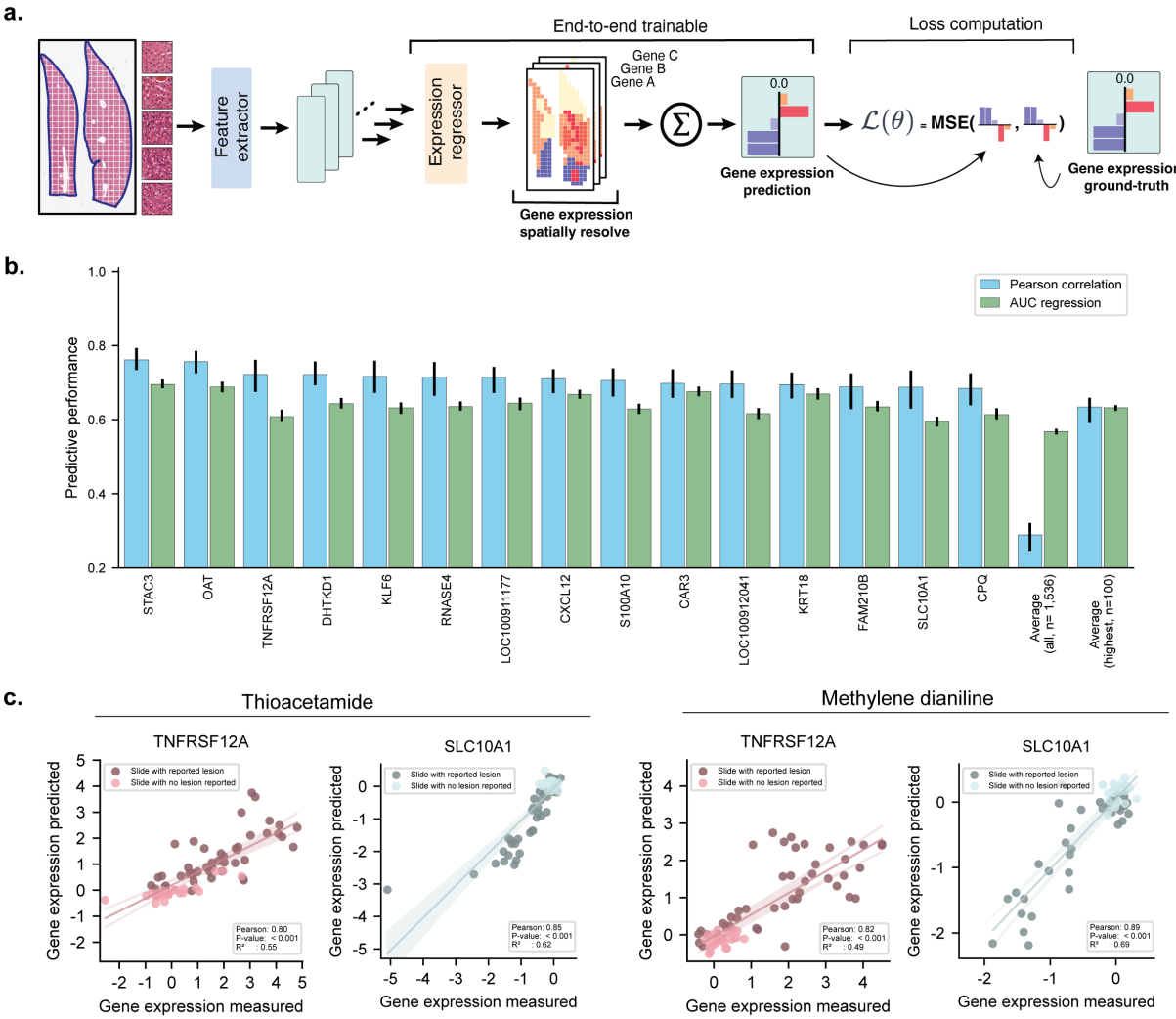

Figure S1: **GEESE model architecture and performance.** **a.** Overview of the gene expression regressor (GEESE) architecture, which can predict expression changes (log2 fold change) from an input tissue section (a whole-slide image). Patch attribution enables deriving a pseudo-spatially-resolve expression map, where each patch becomes associated with a pseudo-expression profile for all predicted genes (n=1,536). **b.** Slide-level gene expression prediction performance evaluated using Pearson correlation and macro-AUC regression on TG-GATEs test set (N=2,002 slides). We report predictive performance for the top 10 predicted genes, the average predictive performance across all genes (n=1,536), and the top 100 best-predicted. Error bars represent 95% confidence intervals using non-parametric bootstrapping (100 iterations). **c.** Expression prediction of *TNFRSF12A* and *SLC10A1* genes on compounds thioacetamide and methylene dianiline. Each dot represents a sample. Samples without lesions are clustered around the origin (small log2 fold change), while samples with lesions are either downregulated (*SLC10A1*) or upregulated (*TNFRSF12A*). P-value derived from testing the two-sided null hypothesis of non-correlation.

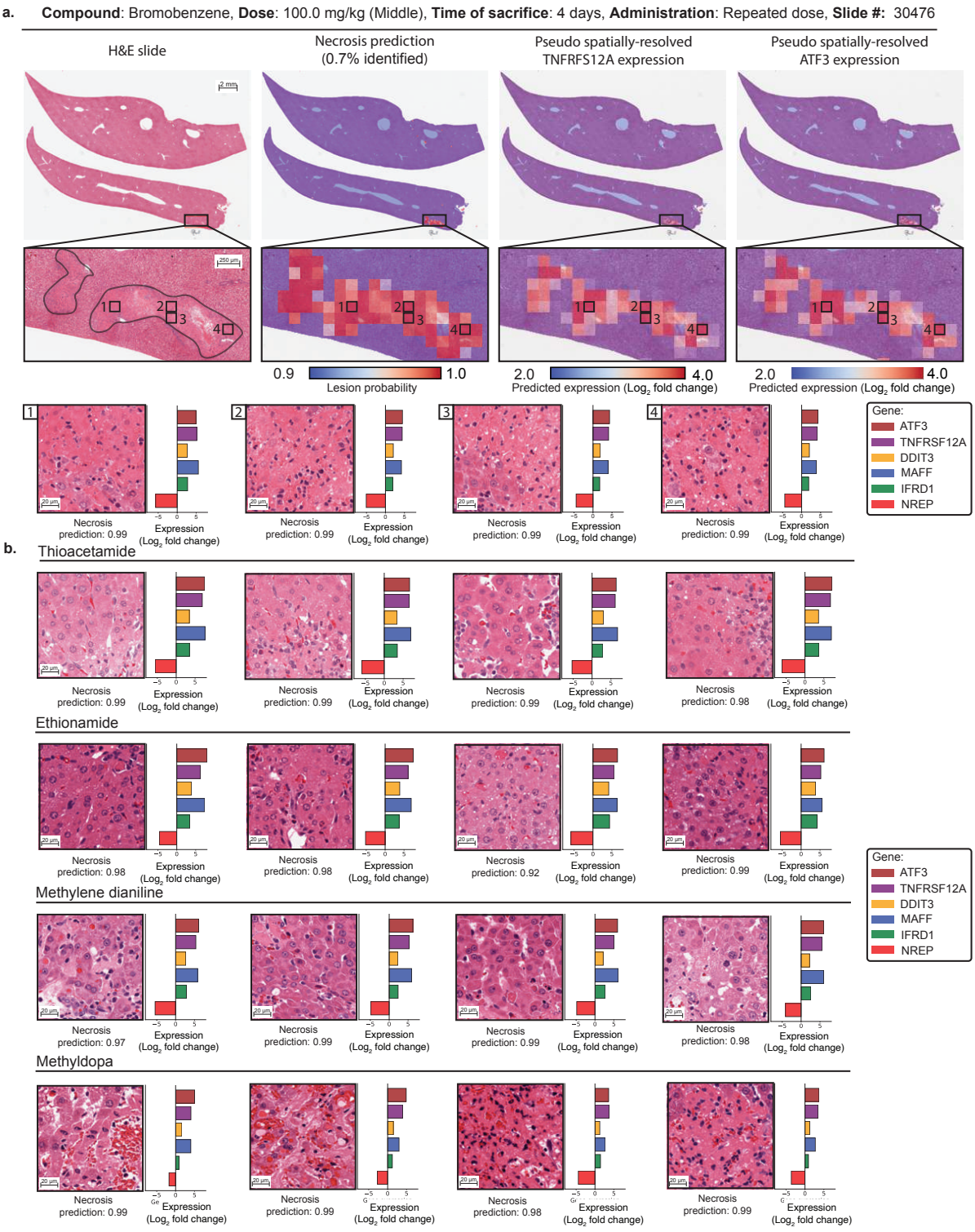

Figure S2: **Caption next page.**

Figure S2: **Pseudo spatially-resolved gene expression and visualization of necrosis. a.** H&E liver section after exposure to bromobenzene (left). Overlay of patch-level necrosis prediction. Predictions below 90% confidence are represented in blue and high-probability predictions are represented in red (center-left). Pseudo-spatially-resolved gene expression heatmaps of genes *TNFRSF12A* (center-right) and *ATF3* (right). **b.** Additional examples of necrotic patches from four studies (thioacetamide, ethionamide, methylene dianiline, and methyldopa) highlight similar upregulation of *TNFRSF12A*, *ATF3*, *DDIT3*, *MAFF* and *IFRD1*, and downregulation of *NREP*.

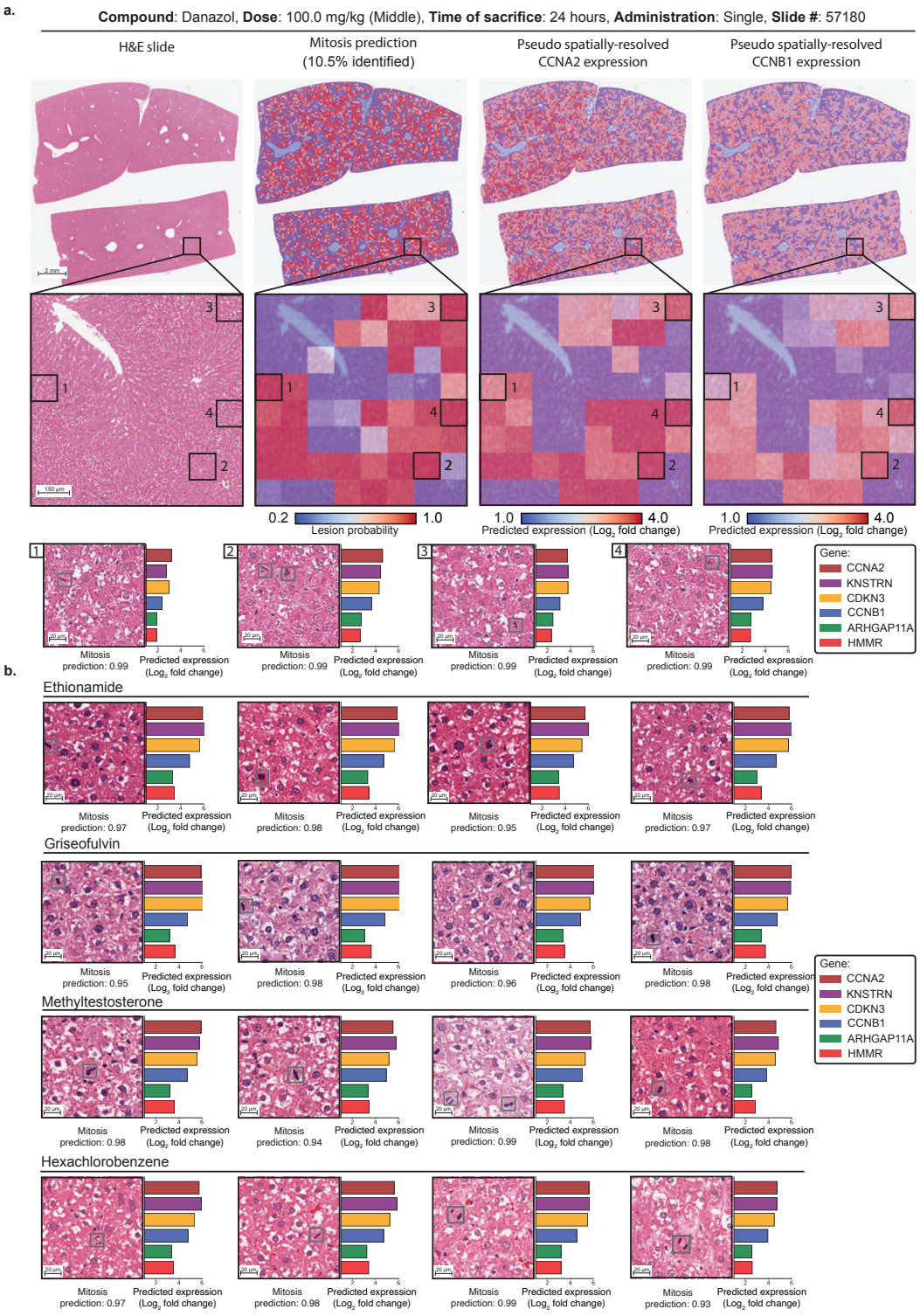

Figure S3: **Caption next page.**

Figure S3: **Pseudo spatially-resolved expression and visualization of mitosis. a.** H&E liver section after exposure to danazol (left). Overlay of patch-level bile duct proliferation prediction. Predictions below 20% confidence are represented in blue, and high-probability predictions are represented in red (center-left). Pseudo-spatially-resolved gene expression heatmaps of *CCNA2* and *CCNB1* (center-right and right). **b.** Additional examples of patches with mitosis from four studies (ethionamide, griseofulvin, methyltestosterone, and hexachlorobenzene) highlight similar upregulation of *CCNA2* and *CCNB1*, *CDKN3*, *KNSTRN*, *ARHGAP11A* and *HMMR*.

