Mutation Research 653 (2008) 23–33

Contents lists available at ScienceDirect

Mutation Research/Genetic Toxicology and Environmental Mutagenesis

journal homepage: www.elsevier.com/locate/gentox
Community address: www.elsevier.com/locate/mutres

Interlaboratory assessment of the GreenScreen HC *GADD45a-GFP* genotoxicity screening assay: An enabling study for independent validation as an alternative method

Nicholas Billinton[a,*], Paul W. Hastwell[b], Dominiek Beerens[c], Louise Birrell[a], Patricia Ellis[b], Sharon Maskell[d], Thomas W. Webster[b], Sam Windebank[d], Filip Woestenborghs[c], Anthony M. Lynch[b], Andrew D. Scott[d], David J. Tweats[e], Jacky van Gompel[c], Robert W. Rees[b], Richard M. Walmsley[a,f]

[a] *Gentronix Ltd, CTF Building, 46 Grafton Street, Manchester M13 9NT, UK*
[b] *GlaxoSmithKline plc, Park Road, Ware, Herts SG12 0DP, UK*
[c] *Johnson&Johnson Pharmaceutical Research and Development, Turnhoutseweg 30, Department of ADME/TOX, B-2340 Beerse, Belgium*
[d] *Safety and Environmental Assurance Centre, Unilever, Sharnbrook MK44 1LQ, UK*
[e] *Genetics Department, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, UK*
[f] *Faculty of Life Sciences, The University of Manchester, Manchester M13 9PT, UK*

ARTICLE INFO

Article history:
Received 21 May 2007
Received in revised form 21 January 2008
Accepted 14 February 2008
Available online 27 March 2008

Keywords:
Compound screening
High-throughput
In vitro assay
Genotoxicity
Genotoxin
Genetic toxicology
GADD45a
Transcriptional induction
Green fluorescent protein
GreenScreen HC
TK6
Human lymphoblastoid
Interlaboratory
Transferability

ABSTRACT

Sixteen coded compounds were blind-tested at 4 laboratories using the recently described *GADD45a-GFP* genotoxicity assay. The compounds were chosen to include non-genotoxic compounds as well as weak and strong genotoxins. None of the compounds required metabolic activation in order to exhibit genotoxic effects. The participating laboratories included 2 global pharmaceutical companies, a global consumer goods company and the Gentronix laboratory in Manchester. Each compound was tested 4 times on different days following a protocol previously described. The tests were carried out after a 3-day training period from the parent lab (Manchester). Following the exclusion of data from tests with positive control failures and data series with 'spikes', 92% of assays gave the correct result: non-genotoxins giving negative results and genotoxins giving positive results. There were no randomly distributed problems suggesting that differences between the results from different sites reflected the use of different instruments, procedural differences and operator experience. In naïve operator laboratories the quality of data improved with operator practice. It was concluded that simple clarification of the protocol would provide the level of reliability required for widespread use of the assay in hazard assessment.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Many new assays and variants of existing assays are introduced each year for use in the discovery and development of novel products. In the pharmaceutical industry, early high-throughput assay methods used to identify 'hits' may only be used at a single location for a particular discovery campaign. They will be designed and validated by a small, dedicated team, and their success is defined by their ability to identify useful new compounds with potential efficacy. This is in contrast to assays used for the safety assessment of 'hits', which have become 'leads' and are being considered for development. Assays for safety assessment are applied for all new chemical products, not just pharmaceuticals. Such assays are used at many locations and are subjected to a rigorous process of validation to establish scientific relevance and reliability to an endpoint of concern. Relevance is established by the testing of

* Corresponding author. Tel.: +44 161 603 7661; fax: +44 161 606 7337.
E-mail address: nick.billinton@gentronix.co.uk (N. Billinton).

1383-5718/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mrgentox.2008.02.011

compounds for which there are useful comparative data related to the endpoint of concern in order to illustrate predictive capacity and applicability. Reliability is established by demonstrating the reproduction of results within a laboratory, as well as the transfer of protocols to, and performance at multiple sites. The process for the validation of new test methods has recently been reviewed by the European Centre for the Validation of Alternative Methods (ECVAM), which proposed a modular approach to make it a more data-driven process [1].

Within the regulatory system to register pharmaceuticals there is a requirement for a battery of genotoxicity tests [2,3], designed to identify compounds that might damage or in other ways alter the genome and hence pose a carcinogenicity and/or heritable mutation hazard. The need for new, validated genotoxicity tests has become apparent from recent reviews of the effectiveness of the existing regulatory *in vitro* mammalian genotoxicity tests (for example, [4–7]). It is clear that the established regulatory tests are effective in the identification of genotoxic carcinogens. However, the *in vitro* mammalian tests lack specificity, which leads to the generation of positive genotoxicity data for well over half of the non-carcinogens tested. A new genotoxicity assay has recently been developed, in which genotoxin-induced transcription of the *GADD45a* gene drives the synthesis of green fluorescent protein (GFP) in a human lymphoblastoid cell line (TK6; [8]). Earlier studies on *GADD45a* originating from work in the Fornace lab [9,10], coupled with a published validation study of the *GADD45a* reporter system [11] have established the scientific relevance of this new genotoxicity screening endpoint. A study of 75 compounds tested with the *GADD45a-GFP* reporter assay revealed that it identified genotoxins with the same high sensitivity as the regulatory tests but, importantly, that it was far more effective in giving negative results with non-genotoxins.

In this paper we report an interlaboratory study designed to provide insight into aspects of the reliability of the new genotoxicity assay. The study followed the general principles of the ECVAM modular criteria [1], but with some key differences, to which attention is drawn.

The first module in the ECVAM guidelines identifies the need to define a new test by the following criteria: protocols; Standard Operating Procedures (SOPs); the endpoint; a training set of data including adequate controls; definition of the predictive model and explanation of the mechanistic basis for the test. The *GADD45a-GFP* genotoxicity test has been described in detail elsewhere by Hastwell *et al.* [11] and their paper defined the test appropriately, as well as providing a validation study based on the assessment of 75 compounds. The endpoint, which may be defined as an indirect measure of genotoxic damage, is described in detail as follows. Agents that damage DNA directly, or affect proteins involved in the critical processes of DNA synthesis and repair, or the mechanisms of chromosome segregation, lead to increased transcription of the *GADD45a* gene. This in turn reflects activation of the appropriate cellular response to these agents, which ultimately result in cell cycle regulation, DNA repair or apoptosis. In the context of the DNA damage response, the role of *GADD45a* in these processes is relatively well described at the mechanistic level, largely as a consequence of the studies originating from the Fornace lab [9,10]. The human lymphoblastoid cell line, TK6 has been genetically modified to contain a reporter system in which the promoter and other key regulatory elements of the *GADD45a* gene are operationally linked to a gene encoding GFP. Following exposure to genotoxins the level of reporter expression is assessed by measuring cell brightness: fluorescence divided by optical density. A positive result for genotoxicity is recorded if the brightness increases by 50% or more. This represents a statistically significant increase in brightness (greater than 3 times the standard deviation in data from untreated cells) and the threshold effectively discriminated between genotoxins and non-genotoxins in the validation study. The assay endpoint therefore reflects a biologically relevant exposure to genotoxins. The test protocol described in the Hastwell *et al.* paper [11] was reduced to an SOP for the study reported here. In one of the participating laboratories the SOP was used as a template to define a protocol compliant with Good Laboratory Practice (GLP).

The second ECVAM module establishes three levels of assessment for reliability. The first is "within-laboratory reliability" and involves an assessment of reproducibility of experimental data in a single laboratory. This was demonstrated in the published study on the training set [11]: 74 of the 75 compounds tested gave either 4 positive results or 4 negative results in repeated experiments. The second level of assessment is "transferability", in which it should be demonstrated that the test can be successfully repeated in a laboratory other than the developing or optimising centre. Transferability is seen as key in evaluating the practicability of the test but is also necessary for determining the degree of training required for a naïve laboratory, as well as in attempting to identify potential sources of variability, both within- and between-laboratory. The third level of assessment is "between-laboratory variability" which entails an assessment of reproducibility of experimental data in 3 or 4 well-trained laboratories with a relatively large number of test compounds. The study presented in the current paper was designed to address transferability but also incorporated elements of a between-laboratory variability study, since experiments were performed by 'experienced' users at one site and naïve users at the other three sites. This significant difference from the ECVAM guidance for a between-laboratory variability study was a deliberate strategy, since the trial presented here was intended as an enabling study to evolve the definition of effective protocols in preparation for a wider, independent trial as well as for the commercial release of the assay.

2. Materials and methods

The *GADD45a-GFP* GreenScreen HC assay protocol, data handling, and decision thresholds are described in full by Hastwell *et al.* [11]. Four compounds were tested in a single experiment (one 96-well microplate). In this study, compounds were coded and provided in two formats; either dissolved in 100% DMSO (13 compounds) and hence requiring aqueous dilution (to 2% DMSO, v/v) prior to addition to the microplate, or dissolved in 2% aqueous DMSO (v/v; 3 compounds) and ready for direct addition to the microplate without further aqueous dilution. Each compound was added to duplicate wells (150 μl per well) of the same microplate before serially diluting each by transferring 75 μl from the first well into 75 μl of 2% DMSO in the next well, mixing and then transferring 75 μl into the next well. In this way, two series of nine serial dilutions (75 μl per well) were created for each compound.

