OXFORD

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

TOXICOLOGICAL SCIENCES, 2017, 1–11

doi: 10.1093/toxsci/kfx034
Advance Access Publication Date: February 10, 2017
Research Article

Exposure-Based Validation of an *In Vitro* Gastrulation Model for Developmental Toxicity Assays

Erica L. L. Warkus and Yusuke Marikawa[1]

Developmental and Reproductive Biology Graduate Program, Institute for Biogenesis Research, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii 96813

[1]To whom correspondence should be addressed at Developmental and Reproductive Biology Graduate Program, Institute for Biogenesis Research, John A. Burns School of Medicine, University of Hawaii, 651 Ilalo Street, Biosciences Building, 163A, Honolulu, HI 96813. Fax: (808) 692-1962. E-mail: marikawa@hawaii.edu.

ABSTRACT

Establishment of effective non-animal alternatives for developmental toxicity screening assays is desirable to ensure maternal and fetal health outcomes. Validation of such assays requires a comparison between the *in vitro* responses to chemical exposures and the *in vivo* impacts of the corresponding compounds at equivalent concentrations. Here, we investigated how the P19C5 gastrulation model responds to 24 compounds at specific concentrations, some of which are categorized as positive exposures based on previously observed detrimental effects on development *in vivo*, whereas others are categorized as negative exposures due to lack of effects *in vivo*. The P19C5 gastrulation model consists of *in vitro* morphogenesis of mouse stem cells aggregated into embryoid bodies (EBs), which recapitulates growth and axial elongation of early embryos during four days of three-dimensional culture. Adverse impacts of chemical exposures were defined as: death, impaired growth, and altered axial elongation of EBs. Ten out of 17 positive exposures caused adverse impacts on EBs. In contrast, only three out of 17 negative exposures adversely affected EBs, although two of the three diminished viability of somatic cell lines (NIH/3T3, HEK293, and JEG3), suggesting general cytotoxicity. Overall, the study showed that 24 out of 34 exposures impacted EB development in a manner concordant with the *in vivo* developmental effects. Validation of other alternative assays using the same set of chemical exposures will provide information on the strengths and weaknesses of each assay, and should help determine the most effective ensemble of assays to detect a wide range of developmentally toxic exposures.

Key words: morphogenesis; gastrulation model; exposure-based validation; stem cell; non-animal alternative; *in vitro*.

In vitro model systems, particularly those using pluripotent stem cells, are promising screening tools for chemical exposures that cause developmental toxicity. Pluripotent stem cells, such as embryonic stem (ES) cells, can differentiate into multiple tissue types *in vitro* and display embryo-like properties. Thus, pluripotent stem cells can recapitulate certain aspects of embryonic development and serve as *in vitro* models to demonstrate the impacts of developmentally toxic chemical exposures. ES cell tests, or ESTs, evaluate the developmental toxicity of compounds based on their inhibitory effects on ES cell differentiation (Riebeling *et al.*, 2012; Theunissen and Piersma, 2012), but each individual EST system recapitulates only a limited aspect of embryonic development, i.e., differentiation of cardiomyocytes, neurons, or osteoblasts. Thus, it is more likely that a panel of complementary *in vitro* systems—each representing distinct aspects of embryonic development—can comprehensively screen a broad range of chemical compounds that inflict developmental toxicity.

To assemble a panel of complementary assays, each assay must undergo a series of validation studies using compounds with known developmental toxicities to assess its applicability and limitations. However, selection of proper reference compounds to validate *in vitro* screening assays has been challenging, partly because developmental impact of compounds varies depending on timing, dose, and duration of exposure to embryos (Friedman, 2010; Jelínek, 2005). For example, a given compound can exert developmental toxicity when exposed at a high dose, but the same compound may pose no risk at lower doses

(Daston *et al.*, 2010). Therefore, dichotomic designation of compounds as developmental toxicants or non-toxicants is ineffectual in properly validating assays. To that end, Daston *et al.* (2014) have proposed an "exposure-based validation list" for developmental toxicity screening assays. This list (referred hereafter as the Daston list) consists of 20 positive exposures known to cause embryo–fetal death or structural malformation in rats, and 19 negative exposures, which have no adverse impacts on rat embryo development. Eleven of these compounds demonstrate both a positive exposure at a high concentration and a negative exposure at a low concentration. The Daston list is a significant step forward in validating *in vitro* assays for developmental toxicity screening.

The objective of the present study is to validate the *in vitro* gastrulation model of P19C5 stem cells using the compounds on the Daston list. P19C5 cells are mouse embryonal carcinoma stem cells that possess properties similar to the pluripotent epiblast lineage of postimplantation embryos. Three-dimensional culture of P19C5 cell aggregates results in spontaneous differentiation of mesendoderm along with steady increase in size and axial elongation (Lau and Marikawa, 2014). This morphogenetic transformation of P19C5 cell aggregates, or embryoid bodies (EBs), resembles gastrulation, the morphogenetic process of early embryonic development that creates the germ layers and elongated body shape along the anterior–posterior axis. Growth and axial elongation of the P19C5 gastrulation model are impaired by pharmacological inhibitors of the major developmental signals that are crucial for embryo body patterning, namely Wnt, Nodal, Fgf, and retinoic acid signaling pathways (Li and Marikawa, 2015). The morphogenesis of P19C5 EBs is also sensitive to various therapeutic drugs that are contraindicated for use during pregnancy (Warkus et al., 2016). Furthermore, valproic acid (VPA) at 0.8 mM, one of the positive exposures in the Daston list, affects the P19C5 morphogenesis, such that VPA-treated EBs are smaller, distorted, and less elongated (Li and Marikawa, 2016). These studies suggest that the P19C5 gastrulation model can serve as an effective *in vitro* tool to detect developmentally toxic chemical exposures.

Here, we validated the P19C5 gastrulation model using the 34 exposures compiled in the Daston list. While the morphology-based assay correctly classified many of the Daston exposures, the present study also revealed some limitations of the assay based on the cases where exposures were misclassified.

MATERIALS AND METHODS

Test compounds. Compounds of the Daston list (Daston et al., 2014) used in the present study are shown in Table 1. Four compounds in the Daston list, namely HEPP (positive exposure), SB-209770 (positive and negative exposures), tapentadol (negative exposure), and VPA (positive exposure), were not evaluated for the following reasons. HEPP and SB-209770 were unavailable from major chemical suppliers. Tapentadol was available from one supplier (Sigma-Aldrich, St Louis, Missouri) as a solution of 1.0 mg/mL dissolved in methanol (Catalog Number T-058). However, to achieve the concentration indicated in the Daston list (252 mg/L), about 25% of the culture medium would be methanol, which is not compatible with our assay (see below). The morphogenetic effect of VPA on the P19C5 gastrulation model has been evaluated previously, including the concentration cited in the Daston list (0.8 mM) (Li and Marikawa, 2016). Note that the concentration of desloratadine indicated in the Daston study (1.5 mM) (Daston et al., 2014) is apparently a typographical error, as the original study referenced therein (FDA, 2001) indicates 1.5 μM instead. Accordingly, desloratadine was evaluated at 1.5 μM as a negative exposure in the present study.

