© 2013 Wiley Periodicals, Inc. Birth Defects Research (Part B) 00:1–21 (2013)

Original Article

Establishment and Assessment of a New Human Embryonic Stem Cell-Based Biomarker Assay for Developmental Toxicity Screening

Jessica A. Palmer,* Alan M. Smith, Laura A. Egnash, Kevin R. Conard, Paul R. West, Robert E. Burrier, Elizabeth L.R. Donley, and Fred R. Kirchner

Stemina Biomarker Discovery, Inc., Madison, Wisconsin

A metabolic biomarker-based in vitro assay utilizing human embryonic stem (hES) cells was developed to identify the concentration of test compounds that perturbs cellular metabolism in a manner indicative of teratogenicity. This assay is designed to aid the early discovery-phase detection of potential human developmental toxicants. In this study, metabolomic data from hES cell culture media were used to assess potential biomarkers for development of a rapid in vitro teratogenicity assay. hES cells were treated with pharmaceuticals of known human teratogenicity at a concentration equivalent to their published human peak therapeutic plasma concentration. Two metabolite biomarkers (ornithine and cystine) were identified as indicators of developmental toxicity. A targeted exposure-based biomarker assay using these metabolites, along with a cytotoxicity endpoint, was then developed using a 9-point dose–response curve. The predictivity of the new assay was evaluated using a separate set of test compounds. To illustrate how the assay could be applied to compounds of unknown potential for developmental toxicity, an additional 10 compounds were evaluated that do not have data on human exposure during pregnancy, but have shown positive results in animal developmental toxicity studies. The new assay identified the potential developmental toxicants in the test set with 77% accuracy (57% sensitivity, 100% specificity). The assay had a high concordance ($\geq$75%) with existing in vivo models, demonstrating that the new assay can predict the developmental toxicity potential of new compounds as part of discovery phase testing and provide a signal as to the likely outcome of required in vivo tests. *Birth Defects Res (Part B)* 00:1–21, 2013. © 2013 Wiley Periodicals, Inc.

Key words: ***in vitro developmental toxicity screen; human embryonic stem cells; biomarker identification; metabolomics; in vitro toxicology; mechanisms of teratogenesis; valproic acid; retinoic acid***

INTRODUCTION

Birth defects are reported in approximately 3% of all human births and are the largest cause of infant mortality in the United States (Hoyert et al., 2006). Exposure to toxic chemicals and physical agents is believed to be responsible for approximately 3% of all birth defects (National Research Council (NCR), 2000). Our goal was to develop an exposure-based human embryonic stem (hES) cell in vitro assay by measuring a metabolic perturbation in the culture media that could be used as an early signal for the potential of developmental toxicity. The teratogenic potential of a compound is associated with the level of exposure to the fetus. Therefore, a compound could be considered both teratogenic and nonteratogenic depending on the exposure level. For example, retinol (vitamin A), when taken at or below the Food and Drug Administration maximum recommended daily allowance (8,000 IU), does not have an adverse effect on the developing fetus. However, high doses of retinol (>25,000 IU/day) have been shown to cause malformations similar to those seen following 13-*cis* retinoic acid (isotretinoin) exposure in both experimental animals and humans (Teratology Society, 1987). Retinol concentrations are homeostatically regulated in plasma and remain constant even when doses as large as 30,000 IU are taken (Blomhoff et al., 2003; Hartmann et al., 2005). In contrast, the concentrations of retinol's teratogenic metabolites (all-*trans* retinoic acid, 13-*cis* retinoic acid) increase with increasing doses of retinol (Hartmann et al., 2005).

The thalidomide tragedy in the 1960s emphasized the importance of preclinical developmental toxicity testing, the significant differences among species in their response

---

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Science Foundation; Grant number: IIP-1058355.

*Correspondence to: Jessica A. Palmer, Stemina Biomarker Discovery, Inc., 504 S. Rosa Road, Suite 150, Madison, WI 53719. E-mail: jpalmer@stemina.com

Received 6 August 2013; Accepted 12 September 2013

to potentially teratogenic compounds, and how the developing fetus can be affected by such compounds. Developmental toxicity testing of thalidomide in rodent models did not indicate the compound's teratogenic potential in humans. Over 10,000 children were born with severe birth defects following in utero exposure. Current preclinical models for detecting developmental toxicity have varying degrees of concordance with observed developmental toxicity in humans, with rats and rabbits (the most commonly used species for developmental toxicity testing) having approximately 70–80% concordance to known human teratogens (Daston and Knudsen, 2010). These decades-old "Segment II" in vivo animal models require large numbers of animals, kilogram quantities of test compound, and are both time consuming and expensive. Due to the cost and complexity of these models, safety assessments often occur too late in the compound's life cycle for the developer to react to a positive developmental toxicity signal, and can result in the termination of the development of the compound or series. Though these animal models are, and have long been, considered the regulatory gold standard, differences in species response to a compound may lead to missed signals of developmental toxicity and biological misinterpretation. As such, the development of a new generation of tools using human cells for assessment of potential developmental toxicity risk related to chemical exposure is needed. The appropriate tests would also reduce product development time, control costs, and respond proactively to the call to decrease animal use. Development of predictive in vitro alternatives using hES cells for developmental toxicity testing could address all of these needs, and focus continued development on compounds with a higher potential for success. In its report, "Toxicity Testing in the 21st Century: A Vision and Strategy" (NCR, 2007), the United States NRC presents a vision for the future wherein toxicity testing is done largely in vitro using human cell lines. There is much work to be done toward achieving this future vision and there is a clear demand for development of highly relevant, predictive, low cost, and rapid human in vitro tests.

hES cells are an innovative in vitro model system that is metabolically similar to embryonic epiblast cells at gastrulation. These cells can be used to predict developmental toxicity of new chemical entities (Ebert and Svendsen, 2010; West et al., 2010; Kleinstreuer et al., 2011; Tandon and Jyoti, 2012). Our unique metabolomics platform profiles change in metabolism that can be measured in the spent cell culture medium from hES cells following compound exposure. This "metabolic footprint" of the cultured medium is a functional measurement of cellular metabolism referred to as the secretome. The "secretome" includes the metabolites present in the spent media (i.e., cell culture supernatant) and is comprised of media components, metabolites passively and actively transported across the plasma membrane, and those produced through extracellular metabolism of enzymes. The change in the secretome elicited by test compound exposure produces a metabolic signature of toxicity that is related to alterations that occur both in the endometabolome (inside the cell) and alteration of the extracellular matrix. The secretome is measured specifically because of several unique qualities for profiling cell culture media: it is very easy to reproducibly sample, minimal handling is required to quench metabolism, it does not destroy the cells that can then be used for other assays, it is amenable to high-throughput evaluation, and strong signals can be measured due to the accumulation of metabolites over time. The ability to measure metabolic changes following compound exposure resulted in the identification of new biomarkers associated with disruption of human development and provided the opportunity to develop highly predictive models of developmental toxicity based on these changes. Our previous work established that an untargeted metabolomics-based evaluation of hES cell spent media following exposure to compounds with known human teratogenicity outcomes produced a predictive signature that could be utilized as a developmental toxicity screen (West et al., 2010; Kleinstreuer et al., 2011). This work led to the development of the targeted biomarker assay described here in Phase 2.

This present research describes the development of a rapid, reproducible, biomarker-based screen for developmental toxicity testing designed to identify the exposure level at which a test compound exhibits teratogenic potential. Perturbation of two metabolites, ornithine and cystine, in response to the test compound was assessed across nine independent experimental replications to ensure repeatability across experiments and liquid chromatography high resolution mass spectrometry (LC-HRMS) systems. Using the ornithine/cystine ratio (o/c ratio), we developed a rapid, targeted assay that measured changes in metabolism and cellular viability across a 9-point dose–response curve to determine the exposure level at which a test compound perturbs metabolism in a manner associated with developmental toxicity potential. To assess the predictivity of the assay for known human teratogens and nonteratogens in the training and test sets of compounds (see Table 1), the exposure level where a compound was predicted to have developmental toxicity potential was scored against the compound's human peak plasma in vivo concentration ($C_{\max}$) following therapeutic doses. The $C_{\max}$ value in this case was used as a benchmark exposure level to aid in interpretation of the performance of the assay as it is the highest concentration a human would normally be exposed to under therapeutic circumstances and we would expect to detect developmental toxicity at this exposure level. However, application of the assay in the discovery stage of a compound's development would not require this $C_{\max}$ information, and a test compound's teratogenic potential would be based on the exposure level at which a test compound perturbs metabolism in a manner indicative of teratogenicity. The design and sensitivity of the assay allows for identification of teratogenic potential at noncytotoxic levels of the test compound by negating the confounding effects of changes in metabolite abundance due strictly to cytotoxicity. The ability to identify developmental toxicity in the absence of cytotoxicity at a variety of exposure levels is a key strength of the assay and distinguishes it from existing in vitro assays.

MATERIALS AND METHODS

Development and evaluation of the targeted biomarker-based assay was conducted in two phases:

Table 1
Useful Terms and Definitions

Term	Definition
Teratogenicity threshold	A threshold of metabolic perturbation that is associated with the potential for teratogenesis. The threshold was empirically determined to be 0.88 for the targeted biomarker assay using the training set results. This threshold was applied to all test and application set compounds evaluated using the assay
Ornithine/cystine ratio (o/c ratio)	The fold change of ornithine (*Orn*) for treatment *x* divided by the fold change of cystine (*Cyss*) for treatment *x* $o/c\ Ratio_x = \frac{Orn_x/Orn_{DMSO}}{Cyss_x/Cyss_{DMSO}}$
Teratogenicity potential	Interpolated exposure level (concentration) of a test compound where the dose–response curve for the o/c ratio or cell viability crosses the teratogenicity threshold. Exposure levels greater than this concentration are associated with teratogenicity
Accuracy	Number of correct predictions divided by the number test compounds evaluated
Sensitivity	Detection of teratogens, true positives/(false negatives + true positives)
Specificity	Detection of nonteratogens, true negatives / (true negatives + false positives)
Training set	Set of compounds that have well established human developmental toxicity information used to identify biomarkers of developmental toxicity. This set of compounds was tested in both phases of the study and used to set the teratogenicity threshold
Test set	Set of compounds that have well-established human developmental toxicity information that were not used to identify the biomarkers, but used to evaluate the predictivity of the biomarkers of developmental toxicity. This set of compounds was used to evaluate the performance of the targeted biomarker assay and the teratogenicity threshold set using the training set
Application set	Set of compounds with poorly defined human developmental toxicity information used to demonstrate application of the assay. These compounds are not classified as a teratogen or nonteratogen based on their $C_{\max}$ since human teratogenicity is unknown at this concentration

- In the first phase, the predictive potential of two previously identified predictive biomarkers (ornithine and cystine, Kleinstreuer et al., 2011) was characterized across nine independent experimental replications (experimental blocks) of the training set using untargeted metabolomic methods.
- In the second phase, the predictive biomarkers were used to develop a rapid turnaround, targeted, exposure-based assay for compound prioritization based on teratogenicity potential. The predictivity of the new assay was evaluated using the original training set as well as an independent test set of compounds.

Test Chemical Selection and Classification

A total of 46 compounds were used to evaluate the ability of ornithine, cystine, and the o/c ratio to predict developmental toxicity in two experimental phases. These 46 compounds were divided into three groups, named the training, test, and application sets (Table 1). The training set consisted of 23 well-characterized pharmaceutical compounds (11 known human nonteratogens and 12 known human teratogens, Table 2) and was previously used to build a computational model and identify biomarkers predictive of teratogenicity (Kleinstreuer et al., 2011). This training set was utilized in both experimental phases. To assess the predictive capacity of the targeted biomarker assay developed in these studies, an additional test set of 13 well-characterized pharmaceutical compounds (six known human nonteratogens and seven known human teratogens, Table 3) was used in the second experimental phase to evaluate the predictivity of the new assay. The final set of compounds (the application set, Table 4) consists of 10 compounds that do not have conclusive developmental toxicity data available on exposure during human pregnancy, but do have animal data available on developmental toxicity potential. A two-class system of compound classification (teratogen and nonteratogen) was applied for assay development, focusing the teratogenicity classification strictly on observed human risk associated with each chemical. Compounds were purchased from Sigma-Aldrich (St. Louis, MO), except for amprenavir, bosentan, entacapone (Toronto Research Chemicals, Toronto, ON, Canada), lapatinib (Chemie Tek, Indianapolis, IN), cidovofir and ramelteon (Selleck Chemicals, Houston, TX).

Undifferentiated hES Cell Line Maintenance (Phases 1 and 2)

WA09 hES cells were obtained from the WiCell Research Institute (Madison, WI) and were maintained in feeder free conditions using mTeSR1 media (StemCell Technologies, Vancouver, BC, Canada) on hESC-qualified Matrigel (BD Biosciences, San Jose, CA) coated 6-well plates. To maintain the undifferentiated stem cell population, differentiated colonies were removed daily through aspiration and media was replaced. Additionally, the hES cells were only used in experiments up to passage 40 and were karyotyped approximately every 10 passages to minimize and monitor the potential for genetic instability. hES cells were passaged at 90–95% confluency (approximately every 7 days) using Versene (Life Technologies, Grand Island, NY). Cell cultures were maintained at 37°C under 5% CO_2.

