Chapter 7

Alternative methods to animal experimentation for testing developmental toxicity

David Pamies[1], Carmen Estevan[2], Eugenio Vilanova[2] and Miguel A. Sogorb[2]

[1]Department of Biomedical Science, Department of Physiology, University of Lausanne, Lausanne, Switzerland; [2]Instituto de Bioingeniería, Universidad Miguel Hernández de Elche, Elche, Spain

Chapter outline

Introduction

The term reproductive toxicology is referred to the adverse effect either on fertility of parental generation or on the development of the progeny. The term developmental toxicology is referred to the adverse effects on the developing organism from the moment of the conception to the time of sexual maturation and therefore the developmental toxicology can be considered as part of the reproductive toxicology. The term embryotoxicity is referred to the toxic effects in progeny in the first period of pregnancy between conception and the fetal stage and therefore is included within developmental toxicology and by extension within reproductive toxicology. Finally, the term teratogenicity refers to the structural malformations or defects in offspring and accounts after the period of embryogenesis and is considered as a developmental toxicology effect.

It seems obvious that the complexity of the reproductive process cannot be studied with a single *in vitro* model and therefore it is necessary to split the whole process in single steps (maturation of gamete, fertilization, implantation, embryogenesis, etc.). This chapter will be mainly focused in the study of the currently available models for testing developmental toxicity (embryotoxicity and teratogenicity). Other parts of the processes as infertility, endocrine disruption, mutagenicity of germinal cells, etc., fall outside the scope of this chapter and are already covered by other chapters of this book.

Reproductive and Developmental Toxicology. https://doi.org/10.1016/B978-0-323-89773-0.00007-2

The necessity of alternative models for testing reproductive and developmental toxicology

Organization for Economic Co-operation and Development (OECD) owns several validated Guidelines for *in vivo* studies of reproductive toxicology (covering in the same assay therefore fertility and developmental issues) and for *in vivo* studies of developmental toxicology (covering teratogenicity since the exposure starts after embryogenesis). These Guidelines will be treated in detail in another chapter of this book. However, it is remarkable that there are no Guidelines for testing only *in vivo* embryotoxicity. This is a relevant gap because a Guideline for this purpose would allow detecting developmental toxicants in early stages of development without waiting for teratogenicity.

Regulations in all developed countries require *in vivo* studies about the toxicity to reproduction in order to assess the necessary risk assessment before to register and authorize the use of chemicals with medium and high volume of production. It supposes the use of a large number of animals with the corresponding ethical, logistical, and economic implications. Rovida and Hartung (2009) have estimated the magnitude of these figures as follows:

- one assay following OECD Guideline 414 for testing prenatal developmental toxicity would require 784 animals and cost 63.100 €;
- one assay following reproductive toxicity in two generations would require 3200 animals and cost 328,000 €;
- one assay following OECD Guideline 421 for running an screening of reproduction/developmental toxicity would require 560 animals and cost 54,600 €;
- one assay following OECD Guideline for testing developmental neurotoxicity would require 1400 animals and cost 1,100,000 €.

Taken into consideration the figures outlined in the above paragraph it is easy to understand that the use of fast, safe, and reliable alternative models for testing reproductive toxicology would be highly appreciated by the Industry. These models might be especially relevant for the process of massive high-throughput screening performed in the early stages of developing molecules as biocide, cosmetic, food additive, or whatever. Also other potential applications of these alternative models would be (Spielman, 2005): (i) to compare the developmental toxicity potential of a new chemical that is only a slight modification of an existing chemical that has already been tested *in vivo*; and (ii) evaluating compounds for which testing is not routinely performed, usually since the anticipated exposure is very low. Moreover, the possibility to use human cells, especially human stem cells (such as induced pluripotent stem cells (iPSCs)), has the potential to reduce extrapolation issues from animal experimentation.

In next sections of this chapter, we review the main alternative models for testing developmental toxicity. The review has been divided into five parts. The first and second one devoted to methods for testing embryotoxicity based on simple (monolayer) cellular models coming from both animals and humans. The third part devoted to other approaches for testing developmental toxicity based on more complex cellular models; i.e., those models cultured in 3D forming organoids. The fourth part will be focused on methods for testing developmental toxicity using different embryos (rodents, amphibians, and fishes). Finally, we will present last tendencies of using batteries of assays and Integrated Approaches to Testing and Assessment (IATA) for testing developmental toxicity based on *in vitro* alternative methods.

Methods for testing embryotoxicity based on animal cells

The embryonic stem cell test

In this method, the embryotoxic potential of chemicals is determined by the evaluation of the inhibition of the differentiation of mouse embryonic stem cells belonging to the D3 line and the loss of viability of these D3 cells and of mouse fibroblast belonging to 3T3 line.

Two permanent cell lines are used in the EST test; D3 cells represent embryonic tissues while non-differentiating tissues are represented by 3T3 cells. Three different endpoints were used to build three different empirical biostatistical functions for discriminating among strong embryotoxicants, weak embryotoxicants, and non-embryotoxicants. These endpoints are (Geschow et al., 2000) (Table 7.1): ID_{50} as the concentration with capability to inhibit 50% the differentiation of D3 cells into beating cardiomyocytes after 10 days of exposure; and IC_{50D3} and IC_{503T3} as the concentrations with capability to reduce to 50% the viability of D3 and 3T3 cells after 10 days of exposure in the MTT assay, respectively. See method protocol for a detailed description of experimental procedures (EURL ECVAM-DB-ALM, 2010a). These three endpoints allow the discrimination of the embryotoxic potential of the tested chemical according to criteria showed in Table 7.2. The testing of the 20 test chemicals employed in the EST validation study provided a 78% of accuracy (correct classifications), with precisions of 83%, 81%, and 70% for strong, weak, and non-embryotoxicants, respectively (Genschow et al., 2004).

Alternative approaches for enhancing performance of EST

Some modifications on the original protocol have been suggested by the scientific community on the EST, in order

TABLE 7.1 Biostatistical discriminant functions for the prediction of embryotoxicity in EST, MM, and WEC methods.

(a) Prediction model of the embryonic stem cell test (EST)

$$\text{Function I} = 5.92\log(IC_{50\ 3T3}) + 3.50\log(IC_{50\ D3}) - 5.31\frac{IC_{50\ 3T3} - ID_{50}}{IC_{50\ 3T3}} - 15.7$$

$$\text{Function II} = 3.65\log(IC_{50\ 3T3}) + 2.39\log(IC_{50\ D3}) - 2.03\frac{IC_{50\ 3T3} - ID_{50}}{IC_{50\ 3T3}} - 6.85$$

$$\text{Function III} = -0.125\log(IC_{50\ 3T3}) - 1.92\log(IC_{50\ D3}) + 1.50\frac{IC_{50\ 3T3} - ID_{50}}{IC_{50\ 3T3}} - 2.67$$

(b) Prediction model of the micromass test (MM)

$$\text{Function I} = 6.65 \times \log(ID_{50}) - 9.49$$
$$\text{Function II} = 6.16 \times \log(ID_{50}) - 8.29$$
$$\text{Function I} = -1.31 \times \log(ID_{50}) - 1.42$$

(c) Prediction model number 2 of the whole embryonic test (WEC)

$$\text{Function I} = 21\frac{IC_{50\ 3T3} - IC_{NOEC\ TMS}}{IC_{50\ 3T3}} + 15.37\log(IC_{max}) - 23.58$$

$$\text{Function II} = 21\frac{IC_{50\ 3T3} - IC_{NOEC\ TMS}}{IC_{50\ 3T3}} + 17.71\log(IC_{max}) - 32.37$$

$$\text{Function III} = 9.3\frac{IC_{50\ 3T3} - IC_{NOEC\ TMS}}{IC_{50\ 3T3}} + 4.21\log(IC_{max}) - 4.23$$

Abbreviations: IC_{503T3}, Concentration that reduces viability of 3T3 cells to 50% after 10 days of exposure; IC_{50D3}, Concentration that reduces viability of D3 cells to 50% after 10 days of exposure; ID_{50}, Concentration that reduces to 50% the differentiation of either D3 cells to cardiomyocytes (EST) or primary culture of limb bud cells to cartilage (MM) after exposure according to the respective EST and MM protocols; $IC_{NOEC\ TMS}$, The lowest concentration without observed effect on the total morphological score (see Table 7.5); IC_{max}, The lowest concentration that causes the maximum malformation rate.

TABLE 7.2 Embryotoxic potential classification criteria according to prediction models displayed in Table 7.1 for EST, WEC, and MM methods.

Classification	Requirements
Strong embryotoxicity	Function III > Function II and Function III > Function I
Weak embryotoxicity	Function II > Function III and Function II > Function I
No embryotoxicity	Function I > Function III and Function I > Function II

to improve its performance and adapt to another areas (such as neurotoxicity). For example, Suzuki et al. (2011) developed a transgenic D3 cell line expressing luciferase gene inserted upstream of the promoter controlling the expression *Hand1*, a biomarker of cardiac morphogenesis. The proposed procedure considers the recording of impairments on D3 cellular differentiation through alterations in bioluminescence expressed by the exposed transgenic D3 cells. This experimental variation allows reducing the test duration and workforce as compared with the standard EST with a total accuracy of 83% (Suzuki et al., 2011, 2012).

Several efforts have been made to detect developmental neurotoxicants by following approaches based on the EST. Theunissen et al. (2010) proposed an abbreviated protocol to promote neural differentiation and to repress mesodermal differentiation using D3 embryoid bodies exposed to retinoic acid. This model has proved useful to study the developmental neurotoxicity of several chemicals like methylmercury (Theunissen et al., 2011), and others such as valproic acid, cyproconazole, hexaconazole (Theunissen et al., 2012a), acetaldehyde, carbamazepine, flusilazole, monoethylhexyl phthalate, penicillin-G, and phenytoin (Theunissen et al., 2012b). It has been also demonstrated that a combined approach, which incorporates a classic EST with the neural EST approach, may improve developmental toxicant detection as compared to individual assays (Theunissen et al., 2013). A similar approach has been used with neural differentiation medium (called DNT-EST) (Hayess et al., 2013) and with the differentiation of D3 cells toward osteoblasts (de Jong et al., 2012).

Romero et al. (2015) developed other interesting variation of EST. They proposed a methodology for detecting

embryotoxicants based on the determination of differentiation impairments through recoding biomarker transcripts. This methodology used three different endpoints (Fig. 7.1) with the intention of substituting the alterations of differentiation to beating cardiomyocytes by alterations of biomarker gene expression (Romero et al., 2015). This methodological approach was able to correctly predict the embryotoxicity of a battery of seven model chemicals (two strong, three weak, and two non-embryotoxicants).

