Machine learning on drug-specific data to predict small molecule teratogenicity

Anup P. Challa1,3,4,5,* , Andrew L. Beam2,3, Min Shen4, Tyler Peryea4, Robert R. Lavieri1, Ethan S. Lippmann5, David M. Aronoff6,7,8

1Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN 37203
2Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115
3Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115
4National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD 20850
5Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN 37212
6Division of Infectious Diseases, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37203
7Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN 37203
8Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37203

Abstract

Pregnant women are an especially vulnerable population, given the sensitivity of a developing fetus to chemical exposures. However, prescribing behavior for the gravid patient is guided on limited human data and conflicting cases of adverse outcomes due to the exclusion of pregnant populations from randomized, controlled trials. These factors increase risk for adverse drug outcomes and reduce quality of care for pregnant populations. Herein, we propose the application of artificial intelligence to systematically predict the teratogenicity of a prescriptible small molecule from information inherent to the drug. Using unsupervised and supervised machine
learning, our model probes all small molecules with known structure and teratogenicity data published in research-amenable formats to identify patterns among structural, meta-structural, and in vitro bioactivity data for each drug and its teratogenicity score. With this workflow, we discovered three chemical functionalities that predispose a drug towards increased teratogenicity and two moieties with potentially protective effects. Our models predict three clinically-relevant classes of teratogenicity with AUC = 0.8 and nearly double the predictive accuracy of a blind control for the same task, suggesting successful modeling. We also present extensive barriers to translational research that restrict data-driven studies in pregnancy and therapeutically “orphan” pregnant populations. Collectively, this work represents a first-in-kind platform for the application of computing to study and predict teratogenicity.

**Keywords**

teratogenicity; drug development; drug exposure; machine learning; informatics; chemical structure; high-throughput screening; translational medicine

---

**Introduction**

1.1. Risky prescriptive behavior in pregnancy

Teratogenicity is the most serious manifestation of iatrogenic fetal toxicity: teratogens lead to fetal malformation and are implicated in lifelong physical and/or mental disabilities. Nonetheless, clinical trial results of drug exposure during pregnancy are often conflicting, and teratogenicity scoring for small molecules is unsystematic and performed outside the clinical environment. The consequences of this subjectivity are seen in the high rate of unintended maternal exposure to a teratogenic agent, reminiscent of the “thalidomide disaster” of the early 1960s. Following this disaster, randomized, controlled trials (RCTs) were modified to exclude pregnant populations, fearing unintended teratogenicity from exposure to unsystematically profiled drugs. This change continues to “orphan” pregnant women, as many diseases in women’s health lack safe and effective drug choices for treatment.

In the wake of the “thalidomide disaster,” the United States Food and Drug Administration (FDA) developed a five-point scale for ranking the teratogenicity of a compound. This scale is presented in Table 1 (Appendix).

A hallmark of the binning within this scale is the absence of definitive human data: at present, teratogenicity scores are established preclinically by pharmacologists, who evaluate biomarkers of fetal toxicity in animal models. This approach is inherently limited, as common in vivo models are not sufficiently representative of human physiology, and human subjects are not included in the teratogenicity scoring process for ethical reasons. Indeed, the limited human data available for teratology scoring are often derived retrospectively from high-profile cases of fetal malformation resulting from drug exposure. While new FDA standards for scoring teratogenicity acknowledge these limitations by providing fewer, more holistic toxicity scores, these standards still suffer from
the absence of robust human data and are not yet integrated in clinical decision-making tools\textsuperscript{18}.

Collectively, the factors above create a significant degree of uncertainty at the point of care (POC), as providers are guided on contradictory, incomplete, and non-human derived information in their choice of prescriptions for pregnant women. This dilemma is of special consequence to expectant mothers with chronic morbidities pre-existing to their pregnancies\textsuperscript{11}.

1.2. Target rationale for teratogenesis

Fetal exposure to a teratogen \textit{in utero} strongly associates with cognitive and/or physical disabilities, resulting from dysregulation of key developmental processes such as neurulation, purine and pyrimidine synthesis, and lipid anabolism\textsuperscript{2,19}.

Broadly, teratogens may be categorized by their mechanism of action (MOA) as either \textquoteleft on-target\textquoteright{} or \textquoteleft off-target\textquoteright{}\textsuperscript{20–22}.” \textquoteleft On-target\textquoteright{} teratogenicity implies the generation of adverse phenotypes from bioactive agents impacting well-defined protein targets that are critically regulated in development. In contrast, \textquoteleft off-target\textquoteright{} teratogenicity implies \textit{mutagenicity}, resulting from DNA damage such as alkylation and thymine dimerization. \textquoteleft Off-target\textquoteright{} teratogenicity involves repeated reactions between a teratogen and newly-synthesized nucleic acid residues, often resulting from the generation of reactive oxygen species (ROS) generated from drug metabolism\textsuperscript{20}.

Thus, teratology is known to converge on few principal MOA classes\textsuperscript{19,23}, which are outlined in Table 2.

1.3. Machine learning in maternal-fetal medicine

The inherent contradiction between the limited target rationale for teratogenesis and the extent of uncertainty that guides prescribing behavior for gravid populations speaks to the need for more rigorous predictions of small molecule teratogenicity. Furthermore, computational modeling on healthcare data is the most accurate method of predicting drug safety in pregnant women, given that phase I trials are unethical for expectant populations and animal models are inherently limited for studying human health\textsuperscript{12,13,24}.

\textit{Classification algorithms} are optimized to identify patterns between associated data sets (such as binding affinity and phenotype data for a cytotoxic target)\textsuperscript{25–28}, suggesting that machine learning (ML) classifiers may play a pivotal role in systematically establishing relationships between maternal drug history and adverse fetal outcomes\textsuperscript{29–31}. While these models are not intended as a replacement for existing physician knowledge of responsible prescriptive practice\textsuperscript{32}, ML classifiers offer an attractive opportunity to discover meaningful relationships within existing biomedical data than could result in meaningful POC conclusions.

There have been few previous studies leveraging this brand of artificial intelligence (AI) for predicting iatrogenic fetal toxicity. Of these select investigations, a majority have focused solely on population-level, patient-derived data to discover adverse outcomes from maternal
medication history and neonatal disease information\textsuperscript{17,30,31,33,34}. In 2017, Boland \textit{et al.} reported on a successful ML algorithm for parsing electronic health record (EHR) data to develop data-driven definitions of adverse drug outcomes associated with class C teratogenicity; the authors focused their modeling on congenital disease and fetal death phenotypes\textsuperscript{35}. Studies with similarly-limited scope that analyzed insurance claims data are also available\textsuperscript{30,33,36}. Recognizing the additional predictive power of chemical data for teratogenicity, Baker \textit{et al.} published a ML model for the identification of compounds implicated in cleft palate formation from existing toxicology high-throughput screening (HTS) bioassay data and information on chemicals implicated in cleft palate phenotype identified from systematic literature review. This allowed the authors to identify biomarkers with high positive predictive value for cleft palate and further elucidate chemical exposure-adverse outcome clusters\textsuperscript{17}.

In this study, we report on a previously-unattempted, unbiased (phenotype-agnostic and target-agnostic) approach to predicting teratogenicity by identifying chemical and biochemical factors that predispose a chemical to increased teratogenic risk. Given significant limitations in established teratogenicity scoring criteria, we propose a novel application of ML to develop a teratogenicity quantitative structure-activity relationship (QSAR)\textsuperscript{37}. By leveraging drug structure, meta-structural elements like molecular energetics, and real-world bioactivity data, we attempt to predict the teratogenic risk of drugs potentially prescribable in pregnancy.