Table 2
Description of the Training Set Compounds

Compound	Pharmacology/ chemical class	FDA pregnancy category[a]	Preclinical in vivo and known human developmental effects[b]
Human nonteratogens			
Ascorbic acid	Vitamin	A	None
Caffeine	Central nervous system stimulant	C	Low doses: none; high doses: limb, craniofacial, embryo toxicity[c]
Diphenhydramine	Antihistamine/H1 histamine receptor antagonist	B	None
Doxylamine	Antihistamine/H1 histamine receptor antagonist	B	None
Folic acid	Vitamin	A	None
Isoniazid	Antibacterial/antitubercular	C	None
Levothyroxine	Synthetic hormone	A	None
Penicillin G	Antibiotic	B	None
Retinol	Vitamin	C	Low doses: none; high doses: craniofacial, central nervous system, cardiovascular, skeletal
Saccharin	Artificial sweetener	A	None
Thiamine	Vitamin	A	None
Human teratogens			
13-*cis* Retinoic acid	RAR/RXR ligand	X	Craniofacial, limb, central nervous system, cardiovascular, skeletal
5-Fluorouracil	Antineoplastic/antimetabolite	D	Craniofacial, central nervous system, skeletal
All-*trans* retinoic acid	RAR/RXR ligand	D	Craniofacial, limb, central nervous system, cardiovascular, skeletal, embryo toxicity[c]
Busulfan	Antineoplastic/alkylating	D	Craniofacial, limb, embryo toxicity[c]
Carbamazepine	Anticonvulsant	D	Craniofacial, central nervous system, cardiovascular
Cytosine arabinoside	Antineoplastic/antimetabolite	D	Limb
Diphenylhydantoin	Anticonvulsant	D	Craniofacial, limb, cardiovascular, neurobehavioral
Hydroxyurea	Antineoplastic/enzyme inhibitor	D	Central nervous system, craniofacial, limb, cardiovascular, embryo toxicity[c]
Methotrexate	Antineoplastic/dihydrofolate acid reductase inhibitor	X	Craniofacial, limb, skeletal, central nervous system, embryo toxicity[c]
Thalidomide	Immunomodulant	X	Craniofacial, cardiovascular, limb, embryo toxicity[c]
Valproic acid	Anticonvulsant/GABA inhibitor	D	Central nervous system, craniofacial, cardiovascular, skeletal, neurobehavioral, embryo toxicity[c]
Warfarin	Anticoagulant	X	Central nervous system, craniofacial, skeletal, embryo toxicity[c]

[a] FDA classification requirements described in Shuren, 2008.

[b] The preclinical in vivo and known human developmental effects were summarized from the Teratogen Information System (TERIS, http://depts.washington.edu/terisweb/teris/) and Briggs et al. (2011).

[c] Embryo toxicity in addition to teratogenic effects (e.g., growth retardation, embryo lethality).

96-Well hES Cell Plating (Phases 1 and 2)

All experimental treatments were carried out in 96-well plates. To minimize plating variability and increase reproducibility, hES cells were plated as a single cell suspension and maintained in an undifferentiated state during compound exposure. Before plating in the 96-well plates, hES cells were removed from a 6-well plate using TrypLE (Life Technologies). The cells were washed with DMEM/F12 (Life Technologies) and resuspended in mTeSR1 containing 10 μM Y27632 Rho-associated kinase inhibitor (Merck KGaA/Calbiochem, Darmstadt, Germany). The rho-associated kinase inhibitor is added to the plating media to increase plating efficiency by decreasing dissociation-induced apoptosis. The inner 60 wells of hESC-qualified Matrigel coated 96-well plates were seeded at a density of 100,000 cells per well. The outer wells of the plate contained an equal volume media to minimize differences in humidity across the plate. Compound exposure began 24 hr after plating.

hES Cell Compound Exposure

(1) Phase 1: hES cells were treated with a test compound at a single concentration equivalent to the compound's published therapeutic, C_{max}. The therapeutic C_{max} was used because it is considered to be a physiologically relevant exposure level and has been correlated with the developmental effect of the compound (NCR, 2000). For six compounds

Table 3
Description of the Test Set Compounds

Compound	Pharmacology/ chemical class	FDA pregnancy category[a]	Preclinical in vivo and known human developmental effects[b]
Human nonteratogens			
Acetaminophen	Analgesic	B	None
Acycloguanosine	Antiviral	B	None
Amoxicillin	Antibiotic	B	None
Loratadine	Antihistamine/H1 histamine receptor antagonist	B	None
Metoclopramide	Antiemetic	B	None
Sitagliptin	Hypoglycemic	B	Low doses: none; high doses: skeletal
Human teratogens			
Aminopterin	Antineoplastic/dihydrofolate acid reductase inhibitor	X	Craniofacial, limb, skeletal, central nervous system
Bosentan	Antihypertensive	X	Craniofacial, cardiovascular
D-penicillamine	Chelator	D	Skeletal
Everolimus	Immunosuppressive	D	Skeletal, embryo toxicity[c]
Lapatinib	Antineoplastic/protein kinase inhibitors	D	Skeletal, embryo toxicity[c]
Lovastatin	Anticholesteremic	X	Skeletal, embryo toxicity[c]
ThioTEPA	Antineoplastic/Alkylating	D	Skeletal, embryo toxicity[c]

[a] FDA classification requirements described in Shuren, 2008.

[b] The preclinical in vivo and known human developmental effects were summarized from the Teratogen Information System (TERIS, http://depts.washington.edu/terisweb/teris/) and Briggs et al. (2011).

[c] Embryo toxicity in addition to teratogenic effects (e.g., growth retardation, embryo lethality).

Table 4
Description of the Application Set Compounds

Compound	Pharmacology/ chemical class	FDA pregnancy category[a]	Preclinical in vivo developmental effects[b]
6-Aminonicotinamide	Nicotinic acid antagonist	NA	Craniofacial
Abacavir	Anti-HIV	C	Skeletal, embryo toxicity[c]
Adefovir dipivoxil	Antiviral	C	None
Amprenavir	Anti-HIV	C	Skeletal, embryo toxicity[c]
Artesunate	Antimalarial	NA	Cardiovascular, skeletal, embryo toxicity[c,d]
Cidofovir	Antiviral	C	None
Entacapone	Antiparkinson	C	Eye defects
Fluoxetine	Serotonin reuptake inhibitor	C	Embryo toxicity[c]
Ramelteon	Sedative/hypnotics	C	None
Rosiglitazone	Hypoglycemic	C	Embryo toxicity[c]

[a] FDA classification requirements described in Shuren (2008).

[b] The preclinical in vivo developmental effects were summarized from the Teratogen Information System (TERIS, http://depts.washington.edu/terisweb/teris/) and Briggs et al. (2011).

[c] Embryo toxicity in addition to teratogenic effects (e.g., growth retardation, embryo lethality).

[d] Clark (2009).

(5-fluorouracil, aminopterin, busulfan, cytosine arabinoside, hydroxyurea, and methotrexate), an experimentally determined IC_{30} was used in place of the C_{max} value due to greater than 30% cytotoxicity at the C_{max} exposure level. This was done to ensure that enough cells were present at the time of sample collection to provide a signal for LC-HRMS analysis. For test compound exposure, all compound stock solutions, with the exception of valproic acid, were made with dimethyl sulfoxide (DMSO, Sigma-Aldrich). Valproic acid was insoluble in DMSO at the concentrations used in this study, so it was diluted in mTeSR1 containing 0.1% DMSO. Each 96-well plate included media controls with and without test compound, 0.1% DMSO solvent control cells, and cells exposed to a single concentration of eight different test compounds (Fig. 1A). Media controls were included on each plate to assess the impact of test compound on the sample matrix. hES cells were exposed to the test compound for 72 hr, with media and test compound replacement every 24 hr. Cells were monitored throughout the treatment period to ensure that no differentiation was occurring. After 72 hr of treatment, the spent media from the final 24-hr treatment period was collected and added to acetonitrile (Sigma-Aldrich, final acetonitrile

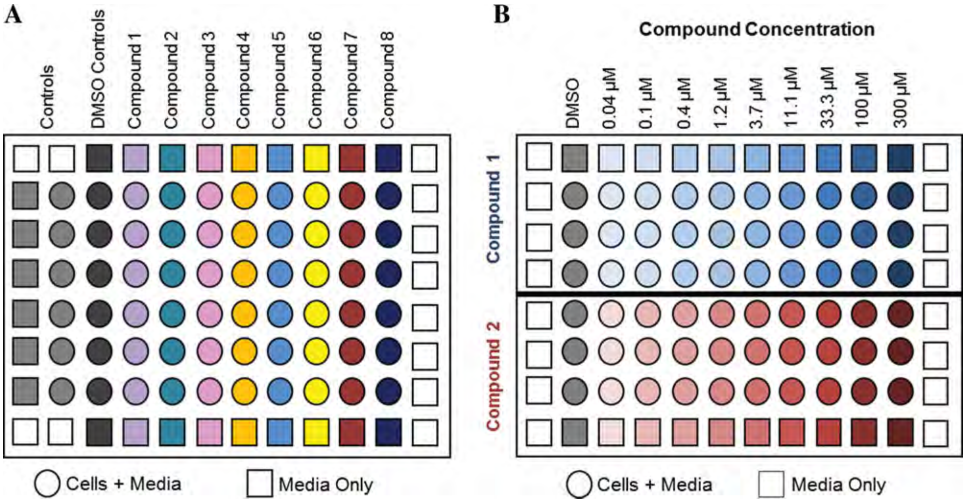

Fig. 1. Plate design for untargeted metabolomics treated at single exposure levels used in Phase 1 experiments (A) and targeted biomarker experiments treated at multiple exposure levels used for Phase 2 experiments (B). Both plates incorporate a reference design where the experimental control or reference treatment (0.1% DMSO) is present on each plate. Media only (lacking cells) controls are used to assess the impact of the test compounds on the sample matrix. Each well is analyzed as an individual sample. Filled circles represent cell samples and filled squares depict media control samples.

concentration 40%) to halt metabolic processes and precipitate proteins from solution. Individual wells from each 96-well plate were collected and analyzed as separate samples. These samples were then stored at −80°C until prepared for LC-HRMS analysis. Cell viability was assessed using the CellTiter-Fluor Cell Viability Assay as per the manufacturer's instructions (Promega, Madison, WI). Quality control parameters were set such that if the coefficient of variation (CV) for the viability relative fluorescent units (RFU) of the six cellular samples in a treatment exceeded 10% and no outliers were identified using the Grubb's test (http://graphpad.com/quickcalcs/Grubbs1.cfm), analysis was halted for that compound and the cell culture experiment was repeated. If outliers were present, the outlier sample was removed from analysis. If the CV for the DMSO control cell samples on a plate were outside of the quality control parameters, the entire plate was repeated. hES cell exposure to each of the 23 compounds was replicated a total of nine times.

(2) Phase 2: The predictivity of the targeted biomarker assay was evaluated in the original training set as well as an independent test set (Tables 2 and 3). The assay was also applied to the application set of compounds (Table 4) to demonstrate utility when human teratogenicity is unknown. The standard compound exposure levels used for most compounds were nine, threefold dilutions ranging from 0.04–300 µM (Fig. 1B). The exposure range for valproic acid was increased to 4–30,000 µM because its therapeutic C_{max} was outside the standard exposure range. Compounds that were cytotoxic at concentrations below 1 µM were repeated at lower exposure levels (0.001–10 µM). A stock solution of each test compound was prepared in 100% DMSO at a concentration of 1,000 times the highest exposure level, with the exception of ascorbic acid, folic acid, and valproic acid. These three compounds were completely insoluble in DMSO and stocks were prepared in mTeSR1 containing 0.1% DMSO. The stock solution was diluted 1:1,000 in mTeSR1 media and subsequent dilutions were performed in mTeSR1 containing 0.1% DMSO such that the final concentration of DMSO was 0.1% in all treatments. hES cells were treated for 72 hr and spent media from the last 24-hr treatment period was collected and added to acetonitrile containing $^{13}C_6$-labeled arginine (Cambridge Isotope Laboratories, Andover, MD) as described under Phase 1. Spent media samples were stored at −80°C until prepared for LC-HRMS analysis. Cell viability was assessed using the CellTiter-Fluor Cell Viability Assay. A quality control step was included with criteria that the CV of the measured viability RFU of the DMSO control cells could not exceed 10% for a plate to undergo LC-HRMS analysis. A dose–response curve was fit to the reference treatment (0.1% DMSO treated control cells) normalized data $\left(\frac{ViabilityRFU_{TrtX}}{ViabilityRFU_{DMSO}}\right)$ using a four-parameter log-logistic model with the R package "drc" (Ritz and Streibig, 2005).

Sample Preparation (Phases 1 and 2)

High molecular weight constituents (>10 KDa) of the spent media samples were removed using a Millipore Multiscreen Ultracel-10 filter plate (EMD Millipore, Billerica, MA). Before sample filtration, the filter plate was washed with 0.1% NaOH to remove a known contaminant polymer. The plate was then rinsed twice with HPLC-grade water to remove residual polymers and NaOH. Spent media samples were added to the washed filter plate. In Phase 1, samples were spiked with $^{13}C_6$-labeled arginine. Samples were centrifuged at 2,000 × *g* at 4°C for 200 min. The filtrate was collected and concentrated overnight in a Savant High Capacity Speed-vac Plus Concentrator. The concentrated sample was resolubilized in a 1:1 0.1% formic acid in water: 0.1% formic acid in acetonitrile mixture containing $^{13}C_5$-labeled glutamic acid (Cambridge Isotope Laboratories). The ^{13}C-labeled compounds were used as internal standards to track preparatory efficiency and LC-HRMS performance.

Mass Spectrometry

(1) Phase 1: LC-HRMS data were acquired for nine biological replications on three separate LC-HRMS systems with three replications evaluated on each system. Each system consisted of an Agilent 1290 Infinity LC system interfaced either with an Agilent G6520A QTOF high-resolution mass spectrometer (QTOF HRMS), an Agilent G6530A QTOF HRMS, or an Agilent G6224A TOF HRMS system (Agilent Technologies, Santa Clara, CA). To facilitate separation of biological small molecules with a wide range of structures and to allow increased retention of hydrophilic species, hydrophilic interaction liquid chromatography (HILIC) was utilized. A Luna HILIC column (Phenomenex, Torrance, CA) with dimensions 3 × 100 mm and 3 μm particle size was used and maintained at 30°C. Sample (2 μl) was injected and the data acquisition time was 23 min at a flow rate of 0.5 ml/min, using a 17-min solvent gradient with 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Electrospray ionization (ESI) was employed using a dual ESI source. The scan range of the instrument was 70–1,600 Da. Data acquisition was performed with MassHunter Acquisition software (version B 04.00, Agilent Technologies) using high-resolution exact mass conditions and each set of samples was run first under ESI-positive polarity then under ESI-negative polarity conditions.