EST has been also combined with toxicokinetic approaches for enhancing its predictivity. Specifically, EST was able to correctly rank the *in vivo* toxicity of 12 different azoles when combined with relative placental transfer rates determined using human placenta BeWo cells culture (Dimopoulou et al., 2018).

EBT (for Embryoid Body Test) is a variation of EST that greatly reduced the exposure time, labor, and required materials. EBT uses another empiric prediction model based on three different endpoints with capability to discriminate between embryotoxic and non-embryotoxic substances (Lee et al., 2019). This is one of the main differences with EST that was designed for discriminating between three different potencies of embryotoxicity. The endpoints considered in EBT are: (i) the concentration of substance that reduces by 50% the area of embryoid bodies of D3 mouse embryonic stem cells; (ii) the concentrations of substance that reduces by 50% the viability of D3 mouse embryonic stem cells; and (iii) the concentrations of substance that reduces by 50% the viability of 3T3 mouse fibroblasts. The performance of EBT in an interlaboratory blind prevalidation study with 21 chemicals was of 95%, 100%, and 83% for accuracy, sensitivity, and specificity, respectively (Lee et al., 2019).

One of the main limitations of EST is that in its original protocol the high-throughput screening is not easily allowed due to the need of a great amount of handmade operations, as the individual generation of hanging-drops or embryoid bodies to the individual assessment of the beating cardiomyocytes. In this sense, several modifications of the EST protocol have been proposed in order to establish its reliability and feasibility for industrial application during drug development. Witt et al. (2020) performed one of the successful attempts in this line. They shortened cytotoxicity assay procedures; based the cytotoxicity assays on determination of ATP; enabled automation compatibility of the workflow for cell seeding, compound dilution, media exchange, and viability assessment; enabled automated embryoid bodies generation in a 96-well format; and introduced flow cytometric quantification of marker expression. These modifications allowed the correct classification for all performed experiments of a battery of model chemicals formed by two non-embryotoxic substances (saccharin and penicillin G), two weak embryotoxic substances (caffeine and dexamethasone), and two strong embryotoxic substances (5-fluorouracil and hydroxyurea) (Witt et al., 2020).

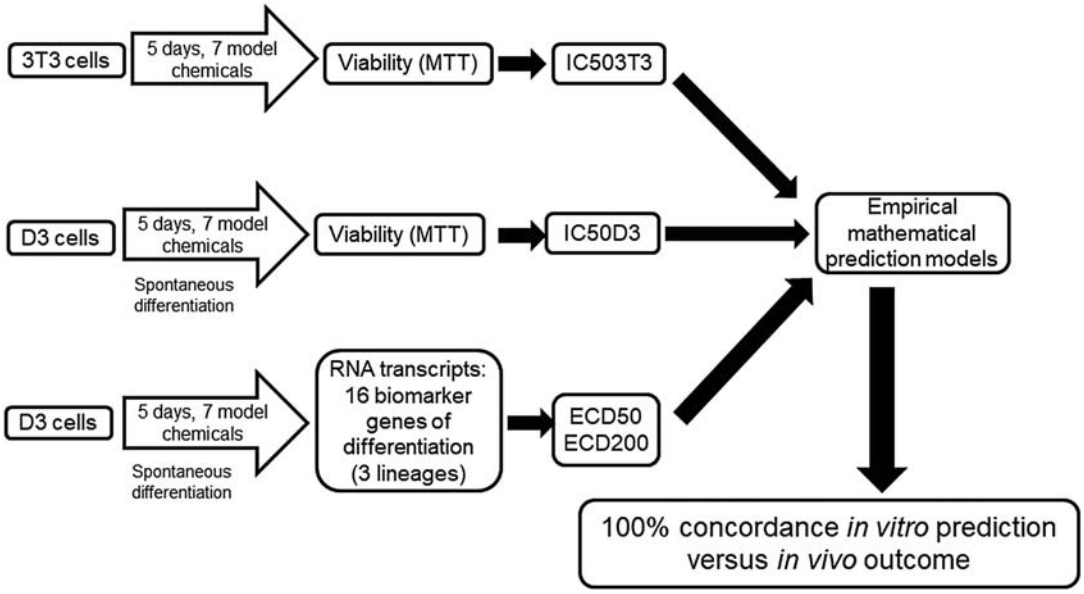

FIGURE 7.1 Overview of a procedure for detecting embryotoxicants based on the detection of alterations of D3 mouse embryonic stem cell differentiation through analysis of RNA transcripts and on alterations on D3 and 3T3 mouse fibroblasts cell viability. *Reproduced from Romero, A.C., Del Río, E., Vilanova, E., Sogorb, M.A., 2015. RNA transcripts for the quantification of differentiation allow marked improvements in the performance of embryonic stem cell test (EST). Toxicol. Lett. 238 (3), 60–69.*

The micromass test

This method uses rat micromass (MM) cultures of limb bud and detects the inhibition of cell differentiation and growth. This method is based on the capability of primary culture of limb bud cells of mammalian origin to reproduce cartilage histogenesis, a fundamental step in morphogenesis of the skeleton, cell proliferation and differentiation, cell-to-cell communication, and cell to extracellular matrix interactions.

See method protocol for a detailed description of experimental procedures (EURL ECVAM-DB-ALM, 2010b). The protocol is based on the isolation of the limb bud cells (Fig. 7.2) of rat embryos on day 14 of gestation that will be further seeded and exposed during 5 days to the tested chemical for being finally assessed for cartilage differentiation using alcian blue (a cartilage-specific proteoglycan stain). The concentration able to reduce the cartilage differentiation by 50% is used as endpoint in empiric biostatistical functions (Table 7.1) for discriminating among strong, weak, and non-embryotoxicants according to criteria shown in Table 7.2.

The performance of MM is lower than the performance of EST and it might be one of the reasons because MM is currently barely used. The predictivity of MM for strong embryotoxic chemicals was 100%, while this record falls to 57% and 71% for non- and weak embryotoxic, respectively (Gleschow et al., 2002).

Methods for testing embryotoxicity based on human cells

All the cellular methods presented in section Methods for testing embryotoxicity based on animal cells of this chapter considers the use of cell cultures from animal origin. The results obtained with these cells raise the problem of the result's extrapolation from animals to humans. The research in developmental toxicity with human cells has been considered during decades unaffordable due to obvious bioethical reasons. However, since around 15–20 years ago, the development of methodologies for the generation of human iPSC (hiPSC) (embryonic-like pluripotent stem cells artificially derived from non-embryonic and non-pluripotent cells as blood, skin, or adipose tissue) has notably reduced these bioethical concerns and has boosted the research of developmental toxicity with human cells. The following paragraphs offer several examples of methodologies based on the use of human cells for testing chemical's embryotoxicity. The main characteristics of these methodologies are also summarized in Table 7.3.

Human iCell Neurons are a mixture of postmitotic GABAergic and glutamatergic neuronlike obtained from hiPSC. When these cells were exposed to 80 chemicals screened for their ability to inhibit neurite outgrowth it was noted that 38 of them were active against neurite outgrowth, being 16 of them selective (capable to inhibit neurite outgrowth in absence of general cytotoxicity). Among these 16 chemicals, 12 of them were annotated as *in vivo* developmental neurotoxicants or neurotoxicants (Ryan et al., 2016). These results allow proposing alterations in neurite outgrowth among the battery of different endpoints for testing developmental neurotoxicity.

In the previous, it was commented a variation on the EST protocol based on reductions of duration of cytotoxicity assays, changes to ATP determination as endpoints of cytotoxicity, automation of the cultures processes, and flow cytometric quantification of marker expression. These same authors also proved that these approaches are also suitable for the use of ZIPi013-B and ZIPi013-E hiPSC (Witt et al., 2020). It is noted that in this approach readouts of hiPSC and rodent fibroblast were mixed for predicting the embryotoxicity of the tested chemicals.

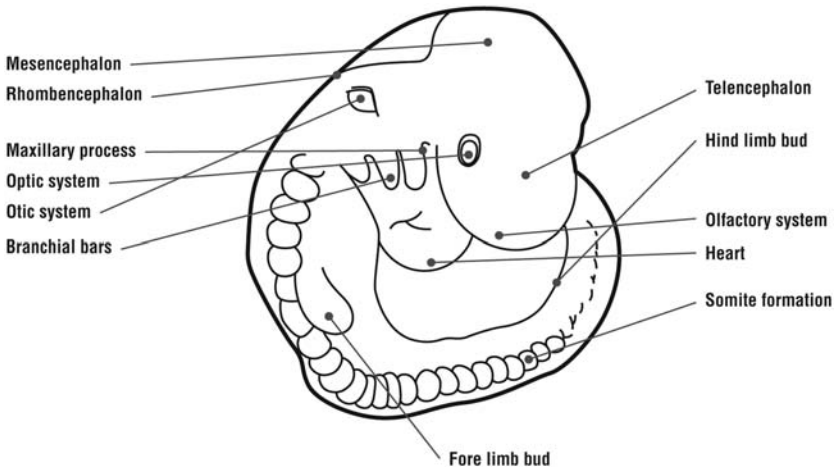

FIGURE 7.2 Morphology of rat embryo.

TABLE 7.3 Summary of methods for testing embryotoxicity using human cells.

Cells	Origin	Endpoint	Reference
iCell® Neurons	hiPSC	Neurite outgrowth and cytotoxicity	Ryan et al. (2016)
ZIPi013-B and ZIPi013-E	hiPSC	Inhibition of cardiomyocyte formation	Witt et al. (2020)
hiPSC	Blood, bone marrow, and mononuclear cells	Inhibition of cardiomyocyte formation	Aikawa (2020)
H9	Embryonic stem cells	Ornithine/cysteine ratio	Palmer et al. (2013) Zurlinden et al. (2020)

hiPSC, human induced pluripotent stem cells.

The suitability of the EST approach using hiPSC derived from blood, bone marrow, and mononuclear cells and human fibroblasts was also demonstrated using as endpoints the inhibitory effects of the drugs on cardiac differentiation and on the proliferation survival of both (hiPSC and human fibroblast) cells. These endpoints were used to develop a biostatistical prediction function for classifying the chemicals within three different categories of toxicity: the non-effective, the embryotoxicants (those causing growth retardation and dysfunction), and the teratogenic (those causing malformations or death). This methodology assigned to aminopterin, methotrexate, retinoic acid, thalidomide, tetracycline, lithium, phenytoin, 5-fluoracil, warfarin, and valproate the classification of embryotoxicants or teratogenic, when the *in vivo* performance of all of them is as teratogenic (Aikawa, 2020). The non-developmentally toxicants ascorbic acid, saccharin, and isoniazid were correctly classified as non-effective substances.

