REVIEW ARTICLE

Toward better assessments of developmental toxicity using stem cell-based in vitro embryogenesis models

Yusuke Marikawa

Department of Anatomy, Biochemistry and Physiology, Institute for Biogenesis Research, University of Hawaii John A. Burns School of Medicine, Honolulu, Hawaii, USA

Correspondence
Yusuke Marikawa, Department of Anatomy, Biochemistry and Physiology, Institute for Biogenesis Research, University of Hawaii John A. Burns School of Medicine, 651 Ilalo Street, Biosciences Building 163A, Honolulu, HI 96813, USA.
Email: marikawa@hawaii.edu

Abstract

In the past few decades, pluripotent stem cells have been explored as non-animal alternatives to assess the developmental toxicity of chemicals. To date, numerous versions of stem cell-based assays have been reported that are allegedly effective. Nonetheless, none of the assays has become the gold standard in developmental toxicity assessment. Why? This article discusses several issues in the hope of facilitating the refinement of stem cell assays and their acceptance as the cornerstone in predictive developmental toxicology. Each stem cell assay is built on a limited representation of embryogenesis, so that multiple assays are needed to detect the diverse effects of various chemicals. To validate and compare the strengths and weaknesses of individual assays, standardized lists of reference chemicals should be established. Reference lists should consist of exposures defined by toxicokinetic data, namely maternal plasma concentrations that cause embryonic death or malformations, and also by the effects on the molecular machineries that control embryogenesis. Although not entirely replacing human or animal tests, carefully selected stem cell assays should serve as practical and ethical alternatives to proactively identify chemical exposures that disturb embryogenesis. To achieve this goal, unprecedented levels of coordination and conviction are required among research and regulatory communities.

KEYWORDS
adverse outcome pathway, Daston list, embryo, embryotoxicity, gastruloid, risk assessment, teratogen

1 | INTRODUCTION

Embryogenesis can be disturbed by various chemicals to which the mother is exposed during or before pregnancy. Disturbance during embryogenesis may result in adverse outcomes, namely embryonic death and malformations. To date, a wide range of chemical exposures that cause developmental toxicity have been discovered through human and animal studies. Nonetheless, there are numerous other chemicals, either pharmaceuticals or nonpharmaceuticals, whose developmental toxicity is still unclear. It is a challenging task to determine whether a particular chemical can disturb embryogenesis. Human epidemiologic studies require many cases of miscarriages and birth defects to pinpoint a specific chemical exposure as the culprit. The use of laboratory animals, such as rats

and rabbits, allows well-controlled experiments to assess the effects of chemical exposures during pregnancy. However, high cost, interspecies differences, and animal welfare issues are some of the major drawbacks of animal-based assays. To ease the burden of human and animal studies and also to facilitate the developmental toxicity screening of many chemicals, nonanimal alternative approaches have been sought out in the past few decades, particularly in vitro assays using pluripotent stem cells.

Pluripotent stem cells are derived from the epiblast, a pluripotent tissue in the pre- and peri-implantation stage embryo, or from somatic cells that have been reprogrammed with the Yamanaka factors (Martello & Smith, 2014; Takahashi & Yamanaka, 2016; Figure 1). The former is commonly referred to as embryonic stem (ES) cells, while the latter as induced pluripotent stem (iPS) cells. ES and iPS cells are maintained in the culture dish as undifferentiated cell populations, while retaining the property of pluripotency, that is, the ability to give rise to derivatives of all three germ layers, namely, ectoderm (e.g., neuron), mesoderm (e.g., cardiomyocyte), and endoderm (e.g., hepatocyte). Pluripotent stem cells can be induced under special culture conditions to differentiate into various types of cells, including neurons and cardiomyocytes (Figure 1). The basic idea behind in vitro assays using pluripotent stem cells is to define the inhibitory impact of a chemical exposure on the differentiation as a sign of developmental toxicity. The first developmental toxicity assay, reported as the embryonic stem cell test (EST), employed the differentiation of contracting cardiomyocytes from mouse ES cells as an in vitro

embryogenesis model to test the impact of chemicals (Laschinski, Vogel, & Spielmann, 1991; Riebeling et al., 2012; Seiler, Buesen, Visan, & Spielmann, 2006). Since then, many versions of developmental toxicity assays have been reported, which employed the differentiation of ES and iPS cells of mouse and human into various cell types, including neurons and osteoblasts (Kim, Che, & Yun, 2019; Luz & Tokar, 2018; Mennen, Oldenburger, & Piersma, 2022). These assays exploit a variety of endpoint analyses using morphological and molecular techniques to measure the adverse impact on cell differentiation. Such alterations and diversifications of the methodologies are intended to improve the sensitivity and specificity to detect various types of chemicals, and also to streamline the assay procedure to enable high-throughput analyses of numerous chemicals in a short period of time.

In recent years, pluripotent stem cells are also used to generate "organoids" to recapitulate the more complex three-dimensional (3D) architectures of tissues and organs in vitro. Organoids are comprised of multiple cell types that are arranged in a spatially organized manner, mimicking the morphology and function of the corresponding organs, such as the intestine, kidney, and brain (Lancaster & Knoblich, 2014; Sasai, 2013). The stem cell-derived organoids have been explored as in vitro models of the organs in toxicology research (Matsui & Shinozawa, 2021; Wang et al., 2021). "Gastruloids" refer to the 3D aggregates of pluripotent stem cells that mimic the key aspects of gastrulation, namely axial patterning and elongation morphogenesis (Baillie-Benson, Moris, &

