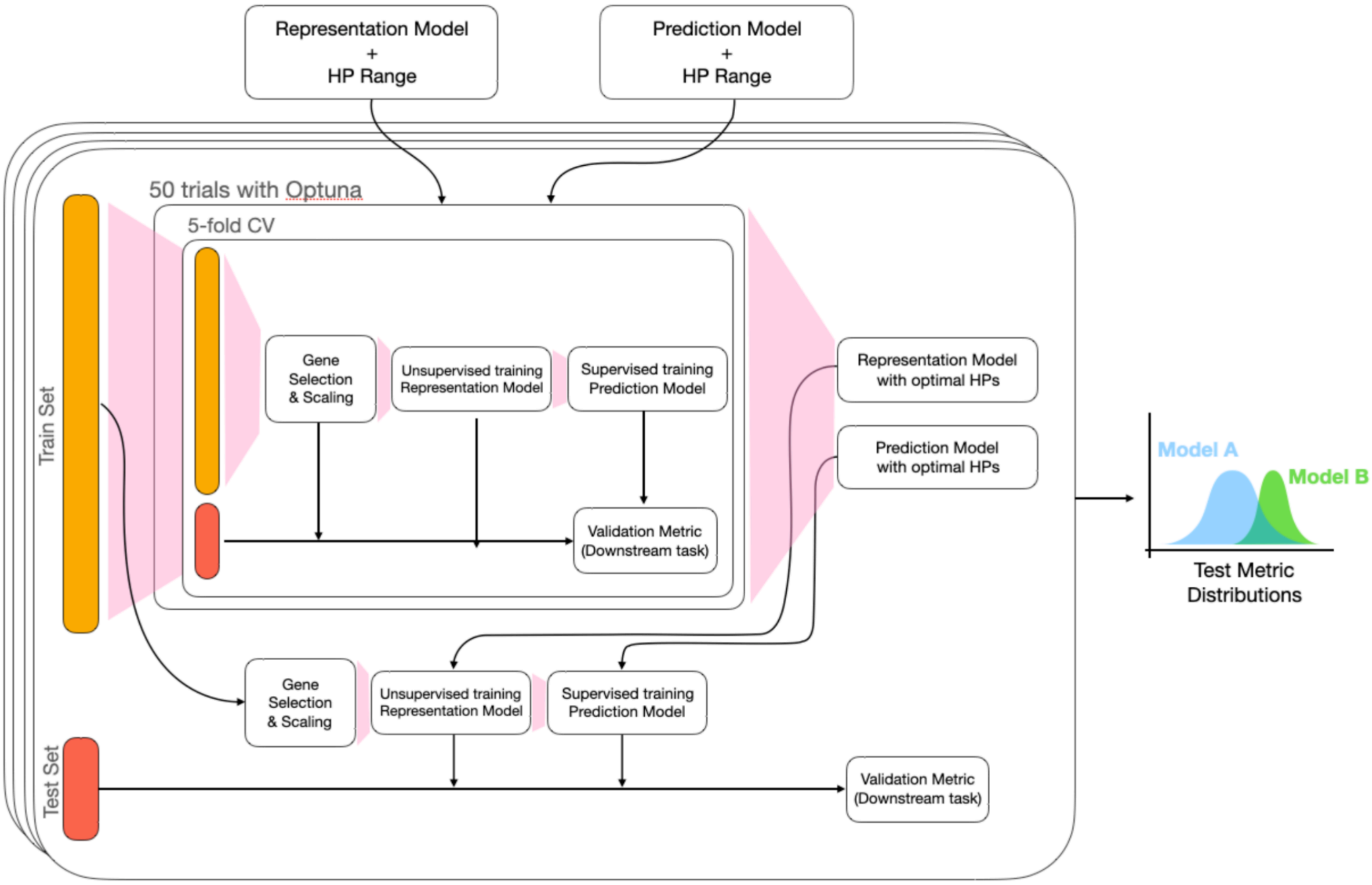
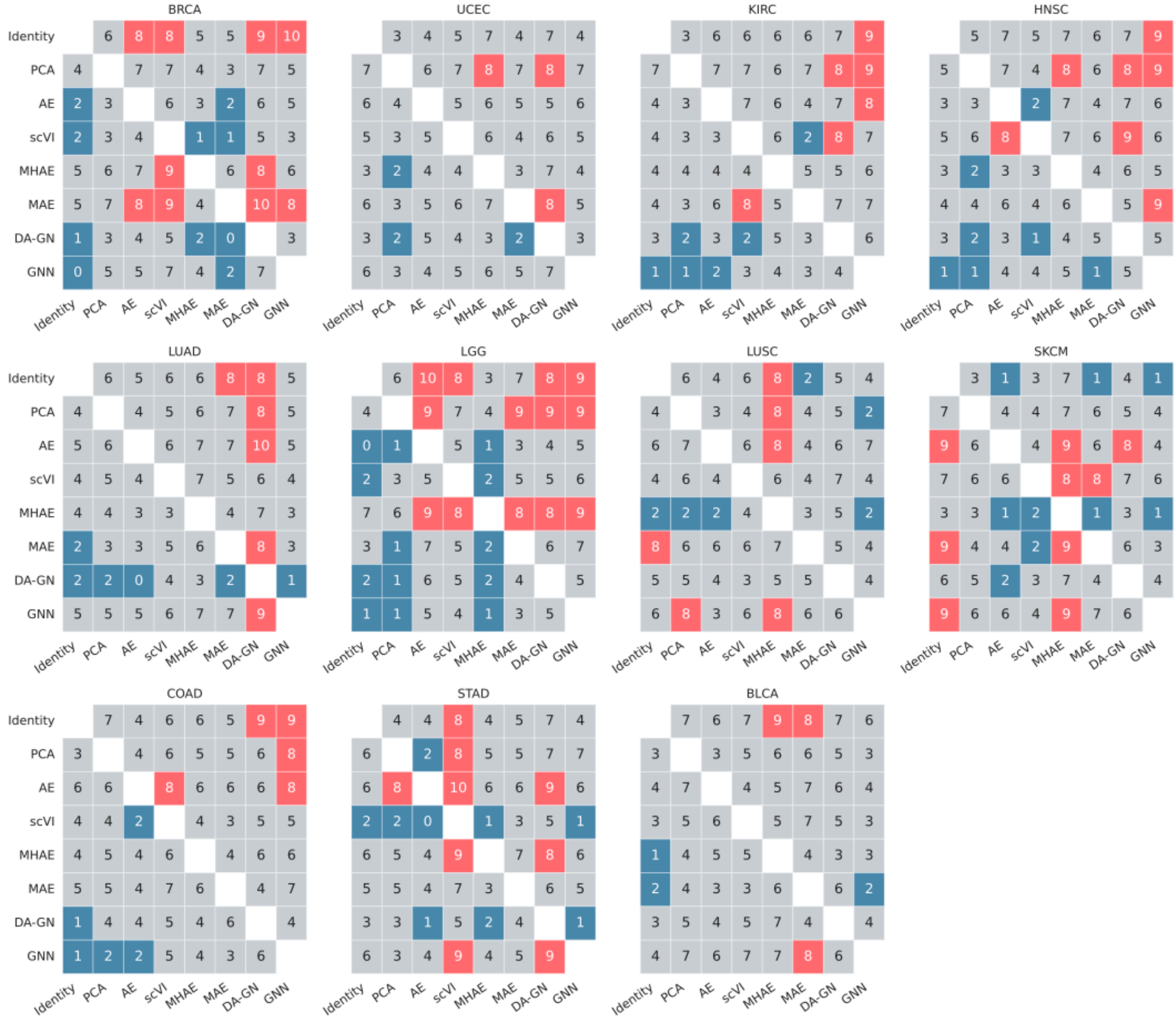
Supplementary Materials

