Major Tissue Type	Num. Slides
Heart	10427
Lung	9846
Kidney	8388
Bowel / Lower GI	8303
Soft Tissue	7863
Brain	7412
Esophagogastric	6705
Endocrine	6138
Female Genital Tract	5796
Lymphatic System	4957
Liver Biliary Tract	4677
Male Genital Tract	4017
Skin	3653
Breast	3364
Bone	2667
Pancreas	2328
Head & Neck	1555
Peritoneum	1210
Bladder	1059
Eye	61
Total	100426

Supplementary Data Table 1: **Major Tissue Type Distribution of Mass-100K.** Mass-100K is a pretraining dataset that consists of 100,130,900 tissue patches from 100,426 diagnostic whole-slide images (WSIs) across 20 major tissue types collected from the Massachusetts General Hospital (MGH), Brigham & Women's Hospital (BWH) and the Genotype-Tissue Expression (GTEx) consortium.

Major Tissue Type	Num. Slides
Heart	4
Lung	3044
Kidney	435
Bowel / Lower GI	2002
Soft Tissue	1275
Brain	1607
Esophagogastric	882
Endocrine	549
Female Genital Tract	1834
Lymphatic System	1342
Liver Biliary Tract	1403
Male Genital Tract	587
Skin	627
Breast	1366
Bone	301
Pancreas	543
Head & Neck	532
Peritoneum	389
Bladder	568
Eye	21
Total	21444

Supplementary Data Table 2: **Major Tissue Type Distribution of Mass-22K.** Mass-22K is a subset of Mass-100K, which contains 16,059,454 histology image patches sampled from 21,444 WSIs across 20 major tissue types with cancerous tissue collected from Massachusetts General Hospital (MGH) and Brigham & Women's Hospital (BWH).

Major Tissue Type	Num. Slides
Heart	3
Lung	136
Kidney	43
Bowel / Lower GI	55
Soft Tissue	176
Brain	219
Esophagogastric	32
Endocrine	34
Female Genital Tract	187
Lymphatic System	134
Liver Biliary Tract	75
Male Genital Tract	33
Skin	43
Breast	42
Bone	32
Pancreas	28
Head & Neck	61
Peritoneum	34
Bladder	30
Eye	7
Total	1404

Supplementary Data Table 3: **Major Tissue Type Distribution of Mass-1K.** Mass-1K is a subset of Mass-22K, which contains 1,064,615 histology image patches sampled from 1,404 WSIs across 20 major tissue types with cancerous tissue collected from Massachusetts General Hospital (MGH) and Brigham & Women's Hospital (BWH).

Cancer Type	OncoTree Code	Num. Slides
Adrenocortical Carcinoma	ACC	51 (31:20)
Ampullary Cancer	AMPCA	42 (16:26)
Anal Cancer	ANSC	65 (51:14)
Appendiceal Cancer	MAAP	40 (23:17)
Bladder Cancer	BLAD	28 (8:20)
	BLCA	65 (20:45)
	UTUC	65 (22:43)
Bone Cancer	CHS	32 (10:22)
	ES	48 (17:31)
	OS	54 (26:28)
Breast Cancer	IDC	65 (56:9)
	ILC	65 (65:0)
	MDLC	65 (62:3)
CNS Cancer	ATM	20 (7:13)
	MNG	65 (47:18)
Cervical Cancer	CESC	60 (60:0)
	ECAD	42 (42:0)
Colorectal Cancer	COAD	65 (35:30)
	MACR	35 (14:21)
	READ	65 (31:34)
Embryonal Tumor	MBL	62 (23:39)
Endometrial Cancer	UCCC	23 (23:0)
	UCS	65 (65:0)
	UEC	65 (65:0)
	UMEC	25 (25:0)
	USC	65 (65:0)
Esophagogastric Cancer	ESCA	64 (19:45)
	ESCC	65 (31:34)
	GEJ	65 (17:48)
	SSRCC	28 (15:13)
	STAD	65 (31:34)
Gastroint. Neuroend. Tumor	GINET	65 (33:32)
Gastroint. Stromal Tumor	GIST	65 (33:32)
Germ Cell Tumor	MGCT	36 (0:36)
	SEM	63 (0:63)

Cancer Type (Cont.)	OncoTree Code	Num. Slides
Glioma	AASTR	54 (20:34)
	AODG	31 (17:14)
	ASTR	65 (30:35)
	EPM	38 (13:25)
	GBM	65 (33:32)
	HGGNOS	65 (26:39)
	LGGNOS	63 (24:39)
	ODG	58 (28:30)
	PAST	35 (18:17)
Head and Neck Cancer	HNSC	65 (19:46)
Hepatobiliary Cancer	CHOL	65 (26:39)
	GBC	65 (47:18)
	HCC	65 (25:40)
Mature B-Cell Neoplasms	CLLSLL	26 (13:13)
	DLBCLNOS	65 (23:42)
	FL	54 (28:26)
Mature T and NK Neoplasms	MYCF	37 (15:22)
Melanoma	MEL	65 (28:37)
Mesothelioma	PLBMESO	63 (6:57)
	PLEMESO	65 (20:45)
Misc. Neuroepithelial Tumor	PGNG	24 (13:11)
Nerve Sheath Tumor	MPNST	40 (17:23)
	SCHW	38 (23:15)
Non-Small Cell Lung Cancer	ALUCA	30 (20:10)
	LUAD	64 (37:27)
	LUAS	32 (19:13)
	LUCA	64 (52:12)
	LUNE	33 (20:13)
	LUSC	65 (30:35)
	NSCLCPD	65 (30:35)
Ovarian Cancer	CCOV	65 (65:0)
	EOV	65 (65:0)
	HGSOC	65 (65:0)
	LGSOC	62 (62:0)
	MOV	33 (33:0)
	MXOV	49 (49:0)
	OCS	42 (42:0)
	SBOV	28 (28:0)
Pancreatic Cancer	PAAD	64 (26:38)
	PANET	65 (31:34)
Peripheral Nervous System	NBL	65 (27:38)
Prostate Cancer	PRAD	65 (0:65)

Cancer Type (Cont.)	OncoTree Code	Num. Slides
Renal Cell Carcinoma	CCRCC	65 (26:39)
	CHRCC	65 (28:37)
	PRCC	65 (22:43)
	ROCY	20 (6:14)
Salivary Gland Cancer	ACYC	33 (21:12)
Sellar Tumor	PTAD	65 (41:24)
Sex Cord Stromal Tumor	GRCT	33 (33:0)
Skin Cancer, Non-Melanoma	CSCC	65 (15:50)
	MCC	65 (23:42)
Small Bowel Cancer	SBC	53 (18:35)
Small Cell Lung Cancer	SCLC	65 (36:29)
Soft Tissue Sarcoma	ANGS	50 (28:22)
	DDLS	65 (28:37)
	DES	41 (27:14)
	ERMS	20 (9:11)
	HEMA	65 (31:34)
	LMS	65 (50:15)
	MFH	60 (24:36)
	MFS	30 (15:15)
	PECOMA	21 (18:3)
	SFT	45 (21:24)
	SYNS	43 (18:25)
	WDLS	23 (12:11)
Thymic Tumor	THYM	24 (14:10)
Thyroid Cancer	THAP	30 (19:11)
	THFO	28 (15:13)
	THME	34 (15:19)
	THPA	65 (39:26)
Uterine Sarcoma	ULMS	51 (51:0)
Vaginal Cancer	VSC	47 (47:0)
Wilms Tumor	WT	63 (29:34)

Supplementary Data Table 4: **Hierarchical label distribution of the coarse and fine-grained tasks in OncoTree (OT) code cancer classification.** The OT-43 and OT-108 tasks is developed from a dataset comprising 5,564 WSIs from 43 cancer types collected from in-house BWH slides, which are further subdivided into 108 OncoTree codes, with at least 20 WSIs per OncoTree code. The OT-43 task is developed using the first level of the hierarchy ("Cancer Type" column), and the OT-108 task is developed using the second level of the hierarchy ("OncoTree Code" column), with both tasks using the same train-test folds. Gender ratio (female:male) is reported for each OncoTree code. Except for BLCA, IDC, ILC, COAD, READ, UEC, STAD, HNSC, DLBCLNOS, MEL, LUAD, LUSC, PAAD, PRAD, CSCC, SCLC, GEJ, and CLLSLL, cancer types in this task are rare cancers designated by RARECARE and NCI-SEER.

Hyper-parameter	Value
Layers	24
Heads	16
Patch size	16
FFN layer	MLP
Head activation	GELU
Embedding dimension	1024
Stochastic dropout rate	0.1
Global crop scale	0.48, 1.0
Global crop number & size	2, 224
Local crop scale	0.16, 0.48
Local crop number & size	8, 96
Max masking ratio	0.5
Min masking ratio	0.1
Gradient clipping max norm	3.0
Normalize last layer	✓
Shared head	✗
AdamW β	(0.9, 0.999)
Batch size	3072
Freeze last layer iterations	1250
Warmup iterations	12500
Warmup teacher temperature iterations	37500
High-resolution finetuning iterations	12500
Max Iterations	125000
Learning rate schedule	Cosine
Learning rate (start)	0
Learning rate (post warmup)	2e-3
Learning rate (final)	1e-6
Teacher temperature (start)	0.04
Teacher temperature (final)	0.4
Teacher momentum (start)	0.992
Teacher momentum (final)	1.000
Weight decay (start)	0.04
Weight decay (end)	0.4
Automatic mixed precision	FP16

Supplementary Data Table 5: **DINOv2 hyperparameters used for ViT-L/16 pretraining**. $4 \times 80GB$ NVIDIA A100 GPUs were used for training. Batch size refers to the total batch size across GPUs.

Hyperparameter	Value
Layers	12
Heads	12
Patch size	16
Head activation	GELU
Embedding dimension	768
Drop path rate	0.1
Global crop scale	0.32, 1.0
Global crop number	2
Local crop scale	0.05, 0.32
Local crop number	10
Partial prediction shape	Block
Partial prediction ratio	0.3
Partial prediction variance	0.2
Gradient clipping max norm	0.3
Normalize last layer	✓
Shared head	✓
AdamW β	(0.9, 0.999)
Batch size	1024
Freeze last layer epochs	3
Warmup epochs	10
Warmup teacher temperature epochs	30
Max epochs	80
Learning rate schedule	Cosine
Learning rate (start)	0
Learning rate (post warmup)	2e-3
Learning rate (final)	2e-6
Teacher temperature (start)	0.04
Teacher temperature (final)	0.4
Teacher momentum (start)	0.996
Teacher momentum (final)	1.000
Weight decay (start)	0.04
Weight decay (end)	0.4
Automatic mixed precision	fp16

Supplementary Data Table 6: **iBOT hyperparameters used for ViT-B/16 pretraining**. 4×80GB NVIDIA A100 GPUs were used for training. Batch size refers to the total batch size across GPUs.

Hyper-parameter	Value
Layers	24
Heads	16
Patch size	16
FFN layer	MLP
Head activation	GELU
Embedding dimension	1024
Stochastic dropout rate	0.1
Feature Dim	256
MLP Dim	4096
Momentum of updating encoder	0.99
Momentum with cosine schedule	✓
Softmax temperature	0.2
Stop gradients conv1	✓
Min crop scale	0.08
AdamW β	(0.9, 0.999)
Batch size	2048
Warmup iterations	20480
Total iterations	156000
Weight decay	0.1
Automatic mixed precision	FP16

Supplementary Data Table 7: **MoCoV3 hyperparameters used in ViT-L/16 pretraining**. $4 \times$ 80GB NVIDIA A100 GPUs were used for training. Batch size refers to the total batch size across GPUs.

Hyper-parameter	Value
Architecture	ResNet-50
Embedding dimension	2048
Feature Dim	256
MLP Dim	4096
Momentum of updating encoder	0.99
Momentum with cosine schedule	✓
Softmax temperature	1.0
Min crop scale	0.2
AdamW β	(0.9, 0.999)
Batch size	2048
Warmup iterations	5120
Total iterations	156000
Weight decay	1e-6
Automatic mixed precision	FP16

Supplementary Data Table 8: **MoCoV3 hyperparameters used in ResNet-50 pretraining**. $4 \times$ 80GB NVIDIA A100 GPUs were used for training. Batch size refers to the total batch size across GPUs.

Hyperparameter	Value
Batch size	1
Weight decay	1e-5
AdamW β	(0.9, 0.999)
Peak learning rate	1e-4
Learning rate schedule	Cosine
Epochs	20

Supplementary Data Table 9: **Hyperparameters used in slide-level supervised classification**. Single 24GB NVIDIA GeForce RTX 3090 GPUs were used for each ABMIL model using weakly-supervised learning and slide-level labels.

Encoder	Arch.	# Params	Inference Speed
ResNet-50$_{IN}$	ResNet-50	23M	1333.8 images/sec
CTransPath	Swin-T/14	28M	729.6 images/sec
REMEDIS	ResNet-152×2	232M	212.5 images/sec
UNI	ViT-L/16	307M	133.1 images/sec

Supplementary Data Table 10: **Architecture details of pretrained vision encoders.** We report architecture details of each encoder. Using 1 NVIDIA GeForce 24GB 3090 TI, inference speed was computed by extracting features from batches of 256×256 RGB images (batch size of 256, 32-bit floating point precision, 16 threads) using the PyTorch `ImageFolder` API (followed by `torchvision.transforms` that performed center-cropping to 256×256 resolution).