FIGURE 1 Schematic diagram depicting the origins of pluripotent stem cells: one from the epiblast of normal embryos and the other from somatic cells that are reprogrammed with the Yamanaka factors. In vivo, the epiblast gives rise to the entire fetal body. In vitro, pluripotent stem cells can be induced to differentiate into various cell types, or to form 3D cell aggregates, known as gastruloids, which undergo axial patterning and elongation morphogenesis. In vitro development of pluripotent stem cells is explored as a model of embryogenesis to assess the developmental toxicity of various chemicals
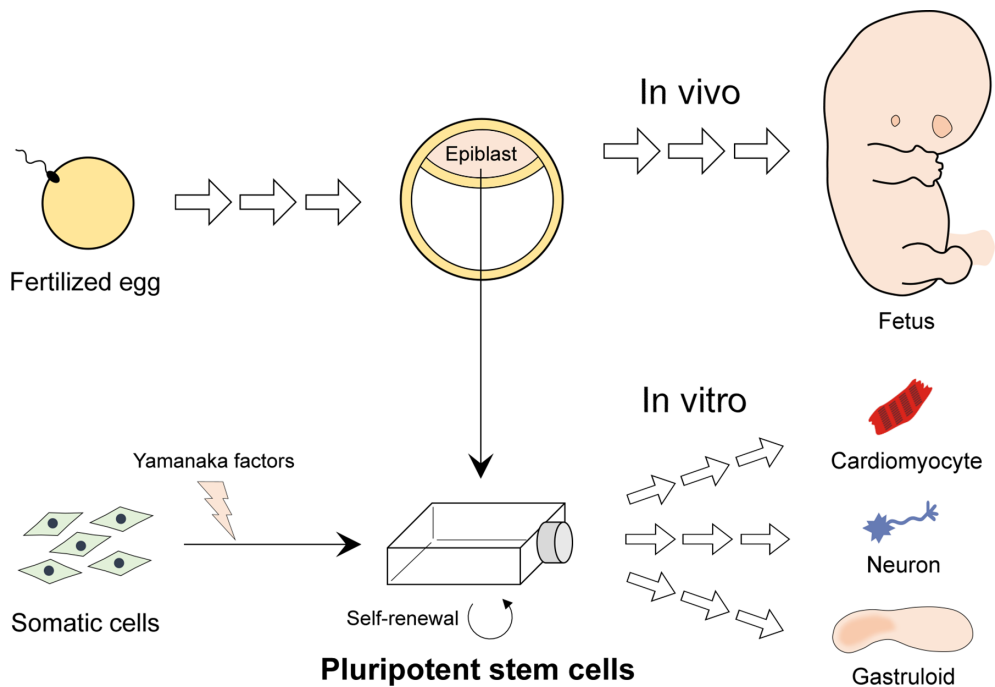
Martinez Arias, 2020; van den Brink & van Oudenaarden, 2021; Figure 1). Because gastrulation is the critical embryological process to construct the basic body plan, gastruloids have been used as in vitro embryogenesis models to test the developmental toxicity of chemical exposures (Mantziou et al., 2021; Marikawa, Chen, Menor, Deng, & Alarcon, 2020; Warkus & Marikawa, 2017; Warkus, Yuen, Lau, & Marikawa, 2016). Whether based on the differentiation of a single cell type or the morphogenesis of gastruloids, most of the reported assays are allegedly effective in detecting various chemical exposures that are known to be developmentally toxic. Therefore, the use of pluripotent stem cells for predictive developmental toxicity screening appears to be warranted.

Nonetheless, assays using pluripotent stem cells have not yet become the gold standard for developmental toxicity screening. No regulatory agency mandates the use of stem cell-based tests to assess the adverse impact of new chemicals on embryogenesis. Although the recent International Council for Harmonization (ICH) Guideline on Detection of Toxicity to Reproduction for Human Pharmaceuticals recommends the use of in vitro assays for developmental toxicity assessment (ICH, 2020), it does not endorse any specific method, including those employing pluripotent stem cells. Does this mean that none of the stem cell-based assays is satisfactory? If so, what are their deficiencies? Among those reported assays, which ones are better than the others? Can they even be compared with each other? Currently, there is no clear consensus among research communities on how to measure the efficacy or accuracy of individual developmental toxicity assays in a comparable manner. Thus, it is difficult to evaluate objectively which versions of assays are more effective than others. This situation discourages stakeholders, including regulatory agencies and chemical industries, to adopt any specific stem cell-based assays as part of the decision-making process in developmental toxicity assessment. Until these issues are resolved, pluripotent stem cells may not become the cornerstone in our collective efforts to identify chemicals that are harmful to embryos.

In this article, several key issues are discussed regarding developmental toxicity assays using pluripotent stem cells, including their need for toxicity screening in addition to other types of alternative assays, strategies to validate and select stem cell-based assays, and limitations that are inherent to stem cell-based assays. It is my sincere hope that these discussions would help improve and invigorate stem cell-based assays to be widely accepted as the gold standard in predictive risk assessment of chemical exposures for developmental toxicity. Note that in this article, the term "developmental toxicity" is mainly used to describe any insult that disturbs embryogenesis to cause death or malformations. It broadly corresponds to "teratogenicity" or "embryotoxicity," which are also used in many literatures to describe insults on embryogenesis. Also, the term "stem cells" used in this article specifically refers to "pluripotent stem cells" for convenience, and does not include tissue- or organ-specific stem cells that do not possess the pluripotency.

2 | THE NEED FOR STEM CELL MODELS

Although pluripotent stem cells have the potential to become any cell type, their in vitro development recapitulates only a small fraction of the entire embryogenesis. This is certainly the case for various stem cell models that have been used for developmental toxicity assays, including cardiomyocyte differentiation and gastruloid morphogenesis. If stem cell models are such incomplete representations of the actual embryo, why are they needed for developmental toxicity assessment? Generally, "models" are used in various fields of biomedical research, including developmental toxicology, for two reasons. One is due to ethical issues. The conduct of experimentations involving human subjects poses a variety of ethical concerns. Clinical trials of pharmaceuticals usually exclude women that are pregnant or are planning to conceive, due to the potential adverse impact on unborn children. Even experimentations with human embryos obtained from the surplus of in vitro fertilization procedures or from elective abortions are highly controversial (Lee, Feeney, Schmainda, Sherley, & Prentice, 2020; Wertz, 2002). Thus, laboratory animals are typically used for experimentations to assess the developmental toxicity of chemicals. However, the sacrifice of a large number of animals for toxicity testing is also a contentious ethical issue. There are great interest and demand among the research and regulatory communities in moving toward nonanimal platforms for toxicity assessment. To circumvent such ethical issues associated with human- and animal-based tests, pluripotent stem cells may be needed as models of embryogenesis even though they are incomplete representations.

The other reason to employ models is for practicality. Embryogenesis is an extremely complex process. The fertilized egg gives rise to diverse arrays of tissues and organs in a spatially and temporally dynamic manner. Mammalian embryos, in particular, develop inside of the mother with physical and physiological association through the placenta. Such complexity often hinders mechanistic investigations into how a chemical exposure disturbs embryogenesis. Stem cell-based models represent

specific structures of the embryo at specific developmental stages. There is no maternal environment that may significantly alter the property and amount of chemicals. Due to the in vitro nature, stem cell models are more accessible to a plethora of experimental interrogations in a controlled fashion. Furthermore, assays using pluripotent stem cells are generally faster, cheaper, less labor-intensive, and more scalable for high-throughput analyses than animal-based assays. For these practical reasons, stem cell assays may be needed to obtain the information on developmental toxicity for a large number of chemicals in a speedy manner, and also to elucidate the molecular mechanisms underlying the developmental toxicity of chemical exposures.

