Bioinformatics
doi.10.1093/bioinformatics/xxxxxx
Advance Access Publication Date: Day Month Year
Manuscript Category

Supplementary Material

Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression

Muhammad Ammad-ud-din [1, 2*], Suleiman A. Khan [1, 2*,], Krister Wennerberg [1] and Tero Aittokallio [1, 2, 3]

[1] Institute for Molecular Medicine Finland FIMM, University of Helsinki, 00014 Helsinki, Finland
[2] Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, 02150 Espoo, Finland
[3] Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland

* Contributed Equally.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Cancer Data Sets

GDSC

Selection of the drug groups primarily on the basis of primary therapeutic targets: We selected the drugs based on the information of primary therapeutic targets available from the GDSC project (Yang *et al.*, 2013). As we are using prior knowledge from the human cancer kinome, we focus our analysis to kinase inhibitors. As a first step, we computed the pairwise correlation of drugs belonging to the same target group and found significant differences in their values, Figure S2. Ideally, the drugs inhibiting same targets have similar responses, but this was not found. Upon closer inspection, we noticed that the number of cell line samples and screening date may have an noisy effect on the drug responses. Therefore, we re-organized the groups of drugs based on three criteria (1) similar target, (2) similar sample size and/or screening date, (3) negative correlation. In this way, we obtained 16 different drug groups as listed in Table S1. We observed high correlated drug responses in each of these groups, Figure S3.

FIMM

Selection of the drug groups on the basis of primary therapeutic targets and responsiveness: Gautam *et al.* (2016) published response measurements of 301 drugs on 19 cell lines. Among these 109 drugs belong to the broader class of kinase inhibitors. As a first step, we choose 109 kinase inhibitors belonging to 12 groups. We subsequently filter 6 drug groups that are responding in more than 80% of the cell lines, as shown in Figure S5 and are highly correlated (shown in Figure S6) as well. The names of FLNs and the number of genes present in each of the FLNs, used for the case study with FIMM data set are illustrated in Figure S7.

Empirical evidence of mean prediction correlation

Here we discuss the use of mean as a baseline prediction metric in drug response analysis, when using the generally used correlation and leave one out cross validation (LOOCV) settings (Azuaje, 2016). We observe that mean prediction of uncentered data, under LOOCV yields a correlation of -1, and should therefore be interpreted accordingly. In order to validate our observation, we generate 10000 random data sets with 100 samples each, and compute the mean prediction using LOOCV. We then compute the prediction performance using Spearman and Pearson correlations. The results confirm that the mean prediction correlation (pearson and spearman) is exactly -1 for all the random data sets. As these settings are used often in drug response prediction analysis, we therefore recommend that highly negative correlations be interpreted accordingly.

References

Azuaje,F. (2016) Computational models for predicting drug responses in cancer research. *Briefings in Bioinformatics,* **22** (8), bbw065.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological),* **57** (1), 289–300.

Benjamini,Y. and Yekutieli,D. (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics,* **29** (4), 1165–1188.

Gautam,P., Karhinen,L., Szwajda,A., Jha,S.K., Yadav,B., Aittokallio,T. and Wennerberg,K. (2016) Identification of selective cytotoxic and synthetic lethal drug responses in triple negative breast cancer cells. *Molecular cancer,* **15** (1), 1.

Yang,W., Soares,J., Greninger,P., Edelman,E.J., Lightfoot,H., Forbes,S., Bindal,N., Beare,D., Smith,J.A., Thompson,I.R. *et al.* (2013) Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.,* **41** (D1), D955–D961.

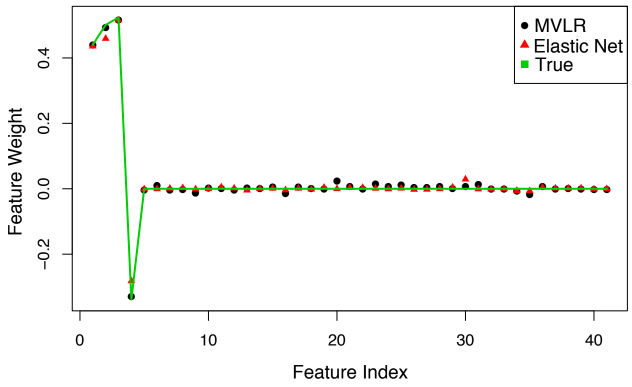

Fig. S1. We also validate our model on single-view data sets, confirming that it performs comparably to the existing methods in identifying analogous and correct set of features in the synthetic data.