Cells were taken from cultures that had been in passage for at least 2 weeks since resuscitation from frozen. The protocol provided required that cells were subcultured into fresh culture medium at least every 3–4 days and that cell counts should not be allowed to exceed 1.2×10^6 cells/ml. Required volumes of cultures were harvested, washed in PBS to remove traces of culture medium and then resuspended in assay medium to give a cell density of 2×10^6 cells/ml. One dilution series for each compound on the microplate was a 'test series' and had TK6 cells containing the *GADD45a-GFP* reporter system (test strain, "GenM-T01") added to a final cell concentration of 1×10^6 cells/ml in 1% (v/v) DMSO (75 μl per well of the 2×10^6 cells/ml culture). The repeat series for each compound was identical but had TK6 cells in which the reporter is unable to express GFP (control strain, "GenM-C01") added to a final cell concentration of 1×10^6 cells/ml in 1% (v/v) DMSO. The latter allows identification of, and correction of data from compounds that are either inherently fluorescent or induce cellular autofluorescence.

In order to provide a measure of the maximum proliferative potential and also the baseline GFP reporter signal, the compound dilution series were always accompanied by wells containing only cells in assay medium and solvent vehicle (4 wells per strain per microplate). Each microplate experiment also included a positive control for both genotoxicity and cytotoxicity (methyl methanesulfonate, MMS; at two concentrations, with each concentration duplicated), as well as additional controls for dilution buffer and growth medium sterility. After microplates were filled, they were sealed using gas-permeable membranes (Breathe-Easy; Diversified Biotech, USA). Following incubation of the microplates in a CO_2 incubator (37 °C, 5% CO_2, 95% humidity), GFP fluorescence (excitation at 485 nm, emission at 535 nm) and cell culture absorbance (620 or 612 nm) measurements were collected at 24 and

48 h using Tecan Ultra384 readers (Tecan UK Ltd, Theale, UK). After data collection at 24 h, microplates were re-sealed and incubated for a further 24 h after which further fluorescence and absorbance measurements were made.

Absorbance data were normalised to the untreated control (=100% growth) and used to give an indication of any inhibitory effect of a compound on cell proliferation ("relative suspension growth"; RSG). Inhibition of cell proliferation is a recommended measurement endpoint for basal cytotoxicity according to guidance issued by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAMs) [12]. 'Brightness units' were determined for each well by dividing the fluorescence data by the absorbance data and then normalised to the untreated control (=1). This allows discrimination between a well with a high number of weakly fluorescing cells and a well with a lower number of strongly fluorescing cells, and was essentially a measure of mean cellular fluorescence. These simple calculations were performed in a Microsoft Excel data-processing template (with macro) which also displayed the data graphically, in the form of dose–response curves. This template allowed the generation of a report containing the genotoxicity classification as well as the 'toxicity' result (from RSG). The software also reported the success or failure of the positive control wells, an important data acceptance criterion for users of the assay. This study presented the opportunity to assess the value of positive control performance as a data acceptance criterion.

The decision thresholds for data interpretation were described in full by Hastwell *et al.* [11]. Briefly, within the data-processing template there are defined numerical thresholds of statistical significance to enable the classification of a test compound by both genotoxicity and cytotoxicity parameters, as well as a data rejection threshold. The genotoxicity threshold (increase in brightness) is set at 1.5 (50% increase) and this is greater than three times the standard deviation of the background brightness. A positive result for genotoxicity is recorded for data that reach and cross the threshold at either 24 or 48 h. The cytotoxicity threshold (decrease in RSG or proliferative potential) is set at 80% of the maximum extent of cell proliferation (for untreated control cells). It must be noted that mortality is not measured in this assay: 80% RSG does not mean that 20% of the cells are dead. There is currently no distinction between cytocidal and cytostatic effects in this assay. The data rejection threshold for genotoxicity is set at 30% RSG, which reflects the issues of confounding optical interference and the inability of the population to complete doubling below that RSG. The optical density giving 30% RSG is approximately equivalent to the optical density of the inoculum. Below this figure cell lysis has occurred, which critically compromises the brightness calculation.

3. Organisation of the study

The exercise was initiated at laboratories in the University of Manchester (UK), housing the research group of Richard Walmsley and Gentronix Ltd. Three further laboratories were invited to participate. These were at GSK (Ware, UK), J&J (Beerse, Belgium) and Unilever (Sharnbrook, UK). All three had participated in previous published [13,14] and unpublished studies of the yeast-based GreenScreen GC assay. In this paper, each participating laboratory has been assigned a location number, from Site 1 to Site 4. The initial aims of the study were akin to those of a 'pre validation' study [15], as follows: to evaluate the protocol and SOP provided; to test the transferability of the assay and protocol; to learn and understand more about performance of the assay, in particular, with a view to proceeding to an independent, broader between-laboratory variability study and commercial launch of the assay.

In order to limit variation between laboratories, a number of components were obtained or prepared as single batches then divided and distributed between the partners. These included the following: two sets of frozen cell cultures; growth medium (RPMI-1640); medium supplements (sodium pyruvate, penicillin/streptomycin mix, hygromycin B, heat-inactivated donor horse serum); 96-well microplates (Matrix ScreenMates 4929 from Thermo Fisher Scientific); microplate sealing membranes (BreathEasy, Diversified Biotech Inc); coded test compounds in either 100% DMSO or 2% aqueous DMSO (v/v with respect to DMSO), along with coded compound information sheets; assay medium; detailed instruction sheets referring to the passage regimes and test method to be used; MS Excel data-processing templates. The latter ensured that the calculation, interpretation and reporting of results would be carried out under standardised procedures, that the results were in a standard format, and that there was no subjective interpretation. The variables in the exercise were as follows: number of laboratories (4); number of personnel involved in testing (6); number of days (4); number of compounds (16); assay duration (24 and 48 h).

The 16 compounds chosen included 8 genotoxins and 8 non-genotoxins that were all obtained at the highest purity available (Sigma, Aldrich, Fluka, and Riedel-de Haën); these are listed in Table 1 along with published results from the bacterial (*Salmonella* test) and mammalian cell tests (*in vitro* and *in vivo* chromosome aberration/micronucleus test and mouse lymphoma assay) that comprise the regulatory battery of tests for genotoxicity [2,3], as well as rodent carcinogenicity and GreenScreen HC data. None of the genotoxins needed exogenous metabolic activation to express genotoxic effects in *in vitro* assays, as an appropriate S9 protocol was only developed after this study [manuscript in preparation]. The highest concentrations to be tested for these compounds were defined by assays performed with the training set of compounds in the published validation exercise [11]. In the Hastwell *et al.* study, these 16 compounds were treated as unknowns and tested at concentrations matching the maximum recommended doses for *in vitro* mammalian cell tests according to ICH Guidelines (10 mM or 5 mg/ml; [3]). Test concentrations were then adjusted according to compound solubility and cytotoxicity as observed in the preliminary assays. The 16 compounds were prepared, coded and distributed by GSK along with assay medium for the trial. The identities of the compounds were not revealed to any of the users until all parties had completed testing and submitted full sets of results to the originating laboratory (Gentronix Ltd). All data were submitted in the standard GreenScreen HC data-processing template Excel files.

One of the authors (Nick Billinton from Gentronix Ltd.) was the Trial Director. He carried out the initial 3-day training at each site, the compilation of results and the preparation of the final report used in the writing of this paper. Each participating laboratory nominated a contact person responsible for communicating to other participating laboratories and that person was required to coordinate the interlaboratory trial within their own organisation. This was achieved through e-mail and telephone communication. The Trial Director ensured that correct test methods were used by visiting the participating laboratories, demonstrating the protocol with 4 test compounds, observing the participants carrying out the assay and offering advice on handling methods. This was a minimal essential training exercise and the programme did not require that the participating laboratories generated a 'perfect' data set for the 4 compounds. This was because part of the aim of the study was to test the independence of the protocol. For similar reasons there was only a small excess of materials supplied. This was restricted to materials for 6 complete repeats of the 16 compounds—accepting that there might be unintended destruction or loss of a set of samples, or handling errors due to lack of familiarity with the protocol.

Several operational aspects had to be agreed in advance. The timing of the study had to fit into the availability of scientists to carry out the trial at each site, but in order to ensure the integrity of the trial, there had to be a defined period over which it would be carried out. In practice the limits were defined as follows. Each compound was tested 4 times on different days at each site over a minimum period of 2 weeks. All the laboratories completed their testing within a 2-month period of fresh material supply. The latter was set because preliminary studies showed that none of the supplied materials had shown any detectable variation over a period of 3 months following preparation.