Encoder	SSL Recipe	Pretrain Source	# Patches	# Slides	Epochs	Images Seen	Batch Size	Time	Resources
CTransPath	MoCoV3	TCGA, PAIP	15.6M	29.8K	100	15.6B	1024	250H	48 V100s
REMEDIS	TCGA	BIT+SimCLR	50.0M	29.0K	1000	50.0B	4096	150H	256 TPUs
UNI	DINOv2	MGB, GTEx	100.1M	100.1K	3.85	384M	3072	32H	32 A100s

Supplementary Data Table 11: **Comparison of implementation details of pretrained vision encoders.** We show differences in self-supervised learning recipe by each pretrained encoder, such as pretraining data source (number of patches and slides) and resources spent (approximate number of epochs and hours spent during training) by the original study. Note that while UNI trained in less time, it was developed with more updated computing environments (GPU resources, CUDA version, software implementation).

Encoder	Full?	Top-1 ACC	Top-3 ACC	Top-5 ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	✓	0.336 (0.312-0.359)	0.557 (0.531-0.581)	0.674 (0.651-0.697)	0.274 (0.251-0.297)	0.869 (0.858-0.879)
CTransPath	✓	0.578 (0.554-0.602)	0.796 (0.777-0.815)	0.876 (0.859-0.891)	0.562 (0.536-0.586)	0.957 (0.950-0.963)
REMEDIS	✓	0.040 (0.030-0.049)	0.157 (0.138-0.175)	0.204 (0.185-0.224)	0.012 (0.007-0.019)	0.727 (0.716-0.736)
UNI	✓	**0.720 (0.698-0.741)**	**0.8907 (0.874-0.906)**	**0.935 (0.924-0.947)**	**0.719 (0.695-0.741)**	**0.976 (0.972-0.981)**
ResNet-50$_{IN}$	✗	0.327 (0.304-0.349)	0.552 (0.527-0.576)	0.657 (0.633-0.680)	0.258 (0.236-0.280)	0.862 (0.850-0.873)
CTransPath	✗	0.569 (0.546-0.592)	0.793 (0.773-0.812)	0.874 (0.858-0.890)	0.555 (0.530-0.578)	0.956 (0.949-0.962)
REMEDIS	✗	0.593 (0.570-0.617)	0.802 (0.783-0.822)	0.875 (0.859-0.890)	0.592 (0.567-0.615)	0.954 (0.946-0.961)
UNI	✗	**0.731 (0.709-0.752)**	**0.894 (0.879-0.910)**	**0.938 (0.926-0.949)**	**0.729 (0.706-0.750)**	**0.976 (0.971-0.981)**

Supplementary Data Table 12: **Weakly-supervised 43-class cancer type classification (OT-43) based on in-house BWH data (43 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1, top-3, and top-5 accuracy, weighted F1 score, and AUROC. We report results for both extracted features using all tissue patches per WSI, as well as extracted features from 200 representative tissue patches per WSI (described in **Online Methods**). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Data	Arch.	Top-1 ACC	Top-3 ACC	Top-5 ACC	Weighted F1	ROCAUC
Mass-1K	ViT-B	0.650 (0.638-0.638)	0.834 (0.828-0.828)	0.892 (0.897-0.897)	0.649 (0.634-0.634)	0.964 (0.966-0.966)
Mass-1K	ViT-L	0.652 (0.628-0.676)	0.835 (0.816-0.852)	0.895 (0.880-0.910)	0.653 (0.627-0.676)	0.962 (0.955-0.968)
Mass-22K	ViT-B	0.714 (0.717-0.717)	0.886 (0.886-0.886)	0.938 (0.940-0.940)	0.711 (0.713-0.713)	0.976 (0.978-0.978)
Mass-22K	ViT-L	0.694 (0.673-0.716)	0.891 (0.875-0.905)	0.931 (0.918-0.943)	0.692 (0.670-0.714)	0.969 (0.963-0.975)
Mass-100K	ViT-B	0.685 (0.700-0.700)	0.865 (0.865-0.865)	0.928 (0.932-0.932)	0.682 (0.695-0.695)	0.973 (0.975-0.975)
Mass-100K	ViT-L	**0.731 (0.709-0.752)**	**0.894 (0.879-0.910)**	**0.938 (0.926-0.949)**	**0.729 (0.706-0.750)**	**0.976 (0.972-0.981)**

Supplementary Data Table 13: **Assessing model scale of UNI (pretrained on Mass-1K, Mass-22K, and Mass-100K) on OT-43 performance**. Comparisons of weakly-supervised performance on OT-43 of UNI pretrained on different data sizes (Mass-1K, Mass-22K, Mass-100K), across different model architectures (arch.), ViT-B/16 using iBOT and ViT-L/16 using DINOv2. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1, top-3, and top-5 accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Data	Iter.	Top-1 ACC	Top-3 ACC	Top-5 ACC	Weighted F1	AUROC
Mass-1K	50K	0.640 (0.616-0.662)	0.843 (0.825-0.860)	0.908 (0.894-0.922)	0.638 (0.612-0.659)	0.968 (0.962-0.973)
Mass-1K	75K	0.646 (0.622-0.669)	0.838 (0.820-0.856)	0.898 (0.883-0.912)	0.644 (0.619-0.667)	0.965 (0.959-0.971)
Mass-1K	100K	0.649 (0.625-0.672)	0.847 (0.829-0.863)	0.905 (0.890-0.919)	0.649 (0.624-0.671)	0.964 (0.957-0.970)
Mass-1K	125K	0.652 (0.628-0.676)	0.835 (0.816-0.852)	0.895 (0.880-0.910)	0.653 (0.627-0.676)	0.962 (0.955-0.968)
Mass-22K	50K	0.707 (0.685-0.730)	0.887 (0.870-0.902)	0.930 (0.918-0.943)	0.704 (0.680-0.728)	0.974 (0.968-0.979)
Mass-22K	75K	0.708 (0.684-0.730)	0.881 (0.865-0.896)	0.930 (0.917-0.941)	0.707 (0.682-0.728)	0.972 (0.967-0.977)
Mass-22K	100K	0.708 (0.686-0.730)	0.888 (0.873-0.904)	0.938 (0.925-0.949)	0.705 (0.682-0.726)	0.973 (0.967-0.978)
Mass-22K	125K	0.694 (0.673-0.716)	0.891 (0.875-0.905)	0.931 (0.918-0.943)	0.692 (0.670-0.714)	0.969 (0.963-0.975)
Mass-100K	50K	0.694 (0.672-0.716)	0.871 (0.855-0.888)	0.927 (0.913-0.939)	0.694 (0.669-0.715)	0.974 (0.968-0.978)
Mass-100K	75K	0.704 (0.682-0.725)	0.889 (0.873-0.904)	0.935 (0.923-0.948)	0.703 (0.679-0.724)	0.977 (0.972-0.981)
Mass-100K	100K	0.720 (0.699-0.742)	0.886 (0.871-0.901)	0.931 (0.919-0.943)	0.719 (0.697-0.741)	**0.977 (0.972-0.981)**
Mass-100K	125K	**0.731 (0.709-0.752)**	**0.894 (0.879-0.910)**	**0.938 (0.926-0.949)**	**0.729 (0.706-0.750)**	0.976 (0.972-0.981)

Supplementary Data Table 14: **Assessing pretraining length of UNI (pretrained on Mass-1K, Mass-22K, and Mass-100K) on OT-43 performance.** Comparisons of weakly-supervised performance on OT-43 of UNI pretrained on different data sizes (Mass-1K, Mass-22K, Mass-100K), across different training iterations ranging from 50K to 125K. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1, top-3, and top-5 accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Full?	Top-1 ACC	Top-3 ACC	Top-5 ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	✓	0.156 (0.138-0.173)	0.309 (0.287-0.333)	0.391 (0.368-0.416)	0.115 (0.101-0.132)	0.874 (0.866-0.882)
CTransPath	✓	0.391 (0.366-0.415)	0.620 (0.597-0.645)	0.715 (0.694-0.736)	0.358 (0.333-0.383)	0.958 (0.954-0.963)
REMEDIS	✓	0.118 (0.102-0.133)	0.225 (0.205-0.244)	0.293 (0.270-0.314)	0.074 (0.060-0.087)	0.857 (0.849-0.867)
UNI	✓	**0.525 (0.501-0.549)**	**0.751 (0.728-0.772)**	**0.829 (0.812-0.847)**	**0.509 (0.483-0.534)**	**0.971 (0.967-0.975)**
ResNet-50$_{IN}$	✗	0.148 (0.130-0.164)	0.299 (0.277-0.322)	0.378 (0.354-0.403)	0.105 (0.091-0.121)	0.869 (0.860-0.877)
CTransPath	✗	0.399 (0.375-0.423)	0.625 (0.602-0.649)	0.723 (0.702-0.746)	0.365 (0.342-0.389)	0.959 (0.955-0.963)
REMEDIS	✗	0.412 (0.387-0.435)	0.654 (0.630-0.678)	0.735 (0.713-0.757)	0.398 (0.372-0.421)	0.952 (0.946-0.956)
UNI	✗	**0.538 (0.514-0.562)**	**0.759 (0.738-0.781)**	**0.843 (0.826-0.860)**	**0.522 (0.498-0.548)**	**0.972 (0.968-0.976)**

Supplementary Data Table 15: **Weakly-supervised 108-class OncoTree code cancer classification (OT-108) based on in-house BWH data (108 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1, top-3, and top-5 accuracy, weighted F1 score, and AUROC. We report results for both extracted features using all tissue patches per WSI, as well as extracted features from 200 representative tissue patches per WSI (described in **Online Methods**). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Data	Arch.	Top-1 ACC	Top-3 ACC	Top-5 ACC	Weighted F1	ROCAUC
Mass-1K	ViT-B	0.453 (0.456-0.456)	0.663 (0.667-0.667)	0.746 (0.743-0.743)	0.436 (0.442-0.442)	0.961 (0.961-0.961)
Mass-1K	ViT-L	0.473 (0.449-0.499)	0.673 (0.652-0.698)	0.757 (0.738-0.780)	0.455 (0.430-0.482)	0.961 (0.956-0.965)
Mass-22K	ViT-B	0.522 (0.514-0.514)	0.754 (0.747-0.747)	0.833 (0.831-0.831)	0.503 (0.495-0.495)	**0.975 (0.976-0.976)**
Mass-22K	ViT-L	0.508 (0.485-0.531)	0.740 (0.719-0.761)	0.825 (0.807-0.842)	0.486 (0.461-0.510)	0.972 (0.968-0.975)
Mass-100K	ViT-B	0.503 (0.507-0.507)	0.729 (0.715-0.715)	0.807 (0.793-0.793)	0.483 (0.490-0.490)	0.973 (0.972-0.972)
Mass-100K	ViT-L	**0.538 (0.514-0.562)**	**0.759 (0.738-0.781)**	**0.843 (0.826-0.860)**	**0.522 (0.498-0.548)**	0.972 (0.968-0.976)

Supplementary Data Table 16: **Assessing model scale of UNI (pretrained on Mass-1K, Mass-22K, and Mass-100K) on OT-108 performance**. Comparisons of weakly-supervised performance on OT-108 of UNI pretrained on different data sizes (Mass-1K, Mass-22K, Mass-100K), across different model architectures (arch.), ViT-B/16 using iBOT and ViT-L/16 using DINOv2. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1, top-3, and top-5 accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Data Size	Iter	Top-1 ACC	Top-3 ACC	Top-5 ACC	Weighted F1	AUROC
Mass-1K	50K	0.459 (0.435-0.486)	0.679 (0.657-0.702)	0.767 (0.748-0.787)	0.436 (0.411-0.462)	0.968 (0.964-0.971)
Mass-1K	75K	0.464 (0.440-0.488)	0.681 (0.660-0.706)	0.757 (0.736-0.778)	0.440 (0.416-0.466)	0.967 (0.963-0.971)
Mass-1K	100K	0.458 (0.434-0.482)	0.675 (0.654-0.699)	0.760 (0.740-0.781)	0.436 (0.412-0.461)	0.964 (0.960-0.968)
Mass-1K	125K	0.473 (0.449-0.499)	0.673 (0.652-0.698)	0.757 (0.738-0.780)	0.455 (0.430-0.482)	0.961 (0.956-0.965)
Mass-22K	50K	0.511 (0.487-0.536)	0.752 (0.732-0.775)	0.830 (0.812-0.848)	0.493 (0.467-0.519)	**0.976 (0.972-0.979)**
Mass-22K	75K	0.522 (0.498-0.549)	0.748 (0.727-0.770)	0.840 (0.821-0.857)	0.498 (0.474-0.524)	0.976 (0.972-0.979)
Mass-22K	100K	0.512 (0.489-0.538)	0.747 (0.726-0.770)	0.830 (0.811-0.848)	0.491 (0.465-0.517)	0.974 (0.971-0.977)
Mass-22K	125K	0.508 (0.485-0.531)	0.740 (0.719-0.761)	0.825 (0.807-0.842)	0.486 (0.461-0.510)	0.972 (0.968-0.975)
Mass-100K	50K	0.511 (0.486-0.536)	0.741 (0.720-0.764)	0.827 (0.809-0.846)	0.491 (0.465-0.516)	0.974 (0.971-0.977)
Mass-100K	75K	0.523 (0.499-0.548)	0.762 (0.741-0.782)	0.837 (0.820-0.854)	0.507 (0.481-0.533)	0.975 (0.972-0.979)
Mass-100K	100K	0.533 (0.508-0.559)	**0.768 (0.748-0.788)**	0.841 (0.823-0.859)	0.516 (0.491-0.542)	0.975 (0.971-0.978)
Mass-100K	125K	**0.538 (0.514-0.562)**	0.759 (0.738-0.781)	**0.843 (0.826-0.860)**	**0.522 (0.498-0.548)**	0.972 (0.968-0.976)

Supplementary Data Table 17: **Assessing pretraining length of UNI (pretrained on Mass-1K, Mass-22K, and Mass-100K) on OT-108 performance.** Comparisons of weakly-supervised performance on OT-108 of UNI pretrained on different data sizes (Mass-1K, Mass-22K, Mass-100K), across different training iterations ranging from 50K to 125K. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1, top-3, and top-5 accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

SSL Method	Arch.	OT-43		OT-108	
		Top-1 ACC	AUROC	Top-1 ACC	AUROC
MoCoV3	ResNet-50	0.600 (0.575-0.623)	0.950 (0.942-0.958)	0.399 (0.376-0.423)	0.957 (0.952-0.962))
MoCoV3	ViT-L	0.499 (0.503-0.503)	0.934 (0.931-0.931)	0.273 (0.283-0.283)	0.940 (0.937-0.937)
DINOV2	ViT-L	**0.652 (0.628-0.676)**	**0.962 (0.955-0.968)**	**0.473 (0.449-0.499)**	**0.961 (0.956-0.965)**

Supplementary Data Table 18: **Assessing impact of SSL pretraining algorithm on OT-43 and OT-108 performance**. Comparing weakly-supervised performance of MoCoV3 and DINOv2 on OT-43 and OT-108 with Mass-1K pretraining. For MoCoV3, we used both the ResNet-50 and ViT-L architecture. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on curated train-test folds (3,944:1,620), with test performance ($n = 1620$ slides) reported using top-1 and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Cancer Subtype	Cancer Diagnosis	Num. Slides
Benign	Normal	44
	Pathological Benign	147
	Usual Ductal Hyperplasia	74
Atypical	Atypical Ductal Hyperplasia	48
	Flat Epithelial Atypia	41
Malignant	Ductal Carcinoma in Situ	61
	Invasive Carcinoma	132

Supplementary Data Table 19: **Hierarchical label distribution of BRCA Coarse- and Fine-Grained Subtyping based on BRACS.** The coarse-grained BRCA subtyping task is developed using the first level of the hierarchy ("Cancer Subtype" column), and the fine-grained BRCA subtyping task is developed using the second level of the hierarchy ("Cancer Diagnosis" column), with both tasks using the official train-validation-test folds of the BRACS dataset.