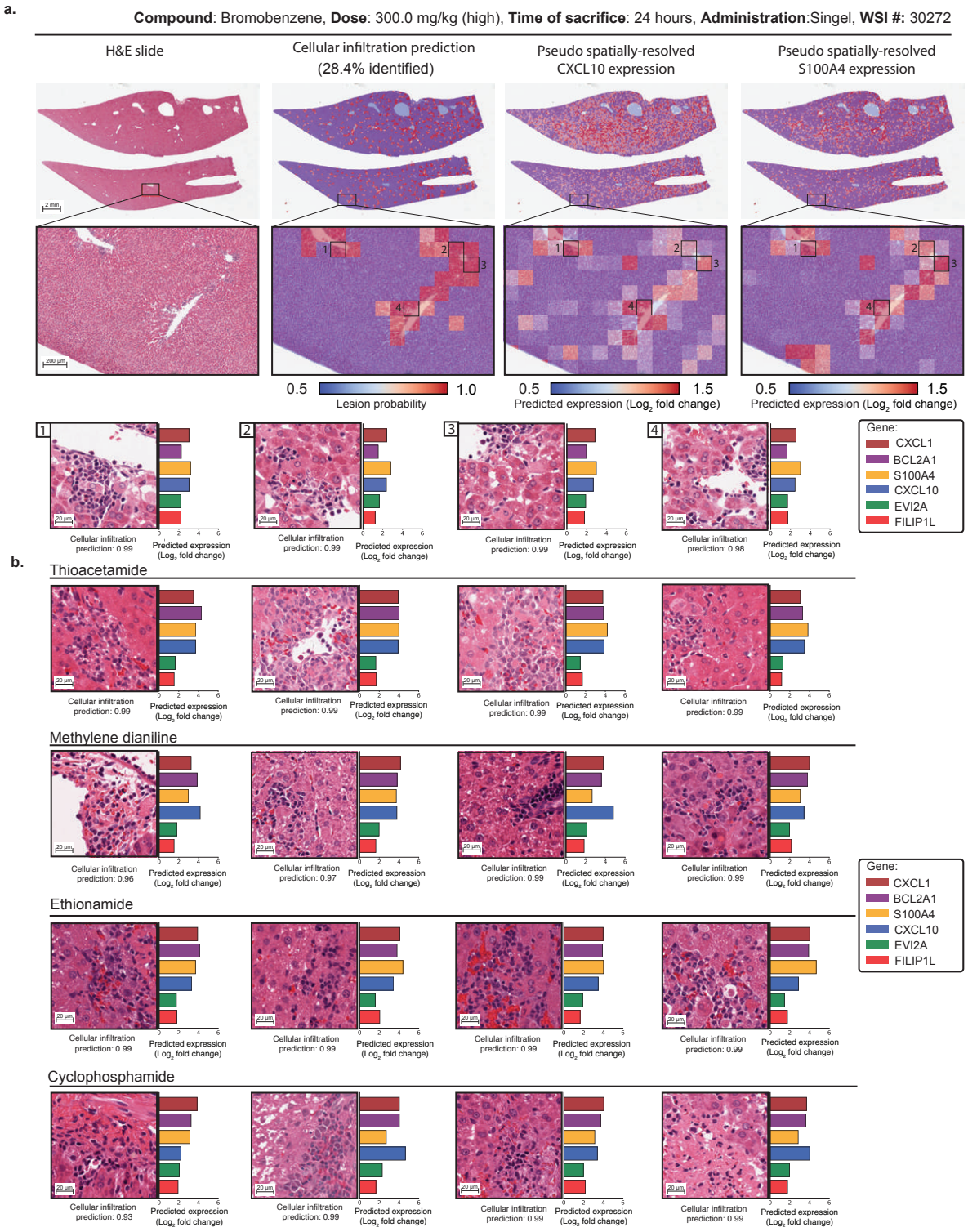

Figure S4: **Caption next page.**

Figure S4: **Pseudo spatially-resolved expression and visualization of cellular infiltration. a.** H&E liver section after exposure to bromobenzene (left). Overlay of patch-level cellular infiltration prediction. Predictions below 50% confidence are represented in blue and high-probability predictions are represented in red (center-left). Pseudo-spatially-resolved gene expression heatmaps of *CXCL10* and *S100A4* (center-right and right). **b.** Additional examples of patches with cellular infiltration from four other studies (thioacetamide, methylene dianiline, ethionamide, and cyclophosphamide) highlight similar upregulation of *CXCL10*, *S100A4*, *CXCL1*, *BCL2A1*, *EVI2A* and *FILIP1L*.

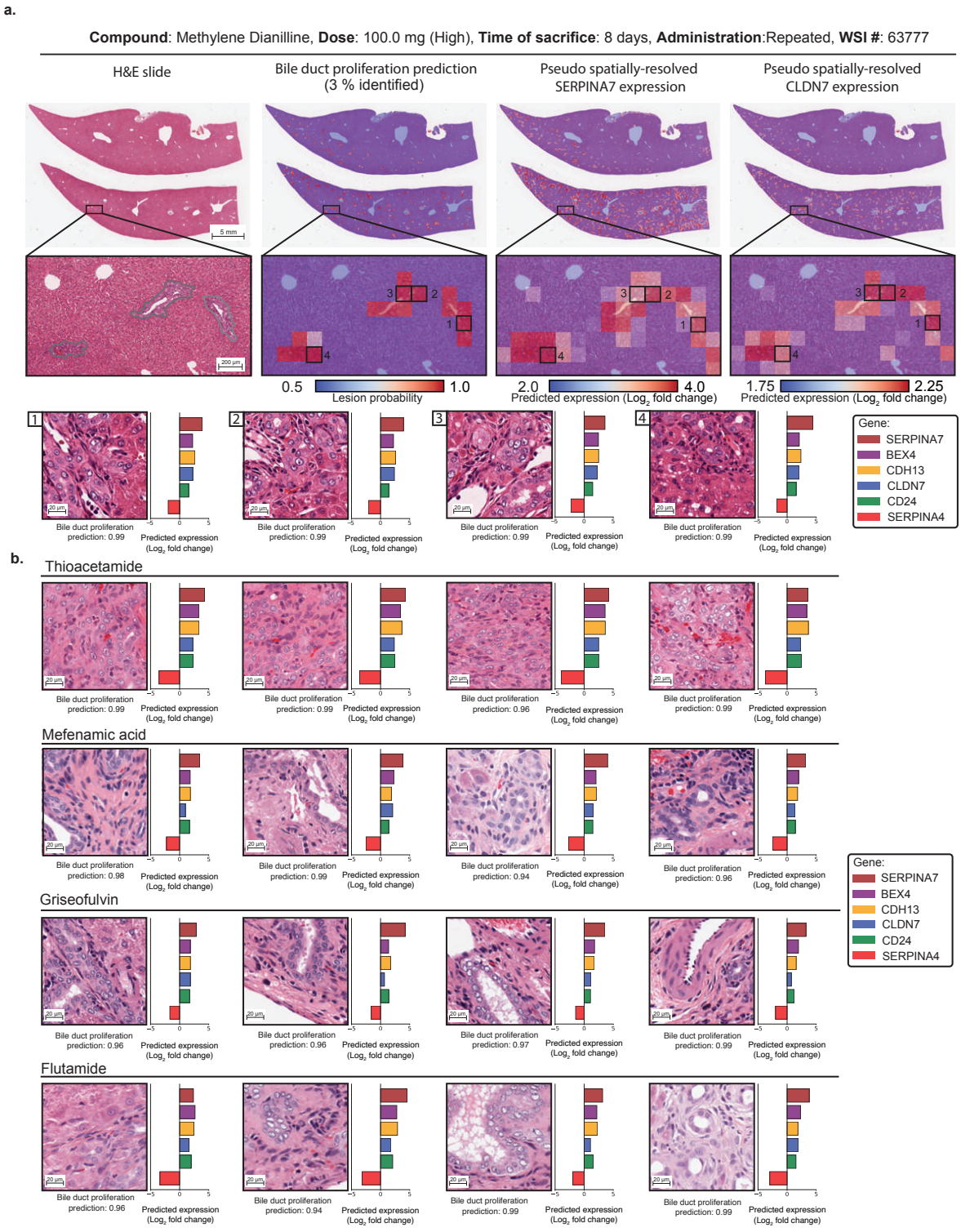

Figure S5: **Caption next page.**

Figure S5: **Pseudo spatially-resolved expression and visualization of bile duct proliferation. a.** H&E liver section after exposure to methylene dianiline (left). Overlay of patch-level bile duct proliferation prediction. Predictions below 50% confidence are represented in blue and high-probability predictions are represented in red (center-left). Pseudo-spatially-resolved gene expression heatmaps of *SERPINA7* and *CLDN7* (center-right and right). **b.** Additional examples of patches with bile duct proliferation from four other studies (thioacetamide, mefenamic acid, griseofulvin and flutamide) highlight similar upregulation of *SERPINA7*, *CLDN7*, *BEX4*, *CDH13* and *CD24*, and downregulation of *SERPINA4*.

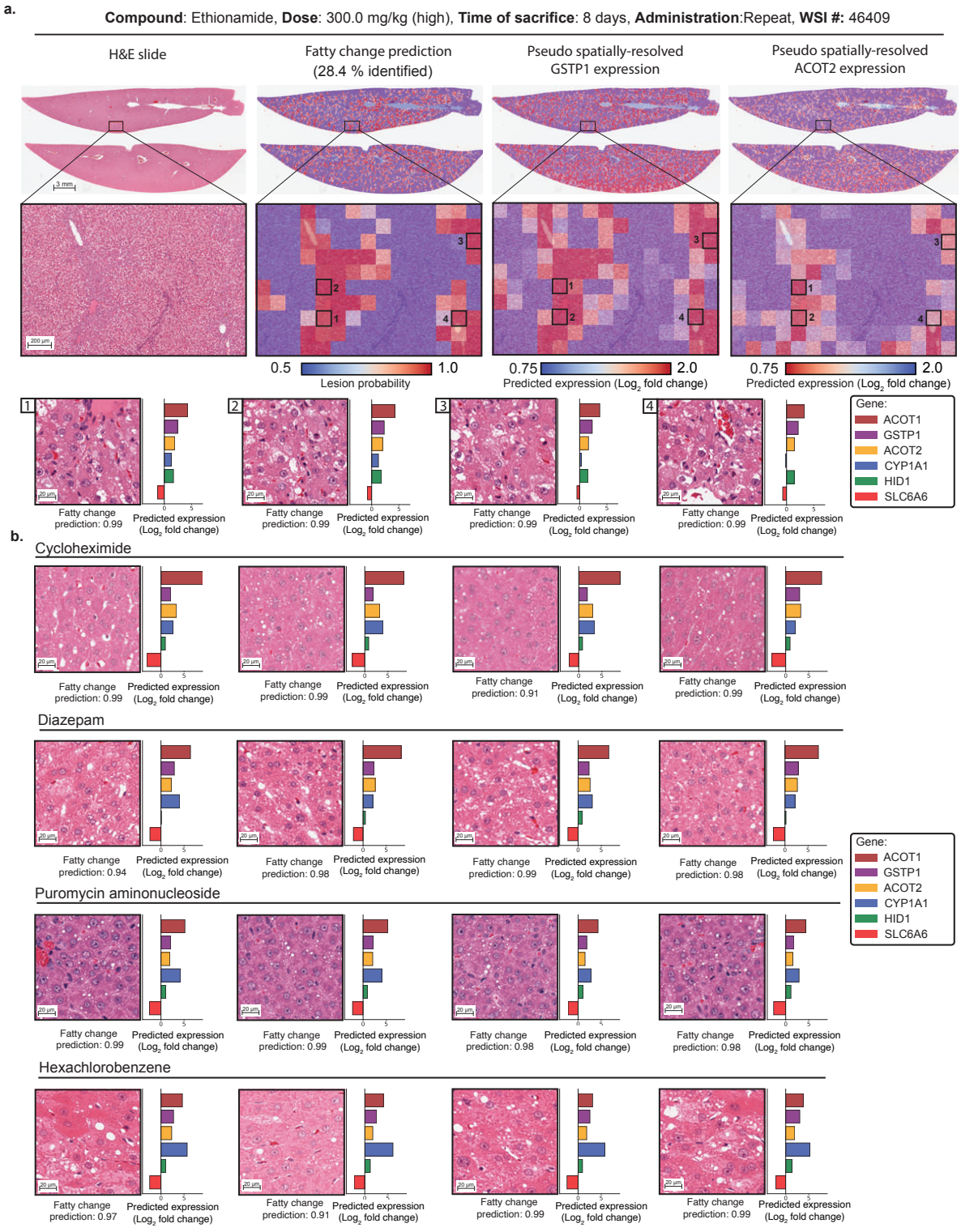

Figure S6: **Caption next page.**

Figure S6: **Pseudo spatially-resolved expression and visualization of fatty change. a.** H&E liver section after exposure to ethionamide (left). Overlay of patch-level fatty change prediction. Predictions below 50% confidence are represented in blue and high-probability predictions are represented in red (center-left). Pseudo-spatially-resolved gene expression heatmaps of *GSTP1* and *ACOT2* (center-right and right). **b.** Additional examples of patches with fatty change from four other studies (cycloheximide, diazepam, puromycin aminonucleoside and hexachlorobenzene) highlight similar upregulation of *GSTP1*, *ACOT2*, *ACOT1*, *HID1* and *CYP1A1* and downregulation of *SLC6A6*.