Supplemental Figures


Representation Model
+
HP Range
Prediction Model
+
HP Range
50 trials with Optuna
5-fold CV
Train Set
Gene Selection & Scaling
Unsupervised training Representation Model
Supervised training Prediction Model
Validation Metric (Downstream task)
Representation Model with optimal HPs
Prediction Model with optimal HPs
Model A
Model B
Test Metric Distributions
Gene Selection & Scaling
Unsupervised training Representation Model
Supervised training Prediction Model
Test Set
Validation Metric (Downstream task)

Supplementary Figure S1. Repeated hold-out pipeline description. The different planes correspond to repetitions of the process, allowing to create the scores distributions per model seen on the right of the figure.



BRCA

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		6	8	8	5	5	9	10
PCA	4		7	7	4	3	7	5
AE	2	3		6	3	2	6	5
scVI	2	3	4		1	1	5	3
MHAE	5	6	7	9		6	8	6
MAE	5	7	8	9	4		10	8
DA-GN	1	3	4	5	2	0		3
GNN	0	5	5	7	4	2	7	

UCEC

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		3	4	5	7	4	7	4
PCA	7		6	7	8	7	8	7
AE	6	4		5	6	5	5	6
scVI	5	3	5		6	4	6	5
MHAE	3	2	4	4		3	7	4
MAE	6	3	5	6	7		8	5
DA-GN	3	2	5	4	3	2		3
GNN	6	3	4	5	6	5	7	

KIRC

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		3	6	6	6	6	7	9
PCA	7		7	7	6	7	8	9
AE	4	3		7	6	4	7	8
scVI	4	3	3		6	2	8	7
MHAE	4	4	4	4		5	5	6
MAE	4	3	6	8	5		7	7
DA-GN	3	2	3	2	5	3		6
GNN	1	1	2	3	4	3	4	

HNSC

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		5	7	5	7	6	7	9
PCA	5		7	4	8	6	8	9
AE	3	3		2	7	4	7	6
scVI	5	6	8		7	6	9	6
MHAE	3	2	3	3		4	6	5
MAE	4	4	6	4	6		5	9
DA-GN	3	2	3	1	4	5		5
GNN	1	1	4	4	5	1	5	

LUAD

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		6	5	6	6	8	8	5
PCA	4		4	5	6	7	8	5
AE	5	6		6	7	7	10	5
scVI	4	5	4		7	5	6	4
MHAE	4	4	3	3		4	7	3
MAE	2	3	3	5	6		8	3
DA-GN	2	2	0	4	3	2		1
GNN	5	5	5	6	7	7	9	

LGG

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		6	10	8	3	7	8	9
PCA	4		9	7	4	9	9	9
AE	0	1		5	1	3	4	5
scVI	2	3	5		2	5	5	6
MHAE	7	6	9	8		8	8	9
MAE	3	1	7	5	2		6	7
DA-GN	2	1	6	5	2	4		5
GNN	1	1	5	4	1	3	5	

LUSC

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		6	4	6	8	2	5	4
PCA	4		3	4	8	4	5	2
AE	6	7		6	8	4	6	7
scVI	4	6	4		6	4	7	4
MHAE	2	2	2	4		3	5	2
MAE	8	6	6	6	7		5	4
DA-GN	5	5	4	3	5	5		4
GNN	6	8	3	6	8	6	6	

SKCM

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		3	1	3	7	1	4	1
PCA	7		4	4	7	6	5	4
AE	9	6		4	9	6	8	4
scVI	7	6	6		8	8	7	6
MHAE	3	3	1	2		1	3	1
MAE	9	4	4	2	9		6	3
DA-GN	6	5	2	3	7	4		4
GNN	9	6	6	4	9	7	6	

COAD

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		7	4	6	6	5	9	9
PCA	3		4	6	5	5	6	8
AE	6	6		8	6	6	6	8
scVI	4	4	2		4	3	5	5
MHAE	4	5	4	6		4	6	6
MAE	5	5	4	7	6		4	7
DA-GN	1	4	4	5	4	6		4
GNN	1	2	2	5	4	3	6	

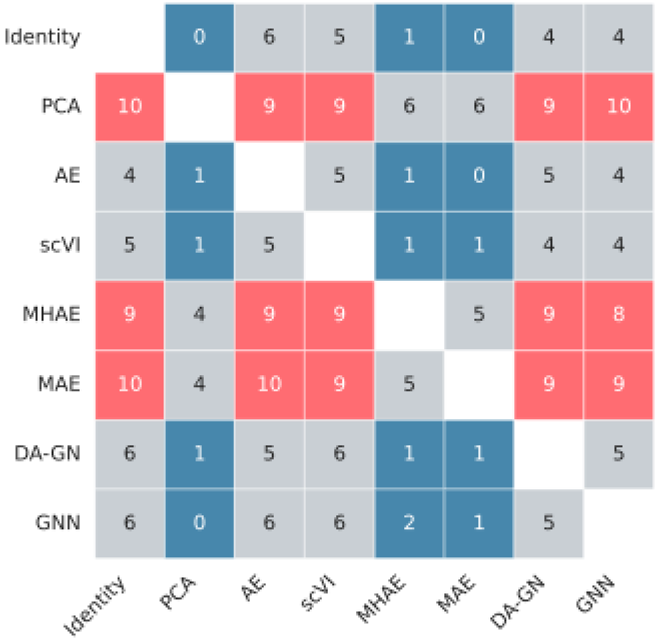
STAD

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		4	4	8	4	5	7	4
PCA	6		2	8	5	5	7	7
AE	6	8		10	6	6	9	6
scVI	2	2	0		1	3	5	1
MHAE	6	5	4	9		7	8	6
MAE	5	5	4	7	3		6	5
DA-GN	3	3	1	5	2	4		1
GNN	6	3	4	9	4	5	9	

