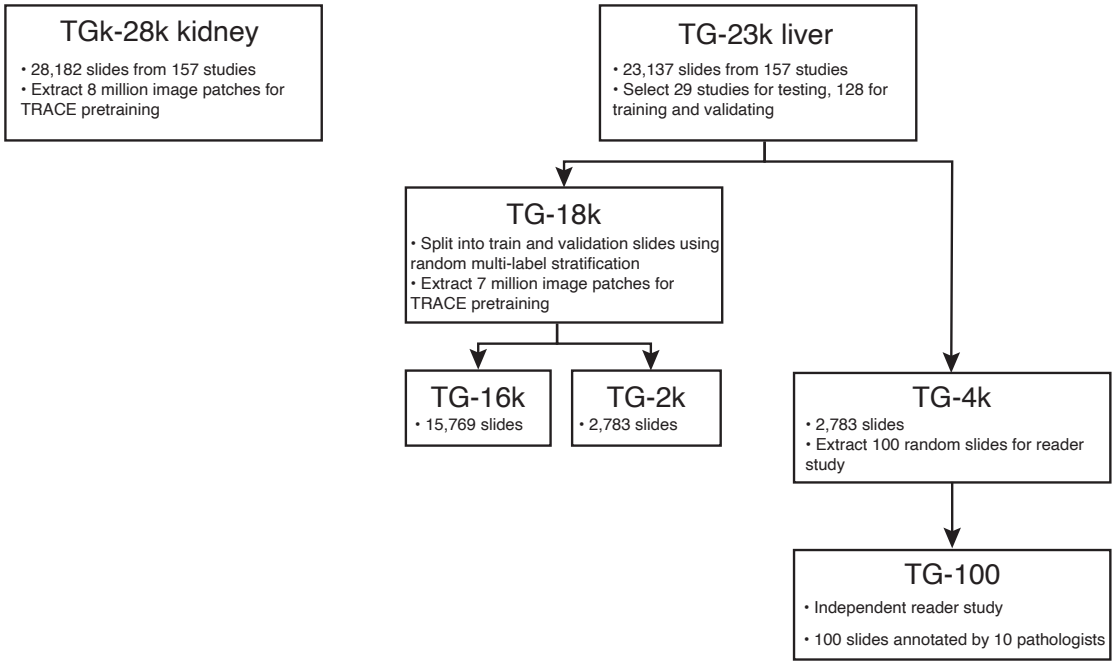

Figure S1: **Overview of data profile.** This study uses the TG-GATEs database for TRACE pretraining (TG-28k, TG-18k), weakly supervised and few-shot classification (TG-4k), TRACE fine-tuning with patch annotations, and conducting the reader study (TG-100).