Cell culture. P19C5 cells, a subline of P19 mouse embryonal carcinoma cell line (Lau and Marikawa, 2014), were propagated in culture medium (Minimum Essential Medium Alpha with nucleosides and GlutaMAX Supplement [LifeTechnologies, Carlsbad, California], 2.5% fetal bovine serum, 7.5% newborn calf serum, 50 units/mL penicillin, and 50 μg/mL streptomycin). Embryoid bodies of P19C5 cells were generated according to the method previously described for P19 cell aggregates (Marikawa et al., 2009). Briefly, P19C5 cells were fully dissociated with Trypsin-EDTA, and suspended in culture medium containing 1% dimethyl sulfoxide (DMSO) at the density of 10 cells/μL with or without specific amount of a test compound (Figure 1). Drops (20 μL each) of cell suspension were spotted on the inner surface of Petri dish lids for hanging drop culture. NIH/3T3 (derived from mouse embryonic fibroblast), HEK293 (derived from human embryonic kidney), and JEG3 (derived from human choriocarcinoma) were obtained from the American Type Culture Collection (Manassas, Virginia), and propagated in culture medium (Dulbecco's Modified Eagle Medium, GlutaMAX Supplement [LifeTechnologies], 10% fetal bovine serum, 50 units/mL penicillin, and 50 μg/mL streptomycin).

Image analysis. Embryoid bodies were removed from hanging drops and grouped together for photography using an AxioCam MRm digital camera connected to an Axiovert 200 microscope with Hoffman modulation-contrast optics (Carl Zeiss, Thornwood, New York). Image files were converted to JPEG format and opened in ImageJ (http://rsb.infonihgov/ij). Morphological parameters of individual EBs were measured on ImageJ by tracing their circumference. In the previous studies (Lau and Marikawa, 2014; Li and Marikawa, 2015; Warkus *et al.*, 2016), tracing of EB circumference was performed manually using the polygon selection tool. In the present study, however, we formulated a series of program operations to enable tracing of EB circumference with less user-dependent variability. The details of the program operations are described in Supplementary Data. Briefly, the first set of operations (e.g., Image Calculator, Find Edges, Brightness/Contrast and Binary) was applied to convert the inside of EBs into solid black while the background became white. The second set of operations (e.g., Analyze Particles and Fit Splines) was applied to detect the outlines of blackened areas, corresponding to the perimeter of EBs. The third set of operations (e.g., Measure) was applied to selected regions of interest (ROI) to determine their morphological parameters, namely area and circularity ($=4\times\pi\times\text{area}/\text{perimeter}^2$). Measurements were exported to Microsoft Excel, where Elongation Distortion Index ($\text{EDI}=1/\text{circularity}-1$) was calculated. As described previously, area was used as a proxy for the size of EB, whereas EDI was used to gauge the extent of EB axial elongation (Warkus *et al.*, 2016).

Viability assay. The impact of drugs on cell proliferation and viability was evaluated using CellTiter-Glo Luminescent Cell Viability Assay system (Promega, Madison, Wisconsin), which determines the number of live cells in culture by measuring the amount of ATP as a quantitative proxy for the number of metabolically active cells. P19C5, NIH/3T3, HEK293, and JEG3 cells were seeded in 96-well plates at the density of 100, 250, 250, and 500 cells/well, respectively, and were cultured in the corresponding medium (100 μL/well) supplemented with 1% DMSO

TABLE 1. Compounds Used in the Present Study

Compound	Vendor	Catalog Number	Stock
Abacavir	Sigma-Aldrich	SML0089	50 mM in water
Acetazolamide	Sigma-Aldrich	A6011	50 mM in DMSO
All-trans retinoic acid	Sigma-Aldrich	R2625	200 μM in DMSO
Artesunate	Sigma-Aldrich	A3731	50 mM in DMSO
Butylparaben	Sigma-Aldrich	PHR1022	50 mM in DMSO
Caffeine	Sigma-Aldrich	C0750	1 M in water
Dabigatran	Santa Cruz Biotech.	sc-351724	20 mM in DMSO
Desloratadine	Sigma-Aldrich	D1069	1 mM in DMSO
Ethylene glycol	Sigma-Aldrich	324558	18 M (100%)
Fingolimod	Sigma-Aldrich	SML0700	20 mM in water
Glycolic acid	Sigma-Aldrich	G8284	1 M in water
Hydroxyurea	Sigma-Aldrich	H8627	100 mM in water
MEHP	Santa Cruz Biotech.	sc-396467	100 mM in DMSO
Methanol	Fisher Scientific	A412P-4	24.7 M (100%)
Methoxyacetate	Sigma-Aldrich	194557	1 M in water
Methylmercury	Sigma-Aldrich	442534	10 mM in DMSO
Nilotinib	Santa Cruz Biotech.	sc-202245	50 mM in DMSO
Oseltamivir	Santa Cruz Biotech.	sc-208135	50 mM in water
Propylene glycol	Sigma-Aldrich	PHR1051	13.62 M (100%)
Ramelteon	Santa Cruz Biotech.	sc-219934	20 mM in DMSO
Saccharin	Sigma-Aldrich	240931	20 mM in water
Salicylic acid	Sigma-Aldrich	247588	1 M in DMSO
Zaleplon	Sigma-Aldrich	Z-004	3.28 mM in methanol
Zidovudine	Sigma-Aldrich	PHR1292	100 mM in DMSO

DMSO, dimethylsulfoxide; MEHP, mono(2-ethylhexyl) phthalate.

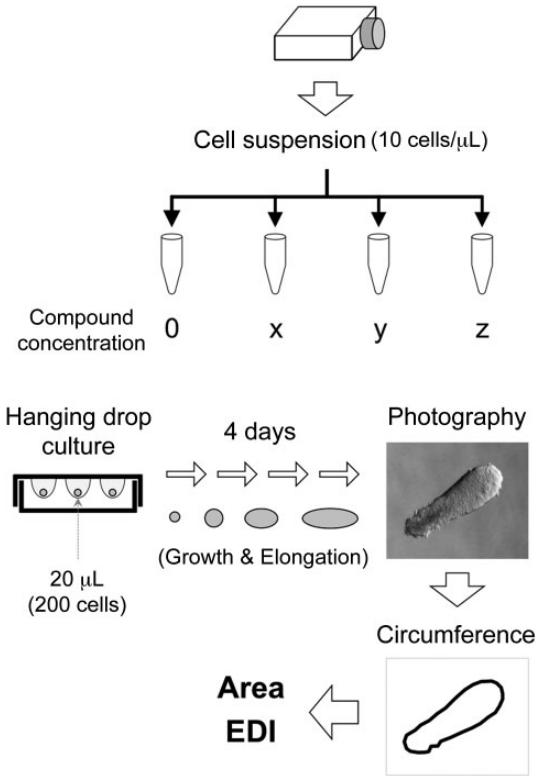

FIG. 1. The experimental scheme to examine morphogenetic impact of the compound exposures. See Materials and Methods section for details.

with or without specific amount of a test compound. After 4 days of culture, cells were treated with CellTiter-Glo Reagent for measurement of luminescence, as a readout of ATP amount, according to the manufacturer's instruction (Promega), using Gene Light 55 Luminometer (Microtech, Chiba, Japan). Cell seeding density was optimized through a series of pilot experiments, to confirm that cell numbers at the end of 4 days of culture were proportionate to intensities of luminescence. Relative cell number was calculated based on ratio of the luminescence intensity in compound-treated cells to that in non-treated cells of the same set of experiments. For each compound exposure, three sets of experiments were conducted as biological replicates.

Statistical analyses. Experiments to assess morphological impacts of compound exposures on EB morphology were conducted in three biological replicates using different collections of cell suspensions. For each replicate, 16 hanging drops were generated per exposure (a specific concentration of compound) in parallel with 16 control (i.e., no compound) hanging drops (Figure 1). Area and EDI of individual Day 4 EBs were normalized against the average of control EBs, and defined as relative area and relative EDI, respectively, expressed in percentages (i.e., averages of relative area and relative EDI of control EBs are 100%). Data from three replicates were compiled, and their averages are shown with 95% confidence intervals (Figure 2). Thus, a total of 45–48 EBs were scored for each exposure (1 or 2 EBs were occasionally lost or damaged during operations). To verify that observed effects on EBs were statistically significant, two-sample t-test was performed between compound-treated group and the matching control group. All morphological impacts that were defined as adverse in the present study (see below) were statistically significant ($P < .01$).