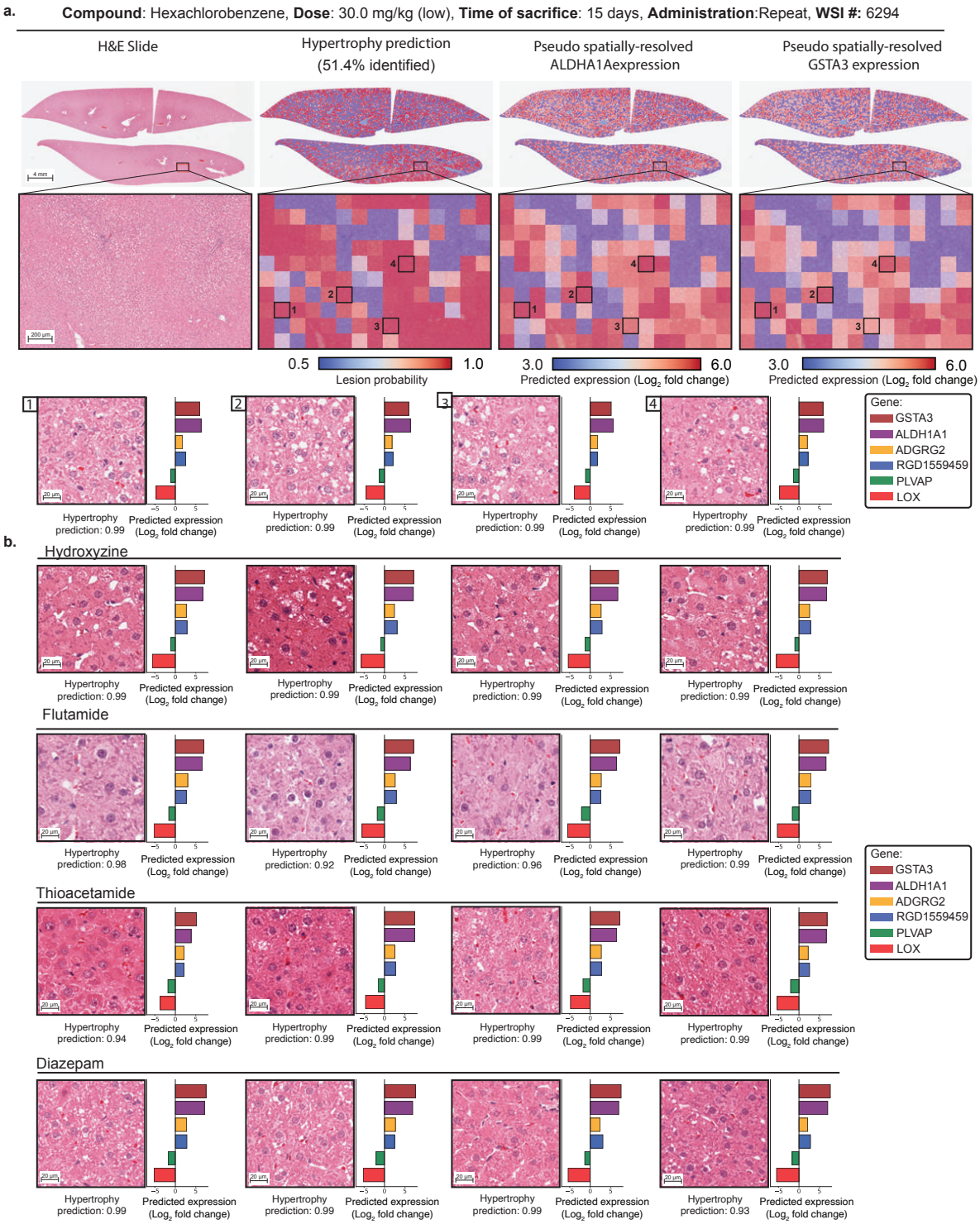

Figure S7: **Caption next page.**

Figure S7: **Pseudo spatially-resolved expression and visualization of hypertrophy. a.** H&E liver section after exposure to hexachlorobenzene (left). Overlay of patch-level hypertrophy prediction. Predictions below 50% confidence are represented in blue and high-probability predictions are represented in red (center-left). Pseudo-spatially-resolved gene expression heatmaps of ALDHA1A and *GSTA3* (center-right and right). **b.** Additional examples of patches with hypertrophy from four other studies (hydroxyzine, flutamide, thioacetamide, and diazepam) highlight similar upregulation of *GSTA3*, *ALDH1A1*, *ADRG2* and *RGD1559459*, and downregulation of *PLAVP* and *LOX*.

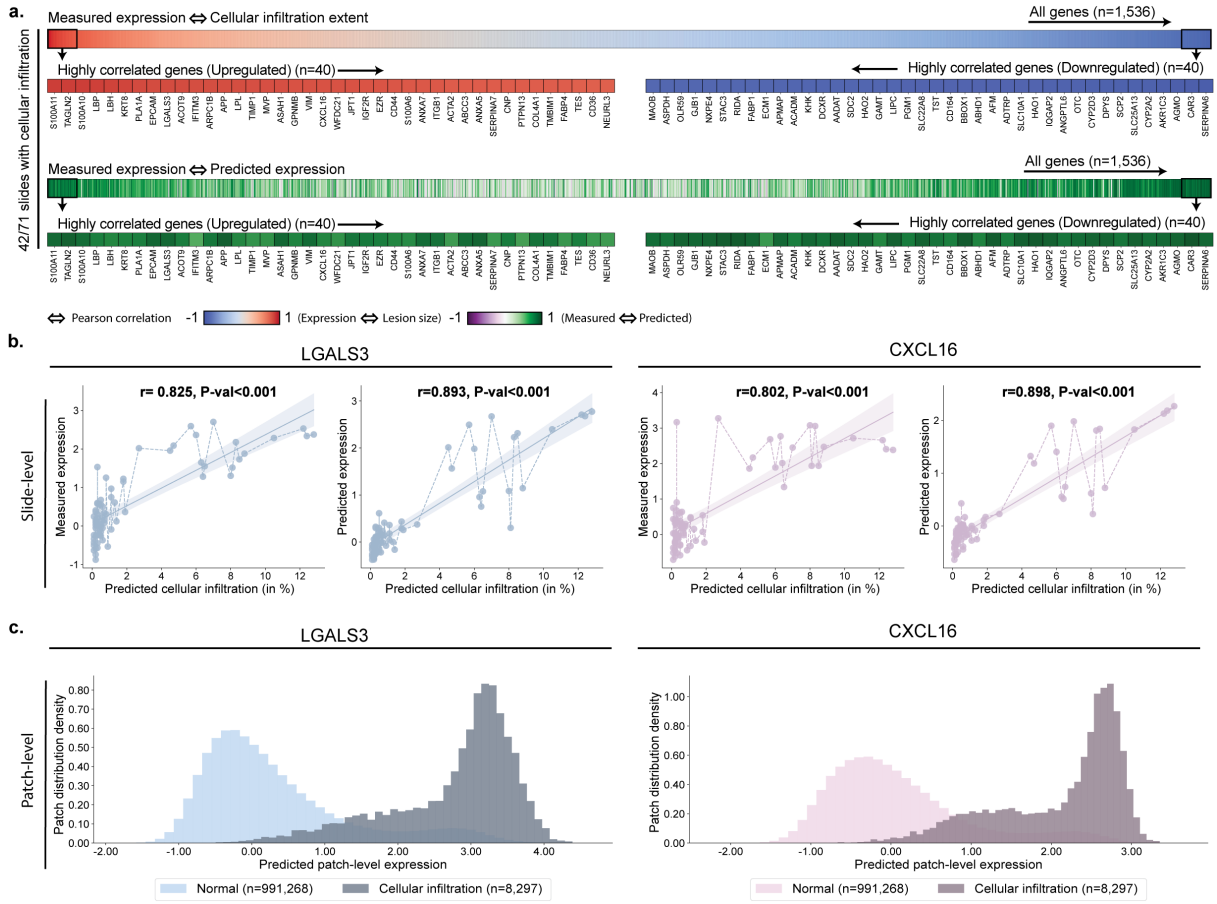

Figure S8: **Morphomolecular analysis of cellular infiltration in methylene dianiline. a.** For each slide-expression pair, we correlate the percentage of cellular infiltration predicted by the lesion classifier with the measured gene expression. Genes with a high correlation between measured expression and lesion size define morphomolecular signatures associated with the compound. **b.** Correlation between the estimated percentage of the slide containing cellular infiltration and the gene expression of *LGALS3* and *CXCL16* (measured with microarray and predicted with GEESE). P-value derived from testing the two-sided null hypothesis of non-correlation. **c.** Distribution of patch-level expression for patches predicted as normal (n=991,268) and patches predicted as containing cellular infiltration (n=8,297).

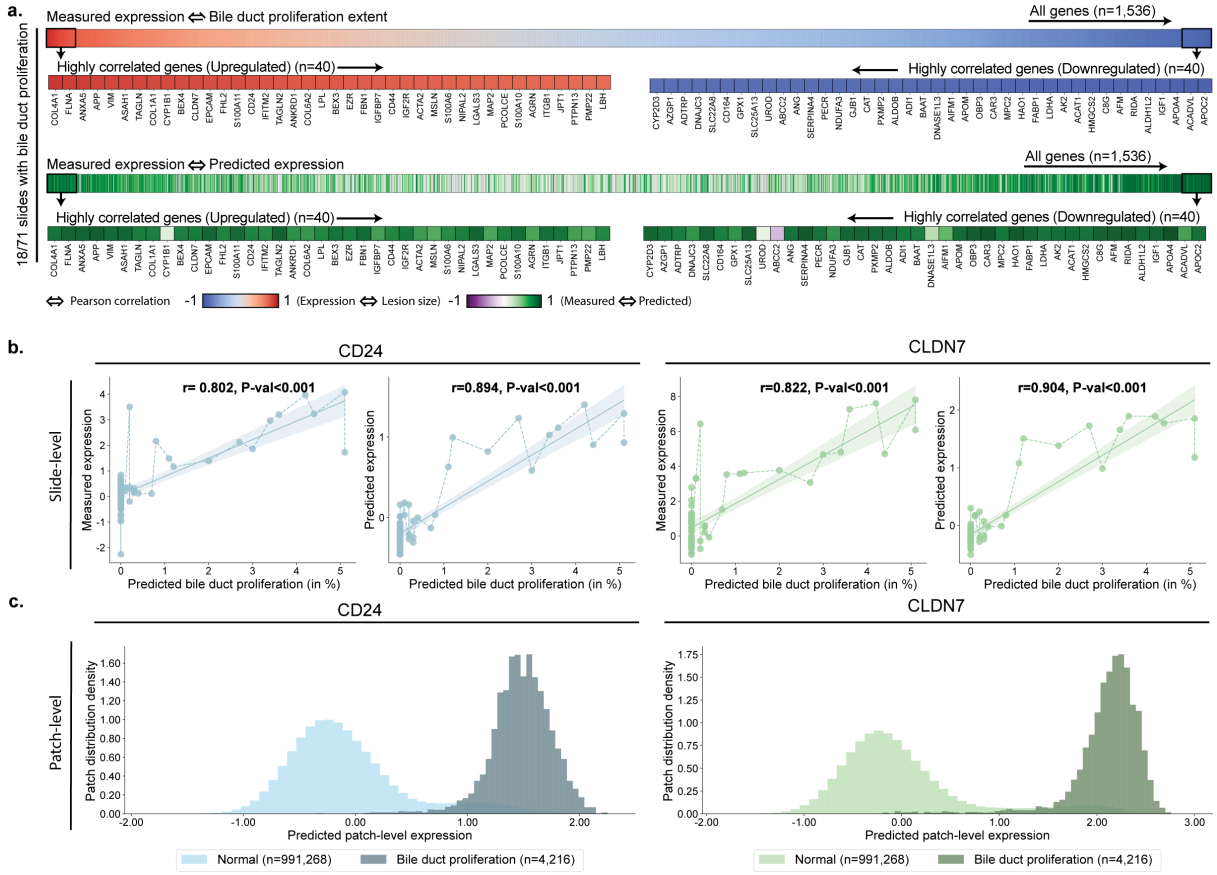

Figure S9: **Morphomolecular analysis of bile duct proliferation detection in methylene dianiline. a.** For each slide-expression pair, we correlate the percentage of bile duct proliferation predicted by the lesion classifier with the measured gene expression. Genes with a high correlation between measured expression and lesion size define morphomolecular signatures associated with the compound. **b.** Correlation between the estimated percentage of the slide containing bile duct proliferation and the gene expression of *CD24* and *CLDN7* (measured with microarray and predicted with GEESE). P-value derived from testing the two-sided null hypothesis of non-correlation. **c.** Distribution of patch-level expression for patches predicted as normal (n=991,268) and patches predicted as containing bile duct proliferation (n=4,216).

