TABLE OF CONTENTS

Administrative Information

1) *Submission Title:* Biomarkers for *in vitro* developmental toxicity screening in a human system

2) *Requesting Organization:*

Name: Stemina Biomarker Discovery, Inc.
Address: 504 S. Rosa Road, Suite 150, Madison, WI 53719
Phone: (608) 204-0104
Website: www.stemina.com

3) *Primary Contact:*

Name: Elizabeth Donley, JD, MBA, MS; Chief Executive Officer
Phone: (608) 577-9209
Email: bdonley@stemina.com

4) *Alternate Contact:*

Name: Jessica Palmer, MS; Associate Director of Toxicology
Phone: (608) 474-1987
Email: jpalmer@stemina.com

5) *LOI Submission Date:* May 4, 2021

Drug Development Need Statement

The current safety assessment requirements for developmental toxicity testing have been in place for over 40 years and require that a new compound be evaluated for effects on embryo-fetal development (EFD) in two animal species, one rodent and one non-rodent. However, these animal models have varying degrees of concordance with observed human outcomes, having approximately 70% concordance to known human developmental toxicants.[1] Further confounding the low level of prediction of human response by either *in vivo* system, the agreement between toxicity predictions in different animal species is only 60%,[2] which calls into question the relevance of using *in vivo* animal models for predicting human developmental toxicity. In addition, a recent study by Braakhuis et al. (2019)[3] found that there can be up to a 25-fold difference in the NOAEL or LOAEL for a given compound when it is tested within the same species as a result of study replication errors, demonstrating the need for newer, more reliable tests to evaluate developmental toxicity and species specific response. Though these animal models are, and have long been, the regulatory standard, differences in species specific response to a test compound may lead to missed signals of developmental toxicity and biological misinterpretation, resulting in compounds being classified as false positives or false negatives with respect to their developmental toxicity potential in humans. Using innovative testing methods centered on human stem cell-based *in vitro* systems at the preclinical stage of drug discovery will provide predictive and protective biomarkers to advance the understanding of human toxicity potential and replace, reduce, and refine animal testing.

Stemina Biomarker Discovery, Inc. has developed both a human embryonic stem cell (hES) and a human induced pluripotent stem cell (iPS) cell-based assay, devTOX *quick*Predict (devTOX^{qP}), for predicting the developmental toxicity potential of compounds that can be used in conjunction with the required *in vivo* tests or as an alternative to one or more of these tests. Development and validation of new approach methods (NAMs) (*in vitro, ex vivo, or in silico*) contributes to the Replacement, Reduction, and Refinement (3Rs) of animal use in toxicology studies. This is important when considering the current regulations and global initiatives to test thousands of chemicals currently in commercial use for their toxicity potential. Additionally, there have been increased efforts in the pharmaceutical and chemical industries to incorporate NAMs earlier in the product development pipeline prior to *in vivo* testing, highlighted by the recent publication from the FDA regarding the agency's perspective on incorporating NAMs into the traditional

nonclinical testing strategies.[4] The assay can replace the need for animal testing or testing in a second species in specific cases under the revised S5 (R3) guideline on Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals recently issued by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Under the revised guidelines, qualified alternative methods can be used to defer or replace conventional *in vivo* testing in certain cases, such as when a pharmaceutical is expected to have an adverse effect on embryo-fetal development and for pharmaceuticals being developed for severely debilitating/life-threatening diseases or late-life onset diseases.

Using human iPS cells eliminates the complex ethical and political issues associated with hES cells while providing insight into species-specific differences in response, which is particularly important since human embryos can have a higher sensitivity to compounds than rodent models (e.g. thalidomide).[5] Stemina has demonstrated that human iPS cells have similar developmental toxicity profiles as hES cells, which are metabolically similar to embryonic epiblast cells at gastrulation. Additionally, iPS cells can recapitulate every cell in the body thereby offering the opportunity to identify disruptions in human development in a human *in vitro* system. The assay predicted the developmental toxicity potential for 65 pharmaceuticals with a balanced accuracy of 90% (87% sensitivity, 92% specificity) when compared to human data and animal developmental toxicity studies. The assay was 80% and 85% concordant with *in vivo* rodent (N=65) and rabbit (N=53) EFD studies, respectively. There is a need for a human-specific endpoint for developmental toxicity testing. The devTOX^{qP} assay provides data on human response and greater accuracy than either the rodent or the rabbit. It is a necessary addition to protect human health, to replace (in certain cases), reduce, and refine animal testing, and to help to reconcile discordant information from the two required *in vivo* endpoints.

Biomarker Information

1) **Biomarker Name**
   a) Metabolite Ratio of Ornithine to Cystine
2) **Biomarker Type**
   a) Other: Metabolite
3) **Analytical Methods**
   a) The levels of ornithine and cystine in the spent media of human induced pluripotent stem cells are measured using ultra-performance liquid chromatography-high resolution mass spectrometry. The method also measures the levels of stable isotope labeled ornithine and cystine, which are added to the samples during sample preparation.
   b) Measurement units and limit(s) of detection: Not Applicable. A quantitative analytical method is not required for conducting the assay.
4) **Post-analytical processing of raw biomarker measurement**
   a) The extracted ion chromatogram (EIC) areas for ornithine and cystine are normalized to the internal standards (ISTD) by dividing the endogenous metabolite signal by the corresponding isotopically labeled ISTD area. Metabolite-specific internal standards are used to correct for changes in the metabolite abundance due to variability in the measurement due to the instrument. They also identify and correct for changes in metabolite abundance due to suppression or interference from the test compound.
   b) Relative fold changes are calculated for each ISTD-normalized metabolite in each sample by dividing the ISTD-normalized value by the median value of the reference treatment samples (cells treated with 0.1% DMSO). Normalization to the plate-specific reference controls accounts for experiment-to-experiment variation and allows the changes in metabolite abundance to be analyzed as a change from the “normal” state of cellular metabolism.
   c) The ratio of ornithine to cystine (o/c ratio) is determined by dividing the reference-normalized value of

ornithine by the reference-normalized value of cystine. Using the metabolites as a ratio of ornithine to cystine was more predictive than when the metabolites were used independently to predict developmental toxicity.[6]