BLCA

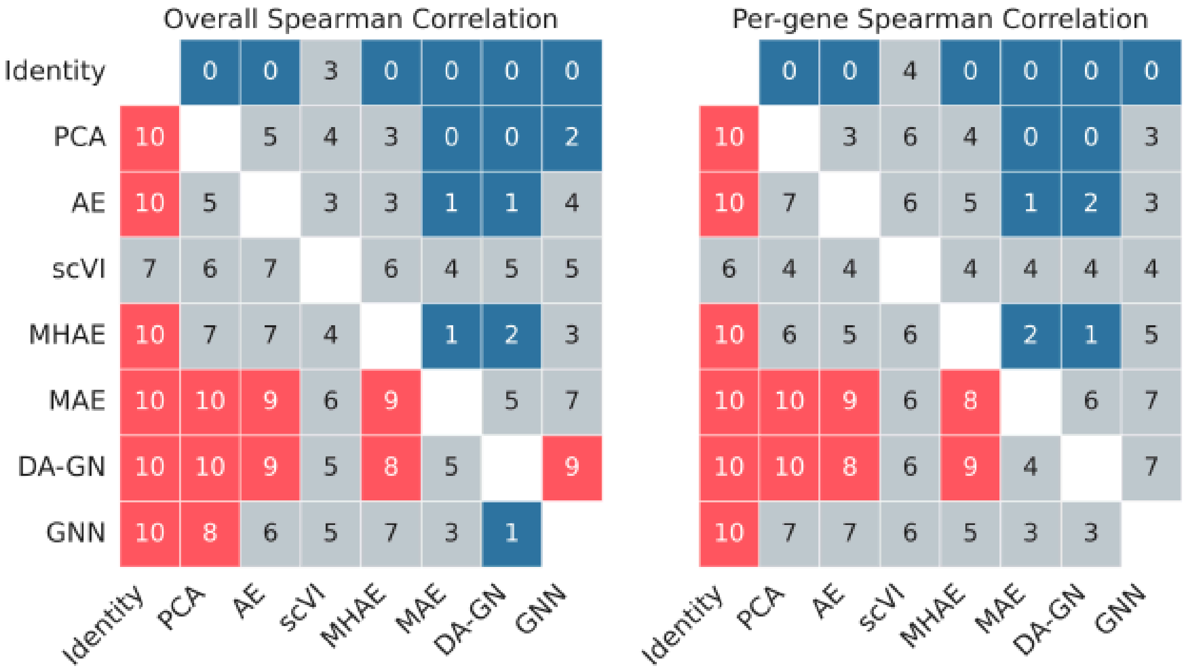
	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		7	6	7	9	8	7	6
PCA	3		3	5	6	6	5	3
AE	4	7		4	5	7	6	4
scVI	3	5	6		5	7	5	3
MHAE	1	4	5	5		4	3	3
MAE	2	4	3	3	6		6	2
DA-GN	3	5	4	5	7	4		4
GNN	4	7	6	7	7	8	6	

Supplementary Figure S2. Comparison of performance on per-cohort OS prediction task on different TCGA cohorts for different bulk RNA-seq representation models. *Number of folds for which the c-index for the model on the y axis is higher than for the model on the x axis. Red (Blue) boxes indicate 75% acceptance criterion on test folds for superiority is satisfied by the model on the y (x) axis.*



	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		0	6	5	1	0	4	4
PCA	10		9	9	6	6	9	10
AE	4	1		5	1	0	5	4
scVI	5	1	5		1	1	4	4
MHAE	9	4	9	9		5	9	8
MAE	10	4	10	9	5		9	9
DA-GN	6	1	5	6	1	1		5
GNN	6	0	6	6	2	1	5	

Supplementary Figure S3. **Comparison of performance on pan-cancer OS prediction task for different bulk RNA-seq representation models.** *Number of folds for which the c-index for the model on the y axis is higher than for the model on the x axis. Red (Blue) boxes indicate 75% acceptance criterion on test folds for superiority is satisfied by the model on the y (x) axis.*



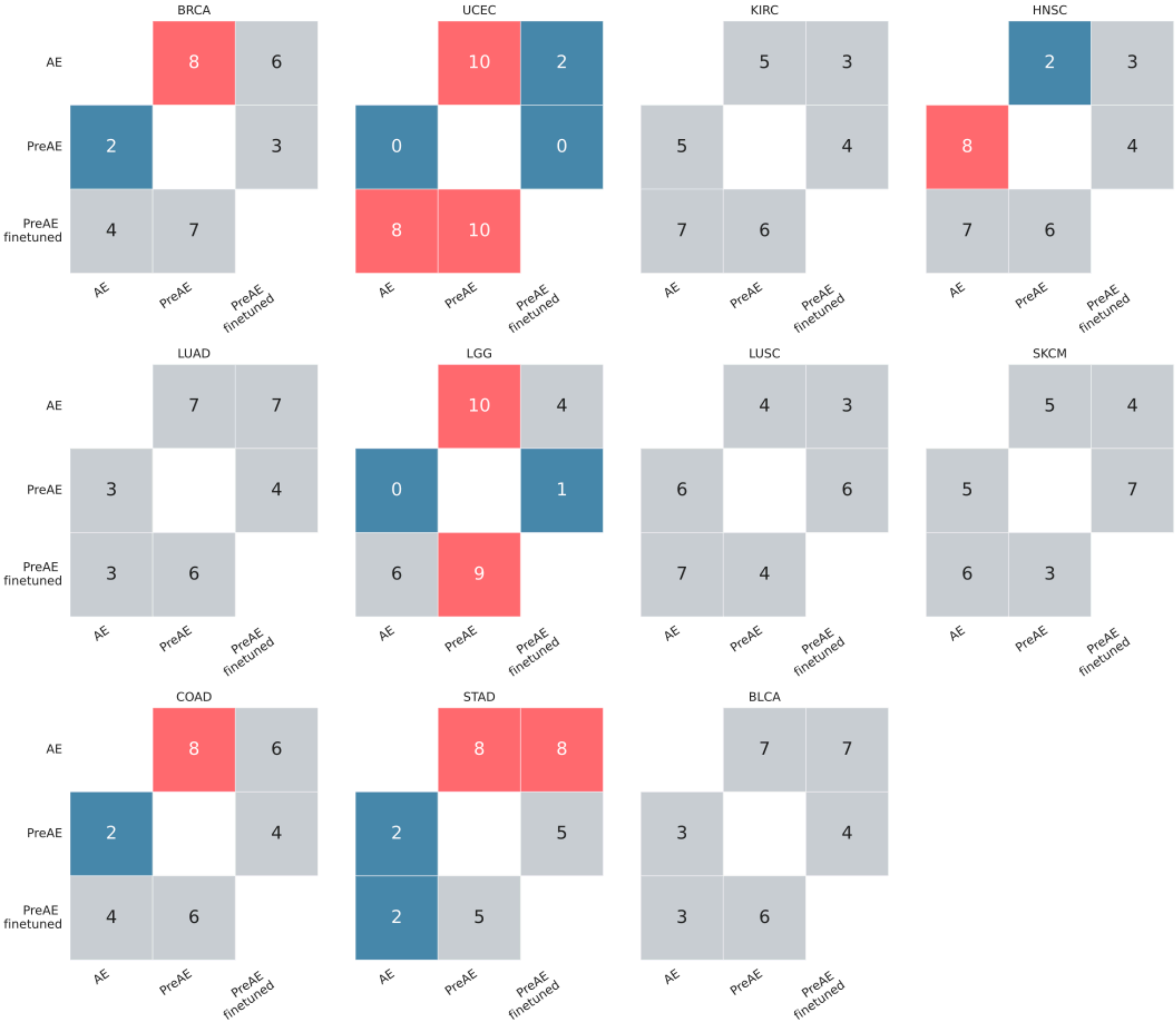
Overall Spearman Correlation

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		0	0	3	0	0	0	0
PCA	10		5	4	3	0	0	2
AE	10	5		3	3	1	1	4
scVI	7	6	7		6	4	5	5
MHAE	10	7	7	4		1	2	3
MAE	10	10	9	6	9		5	7
DA-GN	10	10	9	5	8	5		9
GNN	10	8	6	5	7	3	1	

