Supplementary Materials

MorphADMET links chemical structure and cellular morphology for context dependent ADMET assessment

Md Wasi Ul Kabir [1], Asif Mahmud [1], Sayanto Pal [1], Ryan King [1], Thomas R Caulfield [1,*] and Naeyma N Islam [1,*]
Correspondence: thomas@digitalethercomputing.com; naeyma@digitalethercomputing.com)
Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Table S1. *Cross-validation performance (Development set) for structure-based prediction of strong morphological response across five assay contexts. Values are mean average precision (AP) across grouped development folds, with chemically related compounds kept in the same fold. The ensemble combines candidate models only when averaging their predictions improves validation AP. Bold indicates the best individual model within each assay context. Reserved test set performance is reported in Table S3.*

Context	Logistic regression	Histogram boosting	CatBoost	XGBoost	LightGBM	ExtraTrees	Ensemble CV AP
JUMP	0.366	0.427	0.401	0.386	**0.448**	0.416	0.458
EU-OS	0.382	0.405	**0.431**	0.412	0.424	0.393	0.460
LINCS	0.296	0.312	0.264	0.296	0.299	**0.317**	0.367
RxRx3	0.358	0.391	0.360	0.314	0.362	**0.408**	0.442
RxRx19a	0.385	0.340	0.411	0.345	0.378	**0.433**	0.440

Table S2. *Response-prediction ensembles selected using development data. Each row identifies one ensemble member and its chemical feature representation. Predictions are averaged with equal weights within each assay context. ErG denotes extended reduced graph fingerprints. All models use binary strong-response labels with equal training weight per compound. The two RxRx19a ExtraTrees models differ in their minimum leaf size. OASIS was excluded from this response ensemble-selection analysis because too few strong responders were available within the training and validation folds.*

Context	Ensemble member	Chemical features
JUMP	LightGBM	Morgan fingerprints (count), Avalon count and ErG fingerprints + RDKit molecular descriptors
	ExtraTrees	Morgan fingerprints (binary), atom-pair, topological-torsion, RDKit topological and MACCS fingerprints + RDKit molecular descriptors

Context	Ensemble member	Chemical features
EU-OS	CatBoost	Mordred molecular descriptors + RDKit molecular descriptors
	LightGBM	Morgan fingerprints (binary), atom-pair, topological-torsion, RDKit topological and MACCS fingerprints + RDKit molecular descriptors
	CatBoost	Morgan fingerprints (count), Avalon count and ErG fingerprints + RDKit molecular descriptors
LINCS	ExtraTrees	Morgan fingerprints (binary), atom-pair, topological-torsion, RDKit topological and MACCS fingerprints + RDKit molecular descriptors
	Histogram gradient boosting	Morgan fingerprints (binary) + RDKit molecular descriptors
	XGBoost	Morgan fingerprints (binary), atom-pair, topological-torsion, RDKit topological and MACCS fingerprints + RDKit molecular descriptors
RxRx3	ExtraTrees	Morgan fingerprints (binary) + RDKit molecular descriptors
	Histogram gradient boosting	Morgan fingerprints (count), Avalon count and ErG fingerprints + RDKit molecular descriptors
	Logistic regression	Morgan fingerprints (binary), atom-pair, topological-torsion, RDKit topological and MACCS fingerprints + RDKit molecular descriptors
RxRx19a	ExtraTrees	Morgan fingerprints (binary) + RDKit molecular descriptors
	ExtraTrees	Morgan fingerprints (binary) + RDKit molecular descriptors

Table S3. *Reserved-test performance of MorphADMET response ensembles (Table S2) and a histogram gradient boosting baseline using 14 chemical descriptors across five assay contexts using binary strong-response labels. Bold identifies the higher score for each metric within each assay context. Development n denotes the number of compounds used for model development, Test n the number of compounds in the reserved test set, and Test positives the number of strong-response compounds in that set.*

Context	Development n	Test n	Test positives	Baseline AP	MorphADMET AP	Baseline AUROC	MorphADMET AUROC
JUMP / U2OS	3449	862	86	0.306	**0.407**	0.768	**0.815**
EU-OS / HepG2	1873	453	34	0.326	**0.460**	0.799	**0.859**
LINCS / A549	886	230	26	0.274	**0.472**	0.773	**0.807**
RxRx3 / HUVEC	707	154	21	**0.466**	0.388	0.813	**0.822**
RxRx19a / HRCE	777	169	13	0.230	**0.353**	0.753	**0.876**

Table S4. *Development set cross-validation performance of candidate structure-based ADMET models and the selected ensemble. Values are mean development-fold scores under grouped chemical splits. Average precision (AP) is reported for classification endpoints and mean absolute error (MAE) for solubility in $\log_{10}$ mol/L. Bold indicates the best individual model for each endpoint. Ensemble members are listed in Table S5, and reserved test set results are reported in Table S6.*

Endpoint	Metric	CatBoost	LightGBM	XGBoost	ExtraTrees	Linear	MapLight	Multilayer perceptron (MLP)	Mixture-of-experts neural network	D-MPNN	CheMeleon	Ensemble
hERG	AP	0.880	0.882	**0.883**	0.878	0.821	0.874	0.863	0.858	0.851	0.871	0.891
DICTrank	AP	0.880	0.864	0.874	**0.882**	0.861	0.879	0.868	0.858	0.859	0.872	0.889
DILI	AP	0.782	0.766	0.777	**0.795**	0.759	0.794	0.777	0.772	0.757	0.763	0.809
DIRIL	AP	**0.786**	0.746	0.755	0.764	0.776	0.783	0.775	0.769	0.692	0.758	0.798
Ames	AP	0.905	0.906	**0.907**	0.900	0.882	0.898	0.896	0.895	0.895	0.907	0.917
BBB	AP	0.970	0.972	0.973	0.971	0.973	0.970	0.972	0.969	0.952	**0.974**	0.977

Endpoint	Metric	CatBoost	LightGBM	XGBoost	ExtraTrees	Linear	MapLight	Multilayer perceptron (MLP)	Mixture-of-experts neural network	D-MPNN	CheMeleon	Ensemble
CYP1A2	AP	0.930	**0.934**	0.934	0.924	0.909	0.928	0.925	0.924	0.909	0.933	0.941
CYP2C19	AP	0.881	0.885	**0.885**	0.870	0.850	0.872	0.871	0.871	0.872	0.879	0.893
CYP2C9	AP	0.824	0.827	**0.827**	0.817	0.783	0.823	0.803	0.804	0.806	0.822	0.842
CYP2D6	AP	0.713	0.722	**0.724**	0.702	0.646	0.713	0.687	0.684	0.700	0.713	0.741
CYP3A4	AP	0.875	**0.880**	0.879	0.870	0.835	0.874	0.867	0.865	0.860	0.874	0.890
Solubility	MAE	0.772	**0.771**	0.775	0.796	0.886	0.798	0.787	0.786	0.877	0.773	0.722

Table S5. *Selected members of the structure-based ADMET ensemble for each endpoint. Ensemble membership was determined by improvement in combined development set cross-validation performance. A model could be selected more than once during greedy ensemble selection, increasing its contribution to the final prediction; each retained model is listed once here. The selected ensembles were refitted on the full development sets before test evaluation.*

Endpoint	Selected ensemble models
hERG	CatBoost; LightGBM; XGBoost; ExtraTrees; Regularized linear; MapLight; Multilayer perceptron (MLP); Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon
DICTrank	CatBoost; LightGBM; XGBoost; ExtraTrees; Regularized linear; MapLight; Multilayer perceptron (MLP); D-MPNN; Fine-tuned CheMeleon
DILI	LightGBM; XGBoost; ExtraTrees; Regularized linear; MapLight; Multilayer perceptron (MLP); Fine-tuned CheMeleon
DIRIL	CatBoost; Multilayer perceptron (MLP)
Ames	LightGBM; XGBoost; ExtraTrees; Regularized linear; Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon
BBB	LightGBM; XGBoost; Regularized linear; MapLight; Fine-tuned CheMeleon
CYP1A2	LightGBM; ExtraTrees; MapLight; Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon
CYP2C19	CatBoost; LightGBM; XGBoost; Regularized linear; Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon
CYP2C9	LightGBM; XGBoost; ExtraTrees; MapLight; Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon

Endpoint	Selected ensemble models
CYP2D6	LightGBM; XGBoost; ExtraTrees; MapLight; Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon
CYP3A4	LightGBM; ExtraTrees; MapLight; Multilayer perceptron (MLP); Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon
Solubility	CatBoost; LightGBM; XGBoost; Regularized linear; Multilayer perceptron (MLP); Mixture-of-experts neural network; D-MPNN; Fine-tuned CheMeleon

***Table S6**. Reserved-test performance of structure-based ADMET models. MorphADMET is compared with four benchmark methods using matched endpoint-specific compounds and labels. Higher AP and AUROC indicate better classification performance; lower MAE indicates better solubility prediction. Solubility MAE is expressed in $\log_{10}$ mol/L. Bold identifies the best observed score in each row, not statistical significance. Sample sizes are given in Table S15, and paired confidence intervals in Table S16.*

Endpoint	Metric	MorphADMET	MapLight	D-MPNN	CheMeleon	ADMET-AI
Ames	AP	0.579	0.531	0.503	**0.584**	0.495
Ames	AUROC	**0.772**	0.720	0.683	0.761	0.685
BBBP	AP	**0.913**	0.908	0.898	0.903	0.880
BBBP	AUROC	**0.854**	0.835	0.814	0.832	0.809
CYP1A2	AP	**0.929**	0.888	0.859	0.902	0.917
CYP1A2	AUROC	**0.960**	0.930	0.934	0.952	0.953
CYP2C19	AP	0.574	0.539	0.419	**0.579**	0.463
CYP2C19	AUROC	**0.796**	0.748	0.761	0.779	0.757
CYP2C9	AP	0.230	**0.271**	0.267	0.213	0.226
CYP2C9	AUROC	0.715	**0.756**	0.702	0.676	0.752
CYP2D6	AP	0.745	0.669	0.744	0.767	**0.776**

Endpoint	Metric	MorphADMET	MapLight	D-MPNN	CheMeleon	ADMET-AI
CYP2D6	AUROC	0.869	0.816	0.862	0.867	**0.880**
CYP3A4	AP	0.888	0.863	0.835	0.802	**0.890**
CYP3A4	AUROC	0.958	0.935	0.927	0.934	**0.962**
DICTrank	AP	**0.914**	0.902	0.897	0.909	0.879
DICTrank	AUROC	0.757	0.706	0.707	**0.761**	0.670
DILI	AP	0.782	0.761	**0.801**	0.734	0.770
DILI	AUROC	0.648	0.647	0.639	0.625	**0.672**
DIRIL	AP	0.746	0.702	0.552	**0.747**	0.648
DIRIL	AUROC	**0.665**	0.636	0.470	0.664	0.602
hERG	AP	**0.682**	0.667	0.654	0.650	0.659
hERG	AUROC	0.853	**0.862**	0.835	0.818	0.844
Solubility	MAE	**0.660**	0.722	0.863	0.690	0.723

Tables S7–S13 summarize the cross-validation performance and selected configurations for morphology-based endpoint prediction across seven morphological representations and 12 endpoints. For each endpoint–representation combination, 78 candidate configurations were evaluated across six model families, using feature selection or principal component analysis (PCA) with 16, 32, or 64 features or components. Model hyperparameters were varied within predefined ranges, and the best configuration was selected using three-fold scaffold-grouped development cross-validation. All preprocessing was fitted within training folds, and selected models were refitted for final test evaluation. ET denotes ExtraTrees; ET-selected features were selected using ExtraTrees feature importance.

Table S7*. Morphology-based ADMET model selection for JUMP. Panel (a) compares the best mean cross-validation scores across six model families for each endpoint, while panel (b) presents the selected model configurations, including input transformation, feature dimensions, and hyperparameters. Scores were obtained using three-fold grouped development cross-validation. Bold values identify the selected model family.*

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	0.283	0.316	**0.341**	0.340	0.314	0.295
DICTrank	AP	0.777	**0.792**	0.780	0.783	0.786	0.789
DILI	AP	0.671	0.676	0.685	0.675	**0.693**	0.680
DIRIL	AP	**0.622**	0.608	0.584	0.589	0.598	0.613
Ames	AP	0.352	**0.375**	0.370	0.363	0.360	0.366
BBB	AP	0.688	0.731	**0.746**	0.738	0.719	0.730
CYP1A2	AP	**0.571**	0.541	0.522	0.534	0.536	0.560
CYP2C19	AP	**0.335**	0.294	0.263	0.317	0.282	0.327
CYP2C9	AP	0.333	0.285	0.259	0.242	0.308	**0.334**
CYP2D6	AP	0.291	0.298	**0.325**	0.309	0.295	0.281
CYP3A4	AP	**0.387**	0.375	0.343	0.377	0.358	0.335
Solubility	MAE	1.680	1.658	1.693	1.687	**1.652**	1.762

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	LightGBM	ET-selected features	32	leaves=15
DICTrank	CatBoost	PCA components	64	depth=3
DILI	ExtraTrees	PCA components	16	min. leaf=8
DIRIL	Logistic	ET-selected features	16	C=1
Ames	CatBoost	ET-selected features	64	depth=3

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
BBB	LightGBM	ET-selected features	32	leaves=7
CYP1A2	Logistic	PCA components	64	C=0.01
CYP2C19	Logistic	PCA components	32	C=0.01
CYP2C9	MLP	ET-selected features	16	hidden=(64); α=0.01
CYP2D6	LightGBM	PCA components	16	leaves=7
CYP3A4	Logistic	PCA components	64	C=1
Solubility	ExtraTrees	ET-selected features	64	min. leaf=3

Best development CV score within each model family

Table S8. *Morphology-based ADMET model selection for OASIS CellProfiler. Panel (a) reports the best mean cross-validation performance of six model families across 12 endpoints. Panel (b) provides the selected models and their corresponding input transformations, feature dimensions, and hyperparameters. Bold values identify the selected model family.*

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	**0.522**	0.493	0.455	0.454	0.515	0.522
DICTrank	AP	0.837	0.824	0.813	0.825	0.815	**0.868**
DILI	AP	0.800	0.811	0.805	0.801	0.772	**0.824**
DIRIL	AP	**0.663**	0.639	0.624	0.648	0.642	0.658
Ames	AP	**0.568**	0.524	0.475	0.486	0.534	0.525
BBB	AP	**0.669**	0.635	0.623	0.641	0.645	0.662
CYP1A2	AP	0.402	**0.534**	0.478	0.460	0.534	0.458
CYP2C19	AP	0.227	0.250	0.302	**0.382**	0.293	0.301
CYP2C9	AP	0.223	0.276	0.238	0.253	**0.380**	0.363
CYP2D6	AP	0.422	0.431	**0.441**	0.420	0.408	0.435

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
CYP3A4	AP	0.349	0.363	0.342	0.311	0.318	**0.409**
Solubility	MAE	1.547	1.496	1.555	**1.478**	1.500	1.737

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	Logistic	PCA components	32	C=0.1
DICTrank	MLP	PCA components	64	hidden=(64); α=0.01
DILI	MLP	ET-selected features	32	hidden=(128,64); α=0.1
DIRIL	Logistic	PCA components	32	C=0.01
Ames	Logistic	PCA components	16	C=0.01
BBB	Logistic	ET-selected features	16	C=1
CYP1A2	CatBoost	PCA components	16	depth=5
CYP2C19	XGBoost	PCA components	64	depth=3
CYP2C9	ExtraTrees	ET-selected features	16	min. leaf=3
CYP2D6	LightGBM	ET-selected features	16	leaves=7
CYP3A4	MLP	PCA components	32	hidden=(128,64); α=0.1
Solubility	XGBoost	ET-selected features	64	depth=5

Best development CV score within each model family

Table S9. *Morphology-based ADMET model selection for OASIS DINOv2. Panel (a) compares the best mean development cross-validation scores across six model families and 12 endpoints. Panel (b) summarizes the selected configurations, including model type, input transformation, feature dimensions, and hyperparameters. Bold values identify the selected model family.*

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	0.534	0.542	0.535	**0.565**	0.526	0.552

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
DICTrank	AP	**0.883**	0.845	0.813	0.825	0.852	0.849
DILI	AP	0.794	0.786	0.789	0.788	0.799	**0.805**
DIRIL	AP	0.624	0.621	0.623	0.612	0.615	**0.628**
Ames	AP	0.531	0.494	0.493	0.491	0.508	**0.551**
BBB	AP	0.700	0.696	0.688	0.704	**0.706**	0.685
CYP1A2	AP	0.476	0.486	0.476	0.489	**0.535**	0.523
CYP2C19	AP	0.320	0.285	**0.389**	0.353	0.316	0.322
CYP2C9	AP	0.208	0.299	0.251	0.238	0.307	**0.337**
CYP2D6	AP	**0.369**	0.344	0.323	0.331	0.324	0.295
CYP3A4	AP	**0.522**	0.334	0.319	0.385	0.411	0.423
Solubility	MAE	1.530	**1.500**	1.551	1.570	1.503	1.795

