OXFORD

SOT | Society of Toxicology
www.toxsci.oxfordjournals.org

ToxSci 20 Years

TOXICOLOGICAL SCIENCES, 165(1), 2018, 31–39

doi: 10.1093/toxsci/kfy174
Advance Access Publication Date: June 4, 2018
Contemporary Review

CONTEMPORARY REVIEW

Pluripotent Stem Cells in Developmental Toxicity Testing: A Review of Methodological Advances

Anthony L. Luz and Erik J. Tokar[1]

Stem Cell Toxicology Group, National Toxicology Program Laboratory, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709

[1]To whom correspondence should be addressed at Stem Cell Toxicology Group, National Toxicology Program Laboratory, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, 111 TW Alexander Drive, Building 101, Room E-105, Research Triangle Park, NC 27709. E-mail: erik.tokar@nih.gov.

ABSTRACT

Millions of children are born each year with a birth defect. Many of these defects are caused by environmental factors, although the underlying etiology is often unknown. *In vivo* mammalian models are frequently used to determine if a chemical poses a risk to the developing fetus. However, there are over 80 000 chemicals registered for use in the United States, many of which have undergone little safety testing, necessitating the need for higher-throughput methods to assess developmental toxicity. Pluripotent stem cells (PSCs) are an ideal *in vitro* model to investigate developmental toxicity as they possess the capacity to differentiate into nearly any cell type in the human body. Indeed, a burst of research has occurred in the field of stem cell toxicology over the past decade, which has resulted in numerous methodological advances that utilize both mouse and human PSCs, as well as cutting-edge technology in the fields of metabolomics, transcriptomics, transgenics, and high-throughput imaging. Here, we review the wide array of approaches used to detect developmental toxicants, suggest areas for further research, and highlight critical aspects of stem cell biology that should be considered when utilizing PSCs in developmental toxicity testing.

Key words: pluripotent stem cells; teratogen; developmental toxicity.

Globally, approximately 8 million children are born each year with a birth defect. The causes of only 30% of these birth defects are somewhat understood, leaving the possibility that environmental factors could be playing a significant role (Weinhold, 2009). In support of this, *in utero* exposure to various environmental contaminants, including industrial solvents, metals, and pesticides have been implicated in causing birth defects in humans and in various mammalian models (reviewed in Stillerman *et al.*, 2008). However, there are currently over 80 000 chemicals registered for commercial use in the United States, and there is limited human toxicity and exposure data for the majority of these chemicals (Judson *et al.*, 2009). The prenatal developmental toxicity test, which utilizes pregnant mice, rats, or rabbits, is considered the definitive test to identify chemicals that may pose a risk to the developing fetus, and is regularly used to establish human exposure guidelines. However, these *in vivo* tests are both expensive and time-consuming, making it impractical to test all commercially used chemicals for developmental toxicity, thus necessitating the development of rapid, low-cost *in vitro* methods that can be used to detect potential developmental toxicants, and prioritize chemicals for further *in vivo* testing.

In response, researchers have developed myriad alternative methods for detecting developmental toxicants. These include the *in vivo* zebrafish embryotoxicity test (Chapin *et al.*, 2008), and the frog embryo teratogenesis assay (Bantle *et al.*, 1989), as well as the *in vitro* rat micromass (MM) test, the rat whole-embryo culture assay (WEC test), and the mouse embryonic stem cell (mESC) test (mEST) (Genschow *et al.*, 2002). Of these early embryotoxicity tests, the MM test, WEC test, and the mEST have all undergone thorough validation studies coordinated by the European Center for the Validation of Alternative Methods

(ECVAM) (Genschow *et al.*, 2002). Furthermore, these *in vitro* methods can all properly classify embryotoxic compounds with at least 70% accuracy (Table 1). However, the mEST is the only assay that does not require the sacrifice of pregnant animals, making it a more attractive method for developmental toxicity testing. Since its validation, the mEST has been used extensively by researchers to test the embryotoxic potential of a wide array of compounds, including metals (Stummann *et al.*, 2007, 2008), nanoparticles (Di Guglielmo *et al.*, 2010; Park *et al.*, 2009), cosmetics (Chen *et al.*, 2010), industrial chemicals (de Jong *et al.*, 2009), pharmaceuticals (Eckardt and Stahlmann, 2010; Paquette *et al.*, 2008), and a wide-variety of environmental contaminants (Kamelia *et al.*, 2017; Kong *et al.*, 2013; Zhou *et al.*, 2017). Indeed, the widespread acceptance of the mEST by the toxicological research community over 15 years ago has triggered a burst of research in the field of stem cell toxicology and has led to the development of numerous *in vitro* assays that can be used to detect developmental toxicants. These methodological advances use mESCs and human embryonic stem cells (hESCs), spontaneous and directed ESC differentiation protocols, and cutting-edge technology in the fields of metabolomics, transcriptomics, transgenics, and high-throughput imaging. Here, we review the wide array of approaches used to detect developmental toxicants, as well as highlight critical aspects of stem cell biology that should be considered when utilizing ESCs in developmental toxicity testing.

CHALLENGES OF *IN VITRO* DEVELOPMENTAL TOXICITY TESTING

No *in vitro* assay can completely recapitulate the complexities of fetal development or the fetal-maternal interactions that occur *in vivo*. Many of the challenges involved in using ESCs for developmental toxicity testing have been recently reviewed (Kugler *et al.*, 2017), and thus are only briefly highlighted here.

Fetal-Maternal Interactions

Pluripotent SCs can differentiate into nearly any cell type in the human body, including trophoblasts, which are placenta-specific epithelial cells (Gamage *et al.*, 2016); however, the formation of mature placental tissue from PSCs has not been demonstrated. *In vivo*, the developing fetus is dependent upon the placenta for transfer of nutrients and gases from maternal blood. From a toxicological perspective, the placenta can play a role in transporting toxicants from maternal to fetal blood resulting in fetal exposure, or may act as a barrier, preventing fetal exposure. Recently, researchers have utilized placental BeWo cells to generate toxicokinetic data, and have demonstrated that incorporation of this data into the mEST can increase assay predictivity (Dimopoulou *et al.*, 2018). Maternal metabolism can also play a role in the detoxification or toxification of teratogens. As minimal xenobiotic metabolism occurs in ESCs, it is likely that most proteratogens will be classified as false-negatives in ESC-based assays. To overcome this, metabolic activation systems (MAS) utilizing rat liver microsomes have been incorporated into the zebrafish embryotoxity test (Busquet *et al.*, 2008) and the frog embryo teratogenesis assay (Fort *et al.*, 1989) to great effect; however, MAS have yet to be incorporated into ESC-based assays.

Human Versus Mouse ESCs

To date, most developmental toxicity testing has been performed using mESCs (Table 2), and the implementation of hESCs in developmental toxicity testing has been slow. Ethical and legal issues of performing toxicity testing with hESCs may be partially to blame. Historically, hESCs have been more difficult to culture than mESCs; however, advances in culturing techniques have eliminated this issue (Desai *et al.*, 2015). The most apparent advantage of using hESCs instead of mESCs is to limit the possibility of false-negatives that may arise due to species-specific differences. However, it is yet to be demonstrated that hESC-based assays perform better than mESC-based developmental toxicity assays. Furthermore, species-specific differences are often due to *in vivo* differences in metabolism or toxicokinetics, which may not apply to *in vitro* assays. Finally, it is important to note that mESCs are more naïve than hESCs, meaning the molecular features of mESCs more closely resemble those of pluripotent cells in the early embryo, which may mean mESCs are a more appropriate model for early embryogenesis.