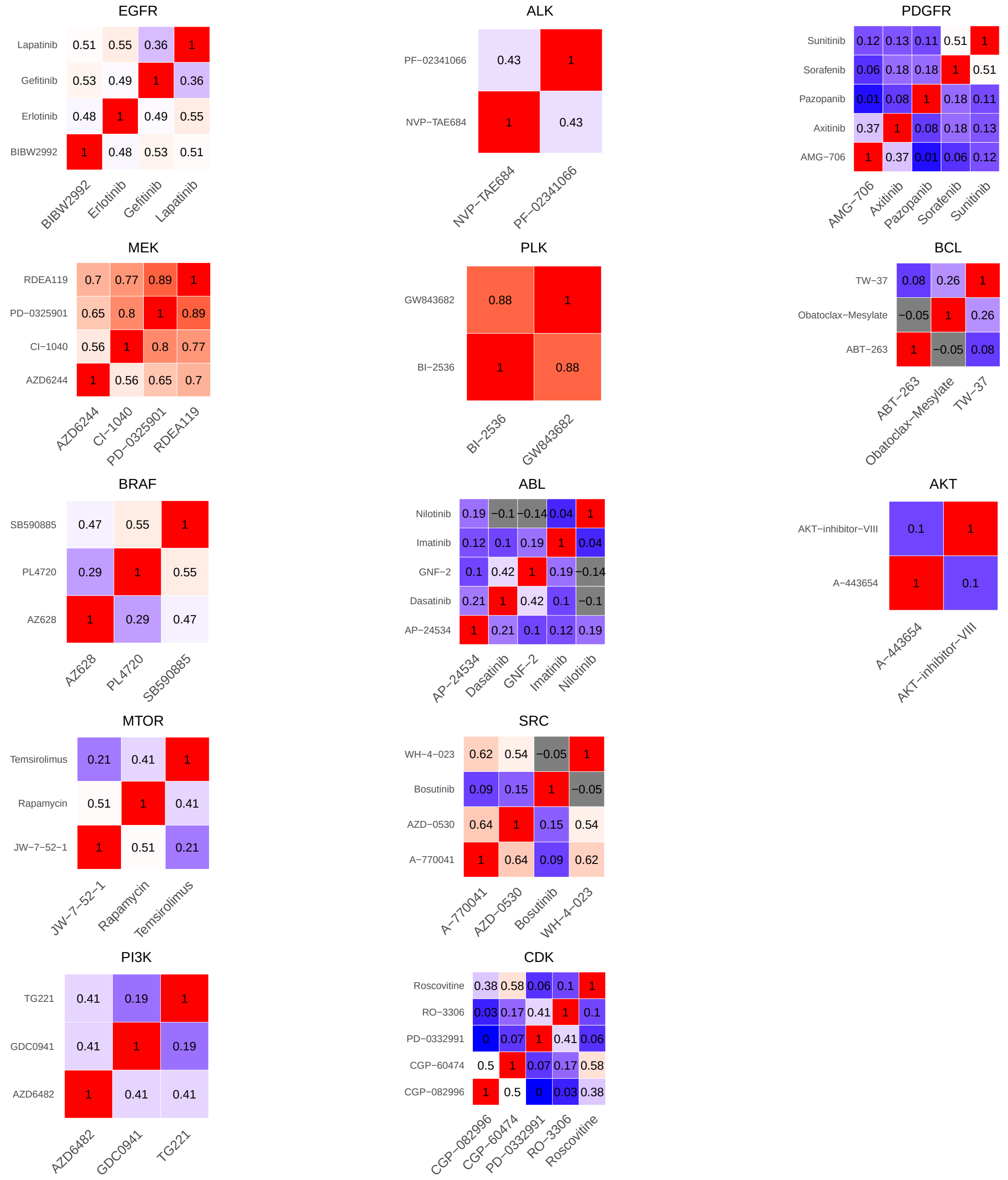

Fig. S2. Pairwise correlation of drug responses grouped by similar primary targets. Among EGFR inhibitors, lapatinib and gefitinib are poorly correlated, compared to lapatinib and erlotinib. Likewise, erlotinib and BIBW2992 are less correlated compared to BIBW2992 and gefitinib. Similar, patterns are seen in CDK, PDGFR, SRC and ABL inhbitors.

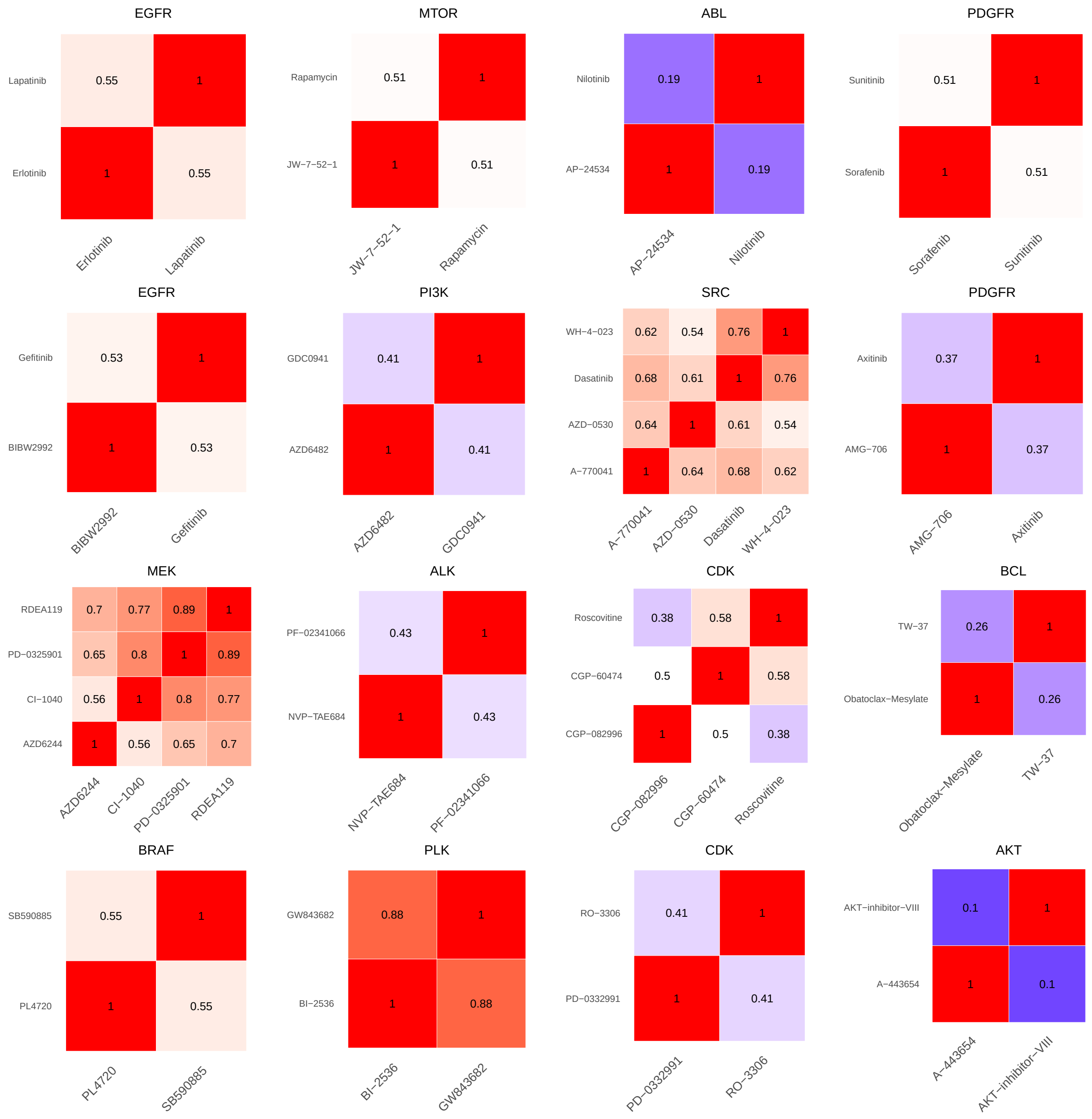

Fig. S3. Pairwise correlation of drug responses grouped by similar primary targets, sample size and screening date. Further details are given in Table S1 and in the text

.

