From Animal to Algorithm: in Silico NAMs at FDA/NCTR

Weida Tong, Ph.D
Division of Bioinformatics and Biostatistics
National Center for Toxicological Research, FDA

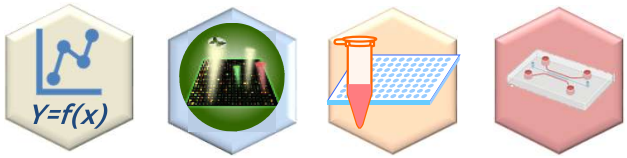
Disclaimer: The views expressed are of the presenter and should not be considered as the official position or policy of U.S. FDA.

Drug-Induced Liver Injury (DILI)

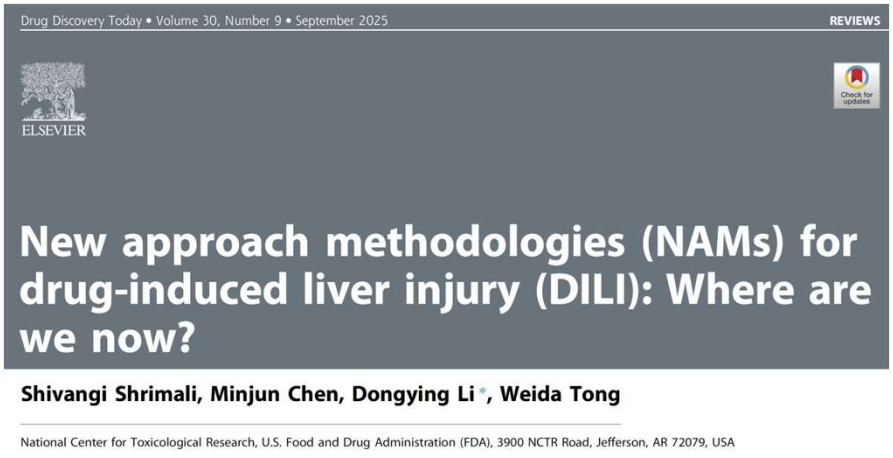
Close to 50% of DILI cases are not detected in preclinical studies

The current preclinical models are not effective

Exploring alternative approaches (NAMs) for improved DILI prediction



AI/ML ***Genomics*** ***In vitro*** ***MPS***



Drug Discovery Today • Volume 30, Number 9 • September 2025 REVIEWS

ELSEVIER

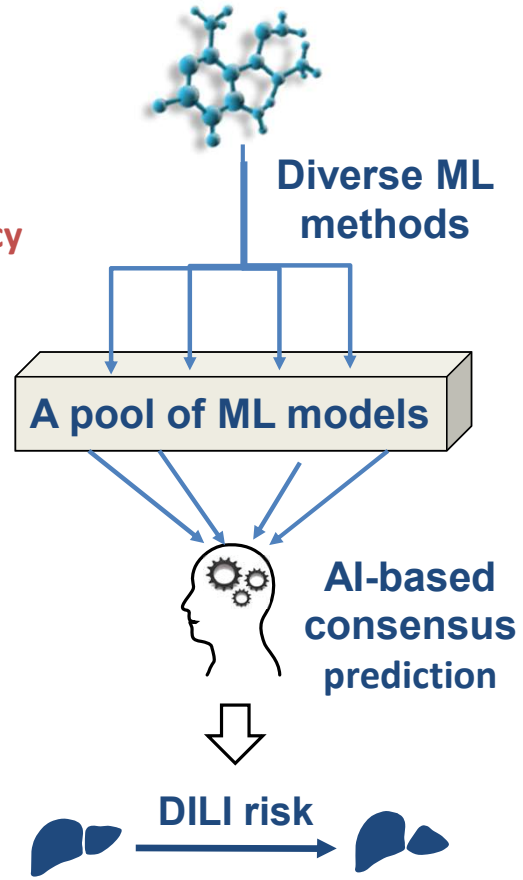
New approach methodologies (NAMs) for drug-induced liver injury (DILI): Where are we now?

