Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Syst Biol Reprod Med. Author manuscript; available in PMC 2018 August 03.

Published in final edited form as:

Syst Biol Reprod Med. 2012 February ; 58(1): 10–22. doi:10.3109/19396368.2011.648302.

Alternative models in developmental toxicology

Hyung-yul Lee[1], **Amy L. Inselman**[1], **Jyotshnabala Kanungo**[2], and **Deborah K. Hansen**[1,*]

[1]Division of Personalized Nutrition and Medicine, FDA/National Center for Toxicological Research, Jefferson, AR, USA

[2]Division of Neurotoxicology, FDA/National Center for Toxicological Research, Jefferson, AR, USA

Abstract

In light of various pressures, toxicologists have been searching for alternative methods for safety testing of chemicals. According to a recent policy in the European Union (Regulation, Evaluation Authorisation and Restriction of Chemicals, REACH), it has been estimated that over the next twelve to fifteen years, approximately 30,000 chemicals may need to be tested for safety, and under current guidelines such testing would require the use of approximately 7.2 million laboratory animals [Hofer et al. 2004]. It has also been estimated that over 80% of all animals used for safety testing under REACH legislation would be used for examining reproductive and developmental toxicity [Hofer et al., 2004]. In addition to REACH initiatives, it has been estimated that out of 5,000 to 10,000 new drug entities that a pharmaceutical company may start with, only one is finally approved by the Food and Drug Administration at a cost of over one billion dollars [Garg et al. 2011]. A large portion of this cost is due to animal testing. Therefore, both the pharmaceutical and chemical industries are interested in using alternative models and *in vitro* tests for safety testing. This review will examine the current state of three alternative models - whole embryo culture (WEC), the mouse embryonic stem cell test (mEST), and zebrafish. Each of these alternatives will be reviewed, and advantages and disadvantages of each model will be discussed. These models were chosen because they are the models most commonly used and would appear to have the greatest potential for future applications in developmental toxicity screening and testing.

Keywords

alternative models; developmental toxicology; embryonic stem cells; *in vitro*; whole embryo culture; zebrafish

Introduction

Experimental paradigms for developmental toxicity testing have changed little over the last 40 years and rely entirely on the use of *in vivo* animal models as guidance for human developmental toxicity. Animal experiments require large numbers of animals and are

*Address correspondence to Deborah K. Hansen, Ph.D., Acting Director, Division of Personalized Nutrition and Medicine, FDA/NCTR, 3900 NCTR Rd., Jefferson, AR USA 72079. deborah.hansen@fda.hhs.gov.

Disclaimer: The views expressed are those of the authors and do not reflect the position of the United States Food and Drug Administration.

Declaration of interest: The authors acknowledge funding for the cited WEC and the mEST studies from FDA/NCTR.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

complex, very costly, and time-consuming. Recently, the need for new approaches was discussed in “Toxicity Testing in the 21st Century: A Vision and a Strategy” [National Research Council 2007]. In this report, the use of *in vitro* systems and pathway analysis was advocated; this would permit more efficient and high-throughput screening methods to detect developmental toxicants.

Since the 1980s the three R’s, reduction, refinement, and replacement have become more central to scientific discussion regarding the humane use of animals. In 1993, the European Union established the European Center for the Validation of Alternative Methods (ECVAM); in 1997, the United States government established the Interagency Coordinating Center for the Validation of Alternative Methods (ICCVAM). More recently, JaCVAM and KoCVAM were formed in Japan and Korea, respectively. The primary focus of these groups has been on reducing the number of animals used for safety testing, and they have supported the development and validation of alternative tests [Spielmann et al. 2008]. A full validation study of three *in vitro* assays was funded by ECVAM and conducted in four laboratories [Genschow et al. 2002]; the assays included the rat limb bud micromass assay, rat WEC, and the mEST. A total of 20 chemicals and a positive control were tested; chemicals were correctly categorized in 78% of experiments using the mEST, 80% in WEC, and 70% in the rat limb bud micromass assay.

The role of *in vitro* assays as screens or as replacements for animal testing is still being determined. Currently, these assays are used to screen compounds for further development as drug candidates or to prioritize chemicals for animal testing; they may also be used for mechanistic studies. However, some believe that these tests should replace animals, particularly the second species required for evaluation of embryo/fetal development. The purpose of this review is to describe three assays that currently appear to have the greatest utility as screens; these assays are rodent WEC, the mEST, and zebrafish. Advantages and disadvantages of each assay, along with a recent comparison of the three assays, will be presented [de Jong et al. 2011]. Finally, the topic of the ‘gold standard’ list of developmental toxicants for validation efforts will be briefly discussed.

Whole Embryo Culture

Rodent whole embryo culture (WEC) as used today was introduced by Denis New in 1978 [New 1978]. This was one of three *in vitro* tests for developmental toxicity that was validated by ECVAM [Genschow et al. 2002]. Briefly a mouse or rat embryo is removed from its dam and cultured for up to 48 hours during early organogenesis. The culture medium is primarily rat serum, and compounds and/or metabolites can be added to the serum during the culture period. Although the culture period can vary, generally the embryos are cultured for no longer than 48 hours. The endpoints usually examined include viability as determined by heart beat and yolk sac circulation, growth as measured by crown-rump length and/or protein content and overall development as scored by the total morphological score. A scoring system was developed by Brown and Fabro [1981] that is used by most who culture embryos; this morphological score evaluates the development of up to 17 different parameters. Since this scoring system is time consuming and requires in-depth

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

knowledge of morphology, simplified alternative scoring systems have been proposed [Klug et al. 1985; Van Maele-Fabry et al. 1990].

In the ECVAM validation study, two prediction models were developed [Genschow et al.. 2002]. For the first prediction model, the endpoints used included the total morphological score which was a modification of the original Brown and Fabro [1981] scoring system and the numbers and types of malformations. Using this prediction model, rat WEC was able to correctly categorize 68% of the 20 ECVAM chemicals with the non- and weakly embryotoxic categories being the most problematic. For the second prediction model, cytotoxicity measured with a mouse 3T3 fibroblast line was included, and the overall accuracy increased to 80%. Differences in the results between the four laboratories involved in the validation study were apparent with WEC; this difference was attributed to the use of a different rat strain in one of the laboratories [Genschow et al. 2002].

Recent studies have shown that by incorporating gene expression profiling, the predictive ability of the WEC model may be improved. In an initial experiment, retinoic acid (RA) was shown to induce effects on gene expression 4 hours following exposure. Multiple genes were identified that had previously been associated with *in vivo* changes observed upon RA-treatment [Luijten et al. 2010]. In a second study, transcriptomic responses to caffeine, methyl-mercury, monobutyl phthalate, and methoxyacetic acid were studied 4 hours after treatment, this was alongside morphological developmental assessments [Robinson et al. 2010]. Despite the induction of common morphological effects, little overlap was found at the level of gene expression resulting in different biological processes being affected. These results demonstrate that distinct chemically induced responses in gene expression precede the morphological effects observed in WEC and that these responses in gene expression are relevant to mechanisms observed *in vivo*.

WEC has a number of advantages, the greatest of which is the use of an intact embryo so that individual embryonic components can interact with each other as they would *in vivo*. Additionally, development over this 48 hour period *in vitro* parallels *in vivo* development over the same time frame. Because there is no maternal component, one can examine the direct effect of a condition on embryonic development. However, this can also be considered a disadvantage, since a variety of changes occurring in the maternal organism may ultimately effect embryonic development. For example, serum from treated rats can be used as the serum source, so that changes induced in a whole organism by the compound can be evaluated [Klein et al. 1980; Flynn et al. 2003]. Although rat serum is commonly used as the serum source, human serum has been used by some investigators [Chatot et al. 1984; Ferrari et al. 1994; Vanaerts et al. 1994].

A strength of WEC is the ability to perform mechanistic studies. The ability of a compound to ameliorate the effects of another chemical can be examined. For example, the ability of folic acid and compounds involved in one-carbon metabolism [Hansen et al. 1995] to decrease the incidence of neural tube defects produced by valproic acid has been examined with both mouse [Hansen 1993] and rat [Hansen and Grafton 1991] embryos cultured *in vitro*. Another example involves homocysteine. This compound is increased in folate deficiency, and epidemiological studies have found an increase in plasma homocysteine

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

levels associated with neural tube defect pregnancies [Bjorke-Monson et al. 1997]. Additionally, Rosenquist et al. [1996] found that treatment of chick embryos with homocysteine thiolactone increased the incidence of neural tube defects. In order to clarify the role of homocysteine in inducing neural tube defects in a mammalian embryo, mouse embryos were treated with homocysteine thiolactone added to the culture medium; no increase in the incidence of neural tube defects was observed [Hansen et al. 2001]. Since it was possible that homocysteine in the culture medium was not taken up by the embryos, homocysteine thiolactone was directly injected into the amniotic sac surrounding the embryos in culture; again, no increase in the incidence of neural tube defects was observed [Hansen et al. 2001]. These *in vitro* results were later confirmed *in vivo* [Greene et al. 2003].