Other nonmammalian organisms are also explored as alternatives in developmental toxicity assessment, such as fish, frog, nematode, and slime mold (Baines, Wolton, & Thompson, 2021; Beekhuijzen et al., 2015; Boyd, Smith, & Freedman, 2012; Song et al., 2004). These whole organism systems are likely to bring about valuable insights into the adverse actions of various chemical exposures. Nonetheless, they are evolutionarily distant from mammals, and their gene sequences (i.e., protein primary structures) are considerably different, which may influence the susceptibility to some chemical exposures. Also, mammals have adopted the viviparous mode of reproduction, in which embryogenesis depends on the nutrients and other chemical building blocks that are provided through the mother. By contrast, for non-mammalian organisms, those essential chemicals are already stored inside of the oocyte, and are sufficient to support the entire embryogenesis. Such fundamental disparities in the mode of chemical usage may contribute to the significant differences in the impact of chemical exposures. For these reasons, assays involving pluripotent stem cells are needed to reflect the unique nature of mammalian embryogenesis.

3 | VALIDATION OF STEM CELL MODELS

Diverse versions of assays have been reported, using different stem cell lines that are differentiated into different cell types or structures. The impact of chemical exposures is also measured differently using various methods, involving specific morphological or molecular features that are linked to the differentiation. Every single one of these assays appears to exhibit fairly good performance according to the validation criteria employed in each study. The question is, can these assays be ranked according to the reported concordance rates (%) to determine which assay is the most effective? The answer is no,

mainly because most of these assays were validated differently, that is, using different combinations of reference chemicals with different criteria to define adverse effects. To enable fair comparisons between the different assays, they need to be validated using a common set of reference chemicals. Such a standardized method of validation should reveal the strengths and weaknesses of individual assays in a comparable manner, and help identify which assays are more suited (or not suited) to detect certain types of developmental toxicity.

However, the choosing of reference chemicals for assay validation has been a highly challenging task with a long history of struggles (Brown, 2002; Daston et al., 2010; Marx-Stoelting et al., 2009; Smith et al., 1983; Webster, Brown-Woodman, & Ritchie,1997). It requires understanding of the principles of teratology (developmental toxicology) as well as a tremendous level of cooperation and coordination among researchers and regulatory bodies. Below, some of the key issues concerning reference chemicals are discussed to help navigate our concerted efforts to establish effective validation strategies.

3.1 | Exposure-based validation

"The dose makes the poison" is perhaps the most important dogma in toxicology, often credited to Paracelsus (Grandjean, 2016). Essentially, any chemical can be toxic at a high enough concentration, but it can also be safe at a low enough concentration. This is certainly the case for developmental toxicity as well. In many studies, assay validations were performed using reference chemicals that are categorized as either "nonembryotoxic" or "embryotoxic," the latter of which may be further divided into multiple groups, such as "weakly embryotoxic" and "strongly embryotoxic" (Genschow et al., 2002; Marx-Stoelting et al., 2009). However, even those chemicals that are categorized as strongly embryotoxic can be safe when concentrations are low enough. Thus, validation of assays using reference chemicals should be performed with great attention to their concentrations.

Daston and colleagues have pointed out this matter most eloquently (Daston et al., 2010), and generated a list of reference "exposures," each of which is a specific concentration of a specific chemical, designated either as developmentally toxic or nontoxic exposure (Daston et al., 2014). These exposures are selected based on in vivo rat studies, where adequate toxicokinetic data are available, namely maternal plasma maximum concentration C_{max} and the incidence of embryonic death or malformations. The list, often dubbed as the Daston list, may be used to validate the effectiveness of stem cell-based

(a)

Cell line: Mouse P19C5 (embryonal carcinoma) stem cell
Culture format: 3D cell aggregation (hanging drop)
Culture period: 4 days
Exposure period: 4 days
Endpoint analysis: Morphology (size, shape) of aggregates

Day 1 Day 2 Day 3 Day 4
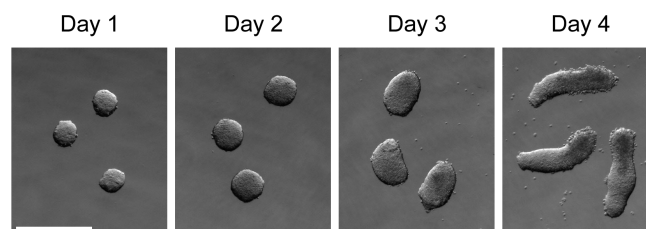
(b)

Developmentally toxic exposure		Nontoxic exposure	
Chemical (concentration)	Adverse effect	Chemical (concentration)	Adverse effect
Abacavir (80 μM)	O	Abacavir (18 μM)	X
Acetazolamide (121 μM)	O	All-*trans* retinoic acid (1.7 nM)	X
All-*trans* retinoic acid (200 nM)	O	Butylparaben (110 μM)	O
Artesunate (20 nM)	X	Caffeine (7.7 μM)	X
Caffeine (325 μM)	O	Dabigatran (1 μM)	X
Dabigatran (7 μM)	O	Desloratadine (1.5 μM)	X
Ethylene glycol (57 mM)	O	Ethylene glycol (1.4 mM)	X
Fingolimod (67 nM)	X	Glycolic acid (275 μM)	X
Glycolic acid (5 mM)	O	Methanol (22 μM)	X
Hydroxyurea (350 μM)	O	Mono-2-ethylhexyl phthalate (1 μM)	X
Methanol (270 mM)	X	Nilotinib (2 μM)	O
Methoxyacetic acid (5 mM)	O	Oseltamivir (12 μM)	X
Methylmercury (5 μM)	O	Propylene glycol (850 mM)	O
Mono-2-ethylhexyl phthalate (146 μM)	O	Ramelteon (19 nM)	X
Nilotinib (28 μM)	O	Saccharin (24 μM)	X
Ramelteon (81 μM)	O	Zaleplon (12 μM)	X
Salicylic acid (3 mM)	O	Zidovudine (227 μM)	X
Valproate (0.8 mM)	O		

(c)

| Molecular effect	Chemical (concentration)	Adverse effect
Wnt signaling inhibition	XAV939 ($\geq$ 0.5 μM)	O
	IWP2 ($\geq$ 0.5 μM)	O
Nodal signaling inhibition	SB431542 ($\geq$ 5 μM)	O
Bmp signaling inhibition	DMH1 ($\geq$ 0.2 μM)	O
Fgf signaling inhibition	PD173074 ($\geq$ 0.1 μM)	O
RA signaling inhibition	BMS493 ($\geq$ 40 nM)	O
Notch signaling inhibition	DAPT (5 μM)	X
HH signaling inhibition	Cyclopamine (5 μM)	X
Wnt signaling activation	CHIR99021 ($\geq$ 1 μM)	O
RA signaling activation	Acitretin ($\geq$ 0.1 μM)	O
PCP pathway inhibition	Y-27632 (10 μM)	O
Mevalonate pathway inhibition	Lovastatin ($\geq$ 1 μM)	O
HDAC inhibition	Trichostatin A ($\geq$ 4 nM)	O

