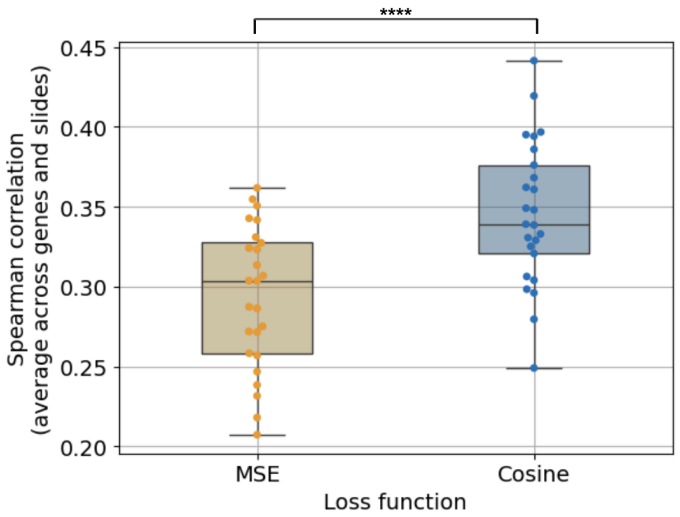

Fig. S1: Cross-validation performances (Spearman correlation per split, averaged over genes and slides) for the baseline MLP architecture, trained with Mean Squared Error (MSE) and loss based on cosine similarity, on n = 25 random train/test splits (box: interquartile range (IQR); horizontal line: median; whiskers: 1.5 times IQR); p-value = 1.1×10^{-7}. P-values per split were obtained from a one-sided t-test, as described in the method section, and aggregation over repeated train/test splits was done by computing the median p-value and multiplying by 2. * : p < 0.05 ; ** : p < 0.01. *** : p < 0.001 ; **** : p < 0.0001. Source data are provided as a Source Data file.

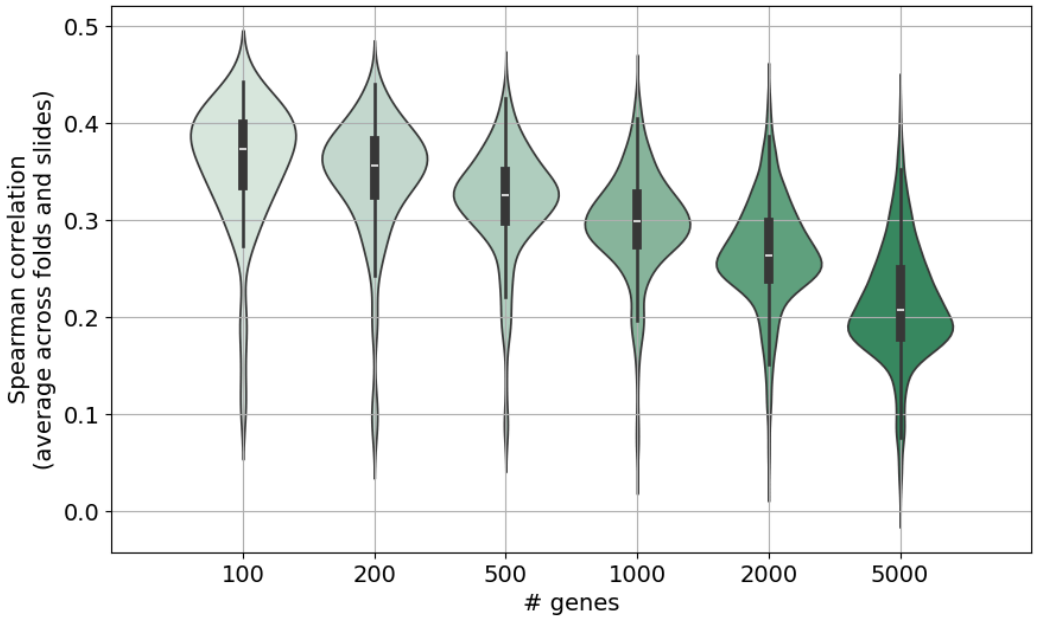

Fig. S2: Distribution of the Spearman correlation per gene (averaged across slides and folds) when varying the number of genes predicted by MISO (box: IQR, white dot: median, whiskers: 1.5 times IQR).