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	XGBoost	PCA components	64	depth=3
DICTrank	Logistic	PCA components	64	C=1
DILI	MLP	ET-selected features	64	hidden=(128,64); α=0.1
DIRIL	MLP	ET-selected features	64	hidden=(128,64); α=0.1
Ames	MLP	PCA components	16	hidden=(64); α=0.01
BBB	ExtraTrees	PCA components	32	min. leaf=3
CYP1A2	ExtraTrees	PCA components	16	min. leaf=3
CYP2C19	LightGBM	ET-selected features	64	leaves=7

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
CYP2C9	MLP	PCA components	16	hidden=(128,64); α=0.1
CYP2D6	Logistic	PCA components	64	C=1
CYP3A4	Logistic	PCA components	64	C=0.01
Solubility	CatBoost	PCA components	16	depth=5

Best development CV score within each model family

Table S10. *Morphology-based ADMET model selection for LINCS. Panel (a) presents the best mean cross-validation scores of six model families across 12 endpoints. Panel (b) lists the selected model configurations, including input transformations, feature dimensions, and hyperparameters. Bold values identify the selected model family.*

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	0.582	0.572	0.568	0.553	**0.602**	0.531
DICTrank	AP	0.884	0.898	0.896	0.889	**0.904**	0.881
DILI	AP	**0.755**	0.726	0.747	0.743	0.727	0.751
DIRIL	AP	0.731	**0.739**	0.710	0.731	0.735	0.717
Ames	AP	0.392	0.389	0.357	0.341	**0.403**	0.292
BBB	AP	**0.825**	0.821	0.801	0.799	0.821	0.797
CYP1A2	AP	0.589	0.532	0.532	0.513	0.556	**0.609**
CYP2C19	AP	0.329	0.332	**0.381**	0.316	0.349	0.379
CYP2C9	AP	0.220	0.175	**0.244**	0.221	0.148	0.236
CYP2D6	AP	0.503	0.506	0.544	**0.568**	0.515	0.468
CYP3A4	AP	0.367	0.423	0.402	0.432	**0.472**	0.376
Solubility	MAE	1.405	1.399	1.467	1.458	**1.395**	1.601

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	ExtraTrees	ET-selected features	32	min. leaf=8
DICTrank	ExtraTrees	PCA components	16	min. leaf=3
DILI	Logistic	PCA components	32	C=1
DIRIL	CatBoost	ET-selected features	16	depth=3
Ames	ExtraTrees	PCA components	64	min. leaf=3
BBB	Logistic	PCA components	64	C=0.01
CYP1A2	MLP	PCA components	16	hidden=(64); α=0.01
CYP2C19	LightGBM	PCA components	64	leaves=7
CYP2C9	LightGBM	PCA components	64	leaves=7
CYP2D6	XGBoost	ET-selected features	32	depth=5
CYP3A4	ExtraTrees	PCA components	16	min. leaf=8
Solubility	ExtraTrees	ET-selected features	32	min. leaf=8

Best development CV score within each model family

Table S11. Morphology-based ADMET model selection for EU-OS. Panel (a) compares the best mean cross-validation scores across six model families for each endpoint. Panel (b) reports the selected models and associated input transformations, feature dimensions, and hyperparameters. Bold values identify the selected model family.

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	**0.491**	0.397	0.404	0.399	0.408	0.427
DICTrank	AP	**0.906**	0.867	0.866	0.877	0.901	0.893
DILI	AP	0.657	**0.685**	0.651	0.648	0.632	0.637
DIRIL	AP	0.665	0.664	0.707	0.663	**0.707**	0.705
Ames	AP	**0.606**	0.483	0.465	0.475	0.539	0.468

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
BBB	AP	0.810	0.778	0.745	0.743	0.789	**0.813**
CYP1A2	AP	0.707	0.676	0.665	0.662	0.670	**0.709**
CYP2C19	AP	**0.407**	0.351	0.293	0.346	0.328	0.385
CYP2C9	AP	**0.384**	0.361	0.308	0.335	0.322	0.368
CYP2D6	AP	0.476	**0.481**	0.429	0.422	0.404	0.428
CYP3A4	AP	0.511	0.483	0.493	0.463	**0.566**	0.513
Solubility	MAE	1.574	**1.496**	1.607	1.585	1.512	1.975

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	Logistic	ET-selected features	64	C=0.1
DICTrank	Logistic	PCA components	16	C=0.1
DILI	CatBoost	PCA components	16	depth=3
DIRIL	ExtraTrees	PCA components	64	min. leaf=8
Ames	Logistic	ET-selected features	64	C=0.01
BBB	MLP	PCA components	64	hidden=(64); α=0.01
CYP1A2	MLP	ET-selected features	64	hidden=(128,64); α=0.1
CYP2C19	Logistic	ET-selected features	16	C=1
CYP2C9	Logistic	ET-selected features	64	C=0.1
CYP2D6	CatBoost	ET-selected features	64	depth=3
CYP3A4	ExtraTrees	ET-selected features	16	min. leaf=8
Solubility	CatBoost	PCA components	16	depth=3

Table S12. Morphology-based ADMET model selection for RxRx3. Panel (a) compares the best mean cross-validation scores across six model families for each endpoint. Panel (b) reports the selected models and associated input transformations, feature dimensions, and hyperparameters. Bold values identify the selected model family.

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	0.505	0.515	0.474	0.469	**0.530**	0.464
DICTrank	AP	**0.853**	0.831	0.826	0.827	0.845	0.834
DILI	AP	**0.795**	0.762	0.754	0.764	0.768	0.776
DIRIL	AP	0.655	0.646	0.641	0.619	0.668	**0.685**
Ames	AP	0.331	**0.335**	0.296	0.322	0.284	0.269
BBB	AP	0.800	0.812	0.787	0.789	**0.815**	0.757
CYP1A2	AP	0.511	0.503	**0.543**	0.521	0.473	0.492
CYP2C19	AP	0.255	0.287	0.282	0.275	0.289	**0.310**
CYP2C9	AP	**0.207**	0.204	0.174	0.156	0.177	0.200
CYP2D6	AP	0.570	0.564	**0.588**	0.565	0.551	0.528
CYP3A4	AP	0.392	0.388	**0.440**	0.424	0.390	0.413
Solubility	MAE	**1.528**	1.573	1.631	1.628	1.544	1.776

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	ExtraTrees	PCA components	16	min. leaf=8
DICTrank	Logistic	ET-selected features	16	C=0.01
DILI	Logistic	ET-selected features	64	C=1
DIRIL	MLP	ET-selected features	32	hidden=(128,64); α=0.1
Ames	CatBoost	PCA components	64	depth=5

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
BBB	ExtraTrees	PCA components	16	min. leaf=3
CYP1A2	LightGBM	ET-selected features	64	leaves=7
CYP2C19	MLP	PCA components	32	hidden=(128,64); α=0.1
CYP2C9	Logistic	PCA components	32	C=1
CYP2D6	LightGBM	PCA components	16	leaves=7
CYP3A4	LightGBM	ET-selected features	32	leaves=7
Solubility	Ridge	PCA components	16	α=100

Best development CV score within each model family

Table S13. Morphology-based ADMET model selection for RxRx19a. Panel (a) compares the best mean cross-validation scores across six model families for each endpoint. Panel (b) reports the selected models and associated input transformations, feature dimensions, and hyperparameters. Bold values identify the selected model family.

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
hERG	AP	0.666	0.690	0.674	0.677	**0.709**	0.653
DICTrank	AP	**0.852**	0.831	0.826	0.835	0.842	0.846
DILI	AP	**0.784**	0.723	0.725	0.752	0.742	0.778
DIRIL	AP	0.713	0.727	0.702	0.715	**0.741**	0.710
Ames	AP	**0.452**	0.397	0.403	0.383	0.436	0.451
BBB	AP	**0.772**	0.769	0.746	0.742	0.767	0.752
CYP1A2	AP	**0.526**	0.488	0.483	0.509	0.518	0.472
CYP2C19	AP	0.224	0.242	0.185	0.235	**0.269**	0.230
CYP2C9	AP	**0.332**	0.209	0.211	0.205	0.275	0.241
CYP2D6	AP	0.538	0.557	0.523	0.545	0.567	**0.569**

Endpoint	Metric	Linear	CatBoost	LightGBM	XGBoost	ExtraTrees	MLP
CYP3A4	AP	0.365	0.329	0.324	0.316	**0.370**	0.257
Solubility	MAE	1.523	**1.504**	1.508	1.515	1.509	1.707

Overall CV-selected single-model configuration

Endpoint	Selected model	Feature representation	Feature dimension (k)	Hyperparameters
hERG	ExtraTrees	PCA components	16	min. leaf=8
DICTrank	Logistic	PCA components	16	C=0.01
DILI	Logistic	PCA components	16	C=0.01
DIRIL	ExtraTrees	PCA components	16	min. leaf=3
Ames	Logistic	PCA components	64	C=1
BBB	Logistic	PCA components	16	C=0.01
CYP1A2	Logistic	PCA components	16	C=1
CYP2C19	ExtraTrees	ET-selected features	64	min. leaf=3
CYP2C9	Logistic	PCA components	64	C=0.1
CYP2D6	MLP	PCA components	32	hidden=(128,64); α=0.1
CYP3A4	ExtraTrees	PCA components	16	min. leaf=3
Solubility	CatBoost	ET-selected features	16	depth=5

Table S14*. Table S14. Development and test compound counts for morphology-based endpoint prediction in main Figure 6. For classification endpoints, parentheses report positive/negative counts; solubility cells report totals only. Counts include compounds with both measured morphology and endpoint labels and therefore vary by endpoint and representation. Some classes are sparse: each OASIS CYP3A4 representation has one positive among 26 test compounds. Table S17 reports the corresponding scores and conditional confidence intervals.*

Dataset / representation	**Endpoint**	**Development n (+/-)**	**Test n (+/-)**	**Development positive (%)**	**Test positive (%)**
JUMP (U2OS)	hERG	470 (96/374)	119 (24/95)	20.430	20.170

Dataset / representation	Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
JUMP (U2OS)	DICTrank	563 (423/140)	142 (107/35)	75.130	75.350
JUMP (U2OS)	DILI	642 (414/228)	162 (105/57)	64.490	64.810
JUMP (U2OS)	DIRIL	140 (72/68)	36 (18/18)	51.430	50.000
JUMP (U2OS)	Ames	484 (111/373)	125 (27/98)	22.930	21.600
JUMP (U2OS)	Solubility	509 (—/—)	309 (—/—)	—	—
JUMP (U2OS)	BBBP	256 (171/85)	63 (43/20)	66.800	68.250
JUMP (U2OS)	CYP1A2	264 (107/157)	69 (27/42)	40.530	39.130
JUMP (U2OS)	CYP2C19	282 (52/230)	69 (13/56)	18.440	18.840
JUMP (U2OS)	CYP2C9	268 (37/231)	68 (10/58)	13.810	14.710
JUMP (U2OS)	CYP2D6	282 (65/217)	71 (17/54)	23.050	23.940
JUMP (U2OS)	CYP3A4	266 (59/207)	68 (15/53)	22.180	22.060
OASIS (PHH, CellProfiler)	hERG	135 (51/84)	43 (12/31)	37.780	27.910
OASIS (PHH, CellProfiler)	DICTrank	192 (143/49)	55 (45/10)	74.480	81.820
OASIS (PHH, CellProfiler)	DILI	261 (201/60)	63 (45/18)	77.010	71.430
OASIS (PHH, CellProfiler)	DIRIL	145 (82/63)	34 (17/17)	56.550	50.000
OASIS (PHH, CellProfiler)	Ames	154 (44/110)	48 (3/45)	28.570	6.250

Dataset / representation	Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
OASIS (PHH, CellProfiler)	Solubility	246 (—/—)	75 (—/—)	—	—
OASIS (PHH, CellProfiler)	BBBP	180 (105/75)	41 (22/19)	58.330	53.660
OASIS (PHH, CellProfiler)	CYP1A2	90 (27/63)	22 (5/17)	30.000	22.730
OASIS (PHH, CellProfiler)	CYP2C19	101 (13/88)	27 (2/25)	12.870	7.410
OASIS (PHH, CellProfiler)	CYP2C9	97 (12/85)	28 (3/25)	12.370	10.710
OASIS (PHH, CellProfiler)	CYP2D6	103 (20/83)	29 (5/24)	19.420	17.240
OASIS (PHH, CellProfiler)	CYP3A4	92 (12/80)	26 (1/25)	13.040	3.850
OASIS (PHH, DINOv2)	hERG	135 (51/84)	43 (12/31)	37.780	27.910
OASIS (PHH, DINOv2)	DICTrank	192 (143/49)	55 (45/10)	74.480	81.820
OASIS (PHH, DINOv2)	DILI	261 (201/60)	63 (45/18)	77.010	71.430
OASIS (PHH, DINOv2)	DIRIL	145 (82/63)	34 (17/17)	56.550	50.000
OASIS (PHH, DINOv2)	Ames	154 (44/110)	48 (3/45)	28.570	6.250

Dataset / representation	Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
OASIS (PHH, DINOv2)	Solubility	246 (—/—)	75 (—/—)	—	—
OASIS (PHH, DINOv2)	BBBP	180 (105/75)	41 (22/19)	58.330	53.660
OASIS (PHH, DINOv2)	CYP1A2	90 (27/63)	22 (5/17)	30.000	22.730
OASIS (PHH, DINOv2)	CYP2C19	101 (13/88)	27 (2/25)	12.870	7.410
OASIS (PHH, DINOv2)	CYP2C9	97 (12/85)	28 (3/25)	12.370	10.710
OASIS (PHH, DINOv2)	CYP2D6	103 (20/83)	29 (5/24)	19.420	17.240
OASIS (PHH, DINOv2)	CYP3A4	92 (12/80)	26 (1/25)	13.040	3.850
LINCS (A549)	hERG	288 (85/203)	84 (17/67)	29.510	20.240
LINCS (A549)	DICTrank	206 (169/37)	69 (62/7)	82.040	89.860
LINCS (A549)	DILI	255 (184/71)	67 (45/22)	72.160	67.160
LINCS (A549)	DIRIL	149 (85/64)	41 (19/22)	57.050	46.340
LINCS (A549)	Ames	122 (28/94)	27 (4/23)	22.950	14.810
LINCS (A549)	Solubility	183 (—/—)	35 (—/—)	—	—
LINCS (A549)	BBBP	218 (154/64)	46 (30/16)	70.640	65.220
LINCS (A549)	CYP1A2	144 (54/90)	32 (8/24)	37.500	25.000

Dataset / representation	Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
LINCS (A549)	CYP2C19	159 (28/131)	42 (4/38)	17.610	9.520
LINCS (A549)	CYP2C9	142 (18/124)	39 (6/33)	12.680	15.380
LINCS (A549)	CYP2D6	153 (43/110)	35 (8/27)	28.100	22.860
LINCS (A549)	CYP3A4	144 (33/111)	35 (5/30)	22.920	14.290
EU-OS (HepG2)	hERG	367 (79/288)	94 (14/80)	21.530	14.890
EU-OS (HepG2)	DICTrank	146 (119/27)	42 (37/5)	81.510	88.100
EU-OS (HepG2)	DILI	167 (107/60)	37 (23/14)	64.070	62.160
EU-OS (HepG2)	DIRIL	101 (60/41)	32 (19/13)	59.410	59.380
EU-OS (HepG2)	Ames	138 (36/102)	26 (11/15)	26.090	42.310
EU-OS (HepG2)	Solubility	212 (—/—)	35 (—/—)	—	—
EU-OS (HepG2)	BBBP	150 (94/56)	36 (22/14)	62.670	61.110
EU-OS (HepG2)	CYP1A2	146 (77/69)	29 (13/16)	52.740	44.830
EU-OS (HepG2)	CYP2C19	160 (33/127)	42 (11/31)	20.620	26.190
EU-OS (HepG2)	CYP2C9	150 (27/123)	40 (9/31)	18.000	22.500
EU-OS (HepG2)	CYP2D6	156 (34/122)	44 (13/31)	21.790	29.550
EU-OS (HepG2)	CYP3A4	150 (51/99)	41 (11/30)	34.000	26.830
RxRx3-core (HUVEC)	hERG	245 (76/169)	58 (15/43)	31.020	25.860
RxRx3-core (HUVEC)	DICTrank	200 (154/46)	55 (47/8)	77.000	85.450
RxRx3-core (HUVEC)	DILI	223 (158/65)	60 (42/18)	70.850	70.000

