JRS (2013) Volume 1: Issue 1 | pages 32-49

TDL.ORG TEXAS DIGITAL LIBRARY

Journal of Regulatory Science

http://journalofregulatoryscience.org

PKP PUBLIC KNOWLEDGE PROJECT

Reevaluation of the embryonic stem cell test

Amy L. Inselman [a,*], Greg T. Nolen [a], Ching-Wei Chang [b], Wafa Harrouk [c], Edward Fisher [d], Melissa S. Tassinari [e], and Deborah K. Hansen [a]

[a] *U.S. Food and Drug Administration, National Center for Toxicological Research, Division of Systems Biology, Jefferson, AR 72079, USA*

[b] *U.S. Food and Drug Administration, National Center for Toxicological Research, Division of Bioinformatics and Biostatistics, Jefferson, AR 72079, USA*

c U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Division of Nonprescription Clinical Evaluation, Silver Spring, MD 20993, USA

d U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Division of Neurology Products, Silver Spring, MD 20993, USA

e U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Pediatric and Maternal Health Staff, Silver Spring, MD 20993 USA

ARTICLE INFORMATION

Article History:
Received Sept 6 2013
Revised received November 18 2013
Accepted November 19 2013

Keywords:
Embryonic stem cell test (EST)
Mouse embryonic stem cells (mESCs)
Developmental toxicity
Embryotoxicity
Alternative test method

ABSTRACT

In recent years there has been a heightened interest in developing alternative toxicity testing methods that enhance the current system that relies almost exclusively on whole-animal testing to include *in vitro* assays focused on defined pathways. The embryonic stem cell test (EST) is one alternative that has been suggested for use in developmental toxicity testing. The EST utilizes the D3 mouse embryonic stem cell line to assess sensitivity of chemicals on differentiating cardiomyocytes. Additionally, a BALB/3T3 line is used to monitor cytotoxicity and compare sensitivities between adult and embryonic cells. We have reevaluated the EST nearly 10 years after formal validation by the European Center for the Validation of Alternative Methods (ECVAM) to test the stability and reliability of the cell lines in predicting developmental toxicity. Eight compounds from the ECVAM validation including the positive control, 5-fluorouracil, and the negative control, penicillin G, were tested. All eight compounds matched the classification reported during validation, indicating comparable responses and transferability of the experimental protocol. However, an increased sensitivity of the cell lines, identified by lower ID_{50} and IC_{50} values, was observed for many of the chemicals when compared to the results from the ECVAM validation.

* Corresponding author. Division of Systems Biology, FDA/NCTR, 3900 NCTR Rd., Jefferson, AR 72079, USA. Tel: (870) 543-7623; fax: (870) 543-7682; E-mail: amy.inselman@fda.hhs.gov (A. Inselman).

Introduction

In the United States pharmaceuticals intended for human use are subjected to extensive premarket testing under the Federal Food, Drug, and Cosmetic (FD&C) Act to ensure safety and effectiveness prior to human use. Testing is typically conducted in at least two animal species and includes, but is not limited to, studies of carcinogenicity, mutagenicity, genotoxicity, toxicokinetics, pharmacokinetics, immunotoxicity, and reproduction as well as subacute, subchronic, and chronic exposures. These studies can be complex, time consuming, costly and require large numbers of animals. Regulatory studies for reproductive toxicity have been noted for their particularly high animal usage (van der Laan et al., 2012). Current International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Guidance S5(R2) (ICH, 2000) and M3(R2) (ICH, 2009) recommend that a compound be tested in two species (a rodent and a non-rodent) to assess potential effects on embryo-fetal development; rats and rabbits are the two species generally used for this testing.

One of the key challenges facing *in vitro* alternatives for predicting embryotoxicity is the complexity of the reproductive cycle, not all of which can be modeled *in vitro* by a single assay (van der Laan et al., 2012). Nevertheless, work continues on numerous assays with the goal of identifying *in vitro* alternative tests that accurately predict *in vivo* toxicity and may reduce animal usage. The embryonic stem cell test (EST) is an alternative test system that has been widely studied since the 1990s and is considered as one of the possible *in vitro* alternatives for predicting embryotoxicity (Spielmann et al., 1997). It is one of the few *in vitro* alternative tests that do not require live animals or embryos. Instead the EST relies on two commercially available cells lines, the D3 mouse embryonic stem cell (mESC) line and a BALB/3T3 fibroblast line, allowing sensitivities between embryonic and adult cells to be assessed. When pluripotent mESCs are allowed to form multicellular aggregates, known as embryoid bodies (EBs), the cells differentiate into derivatives of endodermal, mesodermal and ectodermal origin recapitulating aspects of early embryonic development (Martin et al., 1977).

The EST, as validated by the European Center for the Validation of Alternative Methods (ECVAM), classifies compounds based on three toxicological endpoints: (1) the morphological analysis of beating cardiomyocytes differentiated from mESCs, (2) the cytotoxicity of differentiated mESCs, and the (3) cytotoxicity of BALB/3T3 fibroblasts. These endpoints are measured after 10 days of chemical exposure to identify the ID_{50} (50% inhibition of mESC differentiation into cardiomyocytes) and IC_{50} (50% inhibition of cell viability) values. A biostatistically based prediction model that applies a linear discriminate analysis to the three endpoints is then used to classify a given compound as non-embryotoxic, weakly embryotoxic or strongly embryotoxic (Genschow et al., 2000, 2002, 2004).

Twenty chemicals were selected by ECVAM for testing during formal validation of the assay. The chemicals were selected from a database compiled from previous *in vivo* and *in vitro* developmental toxicity studies on 309 chemicals (Brown, 2002). The chemicals were then grouped into the three classification categories (non-embryotoxic, weakly embryotoxic or strongly embryotoxic) based on previous data. A non-embryotoxicant was defined by ECVAM as "not developmentally toxic at maternally toxic exposures, but which may show some minor embryo/fetal toxicity, which cannot be separated from maternal toxicity" (Brown, 2002). For classification as a strong embryotoxicant a chemical had to be "developmentally toxic in all species tested, inducing multiple developmental effects" (Brown, 2002). The classification of weak embryotoxicants under the ECVAM classification system, however, was the most complex. In general this classification category included "chemicals of intermediate activity" defined further as those "developmentally toxic in multiple (but not all) species" or "developmentally toxic in

multiple species, inducing multiple effects, with exposures that are clearly less than maternally toxic exposures" or those that are "developmentally toxic, inducing effects that are clearly unrelated to maternal toxicity, but with exposures that are close to maternally toxic exposures" (Brown, 2002).

During the validation phase, the EST correctly classified 78% of the twenty selected test chemicals; all six of the strong embryotoxicants were correctly classified (Genschow et al., 2002). The EST was accepted as a validated *in vitro* alternative assay for developmental toxicity by ECVAM in 2008 (Spielmann et al., 2008). Subsequent testing, however, identified limitations with the assay and the prediction model which affected the ability to predict the toxicity of certain classes of compounds (Chapin et al., 2007; Marx-Stoelting et al., 2009). When Pfizer tested 19 of their in-house receptor-mediated pharmaceutical compounds with known *in vivo* developmental toxicity, the overall accuracy was 53% (Paquette et al., 2008). Screening of an additional 29 commercially available pharmaceutical products, the majority of which were known non-developmental toxicants, increased the predictivity of the assay to 75% (Paquette et al., 2008). While questions remain regarding its applicability domain and the predictive value, assays similar to the validated EST are currently being employed in the pharmaceutical industry for prioritizing lead compounds for development (Marx-Stoelting et al., 2009).