(2) Phase 2: Data were acquired to assess the performance of the targeted biomarker assay using two instrument platforms. UPLC-HRMS (where UPLC is Ultra-high performance liquid chromatography) data acquisition for each compound was performed using one of two systems. System 1 consisted of an Agilent 1290 Infinity LC system interfaced with an Agilent G6520A QTOF HRMS. System 2 used the same model LC system interfaced with an Agilent G6224A TOF HRMS. A Waters Acquity UPLC BEH Amide 2.1 × 50 mm 1.7 μm particle size column (Waters, Milford, MA) maintained at 40°C was applied for separation of metabolites. A solvent gradient with 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) at a flow rate of 1.0 ml/min was used and 2 μl of sample was injected. Electrospray ionization was employed using a dual ESI source operated in positive ionization mode only. The mass range of the instrument was set to 60–1,600 Da and data were acquired over 6.5 min using MassHunter Acquisition software (version B 04.00). Identification of cystine and ornithine metabolites in samples was previously confirmed by comparison of their collision-induced dissociation mass spectra to reference standards (Sigma-Aldrich).

Peak Detection (Phases 1 and 2)

Agilent raw data files were converted to the open source mzData file format using MassHunter Qualitative Analysis software version 5.0 (Agilent Technologies). During the conversion process, deisotoping (+1 charge state only) was performed on the centroid data and peaks with an absolute height less than 200 were excluded from analysis. Peak picking and feature creation were each performed using the R package "xcms" (Smith et al., 2006). Mass features (peaks) were detected using the centwave algorithm. Deviations in retention times were corrected using the obiwarp algorithm that is based on a nonlinear clustering approach to align the data from the LC-MS samples. Mass feature bins or groups were generated using a density-based grouping algorithm. After the data had been grouped into mass features, missing features were integrated based on retention time and mass range of a feature bin using iterative peak filling. Feature intensity was based on the Mexican hat integration values of the feature extracted ion chromatograms.

o/c Ratio Calculation

In both phases of the study, every 96-well plate of samples contained a reference treatment (0.1% DMSO) to allow compensation for the differences in LC-MS instrument response over time. Relative fold changes were calculated for each metabolite by dividing the integrated area of each sample within a treatment level by the median integrated area of the reference treatment (DMSO) samples to produce a normalized value for both metabolites in each sample within a plate of cell culture samples. The o/c ratio was calculated for each sample in a treatment by dividing the reference normalized value of ornithine by the reference normalized value of cystine. In Phase 2, a four-parameter log-logistic model of dose response was fit using the mean o/c ratio value of each concentration using the R package "drc" (Ritz and Streibig, 2005).

Teratogenicity Threshold Selection (Phases 1 and 2)

Classification of teratogenicity was based on the premise that a threshold of metabolic perturbation could be identified for individual metabolites that is associated with developmental toxicity. This threshold of metabolic change is called the teratogenicity threshold and is a measure of the magnitude of metabolic perturbation required to differentiate teratogens from nonteratogens. The teratogenicity threshold was empirically generated for ornithine, cystine, and the o/c ratio by iteration through a range from 10 to 25% change to identify a one- or two-sided asymmetric threshold that was able to classify the training set with the greatest accuracy and highest sensitivity. In the case of a tie in classification accuracy and sensitivity between one- and two-sided thresholds, one-sided thresholds were given priority to favor simplicity. A teratogenicity threshold was determined for each phase of the study, since the assays performed in Phase 1 used only a single concentration of each compound and the targeted biomarker assay developed in Phase 2 utilized an exposure-based approach. The teratogenicity threshold was determined in Phase 2 using only the results from the training set. This threshold was then applied to the results from the test and application sets.

Prediction of Developmental Toxicity Potential

(1) Phase 1: A test compound was classified as a developmental toxicant if the mean of the change in the abundance in the treated sample compared to the reference treatment (DMSO) across the nine experimental replications for either metabolite or the o/c ratio exceeded its respective teratogenicity threshold at the concentration tested. The predictive accuracy (correct prediction), sensitivity (true positive rate), and specificity (true negative rate) were based on scoring the predicted result (teratogen or nonteratogen) against the known human teratogenicity of the compound.

(2) Phase 2: For test compounds with unknown developmental toxicity potential, the targeted biomarker assay is utilized to identify the exposure level where a test compound perturbs metabolism in a manner indicative of teratogenicity and does not require any pharmacokinetic information (e.g., C_{max}). Figure 2 illustrates how the assay is applied in this situation. A test compound is considered to be teratogenic at the exposure level where the o/c ratio exceeds the teratogenicity threshold (red box, Fig. 2). The interpolated concentration from the four-parameter log-logistic model of the o/c ratio or cell viability at the teratogenicity threshold is considered to be the teratogenicity potential exposure level of a test compound (Table 1, Fig. 2). Exposure levels greater than the teratogenicity potential concentration are predicted to have developmental toxicity potential.

    To assess the predictivity of the assay in the training and test sets, the teratogenicity potential concentrations determined from the o/c ratio and cell viability were used to classify the teratogenicity of the test compound relative to the human therapeutic C_{max} concentrations. This approach was not applied to the application set since the developmental toxicity potential of these compounds in humans is unknown. The logic of scoring a test compound as a teratogen or nonteratogen using the human therapeutic C_{max} is based on the paradigm that exposure is a critical factor in teratogenesis, and that a known human teratogen would likely perturb cellular metabolism at or below the highest exposure that is likely to occur at the therapeutic circulating levels. If perturbation of the o/c ratio was exhibited at concentrations greater than the compound's C_{max} concentration (Fig. 3A), it was scored as a nonteratogen because perturbation was observed outside of a range likely to be encountered during routine therapy. If a compound exhibited teratogenicity potential at a concentration that was at or below its therapeutic C_{max}, it was classified as a teratogen (Fig. 3B), since a metabolic perturbation indicative of teratogenesis was exhibited within the therapeutic concentration range. The teratogenicity potential concentration from cell viability was used to predict the teratogenicity of a compound using the same paradigm. The predictive accuracy, sensitivity, and specificity of the assay were calculated by comparing the predicted result to the known human teratogenicity of a compound.

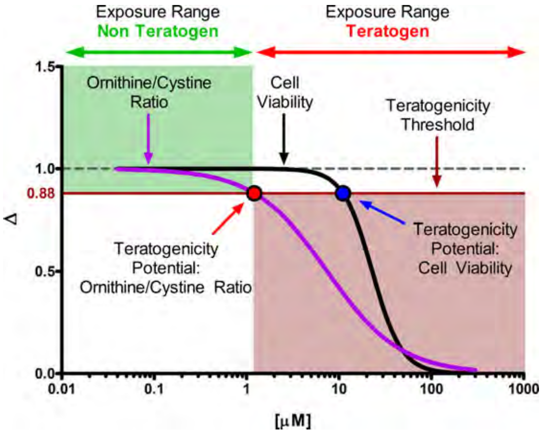

Fig. 2. Graphical representation of the targeted biomarker assay. hES cells were exposed to nine concentrations of a test compound that spanned four log units. The dose–response curve for the ornithine/cystine ratio (o/c ratio; purple curve) and cell viability (solid black curve) was fit using a four-parameter log-logistic model. The concentration predicted by the interpolated point where the dose–response curve of the o/c ratio crosses the teratogenicity threshold (dark red line) indicates the exposure level where a metabolic perturbation has teratogenic potential (i.e., teratogenicity potential: o/c ratio, black bordered red circle). The teratogenicity potential concentration from cell viability (black bordered blue circle) is the interpolated point where the cell viability dose–response curve exceeds the teratogenicity threshold. The teratogenicity potential creates a two-sided toxicity model based on exposure: one where exposure does not perturb metabolism in a manner associated with teratogenicity (green box) and another where exposure may cause a potentially teratogenic shift in metabolism (red box). The *x*-axis is the concentration (µM) of the compound. Both the cell viability measurements and o/c ratio measurements exist on the same scale represented by Δ on the *y*-axis. The *y*-axis value of the o/c ratio is the ratio of the reference treatment normalized (fold change) values (ornithine/cystine). The *y*-axis value of the viability measurement is the treatment cell viability RFU normalized to the reference treatment cell viability RFU.

Comparison of the Targeted Biomarker Assay to Other Developmental Toxicity Tests

A literature review compared the developmental toxicity prediction of the in vivo rodent and rabbit models and three in vitro screens, the European Centre for the Validation of Alternative Methods ECVAM evaluated mouse embryonic stem cell test (mEST), the zebrafish embryotoxicity test (ZET), and the postimplantation rat whole embryo culture (WEC) test for the compounds tested in the targeted biomarker assay. The predictions published for these assays using each original author's classification methods were used for comparison and the data were not reinterpreted. The other in vitro systems employ a three-class classification system (non, weak/moderate, and strong teratogens; Brown, 2002), compared to the two-class system used in this study. Thus, to compare the results from the targeted biomarker assay to other models, the predicted results from these assays needed to be modified to a two class system. Compounds that were predicted to be either weak/moderate or strong teratogens were both labeled as a predicted teratogen. The

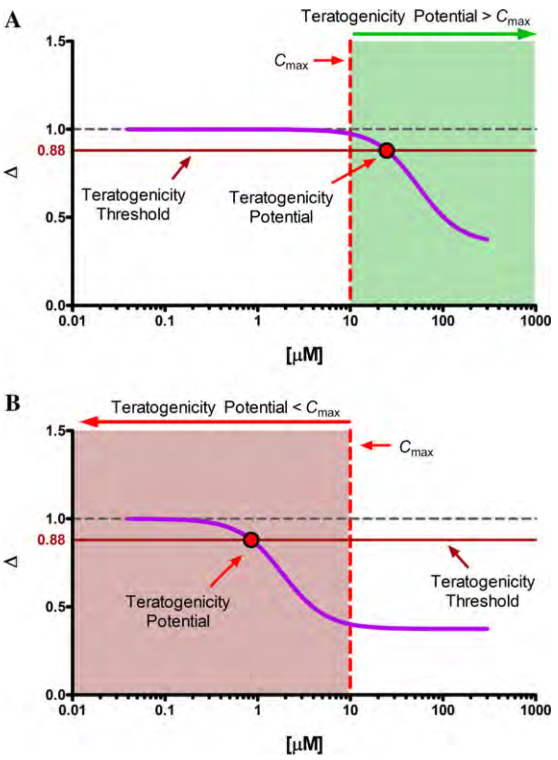

Fig. 3. Graphical representation of the classification scheme for known human teratogens and nonteratogens utilizing the therapeutic C_{max} concentration to set the classification windows. The dose–response curve for the o/c ratio (purple curve) was fit using a four-parameter log-logistic model and used to interpolate the concentration where the o/c ratio crosses the teratogenicity threshold (i.e., teratogenicity potential, black-bordered red circle). A test compound was predicted as a nonteratogen when the teratogenicity potential concentration is higher than the human therapeutic C_{max} (A). A test compound was predicted as a teratogen when the teratogenicity potential concentration is lower than the human therapeutic C_{max} (B). The same logic outlined here is also applied to the viability measurements. The x-axis is the concentration (μM) of the compound. The y-axis value of the o/c ratio is the ratio of the reference treatment normalized (fold change) values (ornithine/cystine).

accuracy, sensitivity, and specificity were calculated for each assay by scoring the predicted result against the known human teratogenicity. These values were additionally calculated for the targeted biomarker assay for the specific set of compounds that had been tested in the other model system. Concordance between the targeted biomarker assay and the other above-mentioned models was evaluated by comparing the classification of teratogen or nonteratogen within the common treatments of each comparison.

RESULTS

Phase 1: Model Confirmation and Characterization of Metabolites Predictive of Developmental Toxicity

The first phase of this study was conducted to confirm the predictivity of individual metabolites identified in previous studies. Characterization of the predictive metabolites led to the development of the new targeted biomarker assay described in the second phase of this study. In previous work, we utilized the training set of 23 pharmaceutical compounds (Table 2) to identify a metabolic signature capable of predicting teratogenicity in vitro (Kleinstreuer et al., 2011). The metabolites that exhibited a statistically significant change upon treatment with teratogens, and lacked a response in nonteratogens, were characterized for their ability to classify developmental toxicants using a simple fold change threshold. Of these metabolites, ornithine and cystine were identified as metabolites that are representative of the previously applied metabolic signature that was highly predictive of developmental toxicity. The capacity of each of these two metabolites to classify developmental toxicants was characterized by determining a teratogenicity threshold (see Table 1) based on the fold change of cells treated with a test compound versus the reference treatment (0.1% DMSO) of each metabolite. The threshold was used to evaluate the classification accuracy of each metabolite within the training set.

Ornithine and cystine each exhibited characteristics amenable to rapid evaluation of the potential for a test compound to perturb metabolism in manner consistent with teratogenicity. Both metabolites are highly abundant in spent cell culture media from hES cells and show changes in their abundance in response to treatment that were reproducibly measured on multiple LC-HRMS instruments. To confirm these initial observations, and the reproducibility of the approach, the metabolites were evaluated in a study that encompassed nine independent experimental replications (blocks) of the training set. The secreted metabolite ornithine was able to distinguish teratogens from nonteratogens with 83% accuracy (Table 5) using a two-sided threshold consisting of either an 18.5% decrease or 20% increase in the accumulation of ornithine (Fig. 4A). Cystine (a media constituent) was the most predictive individual metabolite in classifying teratogens and had an accuracy of 83% (Table 5) using a threshold of a 10% increase relative to the reference treatment (Fig. 4B). Cystine exhibits a significant increase in abundance relative to the reference treatment for most of the teratogens that did not cause cytotoxicity in hES cells (such as hydroxyurea, all-*trans* retinoic acid, 13-*cis* retinoic acid, carbamazepine, and thalidomide). Ornithine decreased with cytotoxic treatments (such as 5-fluorouracil, cytosine arabinoside, methotrexate, and valproic acid), but increased when cells were exposed to the related noncytotoxic teratogens all-*trans* retinoic acid and 13-*cis* retinoic acid.