Experimental design. The overall experimental scheme is shown in Figure 1. In the present study, we examined several concentrations for each compound in addition to those in the Daston

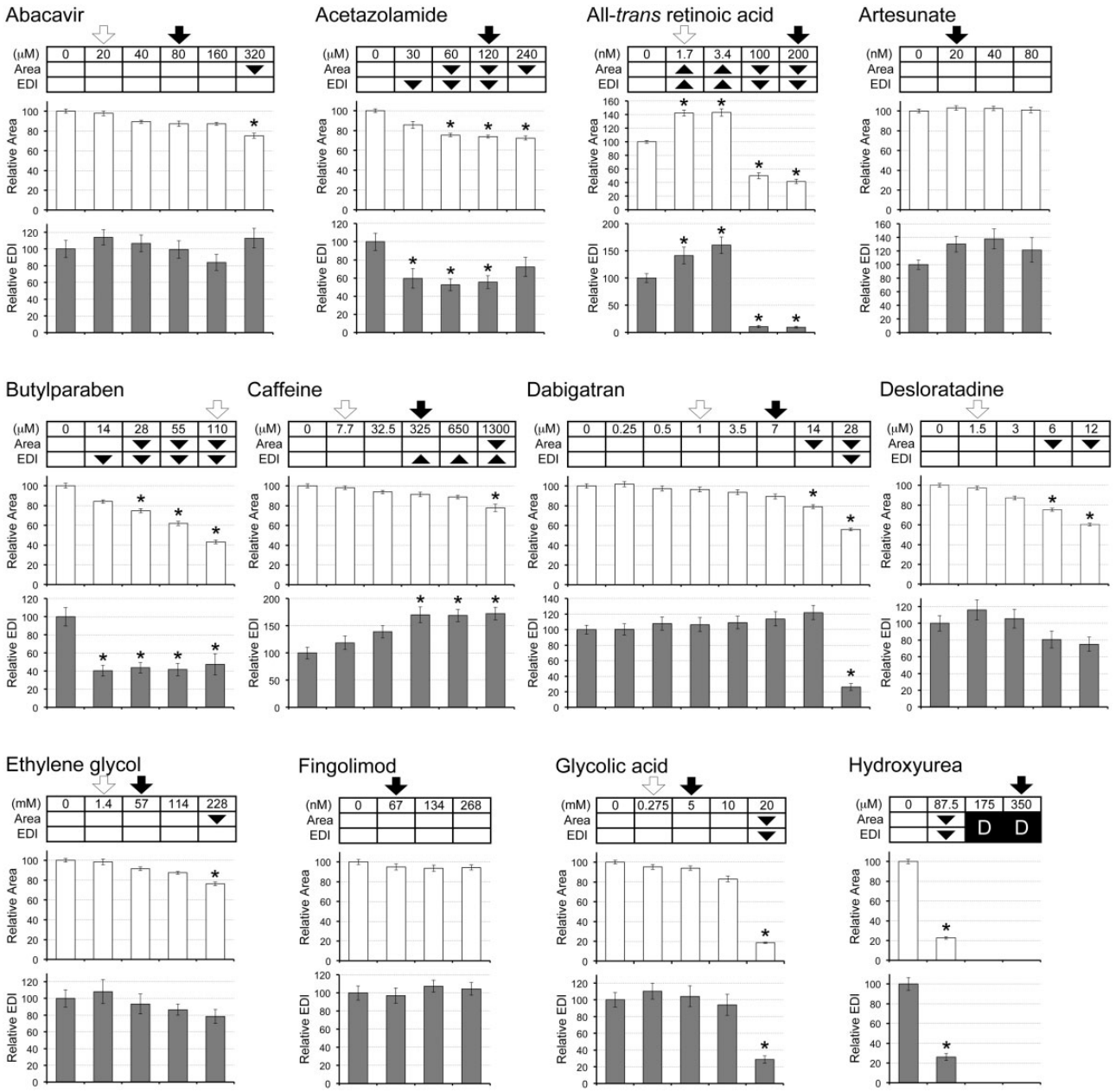

FIG. 2. Impact of the Daston compounds on P19C5 EB morphogenesis. For each compound, concentrations tested are indicated in the top row of the table with a summary of observed morphogenetic impact on EB area and EDI, indicated with upward arrowheads (increase) and downward arrowheads (reduction). No area or EDI value is available when EBs were dead (D) or cells did not aggregate (NA). Arrows above the table indicate negative (white arrow) and positive (black arrow) exposures as cited in the Daston list. Column graphs below the summary tables show averages of relative area (white columns; %) and relative EDI (gray columns; %), for the corresponding compound concentrations indicated above. Error bars show 95% confidence intervals. Asterisks indicate adverse impacts, which are defined in the present study as a change in average area by >20% or a change in average EDI by >40% relative to controls.

list (Daston et al., 2014). Particularly, when positive exposures in the Daston list had no apparent effect on P19C5 EBs, higher concentrations were also evaluated. EBs were observed daily for survival and integrity, and were photographed on the fourth day of culture (Day 4) for morphometric analyses. Impact of each compound exposure was assessed based on changes in area or EDI relative to control EBs. In the present study, a compound exposure was classified as having an adverse effect on EB morphogenesis when it caused either of the following three outcomes: (1) degeneration (i.e., death) of EBs at any time point in culture or cell aggregation failure, (2) a reduction in the average area by more than 20% relative to control EBs, and (3) a decrease or increase in the average EDI by more than 40% relative to control EBs. Note that an increase in the average EDI accompanied by an increase in the average area was not classified as adverse, because such condition suggests that the exposure promoted EB growth and morphogenesis rather than impairing EB development (see Discussion section).

RESULTS

Effects of the Positive and Negative Exposures on EB Morphogenesis

Morphogenetic impacts of compound exposures on P19C5 EBs are shown in Figure 2, and summarized in Tables 2 and 3 with