Mixtures can also be studied with this assay; it is possible to examine an individual compound as well as mixtures of several individual compounds. This approach was used by Andrews et al. [2004] to study water disinfection by-products; a similar approach was used by Kennelly et al. [1999] to examine extracts of blue cohosh. Species differences can be due to inherent differences in the response to a compound or to pharmacokinetic differences between the species. When examining species differences *in vitro*, such differences can be ascribed to inherent differences since pharmacokinetic and metabolic differences can be controlled. This approach was used to examine differences between mice and rats in response to methanol [Andrews et al. 1993] and dexamethasone [Hansen and Grafton 1994].

Recently, WEC using rabbit embryos and a morphological scoring system was developed by Carney et al. [2007]. Using this model system and toxicokinetic data generated in rats and rabbits, Carney et al. [2008] determined that species differences in ethylene glycol-induced developmental toxicity appeared to be due to toxicokinetic differences between the species and further argued that the human appears to be more similar to the rabbit than to the rat in its response to ethylene glycol. Since the rabbit is the non-rodent species usually used in pre-market guideline studies, having such a model available would be advantageous; however, this model suffers from the same disadvantages as rodent WEC.

Recently, an adapted rat hepatic microsomal activation system was evaluated as an adjunct to WEC [Luijten et al. 2008]. In the adapted system, six compounds (cyclophosphamide, valpromide, 2-acetylaminofluorene, 2-methoxyethanol, retinol, and benzo[α]pyrene) were pre-incubated with Aroclor 1254-induced Sprague Dawley rat liver microsomes for 2 h. While three of the compounds (cyclophosphamide, valpromide, and 2-acetylaminofluorene) were successfully activated the other three compounds (methoxyethanol, retinol, and benzo[α]pyrene) were not. This failure was attributed to the limited spectrum of hepatic enzymes present within the microsomes. Attempts to use S9 liver fractions which contain both microsomal enzymes in addition to cytoplasmic enzymes, however, proved toxic in the WEC [Luijten et al. 2008]. Primary hepatocyte cultures have also been used in conjunction with the EST to expand the applicability of the *in vitro* test system [Hettwer et al. 2010]. The proteratogens cyclophosphamide and valpromide were selected for use and cultured with mouse embryonic stems directly as part of the EST or after preincubation with murine hepatocytes for 6 h. Preincubation of cyclophosphamide with hepatocytes resulted in a strong decrease in ID_{50} concentration of differentiating cardiomyocytes. However, similar results were not obtained with valpromide as no inhibition of differentiation was observed.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Further analysis revealed that murine hepatocytes did not metabolize valpromide to valproic acid. However, by switching the hepatocytes to human origin dramatically affected the results as shown by conversion of the proteratogen to the teratogen underscoring the importance of interspecies variations when selecting a metabolic activating system.

Of primary concern is the very limited developmental period that is assayed by WEC. Compounds producing defects at other times during development would not be detected by this model. However, the incorporation of gene expression profiling may allow the detection of embryotoxicants that may not be detected by morphological scoring because their effects occur outside the culture window of WEC [Robinson et al. 2010]. Another disadvantage of WEC is that it fails to recapitulate the maternal-fetal interactions and isolates the embryos from maternal influences (i.e., metabolism) that *in* vivo studies provide [Chapin et al. 2008]. Additionally, the technique is very labor intensive, fairly technically demanding, and requires live animals as a source for both serum and embryos.

Mouse Embryonic Stem Cell Test (mEST)

The mEST is the only test of the three *in vitro* alternative tests validated by ECVAM that does not require live laboratory animals [Genschow et al. 2002]. Embryonic stem cells are able to self-renew as well as to differentiate into all cell types of the developing embryo. They were first derived from the inner cell mass of developing mouse blastocysts [Evans and Kaufman 1981; Martin 1981]. The basic protocol for the mEST was first described by Spielmann et al. [1997] who allowed mouse embryonic stem cells from the D3 cell line to aggregate in small drops of culture media. These aggregates, or embryoid bodies, were formed on the lid of a petri dish in hanging drops and were cultured in this manner for three days. Embryoid bodies resemble the cells of the early embryo in that they are able to differentiate into cells from each of the three lineages, ectodermal, mesodermal, and endodermal. In the mEST after formation, embryoid bodies are cultured for an additional two days in suspension culture before being transferred to a 24-well culture dish where the cells are allowed to adhere and differentiate for an additional five days, making the total culture time 10 days. The cells tend to differentiate to cardiomyocytes under these conditions, and the percentage of wells with beating cardiomyocytes are scored microscopically. This is the protocol that was utilized for the ECVAM validation with the test chemicals present for the entire 10 days of culture.

In addition to the differentiation assay described above, the mEST incorporates assays for cytotoxicity using both the D3 stem cells as well as a mouse fibroblast 3T3 line to serve as a differentiated cell type. This inclusion was an attempt to compare the sensitivity of differentiated and undifferentiated cells to cytotoxicity induced by various chemicals. Cytotoxicity was determined by the MTT assay [Mosmann 1983]; cells were plated in 96-well plates with media changes at days 3 and 5 of culture to coincide with the transfer of cells in the differentiation assay. Dose-response curves were generated for three endpoints; these endpoints are inhibition of growth and viability of D3 cells (IC_{50}D3), inhibition of growth and viability of 3T3 cells (IC_{50}3T3), and inhibition of differentiation of D3 cells into beating cardiomyocytes (ID_{50}). A prediction model was established using the concentrations

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of a chemical which caused a 50% inhibition in each of these endpoints; the prediction model classifies chemicals as non-teratogenic, weakly teratogenic, or strongly teratogenic.

With the 20 chemicals that were used in the ECVAM validation study, the assay was able to correctly predict 78% of the compounds [Genschow et al. 2002]. However, subsequent studies did not produce the same high accuracy. In a study by Marx-Stoelting et al. [2009], only 2 of 13 drugs were correctly categorized. In a study by Paquette et al. [2008], 19 Pfizer compounds that were receptor-mediated were tested with an overall accuracy in a modified version of the mEST of 53%. Compounds that were non- or weakly teratogenic scored especially poorly in this assay.

The mEST has a number of advantages. It requires no live animals and uses only commercially available cell lines. However, the use of human cells in a testing framework would decrease complete reliance on cross-species extrapolation and would allow regulators to better assess the risk to humans from various chemicals. An attempt was made to develop a test similar to the mEST using a human embryonic stem cell line (hESC; [Adler et al. 2008]). These authors used a commercially available hESC line and examined the ability of the cells to differentiate to cardiomyocytes. Gene expression was monitored at 0, 4, 10, 18, and 25 days of culture. The authors observed a progressive decrease in expression of two markers of pluripotency, an early up-regulation in a marker for mesodermal differentiation, and up-regulation of several genes involved in cardiac differentiation. The authors concluded that although this initial assay showed promise, additional research would be needed to standardize a human EST. Mehta et al. [2008] attempted to increase the predictivity of the test by examining gene expression of each of the three germ layers as the endpoint of differentiation in a single hESC line. The authors concluded that the “implementation of the EST in regulatory test guidelines would demonstrate the importance of *in vitro* assays as valuable components of the risk/hazard assessment process” [Mehta et al. 2008].

The mEST is a simple assay and has been standardized and validated by ECVAM. However, modifications to the test continue to be suggested. Various recommended changes can be found in Chapin et al. [2008] and Marx-Stoelting et al. [2009]. The recommendations include alterations in the evaluation of the endpoint as well as inclusion of other cell lineages. The scoring of beating cardiomyocytes requires experience and is subject to observer bias. Evaluation of molecular endpoints to make this endpoint more objective and more high-throughput has been suggested by several groups (reviewed in [Buesen et al. 2004]). Using semi-quantitative RT-PCR, Pellizzer et al. [2004] followed the expression changes of genes involved in cardiac differentiation. They observed that over the ten days of culture exposure to all-transretinoic acid or lithium chloride changed the expression of these genes and that the two teratogens produced different gene expression changes. They suggested that expression profiles of key genes could produce more information on the toxicological mechanism of compounds and further suggested that different gene panels be used for other differentiation endpoints. Another group [zur Nieden et al. 2001] included semi-quantitative RT-PCR for α- and β-myosin heavy chain in the mEST to make the test more quantitative and sensitive. They were able to correctly predict the teratogenicity of seven chemicals that had been used in the ECVAM validation study. Staining cells with antibodies directed against cardiac specific proteins followed by fluorescence-activated cell

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sorting analysis reduced the time needed for the assay from 10 to 7 days and made the endpoint more objective and quantitative [Seiler et al. 2004].