5) Index/scoring system rationale and biomarker interpretation and utility

a) The assay was designed to address the premise that toxicity is dependent, to a large degree, on exposure level. Changes in the o/c ratio following compound exposure are measured over an eight-point dose-response curve and the response is fit with a nonlinear dose-response model. Prediction of developmental toxicity potential is based on the developmental toxicity potential (dTP) concentration, which is defined as the point when the o/c ratio response curve decreases below a critical threshold (referred to as the developmental toxicity threshold, dTT) (Figure 1). This threshold was determined using the dose-response results from the 23-compound training set, based on pharmaceuticals with human developmental toxicity potential established by the FDA. The dTT was optimized for each cell type (hES and iPS cells) by selecting the threshold that produced the highest accuracy of prediction with the greatest sensitivity (the rational and methods used to set the threshold are described in Palmer et al., 2013).[6] Concentration levels greater than the dTP concentration for a given compound are predicted to have developmental toxicity potential in vivo. Cell viability is provided as a secondary endpoint to help interpret the response in the context of embryo toxicity but is not included in the prediction of developmental toxicity potential.

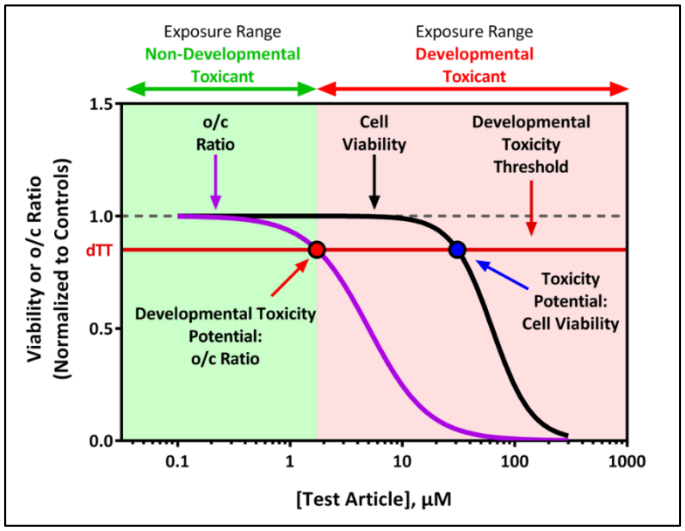

Figure 1: devTOX^{qP} Results Interpretation. The dose-response curves for the o/c ratio and cell viability are illustrated with purple and black lines, respectively. The concentration predicted by the point where the dose-response curve of the o/c ratio crosses the developmental toxicity threshold (dTT, red line) indicates the concentration where a test article (compound) has developmental toxicity potential (Developmental Toxicity Potential: o/c Ratio, red point). The toxicity potential concentration from cell viability (blue point) is the point where the cell viability dose-response curve exceeds the developmental toxicity threshold. The developmental toxicity threshold creates a two-sided toxicity model based on exposure: one where exposure does not perturb metabolism in a manner associated with developmental toxicity (green box) and another where exposure shifts metabolism in manner associated with developmental toxicity (red box). The *x*-axis is the concentration of the test article. The *y*-axis is the reference-normalized (fold change) values for the o/c ratio and viability.

b) Two different scoring methods have been used to assess the accuracy of the assay based on the knowledge of the human therapeutic or *in vivo* animal C_{max} concentration at doses shown to cause developmental toxicity (Figure 2).

i) If pharmacokinetic information is available, the dTP concentration is compared with the C_{max} concentration to score the toxicity potential. This model is also employed when the C_{max} known to cause developmental toxicity in an *in vivo* model is available (typically rodent). If the dTP concentration is lower than the $10\times C_{max}$, the compound is classified as a developmental toxicant. If the dTP concentration is greater than the $10\times C_{max}$, the compound is classified as a non-developmental toxicant. *In vitro* responses observed between 10× and 50× the *in vivo* efficacious exposure are considered to be relevant for prediction of *in vivo* toxicity.[7]

ii) We recognize that exposure information is often not available early in drug development. In these cases, changes in iPSC metabolism can be used as a signal regarding the developmental toxicity potential of the compound. It is also possible to define a concentration threshold for classification using the predicted dTP concentration. Using a broad range of chemotypes, we identified a concentration

threshold of 65 µM as the optimal threshold for separating developmental toxicants from non-developmental toxicants based on balanced sensitivity and specificity. This threshold could be optimized using a training set of only pharmaceuticals or to maximize sensitivity rather than balanced accuracy depending upon an objective to reduce false negatives. Compounds with dTP values below 65 µM are predicted to be developmental toxicants while those with values above 65 µM are predicted to be negative for developmental toxicity *in vivo*. Once data is available on the therapeutic range for the compound, the dTP of the compound can be compared to efficacy ranges to provide insight on the likely toxicity of the compound in that range.