We have reevaluated the EST within our laboratory using a select group of chemicals included in the ECVAM validation of the assay. This was done, in part, so U.S. Food and Drug Administration scientists could gain experience with the assay. More importantly we wanted to test the stability and reliability of the cell lines in predicting developmental toxicity. One of the tenets of using cell lines is that they are stable over time and in this case, would still be able to predict developmental toxicity even ten years after initial development of the assay. If *in vitro* alternative models are to be accepted in a regulatory setting, the performance and stability of the cells lines over an extended period is critical. A total of eight compounds from the validation phase were tested. This included the positive control, 5-fluorouracil, and the negative control, penicillin G. All eight compounds matched the classification reported in the validation assay indicating comparable responses as well as transferability of the experimental protocol. Although the compounds matched the classification, we observed an increased sensitivity of the cell lines to many of the treatment chemicals. This sensitivity generally was identified by lower ID_{50} and IC_{50} values.

2. Materials and methods

The embryonic stem cell test was conducted as described by Seiler & Spielmann (2011) and is summarized below.

2.1. Cell Culture

The D3 mESC line was obtained from ATCC (cat. no. CRL-1934), but its passage number is unknown. The mESCs were thawed and expanded on EmbryoMax Strain CF-1 PMEFs (Millipore, cat. no. PMEF-CFL) that were inactivated by irradiation. Following expansion, the mESCs were transitioned to a feeder independent culture platform following the protocol described by Tremml et al. (2008) and used in the EST. For maintenance, D3 mESCs were cultured in high glucose DMEM (Life Technologies, cat. no. 11965-118) containing 15% defined ES-qualified fetal bovine serum (FBS) (Fisher Scientific, cat. no. SH30070.03EH), 2 mM L-glutamine (Life Technologies, cat. no. 25030-081), 50 U/mL penicillin, 50 µg/mL streptomycin (Life Technologies, cat. no. 15140-122), 1% MEM non-essential amino acids (Life Technologies, cat. no. 11140-050), 0.1 mM β-mercaptoethanol (Life Technologies, cat. no. 21985-023) and 1000 U/mL murine leukemia inhibitory factor (mLIF) (Millipore, cat. no. ESG1107). The D3 mESCs

were routinely passaged every 2-3 days and were plated at 7.5 x 10^5 cells on 0.1% gelatin (Millipore, cat. no. ES-006-B) coated 100 mm tissue culture dishes (BD Biosciences, cat. no. 353003). Cells within passages 12-17 after thawing were used in all experiments.

BALB/3T3 cells (clone A31) were obtained from ATCC (cat. no. CCL-163) at passage number 64. 3T3 cells were grown in high glucose DMEM (Life Technologies), 10% defined FBS (Fisher Scientific, cat. no. SH30070.03), 4 mM L-glutamine (Life Technologies), 50 U/mL penicillin, and 50 µg/mL streptomycin (Life Technologies) for expansion and cytotoxicity assays. The 3T3 cells were passaged every 2-3 days and plated at 1 x 10^6 cells/100 mm tissue culture dish (BD Biosciences).

D3 mESCs and 3T3 cells were grown in a humidified atmosphere at 5% CO_2 and 37°C. Cell lines were passaged using 0.05% trypsin-EDTA (Invitrogen, cat. no. 25300-054) treatment; cell counts and percent viability were determined using the Cellometer Auto T4 cell counter (Nexcelom Bioscience) with 0.4% trypan blue (Life Technologies, cat. no. 15250-061).

2.2. Test Compounds

A subset of the test compounds used in the validation assay (Brown, 2002; Genschow et al., 2004) was selected for testing. This subset included the ECVAM-classified non-embryotoxic compounds acrylamide (Sigma-Aldrich, cat. no. A9099) and diphenhydramine hydrochloride (Sigma-Aldrich, cat. no. D3630); the weakly embryotoxic compounds lithium chloride (Sigma-Aldrich, cat. no. L4408) and valproic acid (Sigma-Aldrich, cat. no. P4543); and the strongly embryotoxic compounds hydroxyurea (Sigma-Aldrich, cat. no. H8627) and methotrexate (Sigma-Aldrich, cat. no. A6770). The negative control, penicillin G (Sigma-Aldrich, cat. no. P3032), and the positive control, 5-fluorouracil (Sigma-Aldrich, cat. no. F6627), were also included.

The chemicals were, in part, selected on their general availability and purity levels. Diphenhydramine hydrochloride, acrylamide and hydroxyurea were included because of their previous misclassification during validation of the EST. Diphenhydramine hydrochloride's previous *in vitro* EST classification did not match the *in vivo* classification; acrylamide and hydroxyurea were each misclassified in at least one independent experiment during validation of the assay. Methotrexate was chosen because of the low concentrations at which it induces embryotoxicity.

Test chemicals were dissolved in the appropriate solvent determined by the known physicochemical properties of each compound. Solvents included Dulbecco's phosphate buffered saline (Life Technologies, cat. no. 14190-144), sterile filtered water (Sigma-Aldrich, cat. no. W3500) and dimethyl sulfoxide (DMSO) (Sigma-Aldrich, cat. no. D2650). The recommended maximum final solvent concentrations as described in Seiler and Spielmann (2011) were used; at these levels the solvents have previously been shown to be non-cytotoxic and have no effects on cell differentiation. Final solvent concentrations were kept constant throughout each experiment.

2.3. Cardiomyocyte Differentiation

For differentiation of D3 mESCs into cardiomyocytes, mLIF was removed from the mESC culture media described above and the 15% defined ES-qualified FBS was replaced with 15% defined FBS (Fisher Scientific, cat. no. SH30070.03), previously determined to provide consistent cardiomyocyte differentiation. EB formation was performed as described in Seiler & Spielmann (2011). Six to seven concentrations of the compounds were applied to the mESCs from day zero through day 10. On day five of culture a single EB was transferred into individual wells of a 24-well tissue culture plate (BD Biosciences, cat. no. 353047). For every test concentration of chemical a single 24-well plate was set up; two plates were set up for vehicle controls. On day 10 the percentage of beating cardiomyocytes was determined by morphological assessment using an Evos XL (Advanced Microscopy Group) digital

inverted microscope. The percentage of wells containing contracting cardiomyocytes for each plate was calculated and compared to the percentage for vehicle controls. An assay was considered valid if at least 21 out of 24 wells on the vehicle control plates contained contracting cardiomyocytes.

2.4. Cell Viability

The cytotoxic effects of the test chemicals on D3 mESCs and BALB/3T3 cells were determined using the thiazolyl blue tetrazolium bromide (MTT) assay (Sigma-Aldrich, cat. no. M5655). Briefly, 500 cells were seeded per well of a 96-well tissue culture plate (BD Biosciences, cat. no. 353072) and incubated with vehicle or varying concentrations of the test chemical for 10 days. On day 10, the MTT assay (validated method) was performed as described in Seiler & Spielmann (2011). Absorbance (570/630 nm) was determined using the SpectraMax M2e Microplate Reader (Molecular Devices).

2.5. Prediction and Classification of Test Compounds

To evaluate and classify the test compounds according to their toxic potential three endpoints were determined: ID_{50} values (50% inhibition of mESC differentiation into cardiomyocytes) for D3 mESCs and IC_{50} values (50% inhibition of cell viability) for D3 mESCs and BALB/3T3 cells (Seiler & Spielmann, 2011). Briefly, dose response curves were generated for each test chemical and fit to a three-parameter logistic function using SigmaPlot 11.0 (Systat Software Inc.). ID_{50} and IC_{50} values were then used in the validated prediction models linear discriminant functions to classify the compounds as non-embryotoxic (class I), weakly embryotoxic (class II) or strongly embryotoxic (class III).

Confidence intervals for each chemical were determined from the standard error of the ID_{50} and IC_{50} values. Confidence interval values were then used to compare whether data generated in our laboratory covered the mean values or overlapped with the confidence intervals reported in Appendix 2 of the validation study (Genschow et al, 2004).