\textbf{Materials and Methods}

Our teratogenicity QSAR accesses chemical and bioassay data to predict a teratogenicity score for compounds that are prescribable in pregnancy and to identify patterns within drug-specific information that predispose a drug towards an increased risk of fetal toxicity.

Broadly, we leverage three layers of drug data to accomplish these tasks:

\begin{enumerate}
\item The inherent structure of each drug, as encoded by several classes of chemical fingerprints\textsuperscript{38} that capture upwards of 1,024 structural features of each molecule
\item Meta-structural features for each drug, including druglikeness, predicted molecular energetics, and mutagenicity—as calculated from the Molecular Operating Environment (MOE)\textsuperscript{39}, an industrial-grade chemical computing software—and mutagenicity data from predictive models of the Ames test\textsuperscript{40}
\item Antagonist-mode toxicology assay data from Tripod, a public-facing collection of HTS data from the Toxicology in the 21\textsuperscript{st} Century (Tox21) initiative of the National Institutes of Health\textsuperscript{41,42}, on all targets listed in Table 2 with simultaneous coverage in Tox21
\end{enumerate}

We employed version 3.5.3 of the R integrated development environment (https://www.r-project.org/)\textsuperscript{43} for parsing all chemical and bioassay data and implementing and tuning unsupervised and supervised ML models for associating these data.
We are committed to open-source science. All source data, code, and output files relevant to the development of our model is available through the following GitHub repository: https://github.com/apchalla/teratogenicity-qsar.

2.1. Mining structure and teratogenicity data

DrugBank 5.1.0 (https://www.drugbank.ca/) is a publically-available drug encyclopedia developed by the University of Alberta. It contains comprehensive entries of more than two hundred (200) data fields for 9,099 small molecules of known structure that have passed phase I of an existing RCT. Each DrugBank entry contains structured information on compound structure, MOA, existing formulations, and drug marketing history, among other clinically-relevant datasets. DrugBank is a self-described cheminformatics resource; therefore, the pharmacopeia provides a highly pliable application programming interface (API), which allows for easy data mining and extraction. Given the comprehensiveness of the DrugBank database, as well as its amenability for data-driven analyses, we extracted the structures of all 9,099 DrugBank entries as three-dimensional spatial data files (3D-SDFs).

To obtain relevant, FDA-compliant teratology data, we interrogated SafeFetus (https://www.safefetus.com/), a registry for expectant mothers hosting the largest publically-available repository of structured, FDA-aligned teratogenicity scores. Therefore, we extracted teratogenicity scores for all 652 eligible drugs from SafeFetus.

Teratology data are not routinely published, and many large pharmacovigilance databanks like FDA’s DailyMed (https://dailymed.nlm.nih.gov/dailymed/) do not present all teratology information in structured fields—as required for computing on this information—and have inflexible APIs for data extraction.

2.2. Layer 1: Leveraging drug structure for predicting teratogenicity

From DrugBank 5.1.0, all 9,099 small molecule structure files were mined in SDF format. To ensure that DrugBank structure files were not corrupted in the extraction process, the SDF set was imported into the LigPrep graphical user interface of Schrödinger 2018-2 (https://www.schrodinger.com/), a suite of chemical computing software that enables predictive modeling in structure-guided pharmacological studies. Validating by visual inspection that all DrugBank files were chemically-valid, the SDF set was imported to R. Then, using the cheminformatics toolkits ChemmineR (CRAN: ChemmineR) and Rcdk (CRAN: Rcdk), the SDF set was converted to twelve (12) classes of chemical fingerprints, encodings of chemical structure as thousand-dimensional matrices that record the presence of absence of distinctive chemical motifs, including topological torsions, R/S stereochemistry, common functional groups, Brønsted-Lowey acidity/basicity, general acid/base catalysts, and other salient chemomarkers. This fingerprinting process is only valid for organic small molecules; therefore, all inorganic agents were automatically parsed from our drug set by the ChemmineR and Rcdk fingerprinting algorithms. Thus, fingerprinting allowed us to access comprehensive, structured information on nearly nine thousand (9,000) small molecules and one-hot encode this information.

As noted above, we obtained FDA-compliant teratogenicity data from SafeFetus, the largest publically-available source of structured FDA teratogenicity scores with an API. Integrating
the data sets for teratogenicity and drug structure in R, we obtained \( N = 611 \) drugs with information on both structure and teratogenicity.

We then developed multiple label classification strategies for teratogenicity scores, based on the nature of FDA teratology scores and a bibliostatistic search. While we acknowledge that the conceptual relevance and quality of labels are of utmost importance for classification tasks, we also argue that the definition of optimal labels \textit{a priori} can be difficult. Therefore, in this manuscript, we detail the procedure we employed to define teratogenicity score embeddings, starting with the application of published rubrics and moving towards literature searches and necessary trial-and-error approaches to define and optimize a more precise set of scores.

Therefore, one set of teratogenicity scores we employed for all 611 drugs was aligned according to native FDA schema. Through consultation with practicing clinicians to discuss the heuristics they employ in prescriptive practice, we redefined a second set of scores as a three-pronged scale of bins: “Clinically Acceptable Risk” (scores A/B), “Moderate Risk” (score C), and “Clinically Unacceptable Risk” (scores D/X). We then defined a third scale by a systematic literature search of the Embase medical library system (https://www.elsevier.com/solutions/embase-biomedical-research)\textsuperscript{51} and a Cochrane review (https://www.cochrane.org/evidence)\textsuperscript{52} for the keyword “teratogenic.” We queried the ~16,000 articles that resulted from this search using simple random sampling, such that we assigned an identifier to each article and sampled \( O(10^3) \) articles by their identifiers. For articles within our random sample and which referenced specific drugs, we observed that the keyword “teratogenic” was associated with a mention within the article of FDA scores of C, D, or X. Therefore, we defined a binary scale of scores as “Non-Teratogenic” (scores A/B) and “Teratogenic” (scores C/D/X) classes.

We emphasize that this \textit{ad hoc} literature review was a non-rigorous—but necessary—step that allowed us to develop a starting point from which we could study and discuss potential tuning of the definition of our labels, per their contextual relevance and the model performance that we observed with these embeddings. We discuss these issues throughout the remainder of our manuscript.

2.2.1. \textbf{Unsupervised modeling}—First, to discover clustering relationships between teratogenicity and drug structure, the Barnes-Hut implementation of the t-Distributed Stochastic Neighbor Embedding (t-SNE) procedure\textsuperscript{53} was enacted on all combinations of fingerprint and teratogenicity score data sets, including a combined, non-redundant, and feature-prioritized set of all chemical fingerprints. t-SNE is a dimensionality reduction procedure, which can plot all dimensions of drug structure against all dimensions of teratogenicity for all drugs included our data sets. The presence of tight clusters in a t-SNE plot indicates dependency between the plotted variables\textsuperscript{54,55}.