Studies utilizing transcriptomic-based methods, whereby gene sets were used to identify differential gene expression in response to toxicant treatment, has further reduced the assay duration and provided an improved, less subjective endpoint. Improved endpoints are necessary to facilitate implantation of the EST into regulatory testing guidelines. These studies investigated gene expression in the early phase of differentiation. Using principal component analysis a 'differentiation track' representing changes in gene expression during differentiation were defined [van Dartel et al. 2009; van Dartel et al. 2010a; van Dartel et al. 2010b; van Dartel et al. 2011a; Pennings et al. 2011]. Toxicant treatment that resulted in deviations from this 'differentiation track' were then used to predict embryotoxicity. In one study, developmental toxicity was predicted for 12 diverse well-characterized positive and negative developmental toxicants using transcriptomics 24 hours after exposure. Evaluation of beating cardiomyocytes was conducted in parallel and scored on day 10 of differentiation. By evaluating deviations in the 'differentiation track', transcriptomics was able to successfully predict 83% (10/12) and 67% (8/12) compounds using two predefined gene sets [van Dartel et al. 2011a]. The responses in gene expression have also been shown to be concentration-dependent in differentiating cultures [van Dartel et al. 2011b]. Further refinement using integrated analysis identified 52 genes that contribute significantly to the predictability of the assay [Pennings et al. 2011].

A slightly different approach was used by Barrier et al. [2011]. They developed a cell adherent assay using mESCs that does not require the formation of embryoid bodies or their transfer; this greatly simplifies cell handling and increases throughput. They also used an in-cell Western blot method to quantify levels of myosin heavy chain protein as a marker of differentiation; a cytotoxicity assay was included that quantitated DNA as well as the number of cells. This novel assay was compared to the traditional mEST employing the J1 cell line using acetic acid, bromochloroacetic acid, and 5-fluorouracil as representative non-teratogenic, weakly teratogenic, and strongly teratogenic compounds, respectively. Both assays were able to discriminate the relative potencies of these three compounds. The authors also assessed the transcriptome of the cells at the end of the differentiation protocol using microarrays and observed that a large number of genes associated with various differentiation lineages were altered during the course of the protocol. They suggested that some of these other genes could also be examined to expand the capability and predictivity of the model. The authors later utilized this ACDC assay to examine the developmental toxicity potential of 309 environmental chemicals from the ToxCast chemical library [Chandler et al. 2011]. Eighteen percent of the chemicals exhibited either cytotoxicity or an inhibition of differentiation. These data were mined using a variety of methods and provided an initial listing of metabolic and regulatory pathways that may be disturbed during embryonic stem cell differentiation.

Differentiation to a single lineage has also been viewed as a shortcoming of the mEST. zur Nieden et al. [2004] altered culture conditions to allow differentiation to osteogenic, chondrogenic, and neural cells; gene expression by PCR for each cell type was the endpoint evaluated. They assessed six chemicals from the three categories defined by ECVAM; a non-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

teratogen, penicillin G, did not alter expression of any of the genes, and 5-fluorouracil, a strong teratogen, did not demonstrate any target cell specificity. Valproic acid which produces neural tube defects in animals and humans specifically affected differentiation of neuronal cells; thalidomide appeared to particularly impair osteogenic and chondrogenic differentiation. Retinoic acid, a strong teratogen, affected multiple cell types and might induce a general dysregulation of gene expression. Phenytoin, a weak teratogen, showed differences in gene expression for all cell types at similar concentrations [zur Neiden et al., 2004]. The authors suggested that the addition of other molecular endpoints and alterations in the culture conditions to include differentiation to multiple cell types could make the mEST more objective and enhance its predictivity.

Some preliminary work from our group has suggested that differentiation to additional endpoints and the use of multiple cell lines may enhance the assay. Using culture conditions reported to drive differentiation toward neural cells [Ying et al. 2003], we have observed that over 14 days in culture the number of cells with neural processes increased dramatically (Fig. 1). Additionally, expression of the pluripotency markers, Oct4, Nanog, and Sox2 decreased from initial levels (Fig. 2), while markers of neural differentiation such as Sox1, Nestin, Pax6, and Olig2 increased over time (Fig. 3). Furthermore, preliminary data suggests that cell lines differ in their ability to differentiate (we examined the ability of two cell lines, the E14TG2a and BK4 lines, to differentiate to osteoblasts). The degree of mineralization was determined by staining the cultures with Alizarin Red S, a calcium specific stain. Under identical culture conditions, the two lines displayed differences in the degree of mineralization (Fig. 4). This suggests that different cell lines or genotypes may play a role in the overall ability of the cells to differentiate to a desired lineage. Kalter [1979] suggested that the use of multiple inbred strains of mice would test a wide spectrum of the mouse genome and give a better idea of the variety of responses of which the species was capable. However, the use of multiple inbred strains of mice to test a single chemical greatly increases the expense of such an evaluation [Inselman et al. 2011] and still does not address the issue of cross species extrapolation. A major disadvantage of including multiple cell lineages is the added expense of the assay as well as the longer culture times required for some of the additional lineages. However, while not especially high-throughput, the addition of multiple cell lines and multiple lineages remains less costly and time-consuming than studies in live animals, and also allows one to test more concentrations of a compound than can be done in live animal studies.

Zebrafish Assay

Zebrafish have been a model for developmental toxicity for only a short period of time, and currently there is no validated or common protocol in use. Several protocols have been described in the literature; these protocols vary in the length of the assay, whether or not the chorion is removed, and the evaluation of the larvae.

A protocol was described by Selderslaghs et al. [2009] in which they examined six compounds with known developmental toxicity activity. Fish were treated from 2 hours post fertilization (hpf; 0 hpf defined as 1 h after the lights were turned on) continuously until evaluation at 24, 48, 72, or 144 hpf. The chorion was not removed in this assay. Compounds

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

used in this evaluation were the teratogens all-transretionic acid, valproic acid, lithium chloride, and caffeine as well as the non-teratogens glucose and saccharin. Endpoints evaluated included embryotoxicity, presence and development, as appropriate, of somites, tail detachment, otic vesicle and otoliths, eyes, heart beat, and blood circulation; larvae were evaluated for skeletal malformations, body position, and ability to swim. Each endpoint was scored in a binomial fashion (normal = 0; abnormal or not present = 1). They observed adverse effects with treatment with each of the teratogenic compounds, and these effects were similar to those observed in live animal models. These effects generally occurred in a concentration-responsive manner. They concluded that the assay could discriminate between teratogenic and non-teratogenic compounds. However, they cautioned that with some compounds, such as lithium chloride, adverse effects were not observed until the later stages of development were reached.

Brannen et al. [2010] described a different assay. In this assay, embryos were treated from 4–6 hpf until the larvae were evaluated at 5 days post fertilization (dpf). The chorions were removed by enzymatic digestion followed by manual dissection, if necessary. A number of morphological features were evaluated, similar to those evaluated in *in vivo* test systems; these included viability, motility, cardiovascular function, pigmentation and morphology of brain, facial structure, jaw and pharyngeal arch, somites, notochord, tail, fins, heart, and intestine. They tested a total of 31 compounds with known *in vivo* developmental toxicity activity; 13 compounds were non-teratogens, and 18 were teratogenic. Several of these compounds had been used in the ECVAM validation study, while others were proprietary compounds (Bristol-Myers Squibb Company; BMS Princeton, NJ, USA). Similar to the validated models for the mEST and WEC, the authors developed a prediction model using the concentration which produced lethality in 25% of larvae at 5 dpf (LC25), and concentration producing no observed adverse effects (NOAEL). They also tested for cytotoxicity in the 3T3 mouse fibroblast line which was also used in the ECVAM validation study for rat WEC and the mEST. Using their prediction model, the authors correctly classified 87% (27/31) of the compounds; two true non-teratogens were misclassified (dimethyl phthalate and a BMS compound), while two true teratogens were also mis-classified (valproic acid and a BMS compound). The authors concluded that the assay as described was very successful in correctly identifying teratogenic and non-teratogenic compounds with low false positive (15%) and false negative (11%) rates.

Another assay was described by Hermsen et al. [2011]. The authors treated embryos beginning at the 4- to 32-cell stage to 72 hpf without removing the chorion. At the end of the treatment protocol, embryos were evaluated for the following features: movement, heart beat, blood circulation, tail detachment, somite formation, eye development, pigmentation development, mouth/jaw development, pectoral fin development, and hatching. These features were scored, and the scores were summed for a general morphological score; this process is similar to the scoring system established by Brown and Fabro [1981] for rodent WEC. Rather than developing a prediction model, these authors determined the benchmark dose (BMD) by fitting a concentration-response curve. They compared the concentration which decreased the general morphological score by 5% (BMC_{GMS}) and the benchmark concentration which increased the percentage of embryos with one or more teratogenic effects by 5% (BMC_T). They evaluated six glycol ether metabolites as well as two parent

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

compounds (ethylene glycol monomethyl ether and ethylene glycol monoethyl ether) and six triazole compounds. For the glycol ether compounds, the authors compared the BMC_{GMS} and BMC_T to the benchmark dose (BMD) for body weight and malformations from whole animal studies (Table 1). The parent compounds were the most potent teratogens from the *in vivo* data, while in the zebrafish assay they produced no adverse effects. However, metabolites of the two parent compounds did produce adverse effects in the zebrafish assay; the defects produced involved the heart, head, and tail. The results with the triazole compounds will be discussed below [de Jong et al. 2011].

