

Interpretable tissue-resolved drug toxicity prediction with pathway-level mechanism attribution

Aleksandr Ianevski

`aleksandr.ianevski@ntnu.no`

Norwegian University of Science and Technology (NTNU)

Erlend Ravlo

Norwegian University of Science and Technology (NTNU)

Jørn Schjølberg

Norwegian University of Science and Technology (NTNU)

Rakel Sæther

Norwegian University of Science and Technology (NTNU)

María Quílez

Norwegian University of Science and Technology (NTNU)

Ine Nordli

Norwegian University of Science and Technology (NTNU)

Hilde Lysvand

Norwegian University of Science and Technology (NTNU)

Synnøve Fjeldstad

Norwegian University of Science and Technology (NTNU) <https://orcid.org/0009-0003-3720-4429>

Marte Skjelle

Norwegian University of Science and Technology (NTNU)

Emil Wiik

Norwegian University of Science and Technology (NTNU)

Anna Lademo

Norwegian University of Science and Technology (NTNU)

Maja Dahle

Norwegian University of Science and Technology (NTNU)

Evgeny Kuleskiy

Institute for Molecular Medicine Finland <https://orcid.org/0000-0001-5965-2917>

Jani Saarela

FIMM, University of Helsinki <https://orcid.org/0000-0001-7306-7175>

Juho Väänänen

Institute for Molecular Medicine Finland

Niklas Andersen

Norwegian University of Science and Technology (NTNU)

Vidar Saasen

Norwegian University of Science and Technology (NTNU)

Anne Mäkelä

Institute for Molecular Medicine Finland

Anni Nieminen

Institute for Molecular Medicine Finland

Nazir Kenneth

Institute for Molecular Medicine Finland

James Booth

Norwegian University of Science and Technology (NTNU)

Emily Helgesen

Norwegian University of Science and Technology (NTNU)

Jing Ye

Norwegian University of Science and Technology (NTNU)

Magnar Bjørås

Norwegian University of Science and Technology (NTNU)

Denis Kainov

Norwegian University of Science and Technology <https://orcid.org/0000-0001-7716-6955>

Article

Keywords: drug toxicity, tissue-specific toxicity, machine learning, organoids, antivirals, antibiotics, antipsychotics, antidepressants, transcriptomics, metabolomics

Posted Date: July 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9996365/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Drug-induced adverse effects often depend on tissue context, yet most computational toxicity models return endpoint-specific scores without mechanisms. We developed ToxiPred, a two-tower neural network that predicts a toxicity-oriented drug sensitivity for each drug-tissue pair. ToxiPred combines chemical structure, target annotations and measured or imputed perturbation profiles with tissue expression and essentiality data. A 70-pathway bottleneck links these inputs to prediction, allowing each toxicity estimate to be traced to pathway-level drug-tissue interactions, while a gated DMPNN/Tanimoto imputer extends scoring to SMILES-only compounds. In public validation, ToxiPred achieved mean AUROC = 0.85 across three HepG2 liver endpoints and AUROC = 0.73 on ClinTox CT_TOX. In anti-infective validation, predicted and measured tissue-wise liabilities correlated at $R = 0.78\text{--}0.84$ across six tissue systems. In the psychotropic panel, ToxiPred predicted high liability for quetiapine and sertraline, with lipid/sterol-related pathway attributions including cholesterol biosynthesis and bile-acid/sterol metabolism. Human organoid validation supported these predictions: cardiac organoids were the most vulnerable system, both drugs stopped visible contractions at 20 μM , and transcriptomic/metabolomic profiling recovered lipid/sterol pathway involvement. ToxiPred therefore provides tissue-conditioned cellular toxicity-liability prediction with pathway-level mechanism hypotheses, enabling confidence-aware prioritization of compounds and tissues for experimental safety testing.

1. Introduction

Drug-induced toxicity is a major limitation of modern pharmacotherapy. Although many therapies are highly effective against infection, cancer, and neuropsychiatric disease, their clinical utility is often constrained by adverse effects that reduce adherence, limit dosing, and compromise patient safety. Because many drugs are administered systemically, toxicity frequently emerges in tissues unrelated to the primary therapeutic target, creating a need for approaches capable of anticipating organ-specific vulnerability before overt injury occurs [1–3].

Current computational methods for toxicity prediction remain limited in biological resolution. Many models rely primarily on molecular fingerprints and are trained for single endpoints such as Tox21 readouts, hepatotoxicity, cardiotoxicity, nephrotoxicity or acute oral toxicity [4, 5]. Drug-response models trained on cancer cell lines have improved cell-line sensitivity prediction [6, 7], yet their direct use for tissue-specific toxicity remains challenging unless tissue biology is explicitly represented and the prediction can be interpreted mechanistically.

These limitations are particularly important for anti-infective and psychotropic drugs. Anti-infective agents span diverse antibacterial and antiviral mechanisms while often requiring systemic exposure, making them a stringent test of whether host-cell toxicity can be separated from therapeutic selectivity. Psychotropic compounds pose a complementary challenge because chronic exposure is associated with metabolic, cardiovascular and neuroendocrine liabilities that vary by tissue context. Both settings

require a framework that estimates not only whether a compound is toxic, but also where toxicity is most likely to occur and which pathways may underlie that liability.

We reasoned that pathway-level representations could provide a transferable and interpretable bridge between drug perturbation data and tissue-specific vulnerability. In ToxiPred, drug evidence is summarized from chemical structure, ChEMBL target annotations and measured or imputed perturbation profiles from LINCS/Tahoe-type resources, whereas tissue evidence is summarized from DepMap expression and CRISPR essentiality profiles [8–13]. A pathway-bottleneck neural network then learns pathway-wise drug-tissue interactions and predicts one toxicity-oriented drug sensitivity score (DSStox) value per tissue. Because many candidate compounds lack measured perturbation profiles, ToxiPred also includes an out-of-distribution chemistry imputer that estimates missing perturbation evidence from SMILES. To support practical deployment, we additionally precomputed and release a DSStox prediction atlas for 53,720 catalogued small molecules across the 24 modeled tissues, distributed through an interactive web interface (<https://toxipred.com>), the project repository, and a self-contained Docker container. The web interface was designed to support an end-to-end practical workflow in which a user submits a compound, receives a tissue-resolved DSStox heatmap, reviews pathway-level attributions and interprets predictions together with confidence information. This deployment setting was evaluated before experimental follow-up using public toxicity benchmarks that separate tissue-routed ToxCast/Tox21 liver assays from compound-level ClinTox clinical-trial toxicity labels.

Here we apply ToxiPred in the same sequence as its intended use: model selection, out-of-distribution deployment, experimental validation and mechanism recovery. We first test whether the pathway bottleneck is a predictive inductive bias rather than a post-hoc interpretation layer by comparing it with linear, boosting, factorization, graph-encoder and pathway-resolution alternatives under drug-disjoint evaluation. We then deploy the trained model to SMILES-only compounds and evaluate high-confidence predictions on public ToxCast/Tox21 liver assays and ClinTox clinical-trial toxicity labels. Next, we test predicted host-cell liability in a chemically diverse anti-infective and reference panel across six tissue-representative cell-line systems. Finally, we apply ToxiPred to psychotropic drugs, using pathway attributions to nominate mechanisms and human organoids with transcriptomic and metabolomic profiling to test them. This sequence links prediction, validation and explanation within a single tissue-aware workflow, with sertraline and quetiapine serving as an out-of-distribution antipsychotic case study connecting tissue-resolved scoring to independent hepatobiliary and metabolic biology. Source code, trained models, inference scripts and benchmark-evaluation workflows are available at <https://github.com/lanevskiAleksandr/ToxiPred>, and an interactive web interface for querying the precomputed atlas is available at <https://toxipred.com>.

2. Materials and Methods

2.1. Drug-response training labels

Per-(drug × cell-line) toxicity-oriented drug-sensitivity scores were assembled from nine public pharmacogenomic drug-response resources accessed through the PharmacDB 2.0 platform: CCLE, CTRPv2, GDSC1, GDSC2, NAIR, NCI60, PRISM, gCSI and gds2combi [14]. For each compound-cell-line pair, a drug sensitivity score (DSStox) was computed as the normalized area above an activity threshold of 10% across the tested dose range from four-parameter log-logistic dose-response fits, following the formulation of Yadav et al. [15] and the high-throughput drug-response processing framework used in Breeze and Breeze 2.0 [16, 17]. Scores were reported on a 0–50 scale, with higher values indicating stronger drug-induced loss of viability. After harmonization on canonical SMILES and DepMap cell-line identifiers, this resulted in 178,117 per-(drug × cell-line) observations across 757 unique drugs and 1,145 cell lines spanning 24 DepMap tissues. Per-cell observations were aggregated by taking the median across replicates and averaging across cell lines within each tissue, with a quality filter requiring at least three cell lines per tissue. This resulted in 9,275 drug-lineage DSStox training labels across 755 drugs and 24 lineages, representing essentially the full set of compounds in public pharmacogenomic resources with sufficient per-tissue coverage for stable label aggregation. In toxicity analyses we refer to this viability-derived DSS as DSStox to distinguish it from the general drug-sensitivity-score formulation. In addition to the 9,275 labelled training pairs, we extended chemical coverage to a 53,720-compound chemistry universe drawn from PubChem-resolvable SMILES across the same nine pharmacogenomic resources. These additional compounds had no measured LINCX or Tahoe perturbation profiles and were not used for model fitting. Instead, the trained ensemble was applied forward to every (drug × tissue) pair using DMPNN- and Tanimoto-imputed pathway features (Methods 2.2), yielding $53,720 \times 24 \approx 1.29 \times 10^6$ precomputed DSStox predictions that constitute the released atlas (Methods 2.3).

2.2. ToxiPred model inputs and architecture

ToxiPred was implemented as a pathway-bottleneck dual-encoder neural network for tissue-specific DSStox prediction. The drug input vector contained 488 features: 424 pathway, perturbation and target features, and 64 chemical-structure features (ECFP4 fingerprints concatenated with RDKit descriptors and reduced by principal-component analysis to 64 dimensions) [18, 19]. The 424 pathway and target features were derived from measured LINCX L1000 (978 landmark plus 11,350 inferred genes) and Tahoe-100M single-cell RNA-sequencing perturbation gene-rank profiles when available, ChEMBL release 35 target annotations (binarized at activity thresholds of 1 μ M and 10 μ M), pathway-source indicators and imputed perturbation profiles for compounds without measured LINCX or Tahoe data [8, 12, 13]. For each compound with measured perturbation data, per-gene rank profiles from these resources were aggregated across cell lines and concentrations by taking the median rank, then summarized within each of 70 named-toxicity pathways using mean, 90th percentile and standard deviation of within-pathway gene ranks. The 70-pathway dictionary was assembled from MSigDB Hallmark, Reactome and curated toxicity-relevant gene sets (full list in Supplementary Data). Rank-based aggregation makes the perturbation representation robust to absolute fold-change magnitudes, normalization differences and cell-line-specific dynamic range while preserving the relative direction and magnitude of pathway perturbation.