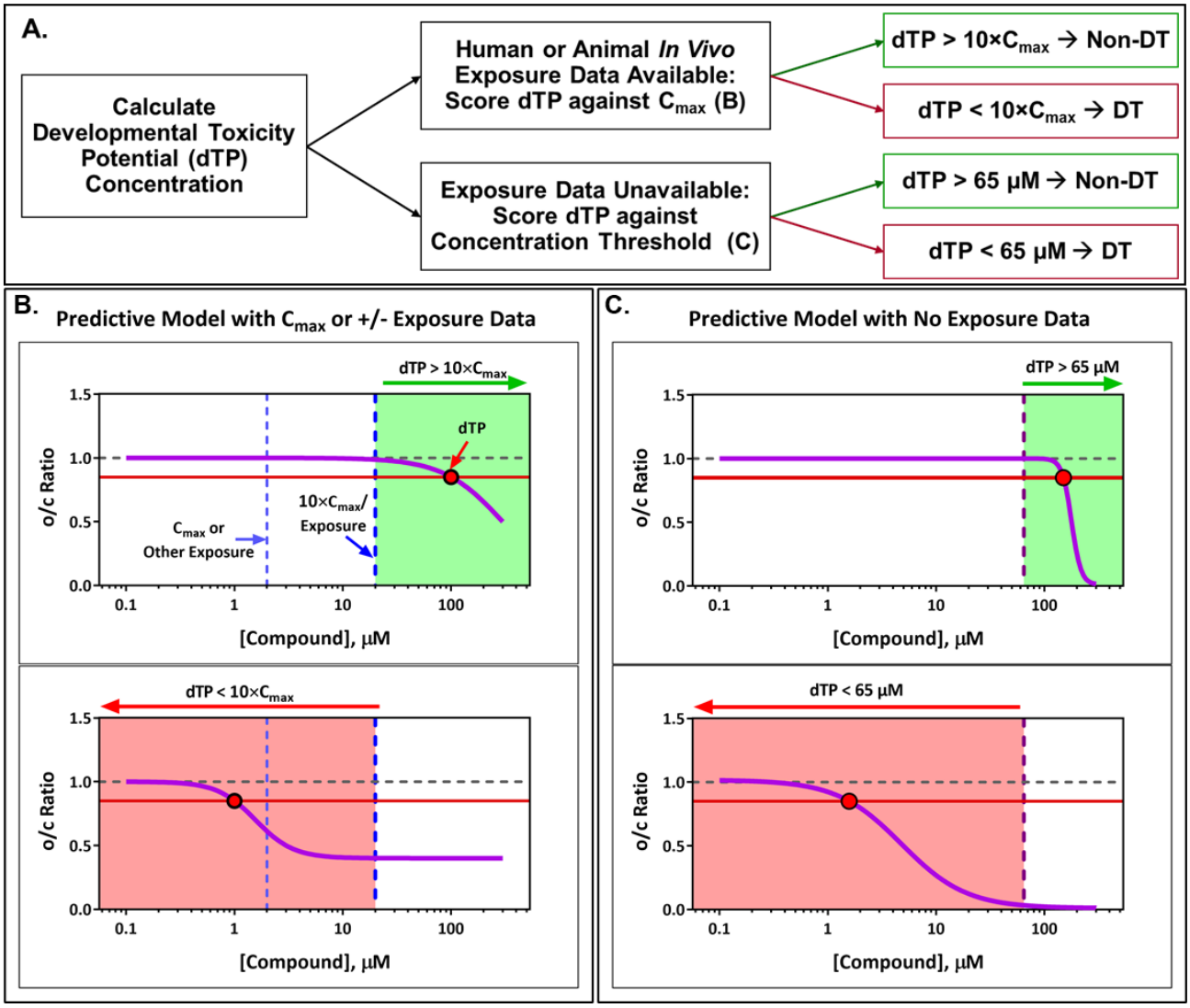

Figure 2: Visual Explanation of devTOX^{qP} Prediction Models. **(A)** Decision tree for determining the scoring method of the prediction model. **(B)** Graphical representation of the classification scheme utilizing the C_{max} concentration to set the classification windows. **(C)** Graphical representation of the classification scheme utilizing a concentration threshold of 65 µM when no human or *in vivo* exposure data is available. In each panel, the purple curve represents the dose–response curve for the o/c ratio, the black-bordered red circle represents the (dTP), and the red line is the dTT. The *x*-axis is the concentration (µM) of the compound. The *y*-axis is the reference-normalized (fold change) values for the o/c ratio. If the dTP concentration is present in the green shading, the compound is classified as a non-developmental toxicant and if the dTP concentration falls in the red shading, the compound is classified as a developmental toxicant.

c) The scoring methods described above and outlined in Figure 2 indicate how we have evaluated the accuracy of the assay and do not necessarily apply to how the assay will be used for new compounds with unknown

toxicity. Directly comparing the dTP concentration to established C_{max} data (animal or human) is the simplest way of interpreting the assay results; however, this data may not be available, or it may be more relevant to interpret the predicted dTP concentration in terms of administered dose. Recent advances in *in vitro* to *in vivo* extrapolation (IVIVE) using reverse-dosimetry physiologically based kinetic (PBK) modeling have provided easier access to "generic" pharmacokinetics (PK) models that can be used for a broad range of compounds. IVIVE incorporates various PK parameters (dependent on PK model) to convert an *in vitro* effect concentration (such as the dTP) to an *in vivo* dose in humans or rodents. These results can be used to interpret the assay results within the context of the *in vivo* EFD studies or the intended human dose. For example, devTOX^{qP} predicts thalidomide to be developmentally toxic, with a dTP concentration of 0.09 µM. This corresponds to a human oral equivalent administered dose (EAD) of 0.01 mg/kg/day (calculated with the a multiple-compartment PBPK model available at https://ice.ntp.niehs.nih.gov/)[8], indicating that thalidomide is expected to be developmentally toxic at therapeutic doses (0.83 mg/kg/day), which are known to cause teratogenicity in humans. In contrast, if a drug does not cause a response in the o/c ratio indicative of developmental toxicity, such as amoxicillin, the top concentration tested can be used to determine the human or rat EAD. The human oral EAD for 300 µM amoxicillin (highest exposure tested) is 195 mg/kg, which indicates that this drug would not be developmentally toxic at therapeutic doses (12.5-29.2 mg/kg).