Another recent zebrafish assay was described by Van den Bulck et al. [2011]. These authors treated zebrafish from the sphere stage until 96 hpf; at the conclusion of the culture, 28 morphological endpoints were evaluated. Each endpoint was given equal weight, so any abnormality caused by a compound was considered to be teratogenic. The authors examined 15 compounds from Janssen Pharmaceutica (Janssen Pharmaceutica, Beerse, Belgium), and attempted to determine the lowest observed effect concentration, the NOAEL and the lethal concentration. Only two of the 15 compounds did not produce any abnormality in the zebrafish, although both of these compounds had produced various malformations *in vivo*. Although the overall predicitivity of this assay was 75%, the false positive and false negative rates were both 40% which is extremely high. This may have been due in part to the equal weighting given each endpoint, and the authors suggested that scoring for severity and/or incidence could improve the model.

The EPA has also developed a zebrafish developmental toxicity assay, but complete results using this assay have not yet been reported. According to a review article, only 55% of the chemicals tested were concordant with *in vivo* mammalian data [Sipes et al. 2011b], but 271 chemicals were tested which is a far greater number than those that have been tested in the other publications. These chemicals were part of the 309 chemicals tested under EPA's ToxCast initiative, and they were primarily environmental chemicals.

The zebrafish model has a number of advantages, the major advantages being the use and evaluation of the whole embryo and the inclusion of all stages of development in the assay, because development progresses rapidly in zebrafish. Other advantages include an embryo that is transparent permitting clear observation of the early stages of development, high fecundity, and development that is similar to that of mammals. It is also easy to modify the genetics of the zebrafish for mechanistic studies. However, this model is not a mammalian model, and it is not clear how applicable data derived from this model will be to mammalian systems. Additionally, little is known concerning the uptake and metabolism of compounds in zebrafish; the role of the chorion as a barrier to uptake of compounds from the medium has not been resolved. The role of strain differences in sensitivity of zebrafish to compounds also has not been examined.

Van den Bulck et al. [2011] attempted to address, in part, the uptake of compounds by zebrafish. At the lowest observed effect concentration or the NOAEL, larvae were processed to quantitate the amount of compound in the larvae. This body burden was taken into account when classifying the compounds as teratogenic and/or embryotoxic. The authors set the exposure threshold at ~50 ng/larva. If adverse findings were present in larvae at body

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

burdens of less than 50 ng/larva, the compound would be classified as being teratogenic and/or embryotoxic. If adverse findings were only present if the body burden was greater than 50 ng/larva, the findings were considered to be due to compound overload, and the compound was considered non-toxic. The authors concluded that the 50 ng/larva threshold will need to be evaluated in additional work, and the threshold may need to be altered.

The recent rise in the popularity of zebrafish embryos as an alternative to animal experimentation for evaluating developmental toxicity, however, has raised questions regarding current animal welfare regulations. Specifically being discussed is the point at which zebrafish should be classified as laboratory animals for research purposes. Under REACH legislation one of the primary goals is to minimize the number of animals used in experimentation by using scientifically sound alternatives, when possible, to reduce, refine, or replace animal models. The European Food Safety Authority (EFSA) previously issued an opinion which considered fish embryos as a suitable replacement or refinement method [EFAS 2005]. This was based upon the opinion that early developmental stage embryos are likely to experience less pain, suffering, and lasting harm. A recently issued EU Directive 2010/63/EU [EU Directive 2010] seeks to harmonize animal welfare regulations across Europe. This directive states that the earliest life-stages of animal development are not protected, and therefore these stages of life are not considered within the regulatory framework for animal experimentation. This directive, however, establishes the ability to feed independently as the criterion for determining when experiments would require prior authorization. Although the definition under the directive is relatively clear, interpretation may vary widely and is dependent on the particular species or strain being investigated. A recent review of the literature which took into account factors such as yolk consumption, formation of the digestive organs, swimming patterns, and the ability to incorporate food, concluded that zebrafish embryos up to 120 hours post-fertilization could be considered as a non-animal model of research since it is not until after 120 hours post-fertilization that the embryos develop the ability to feed independently and thus should be subject to regulation [Strähle et al. 2011]. In the United States independent feeding is also used as the criterion for determining when an animal is protected. However, it is up to each individual institution's Institutional Animal Care and Use Committee (IACUC) to establish the point at which zebrafish transition from the embryonic stage to a free-living organism. However, there has been a call to develop new guidelines for zebrafish research within the United States to better promote the quality and standard of science being conducted and to ensure the highest level of animal care [Lawrence et al. 2009].

Assay Comparison

A recent publication compared these three assays to each other and to the results of *in vivo* tests [de Jong et al. 2011]. The authors used the validated rat WEC and mEST protocols along with the zebrafish assay as conducted by Hermsen et al. [2011]. The chemicals tested included six triazoles for which the *in vivo* developmental toxicity activity was available through the ToxRefDB database from the EPA [http://www.epa.gov/ncct/toxrefdb]. The authors calculated benchmark concentrations decreasing the total morphological score by 5% in rat WEC, the general morphological score by 5% in zebrafish, and the incidence of beating cardiomyocytes by 50% in the mEST; these values were then compared to the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

developmental lowest effect level (dLEL) from *in vivo* studies. When the potencies of the six compounds were compared to each other using these four model systems, identical relative potencies were obtained from *in vivo* studies and the zebrafish assay (Table 2). The authors chose skeletal variations as the *in vivo* endpoint for their comparisons because it was reported to be the most sensitive endpoint; however, the choice of this endpoint compared to other possible endpoints could be questioned.

Disadvantages of these assays as well as other *in vitro* assays involve solubility of the compounds, the lack of metabolic capability, as well as the static nature of the cultures. For these assays, the compounds need to be soluble in aqueous media. Small amounts of ethanol or DMSO can be added to the cultures to increase solubility; Paquette et al. [2008] used 0.25% (v/v) for the upper limit of DMSO and 0.5% (v/v) for 70% ethanol in the mEST; higher concentrations were toxic to the cells. In the zebrafish assays, DMSO is the most commonly used solvent; concentrations used were 0.5% [Brannen et al. 2010] or 0.2% [Hermsen et al. 2011]. Selderslaghs et al. [2009] tested various concentrations of DMSO and found that the no observed effect level was 0.25%. EVAM guidelines for final concentrations of solvents for the mEST are 1% for phosphate-buffered saline, water or DMEM medium, 0.25% for DMSO, and 0.5% for ethanol [Seiler and Spielmann 2011]; guidelines for the WEC are 0.125% DMSO and 0.2% ethanol [http://ecvam-dbalm.jrc.ec.europa.eu]. This can limit the ability to test some compounds. Also, the chemicals chosen for the ECVAM validation study were specifically chosen because they did not require metabolic activation [Brown 2002]. If available, metabolites can be directly tested. However, these are not always available, may be available only in very limited amounts, and due to their transient existence, very reactive intermediates cannot be examined. Some initial efforts were made to include metabolic capability with rodent WEC using either hepatocytes [Oglesby et al. 1986; Piersma et al. 1991] or microsomal fractions [Schmid et al. 1981; Kitchin and Ebron 1983]. Alternatively, some authors treated embryos with cytochrome P-450 inducers in an effort to increase metabolic capability of the cultures [Juchau et al. 1985a; Juchau et al. 1985b].

Finally, pharmacokinetic considerations should be included in these assays. For example, a compound may not be present in the animal at the same concentration over a 2 to 10 day period due to metabolism and excretion. However, generally in the *in vitro* assays, the compound remains in the culture at the same concentration during the entire culture period. When such *in vivo* pharmacokinetic parameters are known, such information should be taken into account in the *in vitro* assay system.

Which Compounds Should be Tested?

