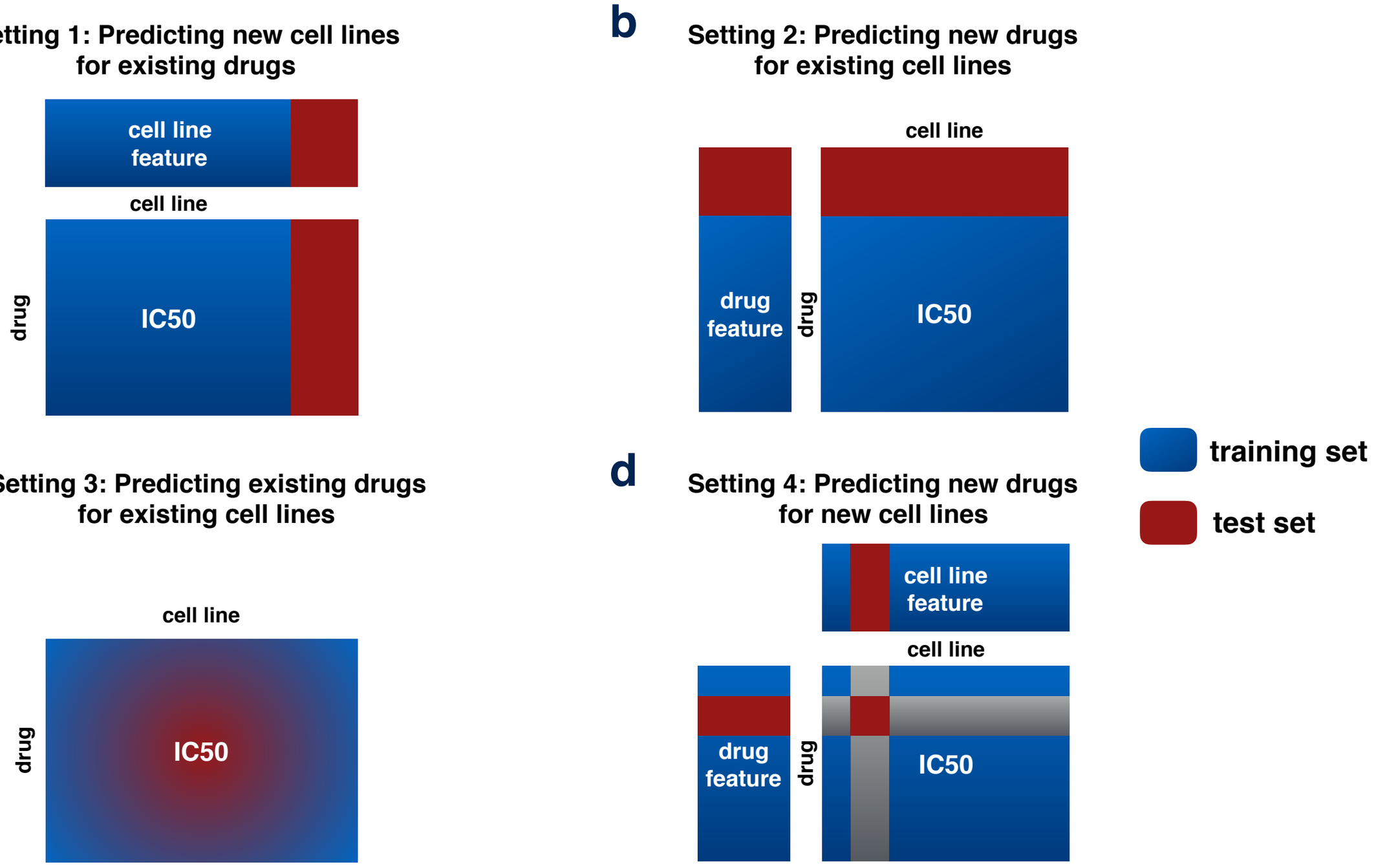
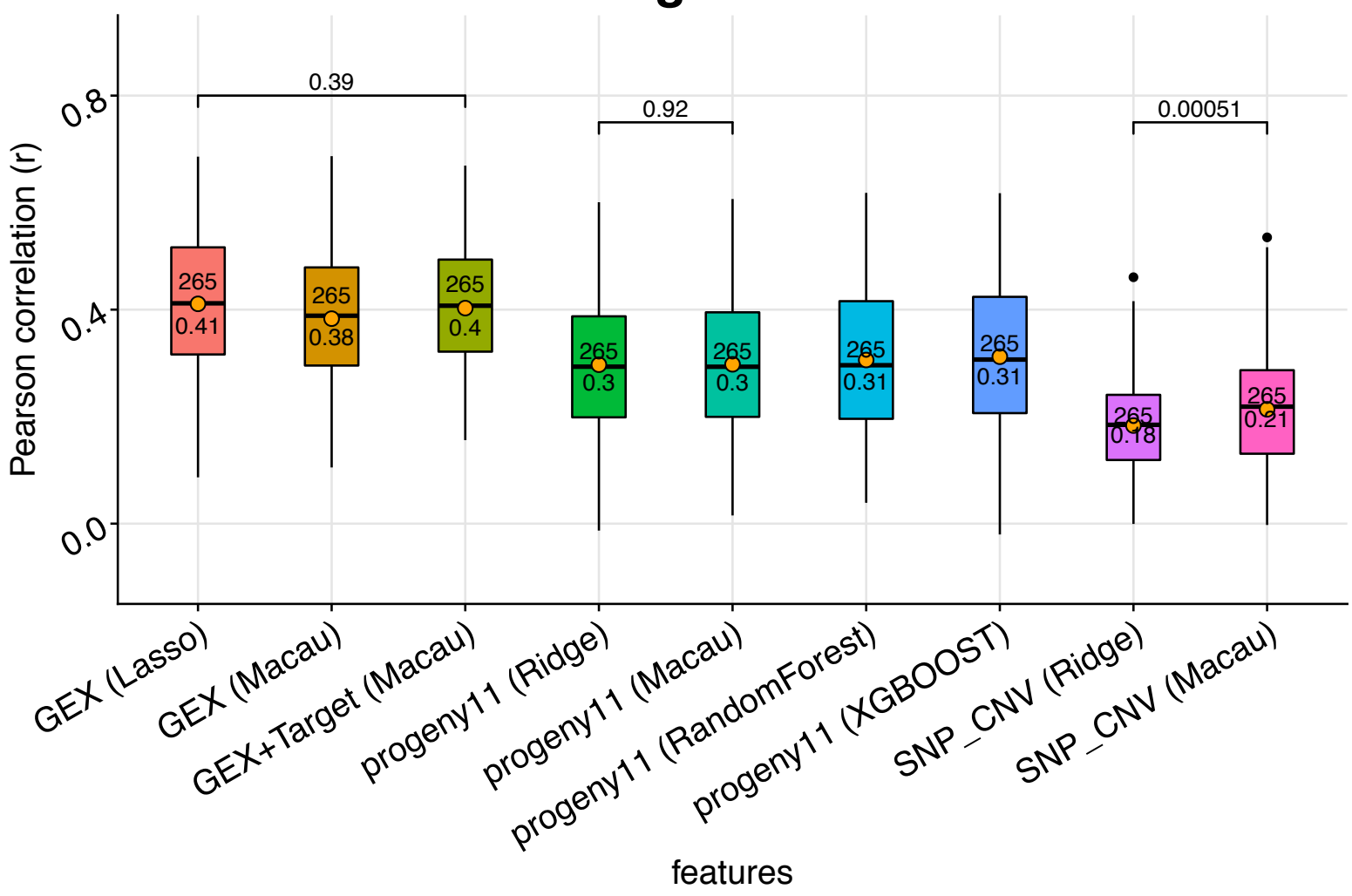
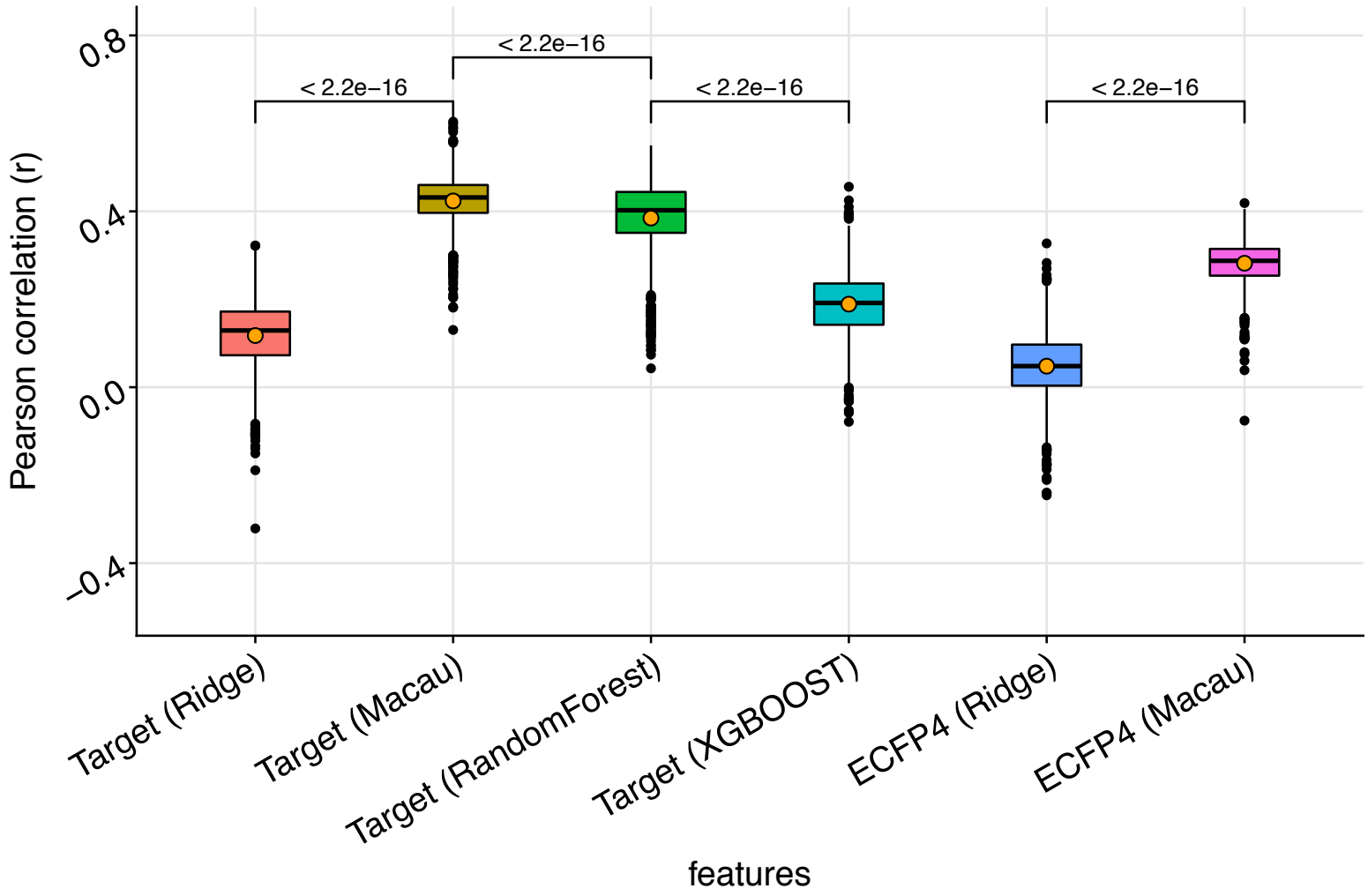


Supplementary Fig. S1. Different settings in drug response prediction. **(a)** Predicting new cell lines for existing drugs. For each drug, we compute the pearson correlation of observed versus predicted IC50 across all cell lines of the test set. **(b)** Predicting new drugs for existing cell lines. For each cell line we compute the pearson correlation of observed versus predicted IC50 across all drugs of the test set. **(c)** Predicting existing drugs for existing cell lines. This is a missing value imputation setting where side information of drug and cell lines are not required, but can be used to improve the result. The test data is defined by a percentage of the whole dataset. We compute the pearson correlation of observed versus predicted IC50 for all randomly chosen drug - cell line pairs of the test set. **(d)** Predicting new drugs for new cell lines. We do 2 simultaneous cross validation on both drug and cell line sides. The test data is defined by association of the test set of the drug side with the test set of the cell lines side. We compute the pearson correlation of observed versus predicted IC50 for all drug - cell line pairs of the test set.