In a joint study with the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), we evaluated a series of ten valproic acid (VPA) analogs with the devTOX^{qP} assay and applied IVIVE to the results to determine if the assay could quantitatively predict the *in vivo* exposure resulting in developmental toxicity (poster provided as Appendix 2). In this study, five different PK models (3 non-pregnancy, 2 human pregnancy models) were used to estimate equivalent administered doses (EADs) that would result in maternal or fetal plasma concentrations equivalent to the *in vitro* dTP concentrations identified with the devTOX^{qP} assay. Five of the analogues had published lowest effect levels (LELs) from *in vivo* rat developmental toxicity studies. For these five analogues, at least one PK model produced a rat EAD within 1.5-fold of the rat LEL range. VPA was the only compound in this study with a human clinical dose. The human EADs from all 5 human models were within or lower than the clinical dose range. The close agreement between EADs and *in vivo* rat LELs suggests that using devTOX^{qP} assay input data and IVIVE can quantitatively predict *in vivo* developmental toxicity potential at relevant concentrations.

Context of Use Statement

The ratio of ornithine to cystine may be used as a safety biomarker for detecting human developmental toxicity potential *in vitro* using human pluripotent stem cells at the nonclinical stage of drug development for small molecule drug candidates expected to be embryo-fetal toxicants as described in Annex 2 of the ICH S5(R3) guideline.

Analytical Considerations

1) Biomarker measurement description

a) The o/c Ratio is a measurement of the change in ornithine and cystine by ultra-performance liquid chromatography-high resolution mass spectrometry relative to the experiment-specific reference controls (cells treated with 0.1% DMSO).

2) Sample matrix and stability

a) The samples used to determine the o/c ratio are spent media from human iPSCs treated with the test article of interest. After collection, samples are filtered to remove high molecular weight constituents (>3KDa) and the filtrate is concentrated. Next the concentrated sample is resolubilized in a 1:1 mixture of 0.1% formic acid in water: 0.1% formic acid in acetonitrile prior to LC-MS analysis.

b) Stability experiments have demonstrated the stability of ornithine, cystine, and the o/c ratio in: 1) samples quenched with acetonitrile and stored at -80°C for up to 2 months; 2) filtered and concentrated samples stored at 4°C for up to 1 month; and 3) resolubilized samples stored at 4°C for up to 2 weeks. There was no difference in the o/c ratio and predicted dTP concentration between samples stored under these conditions compared to samples analyzed immediately after preparation.

3) Quality System

a) Stemina developed a GLP-like quality management plan (QMP) and standard operating procedures (SOPs) as required by the US Environmental Protection Agency (EPA) to qualify as a ToxCast contractor. This QMP and SOPs provide the quality, robustness and reproducibility that govern execution of the devTOX^{qP} assay. The SOPs cover all steps in assay execution from test article receipt through reporting, including maintenance of the equipment required to conduct the assay. These SOPs are available upon request. The SOPs have been refined through continual improvement to broaden the scope of QC metrics applied to cell culture, assay endpoints, and instrumentation as well as to increase transferability. Every new batch is pre-qualified with three drug treatments before use in production and every experimental plate includes reference, positive and negative controls to ensure that iPS cell metabolism is within the assay specifications

b) Stemina's quality system was audited in 2015 by the EPA as part of our ToxCast contract. There were no significant findings during this audit. Additionally, Stemina received an "Exceptional" rating from the EPA for Quality, Schedule, and Management on the Contractor Performance Assessment Report regarding the devTOX^{qP} work that has been performed for our ToxCast contract.

4) Analytical validation plan

a) Stemina has extensive experience with analytical method validation of metabolite biomarkers based on the May 2018 FDA Guidance for Industry – Bioanalytical Method Validation[9] and the Clinical Laboratory Standards Institute's (CLSI) Liquid Chromatography-Mass Spectrometry Methods; Approved Guideline (CLSI document C62-A)[10] and will design a fit-for-purpose analytical method validation plan based on these guidelines. The validation plan will include criteria for assessing precision, sensitivity, specificity, dilution linearity, reinjection reproducibility, extraction recovery, and sample carryover. Sample stability has already been evaluated as described above. Accuracy (spike-in recovery) and analytical measurement range are not included in the validation plan as the data is normalized to the reference controls present on each experimental plate and the assay does not require a quantitative analytical method.

Clinical Considerations

1) Use in drug development

a) The positioning of devTOX^{qP} in the drug development pipeline, and ultimately, its application, is dependent upon multiple factors, including, but not limited to, company policy, drug development program, patient population (i.e., is the drug intended for women of childbearing age), and regulatory requirements or guidelines for developmental toxicity testing. The context of use proposed in this LOI is focused on the use of devTOX^{qP} to predict pharmaceuticals expected to cause MEFL as described in the new ICH S5(R3) guideline (Figure 3A). Annex 2 of the new ICH S5(R3) guideline, Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals , describes how alternative methods, like devTOX^{qP}, can be used in EFD testing strategies for pharmaceuticals expected to be embryo-fetal toxicants (Figure 3B) and pharmaceuticals being developed for certain severely debilitating or life-threatening (SDLT), or late-life onset (LLO) diseases.[11] In these scenarios, a positive response in the devTOX^{qP} assay is sufficient for classifying the compound as positive for malformations or embryo-fetal lethality (MEFL) and no additional *in vivo* testing is required. If the compound is negative or equivocal in the devTOX^{qP} assay, a definitive *in vivo* EFD study would need to be conducted in one or two species (Figure 3B). In addition to these examples, the FDA's Guidance for Industry

on Reproductive Toxicity Testing and Labeling Recommendations for Oncology Pharmaceuticals indicates that alternative assays can be included as part of a weight of evidence (WOE) approach that shows reproductive or developmental toxicity can eliminate the need to conduct dedicated EFD studies for certain pharmaceuticals, many of which would likely fall under the scenarios described in the ICH S5(R3) guideline.[12]

