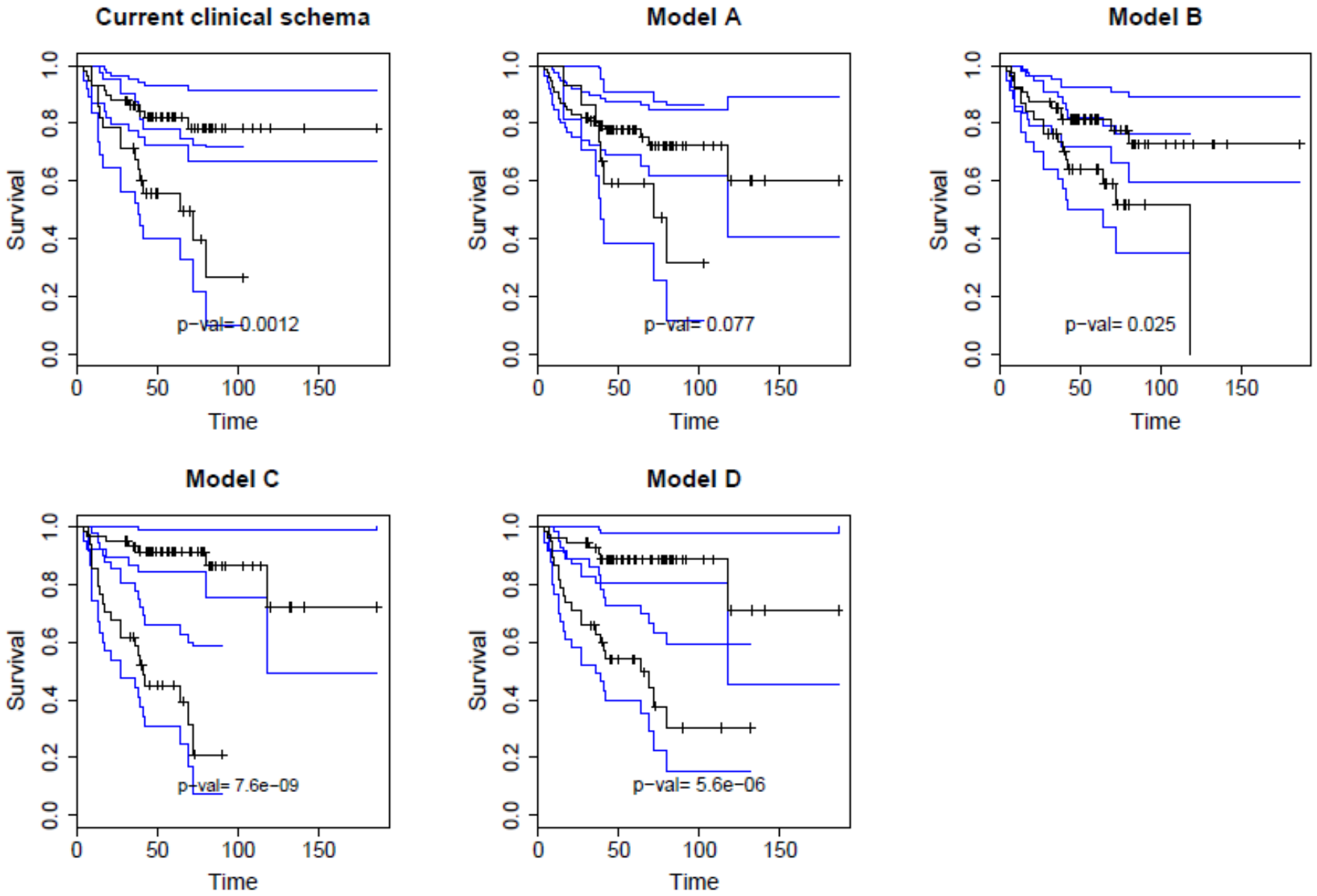


Figure S5-B

Test Set

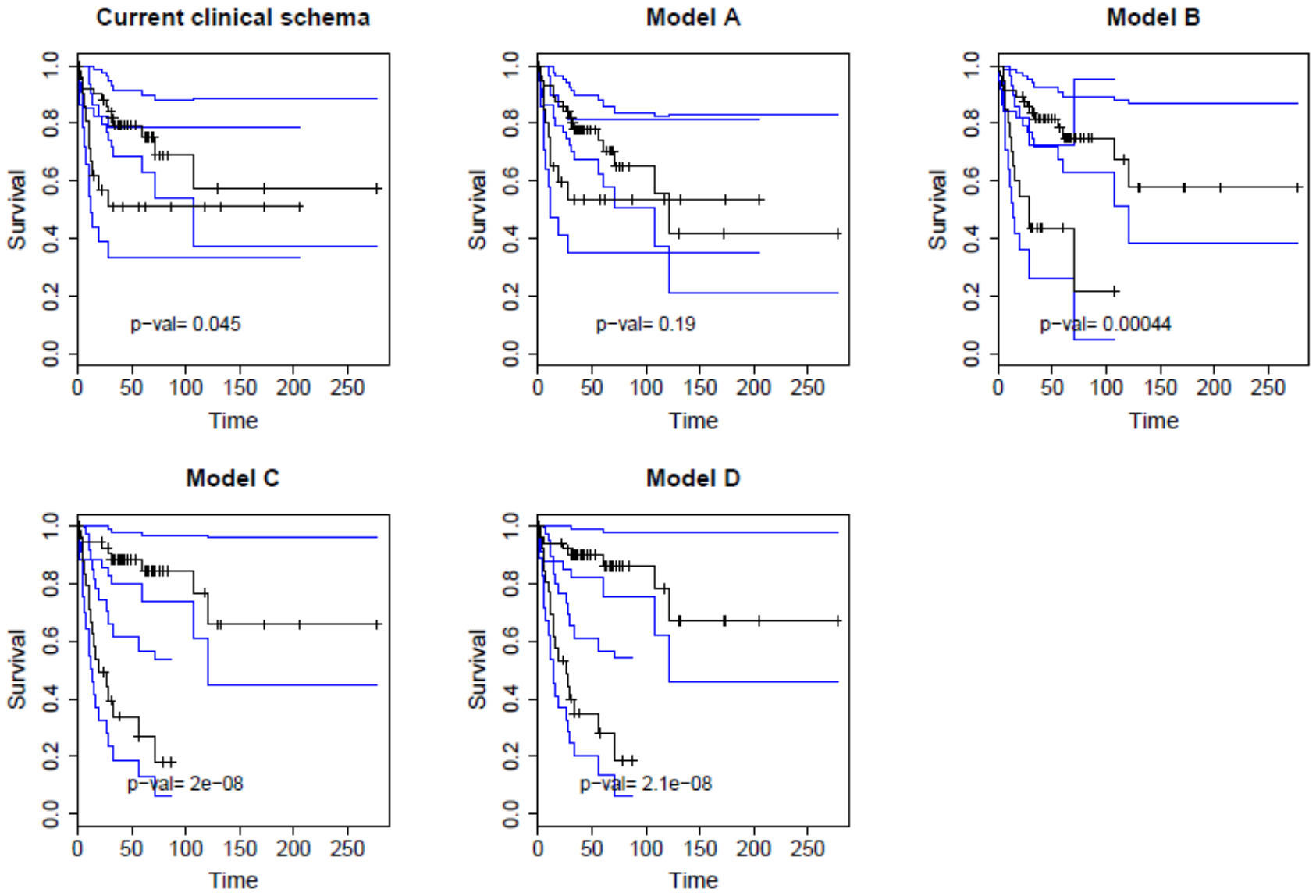


Figure S6

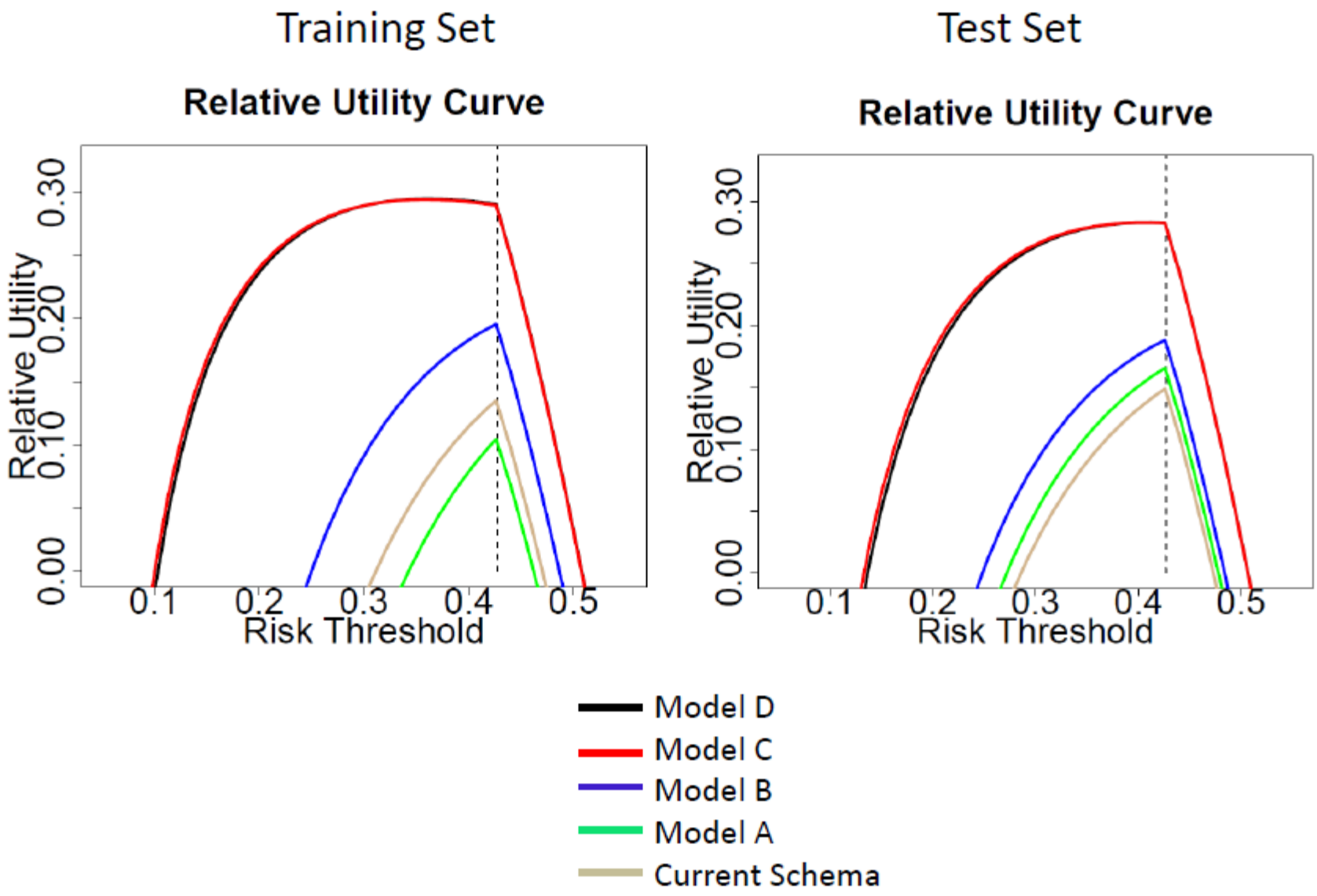

Training Set
Relative Utility Curve
Relative Utility
0.00
0.10
0.20
0.30
0.1
0.2
0.3
0.4
0.5
Risk Threshold
Test Set
Relative Utility Curve
Relative Utility
0.00
0.10
0.20
0.30
0.1
0.2
0.3
0.4
0.5
Risk Threshold
Model D
Model C
Model B
Model A
Current Schema

