- Supplement S2 -
MERIDA: Results

1 Runtime analysis

The runtime analysis was conducted to show how the runtime can be reduced using MERIDA in comparison to LOBICO, which ultimately allows us to fit larger models and iteratively apply our method using the results of previous runs of our method. Moreover, we could demonstrate that the usage of a different weight function can additionally decrease the runtime for both methods.
First, we will explain how we obtained the data set used for the runtime experiment. Then, we will discuss how the experiment was conducted and present the results.

1.1 Data set generation

We performed the runtime analysis on the Rapamycin drug data set and the mutation data of GDSC Version 6.1. Since LOBICO as well as MERIDA need binarized input and output features, we will in the following explain how these can be generated.
The drug data set contains logarithmized IC50 values for approximately 350 samples. For binarization of the continuous values, we employed the threshold provided by the publication of Iorio et al. (Iorio *et al.*, 2016) in which this version of the GDSC data set was introduced. They calculated the threshold with the method described in Knijnenburg et al. (Knijnenburg *et al.*, 2016), which is the same method that we use to calculate the thresholds for the newer version of the GDSC data set.
The mutation data is given as single mutations of specific genes. For this analysis, we decided to establish a gene-wide mutation status, i.e. we call a gene mutated (feature value of 1) iff at least one mutation is present and not mutated (feature value of 0) otherwise. This means that we do not distinguish between different mutations or classes of mutations for this analysis.
Since we wanted to investigate which influence the number of input features has on the time needed to fit a model, we used the IntOGen cancer driver list (Gonzalez-Perez *et al.*, 2013) to filter the genes. In particular, we compile 16 data sets with a varying feature set size from 25 to 400 features by a stepsize of 25. We prioritize the genes through the frequency of mutations within a gene as provided by IntOGen.

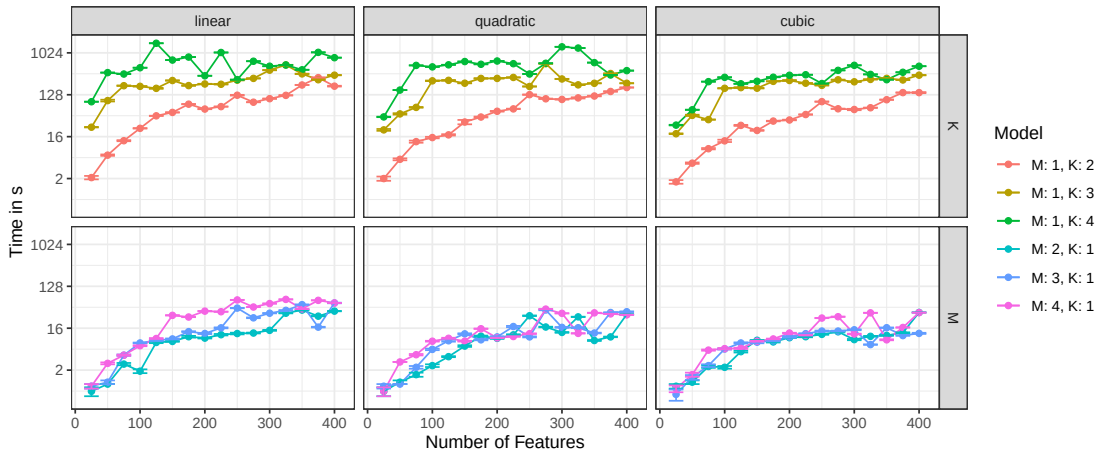


Figure 1: In this figure, we present the runtime of LOBICO with varying hyperparameters M and K for the three different weight functions.

1.2 Runtime analysis additional results

We tested MERIDA and LOBICO with different hyperparameters (see Table 1), different numbers of input features, and the three different weight functions. Note that for LOBICO we could not fit larger models with both $K > 1$ and $M > 1$ in a reasonable amount of time. We performed the computations on a compute server with four Intel(R) Xeon(R) CPU E5-4657L v2 processors with 2.40GHz clock rate. IBM ILOG CPLEX Optimization Studio V12.6.2 for C++ was employed for formulating and solving the ILP. CPLEX was run using 32 cores and a deterministic parallel mode.

We could observe that using MERIDA instead of LOBICO leads to a considerably

Table 1: Parameters

MERIDA parameters	LOBICO parameters
$M \in \{2, 4, 8\}$	$(K, M) \in \{(1, 1), (2, 1), (1, 2), (2, 2), (1, 4), (4, 1)\}$

reduction in runtime (cf. Table 2). Moreover, considering the quadratic or cubic weight function instead of the linear one does also reduce runtime significantly (cf. Table 3). For LOBICO, we note that the two parameters K and M seem to influence the runtime differently, i.e. according to our analyses for different K and M (see Figure 1), it can be seen that the model parameter K increases the runtime more strongly than the model parameter M.

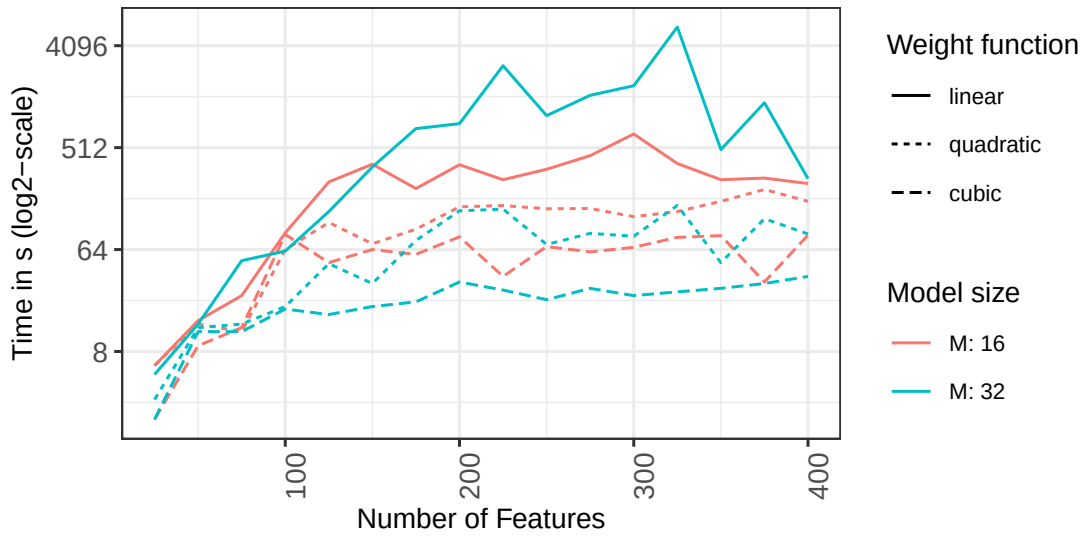


Figure 2: In this figure, we present the runtime of MERIDA for $M = 16$ and $M = 32$ for the different weight functions.

Table 2: Runtime comparison of a four-feature model from LOBICO versus a four-feature model from MERIDA. The values depict the factor by which MERIDA is faster than LOBICO using its original linear weight function. To obtain a four-feature model with LOBICO, three different parameter settings can be used ($(K, M) \in \{(1, 4), (4, 1), (2, 2)\}$).

	MERIDA linear	MERIDA quadratic	MERIDA cubic
$(K, M) = (1, 4)$	3.61	4.85	6
$(K, M) = (4, 1)$	71.79	98.54	125.31
$(K, M) = (2, 2)$	641.97	876.74	1147.6
Mean of the above three rows	239.13	326.71	426.31

Table 3: The table gives the runtime advantage of using a quadratic or cubic weight function instead of using a linear weight function. The values depict the factor by which a model with a quadratic or cubic weight function is faster than a model build with the linear weight function. The given factor is calculated by averaging over all tested numbers of features. The number in brackets are the model parameters. For LOBICO, the first number is the value for K and the second number is the value for M.

	MERIDA (4)	LOBICO $(1, 4)$	LOBICO $(4, 1)$	LOBICO $(2, 2)$
quadratic	1.35	2.25	1.38	2.88
cubic	1.65	2.49	2.55	5.38

Table 4: This table gives the mean and median runtime of MERIDA across all drugs for the different settings, weight functions, and parameter combinations in the pan-cancer analysis. The mean (median) is calculated by summing up the runtime of the cross-validation and the final fit for one specific drug and then averaging across all drugs.

Setting	Weight function	Model size	Mean runtime (in s)	Median runtime (in s)
Setting 1	cubic	1	5.3170	6
Setting 1	cubic	2	15.7073	10
Setting 1	cubic	3	109.6829	96
Setting 1	cubic	4	314.45	265
Setting 1	cubic	5	1128.3076	963
Setting 1	cubic	6	4970.3421	3589.5
Setting 1	quadratic	1	4.6097	6
Setting 1	quadratic	2	17.8780	11
Setting 1	quadratic	3	114.9024	107
Setting 1	quadratic	4	315.6666	241
Setting 1	quadratic	5	1583.2631	1051
Setting 1	quadratic	6	7657.0277	4104
Setting 1	linear	1	4.1463	5
Setting 1	linear	2	26.2195	15
Setting 1	linear	3	128.4390	120
Setting 1	linear	4	409	215
Setting 1	linear	5	2219.1842	1434.5
Setting 1	linear	6	9679.0285	7176
Setting 2	cubic	1	0	0
Setting 2	cubic	2	3.9090	5
Setting 2	cubic	3	9.9090	9
Setting 2	cubic	4	68.6363	62
Setting 2	cubic	5	137.3636	147
Setting 2	cubic	6	428.5454	322
Setting 2	quadratic	1	0	0
Setting 2	quadratic	2	2.3636	1
Setting 2	quadratic	3	11.3636	10
Setting 2	quadratic	4	79.3636	77
Setting 2	quadratic	5	167.7272	174
Setting 2	quadratic	6	695.3636	438
Setting 2	linear	1	0	0
Setting 2	linear	2	2.3636	1
Setting 2	linear	3	24.6363	16
Setting 2	linear	4	102.3636	93
Setting 2	linear	5	376.1818	220
Setting 2	linear	6	1854.9090	1168
Setting 3	cubic	1	3.8181	3

Setting 3	cubic	2	12.9090	12
Setting 3	cubic	3	105.3636	109
Setting 3	cubic	4	265.1818	230
Setting 3	cubic	5	1100.8181	701
Setting 3	cubic	6	4942.1818	2544
Setting 3	quadratic	1	4.8181	6
Setting 3	quadratic	2	15.9090	12
Setting 3	quadratic	3	107	124
Setting 3	quadratic	4	273.5454	221
Setting 3	quadratic	5	1600.6363	1381
Setting 3	quadratic	6	9044.1818	6007
Setting 3	linear	1	3.7272	4
Setting 3	linear	2	28.7272	16
Setting 3	linear	3	114	99
Setting 3	linear	4	452	179
Setting 3	linear	5	2950.4545	1332
Setting 3	linear	6	8749	7584

Table 5: This table gives the mean and median runtime of MERIDA across all drugs for the different settings, weight functions, and parameter combinations in the non-haematological cancer cell line analysis. The mean is calculated by summing up the runtime of the cross-validation and the final fit for one specific drug and then averaging across all drugs.

Setting	Weight function	Model size	Mean runtime (in s)	Median runtime (in s)
Setting 1	cubic	1	1.2926	1
Setting 1	cubic	2	12.9024	10
Setting 1	cubic	3	88.7317	62
Setting 1	cubic	4	233.9268	172
Setting 1	cubic	5	774.1707	476
Setting 1	cubic	6	2433.5609	1759
Setting 1	quadratic	1	1.5853	1
Setting 1	quadratic	2	14.4146	10
Setting 1	quadratic	3	70.1463	62
Setting 1	quadratic	4	188.7560	151
Setting 1	quadratic	5	622.6097	445
Setting 1	quadratic	6	2753.7560	1930
Setting 1	linear	1	2.0487	1
Setting 1	linear	2	23.9024	12
Setting 1	linear	3	67.4634	62
Setting 1	linear	4	177.4146	127
Setting 1	linear	5	700.3170	480
Setting 1	linear	6	5473.6829	2277

Setting 2	cubic	1	0	0
Setting 2	cubic	2	1.4545	1
Setting 2	cubic	3	7.0909	7
Setting 2	cubic	4	40.8181	32
Setting 2	cubic	5	90.1818	79
Setting 2	cubic	6	195	145
Setting 2	quadratic	1	0	0
Setting 2	quadratic	2	0.54545	0
Setting 2	quadratic	3	11.0909	9
Setting 2	quadratic	4	54.5454	59
Setting 2	quadratic	5	110.1818	100
Setting 2	quadratic	6	261.1818	133
Setting 2	linear	1	0	0
Setting 2	linear	2	0.7272	0
Setting 2	linear	3	19.4545	10
Setting 2	linear	4	61.1818	54
Setting 2	linear	5	150.3636	101
Setting 2	linear	6	742.8181	232
Setting 3	cubic	1	0.7272	0
Setting 3	cubic	2	10.9090	9
Setting 3	cubic	3	66.2727	68
Setting 3	cubic	4	179.7272	192
Setting 3	cubic	5	498.9090	435
Setting 3	cubic	6	1724.1818	1539
Setting 3	quadratic	1	1.0909	0
Setting 3	quadratic	2	13.2727	8
Setting 3	quadratic	3	65.9090	63
Setting 3	quadratic	4	166.5454	168
Setting 3	quadratic	5	581.3636	458
Setting 3	quadratic	6	2644.7272	2203
Setting 3	linear	1	1	0
Setting 3	linear	2	23.9090	12
Setting 3	linear	3	67	55
Setting 3	linear	4	225.4545	130
Setting 3	linear	5	900.3636	553
Setting 3	linear	6	9887.0909	2176

Table 6: This table gives the mean and median runtime of MERIDA across all drugs for the different settings, weight functions, and parameter combinations in the haematological cancer cell line analysis. The mean is calculated by summing up the runtime of the cross-validation and the final fit for one specific drug and then averaging across all drugs.

Setting	Weight function	Model size	Mean runtime (in s)	Median runtime (in s)
Setting 1	cubic	1	0	0
Setting 1	cubic	2	0.0243	0
Setting 1	cubic	3	2.1707	1
Setting 1	cubic	4	10.5853	11
Setting 1	cubic	5	20.4878	23
Setting 1	cubic	6	29.3170	31
Setting 1	quadratic	1	0	0
Setting 1	quadratic	2	0.0487	0
Setting 1	quadratic	3	2.9756	2
Setting 1	quadratic	4	11.7804	12
Setting 1	quadratic	5	20.4878	22
Setting 1	quadratic	6	27.3414	27
Setting 1	linear	1	0	0
Setting 1	linear	2	0.2195	0
Setting 1	linear	3	4.7073	4
Setting 1	linear	4	14.3414	16
Setting 1	linear	5	19.9756	21
Setting 1	linear	6	26.5365	27
Setting 2	cubic	1	0	0
Setting 2	cubic	2	0	0
Setting 2	cubic	3	0	0
Setting 2	cubic	4	1.3636	1
Setting 2	cubic	5	8.4545	10
Setting 2	cubic	6	18	21
Setting 2	quadratic	1	0	0
Setting 2	quadratic	2	0	0
Setting 2	quadratic	3	0	0
Setting 2	quadratic	4	1.5454	1
Setting 2	quadratic	5	10	11
Setting 2	quadratic	6	16.0909	20
Setting 2	linear	1	0	0
Setting 2	linear	2	0	0
Setting 2	linear	3	0.0909	0
Setting 2	linear	4	2.2727	2
Setting 2	linear	5	9.4545	9
Setting 2	linear	6	18.6363	22
Setting 3	cubic	1	0	0
Setting 3	cubic	2	0.0909	0
Setting 3	cubic	3	1.3636	1
Setting 3	cubic	4	9.3636	12
Setting 3	cubic	5	19.8181	21
Setting 3	cubic	6	30.4545	35

Setting 3	quadratic	1	0	0
Setting 3	quadratic	2	0	0
Setting 3	quadratic	3	1.9090	1
Setting 3	quadratic	4	11.6363	13
Setting 3	quadratic	5	16.8181	18
Setting 3	quadratic	6	25.2727	28
Setting 3	linear	1	0	0
Setting 3	linear	2	0	0
Setting 3	linear	3	2.9090	2
Setting 3	linear	4	11.2727	13
Setting 3	linear	5	18.0909	19
Setting 3	linear	6	24.3636	25

Table 7: This table provides the runtime (CV + final fit) of LOBICO for each of the 10 analyzed drugs. In total, LOBICO needed approximately 4.5 days (total runtime) for fitting the shown models. For the calculation of the total runtime, we excluded all models where either a single fit in the cross-validation or the final fitting procedure was higher than 6 hours. Note that these models were also not completely fit. Additionally, we did not fit models for specific parameter combinations for the quadratic and linear weight function if the runtime for the cubic weight function already exceeded this threshold. Similarly, we did not fit a model for the linear weight function if the quadratic weight function exceeded this threshold.

Drug	Setting	Weight function	Model size	Runtime (in s)
AZD8055	Setting 1	cubic	1, 1	27
AZD8055	Setting 1	cubic	1, 2	1036
AZD8055	Setting 1	cubic	2, 1	137
AZD8055	Setting 1	cubic	2, 2	42971
AZD8055	Setting 1	quadratic	1, 1	29
AZD8055	Setting 1	quadratic	1, 2	3251
AZD8055	Setting 1	quadratic	2, 1	143
AZD8055	Setting 1	quadratic	2,2	$>$ 21 600
AZD8055	Setting 1	linear	1, 1	28
AZD8055	Setting 1	linear	1, 2	1916
AZD8055	Setting 1	linear	2, 1	136
AZD8055	Setting 1	linear	2,2	-
Omipalisib	Setting 1	cubic	1, 1	27
Omipalisib	Setting 1	cubic	1, 2	1412
Omipalisib	Setting 1	cubic	2, 1	135
Omipalisib	Setting 1	cubic	2, 2	$>$ 21 600
Omipalisib	Setting 1	quadratic	1, 1	28
Omipalisib	Setting 1	quadratic	1, 2	2507

Omipalisib	Setting 1	quadratic	2, 1	137
Omipalisib	Setting 1	quadratic	2,2	-
Omipalisib	Setting 1	linear	1, 1	31
Omipalisib	Setting 1	linear	1, 2	7317
Omipalisib	Setting 1	linear	2, 1	9530
Omipalisib	Setting 1	linear	2, 2	-
Voxtalisib	Setting 1	cubic	1, 1	28
Voxtalisib	Setting 1	cubic	1, 2	421
Voxtalisib	Setting 1	cubic	2, 1	145
Voxtalisib	Setting 1	cubic	2, 2	35849
Voxtalisib	Setting 1	quadratic	1, 1	26
Voxtalisib	Setting 1	quadratic	1, 2	1887
Voxtalisib	Setting 1	quadratic	2, 1	920
Voxtalisib	Setting 1	quadratic	2,2	> 21 600
Voxtalisib	Setting 1	linear	1, 1	29
Voxtalisib	Setting 1	linear	1, 2	2916
Voxtalisib	Setting 1	linear	2, 1	10353
Voxtalisib	Setting 1	linear	2, 2	-
NSC319726	Setting 1	cubic	1, 1	36
NSC319726	Setting 1	cubic	1, 2	8696
NSC319726	Setting 1	cubic	2, 1	3121
NSC319726	Setting 1	cubic	2, 2	> 21 600
NSC319726	Setting 1	quadratic	1, 1	26
NSC319726	Setting 1	quadratic	1, 2	6859
NSC319726	Setting 1	quadratic	2, 1	8616
NSC319726	Setting 1	quadratic	2, 2	-
NSC319726	Setting 1	linear	1, 1	25
NSC319726	Setting 1	linear	1, 2	49149
NSC319726	Setting 1	linear	2, 1	16323
NSC319726	Setting 1	linear	2,2	-
Niraparib	Setting 1	cubic	1, 1	24
Niraparib	Setting 1	cubic	1, 2	708
Niraparib	Setting 1	cubic	2, 1	103
Niraparib	Setting 1	cubic	2, 2	> 21 600
Niraparib	Setting 1	quadratic	1, 1	26
Niraparib	Setting 1	quadratic	1, 2	1428
Niraparib	Setting 1	quadratic	2, 1	102
Niraparib	Setting 1	quadratic	2,2	-
Niraparib	Setting 1	linear	1, 1	23
Niraparib	Setting 1	linear	1, 2	2499
Niraparib	Setting 1	linear	2, 1	168
Niraparib	Setting 1	linear	2, 2	-
Temsirolimus	Setting 3	cubic	1, 1	34
Temsirolimus	Setting 3	cubic	1, 2	2545

Temsirolimus	Setting 3	cubic	2, 1	176
Temsirolimus	Setting 3	cubic	2,2	> 21 600
Temsirolimus	Setting 3	quadratic	1, 1	26
Temsirolimus	Setting 3	quadratic	1, 2	3338
Temsirolimus	Setting 3	quadratic	2, 1	2901
Temsirolimus	Setting 3	quadratic	2, 2	-
Temsirolimus	Setting 3	linear	1, 1	27
Temsirolimus	Setting 3	linear	1, 2	5280
Temsirolimus	Setting 3	linear	2, 1	15451
Temsirolimus	Setting 3	linear	2, 2	-
CX-5461	Setting 3	cubic	1, 1	26
CX-5461	Setting 3	cubic	1, 2	2212
CX-5461	Setting 3	cubic	2, 1	515
CX-5461	Setting 3	cubic	2, 2	> 21 600
CX-5461	Setting 3	quadratic	1, 1	31
CX-5461	Setting 3	quadratic	1, 2	7170
CX-5461	Setting 3	quadratic	2, 1	5794
CX-5461	Setting 3	quadratic	2, 2	-
CX-5461	Setting 3	linear	1, 1	28
CX-5461	Setting 3	linear	1, 2	31725
CX-5461	Setting 3	linear	2, 1	21984
CX-5461	Setting 3	linear	2, 2	-
Rapamycin	Setting 3	cubic	1, 1	25
Rapamycin	Setting 3	cubic	1, 2	44
Rapamycin	Setting 3	cubic	2, 1	115
Rapamycin	Setting 3	cubic	2, 2	9440
Rapamycin	Setting 3	quadratic	1, 1	23
Rapamycin	Setting 3	quadratic	1, 2	337
Rapamycin	Setting 3	quadratic	2, 1	111
Rapamycin	Setting 3	quadratic	2, 2	20823
Rapamycin	Setting 3	linear	1, 1	30
Rapamycin	Setting 3	linear	1, 2	1927
Rapamycin	Setting 3	linear	2, 1	110
Rapamycin	Setting 3	linear	2, 2	> 21 600
Dactolisib	Setting 3	cubic	1, 1	24
Dactolisib	Setting 3	cubic	1, 2	217
Dactolisib	Setting 3	cubic	2, 1	107
Dactolisib	Setting 3	cubic	2, 2	11968
Dactolisib	Setting 3	quadratic	1, 1	25
Dactolisib	Setting 3	quadratic	1, 2	1181
Dactolisib	Setting 3	quadratic	2, 1	110
Dactolisib	Setting 3	quadratic	2, 2	15602
Dactolisib	Setting 3	linear	1, 1	24
Dactolisib	Setting 3	linear	1, 2	1900

Dactolisib	Setting 3	linear	2, 1	119
Dactolisib	Setting 3	linear	2, 2	> 21 600
Talazoparib	Setting 3	cubic	1, 1	24
Talazoparib	Setting 3	cubic	1, 2	1078
Talazoparib	Setting 3	cubic	2, 1	131
Talazoparib	Setting 3	cubic	2,2	> 21 600
Talazoparib	Setting 3	quadratic	1, 1	23
Talazoparib	Setting 3	quadratic	1, 2	956
Talazoparib	Setting 3	quadratic	2, 1	125
Talazoparib	Setting 3	quadratic	2,2	-
Talazoparib	Setting 3	linear	1, 1	23
Talazoparib	Setting 3	linear	1, 2	1284
Talazoparib	Setting 3	linear	2, 1	5258
Talzoparib	Setting 3	linear	2, 2	-

Table 8: This table provides the runtime (CV + final fit) of MERIDA for each of the 10 drugs analyzed in the comparison between LOBICO and MERIDA. In total, MERIDA needed approximately 3 days for fitting the shown models. No models were excluded for the calculation of the total runtime.

Drug	Setting	Weight function	Model size	Runtime (in s)
AZD8055	Setting 1	cubic	1	7
AZD8055	Setting 1	cubic	2	10
AZD8055	Setting 1	cubic	3	137
AZD8055	Setting 1	cubic	4	327
AZD8055	Setting 1	cubic	5	1234
AZD8055	Setting 1	cubic	6	4742
AZD8055	Setting 1	quadratic	1	6
AZD8055	Setting 1	quadratic	2	9
AZD8055	Setting 1	quadratic	3	115
AZD8055	Setting 1	quadratic	4	256
AZD8055	Setting 1	quadratic	5	1043
AZD8055	Setting 1	quadratic	6	4188
AZD8055	Setting 1	linear	1	6
AZD8055	Setting 1	linear	2	22
AZD8055	Setting 1	linear	3	143
AZD8055	Setting 1	linear	4	260
AZD8055	Setting 1	linear	5	1581
AZD8055	Setting 1	linear	6	8986
Omipalisib	Setting 1	cubic	1	6
Omipalisib	Setting 1	cubic	2	14
Omipalisib	Setting 1	cubic	3	53
Omipalisib	Setting 1	cubic	4	203

Omipalisib	Setting 1	cubic	5	317
Omipalisib	Setting 1	cubic	6	1228
Omipalisib	Setting 1	quadratic	1	5
Omipalisib	Setting 1	quadratic	2	19
Omipalisib	Setting 1	quadratic	3	133
Omipalisib	Setting 1	quadratic	4	212
Omipalisib	Setting 1	quadratic	5	728
Omipalisib	Setting 1	quadratic	6	2838
Omipalisib	Setting 1	linear	1	4
Omipalisib	Setting 1	linear	2	57
Omipalisib	Setting 1	linear	3	217
Omipalisib	Setting 1	linear	4	1221
Omipalisib	Setting 1	linear	5	7524
Omipalisib	Setting 1	linear	6	39933
Voxtalisib	Setting 1	cubic	1	6
Voxtalisib	Setting 1	cubic	2	8
Voxtalisib	Setting 1	cubic	3	99
Voxtalisib	Setting 1	cubic	4	219
Voxtalisib	Setting 1	cubic	5	1160
Voxtalisib	Setting 1	cubic	6	6790
Voxtalisib	Setting 1	quadratic	1	6
Voxtalisib	Setting 1	quadratic	2	22
Voxtalisib	Setting 1	quadratic	3	117
Voxtalisib	Setting 1	quadratic	4	241
Voxtalisib	Setting 1	quadratic	5	818
Voxtalisib	Setting 1	quadratic	6	5589
Voxtalisib	Setting 1	linear	1	7
Voxtalisib	Setting 1	linear	2	41
Voxtalisib	Setting 1	linear	3	119
Voxtalisib	Setting 1	linear	4	165
Voxtalisib	Setting 1	linear	5	610
Voxtalisib	Setting 1	linear	6	4254
NSC319726	Setting 1	cubic	1	3
NSC319726	Setting 1	cubic	2	6
NSC319726	Setting 1	cubic	3	32
NSC319726	Setting 1	cubic	4	93
NSC319726	Setting 1	cubic	5	153
NSC319726	Setting 1	cubic	6	470
NSC319726	Setting 1	quadratic	1	4
NSC319726	Setting 1	quadratic	2	12
NSC319726	Setting 1	quadratic	3	93
NSC319726	Setting 1	quadratic	4	135
NSC319726	Setting 1	quadratic	5	467
NSC319726	Setting 1	quadratic	6	1734

NSC319726	Setting 1	linear	1	1
NSC319726	Setting 1	linear	2	30
NSC319726	Setting 1	linear	3	125
NSC319726	Setting 1	linear	4	235
NSC319726	Setting 1	linear	5	1436
NSC319726	Setting 1	linear	6	8995
Niraparib	Setting 1	cubic	1	1
Niraparib	Setting 1	cubic	2	3
Niraparib	Setting 1	cubic	3	15
Niraparib	Setting 1	cubic	4	65
Niraparib	Setting 1	cubic	5	134
Niraparib	Setting 1	cubic	6	317
Niraparib	Setting 1	quadratic	1	0
Niraparib	Setting 1	quadratic	2	5
Niraparib	Setting 1	quadratic	3	28
Niraparib	Setting 1	quadratic	4	79
Niraparib	Setting 1	quadratic	5	142
Niraparib	Setting 1	quadratic	6	372
Niraparib	Setting 1	linear	1	1
Niraparib	Setting 1	linear	2	7
Niraparib	Setting 1	linear	3	52
Niraparib	Setting 1	linear	4	117
Niraparib	Setting 1	linear	5	201
Niraparib	Setting 1	linear	6	1100
Temsirolimus	Setting 3	cubic	1	7
Temsirolimus	Setting 3	cubic	2	16
Temsirolimus	Setting 3	cubic	3	122
Temsirolimus	Setting 3	cubic	4	315
Temsirolimus	Setting 3	cubic	5	1955
Temsirolimus	Setting 3	cubic	6	13281
Temsirolimus	Setting 3	quadratic	1	6
Temsirolimus	Setting 3	quadratic	2	40
Temsirolimus	Setting 3	quadratic	3	184
Temsirolimus	Setting 3	quadratic	4	330
Temsirolimus	Setting 3	quadratic	5	3298
Temsirolimus	Setting 3	quadratic	6	19885
Temsirolimus	Setting 3	linear	1	4
Temsirolimus	Setting 3	linear	2	89
Temsirolimus	Setting 3	linear	3	185
Temsirolimus	Setting 3	linear	4	889
Temsirolimus	Setting 3	linear	5	4035
Temsirolimus	Setting 3	linear	6	31163
CX-5461	Setting 3	cubic	1	5
CX-5461	Setting 3	cubic	2	7

CX-5461	Setting 3	cubic	3	109
CX-5461	Setting 3	cubic	4	198
CX-5461	Setting 3	cubic	5	701
CX-5461	Setting 3	cubic	6	2544
CX-5461	Setting 3	quadratic	1	6
CX-5461	Setting 3	quadratic	2	18
CX-5461	Setting 3	quadratic	3	168
CX-5461	Setting 3	quadratic	4	357
CX-5461	Setting 3	quadratic	5	2725
CX-5461	Setting 3	quadratic	6	6852
CX-5461	Setting 3	linear	1	6
CX-5461	Setting 3	linear	2	51
CX-5461	Setting 3	linear	3	235
CX-5461	Setting 3	linear	4	2132
CX-5461	Setting 3	linear	5	14777
Rapamycin	Setting 3	cubic	1	3
Rapamycin	Setting 3	cubic	2	12
Rapamycin	Setting 3	cubic	3	186
Rapamycin	Setting 3	cubic	4	367
Rapamycin	Setting 3	cubic	5	1873
Rapamycin	Setting 3	cubic	6	6656
Rapamycin	Setting 3	quadratic	1	1
Rapamycin	Setting 3	quadratic	2	8
Rapamycin	Setting 3	quadratic	3	126
Rapamycin	Setting 3	quadratic	4	385
Rapamycin	Setting 3	quadratic	5	1757
Rapamycin	Setting 3	quadratic	6	12140
Rapamycin	Setting 3	linear	1	1
Rapamycin	Setting 3	linear	2	12
Rapamycin	Setting 3	linear	3	101
Rapamycin	Setting 3	linear	4	225
Rapamycin	Setting 3	linear	5	1302
Rapamycin	Setting 3	linear	6	7584
Dactolisib	Setting 3	cubic	1	1
Dactolisib	Setting 3	cubic	2	5
Dactolisib	Setting 3	cubic	3	54
Dactolisib	Setting 3	cubic	4	185
Dactolisib	Setting 3	cubic	5	338
Dactolisib	Setting 3	cubic	6	1719
Dactolisib	Setting 3	quadratic	1	6
Dactolisib	Setting 3	quadratic	2	6
Dactolisib	Setting 3	quadratic	3	81
Dactolisib	Setting 3	quadratic	4	197
Dactolisib	Setting 3	quadratic	5	528

Dactolisib	Setting 3	quadratic	6	2469
Dactolisib	Setting 3	linear	1	6
Dactolisib	Setting 3	linear	2	14
Dactolisib	Setting 3	linear	3	107
Dactolisib	Setting 3	linear	4	179
Dactolisib	Setting 3	linear	5	1045
Dactolisib	Setting 3	linear	6	3689
Talazoparib	Setting 3	cubic	1	3
Talazoparib	Setting 3	cubic	2	5
Talazoparib	Setting 3	cubic	3	21
Talazoparib	Setting 3	cubic	4	94
Talazoparib	Setting 3	cubic	5	172
Talazoparib	Setting 3	cubic	6	317
Talazoparib	Setting 3	quadratic	1	6
Talazoparib	Setting 3	quadratic	2	7
Talazoparib	Setting 3	quadratic	3	27
Talazoparib	Setting 3	quadratic	4	84
Talazoparib	Setting 3	quadratic	5	174
Talazoparib	Setting 3	quadratic	6	365
Talazoparib	Setting 3	linear	1	5
Talazoparib	Setting 3	linear	2	12
Talazoparib	Setting 3	linear	3	68
Talazoparib	Setting 3	linear	4	131
Talazoparib	Setting 3	linear	5	323
Talazoparib	Setting 3	linear	6	3916

1.3 Iterative feature selection

The results of our runtime analysis clearly show that MERIDA has a significantly reduced runtime compared to LOBICO. However, since in our real performance analysis, the number of input features is considerably higher (1500 features), the number of possible logic combinations becomes very large and the parameter M had to be chosen carefully to balance between the runtime of the analysis and the expressiveness of the Boolean formulas. Whereas MERIDA could still identify models with up to M = 10 features in an acceptable amount of time, LOBICO could mostly only handle 2 features. In order to simultaneously increase the parameter M and the number of input features for MERIDA without compromising the performance, we tested an iterative approach. To this end, we used the data sets for OSI-027, PIK-93, and Voxtalisib, which we prepared as described in Supplement S1. Then we randomly drew 400 features for each of the drugs and use these as input features for the analyses described in the following. For each of the three drugs, we trained models with $M \in \{1, \ldots, 16\}$ in one-shot, i.e. without any iterative selection. In addition, we tested an iterative procedure: we first selected the best model among the models with $M = 1$ to $M = 6$ using Youden's J. Then, we iteratively increased M by 4 putative features and selected the best model using Youden's J until we could fit a model with $M = 16$. Figure 3 shows the results of this analysis for the cubic weight function. It can be seen, that the overall runtime (including all necessary cross-validation steps) can be reduced by a factor of 25 on average. Moreover, the selected feature sets for the two approaches resemble each other significantly (Fisher's exact test p-value < 0.05). Based on these tests, we decided to perform the iterative analysis in the main manuscript similarly: we initially fit models with M in the range of 1 to 6 and increase M by 4 putative features in each iteration.