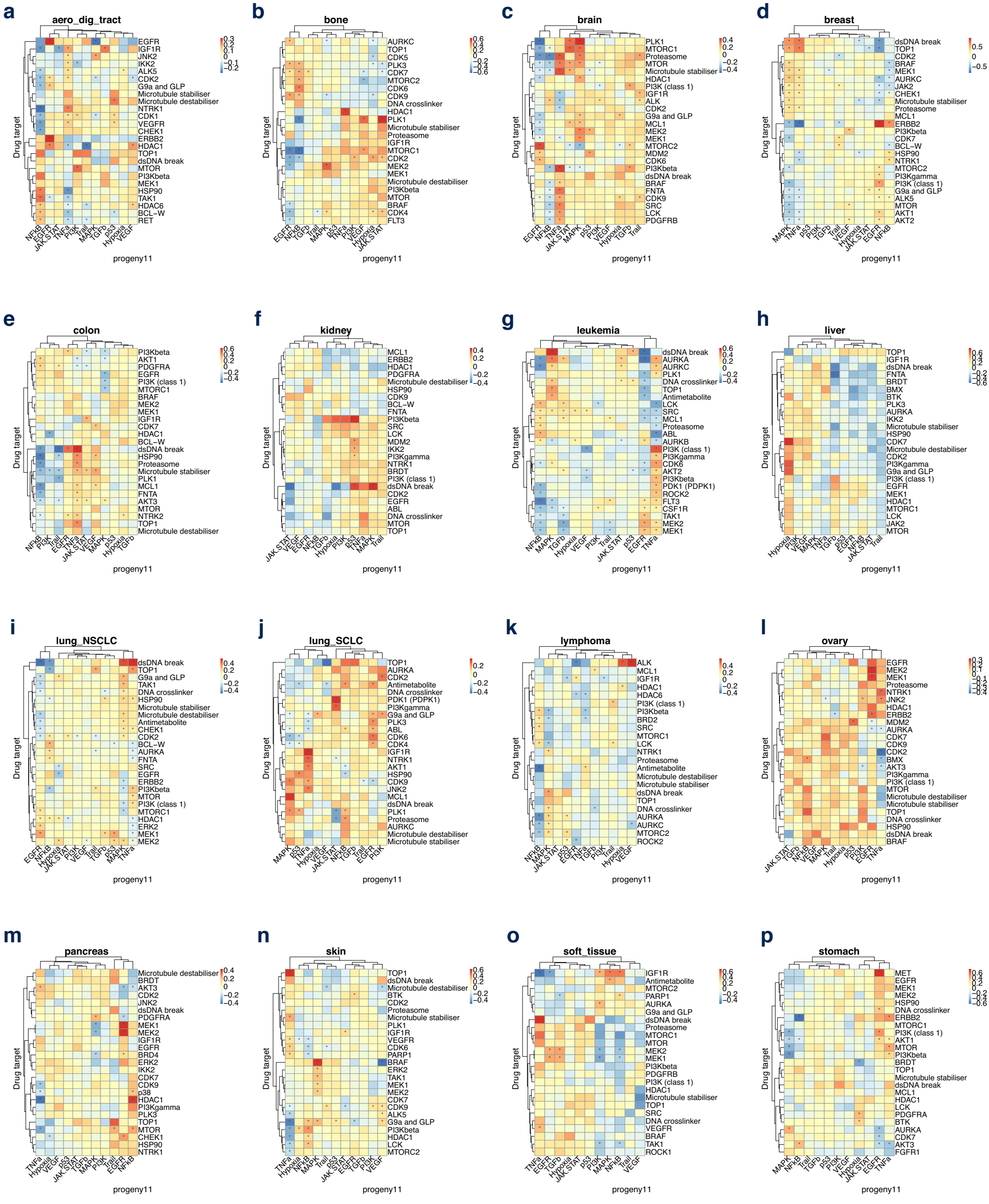
a



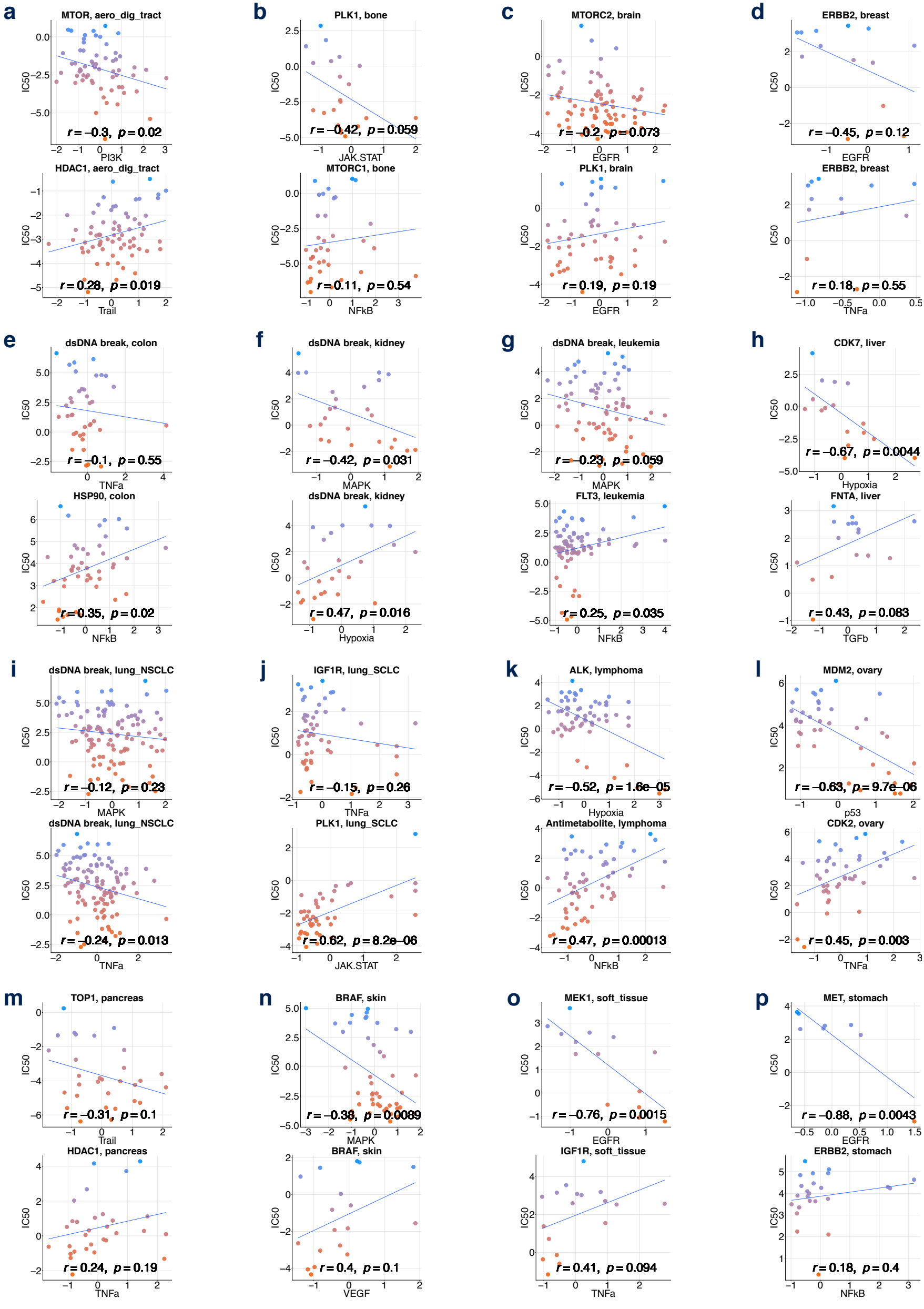
b



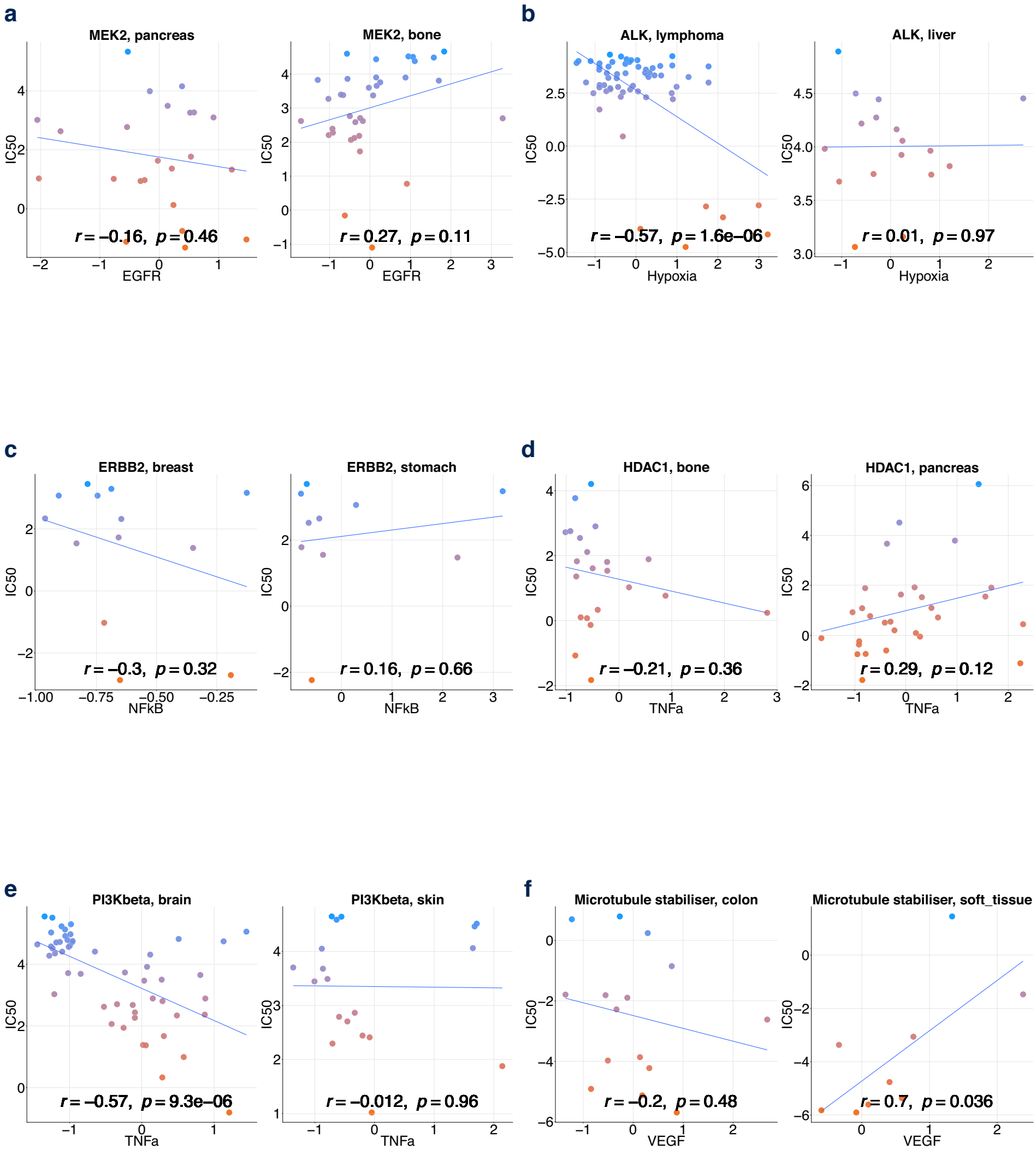
Supplementary Fig. S2. Drug response prediction performance. (a) We compare prediction performance (correlation of observed versus predicted IC50) of existing drugs on new cell lines. We use Macau, ridge/lasso regression, Random Forest and XGBOOST. The features are gene expression, pathway activity, mutation (SNP) and copy number variation (CNV). **(b)** We compare prediction performance of existing cell lines on new drugs. The features are drug protein targets and ECFP4 fingerprint.