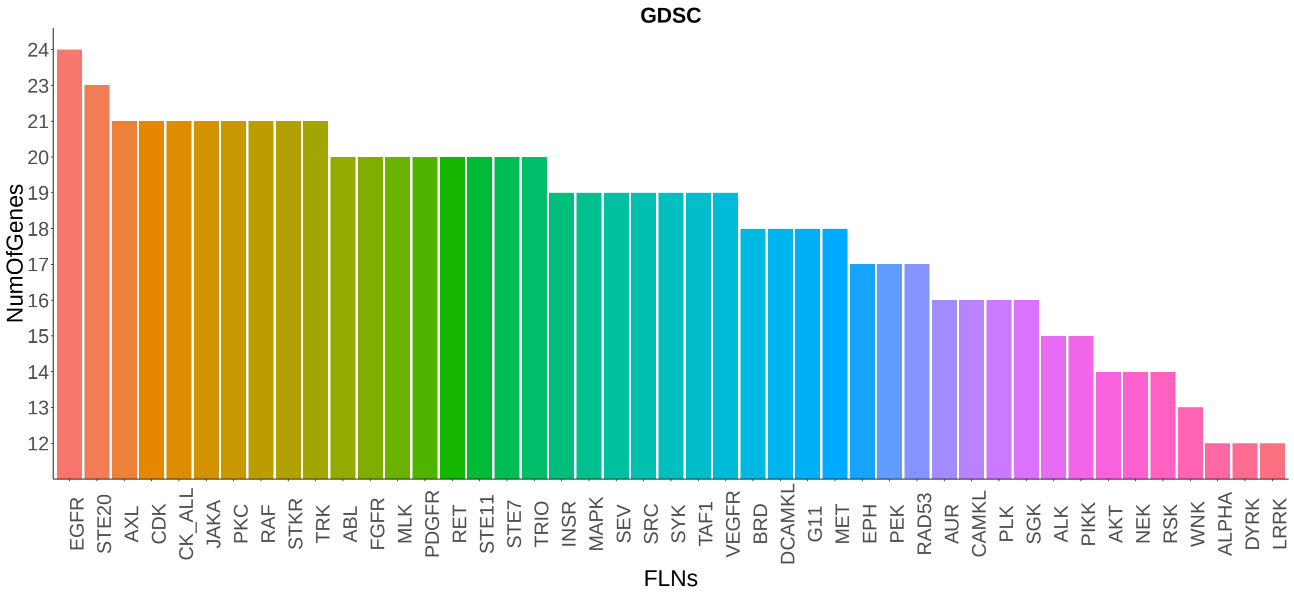

Fig. S4. Number of genes (y-axis) present in each of the FLNs (x-axis) used for the case study with GDSC data set.

.

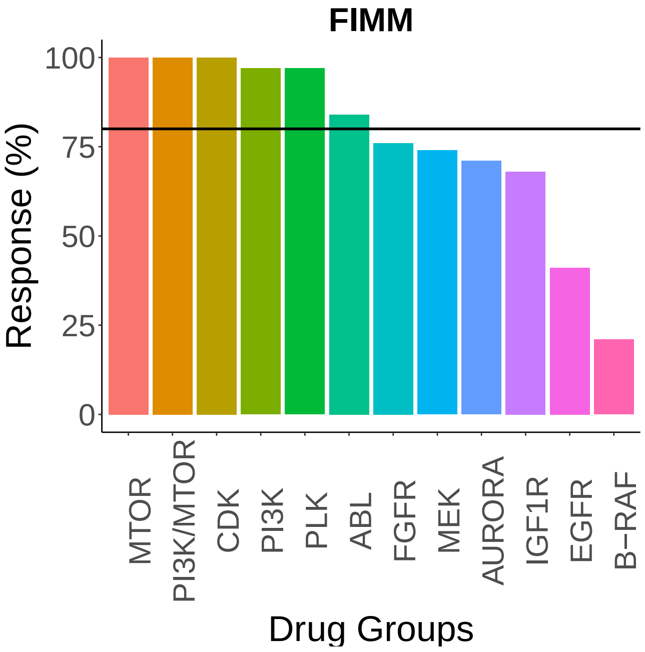

Fig. S5. Percentage of response (y-axis) observed in each of the drug groups (x-axis) for the case of the FIMM data set. The drug group showing response in more than 80% of the cell lines are selected for the analysis.

.