OSI-027

M	No iteration	Iteration	P-value
2	Sensitive: SPOP_high_expr, TJP1_low_expr	Sensitive: SPOP_high_expr, TJP1_low_expr	—
3	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: CSNK2A1_high_expr	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: CSNK2A1_high_expr	1.25e-05
4	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: FAF1_low_expr, DIS3_low_expr	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: CSNK2A1_high_expr, PSIP1_low_expr	0.00044
5	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: FAF1_low_expr, DIS3_low_expr, PSIP1_low_expr	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr	9.37e-06
6	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: FAF1_low_expr, DIS3_low_expr, PSIP1_low_expr, HLF_high_expr	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr	4.31e-10
7	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr Resistant: FAF1_low_expr, DIS3_low_expr, HLF_high_expr, ATRX_low_expr	Sensitive: SPOP_high_expr, TJP1_low_expr Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown	1.44e-06
8	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr Resistant: FAF1_low_expr, DIS3_low_expr, HLF_high_expr, PSIP1_low_expr, HSP90AA1_Unknown	Sensitive: SPOP_high_expr, TJP1_low_expr, FMR1_Unknown Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown	3.69e-08
9	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr Resistant: FAF1_low_expr, DIS3_low_expr, HLF_high_expr, PSIP1_low_expr, HSP90AA1_Unknown, ERBB3_Unknown	Sensitive: SPOP_high_expr, TJP1_low_expr, FMR1_Unknown Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown, EP300_low_expr	1.84e-07
10	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr Resistant: FAF1_low_expr, HLF_high_expr, PSIP1_low_expr, HSP90AA1_Unknown, ERBB3_Unknown, CAD_high_expr, PLXNB2_Unknown	Sensitive: SPOP_high_expr, TJP1_low_expr, FMR1_Unknown Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown, EP300_low_expr, CDK4_high_expr	7.77e-09

M	No iteration	Iteration	P-value
12	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr Resistant: FAF1_low_expr, HLF_high_expr, PSIP1_low_expr, HSP90AA1_Unknown, ERBB3_Unknown, CAD_high_expr, PLXNB2_Unknown, HSP90AA1_low_expr, CDK4_high_expr	Sensitive: SPOP_high_expr, TJP1_low_expr, FMR1_Unknown Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown, EP300_low_expr, CDK4_high_expr, CAD_high_expr, HSP90AA1_Unknown	7.18e-14
14	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr, CHEK2_low_expr Resistant: FAF1_low_expr, HLF_high_expr, PSIP1_low_expr, HSP90AA1_Unknown, ERBB3_Unknown, CAD_high_expr, PLXNB2_Unknown, CDK4_high_expr, RTN4_high_expr, EZH2_low_expr	Sensitive: SPOP_high_expr, TJP1_low_expr, FMR1_Unknown, CHEK2_low_expr Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown, EP300_low_expr, CDK4_high_expr, CAD_high_expr, HSP90AA1_Unknown, RTN4_high_expr	1.41e-16
16	Sensitive: SPOP_high_expr, TJP1_low_expr, ARID1B_high_expr, CHEK2_low_expr, SRN1_high_expr Resistant: RTN4_high_expr, ARFGAP2_low_expr, HSP90AA1_Unknown, FAF1_low_expr, CAD_high_expr, PSIP1_low_expr, PLXNB2_Unknown, CTTN_high_expr, CDK4_high_expr, ACSL6_high_expr, ERBB3_Unknown	Sensitive: SPOP_high_expr, TJP1_low_expr, FMR1_Unknown, CHEK2_low_expr Resistant: CSNK2A1_high_expr, PSIP1_low_expr, HLF_high_expr, FAF1_low_expr, PLXNB2_Unknown, EP300_low_expr, CDK4_high_expr, CAD_high_expr, HSP90AA1_Unknown, RTN4_high_expr, HSP90AA1_low_expr, PPP2R1A_low_expr	2.28e-12
Total Time	33135 s	1144 s	

PIK-93

M	No iteration	Iteration	
2	sensitive: PTPRF_low_expr, CTTN_low_expr	sensitive: PTPRF_low_expr, CTTN_low_expr	—
5	sensitive: NTRK1_CNVgain, PPM1D_Unknown, NCKAP1_low_expr, MED23_high_expr, CTTN_low_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain	9.37e-06
8	sensitive: FRG1_high_expr, NTRK1_CNVgain, ELF1_high_expr, PPM1D_Unknown, NCKAP1_low_expr, CTTN_low_expr resistant: DICER1_low_expr, MMP2_high_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr	1.41e-10
9	sensitive: FRG1_high_expr, NTRK1_CNVgain, ELF1_high_expr, PPM1D_Unknown, NCKAP1_low_expr, CTTN_low_expr resistant: DICER1_low_expr, MMP2_high_expr, PLXNA1_high_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr, DLG1_low_expr	1.26e-09
10	sensitive: ING1_low_expr, NTRK1_CNVgain, PPM1D_Unknown, NCKAP1_low_expr, MED23_high_expr, CTTN_low_expr resistant: DICER1_low_expr, TBX3_high_expr, NTN4_high_expr, ATRX_low_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr, DLG1_low_expr, IRF6_high_expr	7.23e-07
11	sensitive: ING1_low_expr, NTRK1_CNVgain, PPM1D_Unknown, NCKAP1_low_expr, MED23_high_expr, CTTN_low_expr resistant: ANK3_high_expr, DICER1_low_expr, TBX3_high_expr, NTN4_high_expr, ATRX_low_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr, DLG1_low_expr, IRF6_high_expr, MYD88_Unknown	2.37e-06
12	sensitive: HDAC3_low_expr, NCKAP1_low_expr, MED23_high_expr, CTTN_low_expr, ASXL1_Unknown resistant: TBX3_high_expr, ANK3_high_expr, MED12_low_expr, CAD_low_expr, IRF6_high_expr, FXR1_low_expr, ERBB3_high_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr, DLG1_low_expr, IRF6_high_expr, MYD88_Unknown, ATRX_low_expr	0.0039

M	No iteration	Iteration	
13	sensitive: HDAC3_low_expr, NCKAP1_low_expr, MED23_high_expr, CTTN_low_expr, ASXL1_Unknown, MNDA_high_expr resistant: TBX3_high_expr, ANK3_high_expr, MED12_low_expr, CAD_low_expr, IRF6_high_expr, FXR1_low_expr, ERBB3_high_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr, DLG1_low_expr, IRF6_high_expr, MYD88_Unknown, ATRX_low_expr, NRAS_high_expr	0.0063
16	sensitive: ATR_low_expr, ROBO2_high_expr, FRG1_high_expr, NTRK1_CNVgain, FAT1_low_expr, CTTN_low_expr resistant: AMER1_low_expr, PCDH18_high_expr, DLG1_low_expr, IRF6_high_expr, TBX3_high_expr, FN1_high_expr, MYH10_Unknown, ACACA_low_expr, ATRX_low_expr	sensitive: PTPRF_low_expr, CTTN_low_expr, CCND1_high_expr, PPM1D_Unknown, NTRK1_CNVgain, FAT1_low_expr, NCKAP1_low_expr resistant: DICER1_low_expr, MMP2_high_expr, DLG1_low_expr, IRF6_high_expr, MYD88_Unknown, ATRX_low_expr, ELF1_low_expr, ACACA_low_expr, PCDH18_high_expr	9.43e-09
Total Time	35446 s	1318 s	

Voxtalisib

M	No iteration	Iteration	P-value
4	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr	—
8	sensitive: MYC_Unknown, AFF4_low_expr, CREBBP_Loss-of-function, BCL11A_high_expr, SCAI_high_expr resistant: SVEP1_Unknown, PTPRU_high_expr, CHD4_Unknown	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr, STK11_Unknown, FAT2_CNVloss, PRKAR1A_high_expr resistant: AKAP9_Unknown	0.00028
9	sensitive: MYC_Unknown, SPOP_low_expr, CREBBP_Loss-of-function, FAT2_CNVloss, BCL11A_high_expr, SCAI_high_expr resistant: SVEP1_Unknown, AKAP9_Unknown, ANK3_Unknown	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr, STK11_Unknown, FAT2_CNVloss, PRKAR1A_high_expr resistant: AKAP9_Unknown, STK11_low_expr	1.84e-07
10	sensitive: MYC_Unknown, MYH10_high_expr, FAT2_CNVloss, STK11_Unknown, BCL11A_high_expr, SCAI_high_expr, PRKAR1A_high_expr resistant: CHD1L_Unknown, AKAP9_Unknown, STK11_low_expr	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr, STK11_Unknown, FAT2_CNVloss, PRKAR1A_high_expr resistant: AKAP9_Unknown, STK11_low_expr, CHD1L_Unknown	3.8e-20
12	sensitive: MYC_Unknown, MYH10_high_expr, FAT2_CNVloss, STK11_Unknown, BCL11A_high_expr, SCAI_high_expr, PRKAR1A_high_expr resistant: RFC4_high_expr, CHD1L_Unknown, RPL22_low_expr AKAP9_Unknown, STK11_low_expr	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr, STK11_Unknown, FAT2_CNVloss, PRKAR1A_high_expr resistant: AKAP9_Unknown, STK11_low_expr, CHD1L_Unknown, RFC4_high_expr, RPL22_low_expr	3.37e-23
13	sensitive: MYC_Unknown, CREBBP_Loss-of-function, FAT2_CNVloss, MTOR_Neural, STK11_Unknown, BCL11A_high_expr, SCAI_high_expr, PIK3C2B_low_expr resistant: F8_high_expr, AHCTF1_low_expr, MYB_Unknown, RB1_Loss-of-function, CHD4_Unknown	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr, STK11_Unknown, FAT2_CNVloss, PRKAR1A_high_expr, CDK12_Unknown resistant: AKAP9_Unknown, STK11_low_expr, CHD1L_Unknown, RFC4_high_expr, RPL22_low_expr	1.73e-05

M	No iteration	Iteration	P-value
16	sensitive: MYC_Unknown, CREBBP_Loss-of-function, FAT2_CNVloss, MTOR_Neutral, STK11_Unknown, BCL11A_high_expr, SCAI_high_expr, PIK3C2B_low_expr, SF3B1_low_expr resistant: F8_high_expr, DIS3_low_expr, AHCTF1_low_expr, MYB_Unknown, RB1_Loss-of-functiion, DICER1_high_expr, CHD4_Unknown	sensitive: MYC_Unknown, MYH10_high_expr, BCL11A_high_expr, SCAI_high_expr, STK11_Unknown, FAT2_CNVloss, PRKAR1A_high_expr, SF3B1_low_expr resistant: EEF1A1_low_expr, DIS3_low_expr, AKAP9_Unknown, STK11_low_expr, CHD1L_Unknown, RFC4_high_expr, RPL22_low_expr, CHD4_Unknown	9.43e-09
Total Time	10893 s	520 s	

Figure 3: In this figure, we present a comparison between the selected features of a one-shot approach and an iterative selection for OSI-027, PIK-93, and Voxtalisib using the cubic weight function. We show the feature sets for the best M value in each iteration compared to the model of the same size in the one-shot approach. Common features are marked in magenta, blue features only occur in one of the approaches and black features appear in both approaches but not for the given value of M. We start with the best model among $M = 1$ to $M = 6$ and then iteratively add 4 putative features until we reach a possible model size of $M = 16$. For $M = 16$, we simply show the first model of this size, which is reached in the last iteration. This model does not correspond to the best model of the last iteration. The given p-value is the Fisher's exact test p-value for the intersection of the feature sets. The total time refers to the time needed to fit all models including all M values and CV steps.

2 Enrichment analysis

In order to determine whether the selected biomarkers for the drugs that we investigated are of broad importance to various drugs, we decided to perform an enrichment analysis on the gene expression data. Specifically, we performed Kolmogorov-Smirnov tests (Hollander *et al.*, 2013) for sets of cell lines (categories) on a ranked list of cell lines for each drug using the GeneTrail3 C++ library (Gerstner *et al.*, 2020).
In the following, we will first describe how we obtain the categories and the ranked lists of cell lines. Afterwards, we will outline how the enrichment analysis is performed. Lastly, we present the results of the analysis for all of the genes that have been selected by MERIDA.

2.1 Category definition

To derive categories, we use the z-score based definitions for up- and down-regulated genes as described in Section 2.2 of the main text: In particular, we calculate for each gene one z-score and consider the cell lines with a z-score above 1.68 as up-regulated and the cell lines with a z-score below -1.68 as down-regulated. We then define two categories per gene, the first category contains all cell lines that are up-regulated, the second category contains all cell lines that are down-regulated. From these categories, we assembled two category databases, one that contains all up-regulated categories and one that contains all down-regulated categories. In contrast to the analysis in the main manuscript, we do not consider any splitting into training and test set for this analysis, i.e. the categories are defined on the whole data set available per drug.

2.2 Definition of ranked lists

For each drug, we employ the given logarithmized IC50 values from the GDSC database to increasingly sort the cell lines from sensitive to resistant responses. If there is drug response data in GDSC1 and GDSC2, we construct two sorted lists in order to avoid mixing up data from different drug response assays. The positions in the sorted list then serve as ranks for the cell lines.

2.3 Enrichment analysis results

We employ the GeneTrail C++ library for performing a Gene Set Enrichment Analysis (GSEA) like analysis with the custom categories and ranked lists as described above, i.e. we test for an enrichment of a specific category at the top or bottom of a ranked cell line list. We use the Benjamini-Hochberg procedure (Hochberg and Benjamini, 1990) to adjust the resulting p-values. We consider results as significant at the significance level $\alpha = 0.05$.

Tables 9 + 10 show the result of this analysis for the genes that have been selected by MERIDA.

Table 9: The table shows the number of times a lowly expressed gene has been enriched or depleted.

Gene	#Enriched GDSC1	#Enriched GDSC2	#Depleted GDSC1	#Depleted GDSC2
IDH1	8	33	3	1
PIK3CB	65	40	8	0
FRG1	0	0	20	56
CYTH4	0	0	0	0
TJP1	258	156	13	0
PTPRF	205	150	12	1
MAP4K1	0	0	0	0
IREB2	8	0	16	18
MLL2	1	1	0	4
DDX5	3	0	40	4
WIPF1	0	0	0	0
SMARCB1	0	0	20	16
RGS3	0	0	0	0
HDAC9	0	0	0	0
EPC1	1	0	4	13
AHCTF1	1	1	9	2
ADCY1	0	0	0	0
ROBO2	0	0	0	0
ERBB2	0	0	0	0
PPP6C	2	0	38	86
FOXA1	0	0	0	0
TBX3	0	0	0	0
KAT6B	0	0	5	3
SOX9	0	0	0	0
JMY	0	0	3	0
GNG2	0	0	0	0
VHL	6	0	35	46
ACO1	44	63	8	0
HDAC3	0	0	1	0
SFPQ	0	0	2	14
NCKAP1	239	156	10	0
FKBP5	0	0	0	0
NTRK2	0	0	0	0
PPP2R5A	16	15	35	3
HSP90AB1	0	0	9	0

XPO1	0	0	9	1
PTGS1	0	0	0	0
CNOT3	0	0	1	6
STARD13	0	0	0	0
IDH2	6	0	0	0
FMR1	1	5	2	1
ZNF638	0	0	16	10
ADAM10	10	0	24	7
TCF4	0	0	0	0
TCF7L2	0	0	0	0
CDC73	1	0	0	0
PBRM1	3	0	62	42
ARHGEF6	0	0	0	0
CCND1	33	42	14	0
BCL11A	0	0	0	0
KLF6	111	119	7	0
CLASP2	0	0	14	11
CASP8	32	58	30	1
FBXW7	0	0	2	2
HLA-B	34	65	66	5
MYC	16	12	85	7
DNMT3A	0	0	0	0
MFNG	0	0	0	0
RBM10	2	1	67	11
MYH9	15	45	11	2
TCF12	1	0	0	0
UPF3B	0	0	57	38
ARFGEF2	5	9	5	0
ARID2	0	0	56	16
NAP1L1	0	1	52	58
AHNAK	173	144	15	3

Table 10: The table shows the number of times a highly expressed gene has been enriched or depleted.

Gene	#Enriched GDSC1	#Enriched GDSC2	#Depleted GDSC1	#Depleted GDSC2
IDH1	0	1	28	98
PIK3CB	11	5	2	1
FRG1	184	139	6	0
CYTH4	205	152	4	0
TJP1	0	0	0	0
PTPRF	0	0	7	0
MAP4K1	191	151	15	3
IREB2	155	123	5	0
MLL2	143	113	2	0
DDX5	14	0	0	0
WIPF1	193	150	13	1
SMARCB1	193	146	9	2
RGS3	24	4	51	43
HDAC9	10	0	1	2
EPC1	177	129	6	1
AHCTF1	53	20	6	0
ADCY1	22	33	58	6
ROBO2	112	117	5	1
ERBB2	20	9	57	48
PPP6C	38	20	8	1
FOXA1	1	6	184	109
TBX3	3	0	56	30
KAT6B	153	127	7	2
SOX9	0	0	24	16
JMY	40	97	4	0
GNG2	138	122	8	1
VHL	128	140	6	0
ACO1	11	1	70	50
HDAC3	26	17	1	0
SFPQ	80	94	6	0
NCKAP1	0	0	0	0
FKBP5	192	150	10	2
NTRK2	0	0	24	3
PPP2R5A	10	4	2	0
HSP90AB1	6	33	3	0
XPO1	17	12	5	1
PTGS1	37	10	0	1
CNOT3	15	30	1	0
STARD13	28	23	0	0

IDH2	33	58	12	2
FMR1	2	1	20	2
ZNF638	62	44	6	0
ADAM10	2	0	27	21
TCF4	146	131	14	1
TCF7L2	8	2	38	55
CDC73	3	1	9	3
PBRM1	181	119	6	1
ARHGEF6	237	151	9	0
CCND1	8	2	5	1
BCL11A	200	146	8	0
KLF6	6	2	59	41
CLASP2	133	120	10	2
CASP8	105	89	0	0
FBXW7	147	137	11	0
HLA-B	0	0	0	0
MYC	0	0	0	0
DNMT3A	135	126	11	2
MFNG	224	151	13	1
RBM10	76	60	17	2
MYH9	18	1	120	70
TCF12	106	92	20	5
UPF3B	72	93	16	2
ARFGEF2	0	0	107	116
ARID2	162	129	7	1
NAP1L1	153	72	2	0
AHNAK	2	0	23	30

3 Comparison to other machine learning approaches

In the main manuscript, we present a detailed performance evaluation between MERIDA and LOBICO, which are both methods based on the idea of using an ILP formulation for synthesizing a logical rule for drug sensitivity prediction (cf. Supplement S3 for an in-depth comparison). To show that our MERIDA method is also competitive to standard machine learning methods, we fitted models for random forests (Breiman, 2001) and k-nearest neighbors (Fix and Hodges, 1952) using the implementations provided by the R caret package (Kuhn, 2020). In the following, we will first outline how we prepared the data and fitted the models and then present the results.

3.1 Data preparation and model fitting

We used the data sets as prepared for our main analysis and described in our main manuscript (Section 2.2) and Supplement S1. In particular, we consider the same partition into training and test set. We fitted models for Setting 1, i.e. the matrix without any included a priori knowledge.

For random forests, we employ the *rf* method offered by the package caret to fit a model. We use the default settings for this package except for the parameter mtry, the number of considered features to determine the optimal split in a node, for which we do a grid search in the range of $[max(\sqrt{\#Features} - 50, 1), \sqrt{\#Features} + 50]$ with a step size of 2. Here, $\sqrt{\#Features}$ is the default value that is often used in classification for random forests and also in the package used by caret (Liaw and Wiener, 2002). Analogously to our main analysis, we perform a 5-fold CV on the training set and apply the best model fitted on the complete training data to the test set. As model selection criterion for the best model we opted for the area under the ROC curve, which should in general balance between sensitivity and specificity, and sensitivity alone.

For KNN, we use the *knn* method included in the caret package. We optimize the only parameter available for this package: k. It represents the number of neighbors to consider for prediction and we try values in the range of $[1, 50]$ with a step size of 2. Similar to random forests, we conducted a 5-fold CV on the training set and apply the best model to the test set. We also employed the same selection criteria, i.e. area under the ROC curve and sensitivity.

3.2 Results

Figures 4 to 11 depict the performance for random forests and k-nearest neighbors using both of the mentioned model selection criteria. In Figure 12 and 13, the performance of MERIDA for all 41 drugs is shown.

For k-nearest neighbors, the ROC selection criterion leads to models with high specificity and very low sensitivity for all of the investigated 41 drugs. Notably, almost only drugs that are balanced with respect to sensitive and resistant cell lines (CX-5461, NSC319726, Niraparib, Omipalisib, Talazoparib) achieve sensitivities different from 0. By selecting the best model based on sensitivity alone, the average sensitivity could be improved (see

Figure 6 and 7). However, it is still much lower than for our MERIDA method (cf. Figure 12 and 13).
For random forests, it can similarly be seen that the average sensitivity is comparatively low and is also only marginally improved by using sensitivity as selection criterion. In general, drugs with less class imbalance also seem to achieve a better prediction performance.

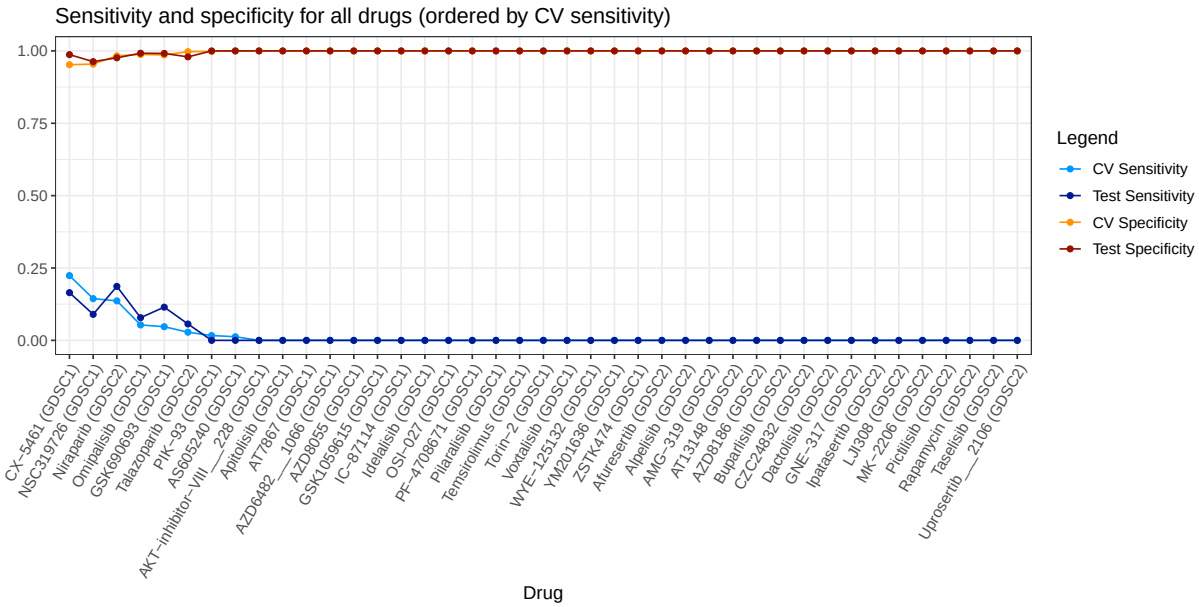


Figure 4: This figure depicts the statistical performance of the best KNN model in Setting 1 for all 41 drugs using the ROC as selection criterion.

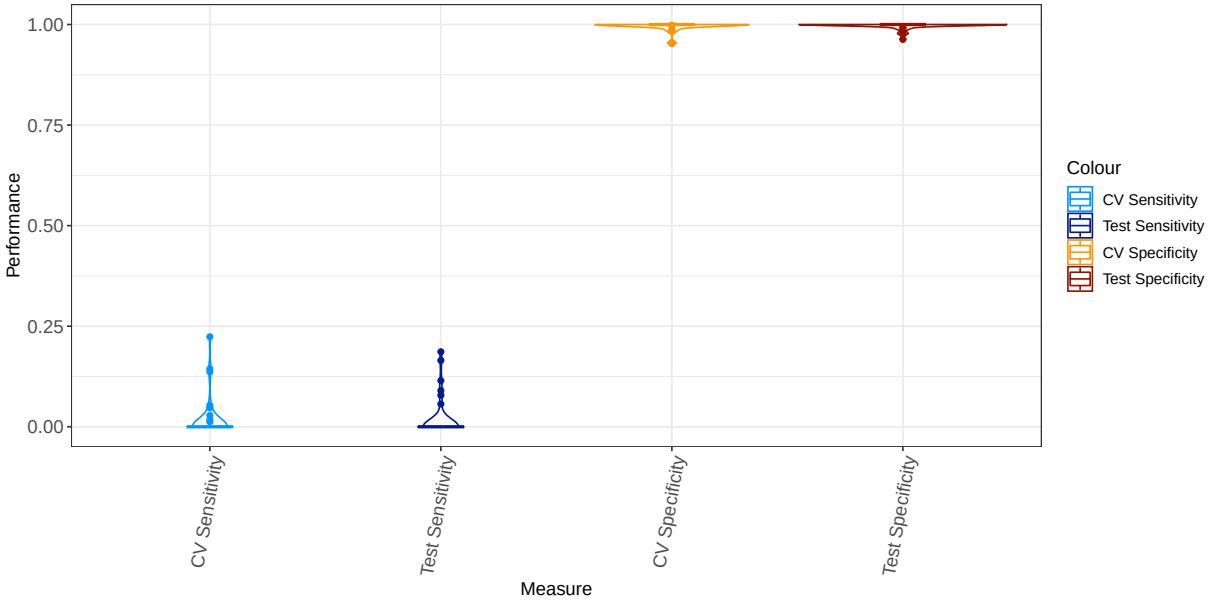


Figure 5: This figure depicts the averaged statistical performance of the best KNN model in Setting 1 for all 41 drugs using the ROC as selection criterion.