Table 1. Alternative Models in Developmental Toxicity Testing

Model	Accuracy	References
Mouse embryonic stem cell test	78%	Genschow *et al.* (2002)
Rat MM test	70%	Genschow *et al.* (2002)
Rat WEC assay	80%	Genschow *et al.* (2002)
Zebrafish embryotoxicity test	72%	Chapin *et al.* (2008)
Frog embryo teratogenesis assay	NA	Bantle *et al.* (1989)

Embryonic Versus Induced Pluripotent

Induced pluripotent stem cells (iPSCs) are derived from somatic cells that are reprogramed back to an embryonic-like state. Although these cells are pluripotent, their use in developmental toxicity testing has remained limited. This may, in part, be due to the fact that many iPSC lines tend to have lineage bias towards the lineage of origin, which may be due to an incomplete reset of DNA methylation back to an "embryonic state" (Liang and Zhang, 2013). However, many commonly used ESC lines have also been shown to have lineage biases (Bock *et al.*, 2011; Tsankov *et al.*, 2015). Additionally, donor age, sex, race, and exposure history may all influence the toxicant response of iPSCs. Therefore, the true power of iPSCs may lie in the field of personalized toxicology, and in their utility to incorporate genetic diversity into *in vitro* developmental toxicology studies [reviewed in (Jennings, 2015; Liu *et al.*, 2017)].

Culture Conditions

Unsurprisingly, growing work has demonstrated that culture conditions can affect SC pluripotency and differentiation. Most notably, the international stem cell initiative (ISCI) compared 8 commonly used culture methods, and found only a few were capable of maintaining PSCs for at least 10 passages (Akopian *et al.*, 2010). Furthermore, the ISCI has also reported subtle genetic differences, including alterations in DNA methylation and SNP occurrence, when they compared over 100 low- and high-passage PSC lines from 38 laboratories (International Stem Cell Initiative, 2011). More recently, Merkle *et al.* (2017) analyzed 140 PSCs lines and identified 6 mutations in the tumor suppressor P53. Furthermore, the allelic fraction increased with passage number suggesting the mutations confer selective advantage. These studies demonstrate that researchers need to carefully consider not only the culture system, but also limit PSC passage number to avoid the potential accumulation of mutations or epigenetic alterations that may increase assay variation.

Table 2. Stem Cell-Based Methods for Developmental Toxicant Screening

Assay	Species	Duration (days)	Compounds in Test Set	Summary	References
mEST	Mouse	10	15-20	Accuracy: 78%–83%	Genschow et al. (2002), Paquette *et al.* (2008)
FACS-EST	Mouse	7	10	Accuracy: 100%	Buesen *et al.* (2009)
Molecular-EST	Mouse	4	65	Accuracy: 72%; Sensitivity: 76%; Specificity: 69%	Panzica-Kelly *et al.* (2013)
EBT	Mouse	10	21	Accuracy: 90.5%	Kang *et al.* (2017)
Untargeted metabolomics	Human	4	8	Accuracy: 88%; Sensitivity: 80%; Specificity: 100%	West *et al.* (2010)
	Human	3	11	Accuracy: 83%; Sensitivity: 92%; Specificity: 75%	Kleinstreuer *et al.* (2011)
Metabolomics (O/C ratio)	Human	3	13	Accuracy: 77%; Sensitivity: 57%; Specificity: 100%	Palmer *et al.* (2013)
High-throughput imaging	Human	3	71	Accuracy: 94%; Sensitivity: 97%; Specificity: 92%	Kameoka *et al.* (2014)
ReProGlo Assay	Mouse	1	17	Accuracy: 76%; Sensitivity: 71%; Specificity: 100%	Uibel *et al.* (2010)
Hand1-EST	Mouse	6	24	Accuracy: 83%; Sensitivity: 93%; Specificity: 63%	Suzuki *et al.* (2011)
Cmya-1-EST	Mouse	6	24	Accuracy: 92%; Sensitivity: 93%; Specificity: 87%	Suzuki *et al.* (2011)

Solvent Effects

Solvents used to solubilize hydrophobic compounds must be chosen carefully. Research has demonstrated that both ethanol and dimethyl sulfoxide (DMSO) can adversely affect pluripotency and differentiation of mouse and human PSCs. For example, DMSO (≥0.1%) can reduce *Oct4* expression and alter DNA methylation patterns in mESCs (Adler *et al.*, 2006; Iwatani *et al.*, 2006), and impair differentiation of hESCs (Czysz *et al.*, 2015; Pal *et al.*, 2012). Similarly, ethanol (≥0.25%) can alter *Oct4* expression in mPSCs (Adler *et al.*, 2006), and alter DNA methylation, induce apoptosis, and impair neuronal differentiation in hESCs (Khalid *et al.*, 2014; Nash *et al.*, 2012). Given these low-dose solvent effects, final solvent concentrations should be kept low, and appropriate solvent controls should always be included.

PSC-BASED METHODS FR DEVELOPMENTAL TOXICITY TESTING

The mESCs Test and Assay Variations

The mEST was one of the first, and mostly thoroughly validated, ESC-based assays used to assess compounds for embryotoxicity (Genschow *et al.*, 2002). Embryotoxicity is based on 3 endpoints: the ability of a compound to (1) impair differentiation of mESCs into contracting cardiomyocytes, and cause cytotoxicity in (2) undifferentiated mESCs, and in (3) differentiated 3T3 fibroblasts (Genschow *et al.*, 2002) (Figure 1). Overall, the accuracy of the mEST is 78%–83% (Genschow *et al.*, 2002; Paquette *et al.*, 2008), and has been used extensively by toxicologists. However, the mEST has limitations. First, scoring cardiomyocyte beating is a subjective process that requires highly trained staff. Second, the assay is relatively low-throughput. Third, the mEST assumes that undifferentiated ESCs will be more sensitive than highly differentiated fibroblasts to developmental toxicants; however, exceptions to this assumption have led to the misclassification of methyl mercury, cadmium, and various arsenicals (Genschow *et al.*, 2002; Stummann *et al.*, 2008).

Given these limitations, a variety of modified versions of the mEST have been developed, although they have not been validated as thoroughly. For example, protocols have been optimized for embryoid body (EB) formation in 96-well plates to increase assay throughput (Peters *et al.*, 2008), while fluorescent-activated cell sorting has been used to assess cardiomyocyte differentiation, thus eliminating subjectivity of scoring contracting cardiomyocytes (Buesen *et al.*, 2009). Innovatively, Dimopoulou *et al.* (2018) incorporated placental BeWo b30 cells into the mEST, and demonstrated that the incorporation of placental transport velocity data can increase assay predictivity. A great deal of effort has also been expended to incorporate transcriptomics into the mEST (>50 studies published [reviewed in van Dartel and Piersma, 2011]). Of note, the molecular EST, in which expression of 12 developmentally regulated genes are assessed, reduced assay time from 10 to 4 days, and had a similar degree of accuracy (72%–83%) as the original mEST (Panzica-Kelly *et al.*, 2012).

