

Supplementary Material

Clinical Drug Response Prediction by Using a Lq Penalized Network-Constrained Logistic Regression Method

Hai-Hui Huang^{a,b} Jing-Guo Dai^a Yong-Liang^b

^aSchool of Information Science and Engineering & Provincial Demonstration Software Institute, Shaoguan University, Shaoguan, ^bFaculty of Information Technology & State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macau, China

Supplementary Tables

Table S1 Biomarkers selected by our method for erlotinib prediction.

Gene symbol	Gene description
EFEMP1	EGF containing fibulin like extracellular matrix protein 1
OSMR	Oncostatin M receptor
EGFR	Epidermal growth factor receptor
S100A2	S100 calcium binding protein A2
SERPINB5	Serpin peptidase inhibitor 5
CLDN1	Claudin 1
OLFML2A	Olfactomedin like 2A
MET	MET proto-oncogene, receptor tyrosine kinase
SFN	Stratifin
PCSK5	Proprotein convertase subtilisin/kexin type 5
TPBG	Trophoblast glycoprotein
TBC1D2	TBC1 domain family member 2
RAF1	Raf-1 proto-oncogene, serine/threonine kinase
COL17A1	Collagen type XVII alpha 1 chain
KRT13	Keratin 13
FAT1	FAT atypical cadherin 1
FERMT1	Fermitin family member 1
MAP1S	Microtubule associated protein 1S
KRT6A	Keratin 6A
TRIB3	Tribbles pseudokinase 3
DGKA	Diacylglycerol kinase alpha
COPS7B	COP9 signalosome subunit 7B

Table S2 Biomarkers selected by our method for sorafenib prediction.

Gene symbol	Gene Description
LAMA5	laminin subunit alpha 5
ATP8B2	ATPase phospholipid transporting 8B2
GPD2	glycerol-3-phosphate dehydrogenase 2
ANP32C	acidic nuclear phosphoprotein 32 family member C
VEGFA	vascular endothelial growth factor A
ZNF222	zinc finger protein 222
SMIM2	small integral membrane protein 2
GABRP	gamma-aminobutyric acid type A receptor pi subunit
PARP6	poly(ADP-ribose) polymerase family member 6
KCNC3	potassium voltage-gated channel subfamily C member 3
GAGE1	G antigen 1
ANGPT1	angiopoietin 1
ERCC6	ERCC excision repair 6, chromatin remodeling factor
FGF3	fibroblast growth factor 3
ATP8B3	ATPase phospholipid transporting 8B3
MMRN1	multimerin 1
TF	transferrin
ZFY	zinc finger protein, Y-linked
MADCAM1	mucosal vascular addressin cell adhesion molecule 1
ZNF544	zinc finger protein 544
CPA3	carboxypeptidase A3
EXTL3	exostosin like glycosyltransferase 3
NUDT11	nudix hydrolase 11
FMO5	flavin containing monooxygenase 5
MCL1	MCL1, BCL2 family apoptosis regulator
HPGD	15-hydroxyprostaglandin dehydrogenase

Table S3 Comparison with some state-of-the-art methods

	Lq-NLR-E	Lq-NLR-S	Lasso	Ridge	ElasticNet	InterNet
Erlotinib	0.841	-	0.492	0.638	0.530	0.689
Sorafenib	-	0.788	-0.045	-0.115	0.372	0.516

This table shows the rho (from a Spearman's correlation test) achieved for each dataset using Lq-NLR-E, Lq-NLR-S and logistic regression with Lasso, Ridge, Elastic net and InterNet. The way of the selection of the regularization parameters was described in Methods. There are two models for Lasso, Ridge, Elastic net and InterNet methods, respectively. One for the prediction of erlotinib, the other is for sorafenib.

Table S4 Prediction results 15 patients by using the Lq-NLR-erlotinib model.

Patient ID	Predicted index	Class	0-2 months	2-4 months	4-6 months
GSM677324	478.9676735	2	36.4%	54.5%	9.1%
GSM677325	455.992511	2	36.4%	54.5%	9.1%
GSM677328	491.9444822	2	36.4%	54.5%	9.1%
GSM677329	416.2344236	1	100.0%	-	-
GSM677330	460.6683775	2	36.4%	54.5%	9.1%
GSM677331	428.4616235	1	100.0%	-	-
GSM677332	443.1020984	1	100.0%	-	-
GSM677334	405.5066253	1	100.0%	-	-
GSM789978	409.5568935	1	100.0%	-	-
GSM789979	512.2505912	3	-	25.0%	75.0%
GSM789981	503.4250838	3	-	25.0%	75.0%
GSM789983	476.8774282	2	36.4%	54.5%	9.1%
GSM789986	438.2009536	1	100.0%	-	-
GSM789987	429.0898433	1	100.0%	-	-
GSM790019	526.7217054	3	-	25.0%	75.0%

These 15 patients are who neither treatment with erlotinib nor sorafenib from GSE33072. We then applied our erlotinib response prediction model to the patient's gene expression data to yield an in vivo drug response to erlotinib. Value of class can be referred to the Table 1.

Table S5 Prediction results 15 patients by using the Lq-NLR-sorafenib model.

Patient ID	Predicted index	Class	0-2 months	2-4 months	4-6 months	6-8 months	8-12 months
GSM677324	-5.08174	1	81.3%	18.8%	-	-	-
GSM677325	30.52874	2	25.0%	50.0%	16.7%	8.3%	-
GSM677328	21.11218	1	81.3%	18.8%	-	-	-
GSM677329	31.19392	2	25.0%	50.0%	16.7%	8.3%	-
GSM677330	36.31196	2	25.0%	50.0%	16.7%	8.3%	-
GSM677331	30.57758	2	25.0%	50.0%	16.7%	8.3%	-
GSM677332	32.93836	2	25.0%	50.0%	16.7%	8.3%	-
GSM677334	50.5941	3	-	33.3%	11.1%	11.1%	44.4%
GSM789978	-0.37824	1	81.3%	18.8%	-	-	-
GSM789979	21.31755	1	81.3%	18.8%	-	-	-
GSM789981	42.84765	3	-	33.3%	11.1%	11.1%	44.4%
GSM789983	39.3777	2	25.0%	50.0%	16.7%	8.3%	-
GSM789986	32.03477	2	25.0%	50.0%	16.7%	8.3%	-
GSM789987	15.26373	1	81.3%	18.8%	-	-	-
GSM790019	24.56605	1	81.3%	18.8%	-	-	-

These 15 patients are who neither treatment with erlotinib nor sorafenib from GSE33072. We then applied our sorafenib response prediction model to the patient's gene expression data to yield an in vivo drug response to sorafenib. Value of class can be referred to the Table 2.