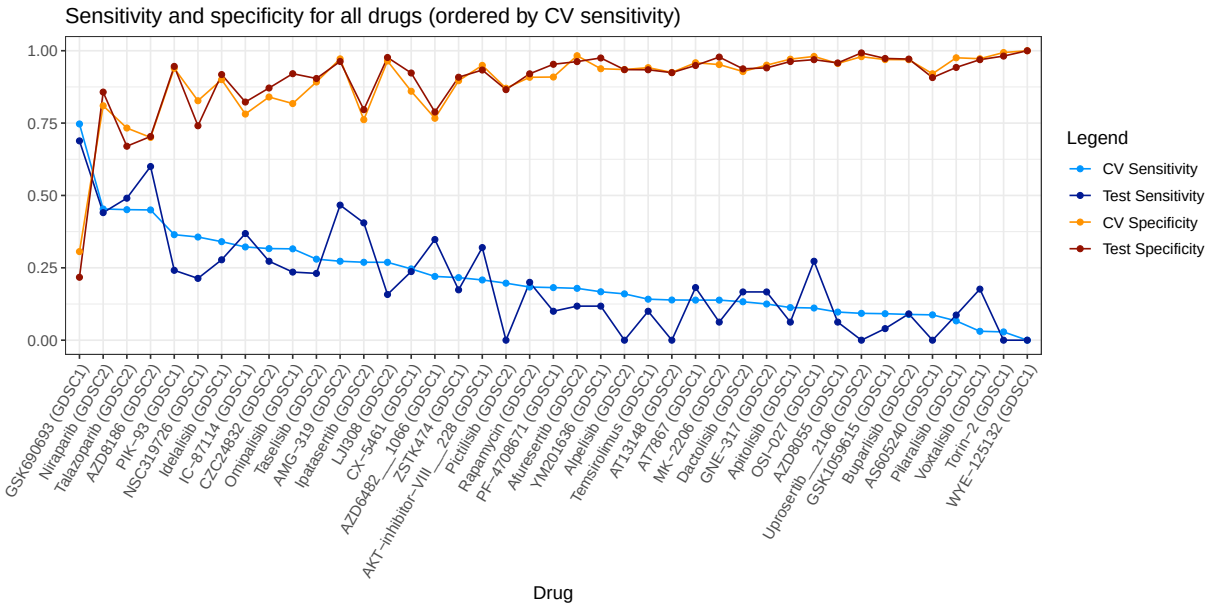


Figure 6: This figure depicts the statistical performance of the best KNN model in Setting 1 for all 41 drugs using sensitivity as selection criterion.

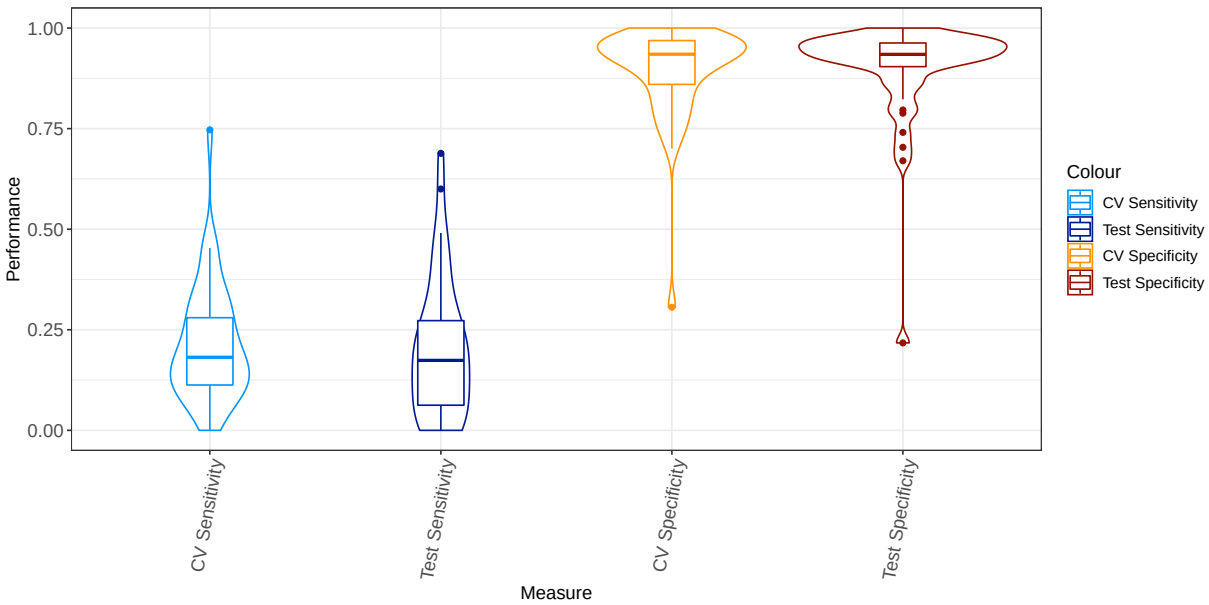


Figure 7: This figure depicts the averaged statistical performance of the best KNN model in Setting 1 for all 41 drugs using sensitivity as selection criterion.

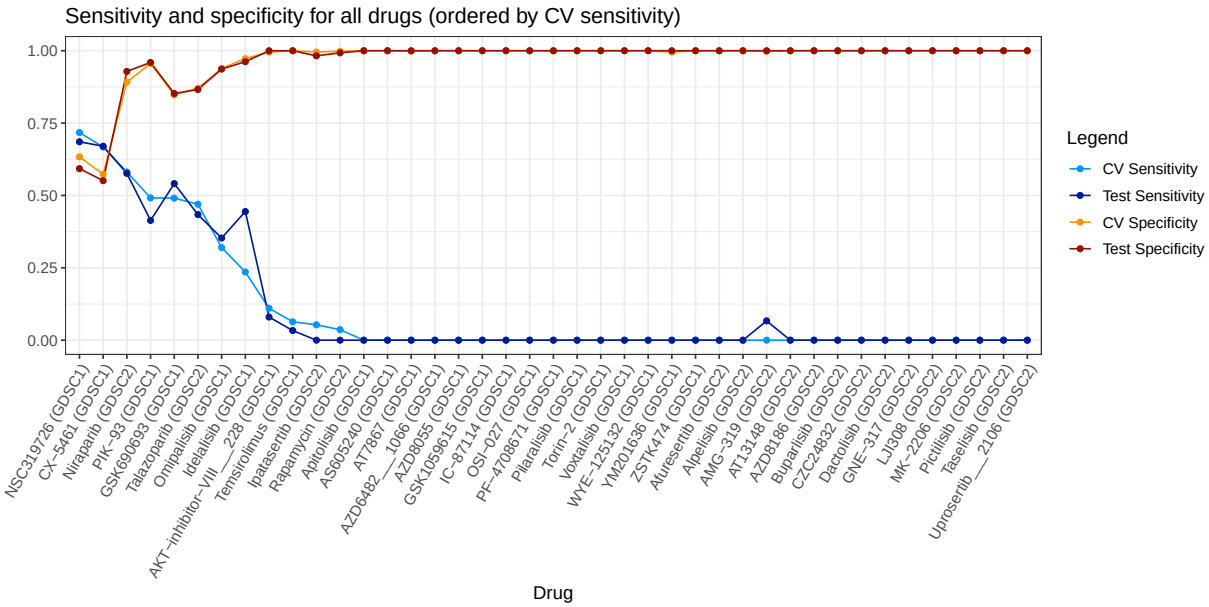


Figure 8: This figure depicts the statistical performance of the best RF model in Setting 1 for all 41 drugs using the ROC as selection criterion.

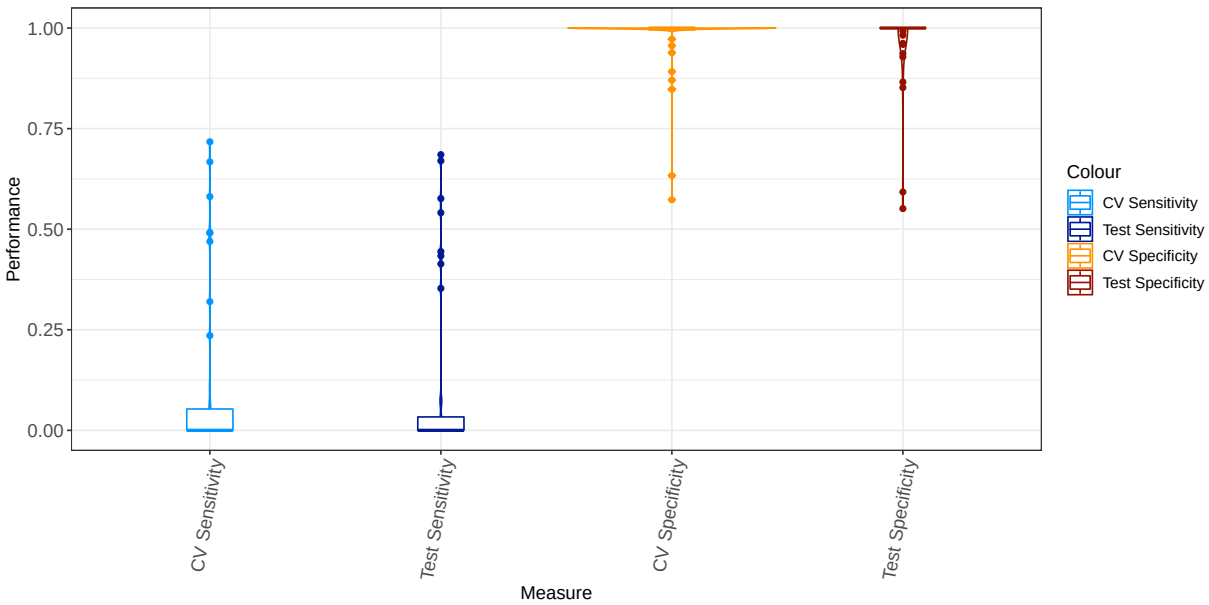


Figure 9: This figure depicts the averaged statistical performance of the best RF model in Setting 1 for all 41 drugs using the ROC as selection criterion.

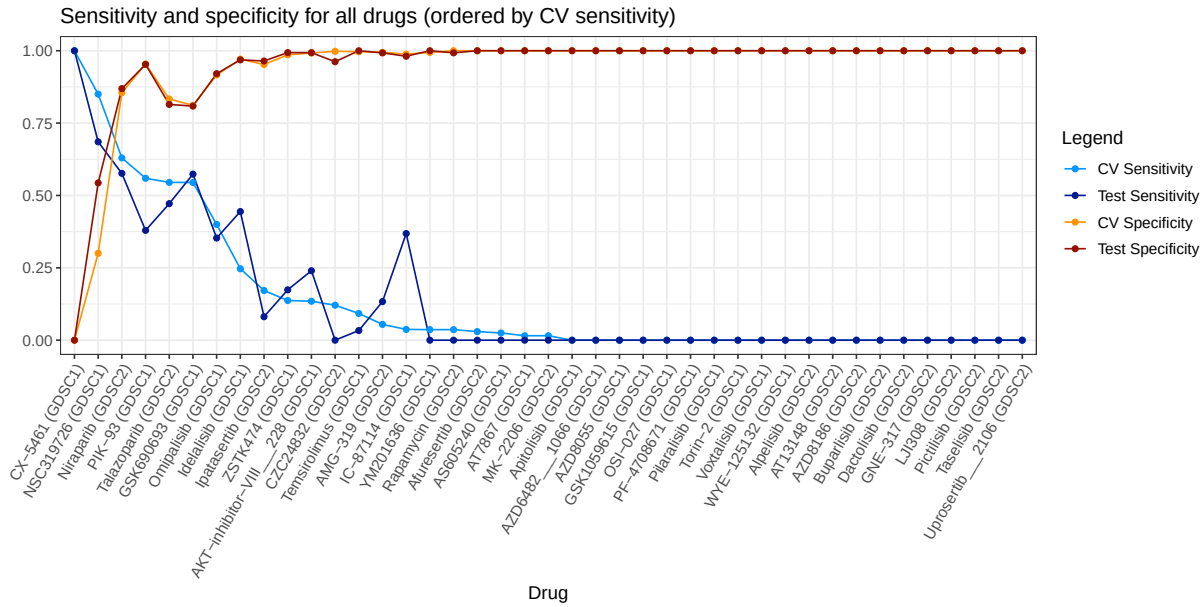


Figure 10: This figure depicts the statistical performance of the best RF model in Setting 1 for all 41 drugs using sensitivity as selection criterion.

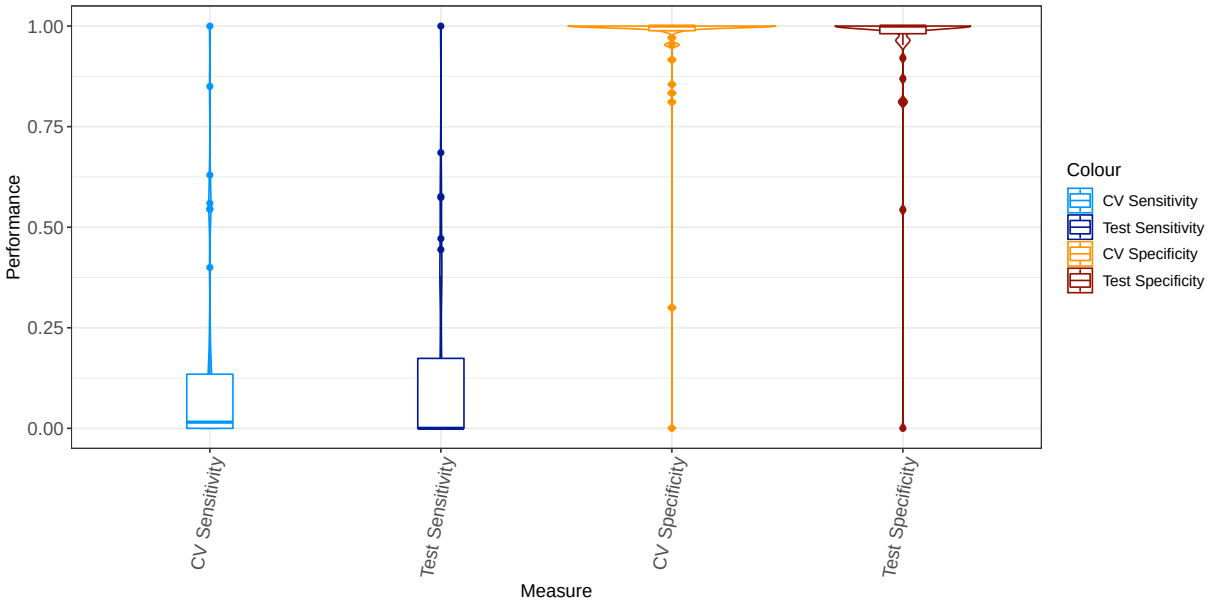


Figure 11: This figure depicts the averaged statistical performance of the best RF model in Setting 1 for all 41 drugs using sensitivity as selection criterion.

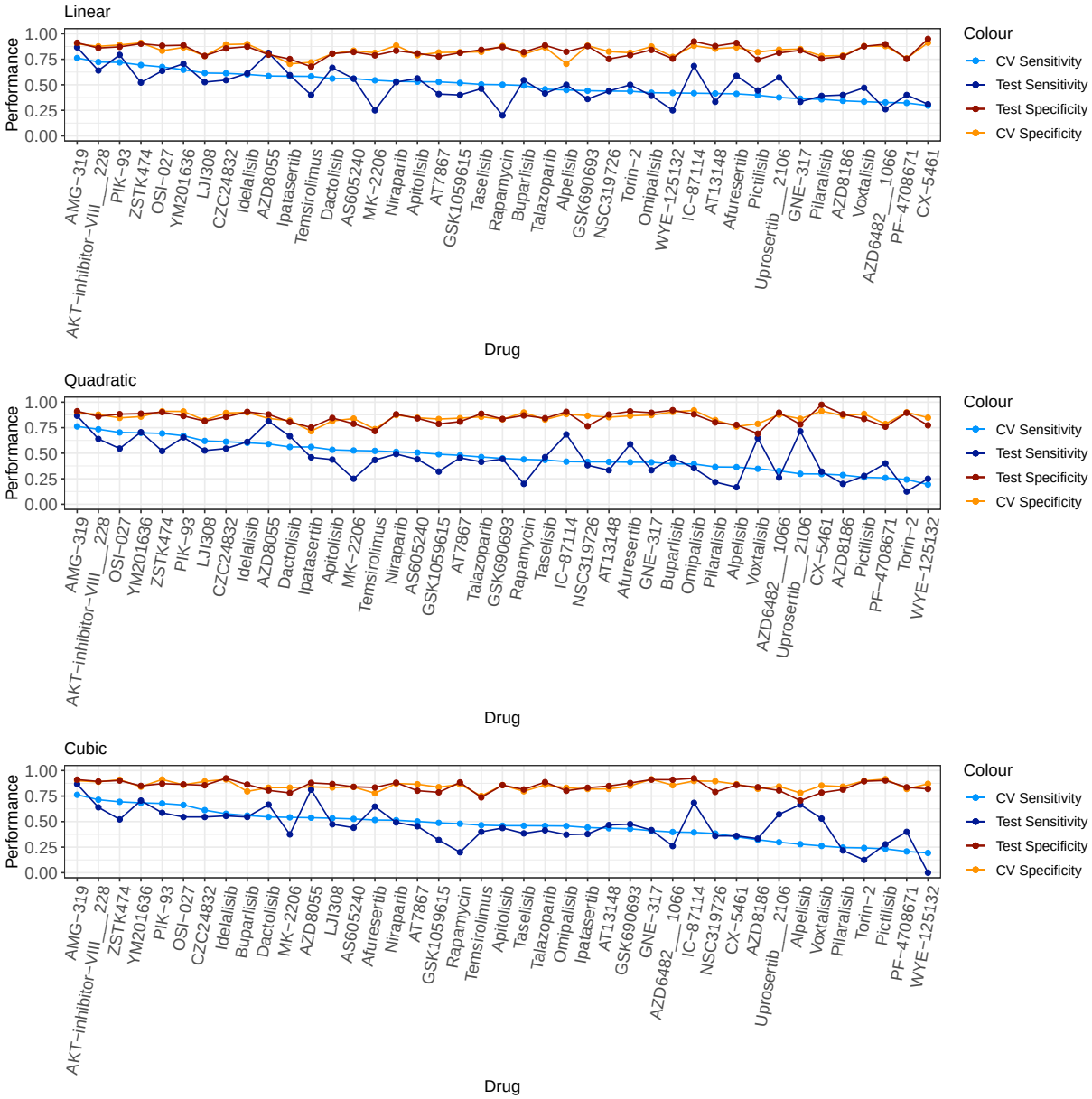


Figure 12: This figure depicts the statistical performance of the best MERIDA model in Setting 1 for all 41 drugs using the three different weight functions.

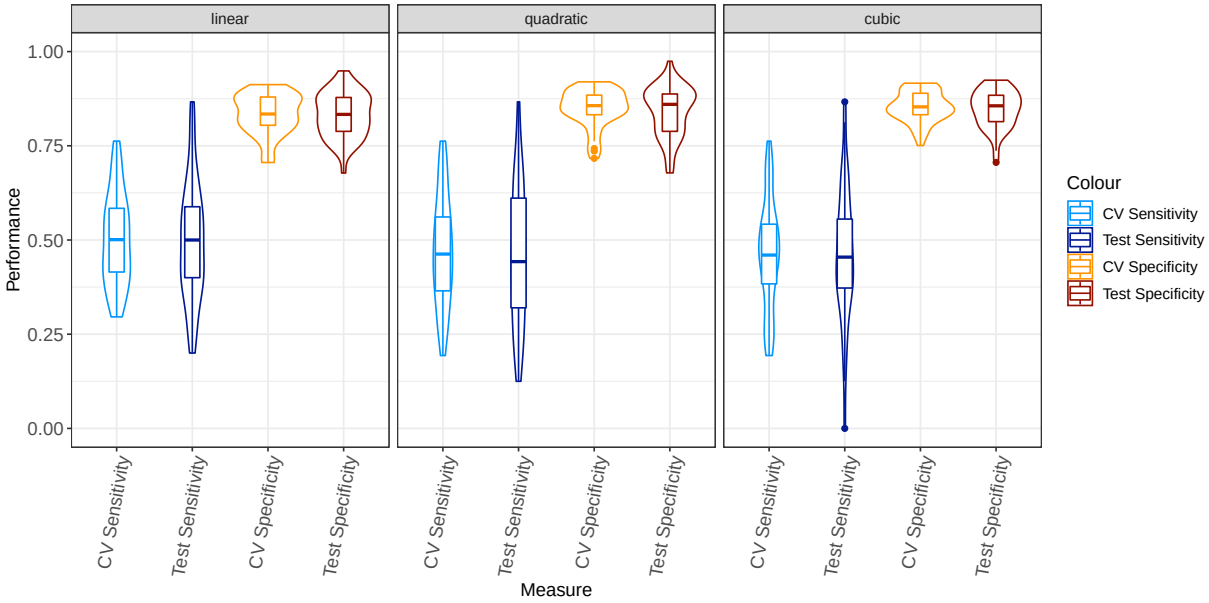


Figure 13: This figure depicts the averaged statistical performance of the best MERIDA model in Setting 1 for all 41 drugs using the three different weight functions.

4 Tissue/site specific analysis

The main analysis presented in the manuscript focuses on the full pan-cancer data set from the GDSC. However, it is well-known that tissue plays a significant role in drug sensitivity and resistance. Therefore, we also applied our method to two interesting subgroups of cell lines: haematological cancer cell lines and non-haematological cancer cell lines. In the following, we will first explain how we prepared the data. Then, we shortly present the results with respect to statistical performance and selected biomarkers.

4.1 Data preparation

In general, the performance of statistical learning methods strongly depends on the number of available samples in comparison to the number of available features. As already discussed in the main manuscript, feature selection or dimension reduction steps are therefore indispensable in this field of bioinformatics. When subdividing the pan-cancer data set into single tissues, most of them will be too small to be analyzed. Therefore, we decided to investigate two groups of cell lines that still have a considerable amount of samples that we determined as described in the following.
When observing the tissue distribution of cell lines across the drug responses of the different drugs, we noted that one particular group of cell lines, the haematological cancer cell lines (i.e. leukemia, lymphoma, and myeloma), is often found to be very sensitive to drugs (cf. Figure 14 and 15 for mTOR pathway inhibitors and Figure 43 and 44 for all drugs). We then use this group containing approximately 160 cell lines per drug and the group of the non-haematological cancer cell lines containing approximately 700 cell lines per drug for further analysis.
We prepared the data sets exactly as described in the main manuscript and Supplement S1. This means that we first divided the available cell lines into the two aforementioned groups and then performed the preparation of the data sets including literature-based feature selection and annotation. We also analyzed these two groups in the first two of the three settings from the main manuscript:

- Setting 1: the a priori knowledge is not included, i.e. the matrix without a priori knowledge is used
- Setting 2: the information about the sensitivity biomarkers is integrated in the input feature matrix as one specific composite feature (cf. Section 2.2. in the main manuscript) and the value of the corresponding ILP feature variable is fixed to 1

4.2 Results

Figures 16 to 23 show the statistical performance for MERIDA on the 41 different drugs for the non-haematological and haematological cell lines. For Setting 1, it can be seen that both specificity and sensitivity are highest for the pan-cancer analysis (cf. 12 and lowest for the haematological cancer analysis. This effect could generally be attributed to the fact that fewer cell lines are used in both, the haematological cancer analysis and

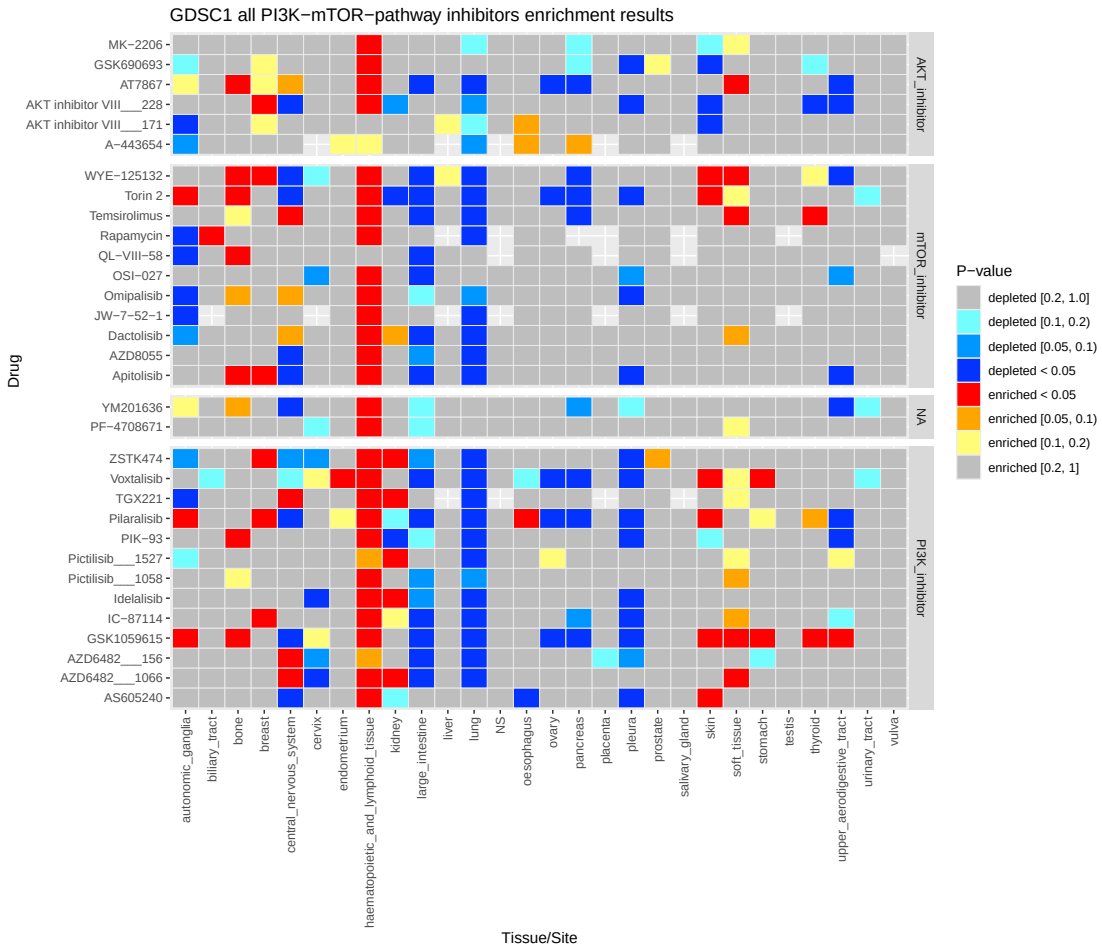


Figure 14: This figure shows the results of an enrichment analysis for the different tissues/sites available in the GDSC database for the mTOR pathway inhibitors in GDSC1. The enrichment analysis was conducted using the respective site as category, and the increasingly sorted logarithmized IC50 lists as ranked lists. The p-values were adjusted using the Benjamini-Yekutieli procedure.

the non-haematological cancer analysis in comparison to the pan-cancer analysis. For Setting 2, it can be noted that the a priori knowledge seems to be more favourable for the pan-cancer cell lines and the non-haematological cancer cell lines in terms of statistical sensitivity, which might indicate that the drug sensitivity of the haematological cancer cell lines is less determined by the used biomarkers.

We also analyzed the importance of the tissue in the resulting models. From our results for Setting 1, we can clearly see that the selected biomarkers from the non-haematological cancer group largely resemble those from the analysis using all cell lines (cf. Figure 24 to 27). There is still some intersection of the selected biomarkers for the haematological cancer cell lines with the biomarkers for some of the drugs (cf. Figure 28 and 29 and

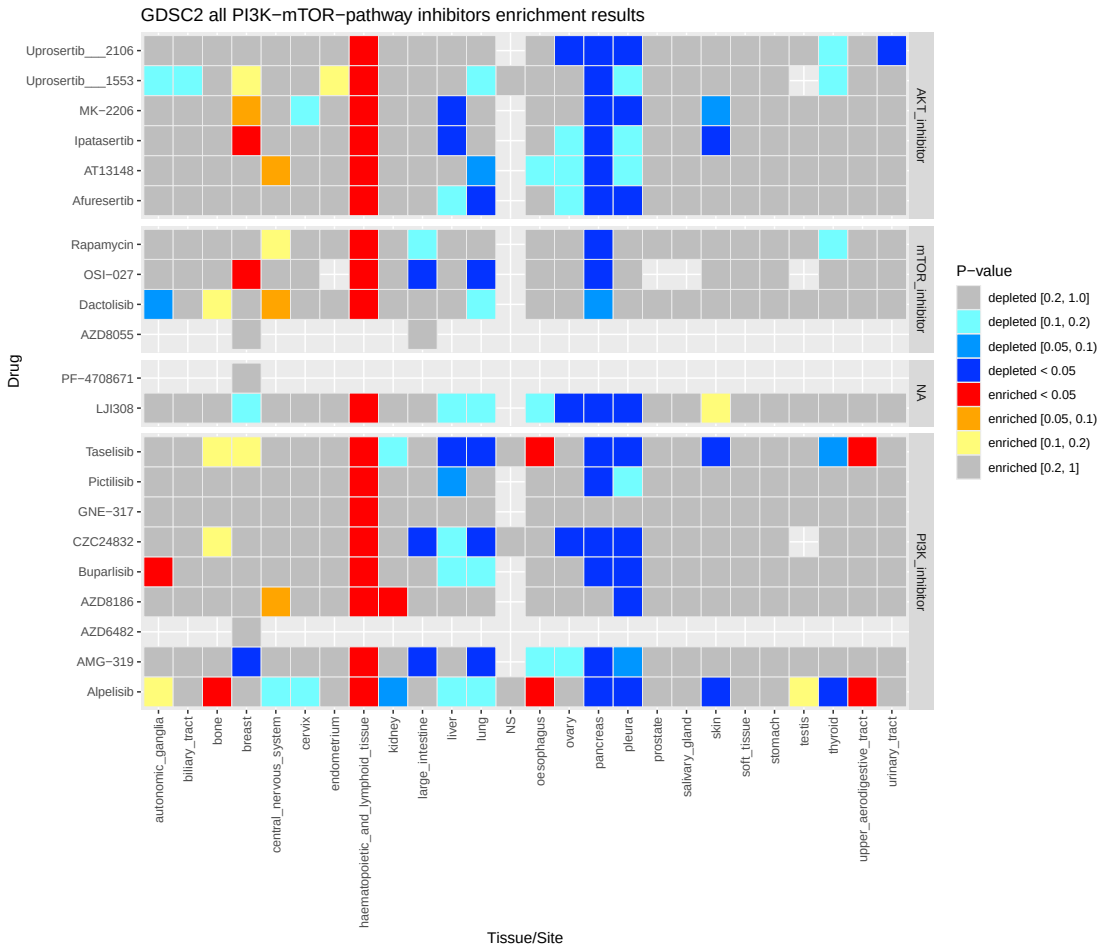


Figure 15: In analogy to Figure 14, this figure shows the results of an enrichment analysis for the different tissues/sites available in the GDSC database for the mTOR pathway inhibitors in GDSC2. The enrichment analysis was conducted using the respective site as category, and the sorted IC50 lists as ranked lists. The p-values were adjusted using the Benjamini-Yekutieli procedure.