For new compounds represented only by SMILES, missing perturbation evidence was estimated using a gated chemistry imputer. One branch used a directed message-passing neural network (DMPNN) trained to predict LINCS/Tahoe-like perturbation ranks from molecular structure [20]. The second branch used a Tanimoto-K10 nearest-neighbour imputer, which estimated pathway features from the ten most chemically similar measured training compounds using ECFP4 Tanimoto similarity [18]. The two pathway representations were combined with a smooth similarity gate, $w_{NN} = \text{sigmoid}(10 \times (\text{top1_Tanimoto} - 0.30))$, so that close analogues contributed more nearest-neighbour evidence whereas chemically distant compounds relied more strongly on DMPNN-imputed perturbation evidence.

The tissue input vector contained 490 features computed by aggregating DepMap 23Q2 CCLE RNA expression and DepMap 23Q2 CRISPR essentiality profiles over the same 70-pathway dictionary across 24 tissues [9–11]. RNA-derived features summarized pathway-level expression as mean, 90th percentile and standard deviation of within-pathway gene expression across cell lines of each tissue. CRISPR-derived features summarized pathway-level dependency and essentiality as mean and 10th percentile of gene-effect scores plus the fraction of essential and strongly essential genes per pathway. Each prediction therefore used a 488-dimensional drug vector and a 490-dimensional tissue vector.

The drug encoder contained a pathway branch with layer sizes 488, 256, 128 and 70 and a shortcut branch with layer sizes 488, 64 and 16. The tissue encoder contained a pathway branch with layer sizes 490, 128, 64 and 70 and a shortcut branch with layer sizes 490, 32 and 16. The two pathway branches projected drug and tissue evidence into a shared 70-dimensional pathway space. Their element-wise Hadamard product, $H = z_{\text{drug}} \times z_{\text{tissue}}$, formed the pathway bottleneck. The final latent representation concatenated H , the 16-dimensional drug shortcut, the 16-dimensional tissue shortcut and a scalar prior, yielding 103 features. These were decoded by a regression head with layer sizes 103, 128, 64 and 1 to predict one DSStox value for each drug–tissue pair.

2.3. Model training, cross-validation and ablation protocol

The model was trained using SmoothL1 loss and AdamW optimization with learning rate 5×10^{-4} , weight decay 1×10^{-3} , cosine learning-rate scheduling, dropout 0.3, batch size 128, gradient clipping at norm 5 and 180 training epochs. Implementation used PyTorch with seeds 0–4 for reproducibility. Performance was estimated using drug-disjoint scaffold cross-validation with five folds and five random seeds; all r and RMSE values reported in Results are out-of-fold from this cross-validation. The final deployment model was a five-seed all-data ensemble that was not evaluated on training data.

Pearson correlation, Spearman correlation, root-mean-square error and mean absolute error were computed between predicted and out-of-fold DSStox values across 9,275 drug-tissue labels (755 drugs, 24 tissues). Model selection used drug-disjoint performance, external public-benchmark transfer and interpretability as joint criteria. Alternative model families and feature configurations were evaluated under the same drug-disjoint selection logic and included Ridge regression, ElasticNet, LightGBM, CatBoost, factorization machines, multi-output dose-anchored predictors, graph-neural-network

chemistry encoders, CRISPR-feature ablation and pathway dictionaries of varying granularity (40–500 pathways). The full ablation history is reported in Supplementary Table 1.

To enable instant inference for clinical and investigational compounds without re-running the network, we precomputed a 53,720-drug × 24-tissue DSStox prediction atlas using the final ensemble. Queries against drugs already in the atlas return precomputed values in milliseconds; queries on novel SMILES trigger a full forward pass of the trained ensemble. The atlas is released alongside the trained model weights at the project repository.

2.4. Public benchmark evaluation

External public validation was performed using ToxCast/Tox21 toxicity assays and ClinTox clinical-trial toxicity labels [21–24]. The main tissue-routed public validation used three HepG2 liver endpoints measured at 24 h from the U.S. EPA ToxCast Toxicity Forecaster: liver DNA damage, liver apoptosis and liver steatosis (assay-source IDs and active/inactive call thresholds as defined in the ToxCast invitrodb release). Liver-tissue DSStox was used as the prediction score. ClinTox CT_TOX was evaluated as a compound-level endpoint using the MoleculeNet/Therapeutics Data Commons release [23, 24]; because no single tissue anchor is defined for ClinTox, the cross-tissue maximum DSStox across the 24 predicted tissues was used as the compound-level score, on the reasoning that a drug is clinically toxic if it is toxic in any tissue.

For each benchmark, AUROC was computed using the public binary endpoint labels and standard scikit-learn implementations. To evaluate ToxiPred as a decision tool rather than a continuous ranking, a fixed confidence rule was applied before endpoint scoring: predicted DSStox below 3.0 was treated as confidently low-liability, DSStox above 5.0 as confidently high-liability, and intermediate scores (DSStox 3.0–5.0) as borderline and excluded from confident calls. The thresholds were fixed before endpoint evaluation.

2.5. Pathway attribution

Pathway attribution was computed directly from the 70-dimensional bottleneck interaction vector H . For each drug-tissue pair, positive H values were interpreted as pathway-wise drug-tissue interactions contributing to higher predicted DSStox. Pathways were ranked by H within each prediction to identify candidate mechanisms. For cross-tissue mechanism summaries, pathway ranks and mean H values were summarized across all 24 tissues. H values were not normalized across drugs or pathways; comparisons are within-prediction ranks rather than absolute magnitudes.

2.6. Psychotropic compound panel

Canonical SMILES for 91 approved psychotropic compounds were retrieved from DrugBank and PubChem to test mechanism recovery in clinically used neuropsychiatric compounds. The panel spanned antipsychotic, antidepressant and related classes (SSRIs, SNRIs, TCAs, MAOIs, NaSSA, SARI, NDRI and melatonergic agents). Six representative compounds spanning the predicted DSStox range were selected for organoid validation: amitriptyline, aripiprazole, loxapine, quetiapine, sertraline and

xanomeline. For descriptive ranking, predicted DSS_{tox} values were aggregated across the 24 model tissues using median, mean and maximum DSS_{tox}. Median DSS_{tox} was used as the primary ranking statistic because it is less dominated by a single sensitive tissue, whereas mean and maximum DSS_{tox} were retained as complementary measures of broad and peak liability.

2.7. Small molecules and bacterial counter-screening

The anti-infective and reference compound panel was assembled and dispensed through the FIMM High Throughput Biomedicine Unit compound-collection and screening infrastructure. Stock solutions were prepared in DMSO or water as appropriate and stored at – 20°C until use. The antibacterial subset included representative quinolones, macrolides, tetracycline-related agents, oxazolidinones, β -lactams, aminoglycoside-related drugs, glycopeptides, lipopeptides, polymyxins and additional antibacterial agents. The antiviral subset included direct-acting antivirals and host-targeting antiviral candidates.

Bacterial activity validation was performed on 93 drugs against four clinical isolates of *Escherichia coli*, *Pseudomonas aeruginosa*, *Acinetobacter baumannii* and *Staphylococcus aureus*. Bacteria were inoculated at standardized starting OD into cation-adjusted Mueller–Hinton broth, exposed to fivefold serial dilutions of compounds and incubated for 18–20 h at 37°C. Growth inhibition was read out by OD₆₀₀ at endpoint. Dose–response curves were fitted with the same four-parameter log-logistic model used for host-cell DSS_{tox}, and bacterial DSS was computed on the same 0–50 scale to enable direct comparison between antibacterial activity and host-cell liability. DSS_{max} denotes the maximum bacterial DSS across the four strains.

2.8. Cell cultures

A549 and MiaPaCa-2 cells were grown in Dulbecco's Modified Eagle's Medium supplemented with 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 10% heat-inactivated fetal bovine serum. Additional cell lines (Caki-1, HCT116, HEK293, Hep3B, HepG2, MOLM-13, HDQ-P1 and DU4475) were maintained under recommended conditions. All cells were cultured at 37°C with 5% CO₂ and 95% humidity, passaged using 0.05% Trypsin/EDTA and confirmed mycoplasma-negative throughout the study.

2.9. Human organoid cultures

Forebrain, retinal and cardiac organoids were generated from human induced pluripotent stem cells (iPSCs) using modified versions of published forebrain, retinal and cardiac organoid protocols [25–27], with adjustments to medium composition, differentiation factors, timing and initial seeding density. iPSCs were reprogrammed from human dermal fibroblasts (obtained under approved donor-consent protocols, see Ethics approval) using the CytoTune-iPS 2.0 Sendai Reprogramming Kit, maintained in Essential 8 medium and passaged every 4–5 days using 0.5 mM EDTA. Orbital shaking was applied throughout differentiation and maintenance to improve nutrient and gas exchange. Forebrain, retinal and cardiac organoids were collected on days 30, 20 and 53 post-induction, respectively, and each organoid was placed in one well of an ultralow-attachment 96-well plate in 100 μ l medium before compound treatment.

2.10. Dose-response scoring and cardiac organoid functional assessment

Drugs were added to cells or organoids as eight-point 5-fold serial dilutions spanning 0.0064 to 100 μ M in 100 μ l medium per well. Plates included untreated and DMSO-only controls; final DMSO concentrations were kept below 0.1%. Organoids were sonicated 72 h after treatment, and viability was measured using CellTiter-Glo, with luminescence normalized to plate-matched untreated controls. Dose-response curves were fitted with the four-parameter log-logistic model

$$y(c) = y_{\min} + (y_{\max} - y_{\min}) / (1 + (c / IC50)^s) \quad (1)$$

where c is drug concentration and s is the slope parameter. Drug sensitivity score was computed from the normalized area above an activity threshold across the tested dose range, following the DSS formulation of Yadav et al. [15] and the Breeze high-throughput drug-response processing framework [16, 17]. An activity threshold of 10% was used and scores were reported on a 0–50 scale. For bacterial assays the term bacterial DSS is retained because the readout reflects antibacterial activity rather than host-cell toxicity.

For cardiac organoids, visible spontaneous contractions were monitored microscopically at 48 h after treatment. The absence or cessation of visible contractions was recorded as a functional cardiac phenotype independent of ATP-based viability. This qualitative readout was used to identify compounds that impaired cardiac organoid function under the tested in vitro conditions.