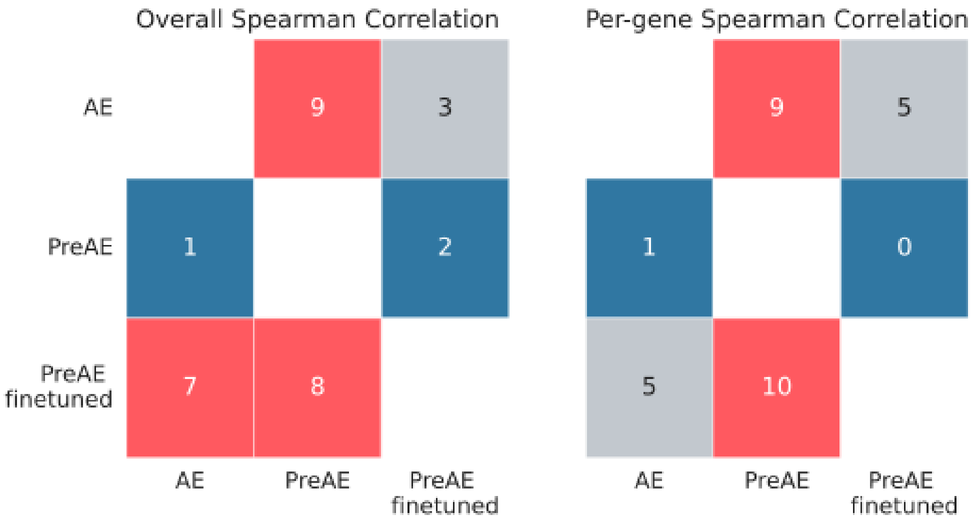
Per-gene Spearman Correlation

	Identity	PCA	AE	scVI	MHAE	MAE	DA-GN	GNN
Identity		0	0	4	0	0	0	0
PCA	10		3	6	4	0	0	3
AE	10	7		6	5	1	2	3
scVI	6	4	4		4	4	4	4
MHAE	10	6	5	6		2	1	5
MAE	10	10	9	6	8		6	7
DA-GN	10	10	8	6	9	4		7
GNN	10	7	7	6	5	3	3	

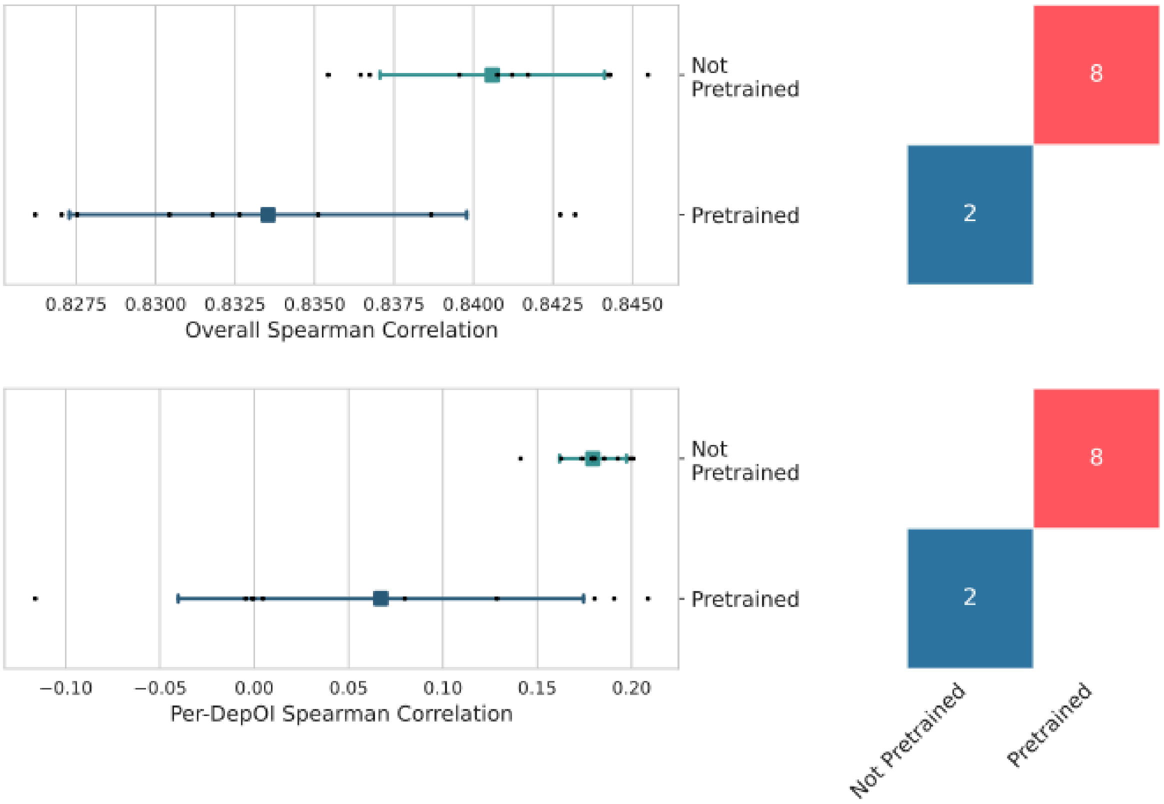
Supplementary Figure S4. **Comparison of performance on gene essentiality prediction task on DepMap dataset for different bulk RNA-seq representation models.** *Top panel) Number of folds for which the overall correlation for the model on the y axis is higher than for the model on the x axis. Red (Blue) boxes indicate 75% acceptance criterion on test folds for superiority is satisfied by the model on the y (x) axis. Bottom panel) Same as Top panel, but correlation computed per-gene.*