Table 11 to Table 13). However, the relationship is much less pronounced highlighting the need for tissue-specific models. The smallest overlap can be observed for the features from the haematological cancer cell lines with the features from the non-haematological cancer cell lines.

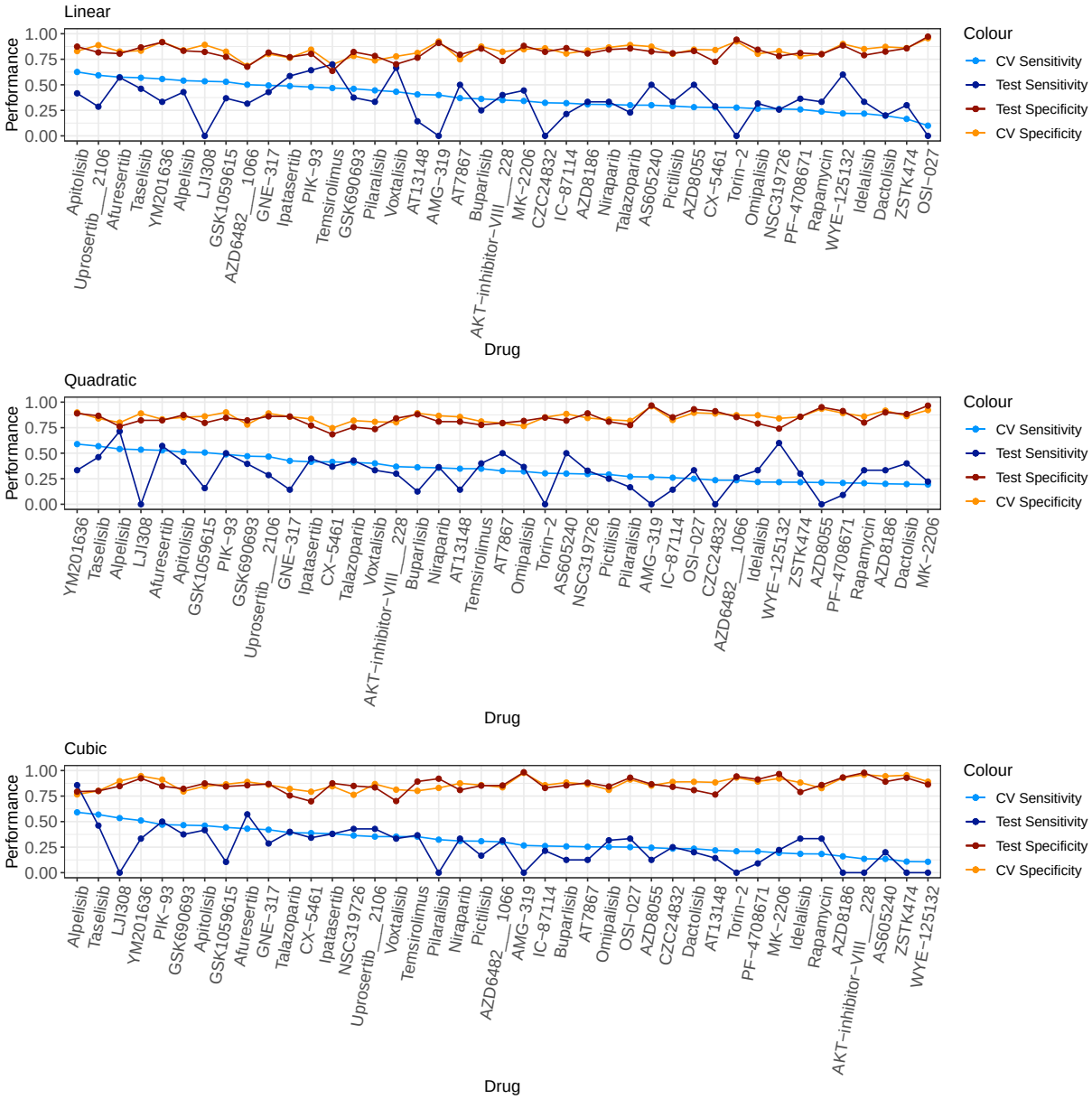


Figure 16: This figure depicts the statistical performance of the best MERIDA model in Setting 1 for all 41 drugs for the non-haematological cell lines.

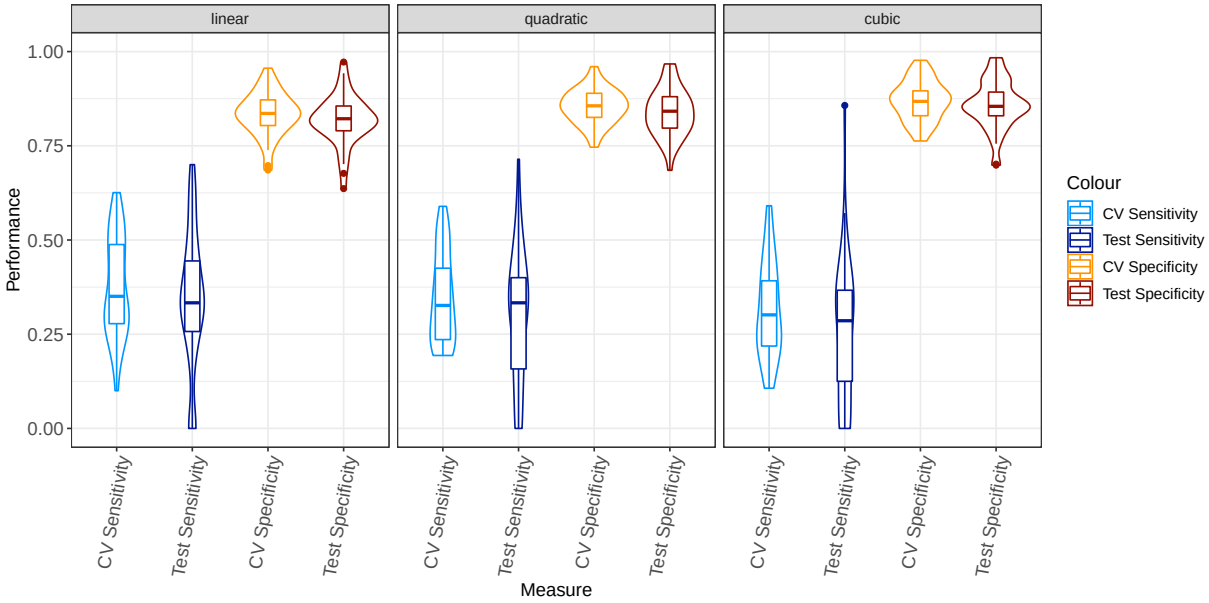


Figure 17: This figure depicts the averaged statistical performance of the best MERIDA model in Setting 1 for all 41 drugs for the non-haematological cell lines.

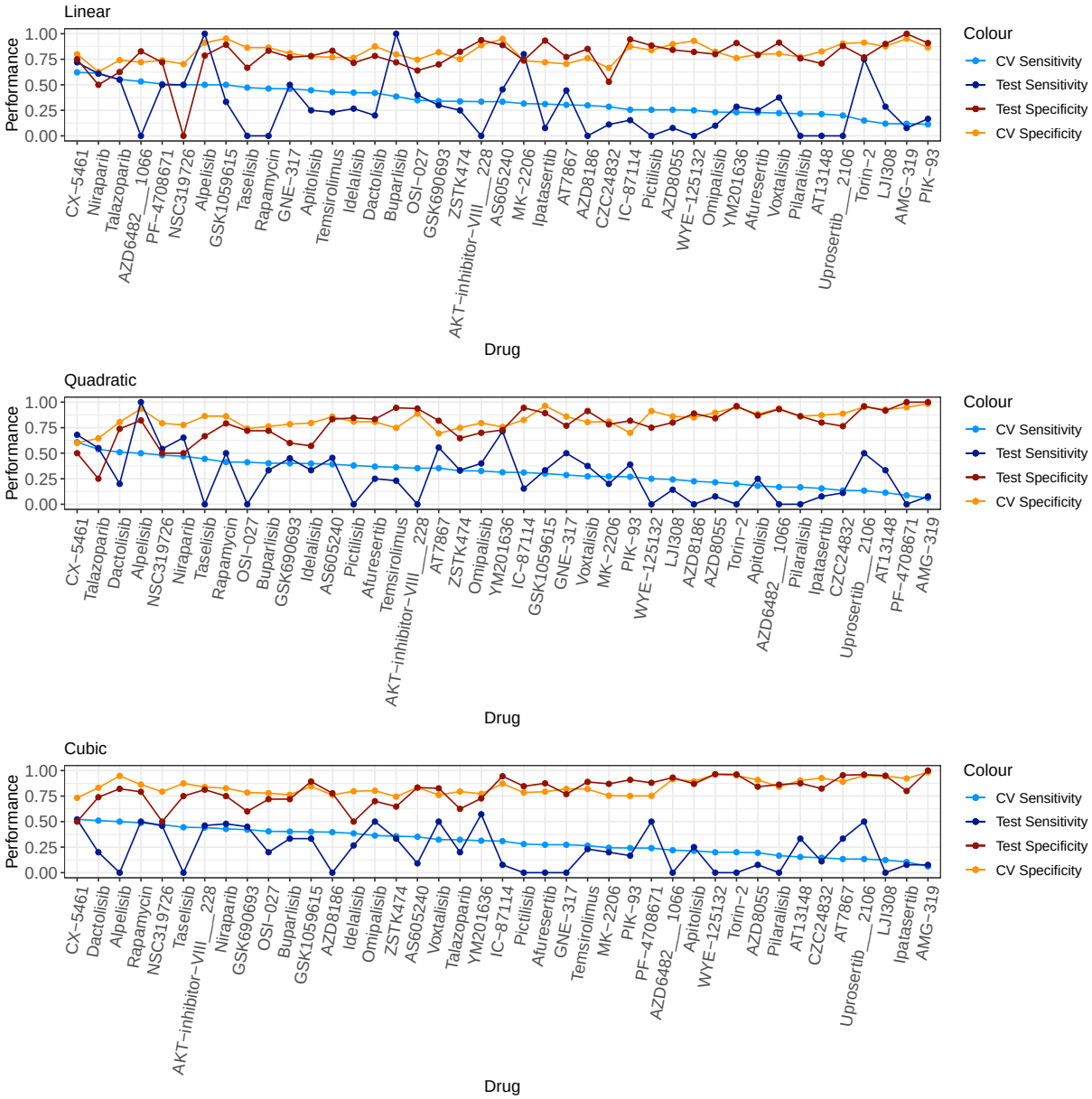


Figure 18: This figure depicts the statistical performance of the best MERIDA model in Setting 1 for all 41 drugs for the haematological cell lines.

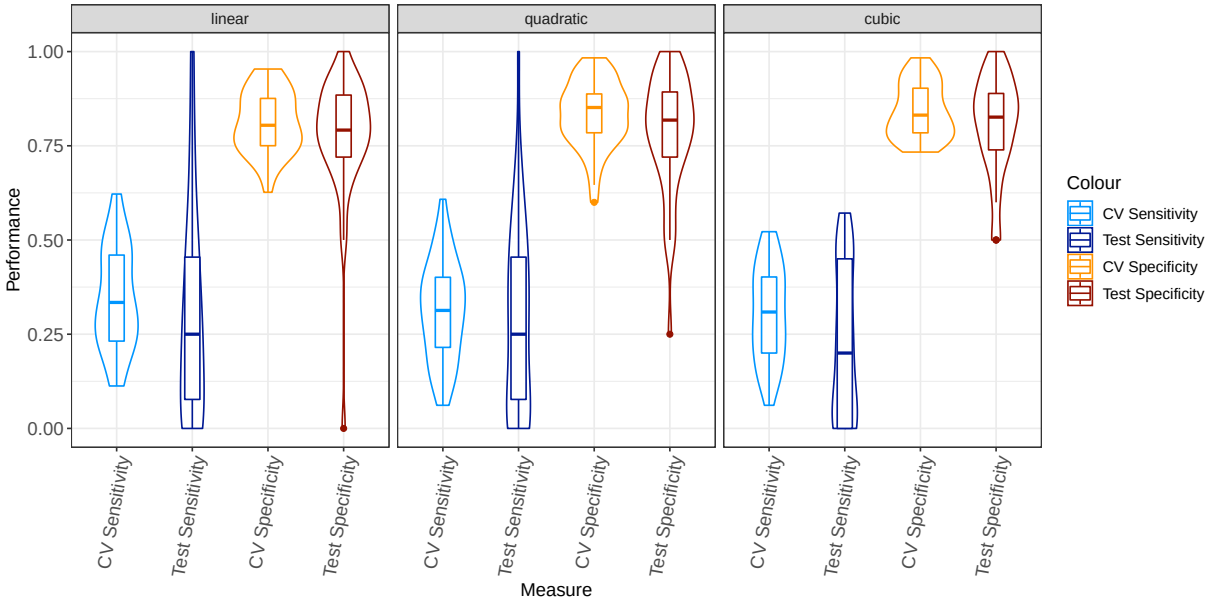


Figure 19: This figure depicts the averaged statistical performance of the best MERIDA model in Setting 1 for all 41 drugs for the haematological cell lines.

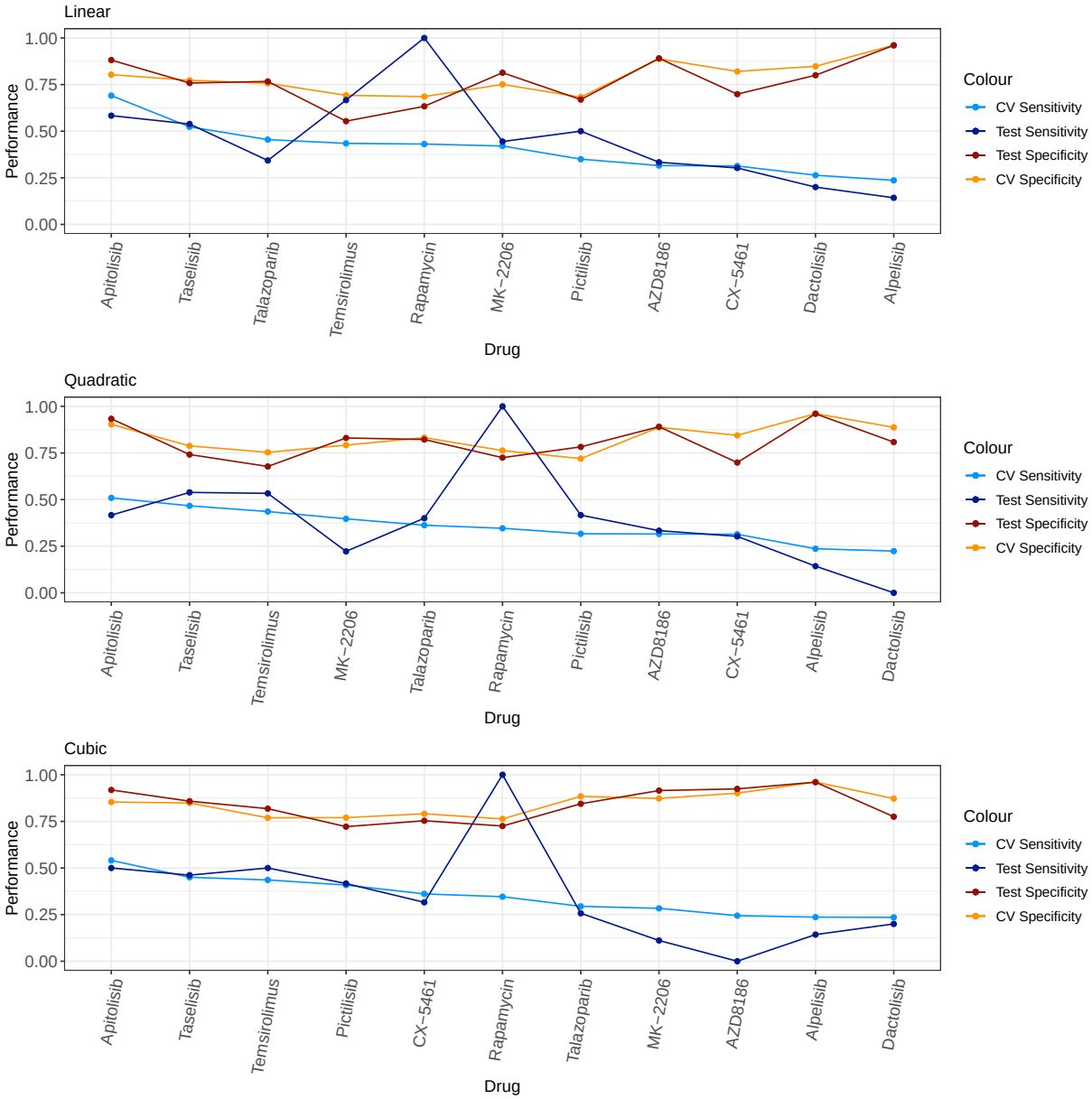


Figure 20: This figure depicts the statistical performance of the best MERIDA model in Setting 2 for non-haematological cell lines.

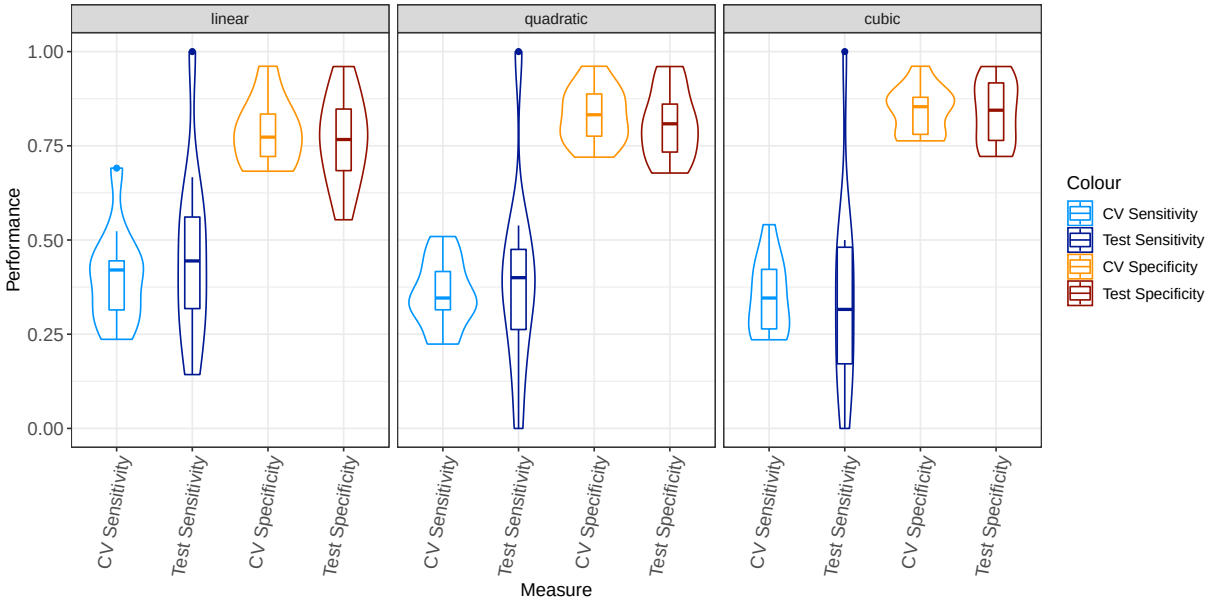


Figure 21: This figure depicts the averaged statistical performance of the best MERIDA model in Setting 2 for non-haematological cell lines.

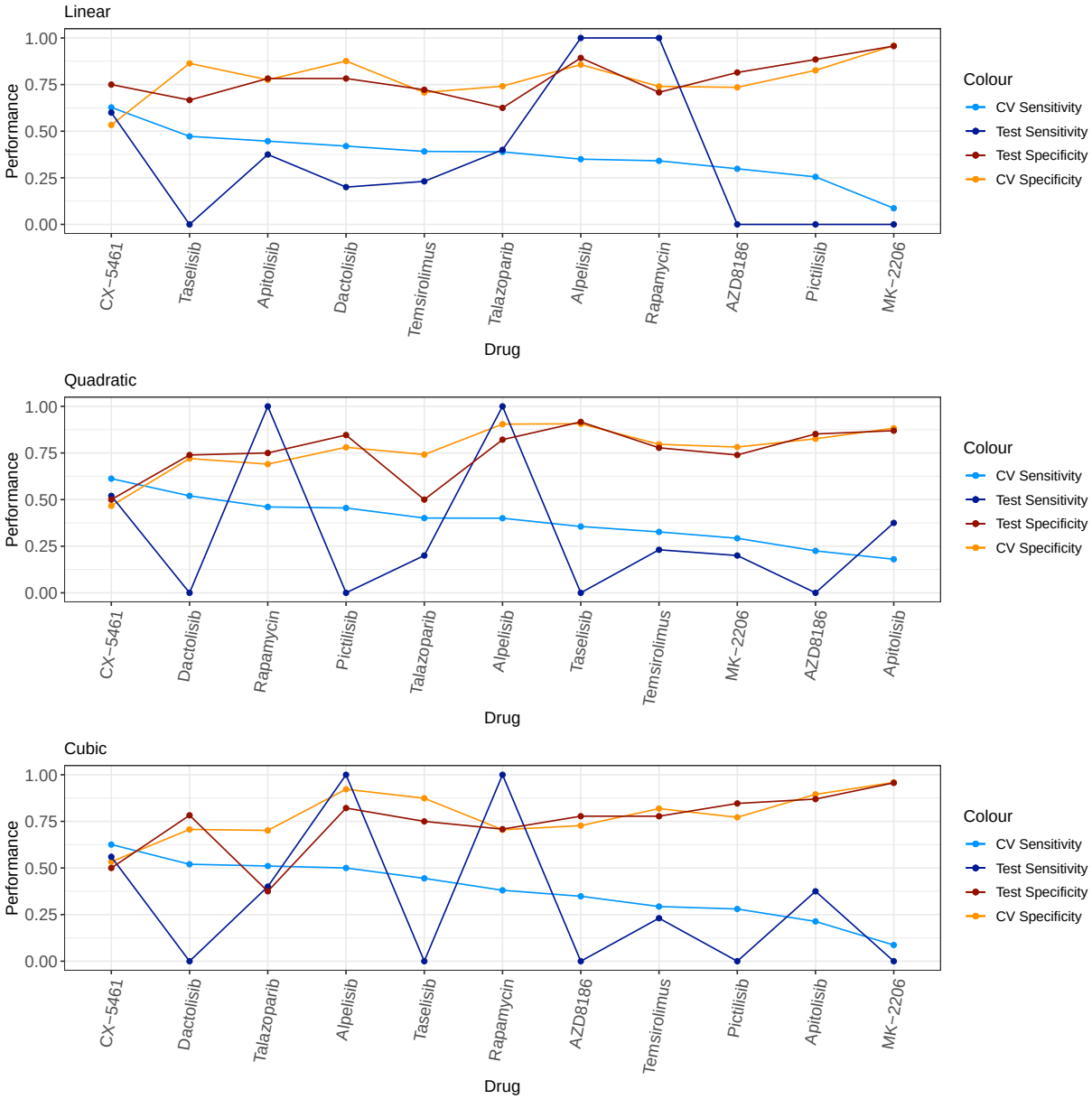


Figure 22: This figure depicts the statistical performance of the best MERIDA model in Setting 2 for haematological cell lines.

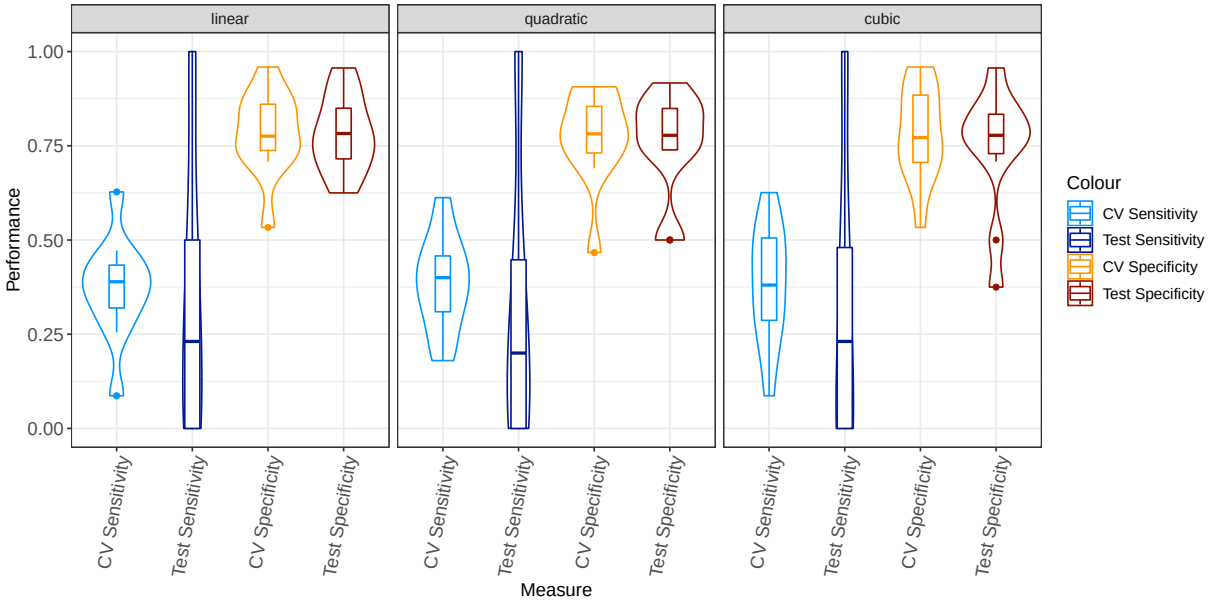


Figure 23: This figure depicts the averaged statistical performance of the best MERIDA model in Setting 2 for the haematological cell lines.

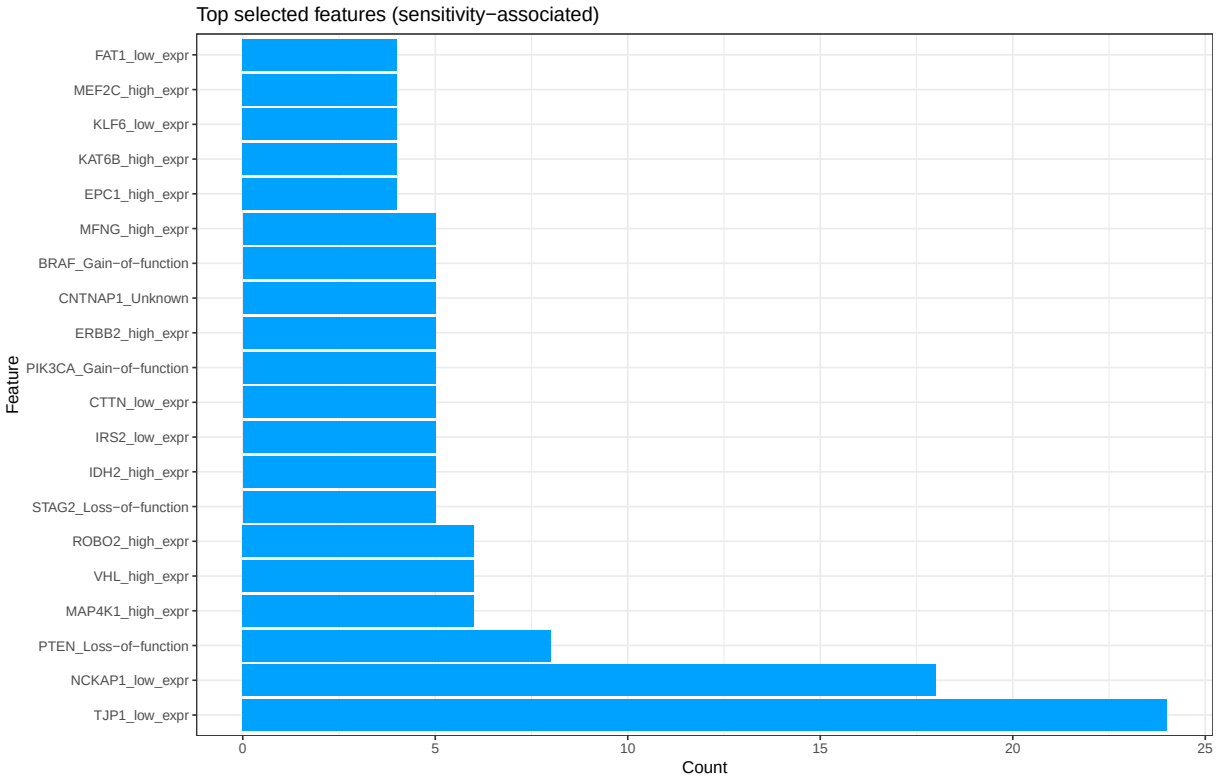


Figure 24: This figure depicts the the frequency of the top 20 selected sensitivity-associated features (using all models of size 1 to 6 per drug) across all mTOR pathway inhibitors (37) for Setting 1 in the pan-cancer analysis.