Control

IWP2 (8 μM)
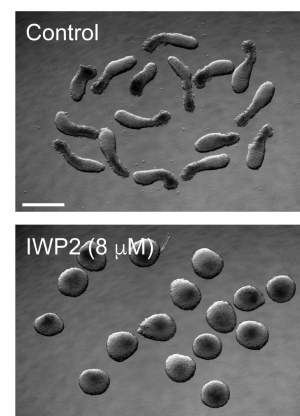
FIGURE 2 Developmental toxicity assay using gastruloids of mouse P19C5 stem cells. (a) Basic characteristics of the P19C5 gastruloid-based assay. Right images show the morphological changes of 3D cell aggregates of P19C5 stem cells over the course of 4 days of hanging drop culture. Aggregates have been removed from hanging drops for photography. Scale bar = 500 μm. (b) Exposure-based validation of the P19C5 gastruloid assay using the Daston list (Daston et al., 2014). Summarized results from the previous studies (Li & Marikawa, 2016; Warkus & Marikawa, 2017) are shown. For adverse effects, circles represent significant morphological impact (i.e., reduced size or altered shape in Day 4 aggregates), and crosses represent no significant morphological impact. (c) Mechanism-based validation of the P19C5 gastruloid assay. Summarized results from the previous studies (Li & Marikawa, 2015, 2016; Warkus, Yuen, Lau, & Marikawa, 2016; Warkus & Marikawa, 2018) are shown. RA (retinoic acid), HH (hedgehog), PCP (planar cell polarity), HDAC (histone deacetylase). Right images show Wnt signaling-inhibited Day 4 aggregates (treated with IWP2 at 8 µM) and the corresponding control (untreated) aggregates. Note that axial elongation is markedly diminished in IWP2-treated aggregates. Scale bar = 500 µm

assays. An effective assay is expected to show an adverse reaction in response to a chemical at the developmentally toxic concentration but not at the nontoxic concentration. Such exposure-based validation is in line with the principles of toxicology, and allows the evaluation of the sensitivity and specificity of individual assays in a manner relevant to "the real world of developmental toxicity" (Daston et al., 2014). Thus far, unfortunately, only a few assays have been evaluated using the Daston list (Cassar et al., 2019; Marikawa, Chen, Menor, Deng, & Alarcon, 2020; Warkus & Marikawa, 2017).

The results of one of the assays that have been evaluated with the Daston list are summarized in Figure 2. The assay is based on the growth and morphogenesis of gastruloids that are made of 3D aggregates of mouse P19C5 stem cells (Lau & Marikawa, 2014). P19C5 cell aggregates grow and elongate with a distinct anterior–posterior axis during 4 days of culture, recapitulating the morphogenetic process of gastrulation (Figure 2a). In this assay, cell aggregates are treated with test chemicals at various concentrations throughout the entire culture period, followed by morphometric measurement of the size and shape of individual aggregates. A significant reduction in size or a significant alteration in shape in comparison to the corresponding control (untreated) gastruloids is defined as the adverse impact of a chemical exposure. Overall, 29 out of the 35 exposures tested (82.9%) are found to affect P19C5 gastruloids in a manner consistent with the in vivo developmental toxicity data (Li & Marikawa, 2016; Warkus & Marikawa, 2017; Figure 2b). However, this also indicates that the assay is not effective for six specific exposures. Namely, three developmentally toxic exposures (artesunate at 20 nM, fingolimod at 67 nM, and methanol at 270 mM) are "false negative," whereas three nontoxic exposures (butylparaben at 110 μM, nilotinib at 2 μM, and propylene glycol at 850 mM) are "false positive." There may be specific reasons why the assay is insensitive or over-sensitive to these six exposures, as discussed previously (Warkus & Marikawa, 2017). Regardless, this exposure-based validation has provided critical information on the sensitivity and specificity of the P19C5 gastruloid assay, which may be compared with other assays (once evaluated with the Daston list) to underscore their strengths and weaknesses.

The Daston list can be expanded further to validate developmental toxicity assays more extensively. The current list is restricted to exposures that are based on rat studies. This is because rats are the most commonly used species in developmental toxicity testing. However, as the major goal is to predict human developmental toxicity, it is ideal to include chemical exposures that are known to disturb human embryogenesis. Even though human toxicokinetic studies are limited, there may be sufficient

information on the plasma concentrations and developmental toxicity for certain pharmaceutical drugs through clinical and epidemiologic studies. The recent ICH Guideline lists a series of reference chemicals to be used for alternative (in vitro, ex vivo, and nonmammalian in vivo) assays (ICH, 2020). Some of the chemicals are listed as "human teratogens" and are provided with the information on human plasma C_{max}. The Guideline also provides the no-observed-adverse-effect level (NOAEL) C_{max} and/or the lowest-observed-adverse-effect level (LOAEL) C_{max} for several chemicals based on mouse, rat, or rabbit studies. These reference chemicals with specific concentrations, including the human teratogens, may be combined with the Daston list to augment the exposure-based validation.

3.2 | Mechanism-based validation

Embryogenesis is characterized by the highly elaborate regulation of cellular behaviors, such as proliferation, differentiation, and migration, which create various tissues and organs in the right place at the right time. These cellular behaviors are genetically controlled and are dependent on the activities of distinct molecular machineries, including signal transduction pathways and transcription factors. In the past several decades, experimental studies using the laboratory mouse as well as other nonmammalian organisms have greatly expanded our knowledge of the molecular machineries that are essential for embryogenesis. In principle, any of these machineries can potentially be targeted by chemical exposures to cause embryonic death or malformations. Thus, another way to validate the effectiveness of developmental toxicity assays is to examine whether they are capable of detecting disturbances in the molecular machineries essential for embryogenesis.

Several core signaling pathways, namely Wnt, Nodal, Bmp, Fgf, Notch, Hedgehog, and Retinoic acid (RA), operate at multiple places and times in the embryo to control distinct cellular behaviors. For example, RA signaling regulates the patterning and morphogenesis of the neural tube, paraxial mesoderm, pharyngeal arches, facial prominences, limbs, heart, and vascular system (Knudsen, Pierro, & Baker, 2021; Piersma, Hessel, & Staal, 2017). Accordingly, a disturbance in RA signaling during embryogenesis, through its inhibition or excessive activation, can cause abnormalities in these structures. Figure 2c summarizes the performance of the P19C5 gastruloid assay in response to the pharmacological inhibitors of the core signaling pathways (Li & Marikawa, 2015). The size and shape of gastruloids are significantly altered by the inhibitors of the Wnt, Nodal, Bmp, Fgf, and RA signaling pathways, indicating that the

gastruloid morphogenesis is controlled by these molecular machineries. However, inhibition of Notch or Hedgehog signaling does not cause significant changes in the morphology of gastruloids. Notch signaling is active during gastruloid development, and the pharmacological inhibitor reduces the expression levels of the Notch signaling target genes (Li & Marikawa, 2015). Nonetheless, the growth and elongation of gastruloids appear to be independent of Notch signaling. By contrast, Hedgehog signaling is apparently inactive during gastruloid development, and the effect of the inhibitor (cyclopamine) is undetectable by either morphological or molecular analyses (Li & Marikawa, 2015). Overall, the mechanism-based validation has provided mechanistic insights into the capability of the gastruloid assay. Namely, it can detect chemical exposures that interfere with Wnt, Nodal, Bmp, Fgf, and RA signaling, but not those that inhibit Notch or Hedgehog signaling.