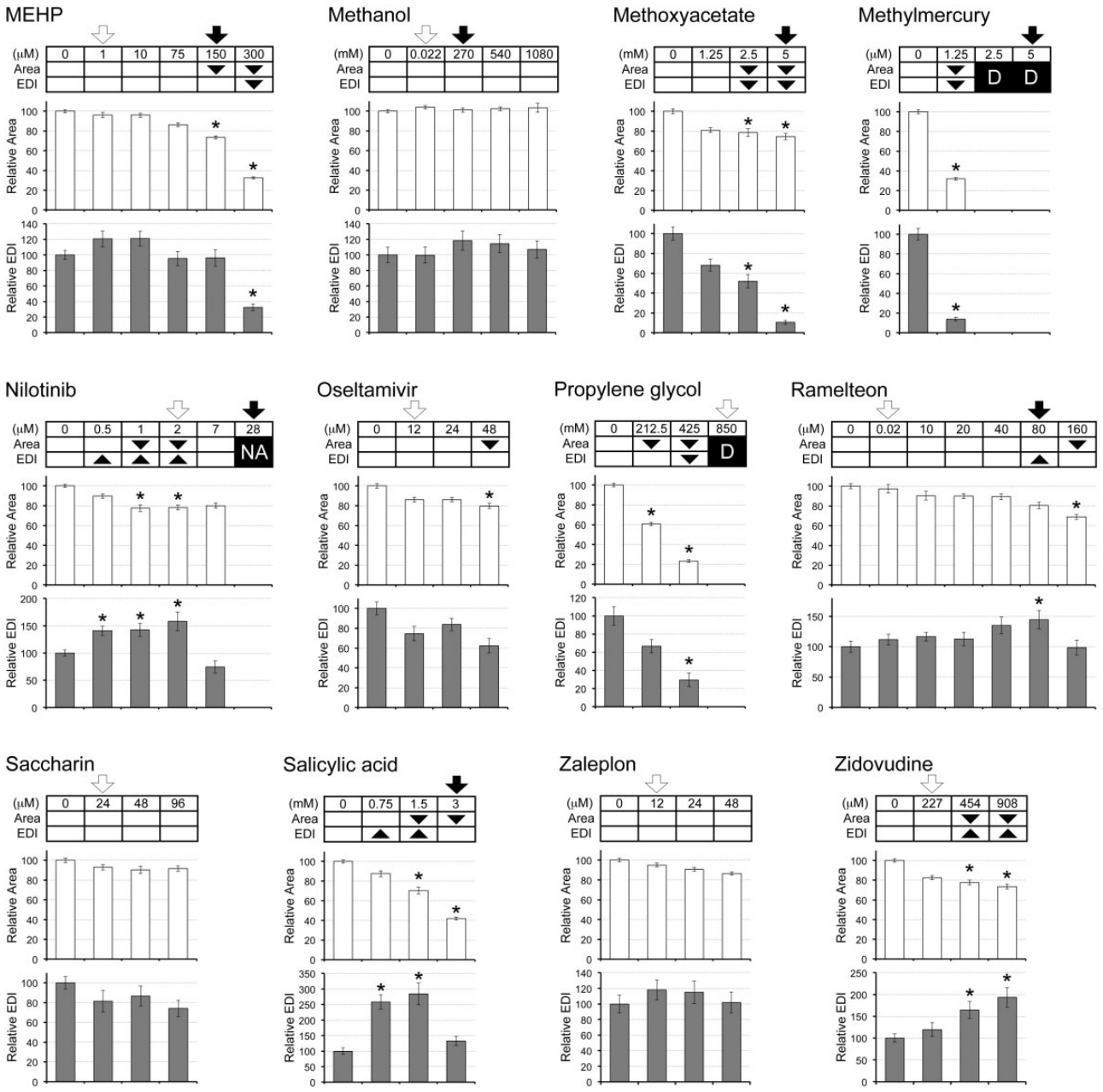

FIG. 2. Continued

respect to specific concentrations indicated in the Daston list. Sample images of EBs, namely those that were distinctly affected by exposures, are shown in Figure 3. Based on the classification criteria described in Materials and Methods section, 58.8% (10 out of 17) of the positive exposures tested had adverse impact on P19C5 EBs (Table 2). Specifically, acetazolamide (120 μM) reduced relative area by 25% and reduced relative EDI by 45%; all-*trans* retinoic acid (ATRA; 200 nM) reduced relative area by 60% and reduced relative EDI by 90%; caffeine (325 μM) increased relative EDI by 70%; hydroxyurea (350 μM) caused death; mono(2-ethylhexyl) phthalate (MEHP; 150 μM) reduced relative area by 25%; methoxyacetate (5 mM) reduced relative area by 25% and reduced relative EDI by 90%; methylmercury (5 μM) caused death; nilotinib (28 μM) prevented cell aggregation (without apparent cell death); ramelteon (80 μM) increased relative EDI by 45%; and salicylic acid (3 mM) reduced relative area by 60%. In contrast, only 17.6% (3 out of 17) of the negative exposures exhibited adverse effects on EB development (Table 3). Namely, butylparaben (110 μM) reduced relative area by 55%; nilotinib (2 μM) reduced relative area by 25% and increased relative EDI by 60%; and propylene glycol (850 mM) caused death. Overall, a significantly higher percentage of the positive exposures exhibited adverse impacts on EBs than the negative exposures ($P = .013$; Chi-square test).

No adverse effect was observed in the P19C5 gastrulation model in response to seven of the Daston list positive exposures, namely abacavir (80 μM), artesunate (20 nM), dabigatran (7 μM), ethylene glycol (57 mM), fingolimod (67 nM), glycolic acid (5 mM), and methanol (270 mM). Thus, we further examined whether EBs could be affected by these compounds at higher

TABLE 2. Adverse Impact of the Daston Positive Exposures on EB Morphology

Compound	Concentration	Adverse Impact	Comments
Abacavir	80 μM	No	Area reduction at 4× higher conc. (320 μM)
Acetazolamide	121 μM	Yes (120 μM)	Area and EDI reduction
ATRA	200 nM	Yes	Area and EDI reduction
Artesunate	20 nM	No	
Caffeine	325 μM	Yes	EDI increase
Dabigatran	7 μM	No	Area reduction at 2× higher conc. (14 μM)
Ethylene glycol	57 mM	No	Area reduction at 4× higher conc. (228 mM)
Fingolimod	67 nM	No	
Glycolic acid	5 mM	No	Area and EDI reduction at 4× higher conc. (20 mM)
Hydroxyurea	350 μM	Yes	Dead
MEHP	146 μM	Yes (150 μM)	Area reduction
Methanol	270 mM	No	
Methoxyacetic acid	5 mM	Yes	Area and EDI reduction
Methylmercury	5 μM	Yes	Dead
Nilotinib	28 μM	Yes	No aggregation
Ramelteon	81 μM	Yes (80 μM)	EDI increase
Salicylic acid	3 mM	Yes	Area reduction

ATRA, all-*trans* retinoic acid; MEHP, mono(2-ethylhexyl) phthalate.

TABLE 3. Adverse Impact of the Daston Negative Exposures on EB Morphology

Compound	Concentration	Adverse Impact	Comments
Abacavir	18 μM	No (20 μM)	
ATRA	1.7 nM	No	EDI increase accompanied by area increase (see text)
Butylparaben	110 μM	Yes	Area and EDI reduction
Caffeine	7.7 μM	No	
Dabigatran	1 μM	No	
Desloratadine	1.5 μM	No	Concentration listed as 1.5 mM in Daston et al. (2014)
Ethylene glycol	1.4 mM	No	
Glycolic acid	275 μM	No	
Methanol	22 μM	No	
MEHP	1 μM	No	
Nilotinib	2 μM	Yes	Area reduction, EDI increase
Oseltamivir	12 μM	No	
Propylene glycol	850 mM	Yes	Dead
Ramelteon	19 nM	No	
Saccharin	24 μM	No	
Zaleplon	12 μM	No	
Zidovudine	227 μM	No	

ATRA, all-*trans* retinoic acid; MEHP, mono(2-ethylhexyl) phthalate.

concentrations than those indicated in the Daston list. Artesunate, fingolimod, or methanol did not show any adverse effect on EB morphogenesis even at four times higher concentrations (Figure 2), suggesting that the P19C5 gastrulation model is unable to detect developmental toxicity of these compounds at physiologically relevant concentrations. However, abacavir at four times higher concentration (320 μM) reduced relative area by 25%; dabigatran at two times higher concentration (14 μM) reduced relative area by 21%; ethylene glycol at four times higher concentration (228 mM) reduced relative area by 25%; and glycolic acid at four times higher concentration (20 mM) reduced relative area by 80% and reduced EDI by 70%. This suggests that the P19C5 gastrulation model may still be able to detect developmentally toxic exposures of these compounds, albeit with slightly less sensitivity than *in vivo* assays.

Ten of the Daston compounds evaluated (abacavir, ATRA, caffeine, dabigatran, ethylene glycol, glycolic acid, MEHP, methanol, nilotinib, and ramelteon) are categorized with both positive and negative exposures depending on concentrations. Such compounds are the most useful to rigorously validate the sensitivity and the specificity of screening assays (Daston et al., 2014). According to the classification criteria described in Materials and Methods section, the P19C5 EB model was able to distinguish between positive and negative exposures for four of these compounds (ATRA, caffeine, MEHP, and ramelteon). If four times higher concentrations were also to be included for positive exposures, then additional four compounds (abacavir, dabigatran, ethylene glycol, and glycolic acid) can also be differentially classified by the P19C5 EB model.