The ornithine/cysteine ratio in cellular media of human embryonic H9 stem cells was developed as a biomarker of embryotoxicity. Specifically, a chemical is considered embryotoxic at the exposure level able to generate an ornithine/cysteine ratio lower than 0.88 (Palmer et al., 2013). This methodology showed a concordance higher than 75% with the existing *in vivo* results of a battery of 46 chemicals. This methodology has been also assessed with the ToxCast phase I and II chemicals showing a positive predictive capability with 79%−82% of accuracy and more than 84% of specificity (Zurlinden et al., 2020).

Three-dimensional cultures and organoids for testing developmental toxicity

The conjunction of advances in microfabrication and stem cells has produced a remarkable change in the development of new *in vitro* models (Pamies and Hartung, 2017; Pamies et al., 2017a). On the one hand, the development of stem cells, especially iPSC (Shi et al., 2017), has allowed not only easy access to human cells with a more stable genetic background, but also a large number of developments related to their genetic modification. These advances have opened numerous doors for the study of the human development and human diseases. On the other hand, the appearance of new culture methods such as 3D cultures, bioreactors, and on-a-chip technologies (Whitesides, 2006), has brought models with greater human physiological relevance.

In the area of developmental toxicity, probably one of the most relevant new *in vitro* models are organoids. These cultures by definition are models, desirably from multicellular human cells, that represent a functional part of an organ or tissue. These models can be obtained using different techniques such as gravity aggregation, bioreactors, suspension cultures, 3D printing, or the combination of them (Fig. 7.3) (Chesnut et al., 2019). Since most of these models are developed through stem cells (generally iPSCs) and the development of these models tries to simulate the normal development of human organ or tissue, these models have the potential to be used as development models. We can find organoids of many organs such as kidney (Freedman et al., 2015; Takasato et al., 2016), stomach (McCracken et al., 2014), lung (Dye et al., 20154), pancreas (Hohwieler et al., 2017), brain (Pasca et al., 2015; Lancaster and Knoblich, 2014), intestine (Leslie et al., 2015), and others (Kim et al., 2020). Although these methods have many advantages, they also bring new challenges, such as greater complexity and cost, lower reproducibility in some cases, the need to adapt some endpoints to 3D cultures, among others. These advantages and disadvantages have already been summarized previously for organoids applied in glioblastoma research (Pamies et al., 2020a), being these shared among other fields. It is important also to take into consideration that quality assurance for these new models and stem-cell-derived models, such as Good Cell Culture Practice, is only now emerging (Pamies et al., 2017b, 2018a, 2020b). The

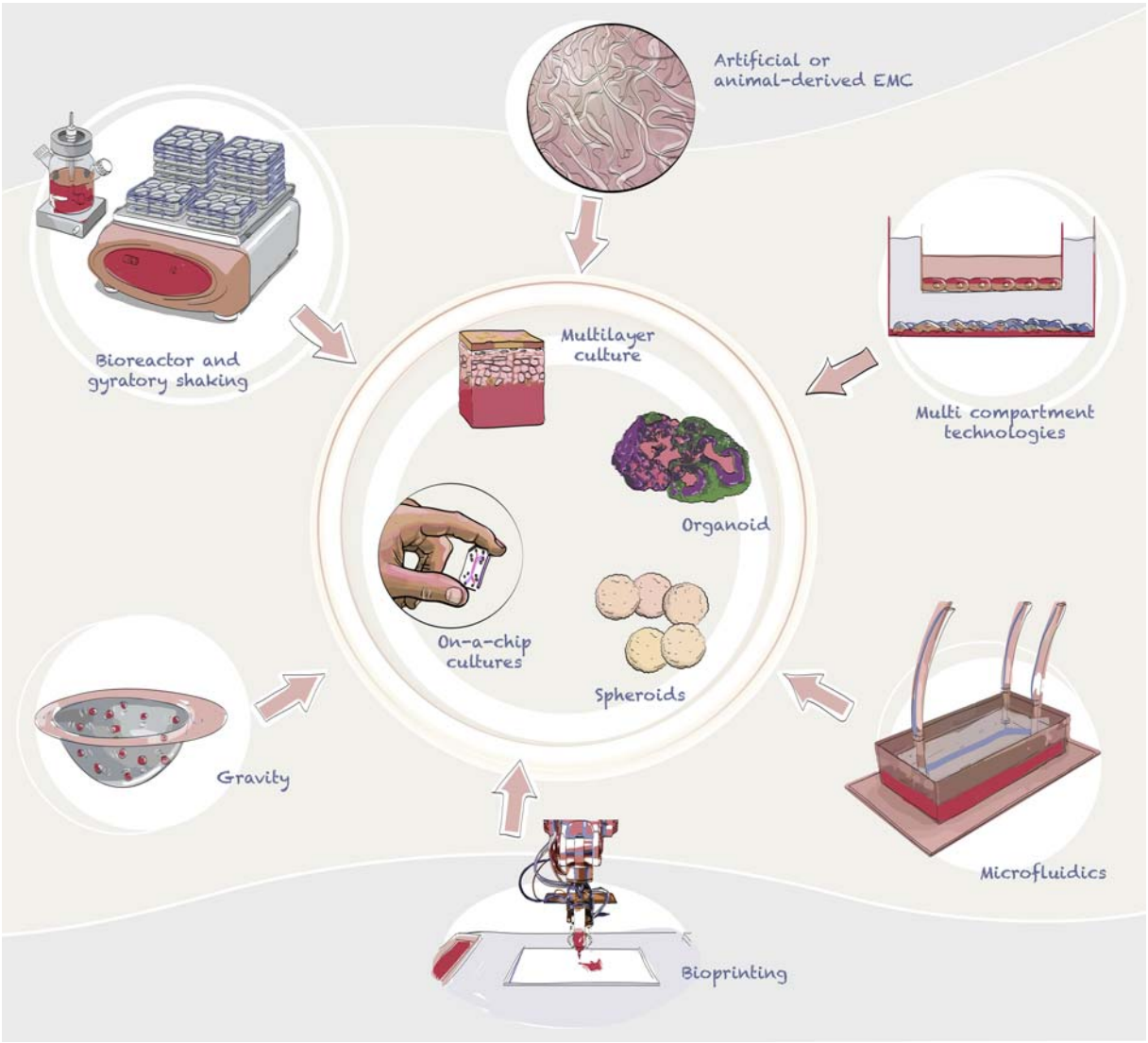

FIGURE 7.3 Advanced cell culture technologies currently employed to develop *in vitro* models of the brain and blood–brain barrier. In the diagram the main ways to generate 3D cultures, aggregating and multilayer technologies, are summarized. Aggregation can be used to generate 3D cultures, for this, gravity or shaking are used to form cell structures (normally spheric). *Reproduced from Chesnut, M., Muñoz, L.S., Harris, G, Freeman, D., Gamam L., Pardo, C.A., Pamies, D., 2019.* In vitro *and in silico models to study mosquito-borne Flavivirus neuropathogenesis, prevention, and treatment. Front. Cell Infect. Microbiol. 9, 223.*

implication of these models in the area of developmental toxicity have been already summarized elsewhere (Luz and Tokar, 2018).

Some models have already been used for developmental toxicity models. For example, a hiPSC-derived brain model (also called BrainSpheres) (Pamies et al., 2017c; Hogberg et al., 2013), developed at the Johns Hopkins Center for Alternative to Animal Testing (CAAT), has been used for several neurotoxicity and developmental neurotoxicity studies (Zhong et al., 2020; Pamies et al., 2018b; Hogberg et al., 2021). More specifically, BrainSpheres have been used to study the effect of an antidepressant paroxetine on brain development. For that, BrainSpheres were exposed to found concentration levels of paroxetine in blood during the development and differentiation of the 3D models. Then, several key events during regular brain development (e.g., neurite outgrowth, synaptogenesis, oligodendrocytes differentiation) were used as an endpoint to assess neurodevelopmental effects (Zhong et al., 2020).

Rotenone was also considered as a potential neurodevelopmental toxicant when caused mitochondrial dysfunction, alterations in Ca^{2+} reabsorption, synaptogenesis, and peroxisome proliferator–activated receptors disruption in this BrainSpheres model (Pamies et al., 2018b). Rat BrainSpheres have been also used for testing the neurodevelopmental toxicity of several organophosphorus flame

retardants. It was found that these substances were able to induce an array of effects as reductions in the levels of glutamate, GABA, and n-acetyl aspartate and in the plasma membrane dopamine active transporter expression, interferences with myelination, increase of cytokine gene and receptor expressions, disruption in transmission of action potentials, cell–cell signaling, synaptic transmission, receptor signaling, immune response, inflammation, defense response, cell cycle, and lipids metabolism and transportation (Hogberg et al., 2021).

In addition to BrainSpheres other 3D-test systems have been developed during recent years. Table 7.4 summarizes the main characteristics of the 3D models presented in this section. CHES6 human embryonic stem cells are able to differentiate into 3D human neural tissues containing mature neurons, astrocytes, and oligodendrocytes. The culture also shows mature myelin sheaths around axons and electrophysiological spontaneous activity. It was noted an increase in markers of astroglial reactivity after exposure of the culture to the *in vivo* teratogens methylmercury and trimethyltin (Sandström et al., 2017).

A three-dimensional human dopaminergic *in vitro* LUHMES cell line has been also used for testing the effects of known neurodevelopmental toxicants as rotenone. Lund human mesencephalic (LUHMES) cells can be differentiated into morphologically and biochemically mature dopamine-like neurons. It was found that this substance was able to decrease complex I activity, ATP, mitochondrial diameter, and neurite outgrowth (Harris et al., 2018).

hiPSC-derived 3D cortical neuron/astrocyte cocultures from a single human donor source was used for testing a library of 87 compounds that included pharmaceutical drugs, pesticides, flame retardants, and other chemicals and it was found that 50 of these compounds were able to significantly alter calcium oscillation rate and peak width, amplitude, and waveform (Sirenko et al., 2019).