The transgenic EST. Transgenic mESCs that can be used to monitor cardiomyocyte differentiation have also been generated (Le Coz *et al.*, 2015; Nagahori *et al.*, 2016; Suzuki *et al.*, 2011). In these models, luciferase expression is driven by either the Hand1 or Cmya1 promoters, both of which are considered indispensable for proper heart development. Importantly, the endpoint of cardiomyocyte beating is replaced with the rapid and highly quantitative endpoint of luciferase expression. Furthermore, both the Hand1-Luc and Cmya1-Luc assays have been shown to properly predict 83% and 92% of compounds, respectively.

The EB stem cell test. Recently, several groups have proposed that EB growth dynamics can be used to predict the embryotoxic potential of chemicals (Flamier *et al.*, 2017; Kang *et al.*, 2017; Warkus and Marikawa, 2017). Kang *et al.* (2017) demonstrate a strong correlation between EB area and cardiomyocyte beating, indicating that a reduction in EB size closely reflects cardiomyocyte differentiation. This led to the development of a novel mEST variation in which cardiomyocyte beating was replaced

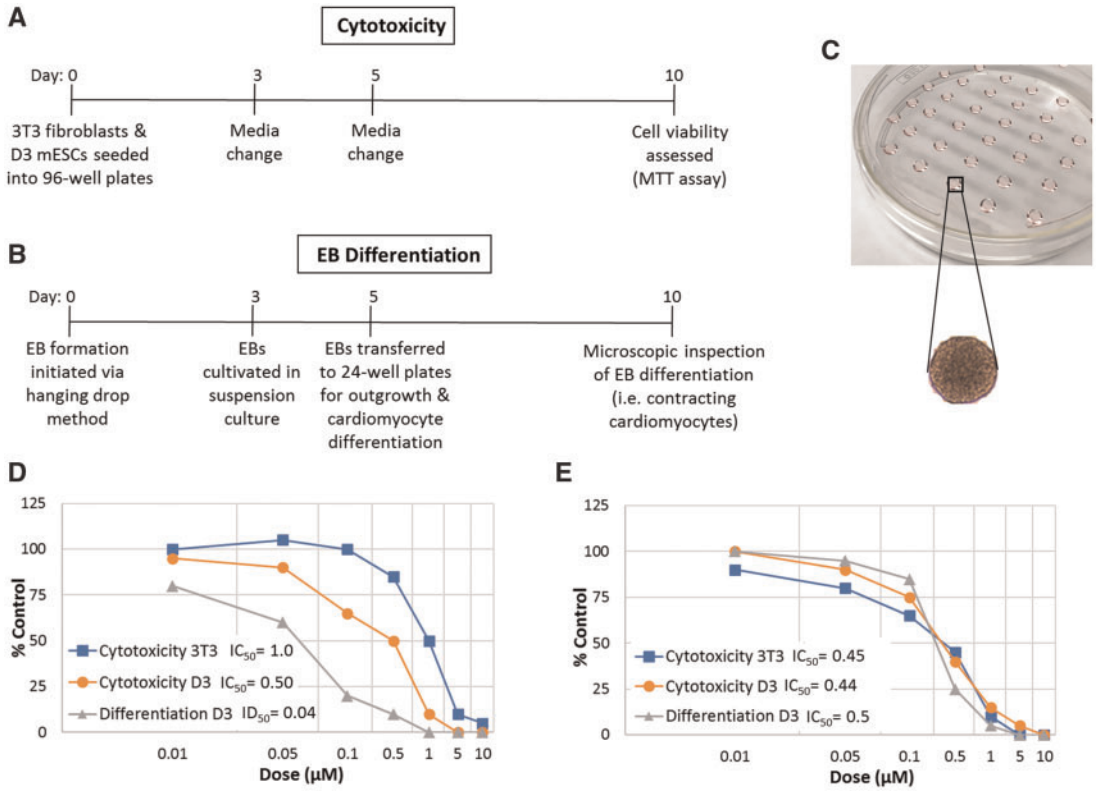

Figure 1. The mESCs test. A, Viability of pluripotent murine D3 ESCs and differentiated 3T3 fibroblasts is compared across an 8-point dose-response following 10 days of toxicant exposure. B, EBs are generated via the hanging drop method, and allowed to differentiate into contracting cardiomyocytes while being exposed for 10 days across an 8-point dose-response. C, EB formation via the hanging drop method. Compound concentrations that reduce cell viability (IC_{50}) and cardiomyocyte differentiation (ID_{50}) by 50% are calculated and input into the EST prediction model to determine the embryotoxic potential of the test compound. D, Hypothetical example of concentration-response curves for an embryotoxic compound that inhibits cardiomyocyte differentiation at non-cytotoxic concentrations. E, Hypothetical example of concentration response curves for a nonembryotoxic compound that only inhibits cardiomyocyte differentiation at cytotoxic concentrations.

with an EB size measurement, termed the EB stem cell test (EBT). The accuracies of the mEST and EBT were found to be similar (86.9% and 90.5%, respectively) when compared across a 21-compound test set. Although the EBT requires further validation, it may represent a significant advancement over the mEST, as it is more adaptable to high-throughput screening, as EB growth can be monitored using high-content imaging (HCI) devices.

Alternative differentiation models. Additional differentiation models have also been incorporated into the mEST in an attempt to expand the assays applicability domain. For example, Adler *et al.* (2008) developed a human EST to limit potential false-negatives that may arise due to species-specific differences. Furthermore, many embryotoxic compounds can adversely affect skeletal development (an endpoint regularly monitored in *in vivo* developmental toxicity studies), which has led researchers to incorporate osteoblast differentiation protocols into the mEST (Chen *et al.*, 2015; zur Nieden *et al.*, 2010a,b). Finally, because the original mEST failed to properly classify the strong neurodevelopmental toxicant methyl mercury, numerous research groups have worked to incorporate neuron differentiation protocols into the mEST (Baek *et al.*, 2012; Stummann *et al.*, 2009; Theunissen *et al.*, 2010), which have resulted in the proper classification of methyl mercury.

Metabolomics-Based Approaches

Metabolomics is the study of small molecules (metabolites) that are the end-product of various cellular processes, including energy metabolism. This highly quantitative approach is being implemented in developmental toxicity testing, and has allowed researchers to identify small molecules that can serve as putative biomarkers of myriad diseases (Cezar *et al.*, 2007; Palmer *et al.*, 2013; West *et al.*, 2010). Cezar *et al.* (2007) first established the utility of metabolomics in developmental toxicity testing by demonstrating that valproate, a neurodevelopmental toxicant, can alter the secretome of hESCs, affecting processes such as tryptophan and glutamate metabolism. In a follow-up study, hESCs were exposed to the ECVAM test set, and the secretome was analyzed using an untargeted metabolomics approach, which led to the identification of 8 small metabolites (dimethylarginine, aspartic acid, arginine, glutamate, GABA, malate, succinate, isoleucine) that correlated with teratogenicity (West *et al.*, 2010). A predictive model was then generated that could accurately classify 88% (7/8) of drugs and 73% (8/11) of environmental toxicants in 2 separate blinded test sets (Kleinstreuer *et al.*, 2011). More recently, it was found that a 12% reduction in the ornithine to cysteine (O/C) ratio in the secretome of hESCs following a 3-day exposure was predictive of developmental toxicity when compared with cell viability across a 9-point dose-response (Palmer *et al.*, 2013). Although this approach could accurately classify 77% of compounds in a 13-compound test set, the overall sensitivity of the assay was low, as only 57% (4/7) of known teratogens were properly classified, while 100% of nonteratogens were correctly classified. Despite this high false-negative rate, changes in the O/C ratio have been used successfully to rank retinoid analogs based on teratogenic

potency using human iPSCs (Palmer *et al.*, 2017). Furthermore, this metabolomics-based method is the only commercially available service that utilizes PSCs for developmental toxicity testing (www.stemina.com).