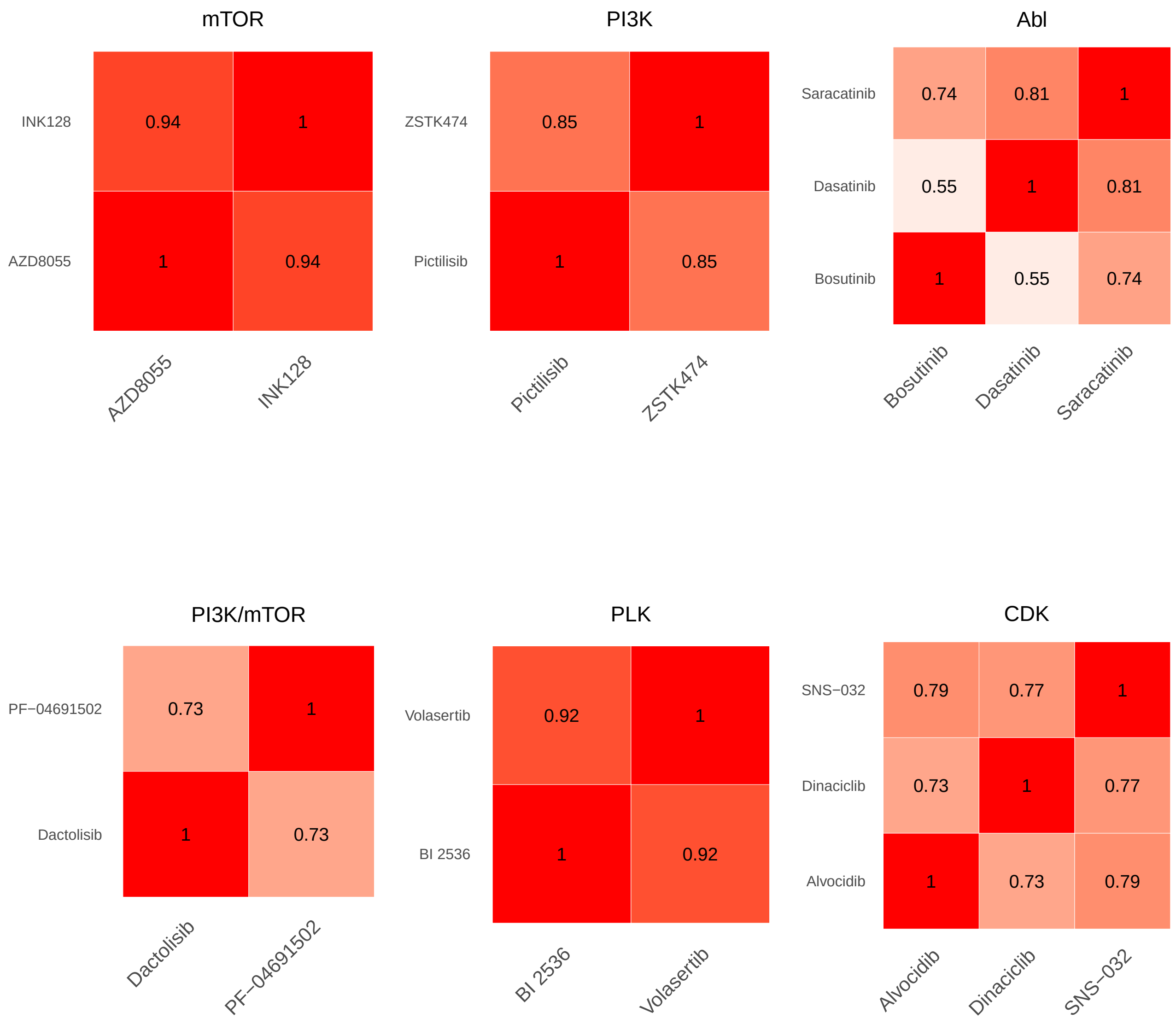

Fig. S6. Pairwise correlation of drug responses grouped by primary targets. The drugs show high correlated response pattern in each of the groups.

.

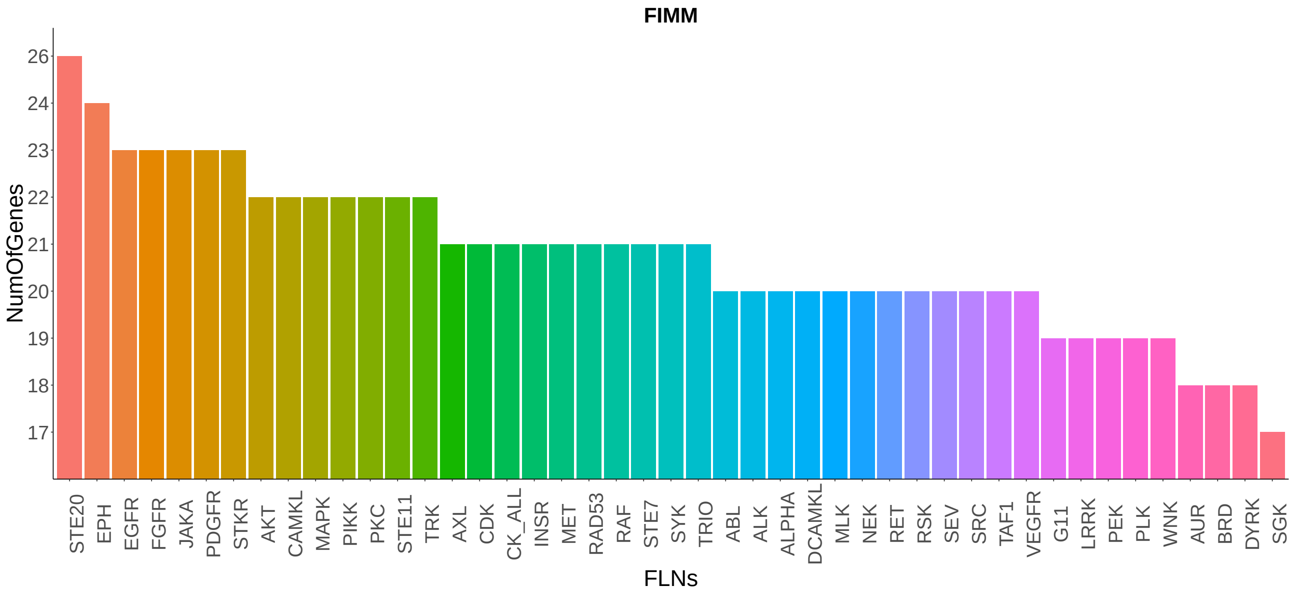

Fig. S7. Number of genes (y-axis) present in each of the FLNs (x-axis) used for the case study with FIMM data set..

.