Mitochondrial membrane impairment, intracellular glutathione level, cell membrane integrity, DNA damage and apoptosis in 3D-cultured ReNcell VM seeded in 384-pillar plates have been used as endpoints for testing the toxicity of the model compounds rotenone, 4-aminopyridine, digoxin, and topotecan. ReNcell VM cells are derived from ventral mesencephalon region of a human fetal brain and are an immortalized human neural progenitor cell with the ability to differentiate into neurons and glial cells. The results yielded a Z′ factor of 0.6 (a result between 0.5 and 1.0 is statistically considered excellent in high-throughput screens) and a coefficient of variation values around 12% (Joshi et al., 2020).

Three-dimensional BIONi010-C, IMR90-1, and IMR90-4 hiPSC cultures were used for developing a testing system called PluriBeat. PluriBeat uses as quantitative readout of the assay the cardiomyocyte contractions in the differentiated embryoid body on day 7. This method was able to

TABLE 7.4 Summary of 3D cellular systems for testing developmental neurotoxicity.

Cellular model	Endpoints	Tested substances	References
Human and brain spheres	Neurite outgrowth, synaptogenesis, oligodendrocytes differentiation, mitochondrial dysfunction, alterations in Ca^{2+} reabsorption, transcriptomics, etc.	Paroxetine Rotenone Organophosphorus flame retardants	Zhong et al. (2020); Pamies et al. (2018b); Hogberg et al. (2021)
CHES6 human embryonic stem cells	Markers of astroglial reactivity	Methylmercury and trimethyltin	Sandström et al. (2017)
Human dopaminergic i*n vitro* LUHMES cell	Complex I activity, ATP, mitochondrial diameter, and neurite outgrowth	Rotenone	Harris et al. (2018)
Human iPSC–derived 3D cortical neuron/astrocyte cocultures from a single human donor	Calcium oscillation rate and peak width, amplitude, and waveform	87 compounds that included pharmaceutical drugs, pesticides, flame retardants, and other chemicals	Sirenko et al. (2019)
ReNcell VM	Mitochondrial membrane impairment, intracellular glutathione level, cell membrane integrity, DNA damage, and apoptosis	Rotenone, 4-aminopyridine, digoxin, and topotecan	Joshi et al. (2020)
BIONi010-C, IMR90-1, and IMR90-4 human induced pluripotent stem cells	Cardiomyocyte contractions in the differentiated embryoid body on day 7	Thalidomide and epoxiconazole	Lauschke et al. (2020)

classify thalidomide and epoxiconazole as embryotoxic substances (Lauschke et al., 2020).

Alternative models based on embryos for testing developmental toxicology

Whole embryo culture

The rationale of this study is based on the *in vitro* exposure of embryos during the time where major aspects of organogenesis occur, as in the case of heart development, closure of neural tube, development of ear and eye, branchial bars, and limb buds. Therefore, it is assumed that interferences during this period may lead to general retardation of growth and development or to specific malformations.

See method protocol for a detailed description of experimental procedures (EURL ECVAM-DB-ALM, 2010c), basically, rat or mouse embryos are cultured *in vitro* for 48 h starting on day 9.5 of gestation in presence of the tested substance. After 48 h, each embryo (Fig. 7.2) is translated to a petri dish and scored according to the parameters described in Table 7.5.

The results obtained during development with model chemicals allowed the development of empirical biostatistical functions (Table 7.1) for classification of chemicals in the three categories (strong developmental, weak developmental, and non-developmental toxicants) according to criteria showed in Table 7.2. Whole embryo culture (WEC) showed predictivities of 70%, 76%, and 100% for non-, weak, and strong developmental toxicants, respectively (Gleschow et al., 2002).

Alternative approaches for enhancing performance of WEC

One interesting proposal for streamlining WEC come from Zhang et al. (2012), who proposed to use only a single test concentration (1 μM) and three structural endpoints (group average morphological scores of spinal cord, heart, and number of somite pairs). This approach allowed developing a prediction model based on preliminary results with 39 different chemicals that yielded no statistically significant differences in the predictivity compared with the validated methodology.

Toxicogenomic-based approaches have been also proposed for complementing morphological WEC endpoints. These approaches are based on observations of the gene expression level associated with specific differential morphological outcomes, which has allowed developing omic signatures that reveal mechanisms of developmental toxicity. These procedures have been successfully applied for the case of retinoic acid (Robinson et al., 2012a) and triazoles (Robinson et al., 2012b).

As was described above WEC has been also combined with toxicokinetic approaches for enhancing its predictivity as regard azoles, finding similar results to those described above (Dimopoulou et al., 2018).

Frog embryo teratogenesis assay

The organogenesis is a process highly conserved in phylogenetic scale and therefore amphibians can be used as model for testing this process in mammalians. In addition to that, amphibian embryos are very sensible to chemicals, easily handled in the laboratory, and the availability of embryos is not seasonal because the ovulation can be induced with chorionic gonadotropin. All these reasons become frog embryos in a good model for testing alterations in the development of vertebrates. Specifically, the first 96 h of embryonic development in *Xenopus laevis* parallel many of the major processes of human organogenesis (NICEATM, 2000). Nevertheless, other authors suggest that other species of *Xenopus* as *Xenopus tropicalis* can also be effectively used (Fort et al., 2004).

The endpoints for the frog embryo teratogenesis (FETAX) assay are (NICEATM, 2000): (1) mortality, expressed as the concentration that causes 50% mortality (LC_{50}); (2) malformations, evaluated and recorded according to the Atlas of Abnormalities (Bantle et al., 1998) and expressed as the concentration that causes malformations in 50% of embryos (EC_{50}); (3) growth, estimated as the distance between head and tail; (4) teratogenic index, estimated as the ratio between LC_{50} and EC_{50}; and (5) minimum concentration to inhibit growth. A chemical ranked with a teratogenic index greater than 1.5 is a great candidate to be teratogenic in absence of significant mortality. In the same way, teratogenic hazard is considered to be present when either growth is significantly inhibited at concentrations below 30% of LC_{50} or when ratio between minimum concentration to inhibit growth and LC_{50} is lower than 0.30.

Several studies of validation for FETAX assays have been run with promising results (Bantle et al., 1996; Fort et al., 1998, 2000; Bantle et al., 1999). Nevertheless, the US EPA asked to Interagency Coordinating Committee on the Validation of Alternative Methods to evaluate the FETAX test. In 2000, an expert scientific panel concluded that FETAX is not sufficiently validated or optimized for regulatory use. Nevertheless and despite this consideration FETAX is an assay of proved scientific validity and is widely used for identification of hazard to human and environmental health (just as examples and among others: Bacchetta et al., 2008; Longo et al., 2008; Bosisio et al., 2009).

TABLE 7.5 Parameters to score after exposure of rat or mouse embryos in the WEC assay.

Growth parameters		**Malformations (0 for normal/1 for malformed)**
Yolk sac diameter (mm)		Yolk sac vessel defect
Crown-rump length (mm)		Allantois nor fused with ectoplacental cone
Head length (mm)		Allantois large size
Functional parameters (1 for normal/0 for abnormal)		Flexion deficient
		Pericardiac sac wide, filled with fluid
Yolk sac circulation		Heart ventrally turned
Allantois circulation		Posterior neuropore open
Heartbeat		Dorsal midline irregular
Somite development		Prosencephalon open
Final somite number		Rhomboencephalon narrow
Final−initial somite number		Cranial neural folds suture line irregular
Morphological scores		Head small and bent backward
A	Yolk sac blood vessels	Craniofacial appearance abnormal
B	Allantois	Neural tube hemorrhagic
C	Flexion	Rhombencephalon large and transparent
D	Heart	Rhombencephalon narrow
E	Caudal neural tube	Otic vesicles deformed
F	Hind brain	Optic vesicles deformed
G	Midbrain	Branchial bars deformed
H	Fore brain	Maxillary process swollen
J	Otic system	Mandibular processes unapproached
K	Optic system	Mandibular process deformed
L	Olfactory system	Somites small
M	Branchial bars	Somites irregular
N	Maxillary process	Tail kinked
P	Mandibular process	Rail short and thickened
Q	Fore limb	Subcutaneous blisters
R	Hind limb	Hemorrhages
Total Morphological Score (A + B + C + … + R)		Other

Developmental toxicity assays with zebrafish

The teleost zebrafish (*Danio rerio*) is a well-known organism frequently used in general and developmental toxicology (Froehlicher et al., 2009) and neurotoxicology (Anichtchik et al., 2004; Bretaud et al., 2004; Linney et al., 2004). The small size, cheap maintenance, easy conditions for breeding, high nativity rate (a single female can lay up to 400 eggs per week (Laale, 1977)), the spawn throughout the year under laboratory conditions, the transparency of its embryos, and the fact they develop outside the mother become zebrafish in an excellent model for research (Yang et al., 2009).

The development of the embryo in zebrafish is particularly fast. The stages of embryonic development of *Danio rerio* are divided into seven periods (Fig. 7.4): the zygote, cleavage, blastula, gastrula, segmentation, pharyngula, and hatching periods. The zygote is formed immediately after union of female with male gametes. Cleavage is produced