While not directly related to the exposures indicated in the Daston list, additional concentrations were also evaluated for 12 compounds to determine their lowest-observed-adverse-effect level (LOAEL) and/or no-observed-adverse-effect level (NOAEL) with respect to morphogenetic impact. The results of

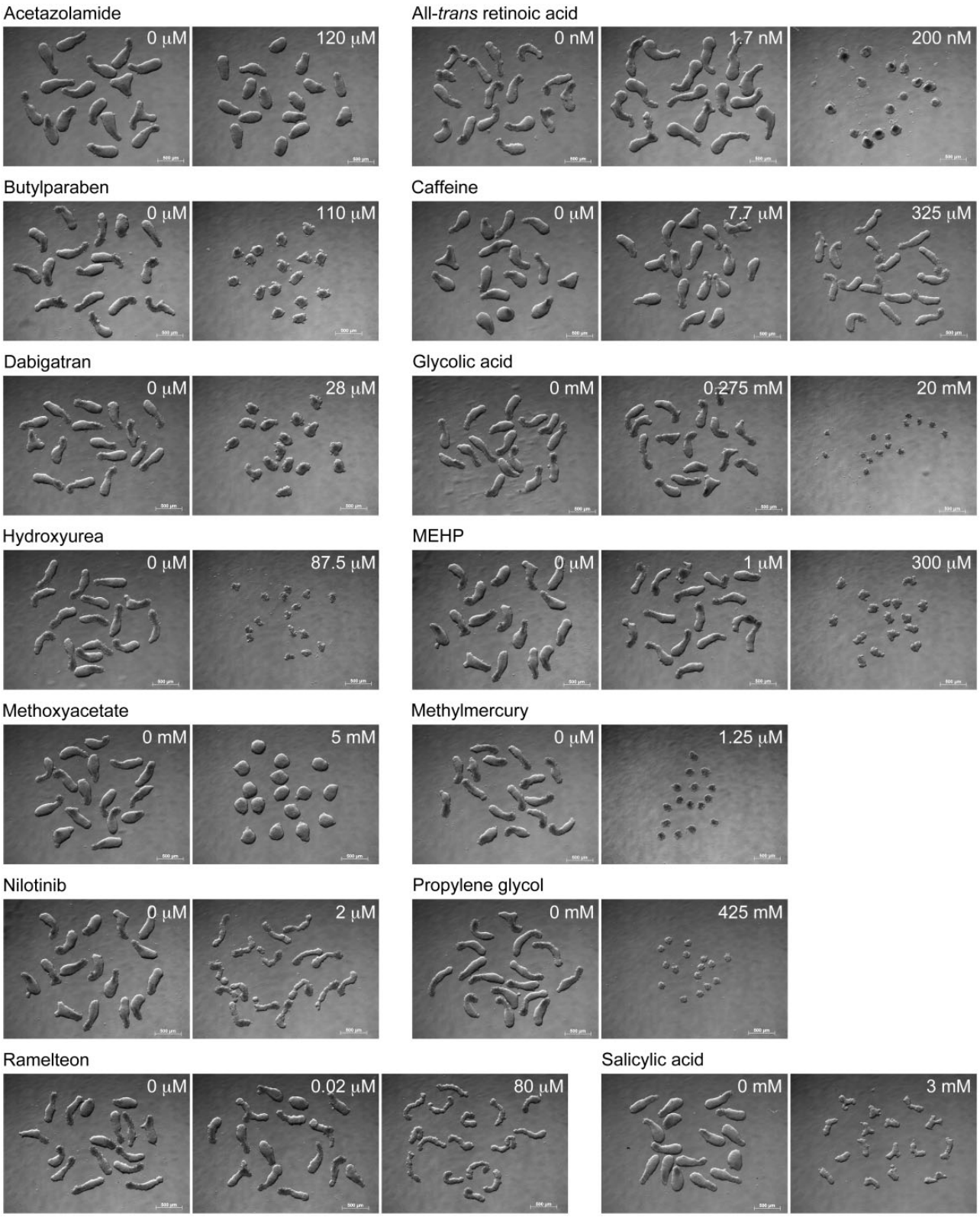

FIG. 3. Representative images of P19C5 EBs. Not all compound exposures are shown (see text). Each set of images shows a control group of EBs (no test compound) and one or two compound-treated groups of EBs, made from the same cell suspension as the control. Scale bars = 500 μm.

morphogenetic analyses of those exposures are shown in Supplementary Figure. LOAEL and NOAEL of all 24 compounds are summarized in Table 4.

Cytotoxicity on Somatic Cell Lines

Three of the negative exposures adversely impacted the P19C5 gastrulation model. Namely, butylparaben and nilotinib reduced EB size, whereas propylene glycol caused death of EBs. Thus, these negative exposures appeared to be detrimental to growth or survival of P19C5 EBs, raising the possibility that the gastrulation model may be overly sensitive to these compound exposures, leading to misclassifications. To test whether harmful effects of these exposures are unique to the P19C5 EBs, we performed cell viability assay using monolayer cultures of P19C5 cells as well as three types of somatic cell lines, NIH/3T3 (mouse embryonic fibroblast-derived), HEK293 (human embryonic kidney-derived), and JEG3 (human trophoblast-derived). These cell lines were cultured for 4 days in the presence of butylparaben, nilotinib, or propylene glycol, and impact on cell proliferation or survival was scored. All cell lines were completely killed by propylene glycol at the negative exposure concentration (850 mM) (Figure 4). Butylparaben at the negative exposure concentration (110 μM) also consistently reduced relative cell numbers of all four cell lines (Figure 4). Thus, on one hand, these two negative exposures appeared to exhibit cytotoxic effects on not only P19C5 cells but also on other cell lines. On the other hand, relative cell numbers were not consistently reduced by nilotinib in any of the cell lines at the negative exposure concentration (2 μM) (Figure 4).

TABLE 4. NOAEL and LOAEL of the Compounds Based on Morphogenetic Impact *In Vitro*

Compound	NOAEL	LOAEL	Morphogenetic Impact[a]
Abacavir	160 μM	320 μM	Area reduction
Acetazolamide	4 μM	20–30 μM	Area or EDI reduction
All-trans retinoic acid	3.4 nM	100 nM	Area and EDI reduction
Artesunate	2 μM	10 μM	Area and EDI reduction
Butylparaben	10 μM	20 μM	EDI reduction
Caffeine	32.5 μM	325 μM	EDI increase
Dabigatran	7 μM	14 μM	Area reduction
Desloratadine	3 μM	6 μM	Area reduction
Ethylene glycol	114 μM	228 μM	Area reduction
Fingolimod	1 μM	2 μM	Area and EDI reduction
Glycolic acid	10 mM	20 mM	Area and EDI reduction
Hydroxyurea	10 μM	20 μM	Area reduction
MEHP	75 μM	150 μM	Area reduction
Methanol	1.5 M	ND	
Methoxyacetate	1.25 mM	2.5 mM	Area and EDI reduction
Methylmercury	0.5 μM	1 μM	Area and EDI reduction
Nilotinib	0.4 μM	0.5–1 μM	EDI increase
Oseltamivir	24 μM	48 μM	Area reduction
Propylene glycol	100 mM	200 mM	Area reduction
Ramelteon	40 μM	80 μM	EDI increase
Saccharin	400 μM	ND	
Salicylic acid	0.25 mM	0.5 mM	EDI increase
Zaleplon	65.6 μM	131 μM	Area reduction
Zidovudine	227 μM	454 μM	Area reduction, EDI increase

NOAEL, no-observed-adverse-effect level; LOAEL, lowest-observed-adverse-effect level; MEHP, mono(2-ethylhexyl) phthalate; ND, not determined; EDI, Elongation Distortion Index.

