Tox-GAN: An AI Approach Alternative to Animal Studies – a Case Study with Toxicogenomics

Xi Chen, Weida Tong*, Zhichao Liu*

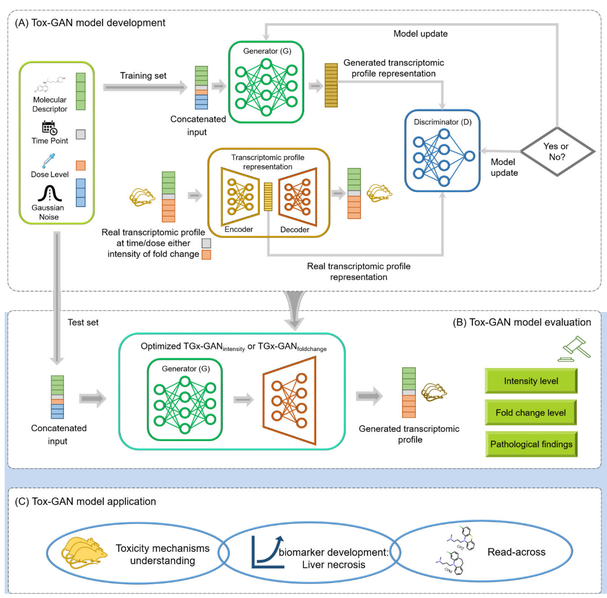
Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, Jefferson, AR 72079

Food and Drug Administration
NCTR
National Center for Toxicological Research

Abstract

Animal study is a critical component in biomedical research, pharmaceutical product development, and regulatory application. Toxicogenomics (TGx) which incorporates emerging genomic technologies into the conventional animal models, has offered an unprecedented opportunity in two areas: inferring toxicity mechanisms based on individual gene activities and developing safety biomarkers based on gene expression profiles. Meanwhile, a worldwide effort has led to a paradigm shift in toxicology towards "reducing, refining and replacing" animal use. Herein, we proposed an artificial intelligence (AI) based approach capable of generating the TGx data from animal studies without using animals. This Tox-GAN approach is developed with a deep generative adversarial network (GAN) to generate both gene activities and expression profiles in TGx involving multiple doses and treatment durations. Using the rat liver TGx data from the Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation System (TG-GATEs), we found that Tox-GAN was an effective alternative to generate the transcriptomic profiles with high similarity (0.997±0.002 in intensity level and 0.740±0.082 in the fold change level) to their corresponding real gene expression profiles without animal consumption. Importantly, we successfully demonstrated the outstanding performance of Tox-GAN in both areas of the TGx application mentioned above. In terms of inferring toxicity mechanisms, over 96% agreement in Gene Ontology was found between the Tox-GAN results and real gene expression data. With respect to the biomarker development, we challenged a necrosis biomarker developed on real gene expression data with a set of studies of both real and generated gene expression profiles. An undistinguishable predictive performance was concluded between the two. We further exemplified the potential utilities of the proposed Tox-GAN in aiding the chemical-based read-across. To the best of our knowledge, the proposed Tox-GAN model is the first attempt to generate the in-vivo transcriptomic profiles at different time and dose settings as long as the chemical structure is provided. Overall, Tox-GAN holds great promise for inferring high-quality toxicological profiles even without drug synthesis and animal treatment to advance the modernized toxicology paradigm.

Workflow of the Tox-GAN



Tox-GAN enabling transcriptomic profile inference

To investigate the performance of the optimized models, we applied the models to infer transcriptomic profiles in the test set. The t-SNE plots depicted the distribution of generated transcriptomic profiles and real ones were well matched within the Gaussian distribution for both Tox-GAN$_{intensity}$ and Tox-GAN$_{foldchange}$ models. Furthermore, the average and standard deviation of Pearson correlation coefficients between the generated transcriptomic profile and their corresponding real ones were 0.997±0.002 and 0.740±0.082 for Tox-GAN$_{intensity}$ and Tox-GAN$_{foldchange}$, suggesting the strong capability of the proposed Tox-GAN for inferring transcriptomic profiles in both intensity and fold change levels.