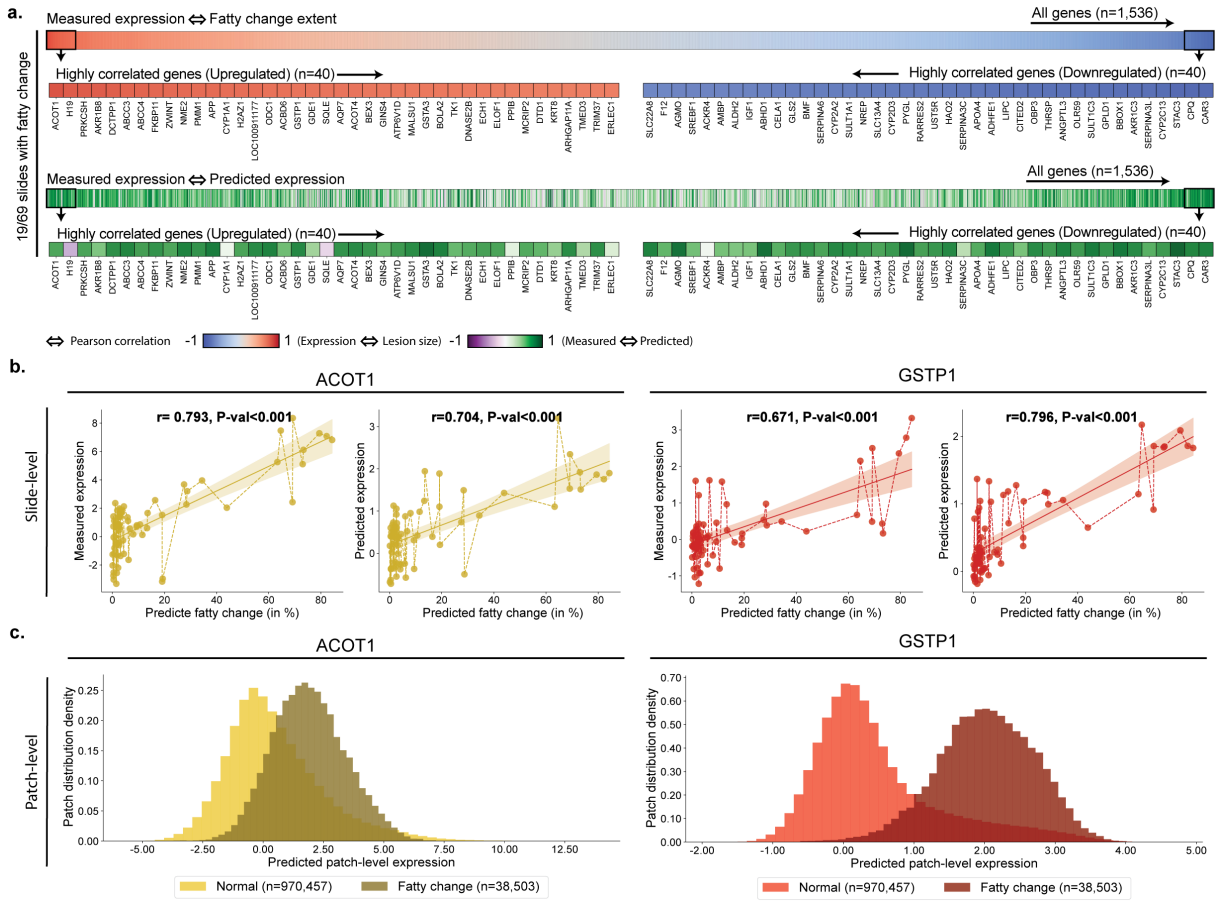

Figure S10: **Morphomolecular analysis of fatty change detection in ethionamide. a.** For each slide-expression pair, we correlate the percentage of fatty change predicted by the lesion classifier with the measured gene expression. Genes with a high correlation between measured expression and lesion size define morphomolecular signatures associated with the compound. **b.** Correlation between the estimated percentage of the slide containing fatty change and the gene expression of *ACOT1* and *GSTP1* (measured with microarray and predicted with GEESE). P-value derived from testing the two-sided null hypothesis of non-correlation. **c.** Distribution of patch-level expression for patches predicted as normal (n=970,457) and patches predicted as containing fatty change (n=38,503).

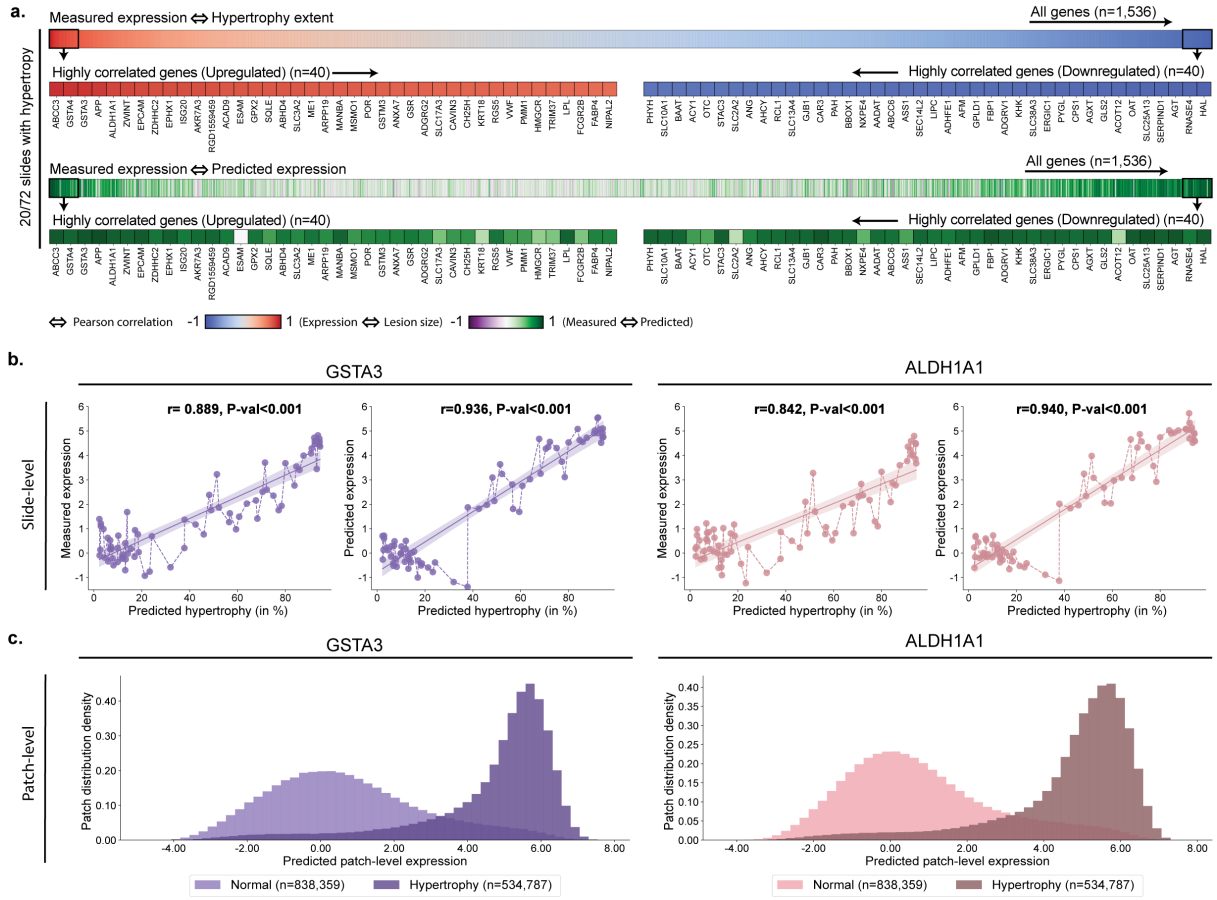

Figure S11: **Morphomolecular analysis of hypertrophy detection in hexachlorobenzene. a.** For each slide-expression pair, we correlate the percentage of hypertrophy predicted by the lesion classifier with the measured gene expression. Genes with a high correlation between measured expression and lesion size define morphomolecular signatures associated with the compound. **b.** Correlation between the estimated percentage of the slide containing hypertrophy and the gene expression of *GSTA3* and *ALDH1A1* (measured with microarray and predicted with GEESE). P-value derived from testing the two-sided null hypothesis of non-correlation. **c.** Distribution of patch-level expression for patches predicted as normal (n=838,359) and patches predicted as containing hypertrophy (n=534,787).

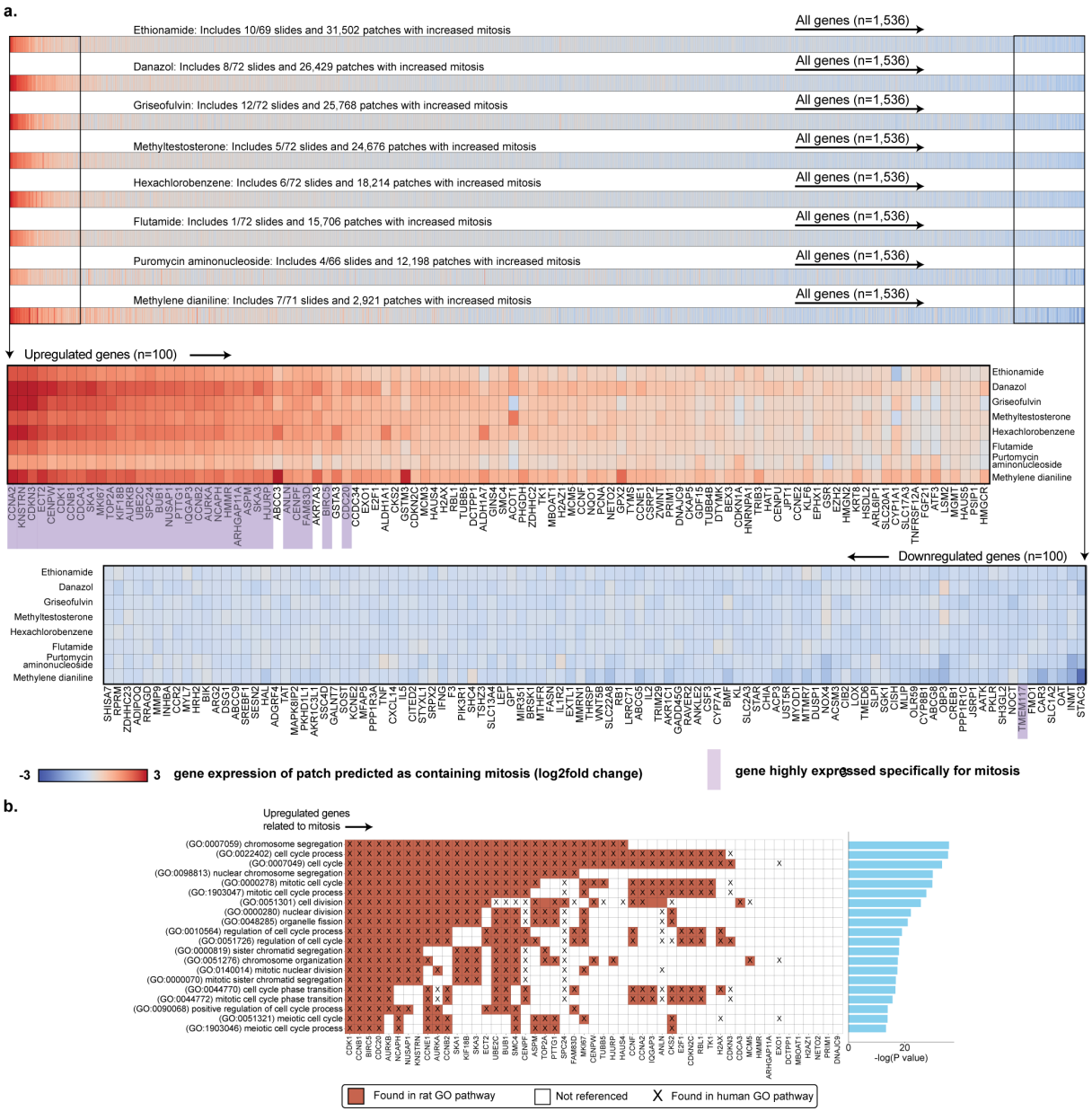

Figure S12: **Caption next page.**

Figure S12: **Cross-study analysis of the molecular signature of mitosis. a.** Average predicted gene expression of patches with mitosis in eight different studies from TG-GATEs test set. Zoom of the top 100 upregulated (top, red) and downregulated genes (bottom, blue). Genes are sorted in decreasing order by their absolute mean gene expression across all considered studies. **b.** Top 20 biological pathways (ranked by highest p-values) of the 51 upregulated genes identified as specific to increased mitosis (**table S9**). The analysis was conducted using the rat genome database (RGD). Identified pathways, including cell cycle, mitotic nuclear division, and DNA replication, are highly relevant to mitosis. Most of the genes found, such as *CCNA2*, *CCNB1*, and *CCNB2* from the cyclin family, are known to play crucial roles in cell division and reproduction pathways are closely related to mitosis. Other genes, such as *HMMR* (involved in microtubule stabilization) and *ARHGAP11A* (associated with Rho GTPase signaling), demonstrate a less direct linkage but are still involved in related processes such as cytoskeletal organization. A lower p-value, obtained using hypergeometric distribution test, signifies a higher degree of over-representation.

Figure S13: **Cross-study analysis of the molecular signature of necrosis. a.** Average predicted gene expression of patches with necrosis in four different studies from TG-GATEs test set. Zoom of the top 100 upregulated (top, red) and downregulated genes (bottom, blue). Genes are sorted in decreasing order by their absolute mean gene expression across all considered studies. **b.** Top 20 biological pathways (ranked by highest p-values) of the 33 upregulated genes identified as specific to necrosis (left), and the top 20 biological pathways (ranked by highest p-values) of the 33 downregulated genes identified as specific to necrosis (right) (**table S10**). The analysis was conducted using the rat genome database (RGD). Most identified biological pathways, including the apoptotic process, cell death, and cellular response to stress, are highly relevant to cellular necrosis. We observe known genes such as *ATF3* (known for modulating the stress response), *TNFRSF12A* (involved in inflammation and cell death), *HMOX1* (associated with the oxidative stress response), and *MYC* (playing roles in the cell cycle, apoptosis, and DNA damage response). Other genes like *DDIT3*, *TRIB3*, *FOSL1*, and *CXCL2* are also upregulated and associated with processes like response to oxidative stress, cellular response to unfolded protein, and cellular response to chemical stress. Conversely, genes such as *ACSS2*, *ACACA*, *CYP7A1*, and *AVPR1A* are downregulated and linked to small molecule biosynthetic processes, lipid biosynthetic processes, and fatty acid metabolic processes. Other genes like *MAFF* (a gene implicated in transcriptional regulation) or *NREP* (a gene known for its involvement in neural development and tissue repair) demonstrate a less known direct linkage to necrosis despite showing high correlation with necrosis in our analysis. A lower p-value, obtained using hypergeometric distribution test, signifies a higher degree of over-representation.