Pairwise scatter plots were also used to assess the interlaboratory deviation of ID_{50} and IC_{50} values as described in Genschow et al. (2004). Comparisons were made for the mean endpoint values determined in our laboratory and those involved in validation of the assay. Outliers reported in Genschow et al. (2004) were excluded from the analysis. When plotted, closely related data will form a straight line; interlaboratory deviation is then assessed by the determination of R^2 values. An R^2 value of zero represents no correlation while a value of one indicates identical endpoint determinations. Genschow et al. (2004) assumed an R^2 value of 0.6 to indicate moderate correlation, 0.8 to indicate good correlation while 0.9 and above were considered to be excellent correlations.

3. Results and discussion

3.1. Quality Control Measures

Prior to screening selected chemicals from the validation assay, batches of defined FBS were tested to ensure consistent cardiomyocyte differentiation - defined as a minimum of 21 out of 24 wells with beating cardiomyocytes on day 10 of differentiation (Seiler & Spielmann, 2011). However, in contrast to the validation protocol in which the same batch of FBS is used for growth and differentiation, we used a defined ES-qualified FBS for the growth and maintenance of the D3 cell line and a defined FBS for cardiomyocyte differentiation and cytotoxicity assays. Different sera were selected based on the rationale that serum that promoted the growth of mESCs in an undifferentiated state would not reliably promote differentiation to cardiomyocytes and serum that promotes differentiation may not support the cell line in an undifferentiated state over extended passaging. This rationale was supported by preliminary tests that used ES-qualified serum to induce differentiation.

Although cardiomyocyte differentiation was achieved, the threshold of 21 out of 24 wells containing contracting cardiomyocytes was not consistently attained (data not shown). ES-qualified serum, available through many manufacturers, is pre-screened to ensure consistent growth of the cell line while minimizing unwanted differentiation. Typically, screening includes plating efficiency, monitoring of colony morphology, cytotoxicity and may in some cases include monitoring of gene expression markers. The advantage of using a defined FBS, whether ES-qualified or not, is that it is biochemically characterized. Once a lot is shown to provide consistent differentiation it can more readily be matched leading to increased reliability and virtually eliminating timely screening of serum lots. Serum concentrations were held constant at 15% for all growth and differentiation experiments performed in our laboratory.

The positive control 5-fluorouracil was tested to ensure consistent results with the historical and validation data (Genschow et al., 2004). Two independent assays were performed, and the mean values are reported in Table 1. In each case the mean value of our data was lower than the reported historical and validation mean values. However, our ID_{50} and IC_{50} D3 values were within the expected ranges published by Seiler & Spielmann (2011), although at the lower end. The greatest amount of variability was observed with the 3T3 line; this was also observed by Genschow et al. (2004) during analysis of the validation data in which two homogenous groups of data were identified between the four laboratories. While the mean values of data from our laboratory were slightly lower than the reported range, the 95% confidence interval (CI) of the mean values overlapped with the lower boundary confidence intervals from the validation study. Overall, 5-fluorouracil was correctly classified as a strong embryotoxicant using the linear discriminate prediction model developed by ECVAM.

3.2. Negative Control – Penicillin G

Penicillin G was included as one of the test compounds as it is commonly used as a negative control in the EST. Two independent assays were performed (Table 2) where the maximum concentration of test chemical, 1000 µg/ml, was used. Penicillin G did not inhibit cardiomyocyte differentiation nor induce cytotoxicity in the mESCs. In one assay, penicillin G did induce some cytotoxicity in the 3T3 cell line. However, this finding was also observed in three of the four laboratories during the original validation of the assay (Genschow et al., 2004; Table 2). Penicillin G was classified as a non-embryotoxicant, matching the *in vivo* prediction as well as the previous EST prediction.

3.3. Classification of Compounds

Six compounds were selected from the twenty originally tested during the validation phase of the EST. Two compounds from each category (non-embryotoxic, weakly embryotoxic and strongly embryotoxic) were selected; selections were based on the ECVAM *in vivo* classification and not on the previous *in vitro* EST predictions. ID_{50} and IC_{50} values were determined for each compound in two independent assays as described in Seiler & Spielmann (2011).

Acrylamide and diphenhydramine hydrochloride were selected as the non-embryotoxicant compounds. During validation of the EST, acrylamide was classified as a non-embryotoxicant in six of eight independent assays performed during the validation phase, matching the *in vivo* classification in a majority of the assays. When acrylamide was tested in our laboratory, it was correctly classified as a non-embryotoxicant, matching both the *in vivo* classification as well as the EST prediction for the compound (Table 3). Increased sensitivity of the 3T3 line was observed with no overlap between the confidence intervals of our data with the validation data. In the cytotoxicity experiments, the D3 mESCs also had lower IC_{50} values than those reported during validation; one value did, however, overlap with the confidence interval from one of our assays.

The EST, however, misclassified the non-embryotoxicant diphenhydramine hydrochloride as a weak embryotoxicant. It was incorrectly classified in our laboratory as well as in the four

Table 1
Comparison of IC_{50} and ID_{50} Values for 5-Fluorouracil in the EST[a]

Endpoint	Laboratory[b]	n	Mean Value (µg/ml)	Lower Boundary (µg/ml)	Upper Boundary (µg/ml)
IC_{50} D3	Historical	22	0.072	0.063	0.081
	Validation	16	0.054	0.042*	0.071
	NCTR	2	0.041	0.036	0.045
IC_{50} 3T3	Historical	19	0.194	0.152	0.249
	Validation	16	0.159	0.114*	0.222
	NCTR	2	0.105	0.085	0.126
ID_{50}	Historical	12	0.052	0.047	0.057
	Validation	16	0.047	0.042*	0.053
	NCTR	2	0.039	0.031	0.046

[a]*Historical and validation values from Genschow et al. (2004).*

[b]*Historical data collected in the lead laboratory over a two year period. Validation values obtained from the compilation of data from the four laboratories involved in the validation study.*

**Indicates the confidence interval of NCTR data overlapped with the confidence intervals from the validation laboratories.*

Table 2
Comparison of IC_{50} and ID_{50} Values for Penicillin G in the EST[a]

Chemical	Laboratory	Run	IC_{50} D3 (µg/ml)	IC_{50} 3T3 (µg/ml)	ID_{50} (µg/ml)	EST Prediction[b]	*In vivo* Prediction[b]
Penicillin G	I	1	>1000	850	>1000	1	1
		2	>1000	1000	>1000	1	1
	J	1	>1000	980	>1000	1	1
		2	>1000	950	>1000	1	1
	K	1	>1000	>1000	1000	1	1
		2	>1000	>1000	1000	1	1
	L	1	>1000	>1000	>1000	1	1
		2	>1000	690	>1000	1	1
	NCTR	1	>1000	490	>1000	1	1
		2	>1000	>1000	>1000	1	1

[a]*Validation data from Genschow et al. (2004).*

[b] *1 = non-embryotoxic; 2 = weakly embryotoxic; 3 = strongly embryotoxic.*

Table 3
Comparison of IC_{50} and ID_{50} Values and Predictions for Select Chemicals in the EST[a]