Spontaneous Differentiation of EBs

EBs are 3D aggregates of PSCs that spontaneously differentiate into all 3 embryonic germ layers (ectoderm, mesoderm, and endoderm), in a process that recapitulates many of the molecular events that occur throughout early embryogenesis (Weitzer, 2008). To test for developmental toxicity, Shinde *et al.* (2015) exposed differentiating EBs to teratogens throughout a 14-day differentiation window, and then assessed EB differentiation using transcriptomics and immunocytochemistry. Although the overall accuracy of this approach has not yet been established, several proof-of-principle experiments with cytosine arabinoside (Jagtap *et al.*, 2011), thalidomide (Meganathan *et al.*, 2012), valproic acid (Krug *et al.*, 2013), and methyl mercury (Shinde *et al.*, 2015) have demonstrated the validity of this approach. For example, thalidomide perturbed the expression of genes associated with limb and heart development (Meganathan *et al.*, 2012), which coincides with clinical observations of thalidomide toxicity in humans.

Despite these promising characteristics, this assay is relatively low-throughput and does not utilize consistently sized or individually cultured EBs. This may prove problematic for high-throughput developmental toxicant screening, as EB size can influence PSC differentiation, likely due to reduced diffusion of nutrients and oxygen into the core of larger EBs, resulting in increased cell death and altered differentiation patterns (Moon *et al.*, 2014; Nath *et al.*, 2017). Furthermore, pooled EBs rapidly (within several hours) fuse, resulting in increased EB size, cell death, and altered differentiation (Dang *et al.*, 2002), which may also contribute to assay variability. However, advances in EB formation protocols may allow researchers to overcome these limitations (reviewed in Pettinato *et al.*, 2015).

Transgenic Reporter Strains

Early embryogenesis is primarily under the control of 6 signaling pathways—the Wnt/β-catenin, transforming growth factor β (TGF-β), Notch, Hedgehog, receptor tyrosine kinase/Ras, and cytokine receptor signaling pathways. The crucial nature of these pathways is demonstrated by the fact that genetic manipulation results in embryonic lethality or developmental defects in mammalian models (Loebel *et al.*, 2003). This has led to the development of several transgenic ESC lines that can be used to monitor pathway activity following toxicant exposure. For example, the ReProGlo Assay utilizes mESCs transfected with the SuperTopFlash luciferase reporter, which can be used to monitor Wnt signaling pathway activity following toxicant exposure (Uibel *et al.*, 2010). In the initial validation study, the ReProGlo assay could properly classify 76% of compounds in a 17-compound test set. Although the low-cost and rapid (24-h) nature of the ReProGlo assay make it an attractive tool for developmental toxicant screening, follow-up studies have reported high false-negative rates, suggesting the applicability domain of the assay needs to be better defined (Uibel and Schwarz, 2015). However, generation of additional transgenic lines capable of assessing the activity of the other signaling pathways (ie, Notch, Hedgehog, TGF-β, receptor tyrosine kinase/Ras, and cytokine receptor signaling) may help overcome this limitation when used in combination.

Another strategy used to generate transgenic ESC lines is to isolate ESCs from transgenic mouse models, which has led to the generation of 2 transgenic mESC lines that can be used to monitor Wnt and TGF-β signaling (Kugler *et al.*, 2015, 2016). Both models have been tested with a small set of developmental toxicants (valproic acid, retinoic acid, and 6-aminonicotinamide), and reporter activity correlates well with subsequent cardiomyocyte differentiation assays, demonstrating the validity of these approaches. However, the overall accuracy of these models is yet to be tested.

High-Content Imaging

HCI, in combination with automated image analysis software, can allow researchers to rapidly screen large suites of compounds for biological activity. Utilizing this technology, Kameoka *et al.* (2014) directed hESCs to differentiate down the mesendoderm lineage in the presence of toxicants throughout a 3-day differentiation window. Cell viability and differentiation were then assessed by staining for DAPI and SOX17, an established mesendoderm marker. Teratogenic risk was based upon a compounds ability to reduce nuclear translocation of the SOX17 transcription factor. Impressively, 94% of pharmaceutical compounds (67/71), and 87% of environmental toxicants (13/15) with known *in vivo* outcomes were properly classified using this approach. Given the rapid (72 h) and automated nature of this approach, it represents a promising advancement in the field of stem cell toxicology; however, the requirement for expensive HCI systems may limit its widespread utility.

In Vitro *Test Battery Approaches*

Embryonic development is a highly complex process that no single *in vitro* assay can completely recapitulate. Furthermore, no single PSC-based assay will be able to detect development toxicants with 100% accuracy. This realization has led several groups to develop a battery of *in vitro* assays that can be used to screen chemicals for developmental toxicity. For example, the ReProTect study employed 14 *in vitro* assays, including the mEST, ReProGlo assay, and WEC discussed earlier. This test was designed to detect adverse effects on male and female fertility, and embryonic development (Schenk *et al.*, 2010). Combined, the 3 *in vitro* assays properly classified >90% of compounds across 2 separate studies (Piersma *et al.*, 2013; Schenk *et al.*, 2010). Alternatively, Augustine-Rauch *et al.* (2016) optimized 3 *in vitro* assays, including the zebrafish embryo culture assay (ZEC), WEC, and the molecular EST, for screening pharmaceuticals for developmental toxicity. Overall, the individual predictivity of the 3 assays ranged from 73% to 82% for a 73-compound library. However, the main advantage of this approach was that it allowed the authors to detect toxicants with diverse mechanisms of action, and more clearly define the applicability domains of each assay. For example, the molecular EST and ZEC frequently misclassified hydrophobic compounds, and compounds with H1 receptor or GABAnergic activity, while the WEC could accurately classify these compounds. Overall, the suite of assays correctly predicted the teratogenicity of 64 out of 73 compounds (88%), while achieving 95% predictivity for the 21 known human teratogens included in the test set.

FUTURE DIRECTIONS

As with any new assay, validation is key to acceptance by the scientific community. To date, the mEST remains the most thoroughly validated ESC-based assay for developmental toxicant testing. Validated by ECVAM, the mEST was tested by 4 independently contracted laboratories against a blinded test set of 20 compounds with known *in vivo* outcomes (Genschow *et al.*, 2002).

Importantly, this process allowed not only assay predictivity to be calculated, but also intra- and interlab variability. Despite the limitations of the mEST, the assay has been widely accepted and used by the toxicological research community, likely in part due to its thorough validation. Alternatively, many of the novel ESC-based assays discussed in this review have not undergone as thorough of a validation process, and are typically only tested for predictive capacity, while assay applicability domains remain poorly defined. Thus, further work is required to validate these cutting-edge approaches. However, it is also important to note that assay validation does not equate to regulatory acceptance, and to date, no SC-based assay is used to make regulatory decisions.