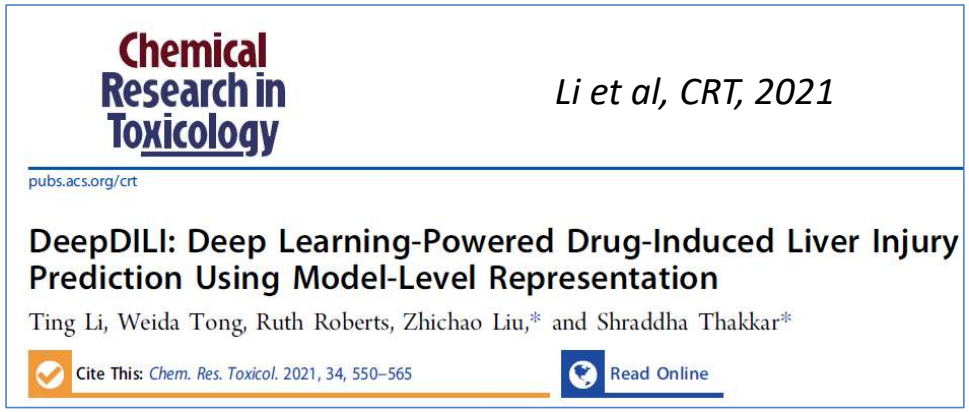
Shivangi Shrimali, Minjun Chen, Dongying Li *, Weida Tong

National Center for Toxicological Research, U.S. Food and Drug Administration (FDA), 3900 NCTR Road, Jefferson, AR 72079, USA

DeepDILI: an AI Model for DILI Assessment

- DeepDILI was developed with a **novel DL approach** that integrates multiple ML models in a deep learning architecture:
  - A binary classification: DILI positive or negative
  - Trained with 753 drugs approved by FDA before 1997
  - Evaluated on 249 drugs approved after 1997 with **~70% prediction accuracy**
  - It is in-house algorithms, thus easy to be implemented in the FDA environment (is available now in RAPID and PrecisionFDA)
  - It has been independently reproduced by NIEHS




Chemical Research in Toxicology

Li et al, CRT, 2021

pubs.acs.org/crt

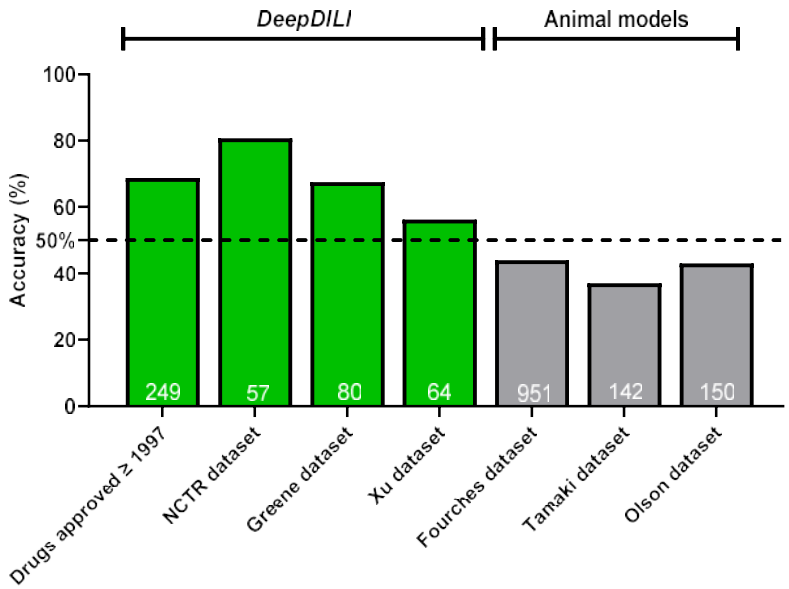
DeepDILI: Deep Learning-Powered Drug-Induced Liver Injury Prediction Using Model-Level Representation