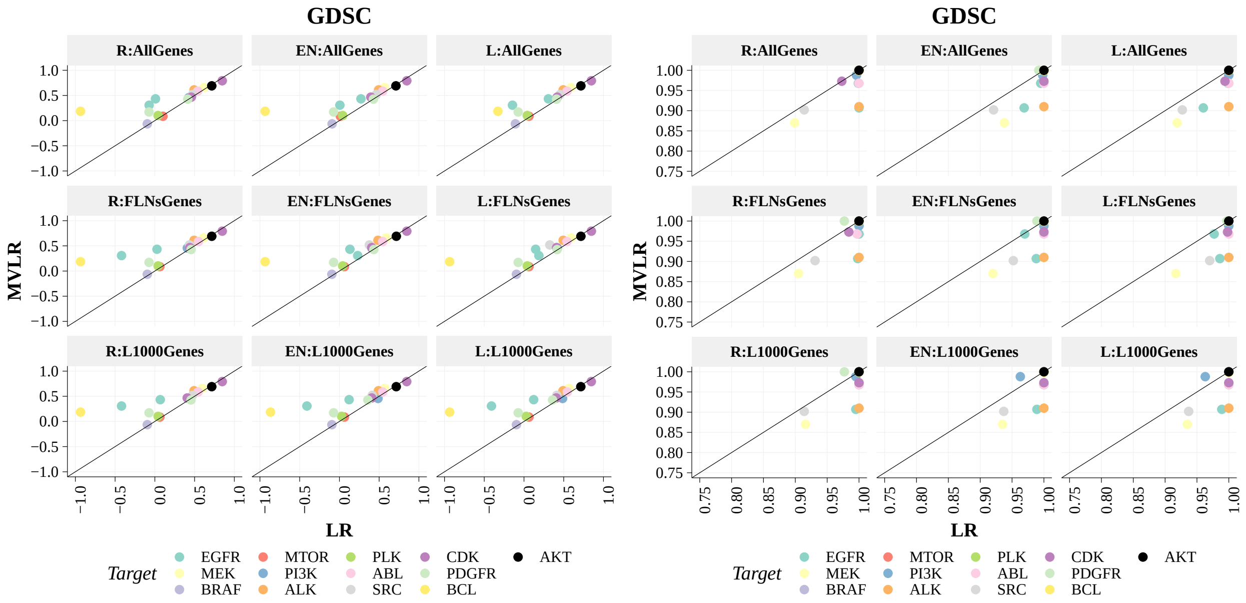

Fig. S8. Prediction Performance of individual drug groups colored according to their primary target, computed across cell lines in GDSC data set. Left: Pearson Correlation, Right: RMSE. Method abbreviations are explained in Table 2 (in the manuscript). The performance obtained by MVLR (shown on y-axis) is found to be significantly higher than the others shown on x-axis ($p<0.01$; one-sided paired Wilcoxon Sign-Rank test corrected for multiple testing). Here -1 denotes the baseline performance, computed using the mean of the training drug response data as predictions.

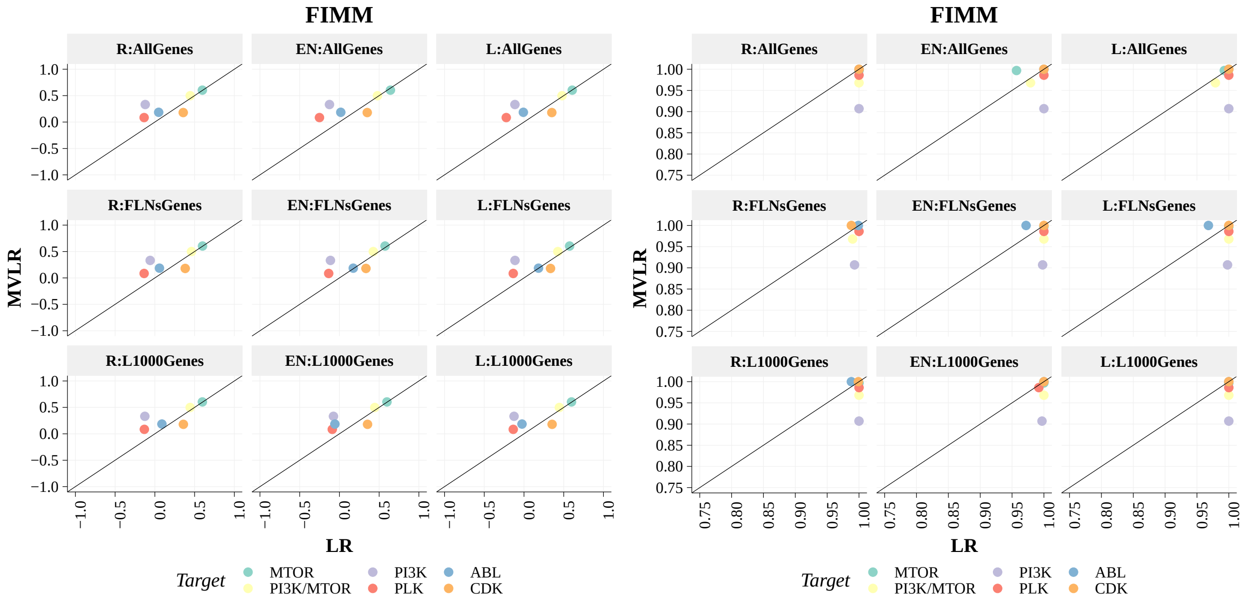

Fig. S9. Prediction Performance of individual drug groups colored according to their primary target, computed across cell lines in FIMM data set. Left: Pearson Correlation, Right: RMSE. Method abbreviations are explained in Table 2 (in the manuscript). Here 1 denotes the baseline performance, computed using the mean of the training drug response data as predictions.

.

Table S1. GDSC data set: drugs and their 16 target-based groups organized by the sample size and screening date. This information is available at the GDSC server.

Name	Targets	Sample.size	Last.screening.date
Erlotinib	EGFR	323	16-MAR-12
Lapatinib	EGFR, ERBB2	348	16-MAR-12
Gefitinib	EGFR	663	05-FEB-13
BIBW2992	EGFR, ERBB2	663	05-FEB-13
RDEA119	MEK1/2	654	05-FEB-13
CI-1040	MEK1/2	659	05-FEB-13
PD-0325901	MEK1/2	654	05-FEB-13
AZD6244	MEK1/2	633	05-FEB-13
PLX4720	BRAF	661	05-FEB-13
SB590885	BRAF	641	05-FEB-13
Rapamycin	MTOR	357	16-MAR-12
JW-7-52-1	MTOR	356	16-MAR-12
GDC0941	PI3K (class 1)	652	05-FEB-13
AZD6482	PI3Kb (P3C2B)	672	14-MAY-12
NVP-TAE684	ALK	357	16-MAR-12
PF-02341066	MET, ALK	357	16-MAR-12
GW843682X	PLK1	356	16-MAR-12
BI-2536	PLK1/2/3	356	16-MAR-12
AP-24534	ABL	672	14-MAY-12
Nilotinib	ABL	645	05-FEB-13
Dasatinib	ABL, SRC, KIT, PDGFR	355	16-MAR-12
A-770041	SRC family	356	16-MAR-12
WH-4-023	SRC family, ABL	355	16-MAR-12
AZD-0530	SRC, ABL1	359	16-MAR-12
CGP-60474	CDK1/2/5/7/9	355	16-MAR-12
CGP-082996	CDK4	355	16-MAR-12
Roscovitine	CDKs	348	16-MAR-12
RO-3306	CDK1	661	05-FEB-13
PD-0332991	CDK4/6	633	05-FEB-13
Sunitinib	PDGFRA, PDGFRB, KDR, KIT, FLT3	355	16-MAR-12
Sorafenib	PDGFRA, PDGFRB, KDR, KIT, FLT3	354	16-MAR-12
Axitinib	PDGFR, KIT, VEGFR	663	05-FEB-13
AMG-706	VEGFR, RET, c-KIT, PDGFR	661	05-FEB-13
TW 37	BCL-2, BCL-XL	653	05-FEB-13
Obatoclax Mesylate	BCL-2, BCL-XL, MCL-1	665	14-MAY-12
AKT inhibitor VIII	AKT1/2	672	14-MAY-12
A-443654	AKT1/2/3	355	16-MAR-12

Table S2. Spearman correlations of predictions for individual drug groups in GDSC data set.