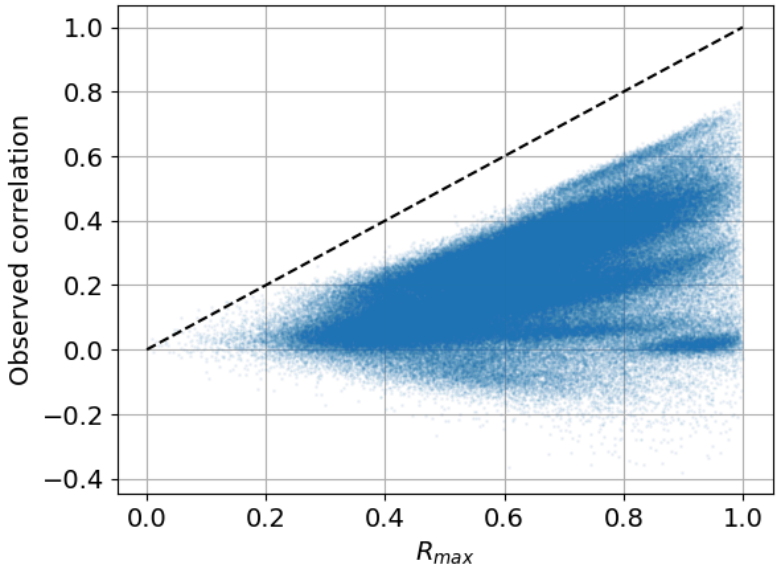

Fig. S3: Performances (Pearson correlation) achieved by the 5,000-gene model for every of the n = 240,000 pairs gene/sample, as a function of the theoretical upper bound R_{max}. Source data are provided as a Source Data file.

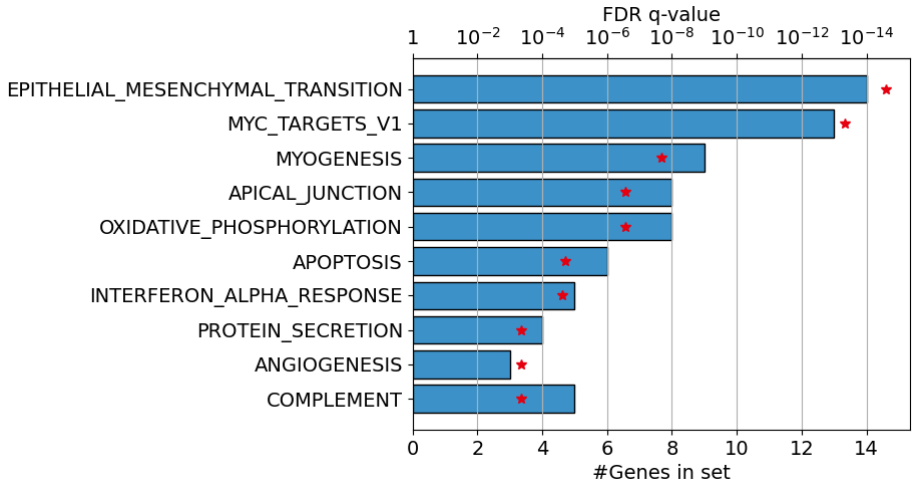

Fig. S4: Gene set enrichment analysis of the 100 genes with the largest average R_{max}. Bars indicate the number of genes among the 100 that belong to each pathway. Red stars indicate the FDR q-value. P-values were obtained from the hypergeometric distribution for (k - 1, K, N - K, 100) where k is the number of genes in the intersection of the 100 genes with a given target set, K is the number of genes in the target set and N is the total number of known human gene symbols. The FDR q-value was obtained after Benjamini-Hochberg correction for multiple hypothesis testing. Source data are provided as a Source Data file.

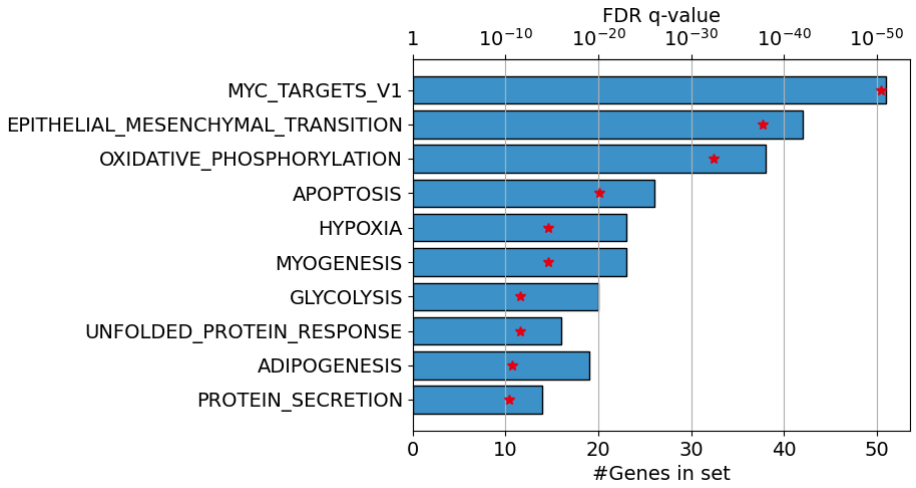

Fig. S5: Gene set enrichment analysis of the 500 genes with the largest average R_{max}. Bars indicate the number of genes among the 500 that belong to each pathway. Red stars indicate the FDR q-value. P-values were obtained from the hypergeometric distribution for (k - 1, K, N - K, 500) where k is the number of genes in the intersection of the 500 genes with a given target set, K is the number of genes in the target set and N is the total number of known human gene symbols. The FDR q-value was obtained after Benjamini-Hochberg correction for multiple hypothesis testing. Source data are provided as a Source Data file.