It was agreed that preliminary handling of the data set would use only original data. All results were included, even those with obvious experimental control failure or user error, since it is necessary to consider all data in the assessment of the transferability

Table 1
Test compounds

Code	Compound name	CAS registry #	Mechanism	GADD45a-GFP	Ames		*In vitro* CA		MLA		*In vivo* MNT/CA		Rodent carc.	
				Result [11]	Result	Ref.	Result	Ref.	Result	Ref.	Result	Ref.	Result	Ref.
A	5-Fluorouracil	51-21-8	Nucleotide synthesis	+	–	[16]	+	[17]	+	[18]	+	[19]	+	[20]
B	3-Amino-1,2,4-triazole	61-82-5	Negative (cytotoxic)	–	–	[21,22]	–	[16,23]	–	[16,21,24–26]	–	[27]	+	[27]
C	4-Nitroquinoline-1-oxide	56-57-5	Direct genotoxin	+	+	[16]	+	[28,29]	+	[24]	+	[19]	+	[21]
D	Paclitaxel	33069-62-4	Aneugen	+	–	[30]	+	[30]	n/a		+	[31]	n/a	
E	Ethylene glycol	107-21-1	Negative	–	–	[16,22]	–	[16]	–	[24]	n/a		–	[16]
F	Aphidicolin	38966-21-1	Nucleotide synthesis	+	–	[32,33]	+	[34]	+	[35]	n/a		n/a	
G	Ampicillin (Na salt)	69-52-3	Negative	–	–	[16]	–	[16,36]	–	[24]	–	[36]	E	[13,16,27,37,38]
H	Phenformin HCl	834-28-6	Negative (cytotoxic)	–	–	[16]	–	[16,29]	–	†	n/a		–	[16]
I	Etoposide	33419-42-0	Topoisomerase inhibitor	+	–	[21,39,40]	+	[41,42]	+	[39]	+	[39]	I	[27]
J	Methylnitrosourea	684-93-5	Direct genotoxin	+	+/–	[43,44]	+	[17]	+	[24]	+	[16,19,45]	+	[27]
K	D-Mannitol	69-65-8	Negative	–	–	[16]	–	[16,29]	–	[24]	–	[16]	–	[16,46]
L	Sodium chloride	7647-14-5	Negative	–	–	[16]	–	[16,29]	–	[47]	–	[29]	–	[37,48,49]
M	Vincristine sulphate	2068-78-2	Aneugen	+	–	[29,50]	+	[17,51]	+	[50]	+	[19,45]	I	[27]
N	Chloramphenicol	56-75-7	Negative (cytotoxic)	–	–	[22]	+	[52]	+	[24]	–	[53]	I	[27]
O	2,4-Dichlorophenol	120-83-2	Negative (cytotoxic)	–	–	[21]	+	[54]	+	[16] †	+	[55]	–	[27]
P	Camptothecin	7689-03-4	Topoisomerase inhibitor	+	–	[56]	+	[42,57]	+	[58]	+	[59]	n/a	

The compounds chosen for this exercise are listed by name along with the code employed for blind testing, the CAS registry number, the mechanism of action and assay result for genotoxicity, mutagenicity and rodent carcinogenicity. Published data from alternative tests are provided in the table: *Salmonella* test (Ames), *in vitro* chromosome aberration test (*in vitro CA*), mouse lymphoma assay (MLA), *in vivo* micronucleus test/chromosome aberration (*in vivo* MNT/CA), and rodent carcinogenicity (Rodent carc.), alongside published data from the *GADD45a-GFP* training set. Key: Negative result (–), positive result (+), equivocal (E), inconclusive (I), in-house data from Site 4 (†), and data not available (n/a).

of an assay. At the follow-up meeting a panel decision approach was used to determine whether for an individual compound there was uncertainty in the results or uncertainty in the measurements. This was necessary as it was recognised that there were a number of factors originating from differences in test equipment and environment that would contribute to the precision of the test procedures. Indeed this would be the case in any new laboratory setting up a new assay. There were no procedures in place to handle outliers as it was decided to handle these case by case at the meeting to discuss the data set. The primary decision-making tool for the assay was the objective data-handling software—no subjective operator judgement was used in producing the original data sets.

4. Results

Each of the 16 coded compounds was blind-tested 4 times at each participating site. Four compounds were tested in each microplate assay and a single compound test produced two sets of data for genotoxicity and cytotoxicity, from nine concentrations across a two-fold serial dilution, at both 24- and 48-h time points. In addition, control data were collected from every microplate assay. Due to this high-throughput format, failure of a control compromises the data for the four compounds tested on that microplate.

The intention of the study was to analyse all aspects of the assay protocol. This of course included the various in-plate controls. In the preliminary analysis of the data set as a whole, it was clear that there was a higher than expected prevalence of positive control failure, i.e. the high dose of MMS did not reach a minimum two-fold GFP induction and/or did not produce a higher induction than the low dose. The control failures were mainly at one trial site. To understand the contribution of the positive control to the outcome of the test, the experimental data sets were analysed both with and without data exclusion from tests with control failures.

The prediction model for the GreenScreen HC *GADD45a-GFP* assay relies on a minimum 50% increase in relative fluorescence induction to identify a genotoxin. Data were assessed using this prediction model and in the initial consideration no data were excluded, i.e. data from microplates where controls did not pass the defined criteria were included. The overall genotoxicity results are summarised in Table 2a for the 24 h time point and in Table 2b for the 48 h endpoint. The summary was created using all 4 repeat tests for each compound at each site. Three or four repeats with the same result lead to a summary result of the same class ('+' for genotoxic, '–' for non-genotoxic) whereas a compound divided between two genotoxic and two non-genotoxic results was classed as an overall equivocal result, '±'. In this summary, all four sites showed good prediction in being able to identify 6 or more of the 8 genotoxins at either the 24 or 48 h time points. At the 48 h time point, three sites were able to identify 7 of the 8 non-genotoxins whilst 'Site 3' identified all 8. In fact, Site 3 also identified all 8 of the genotoxins at 48 h.

Assessment of all data from all tests at all sites at the 48 h time point (Table 3) revealed an overall concordance with expected results of 86.3%. Each individual site produced concordance of >75% from the unadulterated data set, with Site 3 performing the best. Overall sensitivity across the trial was >80% and specificity was >90%. Sites 1 and 2 gave the lowest sensitivities and negative predictive values from the all-inclusive analysis.

Further analysis of the complete dataset from Site 1 revealed a preponderance of positive control failures; of the 12 (from a total of 64) microplate assays that showed positive control failures at the 24 h time point, 11 occurred at Site 1. This control failure rate at Site 1 increased to 14 assays by the 48 h time point. This relatively high control failure rate has not been observed by experienced

Table 2
Overall genotoxicity results

Compounds	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
(a)																
Expected	+	–	+	+	–	+	–	–	+	+	–	–	+	–	–	+
Site 1	+	+	+	–	–	–	–	–	+	+	–	±	+	–	–	+
Site 2	+	+	+	+	–	+	–	–	+	+	–	–	±	–	–	+
Site 3	+	±	+	±	–	+	–	–	+	+	–	–	–	–	–	+
Site 4	+	+	+	+	–	+	–	–	+	+	–	±	–	–	–	+
(b)																
Expected	+	–	+	+	–	+	–	–	+	+	–	–	+	–	–	+
Site 1	+	±	+	+	–	–	–	–	+	+	–	–	–	–	–	+
Site 2	+	±	+	±	–	–	–	–	±	+	–	–	±	–	–	+
Site 3	+	–	+	+	–	+	–	–	+	+	–	–	+	–	–	+
Site 4	+	+	+	+	–	+	–	–	+	+	–	–	–	–	–	+

Genotoxicity results for all assays at all sites are summarised for the 24 h time point (a) and the 48 h time point (b). The quadruplicate results for each compound at each site were interpreted as follows: 3 or 4 tests producing the same result gave an overall interpretation with the same result (e.g. 3 or 4 negative results for a compound gave an overall '–'); compound data divided into 2 positive and 2 negative results produced an equivocal interpretation ('±'). 'Expected' results are provided according to results from the published training set for the *GADD45a-GFP* assay.

users of the assay (e.g. 1 user has recorded 7 positive control failures in 120 assays over 6 months, 4 of which were in the training period of a new operator). The positive control failures were not randomly distributed but instead they largely correlated with the experience of the test operator. For example, Site 1, the least experienced lab both in terms of this assay and the assay format, had the highest level of positive control failure, whereas Site 3, the most experienced lab with this assay format, had only 1 positive control failure.

A similar correlation was found between experience with the assay and variation in the growth of untreated control cells (i.e. the maximum extent of proliferation within each assay). Table 4 summarises averaged absorbance data from untreated cultures at the 48 h time point for all assays at each site. Site 1 had the highest variation in absorbance (Relative Standard Deviations (R.S.D.s) of 11.9% and 14.0%) and Site 3 had the lowest variation (R.S.D.s of 2.8% and 4.3%) and only 1 positive control failure (see further discussion below).

