

Generating crossmodal gene expression from cancer histopathology improves multimodal AI predictions

SUPPLEMENTARY MATERIAL

Supplementary Table 1: Details of data splits across datasets (TCGA, CPTAC) and cancer cohorts used in the experiments. The cohorts include glioblastoma-glioma (TCGA-GBMLGG, CPTAC-GBM), renal (TCGA-KIRC), uterine (TCGA-UCEC, CPTAC-UCEC) and breast (TCGA-BRCA). Data from the CPTAC cohorts were excluded from training. Models trained on the TCGA cohorts were evaluated on the corresponding CPTAC datasets.

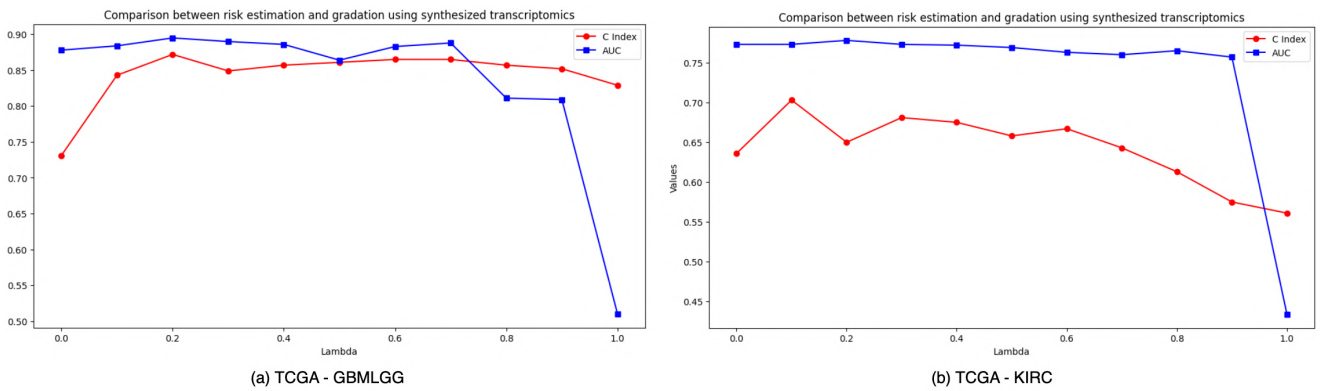
Datasets	Cases				Slides			
	Training	Validation	Calibration	Testing	Training	Validation	Calibration	Testing
TCGA-GBMLGG	532	56	75	82	653	69	86	104
TCGA-KIRC	337	34	45	46	344	42	48	51
TCGA-UCEC	191	19	30	27	210	23	31	30
TCGA-BRCA	676	78	97	95	729	80	102	99
CPTAC-GBM	-	-	32	30	-	-	151	91
CPTAC-UCEC	-	-	38	33	-	-	194	170

Supplementary Table 2: Table for number of genes in each gene group in the transcriptomic data. Same number of genes have been considered in each gene groups for same cohorts from different datasets. We note that there is repetition of genes across gene groups. The last row shows the additive result of the number of genes across gene groups including repetitions, and not the total number of genes considered for prediction.

Gene type	TCGA-GBMLGG	TCGA-KIRC	TCGA-UCEC	TCGA-BRCA	CPTAC-GBM	CPTAC-UCEC
Tumor Suppressor Genes	84	145	152	227	84	152
Oncogenes	314	522	554	959	314	554
Protein Kinases	498	669	700	934	498	700
Cell Differentiation Markers	424	681	693	843	424	693
Transcription Factors	1396	2366	2489	3712	1396	2489
Cytokines and Growth Factors	428	990	965	1142	428	965
All gene groups	3144	5373	5553	7817	3144	5553

Supplementary Table 3: Demographic details of the data used for calibration and testing for all the datasets and cohorts. Ground truth grades are absent for TCGA-BRCA and CPTAC-GBM datasets. Also, patient age is missing in the clinical data for the CPTAC cohorts. ncal denotes the number of samples in the calibration set and ntest denotes the number of samples in the test set.

Category	Group	TCGA - GBMLGG		TCGA - KIRC		TCGA - UCEC		TCGA - BRCA		CPTAC - GBM		CPTAC - UCEC	
		ncal	ntest	ncal	ntest	ncal	ntest	ncal	ntest	ncal	ntest	ncal	ntest
overall	-	86	104	144	153	93	90	306	297	453	273	582	510
gender	male	55	66	102	99	-	-	282	120	282	120	-	-
	female	31	38	42	54	93	90	171	153	171	153	582	510
age	<40	33	43	6	3	-	-	30	9	-	-	-	-
	40 to 60	28	37	51	84	42	39	162	144	-	-	-	-
	>60	25	24	87	66	51	51	114	144	-	-	-	-
magnification	level 1	27	29	48	51	31	30	102	99	151	91	194	170
	level 2	27	38	48	51	31	30	102	99	151	91	194	170
	level 3	32	37	48	51	31	30	102	99	151	91	194	170
censorship	alive	48	67	75	102	81	78	279	279	165	69	555	447
	dead	38	37	69	51	12	12	27	18	288	204	27	63
grade	grade I	-	-	3	3	30	24	-	-	-	-	225	99
	grade II	34	39	57	63	21	24	-	-	-	-	207	270
	grade III	28	36	60	63	42	42	-	-	-	-	150	141
	grade IV	24	29	24	24	-	-	-	-	453	273	-	-
time bin	bin 1	80	82	72	72	66	63	255	249	453	273	474	423
	bin 2	4	13	39	60	27	27	45	45	-	-	108	87
	bin 3	2	9	21	18	-	-	3	3	-	-	-	-
	bin 4	-	-	12	3	-	-	-	-	-	-	-	-



Supplementary Figure 1: Ablation for λ . Plot for AUC and C Index across λ for (a) TCGA-GBMLGG, (b) TCGA-KIRC. $\lambda = 0$ and $\lambda = 1$ correspond to the extreme cases of equation $L = \lambda \times L_{\text{survival}} + (1 - \lambda) \times L_{\text{grade}}$ where the model is trained only using gradation loss and only using survival loss respectively. Using only gradation loss leads to poor performance of risk estimation when $\lambda = 0$, and using only risk estimation loss leads to poor gradation performance. Learning both grade and survival risk together helps the model correlate between the two and learn better. This is expected as one of the losses essentially becomes unused, so we decide to ignore the extreme cases of $\lambda = 0$ and $\lambda = 1$. Also, we observe that for $\lambda = 0.3$ the model reaches an optimal performance where both AUC and C Index scores are close and high together for both cohorts.