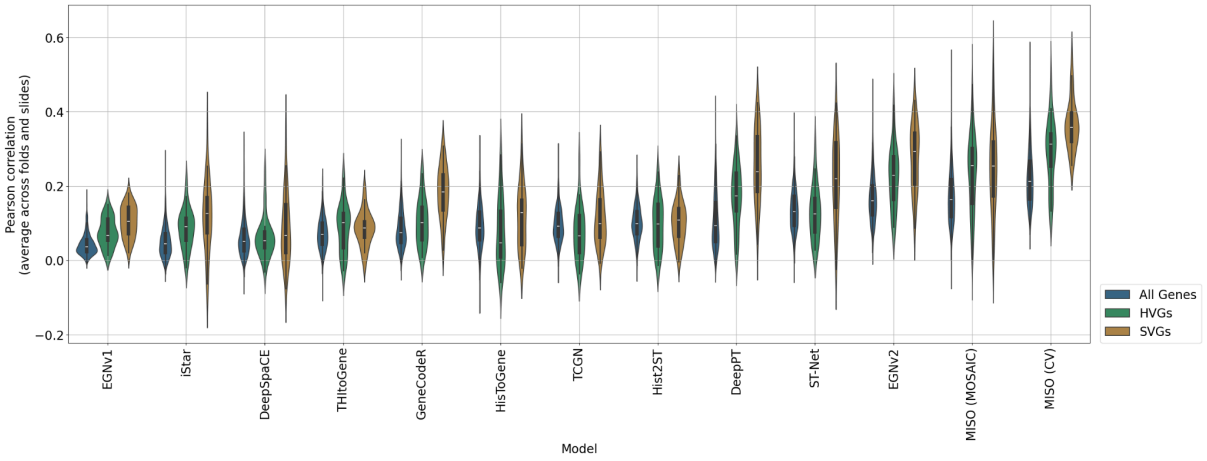

Fig. S6: Distribution of the Pearson correlation per gene (averaged across slides and folds) on the dataset HER2ST. We denote by MISO (MOSAIC) the model trained on MOSAIC data. All other architectures were trained in cross-validation on HER2ST (box: IQR, white dot: median, whiskers: 1.5 times IQR). Source data are provided as a Source Data file.