Based on a previously observed paradigm that metabolic ratios can be used to evaluate teratogenicity (West et al., 2010), we evaluated the possibility that the fold changes in the ratio of ornithine and cystine would be more predictive than their individual fold changes. When the ornithine fold change was divided by the cystine fold change (i.e., the o/c ratio), the resulting ratio was able to correctly classify 91% (Table 5) of the training set (Fig. 4C) using a teratogenicity threshold of a 12% decrease in the o/c ratio, misclassifying only diphenylhydantoin and warfarin. Compared with the accuracy of ornithine and cystine alone, application of the o/c ratio increased the overall prediction accuracy by 8%,

Table 5
Teratogenicity Threshold and Metabolite Model Metrics in the Untargeted Metabolomics-Based Developmental Toxicity Assay

Metabolite	Teratogenicity Threshold	Accuracy	Sensitivity	Specificity
Ornithine	≤81.5% or ≥120%	0.83	0.67	1.00
Cystine	≥110%	0.83	0.83	0.82
Ornithine/Cystine	≤88%	0.91	0.83	1.00

Teratogenicity threshold, a critical threshold of metabolic perturbation that is associated with teratogenesis; accuracy, number of correct predictions divided by the number test compounds evaluated; sensitivity, detection of teratogens; specificity, detection of nonteratogens.

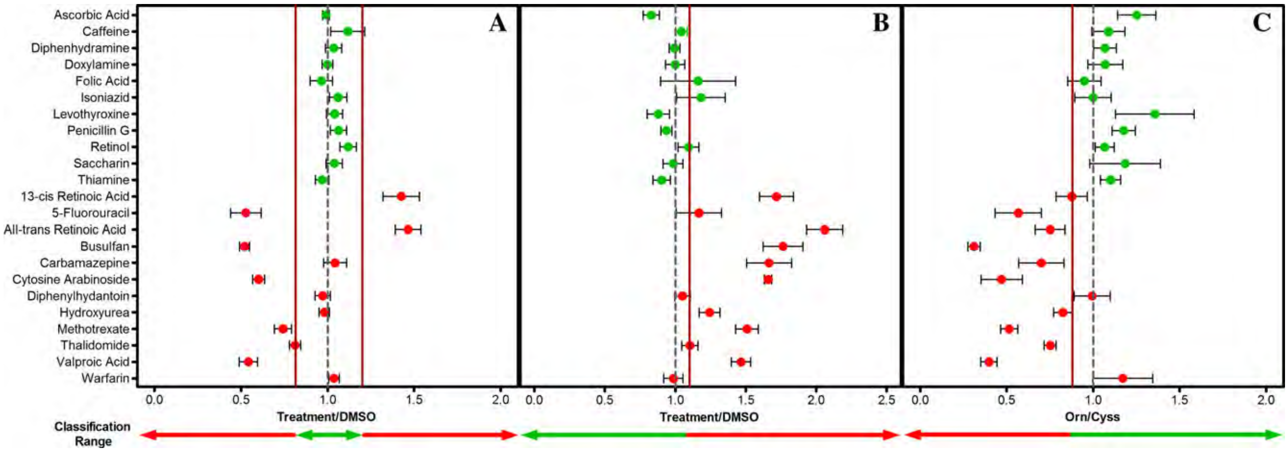

Fig. 4. Metabolic perturbation of ornithine (A), cystine (B), and the o/c ratio (C) measured in experimental Phase 1. Each point represents the mean value of the nine independent experimental blocks. Red points indicate teratogens and green points indicate nonteratogens. Error bars are the standard error of the mean. The vertical dark red line(s) represent the teratogenicity threshold. The *x*-axis is the reference normalized fold change of each metabolite (A–B) or the ratio of ornithine/cystine reference normalized values (C). The *y*-axis is the treatment ordered by nonteratogens and teratogens. Green arrows indicate range where a compound would be classified as a nonteratogen. Red arrows indicate the range where a compound would be classified as a teratogen.

capturing the high specificity of ornithine and high sensitivity of cystine (Table 5) yielding a more accurate classification of teratogenicity.

Phase 2: Development and Evaluation of a Targeted Biomarker Assay to Predict Developmental Toxicity Associated with Exposure

Targeted LC-HRMS method development

In the second phase of this study, we developed a targeted biomarker-based assay using the metabolites confirmed in Phase 1. Since toxicity is a function of both the chemical agent and exposure level, the high level of predictivity associated with a threshold of toxicity of the o/c ratio provided an opportunity for development of a targeted, rapid, teratogenicity assay. To that end, a short and reproducible analysis method was developed and optimized for fast turnaround analysis of relative changes in ornithine and cystine abundance in hES cell spent media samples. In contrast, the untargeted metabolomic methods that had been previously used were designed to analyze a wider breadth of small molecules, and thus required a lengthy chromatographic separation. The prior platform also depended upon two data acquisitions for each sample in positive and negative ionization modes. Focusing on the chromatographic separation, ionization, and detection of ornithine and cystine only, a new, targeted method was designed specifically to more rapidly measure the relative changes of these metabolites observed in the hES cell model system. The new UPLC-HRMS method was developed and assessed using spent media samples (prepared as previously described) for added speed, sensitivity, and retention time reproducibility for measurements of ornithine and cystine. This resulted in a significant reduction in assay turnaround time. The data acquisition time for each sample was reduced from 23 to 6.5 min, providing a fourfold increase in LC-HRMS throughput. The positive ionization mode was preferentially amenable for detection of these metabolites, thereby eliminating the need for the negative mode which further reduced the total analysis time by half for each sample batch, thus, increasing total instrument throughput eightfold. Method reproducibility was evaluated across 17 batches performed over 120 days using reference treatment samples (DMSO treated cells). The average CV for the integrated area of the internal standards and endogenous metabolites was <5 and <8%, respectively, demonstrating that the method performs in a reproducible manner.

Identification of the Teratogenicity Threshold

Based on the high classification accuracy achieved in Phase 1 using a defined teratogenicity threshold, a 9-point concentration curve was used to classify

Table 6
Targeted Biomarker Assay Results: Training Set

Compound	C_{max} (μM)	Teratogenicity potential (μM)		o/c Ratio prediction	Viability prediction	C_{max} reference
		o/c Ratio	Cell Viability			
Nonteratogens						
Ascorbic acid	90	>300	>300	NON	NON	Padayatty et al. (2004)
Caffeine	9.3	>300	>300	NON	NON	Caffeine Pharmacology (n.d.)
Diphenhydramine	0.25	1.8	78.9	NON	NON	Luna et al. (1989)
Doxylamine	0.38	12.9	>300	NON	NON	Luna et al. (1989)
Folic acid	0.035	>300	>300	NON	NON	Ubeda et al. (2011)
Isoniazid	51	165.4	>300	NON	NON	Isoniazid (2000)
Levothyroxine	0.14	43.5	>300	NON	NON	Briggs et al. (2011)
Penicillin G	134.6	>300	>300	NON	NON	(Baxter Healthcare (2012)
Retinol	2.4	42.2	42.8	NON	NON	Mayne Pharma (n.d.)
Saccharin	1.4	>300	>300	NON	NON	(Vaisman et al. (2001)
Thiamine	0.67	>300	>300	NON	NON	Drewe et al. (2003)
Teratogens						
13-*cis* retinoic acid	2.9	**0.0007**	>300	TER	NON	Roche Laboratories (2010)
5-Fluorouracil	4.25	**3**	**2**	TER	TER	Oman et al. (2005)
All-*trans* retinoic acid	1.2	**0.00004**	114.5	TER	NON	Muindi et al. (1992)
Busulfan	49.6	**0.6**	**3**	TER	TER	(Otsuka America Pharmaceutical (2011)
Carbamazepine	47	**0.9**	>300	TER	NON	(Mahmood and Chamberlin (1998)
Cytosine arabinoside	0.6	**0.04**	**0.1**	TER	TER	Weinstein et al. (1982)
Diphenylhydantoin	79.3	263.3	288.7	NON	NON	Pfizer (2012)
Hydroxyurea	565	**5**	**251.6**	TER	TER	Liebelt et al. (2007)
Methotrexate	0.2	**0.05**	**0.05**	TER	TER	Shoda et al. (2007)
Thalidomide	12.4	**0.2**	>300	TER	NON	Thalidomide Pharmacology (n.d.)
Valproic acid	1000	**90.8**	1113.7	TER	NON	AbbVie (2013)
Warfarin	23.4	**6.5**	>300	TER	NON	Welle-Watne et al. (1980)

C_{max}, therapeutic peak plasma in vivo concentration; teratogenicity potential, interpolated concentration when the dose–response curve of the o/c ratio or cell viability crosses the teratogenicity threshold; NON, potential nonteratogen; TER, potential teratogen. Teratogenicity potential values for the o/c ratio and viability measurements that occur at an exposure level below the C_{max} value are bolded.

developmental toxicity potential based on a range of exposures. The teratogenicity threshold was optimized using the Phase 2 training set data by selecting a threshold that produced the highest accuracy of prediction with the greatest sensitivity. The predicted teratogenicity potential concentration was compared to the therapeutic C_{max} to score the performance and classification accuracy of this new assay design (described in Fig. 3, Table 6). With this approach, a 12% decrease in the o/c ratio relative to the reference treatment was the optimum threshold and was able to classify the training set of compounds with 96% accuracy (Table 7, Fig. 5A). The assay correctly classified all of the nonteratogens (100% specificity) and misclassified only one of the known human development toxicants, diphenylhydantoin (92% sensitivity).

Table 7
Model Metrics of the o/c Ratio Compared to Cell Viability from the Targeted Biomarker Assay

Assay	Accuracy	Sensitivity	Specificity
Training set			
o/c Ratio	0.96	0.92	1.00
Cell viability	0.70	0.42	1.00
Test set			
o/c Ratio	0.77	0.57	1.00
Cell viability	0.62	0.29	1.00

Accuracy, number of correct predictions divided by the number test compounds evaluated; sensitivity, detection of teratogens; specificity, detection of nonteratogens.

Evaluation of the Targeted Biomarker Assay Performance Based on the Test Set Predictions

The teratogenicity threshold identified using the training set was applied to the test set of compounds to assess the predictivity of the targeted biomarker assay developed in this study. The test set consisted of 13 compounds not included in the training set with known human teratogenicity, having Food and Drug Administration pregnancy classifications of B, D and X (Table 3). The teratogenicity potential concentration of each compound for the o/c ratio was scored against the compound's therapeutic C_{max}. The test set was classified with 77% accuracy (100% specificity, 57% sensitivity, Table 7). The o/c ratio incorrectly classified the teratogens bosentan, lapatinib, and lovastatin (Table 8, Fig. 5B). Please note that the C_{max} for everolimus is below the lowest exposure level used in the assay and the o/c ratio for this compound begins below the teratogenicity threshold, so it is classified as a

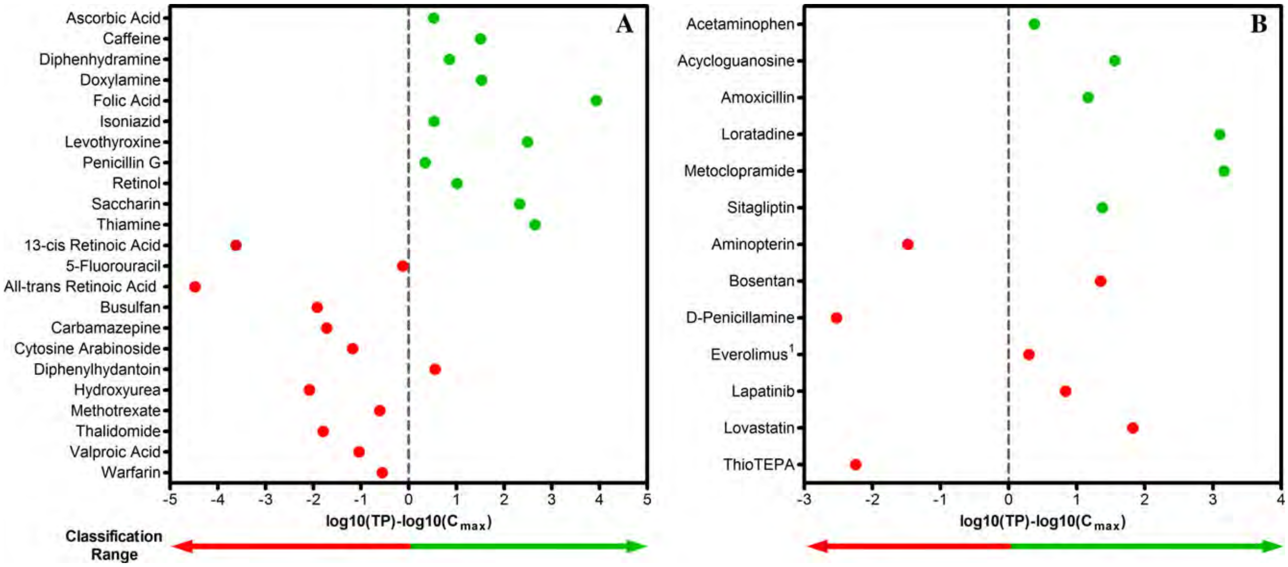

Fig. 5. Visualization of the difference between a compound's teratogenicity potential concentration for the o/c ratio (TP) and C_{max} values for the training set (A) and test set (B) in Phase 2. Red points correspond to teratogens and green points correspond to nonteratogens. Treatments that have a difference between the TP and C_{max} less than 0 are classified as teratogens and treatments with a difference between the TP and C_{max} greater than 0 are classified as nonteratogens. The *x*-axis is the log base 10 transformed TP concentration value subtracted from the log base 10 transformed C_{max} concentration value (see Tables 6 and 8). The *y*-axis is the treatment ordered by nonteratogens and teratogens. Green arrows indicate the range where a compound would be classified as a nonteratogen. Red arrows indicate the range where a compound would be classified as a teratogen. ^{1}The C_{max} for everolimus is below the lowest exposure level used in the assay, the o/c ratio for this compound begins below the teratogenicity threshold, so it is classified as a teratogen.