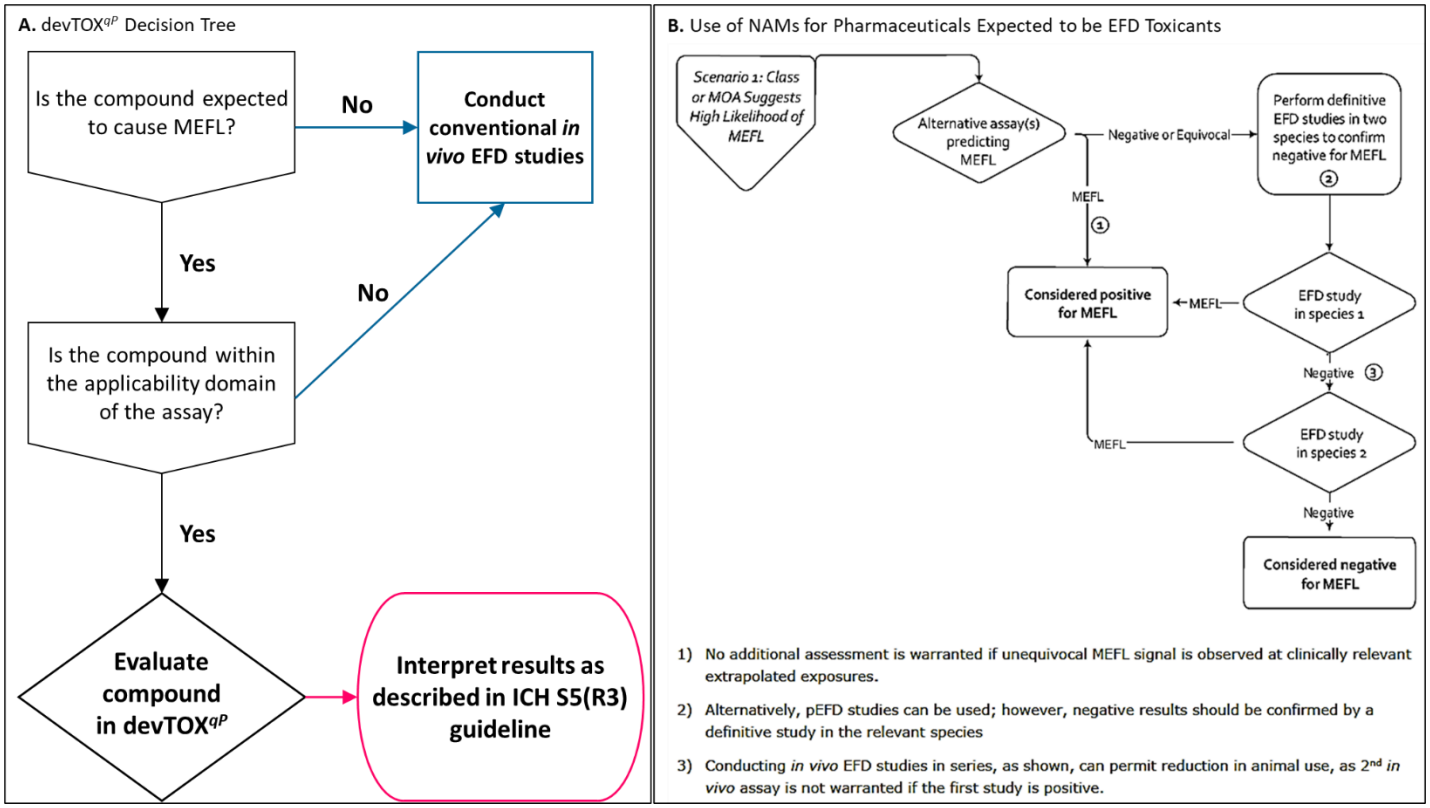

Figure 3: A) Decision tree for using devTOX^{qP} in place of or to support *in vivo* EFD studies for pharmaceuticals that are expected to cause malformations or embryo-fetal lethality (MEFL). **B)** Example ICH S5(R3) testing strategy for New Approach Methods (NAMs) for pharmaceuticals expected to cause MEFL.

2) Benefits and risks of applying the biomarker in drug development

a) Better prediction of human developmental toxicity potential using a human *in vitro* system would benefit consumer health by delivering safer drugs to the market thereby reducing the potential for children born with birth defects and developmental disabilities due to drug exposure *in utero*. The physiology of rodents and rabbits, which are used most often for assessing the potential of drugs to cause birth defects, differs significantly from humans. In its report, "Toxicity Testing in the 21st Century: A Vision and Strategy" (Tox-21c), the United States National Research Council (NRC) presented a vision for the future wherein toxicity testing is done largely *in vitro* using human relevant cell-based models.[13] The devTOX^{qP} assay is aligned with this vision. A significant benefit of the devTOX^{qP} assay is that it is a human system, reducing the risk of false-negatives and false-positives due to inter-species differences in developmental pathways and pharmacokinetics.[14] This is particularly important since human embryos can have a higher sensitivity to compounds than rodent models.[5]

b) The assay can replace the need for animal testing or a second species in specific cases under the revised S5 (R3) guideline on Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals recently issued by the ICH. Under the revised guidelines, qualified, alternative methods can be used to defer or replace conventional *in vivo* testing in certain cases, such as when a pharmaceutical is expected to have an adverse

effect on embryo-fetal development and pharmaceuticals being developed for severely debilitating/life-threatening diseases or late-life onset diseases.

c) As with animal models and other *in vitro* assays, there is a risk of false positives and false negatives. The risk for false negatives can be mitigated by testing negative compounds in an *in vivo* model or other *in vitro* assays prior to human exposure (as shown in the decision tree). False positives are less common with the assay, as demonstrated in our recent publication with the US EPA.[15]

3) Limitations in applying the biomarker in drug development

a) Human iPS cells address the biology of the early developing embryo; however, as with all models, limitations exist in their capacity to predict developmental toxicity.

i) We do not have a complete understanding of all the changes in gene expression or biochemical pathways that can impact ornithine and cystine metabolism.

ii) Toxicities associated with later stage development may not be captured well since the targets of toxicity may not be expressed in undifferentiated cells.

iii) *In vitro* models also lack the mechanisms associated with the effects of absorption, distribution, metabolism, and excretion (ADME), which make it difficult to extrapolate doses, tissue/cellular chemical delivery, and duration of exposure. Prior knowledge of drug metabolism provides an opportunity to assess the toxicity of specific drug metabolites, thereby partially mitigating limitations associated with drug metabolism. Recent advances in modeling IVIVE, which compare the exposure levels where effects are observed with *in vitro* assay to the dose required to observe the specific effect in an *in vivo* animal test, provide promise for dose extrapolation.

iv) Additional aspects of maternal-fetal interactions, environmental, genetics, and prenatal care (nutrition, drug use, etc.) also go beyond the scope of current *in vitro* assays.