The overall assessment of the whole trial data set was repeated after removal of data arising from microplate assays where positive controls had failed (Table 5). Four compounds were tested per microplate assay and so control failure resulted in the removal of the corresponding data for the four compounds. In addition, data were removed where a positive result was recorded as a consequence of a spike (see further comment below). This application of data acceptance criteria resulted in the concordance with expected results for the overall study rising from 86.3 to 92.5%. Correct identification of both genotoxins and non-genotoxins across all sites and repeat experiments was >90%. It should be noted that after applying the acceptance criteria, data from only 8 compound tests remained of the data set from Site 1. However, if Site 1 was completely removed from the analysis, the overall figures reported in Table 5 would only change by ≤1% (sensitivity from 94 to 93.6%, specificity from 90.8 to 91.8%, positive predictivity from 91.8 to 92.4%, negative predictivity from 93.2 to 93.1%, and concordance from 92.5 to 92.7%). It is clear that the positive controls provide good acceptance criteria for data analysis.

There were just three compound assays where an apparently positive result was caused by a single spike (at one or both time points; see for example Fig. 2c, below) and in these cases only the individual compound's data series was removed in the analysis for Table 5. In tests of one compound (compound O; 2,4-dichlorophenol) at Site 3, there was an outlier within the four repeat tests. Visual inspection of the data (see Fig. 1) revealed an anomalous downward trend in the brightness data for the control cells (the lower line in Fig. 1a), which led to an upward trend in the genotoxicity evaluation (control cell brightness subtracted from the test strain brightness), and a single point at the positive threshold (Fig. 1b).

Examples of graphical data for different outcomes of the assay from the ring trial are presented in Fig. 2. Fig. 2a and b show the averaged fluorescence induction data from four tests for a genotoxin (compound C; 4-nitroquinoline-1-oxide)

Table 3
Initial assessment of assay performance

	Site 1	Site 2	Site 3	Site 4	Overall
Sensitivity	75.0	62.5	100.0	90.6	82.0
Specificity	90.6	90.6	93.8	87.5	90.6
Predictive value (+)	88.9	87.0	94.1	87.9	89.7
Predictive value (–)	78.4	70.7	100.0	90.3	83.4
Concordance	82.8	76.6	96.9	89.1	86.3

Preliminary predictivity statistics for *GADD45a-GFP* assay performance at the 48 h time point for all of the sites. The statistics were calculated for prediction of the expected genotoxicity results as defined by the published training set. Calculations for the terms used are defined elsewhere [11].

Table 4
Assay absorbance variation across the exercise

Test site	Control strain			Test strain		
	Ave. Abs.	St. Dev.	R.S.D. (%)	Ave. Abs.	St. Dev.	R.S.D. (%)
1	0.0837	0.0100	11.9	0.0890	0.0124	14.0
2	0.1008	0.0060	6.0	0.0983	0.0072	7.4
3	0.1157	0.0033	2.8	0.1184	0.0051	4.3
4	0.0948	0.0064	6.8	0.0938	0.0063	6.8

The table shows the averaged absorbance ("Ave. Abs.") reading for the untreated control and test TK6 cells from 16 microplate assays at each site (at 48 h); the standard deviation ("St. Dev.") in the mean absorbance at each site; and the relative standard deviation ("R.S.D. (%)") as a percentage of the mean value.

Table 5
Re-assessment of assay performance

	Site 1	Site 2	Site 3	Site 4	Overall
Sensitivity	100.0	86.4	100.0	92.0	94.0
Specificity	66.7	94.1	93.1	88.9	90.8
Predictive value (+)	83.3	95.0	93.9	88.5	91.8
Predictive value (–)	100.0	84.2	100.0	92.3	93.2
Concordance	87.5	89.7	96.7	90.4	92.5

Predictivity statistics re-calculated for all sites using 48 h time-point data after application of data acceptance criteria and removal of obvious data 'spikes'.

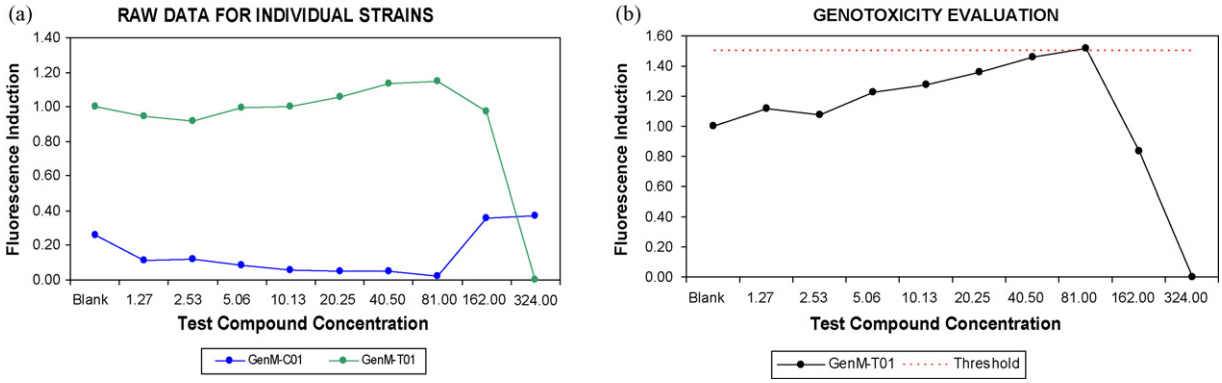

Fig. 1. Anomalous result for 2,4-dichlorophenol (compound O) at Site 3. (a) Fluorescence induction for the control TK6 strain (blue line) with unusual downward trend and the test TK6 strain (green line). (b) The resulting genotoxicity evaluation graph (control strain brightness subtracted from the test strain) with upward trend to the threshold. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

and a non-genotoxin (compound L; sodium chloride), respectively, from each participating site. The averaged dose–response curves in Fig. 2a show similar profiles at each site, with induction peaking at the same compound concentration. The averaged curves in Fig. 2b show no dose response at any of the sites and also reveal very little variation either within or between sites. Fig. 2c shows data with a discontinuity in the genotoxicity response that crosses the genotoxicity threshold (a 'spike' for compound L; sodium chloride). Fig. 2d shows data from a compound without a no-effect dose, leading to difficulty in data interpretation (compound M; vincristine sulphate, see below).

4.1. Summary of panel discussion on practical aspects of the trial

Representatives from the participating laboratories met at GSK (Ware, UK) to discuss all aspects of the trial following the sharing of data. The following section reports on the practical issues raised during the round-table discussions.

All participants considered that the training was an invaluable and essential contribution to the success of the exercise. This view reinforces the importance of recognising that written protocols from originating laboratories often fail to include all the detail needed to fully define a new protocol. It was however recognised that less experienced labs improved performance during the trial.

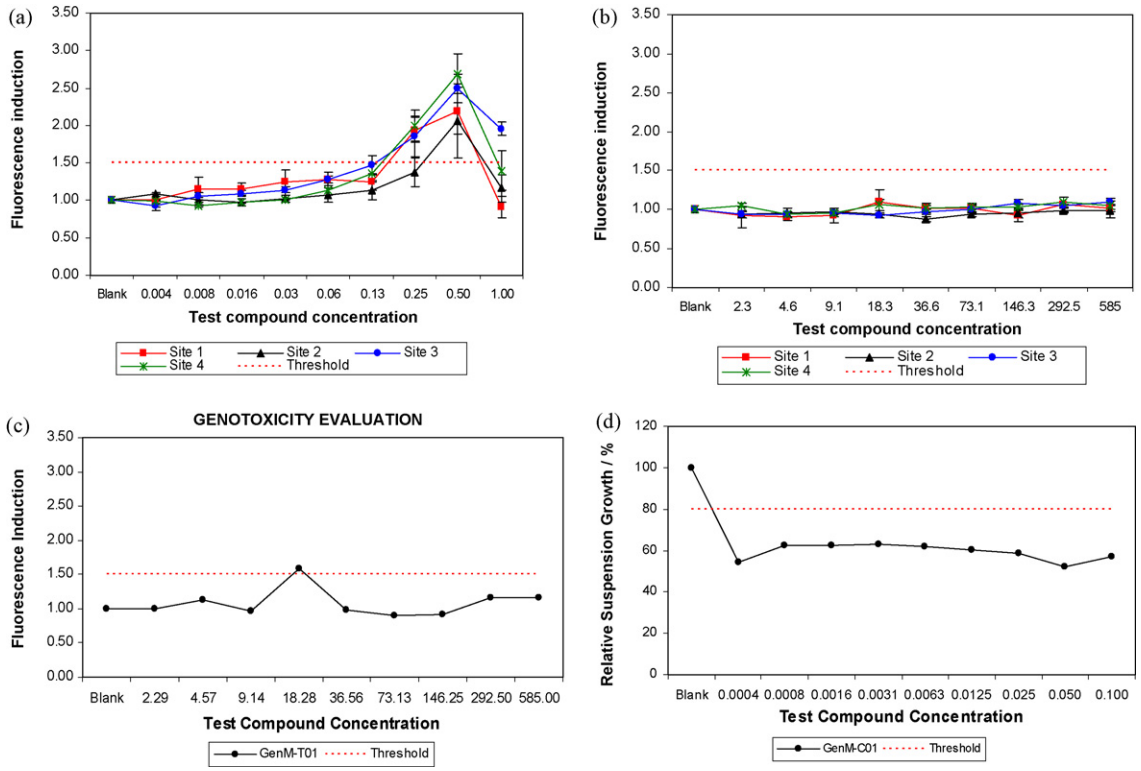

Fig. 2. Example data from the study (all test compound concentrations are in μg/ml). Averaged fluorescence induction data for a genotoxic compound (a), compound C, 4-nitroquinoline-1-oxide; and a non-genotoxic compound (b) compound L, sodium chloride, from each participating site (red = Site 1; black = Site 2; blue = Site 3; green = Site 4; error bars represent the standard error of the mean). (c). An example of an obvious 'spike' in the genotoxicity data at 24 h for compound L (sodium chloride) from the second test at Site 1. (d). Example cytotoxicity data (RSG) illustrating the lack of no-effect dose for compound M (vincristine sulphate; extracted from the fourth test at Site 2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Gentronix reported that since the completion of the trial, new users of the assay had been provided with material to allow 3 repeat plate assays (4 compounds per plate) to be performed following training. This addition enables users to self-certificate assay performance in-house, subsequent to the formal training by the assay provider. The utility of self-certification was reinforced by the observation that experienced microplate users obtained the best results; inexperienced users obtained more variable ('noisier') results (see Table 4). In this trial, the operators at two of the sites were inexperienced microplate users and at a third site a recently appointed technician carried out the testing.