Several attempts have been made to describe a list of compounds that could be utilized as test compounds for *in vitro* assays; the results of *in vivo* assays for these compounds should be known. However, the lists of test compounds designed previously have been criticized. The first attempt was the ‘Smith list’ published in 1983 [Smith et al. 1983]. This list contained 47 compounds; however, the criticism was that many of the negative compounds in the list would be non-toxic under all circumstances. A workshop was held in 1991 to address this issue and to establish a new list of test chemicals. However, after numerous discussions, the group could not agree on a list [Schwetz 1992].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

When ECVAM began their validation study of WEC, the mEST and the rat limb micromass culture models, they started with a database containing 309 chemicals that had either been used in previous validation studies or chemicals with good *in vivo* data or human data [Brown 2002]. This original database was decreased to about 30 chemicals with 20 being selected for use in the ECVAM validation. The chemicals were classified into three classes: 1) strongly embryotoxic chemicals which were developmental toxicants in all species tested; 2) non-embryotoxic chemicals which were not embryotoxic at maternally toxic doses or produced minor embryotoxicity which could not be dissociated from maternal toxicity; and 3) weakly embryotoxic chemicals which were chemicals of intermediate activity. The three *in vitro* tests examined in the ECVAM validation study demonstrated an overall accuracy of 70–80% in correctly classifying these 20 chemicals.

In 2009, the Health and Environmental Sciences Institute (HESI) of the International Life Sciences Institute formed a steering committee to revisit the issue of a consensus list of developmental toxicants. This committee concluded that rather than agreeing on a list of developmental toxic chemicals, a list of developmental toxicant exposures should be developed [Daston et al. 2010]. During their deliberations, they noted that one dose of a compound may be developmentally toxic, while another dose of the same compound is non-toxic; hence their agreement on developmentally toxic exposures rather than chemicals. This approach only considers permanent effects, so a dose of a chemical that causes a decrease in fetal weight may not be considered a developmentally toxic exposure. This approach also does not include a comparison of embryonic to adult sensitivity. Lastly, this approach does require knowledge of internal dose; this information may not be available for all *in vivo* studies. At this time, the committee has not released their list of developmental toxicant exposures.

Conclusions

The *in vitro* assays discussed above as well as numerous other assays have great potential to be used as screening assays to prioritize compounds for further development or testing. They are currently being used in this capacity in industry, and their use should only increase. These assays all have advantages and disadvantages, and no one assay will address all issues. Additionally, all of these assays need further work to increase their utility.

Rather than a single assay, a battery of tests may be a more rewarding approach to *in vitro* assessment of reproductive and developmental toxicity. Such a battery approach is being developed for ReProTect and ToxCast. Among the fourteen assays used to examine a number of different steps in reproductive development, the ReProTect battery also includes the WEC, the mEST, and a ReProGlo assay for developmental toxicology [Schenk et al. 2010]. The ReProGlo assay examines expression of a luciferase reporter gene for the Wnt/β-catenin signaling pathway in mouse embryonic stem cells [Uibel et al. 2010]. Using these three assays for developmental toxicity, ten chemicals were correctly identified as either developmental toxicants, non-developmental toxicants, or developmental toxicants depending upon the route of application [Schenk et al. 2010]. The EPA has utilized a large number of high-throughput assays to examine the toxicity of 309 chemicals. They recently examined the associations between the results of these high-throughput assays (many of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

which are cell-free or cell based assays) and *in vivo* developmental toxicity data from rat and rabbit studies from the ToxRefDB database [Sipes et al. 2011a]. The authors developed species-specific models for developmental toxicity that had a balanced accuracy of over 70%, suggesting that a large battery of tests, many of which have nothing to do with embryonic development, can be used for developing pathway-based models that can predict developmental toxicity.

There is increasing pressure to either greatly diminish or to eliminate safety testing in live animals; this will move *in vitro* assays from screening assays to replacements for animal testing. As mentioned above, none of the reviewed assays are ready to be utilized as a replacement for animal studies. As was written fifteen years ago, “Clearly these tests cannot meet the standard of replacement there are too many limitations, the most significant of which is the failure to adequately represent the full scope of developmental complexity, for these tests to ever replace *in vivo* mammalian testing” [Daston 1996]. Since that time, the assays being discussed have changed, but the limitations remain the same. In the meantime, work should and will continue to improve *in vitro* assays with the hope that someday an assay(s) may be found that will be suitable to replace whole animal testing. Although the pace of investigation on the use of alternative models has increased, much work remains before these tests can be used as replacements for current animal testing.

Abbreviations

MTT	3-(4,5-dimethylthiazol-2yl)-2,5-diphenyltetrazolium bromide
mEST	mouse embryonic stem cell test
WEC	whole embryo culture
ECVAM	European Centre for the Validation of Alternative Methods
hpf	hours post fertilization
dpf	days post fertilization
NOAEL	no observed adverse effect level
LC25	25% lethal concentration
BMS	Bristol-Myers Squibb
BMD	benchmark dose
BMC_{GMS}	benchmark concentration which decreased general morphological score by 5%
BMC_{T}	benchmark concentration which increased the percentage of embryos with one or more teratogenic effects by 5%
dLEL	developmental lowest effect level
ILSI	International Life Sciences Institute

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

HESI Health and Environmental Sciences Institute

References

http://ecvam-dbalm.jrc.ec.europa.eu

http://www.epa.gov/ncct/toxrefdb

Adler S, Pellizzer C, Hareng L, Hartung T, Bremer S. First steps in establishing a developmental toxicity test method based on human embryonic stem cells. Toxicol in vitro. 2008; 22:200–211. [PubMed: 17961973]

Andrews JE, Ebron-McCoy M, Logsdon TR, Mole LM, Kavlock RJ, Rogers JM. Deveopmental toxicity of methanol in whole embryo culture: A comparative study with mouse and rat embryos. Toxicology. 1993; 81:205–215. [PubMed: 8212026]

Andrews JE, Nichols HP, Schmid JE, Mole LM, Hunter ES 3rd, Klinefelter GR. Developmental toxicity of mixtures: The water disinfection by-products dichloro-, dibromo- and bro-mochloror acetic acid in rat embryo culture. Reprod Toxicol. 2004; 19:111–116. [PubMed: 15336719]

Barrier M, Jeffay S, Nichols HP, Chandler KJ, Hoopes MR, Slentz-Kesler K, et al. Mouse embryonic stem cell adherent cell differentiation and cytotoxicity (ACDC) assay. Reprod Toxicol. 2011; 31:383–391. [PubMed: 21296659]

Bjorke-Monsen AL, Ueland PM, Schneede J, Vollset SE, Refsum H. Elevated plasma total homocysteine and C677T mutation of the methylenetetrahydrofolate reductase gene in patients with spina bifida. Q J Med. 1997; 90:593–596.

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

Brown NA. Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. European Centre for the Validation of Alternatives Methods. Altern Lab Anim. 2002; 30:177–198. [PubMed: 11971754]

Brown NA, Fabro S. Quantitation of rat embryonic development in vitro: A morphological scoring system. Teratology. 1981; 24:65–78. [PubMed: 7302873]

Buesen R, Visan A, Genschow E, Slawik B, Spielmann H, Seiler A. Trends in improving the embryonic stem cell test (EST): an overview. ALTEX. 2004; 21:15–22. [PubMed: 14976585]

Carney EW, Tornesi B, Keller C, Findlay HA, Nowland WS, Marshall VA, et al. Refinement of a morphological scoring system for postimplantation rabbit conceptuses. Birth Defects Res B. 2007; 80:213–222.

Carney EW, Tornesi B, Markham DA, Rasoulpour RJ, Moore N. Species-specificity of ethylene glycol-induced developmental toxicity: Toxicokinetic and whole embryo culture studies in the rabbit. Birth Def Res B. 2008; 83:573–581.

Chandler KJ, Barrier M, Jeffay S, Nichols HP, Kleinstreuer NC, Singh AV, et al. Evaluation of 309 environmental chemicals using a mouse embryonic stem cell adherent cell differentiation and cytotoxicity assay. PLoS One. 2011; 6(6):pe18540.

Chapin R, Augustine-Rauch K, Beyer B, Daston G, Finnell R, Flynn T, et al. 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 Def Res B. 2008; 83:446–456.

Chatot CL, Klein NW, Clapper ML, Resor SR, Singer WD, Russman BS, et al. Human serum teratogenicity studied by rat embryo culture: Epilepsy, anticonvulsant drugs, and nutrition. Epilepsia. 1984; 25:205–216. [PubMed: 6705751]

Daston GP. The theoretical and empirical case for in vitro developmental toxicity screens, and potential applications. Teratology. 1996; 53:339–344. [PubMed: 8910979]

Daston GP, Chapin RE, Scialli AR, Piersma AH, Carney EW, Rogers JM, et al. A different approach to validating screening assays for developmental toxicity. Birth Def Res B. 2010; 89:526–530.

De Jong E, Barenys M, Hermsen SAB, Verhoef A, Ossendorp BC, Bessems JGM, et al. Comparison of the mouse embryonic stem cell test, the rat whole embryo culture and the zebrafish embryotoxicity test as alternative methods for developmental toxicity testing of six 1,2,4-triazoles. Toxicol Appl Pharmacol. 2011; 251:103–111.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

EFSA. Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to the aspects of the biology and welfare of animals used for experimental and other scientific purposes (EFSA-Q-2004-105). EFSA J. 2005; 292:1–46.