Due to limitations, no single *in vitro* assay will ever be able to detect developmental toxicants with 100% accuracy. Thus, it is critical that assay limitations are acknowledged and discussed so that more complete *in vitro* approaches to developmental toxicant screening can be developed. In particular, developing a battery of PSC-based assays may prove the best approach for *in vitro* developmental toxicity testing, as several studies have shown promising results (Augustine-Rauch *et al.*, 2016; Schenk *et al.*, 2010). However, it is necessary that assays chosen for inclusion in the battery are thoroughly validated, and are chosen such that the assays address the limitations of one another. For example, inclusion of the mEST and the ReProGlo assay, as done in the ReProTect study, may not be appropriate (Schenk *et al.*, 2010), as the ReProGlo assay detects perturbations in Wnt signaling, but not other signaling networks critical to embryonic development. Instead, inclusion of multiple transgenic reporter strains designed to detect perturbations in the key developmental signaling pathways (ie, Wnt, Notch, TGF-β, receptor tyrosine kinase, cytokine receptor, and Hedgehog), in combination with assays that utilize spontaneously differentiating EBs (Shinde *et al.*, 2015), or ESCs directed to differentiate down a specific lineage (Kameoka *et al.*, 2014) may dramatically expand the applicability domain of the test battery and limit false-negatives. However, it will also be critical to address some of the more conspicuous limitations of PSC-based assays, such as lack of xenobiotic metabolism. This point may be especially important, as maternal metabolism can bioactivate benign compounds to teratogens, and these proteratogens would likely be misclassified as false-negatives by current PSC-based methods. Furthermore, zebrafish and frog teratogenesis assays have made good use of MASs (Busquet *et al.*, 2008; Fort *et al.*, 1989) that have potential to be applicable to PSC-based assays.

There is a growing push to use 3D cell culture models in toxicology, disease modeling, and drug discovery, as mounting evidence indicates that 3D *in vitro* systems are more analogous to *in vivo* events than cells grown in monolayer (reviewed in Ravi *et al.*, 2015; Trosko, 2018). In response, myriad complex 3D *in vitro* organoid models have been generated using human PSCs, including intestinal (Leslie *et al.*, 2015), kidney (Freedman *et al.*, 2015; Takasato *et al.*, 2015), liver (Takebe *et al.*, 2013), lung (Dye *et al.*, 2015), neural (Jo *et al.*, 2016; Sandström *et al.*, 2017) pancreas (Hohwieler *et al.*, 2017), prostate (Calderon-Gierszal and Prins, 2015), and stomach organoids (McCracken *et al.*, 2014). To date, these organoids have primarily been used for disease modeling and drug discovery; however, several promising studies suggest these models may be useful for investigating organ-specific developmental toxicity. For example, Calderon-Gierszal and Prins (2015) demonstrated that low-dose bisphenol A exposure can disrupt prostate organoid morphology and alter stem cell dynamics. Alternatively, Sandström *et al.* (2017) demonstrated that prototypical neurodevelopmental toxicants such as methyl mercury and trimethyltin can cause increased astroglial reactivity in neurospheres generated from hESCs.

Although methodological advances have primarily focused on using ESCs for developmental toxicity testing *in vitro*, iPSCs may allow researchers to incorporate genetic diversity (ie, sex, ethnicity, age, etc.) into their studies (reviewed in Warren and Cowan, 2017). Indeed, several large cohorts of genetically diverse iPSCs have been derived, and are publicly available (Panopoulos *et al.*, 2017; Park *et al.*, 2017). Although, these genetically diverse cohorts have yet to be utilized for developmental toxicity testing, several studies have successfully utilized iPSCs to incorporate genetic diversity into cardiotoxicity testing. For example, iPSC-derived cardiomyocytes can recapitulate the penchant of patients to doxorubicin-induced cardiotoxicity.

Despite many challenges, a great deal of progress has been made in using PSCs for developmental toxicity testing, and this trend will continue with future research efforts and discoveries. In particular, many of the advanced PSC-based assays discussed in this review require further validation and standardization. Further standardization will help reduce variability between similar methods, and increase the reproducibility between different labs and lab settings, which will be ultimately be critical to improving our ability to predict *in vivo* developmental toxicants *in vitro*.

ACKNOWLEDGMENTS

The authors thank Dr Alex Merrick and Dr William Gwinn for their careful review and insightful comments for this article.

FUNDING

This work was supported by the Division of the National Toxicology Program, National Institute of Environmental Health Sciences (NIEHS). This article may be the work product of an employee or group of employees of the NIEHS, NIH; however, the statements contained herein do not necessarily represent the statements, opinions, or conclusions of the NIEHS, NIH of the U.S. Government. The content of this publication does not necessarily reflect the views or the policies of the Department of Health and Human Services nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

Adler, S., Pellizzer, C., Hareng, L., Hartung, T., and Bremer, S. (2008). First steps in establishing a developmental toxicity test method based on human embryonic stem cells. *Toxicol. In Vitro* **22**, 200–211.

Adler, S., Pellizzer, C., Paparella, M., Hartung, T., and Bremer, S. (2006). The effects of solvents on embryonic stem cell differentiation. *Toxicol. In Vitro* **20**, 265–271.

Akopian, V., Andrews, P. W., Beil, S., Benvenisty, N., Brehm, J., Christie, M., Ford, A., Fox, V., Gokhale, P. J., Healy, L., *et al.* (2010). Comparison of defined culture systems for feeder cell free propagation of human embryonic stem cells. *In Vitro Cell. Dev. Biol Anim.* **46**, 247–258.

Augustine-Rauch, K., Zhang, C. X., and Panzica-Kelly, J. M. (2016). A developmental toxicology assay platform for screening teratogenic liability of pharmaceutical compounds. *Birth Defects Res. B Dev. Reprod. Toxicol.* **107**, 4–20.

Baek, D. H., Kim, T. G., Lim, H. K., Kang, J. W., Seong, S. K., Choi, S. E., Lim, S. Y., Park, S. H., Nam, B-h., Kim, E. H., *et al.* (2012). Embryotoxicity assessment of developmental neurotoxicants using a neuronal endpoint in the embryonic stem cell test. *J. Appl. Toxicol.* **32**, 617–626.

Bantle, J. A., Fort, D. J., and James, B. L. (1989). Identification of developmental toxicants using the frog embryo teratogenesis assay-Xenopus (FETAX). *Hydrobiologia* **188-189**, 577–585.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z. D., Ziller, M., Croft, G. F., Amoroso, M. W., Oakley, D. H., *et al.* (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144**, 439–452.

Buesen, R., Genschow, E., Slawik, B., Visan, A., Spielmann, H., Luch, A., and Seiler, A. (2009). Embryonic stem cell test remastered: Comparison between the validated EST and the new molecular FACS-EST for assessing developmental toxicity in vitro. *Toxicol. Sci.* **108**, 389–400.

Busquet, F., Nagel, R., von Landenberg, F., Mueller, S. O., Huebler, N., and Broschard, T. H. (2008). Development of a new screening assay to identify proteratogenic substances using zebrafish danio rerio embryo combined with an exogenous mammalian metabolic activation system (m Dar. T). *Toxicol. Sci.* **104**, 177–188.

Calderon-Gierszal, E. L., and Prins, G. S. (2015). Directed differentiation of human embryonic stem cells into prostate organoids in vitro and its perturbation by low-dose bisphenol A exposure. *PLoS One* **10**, e0133238.