With regards to the materials supplied centrally for the trial, the first set of frozen cells was successfully revived in the growth medium in all laboratories except one. The second aliquot was successfully revived at that site. The volume of assay medium plus cells (4 ml) defined by the protocol did not provide sufficient 'dead volume' for ease of handling. The assay medium contained horse serum, which can cause frothing and adhesion to surfaces, increasing the importance of the excess volume. Participants agreed that 5 ml aliquots would have been more appropriate, and that protocols for studies following this exercise would be amended appropriately.

The panel discussed the utility of shaking the assay microplate prior to measurement. Thirty seconds of vigorous cell dispersion/resuspension on a microplate shaking instrument, before collecting fluorescence and absorbance data produces the best results. One laboratory noted that a microplate, in which there was the appearance of cell aggregation, produced 'noisy' data. The same microplate assay subsequently gave less noisy data after shaking sufficiently to make the well contents appear more homogeneous. Another lab also reported occasional cell aggregation and noted that when this occurred, microplates were shaken prior to data collection. It was agreed that the protocol should emphasize the need for plate shaking before reading. Fig. 3 illustrates this point by showing the cytotoxicity (RSG) data for a compound (MMS) before (Fig. 3a) and after shaking (Fig. 3b). These data were taken from a single, separate experiment and not taken directly from data produced in this study, but they clearly illustrate the impact of microplate shaking on data quality.

Each participating site used its own pipetting devices during the performance of the assay in this study and the question arose as to the applicability of certain devices. One trial site where the data were generally 'noisier' performed a follow-up experiment after the completion of the trial. A set of data was produced at Site 1 using a smaller volume pipetting device than was used during the trial. This resulted in improved data quality. Fig. 4 shows genotoxicity data for both a genotoxin and a non-genotoxin from parallel experiments. One microplate was prepared using an electronic multi-channel pipette with a large tip volume (1000 μl) (Fig. 4a and c) and a second was prepared using a manual multi-channel pipette with a small tip volume (200 μl) (Fig. 4b and d). It is clear from this graphical data that the smaller volume device gave the better quality data. It was agreed that the protocol should be prescriptive about pipette volume or required pipetting accuracy.

4.2. Summary of panel discussion regarding results of the trial

Possible reasons for sources of variation in the trial were discussed. The differences between data sets, both within and between trial sites, seemed to correlate most closely with growth of the cells (including control, untreated cells) during the assay incubation. Good microplate culture growth resulted in smooth data graphs, with low variation between sample repeats. Lower microplate culture growth resulted in more variable results and greater variation between test repeats (see Table 4). Discussion about how this variation might arise suggested that the passage regime recommended in the trial SOP had been interpreted differently at individual sites. This probably resulted in an unforeseen variability in the physiological state of cells entering the assay. It was agreed that the protocol should draw attention to the specific passage instructions for this assay, which may differ from those in routine use for other cell line assays. Conclusions such as these are anticipated from a transferability exercise, which aims to identify possible sources of error between laboratories.

Data from 4 compounds (B, D, F and M) were less consistent than others. Variability was largely linked to positive control failures (and hence level of experience with the assay) and the rare data spikes. The variability was substantially reduced following rejection of data from plates with control failure, or rejection of positive results caused by a data anomaly (Tables 3 and 5). The following are compound-specific comments for the 4 compounds.

4.2.1. Compound B

3-Amino-1,2,4-triazole gave a formal positive result at a single concentration for two sites, though at different concentrations below the highest dose. This compound was a variable inducer of cellular fluorescence which was also variable between the strains; in some instances this type of confounding interference can be reduced by using fluorescence polarisation [60] but this technique was not applied in the current exercise.

4.2.2. Compound D

Paclitaxel has aneugenic effects and cells must undergo mitosis before such effects become apparent. This is the posited reason for the expected variability at 24 h (4/12 data points negative in the filtered dataset) which largely disappeared by the 48 h time point (1/10 data points negative).

4.2.3. Compound F

Aphidicolin provided several unexpected negative results in the unadulterated dataset. However, this was not a compound-specific effect as all of these negatives were effectively removed by applying the acceptance criteria for positive controls. The 'false negative' results correlated with poor growth of the cells.

4.2.4. Compound M

Vincristine sulphate revealed a different problem, relating to the dose range chosen for the study. The cytotoxicity data from 3 of the 4 sites did not show a no-effect dose: none of the doses tested permitted the cultures to reach an RSG of 80% or greater. Two of the sites obtained positive results with an apparent threshold effect.

On the whole, data from the 48 h time point identified all of the genotoxins in the study. At two sites the results from the 24 and 48 h time points were very similar, suggesting that with a modified protocol and self-certification period, it might be possible to successfully operate the assay at a single 24 h time point.

Positive genotoxicity control failures were more frequent than expected, though the majority of these control failures (~60%) occurred at one site, and were associated with poor growth of the target cells and the ensuing variability. The consensus view was that this demonstrated that these controls were acting correctly as an indicator of effective assay performance. It is, in retrospect, not surprising that the users most unfamiliar with the assay did not get perfect data immediately, but it is reassuring that this was apparent from the data-handling software which is set up to give a clear indication of control failure. It was concluded that failure of the positive control incorporated in this assay, in its present form, is a reasonable basis on which to reject data.

All the data generated during validation and in this study were from assays performed using 1% (v/v) aqueous DMSO as the solvent vehicle. This was by design, since early pharmaceutical samples for

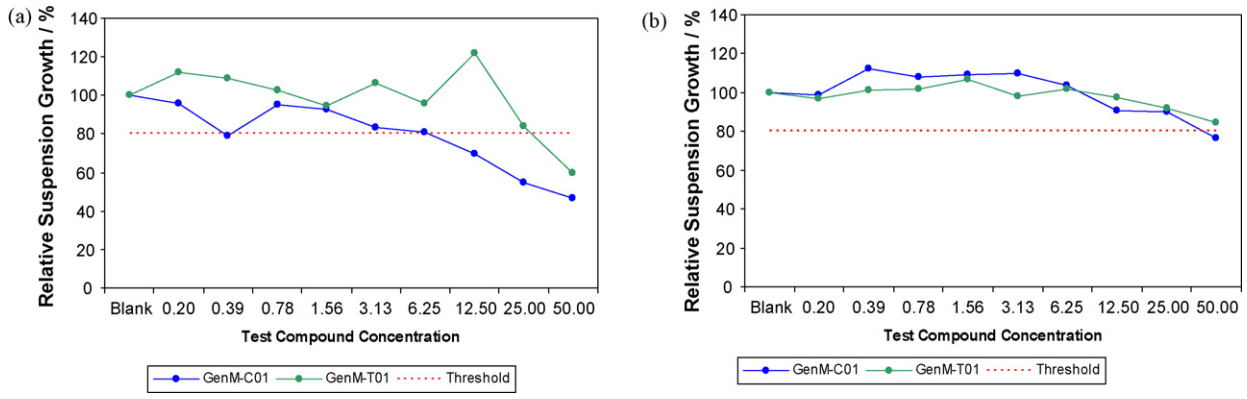

Fig. 3. The effect of vigorous microplate shaking prior to taking measurements on the relative suspension growth. (a) Measurements taken without shaking the microplate. (b). Measurements for the same assay taken after shaking the microplate. Blue = control TK6 strain, green = test TK6 strain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

screening usually come from stock held at or close to 100% DMSO, and 1% is the practical limit of tolerance for the TK6 cells. For more analytical studies, this would not be a constraint. It is anticipated that for some compounds, a purely aqueous solvent would be preferred since there is evidence to suggest that DMSO can interfere with certain compounds (e.g. cisplatin [61]).