2.11. Transcriptomics

Total RNA was isolated from organoids treated for 24 h at 5 μ M, a submaximal concentration selected to capture stress and adaptation programs rather than terminal cytotoxicity, using the NAXtra kit according to the manufacturer's protocol. Messenger RNA was enriched using oligo-dT beads, fragmented and used for cDNA synthesis. Libraries were prepared by adaptor ligation and PCR amplification, circularized, amplified into DNA nanoballs and sequenced on the DNBSEQ platform with paired-end 150 bp reads. Raw sequencing data were filtered to remove low-quality reads and adapter contamination. Reads were aligned to the human GRCh38 reference genome using STAR v2.7.11b [28]. Gene-level counts were generated using quantMode GeneCounts, and gene symbols and descriptions were assigned using biomaRt [29]. Differential expression was performed in R using edgeR v3.42 [30] with the quasi-likelihood pipeline. Counts were filtered with filterByExpr (min.count = 10, min.total.count = 15), normalized by trimmed mean of M-values (TMM), and dispersions were estimated with estimateDisp(robust = TRUE) followed by glmQLFit(robust = TRUE). Two complementary designs were used:

For the Qi-modulates-RA analysis, 36 samples (3 organoid types $\times$ 12 treatment conditions) were fit jointly with the design $\sim$ tissue + treatment. The contrast Qi_RA - RA was tested with glmQLFTest to identify genes whose response to retinoic acid was modified by co-treatment with quetiapine. Tissue was included as a blocking factor to absorb baseline expression differences across organoid types. For

per-tissue psychotropic responses, each organoid type was analyzed separately by combining the six psychotropic-drug samples (amitriptyline, aripiprazole, loxapine, quetiapine, sertraline, xanomeline and a No_drug reference) with five neurotransmitter conditions (carbachol, dopamine, GABA, glutamate, serotonin) collected from a parallel experimental arm on the same organoid batches. The resulting 30 samples per tissue (5×6 factorial of neurotransmitter and drug) were fit jointly with the design $\sim$ neurotransmitter + drug, so that the neurotransmitter factor absorbed its main effect and residual variance was estimated from real biological samples of the same organoid system rather than from technical duplicates of single conditions. The drug coefficient therefore captured the psychotropic-drug effect averaged across neurotransmitter pre-conditioning. Each drug was tested against the No_drug reference with glmQLFTest.

To distinguish quetiapine-driven from retinoic-acid-driven responses, three additional coefficients (RA, Qi, Qi_RA versus No_drug) were extracted from the same joint model and each top-ranked gene was classified by whether its Qi_RA - RA difference reflected RA induction attenuated by Qi, Qi-driven suppression from baseline, or Qi-rescued RA suppression. Significance thresholds were FDR < 0.05 with $|\log_2 \text{ fold change}| > 1$ for volcano plots, and a relaxed threshold of FDR < 0.05 with $|\log_2 \text{ fold change}| > 0.5$ for heatmap and pathway-enrichment input. Pathway enrichment was performed with clusterProfiler v4.8.3 [31] using enrichGO (Biological Process, OrgDb = org.Hs.eg.db) and enrichKEGG (organism = "hsa"), with $p < 0.05$ and $q < 0.1$. GO, MSigDB and GSEA-style gene-set resources were used where additional summaries were needed [31, 32].

2.12. Metabolomics

Metabolite extraction was performed from culture media using 40:40:20 acetonitrile:methanol:water. Samples were vortexed, centrifuged, dried under nitrogen, reconstituted and transferred to LC-MS vials.

Analysis was performed using pHILIC chromatography at pH 9.4 with Orbitrap MS and Vanquish LC instrumentation. Chromatographic peaks were visually inspected and processed with Trace Finder 5.2 SP1. Data were normalized to sample volume or total intensity. Blank extraction controls, pooled media controls and in-house serum controls were used to assess carryover and instrument stability. Metabolites were annotated by accurate mass and retention time matched against an in-house authentic-standard library, corresponding to MSI Level 1 confidence where standards were available and Level 2 (accurate mass + retention time, without authentic standard) otherwise.

Differential abundance was assessed in R with limma [33]. Metabolites detected in fewer than 20 of 36 samples were filtered out; remaining intensities were $\log_2$ -transformed after adding a unit pseudocount, and residual missing values were imputed as half the row minimum to indicate below-detection abundance. The design $\sim$ tissue + treatment was fit with lmFit followed by eBayes, and the contrast Qi_RA - RA was tested with contrasts.fit and eBayes. Significance thresholds were adjusted $P < 0.05$ with $|\log_2 \text{ fold change}| > 0.5$.

2.13. Bioinformatic visualization across organoid types

Heatmaps in Figs. 3 and 4 display logCPM-derived fold changes (transcriptomic) or log₂ abundance changes (metabolomic) against the matched No_drug control within each organoid type, allowing direct comparison of Qi, RA and Qi_RA conditions across tissues without re-normalizing per row. Because gene and metabolite detection coverage differed across organoid types, each tissue was analyzed as a separate block rather than forcing a common feature list across all tissues. Pathway enrichment terms were summarized using GO Biological Process, KEGG and MSigDB gene-set resources [31, 32]. Analyses and figure generation were performed in R (edgeR, limma, clusterProfiler, ggplot2, pheatmap, patchwork) and Python.

2.14. Statistical analysis

Two-group comparisons of predicted DSStox between toxic/active and non-toxic/inactive compounds (Fig. 1f) were performed with the Wilcoxon rank-sum (Mann-Whitney U) test. Significance markers in figures use * $p < 0.01$ and ** $p < 0.001$. Tissue-wise correlations between predicted and measured DSStox (Fig. 2b) were computed using Pearson correlation. AUROC was computed using standard scikit-learn implementations. For transcriptomics and metabolomics, FDR was controlled with the Benjamini-Hochberg procedure as implemented in edgeR (topTags) and limma (topTable). No adjustment for multiple comparisons was applied to the pathway-attribution H values, which are reported as within-prediction ranks rather than as significance tests. Drug-specific tissue-lineage selectivity was assessed by fitting additive compound and lineage effects separately to measured and drug-disjoint held-out predicted DSStox matrices and comparing the remaining interaction residuals. Bootstrap confidence intervals were computed by resampling compounds with replacement, and permutation significance was assessed by breaking drug-lineage pairing.

2.15. Ethics approval

The study involved commercially available human cell lines obtained from ATCC. Human iPSC-derived organoids were generated from iPSCs reprogrammed from donor human dermal fibroblasts obtained under approved donor-consent procedures by the providing institution.

3. Results

3.1. A pathway bottleneck predicts tissue-resolved toxicity and explains it

ToxiPred was designed to predict tissue-resolved toxicity liability from both compound evidence and tissue context rather than from chemical structure alone. The model combines drug evidence from chemical descriptors, ChEMBL target annotations and measured or imputed perturbation profiles (from the LINCS L1000 consortium and the Tahoe-100M cell-line perturbation atlas [8, 13]) with tissue evidence from DepMap CCLE RNA expression and CRISPR essentiality data [9–11] (Fig. 1a). Training labels were tissue-aggregated drug sensitivity scores (DSStox) derived from nine public pharmacogenomic drug-response resources (CCLE, CTRPv2, GDSC1, GDSC2, NAIR, NCI60, PRISM, gCSI

and gds2combi). After harmonizing on canonical SMILES and cell-line identifiers and intersecting with compounds carrying complete drug-side evidence, this resulted in 9,275 drug-tissue labels (full data-processing details in Methods). Drug and tissue vectors are encoded by separate towers and projected into a shared 70-pathway space, where the element-wise interaction $H = z_{\text{drug}} \times z_{\text{tissue}}$ forms the model bottleneck (Fig. 1b). Each prediction therefore yields two coupled outputs: a DSStox value for each of 24 tissues (Fig. 1c, **top**) and an H vector that ranks the 70 pathways by their contribution to that prediction (Fig. 1c, bottom). For drugs with well-characterized pharmacology, the top-H pathway was consistent with their canonical biology: venetoclax mapped to apoptosis (BCL2 inhibition), methotrexate to G2M checkpoint regulation (downstream of its antiproliferative action), doxorubicin to programmed cell death (downstream of DNA damage), and simvastatin to bile-acid metabolism (within the lipid/sterol axis it targets via HMG-CoA reductase) (Fig. 1d), indicating that prediction and pathway attribution share the same internal representation.

We tested ToxiPred on compounds whose chemical scaffolds it had never seen during training using drug-disjoint scaffold cross-validation. Under this strict generalization test, ToxiPred achieved Pearson $r = 0.733$ and RMSE = 5.02 (Fig. 1e). To determine whether this performance reflected only broad compound-level liability or also drug-specific tissue-lineage selectivity, we decomposed the measured and held-out predicted DSStox matrices into additive compound effects, additive lineage effects and residual drug $\times$ lineage interactions (**Suppl. Fig. S1**). ToxiPred captured the dominant compound-level cellular-liability signal (Pearson $r = 0.71$) and lineage-wide sensitivity (Pearson $r = 0.87$). After additive compound and lineage effects were removed, predicted and measured interaction residuals remained correlated (Pearson $r = 0.36$, 95% CI [0.31, 0.41]; permutation $p < 0.001$), indicating that ToxiPred retained drug-specific lineage-selectivity signal beyond global compound potency and lineage-wide sensitivity. Representative held-out profiles recovered established lineage-enriched responses, including quizartinib – myeloid, venetoclax – blood lineage, BRAF/MEK inhibitor – skin and fulvestrant – breast responses.

We also systematically evaluated multiple alternative architectures and feature representations during model development, including Ridge regression, ElasticNet, LightGBM, CatBoost, factorization machines, multi-output dose-anchored predictors, graph-neural-network chemistry encoders and pathway dictionaries of varying granularity (40–500 pathways). Classical machine-learning approaches performed worse under drug-disjoint evaluation even after thorough tuning (Ridge: $r = 0.666$, RMSE = 5.47; Optuna-tuned LightGBM with 11 hyperparameters across 60 trials: $r = 0.686$, RMSE = 5.29; factorization machines: $r = 0.716$, RMSE = 5.09; **Suppl. Table 1**), and replacing the final ECFP4/RDKit chemistry representation with learned graph-based chemistry embeddings also reduced performance (DMPNN: $r = 0.726$ versus 0.733). Increasing pathway resolution beyond approximately 70 pathways degraded performance, suggesting over-fragmentation of the internal pathway representation (**Suppl. Table 1**). The final 70-pathway interaction space therefore provided the best balance of drug-disjoint accuracy, stability and pathway-level interpretability. Because the same bottleneck supported both prediction and attribution, the next test was whether this inductive bias could be deployed for compounds without measured perturbation profiles.

3.2. ToxiPred transfers to new compounds on public toxicity benchmarks

Many candidate compounds do not have measured LINCS or Tahoe perturbation profiles. ToxiPred therefore uses a gated imputer for compounds available only as chemical structures (SMILES strings), combining DMPNN-imputed perturbation evidence with Tanimoto-nearest-neighbour measured-profile evidence before DSStox prediction. This deployment setting was evaluated on public toxicity benchmarks before experimental follow-up.