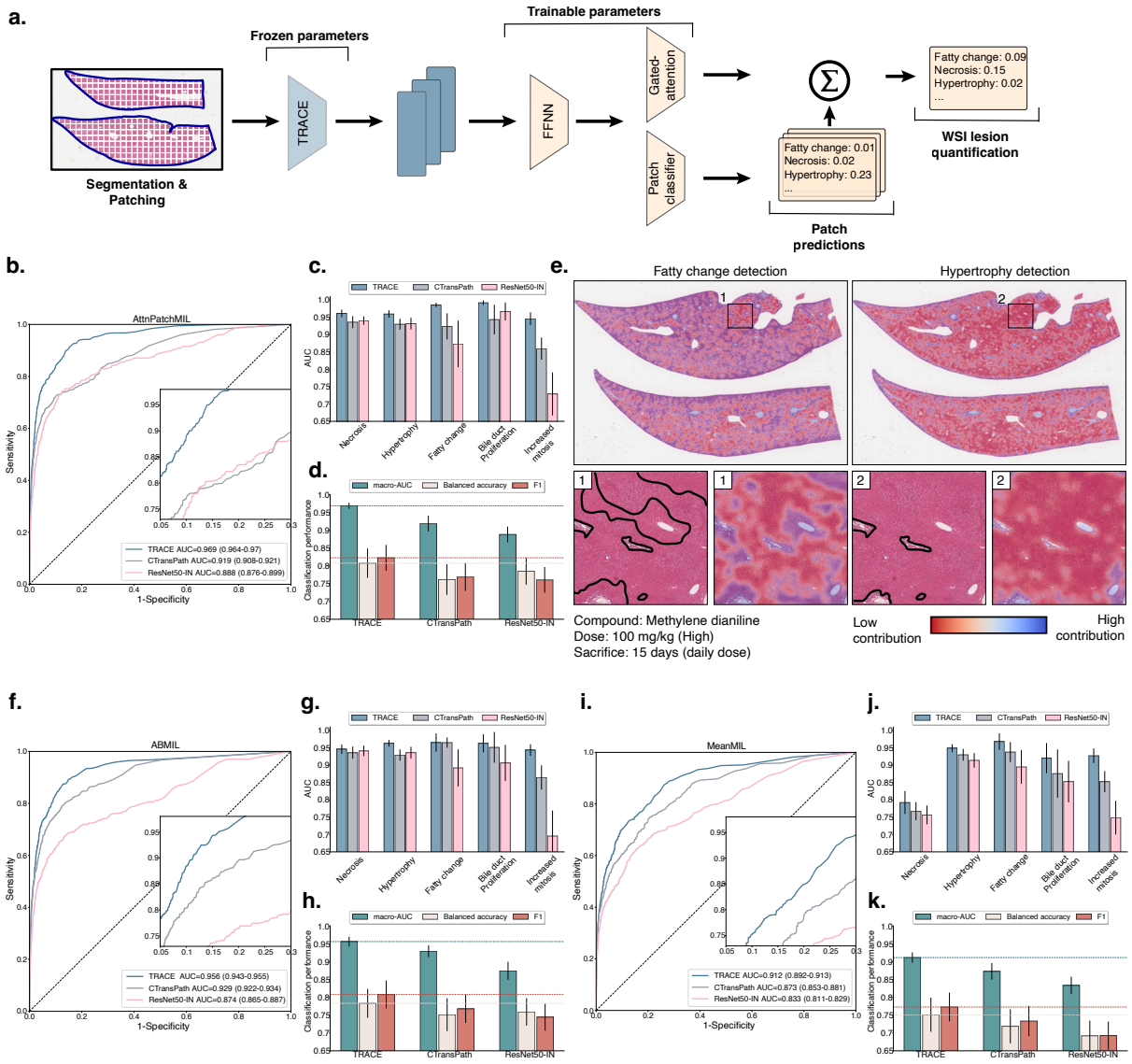

Figure S2: **Weakly supervised slide classification.** **a.** Overview of the proposed multiple instance learning (MIL) architecture, AttnPatchMIL, for joint patch and slide classification using slide-level supervision. **b,c,d.** Evaluation of AttnPatchMIL on 5-class lesion classification comparing TRACE against ResNet50-IN and CTransPath vision encoders. Evaluation based on multi-label ROC and macro-AUC in **b.**; class-wise AUC in **c.** and; overall balanced accuracy, macro-AUC, and F1 in **d.**. **e.** Patch-wise attribution of fatty change and hypertrophy using AttnPatchMIL. **f.,g.,h.** Analogous evaluation using Attention-based MIL (ABMIL). **i.,j.,k.** Analogous evaluation using MeanMIL based on the mean patch embedding. All models are trained on TG-18k and tested on TG-4k. Error bars in **b,c,d,f,g,h,i,j,k.** represent 95% confidence intervals and were computed using non-parametric bootstrapping (100 iterations). ROC: receiver operating characteristic; AUC: area under the ROC curve; FFNN: feed-forward neural network.

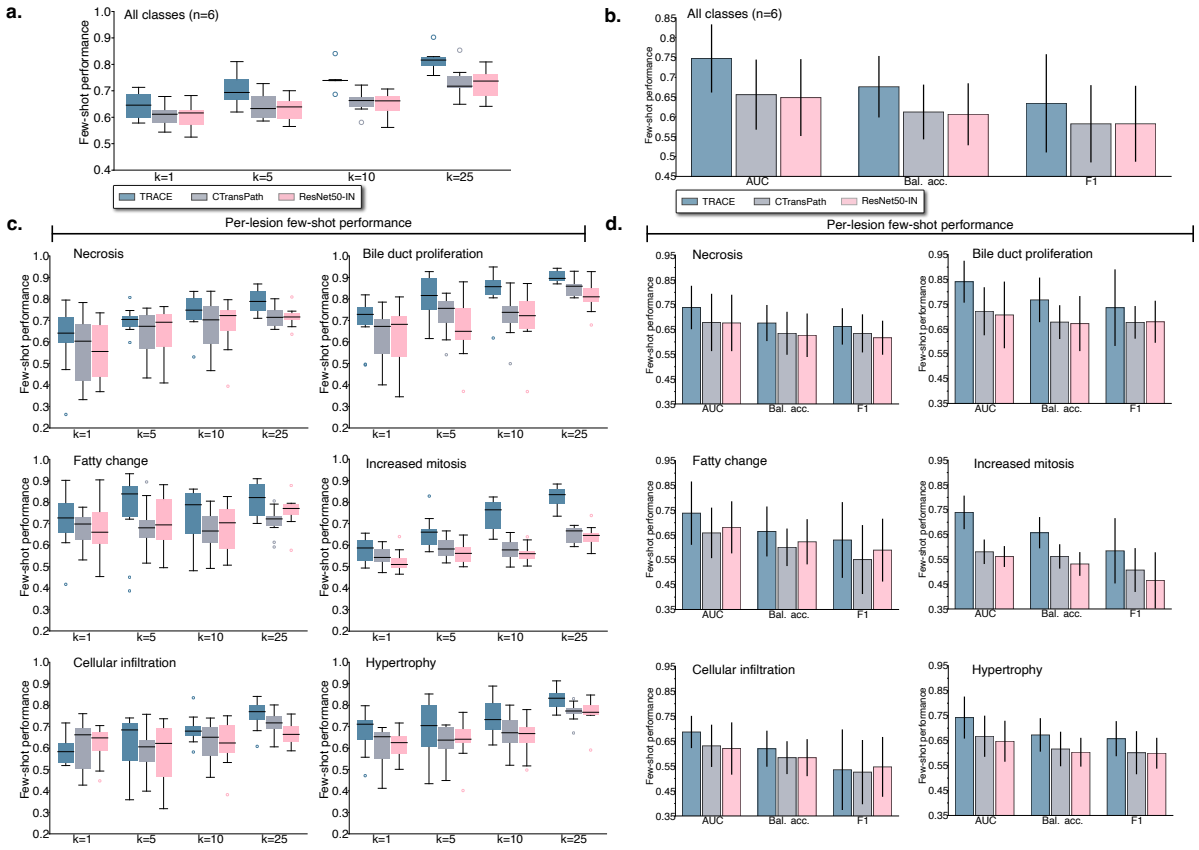

Figure S3: **Few-shot classification. a.** Comparison of TRACE, CTransPath and ResNet50-IN vision encoders for few-shot learning ($k \in 1, 5, 10, 25$) in TG-4k evaluated AUC averaged across six binary classification tasks. **b.** Comparison of TRACE, CTransPath and ResNet50-IN vision encoders for $k = 10$ in TG-4k evaluated using macro-AUC, balanced accuracy and F1 score. **c.** Per-class (n=6 lesions) few-shot performance evaluated using macro AUC. **d.** Per-class (n=6 lesions) few-shot performance evaluated using macro AUC with $k = 10$. Error bars in **a,c.** represent 95% confidence intervals and were computed using non-parametric bootstrapping (100 iterations). Error bars in **b,d.** represent standard deviation and were computed using classification performance repeated over 10 runs, where each run samples a different random set of k training samples per class. AUC: area under the ROC curve; Bal. acc.: balanced accuracy.