Cezar, G. G., Quam, J. A., Smith, A. M., Rosa, G. J., Piekarczyk, M. S., Brown, J. F., Gage, F. H., and Muotri, A. R. (2007). Identification of small molecules from human embryonic stem cells using metabolomics. *Stem Cells Dev.* **16**, 869–882.

Chapin, R., Augustine-Rauch, K., Beyer, B., Daston, G., Finnell, R., Flynn, T., Hunter, S., Mirkes, P., Sue O'Shea, K., Piersma, A., *et al.* (2008). State of the art in developmental toxicity screening methods and a way forward: A meeting report addressing embryonic stem cells, whole embryo culture, and zebrafish. *Birth Defects Res. B Dev. Reprod. Toxicol.* **83**, 446–456.

Chen, R., Chen, J., Cheng, S., Qin, J., Li, W., Zhang, L., Jiao, H., Yu, X., Zhang, X., Lahn, B. T., *et al.* (2010). Assessment of embryotoxicity of compounds in cosmetics by the embryonic stem cell test. *Toxicol. Mech. Methods* **20**, 112–118.

Chen, X., Hansen, D. K., Merry, G., DeJarnette, C., Nolen, G., Sloper, D., Fisher, J. E., Harrouk, W., Tassinari, M. S., and Inselman, A. L. (2015). Developing osteoblasts as an endpoint for the mouse embryonic stem cell test. *Reprod. Toxicol.* **53**, 131–140.

Czysz, K., Minger, S., and Thomas, N. (2015). DMSO efficiently down regulates pluripotency genes in human embryonic stem cells during definitive endoderm derivation and increases the proficiency of hepatic differentiation. *PLoS One* **10**, e0117689.

Dang, S. M., Kyba, M., Perlingeiro, R., Daley, G. Q., and Zandstra, P. W. (2002). Efficiency of embryoid body formation and hematopoietic development from embryonic stem cells in different culture systems. *Biotechnol. Bioeng.* **78**, 442–453.

de Jong, E., Louisse, J., Verwei, M., Blaauboer, B. J., van de Sandt, J. J., Woutersen, R. A., Rietjens, I. M., and Piersma, A. H. (2009). Relative developmental toxicity of glycol ether alkoxy acid metabolites in the embryonic stem cell test as compared with the in vivo potency of their parent compounds. *Toxicol. Sci.* **110**, 117–124.

Desai, N., Rambhia, P., and Gishto, A. (2015). Human embryonic stem cell cultivation: Historical perspective and evolution of xeno-free culture systems. *Reprod. Biol. Endocrinol.* **13**, 9.

Di Guglielmo, C., Lopez, D. R., De Lapuente, J., Mallafre, J. M., and Suarez, M. B. (2010). Embryotoxicity of cobalt ferrite and gold nanoparticles: A first in vitro approach. *Reprod. Toxicol. (Elmsford, N.Y.)* **30**, 271–276.

Dimopoulou, M., Verhoef, A., Gomes, C. A., van Dongen, C. W., Rietjens, I. M., Piersma, A. H., and van Ravenzwaay, B. (2018). A comparison of the embryonic stem cell test and whole embryo culture assay combined with the BeWo placental passage model for predicting the embryotoxicity of azoles. *Toxicol. Lett.* **286**, 10–21.

Dye, B. R., Hill, D. R., Ferguson, M. A. H., Tsai, Y.-H., Nagy, M. S., Dyal, R., Wells, J. M., Mayhew, C. N., Nattiv, R., Klein, O. D., *et al.* (2015). In vitro generation of human pluripotent stem cell derived lung organoids. *Elife* **4**, e05098.

Eckardt, K., and Stahlmann, R. (2010). Use of two validated in vitro tests to assess the embryotoxic potential of mycophenolic acid. *Arch. Toxicol.* **84**, 37–43.

Flamier, A., Singh, S., and Rasmussen, T. P. (2017). A standardized human embryoid body platform for the detection and analysis of teratogens. *PLoS One* **12**, e0171101.

Fort, D. J., James, B. L., and Bantle, J. A. (1989). Evaluation of the developmental toxicity of five compounds with the frog embryo teratogenesis assay: Xenopus (FETAX) and a metabolic activation system. *J. Appl. Toxicol.* **9**, 377–388.

Freedman, B. S., Brooks, C. R., Lam, A. Q., Fu, H., Morizane, R., Agrawal, V., Saad, A. F., Li, M. K., Hughes, M. R., Werff, R. V., *et al.* (2015). Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. *Nat. Commun.* **6**,

Gamage, T. K., Chamley, L. W., and James, J. L. (2016). Stem cell insights into human trophoblast lineage differentiation. *Hum. Reprod. Update* **23**, 77–103.

Genschow, E., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., and Paillard, F. (2002). The ECVAM international validation study on in vitro embryotoxicity tests: Results of the definitive phase and evaluation of prediction models. *Altern Lab Anim* **30**, 151–176.

Hohwieler, M., Illing, A., Hermann, P. C., Mayer, T., Stockmann, M., Perkhofer, L., Eiseler, T., Antony, J. S., Müller, M., Renz, S., *et al.* (2017). Human pluripotent stem cell-derived acinar/ductal organoids generate human pancreas upon orthotopic transplantation and allow disease modelling. *Gut* **66**, 473–486.

International Stem Cell Initiative. (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. *Nat. Biotechnol.* **29**, 1132–1144.

Iwatani, M., Ikegami, K., Kremenska, Y., Hattori, N., Tanaka, S., Yagi, S., and Shiota, K. (2006). Dimethyl sulfoxide has an impact on epigenetic profile in mouse embryoid body. *Stem Cells* **24**, 2549–2556.

Jagtap, S., Meganathan, K., Gaspar, J., Wagh, V., Winkler, J., Hescheler, J., and Sachinidis, A. (2011). Cytosine arabinoside induces ectoderm and inhibits mesoderm expression in human embryonic stem cells during multilineage differentiation. *Br. J. Pharmacol.* **162**, 1743–1756.

Jennings, P. (2015). The future of in vitro toxicology. *Toxicol. In Vitro* **29**, 1217–1221.

Jo, J., Xiao, Y., Sun, A. X., Cukuroglu, E., Tran, H.-D., Göke, J., Tan, Z. Y., Saw, T. Y., Tan, C.-P., Lokman, H., *et al.* (2016). Midbrain-

like organoids from human pluripotent stem cells contain functional dopaminergic and neuromelanin-producing neurons. *Cell Stem Cell* **19**, 248–257.

Judson, R., Richard, A., Dix, D. J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., *et al.* (2009). The toxicity data landscape for environmental chemicals. *Environ. Health Perspect.* **117**, 685.

Kamelia, L., Louisse, J., de Haan, L., Rietjens, I., and Boogaard, P. J. (2017). Prenatal developmental toxicity testing of petroleum substances: Application of the mouse embryonic stem cell test (EST) to compare in vitro potencies with potencies observed in vivo. *Toxicol. In Vitro* **44**, 303–312.

Kameoka, S., Babiarz, J., Kolaja, K., and Chiao, E. (2014). A high-throughput screen for teratogens using human pluripotent stem cells. *Toxicol. Sci.* **137**, 76–90.

Kang, H. Y., Choi, Y. K., Jo, N. R., Lee, J. H., Ahn, C., Ahn, I. Y., Kim, T. S., Kim, K. S., Choi, K. C., Lee, J. K., *et al.* (2017). Advanced developmental toxicity test method based on embryoid body's area. *Reprod. Toxicol.* **72**, 74–85.