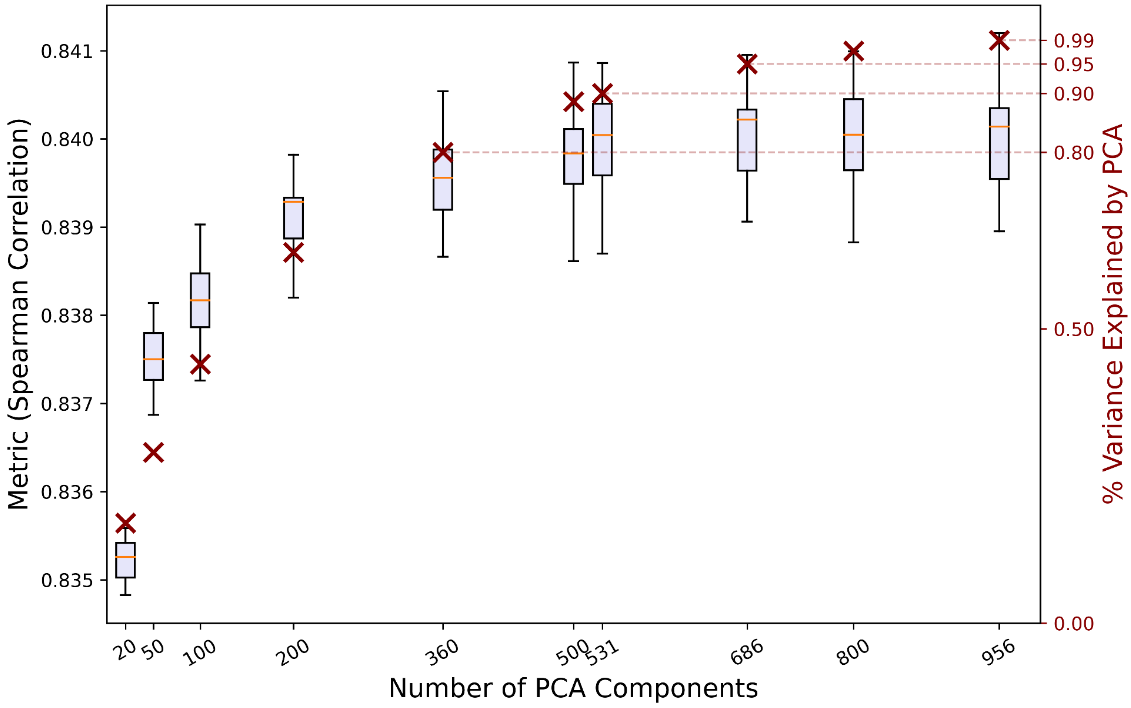
Supplementary Figure S5. Comparison of performance on per-cohort OS prediction task on different TCGA cohorts for pretraining experiments. *Number of folds for which the c-index for the model on the y axis is higher than for the model on the x axis. Red (Blue) boxes indicate 75% acceptance criterion on test folds for superiority is satisfied by the model on the y (x) axis.*



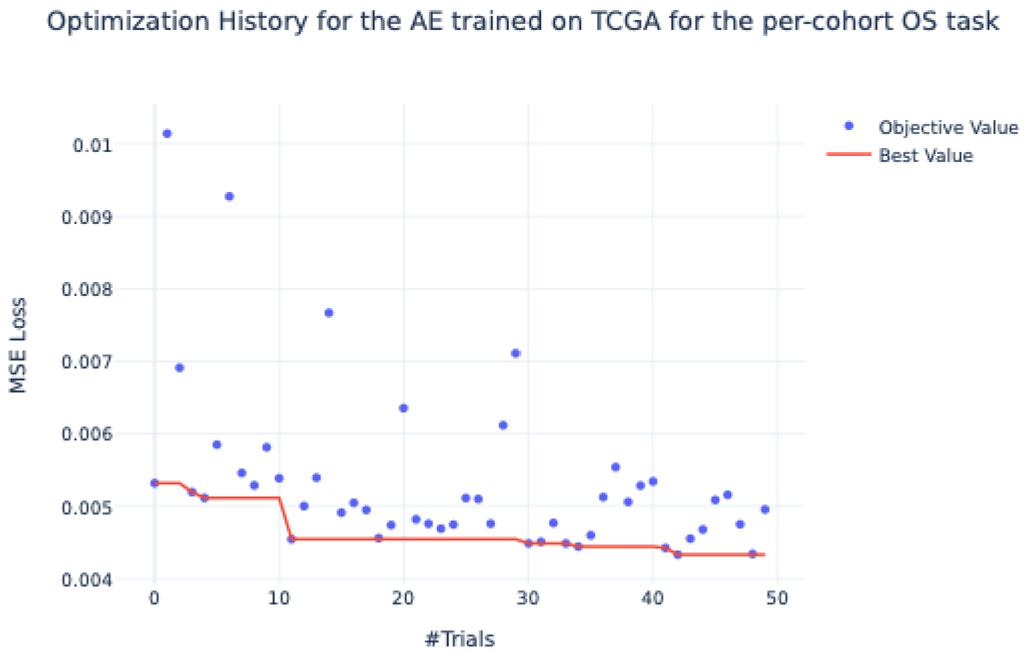
Supplementary Figure S6. Comparison of performance on gene essentiality prediction task on CCLE dataset for pretraining experiments. *Left panel) Number of folds for which the overall correlation for the model on the y axis is higher than for the model on the x axis. Red (Blue) boxes indicate 75% acceptance criterion on test folds for superiority is satisfied by the model on the y (x) axis.* *Right panel) Same as Top panel, but correlation computed per-gene.*



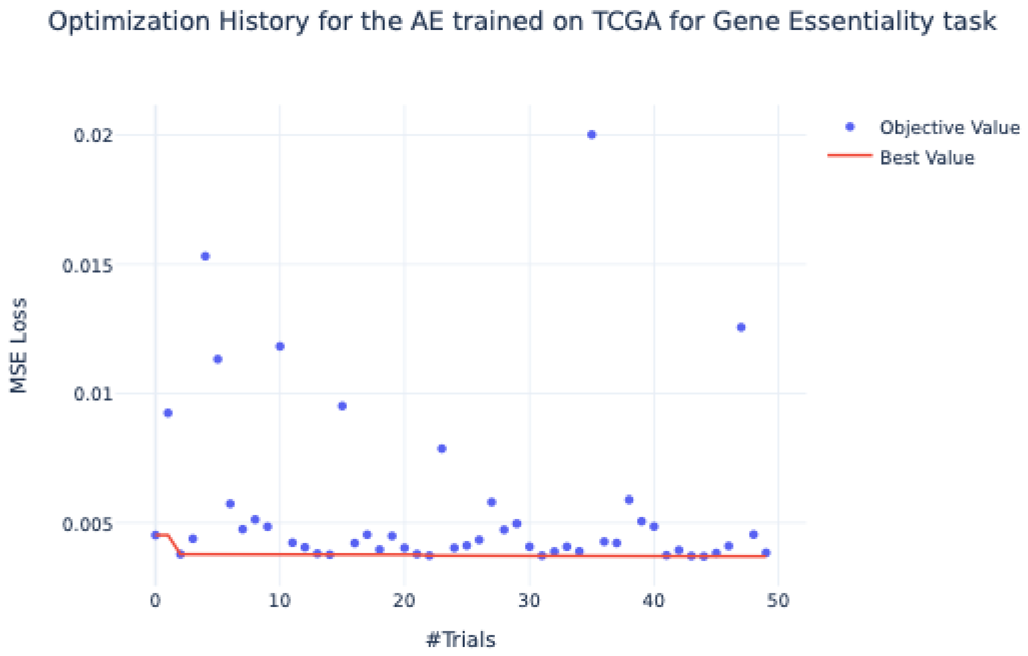
Supplementary Figure S7. Impact of pre-training an Exp-DeepDEP architecture on our data splits using the 5,000 most variable genes from TCGA as features. Skipping the pretraining step of Exp-DeepDEP seems to help reach better performances both in overall Spearman correlation and per-gene Spearman correlation, contrary to experiments performed on the multimodal DeepDEP model.