For tissue-specific liver validation, liver-tissue DSStox was compared with three US EPA ToxCast/Tox21 HepG2 liver endpoints measured at 24 h: DNA damage, apoptosis and steatosis. Across all compounds in each benchmark, predicted DSStox was systematically higher for toxic/active compounds than for non-toxic/inactive compounds (Fig. 1f; Wilcoxon rank-sum test, $p < 0.01$). We also tested transfer to ClinTox, a benchmark in which each compound is labeled as having failed (or not) clinical trials for toxicity reasons, regardless of tissue. Because ClinTox labels are compound-level rather than tissue-specific, we summarized each compound by its maximum DSStox across the 24 predicted tissues, on the reasoning that a drug is clinically toxic if it is toxic in any tissue. Compounds with the toxic CT_TOX label had higher cross-tissue maximum DSStox than non-toxic compounds (Fig. 1f, **right**; Wilcoxon rank-sum test, $p < 0.001$).

To use DSStox as a decision tool rather than just a ranking, we treated DSStox clearly below 3.0 as confidently low-liability and DSStox clearly above 5.0 as confidently high-liability, while flagging intermediate scores (DSStox 3.0 to 5.0) as borderline and excluding them from confident calls. ToxiPred separated toxic from non-toxic compounds with mean AUROC = 0.853 across the three ToxCast liver endpoints (DNA damage 0.83, apoptosis 0.90, steatosis 0.83) and AUROC = 0.73 on ClinTox CT_TOX (Fig. 1g).

ToxiPred is therefore best interpreted as a triage tool rather than a full-coverage classifier: it issues confident low-liability (DSStox < 3.0) and high-liability (DSStox > 5.0) calls and flags borderline compounds (DSStox 3.0 to 5.0) for follow-up. The chosen confidence thresholds align with the empirical DSS interpretation established in the original DSS framework, in which scores below ~ 5 reflect weak or non-responsive drug-cell pairings whereas scores above ~ 5 indicate genuine drug-induced sensitivity [15–17]. Treating DSStox < 3.0 as confidently low-liability and DSStox > 5.0 as confidently high-liability therefore reflects established DSS distributional behaviour rather than thresholds tuned to the present benchmarks. Across both benchmark families, the pathway-bottleneck model and imputer transferred to chemically distant compounds available only as SMILES strings, producing high-confidence (mean AUROC ≈ 0.82 across four benchmarks) toxicity rankings on independent data.

3.3. Predicted liability matches measured anti-infective compound toxicity across six tissue systems

The public benchmarks tested out-of-distribution transfer, but they did not test whether ToxiPred could predict experimentally measured host-cell cytotoxic liability in a controlled tissue panel. We therefore applied ToxiPred to a chemically diverse 93-compound anti-infective and reference panel and compared predicted DSStox values with matched experimental dose-response measurements across nine cell lines spanning six ToxiPred tissue lineages: A549 (lung), HepG2 and Hep3B (liver), HCT116 (large intestine), Caki-1 and HEK293 (kidney), MiaPaCa-2 (pancreas), and HDQ-P1 and DU4475 (breast). For tissues represented by two cell lines, measured DSStox values were averaged within tissue before comparison with predicted DSStox, mirroring the cell-line-to-tissue aggregation used to construct the training labels. Because anti-infective compounds target bacterial or viral biology and are expected to produce minimal effect on human host cells, this panel provides a stringent test of whether ToxiPred correctly assigns low host-cell DSStox to selective anti-infectives while still flagging compounds with genuine off-target tissue liability.

Predicted and measured DSStox heatmaps showed concordant compound ordering across the six tissue systems (Fig. 2a). Compounds predicted to have high host-cell liability, including anisomycin, podofilox, flavopiridol, camptothecin, sirolimus and mitoxantrone, were positioned in the high measured DSStox region. Several anti-infective compounds, including ampicillin, acyclovir, emtricitabine, cefepime and cefotetan, grouped in the low-liability region. Representative dose-response curves confirmed the contrast between anisomycin, a predicted high-liability compound, and ampicillin, a predicted low-liability compound (Fig. 2a, **inset**).

Tissue-wise correlations between predicted and measured DSStox were positive across all six systems, with Pearson R values ranging from 0.78 to 0.84 (all $p < 0.01$; Fig. 2b). These correlations show that ToxiPred recovered experimentally measured host-cell liability across tissue-representative systems. UMAP projection of compound DSStox profiles further separated high-liability compounds from low-liability compounds according to average measured DSStox (Fig. 2c).

We next asked whether high host-cell liability could be explained by non-selective chemical toxicity or broad antimicrobial activity. The same validation panel was profiled against four clinical bacterial isolates, *E. coli*, *P. aeruginosa*, *A. baumannii* and *S. aureus*, using bacterial DSS on the same scale. Bacterial activity was driven mainly by canonical antibacterial scaffolds, including clinafloxacin, ciprofloxacin, rifampicin, cefotetan and cefepime (**Suppl. Table 2**). In contrast, several host-toxic anticancer or antiviral reference compounds, including anisomycin, podofilox, flavopiridol, camptothecin and nelfinavir, were bacterially inactive or weakly active in this assay. This counter-screen supports the interpretation that the host-cell DSStox signal reflects eukaryote-specific liability rather than generic chemical toxicity or broad membrane disruption.

These anti-infective results establish the experimental prediction step in the workflow: after public out-of-distribution benchmarking, ToxiPred recovered measured host-cell cytotoxic liability across a chemically diverse compound panel.

3.4. Psychotropic predictions separate out-of-distribution mechanism axes and match organoid responses

Having established benchmark performance and anti-infective cell-line validation, we next asked whether the pathway bottleneck could nominate mechanisms for clinically used neuropsychiatric drugs. We scored 91 approved psychotropic compounds, spanning antipsychotic, antidepressant and related classes, across the 24 tissues using canonical SMILES from DrugBank and PubChem. Approximately 98% of these compounds were absent from the training set and were therefore evaluated from chemical structure alone by the gated DMPNN/Tanimoto chemistry imputer. The panel was structurally distant from the training data, with a mean nearest-neighbour Tanimoto similarity of 0.26 and 83% of compounds below 0.30.

Predicted liabilities were highest among tricyclic antidepressants, phenothiazine and thioxanthene antipsychotics, QT-labile typical antipsychotics and several atypical antipsychotics (Fig. 3a), drug classes with documented cellular or organ-specific liabilities. Lower predicted liabilities were observed for benzodiazepines and Z-drugs, MAOIs, substituted-benzamide antipsychotics and cognition-enhancing agents, for which major clinical risks more commonly reflect receptor pharmacology, drug-drug interactions, dependence or idiosyncratic effects rather than direct cytotoxicity.

We selected six high-liability compounds for experimental follow-up: amitriptyline, aripiprazole, loxapine, quetiapine, sertraline and xanomeline. These compounds spanned tricyclic antidepressant, atypical antipsychotic, SSRI and muscarinic agonist classes, and all exceeded the DSStox confidence threshold of 5.0. Despite similar predicted liability scores, the pathway bottleneck assigned distinct mechanism families. Aripiprazole was assigned a BH3-only mitochondrial apoptosis program, consistent with reported mitochondrial hyperpolarization and oxidative stress in human liver-cell systems [34, 35]. Quetiapine was assigned BH3-only apoptosis, with bile-acid metabolism as a secondary attribution, consistent with reported apoptotic, cholestatic and lipid/sterol liabilities [36–41]. Sertraline was assigned cholesterol biosynthesis, consistent with SSRI-induced SREBP activation and cholesterologenesis in human liver cells [39–41].

Viability assays in human iPSC-derived forebrain, cardiac and retinal organoids showed heterogeneous dose-response profiles across the six high-liability compounds (Fig. 3b). Sertraline, quetiapine and aripiprazole produced the strongest viability loss in cardiac organoids at the higher concentrations of the dose range. Amitriptyline and xanomeline induced partial viability loss at higher concentrations, and loxapine had the weakest effect across organoid types. These data indicate that elevated predicted liability was reflected in measurable cytotoxicity, but with substantial variation in efficacy and tissue sensitivity across compounds. In addition to ATP-based viability, cardiac organoids provided a functional readout through visible spontaneous contractions. Quetiapine and sertraline stopped visible cardiac organoid contractions at 20 μ M under the tested in vitro conditions (**Movies S1-S3**), indicating that predicted cardiac liability was accompanied by an observable functional phenotype. The tested concentration range and simplified dose-derived estimates are provided as exposure context in

Supplementary Fig. S2. We next profiled transcriptomic responses at a sub-cytotoxic concentration of 5 μ M (Fig. 3c). Aripiprazole, sertraline and quetiapine produced clear transcriptional responses in cardiac organoids, with 42, 74 and 46 differentially expressed genes, respectively, at $FDR < 0.05$ and $|\log_2 \text{fold change}| > 1$. By contrast, loxapine and xanomeline produced no differentially expressed genes at the same thresholds. These two compounds also had the lowest averaged predicted DSStox scores among the six and weaker measured viability effects, suggesting that predictions near the lower end of the high-liability range may not yield detectable transcriptional programs at sub-cytotoxic exposure. Subsequent pathway analyses therefore focused on aripiprazole, sertraline and quetiapine.

The three transcriptionally active compounds separated into two dominant response patterns (Fig. 3d,e). Sertraline and quetiapine induced a shared lipid/sterol program in cardiac organoids, with upregulation of cholesterol and mevalonate-pathway genes including HMGCS1, INSIG1, DHCR24, DHCR7, LDLR, SQLE, FDFT1, FDPS, LSS, CYP51A1, TMEM97, FADS1 and FADS2. This response was concordant with the top-ranked cholesterol-biosynthesis attribution for sertraline and the secondary lipid/sterol attribution for quetiapine. Aripiprazole showed a distinct transcriptional profile, with reduced expression of cardiac-associated structural and stress-response genes, including NPPB, CLIC5 and ACTA1, and enrichment of muscle-development and morphogenesis terms in cardiac organoids. Thus, the pathway bottleneck separated a shared lipid/sterol response for sertraline and quetiapine from a distinct aripiprazole-associated cardiac structural program.

Together, these results indicate that the pathway bottleneck can nominate experimentally recoverable response classes among structurally and pharmacologically distinct neuropsychiatric drugs. They also suggest a graded relationship between predicted liability, measured viability loss and transcriptional detectability. Because quetiapine combined measurable cytotoxicity, a robust cardiac lipid/sterol transcriptional response, model-predicted apoptotic and bile-acid attributions, and documented hepatobiliary injury, we selected it for deeper multi-omics characterization across all three organoid systems.

3.5. Quetiapine multi-omics supports a lipid/sterol liability axis

To resolve the quetiapine response across molecular layers, we used a factorial perturbation design in forebrain, cardiac and retinal organoids. Organoids were treated with quetiapine, retinoic acid or the combination, alongside vehicle controls. Retinoic acid was included as a defined developmental and metabolic perturbation, providing a common transcriptionally responsive background against which quetiapine-specific effects could be isolated. Accordingly, the primary contrast compared quetiapine plus retinoic acid with retinoic acid alone, while the single-treatment arms were used to determine whether the resulting program was driven by quetiapine, retinoic acid or their interaction.