EU Directive. EU Directive 2010/63/EU of the European parliament and the council of 22 September 2010. On the protection of animals used for scientific purposes. Off J EU. 2010; 276:33–79.

Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981; 292:154–156. [PubMed: 7242681]

Ferrari DA, Gilles PA, Klein NW, Nadler D, Weeks BS, Lammi-Keefe CJ, et al. Rat embryo development on human sera is related to numbers of previous spontaneous abortions and nutritional factors. Am J Obstet Gynecol. 1994; 170:228–236. [PubMed: 8296827]

Flynn TJ, Collins TFX, Sprando RL, Black TN, Ruggles DI, Wiesenfeld PW, et al. Developmental effects of serum from flaxseed-fed rats on cultured rat embryos. Food Chem Toxicol. 2003; 41:835–840. [PubMed: 12738188]

Garg RC, Bracken WM, Hoberman AM. Reproductive and developmental safety evaluation of new pharmaceutical compounds. In: Gupta RC, editorReproductive and Developmental Toxicology. Elsevier Inc; Boston, MA, USA: 2011. 89–109.

Genschow E, Spielmann H, Scholz G, Seiler A, Brown N, Piersma A, et al. The ECVAM international validation study on *in vitro* embryotoxicity tests: Results of the definitive phase and evaluation of prediction models. Altern Lab Anim. 2002; 30:151–176.

Greene NDE, Dunlevy LPE, Copp AJ. Homocysteine is embryotoxic but does not cause neural tube defects in mouse embryos. Anat Embryol (Berl). 2003; 206:185–191. [PubMed: 12592569]

Hansen DK. In vitro effects of folate derivatives on valproate-induced neural tube defects in mouse and rat embryos. Toxicol In Vitro. 1993; 7:735–742. [PubMed: 20732274]

Hansen DK, Dial SL, Grafton TF. Lack of attenuation of valproic acid-induced embryotoxicity by compounds involved in one-carbon transfer reactions. Toxicol In Vitro. 1995; 9:615–621. [PubMed: 20650137]

Hansen DK, Grafton TF. Lack of attenuation of valproic acid-induced effects by folinic acid in rat embryos in vitro. Teratology. 1991; 43:575–582. [PubMed: 1882348]

Hansen DK, Grafton TF. Comparison of dexamethasone-induced embryotoxicity in vitro in mouse and rat embryos. Teratogen Carcinogen Mutagen. 1994; 14:281–289.

Hansen DK, Grafton TF, Melnyk S, James SJ. Lack of embryotoxicity of homocysteine thiolactone in mouse embyros in vitro. Reprod Toxicol. 2001; 15:239–244. [PubMed: 11390167]

Hermsen SAB, van den Brandhof EJ, van der Ven LEM, Piersma AH. Relative embryotoxicity of two classes of chemicals in a modified zebrafish embryotoxicity test and comparison with their in vivo potencies. Toxicol In Vitro. 2011; 25:745–753. [PubMed: 21238576]

Hettwer M, Reis-Fernandes MA, Iken M, Ott M, Steinberg P, Nau H. Metabolic activation capacity by primary hepatocytes expands the applicability of the embryonic stem cell test as alternative to experimental animal testing. Reprod Toxicol. 2010; 30:113–120. [PubMed: 20132877]

Hofer T, Gerner I, Gundert-Remy U, Liebsch M, Schulte A, Spielmann H, et al. Animal testing and alternative approaches for the human health risk assessment under the proposed new European chemicals regulation. Arch Toxicol. 2004; 78:549–564. [PubMed: 15170526]

Inselman AL, Hansen DK, Lee HY, Nakamura N, Ning B, Monteiro JP, et al. Assessment of research models for testing gene-environment interactions. Eur J Pharmacol. 2011; 668:S108–S116. [PubMed: 21816149]

Juchau MR, Bark DH, Shewey LM, Greenaway JC. Generation of reactive dysmorphogenic intermediates by rate embryos in culture: effects of cytochrome P-450 inducers. Toxicol Appl Pharmacol. 1985a; 81:533–544. [PubMed: 4082200]

Juchau MR, Giachelli CM, Fantel AG, Greenaway JC, Shepard TH, Faustman-Watts EM. Effects of 3-methylcholanthrene and Phenobarbital on the capacity of embryos to bioactivate teratogens during organogenesis. Toxicol Appl Pharmacol. 1985b; 80:137–146. [PubMed: 4024101]

Kalter H. The structure and uses of genetically homogeneous lines of animals. In: Wilson JG, Fraser FC, editorsHandbook of Teratology, Vol. 4 Research Procedures and Data Analysis. Plenum Press; New York, NY, USA: 1979. 155–190.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kennelly EJ, Flynn TJ, Mazzola EP, Roach JA, McCloud TG, Danford DE, et al. Detecting potential teratogenic alkaloids from blue cohosh rhizomes using an in vitro rat embryo culture. J Nat Prod. 1999; 62:1385–1389. [PubMed: 10543898]

Kitchin KT, Ebron MT. Studies of saccharin and cyclo-hexylamine in a coupled microsomal activating/embryo culture system. Food Chem Toxicol. 1983; 21:537–541. [PubMed: 6686180]

Klein NW, Vogler MA, Chatot CL, Pierro LJ. The use of cultured rat embryos to evaluate the teratogenic activity of serum: cadmium and cyclophosphamide. Teratology. 1980; 21:199–208. [PubMed: 7394724]

Klug S, Lewandowski C, Neubert D. Modification and standardization of the culture of early postimplantation embryos for toxicological studies. Arch Toxicol. 1985; 58:84–88. [PubMed: 4091661]

Lawrence C, Sanders GE, Varga ZM, Baumann DP, Freeman A, Baur B, et al. Regulatory compliance and the zebrafish. Zebrafish. 2009; 6:453–456. [PubMed: 19916799]

Luijten M, van Beelen VA, Verhoef A, Renkens MF, van Herwijnen MH, Westerman A, et al. Transcriptomics analysis of retinoic acid embryotoxicity in rat postimplantation whole embryo culture. Reprod Toxicol. 2010; 30:333–340. [PubMed: 20488242]

Luijten M, Verhoef A, Westerman A, Piersma AH. Application of a metabolizing system as an adjunct to the rat whole embryo culture. Toxicol in Vitro. 2008; 22:1332–1336. [PubMed: 18514478]

Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Nat Acad Sci USA. 1981; 78:7634–7638. [PubMed: 6950406]

Marx-Stoelting P, Adriaens E, Ahr HJ, Bremer S, Garthoff B, Gelbke HP, et al. A review of the implementation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. Altern Lab Anim. 2009; 37:313–328. [PubMed: 19678733]

Mehta A, Konala VBR, Khanna A, Majumdar AS. Assessment of drug induced developmental toxicity using human embryonic stem cells. Cell Biol Internat. 2008; 32:1412–1424.

Mosmann T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J Immunol Meth. 1983; 65:55–63.

National Research Counil. Toxicity Testing in the 21st Century A Vision and A Strategy. The National Academies Press; Washington, DC, USA: 2007.

New DAT. Whole-embryo culture and the study of mammalian embryos during organogenesis. Biol Rev. 1978; 53:81–122. [PubMed: 352414]

Oglesby LA, Ebron MT, Beyer PE, Carver BD, Kavlock RJ. Co-culture of rat embryos and hepatocytes: in vitro detection of a proteratogen. Teratogen Carcinogen Mutagen. 1986; 6:129–138.

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

Pellizzer C, Adler S, Corvi R, Hartung T, Bremer S. Monitoring of teratogenic effects in vitro by analysing a selected gene expression pattern. Toxicol in vitro. 2004; 18:325–335. [PubMed: 15046780]

Pennings JL, van Dartel DA, Robinson JF, Pronk TE, Piersma AH. Gene set assembly for quantitative prediction of developmental toxicity in the embryonic stem cell test. Toxicology. 2011; 284:63–71. [PubMed: 21473898]

Piersma AH, van Aerts LA, Verhoef A, Garbis-Berkvens JM, Robinson JE, Peereboom-Stegeman JH, et al. Biotransformation of cyclophosphamide in post-implantation rat embryo culture using maternal hepatocytes in co-culture. Pharmacol Toxciol. 1991; 69:47–51.