Table 8
Targeted Biomarker Assay Results: Test Set

Compound	C_{max} (μM)	Teratogenicity potential (μM) o/c Ratio	Teratogenicity potential (μM) Cell viability	o/c Ratio prediction	Viability prediction	C_{max} reference
Nonteratogens						
Acetaminophen	116.4	>300	>300	NON	NON	(McNeil Consumer Healthcare (2010)
Acycloguanosine	3	95.8	>300	NON	NON	(Palma-Aguirre et al. (2007)
Amoxicillin	20.5	>300	>300	NON	NON	(Dr Reddy's Laboratories (2011)
Loratadine	0.03	37.8	76.3	NON	NON	(Hilbert et al. (1987)
Metoclopramide	0.15	190.8	>300	NON	NON	Leucuţa et al. (2004)
Sitagliptin	0.95	22.6	>300	NON	NON	Merck (2013)
Teratogens						
Aminopterin	0.3	**0.01**	**0.01**	TER	TER	Cole et al. (2005)
Bosentan	2	44.9	221.9	NON	NON	van Giersbergen et al. (2007)
D-penicillamine	13.4	**<0.04**	>300	TER	NON	Merck (2004)
Everolimus	0.02	**<0.04**	5.2	TER	NON	Novartis Sverige AB (2011)
Lapatinib	4.2	29	20.8	NON	NON	GlaxoSmithKline (2013)
Lovastatin	0.02	1.3	4.1	NON	NON	Andrx Labs (2012)
ThioTEPA	7	**0.04**	**0.5**	TER	TER	Bedford Laboratories (2001)

C_{max}, therapeutic peak plasma in vivo concentration; teratogenicity potential, interpolated concentration when the dose–response curve of the o/c ratio or cell viability crosses the teratogenicity threshold; NON, potential nonteratogen; TER, potential teratogen. Teratogenicity potential values for the o/c ratio and viability measurements that occur at an exposure level below the C_{max} value are bolded.

teratogen even though it groups with the nonteratogens in Figure 5B.

Comparison of the o/c Ratio and Cell Viability

Because the metabolites that make up the o/c ratio are measured in spent cell culture media, the treated cells were available to perform cell viability analysis. The cell viability results were compared to the o/c ratio to determine if the change in the ratio was due to cell death or if it was due to metabolic changes unrelated to changes in cell viability. The viability results were evaluated to determine classification performance using an approach similar to the o/c ratio (Fig. 3). The teratogenicity threshold that was determined using the o/c ratio results from the training set was also used to classify teratogenicity by cell viability based on the interpolated concentration at which the cell viability dose–response curve exceeds the teratogenicity threshold (Tables 6 and 8). This enabled a direct comparison of the o/c ratio and cell viability at equal levels of change from controls. Cell viability had an accuracy of 70% for the training set and 62% for the test set (Table 7). The cell viability assay was successful in correctly classifying all of the nonteratogens in both the training and test sets but performed poorly for the classification of teratogens, correctly classifying only 5 of the 12 compounds in the training set (42% sensitivity, Table 7) and 2 of the 7 teratogens in the test set (29% sensitivity, Table 7). Those that were correctly classified by cell viability are antineoplastic compounds that kill dividing cells.

When applied to the training and test sets, the o/c ratio was 26 and 15% *more* accurate, respectively, than viability alone for the prediction of development toxicity (Table 7). Both the o/c ratio and cell viability assay correctly classify nonteratogens with respect to the $C_{\max}$ having 100% specificity, however, they differ in their ability to discriminate teratogens (Table 7). The o/c ratio is 50% more sensitive in the detection of teratogens than viability alone in the training set and 28% more sensitive in the test set (Table 7). Additionally, the o/c ratio is able to classify both cytotoxic and noncytotoxic teratogens correctly. The decrease in false negatives provided by the o/c ratio is related to the assay's measurement of metabolic perturbation that can occur independent of changes in cell viability.

Highlighted in Figure 6 is a subset of the results that demonstrate several characteristics of the assay with respect to the o/c ratio performance relative to cell viability. Thalidomide (Fig. 6A) and all-*trans* retinoic acid (Fig. 6B) are examples of teratogens that exhibit a change in the o/c ratio indicative of developmental toxicity in the absence of cytotoxicity. The teratogen valproic acid (Fig. 6C) is an example of a cytotoxic teratogen that causes a marked change in the o/c ratio at exposure levels well before cytotoxicity is observed. 5-fluorouracil (Fig. 6D) is an antineoplastic teratogen that yields a change in o/c ratio that is directly correlated with a decrease in cell viability and the change in the metabolite ratio is likely a direct result of cell death. Retinol (Fig. 6E) is an example of a cytotoxic nonteratogen where the o/c ratio is directly correlated with cell death at exposure levels almost 20 times higher than those normally encountered by humans. The nonteratogen saccharin (Fig. 6F) is a compound that yields no change in the o/c ratio or viability at the exposures examined in this study.

Application of the o/c Ratio and Teratogenicity Threshold to Compounds with Unknown Human Teratogenicity

The targeted biomarker assay was applied to an application set of 10 compounds that have unknown human developmental toxicity outcomes. Since the human developmental toxicity of these compounds is unknown, the $C_{\max}$ approach (illustrated in Fig. 3) used to score assay performance was not applied and the compounds were treated as unknowns, as is illustrated in Figure 2. The results are presented as they would be generated by the assay utilized in an industrial setting. The teratogenicity potential concentrations for the o/c ratio and cell viability are summarized in Table 9. All 10 compounds exhibited a change in the o/c ratio indicative of teratogenicity, although concentration at which this change occurred varied greatly between compounds. Nine of the 10 compounds exhibited a change in cell viability within the exposure range tested (Table 9). Seven of the 10 compounds caused a change in the o/c ratio before or in the absence of cytotoxicity (bolded compounds, Table 9). Rodent developmental toxicity testing identified a teratogenic and/or embryotoxic effect in seven of the 10 compounds in the absence of maternal toxicity. The other three compounds (adefovir dipivoxil, cidofovir, and ramelteon) were only embryotoxic at exposure levels that also caused maternal toxicity, so it is unknown if the effect was due to compound exposure.

Assay Performance (Comparison to Other Assays)

The developmental toxicity predictions based on the o/c ratio for the training and test sets were compared to published results from other model systems (Table 10). The developmental toxicity predictions from the model systems presented in Table 10 for the application set are summarized in Supplementary Table S1. For the combined 36 training and test set compounds, comparisons were made on a model system-by-system basis using only the treatments evaluated in both the targeted biomarker assay and each model system it was being compared to. The results of the comparisons (Table 11) indicate that the o/c ratio described here is a more accurate predictor of human developmental toxicants than the other model systems considered. The increase in accuracy is due to a lower false-positive rate (increased specificity) of the o/c ratio in each comparison with a significant increase in specificity over other in vitro systems, such as mEST and WEC, as well as a moderate gain in sensitivity. Interestingly, the o/c ratio is able to correctly classify the nonteratogens caffeine and retinol and teratogens warfarin and D-penicillamine, where the majority of other model systems fail. There is a high degree of concordance ($\geq$75%) between the teratogenicity prediction of the o/c ratio and the in vivo rodent and rabbit models as well as the ZET (Table 11). Concordance is lower between the o/c ratio and the mEST and WEC (67 and 69%, respectively, Table 11). The reason for lower concordance between the o/c ratio and these in vitro models is due to the high accuracy of the targeted biomarker assay.

DISCUSSION

The present assay has been developed to address the need for more accurate, rapid, and less expensive alternatives to animal testing. Our goal was to provide toxicologists with a new and biologically germane tool to aid in compound prioritization before the currently required in vivo testing and as part of emerging multitiered testing strategies. Undifferentiated hES cells represent a simple and elegant test system for modeling a test compound's developmentally toxic effects on human cells at the very earliest stages of development, which in some cases can lead to implications of the compound's ef-

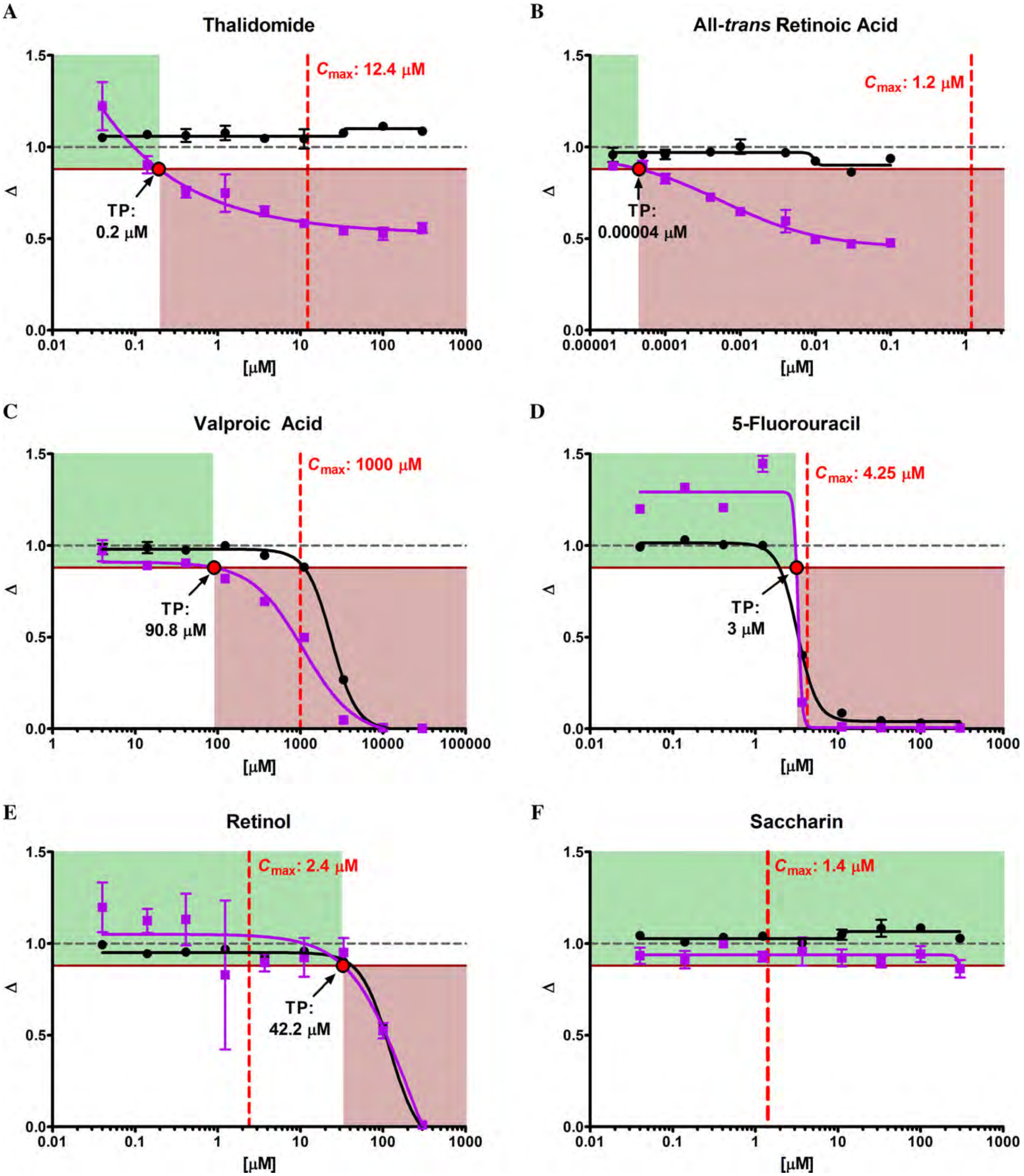

Fig. 6. Targeted biomarker assay results for a representative subset of the training set compounds (Table 6). The dose–response curves for the viability analysis (black curve) and o/c ratio (purple curve) are shown for four known human teratogens: thalidomide (A), all-*trans* retinoic acid (B), valproic acid (C), 5-fluorouracil (D), and two nonteratogens: retinol (E) and saccharin (F). The *x*-axis is the concentration (μM) of the compound. Both the cell viability measurements and o/c ratio measurements exist on the same scale represented by Δ on the *y*-axis. The *y*-axis value of the o/c ratio is the ratio of the reference treatment normalized (fold change) values (ornithine/cystine). The *y*-axis value for the viability measurement is the treatment cell viability RFU normalized to the reference treatment cell viability RFU. The vertical broken red line indicates the compound specific C_{max} and the horizontal dark red line indicates the teratogenicity threshold (0.88). The black-bordered red circle represents the teratogenicity potential concentration (TP) for the o/c ratio. The green- and redshaded areas represent the concentrations where the compound is predicted to be nonteratogenic or teratogenic, respectively. The points are mean values and error bars are the standard error of the mean. Interpretation of these figures is outlined in Figures 2 and 3.