5. Discussion

Genetic toxicology has benefited from several key collaborative trials of new techniques and trials that have attempted to compare the performance of more established *in vitro* and *in vivo* methods for the detection of potential carcinogens and mutagens [62]. Such exercises have provided the basis for the use of the test batteries found in current regulatory guidelines and have revealed some surprising lessons for future trials [63,64]. With relevance to setting up trials, the following should be noted: test chemicals should be as pure as possible and samples should be sent to the participating labs from the same batch; investigators should not be informed of the identity of the test chemicals until results have been dispatched; the genotoxicity/carcinogenicity status of the test chemicals should be established from the literature and results

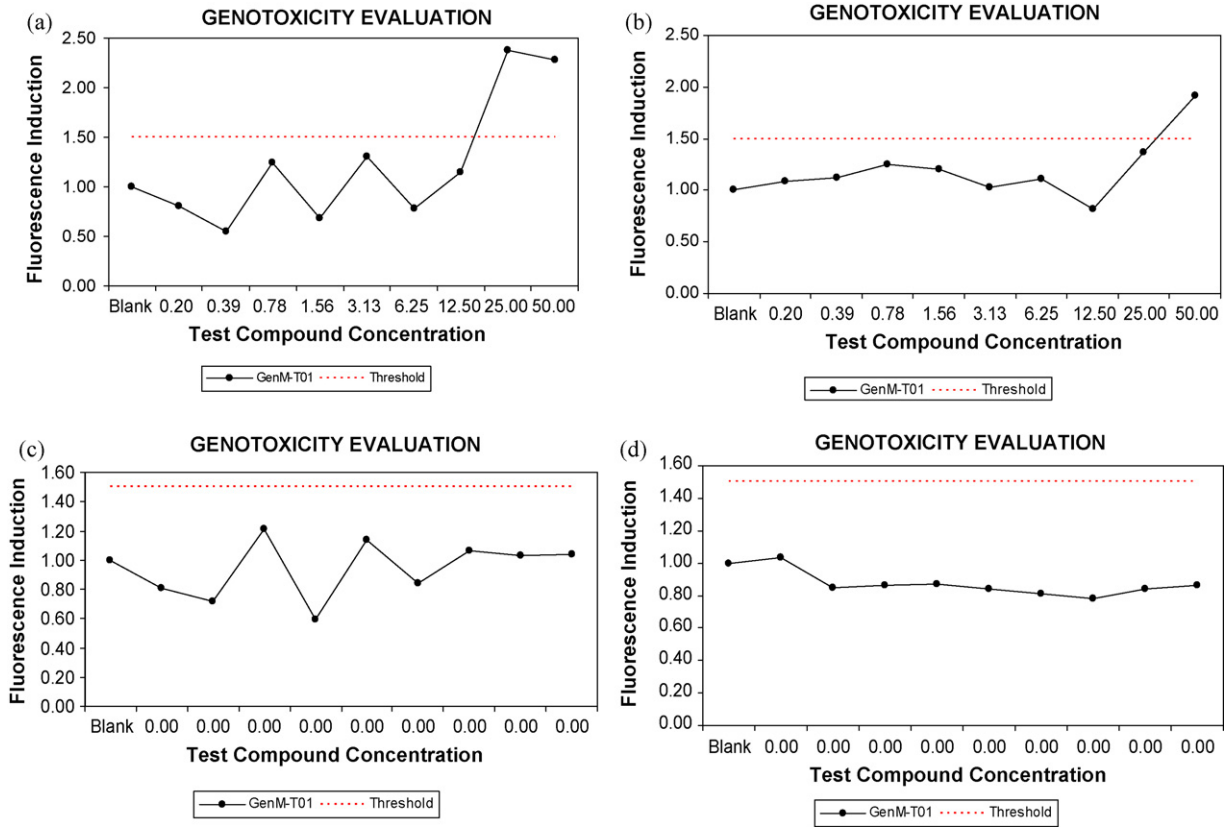

Fig. 4. The importance of the choice of pipetting device in assay microplate preparation (a) and (b) show genotoxicity data for a genotoxin (MMS), (c) and (d) show genotoxicity data for a non-genotoxin (aqueous DMSO, no concentration gradient). Data in (a) and (c) were from a microplate prepared using an electronic multi-channel pipette with a large tip volume (1000 μl per channel), whilst (b) and (d) were from a microplate prepared using a manual multi-channel pipette with a smaller tip volume (200 μl per channel).

should be discussed by the participating scientists after decoding. Comparative results in other collaborative trials have shown that there is a significant site effect with all tests, i.e. small differences in protocol can result in a major difference in assay performance, in particular the shape of dose–response curves and the quantitative response of assays with the same compound. Indeed such trials have resulted in better understanding of the key features of test protocols (e.g. [65]). These lessons have been incorporated into the current exercise.

This study set out to assess the transferability of a new genotoxicity assay in which genotoxin-induced *GADD45a* expression drives the accumulation of GFP in a human lymphoblastoid cell line. A transferable method is one that can be demonstrated to be successfully repeated at a site other than the originating or optimising centre. Transferability is regarded as an important criterion in assessing the practicability of a new assay or method, and the reasons for this are two-fold: (i) it enables the degree of training required for successful transfer to an inexperienced centre or new user to be determined, and (ii) such a study should allow potential sources of both within-laboratory and between-laboratory variation to be identified. Elements of a between-laboratory variability study were also incorporated since the trial was broadened to include a total of four study sites. One site was the centre for the initial transfer of the method away from the originating centre, another was experienced in the assay format but not specifically GreenScreen HC, a third site was the originating laboratory with a newly appointed user, and the final site was the least experienced with the assay format. This study design was intended to more broadly expose areas of the assay protocol that required clarification or amendment and inform on the level of training necessary for naïve users, whilst providing indicative data of the variability of the assay with the SOP used both within and between laboratories.

The individual sites all completed the quadruplicate testing of 16 coded compounds in the time-frame set for the study and submitted full data sets to the Trial Director before test compound identities were revealed to the participants. All data sets were submitted in the required format, i.e. as GreenScreen HC data-processing template files. All further data handling and assessment of results were performed at Gentronix Ltd.

Analysis of a summary of the complete data set (before rejection of data for failing to meet acceptance criteria) showed that all sites were able to identify the majority of genotoxins ($\geq$6 out of 8 at either 24 or 48 h) and non-genotoxins ($\geq$7 out of 8) according to the prediction model. The overall concordance with expected results from this primary data set was 86.3%.

Investigation into the utility of data acceptance criteria, based upon the positive controls included in each microplate assay, revealed that one trial site in particular had a preponderance of control failures. Since the performance of the study reported here, the *GADD45a-GFP* assay is being commercially marketed as GreenScreen HC and is in increasingly wider usage. Experienced users of the assay, both in this study and beyond, have not observed this level of control failure. Indeed, observations from this study and subsequent follow-up experiments at the specific laboratory, suggest that the control failures were caused by a combination of inexperience with the assay and poorer than expected cell growth, compounded by the introduction of further error by the particular pipetting devices used at that lab. The trial data set was re-examined after rejection of compound test data from assays where the intra-assay controls failed or clearly anomalous data points lead to 'spikes' in the dose responses. In this 'filtered' data set, the overall concordance with expected results rose from 86.3 to 92.5%.

The assay data-processing software delivers "genotoxic" and "not genotoxic" indications as well as clear graphical results from the experiments. The appearance of spikes was rare in this study, but their occasional occurrence underlines the importance of inspecting the graphical data from compounds giving positive results. When testing small numbers of compounds, visual inspection is unlikely to be a problem for the user. However, at high throughputs, it would be appropriate to consider software solutions for spike/anomaly identification in data sets determined to give positive results by the software.

Assessment of the SOP used in the study suggested a number of ways of clarifying and improving the assay protocol. It was apparent that there needed to be a greater emphasis on usage of the specific, optimised cell-culture instructions. Cells taken from cultures late in the culture cycle are more likely to fail controls than those earlier in the cycle. This is consistent with the need for cells to be in the exponential phase of growth, and to be actively proceeding through mitosis during the exposure phase. Further clarifications of the SOP should include recommendations on the type of pipetting device and tip volume suitable for performance of this assay, as well as details on the redistribution of microplate well contents by vigorous shaking, prior to data collection.

Minimal training (3 days) was provided to the users at each participating trial site at the beginning of the study. This period only allowed for basic training in setting up the assay and for the new user to have a maximum of two attempts at the assay. It is clear from this study and not wholly unexpected, that successful performance of a new assay requires a strong element of experience. The experience required can be gained through repetitive practise of the assay or from a background experience of a similar assay. All participants in the trial agreed that whilst the minimal training programme was essential, there is an obvious need for a new user to perform repeat assays until reproducible data that meet the acceptance criteria are produced.