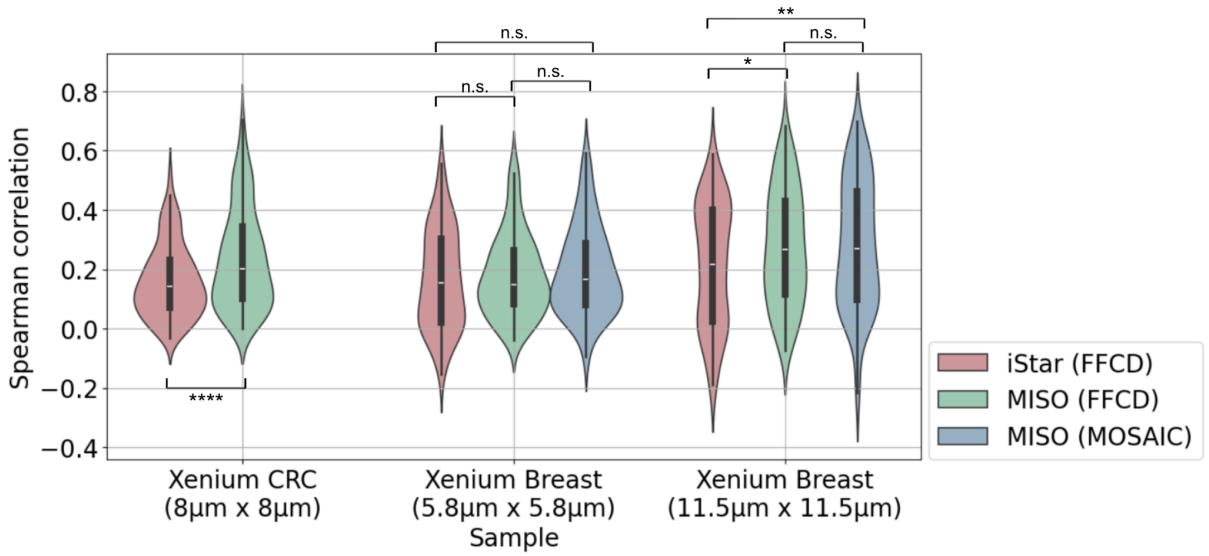

Fig. S7: Distribution of the Spearman correlation per gene achieved by super-resolution approaches on Xenium samples (box: IQR, white dot: median, whiskers: 1.5 times IQR). At a resolution of 0.5 MPP, iStar works with patches of size 8µm x 8µm and MISO with patches of size 7µm x 7µm. For comparison, we used linear interpolation to map MISO prediction to patches of size 8µm x 8µm. The breast cancer sample was scanned at a resolution of 0.36 MPP, leading to patches of size 5µm x 5µm for MISO and 5.8µm x 5.8µm for iStar. We also compared performances with downsampling by a factor of 2; p(MISO vs iStar, CRC) = 4.3 x 10^{-9}, p(MISO-FFCD vs iStar, Breast 5.8 x 5.8) = 0.26, p(MISO-MOSAIC vs MISO-FFCD, Breast 5.8 x 5.8) = 0.21, p(MISO-MOSAIC vs iStar, Breast 5.8 x 5.8) = 0.09, p(MISO-FFCD vs iStar, Breast 11.5 x 11.5) = 0.013, p(MISO-MOSAIC vs MISO-FFCD, Breast 11.5 x 11.5) = 0.37, p(MISO-MOSAIC vs iStar, Breast 11.5 x 11.5) = 0.008. P-values were computed by running a one-sided t-test on the z-transformed distribution of Spearman coefficients. * : $p < 0.05$; ** : $p < 0.01$. *** : $p < 0.001$; **** : $p < 0.0001$. Source data are provided as a Source Data file.

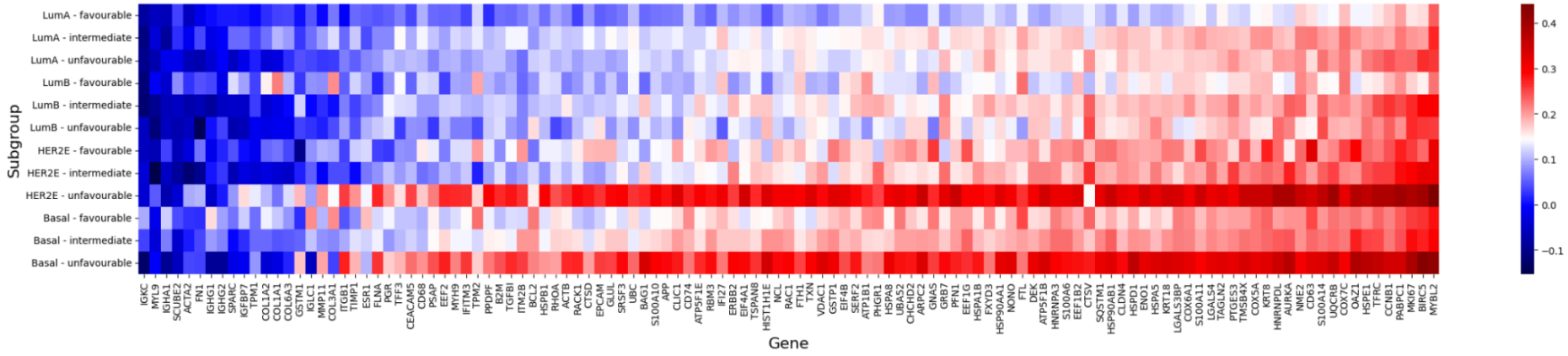

Fig. S8: Heatmap of the DR scores of all considered genes in breast cancer samples from TCGA-BRCA, averaged per molecular subtype and per survival group. Source data are provided as a Source Data file.