Supplementary Table 4: Table of results for evaluation of the use of synthesized transcriptomic data for grade and survival prediction in four different settings- 1) We do not use transcriptomic data at all and attend to the WSI patches randomly; 2) We use the gene expression levels only to co-attend to the WSI patch embeddings; 3) We use the transcriptomic embeddings collectively with co-attended WSI features for survival risk estimation only; 4) We use transcriptomic features for co-attending to WSI patches, and then use the transcriptomic embeddings along with the co-attended features for performing both survival risk estimation and gradation. Setting 3 gives the best mean AUC and C Index for gradation and survival risk estimation across both data cohorts. Further, ANOSIM test for combined risk estimation and gradation between the two cases suggested that the difference in performance between all the cases is significant (p-value < 0.05). Thus, the architecture of our model is aligned to setting 3. $\uparrow$ indicates the larger the better. Best mean scores are marked in **bold**.

	Setting 1 random (not used)		Setting 2 only co-attention		Setting 3 co-attention, risk		Setting 4 co-attention, risk, grade	
λ	Risk C Index $\uparrow$	Grade AUC $\uparrow$	Risk C Index $\uparrow$	Grade AUC $\uparrow$	Risk C Index $\uparrow$	Grade AUC $\uparrow$	Risk C Index $\uparrow$	Grade AUC $\uparrow$
TCGA - GBMLGG								
0.1	0.835	0.829	0.843	0.878	0.863	0.884	0.848	0.868
0.2	0.832	0.863	0.872	0.869	0.852	0.895	0.864	0.809
0.3	0.842	0.823	0.849	0.886	0.861	0.890	0.859	0.849
0.4	0.844	0.828	0.857	0.885	0.864	0.886	0.858	0.829
0.5	0.856	0.847	0.861	0.890	0.846	0.864	0.830	0.844
0.6	0.847	0.873	0.865	0.877	0.847	0.883	0.820	0.733
0.7	0.861	0.836	0.865	0.867	0.855	0.888	0.879	0.843
0.8	0.862	0.745	0.857	0.856	0.836	0.811	0.808	0.731
0.9	0.850	0.810	0.852	0.801	0.845	0.809	0.801	0.724
mean	0.843	0.820	0.848	0.839	0.854	0.847	0.825	0.796
TCGA - KIRC								
0.1	0.626	0.729	0.666	0.713	0.703	0.773	0.576	0.741
0.2	0.709	0.713	0.658	0.768	0.650	0.778	0.516	0.735
0.3	0.671	0.714	0.688	0.763	0.681	0.773	0.548	0.746
0.4	0.701	0.712	0.701	0.770	0.675	0.772	0.513	0.740
0.5	0.662	0.770	0.605	0.766	0.658	0.769	0.509	0.747
0.6	0.672	0.763	0.654	0.767	0.667	0.763	0.516	0.743
0.7	0.677	0.739	0.673	0.768	0.643	0.760	0.512	0.750
0.8	0.708	0.763	0.650	0.774	0.613	0.765	0.551	0.753
0.9	0.607	0.754	0.592	0.760	0.575	0.757	0.573	0.742
mean	0.617	0.742	0.629	0.737	0.639	0.765	0.575	0.742

Supplementary Table 5: Table of results for comparison of synthesized and real transcriptomic features. Normalized MAE results are denoted with nMAE. The correlation coefficient is calculated using Spearman correlation, and all corresponding correlation p-values are found to be < 0.05 , denoting significant correlation. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better. A higher correlation with a high corresponding MAE would denote that the range of gene expression levels is large for the particular group.

Gene type	TCGA-GBMLGG		TCGA-KIRC		TCGA-UCEC	
	nMAE	Correlation	nMAE	Correlation	nMAE	Correlation
Tumor Suppressor Genes	0.184	0.472	0.166	0.629	0.177	0.372
Oncogenes	0.160	0.770	0.136	0.817	0.182	0.397
Protein Kinases	0.148	0.749	0.183	0.785	0.183	0.306
Cell Differentiation Markers	0.137	0.724	0.130	0.777	0.194	0.339
Transcription Factors	0.143	0.615	0.197	0.666	0.139	0.623
Cytokines and Growth Factors	0.107	0.771	0.091	0.610	0.236	0.270
All gene groups	0.141	0.713	0.160	0.717	0.173	0.436

Gene type	TCGA-BRCA		CPTAC-GBM		CPTAC-UCEC	
	nMAE	Correlation	nMAE	Correlation	nMAE	Correlation
Tumor Suppressor Genes	0.194	0.568	0.259	0.445	0.228	0.540
Oncogenes	0.136	0.705	0.043	0.874	0.197	0.582
Protein Kinases	0.131	0.766	0.036	0.923	0.209	0.516
Cell Differentiation Markers	0.166	0.617	0.036	0.941	0.206	0.541
Transcription Factors	0.167	0.629	0.316	0.490	0.146	0.827
Cytokines and Growth Factors	0.136	0.628	0.035	0.965	0.197	0.493
All gene groups	0.155	0.642	0.167	0.662	0.178	0.669

Supplementary Table 6: Table of results for comparing predictive performance on using only WSIs (without transcriptomic data), using synthesized transcriptomic data and WSIs, and using real transcriptomic data and WSIs. The separate p -values for gradation and risk estimation are computed by performing two-sided Wilcoxon rank-sum test, whereas the p -values over gradation and risk estimation together are computed using two-sided ANOSIM test. ANOSIM test p -values = 0.01, for comparison between without vs. synthesized, for TCGA-GBMLGG, TCGA-KIRC, TCGA-UCEC, CPTAC-GBM, and CPTAC-UCEC. However, for real vs. synthesized, ANOSIM test p -values evaluate to 0.929 for TCGA-GBMLGG, 0.327 for TCGA-KIRC, 1.000 for TCGA-UCEC, 0.993 for CPTAC-GBM and 1.000 for CPTAC-UCEC. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better.