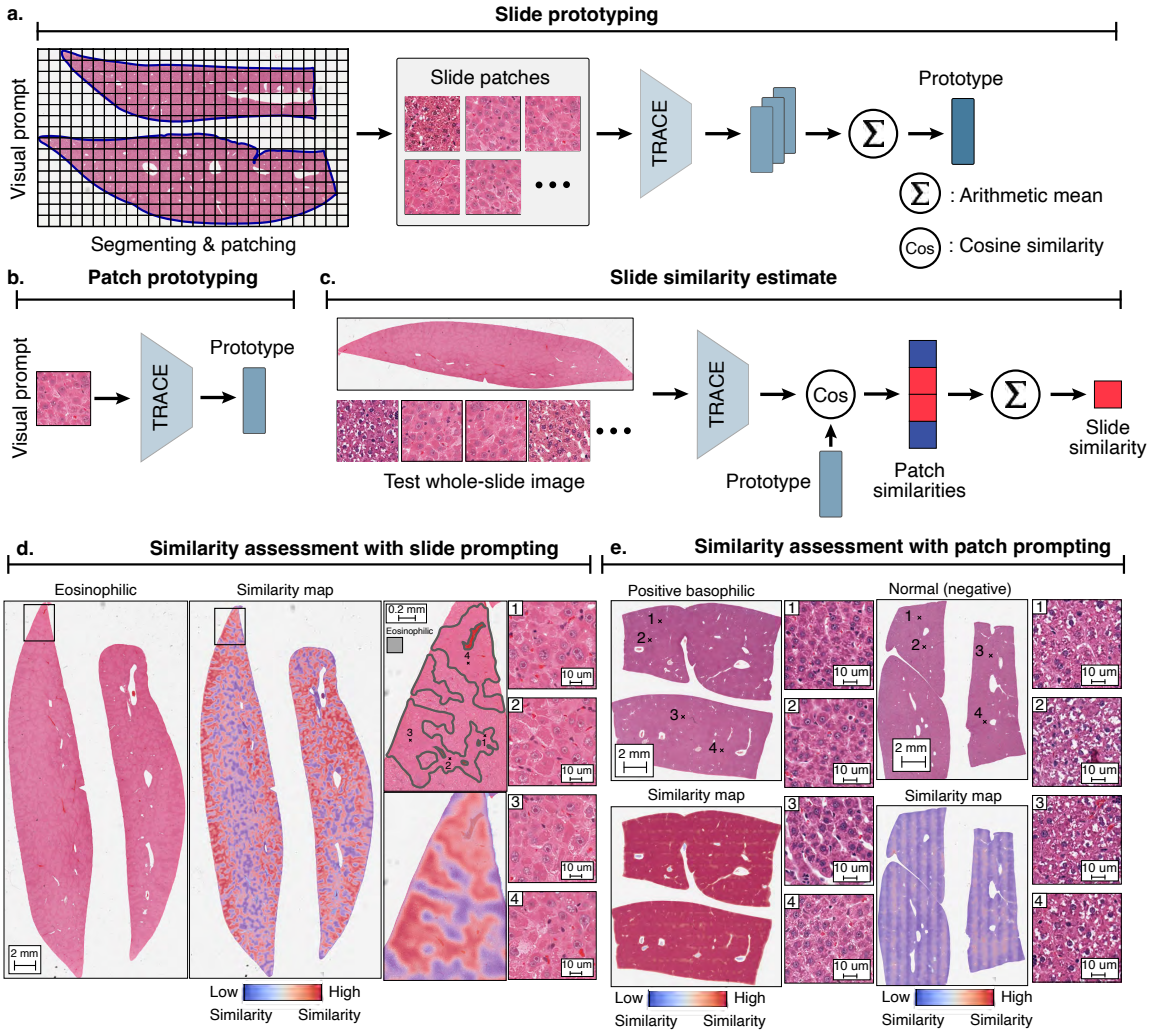

Figure S4: **Patch and slide-level prototyping. a.** A slide-level prototype is defined by taking the average TRACE patch embeddings of a slide that contains a distinct morphology of interest. **b.** A patch-level prototype is defined as a single TRACE patch embedding that contains morphology of interest. **c.** A similarity score is defined by computing the average cosine distance between the prototype (patch or slide) and all patch embeddings from a test slide, with the binary prediction (positive if the slide contains morphology of interest, and negative otherwise) threshold determined by Otsu method. **d.** Similarity assessment using single-slide prompting for detecting eosinophilic cellular alteration in thioacetamide and basophilic cellular alteration in puromycin aminonucleoside. Visualization of the patch-level similarity with the prototype yields a pseudo-segmentation map indicating the presence of the morphology of interest. In a high-dose thioacetamide slide, annotations of eosinophilic regions align almost perfectly with the patch-level similarity. **e.** Similarity assessment using single-patch prompting classification of basophilic cellular alteration. High similarity is identified between the positive basophilic prompt and the positive test slide (center), and low similarity with the negative (normal) slide (right).