Supplementary Figure S8. Overall spearman correlation for different number of components for the PCA representation of gene fingerprints. Red crosses indicate the percentage of variance explained by PCA for the corresponding number of PCA components. For each tested number of PCA components, the overall spearman correlation is obtained over a 5-fold cross validation run with 12 different hyperparameter sets.



Supplementary Figure S9. Optimization History for the AE trained on 22 cohorts of TCGA excluding cohorts used in the downstream task. After 50 trials the final PreAE for the per-cohort OS prediction task has 255 latent dimensions, no hidden layers, a learning rate of 3.1e-4 and a dropout rate of 1.1e-2.



Supplementary Figure S10. Optimization History for the AE trained on the 33 cohorts of TCGA for the Gene Essentiality task. After 50 trials, the final PreAE for the gene essentiality task has 255 latent dimensions, no hidden layers, a learning rate of 1.6e-4 and a dropout rate of 0.13.

Results Tables

Supplementary Table S1 Test sets c-index statistics for the per-cohort OS prediction task.

Cohort	Representation Model	Mean Metric	Median Metric	Standard Deviation
BRCA	Identity	0,632	0.627	0.045
	PCA	0.619	0.628	0.058
	AE	0.606	0.611	0.037
	scVI	0.562	0.550	0.065
	MHAE	**0.635**	0.666	0.060
	MAE	0.632	0.631	0.044
	DA-GN	0.575	0.575	0.044
	GNN	0.595	0.607	0.068
KIRC	Identity	0.723	0.727	0.049
	PCA	**0.734**	0.733	0.040
	AE	0.714	0.718	0.047

	scVI	0.709	0.708	0.049
	MHAE	0.715	0.700	0.059
	MAE	0.717	0.712	0.049
	DA-GN	0.710	0.712	0.047
	GNN	0.698	0.686	0.039
UCEC	Identity	0.631	0.640	0.066
	PCA	**0.653**	0.641	0.072
	AE	0.631	0.633	0.052
	scVI	0.633	0.639	0.062
	MHAE	0.598	0.611	0.059
	MAE	0.633	0.638	0.071
	DA-GN	0.604	0.617	0.067
	GNN	0.632	0.639	0.091
HNSC	Identity	**0.589**	0.585	0.048
	PCA	0.584	0.575	0.032
	AE	0.573	0.571	0.034
	scVI	**0.589**	0.590	0.056
	MHAE	0.561	0.554	0.043
	MAE	0.575	0.571	0.039
	DA-GN	0.553	0.556	0.036
	GNN	0.558	0.563	0.033
LUAD	Identity	0.629	0.626	0.062
	PCA	0.618	0.627	0.075
	AE	0.641	0.653	0.065
	scVI	0.624	0.625	0.074
	MHAE	0.608	0.605	0.048
	MAE	0.621	0.630	0.070
	DA-GN	0.584	0.588	0.045
	GNN	**0.644**	0.660	0.070
LGG	Identity	0.796	0.785	0.034
	PCA	0.794	0.789	0.034
	AE	0.762	0.756	0.043
	scVI	0.766	0.779	0.047
	MHAE	**0.805**	0.803	0.048
	MAE	0.775	0.777	0.038
	DA-GN	0.776	0.771	0.045
	GNN	0.763	0.766	0.041

LUSC	Identity	0.500	0.508	0.044
	PCA	0.502	0.497	0.038
	AE	0.515	0.514	0.042
	scVI	0.507	0.501	0.043
	MHAE	0.474	0.484	0.039
	MAE	0.496	0.517	0.054
	DA-GN	0.493	0.480	0.049
	GNN	0.515	0.513	0.036
COAD	Identity	0.593	0.594	0.062
	PCA	0.578	0.586	0.067
	AE	**0.595**	0.614	0.080
	scVI	0.556	0.562	0.057
	MHAE	0.584	0.556	0.063
	MAE	0.583	0.580	0.050
	DA-GN	0.548	0.548	0.063
	GNN	0.534	0.536	0.047
BLCA	Identity	**0.640**	0.642	0.026
	PCA	0.619	0.625	0.039
	AE	0.633	0.633	0.038
	scVI	0.623	0.627	0.054
	MHAE	0.618	0.615	0.028
	MAE	0.615	0.614	0.039
	DA-GN	0.623	0.619	0.024
	GNN	0.639	0.645	0.029
SKCM	Identity	0.599	0.603	0.041
	PCA	0.609	0.613	0.068
	AE	0.623	0.623	0.045
	scVI	0.628	0.626	0.055
	MHAE	0.572	0.576	0.056
	MAE	0.611	0.604	0.045
	DA-GN	0.609	0.626	0.059
	GNN	0.621	0.615	0.053
STAD	Identity	0.564	0.539	0.078
	PCA	0.584	0.580	0.042
	AE	**0.614**	0.631	0.050
	scVI	0.530	0.522	0.043
	MHAE	0.580	0.573	0.069

	MAE	0.566	0.540	0.102
	DA-GN	0.534	0.533	0.065
	GNN	0.567	0.552	0.059

Best performance based on mean results are shown in boldface.

Supplementary Table S2 Test sets c-index statistics for the pan-cancer OS prediction task.

Representation Model	Mean Metric	Median Metric	Standard Deviation
Identity	0.748	0.748	0.010
PCA	**0.756**	0.756	0.009
AE	0.748	0.746	0.010
scVI	0.747	0.745	0.006
MHAE	0.755	0.756	0.007
MAE	**0.756**	0.755	0.010
DA-GN	0.748	0.747	0.008
GNN	0.749	0.748	0.009

Best performance based on mean results are shown in boldface.

Supplementary Table S3 Test sets overall spearman correlation statistics for the gene essentiality task.

Representation Model	Mean Metric	Median Metric	Standard Deviation
Identity	0.8510	0.8506	0.0028
PCA	0.8543	0.8538	0.0033
AE	0.8545	0.8539	0.0026
scVI	0.8548	0.8549	0.0027
MHAE	0.8548	0.8541	0.0026
MAE	0.8558	0.8553	0.0029
DA-GN	**0.8560**	0.8561	0.0028
GNN	0.8550	0.8538	0.0034

Best performance based on mean results are shown in boldface.

Supplementary Table S4 Test sets per-gene spearman correlation statistics for the gene essentiality task.

Representation Model	Mean Metric	Median Metric	Standard Deviation
Identity	0.237	0.236	0.006
PCA	0.246	0.247	0.006
AE	0.249	0.247	0.008
scVI	0.245	0.243	0.014

MHAE	0.249	0.248	0.003
MAE	0.255	0.255	0.006
DA-GN	**0.256**	0.256	0.006
GNN	0.250	0.249	0.009

Best performance based on mean results are shown in boldface.

Supplementary Table S5 Test sets c-index statistics for the per-cohort OS prediction task for pretrained experiments.