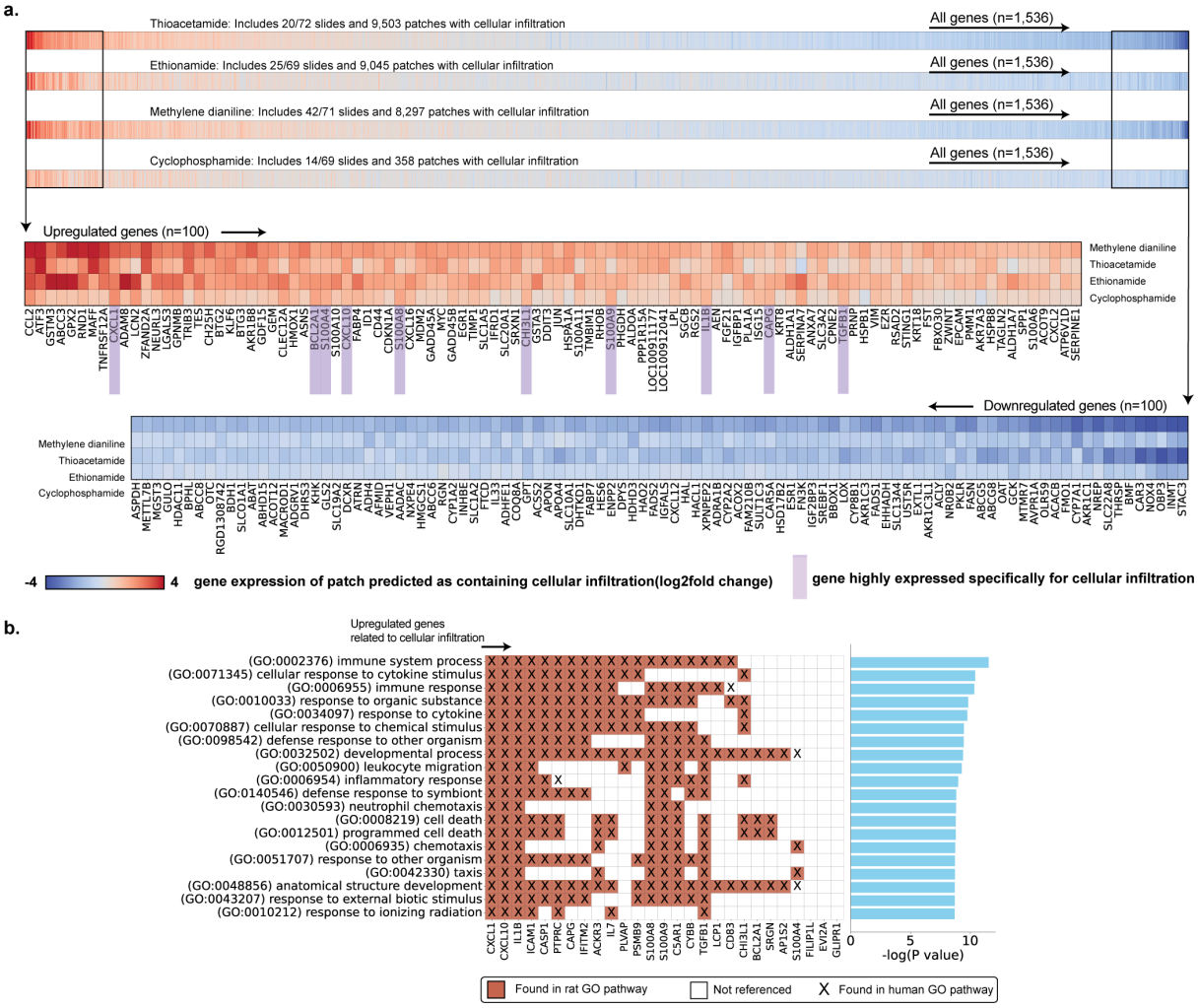

Figure S14: **Cross-study analysis of the molecular signature of cellular infiltration. a.** Average predicted gene expression of patches with cellular infiltration in four different studies from TG-GATEs test set. Zoom of the top 100 upregulated (top, red) and downregulated genes (bottom, blue). Genes are sorted in decreasing order by their absolute mean gene expression across all considered studies. **b.** Top 20 biological pathways (ranked by highest p-values) of the 27 upregulated genes identified as specific to cellular infiltration (**table S11**). Analysis was conducted using the rat genome database (RGD). Genes found such as *CXCL1*, *CXCL10*, noted for their roles in directing leukocytes to the site of infiltration, *BCL2A1*, a gene known for its role in immunity response and the anti-apoptotic role and *S100A4*, involved in modulating the inflammatory response and attracting leukocytes have a biological pathway with links to cellular infiltration. Other genes such as *FILIP1L* and *EVI2A* (gene not currently linked to any specific biological processes in humans and rats according to Gene Ontology Biological Process annotations) demonstrate a less direct linkage to cellular infiltration. A lower p-value, obtained using hypergeometric distribution test, signifies a higher degree of over-representation.

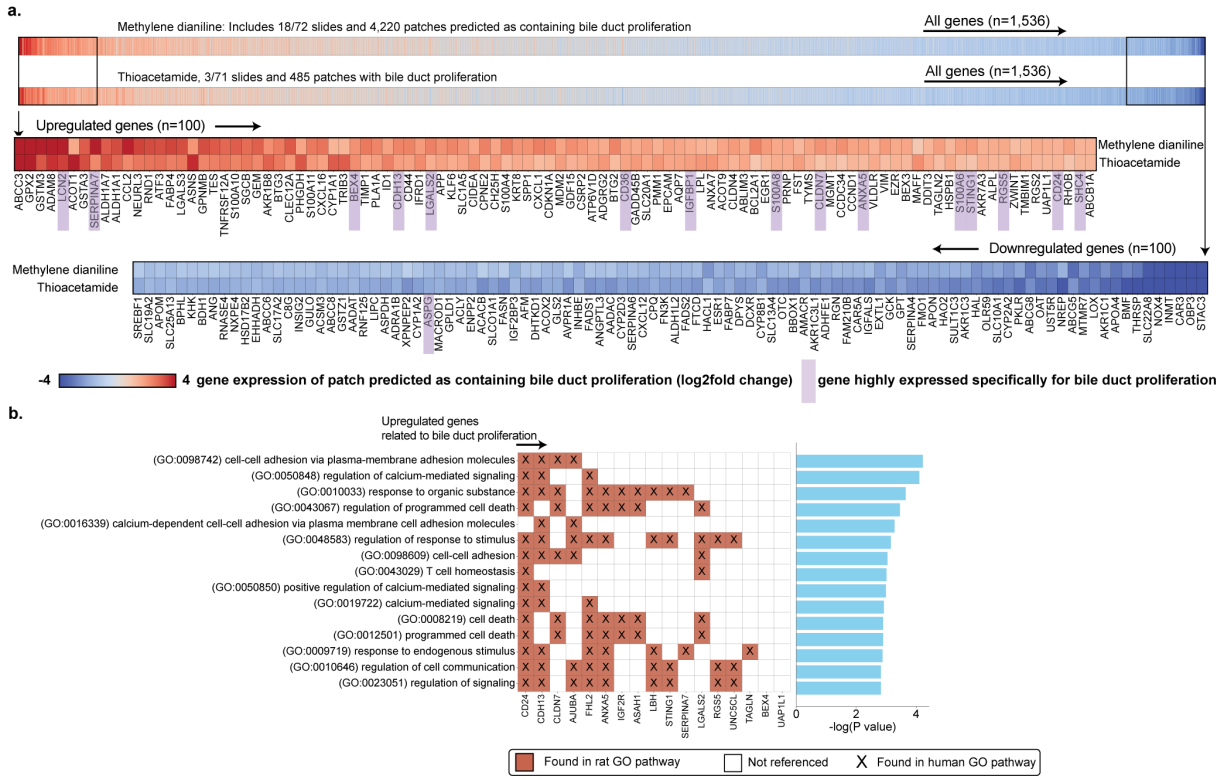

Figure S15: **Cross-study analysis of the molecular signature of bile duct proliferation. a.** Average predicted gene expression of patches with cbile duct proliferation in two studies from TG-GATEs test set. Zoom of the top 100 upregulated (top, red) and downregulated genes (bottom, blue). Genes are sorted in decreasing order by their absolute mean gene expression across all considered studies. **b.** Top 15 biological pathways (highest p-values) of the 17 upregulated genes identified as specific to bile duct proliferation (**table S12**). Analysis was conducted using the rat genome database (RGD). Several genes identified are involved in biological pathways relevant to bile duct proliferation, such as epithelial cell proliferation, extracellular matrix organization, and response to wounding. For instance, genes from the SERPINA family (*SERPINA7* and *SERPINA4*), known for their roles in protease inhibition, may impact inflammation and tissue remodeling processes associated with bile duct proliferation. Other genes like *LGALS2*, involved in cell-cell adhesion, protein-carbohydrate interactions, and immune response modulation, or *CDH13*, known for its role in cell adhesion, also show relevance to the molecular mechanisms underlying bile duct expression. A lower p-value, obtained using hypergeometric distribution test, signifies a higher degree of over-representation.

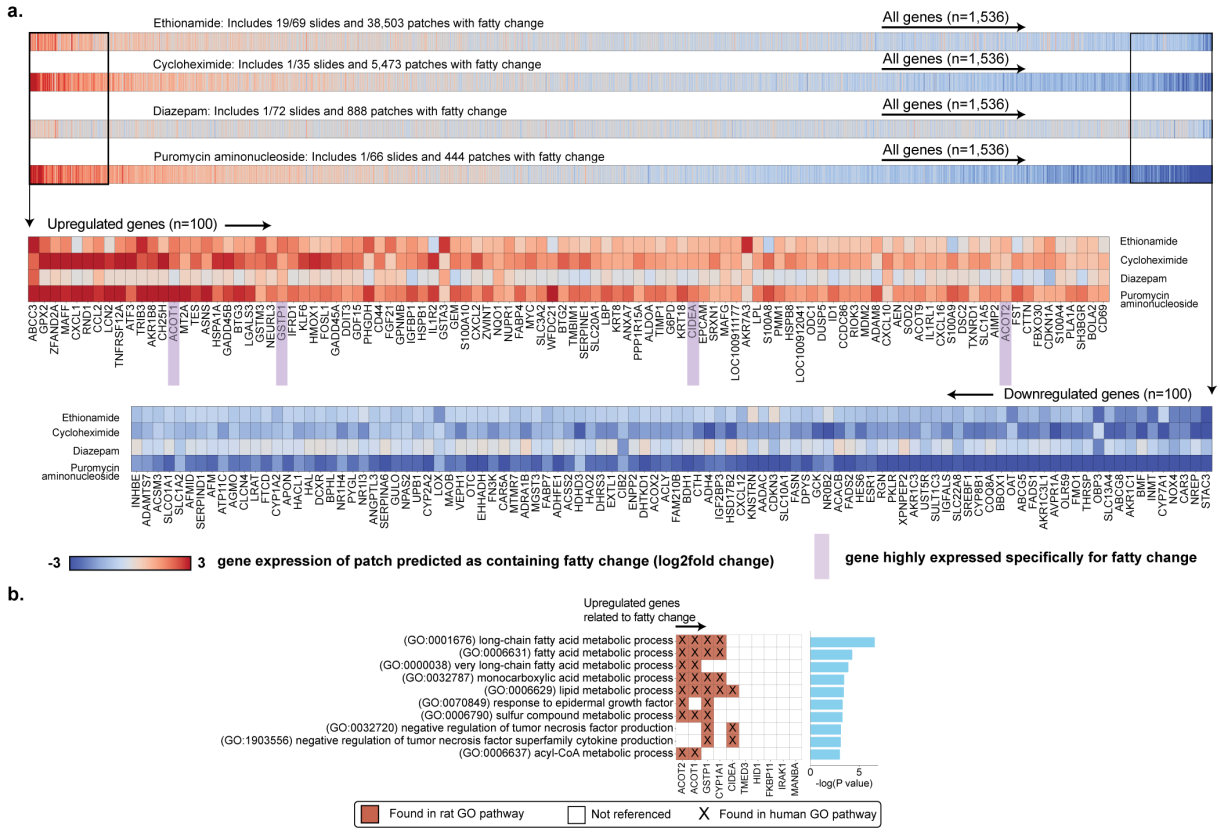

Figure S16: **Cross-study analysis of the molecular signature of fatty change. a.** Average predicted gene expression of patches with fatty change in four different studies from TG-GATEs test set. Zoom of the top 100 upregulated (top, red) and downregulated genes (bottom, blue). Genes are sorted in decreasing order by their absolute mean gene expression across all considered studies. **b.** Top 10 biological pathways (ranked by highest p-values) linked to the 10 upregulated genes identified as specific to fatty change (**table S13**). Analysis was conducted using the rat genome database (RGD). Some identified pathways, including fatty acid metabolic process, lipid metabolic process, and response to oxidative stress, support the relevance of the identified genes to fatty change. Genes such as *ACOT1* and *ACOT2* from the ACOT family, which are involved in the hydrolysis of acyl-CoA thioester compounds, a key process in lipid metabolism are highly relevant to fatty change. Additionally, *GSTP1*, a gene implicated in detoxification and the response to oxidative stress, is also upregulated in fatty change. A lower p-value, obtained using hypergeometric distribution test, signifies a higher degree of over-representation.