This contrast identified 27 differentially expressed genes, of which 23 were downregulated and 4 were upregulated ($FDR < 0.05$; $|\log_2 \text{fold change}| > 0.5$). Transcriptomic heatmaps showed tissue-dependent

remodelling of retinoic-acid-associated expression programs (Fig. 4a), with the largest changes dominated by suppression of apolipoproteins, hepatic-secretory factors, xenobiotic-metabolism genes and coagulation-associated transcripts, including APOA1, APOA2, APOB, APOC3, AFP, ALB, AMBP, FABP1, SERPINA1, CYP3A7, GSTA1, FGB, FGG and F2.

Condition-level decomposition showed that quetiapine alone was sufficient to suppress these genes, whereas retinoic acid alone produced minimal or opposite effects (Fig. 4b). Thus, the combination contrast did not simply reflect retinoic-acid signalling, but captured a quetiapine-driven suppression program revealed on the retinoic-acid background.

Pathway enrichment analysis linked this suppression program to lipid transport and hepatobiliary-associated functions. The most affected terms included triglyceride-rich lipoprotein remodelling, VLDL remodelling, cholesterol homeostasis, sterol homeostasis, cholesterol metabolism, PPAR signalling and coagulation cascades (Fig. 4c). These enrichments were consistent with the coordinated downregulation of apolipoproteins and secretory factors and suggested that quetiapine perturbs lipid handling not only through sterol-biosynthetic activation, as observed in cardiac organoids, but also through suppression of lipoprotein-transport programs.

We next asked whether these transcriptional changes were accompanied by metabolic remodelling. Metabolomic profiling showed that quetiapine altered central carbon and mitochondrial-associated metabolites, including reduced fumarate and malate, together with broad decreases in short- and medium-chain acylcarnitines such as isovalerylcarnitine, hexanoylcarnitine and decanoylcarnitine (Fig. 4d). This pattern is consistent with impaired mitochondrial fatty-acid handling, although direct flux measurements would be required to establish the precise biochemical step affected. Quetiapine also altered tryptophan metabolism and reduced ornithine within the arginine-ornithine axis.

These metabolite changes were accompanied by transcriptional changes in genes involved in carbon flux, mitochondrial function and amino-acid metabolism, including ME1, GOT1, LDHB, BCS1L, CLPP, ARG2 and ASL. Thus, quetiapine produced a coordinated multi-omics response involving lipid transport, sterol homeostasis, mitochondrial metabolism and amino-acid pathways.

Together, these data extend the pathway-bottleneck prediction from a single transcriptomic signature to a multi-layer organoid response. In the psychotropic screen, quetiapine was associated with apoptotic, bile-acid and lipid/sterol attributions; in organoids, it induced cholesterol/mevalonate genes, suppressed apolipoprotein and lipoprotein-remodelling programs, and altered metabolites linked to mitochondrial fatty-acid and central carbon metabolism. These findings support a lipid/sterol-centred liability axis for quetiapine while also indicating broader effects on energy metabolism.

4. Discussion

Drug toxicity is often modeled as a property of a compound, yet adverse cellular liability is shaped by the tissue in which a perturbation occurs. ToxiPred was designed around this principle. It combines

chemical, target and perturbation evidence for each drug with expression and essentiality evidence for each tissue, and represents their interaction through a 70-pathway bottleneck. The central advantage is that the same bottleneck is used for prediction and attribution: each DSStox estimate is accompanied by a ranked set of pathway-level drug-tissue interactions. This makes the model different from approaches in which toxicity scores are generated first and biological explanations are added afterwards.

Under drug-disjoint scaffold cross-validation, ToxiPred predicted tissue-resolved DSStox with Pearson $r = 0.73$, and the final pathway-bottleneck architecture outperformed linear, boosting, factorization and gene-level alternatives, and matched alternative pathway-interaction operators while retaining direct pathway-wise attribution. These comparisons suggest that the pathway bottleneck is not simply an interpretability constraint, but a useful inductive bias for generalization to unseen chemical scaffolds. The result is important because drug-toxicity prediction often faces the hardest deployment setting: predicting liabilities for compounds whose exact chemical series, perturbation profiles and tissue responses were not observed during training.

A second requirement for practical deployment is the ability to score compounds that lack measured perturbation data. ToxiPred addresses this by combining DMPNN-imputed perturbation evidence with Tanimoto-nearest-neighbour evidence through a similarity gate. Public benchmark validation supported this deployment mode. Liver-routed DSStox separated active from inactive compounds across ToxCast/Tox21 HepG2 DNA-damage, apoptosis and steatosis endpoints, whereas cross-tissue maximum DSStox transferred to the compound-level ClinTox CT_TOX label. These benchmarks should not be read as prospective clinical validation. Rather, they show that a viability-derived, tissue-resolved score carries toxicity-relevant signal in independent endpoint families. The fixed confidence rule used here, with DSStox below 3.0 considered low-liability, DSStox above 5.0 considered high-liability and intermediate values withheld as borderline, reflects a triage use case. ToxiPred is therefore best viewed as an abstention-aware prioritization tool rather than as a full-coverage binary classifier.

The anti-infective validation provided the first direct experimental test of this prioritization logic. Anti-infective compounds are selected for activity against microbial or viral biology, not mammalian host-cell pathways, making them a stringent test of whether the model can distinguish therapeutic selectivity from host-cell liability. Across nine cell lines representing six tissue systems, predicted and measured DSStox were concordant, with tissue-wise correlations of $R = 0.78$ to 0.84 . Parallel bacterial profiling further separated antimicrobial activity from mammalian cytotoxicity: canonical antibacterial scaffolds drove bacterial DSS, whereas several host-toxic reference compounds showed little bacterial activity. This supports the interpretation that the host-cell DSStox signal reflects eukaryotic cellular liability rather than broad antimicrobial potency, generic chemical reactivity or non-specific membrane disruption.

The psychotropic analysis tested a complementary question: whether pathway attributions could nominate mechanisms for clinically used drugs that were largely outside the training set. Approximately 98% of the 91 approved psychotropic compounds were absent from model training and were scored from chemical structure alone. The highest predicted liabilities occurred in drug classes with

documented cellular or organ-specific liabilities, including tricyclic antidepressants, phenothiazine and thioxanthene antipsychotics, QT-labile typical antipsychotics and several atypical antipsychotics. Organoid follow-up showed measurable viability loss across the selected high-liability compounds, but also revealed substantial differences in potency and tissue sensitivity. This distinction is important: elevated predicted liability was not equivalent to a uniform experimental phenotype, but instead prioritized compounds for which dose, tissue context and molecular readout could be interrogated.

The cardiac organoid contraction phenotype is particularly important because it provides a functional readout that is independent of ATP-based viability and RNA-seq. Quetiapine and sertraline stopped visible cardiac organoid contractions at 20 μ M, indicating acute functional cardiotoxicity under the tested in vitro conditions. This strengthens the conclusion that the psychotropic validation captures biologically meaningful cardiac liability rather than only molecular perturbation.

Transcriptomic profiling sharpened this interpretation. At sub-cytotoxic exposure, aripiprazole, sertraline and quetiapine produced clear cardiac transcriptional responses, whereas loxapine and xanomeline were transcriptionally inert at the same thresholds. The active compounds separated into two response patterns. Sertraline and quetiapine induced a shared cholesterol and mevalonate program in cardiac organoids, consistent with the cholesterol-biosynthesis attribution for sertraline and the lipid/sterol component of the quetiapine attribution. Aripiprazole, by contrast, showed a distinct cardiac profile involving reduced expression of structural and stress-associated genes and enrichment of muscle-development programs. Thus, the pathway bottleneck did more than rank compounds by global cytotoxicity. It separated structurally and pharmacologically distinct psychotropic drugs into experimentally recoverable response classes.

The quetiapine multi-omics experiment extended this mechanism recovery beyond a single transcriptional signature. Retinoic acid was used as a defined developmental and metabolic perturbation to create a common transcriptionally responsive background, and the quetiapine plus retinoic acid versus retinoic acid contrast isolated the incremental quetiapine response. Condition-level decomposition showed that the dominant suppression program was driven primarily by quetiapine rather than by retinoic acid alone. This program involved apolipoproteins, secretory factors, xenobiotic-metabolism genes and coagulation-associated transcripts, with enrichment of triglyceride-rich lipoprotein remodelling, VLDL remodelling, cholesterol and sterol homeostasis, PPAR signalling and coagulation cascades. Metabolomics added changes in central carbon metabolites, acylcarnitines, tryptophan-related metabolites and the arginine-ornithine axis. Together with the cardiac cholesterol and mevalonate response, these data support a lipid/sterol-centred quetiapine liability axis spanning sterol biosynthesis, lipoprotein handling and mitochondrial-associated metabolism.

This quetiapine result should be interpreted with appropriate caution. The organoid systems used here are forebrain, cardiac and retinal, not liver organoids. Therefore, the suppression of apolipoprotein, secretory and xenobiotic-associated genes does not constitute a direct model of hepatic injury. Rather, it provides a human organoid readout of lipid-transport and hepatobiliary-associated biology that is

consistent with the model attribution and with independent reports of quetiapine-associated metabolic and hepatobiliary liability. Similarly, the acylcarnitine changes indicate altered mitochondrial fatty-acid handling, but direct flux measurements would be required to establish impaired fatty-acid oxidation as the causal biochemical mechanism.

More broadly, tissue-resolved and mechanism-linked predictions may help prioritize compounds for experimental safety testing and pharmacovigilance follow-up. Adverse effects of widely used medicines, including psychotropic drugs, can create regulatory, clinical and legal questions about whether emerging risks were recognized, communicated and monitored appropriately[42, 43]. Quetiapine provides a well-known example of a drug with extensive clinical use, including off-label use, and a history of legal and regulatory scrutiny related to marketing practices and metabolic safety concerns. In such settings, adverse-event reports are often insufficient, whereas objective mechanistic evidence can help clarify whether a claimed toxicity is biologically plausible, tissue-relevant and consistent with independent experimental data.

Several limitations follow from the design of the study. DSStox is derived from viability assays and therefore captures cellular liability rather than patient-level adverse drug reactions. The supervised training data come from public pharmacogenomic resources that are enriched for cancer cell lines and acute drug-response measurements. ToxiPred may therefore underrepresent mechanisms that require mature tissue architecture, immune context, long-term exposure, organism-level pharmacokinetics, reactive metabolites generated in non-modeled tissues or idiosyncratic immune-mediated injury. The 70-pathway dictionary improves interpretability but necessarily compresses biology, and mechanisms absent from the dictionary cannot be directly attributed. For compounds without measured perturbation profiles, predictions depend on imputed LINCS/Tahoe-like evidence, and confidence will vary with chemical distance from measured compounds and with the quality of the imputation.