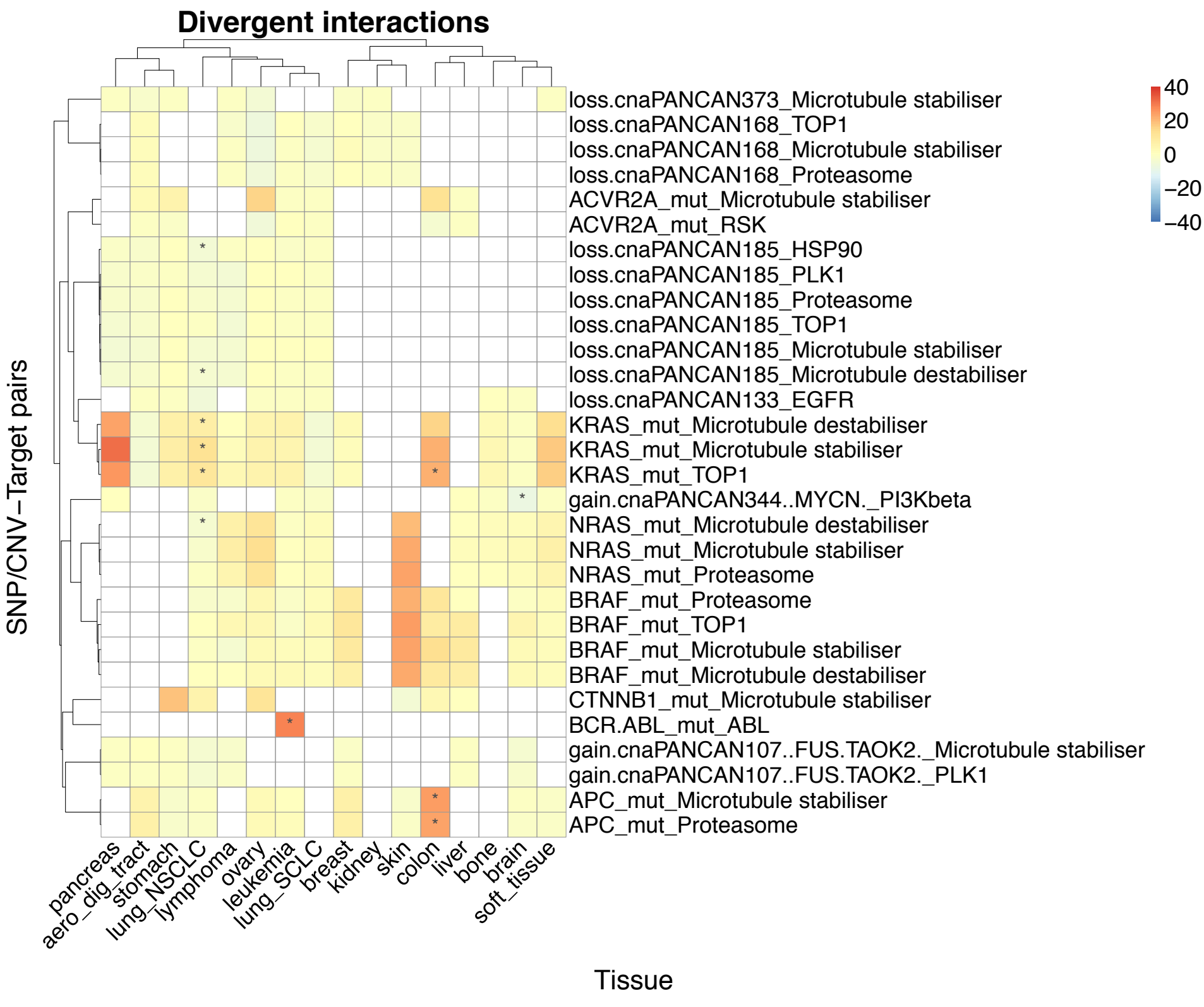
Supplementary Fig. S3. Tissue specific analysis of interaction matrix. We chose 16 tissues in the GDSC panel with at least 20 samples. We kept the targets which have an interaction for at least 1 pathway in the top 5% absolute value. We subset a second time by keeping the top 25 targets with the highest variance across the pathways in term of interaction value.