Figure S17: Caption next page

Figure S17: **Cross-study analysis of the molecular signature of hypertrophy. a.** Average predicted gene expression of patches with hypertrophy in eleven different studies from TG-GATEs test set. Zoom of the top 100 upregulated (top, red) and downregulated genes (bottom, blue). Genes are sorted in decreasing order by their absolute mean gene expression across all considered studies. **b.** Top 20 biological pathways (ranked by highest p-values) linked to the 18 genes upregulated genes identified as specific to hypertrophy (**table S14**). Analysis was conducted using the rat genome database (RGD). Genes involved in managing increased metabolic demands, such as *ACSM2*, *VNN1*, and *ME1*, are relevant in different aspects of metabolism—fatty acid, lipid, and glucose metabolism, respectively. Genes like *GSTA3*, *ALDH1A1*, and *ALDH1A7* are involved in detoxifying oxidative and aldehyde by-products, which are more prevalent as metabolic activities intensify in hypertrophic cells. The enrichment of pathways such as fatty acid metabolic process, oxidation-reduction process, and extracellular matrix organization further supports the relevance of these genes to the cellular adaptations occurring during hypertrophy. Additionally, *LOX*'s role in collagen synthesis and extracellular matrix assembly is necessary for the structural adaptation of tissues undergoing hypertrophy. A lower p-value, obtained using hypergeometric distribution test, signifies a higher degree of over-representation.

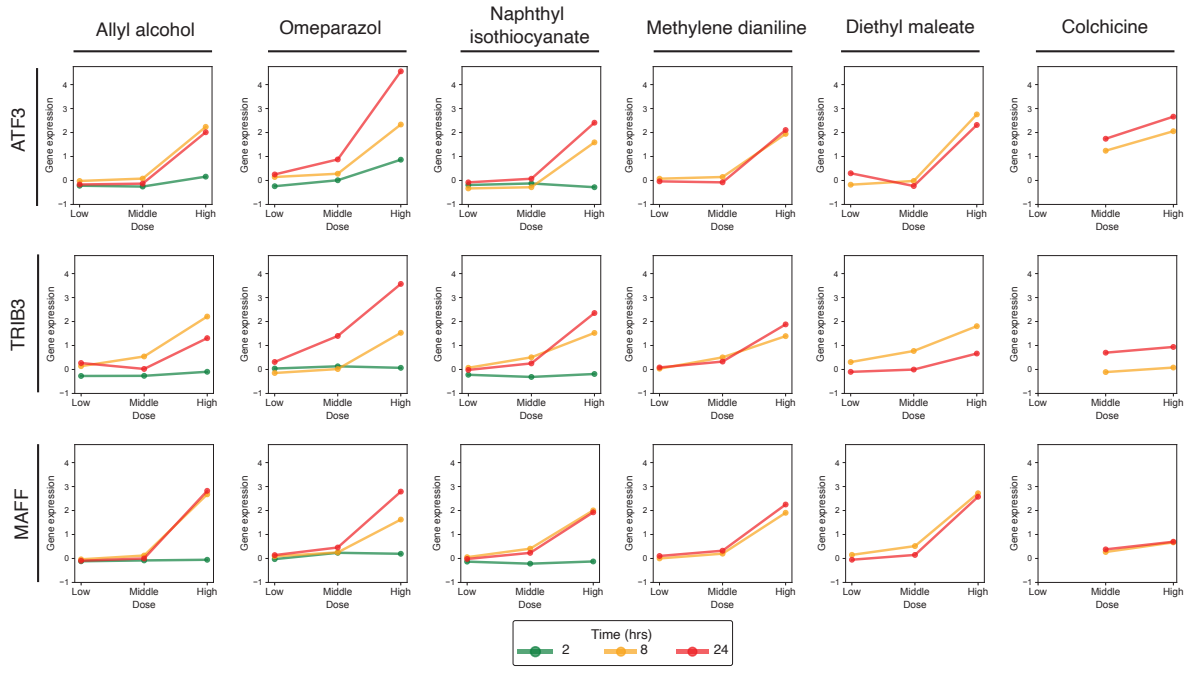

Figure S18: **Translation of necrosis genetic biomarkers to human in vitro primary human hepatocytes cell lines.** Dose-response relationship of six necrosis-inducing selected compounds of in vitro human cell lines at three different time points. Genes tested (*ATF3*, *TRIB3*, and *MAFF*) were identified as connected to necrosis in the in vivo rat experiments.

Name	Number of slides	Number of microarrays	Split
Labetalol	155	71	Train/Val
Glibenclamide	159	71	Train/Val
Methyltestosterone	160	72	Test
Griseofulvin	160	72	Test
Tetracycline	160	72	Train/Val
Perhexiline	160	72	Train/Val
Flutamide	160	72	Test
Lomustine	160	72	Train/Val
Azathioprine	157	71	Train/Val
Methimazole	160	72	Train/Val
Monocrotaline	154	69	Train/Val
Pemoline	159	72	Train/Val
Chlormezanone	159	71	Train/Val
Metformin	160	72	Test
Ethinylestradiol	159	71	Train/Val
Tamoxifen	160	72	Train/Val
Methyldopa	160	72	Test
Vitamin a	158	72	Train/Val
Tacrine	148	65	Train/Val
Ciprofloxacin	160	72	Train/Val
Chloramphenicol	160	72	Train/Val
Nitrofurazone	157	72	Train/Val
Imipramine	159	72	Train/Val
Moxisylyte	160	72	Train/Val
Iproniazid	160	72	Train/Val
Amitriptyline	158	72	Train/Val
Hydroxyzine	160	72	Test
Ibuprofen	157	71	Train/Val
Mefenamic acid	158	72	Test
Furosemide	158	72	Train/Val
Fenofibrate	160	72	Train/Val
Chlorpropamide	160	72	Train/Val
Famotidine	160	72	Train/Val
Nifedipine	160	72	Train/Val
Chlorpheniramine	160	72	Train/Val
Diltiazem	159	72	Train/Val
Quinidine	159	72	Train/Val
Naproxen	153	69	Train/Val
Nicotinic acid	159	72	Train/Val
Ranitidine	160	72	Train/Val

Table S1: **Overview of TG-GATEs compounds.** For each compound, we report the number of slides, microarrays (also corresponding to the number of pairs), and whether the study was used for training/validation or testing.

Name	Number of slides	Number of microarrays	Split
Erythromycin ethylsuccinate	159	72	Train/Val
Tannic acid	158	72	Train/Val
Caffeine	160	72	Train/Val
Captopril	159	72	Train/Val
Clofibrate	160	72	Train/Val
Naphthyl isothiocyanate	160	72	Train/Val
Enalapril	160	72	Train/Val
Phenobarbital	158	72	Train/Val
Allyl alcohol	160	72	Train/Val
Rifampicin	160	72	Train/Val
Carbon tetrachloride	225	72	Train/Val
Phenylbutazone	160	72	Train/Val
Isoniazid	304	72	Test
Acetaminophen	269	74	Train/Val
Indomethacin	153	69	Train/Val
Omeprazole	160	72	Train/Val
Thioacetamide	159	72	Test
Ethionine	160	72	Train/Val
Carbamazepine	160	72	Train/Val
Chlorpromazine	139	60	Train/Val
Coumarin	160	72	Train/Val
Allopurinol	157	72	Train/Val
Diclofenac	160	72	Train/Val
Wy-14643	160	72	Train/Val
Gemfibrozil	160	72	Train/Val
Nitrofurantoin	160	72	Train/Val
Bromobenzene	160	72	Test
Amiodarone	159	72	Train/Val
Sulfasalazine	160	72	Train/Val
Adapin	140	60	Train/Val
Cimetidine	160	72	Train/Val
Cyclophosphamide	160	72	Test
Phenytoin	160	72	Train/Val
Valproic acid	160	72	Train/Val
Ketoconazole	144	66	Train/Val
Ethambutol	158	72	Train/Val
Papaverine	159	72	Train/Val
Penicillamine	160	72	Train/Val
Disopyramide	159	72	Train/Val
Sulindac	156	72	Train/Val

Table S1: **Continuation.**

Name	Number of slides	Number of microarrays	Split
Triamterene	160	72	Train/Val
Mexiletine	159	72	Test
Tolbutamide	159	72	Train/Val
Sulpiride	159	72	Train/Val
Colchicine	159	72	Train/Val
Acarbose	160	72	Train/Val
Triazolam	160	72	Train/Val
Simvastatin	158	72	Train/Val
Clomipramine	160	72	Train/Val
Ajmaline	160	72	Train/Val
Trimethadione	159	72	Train/Val
Dantrolene	160	72	Train/Val
Terbinafine	160	72	Train/Val
Bendazac	159	72	Train/Val
Meloxicam	151	68	Train/Val
Benziodarone	160	72	Train/Val
Lornoxicam	151	69	Train/Val
Etoposide	159	72	Train/Val
Tiopronin	160	72	Train/Val
Ethionamide	153	69	Test
Methapyrilene	79	36	Train/Val
Nimesulide	158	72	Train/Val
Disulfiram	159	72	Train/Val
Ethanol	160	72	Train/Val
Aspirin	160	72	Train/Val
Promethazine	160	72	Train/Val
Phenacetin	160	72	Test
Bucetin	160	72	Train/Val
Chlormadinone	160	72	Train/Val
Danazol	160	72	Test
Phenylanthranilic acid	151	71	Train/Val
Acetamidofluorene	159	75	Train/Val
Cisplatin	156	70	Test
Benzbromarone	157	72	Train/Val
Carboplatin	160	72	Test
Nitrosodiethylamine	151	67	Train/Val
Bromoethylamine	160	72	Test
Ticlopidine	159	71	Train/Val
Cyclosporine a	160	72	Test
Cephalothin	160	72	Test

Table S1: **Continuation.**

Name	Number of slides	Number of microarrays	Split
Puromycin aminonucleoside	149	66	Test
Gentamicin	155	72	Test
Doxorubicin	157	71	Train/Val
Theophylline	155	71	Test
Acetazolamide	160	72	Test
Cycloheximide	77	35	Test
Tunicamycin	80	36	Test
Phalloidin	60	18	Train/Val
Galactosamine	79	35	Train/Val
Phorone	80	36	Train/Val
Tnfa	80	36	Train/Val
Buthionine sulfoximine	80	36	Train/Val
Diethyl maleate	80	36	Train/Val
Diazepam	160	72	Test
Hexachlorobenzene	160	72	Test
Rosiglitazone maleate	160	72	Train/Val
Rotenone	159	72	Train/Val
Dexamethasone	80	36	Train/Val
Fluoxetine hydrochloride	160	72	Train/Val
Methylene dianiline	159	71	Test
Desmopressin acetate	160	72	Train/Val
Butylated hydroxyanisole	160	72	Train/Val
Amphotericin b	132	57	Train/Val
2-nitrofluorene	20	9	Train/Val
N-nitrosomorpholine	20	9	Train/Val
N-methyl-n-nitrosourea	20	9	Train/Val
Aflatoxin b1	20	11	Train/Val
Acetamide	80	36	Train/Val
Propranolol	160	72	Train/Val
Bortezomib	72	31	Train/Val
3-methylcholanthrene	80	36	Train/Val
Gefitinib	80	36	Train/Val
Propylthiouracil	160	72	Train/Val
Haloperidol	39	11	Train/Val
Fluphenazine	159	72	Train/Val
Thioridazine	31	19	Train/Val

Table S1: **Continuation.**

Gene	Pearson correlation	AUC	R2	MSE
STAC3	0.762	0.695	0.546	0.454
OAT	0.757	0.688	0.572	0.428
DHTKD1	0.722	0.643	0.279	0.721
TNFRSF12A	0.722	0.608	0.510	0.490
KLF6	0.717	0.632	0.486	0.514
RNASE4	0.715	0.636	0.464	0.536
LOC100911177	0.714	0.645	0.498	0.502
CXCL12	0.711	0.668	0.450	0.550
S100A10	0.706	0.629	0.487	0.513
CAR3	0.698	0.676	0.456	0.544
LOC100912041	0.696	0.616	0.470	0.530
KRT18	0.695	0.669	0.460	0.540
FAM210B	0.688	0.635	0.416	0.584
SLC10A1	0.688	0.595	0.232	0.768
CPQ	0.685	0.614	0.429	0.571
ATRN	0.683	0.630	0.422	0.578
ABCC6	0.683	0.650	0.450	0.550
ANG	0.680	0.637	0.363	0.637
ALDH1A1	0.680	0.706	0.392	0.608
BMF	0.673	0.673	0.383	0.617
ODC1	0.664	0.675	0.394	0.606
GSTA3	0.659	0.693	0.376	0.624
SERPINA4	0.659	0.617	0.367	0.633
IFRD1	0.658	0.631	0.391	0.609
APOA4	0.657	0.634	0.372	0.628
TIMP1	0.657	0.587	0.385	0.615
NXPE4	0.655	0.606	0.349	0.651
ACOT9	0.655	0.618	0.375	0.625
DPYS	0.652	0.624	0.268	0.732
RIOK3	0.651	0.604	0.406	0.594
APP	0.651	0.621	0.371	0.629
AKR7A3	0.651	0.681	0.364	0.636
CLCN4	0.650	0.622	0.398	0.602
CYP2D3	0.648	0.610	0.367	0.633
SDC2	0.647	0.621	0.316	0.684
SLC19A2	0.645	0.654	0.385	0.615
BTG2	0.645	0.634	0.410	0.590
CDO1	0.644	0.628	0.293	0.707
NEURL3	0.642	0.610	0.390	0.610
BBOX1	0.642	0.643	0.111	0.889
NOX4	0.639	0.690	0.384	0.616
GPLD1	0.638	0.638	0.377	0.623
CCDC86	0.637	0.637	0.380	0.620
AIMP2	0.637	0.644	0.374	0.626
SLC22A8	0.636	0.670	0.375	0.625
RGN	0.628	0.604	0.121	0.879
AGT	0.628	0.651	0.389	0.611
MAOB	0.626	0.639	0.116	0.884
PPP1R15A	0.625	0.596	0.380	0.620
C8A	0.624	0.625	0.266	0.734

Table S2: **Gene expression prediction performance of top-100 best predicted genes.** We report Pearson correlation, AUC, R2, and MSE. Each metric is described in the **Online methods**, section **Metrics**.