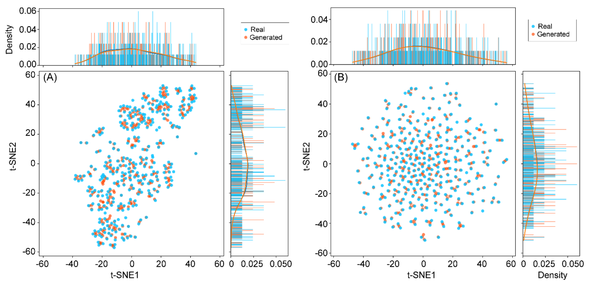


Figure 2. t-SNE visualization and Probability density of the generated and real transcriptomic profiles. The blue and orange represent the real transcriptomic profiles and their corresponding generated ones from (A) Tox-GAN$_{intensity}$ and (B) Tox-GAN$_{foldchange}$, respectively.

The generated transcriptomic profiles and the corresponding real ones had high similarity (0.997±0.002), indicating transcriptomic profiles produced Tox-GAN could well capture the genomic responses in different phenotypes.

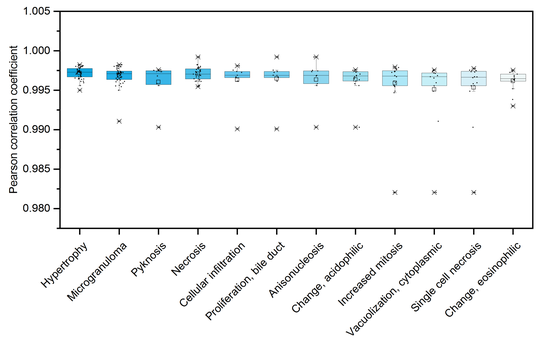


Figure 3. The distribution of Pearson correlation coefficient between generated transcriptomic profiles from Tox-GAN$_{intensity}$ and their corresponding real ones across pathological findings.

Figure 1. Workflow of the Tox-GAN: (A) Tox-GAN model development. First, the concatenated information from the molecular descriptor, duration time, dose, and Gaussian noise in the training set is employed as an input to the Generator G, used to generate the transcriptomic profile representation. Second, the real transcriptomic profile representation is generated from the real transcriptomic profiles through an autoencoder. Third, the discriminator D distinguished the generated and real transcriptomic profile presentation. The process was finalized until the discriminator D could not distinguish the generated and real transcriptomic profile representations. (B) Tox-GAN model evaluation. The optimized Tox-GAN model was used to generate the transcriptomic profiles in the test and evaluated in the intensity, fold change, and pathological findings levels. (C) Tox-GAN model application. The developed Tox-GAN model was applied in 28-day repeated dose toxicity studies, predictive toxicology (e.g., liver necrosis prediction), and read-across.

Tox-GAN facilitating understanding of toxicity mechanisms

The 28-day repeated dose toxicity study is one of the standard toxicity experimental designs to evaluate compounds' adverse effects and uncover the underlying toxicity mechanisms when repeatedly administered to the experimental animals in TGx. The high Pearson correlation coefficients suggested the potential utility of the proposed Tox-GAN to refine the 28- day repeated dose toxicity study. Moreover, the concordance between the Gene Ontology (GO) enrichment analysis results indicated that the generated transcriptomic profiles from the proposed Tox-GAN could reflect the biological processes as the real ones.

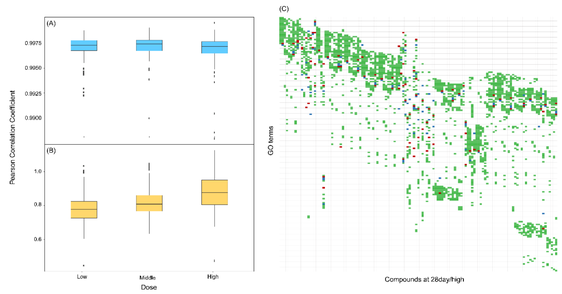


Figure 4. Application of Tox-GAN in 28-day repeated dose toxicity studies: (A) and (B) The Pearson correlation coefficient between the generated transcriptomic profiles and their corresponding real ones in different dose levels in 28-day repeated dose toxicity studies using Tox-GAN$_{intensity}$ and Tox-GAN$_{foldchange}$, respectively. (C) The enriched GO terms between generated transcriptomic profiles from Tox-GAN$_{foldchange}$ and their corresponding real ones. The green, blue and red dots denote the GO terms enriched by both the generated transcriptomic profiles and their corresponding real ones, only generated transcriptomic profiles and real transcriptomic profiles, respectively.

Tox-GAN enhancing the biomarker development

The high correlation (r = 0.951) of possibilities of the test set were obtained between the DNN models for the generated transcriptomic profiles and real ones, suggesting the great potential to utilize the Tox-GAN to facilitate the biomarker development in predictive toxicology.