Of the t-SNE combinations we attempted on our structure and teratology encodings, the t-SNE plot generated with 1,024-dimensional Morgan fingerprints\textsuperscript{56} and a binary classification of teratological risk showed the strongest clustering relationships. Clusters were identified by visual inspection, with each point within a cluster representing a drug.
Hence, we mapped points within each discrete t-SNE cluster in reverse, from t-SNE space to its associated DrugBank entry. Noting that all points within each cluster were consistent with a salient chemical functionality within the component drug structures, and that all cluster component drugs belonged to the same class, we considered our identification of meaningful clusters to be successful. Performing systematic literature review on each drug class identified as strongly associated to the presence or absence of elevated teratogenic risk, we noted that select drug class—teratogenicity score relationships identified by our model were verified in clinical decision-making tools like UpToDate (https://www.uptodate.com/contents/search) and Medscape (https://www.medscape.com). However, several structure-teratology relationships identified by t-SNE appeared contentious in relevant literature: sufficient human data are not available to accurately classify the class of drugs distinguished by the t-SNE-identified chemical functionality as teratogenic or safe. We present a deeper discussion of the contribution of our t-SNE findings to these debates in the “Results and Discussion” section of this publication. The most consistent t-SNE plot and the functionalities it identified as significantly associated to the presence or absence of teratological risk are also shown in Figures 3 and 4 in the “Results and Discussion” section.

Noting that multiple structure-teratogenicity relationships resulting from our t-SNE analysis were validated in the literature, we considered our unsupervised ML model to be a successful proof-of-concept experiment.

2.2.2. Supervised modeling—Given that t-SNE successfully and consistently identified moieties that might predispose a drug towards an increased risk of teratogenicity, we decided to enable a supervised ML model that can prospectively predict a drug’s teratogenicity score from structural information. Using the R package Caret (CRAN: Caret), we developed three (3) models with inherent five (5)-fold cross validation (CV), such that we obtained test set accuracy on running each model. These models included Random Forest, Extreme Gradient Boosting, and Gradient Boosting Machine (GBM). Testing these models with five-pronged, FDA-adherent teratogenicity scores, we found that GBM yielded the highest predictive accuracy. Therefore, we re-trained our GBM model with the trivariate, clinically-oriented teratology scale described above and obtained higher accuracy for this model than for the GBM trained on five-dimensional labels. For all models, we optimized hyperparameters using a large grid search within Caret.

2.3. Layer 2: Curating meta-structural information for exploratory analysis

After deriving a successful model for predicting teratological risk from drug structure, we sought to increase the predictive accuracy of our GBM by supplementing our features with information on “meta-structure.” These factors included the following variables, which were calculated for all 611 sampled drugs within MOE (https://www.chemcomp.com/Products.htm), a suite of industry-grade chemical computing software for computer-aided molecular design. Each of the following meta-structural sets was encoded by chemically-significant cutoffs when available (e.g., druglikeness benchmarks from Lipinski’s Rule of Five (RO5) or cutoffs determined from ROC analysis of extracted data):

- Druglikeness: the adherence of each molecule to Lipinski’s Rule of Five restrictions on the number of hydrogen-bond acceptors, hydrogen-bond donors,
octanol-water partition effects, total polar surface area, molecular weight, and number of rotatable bonds for an attractive drug candidate.

- Energy of the Highest Occupied Molecular Orbital (HOMO): a quantum chemistry metric of the tendency of a molecule to donate an electron, as a proxy for drug stability and tendency to generate mutagenic free radicals.
- Energy of the Lowest Unoccupied Molecular Orbital (LUMO): a quantum chemistry metric of the tendency of a molecule to accept an electron, as a proxy for drug stability and tendency to generate mutagenic free radicals.
- Mutagenicity score, as calculated from in-built predictive models of the Ames test.
- pKa and most basic pKa

2.3.1. Unsupervised modeling—Combining MOE calculations for the above variables and all structural data sets, we performed feature selection within Caret to remove redundancy and highly-correlated features within the integrated descriptor set. Then, we re-executed t-SNE on binary teratogenicity scores, with the hope of identifying new clustering relationships between physiochemical features and teratogenicity.

2.3.2. Supervised modeling—GBM with five-fold CV was re-executed with a three-pronged set of teratogenicity scores and feature-prioritized structural and meta-structural information. Hyperparameters were optimized by large grid search within Caret.

2.4. Layer 3: Repurposing Tox21 HTS Data on Teratogenic Targets

Given that teratogenicity has well-identified target rationale, we decided to leverage existing, real-world bioassay information for all targets implicated in teratogenesis (as described Table 2) and previously screened through the Toxicology in the 21st Century Initiative (Tox21) of the National Institutes of Health. Tox21 leverages HTS of millions of bioactive compounds—including most common pharmaceuticals—in thousand well-plate, cell-based assays. While this HTS platform is not teratogenicity-specific, it does contain information on targets implicated in teratogenesis.

Scoping all information available on Tripod, the public-facing data browser of Tox21, we extracted antagonist-mode RAR and HDAC data for supplementation of our model. RAR data were derived from murine embryo fibroblast cells (C3H10T1/2, American Type Culture Collection, Manassas, Va., USA), and HDAC data were obtained from human colorectal carcinoma cells (HCT-116, American Type Culture Collection, Manassas, Va., USA). Assay protocols are available from the Tripod website specified above.

Data available from Tripod include bioactivity for a given target (encoded as “inactive,” “active”), curve class, IC50, efficacy, and Hill coefficient. Of these variables, we studied curve class, IC50, and efficacy as proxies of binding affinity of each sampled compound for RAR and HDAC. Therefore, all compounds with available structure, teratogenicity score, and RAR/HDAC HTS coverage (N = 128) were probed by t-SNE and GBM. Data were one-
hot encoded using standard bioactivity cutoffs for drug development (i.e., curve class \( \neq 4 \), \( \text{IC}_{50} \leq 20 \mu\text{M} \), efficacy \( \leq -50\% \))\(^{67-69}\).

### 2.4.1. Unsupervised modeling

Combining MOE calculations for the above assay data and all structural and meta-structural data sets, we performed feature selection within Caret to remove redundancy and highly-correlated features within the integrated descriptor set. Then, we re-executed t-SNE on binary teratogenicity scores, with the hope of identifying new clustering relationships between assay data and teratogenicity.

### 2.4.2. Supervised modeling

GBM with five-fold CV was re-executed with a three-pronged set of teratogenicity scores and feature-prioritized structural, meta-structural, and biochemical assay information. Hyperparameters were optimized by large grid search within Caret.

### 2.5. ROC statistics

In evaluating the results of our supervised and unsupervised models, we took special note of the imbalanced nature of our teratogenicity score data set. This is a problem inherent to the subjective nature of teratogenicity scoring by the FDA, as drugs with unclear safety profiles often receive a label of class C\(^{24}\). In accordance with this practice, we observed that 310 of our 611 sampled drugs (51\%) were labelled C. The remainder of our label set was distributed as follows, which is—at large—representative of the FDA’s classification behavior: \( A = 14/611 \) drugs (2\%), \( B = 157/611 \) drugs (26\%), \( D = 91/611 \) drugs (15\%), \( X = 39/611 \) drugs (6\%)\(^{24,47}\).

Therefore, we decided to perform ROC statistics to evaluate the strength of our set of features to predict a drug’s FDA teratogenicity score, since ROC statistics are more resilient to class imbalance than GBM accuracy. ROC statistics for structure-based predictions of teratogenicity (AUC = 0.8) suggested that chemical structure has strong predictive power for a drug’s teratological risk. We describe these results in depth in the following section.

Figure 1 contains a summary of all data sources and modes of ML analysis we considered in the creation of our model. For each feature layer, we ensured that the associated data sources were non-trivial to our model by plotting a feature importance spectrum. In querying these feature importance data, we found that Caret considered all features as having non-zero importance at each implementation of a GBM.