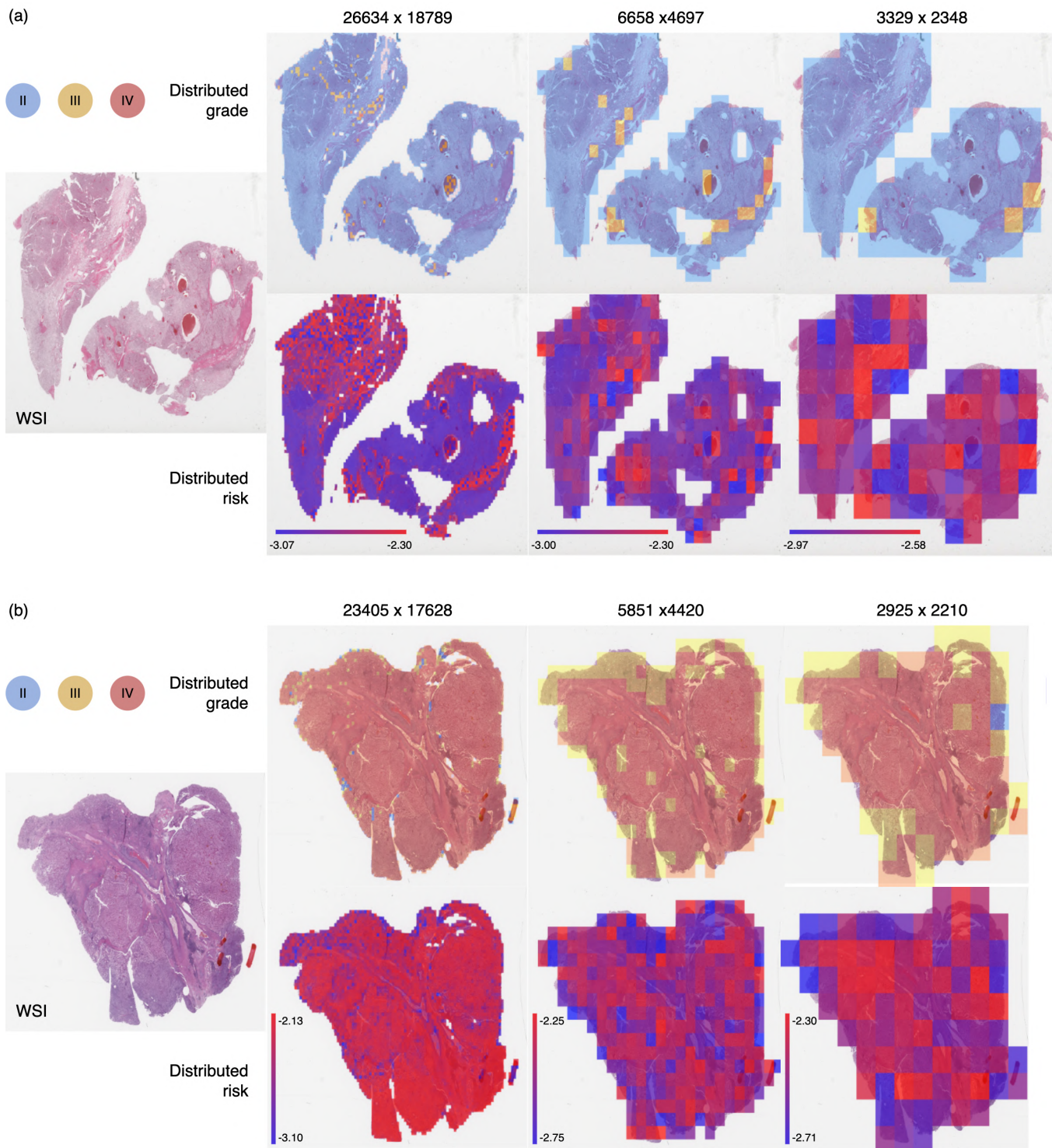
Datasets	Gradation				
	AUC $\uparrow$			Wilcoxon test p-value	
	without	synthesized	real	without vs. synthesized $\downarrow$	real vs. synthesized $\uparrow$
TCGA-GBMLGG	0.823	0.890	0.907	4.23E-02	9.95E-01
TCGA-KIRC	0.714	0.773	0.778	2.21E-04	8.50E-01
TCGA-UCEC	0.796	0.821	0.828	1.28E-02	9.41E-01
CPTAC-GBM	0.736	0.865	0.863	8.40E-03	9.85E-01
CPTAC-UCEC	0.508	0.593	0.593	7.76E-42	1.00E+00
Datasets	Risk Estimation				
	C Index $\uparrow$			Wilcoxon test p-value	
	without	synthesized	real	without vs. synthesized $\downarrow$	real vs. synthesized $\uparrow$
TCGA-GBMLGG	0.842	0.861	0.866	8.20E-07	6.04E-01
TCGA-KIRC	0.671	0.681	0.697	1.23E-11	3.00E-01
TCGA-UCEC	0.663	0.673	0.680	4.81E-31	8.87E-01
TCGA-BRCA	0.603	0.720	0.720	2.35E-03	9.30E-01
CPTAC-GBM	0.547	0.565	0.564	3.90E-33	8.15E-01
CPTAC-UCEC	0.518	0.530	0.533	2.26E-164	9.32E-01

Supplementary Table 7: Table for results for gradation and survival risk estimation using real and synthesized transcriptomic data, separately for the TCGA-GBMLGG cohorts. The p -values are computed using two-sided Wilcoxon rank-sum test.

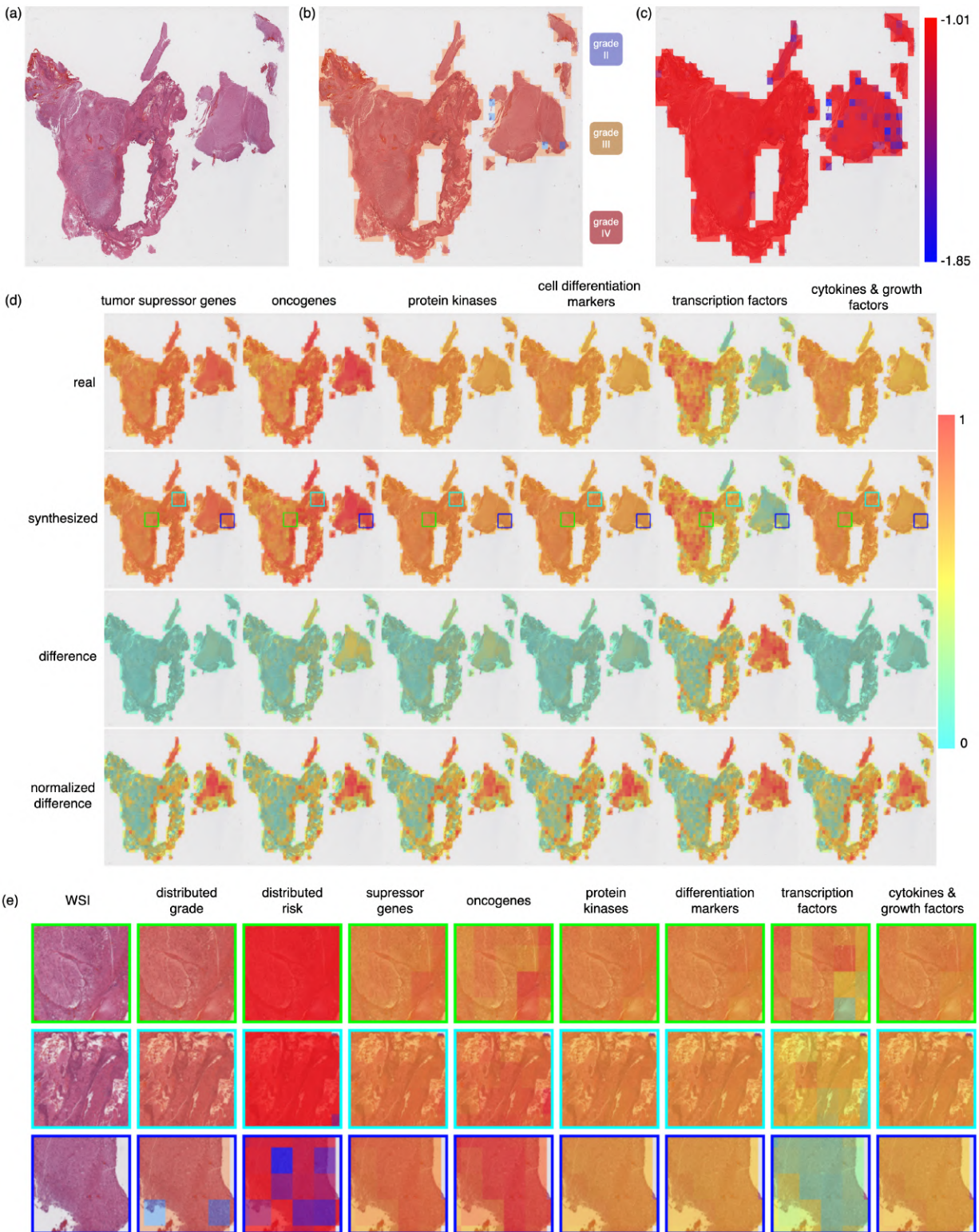
TCGA-GBMLGG cohorts	Gradation (AUC $\uparrow$)			Risk Estimation (C Index $\uparrow$)		
	real	synthesized	Wilcoxon p-value	real	synthesized	Wilcoxon p-value
TCGA-LGG	0.835	0.81	0.775	0.707	0.675	0.472
TCGA-GBM	0.995	0.993	0.822	0.509	0.539	0.489