Robinson JF, van Beelen VA, Verhoef A, Renkens MFJ, Luijten M, van Herwijnen MHM, et al. Embryotoxicant-specific transcriptomic responses in rat postimplantation whole-embryo culture. Toxicol Sci. 2010; 118:675–685. [PubMed: 20864626]

Rosenquist TH, Ratashak SA, Selhub J. Homocysteine induces congenital defects of the heart and neural tube: effect of folic acid. Proc Natl Acad Sci USA. 1996; 93:15227–15232. [PubMed: 8986792]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Schenk B, Weimer M, Bremer S, van der Burg B, Cortvrindt R, Freyberger A, et al. The ReProTect Feasibility Study, a novel comprehensive *in vitro* approach to detect reproductive toxicants. Reprod Toxicol. 2010; 30:200. [PubMed: 20493943]

Schmid BP, Goulding E, Kitchin K, Sanyal MK. Assessment of the teratogenic potential of acrolein and cyclophosphamide in a rat embryo culture system. Toxicology. 1981; 22:235–243. [PubMed: 7340013]

Schwetz BA. Criteria for judging the relative toxicity of chemicals from developmental toxicity data: A workshop summary. Teratology. 1992; 45:337–339. [PubMed: 1350114]

Seiler A, Visan A, Buesen R, Genschow E, Spielmann H. Improvement of an in vitro stem cell assay for developmental toxicity: the use of molecular endpoints in the embryonic stem cell test. Reprod Toxicol. 2004; 18:231–240. [PubMed: 15019721]

Seiler AE, Spielmann H. The validated embryonic stem cell test to predict embryotoxicity in vitro. Nat Protoc. 2011; 6:961–78. [PubMed: 21720311]

Selderslaghs IWT, Van Rompay AR, De Coen W, Witters HE. Development of a screening assay to identify teratogenic and embryotoxic chemicals using the zebrafish embryo. Reprod Toxicol. 2009; 28:308–320. [PubMed: 19447169]

Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV, et al. Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. Toxicol Sci. 2011a; 124:109–124. [PubMed: 21873373]

Sipes NS, Padilla S, Knudsen TB. Zebrafish – As an integrative model for twenty-first century toxicity testing. Birth Def Res C. 2011b; 93:256–267.

Smith MK, Kimmel GL, Kochhar DM, Shepard TH, Speilberg SP, Wilson J. A selection of candidate compounds for in vitro teratogenesis test validation. Teratogen Carcinogen Mutagen. 1983; 3:461–480.

Spielmann H, Grune B, Liebsch M, Seiler A, Vogel R. Successful validation of *in vitro* methods in toxicology by ZEBET, the National Centre for Alternatives in Germany at the BfR (Federal Institute for Risk Assessment). Exp Toxicol Pathol. 2008; 60:225–233. [PubMed: 18440212]

Spielmann H, Pohl I, Doring B, Liebsch M, Moldenhauer F. The embryonic stem cell test (EST), an in vitro embryotoxi-city test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. In Vitro Toxicol. 1997; 10:119–127.

Strähle U, Scholz S, Geisler R, Greiner P, Hollert H, Rastegar S, et al. Zebrafish embryos as an alternative to animal experiments – a commentary on the definition of the onset of protected life stages in animal welfare regulations. Reprod Toxicol. 2011; doi: 10.1016/j.reprotox.2011.06.121

Uibel F, Muhleisen A, Kohle C, Weimer M, Stummann TC, Bremer S, et al. ReproGlo: A new stem cell-based reporter assay aimed to predict embryotoxic potential of drugs and chemicals. Repro Toxicol. 2010; 30:103–112.

Vanaerts LA, Blom HJ, Deabreu RA, Trijbels FJ, Eskes TK, Copius Peereboom-Stegeman JH, et al. Prevention of neural tube defects by and toxicity of L-homocysteine in cultured postimplantation rat embryos. Teratology. 1994; 50:348–360. [PubMed: 7716742]

van Dartel DA, Pennings JL, de la Fonteyne LJ, Brauers KJ, Claessen S, van Delft JH, et al. Evaluation of developmental toxicant identification using gene expression profiling in embryonic stem cell differentiation cultures. Toxicol Sci. 2011a; 119:126–134. [PubMed: 20935163]

van Dartel DA, Pennings JL, de la Fonteyne LJ, Brauers KJ, Claessen S, van Delft JH, et al. Concentration-dependent gene expression responses to flusilazole in embryonic stem cell differentiation cultures. Toxicol Appl Pharmacol. 2011b; 251:110–118. [PubMed: 21192963]

van Dartel DA, Pennings JL, de la Fonteyne LJ, van Herwijnen MH, van Delft JH, van Schooten FJ, et al. Monitoring developmental toxicity in the embryonic stem cell test using differential gene expression of differentiation-related genes. Toxicol Sci. 2010b; 116:130–139. [PubMed: 20421339]

van Dartel DA, Pennings JL, Hendriksen PJ, van Shcooten FJ, Piersma AH. Early gene expression changes during embryonic stem cell differentiation into cardiomyocytes and their modulation by monobutyl phthalate. Reprod Toxicol. 2009; 27:93–102. [PubMed: 19162170]

van Dartel DA, Pennings JL, Hendriksen PJ, van Shcooten FJ, Piersma AH. Transcriptomics-based identification of developmental toxicants through their interference with cardiomyocyte

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

differentiation of embryonic stem cells. Toxicol Appl Pharmacol. 2010a; 243:420–428. [PubMed: 20036270]

Van den Bulck K, Hill A, Mesens N, Diekman H, De Schaepdrijver L. Zebrafish developmental toxicity assay: A fishy solution to reproductive toxicity screening, or just a red herring? Reprod Toxicol. 2011; 32:213–219. [PubMed: 21704152]

Van Maele-Fabry G, Delhaise F, Picard JJ. Morphogenesis and quantification of the development of post-implantation mouse embryos. Toxicol In Vitro. 1990; 4:149–156. [PubMed: 20702276]

Ying QL, Stavridis M, Griffiths D, Li M, Smith A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nature Biotech. 2003; 21:183–186.

zur Nieden NI, Ruf LJ, Kempka G, Hildebrand H, Ahr HJ. Molecular markers in embryonic stem cells. Toxicol In Vitro. 2001; 15:455–461. [PubMed: 11566578]

zur Nieden NI, Kempka G, Ahr HJ. Molecular multiple endpoint embryonic stem cell test – a possible approach to test for the teratogenic potential of compounds. Toxicol Appl Pharmacol. 2004; 194:257–269. [PubMed: 14761682]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

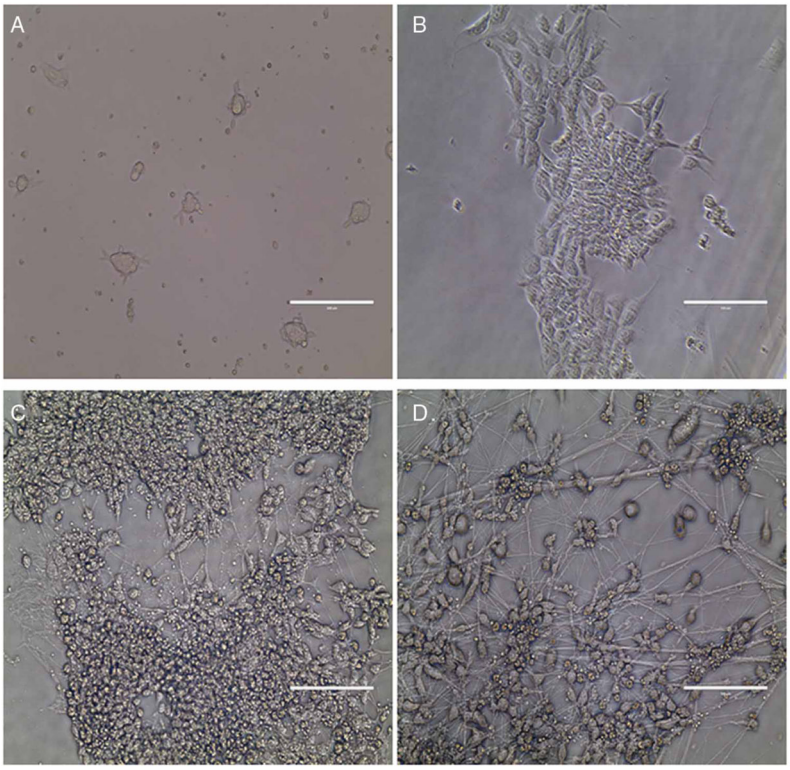

Figure 1.