### Results and Discussion

In this manuscript, we present a first-in-kind application of ML to identify structural, meta-structural, and bioassay performance factors that predispose a drug towards increased teratogenic risk. We developed a model to prospectively score a drug’s teratogenicity from these drug-specific factors. Because our workflow is anchored in computing, our methods apply algorithmic rigor to studying teratogenicity, a contrast to many non-systematic studies which have historically dominated this space.
3.1. Summary of key results

3.1.1. Unsupervised learning outcomes—We found that drug structure is a good predictor of teratogenicity, as multiclass ROC analysis between 1,024-dimensional Morgan fingerprints and a three-pronged teratogenicity metric gave AUC = 0.78 (Figure 2). This result validates our hypothesis that a “form-fits-function” argument is valid for predicting teratogenicity from homology between drug structure and pharmacophore biochemistry among targets implicated in teratogenesis.

From t-SNE analysis between drug structure and a binary encoding of teratogenicity (Figure 3), we discovered clusters of teratogenic risk and the absence thereof, which are partially validated within existing clinical literature (Figure 4). Though t-SNE contains noise across most of the diminished structure-teratogenicity landscape, the clusters we identified by visual inspection were consistent in teratogenic risk. A reason for the limited tightness of the observed clustering behavior may involve dimensionality mismatch between structure and teratogenicity data sets, given that we plotted 1,024 structural motifs against only two (2) teratogenicity scores. However, since generating \(\sim 10^3\) independent teratogenicity scores and reducing chemical structure to \(\sim 10^4\) categories are both unfeasible (this would remove the clinical and chemical significance of the respective data sets), we cannot address this probable cause of loose clustering by adjusting the form of the data we seek to associate. Despite these issues, our t-SNE step was a successful proof-of-concept experiment, as we discovered functionalities that are known to be highly fetal toxic and those that are known to be safe through this procedure.

Beyond these validated associations, we also discovered new structure-teratogenicity relationships that might have application in clarifying cases of suspect toxicity risk in the clinical literature. Indeed, our analysis reveals five motifs that are distinctive among cohorts of molecules identified as “teratogenic” and “non-teratogenic.” Both moieties in the “NO” cluster are components of cephalosporins, which include a group of broad-range antibiotics known to be safe for pregnant mothers (class B)\(^{70–73}\). Two distinctive functionalities distinguish cephalosporins from other classes of drugs: the presence of an azetidinone group and a dihydrothiazine ring\(^{74}\). Therefore, as there features distinctively establish cephalosporin identity—which is non-teratogenic—it is reasonable to assert that the azetidinone functionality and dihydrothiazine ring are non-teratogenic chemomarkers in this case. We recognize that the burden of evidence is significant to claim that these motifs demonstrate protective effects. Instead, we suggest that our results warrant more involved analysis of these potentially protective moieties.

In contrast, similar analysis of “YES” clusters reveals three teratogenic chemomarkers, including corticosteroids, fluoroquinolones, and acetylproline derivatives. While fluoroquinolones are documented teratogens\(^{75–78}\), there is contention on the toxicity of steroid derivatives\(^{79–81}\), as well as prolinated compounds\(^{82–84}\). Our model adds to this discussion by arguing that the safety of steroid derivatives should be more deeply interrogated for potentially teratogenic outcomes.

We reasonably assume that the “YES” functionalities in Figure 4 are the source of teratogenicity within molecules that contain them, given that these moieties are distinctive.
This conclusion requires MOA validation; however, as with fluoroquinolones, available phenotypic data appear to support our conclusions on functional group toxicity.

Drawing on these mappings also allows us to evaluate new trends in drug development; namely, we can extrapolate functional group mappings towards drug development targets in the anti-hypercholesterolemic space. Pregnant women with high cholesterol are not advised to take statins, as these drugs are antagonists of HMG-CoA reductase, restricting fatty acid synthesis in a developing fetus (Table 1)\textsuperscript{19,85–87}. Statins contain a fluorobenzene motif, which our model predicts to be the core teratogenic functionality within these drugs. As of date, only one small-molecule anti-hypercholesterolemic drug, ezetimibe (Zetia), does not belong to the statin class of drugs\textsuperscript{88}. Instead, ezetimibe contains a central azetidinone group and has been noted in reduced teratogenicity across the expectant population, as compared to statins (statins are class D agents; ezetimibe carries a class C score)\textsuperscript{89}. Given that we identify azetidinone-containing drugs to carry potential protective effects, this observation edifies the results from our model and speaks to the potential applicability of structure-teratogenicity relationship modeling similar to that in this paper to inform downstream, data-driven inquiries into drug safety for expectant populations. We emphasize that expansion of this study and downstream mechanistic studies are required to fully substantiate our observations.

### 3.1.2. Supervised learning outcomes

Our GBM predicts three classes of teratogenicity with 64.7% accuracy (SD = 3.0%) when trained on 1,024-dimensional Morgan fingerprints. Thus, our model achieves nearly double the predictive accuracy as a blind, probabilistic control for the same trivariate predictive task; QSAR accuracy enrichment is nearly 32% on these baseline predictions. Model penalization to correct for an imbalance of teratogenicity scores did not increase predictive accuracy. Because there exist no other structure-activity relationships, meta-structure-activity relationships, or structure-assay-activity relationships published in this space, we assert our model as a first attempt at applying drug-inherent information towards predicting teratogenicity.

### 3.2. Ontological limitations and barriers to data-driven studies in pregnancy

While the results above appear promising, the data that we queried in this investigation present significant ontological challenges. These problems drastically reduce the sample size of all drug-specific teratology probes and present significant barriers to translational science, as we explain below.

In this study, we encountered problems with procuring teratogenicity information, given that teratology reference data are not published and updated in the relevant clinical literature very often. Furthermore, existing clinical decision-making tools like UpToDate and Medscape do not have APIs and contain contradictory teratology information that is not available in structured formats—as is required for systematic, retrospective data analysis and ML modeling. FDA resources containing teratology data are also not published in structured formats amenable for computational research, despite the availability an API for FDA pharmacopeias like DailyMed. For this investigation, the consequence of this limitation in available teratogenicity data was a significant reduction in drug sample size, as available to
t-SNE and GBM. Though we used one of the arguably most powerful chemical computing software programs currently available (i.e., MOE), we encountered sparsity in meta-structural predictions within our limited subset of drugs with available structure and teratogenicity information. This restricted the power of our meta-structural t-SNE and GBM probes, resulting in no test power for a feature-selected meta-structural and structural feature set.

Despite the gravity of the inherent uncertainty within available teratogenicity scoring criteria and limited target rationale for teratogenesis, there exist no teratology-specific HTS platforms. Though large toxicology HTS programs like Tox21 have screened targets that overlap with those in Table 2, this intersection remains small: only two (2) targets have coverage through Tox21. Therefore, though real-world bioactivity information is inherently powerful, we were able to access data on only two (2) relevant targets, and for only 128 drugs with structure and available teratogenicity data and assay information. Only sixty-four (64) drugs had information available for both RAR and HDAC, available structure data, and a known teratology score. Hence, a major reason why the addition of Tox21 HTS data did not improve predictive accuracy or t-SNE clustering over a purely structural model was limited sample size. This issue remains intractable, given the inherently limited data resources currently existing available and little action on the part of data providers to address these quality issues.