Figure S7

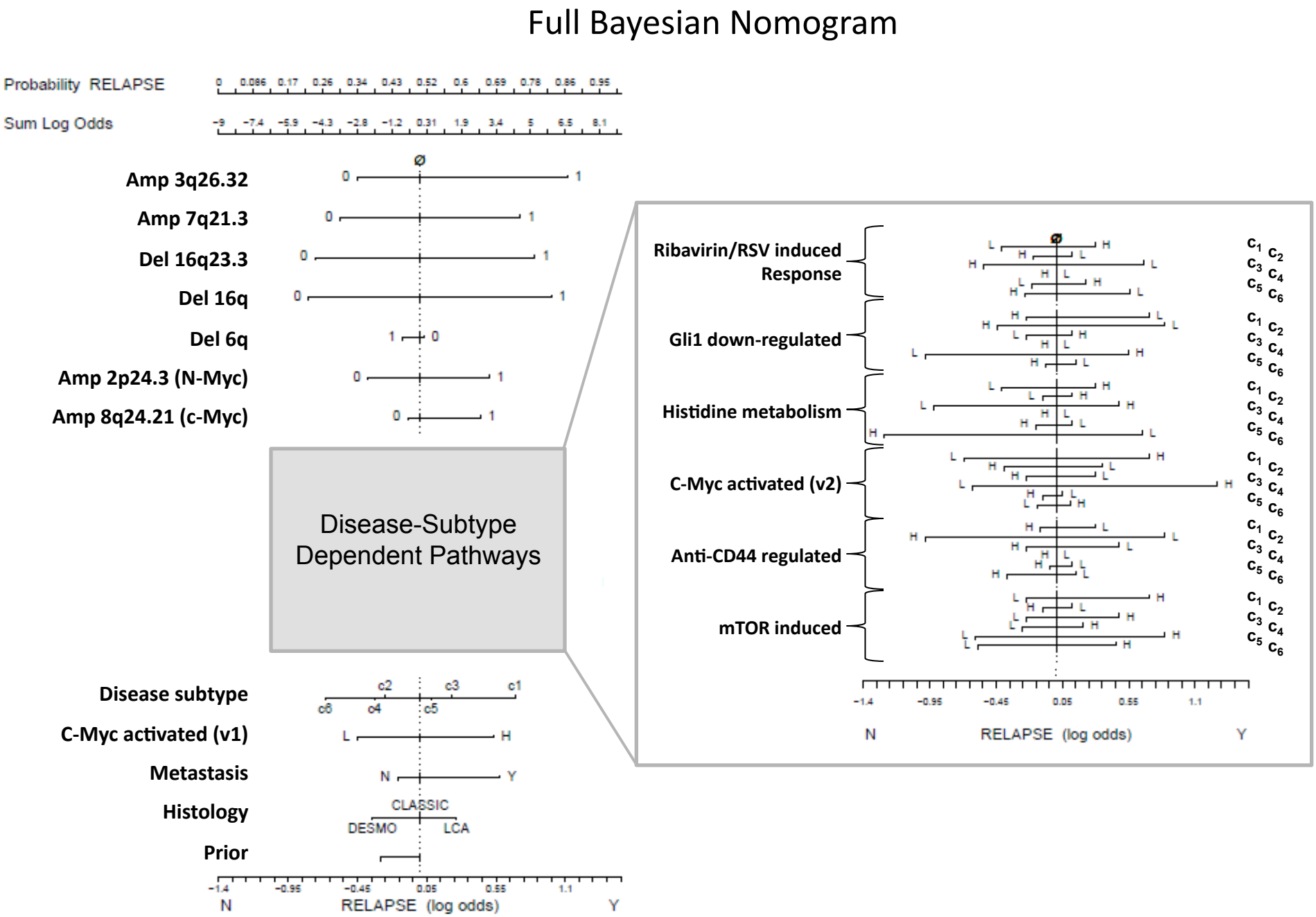