Table 9
Targeted Biomarker Assay Results: Application Set

		Teratogenicity potential (μM)		Rodent in vivo test results[a]		
Compound	C_{max} (μM)	o/c Ratio	Cell viability	Teratogenic[b]	Embryotoxic[c]	C_{max} reference
6-Aminonicotinamide	NA	**<0.04**	24.5	+[d]	-[d]	NA
Abacavir	14.9	95.1	94.1	+	+	GlaxoSmithKline (2012)
Adefovir dipivoxil[e]	0.03	**0.0015**	0.02	-	-	Gilead Sciences (2012)
Amprenavir	15.1	**236.9**	259.5	+	+	GlaxoSmithKline (2005)
Artesunate	73.9	0.64	0.58	+[f]	+[f]	Miller et al. (2012)
Cidofovir[g]	41.2	**0.3**	1.9	-	-	Gilead Sciences (2000)
Entacapone	3.9	**6.7**	127	+	-	Novartis Pharmaceuticals (2010)
Fluoxetine	0.04	25.1	23	-	+	Warner Chilcott (2013)
Ramelteon[h]	0.02	**34**	>300	-	-	Karim et al. (2006)
Rosiglitazone	1.7	**18.9**	21.8	-	+	GlaxoSmithKline (2011)

C_{max}, peak plasma concentration in humans; teratogenicity potential, interpolated concentration when the dose–response curve of the o/c ratio or cell viability crosses the teratogenicity threshold; NA, not available or undetermined. Teratogenicity potential values for the o/c ratio that occur before cell viability are bolded.

[a] Data were compiled from Briggs et al. (2011) unless otherwise noted.

[b] A test compound was considered teratogenic if it caused structural malformations in the absence of maternal toxicity.

[c] This column refers to an embryotoxic effect in the absence of teratogenic effects. A test compound was considered embryotoxic if it caused growth retardation or embryo lethality in the absence of maternal toxicity.

[d] Shepard and Lemire (2007).

[e] Adefovir dipivoxil was teratogenic and embryotoxic at maternally toxic doses.

[f] Clark (2009), Shepard and Lemire (2007).

[g] Cidofovir was embryotoxic at maternally toxic doses.

[h] Ramelteon was teratogenic at maternally toxic doses.

fects in later stage fetal development as well. A developmental toxicity test based on hES cells reduces the risk of false-negatives due specifically to interspecies differences in developmental pathways and pharmacokinetics (Scott et al., 2013). We have modified our untargeted metabolomics-based developmental toxicity assay to decrease complexity and increase throughput by focusing on two biologically relevant metabolites that can accurately model human toxic response over a wide range of exposure levels.

In this study, we explored and demonstrated the concept that a certain degree of metabolic perturbation could be used to predict a test compound's potential to cause developmental toxicity. The new assay uses a multiexposure approach that allows for a look at cellular response over a large range of exposure levels. Application of the teratogenicity threshold to this approach allowed us to use changes in metabolism at increasing exposure levels to identify the concentration at which metabolism was altered in a manner indicative of potential teratogenicity. The model created here allows the comparison of changes in a metabolic ratio of ornithine and cystine to cell viability to identify the exposure level where changes in metabolism are likely to lead to teratogenicity and relate it to cell death. The combined evaluation of cell viability and changes in metabolism allow this assay to also identify when exposure could lead to developmental toxicity due to cell death or possible embryo toxicity. The o/c ratio can discriminate between teratogens and nonteratogens with a combined 89% accuracy in the training and test sets using the teratogenicity threshold set in Phase 2 (Table 11).

Analysis of metabolites is a critical process in understanding mechanisms of toxicity since metabolites play critical roles in the maintenance of homeostasis and signaling. Perturbation of individual metabolites has the ability to disrupt normal developmental processes. Alterations in metabolite abundance can occur via mechanisms independent of protein and transcript abundance such as allosteric interaction of a compound or compound's metabolite with an enzyme, defects in posttranslational modification, disrupted protein–protein interactions, and/or altered transport. Changes in metabolism, as measured in the spent medium of cell culture systems, yield a distinguishable "metabolic footprint," which is a functional measure of cellular metabolism that can be used to evaluate response to treatment. The perturbation of biochemical pathways that contain ornithine and cystine as reactants or products have been experimentally associated with mechanisms of teratogenesis. Extracellularly, or within the secretome measured by our assays, cystine predominates over cysteine due to the oxidative state of the medium. Cystine is rapidly converted to cysteine once it is imported into the intracellular environment and is part of the cystine/cysteine thiol redox couple, a critical component of a cell's regulatory capacity to handle reactive oxygen species (ROS). Its role has been investigated with regard to its capacity to mod-

Table 10
Comparison of Targeted Biomarker Assay Results to Published Developmental Toxicity Assay Results: Training and Test Set

Compound	Humans[a]	Targeted biomarker assay	Rodent[a]	Rabbit[a]	mEST	ZET	WEC
Acetaminophen	NON	NON	NON	NA	NA	NON[b]	**TER**[c]
Acycloguanosine	NON	NON	**TER**	NON	NA	NA	**TER**[d]
Amoxicillin	NON	NON	NON	NA	NA	NA	NA
Ascorbic acid	NON	NON	NON	NA	NON[e]	NON[b,f,g]	NON[h]
Caffeine	NON	NON	**TER**	**TER**	**TER**[e]	**TER**[b]	**TER**[i]
Diphenhydramine	NON	NON	NON	NON	**TER**[e]	**TER**[b]	NON[h]
Doxylamine	NON	NON	NON	NON	**TER**[j]	NA	NON[h]
Folic acid	NON	NON	NON[k]	NA	NA	NA	NON[l]
Isoniazid	NON	NON	NON	NON	NON[e,m]	NON[f,n]	**TER**[h,o]
Levothyroxine	NON	NON	NON	NON	NA	NA	NA
Loratadine	NON	NON	NON	NON	NON[m]	**TER**[g]	NON[h,o]
Metoclopramide	NON	NON	NON	NON	**TER**[j,m]	NON[g]	NON[h,o]
Penicillin G	NON	NON	NON	NON	NON[e,m]	NON[b,f,n]	NON[h,o]
Retinol	NON	NON	**TER**	**TER**	NON[p]	**TER**[f,n]	**TER**[q]
Saccharin	NON	NON	NON	NON	NON[e,m]	NON[b,f]	NON[o]
Sitagliptin	NON	NON	**TER**	NON	NA	NA	NA
Thiamine	NON	NON	NA	NA	NA	NA	NA
13-*cis* retinoic Acid	TER	TER	TER	TER	TER[p]	TER[r]	TER[s]
5-Fluorouracil	TER	TER	TER	TER	TER[e,m]	TER[f]	TER[c,h]
All-*trans* retinoic acid	TER	TER	TER	TER	TER[e,p]	TER[b,f,r]	TER[q,s]
Aminopterin	TER	TER	TER	TER	NA	NA	NA
Bosentan	TER	**NON**	TER	**NON**	NA	NA	NA
Busulfan	TER	TER	TER	TER	TER[m]	NA	TER[o]
Carbamazepine	TER	TER	TER	NA	TER[m]	TER[t]	TER[o]
Cytosine arabinoside	TER	TER	TER	NA	TER[m]	TER[n]	TER[o]
Diphenylhydantoin	TER	**NON**	TER	TER	TER[e,m]	**NON**[n]	TER[h,o]
D-penicillamine	TER	TER	TER	NA	**NON**[j]	**NON**[g]	**NON**[h]
Everolimus	TER	TER	TER	**NON**	NA	NA	NA
Hydroxyurea	TER	TER	TER	TER	TER[e,m]	TER[f]	TER[h,o]
Lapatinib	TER	**NON**	TER	TER	NA	NA	NA
Lovastatin	TER	**NON**	TER	**NON**	TER[j]	TER[g]	NA
Methotrexate	TER	TER	TER	TER	TER[e,m]	TER[g]	TER[h]
Thalidomide	TER	TER	**NON**[u]	TER	NA	TER[g]	TER[h]
ThioTEPA	TER	TER	TER	TER	NA	TER[v]	NA
Valproic acid	TER	TER	TER	TER[u]	TER[e,m]	TER[b,n]	TER[h,o]
Warfarin	TER	TER	TER	**NON**	**NON**[j,m]	TER[g]	**NON**[o]

mEST, mouse embryonic stem cell test; ZET, zebrafish embryotoxicity test; WEC, whole embryo culture; NON, nonteratogen; TER, teratogen; NA, not available. If there were conflicting predictions, the classification from the more recent publication or with more publications in agreement was used. Bolded results indicate predictions that differ from known human developmental toxicity effects.

[a] Human, rodent and rabbit effects summarized from drugs in pregnancy and lactation (Briggs et al., 2011), TERIS and/or the ACToR database (http://actor.epa.gov/actor/faces/ACToRHome.jsp) unless otherwise noted.

[b] Selderslaghs et al. (2012).

[c] Stark et al. (1990).

[d] Klug et al. (1985).

[e] Genschow et al. (2004).

[f] Brannen et al. (2010).

[g] Gustafson et al. (2012).

[h] Zhang et al. (2012).

[i] Robinson et al. (2010).

[j] Marx-Stoelting et al. (2009).

[k] Hansen et al. (1993).

[l] Hansen (1995).

[m] Paquette et al. (2008).

[n] McGrath and Li (2008).

[o] Thomson et al. (2011).

[p] Louisse et al. (2011).

[q] Ritchie et al. (2003).

[r] Herrmann (1995).

[s] Klug et al. (1989).

[t] Madureira et al. (2011).

[u] Jelovsek et al. (1989).

[v] Weigt et al. (2011).

Table 11
Model Metrics of the Targeted Biomarker Assay Predictions Compared to Other Model Predictions Based on Treatments in Common

Model system	*N*	Concordance	Acc	TB_Acc	Sen	TB_Sen	Spec	TB_Spec
Targeted biomarker assay	36	NA	0.89	NA	0.79	NA	1.00	NA
Rodent	35	0.74	0.86	0.89	0.95	0.79	0.75	1.00
Rabbit	28	0.79	0.79	0.86	0.75	0.75	0.83	1.00
mEST	23	0.65	0.74	0.91	0.85	0.85	0.60	1.00
ZET	24	0.75	0.75	0.92	0.86	0.86	0.60	1.00
WEC	26	0.69	0.73	0.96	0.85	0.92	0.62	1.00

N, the number of treatments assayed that were common between the model system and the targeted biomarker assay; TB, the targeted biomarker assay results using the treatments evaluated in that model system; Acc, accuracy of model system; TB˙Acc, accuracy of targeted biomarker assay; Sen, Sensitivity of model system; TB˙Sen, sensitivity of targeted biomarker assay; Spec, specificity of the model system; TB˙Sen, specificity of the targeted biomarker assay.

ulate differentiation, proliferation, apoptosis, and other cellular events that may lead to teratogenesis (Hansen, 2006). A broad spectrum of teratogens, including pharmaceuticals, pesticides, and environmental contaminants, are suspected of creating ROS or disrupting cellular mechanisms that maintain the appropriate balance of a cell's redox state, which can lead to adverse effects on developmental regulatory networks as a mechanism of action of developmental toxicity (Hansen, 2006; Kovacic and Somanathan, 2006). It has been hypothesized that a major mechanism of thalidomide teratogenesis and its species specific manifestation of developmental toxicity is related to ROS-related upregulation of apoptotic pathways during limb formation (Hansen, 2006). The measurement of cystine in this assay provides insight into a cell's redox status. When cystine's uptake is perturbed, it can act as a biomarker, indicating a disruption in the cell's ability to signal using ROS-related pathways.

The second metabolite in this assay is ornithine, which is secreted by the hES cells during culture. Ornithine is formed as a product of the catabolism of arginine into urea, is critical to the excretion of nitrogen, and is a precursor to polyamines. Catabolism of ornithine is impacted by the teratogen all-*trans* retinoic acid, which is a suppressor of the transcription of ornithine decarboxylase (ODC), leading to increased ornithine secretion which in turn inhibits polyamine synthesis (Mao et al., 1993). It is also clear that ODC plays an important role in development, since a mouse model with ODC knocked out leads to disruption of very early embryonic stages and is lethal to the developing embryo (Pegg, 2009). Alterations in ornithine levels could lead to the disruption in polyamine metabolism, which is critical for cellular growth and differentiation during human development (Kalhan and Bier, 2008).

Only 1 of the 23 compounds in the training set (diphenylhydantoin) and 3 of the 13 compounds in the test set (bosentan, lapatinib, and lovastatin) were misclassified in the targeted biomarker assay (Tables 6 and 8). All four of these compounds exhibited a change in the o/c ratio indicative of teratogenicity; however, the teratogenicity potential concentration is higher than the therapeutic $C_{\max}$, which was set as a marker of biological relevance for exposure level. For discovery compounds that will not have an established $C_{\max}$ value, these changes in the o/c ratio can be used as a signal regarding the teratogenic potential of the compound. While epidemiologic studies have shown an association between diphenylhydantoin and birth defects, there have been no such studies describing the incidence of birth defects following bosentan, lapatinib, and lovastatin exposure during pregnancy. No case reports have been published regarding birth defects in infants exposed to bosentan or lapatinib during pregnancy and only a handful of reports describing malformations following lovastatin exposure during early pregnancy (TERIS).

In vivo rat developmental toxicity studies have identified a lowest observed adverse effect level for lovastatin of 100 mg/kg body weight per day during organogenesis (Lankas et al., 2004). Interestingly, this level of exposure results in a $C_{\max}$ around 1.5 μM (Lankas et al., 2004), which is close to the teratogenicity potential identified by the o/c ratio in this study (1.3 μM, Table 7, Fig. 7A). Lapatinib causes rat pup mortality in vivo when given during organogenesis at exposure levels that are about 3.3 times the human clinical exposure based on AUC (Briggs et al., 2011). This level of exposure is approximately equal to the concentration where cell viability decreases in hES cells following lapatinib exposure (Fig. 7B). Animal models are currently used to measure teratogenicity risk, but it is still unknown how well their results correlate to human risk for individual compounds. While the primary goal of the assay is to predict potential for teratogenicity in humans, it is also important to understand concordance with in vivo animal models used for regulatory acceptance. These are a few examples of how the data generated in the targeted biomarker assay can be correlated to in vivo developmental toxicity data.