Finally, we note that we designed this study to remain as translational and open-source as possible, though we encountered significant barriers to model development from the lack of published teratology and HTS data, as well as the lack of granularity and contextual relevance within available teratology scoring protocols like those of the FDA. All data that we employed in our ML models were available publically, either from dedicatedly open-source databases or public disclosures of multi-institutional research initiatives. These databanks are well-referenced in the relevant cheminformatics literature, as they provide high-quality information on the structure, pharmacology, and teratology of small molecules, per what is currently published. To review the clinical applicability of the drugs that we studied, we applied standard-of-care clinical decision support tools like UpToDate and Medscape, which contain peer-reviewed and data-driven documentation for the guidance they present. Within these softwares, users may access the component publications that underlie the clinical decision support that the tools present. Furthermore, these tools—and their component data—are available at no individual charge to most investigators who belong to an institution with an associated patient care facility, as these suites benefit from high-frequency use by staff at most medical centers.

Conclusions

Current standards of evaluating small molecule teratogenicity are inherently unsystematic and driven on a lack of human data. This informs irresponsible prescribing behavior at the POC, reducing the quality of care for pregnant women and their developing fetuses. However, given the rigor of rules-based ML classification algorithms and limited “on-target” rationale for teratogenesis, there is potential to systematically predict a compound’s risk for fetal toxicity by leveraging AI on drug-specific information, such as drug structure, meta-
structure, and existing real-world bioassay data, as a proxy for binding affinity to teratogenic targets.

In our study, we assert that drug structure is a good predictor of teratogenicity, using ROC analysis, unsupervised ML (t-SNE), and a supervised GBM to discover relationships between chemical functionalities within drugs prescriptible in pregnancy and existing teratogenicity information. This allowed us to identify moieties that appear to predispose a drug towards an increased chance of teratogenicity, based on existing use cases that are salient in relevant clinical and drug development literature. We also identify significant barriers to translational research in this space as rationale for the limited utility of existing meta-structural and toxicology HTS platforms for teratogenicity prediction tasks. The importance of these ontological considerations cannot be overstated in considering future research to improve the quality of data-driven maternal-fetal medicine.

Our team of investigators has formed a first-in-kind research collaboration of engineers, informaticians, and clinicians dedicated to the development of computational tools to predict adverse drug outcomes in pregnancy from existing healthcare data on pregnant populations and in vitro drug exposure models that are more representative of pregnant human physiology than the in vivo animal platforms currently employed in this space. This group—called Modeling Adverse Drug Reactions in Embryos (MADRE)—proposes refinement of the teratogenicity QSAR reported in this manuscript by harnessing a more continuous spectrum of relevant phenotype information (Figure 5). Given that data quality and availability issues with teratogenicity scores restricted the scope of this study, we propose a medication history-wide association study (MedWAS) that can leverage billing-encoded, population-level EHR data as a label set. The benefit of MedWAS over QSAR is increased flexibility: associative study model architecture would not necessitate classification of adverse outcomes into rigid bins, as the QSAR requires. Therefore, MedWAS would not be restricted by the limited availability of FDA-encoded teratogenicity data, giving a larger sample size of drugs eligible for analysis and a more continuous spectrum of phenotype information through which to quantify teratogenicity. In turn, this allows for easier validation of associative outcomes in silico and in vitro, as compared to similar hits from QSAR. Indeed, drugs identified as teratogenic through MedWAS may be referred to our QSAR model for validation, and vice versa. We have begun work on this MedWAS and look forward to further exploring its intersections with our teratogenicity QSAR.

Acknowledgements

We thank Asher Schachter, MD, Senior Vice President, Clinical, and Head of Pharmaceutical Sciences at CAMP4 Therapeutics, for sharing teratogenicity data that he extracted from SafeFetus. We also thank Jeffery Goldstein, MD, PhD, Assistant Professor of Pathology at Northwestern University, for providing clinical consultation on our model and reviewing this manuscript.

Research reported in this publication was supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number U54HG007963–05 and the National Center for Advancing Translational Sciences of the National Institutes of Health under Clinical and Translational Science Award Number U54TR02243–02. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.
Appendix

Table 1.
Teratogenicity scoring criteria established by the FDA are driven by a lack of human data, making them dangerously imprecise for application at the bedside.7,13

| Classification | Attributes |
|----------------|------------|
| A              | Generally acceptable. Controlled studies in pregnant women show no evidence of fetal risk. |
| B              | May be acceptable. Either animal studies show no risk but human studies not available or animal studies showed minor risks and human studies done and showed no risk. |
| C              | Use with caution if benefits outweigh risks. Animal studies show risk and human studies not available or neither animal nor human studies done. |
| D              | Use only in life-threatening emergencies when no safer drug available. Positive evidence of human fetal risk. |
| X              | Do not use in pregnancy. Risks involved outweigh potential benefits. Safer alternatives exist. |
| N/A            | Information not available |

Table 2.
Teratogenesis converges on a limited subset of targets.19,44–53

| Target Class | Mechanism of Action |
|--------------|---------------------|
| dihydrofolate reductase (DHFR) | Inhibition of DHFR—both competitively and through antagonism of its folate cofactor—reduces the rate of purine and pyrimidine synthesis and DNA methylation reactions in a developing fetus. This leads to congenital malformations and neural tube defects. |
| retinoic acid receptor (RAR, RXR) | The nuclear ligand-inducible receptors RAR and RXR are mobile; they act as transcription factors for heavily-conserved developmental genes, including the Hox gene. Inhibition of these receptors leads to malformation of the neural crest. |
| androgen receptor (AR), estrogen receptor (ER) | Synthetic estrogens and androgens disrupt natural endocrine homeostasis within the developing fetus, resulting in errors of sexual differentiation. |
| prostaglandin H synthase (PHS), lipoygenase (LPO) | PHS and LPO activation results in increased rates of prototeratogen oxidation, generating ROS with the potential to attack fetal DNA and generate mutagenicity. |
| angiotensin II (ATII) and angiotensin converting enzyme (ACE) | ACE and ATII receptor inhibitors reduce perfusion to developing fetal tissues, which especially affects peripheral structures such as the distal limbs. These agents also decrease the tone of fetal vasculature, leading to cardiovascular morbidity. |
| hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase | HMG-CoA reductase inhibitors downregulate the conversion of HMG-CoA to mevalonic acid, an essential step in cholesterol synthesis. In the developing fetus, cholesterol is an essential progenitor of lipid regulators of the SHH gene, which affects fetal patterning and morphogenesis. Therefore, HMG-CoA inhibition is associated with severe fetal malformation and lipid deficiencies. |
| histone deacetylase (HDAC) | HDAC proteins are essential in regulating gene expression by promoting chromatin unwinding. Therefore, HDAC inhibitors lead to a wide spectrum of morbidities (e.g., axial skeletal malformations) and may be fetal lethal. |
| cyclooxygenase-1 (COX-1) | COX-1 inhibition is associated with cardiac, midline, and diaphragm defects, as the release of prostaglandins required for healthy morphogenesis is reduced by interference within the COX-1 signaling pathway. |
| N-methyl-D-aspartate receptor (NMDAR) | NMDAR inhibition is associated with gross structural defects within the brain, resulting from dysregulation of neuronal migration, synapse formation, and synapse elimination in the developing fetus. |
| 5-hydroxytryptamine (5-HT) receptor, 5-HT transporter | 5-HT is a neurotransmitter critical to craniofacial morphogenesis in development. Agents activating or inhibiting 5-HT—or promoting 5-HT reuptake—disrupt a critical 5-HT concentration, resulting in craniofacial malformations and other structural defects in the fetus. |
Target Class                                      Mechanism of Action
γ-aminobutyric acid (GABA) receptor            GABA is a key inhibitory neurotransmitter that guides healthy testicular, ovarian, pancreatic, enteric, and palatal morphogenesis at a critical concentration. Enhancers of GABA receptor are significantly associated in malformation of these tissues and are therefore implicated in morbidities such as cleft palate and atresia of the gastrointestinal tract.
carbonic anhydrase                             Carbonic anhydrase hydrates carbon dioxide to promote pH homeostasis through the carbonic acid-bicarbonate buffering system. Inhibitors of this target are therefore implicated in pH disruption during development, resulting in metabolic diseases and limb malformation from largescale misfolding of key proteins at non-physiological pH.