	MVLR:FLNs	R:AllGenes	EN:AllGenes	L:AllGenes	R:FLNsGenes	EN:FLNsGenes	L:FLNsGenes	R:L1000Genes	EN:L1000Genes	L:L1000Genes
EGFR	0.408	-0.289	0.302	0.366	-0.085	0.201	0.202	0.074	0.167	0.167
EGFR	0.218	-0.189	-0.002	-0.161	-0.376	0.272	0.233	-0.376	-0.362	-0.362
MEK	0.662	0.631	0.574	0.606	0.627	0.596	0.597	0.613	0.57	0.57
BRAF	-0.09	-0.53	-0.375	-0.531	-0.53	-0.53	-0.53	-0.53	-0.53	-0.53
MTOR	0.079	0.052	-0.233	-0.334	-0.334	-0.334	-0.334	-0.334	-0.334	-0.334
PI3K	0.443	0.385	0.317	0.186	0.041	-0.082	-0.083	0.373	0.451	0.451
ALK	0.562	0.01	0.021	0.01	0.01	0.12	0.01	0.01	0.011	0.011
PLK	0.151	-0.367	-0.367	-0.367	-0.367	-0.367	-0.367	-0.367	-0.367	-0.367
ABL	0.598	0.117	0.116	0.12	0.51	0.123	0.181	0.123	0.116	0.116
SRC	0.444	0.353	0.339	0.334	0.335	0.313	0.286	0.357	0.332	0.332
CDK	0.657	0.523	0.523	0.523	0.523	0.523	0.523	0.523	0.524	0.524
CDK	0.443	0.306	0.088	0.319	0.245	-0.118	-0.077	-0.118	-0.118	-0.118
PDGFR	0.214	-0.425	-0.425	-0.425	-0.425	-0.425	-0.425	-0.425	-0.425	-0.425
PDGFR	0.435	0.321	0.384	-0.003	0.463	0.389	0.31	0.466	0.316	0.31
BCL	0.173	-0.881	-0.881	-0.881	-0.881	-0.881	-0.881	-0.881	-0.874	-0.881
AKT	0.599	0.274	0.287	0.274	0.274	0.274	0.274	0.274	0.289	0.289

Table S3. Pearson correlations of predictions for individual drug groups in GDSC data set.

	MVLR:FLNs	R:AllGenes	EN:AllGenes	L:AllGenes	R:FLNsGenes	EN:FLNsGenes	L:FLNsGenes	R:L1000Genes	EN:L1000Genes	L:L1000Genes
EGFR	0.433	0.008	0.271	0.307	0.027	0.132	0.147	0.068	0.122	0.122
EGFR	0.307	-0.072	0.006	-0.145	-0.419	0.23	0.187	-0.419	-0.409	-0.409
MEK	0.658	0.617	0.568	0.591	0.611	0.592	0.596	0.597	0.569	0.569
BRAF	-0.064	-0.095	-0.092	-0.108	-0.095	-0.095	-0.095	-0.095	-0.095	-0.095
MTOR	0.087	0.102	0.012	0.063	0.063	0.063	0.063	0.063	0.063	0.063
PI3K	0.457	0.428	0.423	0.415	0.409	0.408	0.42	0.429	0.486	0.486
ALK	0.611	0.493	0.493	0.493	0.493	0.485	0.493	0.493	0.491	0.491
PLK	0.102	0.041	0.041	0.041	0.041	0.041	0.041	0.041	0.028	0.028
ABL	0.586	0.539	0.545	0.542	0.549	0.545	0.541	0.544	0.545	0.545
SRC	0.519	0.474	0.464	0.453	0.436	0.377	0.325	0.474	0.413	0.413
CDK	0.792	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.847	0.847
CDK	0.467	0.459	0.395	0.419	0.44	0.407	0.411	0.407	0.407	0.407
PDGFR	0.172	-0.073	-0.073	-0.073	-0.073	-0.073	-0.073	-0.073	-0.073	-0.073
PDGFR	0.427	0.414	0.428	0.41	0.456	0.433	0.417	0.455	0.358	0.36
BCL	0.186	-0.935	-0.935	-0.328	-0.935	-0.935	-0.935	-0.935	-0.867	-0.935
AKT	0.691	0.716	0.711	0.716	0.716	0.716	0.716	0.716	0.715	0.715

Table S4. RMSE of predictions for individual drug groups in GDSC data set. For simplicity, the value greater than 1 have been set to 1 for all methods.

	MVLR:FLNs	R:AllGenes	EN:AllGenes	L:AllGenes	R:FLNsGenes	EN:FLNsGenes	L:FLNsGenes	R:L1000Genes	EN:L1000Genes	L:L1000Genes
EGFR	0.907	1	0.969	0.96	0.998	0.988	0.986	0.995	0.989	0.989
EGFR	0.968	0.999	0.995	0.999	1	0.97	0.977	1	1	1
MEK	0.87	0.899	0.938	0.919	0.905	0.92	0.917	0.916	0.935	0.935
BRAF	1	1	1	1	1	1	1	1	1	1
MTOR	1	1	1	1	1	1	1	1	1	1
PI3K	0.988	0.996	0.998	1	1	1	1	0.995	0.963	0.963
ALK	0.91	1	1	1	1	1	1	1	1	1
PLK	1	1	1	1	1	1	1	1	1	1
ABL	0.968	1	1	1	0.997	1	1	1	1	1
SRC	0.902	0.914	0.921	0.927	0.931	0.952	0.97	0.914	0.937	0.937
CDK	1	1	1	1	1	1	1	1	1	1
CDK	0.973	0.973	1	0.994	0.984	1	0.998	1	1	1
PDGFR	1	1	1	1	1	1	1	1	1	1
PDGFR	1	0.999	0.992	1	0.977	0.989	0.997	0.977	1	1
BCL	0.999	1	1	1	1	1	1	1	1	1
AKT	1	1	1	1	1	1	1	1	1	1