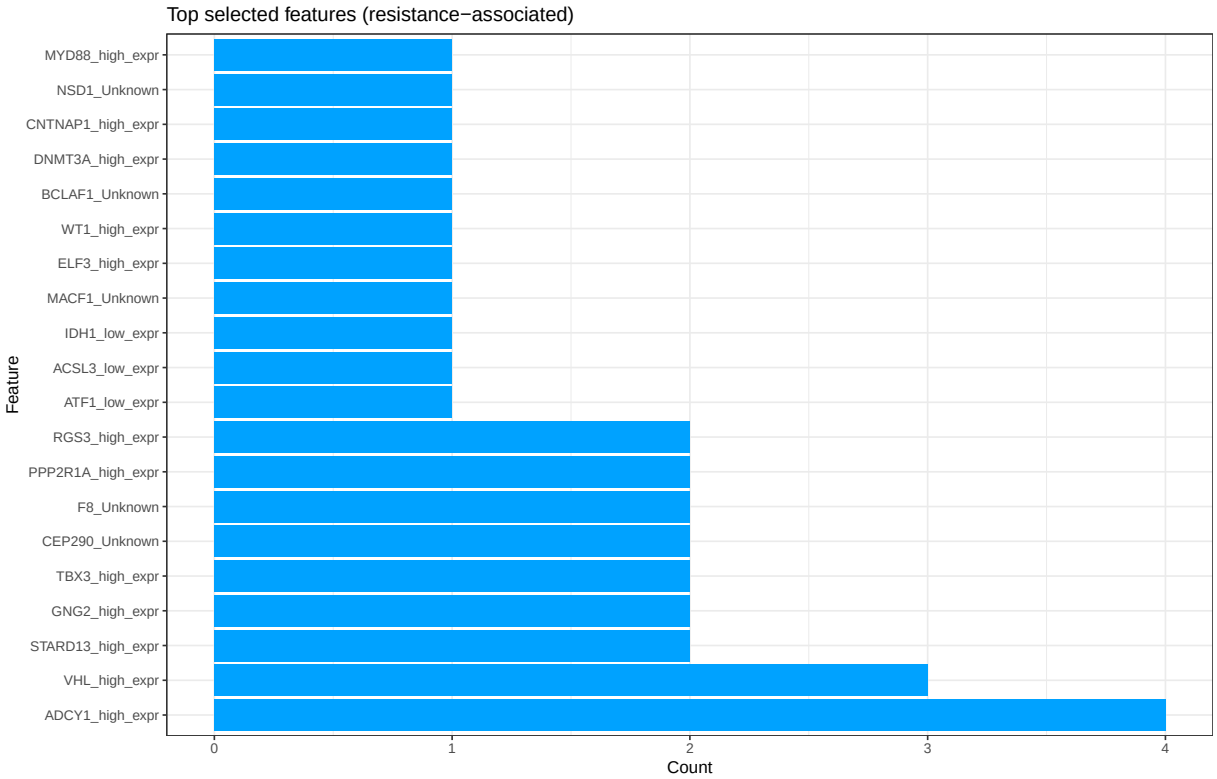


Figure 25: This figure depicts the the frequency of the top 20 selected resistance-associated features (using all models of size 1 to 6 per drug) across all mTOR pathway inhibitors (37) for Setting 1 in the pan-cancer analysis.

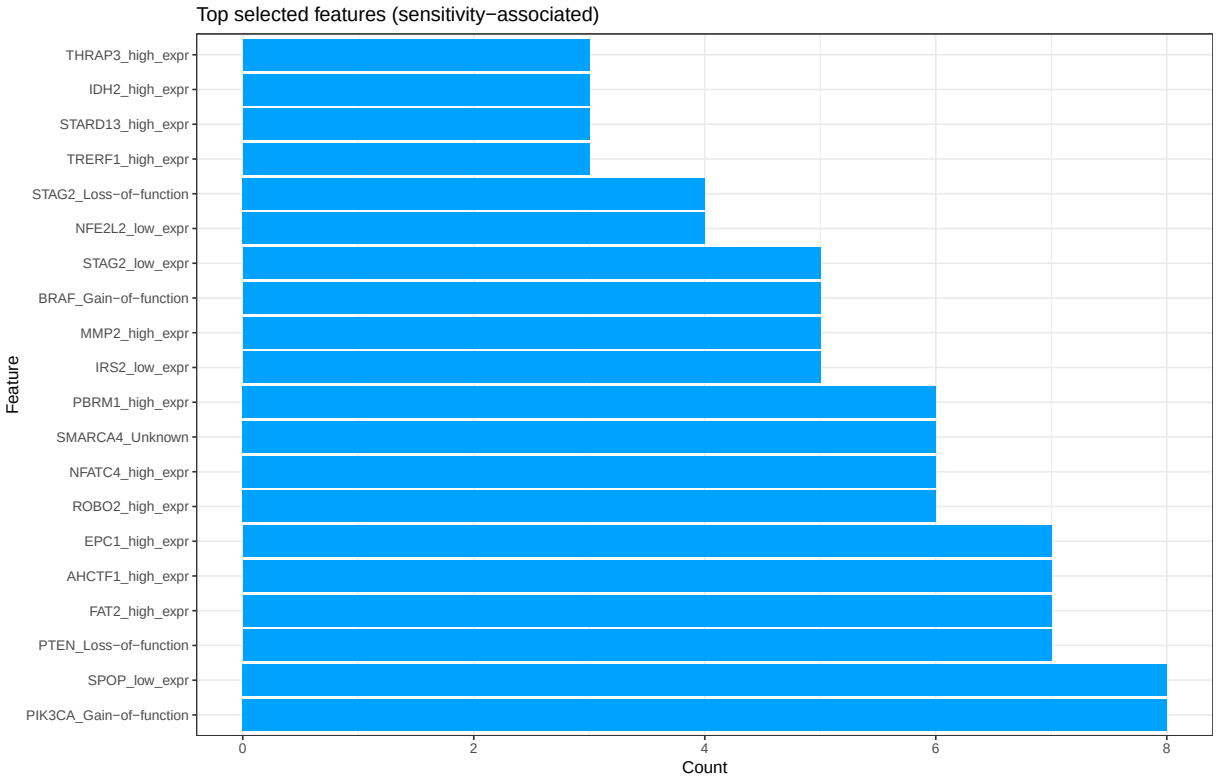


Figure 26: This figure depicts the the frequency of the top 20 selected sensitivity-associated features (using all models of size 1 to 6 per drug) across all mTOR pathway inhibitors (37) for Setting 1 for the non-haematological cancer cell lines.

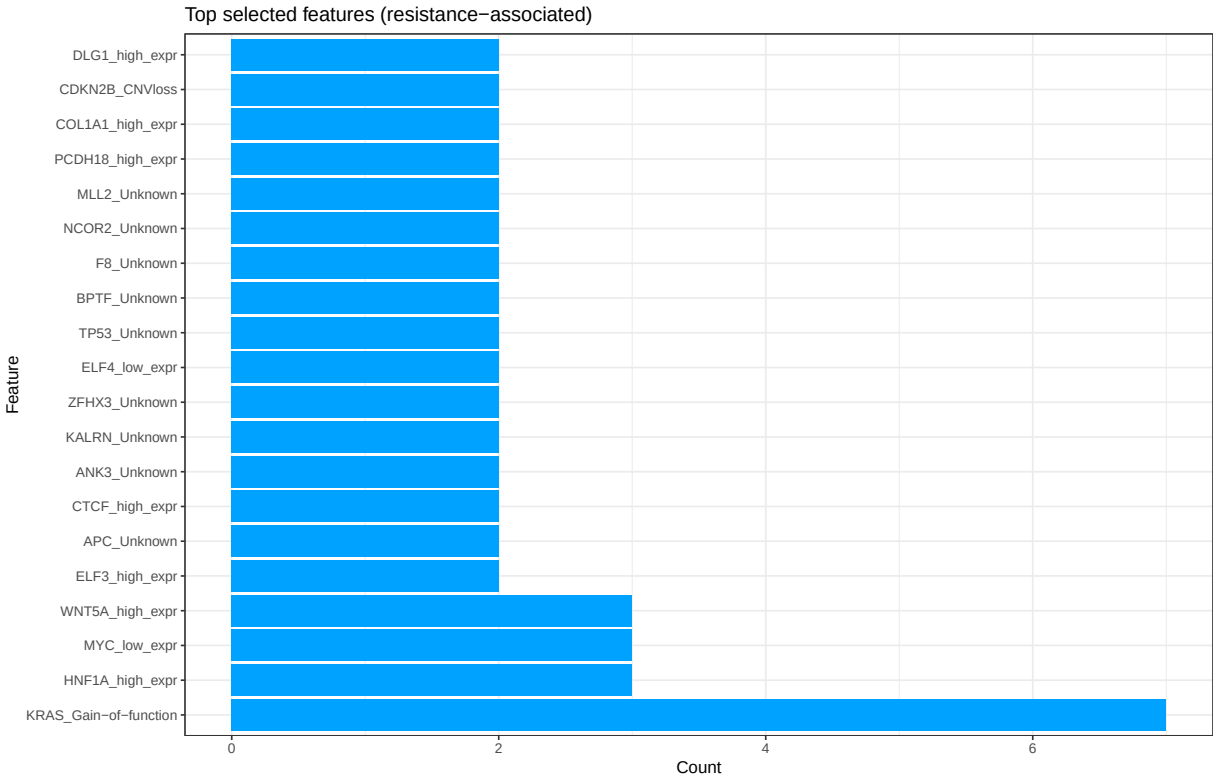


Figure 27: This figure depicts the the frequency of the top 20 selected resistance-associated features (using all models of size 1 to 6 per drug) across all mTOR pathway inhibitors (37) for Setting 1 for the non-haematological cancer cell lines.

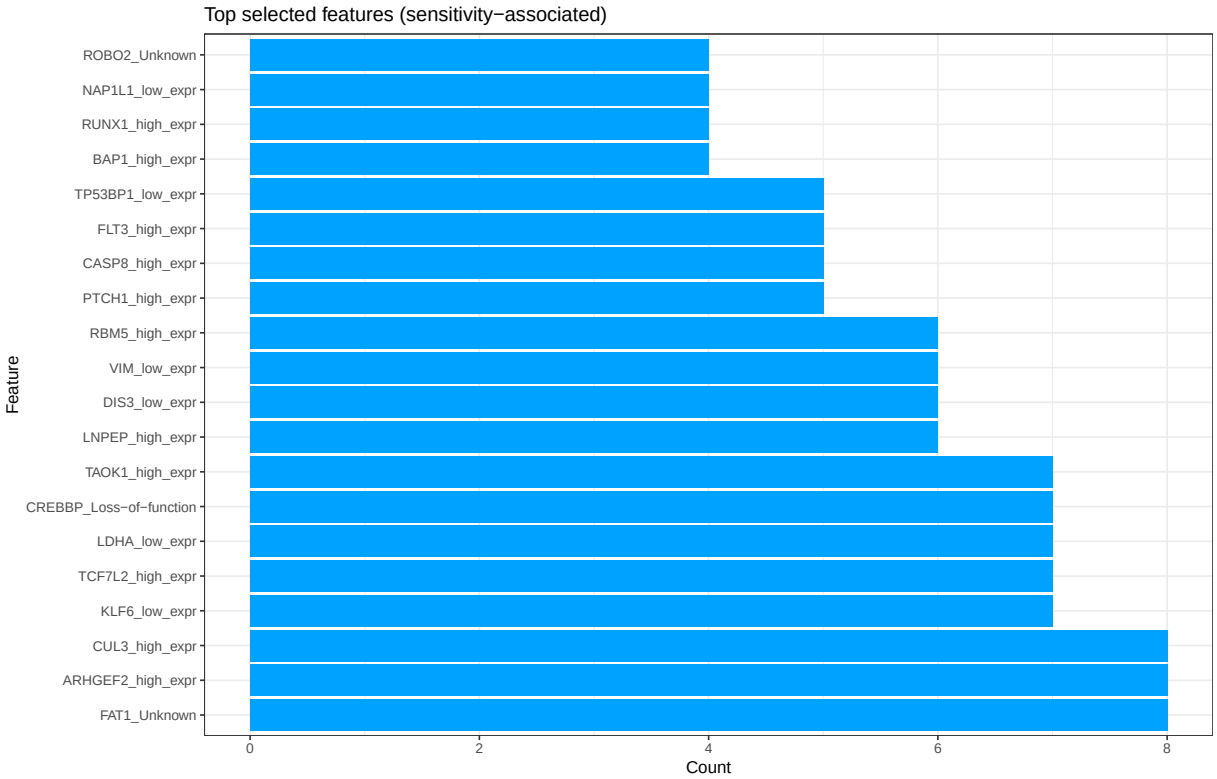


Figure 28: This figure depicts the the frequency of the top 20 selected sensitivity-associated features (using all models of size 1 to 6 per drug) across all mTOR pathway inhibitors (37) for Setting 1 for the haematological cancer cell lines.

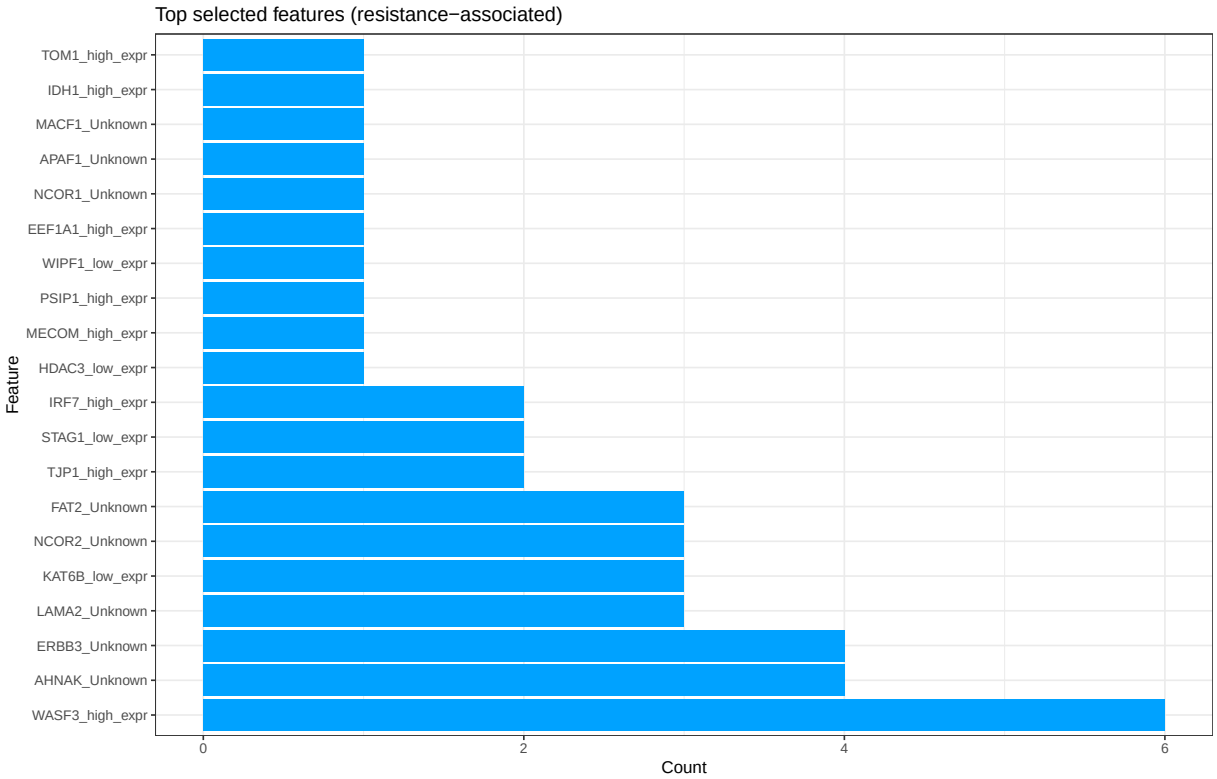


Figure 29: This figure depicts the the frequency of the top 20 selected resistance-associated features (using all models of size 1 to 6 per drug) across all mTOR pathway inhibitors (37) for Setting 1 for the haematological cancer cell lines.

Table 11: Intersection of selected features between non-haematological cell lines and all cell lines

Drug	Database	Setting	P-value (Fisher's exact test)
Afuresertib	GDSC2	1	3.85E-07
AKT-inhibitor-VIII___228	GDSC1	1	1
Alpelisib	GDSC2	1	2.40E-06
AMG-319	GDSC2	1	6.15E-07
Apitolisib	GDSC1	1	2.90E-05
AS605240	GDSC1	1	8.01E-05
AT13148	GDSC2	1	3.21E-03
AT7867	GDSC1	1	6.58E-06
AZD6482___1066	GDSC1	1	8.60E-03
AZD8055	GDSC1	1	2.79E-04
AZD8186	GDSC2	1	1.62E-04
Buparlisib	GDSC2	1	6.55E-07
CX-5461	GDSC1	1	1.64E-03
CZC24832	GDSC2	1	2.35E-01
Dactolisib	GDSC2	1	4.94E-07
GNE-317	GDSC2	1	7.93E-05
GSK1059615	GDSC1	1	5.47E-08
GSK690693	GDSC1	1	2.93E-07
IC-87114	GDSC1	1	2.27E-01
Idelalisib	GDSC1	1	5.11E-02
Ipatasertib	GDSC2	1	1.74E-08
LJI308	GDSC2	1	6.53E-03
MK-2206	GDSC2	1	1.18E-02
Niraparib	GDSC2	1	5.07E-03
NSC319726	GDSC1	1	2.75E-12
Omipalisib	GDSC1	1	6.71E-03
OSI-027	GDSC1	1	5.52E-07
PF-4708671	GDSC1	1	7.27E-08
Pictilisib	GDSC2	1	2.58E-04
PIK-93	GDSC1	1	1.64E-03
Pilaralisib	GDSC1	1	7.78E-09
Rapamycin	GDSC2	1	1.90E-02
Talazoparib	GDSC2	1	7.41E-06
Taselisib	GDSC2	1	2.28E-11
Temsirolimus	GDSC1	1	8.23E-03
Torin-2	GDSC1	1	3.66E-07
Uprosertib___2106	GDSC2	1	2.38E-06
Voxtalisib	GDSC1	1	6.98E-06
WYE-125132	GDSC1	1	7.87E-11
YM201636	GDSC1	1	9.90E-03

ZSTK474	GDSC1	1	1

Table 12: Intersection of selected features between haematological cell lines and all cell lines

Drug	Database	Setting	P-value (Fisher's exact test)
Afuresertib	GDSC2	1	1.66E-01
AKT-inhibitor-VIII___228	GDSC1	1	1
Alpelisib	GDSC2	1	1.81E-01
AMG-319	GDSC2	1	1
Apitolisib	GDSC1	1	1
AS605240	GDSC1	1	1.92E-01
AT13148	GDSC2	1	1
AT7867	GDSC1	1	1.29E-01
AZD6482___1066	GDSC1	1	1
AZD8055	GDSC1	1	1
AZD8186	GDSC2	1	2.03E-02
Buparlisib	GDSC2	1	1
CX-5461	GDSC1	1	1
CZC24832	GDSC2	1	1.39E-01
Dactolisib	GDSC2	1	1.15E-02
GNE-317	GDSC2	1	1
GSK1059615	GDSC1	1	1.45E-01
GSK690693	GDSC1	1	1
IC-87114	GDSC1	1	1
Idelalisib	GDSC1	1	1
Ipatasertib	GDSC2	1	1
LJI308	GDSC2	1	1
MK-2206	GDSC2	1	1
Niraparib	GDSC2	1	1
NSC319726	GDSC1	1	1
Omipalisib	GDSC1	1	1
OSI-027	GDSC1	1	1.46E-01
PF-4708671	GDSC1	1	2.63E-01
Pictilisib	GDSC2	1	3.92E-02
PIK-93	GDSC1	1	1.22E-01
Pilaralisib	GDSC1	1	1
Rapamycin	GDSC2	1	1
Talazoparib	GDSC2	1	1.37E-01
Taselisib	GDSC2	1	1
Temsirolimus	GDSC1	1	1.34E-01
Torin-2	GDSC1	1	1.19E-01
Uprosertib___2106	GDSC2	1	2.56E-01
Voxtalisib	GDSC1	1	1.16E-01

WYE-125132	GDSC1	1	2.93E-02
YM201636	GDSC1	1	1
ZSTK474	GDSC1	1	1

Table 13: Intersection of selected features between non-haematological cell lines and haematological cell lines

Drug	Database	Setting	P-value (Fisher's exact test)
Afuresertib	GDSC2	1	1.76E-01
AKT-inhibitor-VIII_____228	GDSC1	1	1
Alpelisib	GDSC2	1	1.73E-01
AMG-319	GDSC2	1	1
Apitolisib	GDSC1	1	1
AS605240	GDSC1	1	1
AT13148	GDSC2	1	1
AT7867	GDSC1	1	2.22E-01
AZD6482_____1066	GDSC1	1	1
AZD8055	GDSC1	1	1
AZD8186	GDSC2	1	2.51E-01
Buparlisib	GDSC2	1	1
CX-5461	GDSC1	1	1
CZC24832	GDSC2	1	1
Dactolisib	GDSC2	1	1
GNE-317	GDSC2	1	1
GSK1059615	GDSC1	1	1
GSK690693	GDSC1	1	1
IC-87114	GDSC1	1	2.10E-01
Idelalisib	GDSC1	1	1
Ipatasertib	GDSC2	1	1
LJI308	GDSC2	1	1
MK-2206	GDSC2	1	1
Niraparib	GDSC2	1	1
NSC319726	GDSC1	1	1
Omipalisib	GDSC1	1	1
OSI-027	GDSC1	1	3.63E-03
PF-4708671	GDSC1	1	2.41E-01
Pictilisib	GDSC2	1	1
PIK-93	GDSC1	1	1
Pilaralisib	GDSC1	1	1
Rapamycin	GDSC2	1	1
Talazoparib	GDSC2	1	1.76E-01
Taselisib	GDSC2	1	1
Temsirolimus	GDSC1	1	1

Torin-2	GDSC1	1	1
Uprosertib___2106	GDSC2	1	2.24E-01
Voxtalisib	GDSC1	1	8.97E-02
WYE-125132	GDSC1	1	2.52E-01
YM201636	GDSC1	1	1
ZSTK474	GDSC1	1	1

Table 14: MERIDA's output rules for each drug (best model selected based on Youden's J)

Drug	Setting	Weight function	Rule size	Model rule
AZD8055	1	linear	5	FKBP5 high expression ∨ VHL high expression ∨ ROBO2 high expression ∨ TRIO low expression ∨ NCKAP1 low expression
AZD8055	1	quadratic	4	ALK Gain of function ∨ FKBP5 high expression ∨ VHL high expression ∨ NCKAP1 low expression
AZD8055	1 (iteration 4)	quadratic	7	(ALK Gain of function ∨ FKBP5 high expression ∨ VHL high exprression ∨ TRIO low expression ∨ NCKAP1 low expression) ∧¬ (TBX3 high expression ∨ PSME3 low expression)
AZD8055	1	cubic	4	ALK Gain of function ∨ FKBP5 high expression ∨ VHL high expression ∨ NCKAP1 low expression
AZD8055	1 (iteration 4)	cubic	7	(ALK Gain of function ∨ FKBP5 high expression ∨ VHL high expression ∨ PTEN high expression ∨ NCKAP1 low expression) ∧¬ (RBBP7 low expression ∨ HDAC3 low expression)
CX-5461	1	linear	2	CLASP2 high expression ∨ TJP1 low expression
CX-5461	2	linear	2 (5)*	known sensitivity determinants ∨ CLASP2 high expression ∨ TJP1 low expression
CX-5461	3	linear	2	CLASP2 high expression ∨ TJP1 low expression
CX-5461	1	quadratic	2	TCF4 high expression ∨ TJP1 low expression
CX-5461	2	quadratic	2 (5)*	known sensitivity determinants ∨ CLASP2 high expression ∨ TJP1 low expression
CX-5461	3	quadratic	2	TCF4 high expression ∨ TJP1 low expression

CX-5461	1	cubic	4	CLASP2 high expression ∨ PRPF8 high expression ∨ LIMA1 low expression ∨ TJP1 low expression
CX-5461	2	cubic	2 (5)*	known sensitivity determinants ∨ CLASP2 high expression ∨ TJP1 low expression
CX-5461	3	cubic	4	CLASP2 high expression ∨ PRPF8 high expression ∨ LIMA1 low expression ∨ TJP1 low expression
Dactolisib	1	linear	2	MGA Unknown mutation ∨ TJP1 low expression
Dactolisib	2	linear	2 (3)*	MGA Unknown mutation ∨ known sensitivity determinants ∨ TJP1 low expression
Dactolisib	3	linear	2	MGA Unknown mutation ∨ TJP1 low expression
Dactolisib	1	quadratic	2	MGA Unknown mutation ∨ TJP1 low expression
Dactolisib	2	quadratic	2 (3)*	MGA Unknown mutation ∨ known sensitivity determinants ∨ TJP1 low expression
Dactolisib	3	quadratic	2	MGA Unknown mutation ∨ TJP1 low expression
Dactolisib	1	cubic	2	MGA Unknown mutation ∨ TJP1 low expression
Dactolisib	2	cubic	2 (3)*	MGA Unknown mutation ∨ known sensitivity determinants ∨ TJP1 low expression
Dactolisib	3	cubic	2	MGA Unknown mutation ∨ TJP1 low expression
Niraparib	1	linear	5	ARID4B high expression ∨ ARID2 high expression ∨ KLF6 low expression ∨ TJP1 low expression ∨ ARFGEF2 low expression
Niraparib	1	quadratic	4	ARID2 high expression ∨ CASP8 low expression ∨ KLF6 low expression ∨ TJP1 low expression

Niraparib	1 (iteration 4)	quadratic	7	ARID4B high expression ∨ WASF3 high expression ∨ ARID2 high expression ∨ CASP8 low expression ∨ KLF6 low expression ∨ CTTN low expression ∨ TJP1 low expression
Niraparib	1	cubic	4	ARID2 high expression ∨ CASP8 low expression ∨ KLF6 low expression ∨ TJP1 low expression
Niraparib	1 (iteration 4)	cubic	10	(ARID4B high expression ∨ WASF3 high expression ∨ ARID2 high expression ∨ CASP8 low expression ∨ KLF6 low expression ∨ CTTN low expression ∨ TJP1 low expression ∨ ARNTL low expression) ∧¬ (FAT2 high expression ∨ ATF1 low expression)
NSC319726	1	linear	5	CDC73 high expression ∨ TCF12 high expression ∨ TCF4 high expression ∨ TJP1 low expression ∨ ADAM10 low expression
NSC319726	1	quadratic	4	TCF12 high expression ∨ ROBO2 high expression ∨ TJP1 low expression ∨ ADAM10 low expression
NSC319726	1 (iteration 4)	quadratic	9	(TCF12 high expression ∨ TGFBR2 high expression ∨ IDH2 high expression ∨ ROBO2 high expression ∨ KLF6 low expression ∨ TJP1 low expression ∨ ADAM10 low expression ∨ RB1 low expression) ∧¬ RHOA low expression
NSC319726	1	cubic	3	TCF12 high expression ∨ ROBO2 high expression ∨ TJP1 low expression

NSC319726	1 (iteration 4)	cubic	6	TCF12 high expression ∨ TGFBR2 high expression ∨ IDH2 high expression ∨ ROBO2 high expression ∨ TJP1 low expression ∨ ADAM10 low expression
Omipalisib	1	linear	2	PIK3CA Gain of function ∨ TJP1 low expression
Omipalisib	1	quadratic	2	AHCTF1 high expression ∨ TJP1 low expression
Omipalisib	1 (iteration 4)	quadratic	5	(SH2B3 high expression ∨ AHCTF1 high expression ∨ STAG2 low expression ∨ TJP1 low expression) ∧ ¬ BCLAF1 low expression
Omipalisib	1	cubic	4	STAG2 Loss of function ∨ SH2B3 high expression ∨ AHCTF1 high expression ∨ TJP1 low expression
Omipalisib	1 (iteration 4)	cubic	11	(TP53BP1 Loss of function ∨ PPP2R1A Unknown mutation ∨ STAG2 Loss of function ∨ CCNE1 CNV gain ∨ SH2B3 high expression ∨ AHCTF1 high expression ∨ TJP1 low expression ∨ ARID1A low expression ∨ PTPRF low expression) ∧ ¬ (NTN4 Unknown mutation ∨ TAOK2 low expression)
Rapamycin	1	linear	2	MAP4K1 high expression ∨ PTPRF low expression
Rapamycin	2	linear	1 (42)*	known sensitivity determinants ∨ TJP1 low expression
Rapamycin	3	linear	2	MAP4K1 high expression ∨ PTPRF low expression
Rapamycin	1	quadratic	2	MAP4K1 high expression ∨ PTPRF low expression
Rapamycin	2	quadratic	2 (43)*	known sensitivity determinants ∨ NCKAP1 low expression ∨ TJP1 low expression

Rapamycin	3	quadratic	5	(STK4 Unknown mutation ∨ PRKAR1A Unknown mutation ∨ MAP4K1 high expression ∨ PTPRF low expression) ∧¬ ARFGEF2 Unknown mutation
Rapamycin	1	cubic	4	(TBL1XR1 Unknown mutation ∨ CYTH4 high expression ∨ MAP4K1 high expression) ∧¬ WT1 high expression
Rapamycin	2	cubic	2 (43)*	known sensitivity determinants ∨ NCKAP1 low expression ∨ TJP1 low expression
Rapamycin	3	cubic	1	MAP4K1 high expression
Talazoparib	1	linear	3	TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	2	linear	2 (5)*	known sensitivity determinants ∨ TCF12 high expression ∨ TJP1 low expression
Talazoparib	3	linear	3	TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	1	quadratic	3	TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	2	quadratic	3 (6)*	known sensitivity determinants ∨ TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	3	quadratic	3	TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	1	cubic	3	TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	2	cubic	3 (6)*	known sensitivity determinants ∨ TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression
Talazoparib	3	cubic	3	TCF12 high expression ∨ TJP1 low expression ∨ HLA-B low expression

Temsirolimus	1	linear	5	PTEN Loss of function ∨ SMARCB1 high expression ∨ WIPF1 high expression ∨ PRRX1 high expression ∨ DHX9 low expression
Temsirolimus	2	linear	1 (16)*	known sensitivity determinants ∨ TJP1 low expression
Temsirolimus	3	linear	5	PTEN Loss of function ∨ SMARCB1 high expression ∨ WIPF1 high expression ∨ PRRX1 high expression ∨ DHX9 low expression
Temsirolimus	1	quadratic	5	PTEN Loss of function ∨ MMP2 high expression ∨ SMARCB1 high expression ∨ WIPF1 high expression ∨ DHX9 low expression
Temsirolimus	2	quadratic	1 (16)*	known sensitivity determinants ∨ TJP1 low expression
Temsirolimus	3	quadratic	5	PTEN Loss of function ∨ MMP2 high expression ∨ SMARCB1 high expression ∨ WIPF1 high expression ∨ DHX9 low expression
Temsirolimus	1	cubic	4	PTEN Loss of function ∨ MMP2 high expression ∨ SMARCB1 high expression ∨ WIPF1 high expression
Temsirolimus	2	cubic	3 (18)*	(PTEN Loss of function ∨ known sensitivity determinants ∨ MAP4K1 high expression) ∧¬ MAGI2 Unknown mutation
Temsirolimus	3	cubic	4	PTEN Loss of function ∨ MMP2 high expression ∨ SMARCB1 high expression ∨ WIPF1 high expression
Voxtalisib	1	linear	1	TJP1 low expression
Voxtalisib	1	quadratic	5	BRAF Gain of function ∨ PPP2R5A high expression ∨ XPO1 high expression ∨ BCL11A high expression ∨ CSNK1G3 low expression

Voxtalisib	1 (iteration 4)	quadratic	13	(BRAF Gain of function ∨ CDK12 Unknown mutation ∨ TGFBR1 CNV gain ∨ PPP2R5A high expression ∨ CNOT4 high expression ∨ XPO1 high expression ∨ BCL11A high expression ∨ CSNK1G3 low expression) ∧ ¬ (CHD6 Unknown mutation ∨ HDAC9 Unknown mutation ∨ CAT Unknown mutation ∨ SMURF2 high expression ∨ CAST low expression)
Voxtalisib	1	cubic	6	(MITF CNV gain ∨ PPP2R5A high expression ∨ XPO1 high expression ∨ BCL11A high expression ∨ CSNK1G3 low expression) ∧¬ CHD6 Unknown mutation
Voxtalisib	1 (iteration 4)	cubic	10	(MTOR neutral mutations ∨ MITF CNV gain ∨ PPP2R5A high expression ∨ XPO1 high expression ∨ BCL11A high expression ∨ PIK3C2B low expression ∨ CSNK1G3 low expression) ∧¬ (CHD6 Unknown mutation ∨ NCF2 high expression ∨ PIK3R3 high expression)

* Known sensitivity determinants for the respective drugs can be found in Supplement 1. The provided rule size denotes the number of terms without known sensitivity determinants, while the number in brackets includes them.