6. Conclusions from the study

The *GADD45a-GFP* genotoxicity assay (GreenScreen HC) transferred effectively to new laboratories. The materials and the practical demonstrations provided along with practical training were sufficient for successful adoption of the protocol by 3 of the 4 participating laboratories. At the fourth site, issues resulting in data rejection and variability were identifiable and had clear solutions. Cells were revived and cultured successfully from the frozen state at all sites. The SOP performed well, although minor modifications to the written protocol are likely to improve the transferability and reproducibility of the assay (the current version of the protocol can be obtained from the corresponding author). However, the results showed that there was no requirement for substantive change to the materials supplied or methods of interpretation. Modifications subsequently put in place include the following:

- The volumes of assay medium provided have been adjusted to ensure the inclusion of sufficient dead volume for ease of pipetting.
- Recommendations have been included regarding the optimal pipette for microplate preparation.
- Cell cultures should not exceed a cell density of 1.2×10^6 cells/ml in passage prior to assay.
- Microplates require vigorous shaking prior to spectrophotometric data collection.
- New users should 'self-certify' by repeatedly testing 4 compounds until reproducible data are produced that surpass the acceptance criteria.

The overall concordance of the trial results with expected results was high (92.5% after application of the acceptance criteria), especially given the limited training and practice time for naïve operators. Since this ring trial was performed and subsequent improvements made to the protocol, GreenScreen HC has been successfully transferred to more than 8 laboratories in a variety of pharmaceutical, biotech and contract testing companies. With the advent of a new protocol that allows incorporation of S9 metabolic activation, it is hoped that this study will enable further larger validation studies of this assay to occur.

Acknowledgements

R.M.W. is the Founder and Chief Scientific Officer of, D.J.T. is a consultant for, and both N.B. and L.B. are employed by, Gentronix Ltd. P.W.H. is a co-inventor of the *GADD45a-GFP* assay. The remaining authors have no conflicts to declare.

References

[1] T. Hartung, S. Bremer, S. Casati, S. Coecke, R. Corvi, S. Fortaner, L. Gribaldo, M. Halder, S. Hoffmann, A. Janusch Roi, P. Prieto, E. Sabbioni, L. Scott, A. Worth, V. Zuang, A modular approach to the ECVAM principles on test validity, ATLA 32 (2004) 467–472.
[2] EEC Notes for Guidance for the Testing of Medicinal Products for their Mutagenic Potential, Offi. J. Eur. Commun. L73 (1987).
[3] CPMP/ICH (1995), Note for guidance on genotoxicity: Guidance on Specific Aspects of Regulatory Genotoxicity Test for Pharmaceuticals, paper presented to the International Conference on Harmonisation, Topic S2A, CPMP/ICH/141/95 (1995) http://www.emea.eu.int/pdfs/human/ich/014195en.pdf.
[4] D. Kirkland, M. Aardema, L. Henderson, L. Muller, Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity, Mutat. Res. 584 (2005) 1–256.
[5] D. Kirkland, M. Aardema, L. Muller, M. Hayashi, Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens II. Further analysis of mammalian cell results, relative predictivity and tumour profiles, Mutat. Res. 608 (2006) 29–42.
[6] D. Jacobson-Kram, J.F. Contrera, Genetic toxicity assessment: employing the best science for human safety evaluation Part I: Early screening for potential human mutagens, Toxicol. Sci. 96 (2007) 16–20.
[7] E.J. Matthews, N.L. Kruhlak, M.C. Cimino, R.D. Benz, J.F. Contrera, An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints, Reg. Tox. Pharmacol. 44 (2006) 83–96.
[8] T.R. Skopek, H.L. Liber, B.W. Penman, W.G. Thilly, Isolation of a human lymphoblastoid line heterozygous at the thymidine kinase locus: possibility for a rapid human cell mutation assay, Biochem. Biophys. Res. Commun. 84 (1978) 411–416.
[9] A.J. Fornace Jr., J. Jackman, M.C. Hollander, B. Hoffman-Liebermann, D.A. Liebermann, Genotoxic-stress-response genes and growth-arrest genes. GADD, MyD and other genes induced by treatment eliciting growth arrest, Ann. N. Y. Acad. Sci. 663 (1992) 139–153.
[10] M.C. Hollander, I. Alamo, J. Jackman, M.G. Wang, O.W. McBride, A.J. Fornace Jr., Analysis of the mammalian GADD45 gene and its response to DNA damage, J. Biol. Chem. 268 (1993) 24385–24393.
[11] P.W. Hastwell, L.L. Chai, K.J. Roberts, T.W. Webster, J.S. Harvey, R.W. Rees, R.M. Walmsley, High-specificity and high-sensitivity genotoxicity assessment in a human cell line: validation of the GreenScreen HC GADD45a-GFP genotoxicity assay, Mutat. Res. 607 (2006) 160–175.
[12] ICCVAM/NTP/NICEATM Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity—Recommendations from an International Workshop, NIH Publication No: 01-4500, 2001.
[13] P.A. Cahill, A.W. Knight, N. Billinton, M.G. Barker, L. Walsh, P.O. Keenan, C.V. Williams, D.J. Tweats, R.M. Walmsley, The GreenScreen genotoxicity assay: a screening validation programme, Mutagenesis 19 (2004) 105–119.
[14] J. van Gompel, F. Woestenborghs, D. Beerens, C. Mackie, P.A. Cahill, A.W. Knight, N. Billinton, D.J. Tweats, R.M. Walmsley, An assessment of the utility of the yeast GreenScreen assay in pharmaceutical screening, Mutagenesis 20 (2005) 449–454.
[15] R.D. Curren, J.A. Southee, H. Spielmann, M. Liebsch, J.H. Fentem, M. Balls, The role of prevalidation in the development, validation and acceptance of alternative methods, ATLA 23 (1995) 211–217.
[16] National Toxicology Program (NTP) website, http://ntp.niehs.nih.gov.
[17] B. Miller, F. Potter-Locher, A. Seelbach, H. Stopper, D. Utesch, S. Madle, Evaluation of the in vitro micronucleus test as an alternative to the in vitro chromosomal aberration assay: position of the GUM Working Group on the in vitro micronucleus test—IV. Results with 15 chemicals, Mutat. Res. 410 (1998) 81–116.
[18] M. Honma, M. Hayashi, H. Shimada, N. Tanaka, S. Wakuri, T. Awogi, K.I. Yamamoto, N. Kodani, Y. Nishi, M. Nakadate, T. Sofuni, Evaluation of the mouse lymphoma tk assay (microwell method) as an alternative to the in vitro chromosomal aberration test, Mutagenesis 14 (1999) 5–22.
[19] K.H. Mavournin, D.H. Blakey, M.C. Cimino, M.F. Salamone, J.A. Heddle, The in vivo micronucleus assay in mammalian bone marrow and peripheral blood. A report of the US Environmental Protection Agency Gene-Tox Program, Mutat. Res. 239 (1990) 29–80.
[20] A. Cavaliere, P.F. Alberti, R. Vitali, 5-Fluorouracil carcinogenesis in BALB/c mice, Tumori 76 (1990) 179–181.
[21] United States National Library of Medicine Toxicology Data Network (TOXNET), Chemical Carcinogenesis Research Information System (CCRIS); http://toxnet.nlm.nih.gov.
[22] L.D. Kier, D.J. Brusick, A.E. Auletta, E.S. Von Halle, M.M. Brown, V.F. Simmon, V. Dunkel, J. McCann, K. Mortelmans, M. Prival, T.K. Rao, V. Ray, The *Salmonella typhimurium*/mammalian microsomal assay. A report of the US Environmental Protection Agency Gene-Tox Program, Mutat. Res. 185 (1986) 69–240.
[23] T. Meretoja, U. Gripenberg, D. Bamford, I. Laamanen, M. Sorsa, Mutagenicity and toxicity of amitrole. II. Human lymphocyte culture tests, Mutat. Res. 40 (1976) 191–196.
[24] A.D. Mitchell, A.E. Auletta, D. Clive, P.E. Kirby, M.M. Moore, B.C. Myhr, The L5178Y/tk+/− mouse lymphoma specific gene and chromosomal mutation assay a phase III report of the US Environmental Protection Agency Gene-Tox Program, Mutat. Res. 394 (1997) 177–303.
[25] D.B. McGregor, R. Martin, P. Cattanach, I. Edwards, D. McBride, W.J. Caspary, Responses of the L5178Y tk+/tk− mouse lymphoma cell forward mutation assay to coded chemicals. I: Results for nine compounds, Environ. Mutagen. 9 (1987) 143–160.
[26] A.D. Mitchell, C.J. Rudd, W.J. Caspary, Evaluation of the L5178Y mouse lymphoma cell mutagenesis assay: intralaboratory results for sixty-three coded chemicals tested at SRI International, Environ. Mol. Mutagen. 12 (1988) 37–101.
[27] International Agency for Research on Cancer (IARC); http://monographs.iarc.fr or http://www.iarc.fr.
[28] R.J. Preston, W. Au, M.A. Bender, J.G. Brewen, A.V. Carrano, J.A. Heddle, A.F. McFee, S. Wolff, J.S. Wassom, Mammalian in vivo and in vitro cytogenetic assays: a report of the US EPA's Gene-Tox program, Mutat. Res. 87 (1981) 143–188.
[29] M. Ishidate Jr., M.C. Harnois, T. Sofuni, A comparative analysis of data on the clastogenicity of 951 chemical substances tested in mammalian cell cultures, Mutat. Res. 195 (1988) 151–213.
[30] US Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Drugs@FDA Label information section available at: http://www.fda.gov/cder/foi/anda/2001/75-297_Paclitaxel_prntlbl.pdf.
[31] L. Ozkan, U. Egeli, B. Tunca, N. Aydemir, G. Cecener, G. Akpinar, E. Ergul, C. Cimen, S. Ozuysal, A. Kahraman-Cetintas, K. Engin, M.M. Ahmed, Investigation of genotoxic effect of taxol plus radiation on mice bone marrow cells, Teratog. Carcinog. Mutagen. 22 (2002) 1–11.
[32] G. Pedrali-Noy, G. Mazza, F. Focher, S. Spadari, Lack of mutagenicity and metabolic inactivation of aphidicolin by rat liver microsomes, Biochem. Biophys. Res. Commun. 93 (1980) 1094–1103.
[33] W. Suter, F. Romagna, DNA repair induced by various mutagens in rat hepatocyte primary cultures measured in the presence of hydroxyurea, guanazole or aphidicolin, Mutat. Res. 231 (1990) 251–264.
[34] B. Tedeschi, R. Cicchetti, G. Argentin, D. Caporossi, M. Pittaluga, P. Parisi, P. Vernole, Aphidicolin and bleomycin induced chromosome damage as biomarker of mutagen sensitivity: a twin study, Mutat. Res. 546 (2004) 55–64.
[35] D.E. Amacher, G.N. Turner, The mutagenicity of 5-azacytidine and other inhibitors of replicative DNA synthesis in the L5178Y mouse lymphoma cell, Mutat. Res. 176 (1987) 123–131.
[36] G. Stemp, S. Pascoe, D. Gatehouse, In vitro and in vivo cytogenetic studies of three beta-lactam antibiotics (penicillin VK, ampicillin and carbenicillin), Mutagenesis 4 (1989) 439–445.
[37] Gold LS (2005), The carcinogenic potency project; http://potency.berkeley.edu/cpdb.html.
[38] J.F. Contrera, A.C. Jacobs, J.J. DeGeorge, Carcinogenicity testing and the evaluation of regulatory requirements for pharmaceuticals, Regul. Toxicol. Pharmacol. 25 (1997) 130–145.
[39] J. Ashby, H. Tinwell, P. Glover, P. Poorman-Allen, R. Krehl, R.D. Callander, D. Clive, Potent clastogenicity of the human carcinogen etoposide to the mouse bone marrow and mouse lymphoma L5178Y cells: comparison to Salmonella responses, Environ. Mol. Mutagen. 24 (1994) 51–60.
[40] R.S. Gupta, A. Bromke, D.W. Bryant, R. Gupta, B. Singh, D.R. McCalla, Etoposide (VP16) and teniposide (VM26): novel anticancer drugs, strongly mutagenic in mammalian but not prokaryotic test systems, Mutagenesis 2 (1987) 179–186.
[41] C. Vigreux, J.M. Poul, E. Deslandes, P. Lebailly, T. Godard, F. Sichel, M. Henry-Amar, P. Gauduchon, DNA damaging effects of pesticides measured by the single cell gel elctrophoresis assay (comet assay) and the chromosomal aberration test, in CHOK1 cells, Mutat. Res. 419 (1998) 79–90.
[42] H. Suzuki, S. Nakane, Differential inductions of chromosomal aberrations by topoisomerase inhibitors in cultured Chinese hamster cells, Biol. Pharm. Bull. 17 (1994) 222–226.