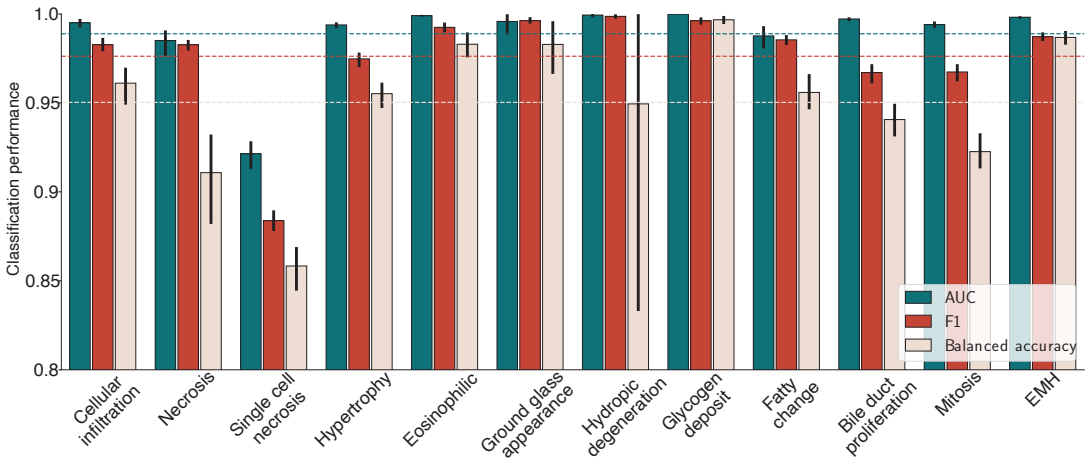

Figure S6: **Patch classification evaluation.** Class-wise macro-AUC, balanced accuracy and F1 score of TRACE (FT). Error bars represent 95% confidence intervals and were computed using non-parametric bootstrapping (100 iterations). EMH: extramedullary hematopoiesis; Eosinophilic: eosinophilic cellular alteration.

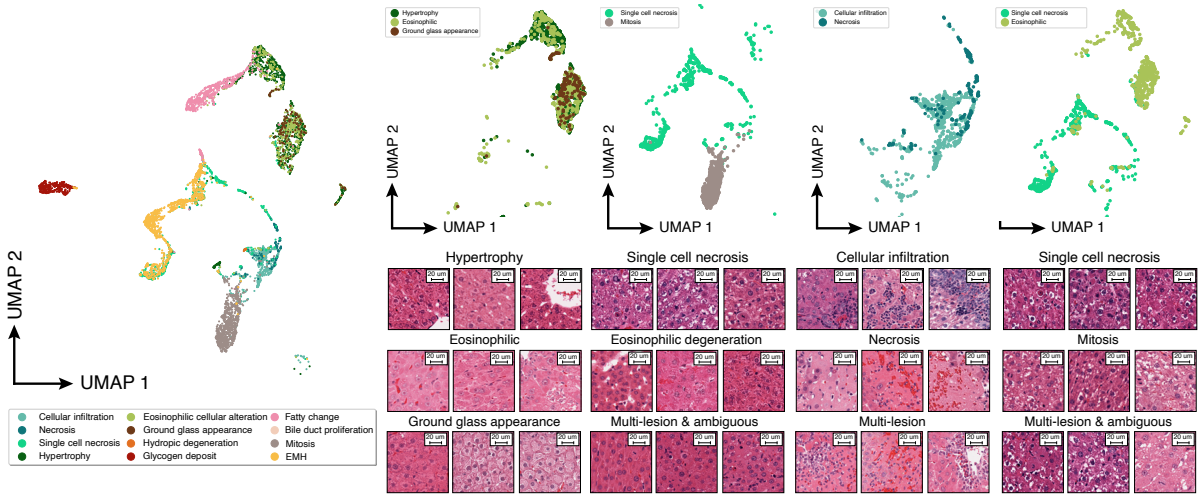

Figure S7: **Visualization of TRACE (FT) embedding space.** Uniform Manifold Approximation and Projection (UMAP) of TRACE (FT) patch embedded colored by their annotation. All shown patches are from TG-4k. Zoom on specific regions of the latent space with patch examples. Each patch is 256×256 pixels at 20×. EMH: extramedullary hematopoiesis. Eosinophilic: eosinophilic cellular alteration.