Ting Li, Weida Tong, Ruth Roberts, Zhichao Liu,* and Shraddha Thakkar*

Cite This: *Chem. Res. Toxicol.* 2021, 34, 550–565 Read Online

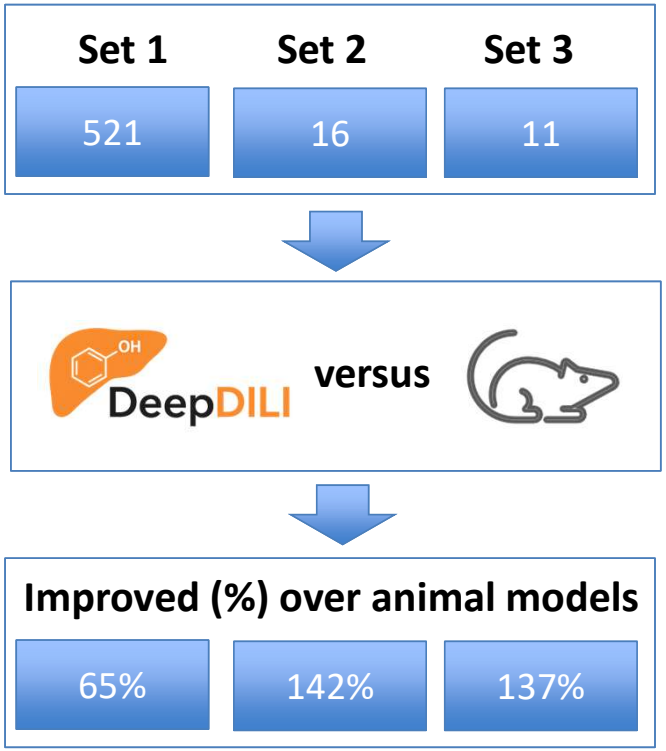
DeepDILI vs Animal Models for DILI Prediction

- **Animal models for DILI in humans**: a survey (1961–present) indicates that rat models demonstrate less than 50% accuracy.
- **DeepDILI for DILI in humans:** ~70% :



Prediction accuracy was reported as described in the original publication of each study

Animal vs DeepDILI: Comparative analyses on three sets of compounds that were not used in DeepDILI development

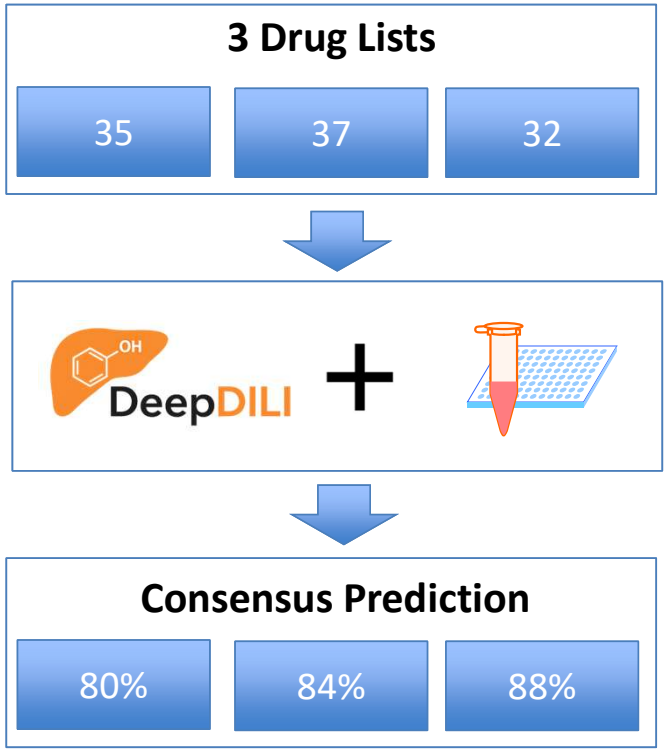


Will combining DeepDILI and in vitro NAMs improve DILI prediction?

DeepDILI + *in vitro* NAMs for DILI Prediction

- Three **3D Spheroid model** were evaluated in combination with DeepDILI, focusing exclusively on drugs that were not included in the DeepDILI training.

	InSphero	Genetech	AZ
Total #drugs	152	110	96
#drugs not by DeepDILI	35	37	32



- **Findings:**
  - **Two methods agreed**: prediction accuracy approached ~90%, representing a 14–30% improvement.
  - **Two methods disagreed:** a confidence-guided integration strategy improved performance at least 50%.
  - **Consensus prediction:** improved DILI prediction by ~15% in average compared to individual approaches.