The experimental validation also remains targeted rather than exhaustive. The anti-infective panel tested host-cell liability across multiple cell-line systems, but immortalized cell lines cannot fully reproduce primary tissue physiology. The psychotropic validation focused on six selected high-liability compounds and one deeper multi-omics case study. The organoid experiments used defined exposure times and concentrations, which may miss slower adaptive responses, delayed toxicity or concentration regimes closer to chronic clinical exposure. Cardiac contraction was monitored as a visible functional phenotype, but future work should quantify beat rate, contraction amplitude, calcium flux, electrophysiology and recovery after washout. These limitations do not weaken the central conclusion, but they define the appropriate scope of use: ToxiPred is a mechanism-guided prioritization framework, not a replacement for pharmacokinetic, animal, clinical or long-term safety studies.

In summary, ToxiPred links tissue-resolved toxicity prediction with pathway-level mechanism attribution in a single reproducible workflow. Across public benchmarks, anti-infective validation and psychotropic organoid profiling, the model prioritized compounds with measurable tissue-specific liability and generated pathway hypotheses that were experimentally recoverable in human cellular systems. The

strongest psychotropic validation came from the convergence of predicted high liability, cardiac organoid vulnerability, contraction arrest, cholesterol/mevalonate transcriptional activation and quetiapine-associated metabolic remodeling. These findings support ToxiPred as a mechanism-guided prioritization framework for identifying biologically plausible tissue-specific liabilities, while emphasizing that its predictions should guide, not replace, pharmacokinetic, clinical and long-term safety studies. As perturbation atlases continue to expand, replacing imputed evidence with measured profiles should further improve chemical coverage, confidence estimation and utility for pharmacovigilance, regulatory review and expert assessment.

Declarations

Competing interests

The authors declare no competing interests.

Funding

This research was funded by Helse Midt-Norge, Felles forskningsutvalg (FFU) grant Nr. 101274100, Helse Sør-Øst Nr. 342266, and NTNU Health grant Nr. 980001250; and by RCN grant Nr. 327005 (to M.B.). FIMM High Throughput Biomedicine unit is supported by the University of Helsinki and Biocenter Finland.

Author contributions

Conceptualization and project administration, A.I.; methodology, A.I., E.R., J.-O.S., M.C.-Q., R.R.S., I.E.O.N., D.E.K. and collaborators; software, A.I.; validation, E.R., I.E.O.N., R.R.S., E.W. and D.E.K.; data curation, A.I. and co-authors; writing - original draft preparation, A.I., D.E.K. and E.R.; interpretation of results, all authors; writing - review and editing, all authors; visualization, A.I.; supervision, M.B. and D.E.K.; funding acquisition, D.E.K. and M.B. All authors read and approved the final manuscript.

Acknowledgements

We thank Toveann Ahlnäs from the FIMM High Throughput Biomedicine Unit for support with drug screening.

Data availability

Raw and processed RNA-seq data for the cardiac and retinal organoid arms are available from GEO under accession GSE334500. Metabolomics tables, dose-response measurements, model predictions, public-benchmark tables, tissue-specificity residual analyses and curated known-lineage recovery tables are provided as Supplementary Data.

Code availability

ToxiPred source code, trained model weights, and Docker container are available at <https://github.com/lanevskiAleksandr/ToxiPred>. An interactive web interface for querying the precomputed atlas and submitting new SMILES is available at <https://toxipred.com>.

References

1. Lazarou J, Pomeranz BH, Corey PN (1998) Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA* 279:1200–1205
2. Pirmohamed M, James S, Meakin S et al (2004) Adverse drug reactions as cause of admission to hospital. *BMJ* 329:15–19
3. Onakpoya IJ, Heneghan CJ, Aronson JK (2016) Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. *Crit Rev Toxicol* 46:477–489
4. Mayr A, Klambauer G, Unterthiner T, Hochreiter S (2016) DeepTox: toxicity prediction using deep learning. *Front Environ Sci* 3:80
5. Pu L, Naderi M, Liu T, Wu HC, Mukhopadhyay S, Brylinski M (2019) eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol Toxicol* 20:2
6. Liu Q, Hu Z, Jiang R, Zhou M (2020) DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 36:i871–i878
7. Kuenzi BM, Park J, Fong SH, Sanchez KS, Lee J, Kreisberg JF et al (2020) Predicting drug response and synergy using a deep learning model of human cancer cells. *Cancer Cell* 38:672–684e6
8. Subramanian A, Narayan R, Corsello SM et al (2017) A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171:1437–1452e17
9. Ghandi M, Huang FW, Jane-Valbuena J et al (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 569:503–508
10. Meyers RM, Bryan JG, McFarland JM et al (2017) Computational correction of copy-number effect improves specificity of CRISPR-Cas9 essentiality screens. *Nat Genet* 49:1779–1784
11. Dempster JM, Pacini C, Pantel S, Behan FM, Green T, Krill-Burger J et al (2019) Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nat Commun* 10:5817
12. Gaulton A, Hersey A, Nowotka M et al (2017) The ChEMBL database in 2017. *Nucleic Acids Res* 45:D945–D954

13. Zhang J, Ubas AA, de Borja R, Svensson V, Thomas N, Thakar N et al (2025) Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. *bioRxiv*. 10.1101/2025.02.20.639398
14. Feizi N, Martin B, Gendoo DMA, Krzykalla J, Ghosh S, Hozumi Y et al (2022) PharmacDB 2.0: improving scalability and transparency of in vitro pharmacogenomics analysis. *Nucleic Acids Res* 50:D1348–D1357. 10.1093/nar/gkab1084
15. Yadav B, Pemovska T, Szwajda A et al (2014) Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies. *Sci Rep* 4:5193
16. Potdar S, Ianevski A, Mpindi JP, Bychkov D, Fiere C, Ianevski P et al (2020) Breeze: an integrated quality control and data analysis application for high-throughput drug screening. *Bioinformatics* 36:3602–3604
17. Potdar S, Ianevski F, Ianevski A, Tanoli Z, Wennerberg K, Seashore-Ludlow B, Kallioniemi O, Östling P, Aittokallio T, Saarela J (2023) Breeze 2.0: an interactive web-tool for visual analysis and comparison of drug response data. *Nucleic Acids Res* 51:W57–W61
18. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. *J Chem Inf Model* 50:742–754
19. Landrum G (2016) RDKit: Open-source cheminformatics software
20. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H et al (2019) Analyzing learned molecular representations for property prediction. *J Chem Inf Model* 59:3370–3388
21. Huang R, Xia M, Nguyen DT et al (2016) Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. *Front Environ Sci* 3:85
22. Richard AM, Judson RS, Houck KA, Grulke CM, Volarath P, Thillainadarajah I et al (2016) ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem Res Toxicol* 29:1225–1251
23. Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, Coley CW, Xiao C, Sun J, Zitnik M (2021) Therapeutics Data Commons: machine learning datasets and tasks for drug discovery and development. *NeurIPS Datasets and Benchmarks*
24. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS et al (2018) MoleculeNet: a benchmark for molecular machine learning. *Chem Sci* 9:513–530
25. Xiang Y, Tanaka Y, Patterson B et al (2018) Generation and fusion of human cortical and medial ganglionic eminence brain organoids. *Curr Protoc Stem Cell Biol* 47:e61
26. Nakano T, Ando S, Takata N et al (2012) Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* 10:771–785
27. Hofbauer P, Jähnel SM, Papai N et al (2021) Cardioids reveal self-organizing principles of human cardiogenesis. *Cell* 184:3299–3317e22
28. Dobin A, Davis CA, Schlesinger F et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21

29. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for genomic datasets with biomaRt. *Nat Protoc* 4:1184–1187
30. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential-expression analysis. *Bioinformatics* 26:139–140
31. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16:284–287
32. Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach. *Proc Natl Acad Sci USA* 102:15545–15550
33. Ritchie ME, Phipson B, Wu D et al (2015) limma powers differential-expression analyses. *Nucleic Acids Res* 43:e47
34. Pirc Marolt T, Kramar B, Vovk A, Podgornik H, Šuput D, Milisav I (2023) Therapeutic dosage of antipsychotic drug aripiprazole induces persistent mitochondrial hyperpolarisation, moderate oxidative stress in liver cells, and haemolysis. *Antioxidants* 12:1930
35. Kramar B, Pirc Marolt T, Yilmaz Göler AM, Monsalve M (2024) Aripiprazole, but not olanzapine, alters the response to oxidative stress in Fao cells by reducing the activation of mitogen-activated protein kinases and promoting cell survival. *Int J Mol Sci* 25:11119
36. National Institute of Diabetes and Digestive and Kidney Diseases. Quetiapine. In: *LiverTox: Clinical and Research Information on Drug-Induced Liver Injury*. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases (2012) Updated 5 June 2023
37. Wright TM, Vandenberg AM (2007) Risperidone- and quetiapine-induced cholestasis. *Ann Pharmacother* 41:1518–1523
38. Klinke JA, Shapira SC, Akbari E, Holmes DT (2010) Quetiapine-associated cholestasis causing lipoprotein-X and pseudohyponatraemia. *J Clin Pathol* 63:741–743
39. Raeder MB, Fernø J, Vik-Mo AO, Steen VM (2006) SREBP activation by antipsychotic- and antidepressant-drugs in cultured human liver cells: relevance for metabolic side-effects? *Mol Cell Biochem* 289:167–173
40. Panizzutti B, Bortolasci CC, Spolding B, Kidnapillai S, Connor T, Richardson MF et al (2021) Biological mechanism(s) underpinning the association between antipsychotic drugs and weight gain. *J Clin Med* 10:4095
41. Laressergues E, Staels B, Valeille K, Majd Z, Hum DW, Duriez P, Cussac D (2010) Antipsychotic drug action on SREBPs-related lipogenesis and cholesterogenesis in primary rat hepatocytes. *Naunyn Schmiedeberg's Arch Pharmacol* 381:427–439
42. American Diabetes Association, American Psychiatric Association, American Association of Clinical Endocrinologists, North American Association for the Study of Obesity. Consensus development conference on antipsychotic drugs and obesity and diabetes. *Diabetes Care* (2004) ;27:596–601
43. Tanne JH (2010) AstraZeneca pays \$520m fine for off-label marketing. *BMJ* 340:c2380