Abbreviations

| Abbreviation | Description                                      |
|--------------|--------------------------------------------------|
| RCT          | randomized controlled trial                      |
| FDA          | United States Food and Drug Administration       |
| POC          | point of care                                    |
| MOA          | mechanism of action                              |
| ROS          | reactive oxygen species                          |
| ML           | machine learning                                 |
| AI           | artificial intelligence                          |
| HER          | electronic health record                         |
| HTS          | high-throughput screening                        |
| QSAR         | quantitative structure-activity relationship     |
| MOE          | Molecular Operating Environment                  |
| Tox21        | Toxicology in the 21st Century Initiative        |
| API          | application programming interface                |
| 3D-SDF       | three-dimensional spatial data file              |
| t-SNE        | t-Distributed Stochastic Neighbor Embedding      |
| GBM          | gradient boosting machine                        |
| CV           | cross-validation                                 |
| RO5          | Lipinski’s Rule of Five                          |
| HOMO         | highest occupied molecular orbital               |
| LUMO         | lowest unoccupied molecular orbital              |
| MADRE        | Modeling Adverse Drug Reactions in Embryos       |
MedWAS medication history-wide association study

References

1. Orna Diav-Citrin MD Human Teratogens: A critical evaluation. https://www.nvp-volumes.org/p2_4.htm.
2. Garry VF & Truran P Chapter 62 - Teratogenicity in Reproductive and Developmental Toxicology (Second Edition) (ed. Gupta RC) 1167–1181 (Academic Press, 2017). 10.1016/B978-0-12-804239-7.00062-7.
3. Teratogenicity, pregnancy complications, and postnatal risks of antipsychotics, benzodiazepines, lithium, and electroconvulsive therapy - UpToDate. https://www.uptodate.com/contents/teratogenicity-pregnancy-complications-and-postnatal-risks-of-antipsychotics-benzodiazepines-lithium-and-electroconvulsive-therapy?search=teratogenicity&source=search_result&selectedTitle=2~150&usage_type=default&display_ranks=2.
4. Antenatal use of antidepressants and risk of teratogenicity and adverse pregnancy outcomes: Selective serotonin reuptake inhibitors (SSRIs) - UpToDate. https://www.uptodate.com/contents/antenatal-use-of-antidepressants-and-risk-of-teratogenicity-and-adverse-pregnancy-outcomes-selective-serotonin-reuptake-inhibitors-ssris?search=teratogenicity&source=search_result&selectedTitle=3~150&usage_type=default&display_ranks=3.
5. Garg RC, Bracken WM, Hoberman AM, Enright B & Tornesi B Chapter 6 - Reproductive and Developmental Safety Evaluation of New Pharmaceutical Compounds in Reproductive and Developmental Toxicology (Second Edition) (ed. Gupta RC) 101–127 (Academic Press, 2017). 10.1016/B978-0-12-804239-7.00006-8.
6. Cohen RL Evaluation of the teratogenicity of drugs. Clinical Pharmacology & Therapeutics 5, 480–514 (1964). [PubMed: 14203841]
7. CFR - Code of Federal Regulations Title 21. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=201.57.
8. Johnson CY Long overlooked by science, pregnancy is finally getting attention it deserves. Washington Post (2019),
9. Vargesson N Thalidomide‐induced teratogenesis: History and mechanisms. Birth Defects Res C Embryo Today 105, 140–156 (2015). [PubMed: 26043938]
10. Kim JH & Scialli AR Thalidomide: The Tragedy of Birth Defects and the Effective Treatment of Disease. Toxicol Sci 122, 1–6 (2011). [PubMed: 21507989]
11. PregOMICS—Leveraging systems biology and bioinformatics for drug repurposing in maternal-child health - Goldstein - 2018 - American Journal of Reproductive Immunology - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/aji.12971.
12. Goldstein JA et al. Calcium Channel Blockers as Drug Repurposing Candidates for Gestational Diabetes: Mining large scale genomic and electronic health records data to repurpose medications. Pharmacol Res 130, 44–51 (2018). [PubMed: 29448118]
13. Pulley JM et al. Using What We Already Have: Uncovering New Drug Repurposing Strategies in Existing Omics Data. Annu. Rev. Pharmacol. Toxicol (2019) 10.1146/annurev-pharmtox-010919-023537.
14. Phelan AL, Kunselman AR, Chuang CH, Raja-Khan NT & Legro RS Exclusion of Women of Childbearing Potential in Clinical Trials of Type 2 Diabetes Medications: A Review of Protocol-Based Barriers to Enrollment. Diabetes Care 39, 1004–1009 (2016). [PubMed: 27208329]
15. van der Graaf R et al. Fair inclusion of pregnant women in clinical trials: an integrated scientific and ethical approach. Trials 19, (2018).
16. Marić I et al. Data-driven queries between medications and spontaneous preterm birth among 2.5 million pregnancies. Birth Defects Research 0,
17. Characterizing cleft palate toxicants using ToxCast data, chemical structure, and the biomedical literature - Baker - - Birth Defects Research - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1002/bdr.2.1581?af=R.