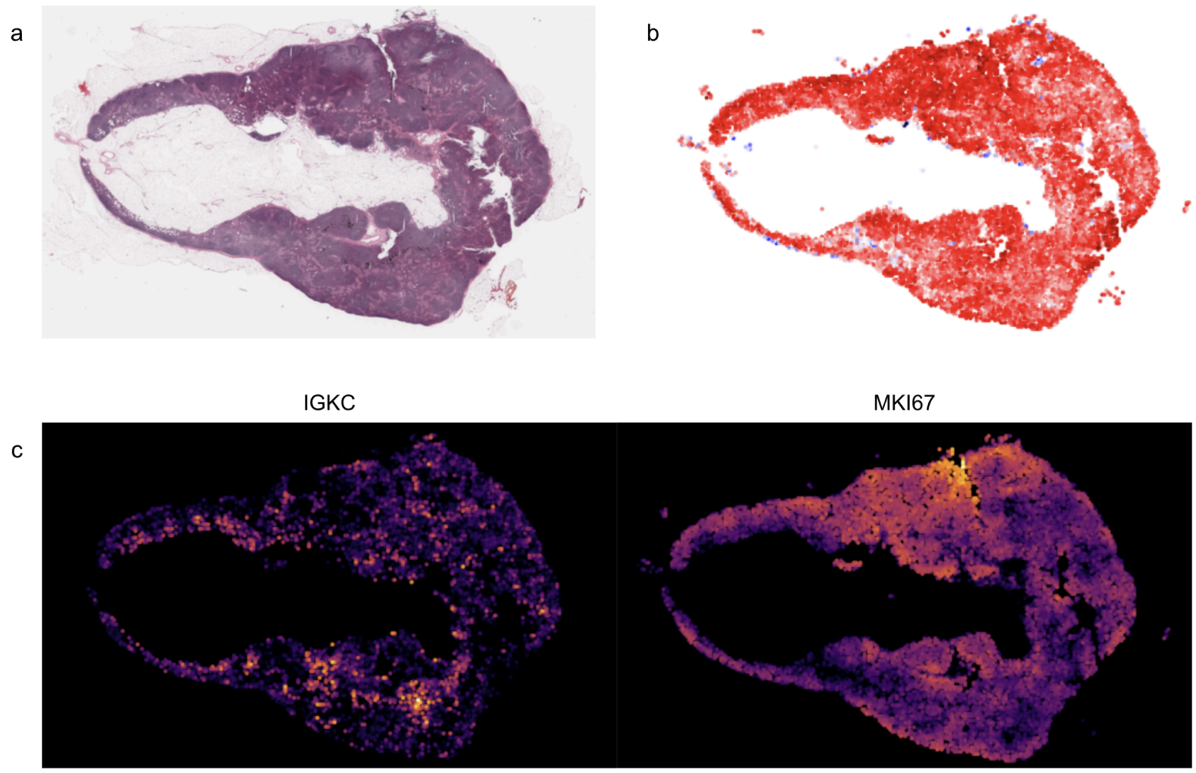

Fig. S9: a. Example slide from a TCGA patient with Basal breast cancer and a favourable predicted prognosis. b. Local survival scores output by the model trained to predict overall survival. Shades of blue indicate areas associated with a lower risk by the model, while red indicates areas associated with a high risk. c. Predicted expression of *IGKC* and *MKI67* on the same slide.