[43] O. Hirai, Y. Miyamae, Y. Hattori, M. Takashima, A. Miyamoto, K. Zaizen, Y. Mine, Microbial mutagenicity and in vitro chromosome aberration induction by FK973, a new antitumor agent, Mutat. Res. 324 (1994) 43–50.
[44] M. Diehl, F. Fort, Spiral *Salmonella* assay: validation against the standard pour-plate assay, Environ. Mol. Mutagen. 27 (1996) 227–236.
[45] R. Gudi, J. Xu, A. Thilagar, Assessment of the in vivo aneuploidy/micronucleus assay in mouse bone marrow cells with 16 chemicals, Environ. Mol. Mutagen. 20 (1992) 106–116.
[46] K.M. Abdo, J.K. Haseman, G. Boorman, D.R. Farnell, J.D. Prejean, R. Kovatch, Absence of carcinogenic response in F344 and B6C3F1 mice given D-mannitol in the diet for two years, Food Chem. Toxicol. 21 (1983) 259–262.
[47] M.M. Moore, K.H. Brock, High concentrations of sodium chloride induce a "positive" response at the TK locus of L5178Y/TK+/− mouse lymphoma cells, Environ. Mol. Mutagen. 12 (1988) 265–268.
[48] J. Wangenheim, G. Bolcsfoldi, Mouse lymphoma L5178Y thymidine kinase locus assay of 50 compounds, Mutagenesis 3 (1988) 193–205.
[49] H.C. Kim, S.W. Cha, C.S. Ha, J.K. Roh, Y.S. Lee, F. Furukawa, A. Nishikawa, M. Takahashi, Reappraisal of eight representative carcinogenic and non-carcinogenic compounds in a new medium-term rat liver bioassay using D-galactosamine, Cancer Lett. 104 (1996) 85–90.
[50] D. Matheson, D. Brusick, R. Carrano, Comparison of the relative mutagenic activity for eight antineoplastic drugs in the Ames Salmonella/microsome and TK+/− mouse lymphoma assays, Drug Chem. Toxicol. 1 (1978) 277–304.
[51] P. Arni, T. Hertner, Chromosomal aberrations in vitro induced by aneugens, Mutat. Res. 379 (1997) 83–93.
[52] I. Sbrana, S. Caretto, G. Rainaldi, N. Loprieno, Induction of chromosomal aberrations and SCE by chloramphenicol, Mutat. Res. 248 (1991) 145–153.
[53] A. Martelli, F. Matioli, G. Pastorino, L. Robbiano, A. Allavena, G. Brambilla, Genotoxicity testing of chloramphenicol in rodent and human cells, Mutat. Res. 260 (1991) 65–72.
[54] C.A. Hilliard, M.J. Armstrong, C.I. Bradt, R.B. Hill, S.K. Greenwood, S.M. Galloway, Chromosome aberrations in vitro related to cytotoxicity of nonmutagenic chemicals and metabolic poisons, Environ. Mol. Mutagen. 31 (1998) 316–326.
[55] S.M. Amer, F.A.E. Aly, Genotoxic effect of 2,4-dichlorophenoxy acetic acid and its metabolite 2,4-dichlorophenol in mouse, Mutat. Res. 494 (2001) 1–12.
[56] Thomson PDR, Physicians Desk Reference Electronic Library™ Online, Thomson Micromedex, Greenwood Village, Colorado (2004). Available online at: http://www.micromedex.com/products/pdrlibrary/.
[57] F. Degrassi, R. De Salvia, C. Tanzarella, F. Palitti, Induction of chromosomal aberrations and SCE by camptothecin, an inhibitor of mammalian topoisomerase I, Mutat. Res. 211 (1989) 125–130.
[58] L.C. Backer, J.W. Allen, K. Harrington-Brock, J.A. Campbell, D.M. DeMarini, C.L. Doerr, D.R. Howard, A.D. Kligerman, M.M. Moore, Genotoxicity of inhibitors of DNA topoisomerase I (camptothecin) and II (m-AMSA) in vivo and in vitro, Mutagenesis 5 (1990) 541–547.
[59] M. Holmstrom, V. Winters, Micronucleus induction by camptothecin and amsacrine in bone marrow of male and female CD-1 mice, Mutagenesis 7 (1992) 189–193.
[60] A.W. Knight, N.J. Goddard, P.R. Fielden, A.L. Gregson, N. Billinton, M.G. Barker, R.M. Walmsley, The application of fluorescence polarisation for the enhanced detection of green fluorescent protein (GFP) in the presence of cellular autofluorescence and other green fluorescent compounds, Analyst 125 (2000) 499–506.
[61] T. Gebel, A. Koenig, Impact of dimethyl sulfoxide and examples of combined genotoxicity in the SOS chromotest, Mutat. Res. 444 (1999) 405–411.
[62] F.H. Sobels, Studies in comparative chemical mutagenesis, Environ. Mutagen. 7 (1985) 759–773.
[63] F.J. De Serres, J. Ashby, Evaluation of Short-term Tests for Carcinogens. Report of the International Collaborative Trial, Elsevier/NorthHolland Inc., 1981.
[64] J.M. Parry, C.F. Arlett, Comparative genetic toxicology, in: The Second UKEMS Collaborative Study, Macmillan Press, 1985.
[65] S.M. Galloway, T. Sofuni, M.D. Shelby, A. Thilager, V. Kumaroo, P. Kaur, D. Gulati, D.L. Putman, H. Murli, R. Marshall, N. Tanaka, B. Anderson, E. Zeiger, M. Ishidate Jr., Multilaboratory comparison of in vitro tests for chromosome aberrations in CHO and CHL cells tested under the same protocols, Environ. Mol. Mutagen. 29 (1997) 189–207.