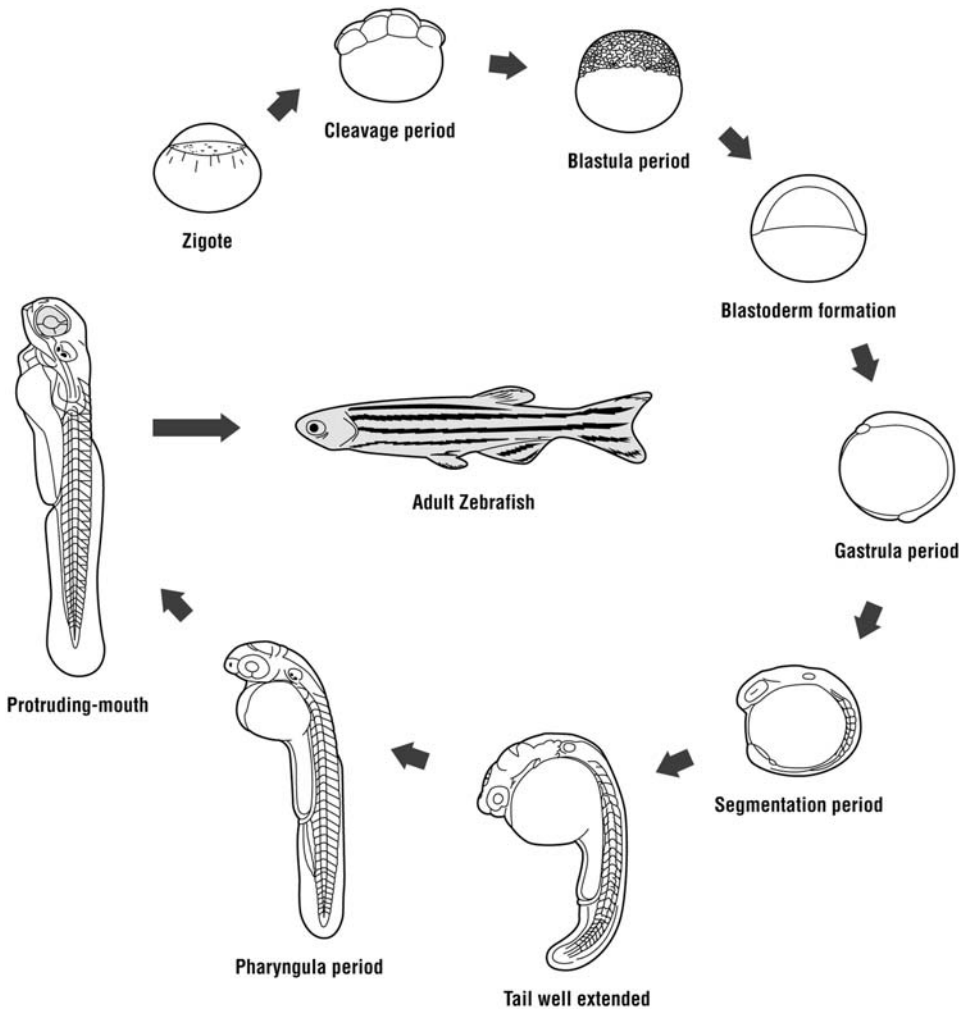

FIGURE 7.4 Embryological stages of zebrafish.

between 45 min and 2 h and consists in a series of mitotic cell differentiation that produce the blastula. The formation of the blastula is produced between 2 and 4 h and is a hollow structure consisting of a single layer of cells. Gastrula is produced around 10 h after fertilization and consists in the migration of the cells forming different structures that result in the formation to the three primary germ layers: ectoderm, endoderm, and mesoderm. Segmentation is a morphogenetic process where the somites develop, and the rudiments of the primary organs start to be patent, the tail bud becomes more prominent, and the embryo of zebrafish elongates. The formation of pharyngula starts at 24 h after fertilization and takes another 24 h. In this stage, all vertebrate embryos show important similitude. However, in this moment notochord and postanal tail are developed, along with rapid cerebellar morphogenesis of the metencephalon. Hatching is a period between 48 and 72 h after fertilization and consists on the formation of the primary organ system, rudiments of the pectoral fins, the jaws, and the gills (Kimmel et al., 1995).

Dario rerio *embryotoxicity test*

Dario rerio embryotoxicity test (DarT) is based in the study of the effects on embryos as consequence of their exposure to the tested chemical for 48 h. DarT is considered an *in vitro* test and is accepted as alternative method to animal experimentation. Twenty embryos per concentration are incubated during 48 h after fertilization in 24-well plate with a 12 h light−dark cycle. After 48 h of exposure different parameters are analyzed: size of the eggs, the eye and the sacculi/otoliths position, the pigmentation, the tail not detached, and the frequency of spontaneous movements. A positive control of 3,4-dichloroaniline is run in parallel with the test compound since the effect of this compound on embryo zebrafish development is well known (Nagel et al., 1991).

Despite the main advantages of this model for testing developmental toxicity, the National Toxicology Program has raised several deficits that hinder the broader adoption of the zebrafish model for toxicological screening. These deficits are: (i) Lack of consistent experimental protocol elements; (ii) Lack of clear understanding of mechanisms of chemical absorption, distribution, metabolism, and excretion in zebrafish; and (iii) Lack of consistent informatics approaches used for classification of outcomes. In order to overcome these problems the National Toxicology Program has launched the Systematic Evaluation of the Application of Zebrafish in Toxicology (SEAZIT) program.[1]

Variants about the DarT

The model called MolDart uses the detection of changes in the expression of specific target genes after 120 h of exposure (Liedtke et al., 2008). This test system uses developing zebrafish and detects changes in mRNA abundance of selected target genes after exposure for 5 days. The aim of this test system is to allow the detection of multiple effects using biomarker analysis. Feasibility of this assay for detection of estrogenic effects by vitellogenin 1 mRNA induction has been demonstrated (Muncke and Eggen, 2006; Muncke et al., 2007).

The zebrafish test of mechanosensory lateral line is normally used in neurotoxicity and yield a very clear idea about the effects of the exposure to chemicals in embryonic development. This method consists on the study of mechanoreceptors found in an interconnected network between head and body. In recent years a large number of genes related to these sensory cells have been reported (Li et al., 2010), which allows through molecular approaches the early detection of dysfunctions and problems. This method may serve as a test for detecting chemicals with effects in development of neurosensory function and detection of variations in gene expression can further be used to discern different mechanisms of action of toxic compounds. Other feasible method proposed for detecting neurodevelopmental toxicants using DarT was based on the estimation of the ratio between concentrations able to induce 50% of defects on primary and secondary motor neurons (recorded through immunostaining) and concentrations able to induce malformations on 50% of zebrafish population because neurodevelopmental endpoints are more sensitive than morphological ones and therefore ratios lower than 1 seem to be indicatives of neurodevelopmental toxicants (Muth-Köhne et al., 2012).

Transcriptomics, as previously reported for EST and WEC, can be also a good tool to enhance DarT performance because gene expression signatures may be relevant points for defining biomarkers of embryotoxicity, as was demonstrated with two independent sets of chemicals as caffeine, carbamazepine, retinoic acid, valproic acid, D-mannitol and saccharin (Hermsen et al., 2013), and 2,4-dichlorophenol, 3,4-dichloroaniline, pentachlorophenol, and cadmium chloride (Sawle at al., 2010).

Batteries of assays and integrated approaches to testing and assessment for testing developmental toxicity

There is a unanimous consensus in the scientific community for considering that such a biologically complex model as embryonic development will be difficult to cover with a single alternative method for testing toxicity. On the contrary, for a full assessment of the developmental toxicity at least one robust well-established alternative method should be addressed for each of the key events that warrants a successful embryonic development. In this context, the integration and translation of the data derived from multiple methods and sources will help for reliable chemical safety assessment. This integratory strategy is called IATA. OECD defines IATA in the website devoted to this concept[2] as *"a pragmatic, science-based approach for chemical hazard characterization that relies on an integrated analysis of existing information coupled with the generation of new information using testing strategies."*

Adverse Outcome Pathway (AOP) is a plausible piece of information of a sequential chain that describes causally linked events at different levels of biological organization (molecular, organelle, cellular, tissue, organ, organism, and population) that lead to an adverse effect of health. AOPs are becoming in a pivotal tool for assessing chemical safety. The concept of AOP should be fully integrated into the development of IATA. Indeed, OECD has published a document entitled "*Guidance document for the use of adverse outcome pathways in developing integrated approaches to testing and assessment (IATA)*" (OECD, 2016). This document clearly establishes that the concept of AOP can be applied as a framework in the development of IATA since it would allow either assessing in a structured way the existing information for getting conclusions on the hazard or, alternatively, to identify the gaps of information that prevents a robust conclusion, suitable for regulatory purposes, about such hazards.

Immediately below of IATA in level of complexity we find the strategy of toxicity testing using a battery of assays and further analysis of the weight of evidence. In other words, this strategy is based on collecting information of

1. https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/dev-tox/seazit/index.html.

2. http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm.

single tests that in an isolated way would not allow a conclusion as regard the hazardousness of a single chemical but when all together the results point into the same direction, the evidence allows to get a robust conclusion as regard the hazardousness. In the next paragraphs of this section, we present two different proposed IATAs based on alternative methods, one for testing developmental toxicity and another for the specific case of neurodevelopmental toxicity.

Example of battery of assays for testing developmental toxicity on the basis of alternative methods

Sogorb et al. (2014) proposed a stepwise methodology for testing developmental toxicity mainly based on methods alternative to animal testing (Fig. 7.5). This bottom-up strategy proposed a battery of assays for testing first embryotoxicity using short cellular tests, afterward, to enlarge the duration of the cellular tests. For these cellular tests, the methods described in former sections of this chapter will be adequate. Once the embryotoxicity has been appropriately assessed, the teratogenicity could be also assessed using a second bottom-up strategy with a test based on whole embryos (mainly zebrafish and WEC, although FETZ should be also appropriate). The careful assessment of the information collected in this battery of assays should allow concluding the lack of necessity of performing OECD Guideline 414 *in vivo* tests for testing teratogenicity, or alternatively, at least to design a reduced OECD 414 test, by example, with a single limit concentration.

Example of IATA for testing neurodevelopmental toxicity

In October 2016, an OECD/EFSA workshop was held in Brussels with attendants from 15 countries across the world, representing non-governmental organizations, academia, industry, and stakeholders from regulatory agencies (Fritsche et al., 2017). This meeting concluded that data requirements for *in vivo* developmental neurotoxicity testing are not sufficient to screen all the list of substances of concern. One of the aims of this workshop was to develop a consensus on a battery of alternative methods for testing developmental neurotoxicity that should be part of an IATA strategy resulting in an OECD guidance document on developmental neurotoxicity testing. This IATA should allow to screen and prioritize hazard assessment of chemicals of concern.