Dataset / representation	Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
RxRx3-core (HUVEC)	DIRIL	152 (94/58)	39 (21/18)	61.840	53.850
RxRx3-core (HUVEC)	Ames	146 (39/107)	34 (5/29)	26.710	14.710
RxRx3-core (HUVEC)	Solubility	195 (—/—)	50 (—/—)	—	—
RxRx3-core (HUVEC)	BBBP	218 (146/72)	48 (31/17)	66.970	64.580
RxRx3-core (HUVEC)	CYP1A2	167 (43/124)	44 (9/35)	25.750	20.450
RxRx3-core (HUVEC)	CYP2C19	190 (21/169)	45 (3/42)	11.050	6.670
RxRx3-core (HUVEC)	CYP2C9	180 (15/165)	53 (4/49)	8.330	7.550
RxRx3-core (HUVEC)	CYP2D6	186 (48/138)	52 (16/36)	25.810	30.770
RxRx3-core (HUVEC)	CYP3A4	171 (27/144)	50 (6/44)	15.790	12.000
RxRx19a (HRCE)	hERG	189 (74/115)	56 (18/38)	39.150	32.140
RxRx19a (HRCE)	DICTrank	207 (158/49)	63 (55/8)	76.330	87.300
RxRx19a (HRCE)	DILI	235 (167/68)	66 (46/20)	71.060	69.700
RxRx19a (HRCE)	DIRIL	159 (98/61)	46 (25/21)	61.640	54.350
RxRx19a (HRCE)	Ames	157 (44/113)	39 (6/33)	28.030	15.380
RxRx19a (HRCE)	Solubility	213 (—/—)	53 (—/—)	—	—
RxRx19a (HRCE)	BBBP	235 (158/77)	55 (36/19)	67.230	65.450
RxRx19a (HRCE)	CYP1A2	180 (42/138)	48 (12/36)	23.330	25.000
RxRx19a (HRCE)	CYP2C19	202 (20/182)	53 (3/50)	9.900	5.660
RxRx19a (HRCE)	CYP2C9	195 (15/180)	59 (4/55)	7.690	6.780

Dataset / representation	Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
RxRx19a (HRCE)	CYP2D6	200 (54/146)	60 (19/41)	27.000	31.670
RxRx19a (HRCE)	CYP3A4	186 (25/161)	55 (5/50)	13.440	9.090

Table S15. *Development and test compound counts fo reach of the endpoint. Counts are reported for 12 endpoints, including positive/negative counts and class prevalence for classification tasks.*

Endpoint	Development n (+/-)	Test n (+/-)	Development positive (%)	Test positive (%)
hERG	12268 (6176/6092)	240 (44/196)	50.340	18.330
DICTrank	719 (514/205)	196 (154/42)	71.490	78.570
DILI	815 (533/282)	211 (143/68)	65.400	67.770
DIRIL	227 (125/102)	73 (38/35)	55.070	52.050
Ames	4659 (2807/1852)	174 (44/130)	60.250	25.290
Solubility	4161 (—/—)	383 (—/—)	—	—
BBBP	1563 (1209/354)	151 (98/53)	77.350	64.900
CYP1A2	11631 (5570/6061)	147 (49/98)	47.890	33.330
CYP2C19	11563 (5597/5966)	168 (27/141)	48.400	16.070
CYP2C9	11079 (3931/7148)	142 (17/125)	35.480	11.970
CYP2D6	12066 (2343/9723)	145 (39/106)	19.420	26.900
CYP3A4	11374 (5011/6363)	135 (29/106)	44.060	21.480

Table S16. *Paired performance differences between MorphADMET and four benchmark models. Differences are MorphADMET minus the comparator for AP/AUROC and the comparator minus MorphADMET for MAE; positive values favor MorphADMET. Pointwise 95% confidence*

intervals were calculated using 2,000 paired scaffold-bootstrap resamples. Bold interpretations identify intervals excluding zero using unrounded bounds.

Endpoint	Model compared	Metric	Difference	Pointwise 95% CI	CI interpretation
Ames	MapLight	AP	0.048	[-0.030, 0.078]	No clear difference
Ames	MapLight	AUROC	0.052	[-0.055, 0.078]	No clear difference
Ames	D-MPNN	AP	0.076	[-0.038, 0.105]	No clear difference
Ames	D-MPNN	AUROC	0.089	[-0.032, 0.113]	No clear difference
Ames	CheMeleon	AP	-0.005	[-0.076, 0.035]	No clear difference
Ames	CheMeleon	AUROC	0.011	[-0.101, 0.037]	No clear difference
Ames	ADMET-AI	AP	0.084	[-0.033, 0.113]	No clear difference
Ames	ADMET-AI	AUROC	0.087	[-0.038, 0.119]	No clear difference
BBBP	MapLight	AP	0.005	[-0.015, 0.027]	No clear difference
BBBP	MapLight	AUROC	0.019	[-0.006, 0.054]	No clear difference
BBBP	D-MPNN	AP	0.015	[-0.011, 0.043]	No clear difference
BBBP	D-MPNN	AUROC	0.039	[0.006, 0.097]	**Favors MorphADMET**
BBBP	CheMeleon	AP	0.010	[-0.010, 0.027]	No clear difference
BBBP	CheMeleon	AUROC	0.022	[-0.008, 0.054]	No clear difference
BBBP	ADMET-AI	AP	0.034	[0.002, 0.078]	**Favors MorphADMET**
BBBP	ADMET-AI	AUROC	0.045	[0.004, 0.115]	**Favors MorphADMET**
CYP1A2	MapLight	AP	0.042	[-0.002, 0.092]	No clear difference

Endpoint	Model compared	Metric	Difference	Pointwise 95% CI	CI interpretation
CYP1A2	MapLight	AUROC	0.030	[0.008, 0.059]	**Favors MorphADMET**
CYP1A2	D-MPNN	AP	0.071	[0.013, 0.147]	**Favors MorphADMET**
CYP1A2	D-MPNN	AUROC	0.026	[0.004, 0.071]	**Favors MorphADMET**
CYP1A2	CheMeleon	AP	0.028	[-0.005, 0.070]	No clear difference
CYP1A2	CheMeleon	AUROC	0.008	[-0.008, 0.023]	No clear difference
CYP1A2	ADMET-AI	AP	0.013	[-0.022, 0.050]	No clear difference
CYP1A2	ADMET-AI	AUROC	0.007	[-0.010, 0.035]	No clear difference
CYP2C19	MapLight	AP	0.035	[-0.016, 0.099]	No clear difference
CYP2C19	MapLight	AUROC	0.048	[0.001, 0.093]	**Favors MorphADMET**
CYP2C19	D-MPNN	AP	0.155	[0.066, 0.235]	**Favors MorphADMET**
CYP2C19	D-MPNN	AUROC	0.036	[-0.013, 0.101]	No clear difference
CYP2C19	CheMeleon	AP	-0.005	[-0.076, 0.068]	No clear difference
CYP2C19	CheMeleon	AUROC	0.018	[-0.023, 0.060]	No clear difference
CYP2C19	ADMET-AI	AP	0.111	[0.008, 0.176]	**Favors MorphADMET**
CYP2C19	ADMET-AI	AUROC	0.039	[-0.001, 0.078]	No clear difference
CYP2C9	MapLight	AP	-0.041	[-0.109, -0.002]	**Favors other model**

Endpoint	Model compared	Metric	Difference	Pointwise 95% CI	CI interpretation
CYP2C9	MapLight	AUROC	-0.041	[-0.129, -0.014]	**Favors other model**
CYP2C9	D-MPNN	AP	-0.037	[-0.204, 0.060]	No clear difference
CYP2C9	D-MPNN	AUROC	0.013	[-0.107, 0.108]	No clear difference
CYP2C9	CheMeleon	AP	0.017	[-0.082, 0.070]	No clear difference
CYP2C9	CheMeleon	AUROC	0.039	[-0.046, 0.081]	No clear difference
CYP2C9	ADMET-AI	AP	0.004	[-0.056, 0.094]	No clear difference
CYP2C9	ADMET-AI	AUROC	-0.037	[-0.092, 0.059]	No clear difference
CYP2D6	MapLight	AP	0.077	[0.020, 0.129]	**Favors MorphADMET**
CYP2D6	MapLight	AUROC	0.053	[0.012, 0.083]	**Favors MorphADMET**
CYP2D6	D-MPNN	AP	0.002	[-0.047, 0.057]	No clear difference
CYP2D6	D-MPNN	AUROC	0.007	[-0.021, 0.042]	No clear difference
CYP2D6	CheMeleon	AP	-0.022	[-0.057, 0.016]	No clear difference
CYP2D6	CheMeleon	AUROC	0.003	[-0.036, 0.026]	No clear difference
CYP2D6	ADMET-AI	AP	-0.030	[-0.091, 0.039]	No clear difference
CYP2D6	ADMET-AI	AUROC	-0.010	[-0.046, 0.021]	No clear difference
CYP3A4	MapLight	AP	0.025	[-0.025, 0.073]	No clear difference
CYP3A4	MapLight	AUROC	0.023	[-0.008, 0.060]	No clear difference
CYP3A4	D-MPNN	AP	0.053	[-0.025, 0.123]	No clear difference
CYP3A4	D-MPNN	AUROC	0.032	[-0.011, 0.077]	No clear difference

Endpoint	Model compared	Metric	Difference	Pointwise 95% CI	CI interpretation
CYP3A4	CheMeleon	AP	0.086	[0.007, 0.172]	**Favors MorphADMET**
CYP3A4	CheMeleon	AUROC	0.024	[0.002, 0.078]	**Favors MorphADMET**
CYP3A4	ADMET-AI	AP	-0.002	[-0.067, 0.056]	No clear difference
CYP3A4	ADMET-AI	AUROC	-0.004	[-0.047, 0.023]	No clear difference
DICTrank	MapLight	AP	0.012	[-0.008, 0.027]	No clear difference
DICTrank	MapLight	AUROC	0.051	[0.001, 0.087]	**Favors MorphADMET**
DICTrank	D-MPNN	AP	0.017	[0.001, 0.060]	**Favors MorphADMET**
DICTrank	D-MPNN	AUROC	0.050	[0.001, 0.102]	**Favors MorphADMET**
DICTrank	CheMeleon	AP	0.005	[-0.011, 0.028]	No clear difference
DICTrank	CheMeleon	AUROC	-0.005	[-0.036, 0.024]	No clear difference
DICTrank	ADMET-AI	AP	0.035	[0.012, 0.066]	**Favors MorphADMET**
DICTrank	ADMET-AI	AUROC	0.087	[0.020, 0.154]	**Favors MorphADMET**
DILI	MapLight	AP	0.021	[-0.012, 0.063]	No clear difference
DILI	MapLight	AUROC	0.001	[-0.025, 0.061]	No clear difference
DILI	D-MPNN	AP	-0.018	[-0.063, 0.030]	No clear difference
DILI	D-MPNN	AUROC	0.009	[-0.056, 0.091]	No clear difference

Endpoint	Model compared	Metric	Difference	Pointwise 95% CI	CI interpretation
DILI	CheMeleon	AP	0.048	[0.007, 0.089]	**Favors MorphADMET**
DILI	CheMeleon	AUROC	0.023	[-0.009, 0.070]	No clear difference
DILI	ADMET-AI	AP	0.013	[-0.059, 0.065]	No clear difference
DILI	ADMET-AI	AUROC	-0.024	[-0.098, 0.065]	No clear difference
DIRIL	MapLight	AP	0.045	[-0.040, 0.140]	No clear difference
DIRIL	MapLight	AUROC	0.029	[-0.061, 0.180]	No clear difference
DIRIL	D-MPNN	AP	0.194	[0.061, 0.282]	**Favors MorphADMET**
DIRIL	D-MPNN	AUROC	0.196	[0.053, 0.353]	**Favors MorphADMET**
DIRIL	CheMeleon	AP	-0.001	[-0.052, 0.067]	No clear difference
DIRIL	CheMeleon	AUROC	0.002	[-0.062, 0.112]	No clear difference
DIRIL	ADMET-AI	AP	0.099	[0.000, 0.210]	**Favors MorphADMET**
DIRIL	ADMET-AI	AUROC	0.064	[-0.040, 0.263]	No clear difference
hERG	MapLight	AP	0.015	[-0.040, 0.071]	No clear difference
hERG	MapLight	AUROC	-0.009	[-0.040, 0.016]	No clear difference
hERG	D-MPNN	AP	0.028	[-0.024, 0.084]	No clear difference
hERG	D-MPNN	AUROC	0.018	[-0.027, 0.059]	No clear difference
hERG	CheMeleon	AP	0.032	[-0.021, 0.095]	No clear difference

Endpoint	Model compared	Metric	Difference	Pointwise 95% CI	CI interpretation
hERG	CheMeleon	AUROC	0.035	[-0.014, 0.087]	No clear difference
hERG	ADMET-AI	AP	0.023	[-0.037, 0.094]	No clear difference
hERG	ADMET-AI	AUROC	0.009	[-0.031, 0.052]	No clear difference
Solubility	MapLight	MAE	0.062	[0.032, 0.143]	**Favors MorphADMET**
Solubility	D-MPNN	MAE	0.203	[0.059, 0.263]	**Favors MorphADMET**
Solubility	CheMeleon	MAE	0.030	[-0.008, 0.052]	No clear difference
Solubility	ADMET-AI	MAE	0.063	[-0.012, 0.094]	No clear difference

Table S17*. Performance and uncertainty of 84 morphology-based ADMET models. The table reports 161 AP, AUROC, and MAE estimates with pointwise 95% scaffold-bootstrap confidence intervals. Binary resamples containing only one class were excluded. Model populations differ across datasets; sample sizes and class distributions are given in Table S14.*