Clearly, embryogenesis is regulated by many more molecular machineries than the seven core signaling pathways mentioned above. Figure 2c also shows the impact of a few additional chemical exposures on the P19C5 gastruloid, which have different mechanisms of actions (Li & Marikawa, 2016; Warkus & Marikawa, 2018). Examination of more machineries should further illuminate on the capability and limitation of the assay. The mechanism-based validation described here is akin to the adverse outcome pathway (AOP) framework, which uses the available mechanistic information on a toxicological response to describe linkages between molecular events and adverse outcomes (Draskau, Spiller, Boberg, Bowles, & Svingen, 2020; Rogers, 2020). The AOP- or mechanism-based approach integrates the information from developmental biology and pharmacology, and should help transform developmental toxicology from an empirical to a more predictive science. One of the skeptical opinions over stem cell-based assays is that their representations of embryogenesis (e.g., cardiomyocyte differentiation or gastrulation) are too narrow to detect a wide range of developmentally toxic chemicals. However, various processes of embryogenesis are controlled by finite sets of molecular machineries, such as the core signaling pathways. Thus, a stem cell model that recapitulates only a limited process of embryogenesis may be able to detect chemical exposures that affect other embryological processes that are not directly represented by the model.

3.3 | Selection of assays

Once evaluated using standardized reference lists, different assays can be objectively compared with each other to reveal their strengths and weaknesses. Only through such fair comparison that the most useful assays may be selected for developmental toxicity screening. Importantly, assays with the highest concordance rates (%) may not necessarily be the most useful ones. This notion is exemplified in the following hypothetical scenario. For 10 toxic reference chemical exposures, assays A, B, and C exhibit the detection capability of 80% (8/10), 60% (6/10), and 40% (4/10), respectively (Figure 3). Intuitively, assay A with 80% concordance may seem to be most useful. However, when looked at closely, a combination of assays B and C yields 100% concordance, which is better than assay A alone or its combination with B or C.

A single type of assay, which is based on a limited representation of embryogenesis, would not be able to detect a whole range of chemicals with various mechanisms of actions that target different embryological processes. However, carefully selected multiple types of assays may be able to detect much more diverse chemicals by complementing the weaknesses of each other. Such selection is only possible when individual assays are evaluated using standardized reference lists. While actual selection must await the results of validations, it may probably require a combination of assays that involve different stem cell lines of different species (e.g., mouse and human), different representations of embryogenesis (e.g., cell differentiation and morphogenesis), and different endpoints of analyses (e.g., morphology and gene expression profile).

Certain practical aspects may also need to be taken into consideration for the selection of assays, if possible. This is because multiple types of assays are to be performed by various investigators. For example, the assay procedures should be relatively easy to conduct by reasonably trained personnel in a consistent manner. The stem cell lines used in the assays should be widely available or freely shared, and the necessary reagents and equipment should be reasonably priced and obtainable.

Reference	Assay A	Assay B	Assay C
Exposure 1	O	O	X
Exposure 2	O	O	X
Exposure 3	O	O	X
Exposure 4	O	O	X
Exposure 5	O	O	X
Exposure 6	X	O	X
Exposure 7	O	X	O
Exposure 8	O	X	O
Exposure 9	O	X	O
Exposure 10	X	X	O
Concordance	**80%**	**60%**	**40%**

FIGURE 3 Comparisons of three hypothetical assays that are validated against 10 hypothetical reference (developmentally toxic) exposures. See the text for details

4 | OVERCOMING THE LIMITATIONS OF STEM CELL MODELS

In a stem cell-based assay, the in vitro development of stem cells is used as a model of embryogenesis to assess the adverse effects of chemical exposures. However, there are crucial differences between embryogenesis in vitro and in vivo (i.e., in the mother), which may pose limitations to the application of stem cell models. Some of such limitations are described below to discuss how they may influence stem cell-based assessment and how they could be overcome.

4.1 | Differences in the surrounding environment

In vivo, the embryo is exposed to the amniotic fluid, which has distinct compositions of organic and inorganic substances that dynamically change over the course of development (Tong et al., 2009; Underwood, Gilbert, & Sherman, 2005). In addition, once the placenta and the embryonic cardiovascular system are established, substances in the maternal circulation are transported into the embryonic circulation, some are passively while others are selectively (Brett, Ferraro, Yockell-Lelievre, Gruslin, & Adamo, 2014). Thus, embryonic tissues in vivo are surrounded by materials that are significantly different in type and amount from the culture medium to which stem cells are exposed. The culture media are usually augmented with an excess amount of basic nutrients (e.g., glucose, amino acids), while they are deficient in a complex array of various macromolecules (e.g., proteins, lipids). The differences in the surrounding environment may potentially influence how a given chemical exposure exerts developmental toxicity.

Protein binding is one of such differences between the in vivo and in vitro environments that may affect the actions of chemicals. Certain chemicals, including pharmaceutical drugs, have a high binding affinity to serum proteins, namely albumin. Generally, only unbound forms can diffuse freely and contribute to their pharmacological actions (Smith, Di, & Kerns, 2010). The level of the serum protein (albumin) is typically much lower in the stem cell culture medium than in the plasma. Thus, even when the total concentration (i.e., protein-bound and unbound forms) of a test chemical is comparable between the plasma and the culture medium, the concentration of the unbound form may be higher in the latter to exert more potent effects. As a result, stem cell-based assays may misleadingly appear "over-sensitive" to certain chemicals, especially those with an extremely high

affinity to albumin. Depending on the binding property of test chemicals and the serum protein concentration in the culture medium used, the exposure-based validation of stem cell assays (discussed in Section 3.1.) may need to be calibrated accordingly.

4.2 | Maternal influences

In vivo, the embryo is under the influence of the mother, which is absent in stem cell models. Maternal influences can significantly affect whether and how chemicals disturb embryogenesis. For example, certain chemicals are known as proteratogens, which by themselves are nontoxic but become developmentally toxic after metabolic conversions by the mother, mainly in the liver (Wells & Winn, 1996). Stem cell-based assays, as well as other in vitro and ex vivo assays, may be unable to detect proteratogens, as they lack maternal metabolisms. To overcome such limitation, several attempts have been made to augment in vitro or ex vivo assays by incorporating exogenous metabolic systems, such as hepatocytes or liver extracts, to simulate the effects of the maternal liver (Hettwer et al., 2010; Luijten, Verhoef, Westerman, & Piersma, 2008; Oglesby, Ebron, Beyer, Carver, & Kavlock, 1986; Ozolins, Oglesby, Wiley, & Wells, 1995; Piersma et al., 1991; Zhao, Krafft, Terlouw, & Bechter, 1993), although it is unclear how effectively such liver substitutes can represent the complexity of maternal metabolisms. For many pharmaceutical drugs, the information on their metabolisms in the human body is often available through pharmacokinetic studies, including their chemical structures and the plasma concentrations of the major metabolites. The testing of individual metabolites using stem cell assays, which would be much easier to do with in vitro systems, may reveal that the parent compounds can act as proteratogens.