Supplementary Table 8: Table for ablation study on percentage contribution of the different gene groups in co-attention with WSI features for different cohorts.

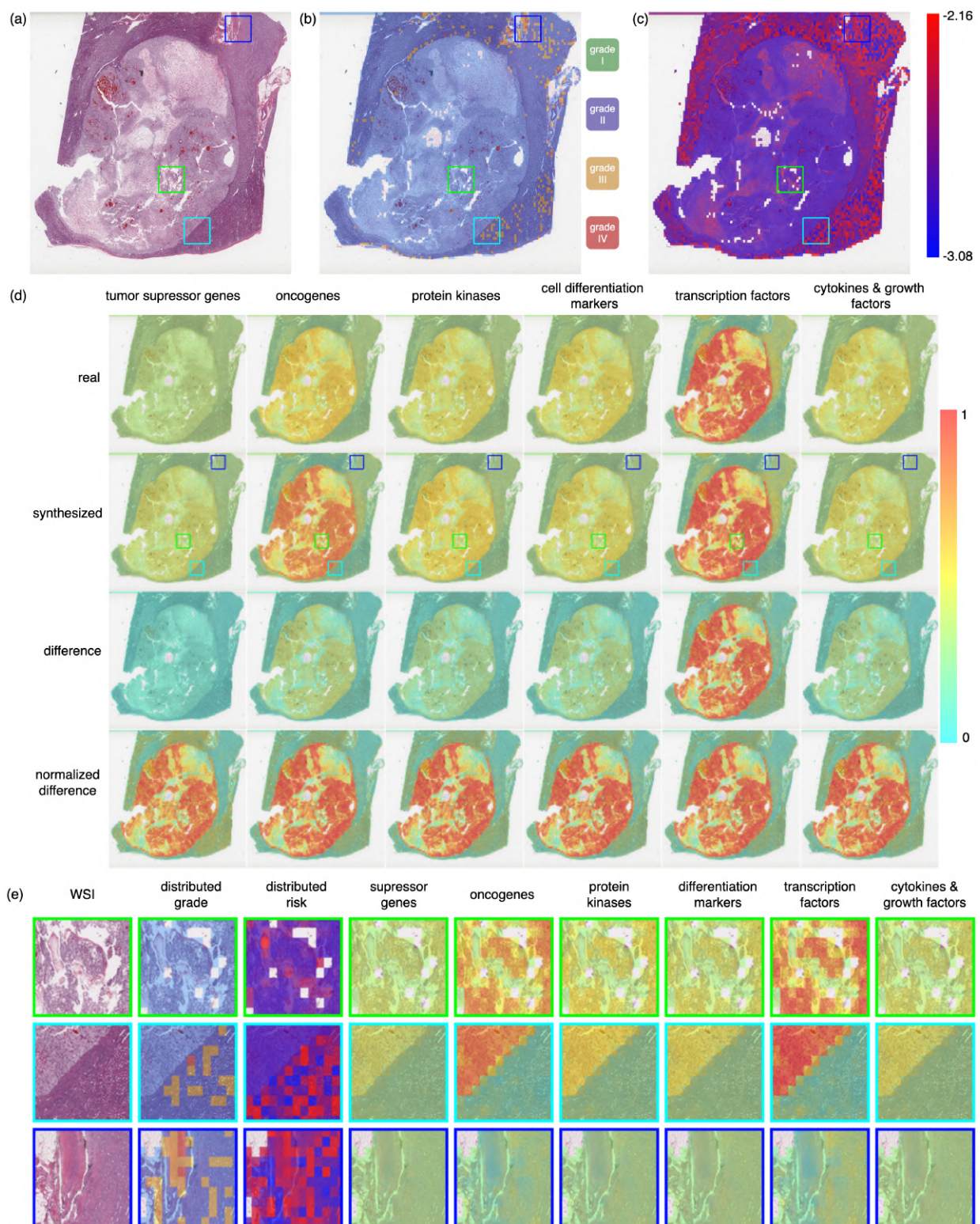
Gene type	Percentage Contribution (in %)					
	TCGA-GBMLGG	TCGA-KIRC	TCGA-UCEC	TCGA-BRCA	CPTAC-GBM	CPTAC-UCEC
Tumor Suppressor Genes	12.51	6.05	5.25	11.56	12.15	4.74
Oncogenes	18.02	15.59	4.33	12.90	13.16	6.35
Protein Kinases	10.40	11.86	38.41	21.12	15.30	29.09
Cell Differentiation Markers	10.00	8.10	10.82	18.14	23.55	18.00
Transcription Factors	42.21	48.80	17.74	23.49	29.06	25.89
Cytokines and Growth Factors	6.85	9.61	23.44	12.80	6.77	15.93