Cohort	Representation Model	Mean Metric	Median Metric	Standard Deviation
BRCA	AE	**0.606**	0.611	0.037
	PreAE	0.575	0.565	0.039
	PreAE finetuned	0.584	0.572	0.044
KIRC	AE	0.714	0.718	0.047
	PreAE	0.724	0.734	0.034
	PreAE finetuned	**0.728**	0.740	0.049
UCEC	AE	0.631	0.633	0.052
	PreAE	0.562	0.576	0.044
	PreAE finetuned	**0.654**	0.652	0.040
HNSC	AE	0.573	0.571	0.034
	PreAE	0.583	0.580	0.047
	PreAE finetuned	**0.587**	0.588	0.043
LUAD	AE	**0.641**	0.653	0.065
	PreAE	0.629	0.638	0.064
	PreAE finetuned	0.632	0.629	0.067
LGG	AE	**0.762**	0.756	0.043
	PreAE	0.718	0.709	0.046
	PreAE finetuned	0.752	0.746	0.047
LUSC	AE	0.515	0.514	0.042
	PreAE	**0.528**	0.534	0.031
	PreAE finetuned	0.524	0.534	0.035
COAD	AE	**0.595**	0.614	0.080
	PreAE	0.531	0.556	0.071
	PreAE	0.570	0.548	0.065

	finetuned			
BLCA	AE	**0.633**	0.633	0.038
	PreAE	0.609	0.625	0.041
	PreAE finetuned	0.622	0.623	0.036
SKCM	AE	0.623	0.623	0.045
	PreAE	**0.633**	0.628	0.038
	PreAE finetuned	0.626	0.610	0.044
STAD	AE	**0.614**	0.631	0.050
	PreAE	0.541	0.537	0.055
	PreAE finetuned	0.544	0.525	0.058

Best performance based on mean results are shown in boldface.

Table S6 Test sets overall spearman correlation statistics for the gene essentiality task for pretrained models.

Representation Model	Mean Metric	Median Metric	Standard Deviation
AE	**0.8554**	0.8558	0.0038
PreAE	0.8539	0.8541	0.0030
PreAE finetuned	0.8551	0.8552	0.0038
Exp-DeepDEP (Task-tuned)	0.8406	0.8410	0.0035
Exp-DeepDEP (Pretrained)	0.8335	0.8322	0.0063

Best performance based on mean results are shown in boldface.

Table S7 Test sets per-gene spearman correlation statistics for the gene essentiality task for pretrained models.

Representation Model	Mean Metric	Median Metric	Standard Deviation
AE	**0.249**	0.247	0.008
PreAE	0.237	0.237	0.007
PreAE finetuned	0.248	0.247	0.006
Exp-DeepDEP (Task-tuned)	0.179	0.180	0.018
Exp-DeepDEP (Pretrained)	0.067	0.042	0.108

Best performance based on mean results are shown in boldface.

Details on representation models implementation

We include here more details on the implementation of different representations models and meaning of certain hyperparameters names. Models not mentioned below are considered described thoroughly in the main text.

Auto-Encoders

In our implementation, Hidden Units First Layers correspond to the number of neurons in the first layer after the input in the AE-based architectures (AE, PreAE, scVI, MAE, MHAE, DA-GN). The Additional Hidden Layers correspond to the number of layers in the Encoder / Decoder excluding the representation layer. The Hidden Decrease Rate controls the bottleneck of the Encoder : a value of 0.5 means that at each additional hidden layer, the number of neurons is divided by 2. The batch size was fixed and not used as a hyperparameter following advice from recent work[1].

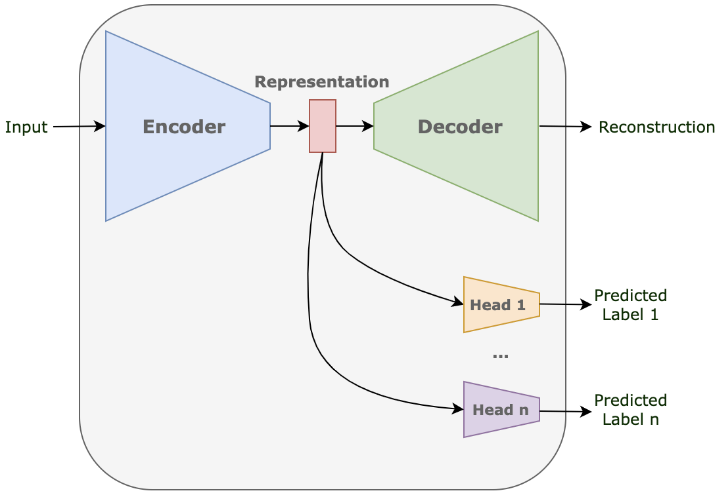
Masking Auto-Encoders

In VIME, the authors introduce an innovative masking scheme (compared to Gaussian noise addition or binary masking), in which they :

1. Generate a permuted variant of the samples
2. Generate a binary mask
3. Compose the binary mask and the permutation to generate a corrupted sample $\hat{x} = m * x_{perm} + (1 - m) * x$, where x is the original samples, m a binary mask sampled from a Bernouilli distribution and x_{perm} the permuted sample.

Multi-Head Auto-Encoders

The MH auto-encoder simplified architecture is depicted below:



This model was trained using a two-term loss function:

$$L = L_{rec} + \beta * L_{aux}$$

where β is the hyperparameter controlling the weight between the two terms, L_{rec} is the auto-encoder reconstruction error (mean squared error), and L_{aux} is the auxiliary head loss function. The latter depends on the predicted endpoint: mean squared error for gene essentiality, and cox loss for overall survival.

[1] https://github.com/google-research/tuning_playbook

Graph Neural Networks

The STRING data was preprocessed by keeping only genes present in our omics data and forming the induced subgraph based on this gene list. Additionally, we retained only the most confident interactions, using the 'combined score' column, by setting a quantile parameter q. The quantile value ranged from 0.7 (as suggested by the STRING db) to 0.99. Nodes not belonging to the largest connected component were discarded to ensure downstream clustering did not include isolated single nodes.

Clustering was then performed on the graph, which would be used in the pooling part of the encoder. Our goal was to define tightly connected gene communities within the graph. The Louvain algorithm implemented in networkX was employed to detect these communities, with the resolution parameter controlling the granularity of the clusters. These clusters, presented as gene lists, were given as input to our GNN encoder, along with the actual graph and the RNA-seq data.

Our GNN model was created using the Pytorch Geometric library. It was built as an auto-encoder, comprising a GNN encoder and a classical MLP decoder. The model aimed to reconstruct the omics signal given the omics signal and the graph as input. The encoder consisted of stacked classical GNN layers (SAGEconv, GraphConv), typically ranging from 1 to 3 layers. The number of channels (i.e., size of the node embedding) increased at each layer, with different values tested, such as [8,8,16], to obtain a deep node embedding of dimension d. To reduce the dimensionality of the embeddings, nodes were grouped per cluster defined in the previous paragraph. Cluster-level representations were generated by applying pooling layers (average pooling or max pooling) to genes within the same cluster. This resulted in $n_{cluster}$ d-dimensional embeddings, which were concatenated into a single $n_{cluster} * d$ dimensional embedding. To ensure comparability with other models and determine the embedding dimension, this embedding was passed through a single MLP layer.

The decoder, derived from our Auto-Encoder architecture, was concatenated with the encoder. The entire model was trained end-to-end to reconstruct the bulk RNA-seq signal when provided with the expression data and graph inputs.