Differentiation of D3 mESCs to neural cells. A) D3 mESCs cultured according to Ying et al. [2003] early in the culture period (day 2) show little differentiation. B–C) As culture progresses, neural processes begin to appear. D) Numerous cells have neural processes by 14 days of culture.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

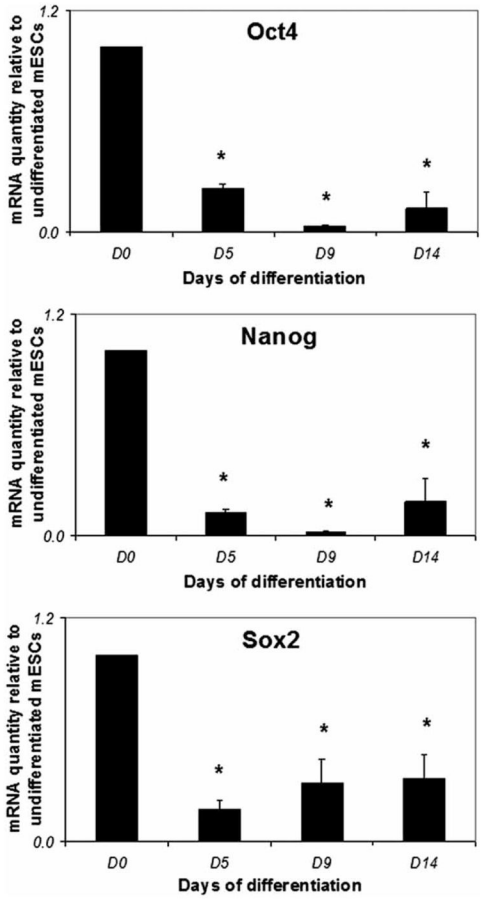

Figure 2.

Expression of the pluripotency markers *Oct4*, *Nanog*, and *Sox2* in D3 mouse embryonic stem cells cultured for 14 days. Decreased expression over time indicates a decrease in pluripotency of the cells.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

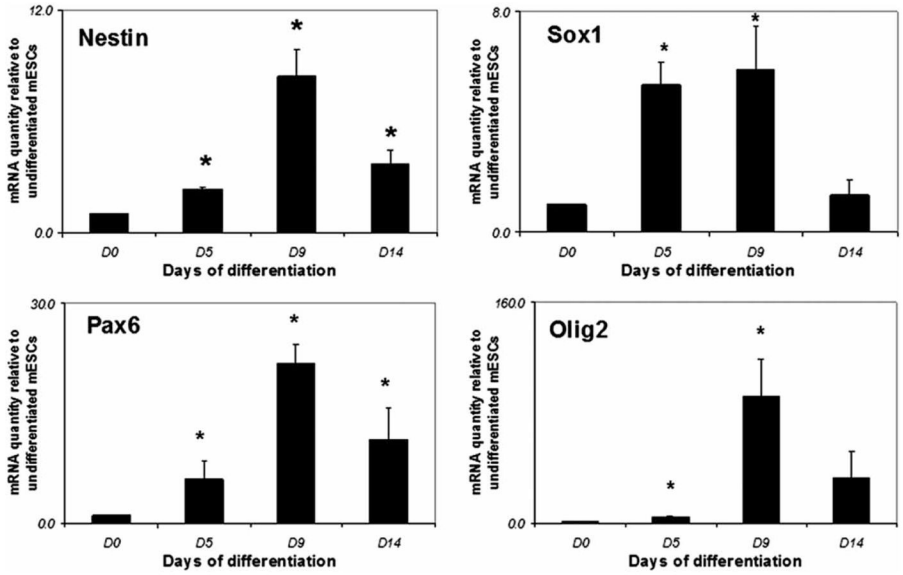

Figure 3.

Expression of the neuronal lineage markers, *Nestin*, *Sox1*, *Pax6*, and *Olig2* in differentiating mESCs. Increased expression of the marker genes indicate differentiation into cells specific for the neuronal lineage.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

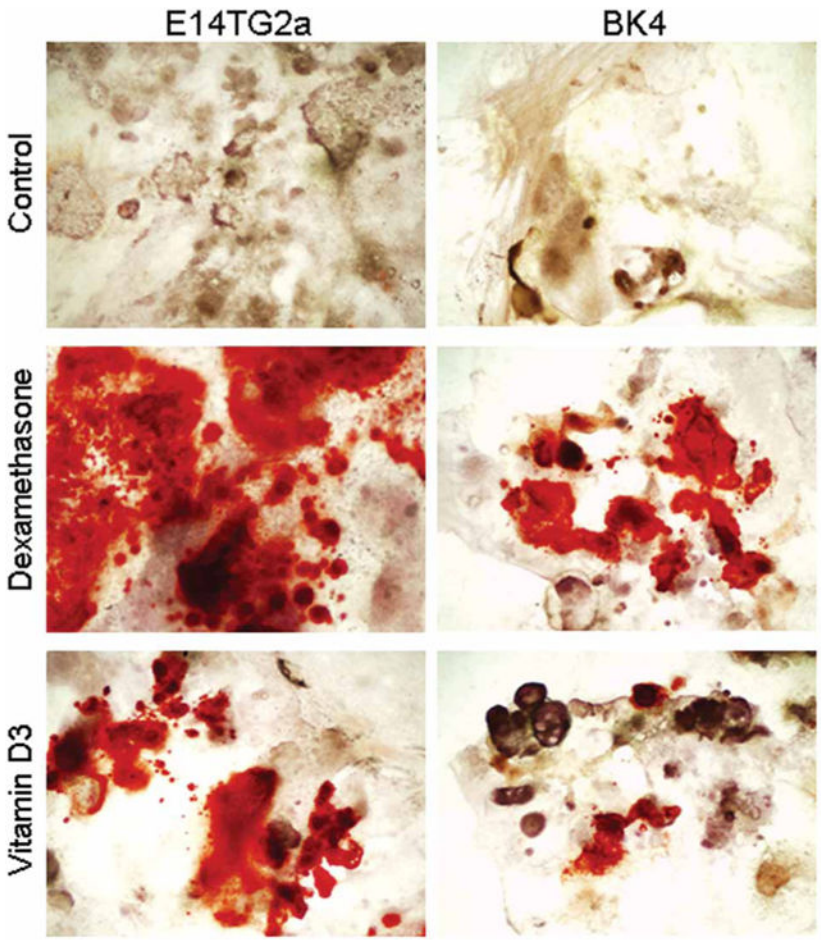

Figure 4.

Differences between cell lines in their differentiation to osteoblasts. Cells from the E14TG2a and BK4 cell lines were cultured under identical conditions to drive differentiation toward osteoblasts. Mineralization was determined by staining with Alizarin Red S, a calcium specific stain. Top panel depicts cells grown in control media, middle panel depicts cells grown in media containing dexamethasone, and lower panel depicts cells grown in media containing vitamin D3.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Comparison of zebrafish teratogenicity assay *in vitro* and whole animal studies *in vivo*.

	BMC_{GMS} (mM)[1]	BMC_T (mM)[2]	BMD_{BW} (mmol/kg bw/d)[3]	BMD_M (mmol/kg bw/d)[4]
MAA - methoxyacetic acid	2.7 (1.9–3.6)	4.6 (2.5–5.7)		
EAA - ethoxyacetic acid	3.1 (2.6–3.7)	2.9 (2.2–3.5)		
BAA - butoxyacetic acid	-	-		
PAA - phenoxyacetic acid	-	-		
MEAA - methoxyethoxyacetic acid	-	-		
BEAA - butoxyethoxyacetic acid	-	-		
EGME - ethylene glycol monomethyl ether	-	-	0.2 (0.1–0.3)	0.5 (0.5–0.7)
EGEE - ethylene glycol monoethyl ether	-	-	0.7 (0.6–0.8)	0.8 (0.8–0.9)

All data from Hermsen et al. [2011].

[1] BMC_{GMS} – Benchmark concentration for general morphological score with a 5% response effect

[2] BMC_T – Benchmark concentration for teratogenic effects with a 5% response effect

[3] BMD_{BW} (mmol/kg bw/d) – Benchmark dose for decreased fetal body weight at 10% response effect

[4] BMD_M (mmol/kg bw/d) – Benchmark dose for increased fetal malformations at 10% response effect

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Comparison of potencies of six triazole compounds *in vivo*, in the mEST, in rat WEC, and in zebrafish.

***In vivo*[1]**	**mEST[2]**	**Rat WEC[3]**	**Zebrafish[4]**
Flusilazole	Flusilazole	Flusilazole	Flusilazole
Hexaconazole	Hexaconazole	Myclobutanil	Hexaconazole
Cyproconazole	Myclobutanil	Hexaconazole	Cyproconazole
Triadimefon	Cyproconazole	Triadimefon	Triadimefon
Myclobutanil	Triadimefon	Triticonazole	Myclobutanil
Triticonazole	Triticonazole	Cyproconazole	Triticonazole

Compounds are listed from most potent to least potent in each model system. All data from de Jong et al. [2011].

[1] *In vivo* potencies are indicated by the developmental lowest effect level (dLEL).

[2] Potencies in the mEST assay are indicated by the BMC_d50 which was the concentration producing a 50% decrease in the incidence of beating cardiomyocytes.

[3] Potencies in the rat WEC assay are indicated by the $BMC05_{TMS}$ which is the concentration producing a 5% decrease in the total morphological score.

[4] Potencies in the zebrafish assay are indicated by the $BMC05_{GMS}$ which is the concentration producing a 5% decrease in the general morphological score.