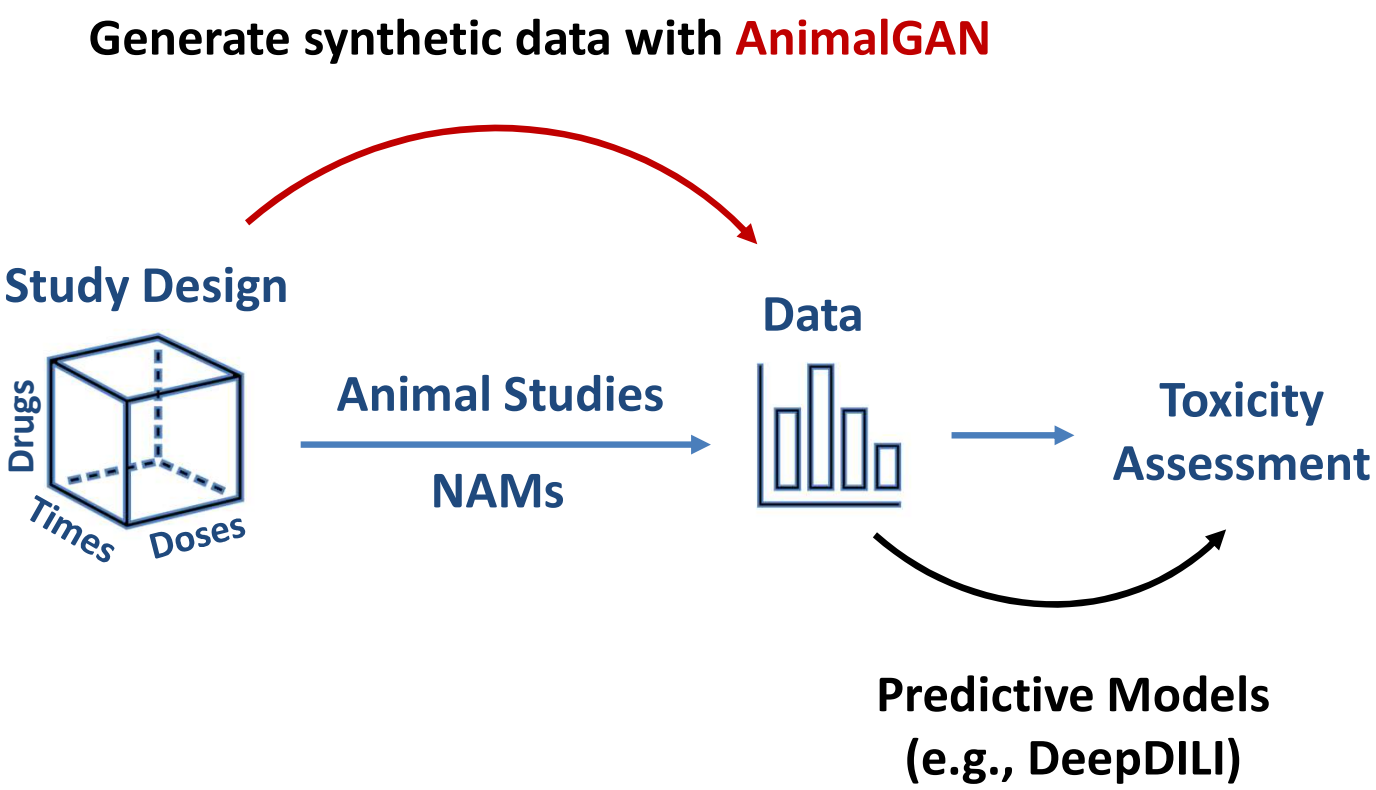
DeepDILI + Spheroid Models Compared to MPS

9 drugs not Used by DeepDILI	Human DILI Risk*	Liver Chip Results	DeepDILI + *InSphero*	DeepDILI + *Genentech*	DeepDILI + *AZ*
FIRU	**0**	**0**			
Fialuridine	**1**	**1**	1		
Ambrisentan**	**0**	**0**		1	1
Sitaxsentan	**1**	**1**		1	1
Tolcapone	**1**	**1**	1	1	1
Pioglitazone	**1**	**1**			1
Troglitazone	**1**	**1**	1	1	1
Trovafloxacin	**1**	**1**		1	
Ximelagatran	**1**	**1**	1	1	1

* Human DILI classification reported by Garside et al;
** Based on FDA labeling documents, its DILI can not be determined (i.e., ambiguous DILI concern in DILIrank)

- MPS (organ-on-a-chip) technologies are low throughput and expensive.
- DeepDILI+Spheroid are comparable with MPS.
- The comparison is based on a small number of drugs and an extensive analysis needs to be done with a large dataset.