Exp-DeepDEP Experiments

In our setting for the Gene Essentiality task, we focused mostly on the cell lines RNA-seq representations but did not fine-tune the fingerprints representation nor created an end-to-end deep learning model specifically for this task such as DeepDEP. DeepDEP is a multimodal model that takes into account not only bulkRNA-seq data but also mutations, copy number alterations, methylation data and fingerprints, integrating them through the combination of different encoder heads in the architecture of the model. They show that, with pretraining on TCGA, this model improves performance in their evaluation framework compared to no pretraining. They also show that a simplified version of their model, Exp-DeepDEP, obtained similar performances as DeepDEP using only RNA-seq and fingerprints. With the pre-trained auto-encoders presented above, we investigated whether a pre-trained representation model on TCGA could provide better embeddings for the expression profiles of the downstream task dataset when concatenated with fixed fingerprints representations and passed through the prediction model (LGBM) directly. Similarly to their work, we selected the top 5,000 variable genes on TCGA and trained an end-to-end Exp-DeepDEP on our data splits, with and without the pretraining step, to investigate the influence of pre-training in the same setting as the original paper when focused only on expression data. The pretraining step was performed similarly as the original Exp-DeepDEP and the same hyperparameters. After pretraining, Ex-DeepDEP models were trained on CCLE following our data splits using again the same HPs as the original paper. To do so, we downloaded the original code from DeepDEP on CodeOcean[2] , modified it and made it available in our GitHub repository.

Details on representation models hyperparameters

Supplementary Table S8 Hyperparameters Range for Representation Models.

Representation Model	Hyperparameter	Ranges	
		OS Tasks	GE Task
PCA	Representation Dimension	⟦4, 256⟧	⟦16, 256⟧
AE	Representation Dimension	⟦4, 256⟧	⟦16, 256⟧
	Hidden Units First Layer	⟦256, 1024⟧	
	Additional Hidden Layers	⟦0, 1⟧	⟦0, 2⟧
	Hidden Decrease Rate	{0.5; 1}	
	Learning Rate	[5e-5, 5e-3]	[5e-6, 5e-4]
	Batch Size	256	1024
	Dropout Rate	[0, 0.2]	
	Maximum Number of Epochs	300	
	Early Stopping Patience	50	
	Early Stopping Delta	0.001	
scVI	Representation Dimension	⟦4, 256⟧	⟦16, 256⟧
	Hidden Units First Layer	⟦256, 1024⟧	
	Additional Hidden Layers	⟦0, 1⟧	⟦0, 2⟧
	Hidden Decrease Rate	{0.5; 1}	
	Learning Rate	0.001	
	Batch Size	256	1024
	Dropout Rate	0.1	
	Maximum Number of Epochs	300	
MAE	Representation Dimension	256	
	Hidden Units First Layer	512	

[2] https://codeocean.com/capsule/7914207/tree/

	Additional Hidden Layers	0	
	Learning Rate	[5e-5, 5e-3]	[5e-5, 5e-4]
	Batch Size	256	1024
	Dropout Rate	0	
	Corruption Probability	[0.1, 0.5]	
	Maximum Number of Epochs	1000	
	Early Stopping Patience	20	
	Early Stopping Delta	0.00001	
MHAE	Representation Dimension	256	
	Hidden Units First Layer	512	
	Additional Hidden Layers	0	
	Learning Rate	[5e-5, 5e-3]	
	Batch Size	256	1024
	Dropout Rate	{0; 0.1}	
	Maximum Number of Epochs	1000	
	Early Stopping Patience	50	
	Early Stopping Delta	0.00001	
	Auxiliary Head Neurons	Per-cohort : {32; 128} Pancancer : 128	32
	Auxiliary Head Dropout Rate	{0; 0.1}	0
	Loss Weight for Aux. Head Beta	[0.001, 1]	[0.1, 10]
GNN	Representation Dimension	128	
	Learning Rate	0.001	
	Batch Size	32	
	Maximum Number of Epochs	50	
	Message Passing	{GraphConv; SAGE}	
	Channels	{[8]; [8, 16]; [8, 16, 16]}	
	Louvain Cluster Resolution	[10, 500]	
	Pooling	{avg, max}	
	StringDB Threshold	[0.7, 0.99]	
	Graph Encoder Dropout	0.3	
	Decoder Hidden Layers	{1, 2}	
	Decoder Hidden Decrease Rate	0.5	
	Decoder Dropout Rate	0.2	

DA-GN	Representation Dimension	256	
	Hidden Units First Layer	512	
	Additional Hidden Layers	0	
	Learning Rate	[5e-5, 5e-3]	[5e-5, 5e-4]
	Batch Size	256	1024
	Dropout Rate	0.1	
	Maximum Number of Epochs	1000	
	Early Stopping Patience	20	
	Early Stopping Delta	0.00001	
	Noise Level	[0.01, 1]	
	Number of copies	4	
PreAE	Representation Dimension	⟦16, 256⟧	
	Hidden Units First Layer	⟦256, 1024⟧	
	Additional Hidden Layers	⟦0, 2⟧	
	Hidden Decrease Rate	{0.5; 1}	
	Learning Rate	[5e-6, 5e-4]	
	Batch Size	1024	
	Dropout Rate	[0, 0.2]	
	Maximum Number of Epochs	1000	
	Early Stopping Patience	20	
	Early Stopping Delta	0.00001	

Brackets represent sets of values, single [ ] represent float intervals and double [[ ]] represent integer ranges.

Details on prediction models hyperparameters

Supplementary Table S9 Hyperparameters Range for Prediction Models.

Prediction Model	Hyperparameter	Ranges
MLP (OS Tasks)	Hidden Layers	[128]
	Learning Rate	[1e-5, 1e-3]
	Batch Size	256
	Dropout Rate	0
	Maximum Number of Epochs	1000
	Early Stopping Patience	50
	Early Stopping Delta	0.00001
LGBM (Gene Essentiality)	Learning Rate	[0.01, 0.3]
	L1 Regularization	[0, 100]

Brackets represent sets of values, single [ ] represent float intervals and double [[ ]] represent integer ranges

Supplementary Table S10. Grid-search results on a single test set using Identity and an MLP for the prediction model in the Gene Essentiality task for overall correlation. The MLP proved harder to train correctly compared to LGBM, reaching at most 0.81 when Identity with LGBM reached 0.85 of correlation (cf main figures).

Test Set Metric	Learning Rate	MLP Architecture
0.657	1e-06	[]
0.666	1e-05	[]
0.665	0.0001	[]
0.662	0.001	[]
0.683	1e-06	[512]
0.814	1e-05	[512]
0.804	0.0001	[512]
0.808	0.001	[512]
0.683	1e-06	[1024, 256]
0.780	1e-05	[1024, 256]
0.784	0.0001	[1024, 256]
0.804	0.001	[1024, 256]
0.727	1e-06	[1024, 1024, 1024, 512, 256]
0.753	1e-05	[1024, 1024, 1024, 512, 256]
0.768	0.0001	[1024, 1024, 1024, 512, 256]
0.798	0.001	[1024, 1024, 1024, 512, 256]