Cancer Subtype	Cancer Diagnosis	Num. Slides
Adult-type diffuse gliomas	Anaplastic astrocytoma, *IDH1*-mutant	47
	Anaplastic astrocytoma, *IDH1*-wildtype	47
	Anaplastic oligodendroglioma, *IDH1*-mutant & 1p/19q codeleted	91
	Glioblastoma, *IDH1*-mutant	34
	Glioblastoma, *IDH1*-wildtype	474
	Gliosarcoma	59
	Oligodendroglioma, *IDH1*-mutant & 1p/19q codeleted	85
Circumscribed astrocytic gliomas	Pilocytic astrocytoma	173
Cranial and paraspinal nerve tumours	Schwannoma	81
Embryonal Tumors	Medulloblastoma, non-WNT/non-SHH	32
Ependymal Tumours	Anaplastic ependymoma	50
	Ependymoma	46
Glioneuronal and neuronal tumours	Ganglioglioma	88
Haematolymphoid tumours involving the CNS	Diffuse large B-cell lymphoma of the CNS	59
	Langerhans cell histiocytosis	32
Meningiomas	Anaplastic meningioma	46
	Angiomatous meningioma	31
	Atypical meningioma	83
	Fibrous meningioma	57
	Meningothelial meningioma	104
	Secretory meningioma	41
	Transitional meningioma	68
Mesenchymal, non-meningothelial tumours involving the CNS	Haemangioblastoma	88
	Haemangioma	30
	Haemangiopericytoma	34
	Lipoma	38
Metastatic tumours	Metastatic tumours	47
Paediatric-type diffuse low-grade gliomas	Diffuse astrocytoma, *IDH1*-mutant	70
Tumours of the sellar region	Adamantinomatous craniopharyngioma	85
	Pituitary adenoma	99

Supplementary Data Table 20: **Hierarchical label distribution of brain tumor coarse- and fine-grained subtyping based on EBRAINS.** The coarse-grained brain tumor subtyping task is developed using the first level of the hierarchy ("Cancer Subtype" column), and the fine-grained brain tumor subtyping task is developed using the second level of the hierarchy ("Cancer Diagnosis" column), with both tasks using same train-validation-test folds created using the EBRAINS Digital Tumor Atlas. All brain tumors in these tasks are designated as rare cancers by RARECARE and NCI-SEER.

IDH1 status	Histomolecular Subtype	Num. Slides
IDH1-mutant	Astrocytoma, IDH1-mutant	257
	Glioblastoma, IDH1-mutant	93
	Oligodendroglioma, IDH1-mutant & 1p/19q codeleted	408
IDH1-wildtype	Astrocytoma, IDH1-wildtype	144
	Glioblastoma, IDH1-wildtype	1094

Supplementary Data Table 21: **Hierarchical label distribution of Glioma *IDH1* Mutation Prediction and Histomolecular Subtyping based on TCGA and EBRAINS.** The glioma *IDH1* mutation prediction task is developed using the first level of the hierarchy ("*IDH1* status" column), and the glioma histomolecular subtyping task is developed using the second level of the hierarchy ("Histomolecular Subtype" column), with both tasks using same train-validation-test folds created using the TCGA, with external evaluation on the EBRAINS Digital Tumor Atlas. All brain tumors in these tasks are designated as rare cancers by RARECARE and NCI-SEER.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.726 (0.652-0.797)	0.765 (0.678-0.837)	0.752 (0.659-0.842)
CTransPath	0.897 (0.836-0.948)	0.907 (0.851-0.953)	0.930 (0.864-0.985)
REMEDIS	0.930 (0.884-0.969)	0.923 (0.877-0.962)	**0.981 (0.947-1.000)**
UNI	**0.957 (0.911-0.991)**	**0.961 (0.922-0.992)**	0.975 (0.937-1.000)

Supplementary Data Table 22: **Weakly-supervised breast metastasis detection based on CAMELYON16 (2 classes).** Pre-extracted patch features of each encoder with ABMIL were trained a custom train-validation split (90:10 ratio) of the official train set (270 slides) and tested on the official test set, with test performance ($n = 129$) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	TCGA	0.857 (0.787-0.929)	0.857 (0.786-0.929)	0.939 (0.893-0.978)
CTransPath	TCGA	0.949 (0.901-0.990)	0.949 (0.898-0.990)	0.993 (0.982-1.000)
REMEDIS	TCGA	0.949 (0.902-0.990)	0.949 (0.898-0.990)	0.985 (0.953-1.000)
UNI	TCGA	**0.969 (0.932-1.000)**	**0.969 (0.929-1.000)**	**0.997 (0.989-1.000)**
ResNet-50$_{\text{IN}}$	CPTAC	0.852 (0.831-0.871)	0.849 (0.829-0.869)	0.924 (0.907-0.939)
CTransPath	CPTAC	0.884 (0.866-0.903)	0.881 (0.862-0.900)	0.953 (0.942-0.963)
REMEDIS	CPTAC	0.854 (0.833-0.873)	0.849 (0.828-0.869)	0.951 (0.938-0.962)
UNI	CPTAC	**0.889 (0.870-0.908)**	**0.891 (0.872-0.909)**	**0.958 (0.946-0.969)**

Supplementary Data Table 23: **Weakly-supervised NSCLC subtyping based on TCGA and CPTAC (2 classes).** Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the TCGA-NSCLC cohort (label-stratified into train-validation-test folds with 80:10:10 ratio, 848:97:98 slides), with external evaluation on slides ($n = 1091$) sourced from CPTAC-LUAD ($n = 578$) and CPTAC-LUSC ($n = 513$). Test performance was reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	TCGA	0.849 (0.734-0.940)	0.877 (0.814-0.938)	0.977 (0.954-0.993)
CTransPath	TCGA	0.892 (0.800-0.969)	0.907 (0.847-0.959)	0.987 (0.970-0.998)
REMEDIS	TCGA	**0.973 (0.937-1.000)**	**0.969 (0.927-1.000)**	**0.997 (0.992-1.000)**
UNI	TCGA	0.947 (0.875-0.994)	0.959 (0.918-0.990)	0.994 (0.984-1.000)
ResNet-50$_{\text{IN}}$	CPTAC-DHMC	0.824 (0.751-0.894)	0.903 (0.886-0.921)	0.972 (0.957-0.984)
CTransPath	CPTAC-DHMC	0.939 (0.888-0.984)	0.968 (0.957-0.979)	**0.996 (0.993-0.998)**
REMEDIS	CPTAC-DHMC	0.790 (0.725-0.858)	0.934 (0.919-0.949)	0.988 (0.982-0.993)
UNI	CPTAC-DHMC	**0.963 (0.930-0.987)**	**0.971 (0.960-0.981)**	0.993 (0.985-0.998)

Supplementary Data Table 24: **Weakly-supervised RCC subtyping based on TCGA and CPTAC-DHMC (3 classes).** Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the TCGA-RCC cohort (label-stratified into train-validation-test folds with 80:10:10 ratio, 736:89:97 slides), with external evaluation on slides ($n = 872$) sourced from CPTAC-CCRCC ($n = 404$) and DHMC-Kidney ($n = 468$) (CCRCC, CHRCC, and PRCC cases only). Test performance was reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.346 (0.306-0.386)	0.478 (0.383-0.565)	0.818 (0.785-0.850)
CTransPath	0.804 (0.743-0.879)	0.883 (0.823-0.937)	0.987 (0.976-0.995)
REMEDIS	0.865 (0.789-0.929)	0.877 (0.818-0.929)	0.973 (0.953-0.988)
UNI	**0.919 (0.852-0.968)**	**0.926 (0.882-0.966)**	**0.993 (0.983-0.999)**

Supplementary Data Table 25: **Weakly-supervised RCC subtyping based on DHMC (5 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on a modified configuration of the official train-validation-test folds (70:4:26 ratio, 393:23:147 slides), with test performance ($n = 147$ slides) reported using balanced accuracy, weighted F1 score, and AUROC. Since no CHRCC slides were included in the validation fold, 8 CHRCC slides from the training fold was moved to the valdiation fold. The test fold was unmodified. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.250 (0.250-0.250)	0.356 (0.209-0.525)	0.671 (0.594-0.754)
CTransPath	0.556 (0.454-0.651)	0.728 (0.603-0.841)	0.845 (0.779-0.898)
REMEDIS	0.604 (0.499-0.696)	0.787 (0.670-0.889)	0.888 (0.814-0.960)
UNI	**0.643 (0.549-0.725)**	**0.824 (0.706-0.931)**	**0.957 (0.908-0.987)**

Supplementary Data Table 26: **Weakly-supervised CRC screening based on HunCRC (4 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the HunCRC cohort (label-stratified into train-validation-test folds with 50:25:25 ratio, 100:50:50 slides). Test performance was reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.552 (0.490-0.607)	0.516 (0.396-0.629)	0.748 (0.675-0.816)
CTransPath	0.639 (0.554-0.728)	0.648 (0.536-0.763)	0.840 (0.776-0.903)
REMEDIS	0.676 (0.578-0.770)	**0.696 (0.591-0.796)**	0.864 (0.805-0.915)
UNI	**0.687 (0.617-0.760)**	0.691 (0.562-0.800)	**0.887 (0.833-0.936)**

Supplementary Data Table 27: **Weakly-supervised BRCA coarse-grained subtyping based on BRACS (3 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the official train-validation-test folds (72:12:16 ratio, 395:65:87 slides), with test performance ($n = 87$ slides) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.248 (0.214-0.276)	0.216 (0.132-0.311)	0.701 (0.645-0.754)
CTransPath	0.360 (0.284-0.440)	0.377 (0.280-0.485)	0.761 (0.706-0.809)
REMEDIS	0.398 (0.307-0.483)	0.428 (0.320-0.530)	0.749 (0.689-0.804)
UNI	**0.468 (0.393-0.550)**	**0.486 (0.366-0.592)**	**0.837 (0.794-0.879)**

Supplementary Data Table 28: **Weakly-supervised BRCA fine-grained subtyping based on BRACS (7 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the official train-validation-test (72:12:16 ratio, 395:65:87 slides), with test performance ($n = 87$ slides) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Quad. Weighted κ	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.574 (0.544-0.603)	0.831 (0.799-0.859)	0.631 (0.599-0.661)	0.885 (0.873-0.897)
CTransPath	0.691 (0.658-0.723)	0.927 (0.912-0.940)	0.752 (0.723-0.779)	0.938 (0.929-0.947)
REMEDIS	0.711 (0.679-0.742)	0.932 (0.918-0.945)	0.766 (0.737-0.794)	0.941 (0.931-0.949)
UNI	**0.757 (0.726-0.785)**	**0.946 (0.933-0.957)**	**0.809 (0.783-0.834)**	**0.956 (0.947-0.963)**

Supplementary Data Table 29: **Weakly-supervised ISUP grading based on PANDA (6 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on label-stratified train-validation-test folds (80:10:10 ratio, 7,647:954:954 slides), with test performance (954 slides) reported using balanced accuracy, Cohen's quadratic weighted κ, weighted F1 core, and ROC AUC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Balanced ACC	Quad. Weighted κ	Weighted F1	AUROC
ResNet-50$_{IN}$	Karolinska	0.499 (0.466-0.531)	0.782 (0.737-0.821)	0.652 (0.620-0.682)	0.846 (0.830-0.861)
CTransPath	Karolinska	0.675 (0.640-0.714)	0.914 (0.889-0.935)	0.809 (0.784-0.833)	0.923 (0.911-0.935)
REMEDIS	Karolinska	**0.688 (0.653-0.723)**	0.929 (0.910-0.944)	**0.814 (0.788-0.839)**	0.928 (0.916-0.939)
UNI	Karolinska	0.664 (0.630-0.697)	**0.935 (0.920-0.946)**	0.807 (0.781-0.833)	**0.947 (0.937-0.956)**
ResNet-50$_{IN}$	Radboud	0.574 (0.545-0.603)	0.805 (0.768-0.840)	0.579 (0.542-0.614)	0.886 (0.876-0.897)
CTransPath	Radboud	0.722 (0.695-0.749)	0.913 (0.894-0.929)	0.731 (0.701-0.759)	0.946 (0.938-0.954)
REMEDIS	Radboud	0.722 (0.695-0.750)	0.918 (0.899-0.933)	0.729 (0.700-0.758)	0.944 (0.937-0.952)
UNI	Radboud	**0.783 (0.757-0.809)**	**0.933 (0.919-0.947)**	**0.791 (0.764-0.817)**	**0.967 (0.962-0.972)**