[a]Morphogenetic effects observed at LOAEL.

DISCUSSION

The present study evaluated our *in vitro* gastrulation model of P19C5 stem cells in reference to the Daston list of compound exposures for developmental toxicity assays. Based on the morphogenetic effects on EBs, 10 out of 17 positive exposures were classified as adverse, whereas 14 out of 17 negative exposures were non-adverse. As shown in the previous study, VPA at 0.8 mM, another positive exposure in the Daston list, also adversely affects P19C5 EB morphogenesis (Li and Marikawa, 2016). Thus, altogether, 25 out of 35 exposures in the Daston list (71.4%) were correctly classified by the morphology-based assay using the P19C5 gastrulation model. When up to four times higher concentrations are included in assessment of the positive exposures, four additional exposures were also classified correctly, bringing the total to 29 out of 35 (82.9%), although the validity of such an *ad hoc* criterion will require additional biological and toxicological justification.

To date, various types of non-animal alternatives have been explored as developmental toxicity screening assays, including those utilizing differentiation, migration, or metabolomics of mouse or human ES cells (Kuske *et al.*, 2012; Palmer *et al.*, 2013; Seiler and Spielmann, 2011; Theunissen and Piersma, 2012; Xing *et al.*, 2015), rodent whole embryo culture (Piersma *et al.*, 2004), micromass culture of limb bud mesenchyme (Pratten *et al.*, 2012), and non-mammalian model systems, namely zebrafish embryos (Sipes *et al.*, 2011). At present, it may not be fruitful to discuss effectiveness of the P19C5 gastrulation model in comparison with these assays, because no validation studies have been reported in reference to the Daston list, to the best of our knowledge. Evaluations of other non-animal alternatives using the Daston list should provide solid framework for direct comparisons between different assays to reveal their strengths and weaknesses, and help to assemble a proper battery of tests to screen a broad range of developmentally toxic chemical exposures.

In the present study, the effects of compound exposures on the P19C5 gastrulation model were evaluated solely based on the morphological features, namely the size and shape of EBs. However, the effects of compounds may also be evaluated through gene expression analyses, because P19C5 EBs exhibit distinct temporal and spatial gene expression patterns that are characteristic of early embryogenesis during axial elongation and patterning (Lau and Marikawa, 2014; Li and Marikawa, 2015; Marikawa *et al.*, 2009). Effects on gene expression are generally considered to be more sensitive endpoints when assessing developmental toxicity, and various types of gene expression analyses have been incorporated into stem cell-based *in vitro* assays, particularly ESTs, to augment their detection of developmental toxicity (Buesen *et al.*, 2009; Gao *et al.*, 2014; Panzica-Kelly *et al.*, 2013; Pennings *et al.*, 2011; Suzuki *et al.*, 2011; Yu *et al.*, 2015; zur Nieden *et al.*, 2001). This notion applies to the response of P19C5 EBs to DAPT, a pharmacological inhibitor of Notch signaling, which markedly down-regulates expression of somitogenesis regulator genes, such as *Hes7*, *Lfng*, and *Nrarp*, without significantly altering EB size or shape (Li and Marikawa, 2015). However, the use of gene expression analyses to detect developmental toxicity requires cautious interpretation, as the heightened sensitivity may increase the rate of false-positive results. Mouse genetic studies, using the targeted gene knockout technology, have demonstrated in various cases that the loss of important developmental regulator genes fails to cause

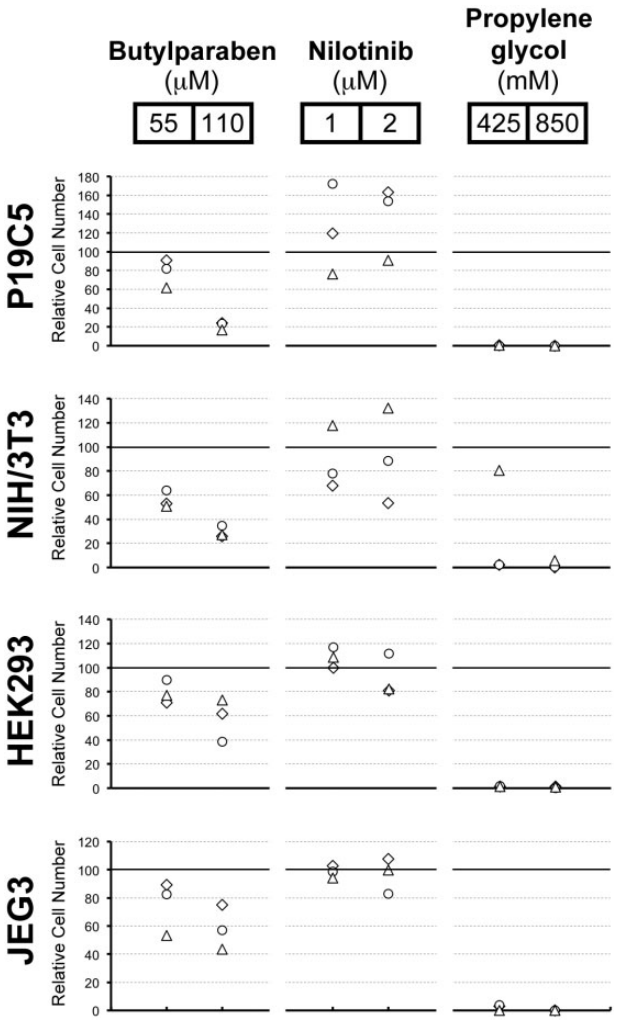

FIG. 4. Relative cell numbers of four cell lines, P19C5, NIH/3T3, HEK293 and JEG3, that are cultured as monolayer in the presence of butylparaben, nilotinib, or propylene glycol, at the negative exposure concentrations indicated in the Daston list. Data points for three biological replicates are shown using different marker styles.

overt phenotypic abnormalities, due to existence of compensatory or redundant mechanisms (Barbaric *et al.*, 2007; Nowak *et al.*, 1997). Thus, alterations in gene expression may not correlate with the dramatic phenotypic effects (e.g. embryonic death or major fetal malformations) referenced in the Daston list.

Among the 34 exposures from the Daston list examined here, the negative exposure of ATRA (1.7 nM) was particularly unique in that it markedly increased relative area of EBs. Because relative EDI was concomitantly increased, ATRA at the low concentration appeared to promote growth and axial elongation of EBs. Although ATRA is teratogenic at excessively high concentrations, it also acts at low physiological concentrations as an endogenous regulator of cell proliferation and embryo body patterning (Clagett-Dame and Knutson, 2011). Inactivation of *Aldh1a2* gene, which encodes retinaldehyde dehydrogenase to synthesize ATRA, impairs growth and axial elongation in mouse embryos (Niederreither *et al.*, 1999). Likewise, treatment of P19C5 EBs with BMS493, a pharmacological antagonist of retinoic acid receptors, reduces relative area and EDI, suggesting that active retinoic acid signaling is essential for growth and axial elongation in the *in vitro* gastrulation model (Li and Marikawa, 2015). The source of retinoic acid or its precursors in the culture medium used in the present study was possibly bovine serum (Materials and Methods section), and it may not have contained sufficient amount to support maximal growth and axial elongation of EBs. Thus, supplementation of a small amount of ATRA (as in the case for 1.7 nM exposure) would be beneficial for more robust EB development *in vitro*.