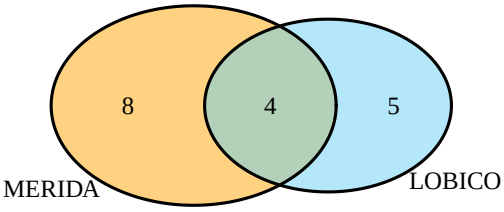
Table 15: LOBICO's output rules for each drug (best model selected based on Youden's J)

Drug	Setting	Weight function	Rule size	Model rule
AZD8055	1	linear	1	TJP1 low expression
AZD8055	1	quadratic	1	NCKAP1 low expression
AZD8055	1	cubic	1	NCKAP1 low expression
CX-5461	3	linear	2	TJP1 low expression ∨ CLASP2 high expression
CX-5461	3	quadratic	2	TJP1 low expression ∨ TCF4 high expression

CX-5461	3	cubic	2	TJP1 low expression $\vee$ TCF4 high expression
Dactolisib	3	linear	2	MGA Unknown mutation $\vee$ TJP1 low expression
Dactolisib	3	quadratic	2	MGA Unknown mutation $\vee$ TJP1 low expression
Dactolisib	3	cubic	2	MGA Unknown mutation $\vee$ TJP1 low expression
Niraparib	1	linear	2	TJP1 low expression $\vee$ CASP8 low expression
Niraparib	1	quadratic	2	TJP1 low expression $\vee$ KLF6 low expression
Niraparib	1	cubic	2	TJP1 low expression $\vee$ KLF6 low expression
NSC319726	1	linear	2	TJP1 low expression $\vee$ TCF12 high expression
NSC319726	1	quadratic	2	TJP1 low expression $\vee$ TCF12 high expression
NSC319726	1	cubic	2	TJP1 low expression $\vee$ TCF12 high expression
Omipalisib	1	linear	2	PIK3CA Gain of function $\vee$ TJP1 low expression
Omipalisib	1	quadratic	2	AHCTF1 high expression $\vee$ TJP1 low expression
Omipalisib	1	cubic	1	TJP1 low expression
Rapamycin	3	linear	2	PTPRF low expression $\vee$ MAP4K1 high expression
Rapamycin	3	quadratic	2	MAP4K1 high expression $\wedge\neg$ PTPRU Unknown mutation
Rapamycin	3	cubic	1	MAP4K1 high expression
Talazoparib	3	linear	2	TCF12 high expression $\vee$ TJP1 low expression
Talazoparib	3	quadratic	2	TCF12 high expression $\vee$ TJP1 low expression
Talazoparib	3	cubic	2	TCF12 high expression $\vee$ TJP1 low expression
Temsirolimus	3	linear	1	TJP1 low expression
Temsirolimus	3	quadratic	1	TJP1 low expression
Temsirolimus	3	cubic	2	MAP4K1 high expression $\vee$ MMP2 high expression
Voxtalisib	1	linear	1	TJP1 low expression
Voxtalisib	1	quadratic	2	SCAI high expression $\vee$ NCKAP1 low expression

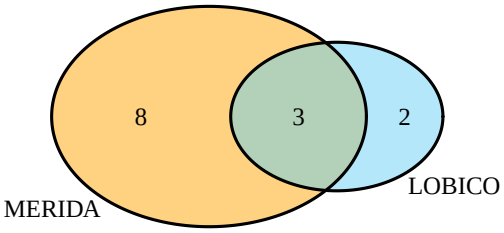
Voxtalisib	1	cubic	2	NCKAP1 low expression $\wedge\neg$ MLH1 high expression

AZD8055 p−value: 2.91e−07



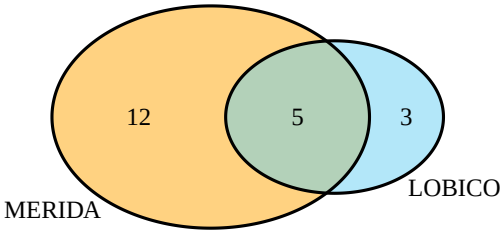
(a) **AZD8055**

Omipalisib p−value: 2.89e−06



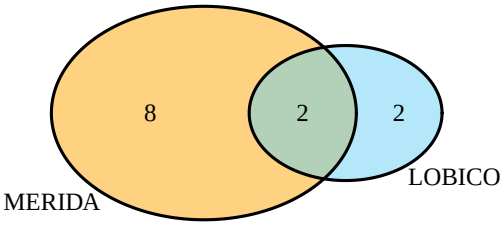
(b) **Omipalisib**

Voxtalisib p−value: 5.37e−09



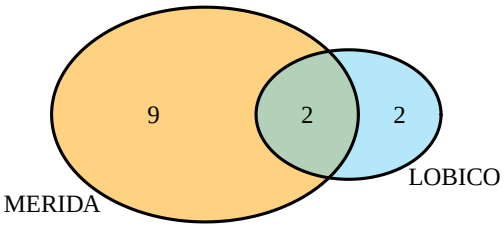
(c) **Voxtalisib**

Niraparib p−value: 2.46e−04



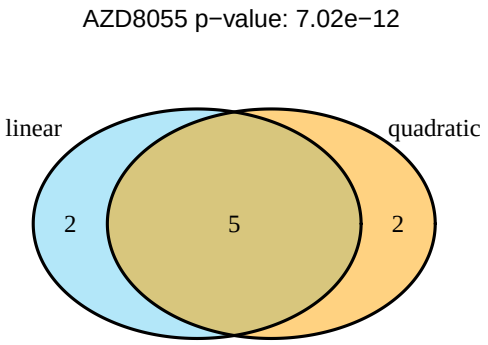
(d) **Niraparib**

NSC319726 p−value: 2.95e−04

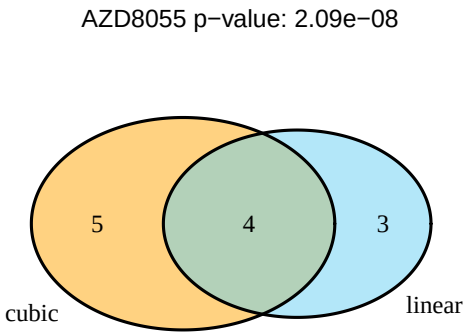


(e) **NSC319726**

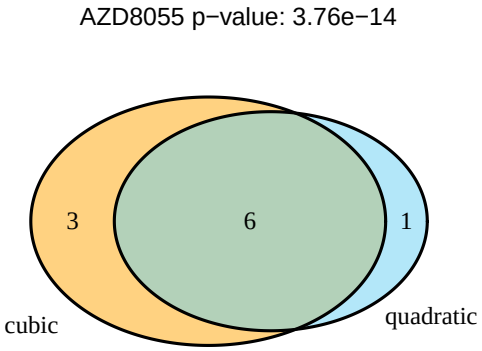
Figure 30: In Sub-Figure (a - e), we present Venn diagrams with the respective Fisher's exact test overlap p-value for the selected features between LOBICO and MERIDA.



(a) **AZD8055 linear vs quadratic**

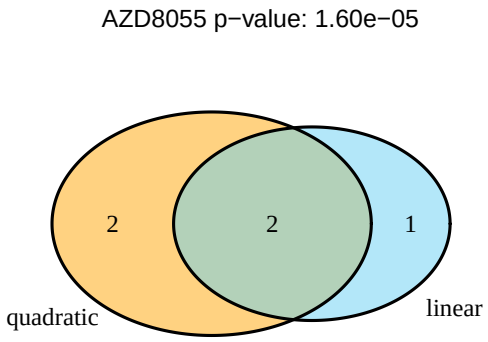


(b) **AZD8055 linear vs cubic**

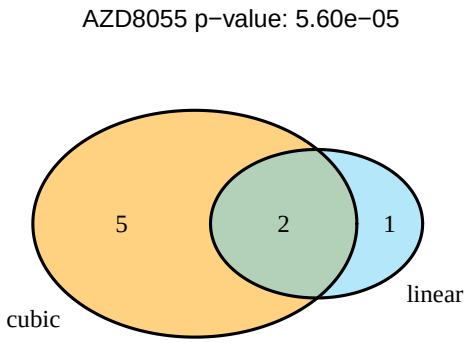


(c) **AZD8055 quadratic vs cubic**

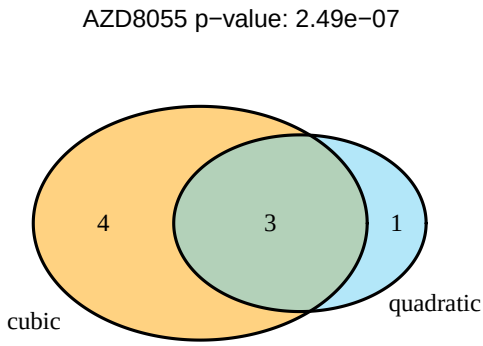
Figure 31: In Sub-Figure (a - c), we present Venn diagrams with the respective Fisher's exact test overlap p-value for the selected features of MERIDA for AZD8055 between the three different weight functions. The remaining overlap p-values for all other drugs can be found in Table 16.



(a) **AZD8055 linear vs quadratic**

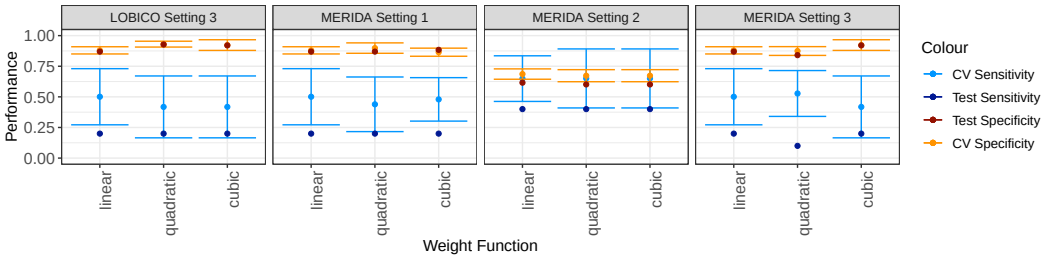


(b) **AZD8055 linear vs cubic**

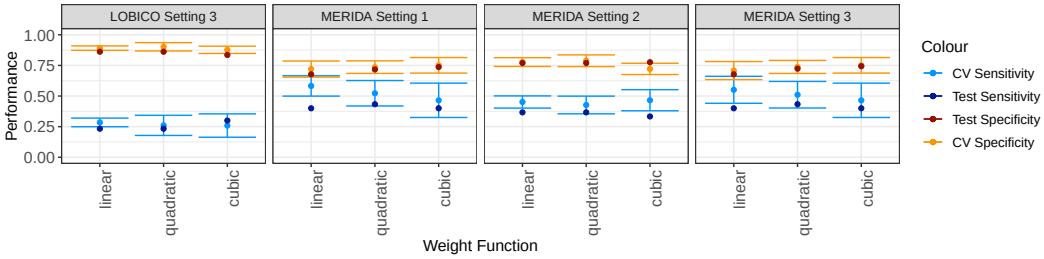


(c) **AZD8055 quadratic vs cubic**

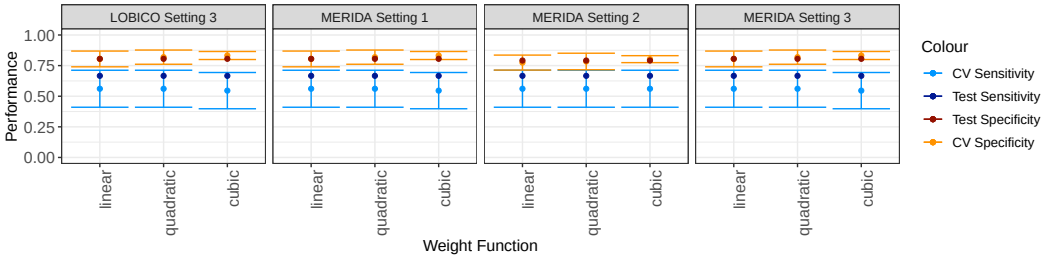
Figure 32: In Sub-Figure (a - c), we present Venn diagrams with the respective Fisher's exact test overlap p-value for the selected features of LOBICO for AZD8055 between the three different weight functions. The remaining overlap p-values for all other drugs can be found in Table 17.



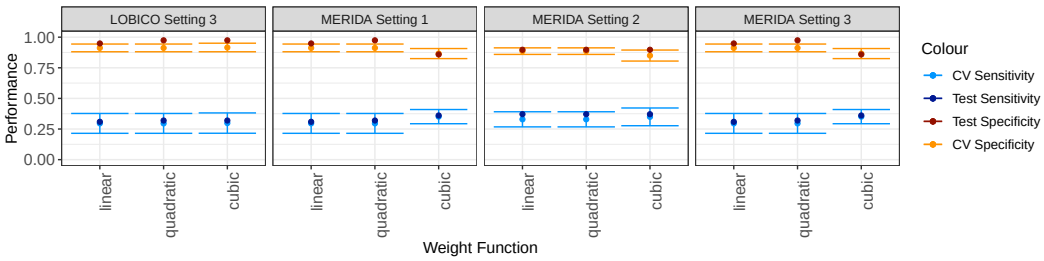
(a) **Rapamycin**



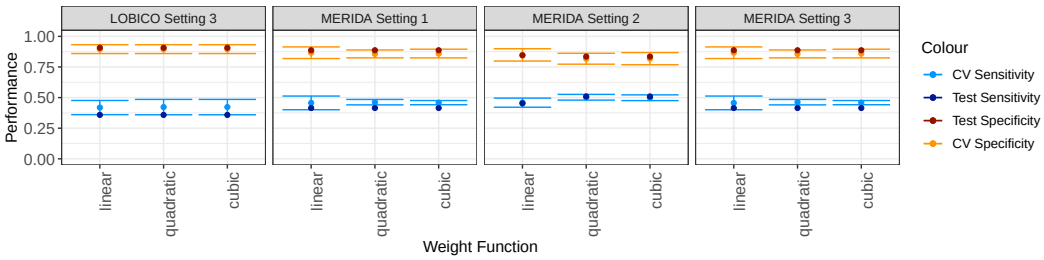
(b) **Temsirolimus**



(c) **Dactolisib**

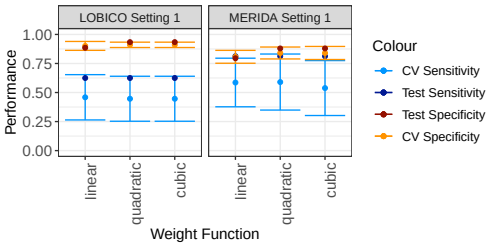


(d) **CX-5461**

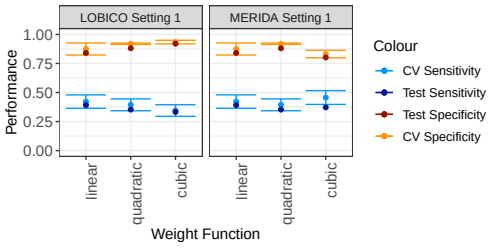


(e) **Talazoparib**

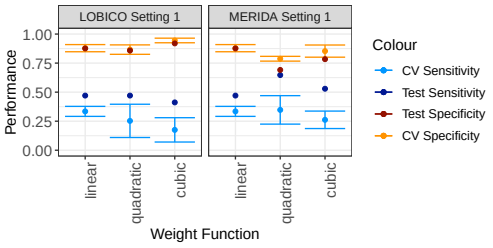
Figure 33: In Sub-Figure (a - e), we present the statistical performance for all drugs with a priori knowledge during CV as well as on a test set.



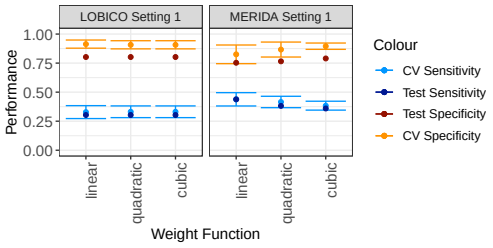
(a) **AZD8055**



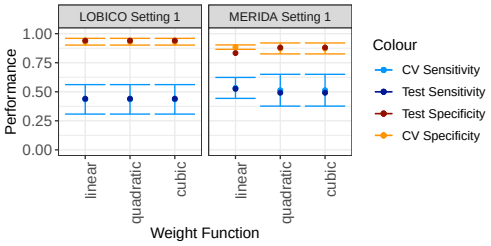
(b) **Omipalisib**



(c) **Voxtalisib**

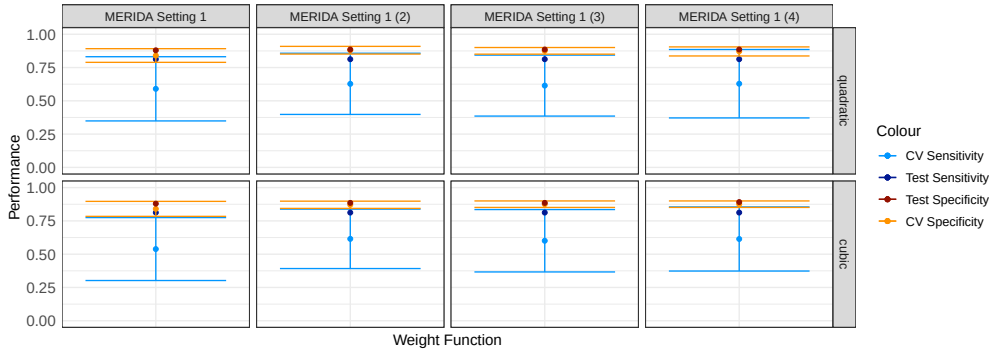


(d) **NSC319726**

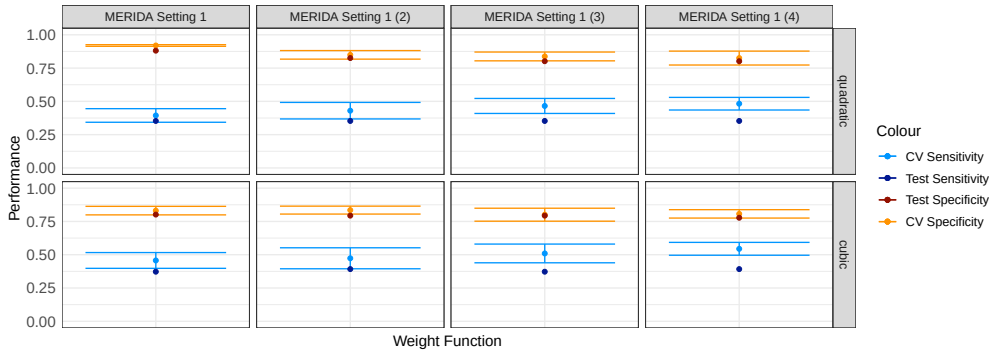


(e) **Niraparib**

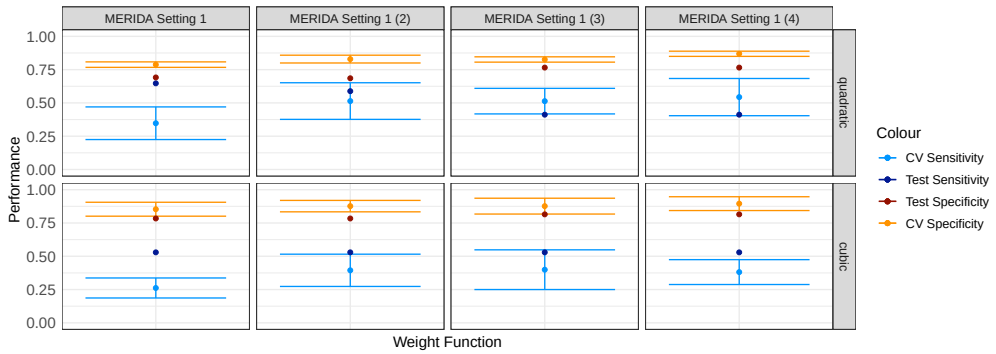
Figure 34: In Sub-Figure (a - e), we present the statistical performance for the drugs without a priori knowledge during CV as well as on a test set.



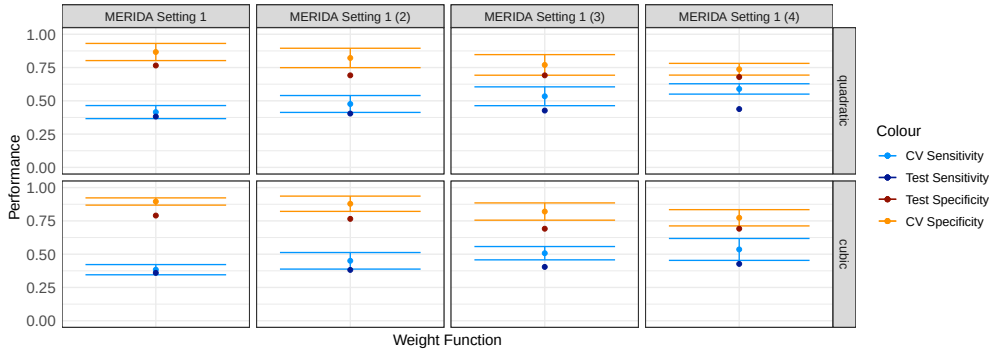
(a) **AZD8055**



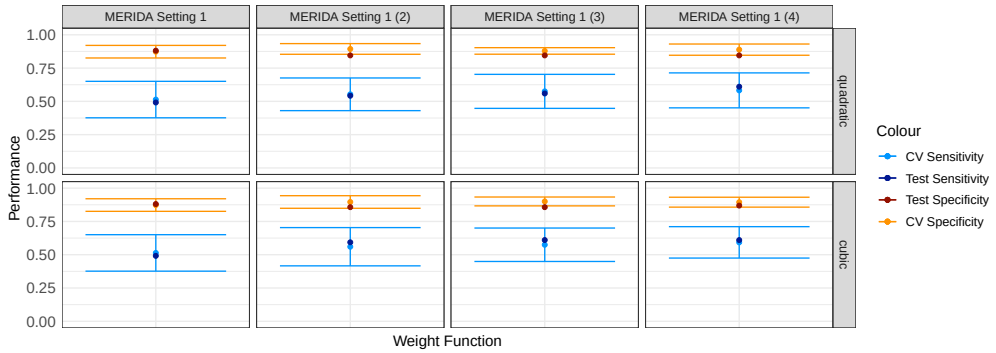
(b) **Omipalisib**



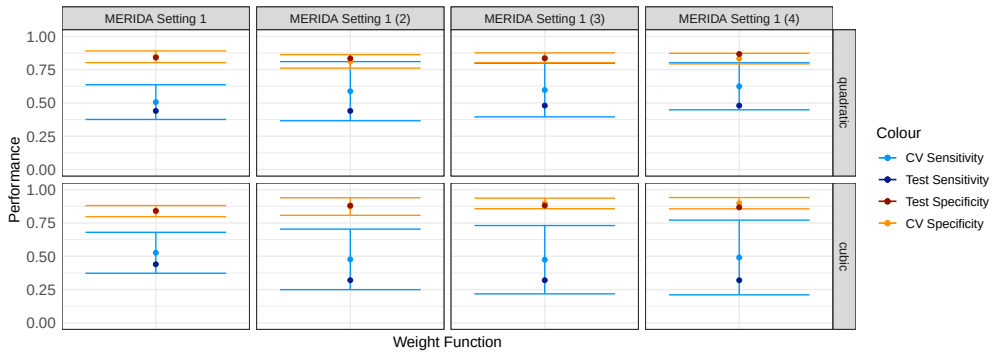
(c) **Voxtalisib**



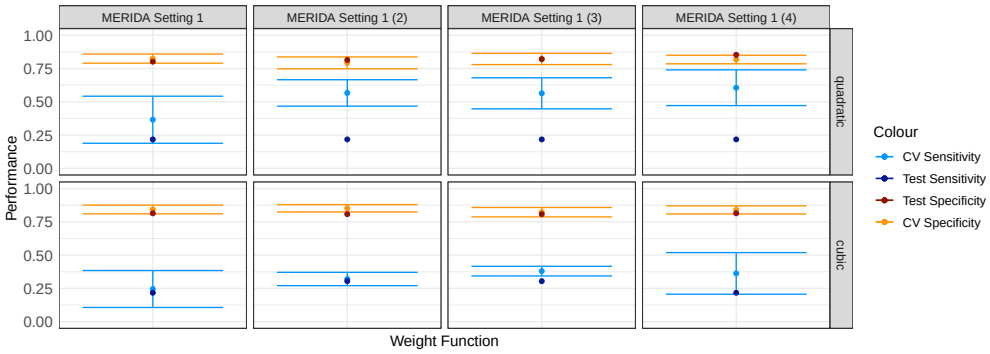
(d) **NSC319726**



(e) **Niraparib**



(f) **AS605240**



(g) **Pilaralisib**

Figure 35: In Sub-Figure (a - g), we present the statistical performance for drugs without a priori knowledge during CV as well as on a test set when iteratively using the features from previous applications of our method as a priori knowledge for a new run.