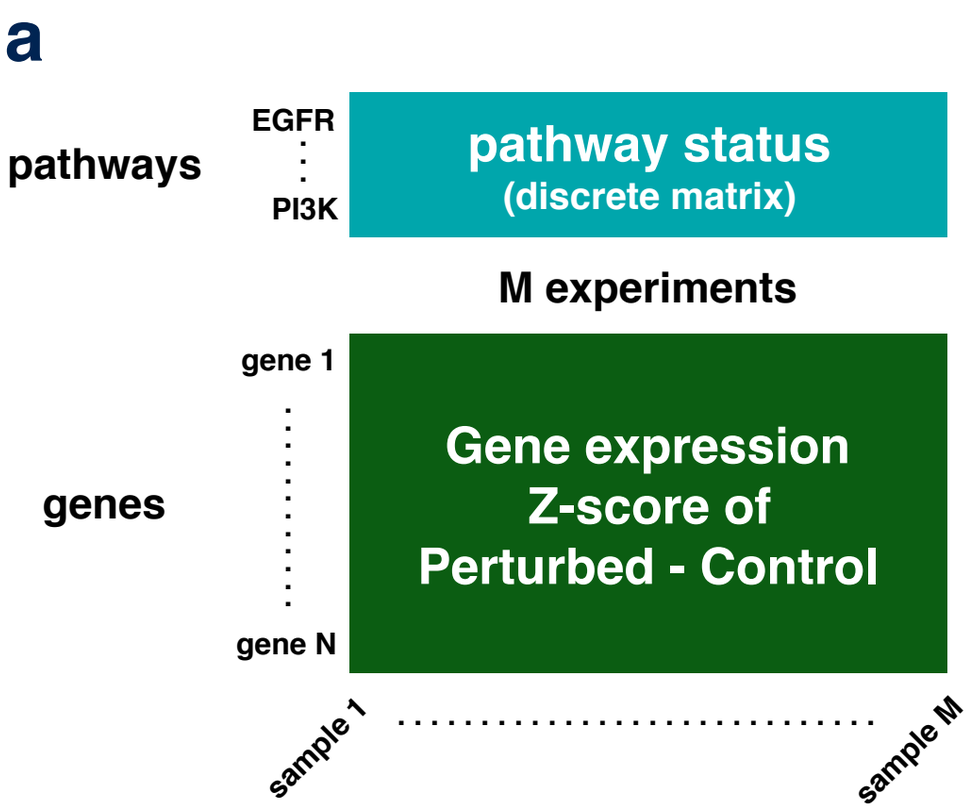
Supplementary Fig. S4. PROGENy as biomarker. For each tissue specific interaction matrix, we select a top positive interaction and a top negative interaction. For both target - pathway pairs, we then find a drug which targets this protein (as described in the manually curated list) and plot its IC50 (log scale) against the corresponding pathway's activity in the specific tissue.

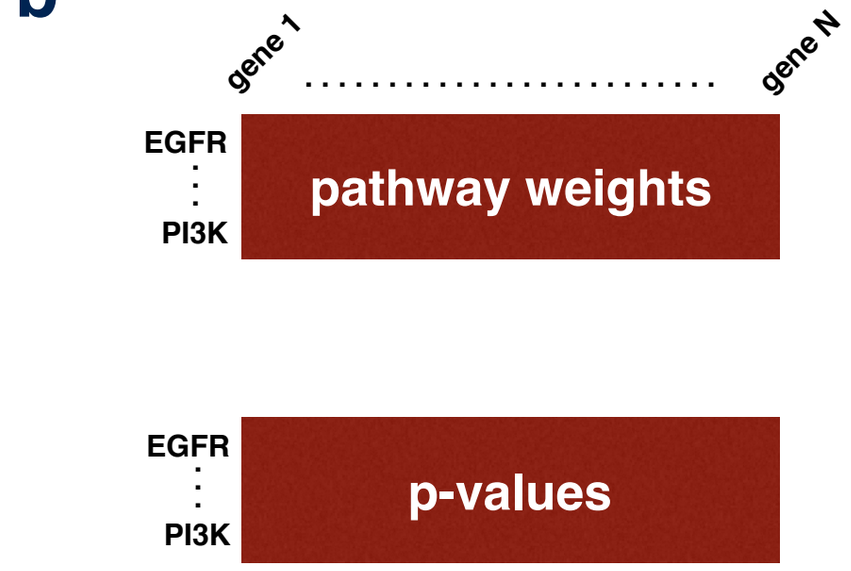


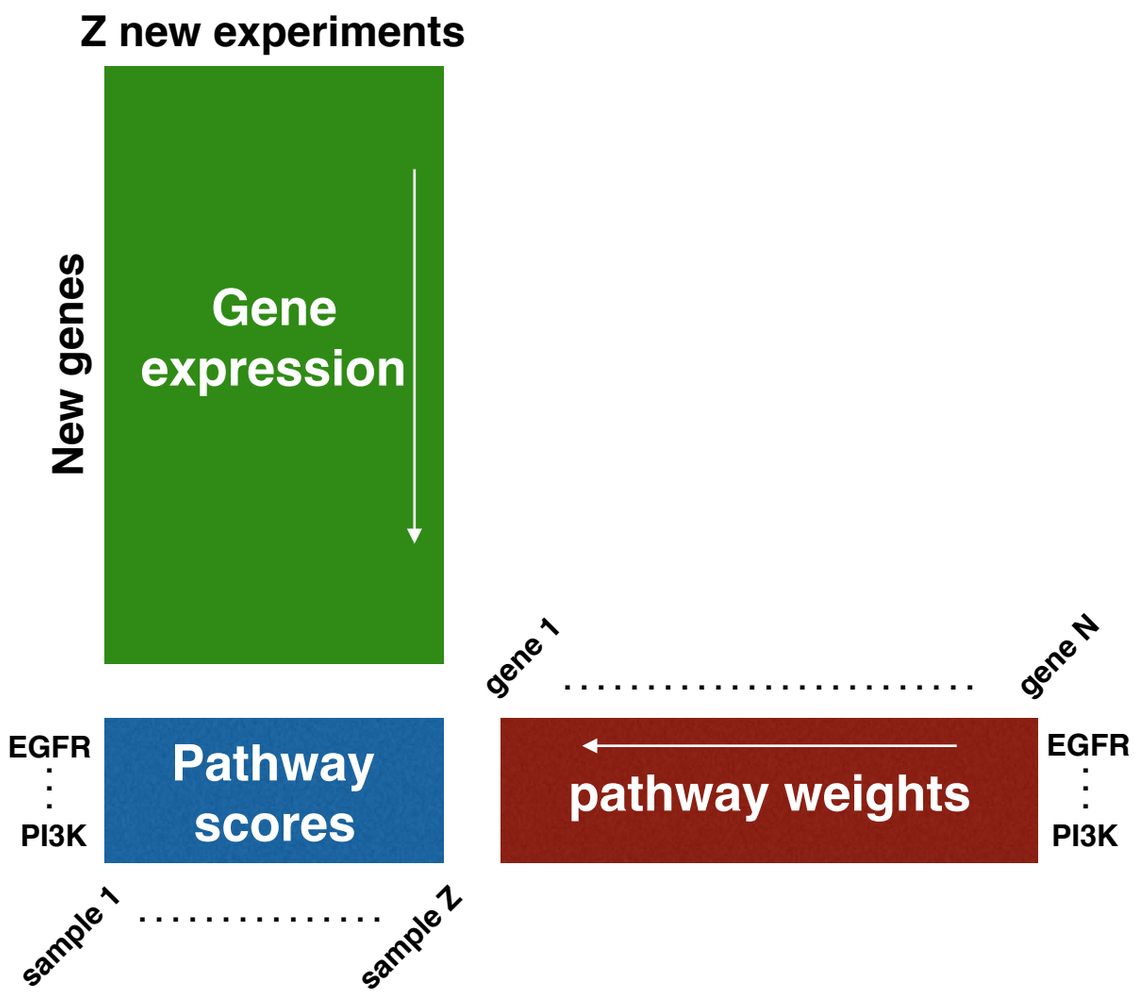
Supplementary Fig. S5. Antagonistic tissues based on target pathway interaction. For all target - pathway pairs which have opposite effect from one tissue to another, we select a drug which specifically targets the protein and plot the drug's IC50 as function of PROGENy activity for the corresponding tissues.