Toxicity Assessment Workflow and The Role of AI

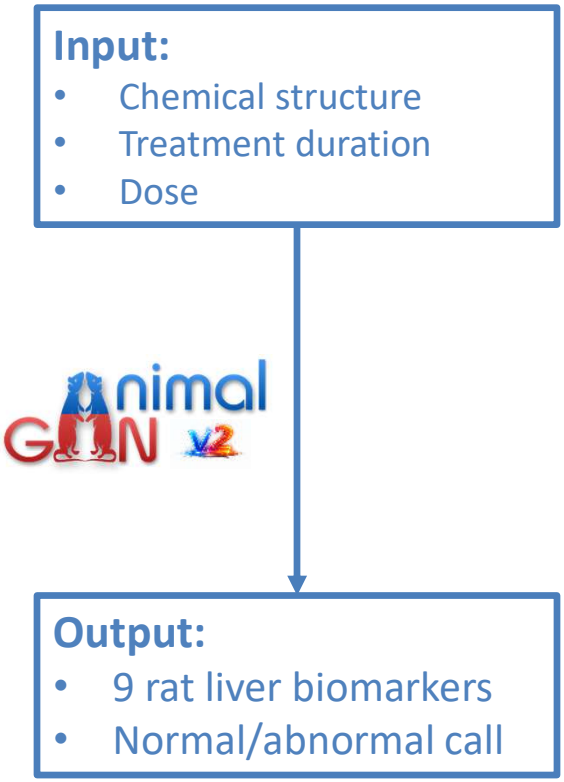


Virtual Animal Models to Generate Animal Study Results with Generative AI

- **AnimalGAN** is a framework to develop virtual animal models using Generative Adversarial Networks (GANs).
- **AnimalGAN v1 model:** Generate 38 rat clinical pathology parameters based on the tested compound, dose level, and treatment duration.

Chen X., Liu Z., and Tong W. ***Nature Communications***. 2023, 14, 7141.

- **AnimalGAN v2 model:** Determine whether individual liver biomarkers fall within the normal range or exceed the abnormal thresholds (e.g., ALP> ≥ 2×ULN [upper limit of normal]) defined by FDA and OECD guidelines for DILI assessment.



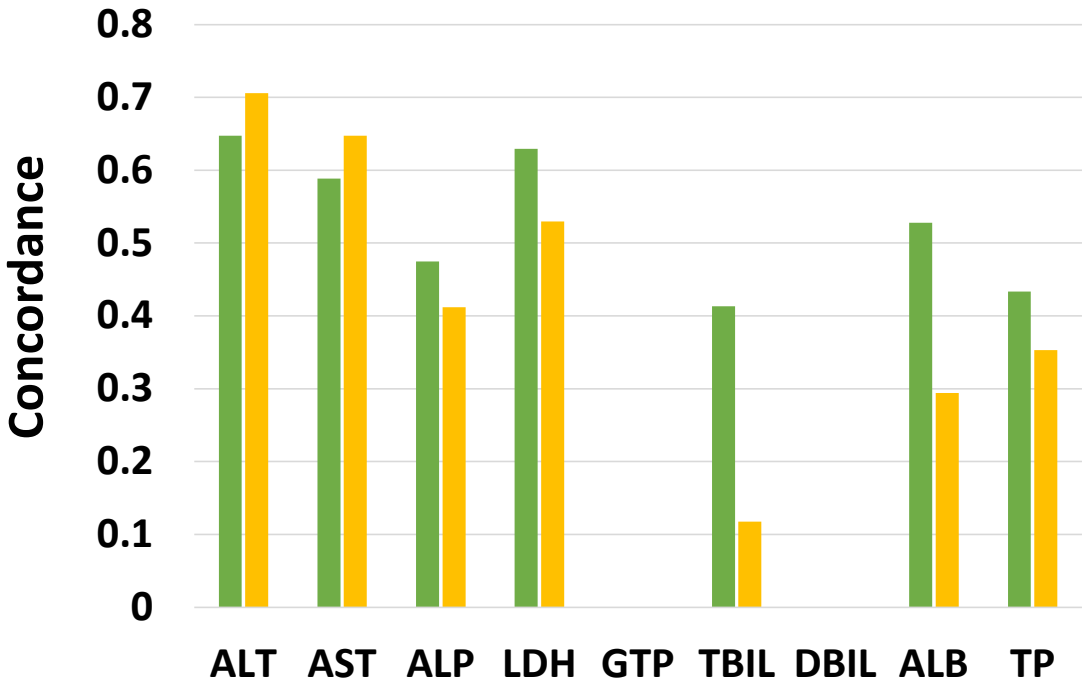
U.S. FOOD & DRUG ADMINISTRATION

Can AnimalGAN2.0 generated DILI biomarkers support DILI assessment under FDA/OECD guideline?

Performance on 355 Market Drugs

355 DrugMatrix compounds (717 treatments) predicted by AnimalGAN2

Cross-Lab replicability based on 70 common compounds (175 treatments) tested by both TG-GATEs and DrugMatrix to serve a baseline



How is the concordance calculated:

- Each method determines whether individual liver biomarkers are within the normal range or exceed abnormal thresholds. Then, agreement between the two methods is calculated.