Chemical	**Laboratory**	**Run**	**IC_{50} D3 (µg/ml)**	**IC_{50} 3T3 (µg/ml)**	**ID_{50} (µg/ml)**	**EST Prediction[b]**	***In vivo* Prediction[b]**
Acrylamide	I	1	40#	38	80*	1	1
		2	59	34	42	**2**	1
	J	1	51	31	32	**2**	1
		2	48	14	80*	1	1
	K	1	97	48	95	1	1
		2	80	35	70	1	1
	L	1	92	51	120	1	1
		2	90	62	140	1	1
	NCTR	1	31	7	82	1	1
		2	36	8	79	1	1
Diphenhydramine hydrochloride	I	1	38	70	18	**2**	1
		2	35	63	9.5	**2**	1
	J	1	45	65	8.5#	**2**	1
		2	48	70	8.5#	**2**	1
	K	1	35	78	10	**2**	1
		2	19	56	16	**2**	1
	L	1	21	21	17	**2**	1
		2	23	42*	20	**2**	1
	NCTR	1	15	39	8.3	**2**	1
		2	9.5	42	4.5	**2**	1
Lithium chloride	I	1	600#	780#	140	2	2
		2	400#	700#	140	2	2
	J	1	>1000	750#	115	2	2
		2	>1000	800#	200#	2	2
	K	1	510#	780#	85	2	2
		2	450#	710#	80	2	2
	L	1	630#	700#	180#	2	2
		2	390#	800#	280	2	2
	NCTR	1	379	760	204	2	2
		2	730	855	170	2	2

Table 3, continued

Chemical	Laboratory	Run	IC_{50} D3 (µg/ml)	IC_{50} 3T3 (µg/ml)	ID_{50} (µg/ml)	EST Prediction[b]	*In vivo* Prediction[b]
Valproic acid	I	1	269*	345	45#	2	2
		2	262*	311	54#	2	2
	J	1	248*	235#	38#	2	2
		2	242*	255	48#	2	2
	K	1	228*	545	48#	2	2
		2	214*	449	52	2	2
	L	1	276*	276	66#	2	2
		2	69#	269	56#	2	2
	NCTR	1	355	223	61	2	2
		2	252	134	42	2	2
Methotrexate hydrate	I	1	0.024	0.022#	0.020	3	3
		2	0.037	0.025#	0.019	3	3
	J	1	0.10	0.013*	0.0006	3	3
		2	0.11	0.010#	0.04	3	3
	K	1	0.045	0.100	0.022	3	3
		2	0.018	0.075	0.020	3	3
	L	1	0.022	0.012#	0.023	3	3
		2	0.080	0.027#	0.047	3	3
	NCTR	1	0.007	0.015	0.009	3	3
		2	0.009	0.018	0.011	3	3
Hydroxyurea	I	1	3.0	4.8#	1.8	3	3
		2	5.1	4.8#	2.1	3	3
	J	1	4.5	2.5	1.0	3	3
		2	5.5	5.0#	1.2	3	3
	K	1	2.4	5.2#	3.1#	3	3
		2	7.0	9.5	12.0	**2**	3
	L	1	7.8	3.8#	0.8	3	3
		2	8.2	3.6#	1.4	3	3
	NCTR	1	3.6	4.0	2.3	3	3
		2	3.6	4.9	2.9	3	3

[a] *Validation data from Genschow et al. (2004).*

[b] *1 = non-embryotoxic; 2 = weakly embryotoxic; 3 = strongly embryotoxic.*

Indicates the confidence interval of one data point covered the reported value from the validation laboratories.

* *Indicates the confidence interval of both data points covered the reported value from the validation laboratories.*

Numbers in bold indicate a different prediction from in vivo.

validation laboratories (Table 3). In comparison to the values previously determined, values from our laboratory were again lower, with the greatest differences seen in the reported IC_{50} D3 and 3T3 values. Interestingly, the IC_{50} 3T3 values reported by laboratory L during the validation were considered interlaboratory outliers (Genschow et al., 2004); however, these values are identical in one instance and nearly identical in another to what was determined in our laboratory. The ID_{50} values showed the least variability.

Lithium chloride and valproic acid were selected as the weak embryotoxicants for testing. In general, a good correlation was seen between the data generated in our laboratory and the validation laboratories with these two chemicals (Table 3). However, the ID_{50} and IC_{50} values had higher standard errors associated with them than many of the other chemicals tested (data not shown) making it appear that lithium chloride and valproic acid had greater overlap with the validation data. Lithium chloride and valproic acid both matched the ECVAM *in vivo* prediction and the EST prediction.

Methotrexate and hydroxyurea were the strong embryotoxicants selected for testing. It is important to note that the hydrate form of methotrexate was used in our laboratory; it also appears that it was used during the validation phase of the assay. However, the CAS number listed in Genschow et al. (2002) is not consistent with the catalog number provided for the hydrate form. The D3 mESCs appeared to be more sensitive to the hydrate form of methotrexate in our laboratory as ID_{50} and IC_{50} values were in general lower (Table 3). There was one reported ID_{50} value from laboratory J that was lower (0.0006 µg/ml), but this was considered an outlier upon statistical analysis (Genschow et al., 2004). The response of 3T3 cells to methotrexate hydrate was very similar to that reported in the validation phase as the confidence intervals of our data overlapped with many of the concentrations determined in the validation study. Methotrexate hydrate matched its *in vivo* classification as a strong embryotoxicant.

The EST also correctly classified hydroxyurea as a strong embryotoxicant in each of two independent runs performed in our laboratory (Table 3). Hydroxyurea was misclassified by one of the laboratories in a single assay during validation (Genschow et al., 2004). Again, our ID_{50} and IC_{50} values matched those reported in the validation assay. However, we did not see as much overlap with the confidence intervals of our data with the reported endpoints in the validation assay. This was because the standard error of the mean for each of these values was small (data not shown).

Although the chemicals identified above matched the previous EST classification, not all the results from the validation study were transferable. Originally, all-*trans*-retinoic acid was selected for testing rather than methotrexate. However, we were unable to grow the D3 mESC line in the presence of all-*trans*-retinoic acid during the initial range finding studies (data not shown). Even at concentrations as low as $5x10^{-6}$ µg/ml we saw an inhibition of growth. The majority of cells died within 24 hours; those that survived treatment appeared to develop a neuronal phenotype as judged by cellular morphology. The reported IC_{50} D3 values for all-*trans*-retinoic acid also varied widely during the validation study, ranging from 0.0008 to 1.9 µg/ml (Genschow et al., 2004). We did not see any inhibition of 3T3 cell growth in the presence of all-*trans*-retinoic acid as the range finding study values were consistent with the final validation values (data not shown).

3.4. Cytotoxicity Assay – Quality Control Measures

On day 10 after the MTT assay had been performed, the absolute optical density ($OD_{550-570}$) of the solvent control wells was checked to ensure normal growth of the cell lines. These values were compared to the 95% confidence intervals previously determined and reported in Seiler & Spielmann (2011). In all assays performed in our

laboratory (n=16) with the 3T3 line the mean 95% confidence interval was met (mean 0.39; validation range for ATCC 3T3 line 0.15-0.6). However, the 95% confidence interval for the ATCC D3 line was not met. As reported in the validation protocol (Seiler & Spielmann, 2011), the $OD_{550-570}$ range for the D3 line should range from 0.50-1.6. The mean $OD_{550-570}$ value from our laboratory was 0.37 (n=16), indicating a change in the growth rate of the cell line. Differences in the growth rate of the cells were also reported in the validation protocol and were dependent upon the source of the cells (Kemler versus ATCC) (Seiler & Spielmann, 2011). It is possible that the D3 mESC line available through ATCC has changed over time from that reported in the original validation assay. It is also possible that the observed change could be attributed to the switch from ES qualified serum to defined serum for the cytotoxicity assay. The switch may initially slow proliferation leading to decreased cell numbers on the final day of the assay.