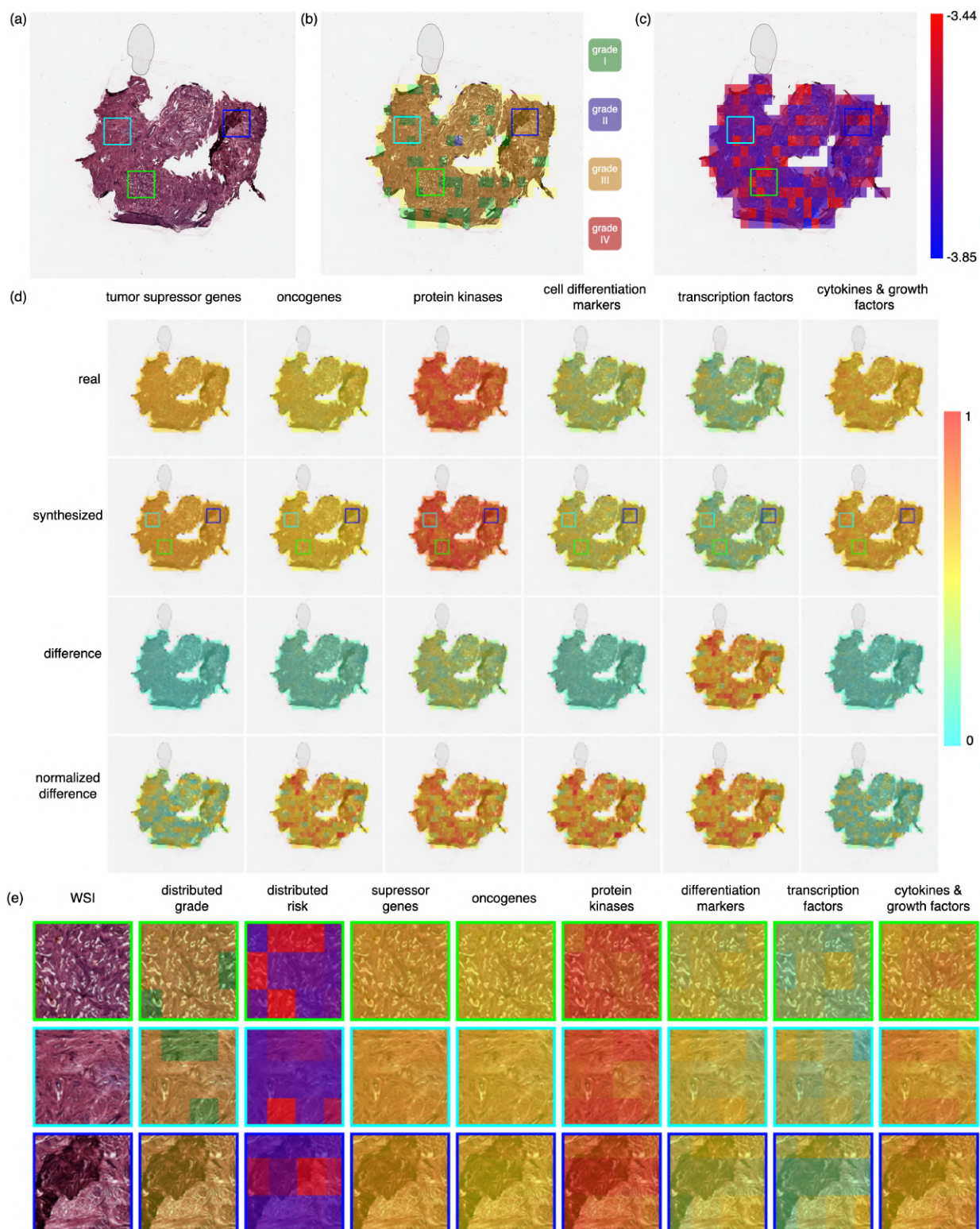
Supplementary Figure 2: Predicted intra-tumour heterogeneity at different magnification levels (a) A whole slide image (WSI) of ground truth grade II, survival time 62.62 months and alive survival status belonging to a male patient of 47 years. (b) A WSI of ground truth grade IV, survival time 54.61 months and dead survival status belonging to a 55 years old female patient. The WSIs belong to TCGA-KIRC. It is observed that for higher resolution images, the distributed predictions are more accurate.



Supplementary Figure 3: Explainability maps for a male patient of age 62, ground truth grade IV, survival time 10.28 months, and survival status deceased from the TCGA-GBM cohort. (a) whole slide image (WSI) (b) intra-tumour gradation heterogeneity (c) intra-tumour survival heterogeneity (d) co-attention maps for real and synthesized transcriptomic data and WSI patches for different gene groups (e) comparative study of intra-tumour heterogeneity and corresponding co-attention maps obtained for prediction using synthesized transcriptomes for chosen regions marked with different coloured rectangles.



Supplementary Figure 4: Explainability maps for a female patient of age 50, ground truth grade II, survival time 22.78 months, and survival status alive from the TCGA-KIRC cohort. (a) whole slide image (WSI) (b) intra-tumour gradation heterogeneity (c) intra-tumour survival heterogeneity (d) co-attention maps for real and synthesized transcriptomic data and WSI patches for different gene groups (e) comparative study of intra-tumour heterogeneity and corresponding co-attention maps obtained for prediction using synthesized transcriptomes for chosen regions marked with different coloured rectangles.



Supplementary Figure 5: Explainability maps for a female patient of ground truth grade III, survival time 58.08 months, and survival status alive from the CPTAC-UCEC cohort. (a) whole slide image (WSI) (b) intra-tumour gradation heterogeneity (c) intra-tumour survival heterogeneity (d) co-attention maps for real and synthesized transcriptomic data and WSI patches for different gene groups (e) comparative study of intra-tumour heterogeneity and corresponding co-attention maps obtained for prediction using synthesized transcriptomes for chosen regions marked with different coloured rectangles.

Supplementary Table 9: Table of uncertainty and fairness evaluation for **TCGA-GBMLGG** dataset. The p -values are computed using two-sided Wilcoxon rank-sum test to compare the grades and survival risks predicted using real and synthesized transcriptomic data, in addition to WSIs. M_{unc} , M_{cov} , C_{unc} and C_{cov} refer to marginal uncertainty, marginal coverage, conditional uncertainty and conditional coverage, respectively. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better. Equivalent or close (± 0.01) performance using real and synthesized data is marked in **bold**.