Supplementary Data Table 30: **Weakly-supervised ISUP grading based on PANDA (6 classes) using the same splits of public MIL baselines [90]**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on label-stratified train-validation-test folds (60:20:20 ratio, 5,736:1,913:(1,026+877) slides) created by WholeSight [90], with test performance on Karolinska cohort (1,026 slides) and Radboud cohort (877 slides) reported separately using balanced accuracy, Cohen's quadratic weighted κ, weighted F1 core, and ROC AUC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	TCGA	0.768 (0.710-0.821)	0.810 (0.767-0.850)	0.864 (0.825-0.901)
CTransPath	TCGA	**0.891 (0.859-0.920)**	**0.862 (0.828-0.894)**	**0.958 (0.938-0.975)**
REMEDIS	TCGA	0.819 (0.770-0.869)	0.860 (0.824-0.897)	0.933 (0.907-0.956)
UNI	TCGA	0.808 (0.760-0.855)	0.831 (0.794-0.867)	0.925 (0.898-0.949)
ResNet-50$_{IN}$	EBRAINS	0.759 (0.731-0.786)	0.777 (0.749-0.804)	0.822 (0.794-0.849)
CTransPath	EBRAINS	0.836 (0.814-0.861)	0.819 (0.795-0.846)	0.924 (0.905-0.941)
REMEDIS	EBRAINS	0.792 (0.765-0.818)	0.822 (0.795-0.850)	0.895 (0.873-0.916)
UNI	EBRAINS	**0.856 (0.836-0.878)**	**0.841 (0.820-0.864)**	**0.941 (0.925-0.955)**

Supplementary Data Table 31: **Weakly-supervised GBMLGG *IDH1* mutation prediction based on TCGA and EBRAINS (2 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the TCGA-GBMLGG cohort (site-stratified into train-validation-test folds with 47:22:33 ratio, 525:243:355 slides), with external evaluation on slides ($n = 873$) sourced from the EBRAINS Digital Tumor Atlas (using slides with available *IDH1* status). Test performance was reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	TCGA	0.412 (0.362-0.462)	0.660 (0.611-0.715)	0.814 (0.783-0.847)
CTransPath	TCGA	0.487 (0.437-0.543)	0.718 (0.671-0.769)	0.866 (0.831-0.905)
REMEDIS	TCGA	0.415 (0.382-0.448)	0.675 (0.628-0.727)	0.837 (0.811-0.864)
UNI	TCGA	**0.673 (0.616-0.732)**	**0.794 (0.751-0.838)**	**0.910 (0.880-0.939)**
ResNet-50$_{IN}$	EBRAINS	0.402 (0.379-0.423)	0.615 (0.581-0.648)	0.684 (0.654-0.713)
CTransPath	EBRAINS	0.498 (0.472-0.523)	0.705 (0.674-0.735)	0.823 (0.798-0.845)
REMEDIS	EBRAINS	0.337 (0.320-0.354)	0.551 (0.517-0.586)	0.751 (0.723-0.777)
UNI	EBRAINS	**0.562 (0.535-0.594)**	**0.728 (0.702-0.755)**	**0.868 (0.848-0.887)**

Supplementary Data Table 32: **Weakly-supervised GBMLGG histomolecular subtyping based on TCGA and EBRAINS (5 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the TCGA-GBMLGG cohort (site-stratified into train-validation-test folds with 47:22:33 ratio, 525:243:355 slides), with external evaluation on slides ($n = 873$) sourced from the EBRAINS Digital Tumor Atlas (using slides with available *IDH1* status). Test performance was reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.302 (0.273-0.334)	0.556 (0.509-0.600)	0.878 (0.860-0.896)
CTransPath	0.666 (0.608-0.722)	0.795 (0.758-0.829)	0.968 (0.957-0.978)
REMEDIS	0.687 (0.638-0.734)	0.789 (0.753-0.824)	0.967 (0.952-0.979)
UNI	**0.883 (0.838-0.924)**	**0.926 (0.902-0.947)**	**0.996 (0.994-0.998)**

Supplementary Data Table 33: **Weakly-supervised coarse-grained brain tumor subtyping based on EBRAINS (12 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the EBRAINS Digital Tumor Atlas (label-stratified into train-validation-test folds with 50:25:25 ratio, 1,151:595:573 slides), with test performance reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.219 (0.195-0.241)	0.300 (0.259-0.339)	0.893 (0.879-0.907)
CTransPath	0.514 (0.473-0.559)	0.597 (0.548-0.638)	0.959 (0.950-0.967)
REMEDIS	0.382 (0.346-0.415)	0.471 (0.428-0.512)	0.917 (0.901-0.931)
UNI	**0.675 (0.633-0.715)**	**0.746 (0.704-0.783)**	**0.976 (0.969-0.982)**

Supplementary Data Table 34: **Weakly-supervised fine-grained brain tumor subtyping based on EBRAINS (30 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on the EBRAINS Digital Tumor Atlas (label-stratified into train-validation-test folds with 50:25:25 ratio, 1151:595:573 slides), with test performance reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.861 (0.822-0.898)	0.861 (0.822-0.898)	0.930 (0.903-0.955)
CTransPath	0.882 (0.847-0.916)	0.882 (0.846-0.916)	0.937 (0.909-0.961)
REMEDIS	0.861 (0.823-0.895)	0.861 (0.822-0.895)	0.933 (0.908-0.958)
UNI	**0.896 (0.864-0.927)**	**0.894 (0.862-0.925)**	**0.962 (0.944-0.979)**

Supplementary Data Table 35: **Weakly-supervised cellular-mediated allograft rejection of endomyocardial biopsies based on in-house BWH data (2 classes)**. Pre-extracted patch features of each encoder with ABMIL were trained and evaluated on case- and label-stratified train-validation-test folds (3,547:484:990 slides, 1,192:164:332 patients), with test performance (332 patients) reported using balanced accuracy, weighted F1 core, and ROC AUC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Method	Encoder	Mag.	Reported By	ACC	AUROC
Avg. Pathologist (w/ TC)[†]	-	-	Bejnordi *et al.* 2017 [78]	-	0.810
Avg. Pathologist (w/o TC)[†]	-	-	Bejnordi *et al.* 2017 [78]	-	0.966
Full Sup. CNN-RF[†]	GoogLeNet	20×	Wang *et al.*2017 [176] [78]	-	0.923
Full Sup. CNN-RF[†]	ResNet-101	20×	Bejnordi *et al.* 2017 [78]	-	0.976
Full Sup. CNN-RF[†]	GoogLeNet	20×	Bejnordi *et al.* 2017 [78]	-	**0.994**
Full Sup. CNN ensemble	-	Multi	Koohbanani *et al.* 2019 [177]	-	0.990
Full Sup. CNN-RF[†,#]	ResNet-34	20×	Campanella *et al.* 2019 [81]	-	0.930
NIC	BiGAN	20×	Tellez *et al.* 2019 [51]	-	0.704
E2E MIL-RNN[‡,#]	ResNet-34	20×	Campanella *et al.* 2019 [81]	-	0.899
CLAM	ResNet-50$_{IN}$	40×	Lu *et al.* 2021 [15]	-	0.936
DSMIL-LC	ResNet-50$_{SimCLR}$	Multi	Li *et al.* 2021 [42]	0.899	0.917
TransMIL	ResNet-50$_{IN}$	20×	Shao *et al.* 2021 [80]	0.884	0.931
Additive-MIL	ShuffleNet$_{IN}$	10×	Javed *et al.* 2022 [178]	0.830	0.846
DTFD-MIL (AFS)	ResNet-50$_{IN}$	20×	Zhang *et al.* 2022 [179]	0.908	0.946
SCL-WC	CTransPath	10×	Wang *et al.* 2022 [112]	0.914	0.957
Bayes-MIL (APCRF)*	ResNet-50$_{IN}$	20×	Cui *et al.* 2023 [180]	0.900	0.948
ILRA-MIL (LRC)	ResNet-50$_{IN}$	10×	Xiang *et al.* 2023 [96]	0.922	0.965
TransMIL (FT)	ResNet-50$_{IN}$	20×	Li *et al.* 2023 [57]	-	0.967
TransMIL (IBMIL)	CTransPath	20×	Lin *et al.* 2023 [128]	**0.961**	0.970
DSMIL (MHIM-MIL) *	ResNet-50$_{IN}$	20×	Tang *et al.* 2023 [181]	0.925	0.965
ABMIL	UNI	20×	Ours	**0.961**	**0.975**

Supplementary Data Table 36: **A comparison of retrospective and current results for metastasis detection in CAMELYON16.** We compare UNI (results from **Extended Data Table 22**, same splits) with the leaderboard taken at the time of the CAMELYON16 challenge (**Top**), as well as a chronological timeline of best-performing models reported in recent CPath and computer vision literature that evaluate CAMELYON16 (**Middle**). In surveying the literature, our reported UNI performance achieves the best performance for a weakly-supervised MIL model (without considering further improvements to itself), and is one of the few MIL results that outperform the average pathologist performance without time constraints. In comparing against other MIL results (**Bottom**), we note that many comparisons are not equivalent due to being developed with ResNet-50$_{IN}$ or other encoders, as well as differences in MIL architecture and training hyper-parameters. † = Used pixel-level annotations. ‡ = MIL-based and trained end-to-end. # = Trained on an external cohort, with C16 test set used as independent test set. * = Did not evaluate on the official test set.

Method	Encoder	Cohort	Quad. Weighted κ
ABMIL	ResNet-50$_{IN}$	Karolinska	0.773
CLAM	ResNet-50$_{IN}$	Karolinska	0.828
TransMIL	ResNet-50$_{IN}$	Karolinska	0.761
Additive-MIL [178]	ResNet-50$_{IN}$	Karolinska	0.786
NAGCN [182]	ResNet-50$_{IN}$	Karolinska	0.655
GraphTransformer [183]	ResNet-50$_{IN}$	Karolinska	0.813
WholeSight [90]	ResNet-50$_{IN}$	Karolinska	0.813
ABMIL	UNI	Karolinska	**0.935**
ABMIL	ResNet-50$_{IN}$	Radboud	0.738
CLAM	ResNet-50$_{IN}$	Radboud	0.786
TransMIL	ResNet-50$_{IN}$	Radboud	0.732
Additive-MIL [178]	ResNet-50$_{IN}$	Radboud	0.760
NAGCN [182]	ResNet-50$_{IN}$	Radboud	0.762
GraphTransformer [183]	ResNet-50$_{IN}$	Radboud	0.818
WholeSight [90]	ResNet-50$_{IN}$	Radboud	0.829
ABMIL	UNI	Radboud	**0.933**

Supplementary Data Table 37: **Comparison with existing baselines for weakly-supervised ISUP grading based on PANDA (6 classes).** We compare UNI (results from **Extended Data Table 30**, same splits) with existing baselines on this task using the same splits as WholeSIGHT [90]. All of the baselines developed in WholeSIGHT used ResNet-50IN features, with evaluation performed on the Karolinska and Radboud cohort separately.