Some chemicals may cause developmental toxicity indirectly through the mother. For example, embryogenesis can be severely disturbed when chemical exposures interfere with the function of the uterus by blocking the blood supply or the endometrial maintenance. Chemicals that cause hyperthermia or hypoglycemia in the mother may also adversely impact embryogenesis indirectly (DeSesso, 1987; Lancaster, 2011; Persson & Hansson, 1993). Such chemicals may not be detectable with stem cell assays nor with any other alternative methods (in vitro, ex vivo, and nonmammalian in vivo).

5 | FUTURE PERSPECTIVES

Our collective efforts are in progress, aiming to transform stem cell-based assays into the gold standard in predictive

developmental toxicity assessment, even though they cannot entirely replace human- or animal-based studies. As discussed above, stem cell models have several inherent limitations and may not detect certain types of chemicals that cause developmental toxicity in vivo. The absence of an adverse effect in stem cell-based assays does not guarantee the safety of the chemical exposures examined. Those exposures need to undergo additional in vivo studies, including animal experimentations, to assure their safety. By contrast, with the use of well-validated and well-selected stem cell assays (as discussed in this article), the presence of an adverse effect can be regarded as a "red flag." Ideally, the data obtained from stem cell assays should be reliable enough to define adverse chemical exposures as developmentally toxic, and to regulate them accordingly. For example, no additional animal test should be required or permitted to evaluate their developmental toxicity in vivo, unless the necessity to do so is scientifically justified. For pharmaceutical drugs, they should be labeled with a pregnancy risk warning. In the case of industrial compounds, their marketing or release to the environment should be restricted by regulatory agencies.

To achieve such an ambitious goal, the imminent task is to build standardized reference lists of chemical exposures, which are essential to validate individual stem cell assays for comparison and selection. It would require enormous levels of collaboration and conviction among communities of toxicologists and developmental biologists, who can provide their expertise in gathering information that are appropriate for exposure-based and mechanism-based validations. It would also be helpful if standardized reference lists are officially endorsed by the Society for Birth Defects Research and Prevention, not only to promote, but also to mandate the usage of the reference lists for the validation of alternative developmental toxicity assays.

Even after the most effective combination of assays is selected, there may still be reluctance or skepticism among research and regulatory communities to accept stem cell-based assays as the gold standard. To build more confidence among skeptics, stronger demonstrations may be necessary to show that stem cell assays can predict the developmental toxicity of new chemicals "before" any in vivo study is performed. For example, a series of new chemicals, such as those that are synthesized during drug development as candidate compounds, may be first tested by stem cell assays for adverse effects. Compounds with adverse effects are then tested in animals to collect toxicokinetics data, namely the plasma concentrations and the incidence of embryonic death or malformations. If the in vivo data are consistent with the results of stem cell assays, then it would unequivocally demonstrate their

predictive power. The use of animals for such purposes may appear contradictory to the goal of nonanimal alternative approaches. However, once accepted as the gold standard, stem cell assays can, in the long run, bring down the animal usage significantly. While other nonanimal alternative approaches, including in silico models (Scialli et al., 2018), can also provide valuable and critical insights into the effects of chemical exposures, it is my belief that pluripotent stem cells will play the major role in predictive developmental toxicity assessment.

ORCID
Yusuke Marikawa https://orcid.org/0000-0001-5720-5806

REFERENCES
Baillie-Benson, P., Moris, N., & Martinez Arias, A. (2020). Pluripotent stem cell models of early mammalian development. *Current Opinion in Cell Biology*, 66, 89–96. https://doi.org/10.1016/j.ceb.2020.05.010

Baines, R. P., Wolton, K., & Thompson, C. R. L. (2021). Dictyostelium discoideum: An alternative nonanimal model for developmental toxicity testing. *Toxicological Sciences*, 183, 302–318. https://doi.org/10.1093/toxsci/kfab097

Beekhuijzen, M., de Koning, C., Flores-Guillén, M. E., de Vries-Buitenweg, S., Tobor-Kaplon, M., van de Waart, B., & Emmen, H. (2015). From cutting edge to guideline: A first step in harmonization of the zebrafish embryotoxicity test (ZET) by describing the most optimal test conditions and morphology scoring system. *Reproductive Toxicology*, 56, 64–76. https://doi.org/10.1016/j.reprotox.2015.06.050

Boyd, W. A., Smith, M. V., & Freedman, J. H. (2012). Caenorhabditis elegans as a model in developmental toxicology. *Methods in Molecular Biology*, 889, 15–24. https://doi.org/10.1007/978-1-61779-867-2_3

Brett, K. E., Ferraro, Z. M., Yockell-Lelievre, J., Gruslin, A., & Adamo, K. B. (2014). Maternal-fetal nutrient transport in pregnancy pathologies: The role of the placenta. *International Journal of Molecular Sciences*, 15, 16153–16185. https://doi.org/10.3390/ijms150916153

Brown, N. A. (2002). Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. European Centre for the validation of alternative methods. *Alternatives to Laboratory Animals*, 30, 177–198. https://doi.org/10.1177/026119290203000205

Cassar, S., Beekhuijzen, M., Beyer, B., Chapin, R., Dorau, M., Hoberman, A., ... Tornesi, B. (2019). A multi-institutional study benchmarking the zebrafish developmental assay for prediction of embryotoxic plasma concentrations from rat embryo-fetal development studies. *Reproductive Toxicology*, *86*, 33–44. https://doi.org/10.1016/j.reprotox.2019.02.004

Daston, G. P., Beyer, B. K., Carney, E. W., Chapin, R. E., Friedman, J. M., Piersma, A. H., ... Scialli, A. R. (2014). Exposure-based validation list for developmental toxicity screening assays. *Birth Defects Research Part B Developmental and Reproductive Toxicology*, *101*, 423–428. https://doi.org/10.1002/bdrb.21132

Daston, G. P., Chapin, R. E., Scialli, A. R., Piersma, A. H., Carney, E. W., Rogers, J. M., & Friedman, J. M. (2010). A different approach to validating screening assays for developmental toxicity. *Birth Defects Research Part B Developmental and Reproductive Toxicology*, *89*, 526–530. https://doi.org/10.1002/bdrb.20276