For the compounds evaluated in this study, the targeted biomarker assay agrees with in vivo rodent and rabbit studies about 75% of the time (Table 11). There is still significant opportunity to improve the understanding of how to translate compound concentrations from in vitro systems to human exposure levels (Bhattacharya et al., 2011). The application set was used to demonstrate how the measurement of toxicity potential across an exposure range can put model response into perspective in terms of the overall compound risk when combined with additional assays conducted during a compound's discovery and development. The 10 compounds in this set have unknown human developmental toxicity outcomes, as would any novel compound. We compared the o/c

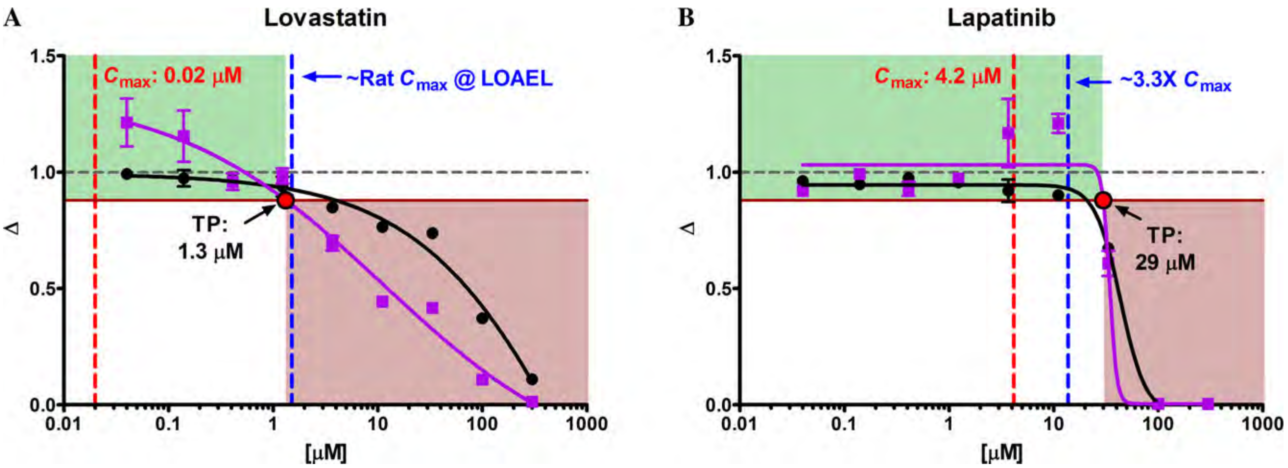

Fig. 7. Targeted biomarker assay results compared to rat in vivo developmental toxicity outcomes for two test set compounds (Table 8): lovastatin (A) and lapatinib (B). The dose–response curves from the targeted biomarker assay for the viability analysis (black curve) and o/c ratio (purple curve) are shown. The *x*-axis is the concentration (µM) of the compound. Both the cell viability measurements and o/c ratio measurements exist on the same scale represented by Δ on the *y*-axis. The *y*-axis value of the o/c ratio is the ratio of the reference treatment normalized (fold change) values (ornithine/cystine). The *y*-axis value for the viability measurement is the treatment cell viability RFU normalized to the reference treatment cell viability RFU. The vertical broken red line indicates the compound specific C_{max} and the horizontal dark red line indicates the teratogenicity threshold (0.88). The black-bordered red represents the teratogenicity potential concentration (TP) for the o/c ratio. The green and redshaded areas represent the concentrations where the compound is predicted to be nonteratogenic or teratogenic, respectively. The vertical broken blue line represents the concentration where a positive result was observed in the rat in vivo developmental toxicity test. The points are mean values and error bars are the standard error of the mean. Interpretation of these figures is outlined in Figures 2 and 3.

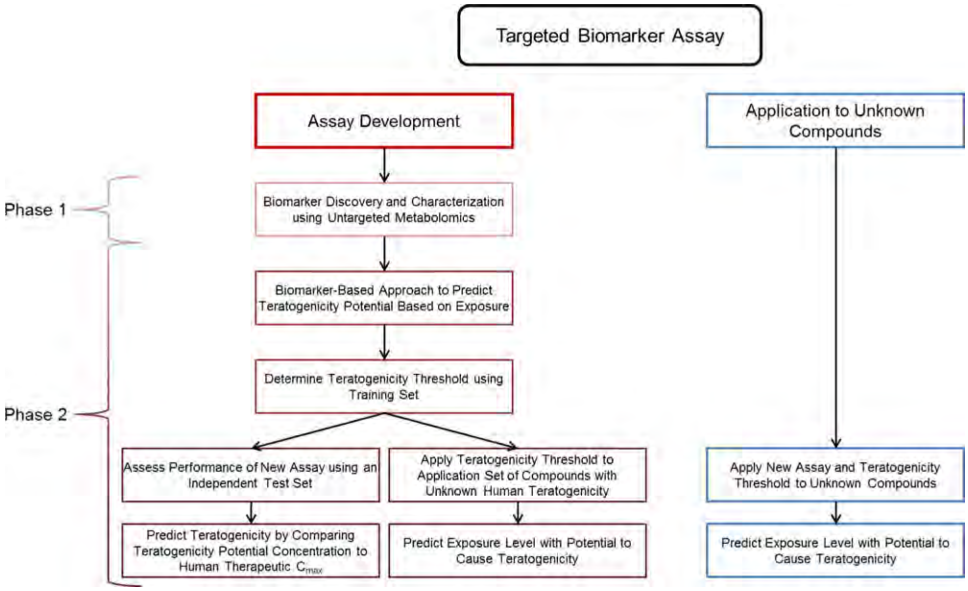

Fig. 8. Diagram outlining the development of the targeted biomarker assay compared to use with unknown compounds.

ratio with the available C_{max} for the application set of compounds to begin to assess the relevance of the signal of teratogenicity potential for each compound (Supplementary Table S1). We used the therapeutic C_{max} to understand the potential exposure level encountered in humans. However, since the human teratogenicity of these compounds is unknown, we did not use the C_{max} to assess the predictivity of the assay. The application set was meant to demonstrate utility of the targeted biomarker assay for unknown compounds in contrast to assessment of assay performance for compounds with known human teratogenicity (Fig. 8). We then used any available preclinical in vivo findings to develop and understanding of each compound and its risk potential. Such an approach could be used in adoption of the assay as part of a traditional compound discovery or preclinical development program, or as part of a new paradigm utilizing a panel of human-cell–based assays aimed at early decision making.

A significant advantage of the targeted biomarker assay is the use of human cells, derived from an embryo, which are able to recapitulate every cell type in the body and have an unlimited capacity to proliferate in culture. The possibility of species-specific differences in developmental toxicity that may be observed in other in vitro developmental toxicity assays is eliminated. In contrast to the European Centre for the Validation of Alternative Methods evaluated mEST, the assay presented here does not require differentiation of the hES cells into specific lineages, such as embryo bodies or cardiomyocytes. Differentiation into specific lineages may limit an assay's potential for predicting teratogens that affect a different developmental lineage. The assay described herein can correctly classify compounds that are known to affect multiple lineages, including cardiovascular, neural, and skeletal (Tables 2 and 3). The targeted biomarker assay provides endpoints that are determined analytically and do not need any subjective interpretation of morphology, as is required by the mEST, postimplantation rat WEC test, and ZET. Recent modifications to the mEST have begun to address these limitations by adding additional developmental endpoints (i.e., neural and osteoblast differentiation) and implementing molecular endpoints in place of subjective evaluation (reviewed in Theunissen and Piersma, 2012). Table 10 presents a comparison of the results of the targeted biomarker assay described here and five other developmental toxicity assays; the targeted biomarker assay has a higher accuracy than the other assays (Table 11). The higher accuracy of the predictions made with the o/c ratio is due to an increase in specificity, or the detection of nonteratogens, over the other assays. It is important to note that differences exist between each of the model systems in the way that compounds are predicted. None of the other assays included in Table 10 classify compounds based on human exposure levels, whereas our classification system directly compares a compound's teratogenicity potential to the known therapeutic $C_{\max}$ for compounds that have known human developmental toxicity outcomes. When making predictions, the actual exposure levels of a compound likely to be encountered by a fetus are critical. Nine of the 17 human nonteratogens tested in the targeted biomarker assay caused a change in the o/c ratio at exposure levels above the therapeutic $C_{\max}$. It is believed that any compound, given at the right dose, at the right time during development, in the right species will be teratogenic (Daston et al., 2010). The ability of the targeted biomarker assay to separate exposure levels that are not indicative of teratogenicity from levels that are indicative of teratogenicity is a key strength of the assay.

Although the targeted biomarker assay described herein shows significant promise in predicting developmental toxicity, hES cells, as with other in vitro models, cannot fully reproduce all events contributing to the disruption of normal human development by exogenous chemicals. In vitro models of toxicity do not include the effects of absorption, distribution, metabolism, and excretion, which may make it difficult to predict how a substance of unknown toxicity will act in vivo. The absence of metabolic activity could partially be overcome by the addition of an exogenous bioactivation system when metabolic activation is required or testing both the parent compound and any known metabolites for developmental toxicity potential. Testing both parent compounds and metabolites can help discern which agent is the proximate teratogen, which is essential to accurately predicting a test compound's developmental toxicity potential. Additionally, maternal–fetal interactions and organogenesis cannot be modeled using an in vitro model. However, one of the advantages of using an in vitro assay is the ability to separate adverse outcomes due to compound versus outcomes due to maternal toxicity from compound exposure. Developmental toxicity testing in cells derived from human embryos is likely to generate more reliable in vitro prediction endpoints than endpoints currently available through the use of animal models, or other in vitro nonhuman assays given the physiologic relevance of hES cells to human development.

This assay can help reduce or eliminate species-specific misinterpretations, reduce need for a second species, and could be included as part of a panel of in vitro assays aimed at defining where potential adverse responses in human populations may exist. Much like other in vitro culture systems that are used to understand potential for target organ toxicity, this assay can assess potential for developmental toxicity. Part of its strength is that this is accomplished across a range of exposure levels. While there is no defined way to project safety margins or fully predict human response based on in vitro data, assays such as this one can help define exposure ranges where response may be expected as well as those where a response would not be expected to occur. Results could then be incorporated into a panel of tests that in aggregate develop an approximation of clinical safety margins. This information could help to drive decisions as to whether a compound should progress along its development path.

ACKNOWLEDGMENTS

The authors acknowledge Agilent Technologies for providing technical assistance, software, and instrumentation. We additionally would like to thank Dr. Gabriela Cezar for her foundational work at the University of Wisconsin Madison in the development of stem-cell–based models for toxicity screening.

REFERENCES

AbbVie. 2013. Depacon [product information]. North Chicago, IL.

ACToR database. 2013. Web site. Available at: http://actor.epa.gov/ actor/faces/ACToRHome.jsp. Accessed July 2013.

Andrx Labs. 2012. Altoprev [product information]. Fort Lauderdale, FL.

Baxter Healthcare. 2012. Penicillin G potassium injection [product information]. Deerfield, IL.

Bedford Laboratories. 2001. Thiotepa [product information]. Bedford, OH.

Bhattacharya S, Zhang Q, Carmichael PL, Boekelheide K, Andersen ME. 2011. Toxicity testing in the 21 century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. PLoS One 6:e20887.

Blomhoff R, Beckman-Sundh U, Brot C, Solvoll K, Steingrimsdottir L, Carlsen MH. 2003. Health risks related to high intake of preformed retinol (vitamin A) in the Nordic countries. Copenhagen: Nordic Council of Ministers.

Brannen KC, Panzica-Kelly JM, Danberry TL, Augustine-Rauch KA. 2010. Development of a zebrafish embryo teratogenicity assay and quantitative prediction model. Birth Defects Res B Dev Reprod Toxicol 89:66–77.

Briggs GG, Freeman RK, Yaffe SJ. 2011. Drugs in pregnancy and lactation. 9th ed. Philadelphia, PA: Lippincott Williams & Wilkins.

Brown NA. 2002. Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. European Centre for the Validation of Alternative Methods. Altern Lab Anim 30: 177–198.

Caffeine Pharmacology. n.d. Medscape web site. Available athttp://reference.medscape.com/drug/cafcit-nodoz-caffeine-342995#10. Accessed August 2011.

Chan LY, Lau TK. 2006. Effect of rosiglitazone on embryonic growth and morphology: a study using a whole rat embryo culture model. Fertil Steril 86:490–492.

Clark RL. 2009. Embryotoxicity of the artemisinin antimalarials and potential consequences for use in women in the first trimester. Reprod Toxicol 28:285–296.

Cole PD, Drachtman RA, Smith AK, Cate S, Larson RA, Hawkins DS, Holcenberg J, Kelly K, Kamen BA. 2005. Phase II trial of oral aminopterin for adults and children with refractory acute leukemia. Clin Cancer Res 11:8089–8096.

Daston GP, Knudsen TB. 2010. Fundamental concepts, current regulatory design and interpretation. In: Knudsen TB, Daston GP, editors. Comprehensive toxicology. 2nd ed. New York: Elsevier. Vol. 12, p 3–9.

Daston GP, Chapin RE, Scialli AR, Piersma AH, Carney EW, Rogers JM, Friedman JM. 2010. A different approach to validating screening assays for developmental toxicity. Birth Defects Res B Dev Reprod Toxicol 89:526–530.

Dr Reddy's Laboratories. 2011. Amoxil [product information]. Bridgewater, NJ.

Drewe J, Delco F, Kissel T, Beglinger C. 2003. Effect of intravenous infusion of thiamine on the disposition kinetics of thiamine and its pyrophosphate. J Clin Pharm Ther 28:47–51.

Ebert AD, Svendsen CN. 2010. Human stem cells and drug screening: opportunities and challenges. Nat Rev Drug Discov 9:367–372.

Genschow E, Spielmann H, Scholz G, Pohl I, Seiler A, Clemann N, Bremer S, Becker K. 2004. Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Altern Lab Anim 32:209–244.

Gilead Sciences. 2000. Vistide [product information]. Foster City, CA.

Gilead Sciences. 2012. Hepsera [product information]. Foster City, CA.

GlaxoSmithKline. 2005. Agenerase [product information]. Research Triangle Park, NC.

GlaxoSmithKline. 2011. Avandia [product information]. Research Triangle Park, NC.