Method	Encoder	ACC	AUROC
CRANE [86]	ResNet-50$_{IN}$	0.893	0.958
ABMIL	UNI	**0.895**	**0.962**

Supplementary Data Table 38: **Comparison with existing baselines for cellular-mediated allograft rejection from BWH-EMB (2 classes).** We compare UNI (results from **Extended Data Table 35**, same splits) with the original baseline on this task, CRANE [86], which used ResNet-50$_{IN}$ features.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_\text{IN}$	0.715 (0.705-0.724)	0.771 (0.761-0.781)	0.958 (0.954-0.961)
CTransPath	0.845 (0.836-0.853)	0.867 (0.859-0.875)	**0.991 (0.990-0.992)**
REMEDIS	0.787 (0.776-0.796)	0.802 (0.793-0.811)	0.980 (0.979-0.982)
UNI	**0.874 (0.866-0.881)**	**0.875 (0.868-0.882)**	0.990 (0.988-0.991)

Supplementary Data Table 39: **Linear probe evaluation for CRC tissue classification based on CRC-100K (9 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on the official train-test folds (100,000:7,180), with test performance (n=7,180 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_\text{IN}$	0.686 (0.676-0.696)	0.698 (0.686-0.708)	0.797 (0.787-0.806)	0.833 (0.825-0.841)
CTransPath	0.815 (0.806-0.824)	0.845 (0.836-0.854)	0.836 (0.828-0.843)	0.848 (0.840-0.857)
REMEDIS	0.855 (0.848-0.863)	0.881 (0.873-0.888)	0.908 (0.901-0.915)	0.924 (0.918-0.931)
UNI	**0.899 (0.892-0.905)**	**0.912 (0.906-0.919)**	**0.924 (0.917-0.931)**	**0.945 (0.940-0.950)**

Supplementary Data Table 40: **SimpleShot and KNN evaluation for CRC tissue classification based on CRC-100K (9 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on the official train-test folds (100,000:7,180 ROIs), with test performance (n=7,180 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Balanced ACC	Weighted F1	AUROC
ResNet-50$_\text{IN}$	Val$_\text{OD}$	0.832 (0.828-0.836)	0.830 (0.826-0.834)	0.936 (0.934-0.938)
CTransPath	Val$_\text{OD}$	0.917 (0.914-0.920)	0.917 (0.914-0.919)	0.983 (0.982-0.984)
REMEDIS	Val$_\text{OD}$	0.959 (0.957-0.961)	0.959 (0.957-0.961)	0.992 (0.992-0.993)
UNI	Val$_\text{OD}$	**0.974 (0.973-0.976)**	**0.974 (0.973-0.976)**	**0.996 (0.995-0.997)**
ResNet-50$_\text{IN}$	Test$_\text{OD}$	0.906 (0.904-0.908)	0.906 (0.904-0.908)	0.964 (0.963-0.965)
CTransPath	Test$_\text{OD}$	0.960 (0.959-0.961)	0.960 (0.959-0.961)	0.991 (0.990-0.992)
REMEDIS	Test$_\text{OD}$	0.926 (0.925-0.928)	0.926 (0.924-0.928)	0.983 (0.983-0.984)
UNI	Test$_\text{OD}$	**0.983 (0.982-0.984)**	**0.983 (0.982-0.984)**	**0.996 (0.996-0.997)**

Supplementary Data Table 41: **Linear probe evaluation for pan-cancer tissue classification based on CAMELYON17-WILDS (2 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on the WILDS train-validation-test folds [164]. The training set contains 302,436 patches from 3 hospitals (50:50 ratio), and the model is evaluated on two out-of-domain validation (Val$_\text{OD}$) and test (Test$_\text{OD}$) folds containing 34,904 patches (50:50 ratio) and 85,054 patches (50:50 ratio) collected from two other hospitals respectively. The in-domain validation fold was not combined with the training set or used for hyper-parameter tuning. The performance is reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_\text{IN}$	Val$_\text{OD}$	0.727 (0.723-0.732)	0.721 (0.717-0.726)	0.809 (0.805-0.813)	0.804 (0.800-0.808)
CTransPath	Val$_\text{OD}$	0.797 (0.794-0.801)	0.789 (0.785-0.794)	0.873 (0.870-0.876)	0.872 (0.868-0.875)
REMEDIS	Val$_\text{OD}$	0.830 (0.826-0.834)	0.827 (0.823-0.831)	0.883 (0.879-0.886)	0.882 (0.878-0.885)
UNI	Val$_\text{OD}$	**0.904 (0.902-0.907)**	**0.904 (0.900-0.907)**	**0.968 (0.966-0.970)**	**0.968 (0.966-0.970)**
ResNet-50$_\text{IN}$	Test$_\text{OD}$	0.725 (0.722-0.728)	0.716 (0.713-0.719)	0.867 (0.865-0.869)	0.865 (0.863-0.867)
CTransPath	Test$_\text{OD}$	0.920 (0.918-0.922)	0.919 (0.918-0.921)	0.917 (0.915-0.919)	0.917 (0.915-0.919)
REMEDIS	Test$_\text{OD}$	**0.958 (0.956-0.959)**	**0.958 (0.956-0.959)**	0.939 (0.938-0.941)	0.939 (0.937-0.941)
UNI	Test$_\text{OD}$	0.955 (0.954 0.957)	0.955 (0.954-0.957)	**0.972 (0.971-0.973)**	**0.972 (0.971-0.973)**

Supplementary Data Table 42: **SimpleShot and KNN evaluation for breast metastasis detection based on CAMELYON17-WILDS (2 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on the WILDS train-validation-test folds [164]. The training set contains 302,436 patches from 3 hospitals (50:50 ratio), and the model is evaluated on two out-of-domain (OD) validation and test folds containing 34,904 patches (50:50 ratio) and 85,054 patches (50:50 ratio) collected from two other hospitals respectively. The in-domain validation fold was not combined with the training set or used for hyper-parameter tuning. The performance is reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	0.770 (0.761-0.779)	0.733 (0.722-0.745)	0.937 (0.932-0.941)
CTransPath	0.802 (0.791-0.812)	0.783 (0.773-0.793)	0.939 (0.934-0.943)
REMEDIS	0.729 (0.717-0.740)	0.744 (0.733-0.754)	0.906 (0.899-0.912)
UNI	**0.890 (0.883-0.898)**	**0.880 (0.871-0.888)**	**0.979 (0.977-0.981)**

Supplementary Data Table 43: **Linear probe evaluation for CCRCC tissue classification based on TCGA and HEL (3 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on the TCGA cohort as the train fold and HEL cohort as the test fold (89:11 ratio, 24,201:2,968 ROIs), with test performance (n=2,968 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{\text{IN}}$	0.765 (0.755-0.775)	0.741 (0.730-0.751)	0.680 (0.670-0.690)	0.641 (0.629-0.653)
CTransPath	0.679 (0.667-0.691)	0.653 (0.639-0.665)	0.697 (0.686-0.708)	0.679 (0.668-0.690)
REMEDIS	0.785 (0.775-0.795)	0.788 (0.778-0.798)	0.833 (0.823-0.843)	0.827 (0.816-0.836)
UNI	**0.801 (0.790-0.810)**	**0.801 (0.791-0.811)**	**0.866 (0.857-0.875)**	**0.860 (0.851-0.868)**

Supplementary Data Table 44: **SimpleShot and KNN evaluation for CCRCC tissue classification based on TCGA and HEL (3 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on the TCGA cohort as the train fold and HEL cohort as the test fold (89:11 ratio, 24,201:2,968 ROIs), with test performance (n=2,968 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Img Res.	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	224	0.738 (0.634-0.824)	0.733 (0.630-0.828)	0.911 (0.854-0.954)
CTransPath	224	0.875 (0.804-0.939)	0.872 (0.796-0.937)	0.982 (0.967-0.994)
REMEDIS	224	0.863 (0.782-0.933)	0.864 (0.786-0.937)	0.981 (0.959-0.997)
UNI	224	**0.925 (0.856-0.976)**	**0.926 (0.866-0.975)**	**0.994 (0.983-1.000)**
ResNet-50$_{\text{IN}}$	448	0.712 (0.610-0.810)	0.714 (0.610-0.815)	0.901 (0.841-0.951)
CTransPath	448	0.863 (0.785-0.933)	0.864 (0.790-0.937)	0.973 (0.946-0.992)
REMEDIS	448	0.913 (0.845-0.970)	0.913 (0.842-0.964)	0.997 (0.990-1.000)
UNI	448	**0.950 (0.898-0.989)**	**0.950 (0.901-0.988)**	**0.998 (0.995-1.000)**
ResNet-50$_{\text{IN}}$	896	0.600 (0.484-0.704)	0.607 (0.496-0.711)	0.848 (0.777-0.909)
CTransPath	896	0.788 (0.688-0.868)	0.791 (0.695-0.866)	0.949 (0.909-0.978)
REMEDIS	896	0.825 (0.741-0.900)	0.830 (0.748-0.902)	0.960 (0.926-0.984)
UNI	896	**0.925 (0.858-0.976)**	**0.926 (0.862-0.975)**	**0.996 (0.990-1.000)**
ResNet-50$_{\text{IN}}$	1344	0.587 (0.464-0.697)	0.593 (0.478-0.697)	0.829 (0.754-0.894)
CTransPath	1344	0.650 (0.547-0.749)	0.656 (0.554-0.753)	0.879 (0.817-0.929)
REMEDIS	1344	0.812 (0.721-0.886)	0.821 (0.728-0.891)	0.966 (0.936-0.987)
UNI	1344	**0.900 (0.829-0.960)**	**0.901 (0.829-0.962)**	**0.987 (0.970-0.997)**

Supplementary Data Table 45: **Linear probe evaluation for BRCA subtyping based on BACH (4 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on label-stratified train-test folds (80:20 ratio, 320:80 ROIs) across multiple image resolutions (without stain normalization), with test performance (n=80 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Img Res.	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{IN}$	224	0.588 (0.484-0.688)	0.579 (0.455-0.685)	0.625 (0.516-0.723)	0.618 (0.504-0.723)
CTransPath	224	0.750 (0.662-0.834)	0.739 (0.635-0.834)	0.750 (0.656-0.837)	0.745 (0.640-0.838)
REMEDIS	224	0.650 (0.550-0.742)	0.628 (0.506-0.741)	0.825 (0.741-0.904)	0.823 (0.735-0.901)
UNI	224	**0.812 (0.725-0.896)**	**0.813 (0.723-0.897)**	**0.900 (0.825-0.962)**	**0.902 (0.830-0.962)**
ResNet-50$_{IN}$	448	0.600 (0.497-0.697)	0.585 (0.471-0.703)	0.613 (0.506-0.717)	0.601 (0.483-0.716)
CTransPath	448	0.725 (0.630-0.814)	0.713 (0.599-0.819)	0.775 (0.682-0.863)	0.771 (0.671-0.862)
REMEDIS	448	0.575 (0.460-0.680)	0.569 (0.452-0.680)	0.687 (0.578-0.784)	0.688 (0.583-0.782)
UNI	448	**0.887 (0.812-0.948)**	**0.888 (0.813-0.950)**	**0.912 (0.841-0.965)**	**0.912 (0.838-0.963)**
ResNet-50$_{IN}$	896	0.587 (0.486-0.688)	0.565 (0.448-0.691)	0.538 (0.429-0.636)	0.520 (0.399-0.638)
CTransPath	896	0.650 (0.545-0.742)	0.641 (0.531-0.745)	0.725 (0.622-0.811)	0.728 (0.624-0.817)
REMEDIS	896	0.562 (0.451-0.667)	0.568 (0.452-0.673)	0.625 (0.514-0.726)	0.638 (0.530-0.732)
UNI	896	**0.850 (0.768-0.925)**	**0.849 (0.766-0.925)**	**0.887 (0.813-0.950)**	**0.885 (0.806-0.950)**
ResNet-50$_{IN}$	1344	0.525 (0.422-0.619)	0.507 (0.390-0.621)	0.475 (0.365-0.574)	0.474 (0.363-0.577)
CTransPath	1344	0.587 (0.478-0.691)	0.591 (0.478-0.701)	0.562 (0.444-0.663)	0.571 (0.460-0.674)
REMEDIS	1344	0.487 (0.371-0.594)	0.503 (0.386-0.613)	0.500 (0.389-0.610)	0.512 (0.398-0.616)
UNI	1344	**0.812 (0.722-0.892)**	**0.808 (0.710-0.889)**	**0.837 (0.749-0.914)**	**0.836 (0.750-0.913)**

Supplementary Data Table 46: **SimpleShot and KNN evaluation for BRCA subtyping based on BACH (4 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on label-stratified train-test folds (80:20 ratio, 320:80 ROIs) across multiple image resolutions (without stain normalization), with test performance (n=80 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	0.265 (0.259-0.271)	0.787 (0.781-0.793)	0.830 (0.822-0.839)
CTransPath	0.422 (0.409-0.435)	0.828 (0.823-0.833)	0.881 (0.874-0.889)
REMEDIS	0.410 (0.395-0.424)	0.816 (0.811-0.821)	0.836 (0.827-0.845)
UNI	**0.465 (0.449-0.481)**	**0.844 (0.838-0.848)**	**0.888 (0.879-0.896)**

Supplementary Data Table 47: **Linear probe evaluation for CRC tissue classification based on HunCRC (9 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on case-stratified train-test folds (77:23 ratio, 76,753:22,655 ROIs), with test performance (n=22,655 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{IN}$	0.421 (0.398-0.446)	0.570 (0.563-0.576)	0.379 (0.363-0.393)	0.775 (0.769-0.781)
CTransPath	**0.496 (0.475-0.517)**	0.695 (0.690-0.701)	0.437 (0.425-0.453)	0.806 (0.800-0.811)
REMEDIS	0.445 (0.426-0.467)	0.681 (0.675-0.687)	0.400 (0.385-0.414)	0.805 (0.800-0.810)
UNI	0.470 (0.451-0.492)	**0.747 (0.741-0.752)**	**0.448 (0.434-0.461)**	**0.831 (0.826-0.836)**

Supplementary Data Table 48: **SimpleShot and KNN evaluation for CRC tissue classification based on Hun-CRC (9 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on case-stratified train-test folds (77:23 ratio, 76,753:22,655 ROIs), with test performance (n=22,655 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	0.576 (0.573-0.579)	0.853 (0.852-0.855)	0.938 (0.936-0.940)
CTransPath	0.728 (0.723-0.734)	0.925 (0.924-0.926)	0.964 (0.963-0.966)
REMEDIS	0.774 (0.768-0.779)	0.879 (0.878-0.881)	0.969 (0.967-0.970)
UNI	**0.829 (0.823-0.834)**	**0.942 (0.941-0.943)**	**0.971 (0.970-0.973)**

Supplementary Data Table 49: **Linear probe evaluation for ESCA subtyping based on UKK, WNS, TCGA and CHA (11 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated using the UKK, WNS and TCGA cohorts as the train fold and the CHA cohort as the test fold (189,142:178,187), with test performance (n=178,187 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{\text{IN}}$	0.641 (0.635-0.646)	0.720 (0.718-0.722)	0.631 (0.626-0.636)	0.837 (0.836-0.839)
CTransPath4	0.821 (0.817-0.826)	0.856 (0.855-0.858)	0.785 (0.780-0.789)	0.918 (0.916-0.919)
REMEDIS	0.793 (0.789-0.798)	0.867 (0.865-0.868)	0.777 (0.773-0.781)	0.918 (0.917-0.919)
UNI	**0.835 (0.830-0.840)**	**0.904 (0.903-0.905)**	**0.836 (0.832-0.840)**	**0.947 (0.946-0.948)**