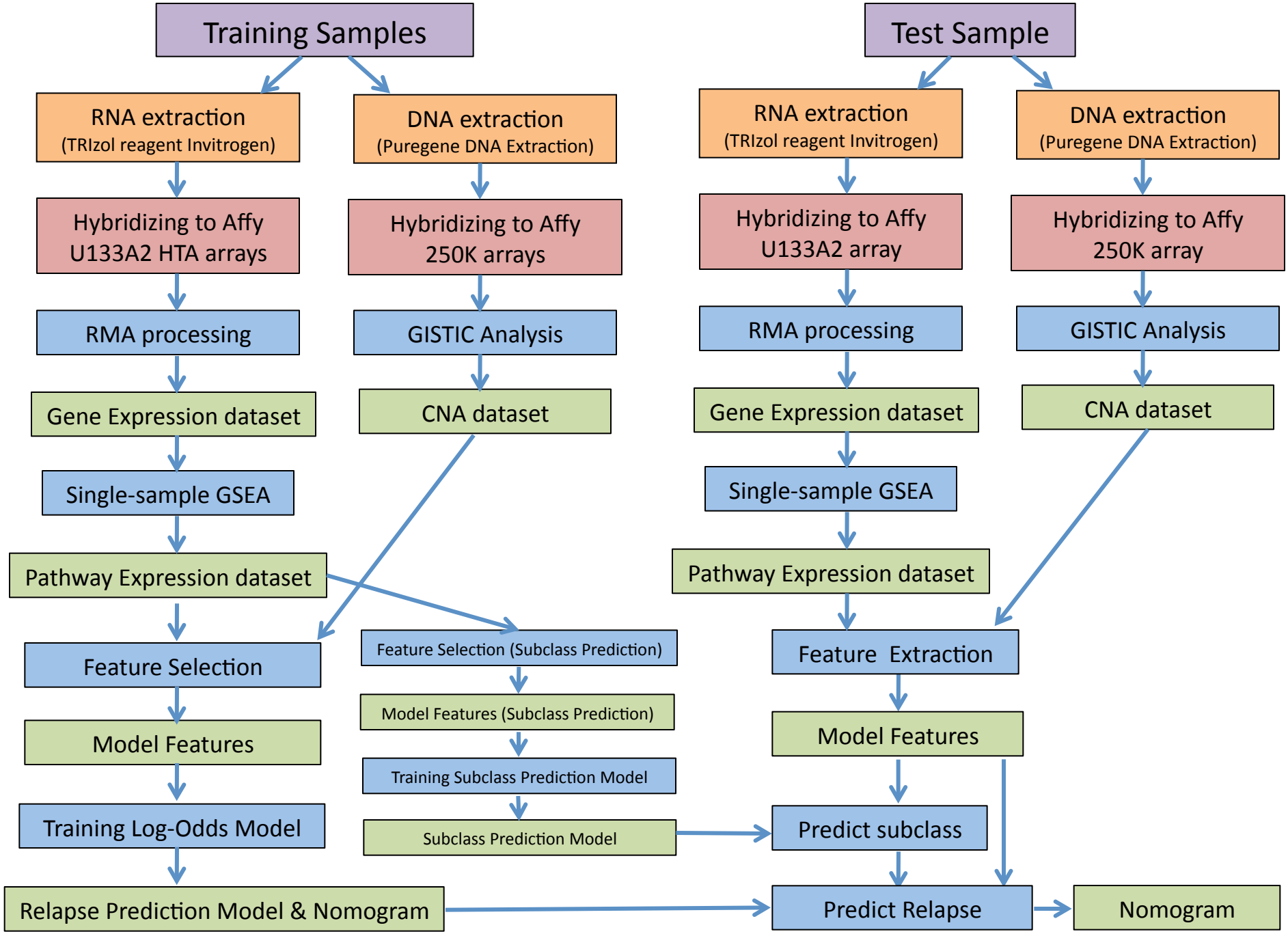