Gene	Pearson correlation	AUC	R2	MSE
ABCC3	0.623	0.656	0.336	0.664
FADS1	0.623	0.632	0.292	0.708
ADHFE1	0.620	0.611	0.368	0.632
SLC25A13	0.620	0.610	0.298	0.702
NREP	0.620	0.657	0.097	0.903
HSPB8	0.618	0.604	0.360	0.640
EPCAM	0.616	0.621	0.373	0.627
LIPC	0.613	0.620	0.244	0.756
NOC2L	0.613	0.630	0.356	0.644
SULT1C3	0.608	0.622	0.271	0.729
GULO	0.608	0.654	0.266	0.734
AVPR1A	0.607	0.637	0.204	0.796
STK17B	0.606	0.636	0.340	0.660
C8B	0.606	0.604	0.344	0.656
GADD45A	0.606	0.634	0.361	0.639
F10	0.605	0.612	0.287	0.713
F12	0.605	0.602	0.347	0.653
OTC	0.602	0.601	0.088	0.912
JUN	0.602	0.631	0.345	0.655
C2CD2	0.602	0.624	0.359	0.641
ICAM1	0.601	0.602	0.328	0.672
HES6	0.601	0.656	0.275	0.725
COQ8A	0.601	0.668	0.084	0.916
PC	0.600	0.623	0.339	0.661
BRIX1	0.600	0.608	0.350	0.650
GSTZ1	0.600	0.610	0.225	0.775
MDM2	0.599	0.621	0.275	0.725
ADRA1B	0.599	0.627	0.314	0.686
BAAT	0.598	0.629	0.012	0.988
GOT1	0.597	0.672	0.337	0.663
AADAT	0.597	0.625	0.325	0.675
OBP3	0.597	0.665	0.331	0.669
AADAC	0.596	0.614	0.147	0.853
CYP2A2	0.595	0.603	0.191	0.809
BDH1	0.593	0.623	0.290	0.710
C8G	0.591	0.577	0.292	0.708
IGFALS	0.590	0.654	0.142	0.858
RHOB	0.590	0.627	0.335	0.665
AGMO	0.588	0.620	0.238	0.762
KRT8	0.587	0.629	0.289	0.711
IQGAP2	0.586	0.619	0.281	0.719
SLCO1A1	0.586	0.587	0.249	0.751
UST5R	0.586	0.671	0.291	0.709
EIF2S1	0.586	0.634	0.326	0.674
ATF3	0.583	0.592	0.327	0.673
CDIP1	0.583	0.616	0.228	0.772
SLC17A2	0.583	0.633	0.258	0.742
SCP2	0.582	0.612	0.260	0.740
HDHD3	0.581	0.620	0.198	0.802
ABCA8A	0.580	0.659	0.169	0.831

Table S2: **Continuation.**

Lesion	Patch Positive	Patch Negative	WSI Positive	WSI Negative
Cellular infiltration	2,232	3,315	984	1,167
Necrosis	6,568	26,400	1023	538
Hypertrophy	4,525	5,451	29	698
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

Table S3: **Number of patch-level annotations per lesion.** Patch positive refers to the number of patches containing the lesion, and Patch negative refers to the number of patches that do not contain it (can be a normal patch or a patch containing another lesion). A description of each lesion is provided in **table S4**.

Lesion	Definition
Cellular infiltration	Infiltrations of inflammatory cells in the liver. Includes neutrophil, mononuclear and peribiliary.
Necrosis	Cell death of hepatocytes. Includes single-cell necrosis (apoptosis), focal/multifocal, and zonal (centrilobular, midzonal, periportal and diffuse).
Hypertrophy	Enlargement of the hepatocyte cytoplasm, secondary to increase in the cytosolic protein or number of organelles.
Fatty change	Hepatocellular vacuolation, consistent with intracytoplasmic lipid accumulation. Includes macro and microvesicular.
Bile duct proliferation	Increased number of small bile ducts arising in portal region. Biliary epithelium appears normal or may show degenerative or atrophic changes.
Increased mitosis	Increased hepatocyte mitoses above normal background levels.

Table S4: **Morphological lesion definition.** Morphological characterization of the lesions considered in this work are based on INHAND guidelines[78], 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.

Lesion	Compound
Necrosis	thioacetamide, ethionamide, methylene-dianiline, bromobenzene
Cellular Infiltration	ethionamide, thioacetamide, cyclophosphamide, methylene-dianiline
Bile duct proliferation	thioacetamide, methylene-dianiline
Fatty change	ethionamide, puromycin-aminonucleoside, cycloheximide, diazepam
Hypertrophy	hexachlorobenzene, diazepam, thioacetamide, hydroxyzine, ethionamide, methylene-dianiline, flutamide, phenacetin, bromobenzene, griseofulvin, methyltestosterone
Increased mitosis	methyltestosterone, griseofulvin, flutamide, ethionamide, puromycin aminonucleoside, danazol, hexachlorobenzene, methylene-dianiline

Table S5: **List of compounds used for inferring morphomolecular signatures for each lesion.**

	Necrosis	Cellular infiltration	Bile duct proliferation	Fatty change	Hypertrophy	Increased mitosis	Lesion (any)	No lesion
Ethionamide	45	25	0	19	2	10	59	59
Methylene dianiline	40	42	18	0	19	7	45	64
Thioacetamide	25	20	3	0	29	0	42	72
Cycloheximide	19	0	0	1	0	0	20	35
Hexachlorobenzene	1	1	0	13	20	6	28	66
Flutamide	2	7	0	0	16	1	25	71
Griseofulvin	2	0	0	0	3	12	17	60
Bromobenzene	3	5	0	0	17	0	18	72
Hydroxyzine	0	0	0	6	17	0	17	72
Cyclophosphamide	0	14	0	0	0	0	14	72
Phenacetin	0	0	0	0	11	0	11	72
Diazepam	0	0	0	1	10	0	11	72
Isoniazid	1	2	0	0	8	0	11	72
Methyltestosterone	0	0	0	0	6	5	10	67
Danazol	0	0	0	0	0	8	8	64
Tunicamycin	4	0	0	0	0	0	4	36
Puromycin aminonucleoside	1	0	0	1	1	4	6	62
Mexiletine	1	0	0	0	5	0	6	72
Bromoethylamine	0	0	0	6	2	0	6	72
Methyldopa	3	0	0	0	0	0	3	72
Theophylline	2	0	0	0	0	0	2	71
Mefenamic acid	1	0	0	0	0	0	1	72
Cisplatin	0	0	0	0	0	0	0	70
Carboplatin	0	0	0	0	0	0	0	72
Cyclosporine A	0	0	0	0	0	0	0	72
Cephalothin	0	0	0	0	0	0	0	72
Gentamicin	0	0	0	0	0	0	0	72
Acetazolamide	0	0	0	0	0	0	0	72
Metformin	0	0	0	0	0	0	0	72

Table S6: **Distribution of lesions in TG-GATEs test set stratified by compound.** Compounds are sorted by their percentage of slides with lesions.

	Necrosis	Cellular infiltration	Bile duct proliferation	Hypertrophy	Fatty change	Increased mitosis	Lesion (any)	No lesion
Hexachlorobenzene	2.53e+03	6.10e+03	3.28e+02	5.35e+05	5.01e+03	1.82e+04	5.60e+05	8.38e+05
Thioacetamide	2.24e+04	9.50e+03	4.85e+02	1.78e+05	1.83e+02	1.17e+04	2.14e+05	8.52e+05
Diazepam	6.35e+02	1.47e+03	7.90e+01	2.20e+05	8.88e+02	7.29e+03	2.29e+05	9.67e+05
Ethionamide	2.55e+04	9.04e+03	3.03e+02	9.02e+04	3.85e+04	3.15e+04	1.79e+05	9.70e+05
Gentamicin	5.38e+02	1.39e+03	2.63e+02	1.19e+04	1.50e+01	8.64e+04	1.00e+05	6.64e+05
Cephalothin	4.91e+02	1.97e+03	4.20e+02	1.17e+04	2.10e+01	8.03e+04	9.47e+04	6.81e+05
Carboplatin	2.13e+03	2.21e+03	1.24e+02	7.78e+04	1.90e+01	1.11e+04	9.31e+04	6.29e+05
Acetazolamide	1.89e+03	1.38e+03	2.30e+02	8.27e+04	1.18e+02	6.34e+03	9.24e+04	6.42e+05
Puromycin aminonucleoside	4.57e+03	3.69e+03	4.71e+02	6.22e+04	4.44e+02	1.22e+04	8.28e+04	6.28e+05
Hydroxyzine	4.40e+01	1.27e+03	3.10e+01	9.60e+04	3.37e+03	2.45e+02	9.76e+04	9.14e+05
Cyclosporine A	4.40e+03	2.77e+03	4.83e+02	4.21e+04	2.07e+02	1.81e+04	6.74e+04	6.68e+05
Cycloheximide	1.43e+04	2.99e+02	6.30e+01	2.24e+03	5.47e+03	1.12e+04	3.11e+04	3.19e+05
Cisplatin	1.33e+03	1.74e+03	3.13e+02	4.30e+04	1.13e+02	7.44e+03	5.35e+04	6.59e+05
Isoniazid	5.23e+03	6.84e+03	6.09e+02	5.64e+04	7.90e+01	6.86e+03	7.51e+04	1.08e+06
Bromoethylamine	8.59e+02	2.07e+03	1.59e+02	3.51e+04	5.80e+01	8.72e+03	4.68e+04	6.88e+05
Methylene dianiline	4.23e+03	8.30e+03	4.22e+03	4.25e+04	3.32e+03	2.92e+03	6.07e+04	9.91e+05
Mexiletine	1.75e+03	5.86e+02	3.30e+01	2.32e+04	6.00e+00	2.33e+04	4.88e+04	8.08e+05
Flutamide	2.09e+02	6.47e+03	2.58e+03	2.87e+04	1.10e+01	1.57e+04	5.16e+04	1.21e+06
Theophylline	7.78e+02	1.96e+02	1.36e+02	1.04e+04	3.10e+01	1.97e+04	3.12e+04	7.43e+05
Tunicamycin	7.99e+02	6.60e+01	6.00e+00	2.54e+03	1.00e+00	1.24e+04	1.58e+04	3.79e+05
Danazol	1.39e+02	2.56e+02	5.60e+01	2.26e+03	1.90e+01	2.64e+04	2.91e+04	7.47e+05
Bromobenzene	1.43e+03	6.70e+03	4.20e+01	1.95e+04	4.00e+01	6.83e+03	3.31e+04	9.27e+05
Phenacetin	1.14e+02	1.04e+03	3.20e+01	2.64e+04	3.90e+01	6.67e+02	2.82e+04	8.00e+05
Griseofulvin	1.16e+02	5.01e+03	1.82e+03	6.82e+03	8.00e+00	2.58e+04	3.82e+04	1.22e+06
Methyltestosterone	1.60e+02	2.05e+03	1.20e+01	2.37e+02	1.50e+01	2.47e+04	2.70e+04	1.06e+06
Mefenamic acid	3.31e+02	6.39e+03	2.68e+03	6.74e+03	3.00e+00	1.70e+04	3.10e+04	1.24e+06
Cyclophosphamide	3.07e+02	3.58e+02	8.00e+00	3.52e+03	2.00e+00	7.19e+03	1.14e+04	8.15e+05
Methyldopa	1.76e+02	1.81e+03	1.02e+02	1.80e+01	1.67e+02	5.56e+03	7.69e+03	9.84e+05
Metformin	1.81e+02	1.75e+03	4.70e+01	1.68e+02	5.10e+01	4.14e+03	6.23e+03	1.01e+06

Table S7: **Number of patches with lesions per test study.** Compounds are sorted by their percentage of slides with lesions.