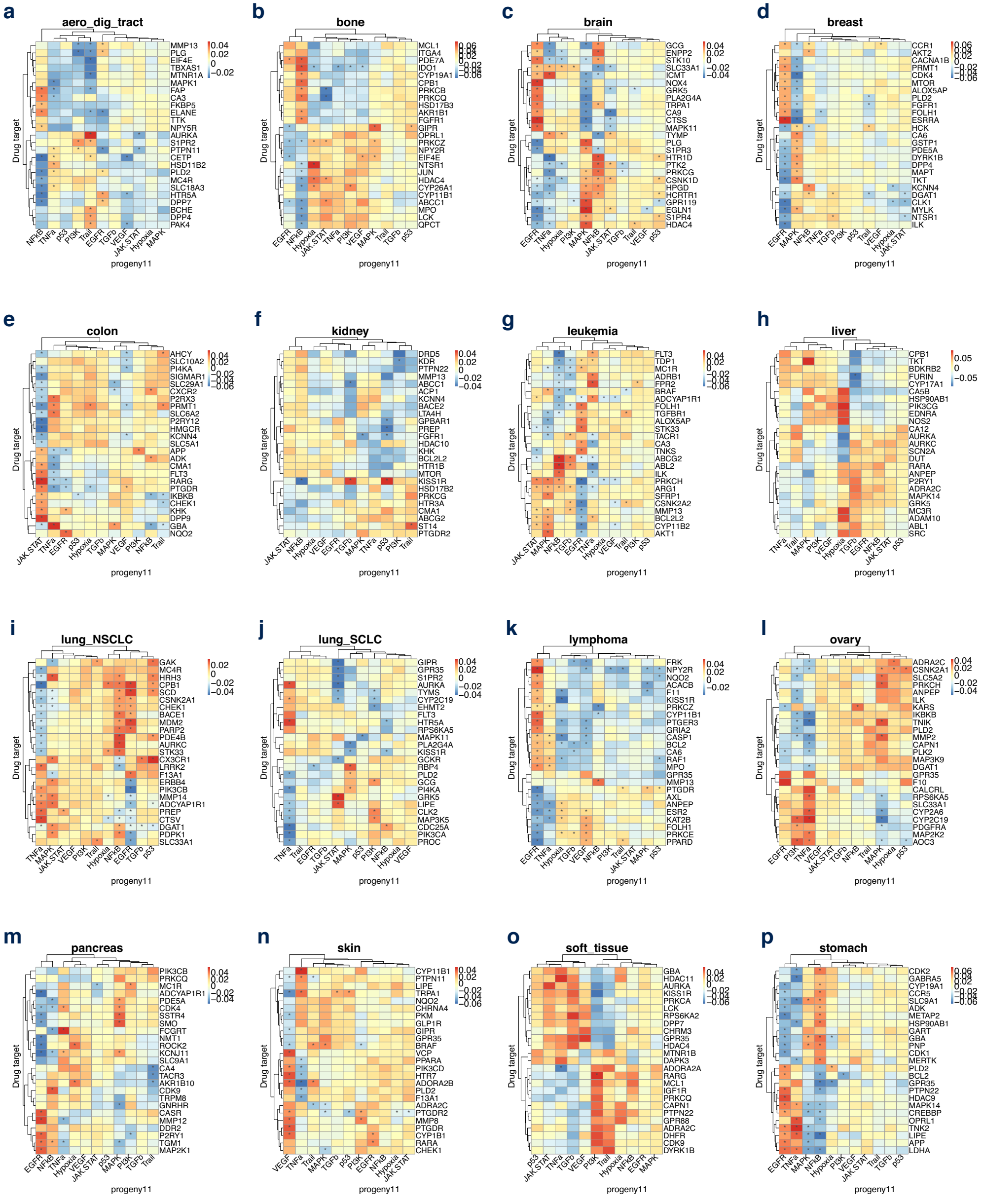
Supplementary Fig. S6. **Feature interaction analysis across tissues for SNP/CNV**. We vectorize all cancer specific interaction matrices between target and SNP/CNV and obtain a matrix of dimension (number of tissues x number of SNP/CNV-target pairs). We do a first subsetting by taking the pairs for which at least one pathway appears in the top 1% highest value, and chose 15 SNP/CNV-target pairs with highest variance of interaction across tissues. We then subset by taking the pairs for which at least one pathway appears in the top 1% lowest value, and chose 15 SNP/CNV-target pairs with highest variance of interaction across tissues. We combine the top hits and then keep the 30 pathway-target pairs. White color indicates when the mutation or CNV is not present.