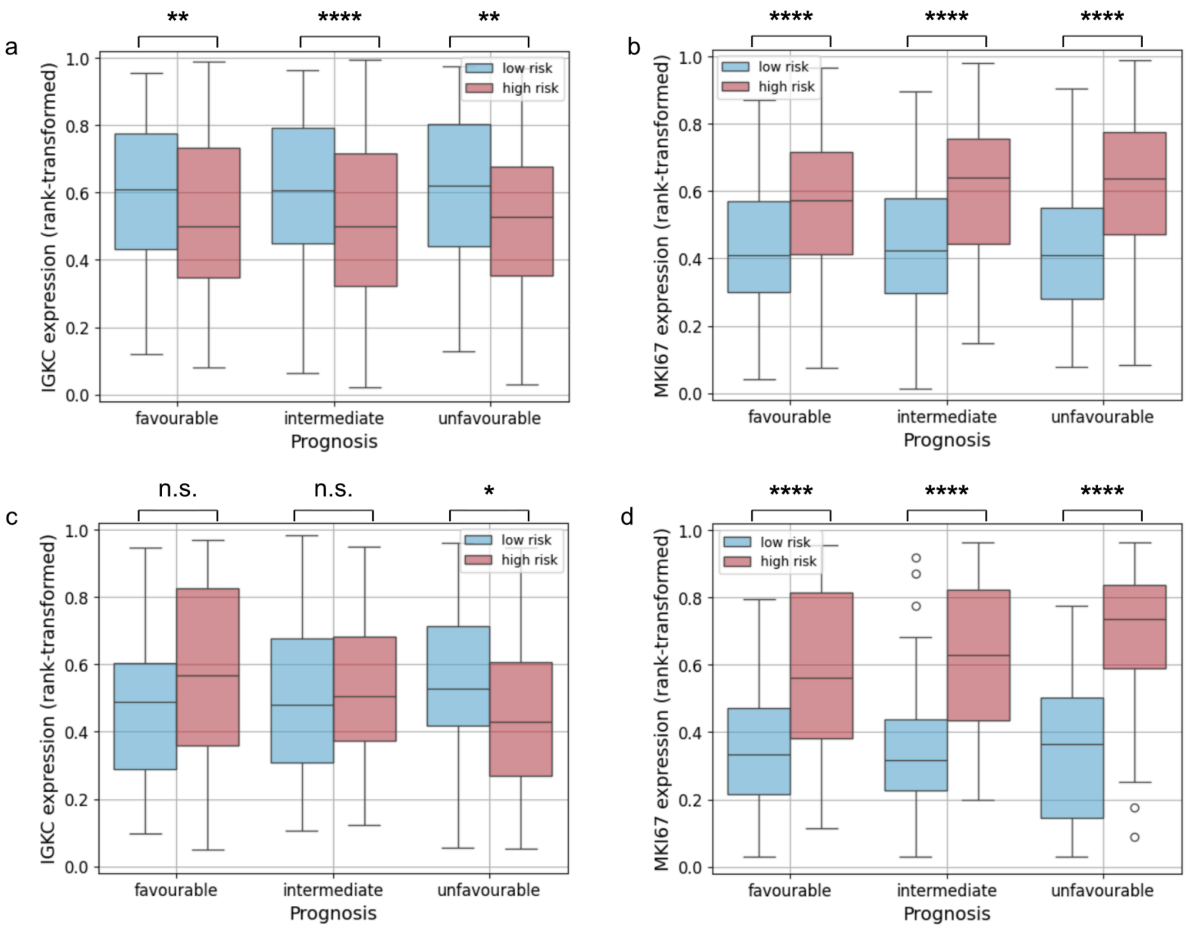

Figure S10: a. Rank-transformed predicted expression of IGKC in low- and high-risk tiles from Luminal A breast cancer patients with favourable (p = 0.007), intermediate (p = 3.3 x 10^{-5}) and unfavourable (p = 0.001) prognosis; n(favourable) = 143, n(intermediate) = 284, n(unfavourable) = 143, (box: interquartile range (IQR); horizontal line: median; whiskers: 1.5 times IQR). b. Same, with MKI67 predicted expression; p(favourable) = 2.7 x 10^{-6}, p(intermediate) = 2.2 x 10^{-18}, p(unfavourable) = 2.6 x 10^{-10}. c. Same as a. in patients with Basal breast cancer; p(favourable) = 0.08, p(intermediate) = 0.16, p(unfavourable) = 0.03. d. Same as b. in patients with Basal breast cancer; n(favourable) = 46, n(intermediate) = 92, n(unfavourable) = 47; p(favourable) = 2.5 x 10^{-5}, p(intermediate) = 1.0 x 10^{-10}, p(unfavourable) = 1.4 x 10^{-7}. P-values were computed with a Wilcoxon signed-rank test (one-tailed). n.s.: non significant, *: p < 0.05, **: p < 0.01, ***: p < 0.001, ****: p < 0.0001. Source data are provided as a Source Data file.

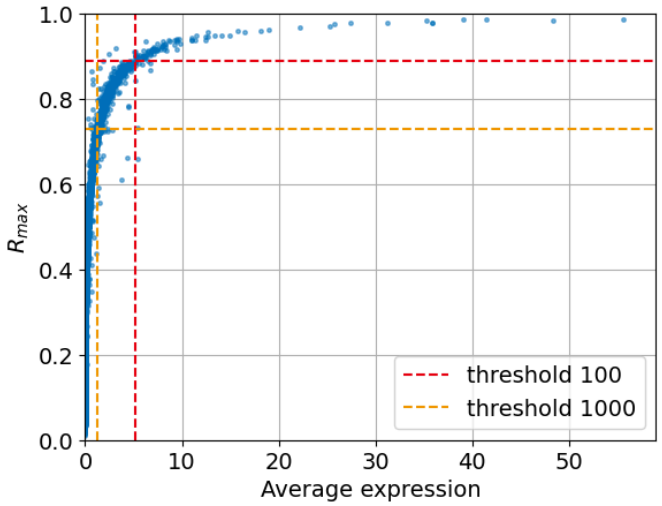

Fig. S11: Relation between gene expression (raw counts, averaged across spots and samples) and the theoretical R_{max} (averaged across samples) derived in the present work, for n = 14,470 genes. Vertical (resp. horizontal) red dashed lines indicate the 100th largest average expression (resp. 100th largest average R_{max}) in the 48 samples of the training set. Orange lines: same, for the 1000th largest values. Source data are provided as a Source Data file.

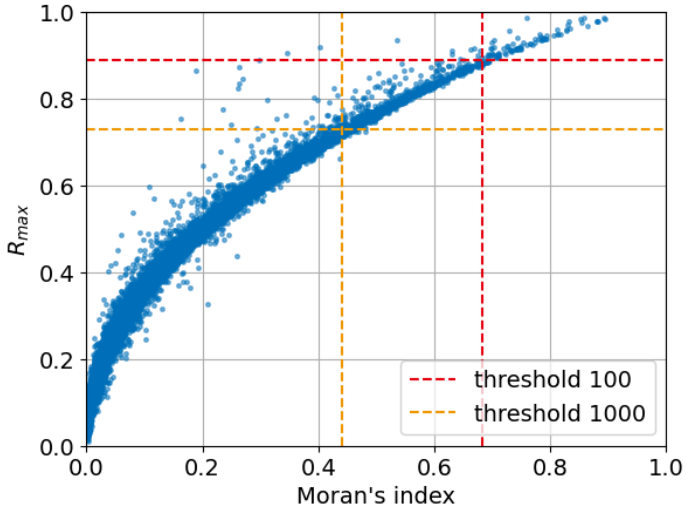

Fig. S12: Relation between Moran's index, that measures spatial autocorrelation, with R_{max}. . Vertical (resp. horizontal) red dashed lines indicate the 100th largest Moran's index (resp. 100th largest average R_{max}) in the 48 samples of the training set, for n = 14,470 genes. Orange lines: same, for the 1000th largest values. Source data are provided as a Source Data file.