Table S5. Statistical significance of the predictive performances on the GDSC data set. P-values from one-sided paired Wilcoxon Sign-Rank test corresponding to the values shown in Figures 3 (in manuscript), Figure S8 (in sup mat), corrected for multiple testing using Benjamini, Hochberg, and Yekutieli's method (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

	R.AllGenes	EN.AllGenes	L.AllGenes	R.FLNsGenes	EN.FLNsGenes	L.FLNsGenes	R.L1000Genes	EN.L1000Genes	L.L1000Genes
Spearman Correlation	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}	3.29×10^{-4}
Pearson Correlation	4.51×10^{-3}	2.61×10^{-3}	2.61×10^{-3}	3.73×10^{-3}	2.61×10^{-3}	2.61×10^{-3}	3.36×10^{-3}	2.61×10^{-3}	2.61×10^{-3}
RMSE	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}	9.49×10^{-3}

Table S6. Spearman correlations of predictions for individual drug groups in FIMM data set.

	MVLR:FLNs	R:AllGenes	EN:AllGenes	L:AllGenes	R:FLNsGenes	EN:FLNsGenes	L:FLNsGenes	R:L1000Genes	EN:L1000Genes	L:L1000Genes
MTOR	0.566	0.262	0.553	0.35	0.289	0.276	0.265	0.262	0.262	0.262
PI3K/MTOR	0.533	0.088	0.235	0.235	0.214	0.121	0.098	0.096	0.089	0.089
PI3K	0.362	-0.274	-0.169	-0.15	-0.051	-0.273	-0.273	-0.266	-0.129	-0.19
PLK	0.132	-0.291	-0.354	-0.369	-0.291	-0.289	-0.289	-0.291	-0.26	-0.291
ABL	0.219	-0.127	-0.072	-0.123	-0.034	0.038	-0.009	-0.018	-0.304	-0.308
CDK	0.202	0.158	0.206	0.175	0.246	0.227	0.227	0.213	0.201	0.201

Table S7. Pearson correlations of predictions for individual drug groups in FIMM data set.

	MVLR:FLNs	R:AllGenes	EN:AllGenes	L:AllGenes	R:FLNsGenes	EN:FLNsGenes	L:FLNsGenes	R:L1000Genes	EN:L1000Genes	L:L1000Genes
MTOR	0.604	0.598	0.642	0.607	0.6	0.574	0.574	0.598	0.598	0.598
PI3K/MTOR	0.499	0.445	0.479	0.479	0.461	0.426	0.426	0.444	0.445	0.445
PI3K	0.332	-0.121	-0.125	-0.114	-0.059	-0.113	-0.113	-0.126	-0.075	-0.125
PLK	0.085	-0.135	-0.251	-0.223	-0.135	-0.135	-0.135	-0.132	-0.09	-0.135
ABL	0.185	0.048	0.017	-0.008	0.058	0.174	0.183	0.089	-0.057	-0.025
CDK	0.18	0.357	0.35	0.35	0.382	0.334	0.334	0.359	0.354	0.354

Table S8. RMSE of predictions for individual drug groups in FIMM data set. For simplicity, the value greater than 1 have been set to 1 for all methods.

r	MVLR:FLNs	R:AllGenes	EN:AllGenes	L:AllGenes	R:FLNsGenes	EN:FLNsGenes	L:FLNsGenes	R:L1000Genes	EN:L1000Genes	L:L1000Genes
MTOR	0.997	1	0.957	0.993	0.998	1	1	1	1	1
PI3K/MTOR	0.968	1	0.979	0.979	0.99	1	1	1	1	1
PI3K	0.907	1	1	1	0.993	0.998	0.998	1	0.997	1
PLK	0.986	1	1	1	1	1	1	1	0.992	1
ABL	1	1	1	1	0.999	0.972	0.968	0.988	1	1
CDK	1	1	1	1	0.988	1	1	0.999	1	1

Table S9. Statistical significance of the predictive performances on the FIMM data set. P-values from one-sided paired Wilcoxon Sign-Rank test corresponding to the values shown in Figures 3 (in the manuscript), Figure S9 (in sup mat), corrected for multiple testing using Benjamini, Hochberg, and Yekutieli's method (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).

	R.AllGenes	EN.AllGenes	L.AllGenes	R.FLNsGenes	EN.FLNsGenes	L.FLNsGenes	R.L1000Genes	EN.L1000Genes	L.L1000Genes
Spearman Correlation	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}	2.96×10^{-2}
Pearson Correlation	1.71×10^{-1}	2.01×10^{-1}	1.71×10^{-1}	1.71×10^{-1}	1.71×10^{-1}	1.71×10^{-1}	1.71×10^{-1}	1.71×10^{-1}	1.71×10^{-1}
RMSE	1.51×10^{-1}	2.92×10^{-1}	1.80×10^{-1}	1.80×10^{-1}	1.80×10^{-1}	1.93×10^{-1}	1.80×10^{-1}	1.51×10^{-1}	1.51×10^{-1}

Table S10. Prediction correlation (mean ± sd) and average standard deviation of the LOOCV estimates over 10 runs of our algorithm. To evaluate the reliability of the model predictions over multiple runs, we compute the variances of the LOOCV estimates on the TNBC data set. Specifically, the model is run with random seeds 10 times in a LOOCV setting to estimate the predictions. The average prediction correlation along with its standard deviation are shown demonstrating that the prediction performance is similar across multiple runs. We also show the average standard deviation of the prediction estimates across each drug group.

	MTOR	PI3K/MTOR	PI3K	PLK	ABL	CDK
Prediction correlation	0.58 ± 0.076	0.54 ± 0.055	0.28 ± 0.119	0.04 ± 0.11	0.3 ± 0.097	0.36 ± 0.092
Average Standard Deviation of LOOCV estimates	0.049	0.09	0.15	0.101	0.193	0.149

Table S11. Spearman correlations of predictions for individual drug groups in GDSC data set in comparison to Random Forest (RF), sparse Partial Least Squares (sPLS), Sparse Group Lasso (SGL) and Support Vector Machine (SVM).