Reprod Toxicol. Author manuscript; available in PMC 2020 October 21.
18. Pernia S & DeMaagd G. The New Pregnancy and Lactation Labeling Rule. P T 41, 713–715 (2016). [PubMed: 27904304]
19. van Gelder MMHJ et al. Teratogenic mechanisms of medical drugs. Hum. Reprod. Update 16, 378–394 (2010). [PubMed: 20061329]
20. Rudmann DG. On-target and off-target-based toxicologic effects. Toxicol Pathol 41, 310–314 (2013). [PubMed: 23085982]
21. Dixit VA. A simple model to solve a complex drug toxicity problem. Toxicol. Res 8, 157–171 (2019).
22. Paiva S-L. Avoiding target misidentification. Nature Reviews Drug Discovery 18, 826–826 (2019).
23. Beedie SL et al. Shared mechanism of teratogenicity of anti-angiogenic drugs identified in the chicken embryo model. Scientific Reports 6, 1–10 (2016). [PubMed: 28442746]
24. Goldstein JA, Bastarache LA, Denny JC, Pulley JM & Aronoff DM. PregOMICS-Leveraging systems biology and bioinformatics for drug repurposing in maternal-child health. Am. J. Reprod. Immunol e12971 (2018) 10.1111/aji.12971. [PubMed: 29726581]
25. Challa AP. Systematically Prioritizing Targets in Genome-Based Drug Repurposing. Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics - BCB ‘18; ACM Press; 2018. 543–543.
26. Ngiam KY & Khor IW. Big data and machine learning algorithms for health-care delivery. The Lancet Oncology 20, e262–e273 (2019). [PubMed: 31044724]
27. Miotto R, Wang F, Wang S, Jiang X & Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 19, 1236–1246 (2018). [PubMed: 28481991]
28. Challa AP, Madu CO & Lu Y. BRCA 1/2 Tumors and Gene Expression Therapy for Breast Cancer Development and Metastasis. Oncomedicine 2, 132–137 (2017).
29. Schachter AD & Kohane IS. Drug target-gene signatures that predict teratogenicity are enriched for developmentally related genes. Reprod Toxicol 31, 562–569 (2011). [PubMed: 21115113]
30. Marić I et al. Data-driven queries between medications and spontaneous preterm birth among 2.5 million pregnancies. Birth Defects Research Part A: Clinical and Molecular Teratology 106, 927–934 (2019). [PubMed: 31433567]
31. Abraham A, Bejan CA, Edwards D & Capra J. Resolving the Preterm Birth Phenotype Using Electronic Health Records and Genomic Biobanks [20A]. Obstetrics & Gynecology 133, 15S (2019).
32. Oh S et al. Physician Confidence in Artificial Intelligence: An Online Mobile Survey. J Med Internet Res 21, (2019).
33. Ailes EC, Simeone RM, Dawson AL, Petersen EE & Gilboa SM. Using insurance claims data to identify and estimate critical periods in pregnancy: An application to antidepressants. Birth Defects Research Part A: Clinical and Molecular Teratology 106, 927–934 (2016). [PubMed: 27891779]
34. Choi L et al. Evaluating statistical approaches to leverage large clinical datasets for uncovering therapeutic and adverse medication effects. Bioinformatics 34, 2988–2996 (2018). [PubMed: 29912272]
35. Boland MR, Polubriaginof F & Tatonetti NP. Development of A Machine Learning Algorithm to Classify Drugs Of Unknown Fetal Effect. Sci Rep 7, 1–15 (2017). [PubMed: 28127051]
36. Yusuf A et al. Use of existing electronic health care databases to evaluate medication safety in pregnancy: Triptan exposure in pregnancy as a case study. Pharmacoepidemiology and Drug Safety 27, 1309–1315 (2018). [PubMed: 30240072]
37. Liu Y, Zhang X, Zhang J & Hu C. Construction of a Quantitative Structure Activity Relationship (QSAR) Model to Predict the Absorption of Cephalosporins in Zebrafish for Toxicity Study. Front. Pharmacol. 10, (2019).
38. Lo Y-C, Rensi SE, Torng W & Altman RB. Machine learning in chemoinformatics and drug discovery. Drug Discovery Today 23, 1538–1546 (2018). [PubMed: 29750902]
39. Molecular Operating Environment (MOE) | MOEsaic | PSILO https://www.chemcomp.com/Products.htm.
40. Ames BN. Identifying environmental chemicals causing mutations and cancer. Science 204, 587–593 (1979). [PubMed: 3731222]
41. Toxicology in the 21st Century (Tox21). National Center for Advancing Translational Sciences [https://ncats.nih.gov/tox21](https://ncats.nih.gov/tox21) (2017).
42. Tox21 Data Browser. [https://tripod.nih.gov/tox21](https://tripod.nih.gov/tox21).
43. R: The R Project for Statistical Computing. [https://www.r-project.org/](https://www.r-project.org/).
44. DrugBank 5.0: a major update to the DrugBank database for 2018 | Nucleic Acids Research | Oxford Academic [https://academic.oup.com/nar/article/46/D1/D1074/4602867](https://academic.oup.com/nar/article/46/D1/D1074/4602867).
45. About DrugBank - DrugBank. [https://www.drugbank.ca/about](https://www.drugbank.ca/about).
46. Medication in Pregnancy and Breastfeeding | SafeFetus.com. [https://www.safefetus.com/](https://www.safefetus.com/).
47. DailyMed. [https://dailymed.nlm.nih.gov/dailymed/](https://dailymed.nlm.nih.gov/dailymed/).
48. Chen I-J & Foloppe N Drug-like bioactive structures and conformational coverage with the LigPrep/ConfGen suite: comparison to programs MOE and catalyst. J Chem Inf Model 50, 822–839 (2010). [PubMed: 20423098]
49. Cao Y, Charisi A, Cheng L-C, Jiang T & Girke T ChemmineR: a compound mining framework for R. Bioinformatics 24, 1733–1734 (2008). [PubMed: 18596077]
50. rdk package | R Documentation. [https://www.rdocumentation.org/packages/rdk/versions/3.4.7.1](https://www.rdocumentation.org/packages/rdk/versions/3.4.7.1).
51. Elsevier. Biomedical research – Embase | Elsevier [https://www.elsevier.com/solutions/embase-biomedical-research](https://www.elsevier.com/solutions/embase-biomedical-research).
52. Sieving PC WHAT IS A COCHRANE REVIEW? ORL Head Neck Nurs 25, 15 (2007). [PubMed: 18728734]
53. Maaten L van der. Accelerating t-SNE using Tree-Based Algorithms. Journal of Machine Learning Research 15, 3221–3245 (2014).
54. Oliveira FHM, Machado ARP & Andrade AO On the Use of t-Distributed Stochastic Neighbor Embedding for Data Visualization and Classification of Individuals with Parkinson’s Disease. Comput Math Methods Med 2018, (2018).
55. Platzer A Visualization of SNPs with t-SNE. PLoS One 8, (2013).
56. Gütlein M & Kramer S Filtered circular fingerprints improve either prediction or runtime performance while retaining interpretability. J Cheminform 8, (2016).
57. Garrison JA UpToDate. J Med Libr Assoc 91, 97 (2003).
58. Roberts J Takayasu Arteritis. MedScape (2016).
59. Kuhn M The caret Package.
60. Couronné R, Probst P & Boulesteix A-L Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 19, (2018).
61. Liu L et al. An interpretable boosting model to predict side effects of analgesics for osteoarthritis. BMC Syst Biol 12, (2018).
62. Zhang Z, Zhao Y, Canes A, Steinberg D & Lyashevska O Predictive analytics with gradient boosting in clinical medicine. Ann Transl Med 7, (2019).
63. TALMACIU MM, BODOKI E & OPREAN R Global chemical reactivity parameters for several chiral beta-blockers from the Density Functional Theory viewpoint. Clujul Med 89, 513–518 (2016). [PubMed: 27857521]
64. Pollastri MP Overview on the Rule of Five. Curr Protoc Pharmacol Chapter 9, Unit 9.12 (2010).
65. Ding LP et al. Understanding the structural transformation, stability of medium-sized neutral and charged silicon clusters. Sci Rep 5, (2015).
66. Huang R et al. Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. Nat Commun 7, 1–10 (2016).
67. Kenny HA et al. Quantitative high throughput screening using a primary human three-dimensional organotypic culture predicts in vivo efficacy. Nature Communications 6, 1–11 (2015).
68. Shen M et al. Identification of Therapeutic Candidates for Chronic Lymphocytic Leukemia from a Library of Approved Drugs. PLOS ONE 8, e75252 (2013). [PubMed: 24073257]
69. Zhu T et al. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based Upon a Critical Literature Analysis. J Med Chem 56, 6560–6572 (2013). [PubMed: 23688234]
70. Czeizel AE, Rockenbauer M, Sørensen HT & Olsen J. Use of cephalosporins during pregnancy and in the presence of congenital abnormalities: A population-based, case-control study. American Journal of Obstetrics and Gynecology 184, 1289–1296 (2001). [PubMed: 11349204]

71. Mylonas I. Antibiotic chemotherapy during pregnancy and lactation period: aspects for consideration. Arch. Gynecol. Obstet. 283, 7–18 (2011). [PubMed: 20814687]

72. Malka I & Ziv M. Safety of Common Medications for Treating Dermatology Disorders in Pregnant Women. Curr Derm Rep 2, 249–257 (2013).