Supplementary Data Table 50: **SimpleShot and KNN evaluation for ESCA subtyping based on UKK, WNS, TCGA and CHA (11 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated using the UKK, WNS and TCGA cohorts as the train fold and the CHA cohort as the test fold (189,142:178,187 ROIs), with test performance (n=189,142 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Img Res.	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	224	0.349 (0.327-0.370)	0.436 (0.412-0.457)	0.830 (0.819-0.840)
CTransPath	224	0.432 (0.410-0.453)	0.481 (0.459-0.505)	**0.843 (0.833-0.851)**
REMEDIS	224	0.446 (0.422-0.472)	0.473 (0.452-0.494)	0.801 (0.786-0.813)
UNI	224	**0.504 (0.483-0.526)**	**0.533 (0.511-0.554)**	0.825 (0.814-0.835)
ResNet-50$_{\text{IN}}$	448	0.363 (0.343-0.383)	0.438 (0.416-0.460)	0.832 (0.821-0.841)
CTransPath	448	0.440 (0.419-0.460)	0.497 (0.475-0.519)	**0.840 (0.829-0.849)**
REMEDIS	448	0.458 (0.435-0.481)	0.534 (0.513-0.555)	0.816 (0.805-0.828)
UNI	448	**0.514 (0.491-0.535)**	**0.565 (0.543-0.588)**	0.836 (0.824-0.847)
ResNet-50$_{\text{IN}}$	896	0.341 (0.322-0.361)	0.438 (0.415-0.461)	0.816 (0.806-0.826)
CTransPath	896	0.436 (0.418-0.456)	0.507 (0.486-0.530)	0.831 (0.820-0.841)
REMEDIS	896	0.502 (0.478-0.525)	0.553 (0.532-0.573)	0.830 (0.819-0.841)
UNI	896	**0.530 (0.507-0.552)**	**0.589 (0.568-0.609)**	**0.853 (0.842-0.863)**
ResNet-50$_{\text{IN}}$	1792	0.308 (0.288-0.328)	0.435 (0.414-0.458)	0.780 (0.770-0.790)
CTransPath	1792	0.396 (0.376-0.414)	0.474 (0.453-0.496)	0.805 (0.795-0.816)
REMEDIS	1792	0.460 (0.436-0.485)	0.561 (0.542-0.583)	0.819 (0.808-0.830)
UNI	1792	**0.528 (0.507-0.549)**	**0.572 (0.550-0.591)**	**0.856 (0.845-0.866)**

Supplementary Data Table 51: **Linear probe evaluation for CRC polyp classification based on UniToPatho (6 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on the official train-test folds (72:28 ratio, 6,270:2,399 ROIs) across multiple image resolutions, with test performance (n=2,399 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Img Res.	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{\text{IN}}$	224	0.468 (0.440-0.493)	0.397 (0.375-0.417)	0.331 (0.309-0.353)	0.432 (0.408-0.453)
CTransPath	224	**0.567 (0.544-0.589)**	**0.488 (0.467-0.510)**	**0.446 (0.425-0.468)**	**0.502 (0.481-0.525)**
REMEDIS	224	0.494 (0.470-0.517)	0.420 (0.400-0.441)	0.353 (0.330-0.377)	0.466 (0.444-0.488)
UNI	224	0.533 (0.508-0.554)	0.429 (0.408-0.451)	0.425 (0.405-0.445)	0.476 (0.455-0.497)
ResNet-50$_{\text{IN}}$	448	0.464 (0.437-0.489)	0.387 (0.364-0.408)	0.351 (0.330-0.373)	0.443 (0.421-0.465)
CTransPath	448	0.592 (0.569-0.614)	0.507 (0.486-0.527)	**0.458 (0.436-0.480)**	**0.514 (0.494-0.535)**
REMEDIS	448	0.517 (0.495-0.540)	0.398 (0.378-0.418)	0.406 (0.383-0.430)	0.464 (0.443-0.487)
UNI	448	**0.602 (0.577-0.627)**	**0.550 (0.530-0.571)**	0.458 (0.438-0.480)	0.503 (0.482-0.523)
ResNet-50$_{\text{IN}}$	896	0.443 (0.420-0.469)	0.393 (0.372-0.413)	0.361 (0.341-0.384)	0.456 (0.436-0.479)
CTransPath	896	0.565 (0.540-0.587)	0.497 (0.477-0.517)	0.440 (0.417-0.462)	0.502 (0.482-0.524)
REMEDIS	896	0.515 (0.490-0.538)	0.426 (0.405-0.446)	0.369 (0.348-0.389)	0.442 (0.420-0.463)
UNI	896	**0.619 (0.597-0.641)**	**0.543 (0.523-0.565)**	**0.471 (0.451-0.491)**	**0.508 (0.486-0.529)**
ResNet-50$_{\text{IN}}$	1792	0.351 (0.326-0.376)	0.323 (0.303-0.343)	0.350 (0.328-0.371)	0.469 (0.448-0.491)
CTransPath	1792	0.494 (0.471-0.516)	0.474 (0.456-0.494)	0.414 (0.394-0.436)	0.494 (0.473-0.517)
REMEDIS	1792	0.472 (0.446-0.495)	0.413 (0.393-0.433)	0.352 (0.330-0.376)	0.464 (0.443-0.484)
UNI	1792	**0.616 (0.594-0.639)**	**0.568 (0.549-0.587)**	**0.476 (0.456-0.498)**	**0.529 (0.506-0.549)**

Supplementary Data Table 52: **SimpleShot and KNN evaluation for CRC polyp classification based on Uni-ToPatho (6 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on the official train-test folds (72:28 ratio, 6,270:2,399 ROIs) across multiple image resolutions, with test performance (n=2,399 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Balanced ACC	Quad. Weighted κ	Weighted F1	AUROC
ResNet-50$_{IN}$	0.506 (0.504-0.507)	0.592 (0.589-0.594)	0.612 (0.610-0.614)	0.862 (0.861-0.862)
CTransPath	0.630 (0.629-0.632)	0.779 (0.777-0.781)	0.723 (0.722-0.725)	0.915 (0.914-0.915)
REMEDIS	0.630 (0.629-0.632)	0.761 (0.759-0.764)	0.716 (0.714-0.717)	0.906 (0.905-0.906)
UNI	**0.658 (0.656-0.660)**	**0.797 (0.796-0.799)**	**0.743 (0.741-0.744)**	**0.922 (0.922-0.923)**

Supplementary Data Table 53: **Linear probe evaluation for PRAD tissue classification based on AGGC (5 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on the label stratified train-test folds (69:31 ratio, 780,619:345,021 ROIs) and test performance (n=345,021 ROIs) reported using balanced accuracy, Cohen's quadratic weighted κ, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{IN}$	0.562 (0.560-0.564)	0.482 (0.481-0.484)	0.527 (0.526-0.529)	0.605 (0.603-0.606)
CTransPath	0.691 (0.690-0.693)	0.630 (0.628-0.631)	0.599 (0.597-0.600)	0.666 (0.664-0.668)
REMEDIS	0.655 (0.653-0.657)	0.600 (0.598-0.602)	0.617 (0.615-0.619)	0.674 (0.672-0.675)
UNI	**0.738 (0.736-0.740)**	**0.697 (0.696-0.699)**	**0.655 (0.653-0.656)**	**0.714 (0.712-0.715)**

Supplementary Data Table 54: **SimpleShot and KNN evaluation for PRAD tissue classification based on AGGC (5 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on label stratified train-test folds (69:31 ratio, 780,619:345,021 ROIs) and test performance (n=345,021 ROIs) reported using balanced accuracy and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	SN	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{IN}$	✓	0.606 (0.598-0.614)	0.686 (0.681-0.691)	0.623 (0.615-0.633)
CTransPath	✓	0.714 (0.707-0.720)	0.772 (0.768-0.776)	0.782 (0.776-0.789)
REMEDIS	✓	0.649 (0.642-0.656)	0.792 (0.788-0.797)	0.740 (0.733-0.747)
UNI	✓	**0.716 (0.708-0.723)**	**0.813 (0.809-0.817)**	**0.797 (0.791-0.804)**

Supplementary Data Table 55: **Linear probe evaluation for CRC MSI prediction based on TCGA (2 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on official train-test folds (38:62 ratio, 19,557:32,361 ROIs), with test performance (n=32,361 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC with stain normalization (SN). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	SN	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{IN}$	✓	0.595 (0.587-0.602)	0.663 (0.658-0.668)	0.596 (0.590-0.604)	0.672 (0.667-0.676)
CTransPath	✓	0.614 (0.608-0.622)	0.669 (0.665-0.674)	0.630 (0.624-0.638)	0.736 (0.732-0.741)
REMEDIS	✓	0.583 (0.576-0.590)	0.669 (0.664-0.673)	0.603 (0.595-0.610)	0.737 (0.733-0.742)
UNI	✓	**0.682 (0.675-0.688)**	**0.726 (0.722-0.731)**	**0.694 (0.687-0.700)**	**0.769 (0.765-0.773)**

Supplementary Data Table 56: **SimpleShot and KNN evaluation for CRC MSI prediction based on TCGA (2 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on official train-test folds (38:62 ratio, 19,557:32,361 ROIs), with test performance (n=32,361 ROIs) reported using balanced accuracy and weighted F1 score with stain normalization (SN). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	SN	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	✗	0.439 (0.434-0.444)	0.525 (0.521-0.529)	0.926 (0.924-0.927)
CTransPath	✗	0.604 (0.599-0.610)	0.678 (0.674-0.682)	0.967 (0.966-0.968)
REMEDIS	✗	0.650 (0.646-0.655)	0.709 (0.705-0.713)	0.959 (0.958-0.961)
UNI	✗	**0.685 (0.680-0.690)**	**0.744 (0.741-0.748)**	**0.978 (0.978-0.979)**
ResNet-50$_{\text{IN}}$	✓	0.413 (0.408-0.418)	0.498 (0.493-0.502)	0.916 (0.914-0.918)
CTransPath	✓	0.561 (0.555-0.566)	0.637 (0.633-0.641)	0.960 (0.959-0.961)
REMEDIS	✓	0.610 (0.605-0.615)	0.678 (0.674-0.682)	0.958 (0.956-0.959)
UNI	✓	**0.657 (0.652-0.662)**	**0.718 (0.714-0.721)**	**0.975 (0.974-0.976)**

Supplementary Data Table 57: **Linear probe evaluation for pan-cancer tissue classification based on TCGA (32 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on label-stratified train-test folds (216,350:55,360 ROIs), with test performance (n=55,360 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	SN	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{\text{IN}}$	✗	0.287 (0.282-0.292)	0.269 (0.265-0.272)	0.347 (0.342-0.351)	0.415 (0.410-0.419)
CTransPath	✗	0.460 (0.454-0.465)	0.468 (0.464-0.472)	0.495 (0.490-0.499)	0.564 (0.560-0.568)
REMEDIS	✗	0.553 (0.548-0.558)	0.594 (0.590-0.598)	0.608 (0.603-0.612)	0.669 (0.665-0.673)
UNI	✗	**0.641 (0.636-0.646)**	**0.675 (0.671-0.679)**	**0.645 (0.640-0.649)**	**0.707 (0.703-0.711)**
ResNet-50$_{\text{IN}}$	✓	0.275 (0.270-0.280)	0.254 (0.249-0.257)	0.318 (0.314-0.323)	0.382 (0.378-0.386)
CTransPath	✓	0.442 (0.436-0.447)	0.438 (0.434-0.442)	0.463 (0.459-0.468)	0.534 (0.529-0.538)
REMEDIS	✓	0.504 (0.499-0.510)	0.529 (0.525-0.533)	0.541 (0.537-0.546)	0.618 (0.614-0.622)
UNI	✓	**0.591 (0.586-0.597)**	**0.611 (0.607-0.615)**	**0.595 (0.590-0.600)**	**0.663 (0.660-0.667)**

Supplementary Data Table 58: **SimpleShot and KNN evaluation for pan-cancer tissue classification based on TCGA (32 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, K=1) and Nearest Neighbors (K=20) were evaluated on label-stratified train-test folds (216,350:55,360 ROIs), with test performance (n=55,360 ROIs) reported using balanced accuracy and weighted F1 score with and without stain normalization (SN). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	SN	Balanced ACC	Weighted F1	AUROC
ResNet-50$_{\text{IN}}$	✗	0.667 (0.662-0.671)	0.845 (0.842-0.849)	0.919 (0.916-0.922)
CTransPath	✗	0.760 (0.756-0.765)	0.893 (0.891-0.896)	0.944 (0.942-0.946)
REMEDIS	✗	**0.891 (0.886-0.894)**	0.937 (0.935-0.939)	0.975 (0.973-0.976)
UNI	✗	0.887 (0.883-0.891)	**0.944 (0.942-0.946)**	**0.979 (0.977-0.980)**
ResNet-50$_{\text{IN}}$	✓	0.651 (0.646-0.655)	0.836 (0.833-0.840)	0.915 (0.911-0.917)
CTransPath	✓	0.758 (0.753-0.762)	0.891 (0.889-0.894)	0.930 (0.928-0.933)
REMEDIS	✓	**0.894 (0.890-0.898)**	**0.944 (0.942-0.946)**	**0.978 (0.977-0.979)**
UNI	✓	0.870 (0.865-0.874)	0.938 (0.936-0.940)	0.975 (0.974-0.977)

Supplementary Data Table 59: **Linear probe evaluation for pan-cancer TIL detection based on TCGA (2 classes)**. Pre-extracted patch features of each encoder with logistic regression were evaluated on official train-validation-test folds (69:13:19 ratio, 209,221:38,601:56,275 ROIs, train-validation combined during training) and test performance (n=56,275 ROIs) reported using balanced accuracy, weighted F1 score, and AUROC with and without stain normalization (SN). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	SN	1-NN Balanced ACC	1-NN Weighted F1	20-NN Balanced ACC	20-NN Weighted F1
ResNet-50$_{IN}$	✗	0.811 (0.806-0.815)	0.842 (0.839-0.845)	0.821 (0.816-0.826)	0.908 (0.905-0.910)
CTransPath	✗	0.806 (0.802-0.811)	0.827 (0.824-0.829)	0.840 (0.836-0.845)	0.919 (0.916-0.921)
REMEDIS	✗	0.804 (0.800-0.808)	0.847 (0.844-0.849)	0.836 (0.832-0.841)	0.905 (0.903-0.908)
UNI	✗	**0.886 (0.883-0.890)**	**0.931 (0.929-0.933)**	**0.865 (0.861-0.870)**	**0.935 (0.933-0.937)**
ResNet-50$_{IN}$	✓	0.812 (0.807-0.816)	0.848 (0.845-0.850)	0.806 (0.801-0.811)	0.900 (0.898-0.903)
CTransPath	✓	0.795 (0.790-0.799)	0.813 (0.810-0.816)	0.831 (0.827-0.836)	0.914 (0.911-0.916)
REMEDIS	✓	0.856 (0.852-0.860)	0.885 (0.883-0.888)	**0.861 (0.857-0.865)**	0.930 (0.928-0.932)
UNI	✓	**0.865 (0.861-0.868)**	**0.907 (0.904-0.909)**	0.855 (0.851-0.859)	**0.931 (0.928-0.933)**

Supplementary Data Table 60: **SimpleShot and KNN evaluation for pan-cancer TIL detection based on TCGA (2 classes)**. Pre-extracted patch features of each encoder with SimpleShot (nearest centroid, $K=1$) and Nearest Neighbors ($K=20$) were evaluated on official train-validation-test folds (69:13:19 ratio, 209,221:38,601:56,275 ROIs, train-validation combined during training) with test performance (n=56,275 ROIs) reported using balanced accuracy and weighted F1 score with and without stain normalization (SN). Best performing model for each metric is bolded. 95% CI is included in parentheses.