DeSesso, J. M. (1987). Maternal factors in developmental toxicity. *Teratogenesis Carcinogenesis and Mutagenesis*, *7*, 225–240. https://doi.org/10.1002/tcm.1770070305

Draskau, M. K., Spiller, C. M., Boberg, J., Bowles, J., & Svingen, T. (2020). Developmental biology meets toxicology: Contributing reproductive mechanisms to build adverse outcome pathways. *Molecular Human Reproduction*, *26*, 111–116. https://doi.org/10.1093/molehr/gaaa001

Genschow, E., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., ... Becker, K. (2002). The ECVAM international validation study on in vitro embryotoxicity tests: Results of the definitive phase and evaluation of prediction models. European Centre for the validation of alternative methods. *Alternatives to Laboratory Animals*, *30*, 151–176. https://doi.org/10.1177/026119290203000204

Grandjean, P. (2016). Paracelsus revisited: The dose concept in a complex world. *Basic & Clinical Pharmacology & Toxicology*, *119*, 126–132. https://doi.org/10.1111/bcpt.12622

Hettwer, M., Reis-Fernandes, M. A., Iken, M., Ott, M., Steinberg, P., & Nau, H. (2010). Metabolic activation capacity by primary hepatocytes expands the applicability of the embryonic stem cell test as alternative to experimental animal testing. *Reproductive Toxicology*, *30*, 113–120. https://doi.org/10.1016/j.reprotox.2010.01.009

The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) (2020). *S5 (R3) guideline on reproductive toxicology: Detection of toxicity to reproduction for human pharmaceuticals.* https://www.ema.europa.eu/en/ich-s5-r3-guideline-reproductive-toxicology-detection-toxicity-reproduction-human-pharmaceuticals

Kim, T. W., Che, J. H., & Yun, J. W. (2019). Use of stem cells as alternative methods to animal experimentation in predictive toxicology. *Regulatory Toxicology and Pharmacology*, *105*, 15–29. https://doi.org/10.1016/j.yrtph.2019.03.016

Knudsen, T. B., Pierro, J. D., & Baker, N. C. (2021). Retinoid signaling in skeletal development: Scoping the system for predictive toxicology. *Reproductive Toxicology*, *99*, 109–130. https://doi.org/10.1016/j.reprotox.2020.10.014

Lancaster, M. A., & Knoblich, J. A. (2014). Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science*, *345*, 1247125. https://doi.org/10.1126/science.1247125

Lancaster, P. A. (2011). Causes of birth defects: Lessons from history. *Congenital Anomalies*, *51*, 2–5. https://doi.org/10.1111/j.1741-4520.2010.00311.x

Laschinski, G., Vogel, R., & Spielmann, H. (1991). Cytotoxicity test using blastocyst-derived euploid embryonal stem cells: A new approach to in vitro teratogenesis screening. *Reproductive Toxicology*, *5*, 57–64. https://doi.org/10.1016/0890-6238(91)90111-r

Lau, C. G., & Marikawa, Y. (2014). Morphology-based mammalian stem cell tests reveal potential developmental toxicity of donepezil. *Molecular Reproduction and Development*, *81*, 994–1008. https://doi.org/10.1002/mrd.22423

Lee, T. S., Feeney, M. B., Schmainda, K. M., Sherley, J. L., & Prentice, D. A. (2020). Human fetal tissue from elective abortions in research and medicine: Science, ethics, and the law. *Issues in law & Medicine*, *35*, 3–61.

Li, A. S., & Marikawa, Y. (2015). An in vitro gastrulation model recapitulates the morphogenetic impact of pharmacological inhibitors of developmental signaling pathways. *Molecular Reproduction and Development*, *82*, 1015–1036. https://doi.org/10.1002/mrd.22585

Li, A. S., & Marikawa, Y. (2016). Adverse effect of valproic acid on an in vitro gastrulation model entails activation of retinoic acid signaling. *Reproductive Toxicology*, *66*, 68–83. https://doi.org/10.1016/j.reprotox.2016.09.015

Luijten, M., Verhoef, A., Westerman, A., & Piersma, A. H. (2008). Application of a metabolizing system as an adjunct to the rat whole embryo culture. *Toxicology in Vitro*, *22*, 1332–1336. https://doi.org/10.1016/j.tiv.2008.04.009

Luz, A. L., & Tokar, E. J. (2018). Pluripotent stem cells in developmental toxicity testing: A review of methodological advances. *Toxicological Sciences*, *165*, 31–39. https://doi.org/10.1093/toxsci/kfy174

Mantziou, V., Baillie-Benson, P., Jaklin, M., Kustermann, S., Arias, A. M., & Moris, N. (2021). In vitro teratogenicity testing using a 3D, embryo-like gastruloid system. *Reproductive Toxicology*, *105*, 72–90. https://doi.org/10.1016/j.reprotox.2021.08.003

Marikawa, Y., Chen, H. R., Menor, M., Deng, Y., & Alarcon, V. B. (2020). Exposure-based assessment of chemical teratogenicity using morphogenetic aggregates of human embryonic stem cells. *Reproductive Toxicology*, *91*, 74–91. https://doi.org/10.1016/j.reprotox.2019.10.004

Martello, G., & Smith, A. (2014). The nature of embryonic stem cells. *Annual Review of Cell and Developmental Biology*, *30*, 647–675. https://doi.org/10.1146/annurev-cellbio-100913-013116

Marx-Stoelting, P., Adriaens, E., Ahr, H. J., Bremer, S., Garthoff, B., Gelbke, H. P., ... Schwarz, M. (2009). A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect workshop. *Alternatives to Laboratory Animals*, *37*, 313–328. https://doi.org/10.1177/026119290903700314

Matsui, T., & Shinozawa, T. (2021). Human organoids for predictive toxicology research and drug development. *Frontiers in Genetics*, *12*, 767621. https://doi.org/10.3389/fgene.2021.767621

Mennen, R. H., Oldenburger, M. M., & Piersma, A. H. (2022). Endoderm and mesoderm derivatives in embryonic stem cell differentiation and their use in developmental toxicity testing. *Reproductive Toxicology*, *107*, 44–59. https://doi.org/10.1016/j.reprotox.2021.11.009

Oglesby, L. A., Ebron, M. T., Beyer, P. E., Carver, B. D., & Kavlock, R. J. (1986). Co-culture of rat embryos and hepatocytes: In vitro detection of a proteratogen. *Teratogenesis Carcinogenesis and Mutagenesis*, *6*, 129–138. https://doi.org/10.1002/tcm.1770060206

Ozolins, T. R., Oglesby, L. A., Wiley, M. J., & Wells, P. G. (1995). In vitro murine embryotoxicity of cyclophosphamide in embryos co-cultured with maternal hepatocytes: Development and application of a murine embryo-hepatocyte co-culture model. *Toxicology*, *102*, 259–274. https://doi.org/10.1016/0300-483x(95)03082-q