Full Bayesian Nomogram
Probability RELAPSE
0 0.086 0.17 0.26 0.34 0.43 0.52 0.6 0.69 0.78 0.86 0.95
Sum Log Odds
-9 -7.4 -5.9 -4.3 -2.8 -1.2 0.31 1.9 3.4 5 6.5 8.1
Amp 3q26.32
Amp 7q21.3
Del 16q23.3
Del 16q
Del 6q
Amp 2p24.3 (N-Myc)
Amp 8q24.21 (c-Myc)
Disease-Subtype Dependent Pathways
Disease subtype
C-Myc activated (v1)
Metastasis
Histology
Prior
c2 c3 c1 c6 c4 c5
CLASSIC DESMO LCA
-1.4 -0.95 -0.45 0.05 0.55 1.1
N RELAPSE (log odds) Y
Ribavirin/RSV induced Response
Gli1 down-regulated
Histidine metabolism
C-Myc activated (v2)
Anti-CD44 regulated
mTOR induced
C1 C2 C3 C4 C5 C6
-1.4 -0.95 -0.45 0.05 0.55 1.1
N RELAPSE (log odds) Y

Figure S8


Training Samples
RNA extraction
(TRIzol reagent Invitrogen)
DNA extraction
(Puregene DNA Extraction)
Hybridizing to Affy
U133A2 HTA arrays
Hybridizing to Affy
250K arrays
RMA processing
GISTIC Analysis
Gene Expression dataset
CNA dataset
Single-sample GSEA
Pathway Expression dataset
Feature Selection
Model Features
Training Log-Odds Model
Relapse Prediction Model & Nomogram
Feature Selection (Subclass Prediction)
Model Features (Subclass Prediction)
Training Subclass Prediction Model
Subclass Prediction Model
Test Sample
RNA extraction
(TRIzol reagent Invitrogen)
DNA extraction
(Puregene DNA Extraction)
Hybridizing to Affy
U133A2 array
Hybridizing to Affy
250K array
RMA processing
GISTIC Analysis
Gene Expression dataset
CNA dataset
Single-sample GSEA
Pathway Expression dataset
Feature Extraction
Model Features
Predict subclass
Predict Relapse
Nomogram