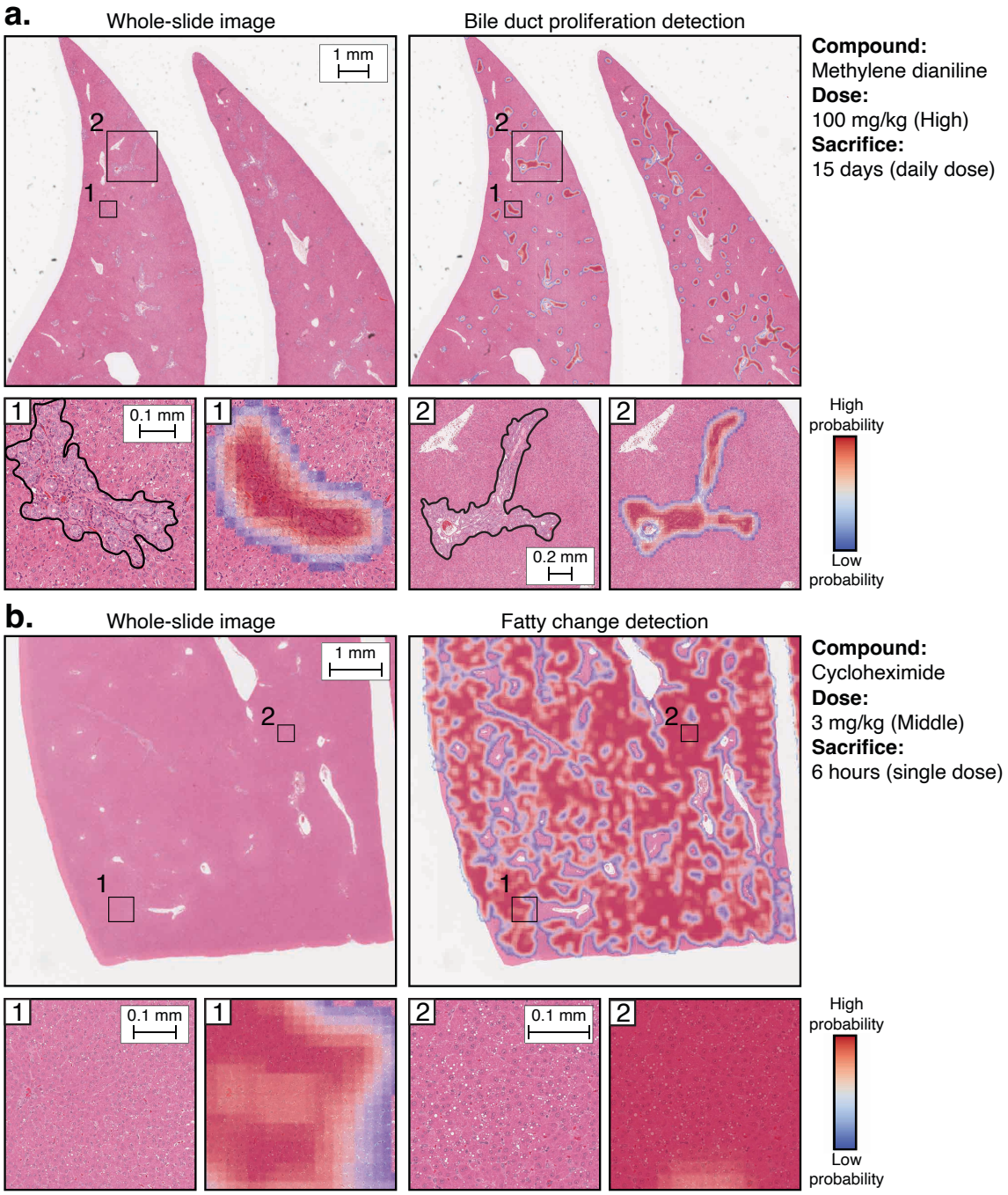

Figure S8: **Visualization of TRACE fine-tuned with patch annotations. a.** Example of lesion detection and segmentation using TRACE (FT) in a high dose sample administered with methylene dianiline. Regions highlight bile duct proliferation. **b.** Example of lesion detection and segmentation using TRACE (FT) in a middle dose sample administered with cycloheximide. Regions highlight hepatocellular fatty change. All results were obtained using 80% patch overlap.