Artesunate, fingolimod, and methanol caused no adverse effects on the P19C5 gastrulation model even at four times higher concentrations than those indicated as positive exposures in the Daston list. These false negatives may exemplify some of the limitations and weaknesses of the model. Artesunate, an anti-malaria medication, is rapidly converted to its active metabolite dihydroartemisinin, which exerts developmental toxicity by generating free radicals that damage embryonic erythroblasts (Clark *et al.*, 2008; Li *et al.*, 2009). Depletion of primitive embryonic erythroblasts likely accounts for the observed cardiovascular and skeletal abnormalities caused by artesunate treatment (Clark, 2009; Clark *et al.*, 2008; White and Clark, 2008; White *et al.*, 2006). The developmental toxicity of fingolimod, an immunomodulator, also manifests as abnormal vascular maturation and cardiac malformation, mediated by inhibition of the sphingosine-1 phosphate receptor (FDA, 2010; Schmid *et al.*, 2007). The P19C5 gastrulation model may not accurately identify such developmental toxicity that selectively disrupts hematologic processes, because *in vitro* development of EBs does not require a vascular system. On the other hand, the developmental toxicity of methanol may be mediated by its metabolic byproduct, formate (Andrews *et al.*, 1995). Compounds that become teratogenic only after chemical modifications by maternal metabolism are termed, "proteratogens" (Wells and Winn, 1996). Detection of proteratogens is often considered a weakness of *in vitro* screens for chemicals that inflict developmental toxicity, because these assays usually lack the maternal metabolic system. Future studies will address whether the P19C5 model is susceptible to formate, and explore the incorporation of an exogenous metabolic activation system, such as co-culture with hepatocytes (Hettwer *et al.*, 2010; Oglesby *et al.*, 1986) or liver microsome fractions (Luijten *et al.*, 2008; Zhao *et al.*, 1993).

Practically speaking, misclassification of negative exposures (i.e., false positives) may be more detrimental for effectiveness of developmental toxicity assays, as compared with misclassification of positive exposures (i.e., false negatives). Unnecessary dismissal of safe compound exposures based on incorrect assay outcomes is disadvantageous for pharmaceutical development and burdensome for regulatory agencies (Waring *et al.*, 2015). In the case of the P19C5 gastrulation model, three negative exposures gave false-positive outcomes: butylparaben at 110 μM (reduction in area by 55%), nilotinib at 2 μM (reduction in area by 25% and increase in EDI by 60%), and propylene glycol at 850 mM (death). The butylparaben and propylene glycol exposures diminished proliferation or survival of three somatic cell lines. Such general cytotoxicity may lead to misclassification of these exposures by other *in vitro* assays as well. It is unclear how such cytotoxic exposures are apparently harmless to developing embryos *in vivo* (Daston *et al.*, 2014). It is important to note that the compound concentrations indicated in the Daston list are C_{max}, and, therefore, developing embryos may not be exposed to such concentrations in a continuous manner, especially when compound clearance by the mother is rapid. Indeed, both butylparaben and propylene glycol have elimination half-lives of approximately 4 h and undergo biotransformations

in vivo (Aubert *et al.*, 2012; Morshed *et al.*, 1988). Additional pharmacokinetics information, namely area under the curve (AUC), may be more applicable to *in vitro* experimental conditions, and may help resolve this potential discrepancy. In contrast, the nilotinib negative exposure did not consistently diminish proliferation of the somatic cell lines or P19C5 cells in monolayer culture, suggesting that the adverse effect was specific to the P19C5 gastrulation model. Interestingly, nilotinib at the high concentration (28 μM; positive exposure in the Daston list), exhibited an unusual impact, i.e., P19C5 cells in hanging drops failed to aggregate while they were still alive and proliferating. Although spontaneous cell aggregation is an essential step in generating EBs, this process does not correspond to normal embryonic development. Thus, nilotinib may be altering cellular properties that are not embryologically relevant, and such non-physiological effects of nilotinib, even at lower concentrations, may have contributed to the false-positive misclassification.

In the future, the exposure-based validation list (Daston et al., 2014) will likely be expanded and refined as additional *in vivo* data on developmental toxicity and pharmacokinetics become available for various compounds. As previously suggested (Wise, 2016), it may be important to categorize positive exposures into distinct groups according to severity and nature of developmental toxicity. For example, the mechanism of an exposure that results in embryonic death is fundamentally different from an exposure that causes spina bifida or microcephaly. Should such differences in *in vivo* impact be reflected in the outcomes of *in vitro* assays? In the present study, the P19C5 gastrulation model exhibited various responses to the positive exposures, ranging from total degeneration to a reduction in axial elongation. The range of phenotypic responses warrants further investigations to examine whether these variations in EB integrity and morphology correlate with the nature and severity of *in vivo* effects observed after exposure to the corresponding developmentally toxic compounds.

SUPPLEMENTARY DATA

Supplementary data are available at *Toxicological Sciences* online.

ACKNOWLEDGMENTS

The authors are grateful to Dr Vernadeth B. Alarcon for reading the article and providing valuable comments.

FUNDING

Johns Hopkins Center for Alternatives to Animal Testing (CAAT) and the Alternatives Research & Development Foundation (ARDF).

REFERENCES

Andrews, J. E., Ebron-McCoy, M., Kavlock, R. J., and Rogers, J. M. (1995). Developmental toxicity of formate and formic acid in whole embryo culture: A comparative study with mouse and rat embryos. *Teratology* **51**, 243–251.

Aubert, N., Ameller, T., and Legrand, J. J. (2012). Systemic exposure to parabens: Pharmacokinetics, tissue distribution, excretion balance and plasma metabolites of [14C]-methyl-, propyl- and butylparaben in rats after oral, topical or subcutaneous administration. *Food Chem. Toxicol.* **50**, 445–454.

Barbaric, I., Miller, G., and Dear, T. N. (2007). Appearances can be deceiving: Phenotypes of knockout mice. *Brief. Funct. Genomic. Proteomic.* **6**, 91–103.

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

Clagett-Dame, M., and Knutson, D. (2011). Vitamin A in reproduction and development. *Nutrients* **3**, 385–428.

Clark, R. L. (2009). Embryotoxicity of the artemisinin antimalarials and potential consequences for use in women in the first trimester. *Reprod. Toxicol.* **28**, 285–296.

Clark, R. L., Lerman, S. A., Cox, E. M., Gristwood, W. E., and White, T. E. K. (2008). Developmental toxicity of artesunate in the rat: Comparison to other artemisinins, comparison of embryotoxicity and kinetics by oral and intravenous routes, and relationship to maternal reticulocyte count. *Birth Defects Res. B Dev. Reprod. Toxicol.* **83**, 397–406.

Daston, G. P., Beyer, B. K., Carney, E. W., Chapin, R. E., Friedman, J. M., Piersma, A. H., Rogers, J. M., and Scialli, A. R. (2014). Exposure-based validation list for developmental toxicity screening assays. *Birth Defects Res. B: Dev. Reprod. Toxicol.* **101**, 423–428.

Daston, G. P., Chapin, R. E., Scialli, A. R., Piersma, A. H., Carney, E. W., Rogers, J. M., and Friedman, J. M. (2010). A different approach to validating screening assays for developmental toxicity. *Birth Defects Res. B: Dev. Reprod. Toxicol.* **89**, 526–530.

FDA. (2001). *Clarinex pharmacology review*. Retrieved from http://www.accessdata.fda.gov/drugsatfda_docs/nda/2001/21-165_Clarinex.cfm. Accessed February 16, 2017.

FDA. (2010). *Gilenya pharmacology review* Retrieved from http://www.accessdata.fda.gov/drugsatfda_docs/nda/2010/022527Orig1s000TOC.cfm. Accessed February 16, 2017.