Fritsche et al. (2015) published a potential alternative developmental neurotoxicity testing strategy where several

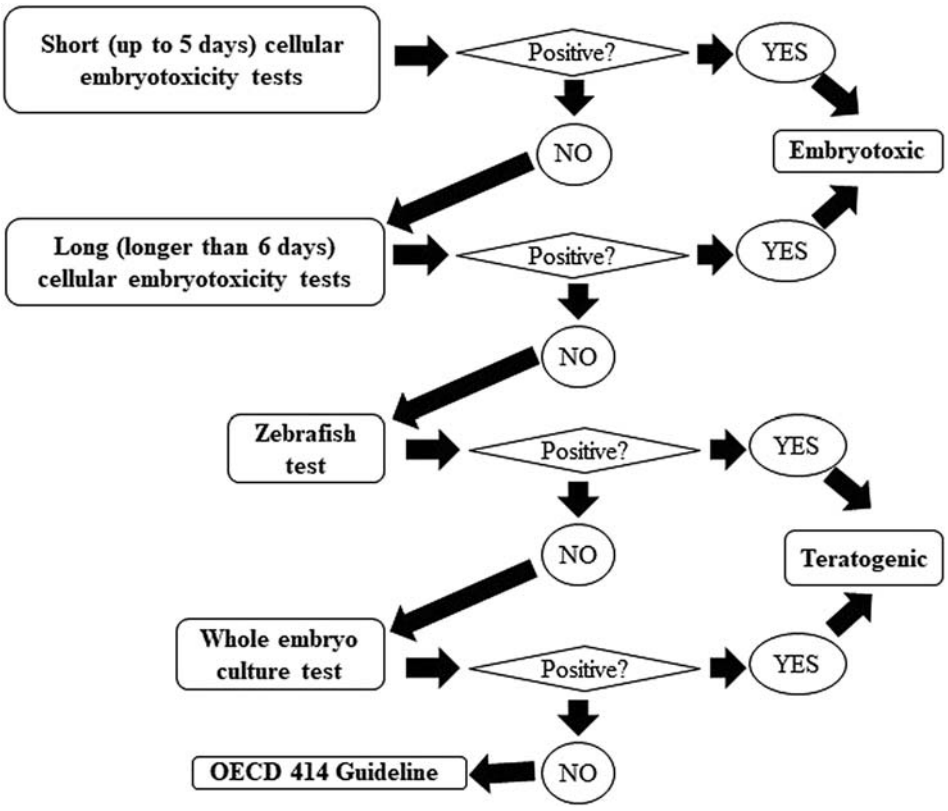

FIGURE 7.5 Battery of assays for testing developmental toxicity based on alternative methods. *Reproduced from Sogorb, M.A., Pamies, D., de Lapuente, J., Estevan, C., Estévez, J., Vilanova, E., 2014. An integrated approach for detecting embryotoxicity and developmental toxicity of environmental contaminants using* in vitro *alternative methods. Toxicol. Lett. 230 (2), 356–367.*

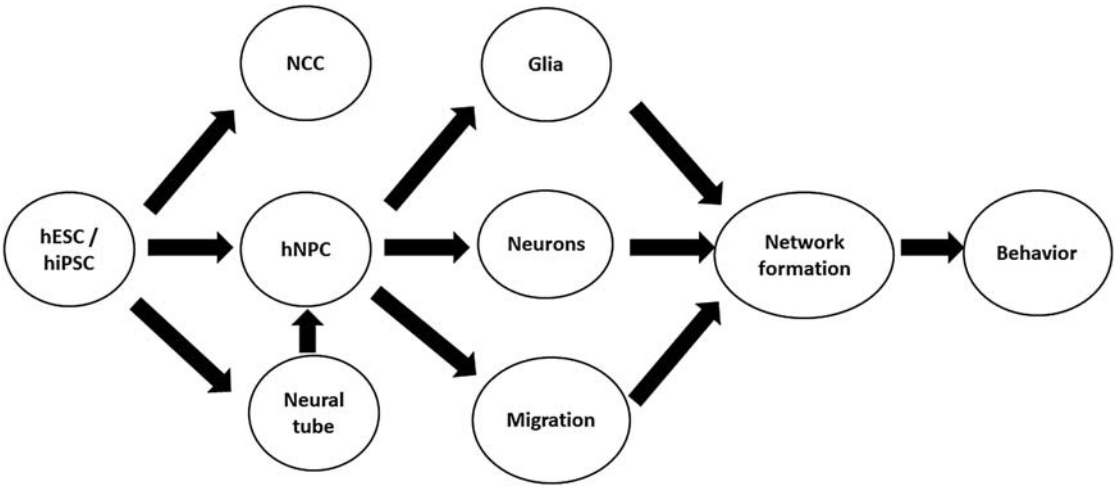

FIGURE 7.6 Potential alternative developmental neurotoxicity testing strategy. Each circle represents a major key event having to be tested by a human/zebrafish-based assay. *hESC*, Human embryonic stem cell; *hiPSC*, Human induced pluripotent stem cell; *hNPC*, Human neural progenitor cell; *NCC*, Neural crest cell. *Adapted from: Fritsche, E., Alm, H., Baumann, J., Geerts, L., Håkansson, H., Masjosthusmann, S., Witters, H., 2015. Literature review on* in vitro *and alternative developmental neurotoxicity (DNT) testing methods. EFSA supporting publication EN778, 186.*

major key events should be tested using human- or zebrafish-based assays (Fig. 7.6). Based on this developmental neurotoxicity testing strategy the above stated workshop adopted a proposal as basis for developing an IATA for testing developmental neurotoxicity as shown in Fig. 7.7 (OECD, 2017). In the moment of drafting this chapter (February 2021), to our knowledge, this IATA is still in process of validation and no official OECD Guideline at this respect has been published.

In the first tier, called tier 0, toxico-/pharmacokinetic information should be modeled in order to determine the concentrations that should be tested in the following tiers, or in other words, the physiologically relevant concentrations. This tier should also provide information about whether some metabolite, in addition to the parental chemical, must be also tested or not.

Tier 1 considers the use of a battery of tests using human cells (Table 7.6). This battery of tests should ideally reveal the most sensitive endpoint that will point toward the most probable mode-of-action of the presumable neurodevelopmental disturbance. This most sensitive endpoint would allow streamlining the test of the tier 2, for example, with zebrafish or even in a future targeted study *in vivo* with rodents in the tier 4, and especially whether the most sensitive endpoint is confirmed before in the *in vitro* assays in rodents considered in tier 3. Obviously, when the effect is specific for humans there is no necessity to take tier 4. However, when tiers 1 and 2 suggest interspecies differences in sensitivity, the *in vivo* rat studies should be taken in order to confirm the relevance of the *in vitro* results in whole organisms.

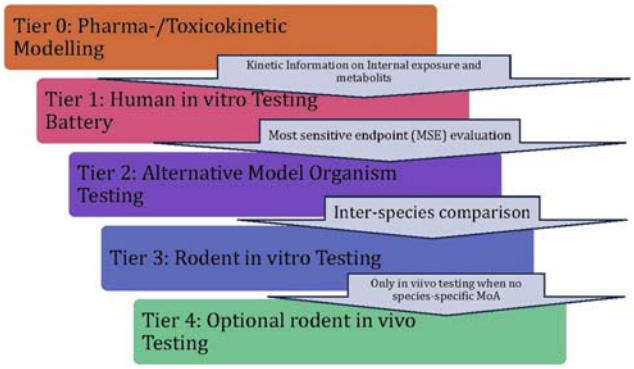

FIGURE 7.7 Proposal for an IATA for testing developmental neurotoxicity. Taken from: OECD. 2017. Report on Integrated Testing Strategies for the identification and evaluation of chemical hazards associated with the developmental neurotoxicity (DNT), to facilitate discussions at the Joint EFSA/OECD Workshop on DNT. Background document on integrated testing strategies for the identification and evaluation of chemical hazards associated with the developmental neurotoxicity (DNT), to facilitate discussions at the Joint EFSA/OECD Workshop on DNT. OECD Environment, Health and Safety Publications Series on Testing and Assessment No. 260.

Concluding remarks and future directions

Embryonic development is a very complex process that includes a number of coordinated steps that account in several stages. The alteration of whatever of these processes due to the action of chemicals might potentially suppose an embryotoxic-teratogenic effect. Due to the complexity of the embryonic development, the whole process cannot be

TABLE 7.6 A selection of assays covering neurodevelopmental endpoints as a ground for tier 1 in the proposal of IATA showed in Fig. 7.6.

Test name	Test system	Endpoint measured
NEP differentiation	hESC differentiation to NEP-containing neural rosettes	NEP differentiation resembling neural tube formation
NPC proliferation (Neurosphere assay)	hNPC	NPC proliferation necessary for brain growth
NPC proliferation	ReNcell CX	NPC proliferation necessary for brain growth
NPC proliferation	hESC	NPC proliferation necessary for brain growth
NPC apoptosis (Neurosphere assay)	hNPC	NPC apoptosis
NPC apoptosis	ReNcell CX	NPC apoptosis
NPC/Radial glia migration (Neurosphere assay)	hNPC	Radial migration
NCC migration (MINC assay)	hESC-derived NCC	NCC migration
Astrocyte differentiation	hESC-NPC astrocyte differentiation	Astrocyte differentiation
Astrocyte differentiation (Neurosphere assay)	hNPC-based astrocyte differentiation	Astrocyte differentiation
Oligodendrocyte differentiation (Neurosphere assay)	hNPC-based oligodendrocyte differentiation	Oligodendrocyte differentiation
Oligodendrocyte differentiation	hESC-based methods without compound testing	Oligodendrocyte differentiation
Neuronal differentiation	hESC-based neuronal differentiation	Neuron differentiation
Neuronal differentiation (Neuroshpere assay)	hNPC-based neuronal differentiation	Young, β−III−Tubulin + neuron differentiation
Neuronal differentiation (LUHMES assay)	LUHMES cell-based neuronal differentiation	Dopaminergic neuron differentiation
Neurite outgrowth	hESC-based neurite assay	Neurite outgrowth
Neurite outgrowth (Neurosphere assay)	hNPC-based neurite assay	Neurite outgrowth
Neurite outgrowth (Neurosphere assay)	hNPC-based neurite assay	Neurite outgrowth
Neuronal network formation	hESC-based neuronal networks	Electrical activity
Peripheral neurotoxicity	hiPSC-derived peripheral neurons	Neurogenesis

NPC, Neural progenitor cell; *NEP,* Neuroepithelial cell; *hESC,* Human embryonic stem cell; *hNPC,* human neural progenitor cell.

covered with a single alternative *in vitro* model and therefore the toxicity to development must be studied with a battery of assays covering each of the stages of the embryonic development (Sogorb et al., 2014).

This approach based on a battery of tests has already probed efficacy with the correct detection of 11 of 12 compounds tested following a battery of assays including (among others) EST and DarT (Piersma et al., 2013) and in a grouping context where 2-methylhexanoic acid, monomethyl phthalate, and monobutyltin trichloride were predicted as non- or weak developmental toxicants and differentiated from the structurally related developmental toxicants valproic acid, monoethylhexyl phthalate, and tributyltin chloride (Kroese et al., 2015). Another example of testing the safety of chemicals using a battery of *in vitro* assays was provided by Harrill et al. (2018). They tested 67 reference chemicals in a battery of high content imaging and microplate reader−based assays that evaluate neural progenitor cell proliferation and apoptosis, neurite maturation, synaptogenesis, and initiation/outgrowth. This assay battery yielded 87% and 71% of sensitivity for categorizing neurodevelopmental and non-developmental toxicants, respectively.