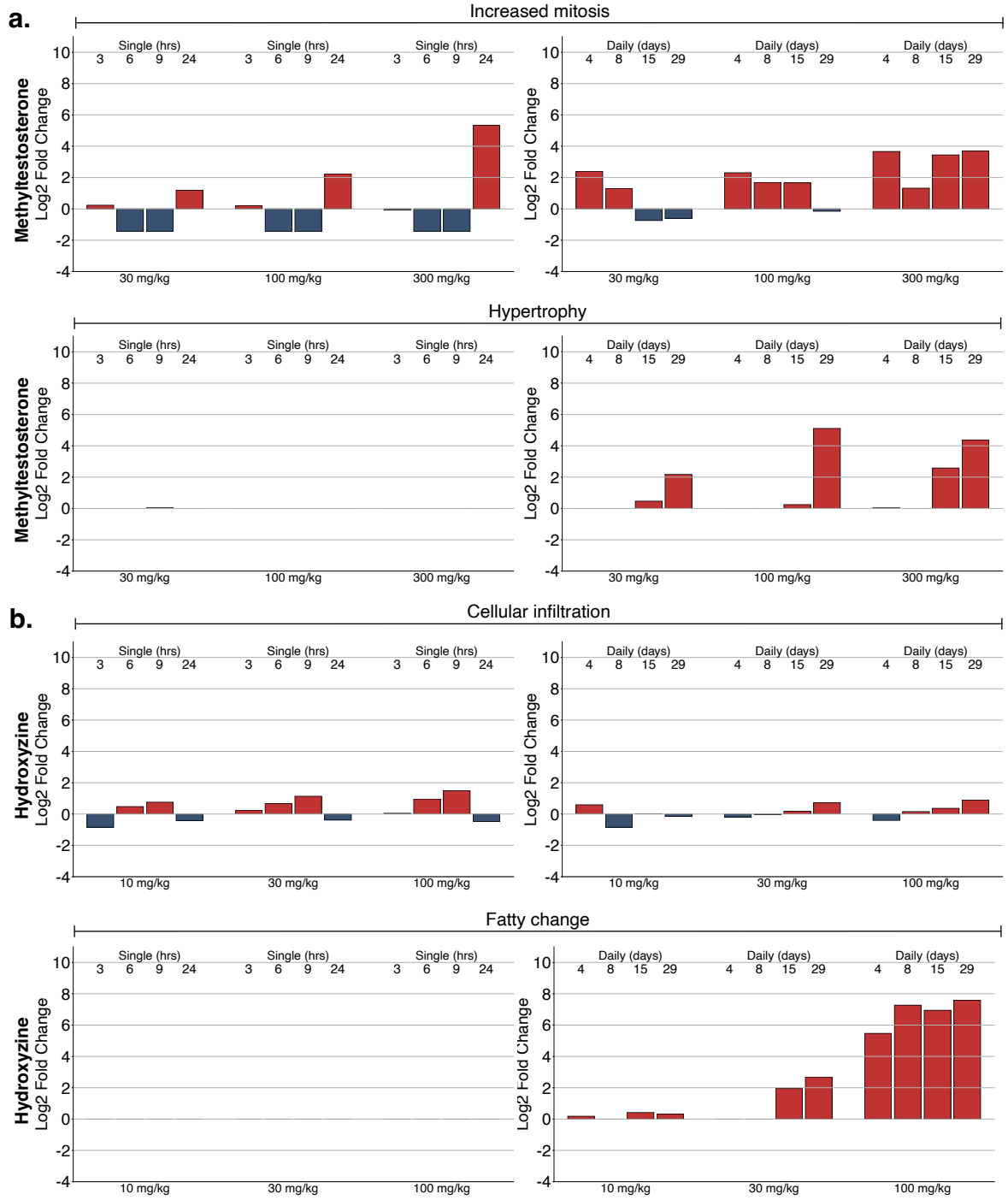

Figure S9: **Automatic dose-response toxicity assessment. a.** Morphological lesion log2 fold change of increased mitosis and hypertrophy in methyltestosterone in single and daily dose sample groups with respect to the control group. **b.** Log2 fold change of cellular infiltration and fatty change in hydroxyzine in single and daily doses with respect to the control group. Plots without bars indicate that no lesion was detected.

	Train/Validation		**Test**	
Lesions	WSIs	Compounds	WSIs	Compounds
Necrosis	748	101	264	17
Hypertrophy	1649	109	331	18
Fatty change	308	33	73	7
Bile duct proliferation	111	10	41	2
Increased mitosis	331	43	94	9

Table S1: **Distribution of lesions used in weakly supervised slide classification.** TG-GATEs includes 23,136 liver WSIs from 157 different pre-clinical studies. Necrosis refers to single-cell, focal/multifocal, or zonal hepatocellular necrosis; hypertrophy refers to enlarged hepatocytes, primarily due to an increase in the cytosolic protein or number of organelles; fatty change includes macro and microvesicular hepatocellular vacuolation; bile duct proliferation refers to bile duct hyperplasia with an increased number of bile ducts; we additionally include oval cell proliferation. Increased mitosis refers to hepatocyte mitosis above normal levels. Complementary information describing each lesion is provided in **table S2** WSIs, whole-slide images.

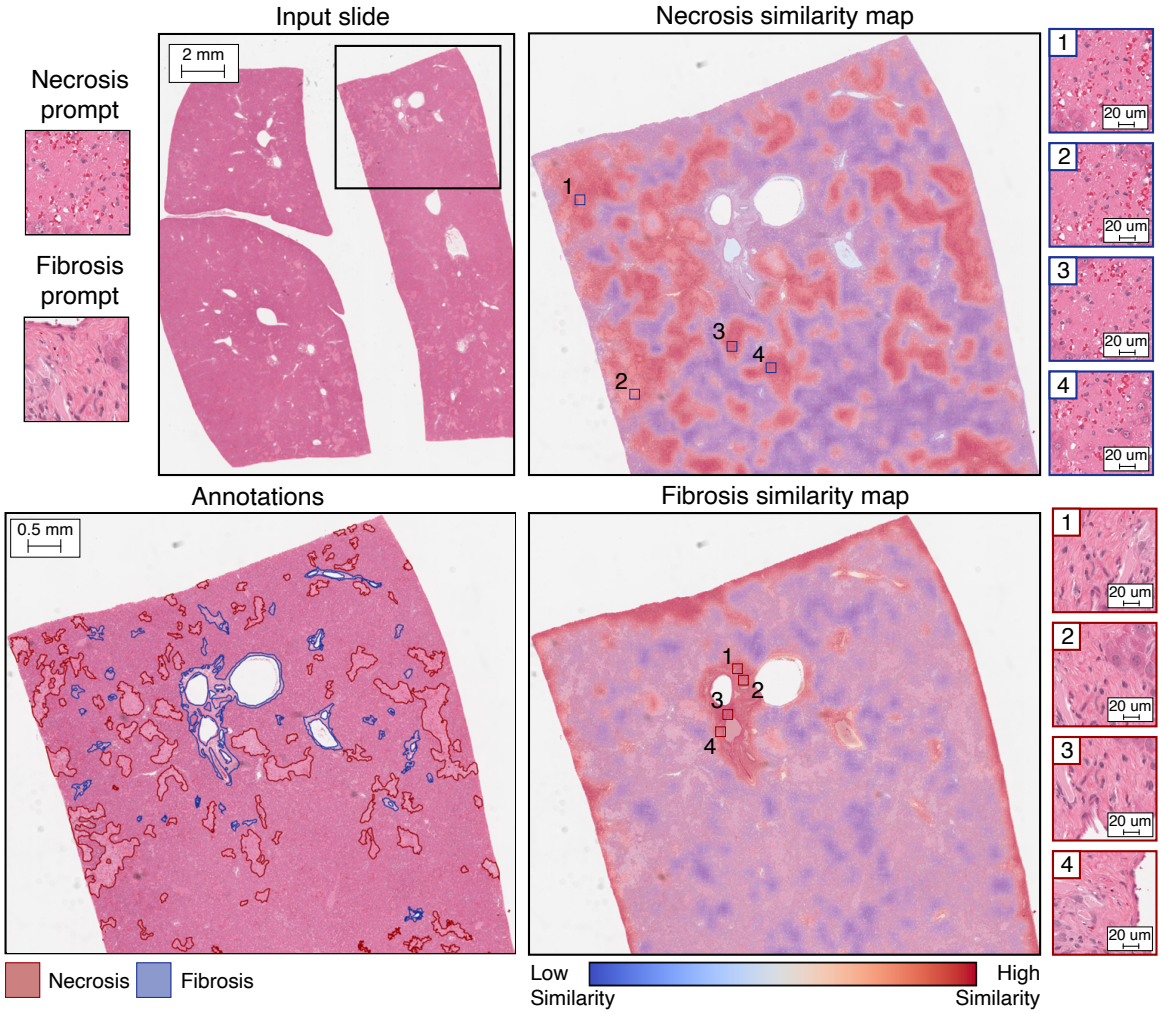

Figure S5: **Similarity assessment with single-patch prompting.** Detection of necrosis and fibrosis using single-patch prompting upon daily administration (15 days) of 30 mg/kg of monocrotaline.