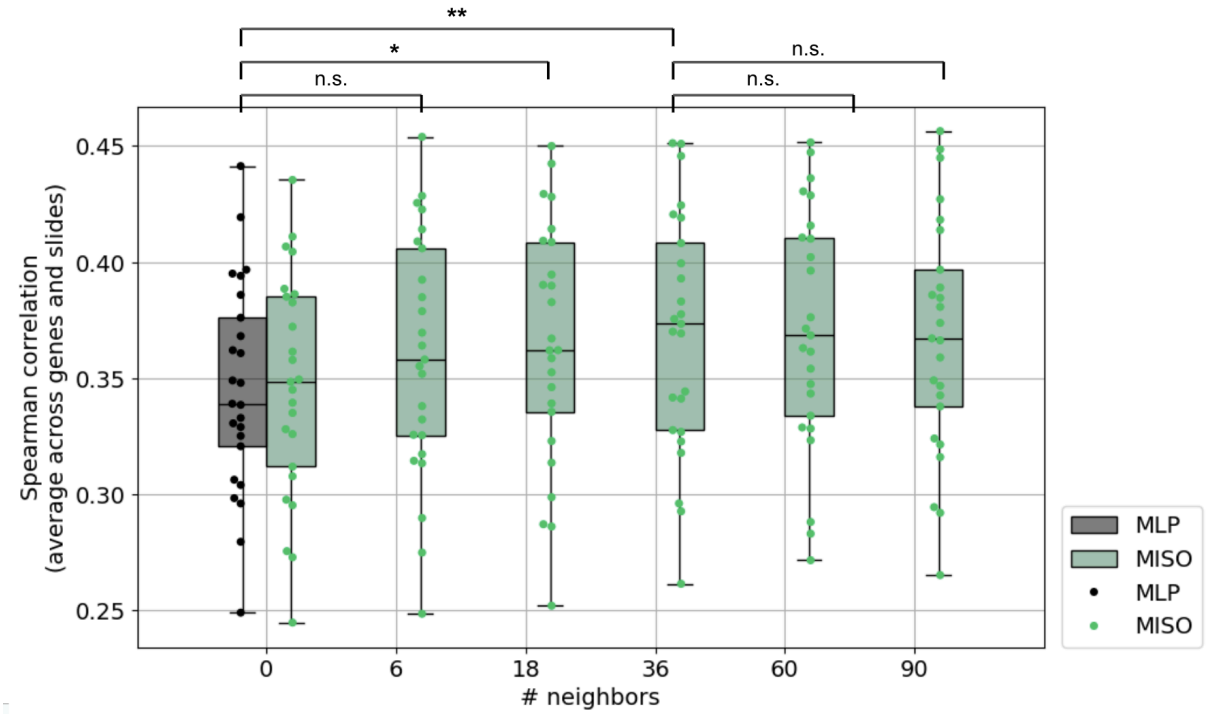

Fig. S13: Cross-validation performances (Spearman correlation per fold, averaged over genes and slides) for the baseline architecture (MLP) and MISO with varying number of neighbors, on n = 25 random train/test splits; p(MISO-6n vs MLP) = 0.20, p(MISO-18n vs MLP) = 0.03, p(MISO-36n vs MLP) = 0.008, p(MISO-60n vs MISO-36n) = 0.25, p(MISO-90n vs MISO-36n) = 0.11. (box: interquartile range (IQR); horizontal line: median; whiskers: 1.5 times IQR). P-values per split were obtained from a one-sided t-test, as described in the method section, and aggregation over repeated train/test splits was done by computing the median p-value and multiplying by 2. * : $p < 0.05$; ** : $p < 0.01$. *** : $p < 0.001$; **** : $p < 0.0001$. Source data are provided as a Source Data file.

# genes	CV (N = 48)		Test (N = 24)	
	Pearson	**Spearman**	**Pearson**	**Spearman**
100	0.351 (0.044)	0.369 (0.051)	0.447 [0.373 - 0.510]	0.457 [0.372 - 0.532]
200	0.339 (0.044)	0.355 (0.051)	0.431 [0.358 - 0.496]	0.437 [0.356 - 0.512]
500	0.318 (0.045)	0.328 (0.052)	0.405 [0.333 - 0.471]	0.410 [0.327 - 0.485]
1,000	0.298 (0.043)	0.305 (0.049)	0.378 [0.311 - 0.442]	0.379 [0.303 - 0.451]
2,000	0.271 (0.038)	0.275 (0.045)	0.341 [0.273 - 0.403]	0.343 [0.270 - 0.411]
5,000	0.222 (0.033)	0.221 (0.039)	0.281 [0.222 - 0.338]	0.279 [0.216 - 0.340]

Table S1: Performances of MISO when predicting a varying number of genes (genes with the highest R_{max} values), in cross-validation or on the test set of PETACC8-Visium. Correlations are averaged across gene and slide. For cross-validation, we report the average and standard deviation across folds. On the test set, we report the 95% confidence interval obtained by bootstrapping 10,000 times with replacement slides from the dataset.