Category	Group	Gradation												
		Real					Synthesized					p-value $\uparrow$		
		AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	grade	M_{unc}	C_{unc}
overall	-	0.907	0.407	0.981	-	-	0.890	0.413	0.981	-	-	0.995	0.935	-
gender	male	0.915	0.439	0.985	0.312	0.953	0.896	0.444	0.985	0.323	0.953	0.975	0.920	0.921
	female	0.896	0.184	0.921			0.886	0.202	0.921			1.000	0.921	
age	<40	0.883	1.000	1.000	0.727	0.986	0.849	1.000	1.000	0.720	0.986	0.736	1.000	0.977
	40 - 60	0.928	0.459	1.000			0.921	0.450	1.000			0.804	0.996	
	>60	0.909	0.722	0.958			0.906	0.708	0.958			1.000	0.934	
magni- fication	level1	0.902	0.195	0.966	0.617	0.989	0.885	0.172	0.931	0.609	0.977	1.000	0.652	0.884
	level2	0.952	0.754	1.000			0.936	0.754	1.000			0.811	1.000	
	level3	0.852	0.901	1.000			0.836	0.901	1.000			0.795	1.000	
vital status	alive	0.865	0.990	1.000	0.612	0.986	0.832	1.000	1.000	0.617	0.986	0.894	0.881	0.941
	dead	0.951	0.234	0.973			0.951	0.234	0.973			0.867	1.000	
grade	grade II	0.914	0.547	0.974	0.527	0.980	0.891	0.598	0.974	0.544	0.980	0.704	0.569	0.856
	grade III	0.755	1.000	1.000			0.728	1.000	1.000			0.866	1.000	
	grade IV	0.995	0.034	0.966			0.993	0.034	0.966			0.822	1.000	
time bin	bin 1	0.898	0.431	0.976	0.335	0.992	0.882	0.407	0.963	0.331	0.988	0.984	0.790	0.766
	bin 2	0.786	0.538	1.000			0.784	0.513	1.000			1.000	0.817	
	bin 3	1.000	0.037	1.000			1.000	0.074	1.000			1.000	0.691	

Category	Group	Risk Estimation												
		Real					Synthesized					p-value $\uparrow$		
		C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	grade	M_{unc}	C_{unc}
overall	-	0.866	0.446	0.933	-	-	0.861	0.438	0.923	-	-	0.604	0.652	-
gender	male	0.808	0.375	0.985	0.514	0.953	0.813	0.376	0.985	0.507	0.953	0.956	0.956	0.678
	female	0.795	0.653	0.921			0.795	0.639	0.921			0.350	0.400	
age	<40	0.524	0.767	0.977	0.541	0.983	0.540	0.727	1.000	0.535	0.991	0.119	0.135	0.546
	40 - 60	0.848	0.480	0.973			0.845	0.492	0.973			0.775	0.770	
	>60	0.738	0.377	1.000			0.742	0.387	1.000			0.741	0.734	
magni- fication	level1	0.855	0.735	0.966	0.586	0.971	0.840	0.725	1.000	0.576	0.982	0.646	0.810	0.787
	level2	0.862	0.484	0.974			0.856	0.486	0.974			0.950	0.975	
	level3	0.873	0.540	0.973			0.858	0.517	0.973			0.578	0.578	
vital status	alive	-	0.368	0.851	0.336	0.912	-	0.359	0.851	0.337	0.925	0.237	0.441	0.502
	dead	0.614	0.304	0.973			0.633	0.314	1.000			0.563	0.563	
grade	grade II	0.823	0.582	0.872	0.431	0.937	0.823	0.544	0.824	0.420	0.920	0.058	0.054	0.471
	grade III	0.590	0.655	0.972			0.526	0.659	0.972			0.901	0.866	
	grade IV	0.509	0.056	0.966			0.539	0.059	0.966			0.489	0.494	
time bin	bin 1	0.827	0.252	0.951	0.684	0.958	0.822	0.259	0.939	0.675	0.954	0.942	0.893	0.465
	bin 2	-	0.804	0.923			-	0.782	0.923			0.209	0.270	
	bin 3	-	0.996	1.000			-	0.984	1.000			0.200	0.233	

Supplementary Table 10: Table of uncertainty and fairness evaluation for **CPTAC-GBM** dataset. p-values are computed using the two-sided Wilcoxon rank-sum test to compare the grades and survival risks predicted using real and synthesized transcriptomic data, in addition to WSIs. M_{unc} , M_{cov} , C_{unc} and C_{cov} refer to marginal uncertainty, marginal coverage, conditional uncertainty and conditional coverage, respectively. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better. Equivalent or close (± 0.01) performance using real and synthesized data is marked in **bold**.

Category	Group	Gradation												
		Real					Synthesized					p-value $\uparrow$		
		AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	grade	M_{unc}	C_{unc}
overall	-	0.863	0.465	0.932		-	0.865	0.465	0.936	-	-	0.985	0.991	-
gender	male	0.847	0.514	0.972	0.495	0.962	0.847	0.515	0.975	0.494	0.964	0.975	0.968	0.961
	female	0.894	0.476	0.953			0.898	0.472	0.953			0.935	0.955	
magni- fication	level1	0.861	0.430	0.940	0.478	0.949	0.864	0.435	0.947	0.483	0.956	0.918	0.958	0.938
	level2	0.865	0.506	0.954			0.867	0.510	0.960			0.931	0.949	
	level3	0.869	0.497	0.954			0.870	0.503	0.960			1.000	0.905	
vital status	alive	0.835	0.463	0.903	0.478	0.932	0.833	0.465	0.903	0.482	0.938	0.970	1.000	0.947
	dead	0.880	0.493	0.962			0.883	0.499	0.972			0.952	0.894	
grade	grade IV	0.863	0.465	0.932	-	-	0.865	0.465	0.936	-	-	0.985	0.991	-
time bin	bin 1	0.863	0.465	0.932	-	-	0.865	0.465	0.936	-	-	0.985	0.991	-

Category	Group	Risk Estimation												
		Real					Synthesized					p-value $\uparrow$		
		C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	risk	M_{unc}	C_{unc}
overall	-	0.564	0.206	0.971	-	-	0.565	0.206	0.971	-	-	0.815	0.811	-
gender	male	0.653	0.223	0.972	0.205	0.971	0.651	0.222	0.975	0.204	0.973	0.809	0.811	0.870
	female	0.430	0.186	0.971			0.432	0.185	0.971			0.931	0.929	
magni- fication	level1	0.544	0.208	0.980	0.208	0.971	0.543	0.207	0.987	0.208	0.976	0.718	0.719	0.852
	level2	0.574	0.209	0.967			0.578	0.209	0.974			0.931	0.937	
	level3	0.570	0.208	0.967			0.572	0.209	0.967			0.900	0.899	
vital status	alive	-	0.183	0.970	0.195	0.964	-	0.184	0.982	0.195	0.974	0.821	0.834	0.853
	dead	0.582	0.207	0.958			0.581	0.206	0.965			0.870	0.873	
grade	grade IV	0.564	0.206	0.971	-	-	0.565	0.206	0.971	-	-	0.815	0.811	-
time bin	bin 1	0.564	0.206	0.971	-	-	0.565	0.206	0.971	-	-	0.815	0.811	-

Supplementary Table 11: Table of uncertainty and fairness evaluation for **TCGA-KIRC** dataset. p-values are computed using the two-sided Wilcoxon rank-sum test to compare the grades and survival risks predicted using real and synthesized transcriptomic data, in addition to WSIs. M_{unc} , M_{cov} , C_{unc} and C_{cov} refer to marginal uncertainty, marginal coverage, conditional uncertainty and conditional coverage, respectively. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better. Equivalent or close (± 0.01) performance using real and synthesized data is marked in **bold**.