Khalid, O., Kim, J. J., Kim, H.-S., Hoang, M., Tu, T. G., Elie, O., Lee, C., Vu, C., Horvath, S., Spigelman, I., *et al.* (2014). Gene expression signatures affected by alcohol-induced DNA methylomic deregulation in human embryonic stem cells. *Stem Cell Res.* **12**, 791–806.

Kleinstreuer, N. C., Smith, A. M., West, P. R., Conard, K. R., Fontaine, B. R., Weir-Hauptman, A. M., Palmer, J. A., Knudsen, T. B., Dix, D. J., Donley, E. L. R., *et al.* (2011). Identifying developmental toxicity pathways for a subset of ToxCast chemicals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol.* **257**, 111–121.

Kong, D., Xing, L., Liu, R., Jiang, J., Wang, W., Shang, L., Wei, X., and Hao, W. (2013). Individual and combined developmental toxicity assessment of bisphenol A and genistein using the embryonic stem cell test in vitro. *Food Chem. Toxicol.* **60**, 497–505.

Krug, A. K., Kolde, R., Gaspar, J. A., Rempel, E., Balmer, N. V., Meganathan, K., Vojnits, K., Baquié, M., Waldmann, T., Ensenat-Waser, R., *et al.* (2013). Human embryonic stem cell-derived test systems for developmental neurotoxicity: A transcriptomics approach. *Arch. Toxicol.* **87**, 123–143.

Kugler, J., Huhse, B., Tralau, T., and Luch, A. (2017). Embryonic stem cells and the next generation of developmental toxicity testing. *Exp. Opin. Drug Metab. Toxicol.* **13**, 833–841.

Kugler, J., Kemler, R., Luch, A., and Oelgeschlager, M. (2016). Editor's highlight: Identification and characterization of teratogenic chemicals using embryonic stem cells isolated from a Wnt/beta-catenin-reporter transgenic mouse line. *Toxicol. Sci.* **152**, 382–394.

Kugler, J., Tharmann, J., Chuva de Sousa Lopes, S. M., Kemler, R., Luch, A., and Oelgeschlager, M. (2015). A Bmp reporter transgene mouse embryonic stem cell model as a tool to identify and characterize chemical teratogens. *Toxicol. Sci.* **146**, 374–385.

Le Coz, F., Suzuki, N., Nagahori, H., Omori, T., and Saito, K. (2015). Hand1-Luc embryonic stem cell test (Hand1-Luc EST): A novel rapid and highly reproducible in vitro test for embryotoxicity by measuring cytotoxicity and differentiation toxicity using engineered mouse ES cells. *J. Toxicol. Sci.* **40**, 251–261.

Leslie, J. L., Huang, S., Opp, J. S., Nagy, M. S., Kobayashi, M., Young, V. B., and Spence, J. R. (2015). Persistence and toxin production by Clostridium difficile within human intestinal organoids result in disruption of epithelial paracellular barrier function. *Infect. Immun.* **83**, 138–145.

Liang, G., and Zhang, Y. (2013). Genetic and epigenetic variations in iPSCs: Potential causes and implications for application. *Cell Stem Cell* **13**, 149–159.

Liu, S., Yin, N., and Faiola, F. (2017). Prospects and frontiers of stem cell toxicology. *Stem Cells Dev.* **26**, 1528–1539.

Loebel, D. A., Watson, C. M., De Young, R. A., and Tam, P. P. (2003). Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev. Biol.* **264**, 1–14.

McCracken, K. W., Catá, E. M., Crawford, C. M., Sinagoga, K. L., Schumacher, M., Rockich, B. E., Tsai, Y.-H., Mayhew, C. N., Spence, J. R., Zavros, Y., *et al.* (2014). Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. *Nature* **516**, 400–404.

Meganathan, K., Jagtap, S., Wagh, V., Winkler, J., Gaspar, J. A., Hildebrand, D., Trusch, M., Lehmann, K., Hescheler, J., Schluter, H., *et al.* (2012). Identification of thalidomide-specific transcriptomics and proteomics signatures during differentiation of human embryonic stem cells. *PLoS One* **7**, e44228.

Merkle, F. T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M., *et al.* (2017). Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. *Nature* **545**, 229–233.

Moon, S.-H., Ju, J., Park, S.-J., Bae, D., Chung, H.-M., and Lee, S.-H. (2014). Optimizing human embryonic stem cells differentiation efficiency by screening size-tunable homogenous embryoid bodies. *Biomaterials* **35**, 5987–5997.

Nagahori, H., Suzuki, N., Le Coz, F., Omori, T., and Saito, K. (2016). Prediction of in vivo developmental toxicity by combination of Hand1-Luc embryonic stem cell test and metabolic stability test with clarification of metabolically inapplicable candidates. *Toxicol. Lett.* **259**, 44–51.

Nash, R., Krishnamoorthy, M., Jenkins, A., and Csete, M. (2012). Human embryonic stem cell model of ethanol-mediated early developmental toxicity. *Exp. Neurol.* **234**, 127–135.

Nath, S. C., Horie, M., Nagamori, E., and Kino-oka, M. (2017). Size- and time-dependent growth properties of human induced pluripotent stem cells in the culture of single aggregate. *Journal of Bioscience and Bioengineering*

Pal, R., Mamidi, M. K., Das, A. K., and Bhonde, R. (2012). Diverse effects of dimethyl sulfoxide (DMSO) on the differentiation potential of human embryonic stem cells. *Arch. Toxicol.* **86**, 651–661.

Palmer, J. A., Smith, A. M., Egnash, L. A., Colwell, M. R., Donley, E. L. R., Kirchner, F. R., and Burrier, R. E. (2017). 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.* **73**, 350–361.

Palmer, J. A., Smith, A. M., Egnash, L. A., Conard, K. R., West, P. R., Burrier, R. E., Donley, E. L., and Kirchner, F. R. (2013). Establishment and assessment of a new human embryonic stem cell-based biomarker assay for developmental toxicity screening. *Birth Defects Res. B Dev. Reprod. Toxicol.* **98**, 343–363.

Panopoulos, A. D., D'Antonio, M., Benaglio, P., Williams, R., Hashem, S. I., Schuldt, B. M., DeBoever, C., Arias, A. D., Garcia, M., Nelson, B. C., *et al.* (2017). iPSCORE: A resource of 222 iPSC lines enabling functional characterization of genetic variation across a variety of cell types. *Stem Cell Rep.* **8**, 1086–1100.

Panzica-Kelly, J. M., Brannen, K. C., Ma, Y., Zhang, C. X., Flint, O. P., Lehman-McKeeman, L. D., and Augustine-Rauch, K. A. (2013). Establishment of a molecular embryonic stem cell developmental toxicity assay. *Toxicol. Sci.* **131**, 447–457.

Paquette, J. A., Kumpf, S. W., Streck, R. D., Thomson, J. J., Chapin, R. E., and Stedman, D. B. (2008). Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. *Birth Defects Res. B Dev. Reprod. Toxicol.* **83**, 104–111.

Park, M. V. D. Z., Annema, W., Salvati, A., Lesniak, A., Elsaesser, A., Barnes, C., McKerr, G., Howard, C. V., Lynch, I., Dawson, K A., *et al.* (2009). In vitro developmental toxicity test detects inhibition of stem cell differentiation by silica nanoparticles. *Toxicol. Appl. Pharmacol.* **240**, 108–116.