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
JUMP (U2OS)	hERG	LightGBM	AP	0.469	[0.267, 0.706]
JUMP (U2OS)	hERG	LightGBM	AUROC	0.696	[0.558, 0.851]
JUMP (U2OS)	DICTrank	CatBoost	AP	0.775	[0.678, 0.845]
JUMP (U2OS)	DICTrank	CatBoost	AUROC	0.456	[0.365, 0.612]
JUMP (U2OS)	DILI	Extra Trees	AP	0.624	[0.569, 0.783]
JUMP (U2OS)	DILI	Extra Trees	AUROC	0.460	[0.313, 0.537]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
JUMP (U2OS)	DIRIL	Linear	AP	0.554	[0.364, 0.910]
JUMP (U2OS)	DIRIL	Linear	AUROC	0.457	[0.313, 0.738]
JUMP (U2OS)	Ames	CatBoost	AP	0.294	[0.215, 1.000]
JUMP (U2OS)	Ames	CatBoost	AUROC	0.621	[0.561, 1.000]
JUMP (U2OS)	Solubility	Extra Trees	MAE	1.458	[1.341, 1.763]
JUMP (U2OS)	BBBP	LightGBM	AP	0.693	[0.547, 0.836]
JUMP (U2OS)	BBBP	LightGBM	AUROC	0.422	[0.262, 0.591]
JUMP (U2OS)	CYP1A2	Linear	AP	0.607	[0.419, 0.793]
JUMP (U2OS)	CYP1A2	Linear	AUROC	0.654	[0.531, 0.797]
JUMP (U2OS)	CYP2C19	Linear	AP	0.261	[0.135, 0.513]
JUMP (U2OS)	CYP2C19	Linear	AUROC	0.552	[0.387, 0.712]
JUMP (U2OS)	CYP2C9	MLP	AP	0.228	[0.064, 0.434]
JUMP (U2OS)	CYP2C9	MLP	AUROC	0.429	[0.204, 0.792]
JUMP (U2OS)	CYP2D6	LightGBM	AP	0.289	[0.180, 0.693]
JUMP (U2OS)	CYP2D6	LightGBM	AUROC	0.573	[0.404, 0.724]
JUMP (U2OS)	CYP3A4	Linear	AP	0.219	[0.133, 0.655]
JUMP (U2OS)	CYP3A4	Linear	AUROC	0.453	[0.167, 0.575]
OASIS (PHH, CellProfiler)	hERG	Linear	AP	0.491	[0.303, 0.798]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
OASIS (PHH, CellProfiler)	hERG	Linear	AUROC	0.632	[0.378, 0.831]
OASIS (PHH, CellProfiler)	DICTrank	MLP	AP	0.856	[0.765, 0.960]
OASIS (PHH, CellProfiler)	DICTrank	MLP	AUROC	0.520	[0.365, 0.791]
OASIS (PHH, CellProfiler)	DILI	MLP	AP	0.706	[0.588, 0.889]
OASIS (PHH, CellProfiler)	DILI	MLP	AUROC	0.454	[0.042, 0.571]
OASIS (PHH, CellProfiler)	DIRIL	Linear	AP	0.652	[0.475, 0.866]
OASIS (PHH, CellProfiler)	DIRIL	Linear	AUROC	0.564	[0.364, 0.769]
OASIS (PHH, CellProfiler)	Ames	Linear	AP	0.089	[0.080, 0.100]
OASIS (PHH, CellProfiler)	Ames	Linear	AUROC	0.526	[0.482, 0.570]
OASIS (PHH, CellProfiler)	Solubility	XGBoost	MAE	1.634	[1.483, 2.911]
OASIS (PHH, CellProfiler)	BBBP	Linear	AP	0.635	[0.546, 0.908]
OASIS (PHH, CellProfiler)	BBBP	Linear	AUROC	0.534	[0.264, 0.740]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
OASIS (PHH, CellProfiler)	CYP1A2	CatBoost	AP	0.265	[0.111, 0.444]
OASIS (PHH, CellProfiler)	CYP1A2	CatBoost	AUROC	0.365	[0.000, 0.586]
OASIS (PHH, CellProfiler)	CYP2C19	XGBoost	AP	0.057	[0.037, 0.273]
OASIS (PHH, CellProfiler)	CYP2C19	XGBoost	AUROC	0.020	[0.000, 0.033]
OASIS (PHH, CellProfiler)	CYP2C9	Extra Trees	AP	0.168	[0.100, 0.300]
OASIS (PHH, CellProfiler)	CYP2C9	Extra Trees	AUROC	0.573	[0.172, 0.759]
OASIS (PHH, CellProfiler)	CYP2D6	LightGBM	AP	0.168	[0.079, 0.709]
OASIS (PHH, CellProfiler)	CYP2D6	LightGBM	AUROC	0.425	[0.143, 0.679]
OASIS (PHH, CellProfiler)	CYP3A4	MLP	AP	0.050	[0.021, 0.500]
OASIS (PHH, CellProfiler)	CYP3A4	MLP	AUROC	0.240	[0.124, 0.667]
OASIS (PHH, DINOv2)	hERG	XGBoost	AP	0.615	[0.363, 0.870]
OASIS (PHH, DINOv2)	hERG	XGBoost	AUROC	0.739	[0.564, 0.913]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
OASIS (PHH, DINOv2)	DICTrank	Linear	AP	0.816	[0.697, 0.936]
OASIS (PHH, DINOv2)	DICTrank	Linear	AUROC	0.400	[0.268, 0.639]
OASIS (PHH, DINOv2)	DILI	MLP	AP	0.623	[0.534, 0.864]
OASIS (PHH, DINOv2)	DILI	MLP	AUROC	0.369	[0.035, 0.451]
OASIS (PHH, DINOv2)	DIRIL	MLP	AP	0.652	[0.440, 0.876]
OASIS (PHH, DINOv2)	DIRIL	MLP	AUROC	0.602	[0.361, 0.782]
OASIS (PHH, DINOv2)	Ames	MLP	AP	0.093	[0.086, 0.104]
OASIS (PHH, DINOv2)	Ames	MLP	AUROC	0.578	[0.541, 0.615]
OASIS (PHH, DINOv2)	Solubility	CatBoost	MAE	1.548	[1.395, 2.740]
OASIS (PHH, DINOv2)	BBBP	Extra Trees	AP	0.523	[0.389, 0.916]
OASIS (PHH, DINOv2)	BBBP	Extra Trees	AUROC	0.476	[0.319, 0.822]
OASIS (PHH, DINOv2)	CYP1A2	Extra Trees	AP	0.474	[0.421, 1.000]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
OASIS (PHH, DINOv2)	CYP1A2	Extra Trees	AUROC	0.635	[0.545, 1.000]
OASIS (PHH, DINOv2)	CYP2C19	LightGBM	AP	0.068	[0.039, 0.375]
OASIS (PHH, DINOv2)	CYP2C19	LightGBM	AUROC	0.200	[0.039, 0.556]
OASIS (PHH, DINOv2)	CYP2C9	MLP	AP	0.100	[0.069, 0.273]
OASIS (PHH, DINOv2)	CYP2C9	MLP	AUROC	0.293	[0.091, 0.455]
OASIS (PHH, DINOv2)	CYP2D6	Linear	AP	0.373	[0.090, 1.000]
OASIS (PHH, DINOv2)	CYP2D6	Linear	AUROC	0.642	[0.333, 1.000]
OASIS (PHH, DINOv2)	CYP3A4	Linear	AP	0.200	[0.091, 1.000]
OASIS (PHH, DINOv2)	CYP3A4	Linear	AUROC	0.840	[0.667, 1.000]
LINCS (A549)	hERG	Extra Trees	AP	0.601	[0.333, 0.814]
LINCS (A549)	hERG	Extra Trees	AUROC	0.786	[0.601, 0.916]
LINCS (A549)	DICTrank	Extra Trees	AP	0.944	[0.827, 1.000]
LINCS (A549)	DICTrank	Extra Trees	AUROC	0.721	[0.309, 1.000]
LINCS (A549)	DILI	Linear	AP	0.731	[0.656, 0.965]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
LINCS (A549)	DILI	Linear	AUROC	0.514	[0.386, 0.815]
LINCS (A549)	DIRIL	CatBoost	AP	0.544	[0.364, 0.796]
LINCS (A549)	DIRIL	CatBoost	AUROC	0.512	[0.203, 0.671]
LINCS (A549)	Ames	Extra Trees	AP	0.621	[0.271, 1.000]
LINCS (A549)	Ames	Extra Trees	AUROC	0.728	[0.529, 1.000]
LINCS (A549)	Solubility	Extra Trees	MAE	1.389	[1.231, 1.801]
LINCS (A549)	BBBP	Linear	AP	0.729	[0.607, 0.962]
LINCS (A549)	BBBP	Linear	AUROC	0.523	[0.214, 0.640]
LINCS (A549)	CYP1A2	MLP	AP	0.205	[0.118, 0.421]
LINCS (A549)	CYP1A2	MLP	AUROC	0.328	[0.080, 0.487]
LINCS (A549)	CYP2C19	LightGBM	AP	0.176	[0.050, 0.627]
LINCS (A549)	CYP2C19	LightGBM	AUROC	0.684	[0.395, 0.870]
LINCS (A549)	CYP2C9	LightGBM	AP	0.117	[0.061, 0.330]
LINCS (A549)	CYP2C9	LightGBM	AUROC	0.253	[0.074, 0.556]
LINCS (A549)	CYP2D6	XGBoost	AP	0.212	[0.111, 0.635]
LINCS (A549)	CYP2D6	XGBoost	AUROC	0.394	[0.143, 0.672]
LINCS (A549)	CYP3A4	Extra Trees	AP	0.326	[0.088, 0.968]
LINCS (A549)	CYP3A4	Extra Trees	AUROC	0.747	[0.529, 0.979]
EU-OS (HepG2)	hERG	Linear	AP	0.418	[0.184, 0.661]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
EU-OS (HepG2)	hERG	Linear	AUROC	0.687	[0.501, 0.839]
EU-OS (HepG2)	DICTrank	Linear	AP	0.860	[0.689, 0.975]
EU-OS (HepG2)	DICTrank	Linear	AUROC	0.378	[0.112, 0.761]
EU-OS (HepG2)	DILI	CatBoost	AP	0.738	[0.624, 0.987]
EU-OS (HepG2)	DILI	CatBoost	AUROC	0.649	[0.541, 0.942]
EU-OS (HepG2)	DIRIL	Extra Trees	AP	0.687	[0.529, 0.906]
EU-OS (HepG2)	DIRIL	Extra Trees	AUROC	0.668	[0.427, 0.820]
EU-OS (HepG2)	Ames	Linear	AP	0.455	[0.347, 1.000]
EU-OS (HepG2)	Ames	Linear	AUROC	0.527	[0.397, 1.000]
EU-OS (HepG2)	Solubility	CatBoost	MAE	1.318	[0.587, 1.771]
EU-OS (HepG2)	BBBP	MLP	AP	0.741	[0.640, 0.976]
EU-OS (HepG2)	BBBP	MLP	AUROC	0.685	[0.555, 0.932]
EU-OS (HepG2)	CYP1A2	MLP	AP	0.661	[0.533, 0.980]
EU-OS (HepG2)	CYP1A2	MLP	AUROC	0.654	[0.485, 0.950]
EU-OS (HepG2)	CYP2C19	Linear	AP	0.314	[0.159, 0.715]
EU-OS (HepG2)	CYP2C19	Linear	AUROC	0.563	[0.361, 0.729]
EU-OS (HepG2)	CYP2C9	Linear	AP	0.341	[0.169, 0.527]
EU-OS (HepG2)	CYP2C9	Linear	AUROC	0.566	[0.363, 0.741]
EU-OS (HepG2)	CYP2D6	CatBoost	AP	0.571	[0.391, 0.852]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
EU-OS (HepG2)	CYP2D6	CatBoost	AUROC	0.668	[0.481, 0.839]
EU-OS (HepG2)	CYP3A4	Extra Trees	AP	0.501	[0.336, 0.845]
EU-OS (HepG2)	CYP3A4	Extra Trees	AUROC	0.609	[0.385, 0.821]
RxRx3-core (HUVEC)	hERG	Extra Trees	AP	0.666	[0.447, 0.860]
RxRx3-core (HUVEC)	hERG	Extra Trees	AUROC	0.760	[0.581, 0.893]
RxRx3-core (HUVEC)	DICTrank	Linear	AP	0.898	[0.780, 0.973]
RxRx3-core (HUVEC)	DICTrank	Linear	AUROC	0.569	[0.362, 0.742]
RxRx3-core (HUVEC)	DILI	Linear	AP	0.801	[0.731, 0.986]
RxRx3-core (HUVEC)	DILI	Linear	AUROC	0.631	[0.539, 0.925]
RxRx3-core (HUVEC)	DIRIL	MLP	AP	0.576	[0.407, 0.806]
RxRx3-core (HUVEC)	DIRIL	MLP	AUROC	0.524	[0.331, 0.701]
RxRx3-core (HUVEC)	Ames	CatBoost	AP	0.149	[0.106, 0.830]
RxRx3-core (HUVEC)	Ames	CatBoost	AUROC	0.448	[0.219, 0.857]
RxRx3-core (HUVEC)	Solubility	Linear	MAE	1.189	[0.787, 1.802]
RxRx3-core (HUVEC)	BBBP	Extra Trees	AP	0.677	[0.547, 0.900]
RxRx3-core (HUVEC)	BBBP	Extra Trees	AUROC	0.454	[0.291, 0.681]
RxRx3-core (HUVEC)	CYP1A2	LightGBM	AP	0.226	[0.152, 0.634]
RxRx3-core (HUVEC)	CYP1A2	LightGBM	AUROC	0.537	[0.357, 0.786]
RxRx3-core (HUVEC)	CYP2C19	MLP	AP	0.163	[0.075, 0.651]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
RxRx3-core (HUVEC)	CYP2C19	MLP	AUROC	0.659	[0.222, 0.917]
RxRx3-core (HUVEC)	CYP2C9	Linear	AP	0.192	[0.152, 1.000]
RxRx3-core (HUVEC)	CYP2C9	Linear	AUROC	0.607	[0.455, 1.000]
RxRx3-core (HUVEC)	CYP2D6	LightGBM	AP	0.579	[0.435, 0.891]
RxRx3-core (HUVEC)	CYP2D6	LightGBM	AUROC	0.644	[0.375, 0.833]
RxRx3-core (HUVEC)	CYP3A4	LightGBM	AP	0.134	[0.056, 0.494]
RxRx3-core (HUVEC)	CYP3A4	LightGBM	AUROC	0.466	[0.143, 0.693]
RxRx19a (HRCE)	hERG	Extra Trees	AP	0.599	[0.388, 0.825]
RxRx19a (HRCE)	hERG	Extra Trees	AUROC	0.762	[0.617, 0.874]
RxRx19a (HRCE)	DICTrank	Linear	AP	0.907	[0.809, 0.972]
RxRx19a (HRCE)	DICTrank	Linear	AUROC	0.552	[0.375, 0.733]
RxRx19a (HRCE)	DILI	Linear	AP	0.739	[0.661, 0.978]
RxRx19a (HRCE)	DILI	Linear	AUROC	0.576	[0.470, 0.875]
RxRx19a (HRCE)	DIRIL	Extra Trees	AP	0.593	[0.438, 0.831]
RxRx19a (HRCE)	DIRIL	Extra Trees	AUROC	0.503	[0.361, 0.721]
RxRx19a (HRCE)	Ames	Linear	AP	0.205	[0.131, 0.797]
RxRx19a (HRCE)	Ames	Linear	AUROC	0.434	[0.000, 0.686]
RxRx19a (HRCE)	Solubility	CatBoost	MAE	1.123	[0.938, 1.710]
RxRx19a (HRCE)	BBBP	Linear	AP	0.731	[0.580, 0.878]

Dataset / representation	Endpoint	Method	Metric	Score	Pointwise 95% CI
RxRx19a (HRCE)	BBBP	Linear	AUROC	0.582	[0.292, 0.700]
RxRx19a (HRCE)	CYP1A2	Linear	AP	0.276	[0.200, 0.640]
RxRx19a (HRCE)	CYP1A2	Linear	AUROC	0.352	[0.242, 0.800]
RxRx19a (HRCE)	CYP2C19	Extra Trees	AP	0.256	[0.059, 0.368]
RxRx19a (HRCE)	CYP2C19	Extra Trees	AUROC	0.720	[0.263, 0.915]
RxRx19a (HRCE)	CYP2C9	Linear	AP	0.059	[0.044, 0.231]
RxRx19a (HRCE)	CYP2C9	Linear	AUROC	0.314	[0.247, 0.667]
RxRx19a (HRCE)	CYP2D6	MLP	AP	0.491	[0.334, 0.813]
RxRx19a (HRCE)	CYP2D6	MLP	AUROC	0.592	[0.410, 0.813]
RxRx19a (HRCE)	CYP3A4	Extra Trees	AP	0.080	[0.034, 0.470]
RxRx19a (HRCE)	CYP3A4	Extra Trees	AUROC	0.348	[0.233, 0.734]

Table S18. *Hyperparameter search settings for morphology-based ADMET models. Six model families were evaluated using feature selection or PCA with 16, 32, or 64 features/components, yielding 78 configurations per endpoint–representation pair. Test scores were excluded from configuration selection.*

Model	Hyperparameters varied	Hyperparameters held constant
Linear models	Logistic inverse regularization (C): 0.010, 0.100, 1 Ridge regularization (alpha): 1, 10, 100	Logistic solver: liblinear Maximum logistic iterations: 3,000
CatBoost	Tree depth: 3 or 5	Iterations: 200 Learning rate: 0.040 L2 regularization: 10 Threads: 1

Model	Hyperparameters varied	Hyperparameters held constant
LightGBM	Leaves per tree: 7 or 15	Trees: 200 Learning rate: 0.040 Minimum samples per leaf: 10 L2 regularization: 10 Deterministic mode: enabled Column-wise fitting: enabled
XGBoost	Tree depth: 3 or 5	Trees: 200 Learning rate: 0.040 L2 regularization: 10 Tree method: histogram, CPU Objective: binary logistic for classification; squared error for regression
Extra Trees	Minimum samples per leaf: 3 or 8	Trees: 300 Features per split: square root of feature count
Neural network (MLP)	Configuration 1: one hidden layer of 64 units; alpha = 0.010 Configuration 2: two hidden layers of 128 and 64 units; alpha = 0.100	Maximum iterations: 200 Batch size: 64 Initial learning rate: 0.001 Early stopping: disabled

Table S19. *Software versions used for morphology-based ADMET model training on Linux.*

Software	Version
Python	3.11.15
NumPy	2.4.6
pandas	3.0.5
SciPy	1.17.1
scikit-learn	1.9.1
CatBoost	1.2.10

Software	Version
LightGBM	4.7.0
XGBoost	3.2.0
joblib	1.6.0
Matplotlib	3.11.2

Table S20. *hERG label agreement for morphology-selected and random references across six imaging contexts. Agreement is class-balanced, with ten references per query. The random baseline is the exact expectation from the eligible reference pool. Gains are reported in percentage points with pointwise 95% confidence intervals from 2,000 query-scaffold bootstrap resamples.*

Dataset / representation	Queries n (+/-)	Morphology (%)	Random (%)	Gain (pp)	Pointwise 95% CI (pp)
JUMP (U2OS)	589 (120/469)	**52.100**	50.000	2.100	[0.630, 3.620]
OASIS (PHH)	178 (63/115)	**52.550**	49.690	2.860	[0.800, 4.960]
LINCS (A549)	372 (102/270)	**58.020**	49.770	8.250	[5.760, 10.690]
EU-OS (HepG2)	461 (93/368)	**53.870**	49.910	3.960	[1.880, 5.940]
RxRx3-core (HUVEC)	303 (91/212)	**55.290**	49.720	5.570	[3.070, 8.340]
RxRx19a (HRCE)	245 (92/153)	**57.580**	49.640	7.940	[5.120, 10.840]

Dataset characteristics and overlap

Table S21. *Compound counts during structure curation. The table reports input records, unique compounds after cleaning, and development/test populations. Cleaning includes structure validation, duplicate consolidation, and label-conflict handling. Compounds sharing a chemical parent or scaffold with test compounds were excluded from development.*

Endpoint	Input records	After cleaning	Development compounds	Test compounds
hERG	13445	12764	12268	240

Endpoint	Input records	After cleaning	Development compounds	Test compounds
DICTrank	1020	996	719	196
DILI	1111	1096	815	211
DIRIL	317	317	227	73
Ames	7278	7176	4659	174
Solubility	9982	9306	4161	383
BBBP	2030	1940	1563	151
CYP1A2	12579	12519	11631	147
CYP2C19	12665	12533	11563	168
CYP2C9	12092	11986	11079	142
CYP2D6	13130	13002	12066	145
CYP3A4	12328	12260	11374	135

Table S22. Endpoint panel evaluated in this study.