73. Cephalexin: Drug information - UpToDate. https://www.uptodate.com/contents/cephalexin-drug-information?search=cephalosporin%20pregnancy&source=search_result&selectedTitle=2~150&usage_type=default&display_rank=2#F3017197.

74. Cephalosporin - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/neuroscience/cephalosporin.

75. Acar S, Keskin-Arslan E, Erol-Coskun H, Kaya-Temiz T & Kaplan YC. Pregnancy outcomes following quinolone and fluoroquinolone exposure during pregnancy: A systematic review and meta-analysis. Reprod. Toxicol 85, 65–74 (2019). [PubMed: 30738954]

76. Bar-Oz B, Moretti ME, Boskovic R, O’Brien L & Koren G. The safety of quinolones--a meta-analysis of pregnancy outcomes. Eur. J. Obstet. Gynecol. Reprod. Biol 143, 75–78 (2009). [PubMed: 19181435]

77. Aboubakr M, Elbadawy M, Soliman A & El-Hewaity M. Embryotoxic and teratogenic effects of norfloxacin in pregnant female albino rats. Adv Pharmacol Sci 2014, 924706 (2014). [PubMed: 24639869]

78. Walker RC & Wright AJ. The Fluoroquinolones. Mayo Clinic Proceedings 66, 1249–1259 (1991). [PubMed: 1749294]

79. Bandoli G, Palmsten K, Forbess Smith CJ & Chambers CD. A review of systemic corticosteroid use in pregnancy and the risk of select pregnancy and birth outcomes. Rheum Dis Clin North Am 43, 489–502 (2017). [PubMed: 28711148]

80. Clinical use of corticosteroids in pregnancy | Human Reproduction Update | Oxford Academic https://academic.oup.com/humupd/article/22/2/240/2457843.

81. Major side effects of systemic glucocorticoids - UpToDate. https://www.uptodate.com/contents/major-side-effects-of-systemic-glucocorticoids?search=steroid%20pregnancy&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1#H3598344855.

82. Ji Y, Guo Q, Yin Y, Blachier F & Kong X. Dietary proline supplementation alters colonic luminal microbiota and bacterial metabolite composition between days 45 and 70 of pregnancy in Huanjiang mini-pigs. J Anim Sci Biotechnol 9, (2018).

83. Proline - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/nursing-and-health-professions/proline.

84. Liu N et al. Maternal L-proline supplementation enhances fetal survival, placental development, and nutrient transport in mice. Biol. Reprod 100, 1073–1081 (2019). [PubMed: 30418498]

85. Karalis DG, Hill AN, Clifton S & Wild RA. The risks of statin use in pregnancy: A systematic review. Journal of Clinical Lipidology 10, 1081–1090 (2016). [PubMed: 27678424]

86. Re: Statins and The BMJ (Statin use in pregnancy). (2019).

87. HMG-CoA Reductase - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/hmg-coa-reductase.

88. Patel J, Sheehan V & Gurk-Turner C. Ezetimibe (Zetia): a new type of lipid-lowering agent. Proc (Baylor Univ Med Cent) 16, 354–358 (2003). [PubMed: 16278712]

89. Ezetimibe: Drug information - UpToDate. https://www.uptodate.com/contents/ezetimibe-drug-information?search=ZETIA&source=panel_search_result&selectedTitle=1~51&usage_type=panel&kp_tab=drug_general&display_rank=1#F169554.

90. Snyder B. Effort seeks to improve safety of drugs given during pregnancy. Vanderbilt University http://news.vumc.org/2019/07/18/effort-seeks-to-improve-safety-of-drugs-given-during-pregnancy/.

Reprod Toxicol. Author manuscript; available in PMC 2020 October 21.
91. Identifying Drugs Safe for Use During Pregnancy. Vanderbilt Discover https://discover.vumc.org/2019/07/identifying-drugs-safe-for-use-during-pregnancy/ (2019).

92. Schnell JR, Dyson HJ & Wright PE Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. Annual Review of Biophysics and Biomolecular Structure 33, 119–140 (2004).

93. Kam RKT, Deng Y, Chen Y & Zhao H Retinoic acid synthesis and functions in early embryonic development. Cell Biosci 2, 11 (2012). [PubMed: 22439772]

94. Hussey HH Teratogenic Effects of Progestogen/Estrogen. JAMA 230, 1019–1020 (1974). [PubMed: 4479553]

95. Smith AS, Birnie AK & French JA Maternal Androgen Levels During Pregnancy are Associated with Early-life Growth in Geoffroy’s Marmosets, Callithrix geoffroyi. General and comparative endocrinology 166, 307 (2010). [PubMed: 19854190]

96. Yu WK & Wells PG Evidence for Lipoxygenase-Catalyzed Bioactivation of Phenytoin to a Teratogenic Reactive Intermediate: In Vitro Studies Using Linoleic Acid-Dependent Soybean Lipoxygenase, and in Vivo Studies Using Pregnant CD-1 Mice. Toxicology and Applied Pharmacology 131, 1–12 (1995). [PubMed: 7878664]

97. Fitton C et al. In-utero exposure to antihypertensive medication and neonatal and child health outcomes: a systematic review. Journal of Hypertension 35, 2123–2137 (2017). [PubMed: 28661961]

98. Andaloro VJ, Monaghan DT & Rosenquist TH Dextromethorphan and Other N-Methyl-D-Aspartate Receptor Antagonists Are Teratogenic in the Avian Embryo Model. Pediatric Research 43, 1–7 (1998). [PubMed: 9432105]

99. Licheri V et al. Plasticity of GABAA Receptors during Pregnancy and Postpartum Period: From Gene to Function. Neural Plasticity 2015, (2015).

100. Maren TH Teratology and Carbonic Anhydrase Inhibition. Arch Ophthalmol 85, 1–2 (1971). [PubMed: 5099659]

101. Robin X et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77 (2011). [PubMed: 21414208]

102. Rtsne function | R Documentation. https://www.rdocumentation.org/packages/Rtsne/versions/0.15/topics/Rtsne.
Figure 1:
To develop our QSAR model, we synthesized data from several pharmacological and clinical databases. However, as we describe in “Materials and Methods,” the quantity and quality of available data differed significantly across the ontologies, with teratogenicity data from SafeFetus as the most limiting. To reflect this heterogeneity, data sources and conclusions that we consider substantial are colored green, while yellow-colored data elements are intermediate in their contribution of meaningful information. Sources in red are most limiting.
Figure 2:
ROC analysis suggests that 1,024-bit Morgan fingerprints have good predictive accuracy for teratogenicity (AUC = 0.78). This plot was generated using the R package pROC (CRAN: pROC).
t-SNE—when enacted on a 1,024-bit representation of the Morgan class of chemical fingerprints and a binary classification of teratogenicity (“YES” (class A/B), “NO” (class C/D/X))—reveals small clusters that indicate potential structure-teratogenicity relationships. This plot was generated using the R package Rtsne (CRAN: Rtsne)\textsuperscript{102}.
Figure 4:
We discovered relationships between teratogenic risk ("YES", "NO") and the presence of distinct chemical functionalities from consistent structure-teratogenicity points within each discrete t-SNE cluster.
Figure 5:
Our team—dubbed Modeling Adverse Drug Reactions in Embryos (MADRE)—leverages a broad knowledge base across the basic, applied, and clinical sciences to develop predictive models of adverse drug outcomes in pregnancy. We leverage the strengths of all sites within our network to optimize both the quantity and quality of data and analytical expertise that are essential to our QSAR and MedWAS models.