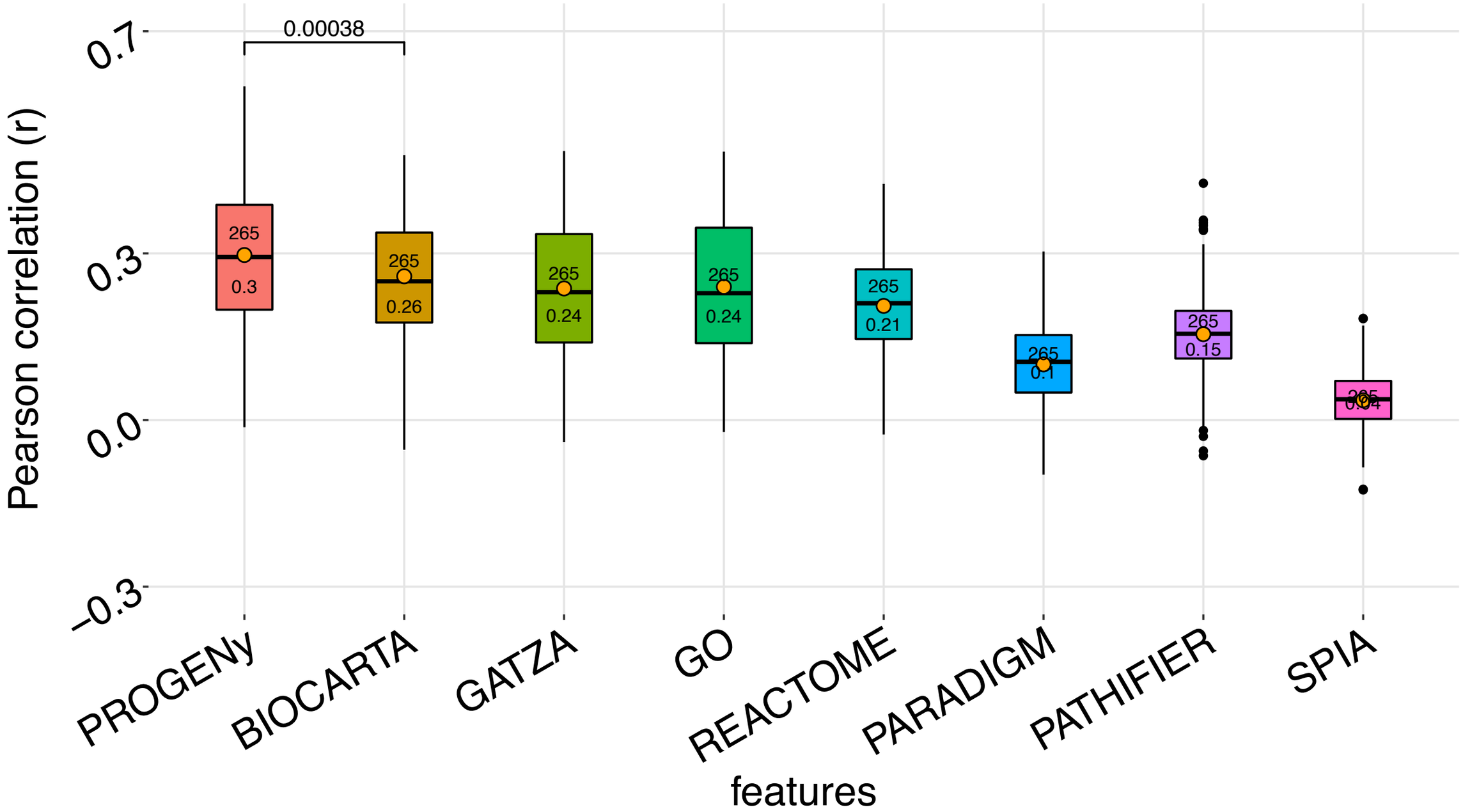




Supplementary Fig. S7. Workflow to produce PROGENy scores. (a) We fit a linear model for each z-score of the perturbation in function of the pathway status. **(b)** We select for each pathway, the top 100 genes with smallest p-values. **(c)** We compute pathway scores for new gene expression dataset by a matrix multiplication with the weight matrix.



Supplementary Fig. S8. Tissue specific analysis of interaction matrix using predicted drug target. We chose 16 tissues in the GDSC panel with at least 20 samples. We kept the targets which have an interaction for at least 1 pathway in the top 5% absolute value. We subset a second time by keeping the top 25 targets with the highest variance across the pathways in term of interaction value.



Supplementary Fig. S9. Drug response prediction performance of different pathway methods. We compare prediction performance (correlation of observed versus predicted IC50) of existing drugs on new cell lines. We use elastic net regression. The features are pathway scores derived from different genesets, as described in Schubert et al.