Figures

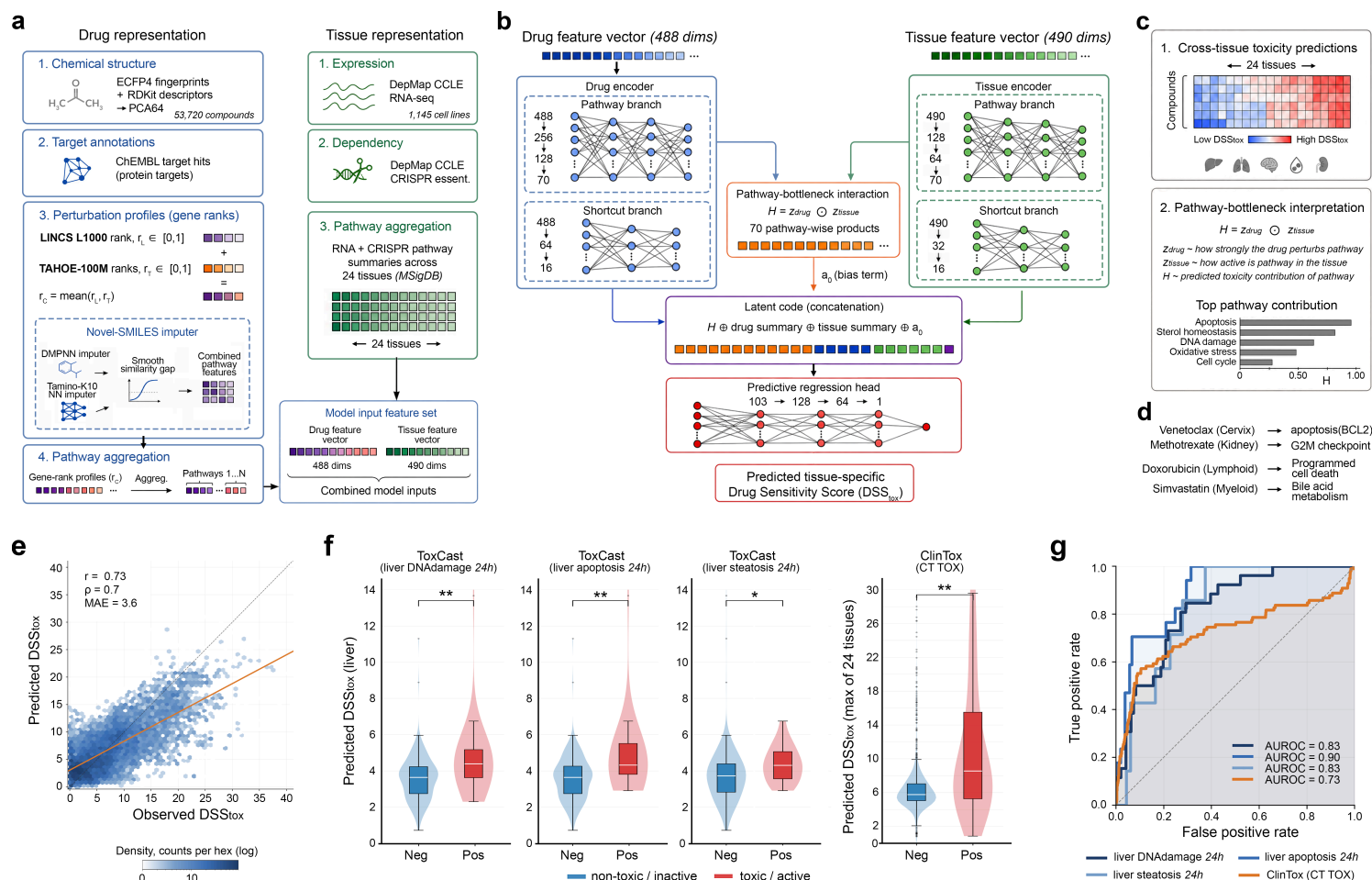


Figure 1

ToxiPred architecture, pathway bottleneck and public transfer. **(a)** Input representations. Drug features (left) combine ECFP4 fingerprints and RDKit descriptors reduced by PCA to 64 dimensions, ChEMBL target annotations, measured LINCS L1000 and Tahoe-100M perturbation gene-rank profiles where available, and DMPNN- and Tanimoto-K10-imputed pathway features for SMILES-only compounds combined through a smooth similarity gate. Tissue features (right) are derived from DepMap CCLE RNA expression and CRISPR essentiality data summarized over the same 70-pathway dictionary across 24 tissues. The two vectors (488 and 490 dimensions) form the model inputs. **(b)** Dual-encoder architecture. Drug and tissue encoders each contain a pathway branch (projecting inputs into a shared 70-dimensional pathway space) and a parallel shortcut branch (16-dimensional residual). The element-wise Hadamard product $H = z_{\text{drug}} \odot z_{\text{tissue}}$ of the two pathway branches forms the pathway-bottleneck interaction (70 pathway-wise products). H is concatenated with the drug and tissue shortcuts and a scalar bias into a 103-dimensional latent code, which is decoded by a regression head to one DSS_{tox} value per drug-tissue pair. **(c)** Two coupled outputs per compound. Top: predicted DSS_{tox} across 24 tissues, displayed as a compounds × tissues heatmap. Bottom: pathway-bottleneck interpretation. The same H vector ranks the 70 pathways by their contribution to that prediction; representative top contributors are shown. **(d)** Canonical-pharmacology sanity checks. **(e)** Drug-disjoint scaffold cross-validation. Density hexbin of predicted versus measured DSS_{tox} across 9,275 drug-tissue labels

(Pearson $r = 0.73$, Spearman $\rho = 0.70$, MAE = 3.6). Diagonal indicates perfect prediction; orange line is the linear fit. **(f)** Out-of-distribution discrimination on public benchmarks. Violin and box plots of predicted DSStox in non-toxic/inactive (Neg, blue) versus toxic/active (Pos, red) compounds across three U.S. EPA ToxCast HepG2 liver endpoints at 24 h (DNA damage, apoptosis, steatosis; liver-tissue DSStox) and the ClinTox CT_TOX clinical-trial toxicity benchmark (cross-tissue maximum DSStox across 24 tissues). Two-sided Wilcoxon rank-sum test; * $p < 0.01$, ** $p < 0.001$. **(g)** ROC curves for the same four benchmarks after applying the fixed confidence rule (DSStox < 3.0, confident low-liability; DSStox > 5.0, confident high-liability; intermediate scores 3.0-5.0 flagged as borderline and excluded). AUROC: liver DNA damage 24 h, 0.83; liver apoptosis 24 h, 0.90; liver steatosis 24 h, 0.83; ClinTox CT_TOX, 0.73. Dashed diagonal indicates chance performance.

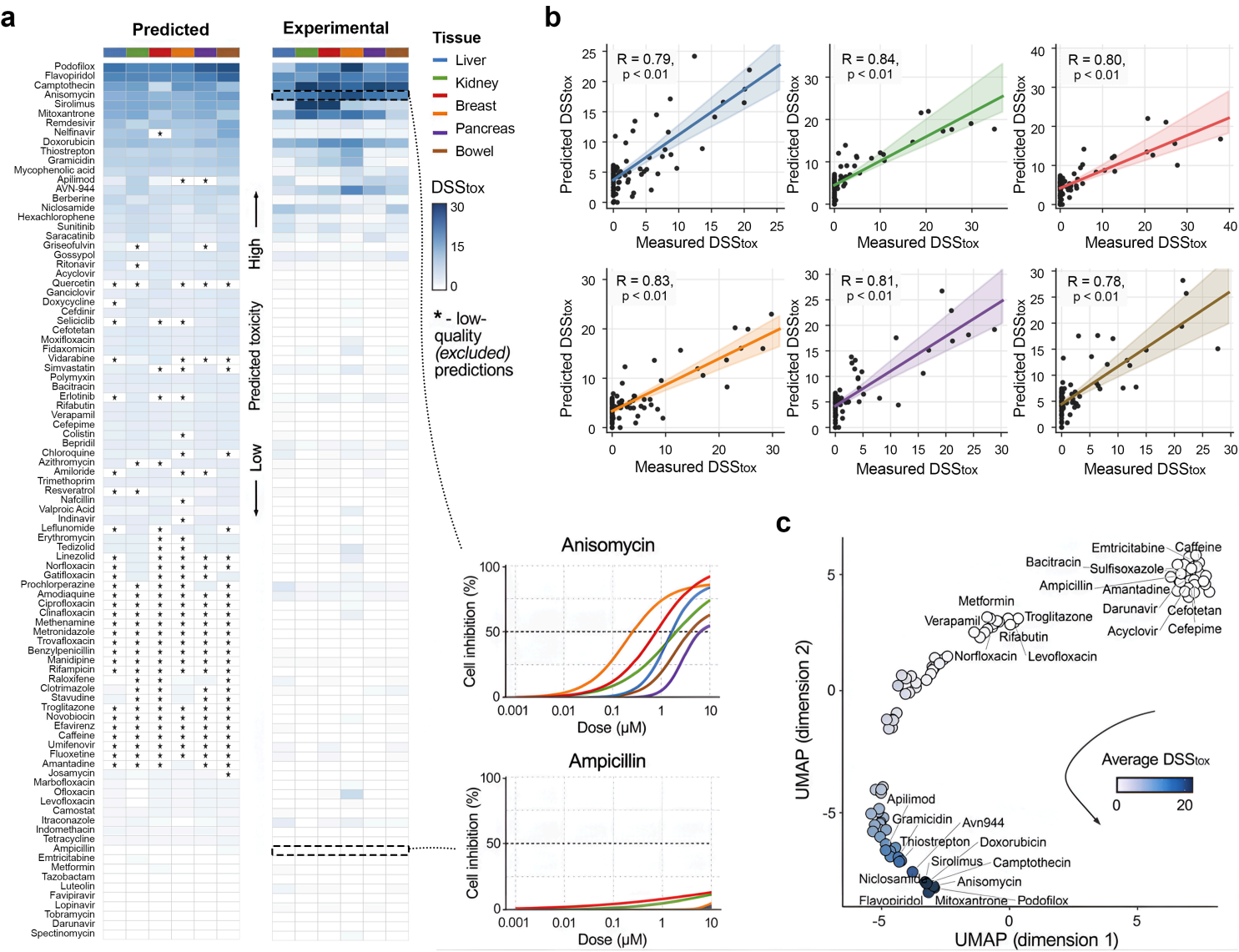


Figure 2

Host-cell cytotoxic liability validation and antimicrobial selectivity control. (a)Side-by-side heatmaps of predicted (left) and experimentally measured (right) DSStox values across six tissue-representative systems: lung (A549), liver (HepG2 and Hep3B), large intestine (HCT116), kidney (Caki-1 and HEK293),

pancreas (MiaPaCa-2) and breast (HDQ-P1 and DU4475). For tissues with two cell lines, measured DSStox values are shown as the within-tissue average. Compounds are ordered top-to-bottom from high to low predicted host-cell liability. Asterisks mark compounds excluded from downstream interpretation due to low-quality dose-response fits. Inset dose-response curves show a predicted high-liability compound (anisomycin) and a predicted low-liability compound (ampicillin) as reference cases. **(b)** Tissue-wise correlations between predicted DSStox and tissue-averaged measured DSStox, one panel per tissue. Each point is one compound; Pearson R and two-sided p values are shown in each panel. Diagonal line indicates perfect agreement. **(c)** UMAP projection of the 93-compound panel based on cross-tissue predicted DSStox profiles, with points colored by average measured DSStox. High- and low-liability compounds separate in the embedding, indicating that predicted tissue profiles carry compound-level liability information beyond a single aggregate score.