The positive control 5-fluorouracil was also included at fixed concentrations in each cytotoxicity assay. The concentration of 5-fluorouracil included for the 3T3 cell line was 0.29 µg/ml and was 0.06 µg/ml for the D3 mESC line. Analysis of historical data from the validation study revealed a reduction in growth by approximately 50% when the above concentrations of 5-fluorouracil were used, although reported values have ranged from 20-80% (Seiler & Spielmann, 2011). At the indicated concentrations we continued to see an increased sensitivity of the cell lines to 5-fluorouracil with the greatest sensitivity again observed for the 3T3 line. After treatment for 10 days with 5-fluorouracil only 9.5% $\pm$ 2.4 (SEM) of 3T3 cells survived (n=16) while 35.3% $\pm$ 4.9 (SEM) of the D3 mESCs survived (n=16).

Although an increased sensitivity was observed with the D3 and BALB/3T3 cells in our laboratory, there was an overall strong correlation between our data and those published by ECVAM when analyzed using pairwise scatter plots (Figures 1-3). The R^2 value provide an indication of correlation between the values generated in each laboratory; a value of zero indicating no correlation while an R^2 value of one indicates identical values (Genschow et al., 2004). The lowest R^2 value observed was 0.91 (Figure 1); this was for the IC_{50} D3 endpoint between our laboratory and laboratory J. The correlation data for the IC_{50} 3T3 endpoints are presented in Figure 2 and those for the ID_{50} endpoints in Figure 3. The R^2 values were all consistently high (>0.90) demonstrating a strong relationship between our data and those published by ECVAM as well as evidence of transferability and reproducibility of the experimental protocol.

3.5. Discussion

Objectives of the current study were to reevaluate the validated EST to verify if the experimental protocol could easily be transferred to a different laboratory, but more importantly to determine whether the cell lines continued to respond consistently over time. We tested eight of the chemicals used by ECVAM in the original validation study and found that all eight were classified as previously reported (Genschow et al. 2004). However, the D3 mESCs appeared to be more sensitive (as determined by lower ID_{50} and IC_{50} concentrations) to the compounds in our hands than in the earlier validation study.

The D3 mESC line was derived nearly 30 years ago (Doetschman et al., 1985), and while still available through ATCC (cat. no. CRL-1934), the passage number is unknown. However, given the age of the line one can assume it is a relatively high passage number line. In addition, this particular D3 line is no longer germline competent (ATCC Product Sheet for ES-D3 CRL-1934; http://www.atcc.org/Products/All/CRL-1934.aspx #357C3571006A4259B64650D34DF19048) and thus no longer meets the definition of a pluripotent line. For this reason, scientists at Pfizer used an in-house DBA/1lacJ derived line in their assessment of the EST using receptor-mediated compounds (Paquette et al., 2008). The limited differentiation potential of the D3 line could influence the endpoint values of the assay as

well as the proliferation rate of the cell line, impacting the predictive value of the assay. We observed a decrease in the proliferation rate upon repeated passaging of the cells, suggesting

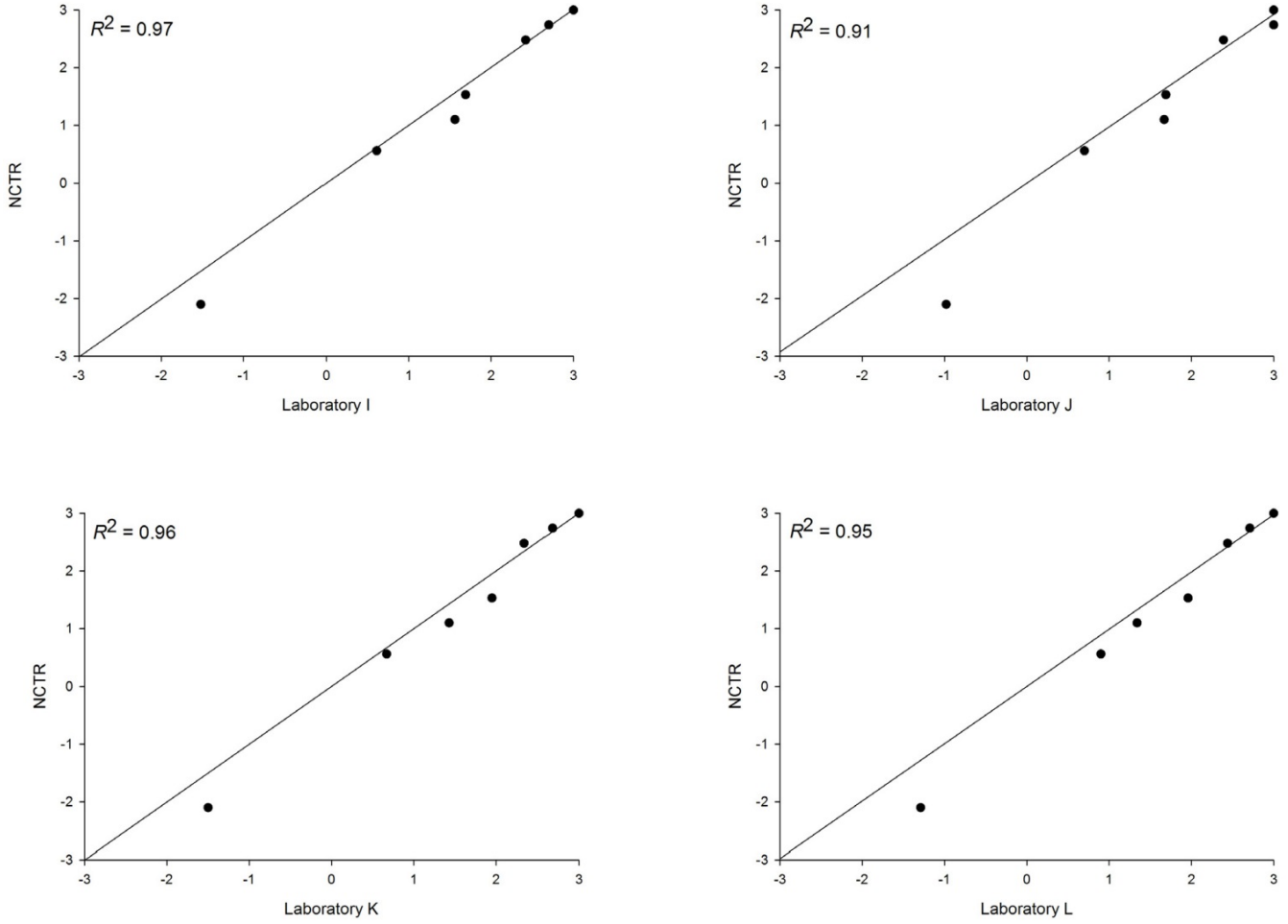

Figure 1. Interlaboratory comparisons of IC_{50} D3 values. Mean IC_{50} D3 endpoint values as reported in Appendix 2 of Genschow et al. (2004) were compared by pairwise scatter plots to the mean IC_{50} D3 endpoint values determined in our laboratory. For comparison, endpoint values were expressed as the log of the concentration (µg/ml) on both axes. Interlaboratory deviation was assessed by the determination of R^2 values. Comparisons were made for seven of the eight chemicals tested; 5-fluorouracil was excluded from analysis.

decreased performance/reliability of the line (data not shown). In the validation protocol, Seiler and Spielmann (2011) cautioned against extensive passaging, which they defined as greater than 25 passages, to prevent inconsistent differentiation. The change we observed in the proliferation rate, however, occurred much earlier than passage 25, bringing into question the reliability of the line. This may have also contributed to the increased sensitivity we observed in the IC_{50} D3 values. In an earlier analysis, Barrier et al. (2011) found the D3 mESC line to be genetically unstable. At passage 25 they reported that out of 11 cells analyzed none had a normal karyotype. ATCC has another D3 mESC line available (cat. no. CRL-11632). Although, this line is also assumed to be a high passage number line, it is still listed as being germline competent and might serve as a better, more stable, line for the EST. Continued performance and reliability of the cell lines are crucial elements for the continued development of *in vitro* alternatives; therefore, it is important that the cell lines demonstrate stability over extended periods. The J1 mESC line (ATCC, cat. no. SCRC-1010) has previously been used in

an adherent cell differentiation and cytotoxicity (ACDC) assay and after 24 passages 91% of the