	MVLR	RF	sPLS	SVM	SGL
EGFR	0.408	0.400	0.414	0.352	0.278
EGFR	0.218	0.128	0.287	0.156	0.121
MEK	0.662	0.685	0.680	0.661	0.643
BRAF	-0.090	0.089	-0.111	-0.042	0.133
MTOR	0.079	0.119	-0.039	0.221	0.086
PI3K	0.443	0.407	0.426	0.461	0.338
ALK	0.562	0.480	0.380	0.419	0.411
PLK	0.151	0.077	0.067	0.055	0.039
ABL	0.598	0.549	0.517	0.578	0.558
SRC	0.444	0.369	0.290	0.347	0.348
CDK	0.657	0.716	0.667	0.711	0.661
CDK	0.443	0.423	0.474	0.478	0.451
PDGFR	0.214	0.023	0.190	0.053	0.018
PDGFR	0.435	0.500	0.371	0.467	0.419
BCL	0.173	0.160	0.102	0.235	0.226
AKT	0.599	0.618	0.570	0.659	0.674
Average Correlation	0.375	0.359	0.330	0.363	0.338

Table S12. Pearson correlations of predictions for individual drug groups in GDSC data set in comparison to Random Forest (RF), sparse Partial Least Squares (sPLS), Sparse Group Lasso (SGL) and Support Vector Machine (SVM).

	MVLR	RF	sPLS	SVM	SGL
EGFR	0.433	0.352	0.385	0.352	0.318
EGFR	0.307	0.319	0.342	0.001	0.162
MEK	0.658	0.666	0.666	0.657	0.643
BRAF	-0.064	0.13	-0.121	-0.027	0.105
MTOR	0.087	0.091	-0.013	0.16	0.092
PI3K	0.457	0.435	0.44	0.482	0.354
ALK	0.611	0.513	0.422	0.459	0.419
PLK	0.102	0.085	0.063	0.158	-0.001
ABL	0.586	0.549	0.483	0.559	0.493
SRC	0.519	0.505	0.409	0.475	0.437
CDK	0.792	0.846	0.819	0.849	0.787
CDK	0.467	0.46	0.49	0.504	0.478
PDGFR	0.172	0.004	0.182	-0.026	0.011
PDGFR	0.427	0.508	0.35	0.463	0.404
BCL	0.186	0.178	0.111	0.001	0.189
AKT	0.691	0.704	0.659	0.652	0.687
Average Correlation	0.402	0.397	0.355	0.357	0.349

Table S13. RMSE of predictions for individual drug groups in GDSC data set in comparison to Random Forest (RF), sparse Partial Least Squares (sPLS), Sparse Group Lasso (SGL) and Support Vector Machine (SVM).

	MVLR	RF	sPLS	SVM	SGL
EGFR	0.907	0.933	0.925	0.933	1.018
EGFR	0.968	0.943	0.952	10.154	1.119
MEK	0.87	0.852	0.856	0.857	0.893
BRAF	1.112	0.993	1.142	480.971	1.102
MTOR	1.07	1.01	1.137	1.011	1.14
PI3K	0.988	0.994	0.995	0.972	1.103
ALK	0.91	0.988	1.059	1.03	1.094
PLK	1.108	1.011	1.132	112.558	1.197
ABL	0.968	0.998	1.058	0.994	1.099
SRC	0.902	0.886	0.958	0.904	0.977
CDK	1.167	1.006	1.087	0.997	1.19
CDK	0.973	0.972	0.97	0.949	0.998
PDGFR	1.032	1.023	1.001	45.706	1.175
PDGFR	1.005	0.946	1.062	0.98	1.075
BCL	0.999	0.978	1.028	25.377	1.082
AKT	1.037	1.017	1.088	1.147	1.058
Average RMSE	1.001	0.972	1.028	42.846	1.083

Table S14. Statistical significance of the predictive performances on the GDSC data set. P-values from one-sided paired Wilcoxon Sign-Rank test corresponding to the values shown in Table S11, S12 and S13.

	RF	sPLS	SVM	SGL
Spearman Correlation	0.17	0.01	0.3	0.05
Pearson Correlation	0.26	0.01	0.14	0.01
RMSE	0.96	0.03	0.04	0

Table S15. Spearman correlations of predictions for individual drug groups in FIMM data set in comparison to Random Forest (RF), sparse Partial Least Squares (sPLS), Sparse Group Lasso (SGL) and Support Vector Machine (SVM).

	MVLR	RF	sPLS	SVM	SGL
MTOR	0.566	0.494	0.518	0.51	0.466
PI3K/MTOR	0.533	0.501	0.438	0.462	0.427
PI3K	0.362	-0.017	0.286	0.366	0.007
PLK	0.132	0.229	-0.275	0.056	0.231
ABL	0.219	0.084	0.251	0.303	0.304
CDK	0.202	0.477	0.421	0.397	0.366
Average Correlation	0.336	0.295	0.273	0.349	0.300

Table S16. Pearson correlations of predictions for individual drug groups in FIMM data set in comparison to Random Forest (RF), sparse Partial Least Squares (sPLS), Sparse Group Lasso (SGL) and Support Vector Machine (SVM).

	MVLR	RF	sPLS	SVM	SGL
MTOR	0.604	0.573	0.484	0.497	0.54
PI3K/MTOR	0.499	0.521	0.381	0.409	0.36
PI3K	0.332	-0.005	0.278	0.368	0.079
PLK	0.085	0.138	-0.169	0.057	0.251
ABL	0.185	0.079	0.193	0.224	0.231
CDK	0.18	0.454	0.451	0.356	0.285
Average Correlation	0.314	0.293	0.270	0.319	0.291

Table S17. RMSE of predictions for individual drug groups in FIMM data set in comparison to Random Forest (RF), sparse Partial Least Squares (sPLS), Sparse Group Lasso (SGL) and Support Vector Machine (SVM).

	MVLR	RF	PLS	SVM	SGL
MTOR	0.997	1.028	1.123	1.126	1.069
PI3K/MTOR	0.968	0.956	1.074	1.055	1.106
PI3K	0.907	0.991	0.925	0.947	1.001
PLK	0.986	0.954	1.51	1.047	1.05
ABL	1.013	1.006	1.008	1.028	1.002
CDK	1.393	0.95	0.994	1.083	1.177
Average RMSE	1.044	0.981	1.106	1.048	1.068