Category	Group	Gradation												
		Real					Synthesized					p-value $\uparrow$		
		AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	grade	M_{unc}	C_{unc}
overall	-	0.778	0.292	0.922	-	-	0.773	0.334	0.948	-	-	0.850	0.068	-
gender	male	0.726	0.303	0.909	0.290	0.936	0.721	0.337	0.929	0.337	0.955	0.905	0.246	0.184
	female	0.864	0.278	0.963			0.864	0.336	0.981			0.878	0.123	
age	<40	0.863	0.056	1.000	0.248	0.952	0.852	0.111	1.000	0.273	0.953	0.513	0.513	0.643
	40 - 60	0.780	0.393	0.976			0.778	0.389	0.964			1.000	0.894	
	>60	0.770	0.295	0.879			0.761	0.318	0.894			0.886	0.521	
magni- fication	level1	0.784	0.369	0.980	0.341	0.961	0.780	0.337	0.961	0.358	0.961	1.000	0.393	0.338
	level2	0.780	0.317	0.961			0.774	0.376	0.980			0.867	0.125	
	level3	0.771	0.337	0.941			0.769	0.363	0.941			0.867	0.495	
vital status	alive	0.826	0.239	0.931	0.320	0.926	0.822	0.288	0.951	0.358	0.946	0.905	0.076	0.286
	dead	0.661	0.402	0.922			0.656	0.428	0.941			0.880	0.495	
grade	grade I	0.222	0.500	0.667	0.356	0.859	0.111	0.500	0.333	0.349	0.779	1.000	1.000	0.625
	grade II	0.987	0.016	0.905			0.978	0.000	0.873			0.880	0.878	
	grade III	0.647	0.283	0.905			0.658	0.333	0.952			0.645	0.164	
	grade IV	0.504	0.625	0.958			0.499	0.563	0.958			0.711	0.458	
time bin	bin 1	0.718	0.366	0.944	0.247	0.874	0.713	0.384	0.944	0.279	0.961	0.889	0.565	0.552
	bin 2	0.837	0.214	0.883			0.828	0.233	0.900			0.885	0.560	
	bin 3	0.848	0.296	1.000			0.835	0.333	1.000			1.000	0.569	
	bin 4	0.667	0.111	0.667			0.667	0.167	1.000			1.000	0.513	

Category	Group	Risk Estimation												
		Real					Synthesized					p-value $\uparrow$		
		C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	risk	M_{unc}	C_{unc}
overall	-	0.697	0.898	0.882	-	-	0.681	0.909	0.928	-	-	0.300	0.307	-
gender	male	0.705	0.909	0.909	0.882	0.834	0.652	0.921	0.960	0.893	0.878	0.669	0.654	0.483
	female	0.710	0.855	0.759			0.771	0.865	0.796			0.219	0.311	
age	<40	-	0.955	1.000	0.919	0.934	-	0.942	1.000	0.923	0.971	0.275	0.275	0.440
	40 - 60	0.612	0.936	0.893			0.591	0.935	0.929			0.266	0.265	
	>60	0.672	0.864	0.909			0.681	0.891	0.985			0.781	0.781	
magni- fication	level1	0.706	0.932	0.961	0.903	0.908	0.676	0.947	1.000	0.915	0.948	0.563	0.588	0.452
	level2	0.734	0.899	0.902			0.702	0.908	0.961			0.248	0.277	
	level3	0.694	0.878	0.863			0.676	0.890	0.882			0.408	0.493	
vital status	alive	-	0.934	0.912	0.889	0.907	-	0.934	0.941	0.904	0.961	0.083	0.089	0.343
	dead	0.487	0.845	0.902			0.499	0.874	0.980			0.595	0.597	
grade	grade I	-	0.976	0.667	0.927	0.847	-	0.969	1.000	0.939	0.948	0.275	0.513	0.358
	grade II	0.835	0.926	0.905			0.841	0.919	0.937			0.139	0.149	
	grade III	0.612	0.906	0.857			0.573	0.904	0.857			0.550	0.659	
	grade IV	0.481	0.901	0.958			0.439	0.964	1.000			0.143	0.112	
time bin	bin 1	0.690	0.922	0.861	0.738	0.961	0.714	0.938	0.944	0.746	0.982	0.441	0.474	0.744
	bin 2	0.842	0.864	0.983			0.836	0.881	0.983			0.557	0.502	
	bin 3	-	0.389	1.000			-	0.389	1.000			0.411	1.000	
	bin 4	-	0.775	1.000			-	0.775	1.000			0.275	1.000	

Supplementary Table 12: Table of uncertainty and fairness evaluation for TCGA-UCEC dataset. p-values are computed using the two-sided Wilcoxon rank-sum test to compare the grades and survival risks predicted using real and synthesized transcriptomic data, in addition to WSIs. M_{unc} , M_{cov} , C_{unc} and C_{cov} refer to marginal uncertainty, marginal coverage, conditional uncertainty and conditional coverage, respectively. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better. Equivalent or close (± 0.01) performance using real and synthesized data is marked in **bold**.

Category	Group	Gradation												
		Real					Synthesized					p-value $\uparrow$		
		AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	grade	M_{unc}	C_{unc}
overall	-	0.828	0.337	0.854	-	-	0.821	0.337	0.854	-	-	0.941	1.000	-
age	40 - 60	0.739	0.744	0.949	0.489	0.916	0.739	0.744	0.949	0.489	0.916	1.000	1.000	1.000
	>60	0.885	0.235	0.882			0.874	0.235	0.882			0.912	1.000	
magni- fication	level1	0.881	0.567	0.933	0.519	0.915	0.870	0.544	0.933	0.511	0.915	0.912	0.877	0.959
	level2	0.810	0.278	0.812			0.810	0.278	0.812			1.000	1.000	
	level3	0.775	0.711	1.000			0.775	0.711	1.000			1.000	1.000	
vital status	alive	0.811	0.350	0.866	0.300	0.850	0.803	0.350	0.866	0.300	0.850	0.941	1.000	1.000
	dead	0.917	0.250	0.833			0.917	0.250	0.833			1.000	1.000	
grade	grade I	0.971	1.000	1.000	0.651	0.984	0.971	1.000	1.000	0.651	0.984	1.000	1.000	1.000
	grade II	0.805	0.556	1.000			0.768	0.556	1.000			0.893	1.000	
	grade III	0.823	0.397	0.952			0.823	0.397	0.952			1.000	1.000	
time bin	bin1	0.761	0.661	0.968	0.470	0.984	0.752	0.651	0.968	0.464	0.984	0.932	0.901	0.950
	bin2	0.955	0.278	1.000			0.955	0.278	1.000			1.000	1.000	