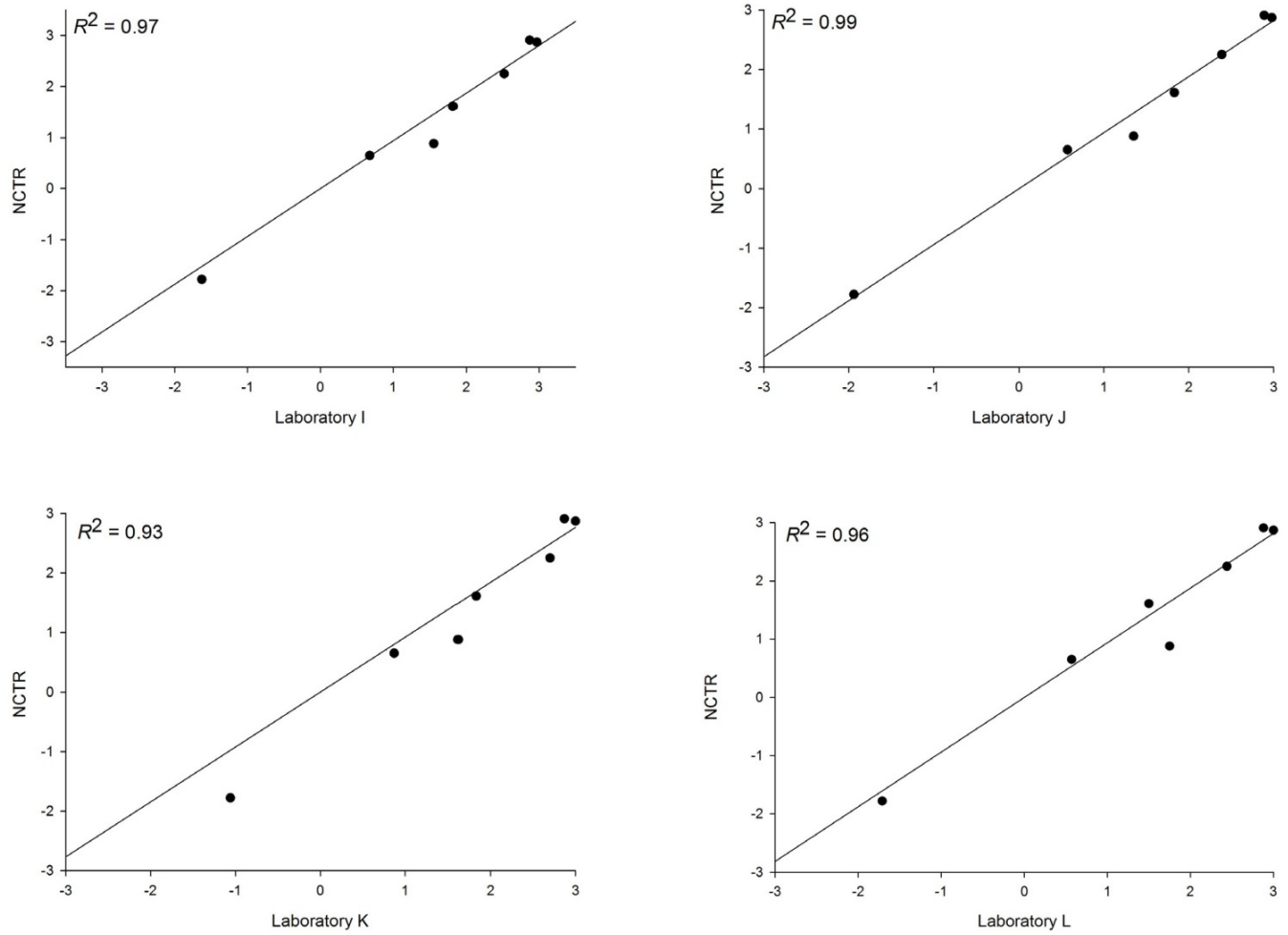

Figure 2. Interlaboratory comparisons of IC_{50} 3T3 values. Mean IC_{50} 3T3 endpoint values as reported in Appendix 2 of Genschow et al. (2004) were compared by pairwise scatter plots to the mean IC_{50} 3T3 endpoint values determined in our laboratory. For comparison, endpoint values were expressed as the log of the concentration (µg/ml) on both axes. Interlaboratory deviation is expressed by the R^2 values. Comparisons were made for seven of the eight chemicals tested; 5-fluorouracil was excluded from analysis.

J1 mESCs maintained a normal karyotype (Barrier et al., 2011).

We chose to follow the EST protocol that was utilized in the validation study. A more recent modification involves shortening the exposure time from 10 days to seven days, with exposure occurring on days three through 10 (Schulpen & Piersma, 2013). This change tends to decrease effects that some compounds may have on proliferation and may increase specificity of the assay for differentiation effects (van Dartel et al., 2009). It is unclear whether the modified protocol would have altered the sensitivity of our D3 cells to these compounds, especially since the growth rate of our cells may have been decreased compared to that of the line used during the validation study.

Regulatory acceptance of an *in vitro* or other alternative test for hazard identification or risk assessment will depend, in part, on the applicability domain of the test system as well as the predictivity of the test (van der Laan et al.,

2012). As is true with nearly all *in vitro* test systems, the EST is limited due to the lack of bioactivation capability (Hartung & Daston, 2009) and the solubility of test compounds, many of