GlaxoSmithKline. 2012. Ziagen [product information]. Research Triangle Park, NC.

GlaxoSmithKline. 2013. Tykerb [product information]. Research Triangle Park, NC.

Gustafson AL, Stedman DB, Ball J, Hillegass JM, Flood A, Zhang CX, Panzica-Kelly J, Cao J, Coburn A, Enright BP, Tornesi MB, Hetheridge M, Augustine-Rauch KA. 2012. Inter-laboratory assessment of a harmonized zebrafish developmental toxicology assay—progress report on phase I. Reprod Toxicol 33:155–164.

Hansen DK. 1995. Folates in reproduction: in vitro studies. Teratology 51:12A.

Hansen DK, Grafton TF, Dial SL. 1993. Effect of dietary supplementation with folic acid on valproate-induced neural tube defects. Teratology 47:420.

Hansen JM. 2006. Oxidative stress as a mechanism of teratogenesis. Birth Defects Res C Embryo Today 78:293–307.

Hartmann S, Brørs O, Bock J, Blomhoff R, Bausch J, Wiegand UW, Hartmann D, Hornig DH. 2005. Exposure to retinyl esters, retinol, and retinoic acids in non-pregnant women following increasing single and repeated oral doses of vitamin A. Ann Nutr Metab 49:155–164.

Herrmann K. 1995. Teratogenic effects of retinoic acid and related substances on the early development of the zebrafish (Brachydanio rerio) as assessed by a novel scoring system. Toxicol In Vitro 9:267–283.

Hilbert J, Radwanski E, Weglein R, Luc V, Perentesis G, Symchowicz S, Zampaglione N. 1987. Pharmacokinetics and dose proportionality of loratadine. J Clin Pharmacol 27:694–698.

Hoyert DL, Mathews TJ, Menacker F, Strobino DM, Guyer B. 2006. Annual summary of vital statistics: 2004. Pediatrics 117:168–183.

Isoniazid (Systemic). n.d. Drugs.com Web site. Available at: http://www.drugs.com/mmx/isoniazid.html. Updated June 30, 2000. Accessed August 2011.

Jelovsek FR, Mattison DR, Chen JJ. 1989. Prediction of risk for human developmental toxicity: how important are animal studies for hazard identification? Obset Gynecol 74:624–636.

Kalhan SC, Bier DM. 2008. Protein and amino acid metabolism in the human newborn. Annu Rev Nutr 28:389–410.

Karim A, Tolbert D, Cao C. 2006. Disposition kinetics and tolerance of escalating single doses of ramelteon, a high-affinity MT1 and MT2 melatonin receptor agonist indicated for treatment of insomnia. J Clin Pharmacol 46:140–148.

Kleinstreuer NC, Smith AM, West PR, Conard KR, Fontaine BR, Weir-Hauptman AM, Palmer JA, Knudsen TB, Dix DJ, Donley EL, Cezar GG. 2011. Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 257:111–121.

Klug S, Lewandowski C, Blankenburg G, Merker HJ, Neubert D. 1985. Effect of acyclovir on mammalian embryonic development in culture. Arch Toxicol 58:89–96.

Klug S, Creech Kraft J, Wildi E, Merker HJ, Persaud TV, Nau H, Neubert D. 1989. Influence of 13-cis and all-trans retinoic acid on rat embryonic development in vitro: correlation with isomerisation and drug transfer to the embryo. Arch Toxicol 63:185–192.

Kovacic P, Somanathan R. 2006. Mechanism of teratogenesis: electron transfer, reactive oxygen species, and antioxidants. Birth Defects Res C Embryo Today 78:308–325.

Lankas GR, Cukierski MA, Wise LD. 2004. The role of maternal toxicity in lovastatin-induced developmental toxicity. Birth Defects Res B Dev Reprod Toxicol 71:111–123.

Leucuţa A, Vlase L, Farcău D, Nanulescu M. 2004. Pharmacokinetic interaction study between ranitidine and metoclopramide. Rom J Gastroenterol 13:211–214.

Liebelt EL, Balk SJ, Faber W, Fisher JW, Hughes CL, Lanzkron SM, Lewis KM, Marchetti F, Mehendale HM, Rogers JM, Shad AT, Skalko RG, Stanek EJ. 2007. NTP-CERHR expert panel report on the reproductive and developmental toxicity of hydroxyurea. Birth Defects Res B Dev Reprod Toxicol 80:259–366.

Louisse J, Gönen S, Rietjens IM, Verwei M. 2011. Relative developmental toxicity potencies of retinoids in the embryonic stem cell test compared with their relative potencies in in vivo and two other in vitro assays for developmental toxicity. Toxicol Lett 203:1–8.

Luna BG, Scavone JM, Greenblatt DJ. 1989. Doxylamine and diphenhydramine pharmacokinetics in women on low-dose estrogen oral contraceptives. J Clin Pharmacol 29:257–260.

Madureira TV, Cruzeiro C, Rocha MJ, Rocha E. 2011. The toxicity potential of pharmaceuticals found in the Douro River estuary (Portugal)–experimental assessment using a zebrafish embryo test. Environ Toxicol Pharmacol 32:212–217.

Mahmood I, Chamberlin N. 1998. A limited sampling method for the estimation of AUC and Cmax of carbamazepine and carbamazepine epoxide following a single and multiple dose of a sustained-release product. Br J Clin Pharmacol 45:241–246.

Mao Y, Gurr JA, Hickok NJ. 1993. Retinoic acid regulates ornithine decarboxylase gene expression at the transcriptional level. Biochem J 295:641–644.

Marx-Stoelting P, Adriaens E, Ahr HJ, Bremer S, Garthoff B, Gelbke HP, Piersma A, Pellizzer C, Reuter U, Rogiers V, Schenk B, Schwengberg S, Seiler A, Spielmann H, Steemans M, Stedman DB, Vanparys P, Vericat JA, Verwei M, van der Water F, Weimer M, Schwarz M. 2009. A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. Altern Lab Anim 37:313–328.

Mayne Pharma. n.d. Aquasol A [product information]. Paramus, NJ.

McGrath P, Li CQ. 2008. Zebrafish: a predictive model for assessing drug-induced toxicity. Drug Discov Today 13:394–401.

McNeil Consumer Healthcare. 2010. Tylenol [Product Information]. Fort Washington, PA.

Merck. 2004. Cuprimine [product information]. Whitehouse Station, NJ.

Merck. 2013. Januvia [product information]. Whitehouse Station, NJ.

Miller RS, Li Q, Cantilena LR, Leary KJ, Saviolakis GA, Melendez V, Smith B, Weina PJ. 2012. Pharmacokinetic profiles of artesunate following multiple intravenous doses of 2, 4, and 8 mg/kg in healthy volunteers: phase 1b study. Malar J 11:255.

Muindi JR, Frankel SR, Huselton C, DeGrazia F, Garland WA, Young CW, Warrell RP Jr. 1992. Clinical pharmacology of oral all-trans retinoic acid in patients with acute promyelocytic leukemia. Cancer Res 52:2138–2142.

National Research Council (NRC). 2000. Scientific frontiers in developmental toxicology and risk assessment. Washington, DC: The National Academies Press.

National Research Council (NRC). 2007. Toxicity testing in the 21st century. A vision and a strategy. Washington, DC: The National Academies Press.

Novartis Pharmaceuticals. 2010. Comtan [product information]. East Hanover, NJ.

Novartis Sverige AB. 2011. Everolimus [product information]. Täby, Sweden.

Oman M, Lundqvist S, Gustavsson B, Hafström LO, Naredi P. 2005. Phase I/II trial of intraperitoneal 5-Fluorouracil with and without intravenous Vasopressin in non-resectable pancreas cancer. Cancer Chemother Pharmacol 56:603–609.

Otsuka America Pharmaceutical. 2011. Busulfex [product information]. Rockville, MD.

Padayatty SJ, Sun H, Wang Y, Riordan HD, Hewitt SM, Katz A, Wesley RA, Levine M. 2004. Vitamin C pharmacokinetics: implications for oral and intravenous use. Ann Intern Med 140:533–537.

Palma-Aguirre JA, Absalón-Reyes JA, Novoa-Heckel G, de Lago A, Oliva I, Rodríguez Z, González-de la Parra M, Burke-Fraga V, Namur S. 2007. Bioavailability of two oral suspension and two oral tablet formulations of acyclovir 400 mg: two single-dose, open-label, randomized, two-period crossover comparisons in healthy Mexican adult subjects. Clin Ther 29:1146–1152.

Paquette JA, Kumpf SW, Streck RD, Thomson JJ, Chapin RE, Stedman DB. 2008. Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. Birth Defects Res B Dev Reprod Toxicol 83:104–111.

Pegg A. 2009. Mammalian polyamine metabolism and function. IUBMB Life 61:880–894.

Pfizer. 2012. Dilantin [product information]. New York, NY.

Piersma AH, Attenon P, Bechter R, Govers MJAP, Krafft N, Schmid BP, Stadler J, Verhoef A, Verseil C. 1995. Interlaboratory evaluation of embryotoxicity in the postimplantation rat embryo culture. Reprod Toxicol 9:275–280.

Ritchie HE, Brown-Woodman PD, Korabelnikoff A. 2003. Effect of co-administration of retinoids on rat embryo development in vitro. Birth Defects Res A Clin Mol Teratol 67:444–451.

Ritz C, Streibig JC. 2005. Bioassay analysis using R. J Statistical Software 12:1–22.

Robinson JF, van Beelen VA, Verhoef A, Renkens MF, Luijten M, van Herwijnen MH, Westerman A, Pennings JL, Piersma AH. 2010. Embryotoxicant-specific transcriptomic responses in rat postimplantation whole-embryo culture. Toxicol Sci 118:675–685.

Roche Laboratories. 2010. Accutane [product information]. Nutley, NJ.

Scott CW, Peters MF, Dragan YP. 2013. Human induced pluripotent stem cells and their use in drug discovery for toxicity testing. Toxicol Lett 219:49–58.

Selderslaghs IW, Blust R, Witters HE. 2012. Feasibility study of the zebrafish assay as an alternative method to screen for developmental toxicity and embryotoxicity using a training set of 27 compounds. Reprod Toxicol 33:142–154.

Shepard TH, Lemire RJ. 2007. Catalog of teratogenic agents. 12th ed. Baltimore, MD: The Johns Hopkins University Press.

Shoda H, Inokuma S, Yajima N, Tanaka Y, Oobayashi T, Setoguchi K. 2007. Higher maximal serum concentration of methotrexate predicts the incidence of adverse reactions in Japanese rheumatoid arthritis patients. Mod Rheumatol 17:311–316.

Shuren J. 2008. Content and format of labeling for human prescription drug and biological products; requirements for pregnancy and lactation labeling. Fed Regist 73:30831–30868.

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787.

Stark KL, Lee QP, Namkung MJ, Harris C, Juchau MR. 1990. Dysmorphogenesis elicited by microinjected acetaminophen analogs and metabolites in rat embryos cultured in vitro. J Pharmacol Exp Ther 255: 74–82.

Tandon S, Jyoti S. 2012. Embryonic stem cells: an alternative approach to developmental toxicity testing. J Pharm Bioallied Sci 4:96–100.

Teratogen Information System (TERIS). 2013. Web site. Available at: http://depts.washington.edu/terisweb/teris/. Accessed July 2013.

Teratology Society. 1987. Teratology Society position paper: recommendations for vitamin A use during pregnancy. Teratology 35:269–275.

Thalidomide Pharmacology. n.d. Medscape web site. Available at: http://reference.medscape.com/drug/thalomid-thalidomide-343211#10. Accessed August 2011.

Theunissen PT, Piersma AH. 2012. Innovative approaches in the embryonic stem cell test (EST). Front Biosci 17:1965–1975.

Thomson J, Johnson K, Chapin R, Stedman D, Kumpf S, Ozolinš TR. 2011. Not a walk in the park: the ECVAM whole embryo culture model challenged with pharmaceuticals and attempted improvements with random forest design. Birth Defects Res B Dev Reprod Toxicol 92:111–121.

Ubeda N, Reyes L, González-Medina A, Alonso-Aperte E, Varela-Moreiras G. 2011. Physiologic changes in homocysteine metabolism in pregnancy: a longitudinal study in Spain. Nutrition 27:925–930.

Vaisman M, Spina LD, Eksterman LF, dos Santos MJ, Lima JS, Volpato NM, da Silva RL, de Brito AP, Noël F. 2001. Comparative bioavailability of two oral L-thyroxine formulations after multiple dose administration in patients with hypothyroidism and its relation with therapeutic endpoints and dissolution profiles. Arzneimittelforschung 51:246–252.

van Giersbergen PL, Treiber A, Schneiter R, Dietrich H, Dingemanse J. 2007. Inhibitory and inductive effects of rifampin on the pharmacokinetics of bosentan in healthy subjects. Clin Pharmacol Ther 81:414–419.

Warner Chilcott. 2013. Sarafem [product information]. Rockaway, NJ.

Weigt S, Huebler N, Strecker R, Braunbeck T, Broschard TH. 2011. Zebrafish (Danio rerio) embryos as a model for testing proteratogens. Toxicology 281:25–36.

Weinstein HJ, Griffin TW, Feeney J, Cohen HJ, Propper RD, Sallan SE. 1982. Pharmacokinetics of continuous intravenous and subcutaneous infusions of cytosine arabinoside. Blood 59:1351–1353.

Welle-Watne A, Wiik I, Waaler T. 1980. Bioavailability of warfarin from three tablet preparations. Medd Norsk Farm Selsk 42:103–114.

West PR, Weir AM, Smith AM, Donley EL, Cezar GG. 2010. Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. Toxicol Appl Pharmacol 247: 18–27.

Zhang C, Cao J, Kenyon JR, Panzica-Kelly JM, Gong L, Augustine-Rauch K. 2012. Development of a streamlined rat whole embryo culture assay for classifying teratogenic potential of pharmaceutical compounds. Toxicol Sci 127:535–546.