AnimalGAN2 Results

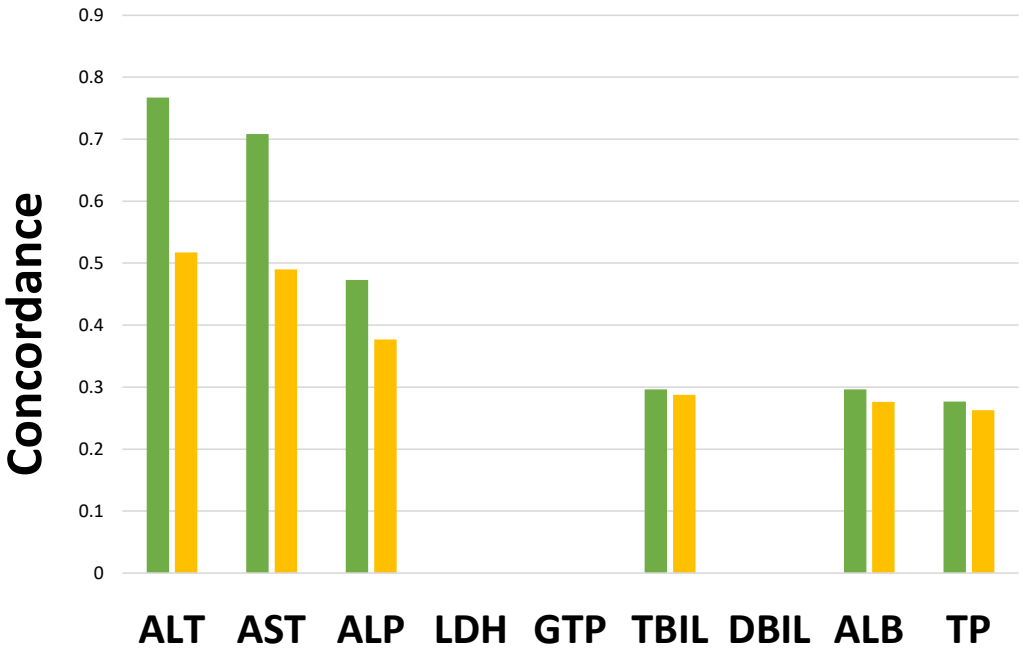
Exp Results	ALT	Normal	Abnormal
	Normal	a	c
	Abnormal	d	b

Concordance = (a+b)/(a+b+c+d)

Performance on 28 Preclinical Drug Candidates

28 preclinical drug candidates (135 treatments) predicted by AnimalGAN2

Intra-Study Replicability that provides a baseline if a study is repeated several times in the same lab.



How is the concordance calculated:

- Each method determines whether individual liver biomarkers are within the normal range or exceed abnormal thresholds. Then, agreement between the two methods is calculated.

AnimalGAN2 Results

Exp Results	ALT	Normal	Abnormal
	Normal	a	c
	Abnormal	d	b

Concordance = (a+b)/(a+b+c+d)

Summary

- How DeepDILI compares with animal studies in assessing DILI risk in humans?
  - DeepDILI outperformed animal models.

- Whether DeepDILI in combination with in vitro NAMs improves DILI prediction?
  - The combined models outperformed individual models alone with the comparable results compared to MPS.

- Can AI generate animal study data in preclinical application?
  - Synthetic DILI biomarkers generated by AniamlGAN2.0 has similar performance as real experiments in accordance with the FDA/OECD guideline.

Acknowledgment

AIRForce Team at NCTR

- Dongying Li (QKAR)
- Ting Li (DeepDILI, TransOrgan, and TransTox)
- Xi Chen (AnimalGAN and DigitalLiver)
- Leihong Wu (askDrugTox)
- Joshua Xu (PathologAI and R2R Branch Chief)
- Mansi Chandra (AIVIVE)
- Yanyan Qu (CardioTox and QSARs)
- Shivangi Shrimali (MPS and women's health)

CDER Collaborators on AnimalGAN

- Shraddha Thakkar (OCS and SafetAI)
- CDER/OCS and OTS leadership

FDA Leadership:

- Steve Kozlowski (FDA Chief Scientist)
- Tucker Patterson (NCTR Director)

External Collaborators:

- AbbVie, BI, and etc