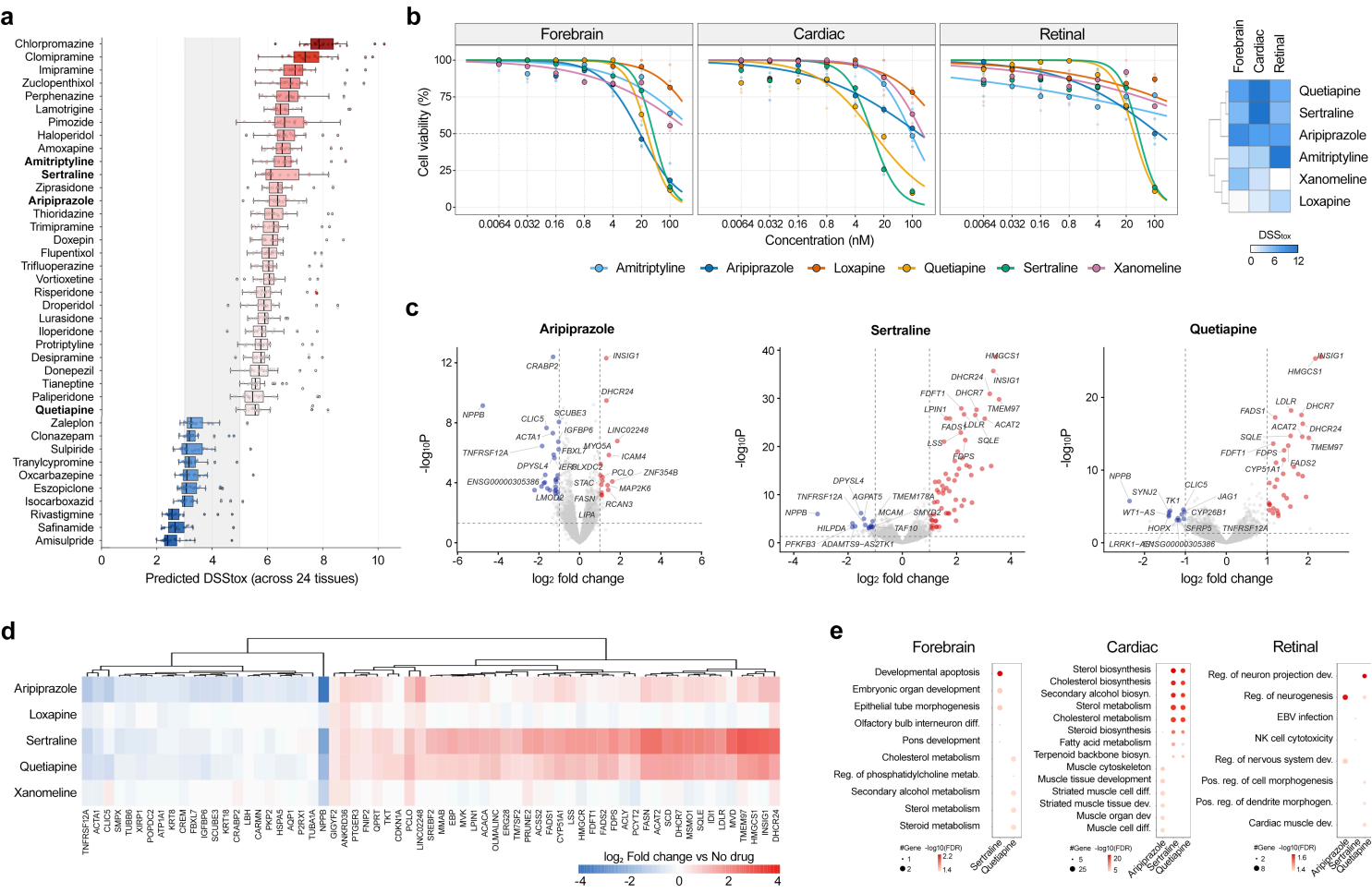


Figure 3

Psychotropic prioritization and organoid screen. **(a)** Predicted DSStox across 24 ToxiPred tissues for the highest- and lowest-ranked compounds of the 91-compound approved psychotropic panel. For clarity the figure displays the 29 highest- and 10 lowest-ranked compounds by median predicted DSStox. The compounds selected for organoid validation are shown in bold. **(b)** Dose-response viability curves for amitriptyline, aripiprazole, loxapine, quetiapine, sertraline and xanomeline in forebrain, cardiac and retinal organoids (5-fold serial dilutions, 0.0064-100 μ M, 72 h). Inset heatmap summarizes DSStox values

across the three organoid systems. (c) Volcano plots of differential expression in cardiac organoids treated for 24 h at 5 μ M with aripiprazole, sertraline or quetiapine, against DMSO controls (edgeR quasi-likelihood pipeline, FDR < 0.05 and $|\log_2$ fold change| > 1; Methods 2.11). Red, upregulated; blue, downregulated; grey, not significant. (d) Heatmap of top drug-regulated genes in cardiac organoids, shown as $\log_2$ fold change relative to DMSO controls. (e) Pathway-enrichment analysis (clusterProfiler enrichGO Biological Process and enrichKEGG, $q < 0.1$) across forebrain, cardiac and retinal organoids. Dot size, gene count; color, $-\log_{10}$ adjusted P value. Sertraline and quetiapine converge on sterol and cholesterol biosynthesis programs in cardiac organoids; aripiprazole enriches muscle-development programs in cardiac organoids and neurodevelopmental programs in retinal organoids.

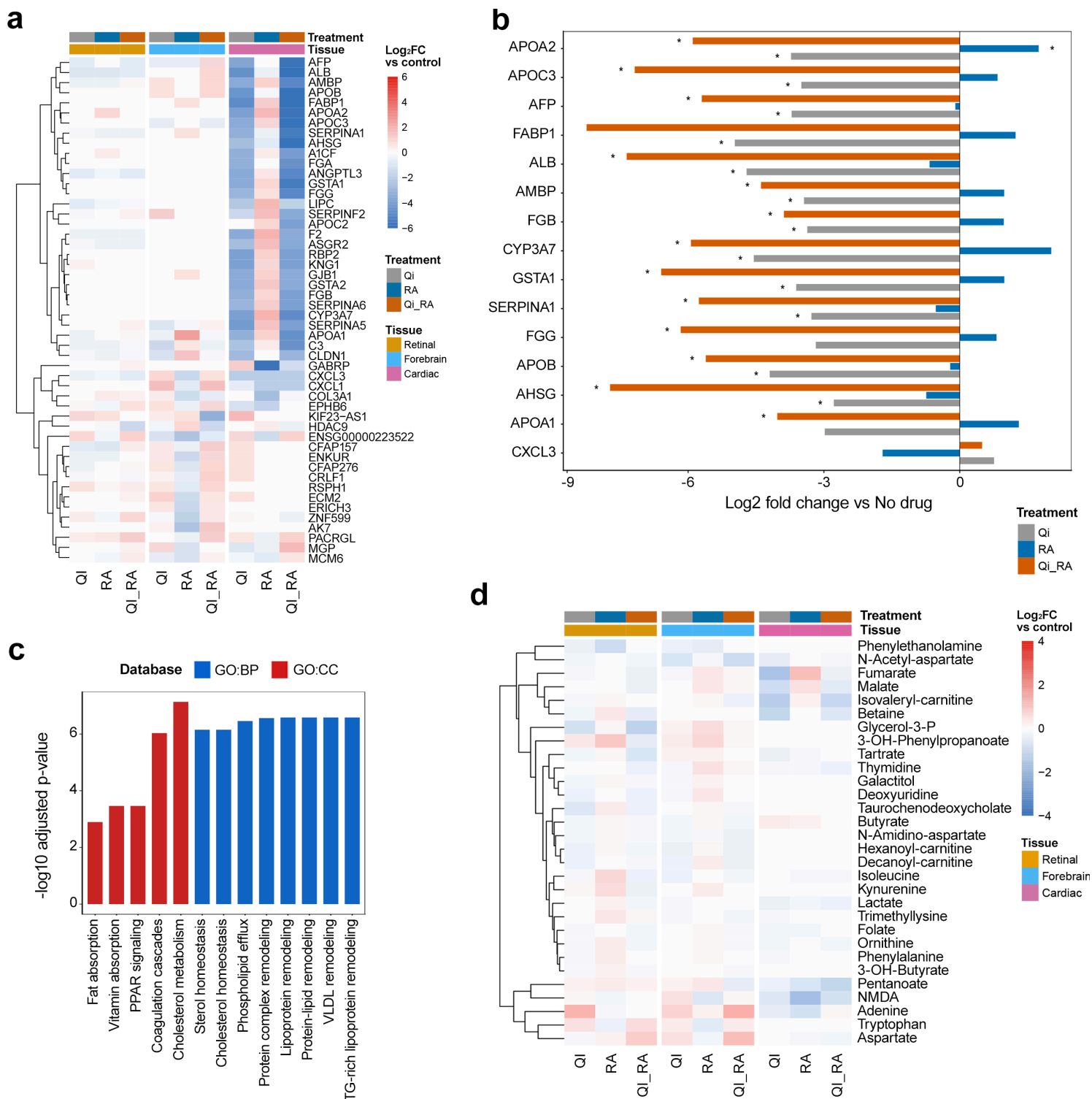


Figure 4

Quetiapine multi-omics follow-up of the predicted lipid/sterol attribution. (a) Transcriptomic heatmap of log2 fold changes versus untreated controls (Qi, quetiapine 5 μ M; RA, retinoic acid 5 μ M; Qi_RA, combined treatment, all 24 h) in retinal, forebrain and cardiac organoids. Rows are top-ranked genes selected by absolute log2 fold-change difference between the Qi_RA and RA conditions, hierarchically clustered (Euclidean distance, complete linkage); columns are tissue $\times$ treatment combinations. Color scale, log2 fold change versus control. (b) Per-gene decomposition of the top-suppressed cardiac

transcripts from (a), showing log₂ fold change versus untreated control for each of the three treatments (Qi, RA, Qi_RA). For every gene shown, the Qi-alone (grey) and Qi_RA (orange) bars are large and concordant in direction, whereas the RA-alone (blue) bar is small or oppositely signed. This indicates that the Qi_RA versus RA contrast in (a) is driven by quetiapine, not by retinoic acid. Asterisks denote significant effects (FDR < 0.05). **(c)** Pathway enrichment of the top-ranked cardiac transcriptomic response from (a) and (b). Bars show $-\log_{10}$ adjusted P value; color denotes annotation source (GO Biological Process, blue; GO Cellular Component, red). **(d)** Metabolomic heatmap of log₂ fold changes versus untreated controls across the same three organoid types and three treatments, for the top-ranked metabolites selected and clustered as in (a). Color scale, log₂ fold change versus control.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ToxiPredSuppl.docx](#)