Endpoint	Task	Positive label or target	Dataset/source	Number of unique compounds	Matched to JUMP	Matched to OASIS	Matched to LINCS	Matched to EU-OS	Matched to RxRx3-core	Matched to RxRx19a
hERG	Binary classification	hERG inhibition	hERG Karim dataset, via TDC (Huang et al., 2022) [15]; Karim et al. (2021) [16].	12,764	591	178	368	459	299	240
DICTrank	Binary classification	Cardiotoxicity category under the study mapping	FDA DICTrank; Qu et al. (2023) [17].	996	710	247	275	188	255	270

Endpoint	Task	Positive label or target	Dataset/source	Number of unique compounds	Matched to JUMP	Matched to OASIS	Matched to LINCS	Matched to EU-OS	Matched to RxRx3-core	Matched to RxRx19a
DILI	Binary classification	Drug-induced liver injury annotation	DILI Goldstandard (DILIst/DILIrank-derived); Seal et al. (2024) [18].	1,096	810	324	322	204	283	301
DIRIL	Binary classification	Drug-induced renal injury annotation	FDA Drug-Induced Renal Injury List (DIRIL); Connor et al. (2024) [19].	317	176	179	190	133	191	205
Ames	Binary classification	Ames mutagenicity	Ames mutagenicity dataset, via TDC (Huang et al., 2022) [15]; Xu et al. (2012) [20].	7,176	617	202	149	164	180	196
BBB	Binary classification	Blood–brain barrier penetration	BBBP dataset, via TDC (Huang et al., 2022) [15]; Martins et al. (2012) [21].	1,940	323	218	262	185	264	288
CYP1A2	Binary classification	Inhibition of the specified CYP isoform	CYP1A2 inhibition dataset, via TDC (Huang et al., 2022) [15]; Veith et al. (2009) [22].	12,519	335	112	176	175	211	228
CYP2C19	Binary classification	Inhibition of the specified CYP isoform	CYP2C19 inhibition dataset, via TDC (Huang et al., 2022) [15]; Veith et al. (2009) [22].	12,533	354	126	199	200	233	253

Endpoint	Task	Positive label or target	Dataset/source	Number of unique compounds	Matched to JUMP	Matched to OASIS	Matched to LINCS	Matched to EU-OS	Matched to RxRx3-core	Matched to RxRx19a
CYP2C9	Binary classification	Inhibition of the specified CYP isoform	CYP2C9 inhibition dataset, via TDC (Huang et al., 2022) [15]; Veith et al. (2009) [22].	11,986	338	124	180	189	232	253
CYP2D6	Binary classification	Inhibition of the specified CYP isoform	CYP2D6 inhibition dataset, via TDC (Huang et al., 2022) [15]; Veith et al. (2009) [22].	13,002	354	130	185	197	236	258
CYP3A4	Binary classification	Inhibition of the specified CYP isoform	CYP3A4 inhibition dataset, via TDC (Huang et al., 2022) [15]; Veith et al. (2009) [22].	12,260	336	118	179	190	221	241
Aqueous solubility	Regression	Measured log10 solubility in mol/L	AqSolDB, via TDC (Huang et al., 2022) [15]; Sorkun et al. (2019) [23].	9,306	831	321	218	247	245	266

Table S23. *Compound counts for the six imaging datasets.*

Dataset / representation	Prepared input n
JUMP (U2OS)	6090
OASIS (PHH, CellProfiler)	939
LINCS (A549)	1504

Dataset / representation	Prepared input n
EU-OS (HepG2)	2455
RxRx3-core (HUVEC)	1650
RxRx19a (HRCE)	1650

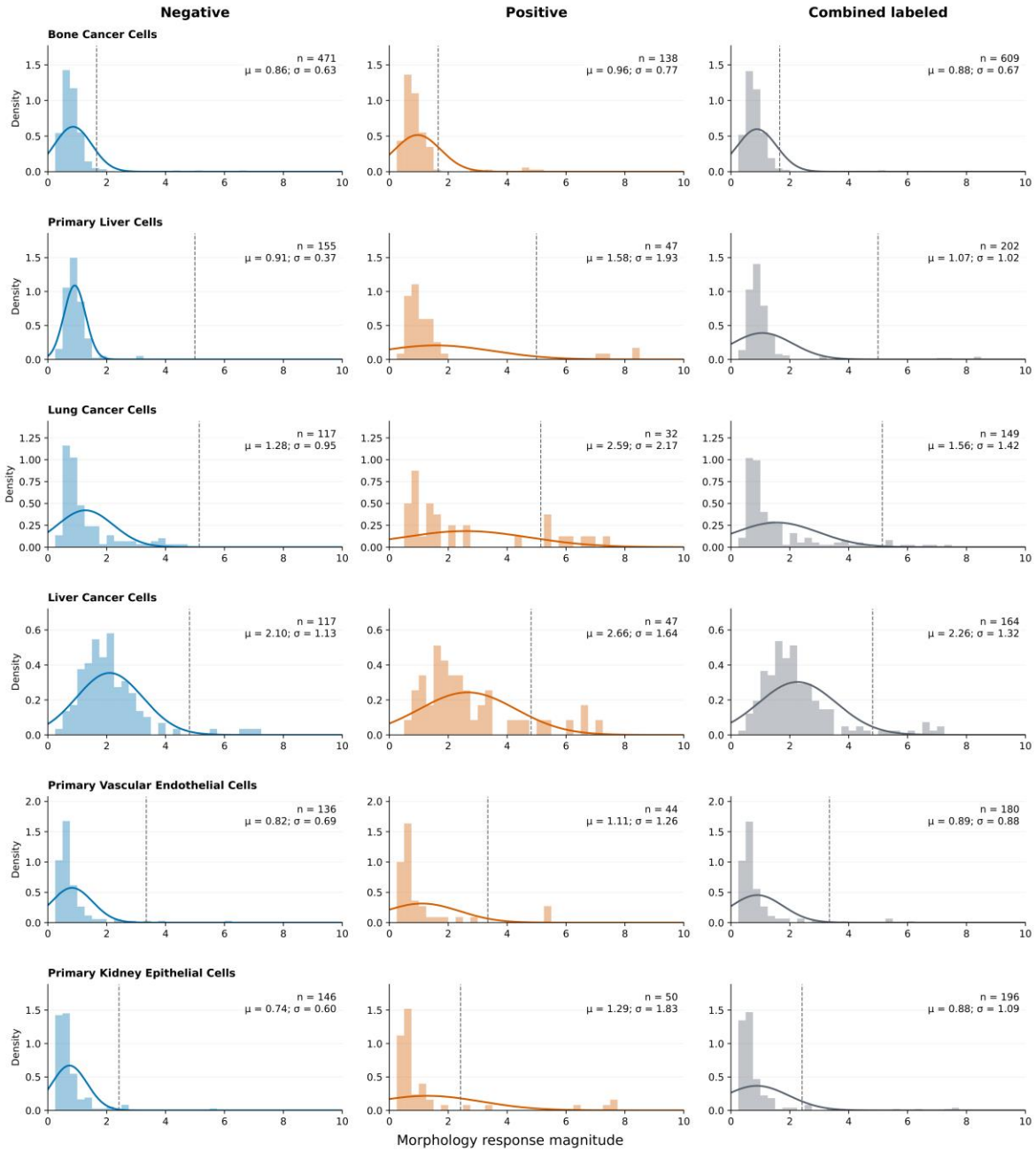


Fig S1a. Distributions of morphology response magnitude stratified by Ames response status across six cell types.

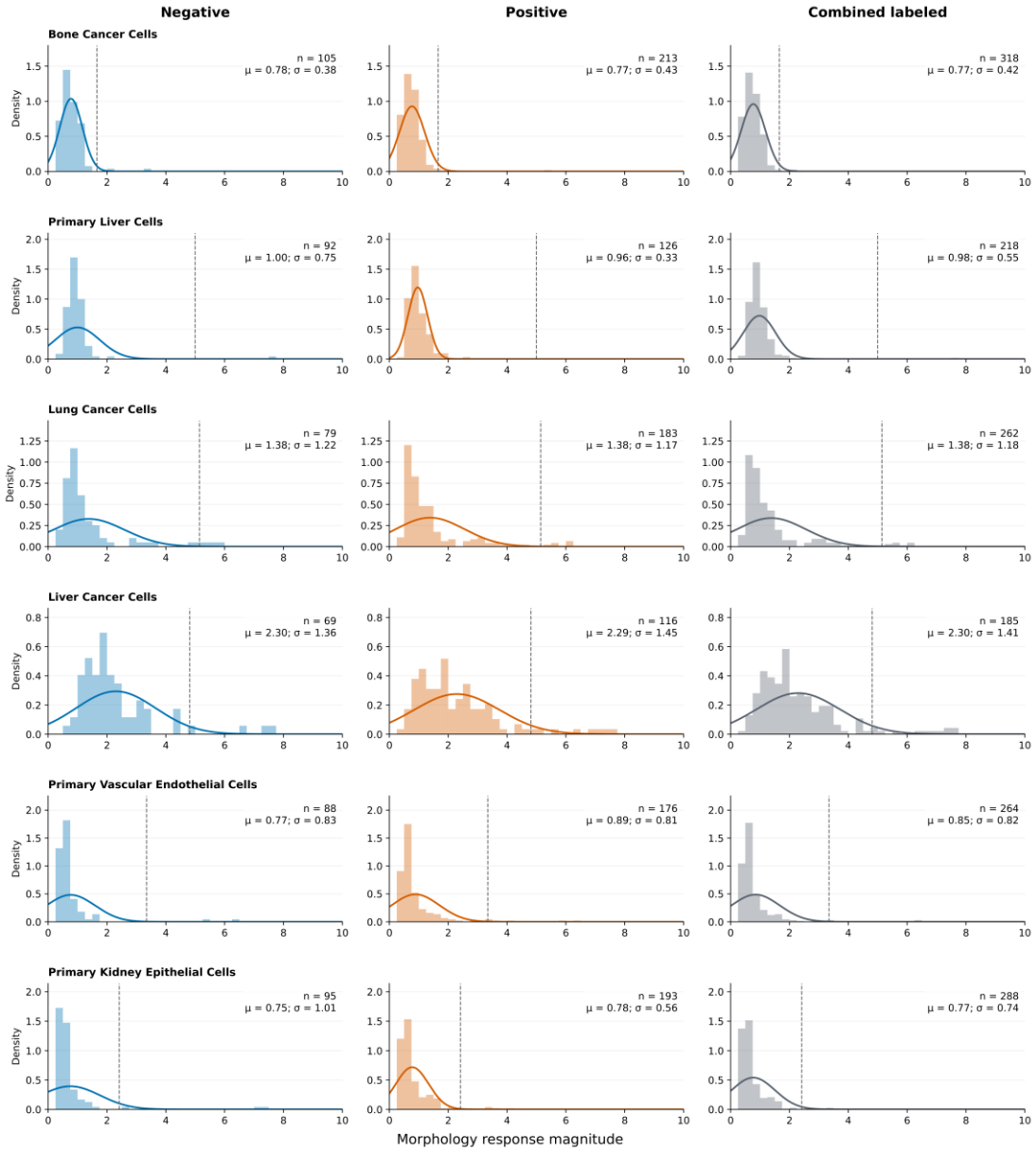


Fig S1b. Distributions of morphology response magnitude stratified by BBB penetration response status across six cell types.

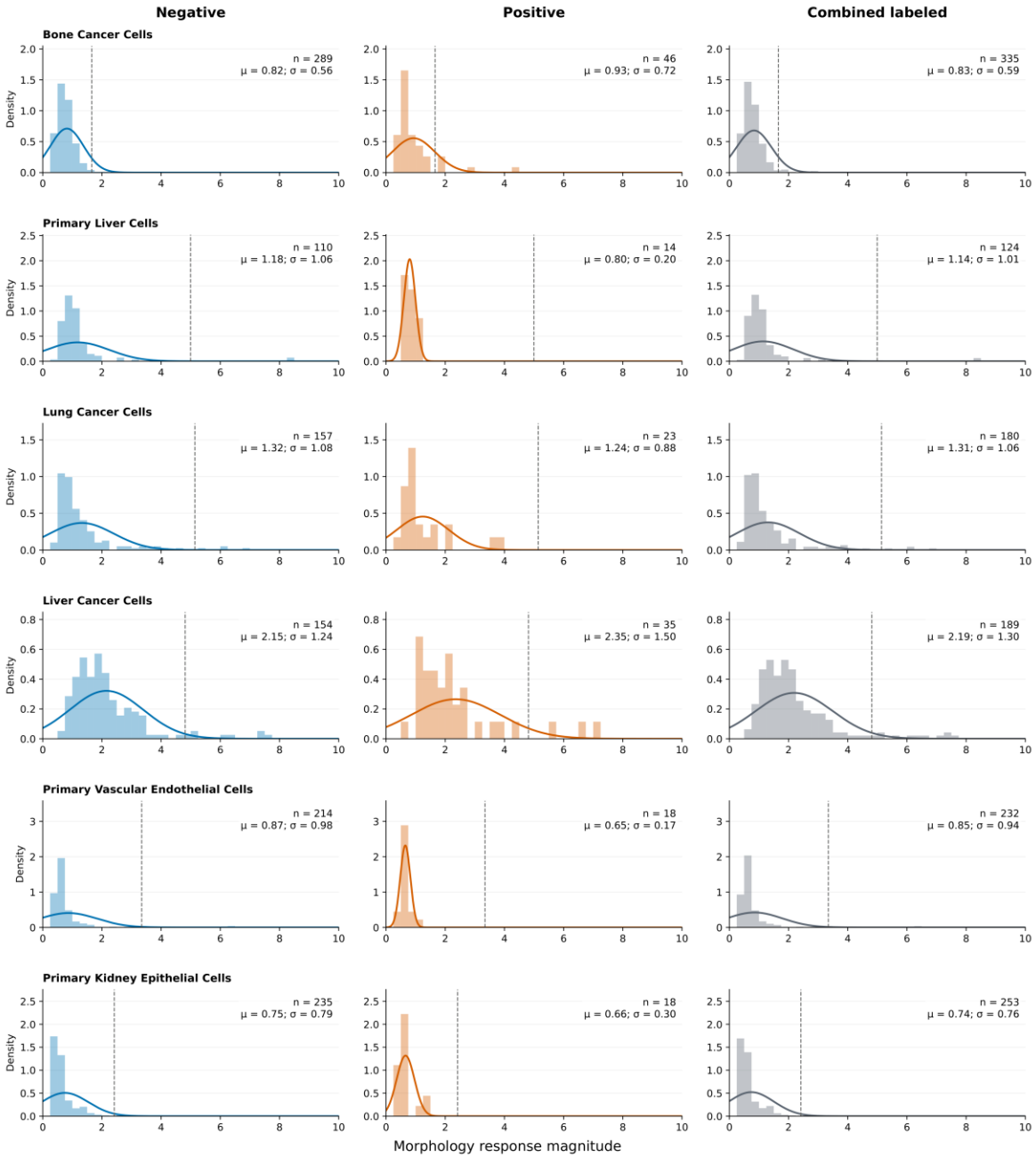


Fig S1c. Distributions of morphology response magnitude stratified by CYP1A2 response status across six cell types.