Supporting Information

1) Biological Rationale

a) Cystine and ornithine have many roles during early human development and are involved in metabolic pathways important for normal cell proliferation and differentiation, which is the biological basis of the assay. These two biomarkers provide important predictive information, using divergent pathways that provide an accurate evaluation of a compound's developmental toxicity potential. Cystine is initially present in the medium and is the extracellular constituent of the cysteine/cystine thiol redox couple, a critical component of a cell's signaling and regulatory pathways. Cysteine is used by cells in glutathione production; as such, the change connected to decreased cystine uptake likely reflects a change in cellular glutathione synthesis and redox balance. Perturbation of cystine uptake indicates a disruption in the cell's ability to use reactive oxygen species (ROS) related pathways or remove ROS byproducts that can have direct consequences on embryonic development. Previous work has shown that ROS production within the cellular and mitochondrial compartments can lead to developmental toxicity.[16–18] Cystine's role has been investigated regarding its capacity to modulate differentiation, proliferation, apoptosis, and other cellular events during development that may lead to developmental toxicity, demonstrating that changes in this metabolite is an important cellular event in multiple adverse outcome pathways (AOPs).[18,19] During early development, embryonic epiblast-like cells, such as human iPS cells, are highly glycolytic and a have a reduced capacity to adapt to altered ROS states, which makes them sensitive to compounds that may increase ROS or disrupt ROS regulation.[20] A broad spectrum of developmental toxicants, 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 as a mechanism of action of developmental toxicity.[18,21] By measuring cystine, this assay provides insight into a cell's redox status. When its transport is perturbed,

cystine acts as a biomarker indicative of a disruption in a cell's ability to properly mitigate and signal through ROS-related mechanisms.

b) Ornithine is formed as a product of the catabolism of arginine into urea. It functions in several biochemical pathways including ammonia detoxification in the urea cycle, pyrimidine synthesis via ornithine transcarbamylase, and polyamine synthesis via ornithine decarboxylase. Ornithine is initially absent from the medium and is secreted by viable cells; as such, increased or decreased cellular release reflects the general metabolic states of these pathways. Catabolism of ornithine is impacted by developmental toxicants such as all-*trans* retinoic acid, which is a suppressor of the transcription of ornithine decarboxylase (ODC). ODC suppression results in increased ornithine secretion which in turn inhibits polyamine synthesis, an important pathway for normal cell growth and development during embryogenesis, leading to birth defects.[22] Similar to cystine, changes in ornithine can provide insight into the key cellular events of AOPs.

c) We have evaluated the o/c ratio as a marker for developmental toxicity with a broad range of compound classes with varying mechanisms of developmental toxicity. The changes in ornithine and/or cystine concentrations in the media likely represent the convergence of many upstream pathways. To determine if retinoic acid receptor (RAR) activation is one of these upstream pathways, seven retinoids (all-*trans*-retinoic acid (ATRA), 13-*cis*-retinoic acid (13-*cis*-RA), 9-*cis*-retinoic acid (9-*cis*-RA), etretinate, acitretin, retinol, and arotinoid acid (TTNPB)) were evaluated in the devTOX^{qP} assay alone and in the presence of the RARα-selective antagonist Ro 41–5253.[23] All seven retinoids caused a concentration-dependent decrease in the o/c ratio, which corresponded to a concentration-dependent increase in cystine. Five of the retinoids caused a modest concentration-dependent increase in ornithine (ATRA, TTNPB, 13-cis-RA, 9-cis-RA, and acitretin). The presence of Ro 41-5253 inhibited the decrease in the o/c ratio for five of the seven retinoids (ATRA, TTNPB, 13-cis-RA, 9-cis-RA, and acitretin), which was driven by inhibiting the observed increase in cystine induced by these retinoids. As expected, Ro 41-5253 co-treatment did not affect the changes in ornithine and cystine metabolism observed following etretinate and retinol exposure. These data indicate that RAR activation is one of the signaling pathways that can alter cystine transport in iPS cells.

2) Summary of existing preclinical data to support the biomarker in its COU

a) The assay predicted the developmental toxicity potential for 65 pharmaceuticals with a balanced accuracy of 90% (87% sensitivity, 92% specificity) when compared to human data and animal developmental toxicity studies (Appendix 1). To better understand the applicability domain of the assay, these results were separated into different pharmacological categories and performance was assessed. The assay's sensitivity in the different pharmacological categories ranged from 50% to 100% and provides insight into the assay's biological applicability domain (Table 1). For example, developmental toxicants classified as channel, kinase, and transcription modulators and DNA modifiers were predicted as developmentally toxic with 100% sensitivity. In contrast, receptor modulators were predicted with 50% sensitivity, and accurate prediction was highly dependent upon whether the iPS cells expressed the specific receptor being modulated.