Table 16: Similarity of selected features for different weight functions for MERIDA

Drug	Database	Setting	Compared weight functions	P-value (Fisher's exact test)
AZD8055	GDSC1	1	linear, quadratic	7.02E-12
AZD8055	GDSC1	1	linear, cubic	2.09E-08
AZD8055	GDSC1	1	quadratic, cubic	3.76E-14
CX-5461	GDSC1	2	linear, quadratic	4.95E-14
CX-5461	GDSC1	2	linear, cubic	1.15E-08
CX-5461	GDSC1	2	quadratic, cubic	1.24E-14
CX-5461	GDSC1	3	linear, quadratic	4.21E-06
CX-5461	GDSC1	3	linear, cubic	2.77E-04
CX-5461	GDSC1	3	quadratic, cubic	9.29E-14
CX-5461	GDSC1	1	linear, quadratic	1.98E-04
CX-5461	GDSC1	1	linear, cubic	2.77E-04
CX-5461	GDSC1	1	quadratic, cubic	4.42E-16
Dactolisib	GDSC2	2	linear, quadratic	6.19E-22
Dactolisib	GDSC2	2	linear, cubic	2.72E-18
Dactolisib	GDSC2	2	quadratic, cubic	3.40E-20
Dactolisib	GDSC2	3	linear, quadratic	2.47E-20
Dactolisib	GDSC2	3	linear, cubic	4.07E-17
Dactolisib	GDSC2	3	quadratic, cubic	3.34E-22
Dactolisib	GDSC2	1	linear, quadratic	2.49E-20
Dactolisib	GDSC2	1	linear, cubic	4.09E-17
Dactolisib	GDSC2	1	quadratic, cubic	3.36E-22
Niraparib	GDSC2	1	linear, quadratic	1.30E-11
Niraparib	GDSC2	1	linear, cubic	1.30E-11
Niraparib	GDSC2	1	quadratic, cubic	7.01E-17
NSC319726	GDSC1	1	linear, quadratic	2.02E-06
NSC319726	GDSC1	1	linear, cubic	2.02E-06
NSC319726	GDSC1	1	quadratic, cubic	6.62E-17
Omipalisib	GDSC1	1	linear, quadratic	5.74E-09
Omipalisib	GDSC1	1	linear, cubic	2.15E-06
Omipalisib	GDSC1	1	quadratic, cubic	3.08E-15
Rapamycin	GDSC2	2	linear, quadratic	1.41E-14
Rapamycin	GDSC2	2	linear, cubic	1.69E-10
Rapamycin	GDSC2	2	quadratic, cubic	9.03E-16
Rapamycin	GDSC2	3	linear, quadratic	1.31E-05
Rapamycin	GDSC2	3	linear, cubic	1.17E-03
Rapamycin	GDSC2	3	quadratic, cubic	1.31E-05
Rapamycin	GDSC2	1	linear, quadratic	7.73E-06
Rapamycin	GDSC2	1	linear, cubic	1.23E-05
Rapamycin	GDSC2	1	quadratic, cubic	1.37E-14
Talazoparib	GDSC2	2	linear, quadratic	9.17E-11

Talazoparib	GDSC2	2	linear, cubic	3.73E-08
Talazoparib	GDSC2	2	quadratic, cubic	7.67E-12
Talazoparib	GDSC2	3	linear, quadratic	1.14E-22
Talazoparib	GDSC2	3	linear, cubic	8.24E-18
Talazoparib	GDSC2	3	quadratic, cubic	2.50E-20
Talazoparib	GDSC2	1	linear, quadratic	1.13E-22
Talazoparib	GDSC2	1	linear, cubic	8.20E-18
Talazoparib	GDSC2	1	quadratic, cubic	2.49E-20
Temsirolimus	GDSC1	2	linear, quadratic	1.24E-14
Temsirolimus	GDSC1	2	linear, cubic	1.86E-11
Temsirolimus	GDSC1	2	quadratic, cubic	1.90E-17
Temsirolimus	GDSC1	3	linear, quadratic	6.98E-12
Temsirolimus	GDSC1	3	linear, cubic	1.16E-08
Temsirolimus	GDSC1	3	quadratic, cubic	1.24E-14
Temsirolimus	GDSC1	1	linear, quadratic	6.93E-12
Temsirolimus	GDSC1	1	linear, cubic	1.15E-08
Temsirolimus	GDSC1	1	quadratic, cubic	1.24E-14
Voxtalisib	GDSC1	1	linear, quadratic	2.44E-12
Voxtalisib	GDSC1	1	linear, cubic	2.89E-07
Voxtalisib	GDSC1	1	quadratic, cubic	1.15E-19

Table 17: Similarity of selected features for different weight functions for LOBICO

Drug	Database	Setting	Compared weight functions	P-value (Fisher's exact test)
AZD8055	GDSC1	1	linear, quadratic	1.60E-05
AZD8055	GDSC1	1	linear, cubic	5.60E-05
AZD8055	GDSC1	1	quadratic, cubic	2.49E-07
CX-5461	GDSC1	3	linear, quadratic	7.98E-06
CX-5461	GDSC1	3	linear, cubic	5.98E-03
CX-5461	GDSC1	3	quadratic, cubic	7.98E-06
Dactolisib	GDSC2	3	linear, quadratic	1.87E-08
Dactolisib	GDSC2	3	linear, cubic	1.87E-08
Dactolisib	GDSC2	3	quadratic, cubic	1.73E-14
Niraparib	GDSC2	1	linear, quadratic	8.25E-06
Niraparib	GDSC2	1	linear, cubic	8.25E-06
Niraparib	GDSC2	1	quadratic, cubic	1.87E-09
NSC319726	GDSC1	1	linear, quadratic	1.81E-09
NSC319726	GDSC1	1	linear, cubic	8.10E-06
NSC319726	GDSC1	1	quadratic, cubic	8.10E-06
Omipalisib	GDSC1	1	linear, quadratic	5.98E-03
Omipalisib	GDSC1	1	linear, cubic	5.98E-03
Omipalisib	GDSC1	1	quadratic, cubic	1.77E-09
Rapamycin	GDSC2	3	linear, quadratic	1.80E-10

Rapamycin	GDSC2	3	linear, cubic	2.64E-07
Rapamycin	GDSC2	3	quadratic, cubic	3.52E-15
Talazoparib	GDSC2	3	linear, quadratic	8.30E-06
Talazoparib	GDSC2	3	linear, cubic	8.30E-06
Talazoparib	GDSC2	3	quadratic, cubic	1.88E-09
Temsirolimus	GDSC1	3	linear, quadratic	7.99E-06
Temsirolimus	GDSC1	3	linear, cubic	1
Temsirolimus	GDSC1	3	quadratic, cubic	1
Voxtalisib	GDSC1	1	linear, quadratic	1
Voxtalisib	GDSC1	1	linear, cubic	1.19E-02
Voxtalisib	GDSC1	1	quadratic, cubic	3.55E-08

Table 18: Comparison between selected features for LOBICO vs MERIDA

Drug	Database	Setting	P-value (Fisher's exact test)
AZD8055	GDSC1	1	2.91E-07
CX-5461	GDSC1	3	6.39E-06
Dactolisib	GDSC2	3	8.42E-06
Niraparib	GDSC2	1	2.46E-04
NSC319726	GDSC1	1	2.95E-04
Omipalisib	GDSC1	1	2.89E-06
Rapamycin	GDSC2	3	2.11E-13
Talazoparib	GDSC2	3	4.27E-04
Temsirolimus	GDSC1	3	1.53E-10
Voxtalisib	GDSC1	1	5.37E-09

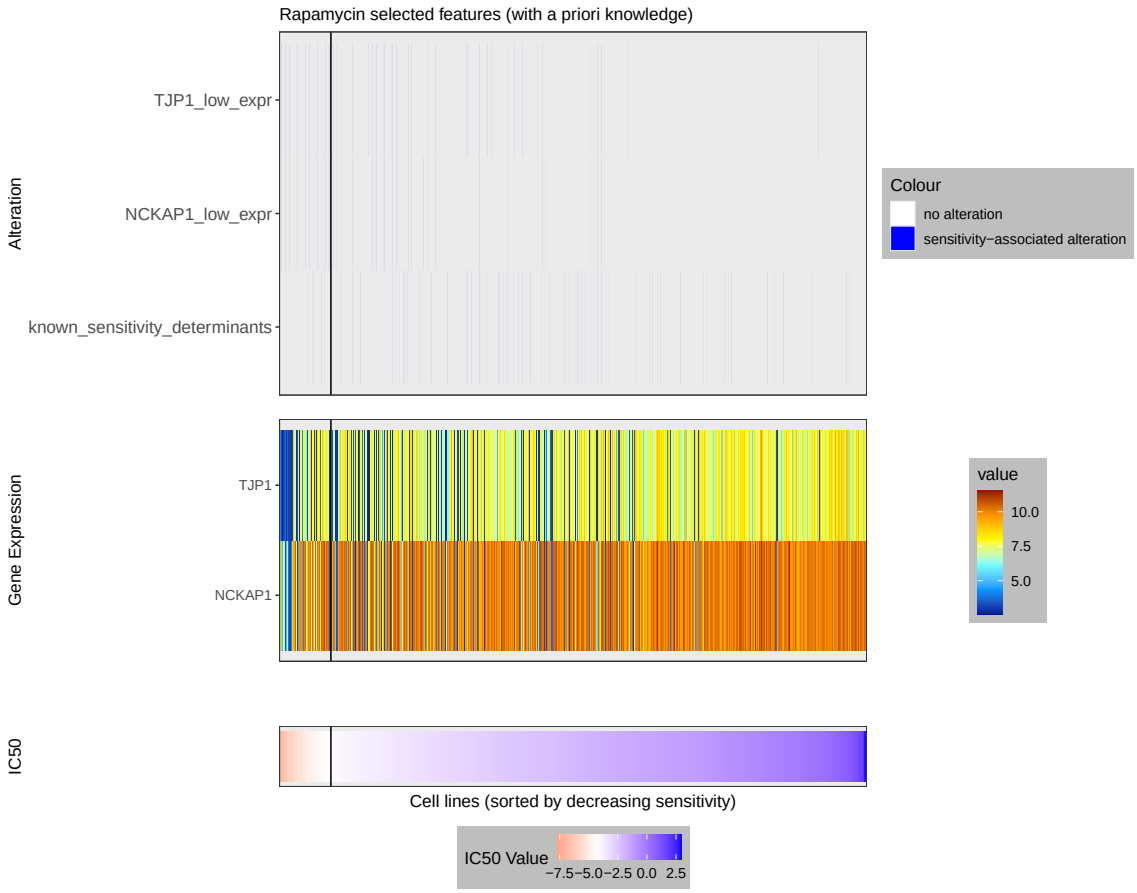


Figure 36: This figure depicts the selected features for the best model (selected with Youden's J) for Rapamycin using the quadratic weight function. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value.

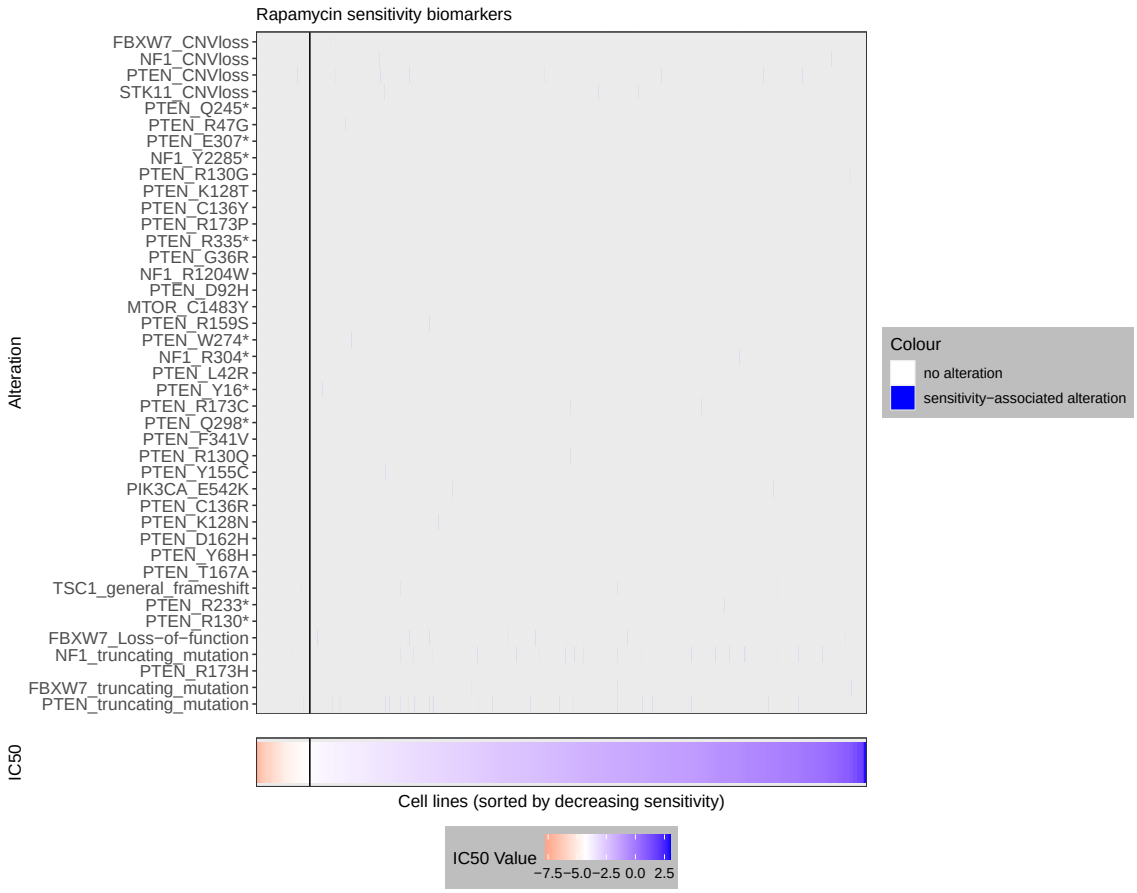


Figure 37: This figure depicts the a priori knowledge (i.e. the sensitivity biomarkers) for Rapamycin and its distribution throughout the cell lines. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value.

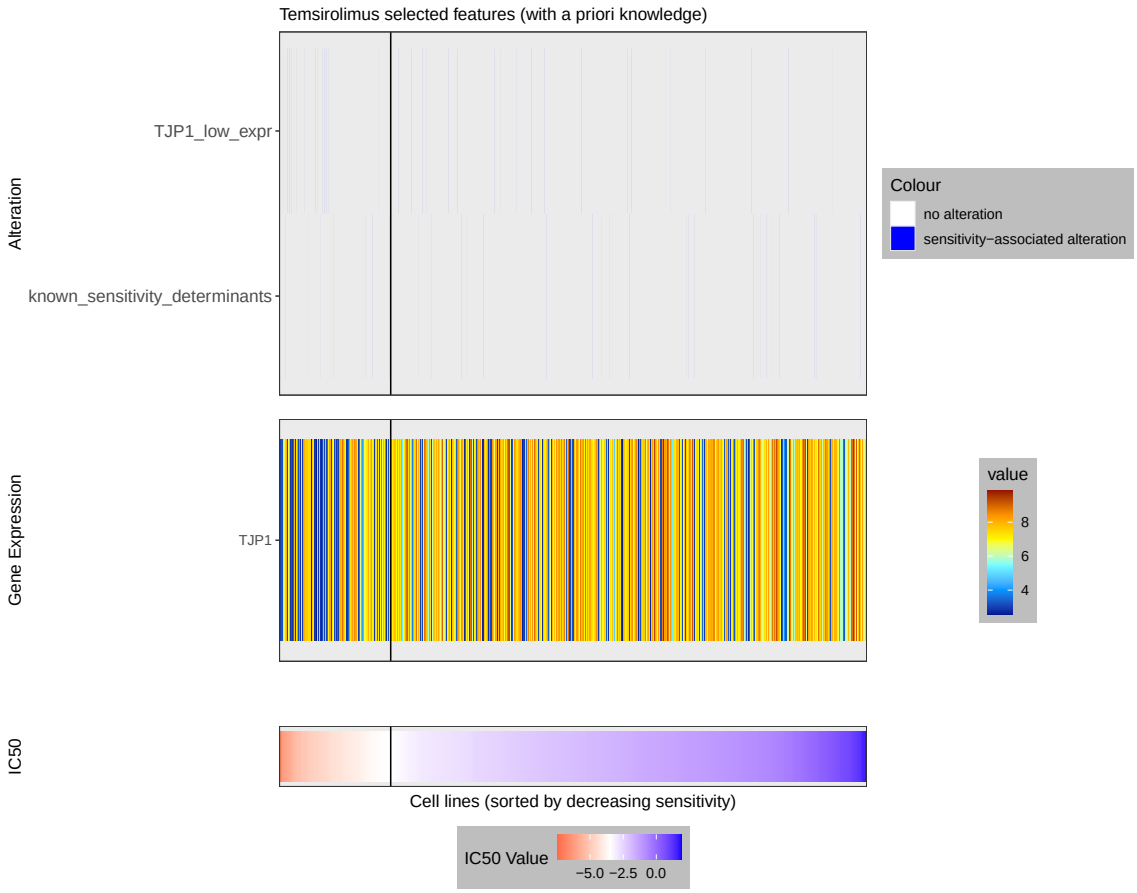


Figure 38: This figure depicts the selected features for the best model (selected with Youden's J) for Temsirolimus using the quadratic weight function. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value.

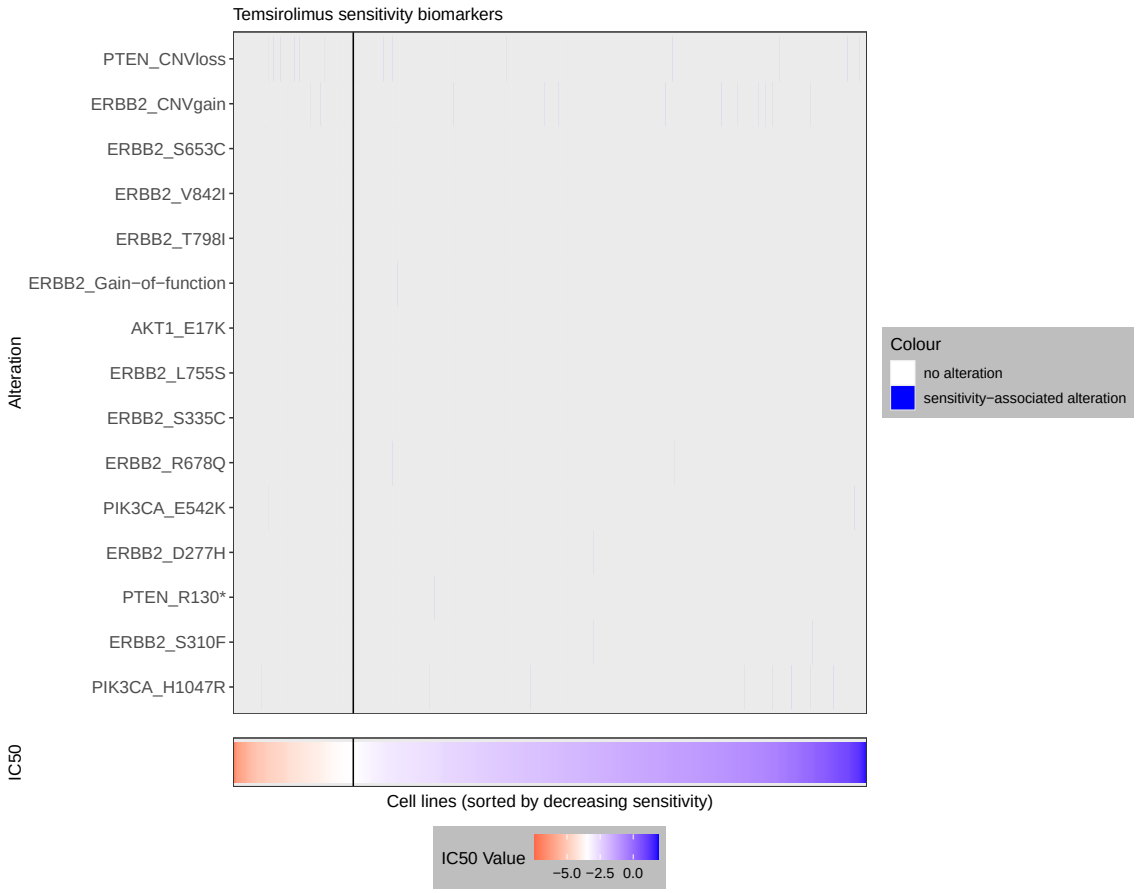


Figure 39: This figure depicts the a priori knowledge (i.e. the sensitivity biomarkers) for Temsirolimus and its distribution throughout the cell lines. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value.

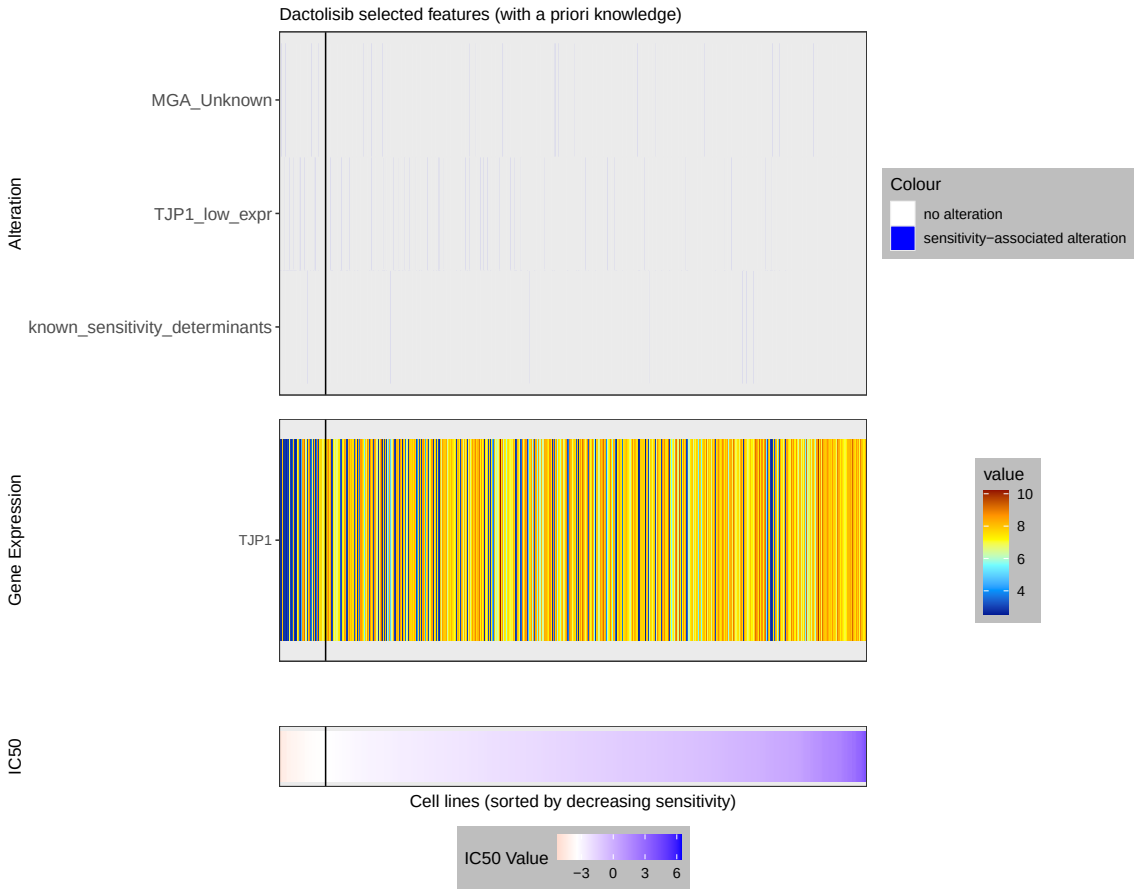


Figure 40: This figure depicts the selected features for the best model (selected with Youden's J) for Dactolisib using the quadratic weight function. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value. Note that the a priori knowledge consists only of one mutation (PIK3CA H1047R).

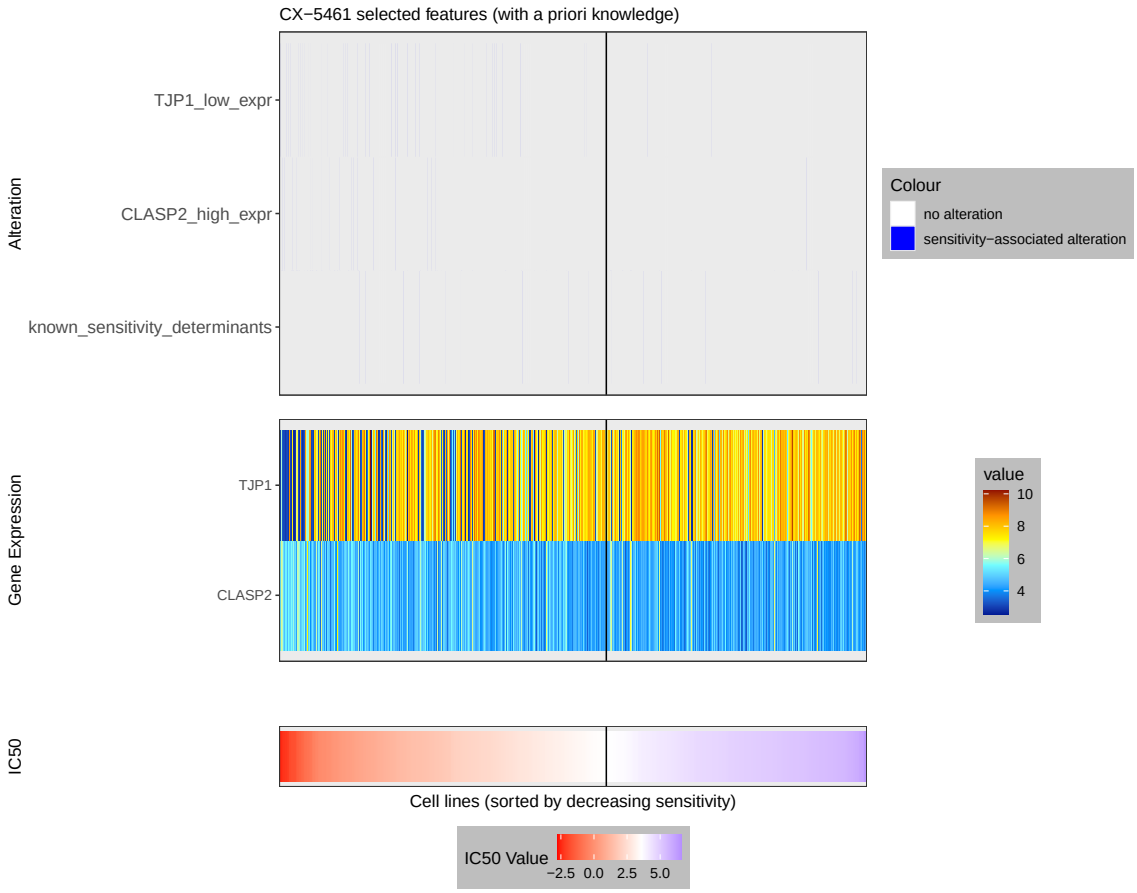


Figure 41: This figure depicts the selected features for the best model (selected with Youden's J) for CX-5461 using the quadratic weight function. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value.

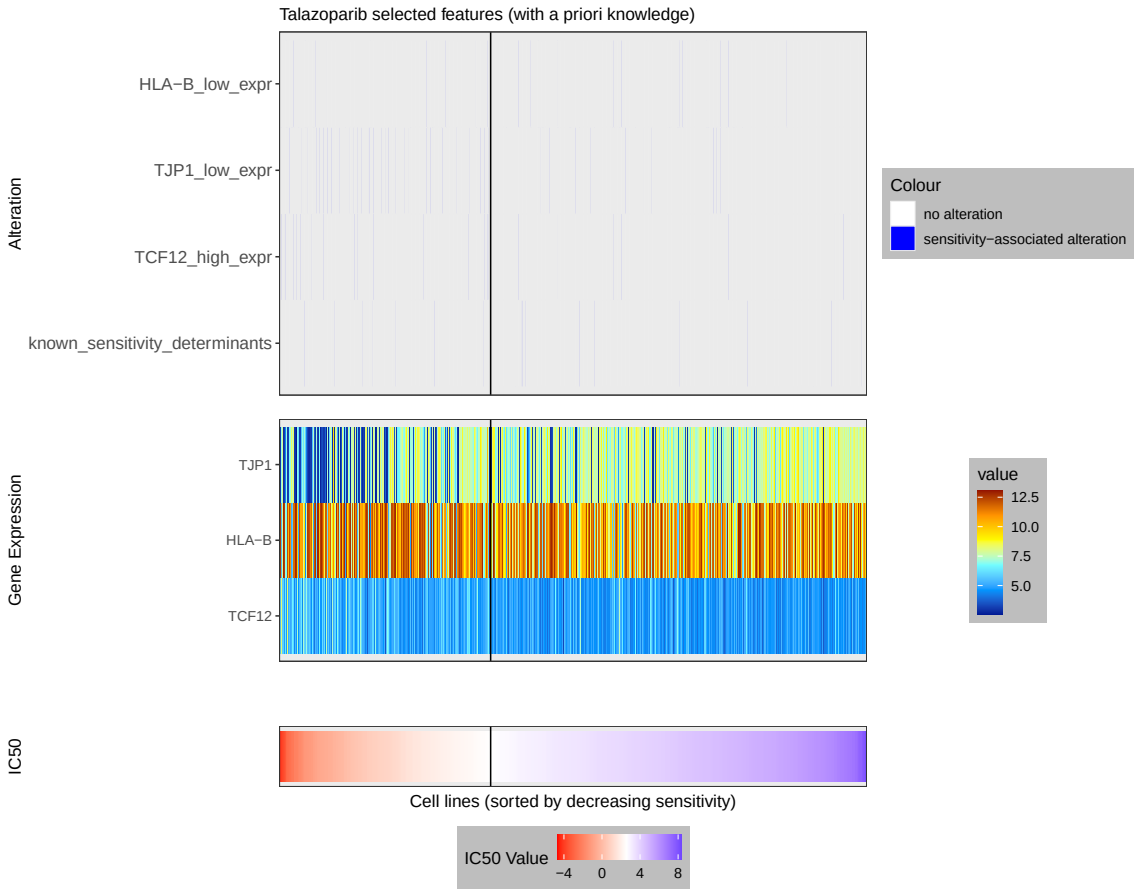
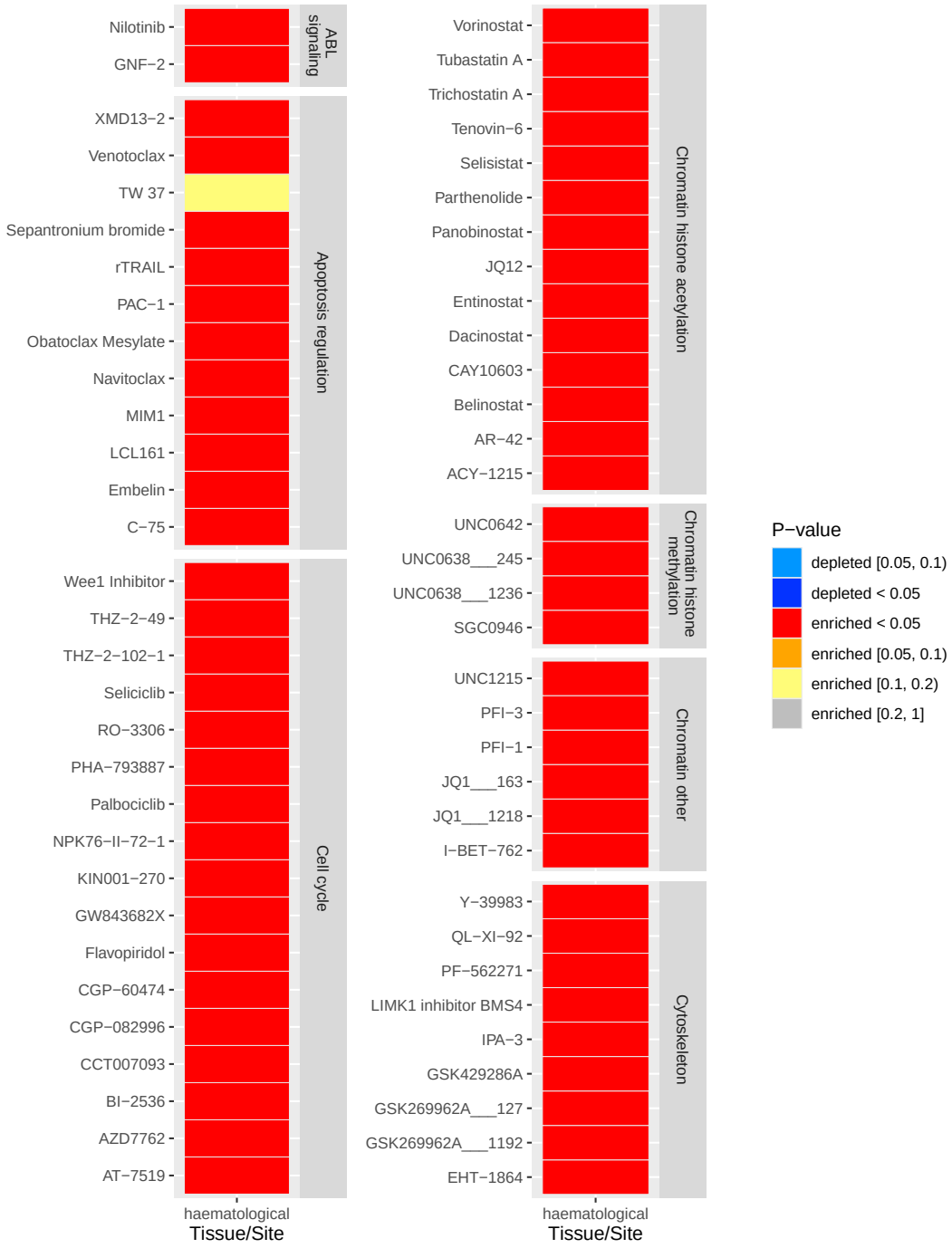
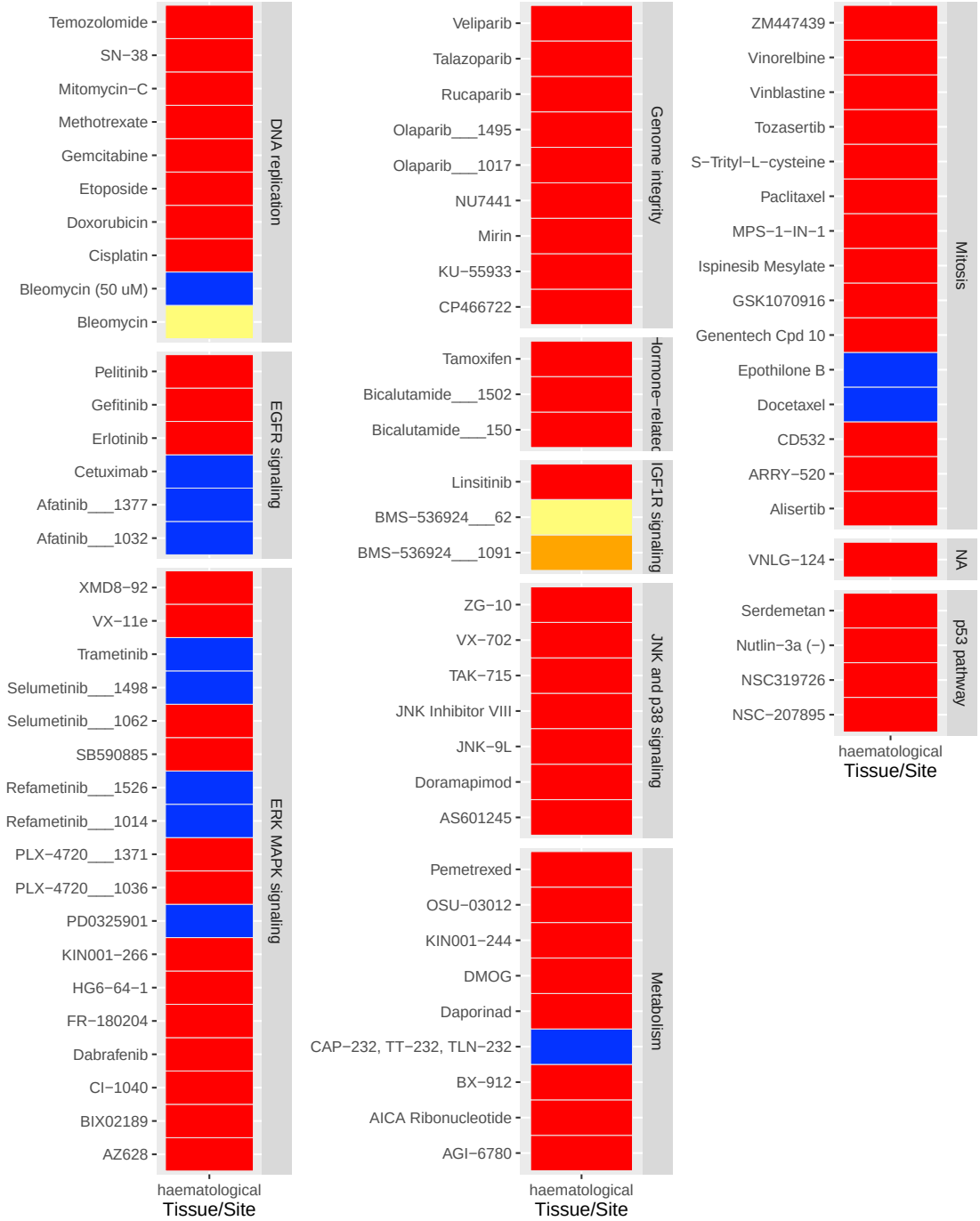
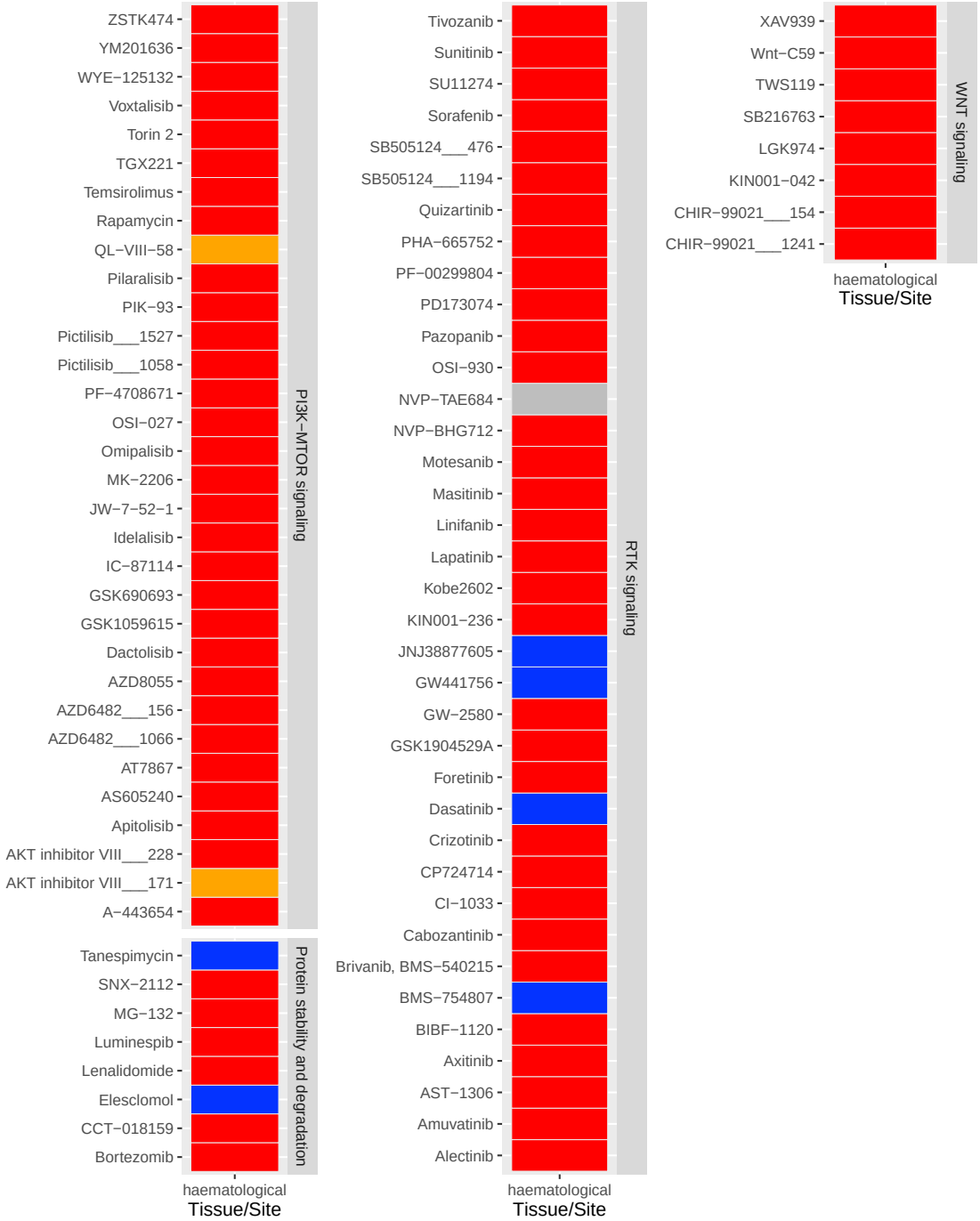