Method	Encoder	AUROC	FNR
FT (AdamW)	ResNet-18	0.970	0.252
FT (AdamW)	ResNet-18$_{IN}$	0.969	0.256
FT (AdamW)	ResNet-18$_{SimCLR}$	0.967	0.275
FT (AdamW)	ResNet50	0.969	0.241
FT (AdamW)	ResNet50$_{IN}$	0.968	0.253
FT (AdamW)	R26-ViT$_{IN}$	0.968	0.253
LP (L-BFGS), SN	UNI	0.975	0.231
KNN ($K=20$), SN	UNI	-	0.281
SimpleShot ($K=1$), SN	UNI	-	0.189
LP (L-BFGS), No SN	UNI	**0.979**	0.198
KNN ($K=20$), No SN	UNI	-	0.252
SimpleShot ($K=1$), No SN	UNI	-	**0.158**

Supplementary Data Table 61: **Comparison with existing baselines for pan-cancer TIL detection based on TCGA (2 classes).** We compare UNI (results from **Extended Data Table 59,60**, same splits) with existing baselines on this task, which were taken from the ChampKit [54] leaderboard. All of the baselines developed in ChampKit are end-to-end finetuned CNN models which explored the effects of random initialization, transfer learning from ImageNet, or using a self-supervised encoder. FT = finetuned. LP = linear probe. SN = Macenko stain normalization. SimpleShot (nearest centroid, $K=1$) is evaluated using all training data for computing class prototypes.

Method	Encoder	Cohort	ACC
FT (SGD, freeze-embed) [184]	CLIP ViT-L/14 [185]	Val$_{OD}$	0.952
ContriMix [186]	DenseNet121 [187]	Val$_{OD}$	0.919
MBDG [188]	DenseNet121 [187]	Val$_{OD}$	0.881
ERM w/ targeted aug [189]	DenseNet121 [187]	Val$_{OD}$	0.927
ERM w/ H&E jitter [190]	ResNeXt101-32x4d [191]	Val$_{OD}$	0.880
KNN (K=20)	UNI	Val$_{OD}$	0.904
SimpleShot (nearest centroid, K=1)	UNI	Val$_{OD}$	0.968
LP (L-BFGS)	UNI	Val$_{OD}$	**0.974**
FT (SGD) [184]	CLIP ViT-L/14 [185]	Test$_{OD}$	0.832
FT (AdamW) [184]	CLIP ViT-L/14 [185]	Test$_{OD}$	0.959
FT (SGD, freeze-embed) [184]	CLIP ViT-L/14 [185]	Test$_{OD}$	0.965
FT (SGD, freeze-embed, no momentum) [184]	CLIP ViT-L/14 [185]	Test$_{OD}$	0.967
ContriMix [186]	DenseNet121 [187]	Test$_{OD}$	0.946
MBDG [188]	DenseNet121 [187]	Test$_{OD}$	0.933
ERM w/ targeted aug [189]	DenseNet121 [187]	Test$_{OD}$	0.921
ERM w/ H&E jitter [190]	ResNeXt101-32x4d [191]	Test$_{OD}$	0.926
KNN (K=20)	UNI	Test$_{OD}$	0.955
SimpleShot (nearest centroid, K=1)	UNI	Test$_{OD}$	0.972
LP (L-BFGS)	UNI	Test$_{OD}$	**0.983**

Supplementary Data Table 62: **Comparison with existing baselines for breast metastasis detection in CAMELYON17-WILDS (2 classes).** We compare UNI (results from **Extended Data Table 41,42**, same splits) with current top-5 solutions from the WILDS [164] leaderboard, which evaluate both the out-of-domain validation and test set. All comparisons are end-to-end finetuned CNN or ViT architectures, with additional finetuning protocols specific to their study. In addition, we also show comparisons with end-to-end finetuned ViT-L models that were developed by the 1st place solution [184]. FT = finetuned. LP = linear probe. OD = out-of-domain. SimpleShot (nearest centroid, $K = 1$) is evaluated using all training data for computing class prototypes.

Encoder	Acc@1	Acc@3	Acc@5	MVAcc@5
ResNet-50$_{\text{IN}}$	0.697 (0.685, 0.707)	0.789 (0.779, 0.798)	0.896 (0.889, 0.903)	0.732 (0.721, 0.742)
CTransPath	0.819 (0.811, 0.828)	0.850 (0.842, 0.858)	0.877 (0.870, 0.885)	0.831 (0.823, 0.839)
REMEDIS	0.879 (0.871, 0.886)	0.942 (0.937, 0.947)	0.956 (0.952, 0.961)	0.917 (0.911, 0.924)
UNI	**0.910 (0.903, 0.916)**	**0.943 (0.938, 0.949)**	**0.962 (0.957, 0.966)**	**0.923 (0.917, 0.929)**

Supplementary Data Table 63: **Image retrieval for CRC tissue classification based on CRC-100K (9 classes).** Pre-extracted patch features of each encoder with retrieval were evaluated on the official train-test folds (100,000:7,180), with test performance (n=7,180 patches) reported using Acc@K for K $\in 1, 3, 5$ and MVAcc@5. Best performing model for each model is bolded. 95% CI is included in parentheses.

Encoder	Acc@1	Acc@3	Acc@5	MVAcc@5
ResNet-50$_{\text{IN}}$	0.716 (0.710, 0.721)	0.868 (0.864, 0.872)	0.912 (0.908, 0.915)	0.769 (0.764, 0.774)
CTransPath	0.764 (0.759, 0.770)	0.877 (0.873, 0.882)	0.909 (0.906, 0.913)	0.801 (0.796, 0.806)
REMEDIS	0.766 (0.760, 0.771)	0.884 (0.880, 0.888)	0.916 (0.913, 0.920)	0.805 (0.800, 0.810)
UNI	**0.796 (0.791, 0.801)**	**0.899 (0.895, 0.902)**	**0.926 (0.923, 0.930)**	**0.828 (0.823, 0.833)**

Supplementary Data Table 64: **Image retrieval for CRC tissue classification based on HunCRC (9 classes).** Pre-extracted patch features of each encoder with retrieval were evaluated on case-stratified train-test folds (76,753:22,655), with test performance (n=22,655 patches) reported using Acc@K for K $\in 1, 3, 5$ and MVAcc@5. Best performing model for each model is bolded. 95% CI is included in parentheses.

Encoder	Acc@1	Acc@3	Acc@5	MVAcc@5
ResNet-50$_{\text{IN}}$	0.756 (0.754, 0.758)	0.858 (0.856, 0.859)	0.887 (0.886, 0.889)	0.793 (0.791, 0.795)
CTransPath	0.859 (0.857, 0.861)	0.926 (0.924, 0.927)	0.944 (0.943, 0.945)	0.883 (0.882, 0.885)
REMEDIS	0.872 (0.870, 0.873)	0.936 (0.935, 0.937)	0.953 (0.952, 0.954)	0.898 (0.896, 0.899)
UNI	**0.913 (0.912, 0.914)**	**0.960 (0.959, 0.960)**	**0.971 (0.970, 0.972)**	**0.929 (0.928, 0.931)**

Supplementary Data Table 65: **Image retrieval for ESCA subtyping based on UKK, WNS, TCGA and CHA (11 classes).** Pre-extracted patch features of each encoder with retrieval were evaluated using the UKK, WNS and TCGA cohorts as the train fold and the CHA cohort as the test fold (189,142:178,187), with test performance (n=178,187 patches) reported using Acc@K for K $\in 1, 3, 5$ and MVAcc@5. Best performing model for each model is bolded. 95% CI is included in parentheses.

Encoder	Acc@1	Acc@3	Acc@5	MVAcc@5
ResNet-50$_{\text{IN}}$	0.553 (0.551, 0.555)	0.763 (0.762, 0.765)	0.833 (0.832, 0.834)	0.623 (0.621, 0.624)
CTransPath	0.621 (0.620, 0.623)	0.793 (0.791, 0.794)	0.847 (0.846, 0.848)	0.670 (0.668, 0.671)
REMEDIS	0.624 (0.622, 0.625)	0.803 (0.802, 0.805)	0.858 (0.857, 0.859)	0.677 (0.676, 0.679)
UNI	**0.664 (0.662, 0.665)**	**0.825 (0.823, 0.826)**	**0.872 (0.871, 0.873)**	**0.710 (0.709, 0.712)**

Supplementary Data Table 66: **Image retrieval for PRAD tissue classification based on AGGC (5 classes).** Pre-extracted patch features of each encoder with retrieval were evaluated on case-stratified train-test folds (780,619:345,021), with test performance (n=345,021 patches) reported using Acc@K for K $\in 1, 3, 5$ and MVAcc@5. Best performing model for each model is bolded. 95% CI is included in parentheses.

Encoder	Acc@1	Acc@3	Acc@5	MVAcc@5
ResNet-50$_{\text{IN}}$	0.433 (0.413, 0.454)	0.636 (0.616, 0.654)	0.712 (0.695, 0.729)	0.516 (0.495, 0.536)
CTransPath	0.478 (0.457, 0.497)	0.625 (0.605, 0.644)	0.684 (0.665, 0.702)	0.531 (0.510, 0.550)
REMEDIS	0.434 (0.413, 0.453)	0.592 (0.572, 0.609)	0.659 (0.639, 0.679)	0.486 (0.466, 0.504)
UNI	**0.509 (0.489, 0.529)**	**0.671 (0.652, 0.690)**	**0.732 (0.714, 0.750)**	**0.551 (0.532, 0.570)**

Supplementary Data Table 67: **Image retrieval for CRC polyp classification based on UniToPatho.** Pre-extracted patch features of each encoder with retrieval were evaluated on the official case-stratified train-test folds (6,270:2,399) with resized 1792^2 image resolutions, with test performance (n=345,021 patches) reported using Acc@K for K $\in 1, 3, 5$ and MVAcc@5. Best performing model for each model is bolded. 95% CI is included in parentheses.

Encoder	SN	Acc@1	Acc@3	Acc@5	MVAcc@5
ResNet-50$_{\text{IN}}$	✗	0.379 (0.375, 0.383)	0.536 (0.532, 0.540)	0.609 (0.606, 0.614)	0.455 (0.451, 0.458)
CTransPath	✗	0.529 (0.525, 0.533)	0.659 (0.655, 0.663)	0.717 (0.713, 0.720)	0.578 (0.574, 0.582)
REMEDIS	✗	0.637 (0.633, 0.642)	0.741 (0.738, 0.745)	0.785 (0.781, 0.788)	0.676 (0.672, 0.680)
UNI	✗	**0.671 (0.667, 0.675)**	**0.778 (0.775, 0.782)**	**0.821 (0.818, 0.824)**	**0.711 (0.707, 0.715)**
ResNet-50 (IN)	✓	0.349 (0.345, 0.353)	0.510 (0.506, 0.514)	0.592 (0.587, 0.596)	0.434 (0.430, 0.438)
CTransPath	✓	0.492 (0.488, 0.497)	0.638 (0.634, 0.642)	0.705 (0.701, 0.709)	0.555 (0.551, 0.559)
REMEDIS	✓	0.574 (0.570, 0.578)	0.704 (0.700, 0.707)	0.758 (0.754, 0.761)	0.630 (0.626, 0.635)
UNI	✓	**0.620 (0.616, 0.624)**	**0.749 (0.745, 0.752)**	**0.801 (0.797, 0.804)**	**0.671 (0.667, 0.675)**

Supplementary Data Table 68: **Image retrieval for pan-cancer tissue classification based on TCGA (32 classes).** Pre-extracted patch features of each encoder with retrieval were evaluated on case-stratified train-test folds (216,350:55,360), with test performance (n=55,360 patches) reported using Acc@K for K $\in 1, 3, 5$ and MVAcc@5 with and without stain normalization (SN). Best performing model for each model is bolded. 95% CI is included in parentheses.