Table 1: devTOX^{qP} Predictivity by Pharmacological Class

Pharmacological Classification	# Positive	# Negative	Balanced Accuracy	Sensitivity	Specificity
Channel Modulators	3	1	100%	100%	100%
DNA Modifiers	2	0	100%	100%	N/A
Enzyme Modulators	7	6	85%	86%	83%
Kinase Modulator	4	0	100%	100%	N/A
Nucleoside Modulator/ Central Metabolite Inhibitor	6	0	100%	100%	N/A
Receptor Modulator	4	10	75%	50%	100%

Table 1: devTOX^{qP} Predictivity by Pharmacological Class

Pharmacological Classification	# Positive	# Negative	Balanced Accuracy	Sensitivity	Specificity
Second Messenger Modulator	0	1	100%	N/A	100%
Transcription Modulator	6	1	100%	100%	100%
Other	7	7	86%	86%	100%

b) Assay endpoint (o/c ratio, cell viability, ornithine, and cystine) reproducibility has been evaluated using the positive control, negative control, and reference treatments from over 50 independently cultured 96-well plates analyzed over a 3-year period. The average intraplate coefficient of variation (CV) across endpoints is 4.1% and the average interplate CV across endpoints is 9.9% for the assay. Additionally, the reproducibility of the predictive model has been evaluated using independent replicates of three drug treatments (carbamazepine, n=40; methotrexate, n=40; thalidomide, n=12) conducted by multiple technicians with multiple iPS cell lines, freeze lots and reagents over the course of 5 years. The interpolated developmental toxicity potential (dTP) values were within two standard deviations of the mean for each of the compounds, demonstrating that the assay endpoints are reproducible over time.

3) Summary of planned studies to support the biomarker and COU

a) Annex 2 of the new ICH S5(R3) guideline outlines the requirements for qualifying alternative methods for predicting MEFL, including a list of 29 positive and 3 negative reference compounds that have been shown to induce MEFL in nonclinical studies and/or humans. We have previously evaluated fifteen of these compounds (data included in the results provided in Appendix 1). We are testing the remaining 17 compounds and plan to prepare the necessary documents to qualify the assay according to the ICH S5(R3) guideline.

4) Description of alternative comparator, current standard(s), or approaches

a) Historically, the regulatory guidelines for developmental toxicity required that a new compound be evaluated for effects on EFD in two animal species, one rodent and one non-rodent. While the new ICH S5(R3) guideline allow the use of NAMs in specific scenarios, the majority of new compounds still require these same *in vivo* studies. Unfortunately, no animal species replicates human development well in terms of developmental toxicity testing. These models have varying degrees of concordance with observed human outcomes, having approximately 70% concordance to known human developmental toxicants.[1] The devTOX^{qP} assay was 80% and 85% concordant with *in vivo* rodent (N=65) and rabbit (N=53) EFD studies, respectively. Forty-eight of the pharmaceuticals that have been tested in the iPS cell-based devTOX^{qP} assay have sufficient evidence to classify them as developmentally toxic or as safe in humans (based on epidemiological studies or known mechanisms of developmental toxicity in humans). In this set of compounds, devTOX^{qP} was more accurate than rodents or rabbits at predicting human response (Table 2).

Table 2: Rodent, Rabbit and devTOX^{qP} Accuracy for Pharmaceuticals with Known (or Expected) Human Response

	devTOX^{qP}	Rodents	Rabbits
N (#DT, #Non-DT)	48 (28, 20)	48 (28, 20)	40 (25, 15)
Balanced Accuracy	94%	86%	87%
Sensitivity	93%	93%	88%
Specificity	95%	80%	87%

The pharmaceuticals used for these calculations are the 48 drugs with data in the “Humans” column of Appendix 1.

Previous Qualification Interactions

1) Stemina had a face-to-face Pre-LOI meeting with members of the Biomarker Qualification Program team on August 28, 2019.
2) The LOI was originally submitted on July 22, 2020 (DDT BMQ000109). The FDA notified us that the submission was non-reviewable on December 16, 2020. Stemina met with the Biomarker Qualification Program team to discuss this decision on February 1, 2021. Our current submission contains an updated LOI that addresses the FDA's comments in the Reviewability Memorandum.

Attachments

1) List of Relevant Publications

a) Palmer JA, Smith AM, Egnash LA, Conard KR, West PR, Burrier RE, Donley EL, Kirchner FR. Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. Birth Defects Res B Dev Reprod Toxicol. 2013;98(4):343-63. doi: 10.1002/bdrb.21078.

b) Palmer JA, Burrier RE, Egnash LA, Donley EL. Predictive Toxicology: From Vision to Reality. Pfannkuch F, Suter-Dick L, editors. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2014. Chapter 14, Stem Cell-Based Methods for Identifying Developmental Toxicity Potential; p.321-346. doi: 10.1002/9783527674183.ch14.

c) Palmer JA, Smith AM, Egnash LA, Colwell MR, Donley ELR, Kirchner FR, Burrier RE. A human induced pluripotent stem cell-based *in vitro* assay predicts developmental toxicity through a retinoic acid receptor-mediated pathway for a series of related retinoid analogues. Reprod Toxicol. 2017;73:350-361. doi: 10.1016/j.reprotox.2017.07.011.

d) Zurlinden TJ, Saili KS, Rush, N, Kothiya P, Judson RS, Houck KA, Hunter ES, Baker NC, Palmer JA, Thomas RS, Knudsen TB. Profiling the ToxCast library with a pluripotent human (H9) stem cell line-based biomarker assay for developmental toxicity. Toxicol Sci. 2020;174(2):189-209. doi: 10.1093/toxsci/kfaa014.

e) Simms L, Rudd K, Palmer J, Czekala L, Yu F, Chapman F, Trelles Sticken E, Wieczorek R, Bode LM, Stevenson M, et al. 2020. The use of human induced pluripotent stem cells to screen for developmental toxicity potential indicates reduced potential for non-combusted products, when compared to cigarettes. Curr Res Toxicol. 1:161–173. doi:10.1016/j.crtox.2020.11.001.