Lesion	Definition	Used in
Hepatocellular responses, cellular degeneration, injury, and death		
Fatty change	Hepatocellular vacuolation, consistent with intracytoplasmic lipid accumulation. Includes macro and microvesicular.	WSC, FSL, PC, RS
Hydropic degeneration	Hepatocellular vacuolation, consistent with intracytoplasmic fluid accumulation.	PC
Necrosis	Cell death of hepatocytes. Includes focal/multifocal and zonal (centrilobular, midzonal, periportal and diffuse).	WSC, FSL, MR, PC, RS
Single-cell necrosis	Necrosis (or apoptosis) of single hepatocytes.	PC, RS
Hypertrophy	Enlargement of the hepatocyte cytoplasm, secondary to increase in the cytosolic protein or number of organelles.	WSC, FSL, PC, RS
Glycogen deposit	Hepatocellular cytoplasmic alteration, consistent with glycogen accumulation.	PC, RS
Inflammatory cell infiltrates, inflammatory cell infiltrates and hepatic		
Cellular infiltration	Infiltrations of inflammatory cells in the liver. Includes neutrophil, mononuclear and peribiliary.	FSL, PC, RS
Fibrosis	The presence of fibrous connective tissue in the liver above normal levels. Includes pericellular, peribiliary, and postnecrotic fibrosis.	MR
Non-neoplastic proliferative lesions		
Basophilic cellular alteration	Hepatocellular cytoplasmic basophilia, due to free ribosomes or rough endoplasic reticulum.	MR
Eosinophilic cellular alteration	Hepatocellular cytoplasmic eosinophilia, due to an increase in cytoplasmic organelles.	PC
Ground glass appearance	Hepatocytes with glassy and hypereosinophilic appearing cytoplasm	PC
Bile duct proliferation	Increased number of small bile ducts arising in portal region. Biliary epithelium appears normal or may show degenerative or atrophic changes.	WSC, FSL, PC, RS
Increased mitosis	Increased hepatocyte mitoses above normal background levels ($>$1-2 mitotic figures per 10 (400x) HPF).	WSC, FSL, PC, RS
Other lesions		
Extramedullary hematopoiesis	Hematopoietic cell proliferation in the liver. Aggregates of hematopoietic cells are distributed in the hepatic sinusoids as well as around central veins and portal areas.	PC, RS

Table S2: **Lesion definition.** Definitions and scope are based on the INHAND guidelines[40]. INHAND is the International Harmonization of Nomenclature and Diagnostic Criteria, a publicly accessible resource that defines guidelines to diagnose lesions found in rodent toxicity and carcinogenicity studies. When INHAND lacked specific guidelines regarding diagnosing, such as for scoring "increased mitosis", we relied on the National Toxicology Program guidelines available online `https://ntp.niehs.nih.gov/atlas/nnl/hepatobiliary-system/liver`. WSC: weakly supervised classification; FSL: few-shot learning; MR: morphological retrieval; PC: patch classification, RS: reader study.

Lesion	Patch Positive	Patch Negative	WSI Positive	WSI Negative
Cellular infiltration	2,232	3,315	984	1,167
Necrosis	2,449	14,453	243	538
Single cell necrosis	4,119	11,947	780	1,393
Hypertrophy	4,525	5,451	29	698
Eosinophilic cellular alteration	1,231	3,123	168	632
Ground glass appearance	1,440	4,294	91	698
Hydropic degeneration	418	868	46	145
Glycogen deposit	2,730	320	608	73
Fatty change	2,127	6,935	144	1,019
Bile duct proliferation	1,564	1,630	94	96
Increased mitosis	3,496	8,978	404	1,079
Extramedullary hematopoiesis	3,111	9,400	1,177	2,954

Table S3: **Distribution of patch annotations.** As many patches include several labels, such as fatty change and hypertrophy, we report positive patches (lesion is present) and the number of hard negatives (lesion is not present). Positive patches may include more than one lesion.

Slide ID	Compound	Dose	Single or Repeat	Sacrifice
10921	Griseofulvin	1000	Single	3 hrs
11089	Griseofulvin	1000	Single	24 hrs
13326	Metformin	0	Single	9 hrs
14012	Methyldopa	200	Single	3 hrs
14430	Methyldopa	0	Repeat	15 days
14478	Methyldopa	60	Repeat	15 days
19714	Hydroxyzine	30	Repeat	15 days
20367	Mefenamic acid	300	Repeat	8 days
20422	Mefenamic acid	300	Repeat	15 days
27539	Thioacetamide	0	Single	6 hrs
30270	Bromobenzene	300	Single	24 hrs
30272	Bromobenzene	300	Single	24 hrs
30274	Bromobenzene	300	Single	24 hrs
31778	Cyclophosphamide	15	Repeat	15 days
33012	Metformin	300	Repeat	4 days
33056	Metformin	1000	Repeat	4 days
33057	Metformin	1000	Repeat	4 days
46078	Ethionamide	300	Single	9 hrs
46360	Ethionamide	100	Repeat	4 days
48472	Phenacetin	0	Single	24 hrs
48744	Phenacetin	2000	Single	6 hrs
48936	Phenacetin	100	Repeat	4 days
49066	Phenacetin	1000	Repeat	8 days
50226	Danazol	100	Repeat	15 days
50987	Cisplatin	0	Single	6 hrs
51434	Cisplatin	1	Repeat	4 days
51604	Carboplatin	0	Single	6 hrs
53169	Bromoethylamine	6	Repeat	29 days
53257	Bromoethylamine	20	Repeat	29 days
53335	Mexiletine	40	Single	9 hrs
54187	Cephalothin	1000	Single	3 hrs
54409	Cyclosporine a	0	Repeat	4 days
54597	Cyclosporine a	10	Repeat	15 days
54816	Cyclosporine a	100	Repeat	8 days
54822	Cyclosporine a	100	Repeat	15 days

Table S4: **Cases used in the reader study.** A total of 100 tissue sections were randomly extracted from TG-GATEs test set (TG-4k).