Center	Cancer	N	Pearson	Spearman
CHUV	Bladder	63	0.353 (0.040)	0.373 (0.060)
CHUV	Breast	8	0.280 (0.113)	0.307 (0.114)
CHUV	Ovarian	11	0.393 (0.082)	0.434 (0.091)
CHUV	Mesothelioma	18	0.372 (0.041)	0.390 (0.043)
CHUV	NSCLC	17	0.271 (0.066)	0.298 (0.063)
Charité	Bladder	24	0.448 (0.122)	0.473 (0.123)
Charité	Breast	3	0.004 (0.117)	0.065 (0.082)
Charité	Ovarian	42	0.406 (0.035)	0.430 (0.040)
Charité	Mesothelioma	15	0.403 (0.096)	0.432 (0.098)
Charité	NSCLC	10	0.345 (0.136)	0.401 (0.164)
Erlangen	Bladder	33	0.368 (0.087)	0.354 (0.084)
Erlangen	Breast	2	0.391 (0.048)	0.462 (0.064)
Erlangen	Ovarian	14	0.297 (0.152)	0.321 (0.173)
Erlangen	Mesothelioma	14	0.277 (0.098)	0.283 (0.104)
Erlangen	NSCLC	11	0.424 (0.064)	0.472 (0.120)
Pittsburgh	NSCLC	24	0.371 (0.067)	0.401 (0.066)
GR	Breast	2	0.423 (0.167)	0.468 (0.188)
GR	Ovarian	37	0.301 (0.038)	0.317 (0.036)
All		*348*	*0.356 (0.014)*	*0.378 (0.022)*

Table S2: Cross-validation performances in MOSAIC, to predict the expression of 94 genes taken from the list of 100 genes with highest R_{max} in PETACC8-Visium. For each center and cancer type, we report the mean correlation per gene and per slide, averaged across folds, as well as the standard deviation across folds.

Sample	Training data	Patch_size	gene	Spearman	Pearson
CRC	PETACC8-Visium	7µm x 7µm	GPX2	0.685	0.589
			EPCAM	0.644	0.562
			ERBB3	0.632	0.514
			RNF43	0.602	0.504
			CDX2	0.568	0.460
			GPRC5A	0.541	0.428
			MET	0.524	0.419
			SOX9	0.514	0.426
			ERBB2	0.505	0.425
Breast	PETACC8-Visium	5µm x 5µm	EPCAM	0.518	0.466
			KRT8	0.511	0.470
			SCD	0.493	0.447
			ERBB2	0.472	0.502
			FASN	0.456	0.431
			MLPH	0.451	0.434
			CD9	0.430	0.411
Breast	MOSAIC	5µm x 5µm	EPCAM	0.562	0.555
			SCD	0.531	0.538
			KRT8	0.529	0.508
			ERBB2	0.487	0.553
			FASN	0.483	0.481
			MLPH	0.468	0.459
			CD9	0.457	0.464
			CDH1	0.415	0.414
Breast	PETACC8-Visium	10µm x 10µm	KRT8	0.653	0.664
			EPCAM	0.591	0.600
			FASN	0.586	0.603
			CD9	0.577	0.592
			MLPH	0.576	0.574
			SCD	0.566	0.563
			CDH1	0.541	0.544
			ERBB2	0.534	0.582
			DSP	0.533	0.523
			MYO5B	0.525	0.520
			S100A14	0.515	0.523

Breast	MOSAIC	10μm x 10μm	KRT8	0.670	0.701
			EPCAM	0.620	0.679
			FASN	0.620	0.666
			CD9	0.600	0.648
			SCD	0.599	0.650
			MLPH	0.593	0.630
			MYO5B	0.565	0.595
			NARS	0.564	0.557
			TPD52	0.557	0.547
			CDH1	0.553	0.588
			DSP	0.552	0.569
			ERBB2	0.531	0.626
			CCDC6	0.521	0.519
			CTTN	0.518	0.528
			S100A14	0.517	0.544

Table S3: Best correlations obtained with the super-resolution approach, depending on the sample, training data and resolution (patch size). For the colorectal cancer sample and the breast cancer sample with patches of size 10μm x 10μm, we display genes with a Spearman correlation above 0.5. For the breast cancer samples with patches of size 5μm x 5μm, we report genes with a Spearman correlation above 0.4