Encoder	Cell type	Dice	Precision	Recall
Komura *et al.* [102]	Endothelium	N/A (0.350-0.400)	N/A	N/A
ResNet-50_IN	Endothelium	0.665 (0.658-0.674)	0.709 (0.697-0.722)	0.658 (0.654-0.663)
CTransPath	Endothelium	0.658 (0.644-0.663)	0.695 (0.682-0.702)	0.656 (0.641-0.663)
REMEDIS	Endothelium	0.684 (0.677-0.694)	0.679 (0.672-0.688)	**0.729 (0.724-0.736)**
UNI	Endothelium	**0.696 (0.686-0.703)**	**0.718 (0.707-0.726)**	0.709 (0.700-0.716)
Komura *et al.* [102]	Epithelium	N/A (0.800-0.805)	N/A	N/A
ResNet-50_IN	Epithelium	0.812 (0.808-0.814)	0.840 (0.837-0.841)	0.816 (0.813-0.819)
CTransPath	Epithelium	0.815 (0.811-0.817)	0.842 (0.839-0.845)	0.816 (0.812-0.818)
REMEDIS	Epithelium	0.824 (0.819-0.829)	0.843 (0.840-0.846)	0.834 (0.829-0.838)
UNI	Epithelium	**0.827 (0.823-0.830)**	**0.849 (0.847-0.851)**	**0.834 (0.831-0.838)**
Komura *et al.* [102]	Leukocyte	N/A (0.600-0.650)	N/A	N/A
ResNet-50_IN	Leukocyte	0.691 (0.689-0.693)	0.726 (0.724-0.727)	0.692 (0.690-0.695)
CTransPath	Leukocyte	0.686 (0.684-0.688)	0.710 (0.707-0.711)	0.699 (0.696-0.702)
REMEDIS	Leukocyte	0.706 (0.704-0.707)	0.725 (0.722-0.727)	**0.719 (0.717-0.721)**
UNI	Leukocyte	**0.706 (0.705-0.709)**	**0.736 (0.734-0.737)**	0.713 (0.712-0.716)
Komura *et al.* [102]	Lymphocyte	N/A (0.600-0.650)	N/A	N/A
ResNet-50_IN	Lymphocyte	0.631 (0.625-0.638)	0.693 (0.684-0.701)	0.626 (0.618-0.634)
CTransPath	Lymphocyte	0.629 (0.623-0.635)	0.677 (0.671-0.682)	0.627 (0.620-0.632)
REMEDIS	Lymphocyte	**0.653 (0.650-0.660)**	**0.686 (0.680-0.696)**	0.665 (0.659-0.669)
UNI	Lymphocyte	0.651 (0.647-0.658)	0.686 (0.680-0.693)	**0.665 (0.657-0.673)**
Komura *et al.* [102]	Smooth Muscle	N/A (0.600-0.650)	N/A	N/A
ResNet-50_IN	Smooth Muscle	0.650 (0.646-0.652)	0.699 (0.696-0.703)	0.668 (0.664-0.672)
CTransPath	Smooth Muscle	0.655 (0.648-0.659)	0.703 (0.698-0.707)	0.670 (0.665-0.675)
REMEDIS	Smooth Muscle	0.674 (0.668-0.678)	0.724 (0.719-0.728)	0.687 (0.684-0.693)
UNI	Smooth Muscle	**0.690 (0.685-0.694)**	**0.736 (0.731-0.740)**	**0.704 (0.700-0.708)**
Komura *et al.* [102]	Myeloid Cell	N/A (0.300-0.350)	N/A	N/A
ResNet-50_IN	Myeloid Cell	0.615 (0.611-0.620)	0.698 (0.690-0.703)	0.596 (0.592-0.602)
CTransPath	Myeloid Cell	0.621 (0.618-0.627)	0.697 (0.692-0.700)	0.604 (0.599-0.610)
REMEDIS	Myeloid Cell	0.652 (0.647-0.659)	**0.728 (0.719-0.733)**	0.630 (0.625-0.638)
UNI	Myeloid Cell	**0.656 (0.655-0.662)**	0.726 (0.724-0.730)	**0.637 (0.634-0.643)**
Komura *et al.* [102]	Plasma Cell	N/A (0.400-0.450)	N/A	N/A
ResNet-50_IN	Plasma Cell	0.703 (0.692-0.706)	0.761 (0.745-0.769)	0.692 (0.683-0.696)
CTransPath	Plasma Cell	0.713 (0.703-0.720)	0.754 (0.743-0.760)	0.709 (0.700-0.716)
REMEDIS	Plasma Cell	**0.742 (0.734-0.747)**	0.784 (0.775-0.786)	**0.736 (0.725-0.742)**
UNI	Plasma Cell	0.737 (0.730-0.743)	**0.788 (0.778-0.794)**	0.728 (0.722-0.732)
Komura *et al.* [102]	RBC	N/A (0.600-0.650)	N/A	N/A
ResNet-50_IN	RBC	0.797 (0.793-0.802)	0.827 (0.825-0.831)	0.808 (0.800-0.812)
CTransPath	RBC	0.786 (0.784-0.792)	0.818 (0.815-0.821)	0.794 (0.789-0.800)
REMEDIS	RBC	0.795 (0.791-0.800)	0.824 (0.821-0.827)	0.807 (0.802-0.815)
UNI	RBC	**0.803 (0.800-0.808)**	**0.839 (0.837-0.843)**	**0.810 (0.804-0.817)**
ResNet-50_IN	Average	0.696	0.744	0.695
CTransPath	Average	0.695	0.737	0.700
REMEDIS	Average	0.716	0.749	**0.726**
UNI	Average	**0.721**	**0.760**	0.725

Supplementary Data Table 69: **Pan-cancer cell type segmentation based on SegPath (8 cell types treated as individual tasks**. ROI-level cell segmentation of eight major cell types using their official train-test folds in the SegPath dataset [102]. We finetune each pretrained encoder using the Mask2Former framework [103], a flexible framework used commonly for adapting self-supervised models as segmentation backbones for dense prediction tasks. All encoders were trained and evaluated on the official train-validation-test folds, with test performance reported using dice score, precision, and recall. Comparisons to the best model in SegPath [102] is also reported. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Top K	Balanced ACC	Weighted F1
ResNet-50$_{IN}$	TCGA	5	0.531 (0.500-0.566)	0.398 (0.280-0.522)
CTransPath	TCGA	5	0.796 (0.705-0.869)	0.796 (0.712-0.868)
REMEDIS	TCGA	5	0.796 (0.707-0.874)	0.795 (0.709-0.877)
UNI	TCGA	5	**0.908 (0.847-0.960)**	**0.908 (0.847-0.959)**
ResNet-50$_{IN}$	TCGA	50	0.531 (0.488-0.574)	0.412 (0.295-0.526)
CTransPath	TCGA	50	0.806 (0.724-0.878)	0.806 (0.724-0.878)
REMEDIS	TCGA	50	0.786 (0.701-0.868)	0.785 (0.701-0.867)
UNI	TCGA	50	**0.918 (0.863-0.970)**	**0.918 (0.867-0.969)**
ResNet-50$_{IN}$	CPTAC	5	0.530 (0.517-0.544)	0.444 (0.407-0.481)
CTransPath	CPTAC	5	0.845 (0.823-0.865)	0.847 (0.825-0.867)
REMEDIS	CPTAC	5	0.771 (0.748-0.793)	0.752 (0.723-0.778)
UNI	CPTAC	5	**0.902 (0.882-0.918)**	**0.902 (0.883-0.919)**
ResNet-50$_{IN}$	CPTAC	50	0.535 (0.520-0.551)	0.461 (0.426-0.497)
CTransPath	CPTAC	50	0.827 (0.805-0.848)	0.830 (0.808-0.852)
REMEDIS	CPTAC	50	0.806 (0.783-0.825)	0.793 (0.767-0.816)
UNI	CPTAC	50	**0.901 (0.882-0.918)**	**0.901 (0.883-0.918)**

Supplementary Data Table 70: **Prototypical NSCLC subtyping based on TCGA and CPTAC (2 classes)**. Class prototypes for MI-SimpleShot were developed using annotated ROIs from the TCGA Uniform Tumor Dataset, which we evaluate on the NSCLC subtyping task using the same folds in weakly-supervised ABMIL evaluation (internal evaluation on slides (n=98) in TCGA), with external evaluation on slides (n=1091) sourced from CPTAC-LUAD (n=578) and CPTAC-LUSC (n=513). Test performance was reported using balanced accuracy, and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Encoder	Cohort	Top K	Balanced ACC	Weighted F1
ResNet-50$_{IN}$	TCGA	5	0.636 (0.530-0.725)	0.501 (0.394-0.601)
CTransPath	TCGA	5	0.896 (0.814-0.950)	0.867 (0.803-0.928)
REMEDIS	TCGA	5	0.769 (0.712-0.828)	0.728 (0.625-0.826)
UNI	TCGA	5	**0.928 (0.886-0.964)**	**0.888 (0.827-0.940)**
ResNet-50$_{IN}$	TCGA	50	0.700 (0.597-0.791)	0.607 (0.502-0.707)
CTransPath	TCGA	50	0.902 (0.824-0.956)	0.878 (0.813-0.930)
REMEDIS	TCGA	50	0.757 (0.696-0.820)	0.710 (0.609-0.809)
UNI	TCGA	50	**0.938 (0.903-0.969)**	**0.897 (0.837-0.949)**
ResNet-50$_{IN}$	CPTAC	5	0.701 (0.654-0.741)	0.439 (0.405-0.473)
CTransPath	CPTAC	5	0.919 (0.905-0.930)	0.828 (0.804-0.849)
REMEDIS	CPTAC	5	**0.957 (0.940-0.973)**	**0.939 (0.924-0.954)**
UNI	CPTAC	5	0.952 (0.940-0.961)	0.896 (0.878-0.914)
ResNet-50$_{IN}$	CPTAC	50	0.744 (0.694-0.791)	0.575 (0.543-0.606)
CTransPath	CPTAC	50	0.891 (0.846-0.925)	0.829 (0.807-0.849)
REMEDIS	CPTAC	50	**0.957 (0.937-0.975)**	**0.951 (0.937-0.965)**
UNI	CPTAC	50	0.957 (0.946-0.967)	0.909 (0.892-0.925)

Supplementary Data Table 71: **Prototypical RCC subtyping based on TCGA and CPTAC-DHMC (3 classes)**. Class prototypes for MI-SimpleShot were developed using annotated ROIs from the TCGA Uniform Tumor Dataset, which we evaluate on the RCC subtyping task using the same folds in weakly-supervised ABMIL evaluation (internal evaluation on slides (n=97) in TCGA), with external evaluation on slides (n=872) sourced from CPTAC-CCRCC (n=404) and DHMC-Kidney (n=468) (CCRCC, CHRCC, and PRCC cases only). Test performance was reported using balanced accuracy, and weighted F1 score. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Pretrain Strategy	Dataset	CAMELYON17-WILDS	UniToPatho	TCGA-TILS	EBRAINS (Fine)
Supervised	TCGA	0.909 (0.907-0.911)	0.398 (0.375-0.422)	0.851 (0.847-0.856)	0.569 (0.528-0.612)
DINOv2	Mass-100K	**0.983 (0.982-0.984)**	**0.528 (0.507-0.549)**	**0.870 (0.865-0.874)**	**0.675 (0.633-0.715)**

Supplementary Data Table 72: **Comparison against transfer learning from a general pathology task**. We additionally compare UNI with a ViT-L/16 (with ImageNet-22K transfer) trained end-to-end on the 32-class pan-cancer tissue classification task. This model was evaluated on the same folds as UNI on breast metastasis detection in CAMELYON17-WILDS (official test fold, 302,436:85,054 ROIs), CRC polyp classification in UNITOPATHO (6,270:2,399 ROIs), pancancer TIL detection in TCGA (209,221:38,601:56,275 ROIs), and fine-grained brain tumor subtyping in EBRAINS (1151:595:573 slides). We chose these representative tasks due to evaluating unique challenges in out-of-domain generalization, adaption to high resolutions, evaluation on diverse tissue types (ROI and slide-level) respectively. ROI and slide-level tasks were evaluated using linear probing and weakly-supervised ABMIL respectively. Test performance was reported using balanced accuracy. Best performing model for each metric is bolded. 95% CI is included in parentheses.

Dataset	Link
TCGA	https://portal.gdc.cancer.gov
CPTAC	https://proteomic.datacommons.cancer.gov/pdc
GTEx	https://www.gtexportal.org/home/
CRC-100K	https://zenodo.org/record/1214456
HunCRC (ROIs)	https://doi.org/10.6084/m9.figshare.c.5927795.v1
HunCRC (Slides)	https://doi.org/10.7937/tcia.9cjf-0127
BACH	https://iciar2018-challenge.grand-challenge.org/Dataset/
TCGA CRC-MSI	https://zenodo.org/record/3832231
CCRCC Tissue Classification	https://zenodo.org/record/7898308
TCGA-TILs	https://zenodo.org/record/6604094
TCGA Uniform Tumor	https://zenodo.org/record/5889558
UniToPatho	https://zenodo.org/record/4643645/
ESCA	https://zenodo.org/record/7548828
CAMELYON17-WILDS	https://wilds.stanford.edu/datasets/
EBRAINS	https://doi.org/10.25493/WQ48-ZGX
DHMC	https://bmirds.github.io/KidneyCancer/
BRACS	https://www.bracs.icar.cnr.it/
PANDA	https://panda.grand-challenge.org/data/
SegPath	https://zenodo.org/record/7412731
AGGC	https://zenodo.org/record/6460100

Supplementary Data Table 73: **Summary of publicly available datasets.**