	Necrosis	Cellular infiltration	Bile duct proliferation	Fatty change	Hypertrophy	Increased mitosis	Macro average
ABCC3	3.20	3.00	3.28	4.39	5.17	2.20	3.54
GSTM3	2.64	2.71	2.93	3.10	3.12	1.29	2.63
GPX2	3.60	2.77	3.11	2.42	2.75	1.00	2.61
GSTA3	1.50	1.71	2.15	3.00	4.02	1.73	2.35
ALDH1A1	0.84	1.43	2.28	2.58	3.64	1.58	2.06
ALDH1A7	0.99	1.35	2.15	2.32	3.31	1.36	1.91
ADAM8	2.48	2.22	2.81	1.57	1.37	0.40	1.81
ATF3	3.27	2.64	1.96	1.44	0.98	0.50	1.80
ACOT1	2.25	1.40	2.27	1.35	1.15	1.97	1.73
TRIB3	2.84	2.01	1.31	1.50	1.65	0.65	1.66
AKR7A3	1.26	1.30	1.02	2.19	2.56	1.59	1.66
CCL2	3.12	2.54	2.15	1.29	0.70	0.02	1.64
TNFRSF12A	2.97	2.21	1.54	1.46	1.10	0.51	1.63
AKR1B8	2.67	1.91	1.61	1.44	1.40	0.61	1.61
ASNS	2.33	1.76	1.87	1.29	1.69	0.59	1.59
RND1	3.13	2.28	1.95	1.18	0.87	0.06	1.58
PHGDH	1.85	1.38	1.49	1.57	1.99	0.98	1.54
OLR59	-2.45	-1.74	-1.68	-1.27	-1.16	-0.71	-1.50
ABCG8	-2.06	-1.54	-1.80	-1.47	-1.38	-0.86	-1.52
PKLR	-2.42	-1.75	-1.91	-1.34	-1.40	-0.78	-1.60
MTMR7	-2.09	-1.76	-2.05	-1.50	-1.78	-0.69	-1.64
AKR1C1	-2.73	-1.95	-1.97	-1.46	-1.44	-0.77	-1.72
NREP	-3.33	-2.02	-1.89	-1.38	-1.47	-0.43	-1.75
THRSP	-2.90	-2.18	-2.35	-1.40	-1.28	-0.57	-1.78
OAT	-2.23	-1.86	-2.02	-1.86	-2.09	-1.08	-1.85
LOX	-1.49	-1.53	-2.13	-2.08	-3.11	-0.84	-1.86
SLC22A8	-2.43	-2.07	-2.52	-1.71	-1.65	-0.82	-1.87
BMF	-3.01	-2.17	-2.21	-1.66	-1.56	-0.68	-1.88
INMT	-3.13	-2.63	-3.16	-1.77	-1.57	-1.22	-2.25
NOX4	-3.21	-2.62	-2.96	-2.02	-2.00	-0.98	-2.30
CAR3	-3.30	-2.50	-3.16	-2.22	-2.10	-0.95	-2.37
OBP3	-3.02	-2.70	-3.66	-2.21	-2.15	-1.26	-2.50
STAC3	-5.14	-4.07	-4.97	-3.91	-4.16	-2.04	-4.05

Table S8: **Genes indicative of general toxic exposure.** Listed genes with the highest mean absolute predicted gene expression across patches with lesions. Analysis was conducted using five studies that report a large range of lesions: thioacetamide, methylene dianiline, ethionamide, bromobenzene, hexachlorobenzene.

	Increased mitosis	Mean expression	Correlation prediction
CCNA2	2.52	0.33	0.40
KNSTRN	2.50	-0.01	0.42
CDKN3	2.41	0.01	0.27
ECT2	2.32	0.31	0.51
CENPW	2.10	0.10	0.46
CDK1	2.09	0.21	0.44
CCNB1	2.05	0.20	0.45
CDCA3	2.05	0.17	0.51
SKA1	2.04	0.31	0.60
MKI67	2.01	0.44	0.58
TOP2A	1.90	0.17	0.46
KIF18B	1.84	0.37	0.40
AURKB	1.81	0.39	0.59
UBE2C	1.80	0.25	0.47
SPC24	1.77	0.12	0.70
BUB1	1.73	0.27	0.57
NUSAP1	1.73	0.13	0.52
PTTG1	1.72	0.17	0.41
IQGAP3	1.67	0.29	0.53
CCNB2	1.67	0.15	0.39
AURKA	1.66	0.27	0.57
NCAPH	1.66	0.34	0.53
HMMR	1.52	0.26	0.53
ARHGAP11A	1.51	0.53	0.70
ASPM	1.49	0.22	0.56
SKA3	1.48	0.27	0.68
HJURP	1.45	0.46	0.61
ANLN	1.30	0.29	0.49
CENPF	1.30	0.21	0.49
FAM83D	1.28	-0.06	0.33
BIRC5	1.17	0.17	0.37
CDC20	1.07	-0.09	0.61
EXO1	0.99	0.21	0.35
E2F1	0.96	0.37	0.57
CKS2	0.93	0.24	0.53
CDKN2C	0.91	0.00	0.56
HAUS4	0.88	0.10	0.49
H2AX	0.88	0.32	0.52
RBL1	0.87	0.36	0.70
TUBB5	0.84	0.41	0.75
DCTPP1	0.83	0.32	0.41
SMC4	0.75	0.14	0.33
TK1	0.74	0.16	0.40
MBOAT1	0.72	0.35	0.67
H2AZ1	0.70	0.18	0.47
MCM5	0.69	0.28	0.62
CCNF	0.68	-0.01	0.51
NETO2	0.67	-0.09	0.50
CCNE1	0.65	0.29	0.70
PRIM1	0.58	0.15	0.65
DNAJC9	0.54	0.13	0.47

Table S9: **List of 51 genes identified as being linked to increased mitosis**. Analysis was conducted using eight compounds that reported increased mitosis. Patch-level average gene expression for patches containing mitosis compared to those identified with any of the five other lesions.

	necrosis	mean expression lesion	correlation prediction
ATF3	3.37	1.40	0.65
TRIB3	3.23	1.47	0.62
MAFF	3.10	0.95	0.66
TNFRSF12A	3.09	1.30	0.79
ZFAND2A	2.87	0.97	0.75
IFRD1	2.11	0.98	0.60
MDM2	2.01	0.99	0.71
FGF21	2.01	0.93	0.46
MYC	1.88	0.77	0.77
CDKN1A	1.87	0.92	0.35
GADD45A	1.87	0.80	0.62
HSPB1	1.86	0.88	0.61
SRXN1	1.84	0.84	0.63
HSPA1A	1.78	0.52	0.41
SLC3A2	1.71	0.80	0.72
DDIT3	1.70	0.75	0.49
HMOX1	1.65	0.64	0.75
AEN	1.62	0.60	0.71
TMBIM1	1.62	0.61	0.73
LOC100911177	1.56	0.64	0.74
LOC100912041	1.53	0.56	0.63
FBXO30	1.37	0.57	0.68
PPP1R15A	1.36	0.46	0.74
CCDC86	1.28	0.46	0.56
JUN	1.25	0.62	0.65
HSPB8	1.24	0.45	0.75
FOSL1	1.21	0.02	0.59
AIMP2	1.20	0.53	0.80
SLC7A11	1.15	0.43	0.38
G6PD	1.13	0.56	0.41
TXNRD1	1.13	0.51	0.57
RIOK3	1.07	0.36	0.50
CXCL2	1.01	0.05	0.47
ALDH4A1	-1.02	-0.37	0.80
ACACA	-1.03	-0.32	0.66
AQP9	-1.09	-0.44	0.47
RARB	-1.10	-0.42	0.67
CAR14	-1.11	-0.38	0.78
FDFT1	-1.12	-0.18	0.40
BHMT2	-1.12	-0.34	0.80
IGFBP3	-1.12	-0.49	0.72
CADPS2	-1.13	-0.47	0.66
EBP	-1.13	-0.39	0.74
HGD	-1.16	-0.49	0.72
SUOX	-1.18	-0.48	0.83
GK	-1.21	-0.57	0.74
ARHGEF19	-1.22	-0.47	0.50
ALBFM1	-1.23	-0.57	0.73
VEPH1	-1.23	-0.27	0.63
ACKR4	-1.27	-0.38	0.83
ADAMTS7	-1.30	-0.63	0.65
CTH	-1.31	-0.33	0.54
HMGCS1	-1.36	-0.38	0.32
PRLR	-1.41	-0.29	0.33
HDHD3	-1.43	-0.58	0.38
HES6	-1.53	-0.55	0.64
ABCA8A	-1.55	-0.72	0.86
ACSS2	-1.75	-0.45	0.66
COQ8A	-1.86	-0.74	0.82
CXCL12	-2.04	-0.91	0.79
FADS1	-2.07	-0.64	0.63
XPNPEP2	-2.15	-0.93	0.72
ACACB	-2.18	-0.65	0.58
CYP7A1	-2.36	-0.67	0.66
AVPR1A	-2.52	-0.93	0.83
NREP	-3.30	-1.45	0.80

Table S10: **Caption next page.**

Table S10: **List of 66 genes identified as being linked to necrosis**. Analysis was conducted using five compounds that reported necrosis. Patch-level average gene expression for patches containing necrosis compared to those identified with any of the five other lesions.

	cellular infiltration	mean expression lesion	correlation prediction
CXCL1	2.38	0.92	0.41
BCL2A1	1.67	0.71	0.48
S100A4	1.66	0.78	0.41
CXCL10	1.63	0.72	0.31
S100A8	1.49	0.56	0.50
CHI3L1	1.32	0.54	0.58
S100A9	1.23	0.38	0.36
IL1B	1.17	0.41	0.51
CAPG	1.11	0.54	0.55
TGFB1	1.06	0.35	0.59
PTPRC	0.91	0.35	0.37
ACKR3	0.88	0.35	0.46
FILIP1L	0.83	0.35	0.47
C5AR1	0.77	0.37	0.42
CD83	0.76	0.29	0.40
EVI2A	0.74	0.26	0.48
LCP1	0.73	0.29	0.55
ICAM1	0.72	0.33	0.56
CYBB	0.70	0.24	0.43
IL7	0.69	0.14	0.44
CASP1	0.67	0.28	0.49
PSMB9	0.64	0.30	0.41
GLIPR1	0.62	0.29	0.55
AP1S2	0.57	0.23	0.40
PLVAP	0.54	-0.04	0.31
SRGN	0.53	0.22	0.39
IFITM2	0.52	0.26	0.68

Table S11: **List of 27 genes identified as being linked to cellular infiltration**. Analysis was conducted using five compounds that reported cellular infiltration. Patch-level average gene expression for patches containing cellular infiltration compared to those identified with any of the five other lesions.

	bile duct proliferation	mean expression lesion	correlation prediction
SERPINA7	3.14	1.08	0.58
BEX4	2.20	1.01	0.78
CDH13	2.12	0.75	0.34
LGALS2	2.01	0.69	0.52
CLDN7	1.60	0.66	0.68
ANXA5	1.57	0.65	0.75
STING1	1.52	0.60	0.60
RGS5	1.50	0.58	0.53
UAP1L1	1.47	0.71	0.63
CD24	1.45	0.49	0.70
TAGLN	1.34	0.38	0.78
AJUBA	1.21	0.58	0.63
IGF2R	1.17	0.53	0.76
ASAH1	1.17	0.47	0.67
LBH	1.16	0.40	0.77
FHL2	1.05	0.27	0.77
UNC5CL	1.01	0.49	0.38
EDEM1	-1.06	-0.52	0.84
AGXT	-1.37	-0.69	0.70
DPT	-1.38	-0.60	0.89
C8G	-1.69	-0.80	0.85
ASPG	-1.79	-0.78	0.65
GPT	-2.25	-1.04	0.61
SERPINA4	-2.32	-1.18	0.88

Table S12: **List of 24 genes identified as being linked to bile duct proliferation.** Analysis was conducted using two compounds that reported bile duct proliferation. Patch-level average gene expression for patches containing bile duct proliferation compared to those identified with any of the five other lesions.

	fatty change	mean expression lesion	correlation prediction
ACOT1	1.93	0.47	0.59
GSTP1	1.48	0.58	0.40
CIDEA	1.04	0.38	0.30
ACOT2	0.83	0.13	0.31
TMED3	0.75	0.35	0.53
CYP1A1	0.75	-0.06	0.63
HID1	0.71	0.18	0.54
FKBP11	0.65	0.25	0.72
IRAK1	0.63	0.26	0.46
MANBA	0.56	0.17	0.65
ACP3	-0.52	-0.21	0.35
EPSTI1	-0.58	-0.15	0.42
C5	-0.61	-0.22	0.30
SLC6A6	-1.01	-0.48	0.48

Table S13: **List of 14 genes identified as being linked to fatty change.** Analysis was conducted using four compounds that reported fatty change. Patch-level average gene expression for patches containing fatty change compared to those identified with any of the five other lesions.

.

	hypertrophy	mean expression lesion	correlation prediction
GSTA3	3.90	1.52	0.72
ALDH1A1	3.56	1.43	0.76
CYP1A1	3.23	0.25	0.43
ALDH1A7	3.19	1.31	0.43
ACSM2	1.44	0.33	0.35
RGD1559459	1.33	0.44	0.77
ADGRG2	1.32	0.45	0.59
VNN1	1.24	0.28	0.51
ME1	1.18	-0.26	0.43
EPHX1	1.04	0.31	0.46
POR	0.91	0.11	0.54
GCLM	0.71	0.00	0.53
GCLC	0.62	0.05	0.46
GSTA4	0.60	-0.06	0.72
TFRC	0.58	0.04	0.46
ADH4	0.51	-0.29	0.45
MCRIP2	0.51	0.08	0.68
GSTM2	0.50	-0.16	0.64
NAPSA	-0.54	0.45	0.31
PLVAP	-0.85	0.23	0.37
SLC34A2	-1.08	-0.25	0.37
SLC6A6	-1.28	-0.38	0.44
LOX	-2.97	-1.32	0.56

Table S14: **List of 23 genes identified as being linked to hypertrophy.** Analysis was conducted using eleven compounds that reported hypertrophy. Patch-level average gene expression for patches containing hypertrophy compared to those identified with any of the 5 other lesions.

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

Table S15: **iBOT hyperparameters used SSL 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

Table S16: **Hyperparameters used in GEESE classification**. A single 24GB NVIDIA GeForce RTX 3090 GPU was used for each MIL model using weakly-supervised learning and slide-level labels.