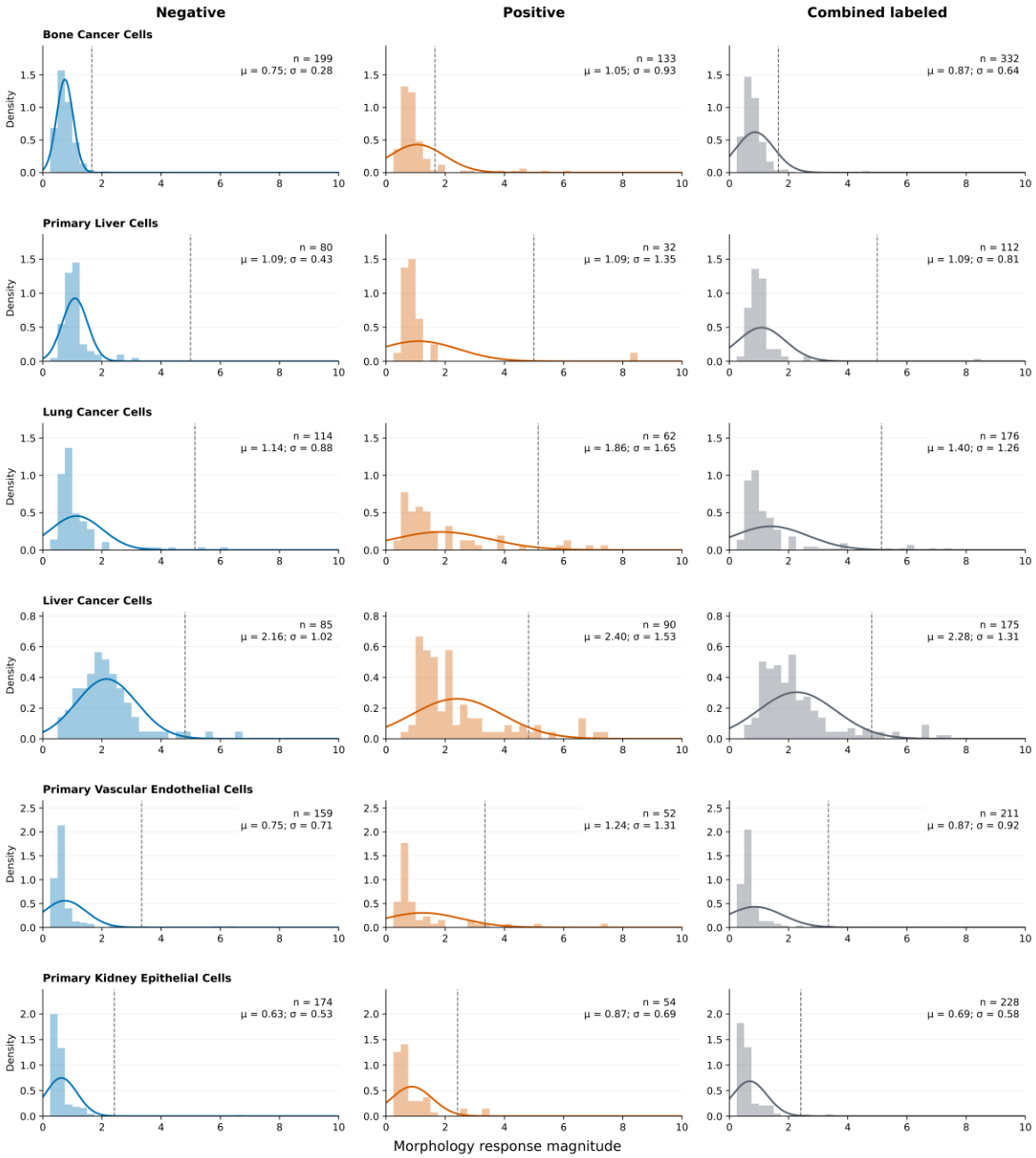


Fig S1d. Distributions of morphology response magnitude stratified by CYP2C9 response status across six cell types.

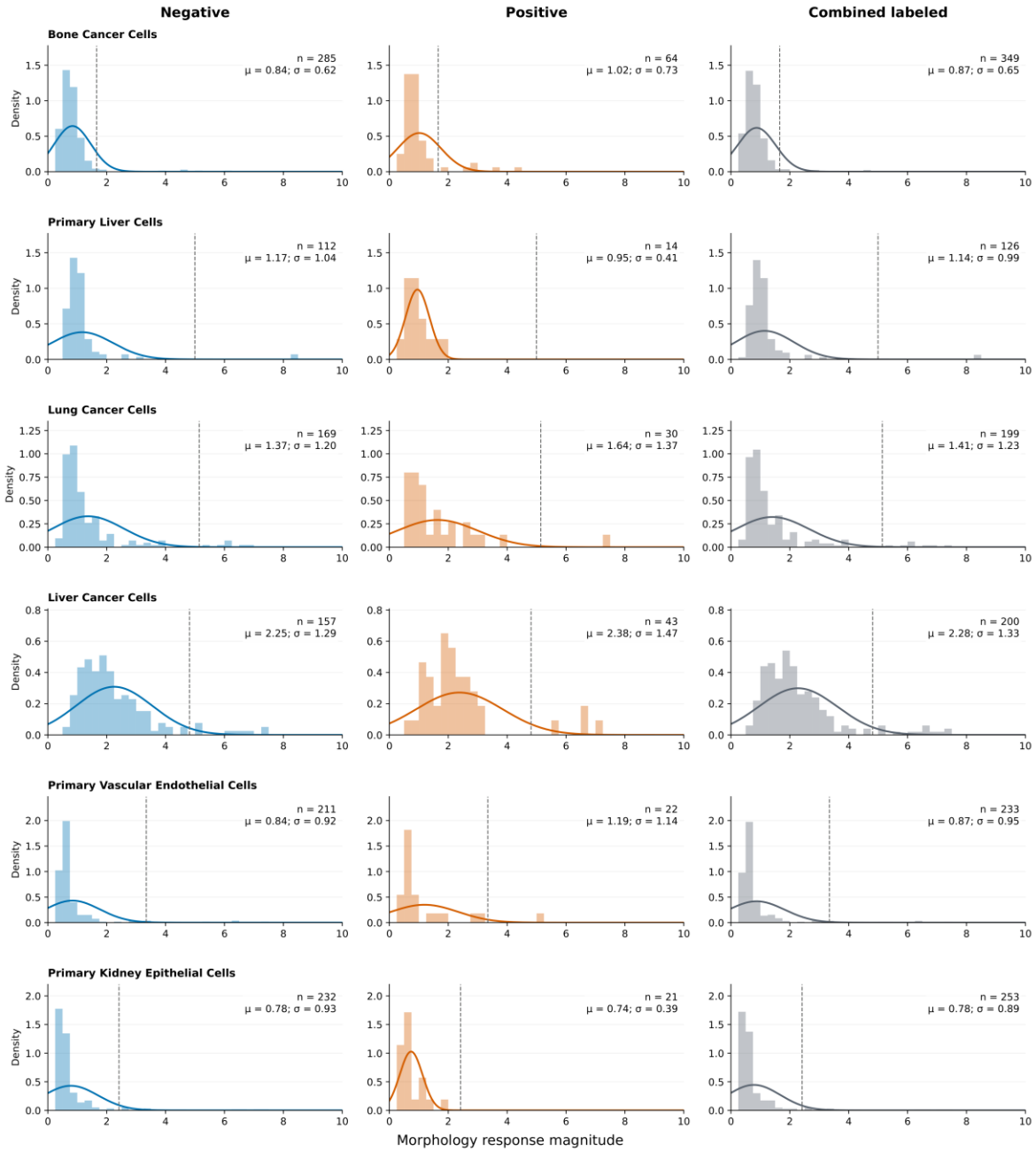


Fig S1e. Distributions of morphology response magnitude stratified by CYP2C19 response status across six cell types.

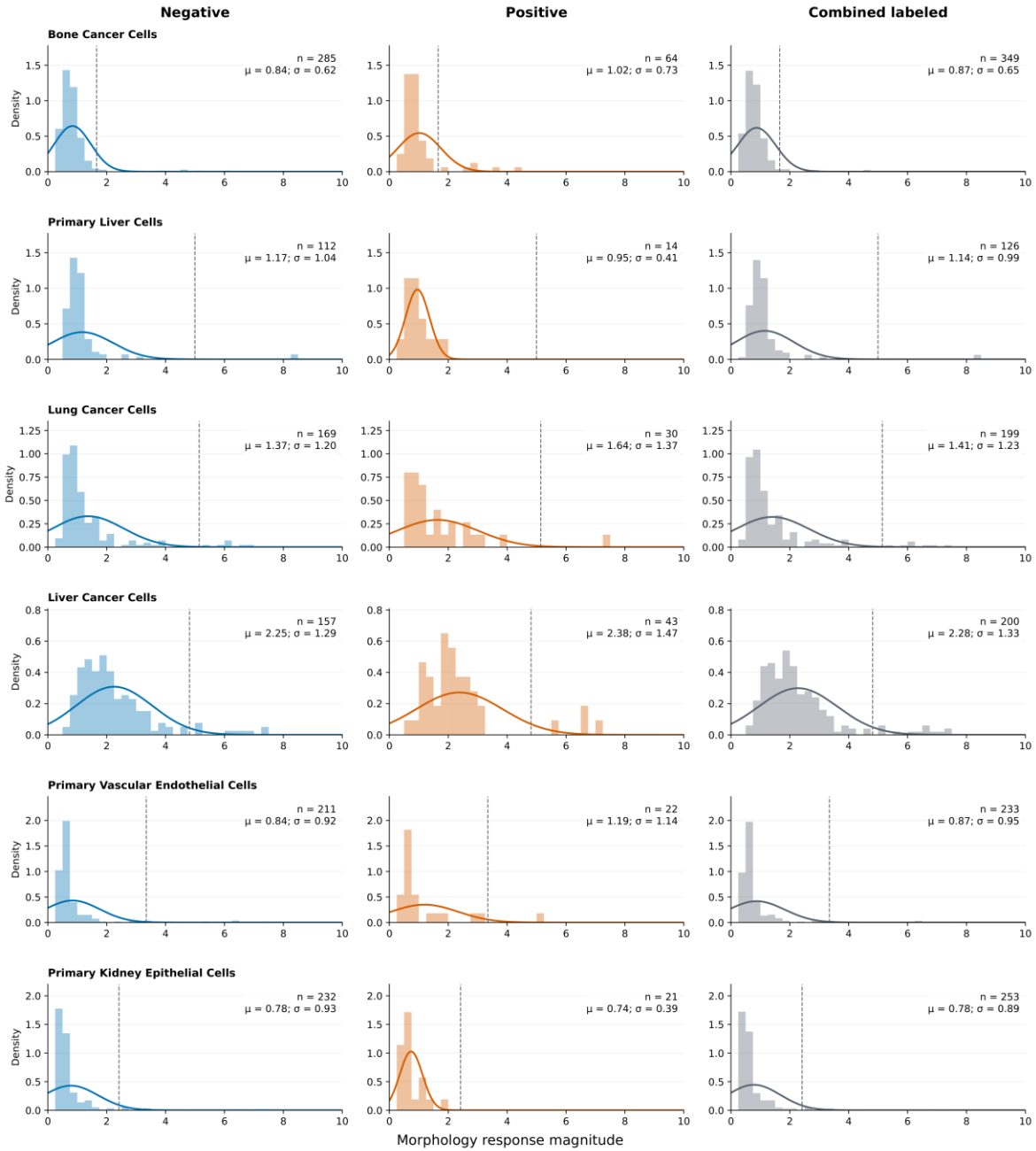


Fig S1f. Distributions of morphology response magnitude stratified by CYP2D6 response status across six cell types.

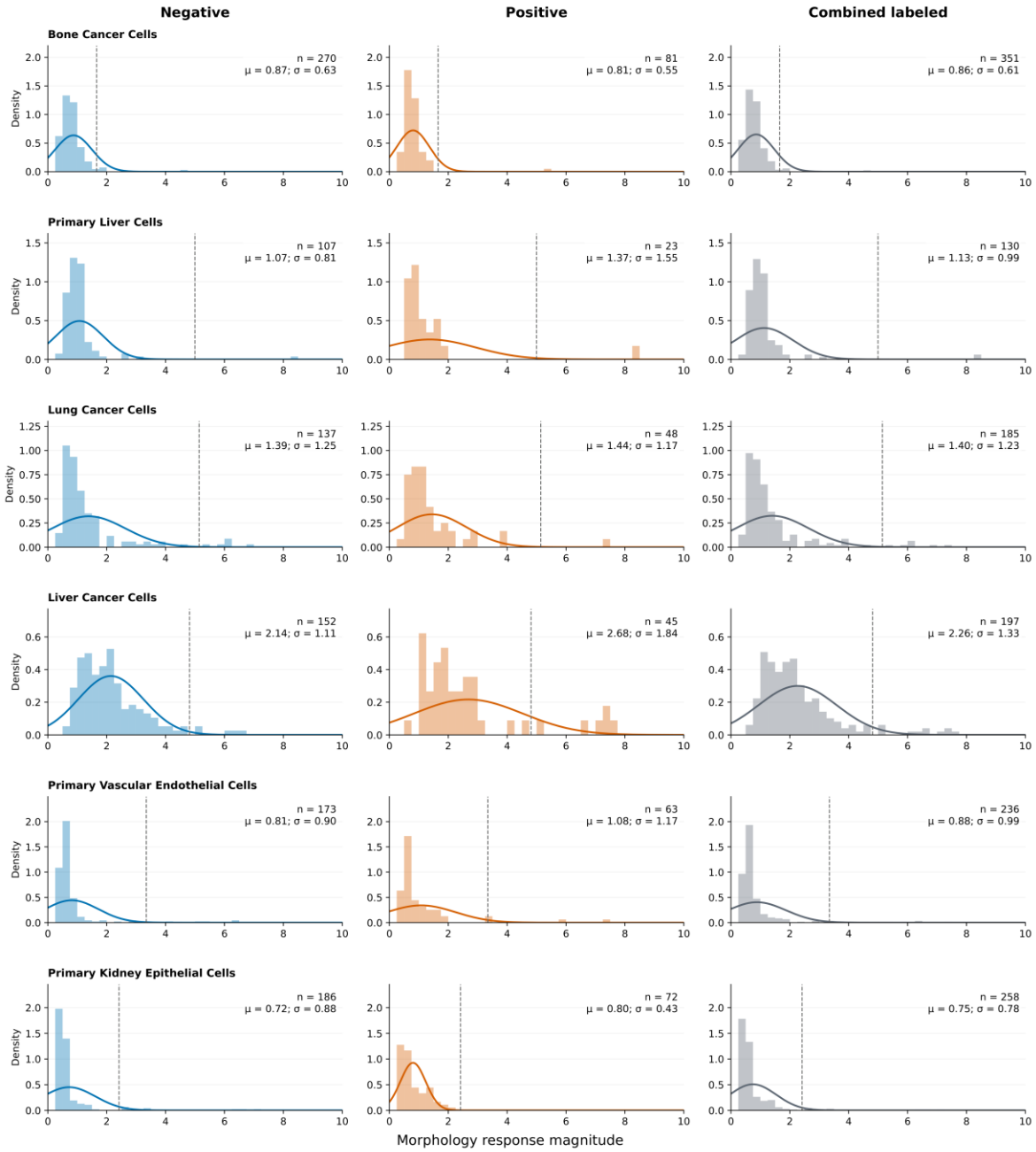


Fig S1g. Distributions of morphology response magnitude stratified by CYP3A4 response status across six cell types.