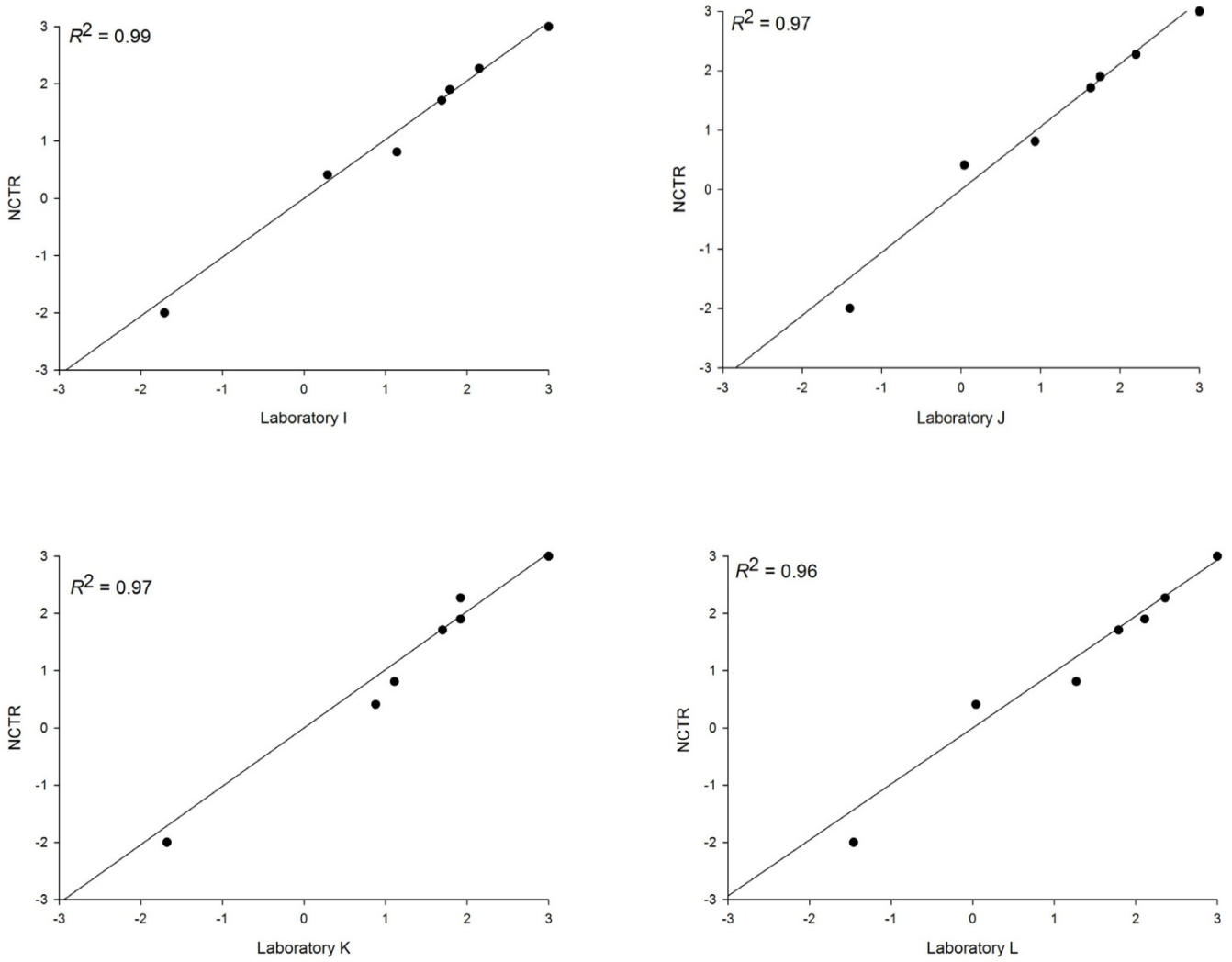

Figure 3: Interlaboratory comparisons of ID_{50} values. Pairwise scatter plots were used to compare mean ID_{50} endpoint values as reported in Appendix 2 of Genschow et al. (2004) to the mean ID_{50} endpoint values determined in our laboratory. Endpoint values were expressed as the log of the concentration in µg/ml and plotted on both axes. Interlaboratory deviation was assessed by the determination of R^2 values. Comparisons were made for seven of the eight chemicals tested; 5-fluorouracil was excluded from analysis.

which require solubilization in DMSO or ethanol (Hartung, 2007).

Several authors have pointed out that the prediction model developed by ECVAM in the validation study may need to be modified since it was developed using a particular set of chemicals and is thereby limited by the chemical and biological activity of those compounds. When used for other chemicals, the predictivity of the ECVAM model is not as high. For example, Paquette et al. (2008) found that a modified EST correctly identified only 53% of their in-house receptor-mediated compounds even though they were able to reproduce the results with the ECVAM validation compounds. The authors commented that the non-embryotoxic compounds used in the validation study were not very cytotoxic, while many of the low and moderate

risk compounds that were tested by Paquette et al. (2008) had lower IC_{50} values in 3T3 cells than either the ID_{50} or IC_{50} values in the D3 cells, leading to misclassification of these compounds. The assay correctly predicted only two of nine non-embryotoxic compounds.

Marx-Stoelting et al. (2009) employed a set of thirteen chemicals, only two of which were correctly predicted using the prediction model from the ECVAM validation study. Riebeling et al. (2012) examined the reason for the low predictivity by looking at several of the compounds tested by Marx-Stoelting et al. (2009). They identified six reasons for misclassification. One reason that up to 55% of these chemicals was misclassified was due to the "substance acting on a different tissue or at a later developmental stage". The authors further noted that the two major endpoints leading to the misclassification were neurotoxicity and bone development. Modifications to the EST to include both of these endpoints have been developed (zur Nieden et al., 2010; de Jong et al., 2012; Theunissen et al., 2010). However, these modified assays have not undergone the same type of validation process as the traditional EST. Another reason that up to 45% of the compounds were misclassified was thought to be due to the different nutrient composition of the culture media. For example, high levels of folic acid in the medium could explain the misclassification of ochratoxin A, and excess vitamin K may have played a role in the misclassification of warfarin. The additional possible causes for misclassification were generally limited to individual compounds.

Another reason for the misclassification of several chemicals may be due to the premise that the alternative tests should be compared to a defined "gold standard" (Wilcox & Goldberg, 2011). In most cases, the results of alternative tests have been measured against *in vivo* whole animal studies. Different groups of experts have selected different lists of compounds to be used as positives (embryotoxicants) and negatives (non-embryotoxicants) (reviewed in Lee et al., 2012). Recently, Daston et al. (2010) have suggested that instead of compounds classified as positive or negative, developmental toxicity exposures should be considered as a gold standard for assay validation. Since a compound may be non-embryotoxic at one concentration but embryotoxic at a higher concentration, Daston et al. (2010) proposed to develop lists of exposures consisting of chemicals at specific concentrations that would be expected to produce developmental toxicity and a second list of chemicals at specific concentrations that would not be expected to produce developmental toxicity. Thus far, such lists have not been produced. There has not been agreement on whether three categories (as utilized by ECVAM) or another number is appropriate. Some scientists believe that two categories (embryotoxic or non-embryotoxic) is sufficient (Marx-Stoelting et al., 2009), while the ReProTect study utilized four categories (strongly teratogenic, moderately teratogenic, mildly teratogenic and non-teratogenic; Marx-Stoelting et al., 2009). Additionally, experts have not always agreed on the *in vivo* classification of compounds. Following publication of the "Smith list" in 1983 (Smith et al., 1983), a workshop was held in 1991 to address the issue that many of the non-embryotoxicants were non-toxic under all circumstances and to develop a new list of test chemicals. However, after a number of discussions, this group of experts could not agree on a list of chemicals (Schwetz, 1992). Toxicity Testing in the 21st Century (National Research Council, 2007) suggested that the emphasis be placed on mechanistic and pathway evaluation rather than on empirical observations. To use data from whole animal studies as the gold standard for such analyses when the toxicological mechanism is rarely known would be inappropriate and would require a new validation model (Wilcox & Goldberg, 2011).

It is important to stress that just because an assay has gone through a formal validation process, this does not guarantee its regulatory acceptance. For regulatory applications the method must be shown to perform to its intended measurement of endpoints for acceptance (Wilcox & Goldberg, 2011). The authors also point out that while a standardized and validated alternative

test may be applicable and even preferred for hazard identification, it generally is not appropriate for exposure assessment and risk characterization, both of which are important elements in performing risk assessment. For example, a compound may be found by an *in vitro* test to be "safe" (that is, it is not a hazard) according to the concentration and duration of exposure used in the test system. However, these characteristics must then be compared to exposure criteria for the intended use of the compound; data that is derived currently from an *in vivo* study. This is simpler with drugs where exposure conditions are generally better known and controlled than with environmental chemicals where the exposure conditions are less well defined.

While the EST appears to be a valuable addition for the evaluation of possible developmental toxicants, there are still shortcomings with the test (Lee et al., 2012). The results from a workshop held in 2011 suggested that one step toward reducing the number of animals used in reproductive and developmental toxicity testing could involve the use of alternative tests (van der Laan et al., 2012). Many laboratories seem to be leaning toward the use of a battery of alternative tests; this battery will have to include all of the key stages of development and all of the endpoints currently evaluated in whole animal tests. Although modifications to the EST have been recommended to increase its utility and make it more quantifiable, generally the test appears to correctly classify approximately 75-80% of compounds, and the assay protocol can be transferred to other laboratories reliably. The assay also has the advantage of being a test that requires the sacrifice of no animals. To date the EST has generally not yet been used as a stand-alone method for assessing developmental toxicity, but overall, the strengths of the EST could make it one of the tests to be included in such a battery approach.