Category	Group	Risk Estimation												
		Real					Synthesized					p-value $\uparrow$		
		C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	risk	M_{unc}	C_{unc}
overall	-	0.680	1.000	0.933	-	-	0.673	1.000	0.933	-	-	0.887	0.992	-
age	40 - 60	-	1.000	0.897	1.000	0.863	-	1.000	0.897	1.000	0.863	0.912	0.980	0.986
	>60	0.606	0.999	0.829			0.592	0.999	0.829			0.823	0.992	
magni- fication	level1	0.720	1.000	0.900	1.000	0.922	0.680	1.000	0.900	1.000	0.922	0.767	0.971	0.982
	level2	0.707	1.000	0.933			0.707	1.000	0.933			0.900	0.982	
	level3	0.547	1.000	0.933			0.547	1.000	0.933			0.918	0.994	
vital status	alive	-	1.000	0.936	0.998	0.968	-	1.000	0.936	0.998	0.968	0.897	0.992	0.961
	dead	0.204	0.996	1.000			0.204	0.996	1.000			0.729	0.931	
grade	grade I	-	1.000	0.958	1.000	0.962	-	1.000	0.958	1.000	0.962	0.926	0.992	0.991
	grade II	0.778	1.000	1.000			0.746	1.000	1.000			0.805	1.000	
	grade III	0.503	1.000	0.929			0.509	1.000	0.929			0.936	0.982	
time bin	bin1	0.714	1.000	0.921	0.996	0.960	0.700	1.000	0.921	0.996	0.960	0.880	0.990	0.970
	bin2	-	0.992	1.000			-	0.992	1.000			0.902	0.951	

Supplementary Table 13: Table of uncertainty and fairness evaluation for **CPTAC-UCEC** dataset. p-values are computed using the two-sided Wilcoxon rank-sum test to compare the grades and survival risks predicted using real and synthesized transcriptomic data, in addition to WSIs. M_{unc} , M_{cov} , C_{unc} and C_{cov} refer to marginal uncertainty, marginal coverage, conditional uncertainty and conditional coverage, respectively. $\uparrow$ indicates the larger the better, and $\downarrow$ indicates the smaller the better. Equivalent or close (± 0.01) performance using real and synthesized data is marked in **bold**.

Category	Group	Gradation												
		Real					Synthesized					p-value $\uparrow$		
		AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	AUC $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	grade	M_{unc}	C_{unc}
overall	-	0.593	0.821	0.984	-	-	0.593	0.818	0.980	-	-	1.000	0.914	-
magnification	level1	0.595	0.761	0.988	0.805	0.976	0.595	0.761	0.988	0.805	0.976	1.000	1.000	1.000
	level2	0.601	0.824	0.971			0.600	0.824	0.971			1.000	1.000	
	level3	0.593	0.831	0.971			0.592	0.831	0.971			1.000	1.000	
vital status	alive	0.575	0.826	0.982	0.717	0.959	0.575	0.823	0.978	0.715	0.957	1.000	0.908	0.954
	dead	0.728	0.608	0.937			0.728	0.608	0.937			1.000	1.000	
grade	grade I	0.469	0.973	0.960	0.644	0.928	0.465	0.973	0.960	0.643	0.926	0.939	1.000	0.929
	grade II	0.454	0.746	0.956			0.454	0.751	0.956			0.952	0.887	
	grade III	0.887	0.213	0.870			0.889	0.206	0.863			1.000	0.899	
time bin	bin1	0.496	0.812	0.983	0.841	0.986	0.503	0.809	0.979	0.840	0.984	1.000	0.905	0.953
	bin2	0.608	0.870	0.989			0.608	0.870	0.989			1.000	1.000	

Category	Group	Risk Estimation												
		Real					Synthesized					p-value $\uparrow$		
		C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	C Index $\uparrow$	$M_{unc}\downarrow$	$M_{cov}\uparrow$	$C_{unc}\downarrow$	$C_{cov}\uparrow$	risk	M_{unc}	C_{unc}
overall	-	0.533	1.000	0.908	-	-	0.530	1.000	0.908	-	-	0.932	0.918	-
magnification	level1	0.393	0.999	0.912	0.999	0.920	0.387	0.999	0.912	0.999	0.920	0.891	0.997	0.950
	level2	0.594	1.000	0.941			0.592	1.000	0.941			0.975	0.925	
	level3	0.586	1.000	0.906			0.583	1.000	0.906			0.980	0.927	
vital status	alive	-	1.000	0.899	0.582	0.934	-	1.000	0.895	0.583	0.932	0.926	0.993	0.695
	dead	0.509	0.165	0.968			0.510	0.166	0.968			0.996	0.397	
grade	grade I	0.471	1.000	0.949	0.999	0.921	0.448	1.000	0.960	0.999	0.924	0.871	0.733	0.909
	grade II	0.507	1.000	0.933			0.507	1.000	0.933			0.994	0.996	
	grade III	0.445	0.996	0.879			0.444	0.996	0.879			0.930	0.999	
time bin	bin1	0.527	1.000	0.915	0.895	0.957	0.527	1.000	0.915	0.895	0.952	0.988	0.999	0.950
	bin2	-	0.791	1.000			-	0.790	0.989			0.901	0.901	

Supplementary Table 14: Table of expected deviation from the desired coverage for a given number of samples in the calibration set and conformal error rate $\alpha = 0.1$. ncal denotes the calibration set size, and ε denotes the deviation. The less the calibration set size, the greater the expected deviation.

ncal	ε	ncal	ε	ncal	ε	ncal	ε	ncal	ε	ncal	ε
9	0.183	33	0.079	67	0.060	104	0.047	194	0.035	373	0.025
12	0.121	34	0.076	69	0.064	111	0.048	201	0.036	375	0.026
13	0.106	36	0.077	72	0.058	120	0.047	204	0.034	399	0.026
18	0.097	37	0.078	75	0.055	126	0.044	207	0.035	447	0.024
21	0.107	38	0.078	78	0.057	132	0.043	225	0.033	453	0.023
24	0.085	39	0.087	80	0.058	138	0.043	228	0.034	501	0.022
25	0.086	42	0.074	81	0.056	141	0.043	231	0.033	510	0.022
27	0.087	43	0.071	82	0.054	144	0.040	252	0.031	531	0.022
28	0.087	48	0.071	84	0.052	150	0.042	270	0.031	550	0.022
29	0.102	51	0.071	86	0.053	151	0.041	273	0.030	555	0.021
30	0.095	54	0.063	87	0.053	153	0.039	279	0.031	582	0.021
31	0.089	55	0.063	90	0.055	165	0.038	282	0.030	693	0.019
32	0.084	57	0.065	91	0.053	170	0.039	288	0.030	700	0.019
33	0.079	60	0.067	93	0.050	171	0.039	300	0.029	780	0.018
34	0.076	63	0.060	99	0.053	177	0.038	309	0.029	918	0.017
32	0.084	66	0.060	102	0.049	192	0.036	339	0.028	1119	0.015