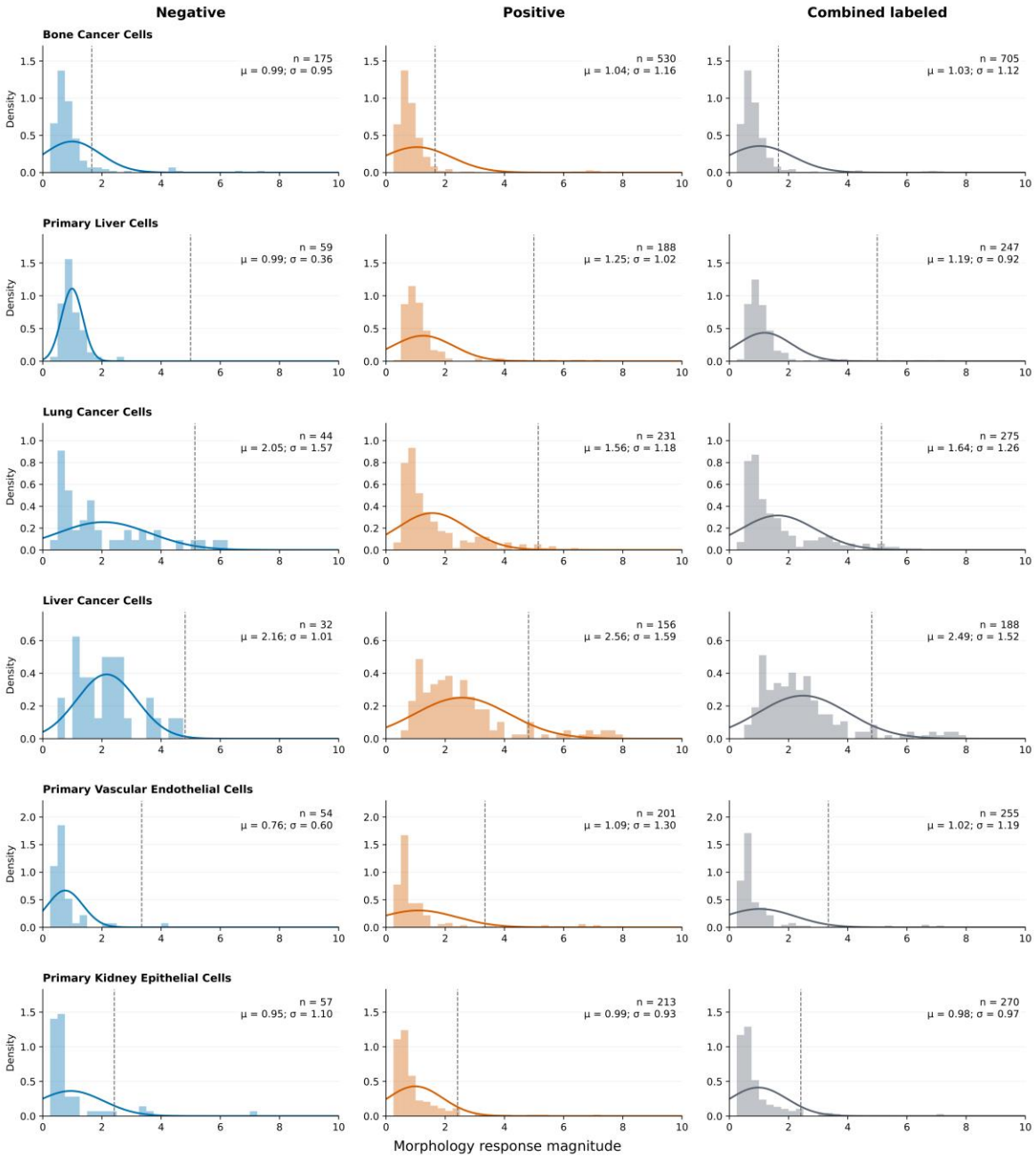


Fig S1h. Distributions of morphology response magnitude stratified by DICTrank response status across six cell types.

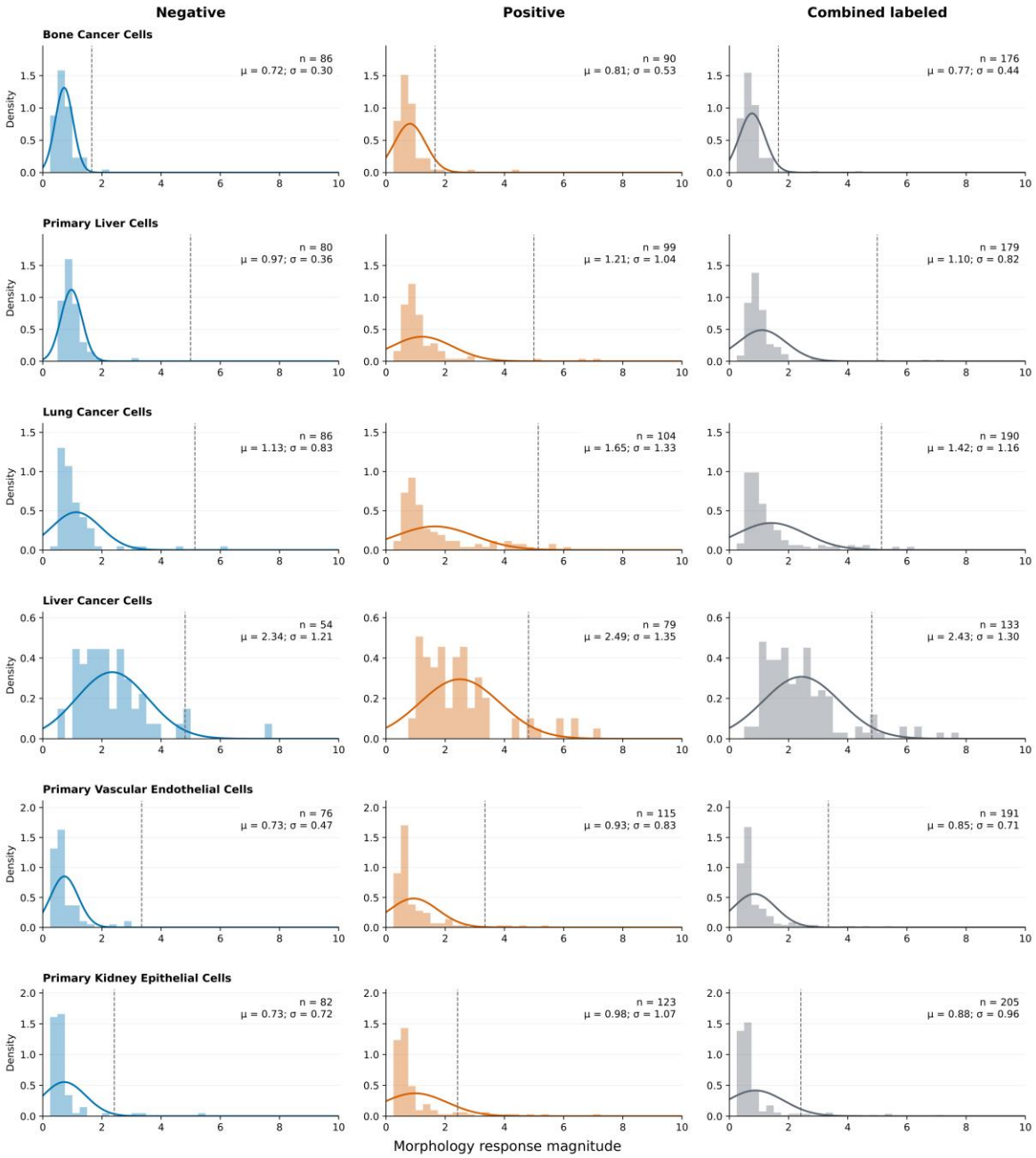


Fig S1i. Distributions of morphology response magnitude stratified by DILI response status across six cell types.

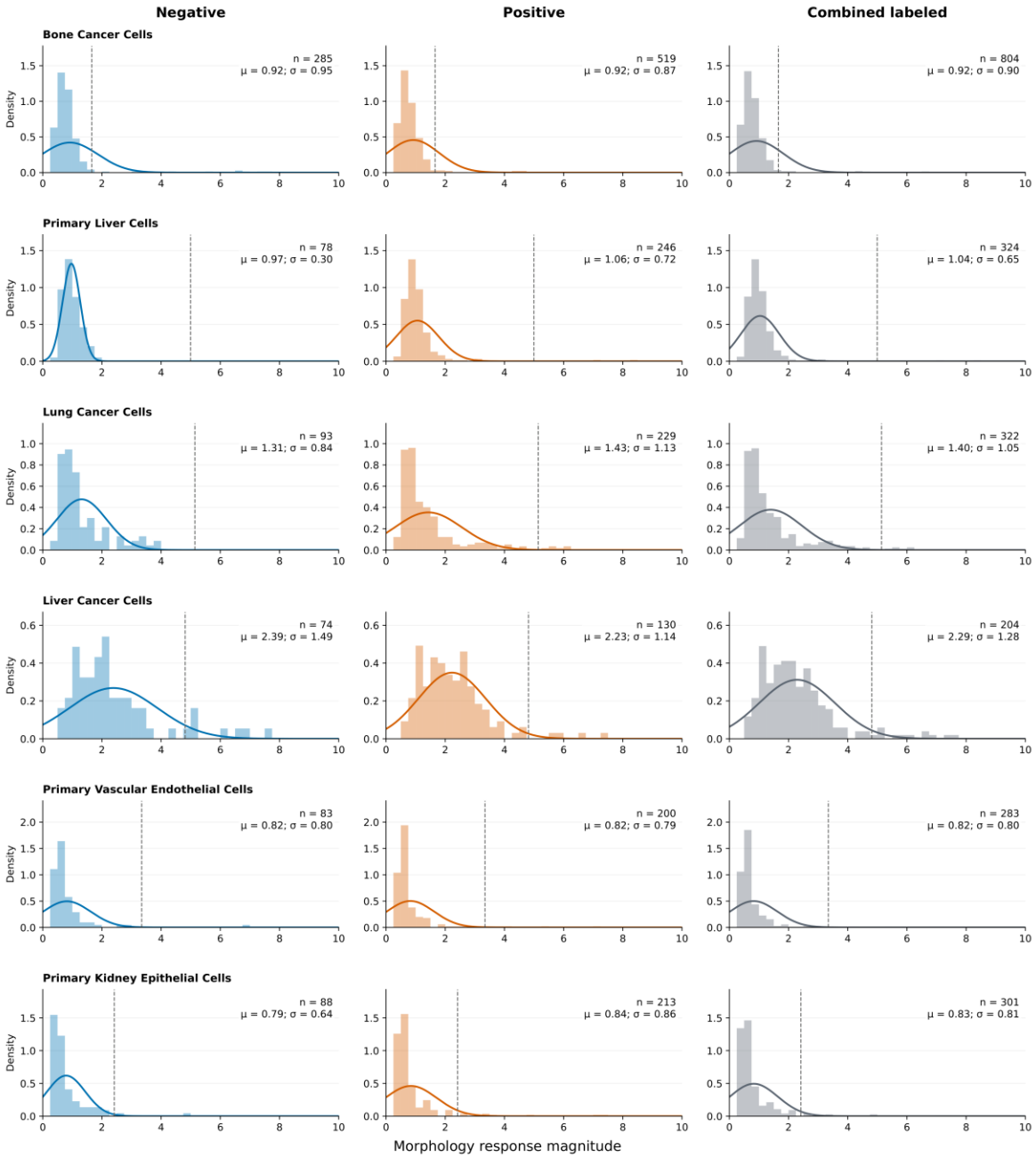


Fig S1j. Distributions of morphology response magnitude stratified by DIRIL response status across six cell types.

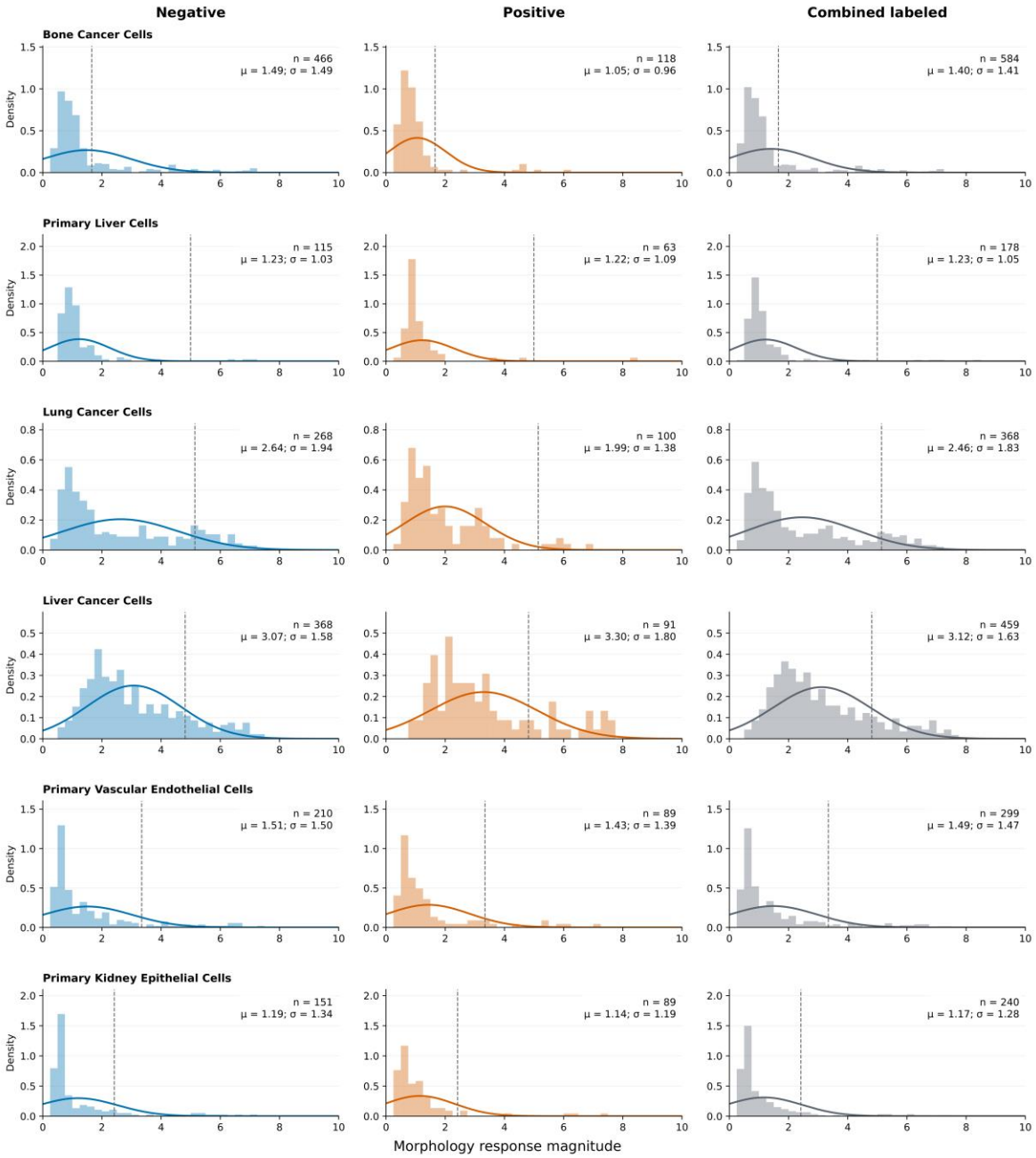


Fig S1k. Distributions of morphology response magnitude stratified by hERG response status across six cell types.