Persson, B., & Hansson, U. (1993). Hypoglycaemia in pregnancy. *Baillière's Clinical Endocrinology and Metabolism*, *7*, 731–739. https://doi.org/10.1016/s0950-351x(05)80216-2

Piersma, A. H., Hessel, E. V., & Staal, Y. C. (2017). Retinoic acid in developmental toxicology: Teratogen, morphogen and biomarker. *Reproductive Toxicology*, *72*, 53–61. https://doi.org/10.1016/j.reprotox.2017.05.014

Piersma, A. H., van Aerts, L. A., Verhoef, A., Garbis-Berkvens, J. M., Robinson, J. E., Peereboom-Stegeman, J. H., & Peters, P. W. (1991). Biotransformation of cyclophosphamide in post-implantation rat embryo culture using maternal hepatocytes in co-culture. *Pharmacology & Toxicology*, *69*, 47–51. https://doi.org/10.1111/j.1600-0773.1991.tb00408.x

Riebeling, C., Hayess, K., Peters, A. K., Steemans, M., Spielmann, H., Luch, A., & Seiler, A. E. (2012). Assaying embryotoxicity in the test tube: Current limitations of the embryonic stem cell test (EST) challenging its applicability domain. *Critical Reviews in Toxicology*, *42*, 443–464. https://doi.org/10.3109/10408444.2012.674483

Rogers, J. M. (2020). Adverse outcome pathways for developmental toxicity. In R. Kishi & P. Grandjean (Eds.), *Health impacts of developmental exposure to environmental chemicals. Current topics in environmental health and preventive medicine*. Singapore: Springer. https://doi.org/10.1007/978-981-15-0520-1_17

Sasai, Y. (2013). Next-generation regenerative medicine: Organogenesis from stem cells in 3D culture. *Cell Stem Cell*, *12*, 520–530. https://doi.org/10.1016/j.stem.2013.04.009

Scialli, A. R., Daston, G., Chen, C., Coder, P. S., Euling, S. Y., Foreman, J., ... Thompson, K. E. (2018). Rethinking developmental toxicity testing: Evolution or revolution? *Birth Defects Research*, *110*, 840–850. https://doi.org/10.1002/bdr2.1212

Seiler, A. E., Buesen, R., Visan, A., & Spielmann, H. (2006). Use of murine embryonic stem cells in embryotoxicity assays: The embryonic stem cell test. *Methods in Molecular Biology*, *329*, 371–395. https://doi.org/10.1385/1-59745-037-5:371

Smith, D. A., Di, L., & Kerns, E. H. (2010). The effect of plasma protein binding on in vivo efficacy: Misconceptions in drug discovery. *Nature Reviews Drug Discovery*, *9*, 929–939. https://doi.org/10.1038/nrd3287

Smith, M. K., Kimmel, G. L., Kochhar, D. M., Shepard, T. H., Spielberg, S. P., & Wilson, J. G. (1983). A selection of candidate compounds for in vitro teratogenesis test validation. *Teratogenesis Carcinogenesis and Mutagenesis*, *3*, 461–480. https://doi.org/10.1002/1520-6866(1990)3:6<461::aid-tcm1770030603>3.0.co;2-i

Song, M. O., Fort, D. J., McLaughlin, D. L., Rogers, R. L., Thomas, J. H., Buzzard, B. O., ... Myers, N. K. (2004). Evaluation of Xenopus tropicalis as an alternative test organism for frog embryo teratogenesis assay--Xenopus (FETAX). *Drug and Chemical Toxicology*, *26*, 177–189. https://doi.org/10.1081/dct-120022647

Takahashi, K., & Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews Molecular Cell Biology*, *17*, 183–193. https://doi.org/10.1038/nrm.2016.8

Tong, X. L., Wang, L., Gao, T. B., Qin, Y. G., Qi, Y. Q., & Xu, Y. P. (2009). Potential function of amniotic fluid in fetal development—Novel insights by comparing the composition of human amniotic fluid with umbilical cord and maternal serum at mid and late gestation. *Journal of the Chinese Medical Association*, *72*, 368–373. https://doi.org/10.1016/S1726-4901(09)70389-2

Underwood, M. A., Gilbert, W. M., & Sherman, M. P. (2005). Amniotic fluid: Not just fetal urine anymore. *Journal of Perinatology*, *25*, 341–348. https://doi.org/10.1038/sj.jp.7211290

van den Brink, S. C., & van Oudenaarden, A. (2021). 3D gastruloids: A novel frontier in stem cell-based in vitro modeling of mammalian gastrulation. *Trends in Cell Biology*, *31*, 747–759. https://doi.org/10.1016/j.tcb.2021.06.007

Wang, H., Brown, P. C., Chow, E. C. Y., Ewart, L., Ferguson, S. S., Fitzpatrick, S., ... Huang, S. M. (2021). 3D cell culture models: Drug pharmacokinetics, safety assessment, and regulatory consideration. *Clinical and Translational Science*, *14*, 1659–1680. https://doi.org/10.1111/cts.13066

Warkus, E. L., Yuen, A. A., Lau, C. G., & Marikawa, Y. (2016). Use of in vitro morphogenesis of mouse embryoid bodies to assess developmental toxicity of therapeutic drugs contraindicated in pregnancy. *Toxicological Sciences*, *149*, 15–30. https://doi.org/10.1093/toxsci/kfv209

Warkus, E. L. L., & Marikawa, Y. (2017). Exposure-based validation of an in vitro gastrulation model for developmental toxicity assays. *Toxicological Sciences*, *157*, 235–245. https://doi.org/10.1093/toxsci/kfx034

Warkus, E. L. L., & Marikawa, Y. (2018). Fluoxetine inhibits canonical Wnt signaling to impair embryoid body morphogenesis: Potential teratogenic mechanisms of a commonly used antidepressant. *Toxicological Sciences*, *165*, 372–388. https://doi.org/10.1093/toxsci/kfy143

Webster, W. S., Brown-Woodman, P. D., & Ritchie, H. E. (1997). A review of the contribution of whole embryo culture to the determination of hazard and risk in teratogenicity testing. *The International Journal of Developmental Biology*, *41*, 329–335.

Wells, P. G., & Winn, L. M. (1996). Biochemical toxicology of chemical teratogenesis. *Critical Reviews in Biochemistry and Molecular Biology*, *31*, 1–40. https://doi.org/10.3109/10409239609110574

Wertz, D. C. (2002). Embryo and stem cell research in the United States: History and politics. *Gene Therapy*, *9*, 674–678. https://doi.org/10.1038/sj.gt.3301744

Zhao, J., Krafft, N., Terlouw, G. D., & Bechter, R. (1993). A model combining the whole embryo culture with human liver S-9 fraction for human teratogenic prediction. *Toxicology In Vitro*, *7*, 827–831. https://doi.org/10.1016/0887-2333(93)90087-l