Figure 42: This figure depicts the selected features for the best model (selected with Youden's J) for Talazoparib using the quadratic weight function. The cell lines on the x-axis are sorted by decreasing sensitivity, i.e. increasing logarithmized IC50 value.


Nilotinib
GNF−2
ABL signaling
XMD13−2
Venotoclax
TW 37
Sepantronium bromide
rTRAIL
PAC−1
Obatoclax Mesylate
Navitoclax
MIM1
LCL161
Embelin
C−75
Apoptosis regulation
Wee1 Inhibitor
THZ−2−49
THZ−2−102−1
Seliciclib
RO−3306
PHA−793887
Palbociclib
NPK76−II−72−1
KIN001−270
GW843682X
Flavopiridol
CGP−60474
CGP−082996
CCT007093
BI−2536
AZD7762
AT−7519
Cell cycle
haematological
Tissue/Site
Vorinostat
Tubastatin A
Trichostatin A
Tenovin−6
Selisistat
Parthenolide
Panobinostat
JQ12
Entinostat
Dacinostat
CAY10603
Belinostat
AR−42
ACY−1215
Chromatin histone acetylation
UNC0642
UNC0638___245
UNC0638___1236
SGC0946
Chromatin histone methylation
UNC1215
PFI−3
PFI−1
JQ1___163
JQ1___1218
I−BET−762
Chromatin other
Y−39983
QL−XI−92
PF−562271
LIMK1 inhibitor BMS4
IPA−3
GSK429286A
GSK269962A___127
GSK269962A___1192
EHT−1864
Cytoskeleton
haematological
Tissue/Site
P−value
depleted [0.05, 0.1)
depleted < 0.05
enriched < 0.05
enriched [0.05, 0.1)
enriched [0.1, 0.2)
enriched [0.2, 1]


Temozolomide
SN−38
Mitomycin−C
Methotrexate
Gemcitabine
Etoposide
Doxorubicin
Cisplatin
Bleomycin (50 uM)
Bleomycin
DNA replication
Pelitinib
Gefitinib
Erlotinib
Cetuximab
Afatinib___1377
Afatinib___1032
EGFR signaling
XMD8−92
VX−11e
Trametinib
Selumetinib___1498
Selumetinib___1062
SB590885
Refametinib___1526
Refametinib___1014
PLX−4720___1371
PLX−4720___1036
PD0325901
KIN001−266
HG6−64−1
FR−180204
Dabrafenib
CI−1040
BIX02189
AZ628
ERK MAPK signaling
haematological
Tissue/Site
Veliparib
Talazoparib
Rucaparib
Olaparib___1495
Olaparib___1017
NU7441
Mirin
KU−55933
CP466722
Genome integrity
Tamoxifen
Bicalutamide___1502
Bicalutamide___150
Hormone−relatec
Linsitinib
BMS−536924___62
BMS−536924___1091
IGF1R signaling
ZG−10
VX−702
TAK−715
JNK Inhibitor VIII
JNK−9L
Doramapimod
AS601245
JNK and p38 signaling
Pemetrexed
OSU−03012
KIN001−244
DMOG
Daporinad
CAP−232, TT−232, TLN−232
BX−912
AICA Ribonucleotide
AGI−6780
Metabolism
haematological
Tissue/Site
ZM447439
Vinorelbine
Vinblastine
Tozasertib
S−Trityl−L−cysteine
Paclitaxel
MPS−1−IN−1
Ispinesib Mesylate
GSK1070916
Genentech Cpd 10
Epothilone B
Docetaxel
CD532
ARRY−520
Alisertib
Mitosis
VNLG−124
NA
Serdemetan
Nutlin−3a (−)
NSC319726
NSC−207895
p53 pathway
haematological
Tissue/Site


ZSTK474
YM201636
WYE−125132
Voxtalisib
Torin 2
TGX221
Temsirolimus
Rapamycin
QL−VIII−58
Pilaralisib
PIK−93
Pictilisib___1527
Pictilisib___1058
PF−4708671
OSI−027
Omipalisib
MK−2206
JW−7−52−1
Idelalisib
IC−87114
GSK690693
GSK1059615
Dactolisib
AZD8055
AZD6482___156
AZD6482___1066
AT7867
AS605240
Apitolisib
AKT inhibitor VIII___228
AKT inhibitor VIII___171
A−443654
PI3K−MTOR signaling
Tanespimycin
SNX−2112
MG−132
Luminespib
Lenalidomide
Elesclomol
CCT−018159
Bortezomib
Protein stability and degradation
haematological
Tissue/Site
Tivozanib
Sunitinib
SU11274
Sorafenib
SB505124___476
SB505124___1194
Quizartinib
PHA−665752
PF−00299804
PD173074
Pazopanib
OSI−930
NVP−TAE684
NVP−BHG712
Motesanib
Masitinib
Linifanib
Lapatinib
Kobe2602
KIN001−236
JNJ38877605
GW441756
GW−2580
GSK1904529A
Foretinib
Dasatinib
Crizotinib
CP724714
CI−1033
Cabozantinib
Brivanib, BMS−540215
BMS−754807
BIBF−1120
Axitinib
AST−1306
Amuvatinib
Alectinib
RTK signaling
haematological
Tissue/Site
XAV939
Wnt−C59
TWS119
SB216763
LGK974
KIN001−042
CHIR−99021___154
CHIR−99021___1241
WNT signaling
haematological
Tissue/Site

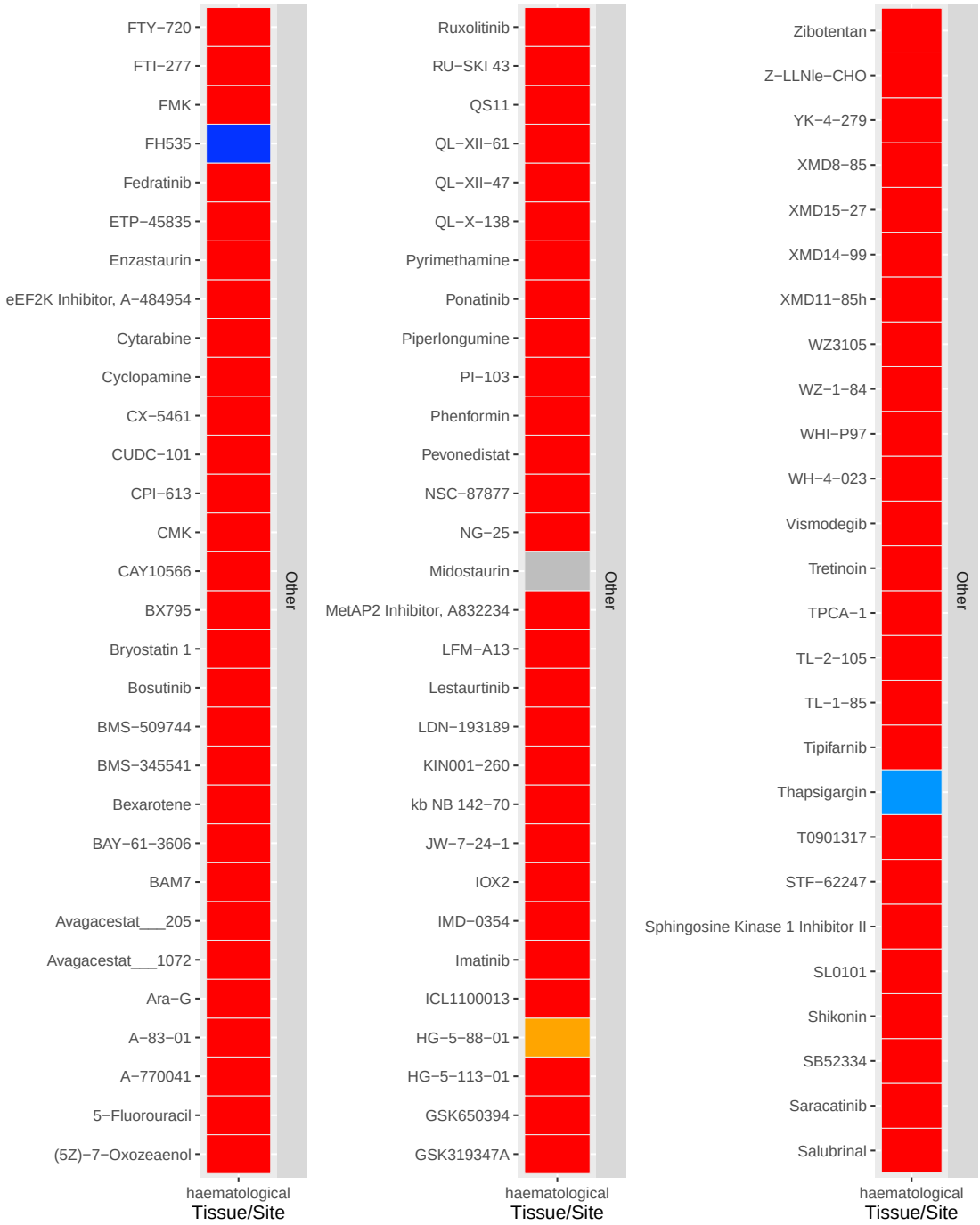
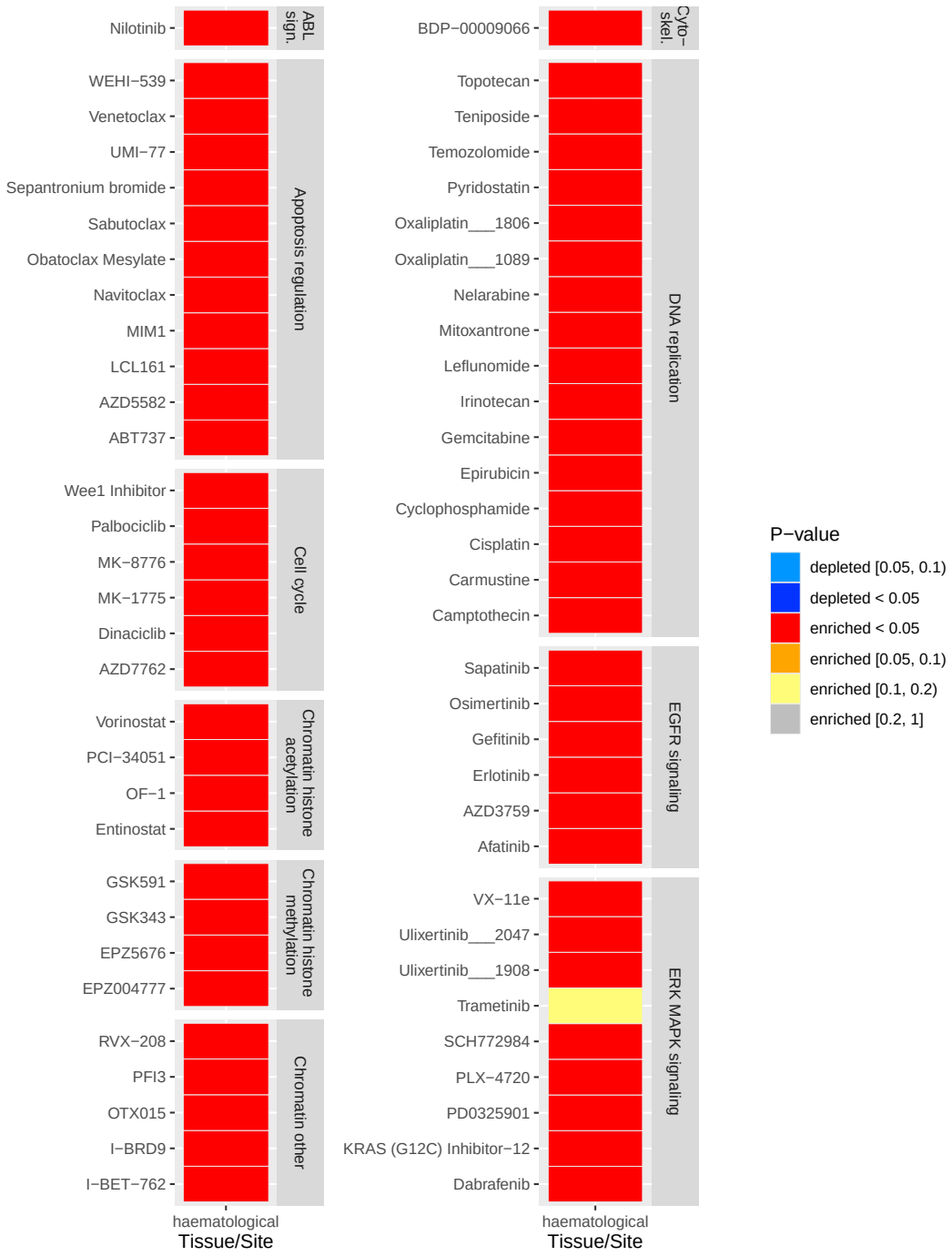
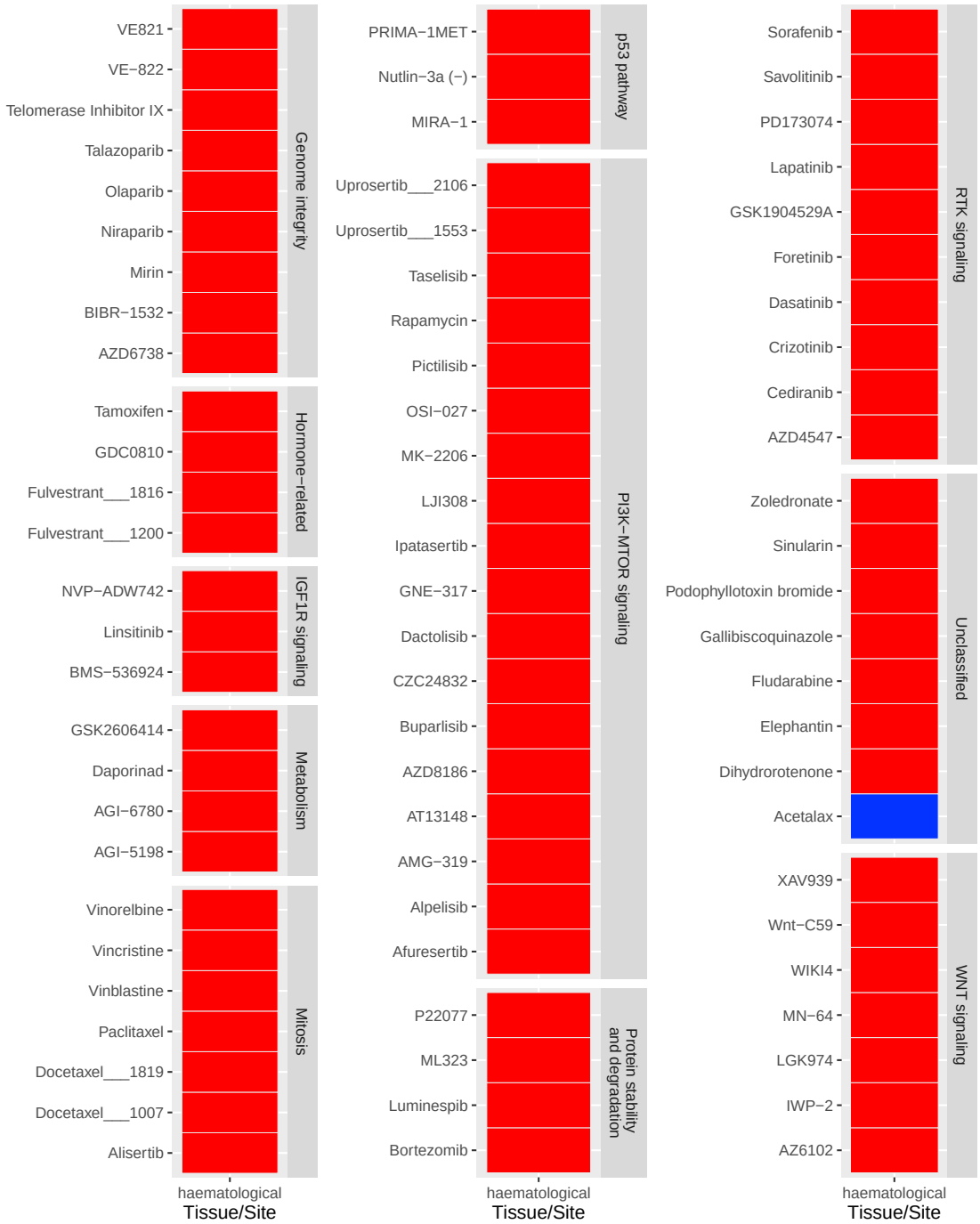


Figure 43: This figure depicts enrichment results for all drugs in the GDSC1 data set using the haematological cancer cell lines as category and the increasingly sorted logarithmized IC50 lists as ranked lists. It can be clearly seen that the haematological tissue is enriched for almost all drugs indicating that haematological cancer cell lines react more sensitive to drug treatment.


Nilotinib
ABL sign.
WEHI−539
Venetoclax
UMI−77
Sepantronium bromide
Sabutoclax
Obatoclax Mesylate
Navitoclax
MIM1
LCL161
AZD5582
ABT737
Apoptosis regulation
Wee1 Inhibitor
Palbociclib
MK−8776
MK−1775
Dinaciclib
AZD7762
Cell cycle
Vorinostat
PCI−34051
OF−1
Entinostat
Chromatin histone acetylation
GSK591
GSK343
EPZ5676
EPZ004777
Chromatin histone methylation
RVX−208
PFI3
OTX015
I−BRD9
I−BET−762
Chromatin other
haematological
Tissue/Site
BDP−00009066
Cyto−skel.
Topotecan
Teniposide
Temozolomide
Pyridostatin
Oxaliplatin___1806
Oxaliplatin___1089
Nelarabine
Mitoxantrone
Leflunomide
Irinotecan
Gemcitabine
Epirubicin
Cyclophosphamide
Cisplatin
Carmustine
Camptothecin
DNA replication
Sapatinib
Osimertinib
Gefitinib
Erlotinib
AZD3759
Afatinib
EGFR signaling
VX−11e
Ulixertinib___2047
Ulixertinib___1908
Trametinib
SCH772984
PLX−4720
PD0325901
KRAS (G12C) Inhibitor−12
Dabrafenib
ERK MAPK signaling
haematological
Tissue/Site
P−value
depleted [0.05, 0.1)
depleted < 0.05
enriched < 0.05
enriched [0.05, 0.1)
enriched [0.1, 0.2)
enriched [0.2, 1]


VE821
VE−822
Telomerase Inhibitor IX
Talazoparib
Olaparib
Niraparib
Mirin
BIBR−1532
AZD6738
Genome integrity
Tamoxifen
GDC0810
Fulvestrant___1816
Fulvestrant___1200
Hormone−related
NVP−ADW742
Linsitinib
BMS−536924
IGF1R signaling
GSK2606414
Daporinad
AGI−6780
AGI−5198
Metabolism
Vinorelbine
Vincristine
Vinblastine
Paclitaxel
Docetaxel___1819
Docetaxel___1007
Alisertib
Mitosis
haematological
Tissue/Site
PRIMA−1MET
Nutlin−3a (−)
MIRA−1
p53 pathway
Uprosertib___2106
Uprosertib___1553
Taselisib
Rapamycin
Pictilisib
OSI−027
MK−2206
LJI308
Ipatasertib
GNE−317
Dactolisib
CZC24832
Buparlisib
AZD8186
AT13148
AMG−319
Alpelisib
Afuresertib
PI3K−MTOR signaling
P22077
ML323
Luminespib
Bortezomib
Protein stability and degradation
haematological
Tissue/Site
Sorafenib
Savolitinib
PD173074
Lapatinib
GSK1904529A
Foretinib
Dasatinib
Crizotinib
Cediranib
AZD4547
RTK signaling
Zoledronate
Sinularin
Podophyllotoxin bromide
Gallibiscoquinazole
Fludarabine
Elephantin
Dihydrorotenone
Acetalax
Unclassified
XAV939
Wnt−C59
WIKI4
MN−64
LGK974
IWP−2
AZ6102
WNT signaling
haematological
Tissue/Site

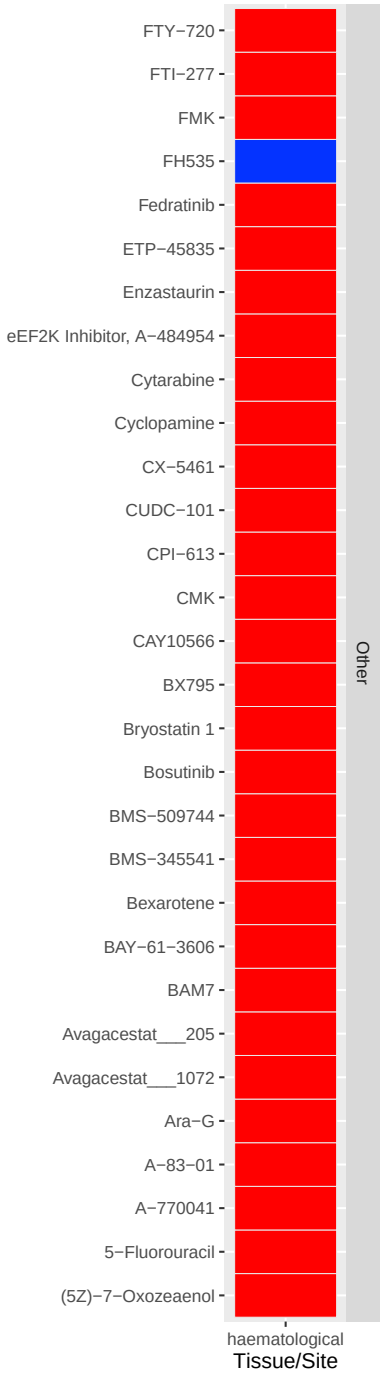


Figure 44: This figure depicts enrichment results for all drugs in the GDSC2 data set using the haematological cancer cell lines as category and the increasingly sorted logarithmized IC50 lists as ranked lists. It can be clearly seen that the haematological tissue is enriched for almost all drugs indicating that haematological cancer cell lines react more sensitive to drug treatment.

References

Iorio, F., Knijnenburg, T., Vis, D. et al. (2016) A landscape of pharmacogenomic interactions in cancer. *Cell*, **166**, 740-754. https://doi.org/10.1016/j.cell.2016.06.017

Knijnenburg, T., Klau, G., Iorio, F. et al. (2016) Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. *Sci Rep*, **6**, 1-14. https://doi.org/10.1038/srep36812

Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J. et al. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods*, **1**0, 1081-1082. https://doi.org/10.1038/nmeth.2642

Hollander, M., Wolfe, D. A., and Chicken, E. (2013) Nonparametric statistical methods. (Vol. 751). John Wiley & Sons.

Gerstner, N., Kehl, T., Lenhof, K. et al. (2020) GeneTrail 3: advanced high-throughput enrichment analysis. *Nucleic Acids Res*, **48 (W1)**, W515–W520. https://doi.org/10.1093/nar/gkaa306

Hochberg, Y., and Benjamini, Y. (1990) More powerful procedures for multiple significance testing. *Stat Med*, **9(7)**, 811-818.

Breiman, L. (2001) Random Forests. *Machine Learning 45*, 5–32. https://doi.org/10.1023/A:1010933404324

Fix, E., and Hodges, J.L. Jr. (1952) Discriminatory analysis-nonparametric discrimination: Small sample performance. *CALIFORNIA UNIV BERKELEY*

Kuhn M. (2020) caret: Classification and Regression Training. R package version 6.0-87. https://CRAN.R-project.org/package=caret

Liaw, A. and Wiener M.. (2002) Classification and Regression by randomForest *R News 2(3)*, 18-22.