Park, S., Gianotti-Sommer, A., Molina-Estevez, F. J., Vanuytsel, K., Skvir, N., Leung, A., Rozelle, S. S., Shaikho, E. M., Weir, I., Jiang, Z., *et al.* (2017). A comprehensive, ethnically diverse library of sickle cell disease-specific induced Pluripotent stem cells. *Stem Cell Rep.* **8**, 1076–1085.

Peters, A. K., Steemans, M., Hansen, E., Mesens, N., Verheyen, G. R., and Vanparys, P. (2008). Evaluation of the embryotoxic potency of compounds in a newly revised high throughput embryonic stem cell test. *Toxicol. Sci.* **105**, 342–350.

Pettinato, G., Wen, X., and Zhang, N. (2015). Engineering strategies for the formation of embryoid bodies from human pluripotent stem cells. *Stem Cells Dev.* **24**, 1595–1609.

Piersma, A. H., Bosgra, S., van Duursen, M. B. M., Hermsen, S. A. B., Jonker, L. R. A., Kroese, E. D., van der Linden, S. C., Man, H., Roelofs, M. J. E., Schulpen, S. H. W., *et al.* (2013). Evaluation of an alternative in vitro test battery for detecting reproductive toxicants. *Reprod. Toxicol.* **38**, 53–64.

Ravi, M., Paramesh, V., Kaviya, S., Anuradha, E., and Solomon, F. (2015). 3D cell culture systems: Advantages and applications. *J. Cell. Physiol.* **230**, 16–26.

Sandström, J., Eggermann, E., Charvet, I., Roux, A., Toni, N., Greggio, C., Broyer, A., Monnet-Tschudi, F., and Stoppini, L. (2017). Development and characterization of a human embryonic stem cell-derived 3D neural tissue model for neurotoxicity testing. *Toxicol. In Vitro* **38**, 124–135.

Schenk, B., Weimer, M., Bremer, S., van der Burg, B., Cortvrindt, R., Freyberger, A., Lazzari, G., Pellizzer, C., Piersma, A., and Schäfer, W. R. (2010). The ReProTect Feasibility Study, a novel comprehensive in vitro approach to detect reproductive toxicants. *Reprod. Toxicol.* **30**, 200–218.

Shinde, V., Klima, S., Sureshkumar, P. S., Meganathan, K., Jagtap, S., Rempel, E., Rahnenführer, J., Hengstler, J. G., Waldmann, T., and Hescheler, J. (2015). Human pluripotent stem cell based developmental toxicity assays for chemical safety screening and systems biology data generation. *J. Vis. Exp.* **100**, e52333–e52333.

Stillerman, K. P., Mattison, D. R., Giudice, L. C., and Woodruff, T. J. (2008). Environmental exposures and adverse pregnancy outcomes: A review of the science. *Reprod. Sci.* **15**, 631–650.

Stummann, T. C., Hareng, L., and Bremer, S. (2008). Embryotoxicity hazard assessment of cadmium and arsenic compounds using embryonic stem cells. *Toxicology* **252**, 118–122.

Stummann, T. C., Hareng, L., and Bremer, S. (2007). Embryotoxicity hazard assessment of methylmercury and chromium using embryonic stem cells. *Toxicology* **242**, 130–143.

Stummann, T. C., Hareng, L., and Bremer, S. (2009). Hazard assessment of methylmercury toxicity to neuronal induction in embryogenesis using human embryonic stem cells. *Toxicology* **257**, 117–126.

Suzuki, N., Ando, S., Yamashita, N., Horie, N., and Saito, K. (2011). Evaluation of novel high-throughput embryonic stem cell tests with new molecular markers for screening embryotoxic chemicals in vitro. *Toxicol. Sci.* **124**, 460–471.

Takasato, M., Er, P. X., Chiu, H. S., Maier, B., Baillie, G. J., Ferguson, C., Parton, R. G., Wolvetang, E. J., Roost, M. S., Chuva de Sousa Lopes, S. M., *et al.* (2015). Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. *Nature* **526**, 564–568.

Takebe, T., Sekine, K., Enomura, M., Koike, H., Kimura, M., Ogaeri, T., Zhang, R.-R., Ueno, Y., Zheng, Y.-W., Koike, N., *et al.* (2013). Vascularized and functional human liver from an iPSC-derived organ bud transplant. *Nature* **499**, 481–484.

Theunissen, P., Schulpen, S., Van Dartel, D., Hermsen, S., van Schooten, F., and Piersma, A. (2010). An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. *Reprod.ctive Toxicol.* **29**, 383–392.

Trosko, J. E. (2018). Mechanistic based 3-dimensional use of human adult stem cells in toxicology. *Toxicol. Sci.*

Tsankov, A. M., Akopian, V., Pop, R., Chetty, S., Gifford, C. A., Daheron, L., Tsankova, N. M., and Meissner, A. (2015). A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. *Nat. Biotechnol.* **33**, 1182.

Uibel, F., Muhleisen, A., Kohle, C., Weimer, M., Stummann, T. C., Bremer, S., and Schwarz, M. (2010). ReProGlo: A new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. *Reprod. Toxicol.* **30**, 103–112.

Uibel, F., and Schwarz, M. (2015). Prediction of embryotoxic potential using the ReProGlo stem cell-based Wnt reporter assay. *Reprod. Toxicol.* **55**, 30–49.

van Dartel, D. A., and Piersma, A. H. (2011). The embryonic stem cell test combined with toxicogenomics as an alternative testing model for the assessment of developmental toxicity. *Reprod. Toxicol.* **32**, 235–244.

Warkus, E. L., and Marikawa, Y. (2017). Exposure-based validation of an in vitro gastrulation model for developmental toxicity assays. *Toxicol. Sci.* **157**, 235–245.

Warren, C. R., and Cowan, C. A. (2017). Humanity in a dish: Population genetics with iPSCs. *Trends Cell Biol.* **28**(1), 46–57.

Weinhold, B. (2009). Environmental factors in birth defects: What we need to know. *Environ. Health Perspect.* **117**, A440–A447.

Weitzer, W. (2008). Embryonic stem cell-derived embryoid bodies: An in vitro model of eutherian pregastrulation development and early gastrulation. In *Stem Cells*, pp. 21–51. Springer, Berlin, Heidelberg.

West, P. R., Weir, A. M., Smith, A. M., Donley, E. L., and Cezar, G. G. (2010). Predicting human developmental toxicity of pharmaceuticals using human embryonic stem cells and metabolomics. *Toxicol. Appl. Pharmacol.* **247**, 18–27.

Zhou, R., Cheng, W., Feng, Y., Wei, H., Liang, F., and Wang, Y. (2017). Interactions between three typical endocrine-disrupting chemicals (EDCs) in binary mixtures exposure on myocardial differentiation of mouse embryonic stem cell. *Chemosphere* **178**, 378–383.

zur Nieden, N. I., Davis, L. A., and Rancourt, D. E. (2010a). Comparing three novel endpoints for developmental osteotoxicity in the embryonic stem cell test. *Toxicol. Appl. Pharmacol.* **247**, 91–97.

zur Nieden, N. I., Davis, L. A., and Rancourt, D. E. (2010b). Monolayer cultivation of osteoprogenitors shortens duration of the embryonic stem cell test while reliably predicting developmental osteotoxicity. *Toxicology* **277**, 66–73.