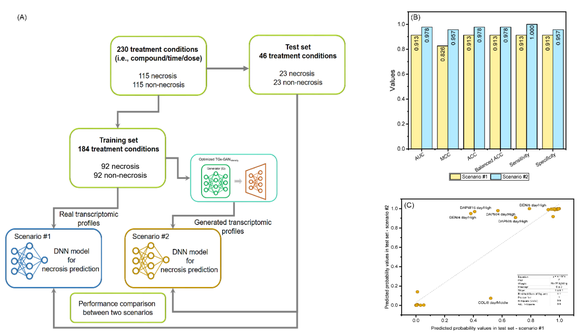
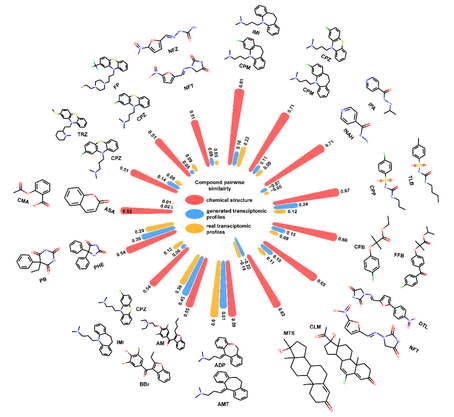


Figure 5. Application of Tox-GAN in predictive toxicology of liver necrosis: (A) the designed workflow for predictive model development for liver necrosis. Specifically, the 230 treatment conditions were divided into the training and test sets with a ratio of 8:2. The DNN models were developed using two scenarios: real transcriptomic profiles and generated transcriptomic profiles by Tox-GAN$_{intensity}$. The test set is further used to evaluate the trained DNN models based on the two scenarios. (B) the performance metrics of the developed DNN models in the test set with the proposed two scenarios. The light yellow and blue colors represent scenarios #1 (i.e., DNN with real transcriptomic profiles) and #2 (i.e., DNN with generated transcriptomic profiles). (C) The correlation between predicted probabilistic values of samples in the test sets yielded from the two scenarios.

Tox-GAN aiding the read-across

The similarities between transcriptomic profiles (in the range of -0.22 ~ 0.45) were much smaller than those of chemical structure, suggesting the better discrimination power of transcriptomic profile-based read-across. Furthermore, the similarity based on the generated transcriptomic profiles was very similar to the real transcriptomic profiles, demonstrating the utility of the Tox-GAN model in the biological profile-based read-across.



clomipramine: CPM imipramine: IMI chlorpromazine: CPZ isoniazid: INAH iproniazid: IPA chlorpropamide: CPP
tolbutamide: TLB clofibrate: CFB fenofibrate: FFB dantrolene: DTL nitrofurantoin: NFT chlormadinone: CLM
methyltestosterone: MTS adapin: ADP amitriptyline: AMT amiodarone: AM benzbromarone: BBr phenobarbital: PB
phenytoin: PHE aspirin: ASA coumarin: CMA thioridazine: TRZ fluphenazine: FP nitrofurazone: NFZ

Figure 6. Application of Tox-GAN in read-across. The circle bar plot illustrates the similarities between the top-ten drug pairs in the chemical space, generated transcriptomic profiles, and corresponding real ones with red, blue, and yellow colors, respectively. Meanwhile, the chemical structures of drug pairs were illustrated.

Summary

- We proposed a novel TGx-based deep generative adversarial network (GAN) model (Tox-GAN) to infer transcriptomic response in rat liver based on their chemical structure information only.
- The developed Tox-GAN models could generate transcriptomic profiles in both intensity and fold changes levels, highly correlated with their corresponding real ones.
- Furthermore, we exemplified the potential utility of Tox-GAN in facilitating the understanding of toxicity mechanisms, enhancing the biomarker development in predictive toxicology (i.e., liver necrosis prediction), and aiding the chemical-based read-across.
- Moreover, the developed Tox-GAN models were openly accessible through https ://github.com/XC-NCTR/Tox-GAN, which could be utilized for estimating in vivo transcriptomic profiles without any animal studies needed.
- To the best of our knowledge, this is the first attempt to explore GAN for toxicology as an alternative to animal models.

Disclaimer

The views presented in this article do not necessarily reflect those of the U.S. Food and Drug Administration. Any mention of commercial products is for clarification and is not intended as an endorsement.

Correspondence

- Weida Tong Weida.Tong@fda.hhs.gov
- Zhichao Liu Zhichao.Liu@fda.hhs.gov