Slide ID	Compound	Dose	Single or Repeat	Sacrifice
55649	Gentamicin	30	Repeat	8 days
57174	Danazol	1000	Single	24 hrs
57236	Danazol	2000	Single	6 hrs
57336	Theophylline	0	Single	6 hrs
58909	Cycloheximide	10	Single	3 hrs
59123	Tunicamycin	300	Single	3 hrs
6009	Diazepam	0	Single	3 hrs
6013	Diazepam	0	Single	6 hrs
60435	Isoniazid	2000	Single	6 hrs
62014	Hexachlorobenzene	0	Single	3 hrs
6224	Hexachlorobenzene	0	Repeat	15 days
63748	Methylene dianiline	30	Repeat	8 days
63941	Methylene dianiline	100	Repeat	29 days
63952	Methylene dianiline	100	Repeat	29 days
63955	Methylene dianiline	100	Repeat	29 days
10913	Griseofulvin	300	Single	3 hrs
10969	Griseofulvin	300	Single	6 hrs
11080	Griseofulvin	1000	Single	24 hrs
14008	Methyldopa	200	Single	3 hrs
14023	Methyldopa	200	Single	6 hrs
14211	Methyltestosterone	300	Single	24 hrs
14554	Methyldopa	200	Repeat	29 days
19318	Hydroxyzine	10	Single	24 hrs
19764	Hydroxyzine	100	Repeat	15 days
19773	Hydroxyzine	100	Repeat	15 days
19777	Hydroxyzine	100	Repeat	29 days
19781	Hydroxyzine	100	Repeat	29 days
19785	Hydroxyzine	100	Repeat	29 days
20370	Mefenamic acid	300	Repeat	8 days
20414	Mefenamic acid	300	Repeat	15 days
20477	Mefenamic acid	300	Repeat	29 days
20483	Mefenamic acid	300	Repeat	29 days
2421	Isoniazid	100	Single	6 hrs
2589	Isoniazid	0	Repeat	4 days
30278	Bromobenzene	300	Single	24 hrs

Table S4: Continuation of table S4.

Slide ID	Compound	Dose	Single or Repeat	Sacrifice
30465	Bromobenzene	30	Repeat	15 days
33006	Metformin	100	Repeat	29 days
33077	Metformin	1000	Repeat	15 days
46792	Thioacetamide	45	Repeat	4 days
48569	Phenacetin	300	Single	9 hrs
48685	Phenacetin	1000	Single	9 hrs
48998	Phenacetin	100	Repeat	15 days
49064	Phenacetin	1000	Repeat	8 days
51220	Cisplatin	3	Single	24 hrs
52036	Carboplatin	0	Repeat	15 days
53036	Bromoethylamine	2	Repeat	29 days
53941	Cephalothin	0	Single	3 hrs
53997	Cephalothin	0	Single	9 hrs
54300	Cephalothin	2000	Single	3 hrs
54937	Puromycin aminonucleoside	0	Single	9 hrs
54991	Gentamicin	0	Single	24 hrs
55355	Puromycin aminonucleoside	120	Single	6 hrs
55696	Gentamicin	30	Repeat	15 days
55728	Puromycin aminonucleoside	40	Repeat	8 days
56895	Danazol	0	Single	6 hrs
57802	Theophylline	0	Repeat	8 days
58319	Acetazolamide	200	Single	9 hrs
58819	Cycloheximide	3	Single	6 hrs
59224	Tunicamycin	300	Single	24 hrs
60413	Isoniazid	200	Single	6 hrs
60648	Cyclophosphamide	50	Single	9 hrs
62091	Hexachlorobenzene	1000	Single	6 hrs
6302	Hexachlorobenzene	30	Repeat	29 days
63875	Methylene dianiline	100	Repeat	15 days
70760	Phenacetin	1000	Repeat	8 days

Table S4: Continuation of table S4.

Hyper-parameter	Value
Layers	12
Heads	12
Patch size	16
FFN layer	MLP
Head activation	GELU
Embedding dimension	768
Stochastic dropout 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
Max masking ratio	0.3
Min masking ratio	0.0
Gradient clipping max norm	0.3
Normalize last layer	✓
Shared head	✓
head output dimension	8192
Optimizer	AdamW
Batch size	1024
Freeze last layer (it, ep)	44124, 3
Warmup (it, ep)	73540, 5
Warmup teacher temperature (it, ep)	441240, 30
Max training (it, ep)	1176640, 80
Number of images	15061790
Learning rate schedule	Cosine
Learning rate (start)	0.0
Learning rate (post warmup)	0.0005
Learning rate (final)	0.000002
Teacher temperature (start)	0.04
Teacher temperature (final)	0.07
Teacher momentum	0.996
Weight decay (start)	0.04
Weight decay (end)	0.4
Automatic mixed precision	fp16

Table S5: **iBOT hyperparameters used in TRACE pretraining**. The training converged after 80 epochs for a total training time of 208 hours using 8 $\times$ 80GB NVIDIA A100 GPUs.

Hyperparameter	Value
Batch size	48
Weight decay	1e-5
AdamW β	(0.9, 0.999)
Peak learning rate	1e-4
Learning rate schedule	Cosine
Epochs	20

Table S6: **Hyperparameters used in slide-level supervised classification**. A single 24GB NVIDIA GeForce RTX 3090 GPU was used to train the multiple instance learning models.