Friedman, J. M. (2010). The principles of teratology: Are they still true? *Birth Defects Res. A Clin. Mol. Teratol.* **88**, 766–768.

Gao, X., Yourick, J. J., and Sprando, R. L. (2014). Transcriptomic characterization of C57BL/6 mouse embryonic stem cell differentiation and its modulation by developmental toxicants. *PLoS One* **9**, e108510.

Hettwer, M., Reis-Fernandes, M. A., Iken, M., Ott, M., Steinberg, P., and Nau, H. (2010). Metabolic activation capacity by primary hepatocytes expands the applicability of the embryonic stem cell test as alternative to experimental animal testing. *Reprod. Toxicol.* **30**, 113–120.

Jelínek, R. (2005). The contribution of new findings and ideas to the old principles of teratology. *Reprod. Toxicol.* **20**, 295–300.

Kuske, B., Pulyanina, P. Y., and zur Nieden, N. I. (2012). Embryonic stem cell test: Stem cell use in predicting developmental cardiotoxicity and osteotoxicity. *Methods Mol. Biol.* **889**, 147–179.

Lau, C. G. Y., and Marikawa, Y. (2014). Morphology-based mammalian stem cell tests reveal potential developmental toxicity of donepezil. *Mol. Reprod. Dev.* **81**, 994–1008.

Li, A. S. W., and Marikawa, Y. (2015). An in vitro gastrulation model recapitulates the morphogenetic impact of pharmacological inhibitors of developmental signaling pathways. *Mol. Reprod. Dev.* **82**, 1015–1036.

Li, A. S. W., and Marikawa, Y. (2016). Adverse effect of valproic acid on an in vitro gastrulation model entails activation of retinoic acid signaling. *Reprod. Toxicol.* **66**, 68–83.

Li, Q., Si, Y., Xie, L., Zhang, J., and Weina, P. (2009). Severe embryolethality of artesunate related to pharmacokinetics following intravenous and intramuscular doses in pregnant rats. *Birth Defects Res. B Dev. Reprod. Toxicol.* **86**, 385–393.

Luijten, M., Verhoef, A., Westerman, A., and Piersma, A. H. (2008). Application of a metabolizing system as an adjunct to the rat whole embryo culture. *Toxicol. In Vitro* **22**, 1332–1336.

Marikawa, Y., Tamashiro, D. A. A., Fujita, T. C., and Alarcón, V. B. (2009). Aggregated P19 mouse embryonal carcinoma cells as a simple in vitro model to study the molecular regulations of mesoderm formation and axial elongation morphogenesis. *Genesis* **47**, 93–106.

Morshed, K. M., Nagpaul, J. P., Majumdar, S., and Amma, M. K. P. (1988). Kinetics of propylene glycol elimination and metabolism in rat. *Biochem. Med. Metab. Biol.* **39**, 90–97.

Niederreither, K., Subbarayan, V., Dolle, P., and Chambon, P. (1999). Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat. Genet.* **21**, 444–448.

Nowak, M. A., Boerlijst, M. C., Cooke, J., and Smith, J. M. (1997). Evolution of genetic redundancy. *Nature* **388**, 167–171.

Oglesby, L. A., Ebron, M. T., Beyer, P. E., Carver, B. D., and Kavlock, R. J. (1986). Co-culture of rat embryos and hepatocytes: In vitro detection of a proteratogen. *Teratog. Carcinog. Mutagen.* **6**, 129–138.

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

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

Pennings, J. L. A., van Dartel, D. A. M., Robinson, J. F., Pronk, T. E., and Piersma, A. H. (2011). Gene set assembly for quantitative prediction of developmental toxicity in the embryonic stem cell test. *Toxicology* **284**, 63–71.

Piersma, A. H., Genschow, E., Verhoef, A., Spanjersberg, M. Q. I., Brown, N. A., Brady, M., Burns, A., Clemann, N., Seiler, A., and Spielmann, H. (2004). Validation of the postimplantation rat whole-embryo culture test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Altern. Lab. Anim.* **32**, 275–307.

Pratten, M., Ahir, B. K., Smith-Hurst, H., Memon, S., Mutch, P., and Cumberland, P. (2012). Primary cell and micromass culture in assessing developmental toxicity. *Methods Mol. Biol.* **889**, 115–146.

Riebeling, C., Hayess, K., Peters, A. K., Steemans, M., Spielmann, H., Luch, A., and Seiler, A. E. M. (2012). Assaying embryotoxicity in the test tube: Current limitations of the embryonic stem cell test (EST) challenging its applicability domain. *Crit. Rev. Toxicol.* **42**, 443–464.

Schmid, G., Guba, M., Ischenko, I., Papyan, A., Joka, M., Schrepfer, S., Bruns, C. J., Jauch, K. W., Heeschen, C., and Graeb, C. (2007). The immunosuppressant FTY720 inhibits tumor angiogenesis via the sphingosine 1-phosphate receptor 1. *J. Cell. Biochem.* **101**, 259–270.

Seiler, A. E. M., and Spielmann, H. (2011). The validated embryonic stem cell test to predict embryotoxicity in vitro. *Nat. Protoc.* **6**, 961–978.

Sipes, N. S., Padilla, S., and Knudsen, T. B. (2011). Zebrafish: As an integrative model for twenty-first century toxicity testing. *Birth Defects Res. C: Embryo Today* **93**, 256–267.

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

Theunissen, P. T., and Piersma, A. H. (2012). Innovative approaches in the embryonic stem cell test (EST). *Front. Biosci.* **17**, 1965–1975.

Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Pairaudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., *et al.* (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* **14**, 475–486.

Warkus, E. L. L., Yuen, A. A. Y. Q., Lau, C. G. Y., and Marikawa, Y. (2016). Use of in vitro morphogenesis of mouse embryoid bodies to assess developmental toxicity of therapeutic drugs contraindicated in pregnancy. *Toxicol. Sci.* **149**, 15–30.

Wells, P. G., and Winn, L. M. (1996). Biochemical toxicology of chemical teratogenesis. *Crit. Rev. Biochem. Mol. Biol.* **31**, 1–40.

White, T. E. K., Bushdid, P. B., Ritter, S., Laffan, S. B., and Clark, R. L. (2006). Artesunate-induced depletion of embryonic erythroblasts precedes embryolethality and teratogenicity in vivo. *Birth Defects Res. B: Dev. Reprod. Toxicol.* **77**, 413–429.

White, T. E. K., and Clark, R. L. (2008). Sensitive periods for developmental toxicity of orally administered artesunate in the rat. *Birth Defects Res. B: Dev. Reprod. Toxicol.* **83**, 407–417.

Wise, L. D. (2016). Numeric estimates of teratogenic severity from embryo-fetal developmental toxicity studies. *Birth Defects Res. B: Dev. Reprod. Toxicol.* **107**, 60–70.

Xing, J., Toh, Y. C., Xu, S., and Yu, H. (2015). A method for human teratogen detection by geometrically confined cell differentiation and migration. *Sci. Rep.* **5**, 10038.

Yu, R., Miyamura, N., Okamoto-Uchida, Y., Arima, N., Ishigami-Yuasa, M., Kagechika, H., and Nishina, H. (2015). A modified murine embryonic stem cell test for evaluating the teratogenic effects of drugs on early embryogenesis. *PLoS One* **10**, e0145286.

Zhao, J., Krafft, N., Terlouw, G. D., and Bechter, R. (1993). A model combining the whole embryo culture with human liver S-9 fraction for human teratogenic prediction. *Toxicol. In Vitro* **7**, 827–831.

zur Nieden, N. I., Ruf, L. J., Kempka, G., Hildebrand, H., and Ahr, H. J. (2001). Molecular markers in embryonic stem cells. *Toxicol. In Vitro* **15**, 455–461.