2) Other Supporting Information *(<u>Optional</u> – not for public posting)*

a) Appendix 1. List of pharmaceuticals that have been tested with devTOX^{qP} with published *in vivo* outcomes and devTOX^{qP} prediction.

b) Appendix 2. Chang X, Lumen A, Palmer J, Lee UJ, Ceger P, Donley B, Bell S, Allen D, Casey W, Kleinstreuer NC. *In Vitro* to *e* Extrapolation for Developmental Toxicity Potency of Valproic Acid Analogues. Poster presented at: 60th Annual Meeting of the Society for Birth Defects Research and Prevention; June 25-July 2, 2020; Virtual Meeting.

References

1. Daston GP, Knudsen TB. Fundamental concepts, current regulatory design and interpretation. In: Daston G, Knudsen T, eds. *Comprehensive Toxicology, Vol. 12*. 2nd ed. New York: Elsevier; 2010:3-9.
2. Basketter DA, Clewell H, Kimber I, et al. A roadmap for the development of alternative (non-animal) methods for systemic toxicity testing. *ALTEX*. 2012;29(1):3-91. http://kops.ub.uni-konstanz.de/handle/id-15/browse?offset=80&type=ddc&value=570. Accessed October 10, 2013.
3. Braakhuis HM, Theunissen PT, Slob W, Rorije E, Piersma AH. Testing developmental toxicity in a second species: are the differences due to species or replication error? *Regul Toxicol Pharmacol*. 2019;107(June):104410. doi:10.1016/j.yrtph.2019.104410
4. Avila AM, Bebenek I, Bonzo JA, et al. An FDA/CDER perspective on nonclinical testing strategies: Classical

toxicology approaches and new approach methodologies (NAMs). *Regul Toxicol Pharmacol*. 2020;114(January):104662. doi:10.1016/j.yrtph.2020.104662

5. Liu W, Deng Y, Liu Y, Gong W, Deng W. Stem cell models for drug discovery and toxicology studies. *J Biochem Mol Toxicol*. 2013;27(1):17-27. doi:10.1002/jbt
6. Palmer JA, Smith AM, Egnash LA, et al. Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. *Birth Defects Res B Dev Reprod Toxicol*. 2013;98(4):343-363. doi:10.1002/bdrb.21078
7. Talbert DR, Doherty KR, Trusk PB, Moran DM, Shell SA, Bacus S. A multi-parameter in vitro screen in human stem cell-derived cardiomyocytes identifies ponatinib-induced structural and functional cardiac toxicity. *Toxicol Sci*. 2015;143(1):147-155. doi:10.1093/toxsci/kfu215
8. Bell SM, Phillips J, Sedykh A, et al. An integrated chemical environment to support 21st-century toxicology. *Environ Health Perspect*. 2017;125(5):054501. doi:10.1289/EHP1759
9. U. S. Food and Drug Administration. Center for Drug Evaluation and Research. *Bioanalytical Method Validation*.; 2018. https://www.fda.gov/files/drugs/published/Bioanalytical-Method-Validation-Guidance-for-Industry.pdf.
10. Clinical and Laboratory Standards Institute. *Liquid Chromatography-Mass Spectrometry Methods; Approved Guideline*.; 2014.
11. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *S5(R3): Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals*.; 2020. https://www.ich.org/page/safety-guidelines.
12. U. S. Food and Drug Administration. Center for Drug Evaluation and Research. *Oncology Pharmaceuticals: Reproductive Toxicity Testing and Labeling Recommendations Guidance for Industry*.; 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/oncology-pharmaceuticals-reproductive-toxicity-testing-and-labeling-recommendations-guidance.
13. National Research Council. *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press; 2007.
14. Wobus AM, Löser P. Present state and future perspectives of using pluripotent stem cells in toxicology research. *Arch Toxicol*. 2011;85(2):79-117. doi:10.1007/s00204-010-0641-6
15. Zurlinden TJ, Saili KS, Rush N, et al. Profiling the ToxCast Library with a Pluripotent Human (H9) Stem Cell Line-Based Biomarker Assay for Developmental Toxicity. *Toxicol Sci*. 2020;174(2):189-209. doi:10.1093/toxsci/kfaa014
16. Beck MJ, McLellan C, Lightle RL, Philbert MA, Harris C. Spatial glutathione and cysteine distribution and chemical modulation in the early organogenesis-stage rat conceptus in utero. *Toxicol Sci*. 2001;62(1):92-102. doi:10.1093/TOXSCI/62.1.92
17. Harris C, Jilek JL, Sant KE, Pohl J, Reed M, Hansen JM. Amino acid starvation induced by protease inhibition produces differential alterations in redox status and the thiol proteome in organogenesis-stage rat embryos and visceral yolk sacs. *J Nutr Biochem*. 2015;26(12):1589-1598. doi:10.1016/j.jnutbio.2015.07.026
18. Hansen JM. Oxidative stress as a mechanism of teratogenesis. *Birth defects Res Part C, Embryo today Rev*. 2006;78(4):293-307. doi:10.1002/bdrc.20085
19. Hansen JM, Harris C. Redox control of teratogenesis. *Reprod Toxicol*. 2013;35:165-179. doi:10.1016/j.reprotox.2012.09.004
20. Ito K, Suda T. Metabolic requirements for the maintenance of self-renewing stem cells. *Nat Rev Mol Cell Biol*. 2014;15(4):243-256. doi:10.1038/nrm3772
21. Kovacic P, Somanathan R. Mechanism of teratogenesis: electron transfer, reactive oxygen species, and antioxidants. *Birth defects Res Part C, Embryo today Rev*. 2006;78(4):308-325.
22. Mao Y, Gurr JA, Hickok NJ. Retinoic acid regulates ornithine decarboxylase gene expression at the transcriptional level. *Biochem J*. 1993;295(3):641-644. doi:10.1042/bj2950641
23. Palmer JA, Smith AM, Egnash LA, et al. A human induced pluripotent stem cell-based in vitro assay predicts developmental toxicity through a retinoic acid receptor-mediated pathway for a series of related retinoid analogues. *Reproductive Toxicology*. http://www.sciencedirect.com/science/article/pii/S0890623817301326. Published July 23, 2017. Accessed October 12, 2017.