Declaration of conflicting interest

The authors state that there are no actual or potential conflicts of interest. Research was funded by the National Center for Toxicological Research, U.S. Food and Drug Administration.

Disclaimer

The views expressed are those of the authors and should not be construed to represent the views or policies of the U.S. Food and Drug Administration. Any reference to a specific commercial product, manufacturer, or otherwise, is for the information and convenience of the public and does not constitute an endorsement, recommendation or favoring by the U.S. Food and Drug Administration.

Acknowledgments

Research was funded by the National Center for Toxicological Research, U.S. Food and Drug Administration.

References

Barrier, M., Jeffay, S., Nichols, H.P., Chandler, K.J., Hoopes, M.R., Slentz-Kesler, K., & Hunter, E.S. III (2011). Mouse embryonic stem cell adherent cell differentiation and cytotoxicity (ACDC) assay. *Reproductive Toxicology, 31,* 383-391.

Brown, N.A. (2002). Selection of test chemicals for the ECVAM international validation study on in vitro embryotoxicity tests. *Alternatives to Laboratory Animals, 30,* 177-198.

Chapin, R., Stedman, D., Paquette, J., Streck, R., Kumpf, S., & Deng, S. (2007). Struggles for equivalence: *in vitro* developmental toxicity model evolution in pharmaceuticals in 2006. *Toxicology In Vitro, 21,* 1545-1551.

Daston, G.P., Chapin, R.E., Scialli, A.R., Piersma, A.H., Carney, E.W., Rogers, J.M., & Friedman, J.M. (2010). A different approach to validating screening assays for developmental toxicity. *Birth Defects Research, Part B, 89,* 526-530.

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. *Toxicology In Vitro, 26,* 970-978.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W., & Kemler, R. (1985). The in vitro development of

blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *Journal of Embryology and Experimental Morphology, 87,* 27-45.

Genschow, E., Scholz, G., Brown, N., Piersma, A., Brady, M., Clemann, H., Huuskonen, H., Paillard, F., Bremer, S., Becker, K., & Spielmann, H. (2000). Development of prediction models for three in vitro embryotoxicity tests in an ECVAM validation study. *In Vitro Molecular Toxicology, 13,* 51-66.

Genschow, E., Spielmann, H., Scholz, G., Pohl, I., Seiler, A., Clemann, N., Bremer, S., & Becker, K. (2004). Validation of the embryonic stem cell test in the international ECVAM validation study on three in vitro embryotoxicity tests. *Alternatives to Laboratory Animals, 32,* 209-244.

Genschow, E., Spielmann, H., Scholz, G., Seiler, A., Brown, N., Piersma, A., Brady, M., Clemann, N., Huuskonen, H., Paillard, F., Bremer, S., & Becker, K. (2002). The ECVAM international validation study on in vitro embryotoxicity tests: results of the definitive phase and evaluation of prediction models. *Alternatives to Laboratory Animals, 30,* 151-176.

Hartung, T. (2007). Food for thought…on cell culture. *Alternatives to Animal Experimentation, 24,* 143-147.

Hartung, T. & Daston, G. (2009). Are *in vitro* tests suitable for regulatory use? *Toxicological Sciences, 111,* 233-237.

International Conference on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2000). Detection of toxicity to reproduction for medicinal products and toxicity to male fertility S5(R2). Retrieved from http://www/ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Safety/S5_R2/Step4/S5_R2__Guideline.pdf.

International Conference on Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (2009). Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals M3(R2). Retrieved from http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Multidisciplinary/M3_R2/Step4/M3_R2__Guideline.pdf.

Lee, H.Y., Inselman, A.L., Kanungo, J., & Hansen, D.K. (2012). Alternative models for developmental toxicology. *Systems Biology in Reproductive Medicine, 58,* 10-22.

Martin, G.R., Wiley, L.M., & Damjanov, I. (1977). The development of cystic embryoid bodies *in vitro* from clonal teratocarcinoma stem cells. *Developmental Biology, 61,* 220-244.

Marx-Stoelting, P., Adriaens, E., Ahr, H.J., Bremer, S., Garthoff, B., Gelbke, H.P., Piersma, A., Pellizzer, C., Reuter, U., Rogiers, V., Schenk, B., Schwengberg, S., Seiler, A., Spielmann, H., Steemans, M., Stedman, D.B., Vanparys, P., Vericat, J.A., Verwei, M., van der Water, F., Weimer, M., & Schwarz, M. (2009). A review of the implantation of the embryonic stem cell test (EST). The report and recommendations of an ECVAM/ReProTect Workshop. *Alternatives to Laboratory Animals, 37,* 313-328.

National Research Council (2007). *Toxicity testing in the 21st century: a vision and a strategy*. Washington, D.C.: The National Academies Press.

Paquette, J.A., Kumpf, S.W., Streck, R.D., Thomson, J.J., Chapin, R.E., & Stedman D.B. (2008). Assessment of the embryonic stem cell test and application and use in the pharmaceutical industry. *Birth Defect Research Part B, 83,* 104-111.

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

Schulpen, S. H.W., & Piersma, A.H. (2013). The embryonic stem cell test. *Methods in Molecular Biology, 947,* 375-382.

Schwetz, B. A. (1992). Criteria for judging the relative toxicity of chemicals from developmental toxicity data: A workshop summary. *Teratology 45*, 337-339.

Seiler, A.E.M., & Spielmann, H. (2011). The validated embryonic stem cell test to predict embryotoxicity *in vitro*. *Nature Protocols, 6,* 961-978.

Smith, M.K., Kimmel, G.L., Kochhar, D.M., Shepard, T.H., Speilberg, S.P., & Wilson, J. (1983). A selection of candidate compounds for in vitro teratogenesis test validation. *Teratogenesis, Carcinogenesis & Mutagenesis, 3,* 461-480.

Spielmann, H., Gurne, B., Liebsch, M., Seiler, A., & Vogel, R. (2008). Successful validation of *in vitro* methods in toxicology by ZEBET, the National Centre for Alternatives in Germany at the BfR (Federal Institute for Risk Assessment). *Experimental Toxicology Pathology, 60,* 225-233.

Spielmann, H., Pohl, I., Döring, B., Liebsch, M., & Moldenhauer, F. (1997). The embryonic stem cell test (EST), an *in vitro* embryotoxicity test using two permanent mouse cell lines: 3T3 fibroblasts and embryonic stem cells. *In Vitro Toxicology, 10,* 119-127.

Theunissen, P.T., Schulpen, S.H.W., van Dartel, D.A.M., Hermsen, S.A.B., van Schooten, F.J., & Piersma, A.H. (2010). An abbreviated protocol for multilineage neural differentiation of murine embryonic stem cells and its perturbation by methyl mercury. *Reproductive Toxicology, 29,* 383-392.

Tremml, G., Singer, M., & Malavarca, R. (2008). Culture of mouse embryonic stem cells. *Current Protocols in Stem Cell Biology, 5,* 1C.4.1-1C.4.19.

van Dartel, D.A.M., Zeijen, N.J.L., de la Fonteyne, L.J.J., van Schooten, F.J., & Piersma, A.H. (2009). Disentangling cellular proliferation and differentiation

in the embryonic stem cell test, and its impact on the experimental protocol. *Reproductive Toxicology, 28*, 254-261.

van der Laan, J.W., Chapin, R.E., Haenen, B., Jacobs, A.C., & Piersma, A. (2012). Testing strategies for embryo-fetal toxicity of human pharmaceuticals. Animal models vs. in vitro approaches. *Regulatory Toxicology and Pharmacology, 63*, 115-123.

Wilcox, N., & Goldberg, A. (2011). Food for thought . . . on validation. A puzzle or a mystery: an approach founded on new science. *Alternatives to Animal Experimentation, 28*, 3-8.

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