Organ-on-a-chip models are devices that allow relevant dynamic coculture systems. This methodology is a novel technology reproducing physiological functions of *in vivo* tissue more accurately than conventional cell-based model systems. Several US and European companies trade organs-on-a-chip for studies with organs as lung, liver, kidney,

intestine, blood vessel, or blood brain barrier. These organs-on-a-chip allow the study of physiological, pharmacological, and toxicological studies with these organs. However, to our knowledge, no organ-on-a-chip was marketed for study of the fetal and embryonic development. The development of fetal-placental organ-on-a-chip with immortalized human cells would be strongly desirable since it would allow studying the effects of chemicals on the process in physiologically relevant conditions, considering, by example, the bioavailability of chemicals after crossing placenta barrier or the role of placenta in biotransformation of chemicals.

There are several cases illustrating the power of batteries of alternative assays for assessing chemical safety. However, there is still work to do before the results of these batteries of assays could have regulatory impact. In this sense, the development and validation of AOP-based IATAs as shown above for the case of developmental neurotoxicity are strongly advisable and could have a notable economic and bioethical impact in the costs of safety assessment. The coordinated work of European and US Agencies involved in the assessment of safety as EPA, FDA, ECHA, EMA, EFSA, and others must play a pivotal role in the development of these demanded IATAs based on alternative methods with regulatory acceptance.

References

Aikawa, N., 2020. A novel screening test to predict the developmental toxicity of drugs using human induced pluripotent stem cells. J. Toxicol. Sci. 45 (4), 187−199.

Anichtchik, O.V., Kaslin, J., Peitsaro, N., et al., 2004. Neurochemical and behavioural changes in zebrafish *Danio rerio* after systemic administration of 6-hydroxydopamine and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. J. Neurochem. 88, 443−453.

Bacchetta, R., Mantecca, P., Andrioletti, M., et al., 2008. Axial-skeletal defects caused by Carbaryl in *Xenopus laevis* embryos. Sci. Total Environ. 392, 110−118.

Bantle, J.A., Finch, R.A., Burton, D.T., et al., 1996. FETAX interlaboratory validation study: phase III-Part 1 testing. J. Appl. Toxicol. 16, 517−528.

Bantle, J.A., Dumont, J.N., Finch, R.A., et al., 1998. Atlas of Abnormalities: A Guide for the Performance of FETAX, second ed. Oklahoma State University Press.

Bantle, J.A, Finch, R.A, Fort, D.J, et al., 1999. Phase III interlaboratory study of FETAX. Part 3. FETAX validation using 12 compounds with and without an exogenous metabolic activation system. J. Appl. Toxicol. 19, 447−472.

Bosisio, S., Fortaner, S., Bellinetto, S., et al., 2009. Developmental toxicity, uptake and distribution of sodium chromate assayed by frog embryo teratogenesis assay-Xenopus (FETAX). Sci. Total Environ. 407, 5039−5045.

Bretaud, S., Lee, S., Guo, S., 2004. Sensitivity of zebrafish to environmental toxins implicated in Parkinson's disease. Neurotoxicol. Teratol. 26, 857−864.

Chesnut, M., Muñoz, L.S., Harris, G., et al., 2019. In vitro and in silico models to study mosquito-borne Flavivirus neuropathogenesis, prevention, and treatment. Front. Cell Infect. Microbiol. 9, 223.

de Jong, E., van Beek, L., Piersma, A.H., 2012. Osteoblast differentiation of murine embryonic stem cells as a model to study the embryotoxic effect of compounds. Toxicol. In Vitro 26, 970−978.

Dimopoulou, M., Verhoef, A., Gomes, C.A., et al., 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., et al., 2015. In vitro generation of human pluripotent stem cell derived lung organoids. Elife 24 (4), e05098.

EURL ECVAM-DB-ALM, 2010a. Method Summary No. 113: Embryonic Stem Cell Test (EST). Available in: http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/DBALM/LATEST/online/DBALM_docs/113_M_Embryonic%20Stem%20Cell%20Test.pdf.

EURL ECVAM-DB-ALM, 2010b. Method Summary No. 122: The Micromass Test. Available in: http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/DBALM/LATEST/online/DBALM_docs/122_M_The%20Micromass%20Test.pdf.

EURL ECVAM-DB-ALM, 2010c. Method Summary No. 123: Embryotoxicity Testing in Post-Implantation Whole Embryo Culture (WEC) - Method of Piersma. Available in: http://cidportal.jrc.ec.europa.eu/ftp/jrc-opendata/EURL-ECVAM/datasets/DBALM/LATEST/online/DBALM_docs/123_P_Embryotoxicity%20Testing%20in%20Post-Implantation%20Whole%20Embryo%20Culture.pdf.

Fort, D.J., Stover, E.L., Bantle, J.A., et al., 1998. Phase III interlaboratory study of FETAX, Part 2: interlaboratory validation of an exogenous metabolic activation system for frog embryo teratogenesis assay-Xenopus (FETAX). Drug Chem. Toxicol. 21, 1−14.

Fort, D.J., Stover, E.L., Farmer, D.R., et al., 2000. Assessing the predictive validity of frog embryo teratogenesis assay-Xenopus (FETAX). Teratog. Carcinog. Mutagen. 20, 87−98.

Fort, D.J., Rogers, R.L., Thomas, J.H., et al., 2004. Comparative sensitivity of Xenopus tropicalis and *Xenopus laevis* as test species for the FETAX model. J. Appl. Toxicol. 24, 443−457.

Freedman, B.S., Brooks, C.R., Lam, A.Q., et al., 2015. Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids. Nat. Commun. 6, 8715.

Fritsche, E., Alm, H., Baumann, J., et al., 2015. Literature Review on in Vitro and Alternative Developmental Neurotoxicity (DNT) Testing Methods. EFSA supporting publication 2015:EN-778, p. 186. Available in: https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/sp.efsa.2015.EN-778.

Fritsche, E., Crofton, K.M., Hernandez, A.F., et al., 2017. OECD/EFSA workshop on developmental neurotoxicity (DNT): the use of non-animal test methods for regulatory purposes. ALTEX 34 (2), 311−315.

Froehlicher, M., Liedtke, A., Groh, K.J., et al., 2009. Zebrafish (*Danio rerio*) neuromast: promising biological endpoint linking developmental and toxicological studies. Aquat. Toxicol. 95, 307−319.

Genschow, E., Scholz, G., Brown, N., et al., 2000. Development of prediction models for three in vitro embryotoxicity tests in an ECVAM validation study. Vitro Mol. Toxicol. 13, 51−66.

Genschow, E., Spielmann, H., Scholz, G., et al., 2002. The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. European

Centre for the Validation of Alternative Methods. Altern. Lab. Anim. 30, 151–176.

Genschow, E., Spielmann, H., Scholz, G., et al., 2004. Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. Altern. Lab. Anim. 32, 209–244.

Harrill, J.A., Freudenrich, T., Wallace, K., et al., 2018. Testing for developmental neurotoxicity using a battery of in vitro assays for key cellular events in neurodevelopment. Toxicol. Appl. Pharmacol. 354, 24–39.

Harris, G., Eschment, M., Orozco, S.P., et al., 2018. Toxicity, recovery, and resilience in a 3D dopaminergic neuronal in vitro model exposed to rotenone. Arch. Toxicol. 92 (8), 2587–2606.

Hayess, K., Riebeling, C., Pirow, R., et al., 2013. The DNT-EST: a predictive embryonic stem cell-based assay for developmental neurotoxicity testing in vitro. Toxicology 314, 135–147.

Hermsen, S.A., Pronk, T.E., van den Brandhof, E.J., et al., 2013. Transcriptomic analysis in the developing zebrafish embryo after compound exposure: individual gene expression and pathway regulation. Toxicol. Appl. Pharmacol. 272, 161–171.

Hogberg, H.T., Bressler, J., Christian, K.M., et al., 2013. Toward a 3D model of human brain development for studying gene/environment interactions. Stem Cell Res. Ther. 4 (Suppl. 1), S4.

Hogberg, H.T., de Cássia da Silveira E Sá, R., Kleensang, A., et al., 2021. Organophosphorus flame retardants are developmental neurotoxicants in a rat primary brainsphere in vitro model. Arch. Toxicol. 95 (1), 207–228.

Hohwieler, M., Illing, A., Hermann, P.C., et al., 2017. Human pluripotent stem cell-derived acinar/ductal organoids generate human pancreas upon orthotopic transplantation and allow disease modelling. Gut 66 (3), 473–486.

Joshi, P., Kang, S.Y., Yu, K.N., et al., 2020. High-content imaging of 3D-cultured neural stem cells on a 384-pillar plate for the assessment of cytotoxicity. Toxicol. In Vitro 65, 104765.

Kim, J., Koo, B.K., Knoblich, J.A., 2020. Human organoids: model systems for human biology and medicine. Nat. Rev. Mol. Cell Biol. 21 (10), 571–584.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., et al., 1995. Stages of embryonic development of the zebrafish. Dev. Dynam. 203, 253–310.

Kroese, E.D., Bosgra, S., Buist, H.E., et al., 2015. Evaluation of an alternative in vitro test battery for detecting reproductive toxicants in a grouping context. Reprod. Toxicol. 55, 11–19.

Laale, H.W., 1977. Biology and use of zebrafish, brachydanio-rerio in fisheries research. J. Fish. Biol. 10, 121–173.

Lancaster, M.A., Knoblich, J.A., 2014. Generation of cerebral organoids from human pluripotent stem cells. Nat. Protoc. 9 (10), 2329–2340.

Lauschke, K., Rosenmai, A.K., Meiser, I., et al., 2020. A novel human pluripotent stem cell-based assay to predict developmental toxicity. Arch. Toxicol. 94 (11), 3831–3846.

Lee, J.H., Park, S.Y., Ahn, C., et al., 2019. Pre-validation study of alternative developmental toxicity test using mouse embryonic stem cell-derived embryoid bodies. Food Chem. Toxicol. 123, 50–56.

Leslie, J.,L., Huang, S., Opp, J.S., et al., 2015. Persistence and toxin production by *Clostridium difficile* within human intestinal organoids result in disruption of epithelial paracellular barrier function. Infect. Immun. 83 (1), 138–145.