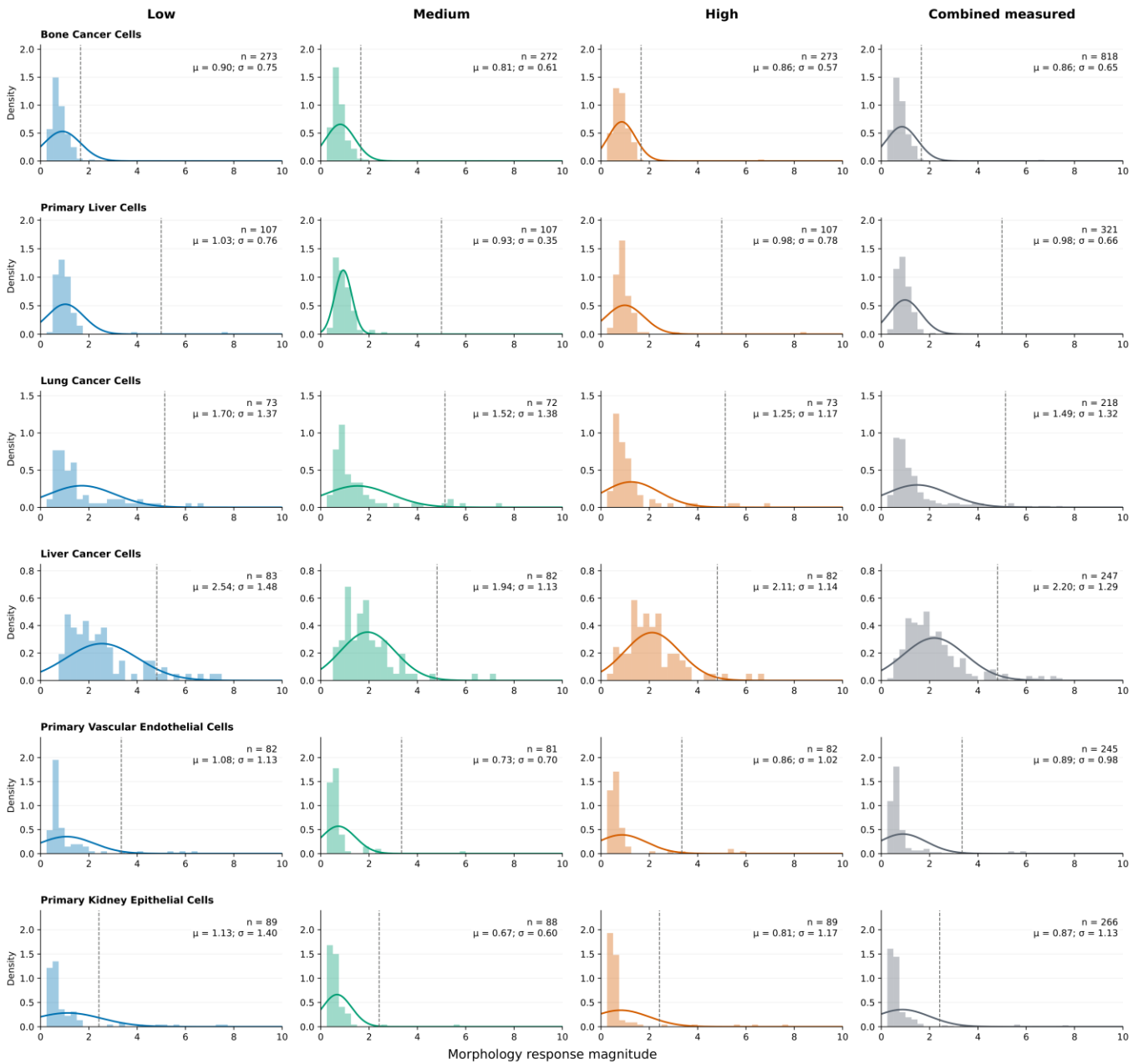


Fig S1l. Distributions of morphology response magnitude stratified by Solubility response status across six cell types.

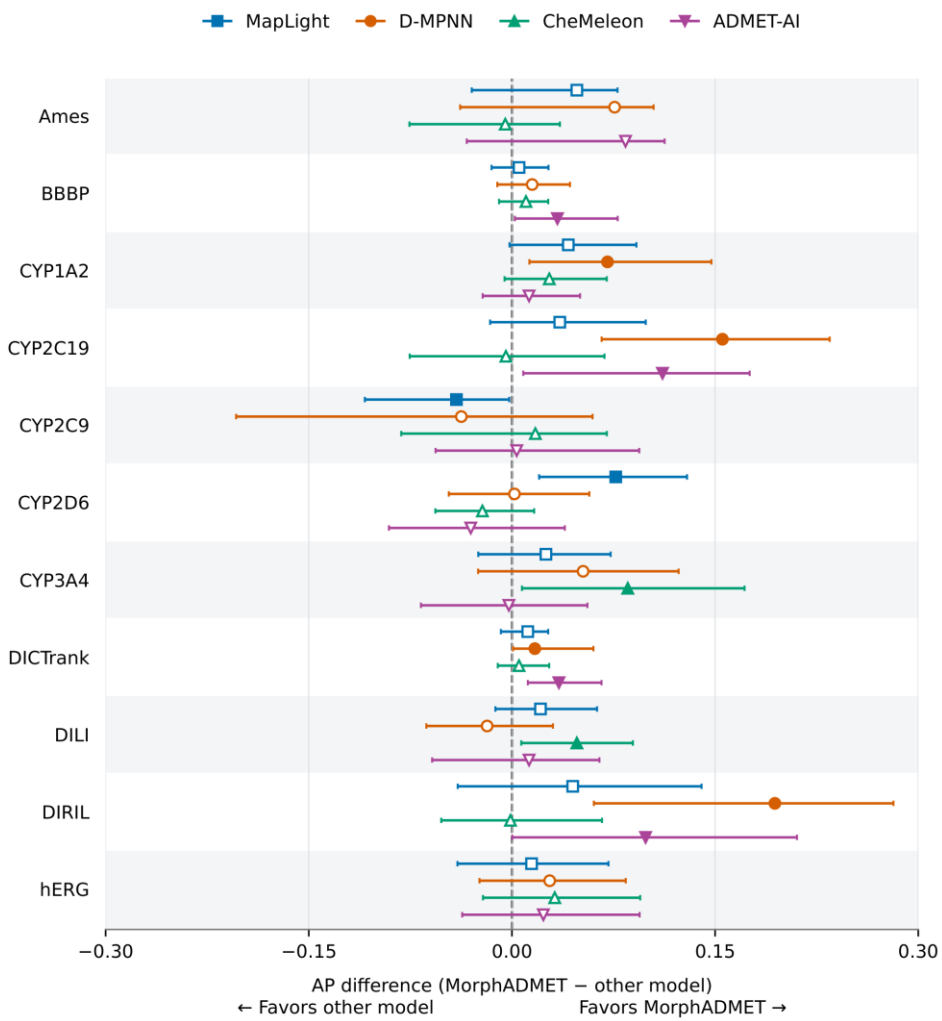


Figure S2. Paired differences in benchmark average precision (AP). *Markers show MorphADMET minus comparator AP, with pointwise 95% confidence intervals. Positive differences favor MorphADMET. Filled markers indicate intervals excluding zero; hollow markers indicate intervals containing zero. The 44 comparisons are reported in Table S16.*

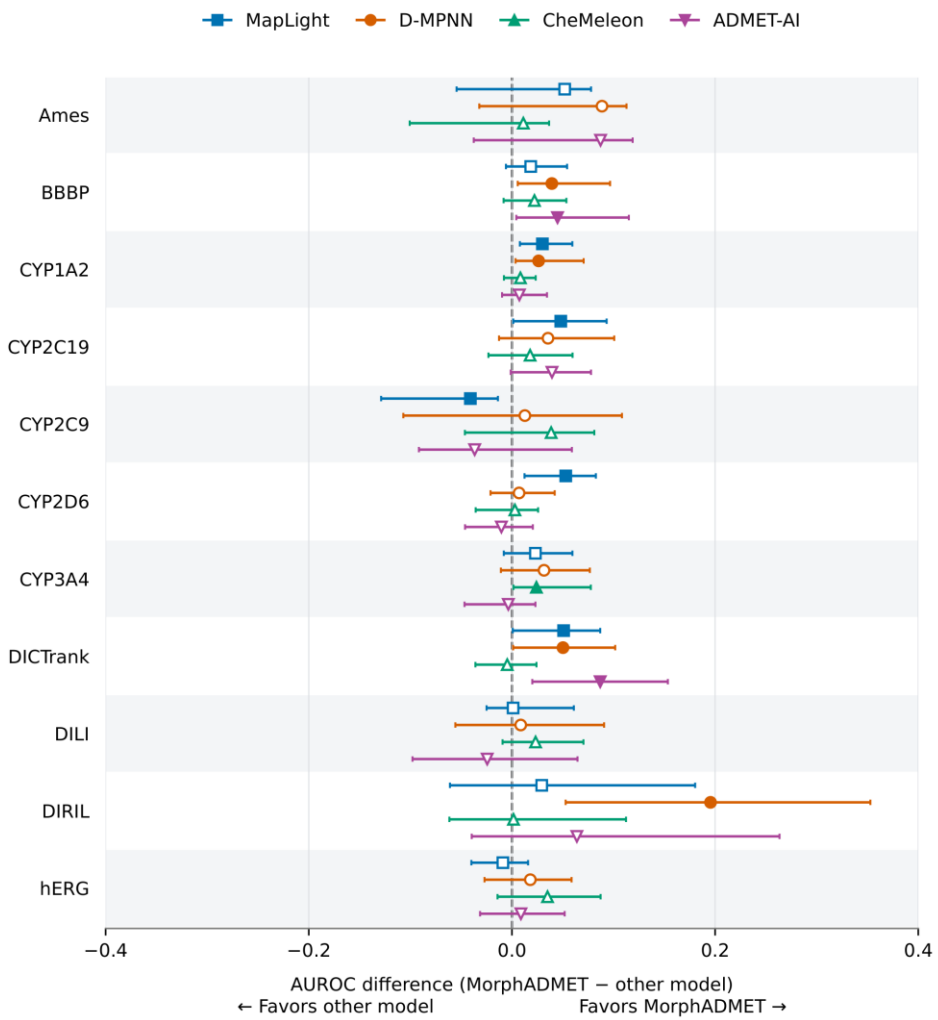


Figure S3. Paired differences in benchmark AUROC. *Markers show MorphADMET minus comparator AUROC, with pointwise 95% confidence intervals. Positive differences favor MorphADMET. Filled markers indicate intervals excluding zero; hollow markers indicate intervals containing zero. The 44 comparisons are reported in Table S16.*

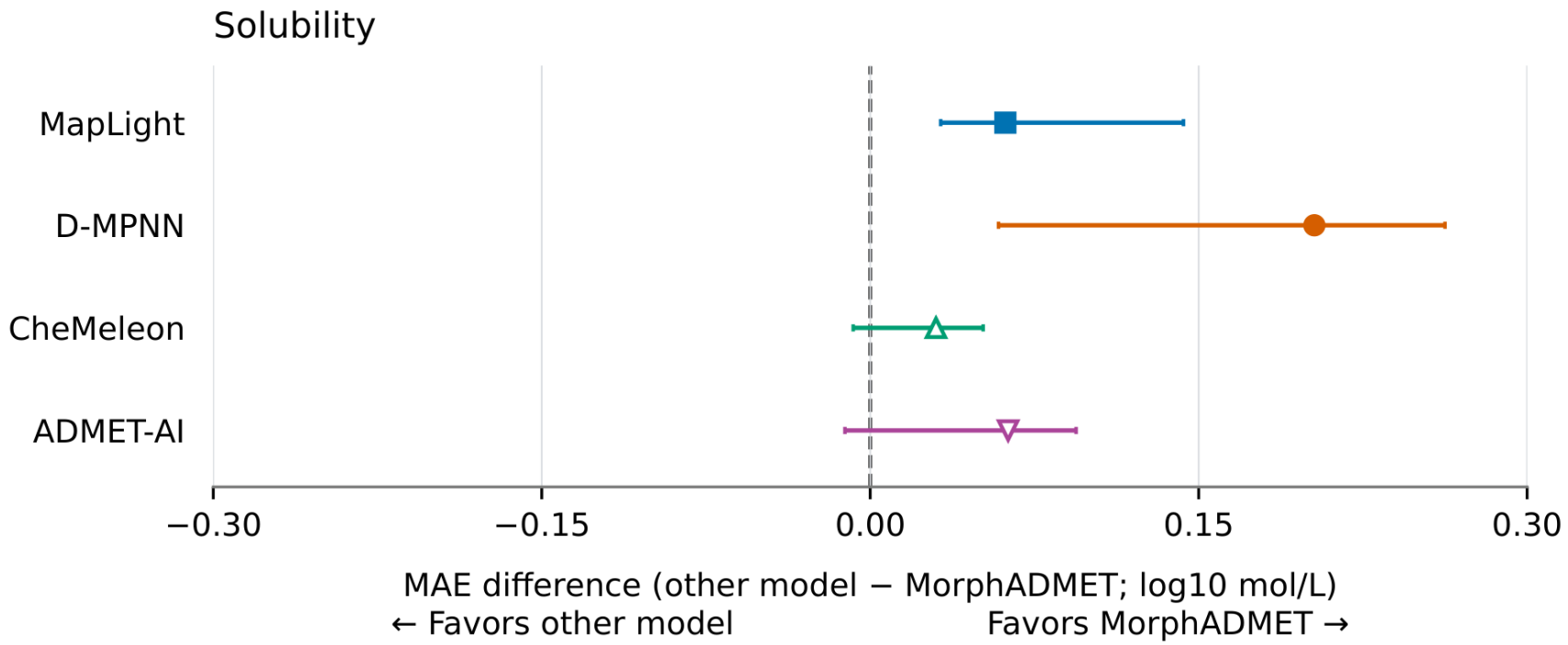


Figure S4. Paired differences in solubility prediction error. Markers show comparator MAE minus MorphADMET MAE (log $_{10}$ mol/L), with pointwise 95% confidence intervals. Positive differences favor MorphADMET. Filled markers indicate intervals excluding zero; hollow markers indicate intervals containing zero. All four comparisons are reported in Table S16

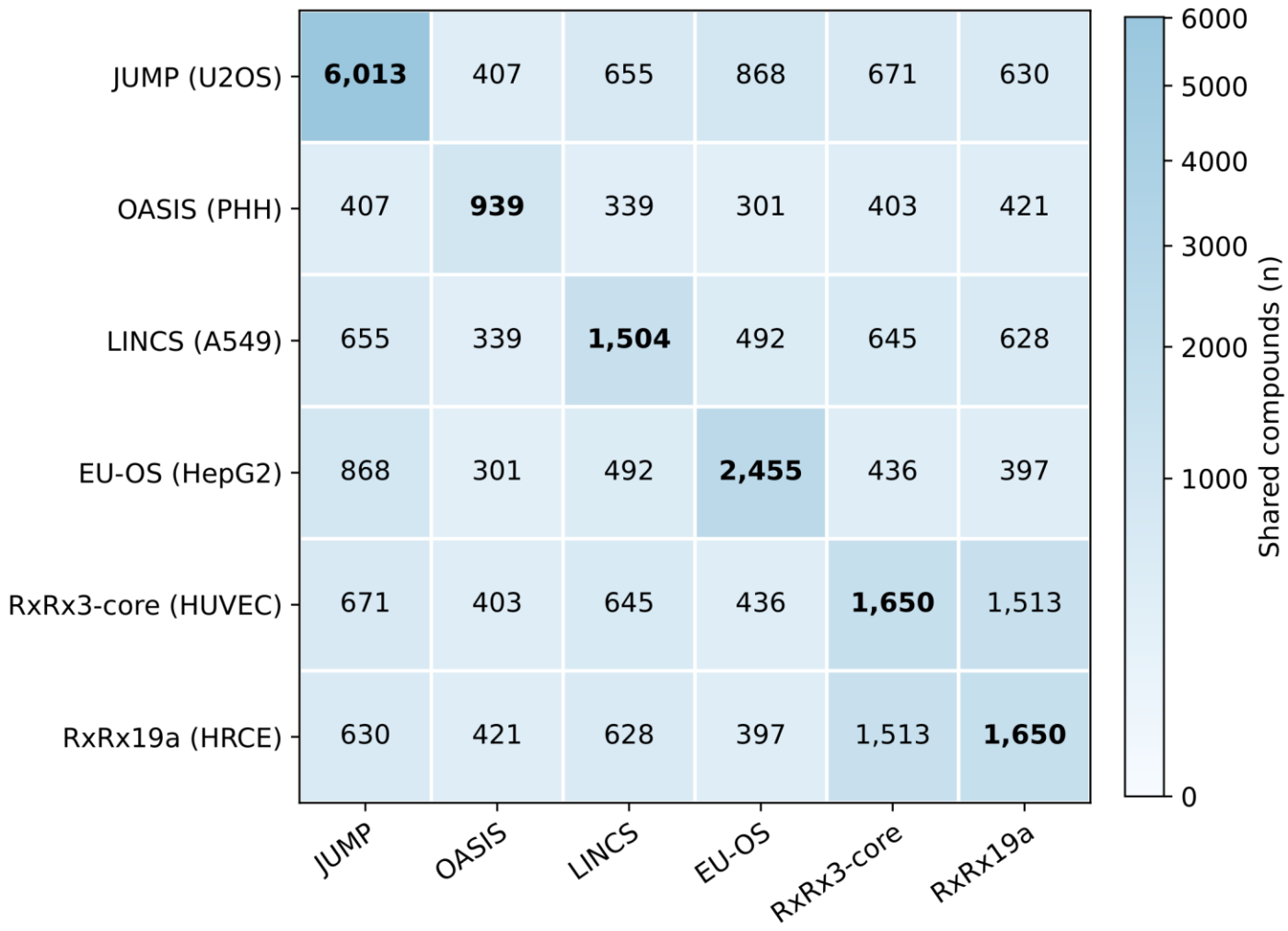


Figure S5. Compound overlap across six imaging datasets. *The matrix shows shared compounds based on exact standardized InChIKeys. Diagonal values represent retained dataset sizes; off-diagonal values represent pairwise overlap. The combined population contains 9,161 unique compounds, including 79 shared across all six datasets. OASIS is counted once because both representations use the same compounds.*