Li, H., Kloosterman, W., Fekete, D.M., 2010. MicroRNA-183 family members regulate sensorineural fates in the inner ear. J. Neurosci. 30, 3254–3263.

Liedtke, A., Muncke, J., Rüfenacht, K., Eggen, R.I., 2008. Molecular multi-effect screening of environmental pollutants using the MolDarT. Environ. Toxicol. 23, 59–67.

Linney, E., Upchurch, L., Donerly, S., 2004. Zebrafish as a neurotoxicological model. Neurotoxicol. Teratol. 26, 709–718. Erratum in: Neurotoxicol. Teratol. 2005 27, 175.

Longo, M., Zanoncelli, S., Della Torre, P., et al., 2008. Investigations of the effects of the antimalarial drug dihydroartemisinin (DHA) using the Frog Embryo Teratogenesis Assay-Xenopus (FETAX). Reprod. Toxicol. 25, 433–441.

Luz, A.L., Tokar, E.J., 2018. Pluripotent stem cells in developmental toxicity testing: a review of methodological advances. Toxicol. Sci. 165 (1), 31–39.

McCracken, K.W., Catá, E.M., Crawford, C.M., et al., 2014. Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nature 516 (7531), 400–404.

Muncke, J., Eggen, R.I., 2006. Vitellogenin 1 mRNA as an early molecular biomarker for endocrine disruption in developing zebrafish (*Danio rerio*). Environ. Toxicol. Chem. 25, 2734–2741.

Muncke, J., Junghans, M., Eggen, R.I., 2007. Testing estrogenicity of known and novel (xeno)estrogens in the MolDarT using developing zebrafish (*Danio rerio*). Environ. Toxicol. 22, 185–193.

Muth-Köhne, E., Wichmann, A., Delov, V., et al., 2012. The classification of motor neuron defects in the zebrafish embryo toxicity test (ZFET) as an animal alternative approach to assess developmental neurotoxicity. Neurotoxicol. Teratol. 34, 413–424.

Nagel, R., Bresch, H., Caspers, N., et al., 1991. Effect of 3,4-dichloroaniline on the early life stages of the zebrafish (Brachydanio rerio): results of a comparative laboratory study. Ecotoxicol. Environ. Saf. 21, 157–164.

NICEATM (National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods), 2000. Background Review Document Frog Embryo Teratogenesis Assay - Xenopus (FETAX). Available in: http://ecvam.niehs/nih/gov/methods/fetaxdoc.

Organization for Economic Cooperation and Development (OECD), 2016. Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA). OECD Environment. Health and Safety Publications Series on Testing & Assessment No. 260. Available in: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO(2016)67/&doclanguage=en.

Organization for Economic Co-operation and Development (OECD), 2017. Report on Integrated Testing Strategies for the Identification and Evaluation of Chemical Hazards Associated With the Developmental Neurotoxicity (DNT), to Facilitate Discussions at the Joint EFSA/OECD Workshop on DNT. Background Document on Integrated Testing Strategies for the Identification and Evaluation of Chemical Hazards Associated with the Developmental Neurotoxicity (DNT), to Facilitate Discussions at the Joint EFSA/OECD Workshop on DNT. OECD Environment. Health and Safety Publications Series on Testing and Assessment No. 260. Available in: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=ENV/JM/MONO%282017%294/ANN1&docLanguage=En.

Palmer, J.A., Smith, A.M., Egnash, L.A., et al., 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 (4), 343−363.

Pamies, D., Hartung, T., 2017. 21st century cell culture for 21st century toxicology. Chem. Res. Toxicol. 30 (1), 43−52.

Pamies, D., Hartung, T., Hogberg, H.T., 2017a. Biological and medical applications of a brain-on-a-chip. Exp. Biol. Med. 239 (9), 1096−1107.

Pamies, D., Bal-Price, A., Simeonov, A., et al., 2017b. Good Cell Culture Practice for stem cells and stem-cell-derived models. ALTEX 34 (1), 95−132.

Pamies, D., Barreras, P., Block, K., et al., 2017c. A human brain microphysiological system derived from induced pluripotent stem cells to study neurological diseases and toxicity. ALTEX 34 (3), 362−376.

Pamies, D., Bal-Price, A., Chesné, C., et al., 2018a. Advanced Good Cell Culture Practice for human primary, stem cell-derived and organoid models as well as microphysiological systems. ALTEX 35 (3), 353−378.

Pamies, D., Block, K., Lau, P., et al., 2018b. Rotenone exerts developmental neurotoxicity in a human brain spheroid model. Toxicol. Appl. Pharmacol. 354, 101−114.

Pamies, D., Zurich, M.G., Hartung, T., 2020a. Organotypic models to study human glioblastoma: studying the beast in its ecosystem. iScience 23 (10), 101633.

Pamies, D., Leist, M., Coecke, S., et al., 2020b. Good Cell and Tissue Culture Practice 2.0 (GCCP 2.0) - draft for stakeholder discussion and call for action. ALTEX 37 (3), 490−492.

Paşca, A.M., Sloan, S.A., Clarke, L.E., et al., 2015. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. Nat. Methods 12 (7), 671−678.

Piersma, A.H., Bosgra, S., van Duursen, M.B., et al., 2013. Evaluation of an alternative in vitro test battery for detecting reproductive toxicants. Reprod. Toxicol. 38, 53−64.

Robinson, J.F., Verhoef, A., Pennings, J.L., et al., 2012a. A comparison of gene expression responses in rat whole embryo culture and in vivo: time-dependent retinoic acid-induced teratogenic response. Toxicol. Sci. 126, 242−254.

Robinson, J.F., Tonk, E.C., Verhoef, A., et al., 2012b. Triazole induced concentration-related gene signatures in rat whole embryo culture. Reprod. Toxicol. 34, 275−283.

Romero, A.C., Del Río, E., Vilanova, E., et al., 2015. RNA transcripts for the quantification of differentiation allow marked improvements in the performance of embryonic stem cell test (EST). Toxicol. Lett. 238, 60−69.

Rovida, C., Hartung, T., 2009. Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals - a report by the transatlantic think tank for toxicology (t(4)). ALTEX 26, 187−208.

Ryan, K.R., Sirenko, O., Parham, F., et al., 2016. Neurite outgrowth in human induced pluripotent stem cell-derived neurons as a high-throughput screen for developmental neurotoxicity or neurotoxicity. Neurotoxicology 53, 271−281.

Sandström, J., Eggermann, E., Charvet, I., et al., 2017. Development and characterization of a human embryonic stem cell-derived 3D neural tissue model for neurotoxicity testing. Toxicol. In Vitro 38, 124−135.

Sawle, A.D., Wit, E., Whale, G., et al., 2010. An information-rich alternative, chemicals testing strategy using a high definition toxicogenomics and zebrafish (*Danio rerio*) embryos. Toxicol. Sci. 118, 128−139.

Shi, Y., Inoue, H., Wu, J.C., et al., 2017. Induced pluripotent stem cell technology: a decade of progress. Nat. Rev. Drug Discov. 16 (2), 115−130.

Sirenko, O., Parham, F., Dea, S., et al., 2019. Functional and mechanistic neurotoxicity profiling using human iPSC-derived neural 3D cultures. Toxicol. Sci. 167 (1), 58−76.

Sogorb, M.A., Pamies, D., de Lapuente, J., et al., 2014. An integrated approach for detecting embryotoxicity and developmental toxicity of environmental contaminants using in vitro alternative methods. Toxicol. Lett. 230, 356−367.

Spielmann, H., 2005. Predicting the risk of developmental toxicity from in vitro assays. Toxicol. Appl. Pharmacol. 207 (2 Suppl. l), 375−380.

Suzuki, N., Ando, S., Yamashita, N., et al., 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.

Suzuki, N., Yamashita, N., Koseki, N., et al., 2012. Assessment of technical protocols for novel embryonic stem cell tests with molecular markers (Hand1- and Cmya1-ESTs): a preliminary cross-laboratory performance analysis. J. Toxicol. Sci. 37, 845−851.

Takasato, M., Er, P.X., Chiu, H.S., et al., 2016. Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. Nature 536 (7615), 238.

Theunissen, P.T., Schulpen, S.H., van Dartel, D.A., et al., 2010. An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. Reprod. Toxicol. 29, 383−392.

Theunissen, P.T., Pennings, J.L., Robinson, J.F., et al., 2011. Time-response evaluation by transcriptomics of methylmercury effects on neural differentiation of murine embryonic stem cells. Toxicol. Sci. 122, 437−447.

Theunissen, P.T., Robinson, J.F., Pennings, J.L., et al., 2012a. Transcriptomic concentration-response evaluation of valproic acid, cyproconazole, and hexaconazole in the neural embryonic stem cell test (ESTn). Toxicol. Sci. 125, 430−438.

Theunissen, P.T., Robinson, J.F., Pennings, J.L., et al., 2012b. Compound-specific effects of diverse neurodevelopmental toxicants on global gene expression in the neural embryonic stem cell test (ESTn). Toxicol. Appl. Pharmacol. 262, 330−340.

Theunissen, P.T., Pennings, J.L., van Dartel, D.A., et al., 2013. Complementary detection of embryotoxic properties of substances in the neural and cardiac embryonic stem cell tests. Toxicol. Sci. 132, 118−130.

Whitesides, G.M., 2006. The origins and the future of microfluidics. Nature 442 (7101), 368−373.

Witt, G., Keminer, O., Leu, J., et al., 2020. An automated and high-throughput-screening compatible pluripotent stem cell-based test platform for developmental and reproductive toxicity assessment of small molecule compounds. Cell Biol. Toxicol. https://doi.org/10.1007/s10565-020-09538-0.

Yang, L., Ho, N.Y., Alshut, R., Legradi, J., et al., 2009. Zebrafish embryos as models for embryotoxic and teratological effects of chemicals. Reprod. Toxicol. 28, 245−253.

Zhang, C., Cao, J., Kenyon, J.R., et al., 2012. Development of a streamlined rat whole embryo culture assay for classifying teratogenic potential of pharmaceutical compounds. Toxicol. Sci. 127 (2), 535−546.

Zhong, X., Harris, G., Smirnova, L., et al., 2020. Antidepressant paroxetine exerts developmental neurotoxicity in an iPSC-derived 3D human brain model. Front. Cell. Neurosci. 14, 25.

Zurlinden, T.J., Saili, K.S., Rush, N., et al., 2020. Profiling the ToxCast library with a pluripotent human (H9) stem cell line-based biomarker assay for developmental toxicity. Toxicol. Sci. 174 (2), 189−209.